

**Identification of novel genes involved in the pathogenesis of
inflammatory bowel disease using high-throughput
microarray expression screening**

Dissertation zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Nancy Mah

B.Sc., M.Sc.

Kiel

December 12, 2006

Referentin: Prof. Dr. Karin Krupinska, Kiel (D)

Korreferent: Prof. Dr. Stefan Schreiber, Kiel (D)

Tag der mündlichen Prüfung: 11. Dez 2006

Zum Druck genehmigt: 12. Dez 2006

gez. Prof. Dr. (Dekan)

To my parents

Table of contents

1	Introduction	1
1.1	Description of inflammatory bowel disease	1
1.2	Interacting factors in inflammatory bowel disease	2
1.2.1	Genetic factors.....	2
1.2.2	Environmental factors	4
1.2.3	Immunological interaction.....	5
1.2.4	Epithelial barrier function	8
1.2.5	Wnt signaling pathway.....	8
1.3	Microarray expression screening	11
1.4	Aims of the study.....	13
2	Methods	16
2.1	Microarray comparison.....	16
2.1.1	Patient cohort for microarray expression screening	16
2.1.2	Isolation of total RNA from biopsies	16
2.1.2.1	Homogenization of colon mucosa biopsies	17
2.1.2.2	DNase treatment of RNA and final isolation of RNA	17
2.1.2.3	Quality control of RNA	18
2.1.3	Production of cDNA microarray: Human Unigene Set RZPD1.....	19
2.1.4	Hybridization of cDNA microarrays.....	19
2.1.5	Normalization of clone-based microarray data using Zipf's Law	20
2.1.6	Isolation of plasmid DNA	21
2.1.7	Sequence verification of bacterial clones	22
2.1.8	Hybridization and data analysis of Affymetrix U 95Av2 microarrays.....	23
2.1.8.1	Hybridization of Affymetrix U 95Av2 microarrays	23
2.1.8.2	Microarray suite software (MAS) 5.0.....	23
2.1.9	Matching of clone-based PCR products to corresponding oligonucleotide probes....	24
2.1.10	Statistical analysis of expression data for microarray comparison	25
2.1.11	Real-time PCR quantitation of selected transcripts	26
2.2	Identification of genes involved in IBD using cDNA microarrays.....	27
2.2.1	cDNA microarray expression screening patient cohorts	27
2.2.1.1	Patient cohort for cDNA microarray expression screening.....	27
2.2.1.2	Patient cohort for real-time PCR verification.....	28
2.2.2	Microarray data analysis.....	29
2.2.3	Prediction of function in unknown genes	30
2.2.4	Verification of cDNA microarray results by fluorogenic real-time PCR	31
2.2.5	Immunohistochemistry.....	31
2.3	Identification of genes involved in inflamed/non-inflamed IBD using Affymetrix HG-U133 arrays.....	32
2.3.1	Patient cohorts for microarray analysis on HG-U133 Affymetrix arrays.....	32
2.3.1.1	Inflamed and non-inflamed UC cohort	32

2.3.1.2	Inflamed and non-inflamed CD cohort	33
2.3.2	Microarray hybridization and image processing	33
2.3.3	Affymetrix HG-U133 microarray data analysis	34
2.4	Real-time PCR analysis	37
2.4.1	Patients for real-time PCR analysis	37
2.4.2	cDNA synthesis	37
2.4.3	Plate production	38
2.4.4	Gene expression assays on ABI prism 7900HT	39
2.4.5	Real-time PCR data processing	40
3	Results	41
3.1	Microarray comparison	41
3.1.1	Variation in expression signals	41
3.1.2	Overlap between gene detection on both platforms	43
3.1.3	Comparison of expression levels between platforms	47
3.2	Molecular profiling of IBD using cDNA microarrays	51
3.2.1	Expression profiles of the IBD colonic mucosa	51
3.2.2	Functional prediction of unknown genes	56
3.2.3	Independent quantitation by fluorescent real-time PCR	56
3.2.4	Expression profiles in non-IBD disease samples	61
3.2.5	Localization by immunohistochemistry	62
3.3	Molecular profiling of IBD samples using Affymetrix microarrays	64
3.3.1	Microarray data pre-processing	64
3.3.2	Expression profiling of the non-/inflamed UC mucosa	66
3.3.2.1	Differentially regulated genes in the UC dataset	67
3.3.2.2	Functional grouping of differentially regulated transcripts for UC dataset	74
3.3.3	Expression profiling of the non-/inflamed CD mucosa	75
3.3.3.1	Differentially regulated genes in the CD dataset	76
3.3.3.2	Functional grouping of differentially regulated transcripts for CD dataset	90
3.3.4	Functional themes supported by microarray analysis in IBD	92
3.3.5	mRNA expression of genes involved in Wnt signaling in IBD	101
3.3.6	mRNA expression of genes involved in barrier function	106
4	Discussion	112
4.1	Microarray comparison	112
4.2	Expression profiling in inflammatory bowel disease	115
4.2.1	Broad functional themes in the pathology of IBD	117
4.2.1.1	Immune and inflammatory response	118
4.2.1.2	Oncogenesis, cell proliferation and growth	120
4.2.1.3	Structure and permeability	122
4.2.2	Specific themes in the pathology of IBD	124
4.2.2.1	Wnt pathway	124
4.2.2.2	Cell-cell adhesion	130

4.3	Conclusions.....	134
5	Summary.....	137
6	Zusammenfassung	138
7	References.....	140
8	Appendix.....	152
8.1	Materials.....	152
8.2	Equipment.....	153
8.3	Solutions and media.....	154
8.4	Primer and probe sequences.....	154
8.5	Abbreviations	158
9	Curriculum vitae.....	161
10	Declaration (Erklärung).....	163
11	Acknowledgements.....	165

1 Introduction

1.1 Description of inflammatory bowel disease

Inflammatory bowel disease (IBD) is a chronic, relapsing inflammatory disorder of the gastrointestinal tract. The exact origin of IBD is unknown, but identified factors that influence IBD pathogenesis involve the interaction of genetic background, environment, and immune mechanisms. Symptoms of IBD include abdominal pain, diarrhea, weight loss, and fever (Yang et al., 2001). The incidence of IBD is about 0.1% worldwide, with a higher rate in developed countries. Onset of IBD typically occurs in the late-teens/early twenties, and there are two main subtypes of IBD, Crohn's disease (CD) and ulcerative colitis (UC).

Distinct clinical features separate the two subtypes, although mucosal inflammation, ulceration, and chronic mucosal damage are common to both CD and UC. In CD, any part of the gastrointestinal tract may be affected in a discontinuous fashion, resulting in 'skip lesions'. In contrast, the mucosal damage in UC is restricted to the colon and extends continuously from the rectum, resulting in proctitis (inflammation of the rectum), left-sided colitis (inflammation extending from rectum to left flexure), or pancolitis (inflammation of the entire colon). Histologically, CD is characterized by thickened intestinal walls, fissures or fistulae, fibrosis, the presence of granulomas in all tissue layers, and inflammation affecting all layers of the intestinal mucosa. In UC, affected areas are primarily restricted to the superficial layers (mucosa), goblet cells are depleted, and ulcers penetrate deeply into the submucosa (Hendrickson et al., 2002).

Complications arising from IBD are diverse. It is considered that general defects in the immune response can lead to extraintestinal inflammation causing complications such as joint disorders (ankylosing spondylitis), skin disorders (erythema nodosum, epidermolysis bullosa acquisita), hepatobiliary diseases (primary sclerosing cholangitis), and eye diseases (uveitis). UC is often associated with ankylosing spondylitis, arthritis, primary sclerosing cholangitis, and an increased risk of colon carcinoma (1998), while CD is more often associated with psoriasis and thrombotic vascular complications (Yang et al., 2001).

1.2 Interacting factors in inflammatory bowel disease

The etiology of IBD involves the interaction of genetic, environmental and immunological factors.

1.2.1 Genetic factors

There are many studies that support the influence of genetic background in IBD; in some cases genetic factors differentially influence the two disease subtypes, UC and CD. The fact that IBD is a complex polygenic disease, rather than a monogenic disease, is best highlighted by differential concordance rates in twins. In CD, monozygotic twins have a concordance rate of 58%, while dizygotic twins have rates similar to siblings. In UC, the concordance rates are lower than in CD (monozygotic twins, 6-17%; dizygotic twins, 0-5%) (Orholm et al., 2000; Thompson et al., 1996; Tysk et al., 1988). It is thought that non-genetic factors, such as the environment, have an influence on the incidence of IBD, since the concordance rate for monozygotic twins is not 100% (Yang et al., 2001). Further evidence for the genetic component of IBD is reflected in the nine susceptibility loci that have been identified for IBD, as outlined in Table 1.1. Most prominently, variations in *CARD15* (IBD1), an innate immunity receptor, have been strongly associated solely with CD pathogenesis. Homozygotes carrying two risk mutations have a 44 times greater probability of contracting CD than non-carriers of the mutations (Hugot et al., 2001). In the case of IBD5, a 250 kb haplotype block on 5q31 was identified to have an association with CD (Rioux et al., 2001) and subsequently variants in the organic cation transporter (OCTN) family (*SLC22A4*, *SLC22A5*) on 5q31 were found to confer susceptibility to CD and UC (Dring et al., 2006; Peltekova et al., 2004). Outside the major susceptibility loci, variants in *discs, large homolog 5* (*DLG5*; 10q23) were found to be associated with IBD (Stoll et al., 2004) and polymorphisms in *mucin 3* (*MUC3*; 7q22) and *anterior gradient 2* (*AGR2*; 7p21.3) were associated with UC (Kyo et al., 1999; Zheng et al., 2006).

Table 1.1. Susceptibility loci for inflammatory bowel disease

Locus	Chromosomal Location	Candidate genes located in or near locus
IBD1	16p12-q13	<i>CARD15</i> , CD11 integrin cluster (<i>ITGAM</i> , <i>ITGAX</i> , <i>ITGAL</i> , <i>ITGAD</i>), <i>CD19</i> , <i>IL4R</i> , <i>SPN</i>
IBD2 (Satsangi et al., 1996)	12p13.2-q24.1	<i>IFNG</i> (12q14), <i>SLC11A2</i> (<i>NRAMP2</i> : 12q13)
IBD3 (Hampe et al., 1999b)	6p	Major histocompatibility complex (MHC) region, <i>TNF</i>
IBD4 (Duerr et al., 2000)	14q11-q12	T-cell receptor alpha/gamma complex
IBD5 (Rioux et al., 2000)	5q31	<i>SLC22A4</i> , <i>SLC22A5</i> , cytokine gene cluster (<i>IL3</i> , <i>IL4</i> , <i>IL5</i> , <i>IL13</i> , <i>CSF2</i>), <i>SPINK5</i>
IBD6 (Rioux et al., 2000)	19p13	<i>ICAM1</i> , complement component 3, thromboxane A2 receptor, leukotriene B4 hydroxylase, <i>TYK2</i> , <i>JAK3</i>
IBD7 (Cho et al., 2000)	1p36	
IBD8 (Hampe et al., 2002)	16p	
IBD9 (Duerr et al., 2000; Hampe et al., 2001)	3p26	
unnamed	10q23 (Hampe et al., 1999a)	<i>DLG5</i> (Stoll et al., 2004)
unnamed	7q21.1	<i>ABCB1</i> (<i>MDR1</i>)
unnamed	7q22 (Satsangi et al., 1996)	<i>MUC3</i> , <i>HGF</i> , <i>EGFR</i>
unnamed	7p21.3	<i>AGR2</i> (Zheng et al., 2006)
unnamed	3p21.2 (Satsangi et al., 1996)	<i>GNAI2</i>

Animal models also strongly support the genetic component of IBD. The mice strains Samp/Yit and C3H/HeJBir are murine models of intestinal inflammation and spontaneous colitis, respectively (Kosiewicz et al., 2001; Kozaiwa et al., 2003). In particular, the C3H/HeJBir strain was also shown to be sensitive to chemically induced colitis by administration of dextran sulfate sodium (DSS) in comparison to other strains of mice (Mahler et al., 1998). Furthermore, though no one single model manifests all the characteristics of human IBD, different genetic defects in transgenic animal models (knock-in or knock-out) can reproduce the same clinical phenotype. For example, knocking-out interleukin 10 (IL10), a cytokine that suppresses macrophage function, yields a mouse model with an over-production of type 1 helper T cell (Th1) cytokines, resulting in intestinal inflammation (Mahler et al., 2002; Weinstock et al., 2003). An interleukin 2 (IL2; stimulates T-cell proliferation) knock-out mouse results in increased transforming growth factor-beta 1 (TGFB1), CD14 and inducible NOS, and ultimately, colitis (Harren et al., 1998; Meijssen et al., 1998; Sadlack et al., 1993; Schultz et al., 1999). Finally, cross-breeding a DSS-susceptible mouse strain (C3H/HeJ) with a partially resistant strain (C57BL/6) yielded significant genetic linkage to colitis at marker D5Mit216, a locus which is syntenic with the human linkage on chromosome 5 (Ma et al., 1999).

Together, these examples illustrate that genetic changes have the ability to cause colonic inflammation.

1.2.2 Environmental factors

A wide variety of environmental factors has been suggested as risk factors to developing IBD, including smoking, diet, hygiene, and technological advances that cause changes in bacterial flora. Indeed, IBD may generally be described as a side effect of 20th century industrialization, since incidence of IBD has increased within industrialized nations since the 1950's (Mayberry and Rhodes, 1984).

Of all environmental factors, smoking has the strongest influence on IBD, with opposite effects in CD and UC. Whereas smoking increases the risk of development, relapse, and the need for surgical intervention in CD, smoking is viewed as beneficial against UC (Lindberg et al., 1988; Regueiro et al., 2005). The mechanism by which smoking influences different effects in CD and UC is unknown; however, one possibility is that smoking may impair antimicrobial response in CD, while promoting protective effects of luminal bacteria in UC (Birrenbach and Bocker, 2004).

With respect to diet, breastfeeding is suggested to be protective against IBD, (Klement et al., 2004), whereas increased intake of simple carbohydrates and fast food are risk factors for CD (Persson et al., 1992). High sucrose intake is a risk factor for IBD, while high fat intake is a risk factor for UC (Reif et al., 1997). Factors which affect availability, transport or metabolism of the major energy source of the colonocytes, the short chain fatty acids (SCFA) (and in particular, butyrate), may introduce a risk for IBD by causing colonocyte starvation. For example, increased ingestion of food containing sulfur may be associated with UC, since sulfur compounds inhibit metabolism of butyrate (Roediger et al., 1993). Lack of dietary fiber to breakdown into SCFA by bacterial flora or impaired transport of SCFAs can cause colon pathology (Cuff et al., 2005; Scheppach et al., 1997).

Changes in hygiene have also been hypothesized to be related to the increasing rate of IBD. It has been suggested that overall improvements in public sanitation may predispose to IBD by removing the stimulus for a mucosal immune response, which would normally be required in an unsanitary

environment (Rotter, 1994). It is postulated that the predisposing immune response genes, when not sufficiently activated in a 'clean' environment, may lead to hyperstimulation of the immune response if challenged by a microbial antigen later in life. Alternatively, the lack of exposure to microbial antigens may leave the mucosal immunological system in a perpetually primed state which makes the system vulnerable to dysregulation in the form of an autoimmunity reaction (Wills-Karp et al., 2001). The loss of the intestinal parasite, the helminth, is an example of the hygiene hypothesis. As an integral part of the intestinal flora, helminths affect T regulatory cells to prevent over-activation of T cells. The absence of the helminth upsets T cell regulation, a process that is probably common in diseases with an autoimmune component (IBD, asthma, multiple sclerosis, allergies) (Korzenik, 2005; Weinstock et al., 2004).

Interestingly, the advent of refrigeration has been proposed to cause CD in genetically susceptible individuals. The use of refrigeration in commercial and domestic food preparation encourages the growth of psychrotrophic bacteria, such as *Yersinia enterocolitica* and *Listeria monocytogenes*, which are able to grow at cold temperatures (between -1°C and 10°C). The "cold chain hypothesis" proposes that chronic ingestion of psychrotrophic bacteria from refrigerated foods invokes a over-active host reaction in genetically susceptible individuals who have lost the ability to tolerate the bacteria (Hugot et al., 2003).

Lifestyle changes in the last century have influenced hygienic practices, what we eat and how we store our food. Though all of these environmental factors are thought to contribute to IBD susceptibility, smoking remains the most dominant factor.

1.2.3 Immunological interaction

The gut immune system must maintain a delicate balance between tolerating food antigens and commensal bacteria, while being able to respond appropriately to pathogens. Oral tolerance is thought to occur by three mechanisms: 1) clonal selection takes place to eliminate the effector T cells that recognize commensal bacterial as 'self-antigen'; 2) clonal anergy is induced (i.e. lymphocytes are rendered unresponsive to the oral antigen); and 3) suppressor T-cells (Th3 cells) are induced (Bouma and Strober, 2003). In IBD, different defects in the maintenance of oral tolerance have the potential to cause chronic inflammation. Though there are no prospective

studies to definitively show that oral tolerance is defective in IBD, there are studies that suggest that there is some truth to the hypothesis. In one example, a non-prospective study has shown that T-cells from IBD patients proliferated and produced cytokines in response to their own microflora, whereas T-cells from control subjects did not (Duchmann et al., 1995). In a more recent example, IBD patients and controls were administered an oral antigen (keyhole limpet hemocyanin) before initial immunization and booster (Kraus et al., 2004). In contrast to the control group, tolerance was not induced in CD and UC patients, as shown by enhanced levels of T-cell proliferation after immunization, suggesting a defect in suppression response in IBD. In a follow-up study, the same authors repeated the protocol on IBD patients and first-degree relatives and found that intestinal permeability could not be faulted for the lack of tolerance induction and suggested that there is a genetic defect responsible for oral tolerance in IBD (Kraus et al., 2006).

An upset in balance between effector T-cells (Th1 or Th2), which cause inflammation, and regulatory T-cells, which prevent inflammation, can also result in mucosal inflammation. According to mouse models, inflammation may result from either an over-active effector T-cell response or a weakened regulatory T-cell response (Bouma and Strober, 2003). Within the effector T-cells, mucosal inflammation can be a consequence of excessive T helper 1 cell (Th1) response, which induces secretion of interleukin 12 (IL12), interferon γ (IFNG) and/or tumor necrosis factor alpha (TNF). Alternatively, inflammation may result from an excessive Th2 response, which induces secretion of interleukins 4, 5 and/or 13 (IL4, IL5, IL13). These cytokine profiles are clear distinguishing characteristics of the two subtypes of IBD. The cytokine profile for CD is similar to that observed in Th1-mediated inflammation. Many studies have showed that macrophages isolated from CD, but not UC patients, produce increased amounts of IL12, which activates Th1 cells (Monteleone et al., 1997; Parronchi et al., 1997). In addition, T cells from CD patients show increased levels of IFNG and decreased levels of IL4 in comparison to controls (Fuss et al., 1996; Monteleone et al., 1997; Parronchi et al., 1997). A final piece of evidence supporting the role of Th1 response in CD is the finding that CD patients treated with an antibody against the p40 chain of IL12 show an immediate reduction in inflammation, which is paralleled by decreased IL12 and IFNG levels in mononuclear cells from colonic lamina propria (Mannon et al., 2004). In contrast to CD, the UC cytokine profile resembles that of a Th2-mediated inflammation. Although the hallmark cytokine of Th2 response, IL4, has not been shown to be increased in UC Th2 cells, studies

suggest that UC is driven by Th2-mediated inflammation (Bouma and Strober, 2003). For example, the development of autoantibodies, such as anti-neutrophil cytoplasmic antibody (pANCA), is more prevalent in UC in contrast to CD (Saxon et al., 1990), while the anti-Saccharomyces cerevisiae antibody (ASCA) is more prevalent in CD than in UC, which provides a possibility to distinguish UC and CD by serological testing (Beaven and Abreu, 2004). Moreover, Th2-related subclasses of IgG (IgG1 and IgG4) are associated with UC, as opposed to Th1-related subclasses (IgG2), which are more associated with CD (Kett et al., 1987). Together, these studies support the hypothesis of distinct Th-responses in CD and UC.

Abnormalities in any aspect of the innate immune response may ultimately lead to inflammation in IBD. The most prominent case in point is the association of mutations in *CARD15* with CD. The protein encoded by this gene, NOD2, is an intracellular receptor for bacterial peptidoglycan, in particular, muramyl dipeptide. NOD2 is expressed in epithelial cells, dendritic cells, granulocytes and monocytes. There are three hypotheses as to how defects in NOD2 act to cause CD (Bouma and Strober, 2003): 1) the lack of functional NOD2 leads to impaired macrophage activation and infection; 2) lack of functional NOD2 in epithelial cells negates activation of an epithelial cell defense response (in the form of chemokine and defensin secretion), allowing bacterial colonization of the crypt and penetration through the epithelial barrier; 3) impaired recognition of bacterial antigen leads to inappropriate activation of antigen presenting cells, resulting in an imbalance between effector and regulatory T cells.

Various defects in the regulation of the immune response are observed in IBD patients. Tolerance to an external antigen appears to be different in IBD patients compared to normal controls, while other studies identify defects that appear to characterize CD or UC. CD is typified by a dysregulation of innate immune response, such as Th-1-mediated immune response or impaired function of the bacterial receptor NOD2. On the other hand, the IgG subclass profile and the prevalence of autoantibodies support UC as a Th-2 mediated-mediated immune response.

1.2.4 Epithelial barrier function

The mucosal epithelium provides a physical barrier separating the outside environment (food antigens, microflora, pathogens) from the inner workings of the body. This physical barrier consists of a thick glycocalyx mucous layer coating the apical surface of the epithelial cell layer and the epithelial cells themselves, which are joined laterally by tight and adherens junctions and basally to the extracellular matrix of the lamina propria. Breaches of the epithelial barrier can allow components of the outside environment to come into direct contact with the gut immunological machinery, thereby causing an overwhelming immune response. In the case of IBD, it is not clear whether impaired barrier function first causes a chronic inflammation or if an impaired immune response leads to degradation of the barrier. Due to the multifactorial nature of IBD, it is possible that one or both events contribute to the initiation of chronic inflammation, depending on the characteristics of the individual patient. Many animal models support the hypothesis that the loss of barrier function causes inflammation. In one example, severe inflammation is observed in transgenic mice expressing a dominant negative N-cadherin transgene that impairs intra-epithelial cell adhesion (Hermiston and Gordon, 1995). Not only does the intestinal epithelial layer provide a physical barrier, it also plays an active role in immune response. Epithelial cells are able to detect microbial antigens and further transmit the appropriate response. For example, Toll-like receptor 5 (TLR5) is specifically expressed on the basolateral side of the epithelial cell, where, if it is activated by bacterial flagellin, it initiates an inflammatory response to combat the bacterial breach (Gewirtz et al., 2001). And finally, recent *in vitro* studies show that the epithelial tight junction can alter its structure in response to cytokines such as TNF (Ye et al., 2006) or IFNG (Utech et al., 2005).

1.2.5 Wnt signaling pathway

The Wnt signaling pathway has been widely studied for its role in developmental processes of model organisms such as *Drosophila*, *Xenopus* and *C. elegans*. Mammalian homologues of the Wnt pathway have been found to play roles in stem cell development, embryonic development, and cancer. The role of the WNT pathway in the pathology of IBD is potentially interesting for two reasons: 1) IBD patients, and in particular, UC patients, show an increased risk for colon cancer (van Hogezand et al., 2002); 2) Wnt pathway is instrumental in gastrointestinal stem cell development and maintenance of the colon (Batlle et al., 2002; van de Wetering et al., 2002).

The Wnt genes, of which there are 19 in humans (Miller, 2002), encode secreted glycoproteins that bind and activate a transmembrane receptor complex formed by the lipoprotein-receptor-related proteins 5 and 6 (LRP5/6) and Frizzled (FZ) transmembrane family, of which there are 10 members. Wnt binding may be enhanced by binding to heparan sulphate proteoglycans (HSPG), which act as co-receptors for Wnt binding, and cell surface sulfatases that have been shown to modify HSPG (Ai et al., 2003). Secreted antagonists, such as Dickkopf (DKK) family, secreted Frizzled-related protein (sFRP) family and Wnt-inhibitory factor-1 (WIF1), prevent Wnt activation by preventing Wnt ligands from binding to the LRP/FZ receptor complex (Moon et al., 2004). The wide range of tissue specific functions of Wnt can be attributed to the specific combinations of Wnt ligands, receptors, and effectors.

Wnt signaling pathways can be divided into two types: 1) β -catenin-dependent signaling (canonical Wnt pathway); and 2) β -catenin-independent signaling, which includes non-canonical pathways such as the planar polarity and Wnt/Ca²⁺ pathways. The Wnt ligands can be classified according to the Wnt pathway that is activated when expressed in *Xenopus* embryos (Strutt, 2003). The first group, including Wnt1, Wnt3a, Wnt8 and Wnt8b, activates signals through the canonical Wnt pathway and is able to transform mammalian cells. The second group, including Wnt4, Wnt5a, and Wnt11, transmits signals through the non-canonical Wnt/Ca²⁺ pathway, but cannot transform cells.

Canonical Wnt signaling is well-characterized and is summarized in Figure 1.1. In the inactive state, free cytoplasmic β -catenin protein levels are minimized by protein degradation (Nelson and Nusse, 2004). A pro-degradation complex of axin, adenomatous polyposis coli (APC), casein kinase I (CKI) and glycogen synthase-3 (GSK3) phosphorylates β -catenin and targets it for ubiquitination and degradation. Upon activation of Wnt signaling by binding of Wnt ligands to the LRP/FZ complex, dishevelled (DSH or DVL) becomes activated and is recruited to the transmembrane receptor complex, which then attracts axin and the degradation complex to the membrane. From this point on, a combination of axin degradation and inhibition of GSK3 by DSH decreases degradation of β -catenin. Activation of the Wnt pathway stabilizes β -catenin by inhibiting GSK3B activity. As a result, cytoplasmic β -catenin protein accumulates and translocates from the cytoplasm to the nucleus, where it interacts with T-cell-specific (TCF) and lymphoid enhancer-

binding (LEF) transcription factors. The transcription factors then activate transcription of Wnt target genes, such as cell cycle regulators that control cell growth or proliferation (*cyclin D1*, *MYC*, *FOS* and *JUN*). Of particular interest in the canonical Wnt pathway is the involvement of β -catenin and APC, which take part in both gene transcription of Wnt signals and cell-cell adhesion and cytoskeletal rearrangements (Logan and Nusse, 2004).

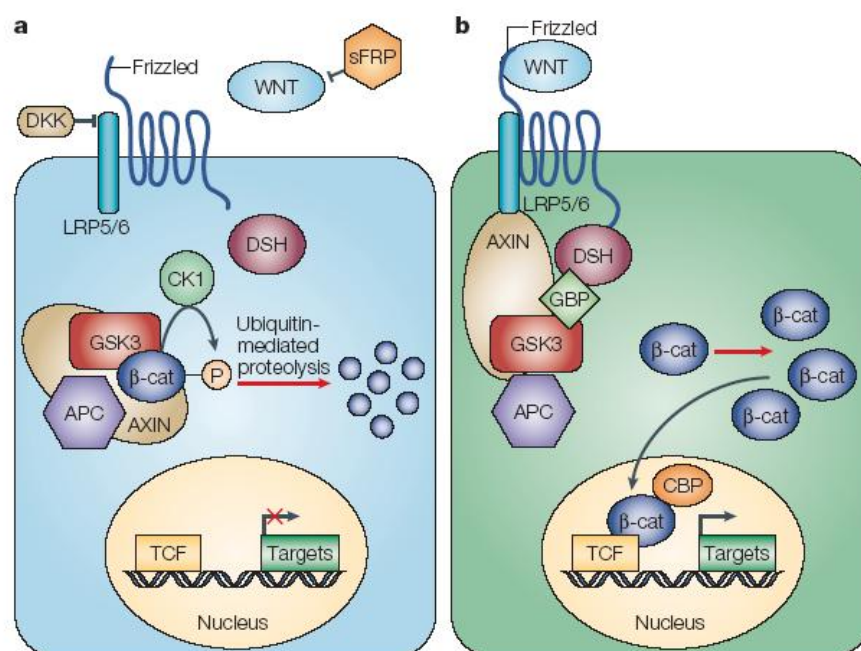


Figure 1.1 Canonical Wnt signaling pathway.

Panel A. In the inactivated state, extracellular inhibitors sFRP and DKK prevent binding of Wnt ligands to the receptor complex LRP5/6 and Frizzled. The degradation complex consisting of axin, APC, GSK3, and CK1 phosphorylates β -catenin, which marks it for ubiquitin-mediated degradation. Panel B. Wnt ligand binds to the receptor complex (LRP5/6 and Frizzled), activating DSH and recruiting the degradation complex to the plasma membrane. This inhibits phosphorylation of β -catenin, which allows β -catenin to escape ubiquitin-mediated degradation and accumulate in the cytoplasm. β -catenin then translocates to the nucleus, where it binds the co-activator CBP (CREB-binding protein) and the transcription factor TCF to initiate transcription of Wnt target genes. This figure was taken from a Wnt review (Moon et al., 2004).

As IBD patients demonstrate an increased risk for colon cancer, it is of interest that mutations which cause the constitutive activation of the canonical Wnt pathway are responsible for most colon cancers (Peifer and Polakis, 2000). Loss-of-function mutations in APC can be attributed to 85% of all sporadic colorectal cancers, whereas mutations that stabilize β -catenin accumulation in the cytoplasm can be attributed to 10% of sporadic colorectal cancers (Giles et al., 2003). Cell proliferation leading to tumour formation is thought to be driven by the constant Wnt activation of

cyclin D1 and the proto-oncogene MYC. In addition to its role in cancer, the Wnt pathway regulates normal colon cell proliferation in a space-dependent manner, a process which is essential for the constant regeneration of the intestinal epithelial cells, which shed every three to five days (Peifer, 2002). Wnt pathway is activated in the base of proliferative colon crypts. Conversely, Wnt pathway is deactivated in colon epithelial cells that undergo differentiation and migrate to the villi tips. Protein p21, a cell-cycle inhibitor, was identified to be the pivotal target of Wnt pathway in the regulation of colon proliferation (van de Wetering et al., 2002). In the colon crypts, Wnt activation down-regulated p21, thus allowing proliferation. Away from the crypts, in the absence of Wnt activation, p21 expression induced cell-cycle arrest and differentiation. Therefore, inappropriate activation or deactivation of the Wnt pathway has the potential to disrupt orderly formation of the colon epithelial barrier or to encourage tumour formation.

Less has been elucidated about the β -catenin independent Wnt pathways, namely the planar-cell-polarity or the Wnt/calcium pathways, which may prove to be overlapping pathways (Sheldahl et al., 2003). In these alternate pathways, Wnt binds Frz receptors and activates DSH, but downstream signals are mediated by different components, such as the small GTPase proteins Rac and Rho, Rho kinase (ROK), c-Jun N-terminal kinase (JNK), or activation of calcium/calmodulin-regulated protein kinase II and PKC, mediated by an increase in intracellular Ca^{2+} . The β -catenin independent Wnt pathways in vertebrates function to regulate cytoskeletal organization and polarity during embryonic development (Miller, 2002).

1.3 Microarray expression screening

Given that the pathogenesis of IBD results from a complex interaction of different factors, it is fitting that a battery of screening methods, such as genome-wide linkage and association analysis and microarray expression screening, are employed to investigate this disease at the DNA or RNA level. Genetic analysis has identified nine IBD loci and markers within numerous genes (*CARD15*, *SLC22A4*, *SLC22A5*, *DLG5*, *AGR2*) that are associated with one or both IBD subtypes. While gene linkage approaches have been in use for over a two decades, microarray expression screening has emerged within the last six years to become a mainstay technology in the research laboratory. Microarray technology has been used in various applications, such as tumour classification and

prediction of clinical outcomes (Beer et al., 2002; Middleton et al., 2002; Nielsen et al., 2002), identification of genes involved in various diseases, cellular processes or induced responses to external stimuli (McDonald and Rosbash, 2001; Middleton et al., 2002; Ohki et al., 2002), and deconvolution of biological pathways (Yoshimoto et al., 2002), to name but a few.

The power of microarray expression screening is its ability to monitor the gene transcript levels of tens of thousands of genes in a single experiment. A typical microarray study involves comparing global mRNA expression profiles of different groups of samples, for example, different time points or stages of disease. Needless to say, the advent of microarrays has resulted in a rapid accumulation of raw data and necessitated the development of analysis methods (i.e. normalization techniques, application of statistical tests to determine regulated genes, classification methods) and data management structures to handle the immense volume of microarray data. To manage the flood of microarray data, there is a growing consensus in the scientific community of the benefit in establishing public repositories for gene expression data, analogous to freely available databases for sequence information, such as GenBank and EMBL. This endeavor has already been undertaken with the creation of public depositories for expression data, such as ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>) and Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>). Common standards for data input, annotation, information on experimental design, and data normalization are implemented so that the data is in a comprehensible format for biologists and bioinformatic specialists alike (Stoeckert et al., 2002).

The volume and diversity of microarray data available from public repositories offers tremendous opportunities for large-scale data analysis. Mining of public data sets can increase the confidence in a particular result (Geschwind, 2001) and conserve resources by redirecting research efforts in other directions, if meaningful data can already be extracted from existing data sets. For example, a statistical model for performing meta-analysis of independent microarray data sets has been developed using prostate cancer microarray data (Rhodes et al., 2002). Though the opportunity for meta-analysis is promising, few studies have addressed the compatibility of data sets generated by different expression screening platforms.

1.4 Aims of the study

Since the inception of DNA microarrays (Gress et al., 1992; Lennon and Lehrach, 1991; Schena et al., 1995), technological advances have led to the development of two main types of arrays, namely cDNA clone-based and oligonucleotide-based arrays (Jordan, 2002). Both expression systems, each with different experimental designs, are routinely used to compile comparative global mRNA expression profiles of tissues or cell lines. Various technical aspects of microarrays have been examined, including probe length and composition, cross-hybridization, and hybridization effects from immobilized substrate (Kane et al., 2000; Okamoto et al., 2000; Stillman and Tonkinson, 2001). To date, direct comparisons between the two systems have only been carried out using publicly available data sets (Kuo et al., 2002) or differential expression of defined model systems in smaller genes sets, all using fluorescent labeling methods (Yuen et al., 2002). The first aim of the present study (Figure 1.2, Panel 1) was to evaluate how the performance of the two most widely used platforms compare with regard to genes identified as expressed.

The second objective of this study (Figure 1.2, Panel 2) was to use microarrays to investigate differential gene expression in both inflamed and non-inflamed IBD samples and normal controls. Within the field of IBD research, two main studies have been published, in which microarrays were used to detect differential gene expression between: 1) controls (3 normal and 3 inflamed control individuals) and UC (7 patients) (Dieckgraefe et al., 2000); or 2) normal controls (6 patients) and IBD samples (6 patients of each subtype), which were pooled for microarray analysis (Lawrance et al., 2001). Both studies used arrays that were capable of monitoring ~7000 transcripts. In contrast, the present work expands the microarray analysis to > 22 000 transcripts, includes more samples and does not pool the samples for microarray analysis. In order to decipher changes in the molecular profiles of inflamed and non-inflamed IBD, three distinct and comprehensive microarray experiments were carried out. In the first part (Figure 1.2, Panel 2.A; Section 3.2), cDNA microarrays (Human UniGene Set RZPD 1 clone set) were used to detect differential expression between 11 samples from normal controls and 10 samples from each of inflamed UC and inflamed CD. The second part (Figure 1.2, Panel 2.B; Section 3.3.2) involves using oligonucleotide arrays (Affymetrix Human Genome U133A) to examine differences between 10 matched UC non-/inflamed samples from the same patient and 10 normal controls. The last part (Figure 1.2, Panel

2.C; Section 3.3.3) involves using oligonucleotide arrays (Affymetrix Human Genome U133A and B chips) in the expression analysis of 7 inflamed CD, 8 non-inflamed CD (4 of which are matched samples from the same patient) and 25 normal controls. The results from the three approaches contribute to our understanding of molecular mechanisms in IBD in a complementary manner.

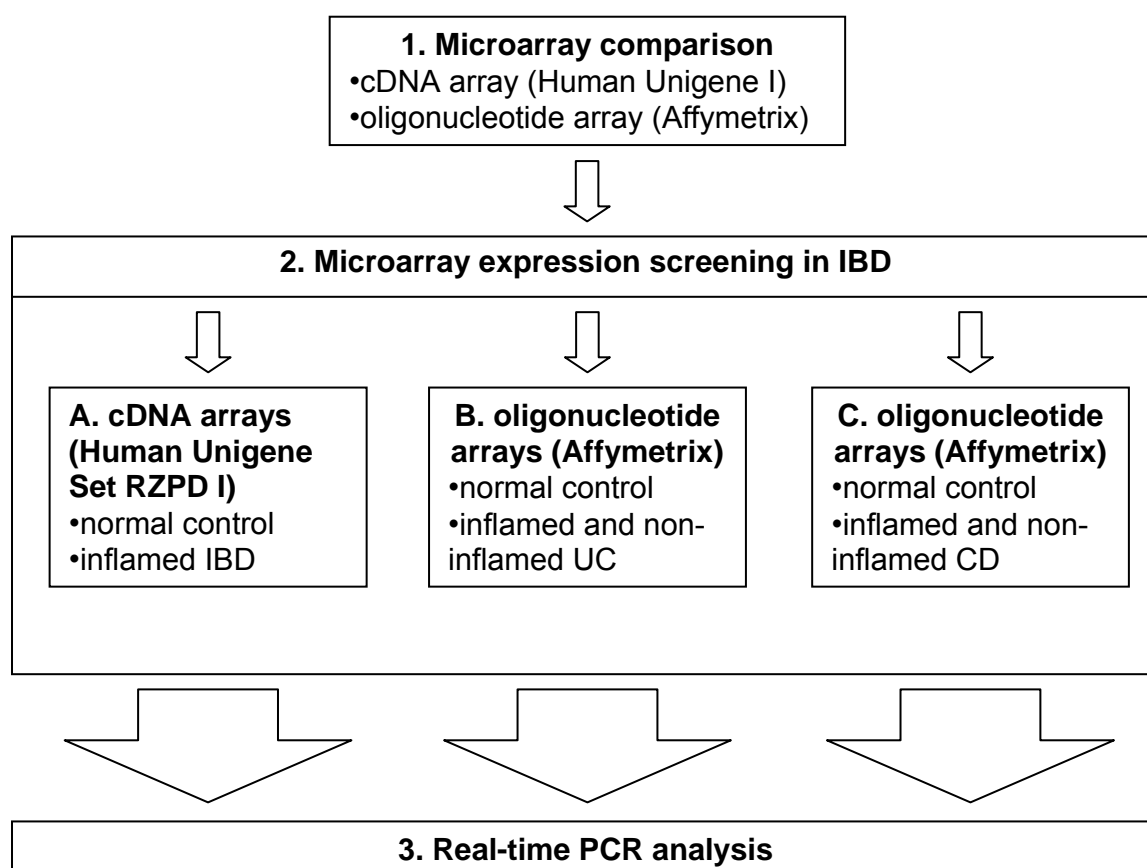


Figure 1.2 Flow diagram illustrating the methodology used in this study.

As a follow-up to the microarray results, high-throughput 384-well plate format real-time PCR (Figure 1.2, Panel 3) was used to further verify the gene transcript expression in a larger number of patient samples (>300 samples, including normal controls, IBD samples and disease specificity controls). To date, this cohort represents the largest real-time cohort ever collected for the purposes of verifying differential transcript expression in IBD.

In summary, the objectives of this study were three-fold: **1)** to assess the comparability of data generated from two microarray expression platforms; **2)** to use microarray expression screening (both cDNA and oligonucleotide platforms) to identify novel genes involved in the pathogenesis of

IBD; **3)** to focus on processes identified by microarray screening and to analyze transcript expression in a larger number of patients using quantitative real-time PCR.

2 Methods

The work contained in this thesis was completed as several separate microarray projects. To facilitate the description of methods, methods for each project will be described as a separate unit. General materials, laboratory equipment, and solutions used in the course of this study are listed in Sections 8.1, 8.2, and 8.3. For all parts of the study, patients were recruited for the studies through the First Medical Clinic, University Clinic of Schleswig-Holstein, Campus Kiel. The study procedures were approved by the hospital ethical committee, and patients consented in writing to the additional research biopsies being taken 24 h prior to endoscopy. All colonic mucosa biopsies were obtained from the lower bowel unless otherwise indicated.

2.1 Microarray comparison

2.1.1 Patient cohort for microarray expression screening

In order to compare two different microarray platforms, five patients (2 females, 3 males; age range: 55-70 y) undergoing colonoscopy for routine cancer screening were selected. All biopsies were obtained from the sigmoid colon. An additional two biopsies from each patient, from the same region of the colon, were formalin-fixed and paraffin embedded. Pathohistological examination was performed by an independent pathologist who was given no information about the patients' health status. Clinical evaluation yielded no significant pathological findings in this group; therefore, these patients were classified as a normal population.

2.1.2 Isolation of total RNA from biopsies

All reagents, glassware and laboratory utensils to be used for RNA isolation were specially treated in order to minimize RNA degradation by RNAses. Solutions were prepared with 0.1% DEPC-treated distilled water and sterilized by autoclaving. All glassware, ceramic mortar and pestles, Teflon pestles and metal spatulas were cleaned with common laboratory washing detergent, rinsed thoroughly in distilled water and air-dried before wrapping in aluminum foil and baking at 180°C for 12-16 h before use, in order to inactivate any contaminating RNAses. All plasticware was

purchased as UV-sterilized consumables (50 mL conical tubes, pipet tips with aerosol filters) or RNase-free consumables (microfuge tubes).

2.1.2.1 Homogenization of colon mucosa biopsies

Mucosal biopsies were stored in liquid nitrogen prior to use. To minimize premature thawing of the biopsies during RNA isolation, homogenization equipment (mortar, pestle, polyethylene tubes) was cooled with liquid nitrogen prior to the start of the procedure. A commercial kit was used to isolate total RNA from biopsy material (RNeasy mini-kit, Qiagen) according to the manufacturer's protocol. Briefly, one mucosal biopsy was ground into a fine powder using a cooled Teflon mortar in a 1.5 mL microfuge tube. The powder was then mixed with 650 μ L of RLT lysis solution, incubated at room temperature for 10 min and vortexed for 30 s. The sample was then further treated by centrifuging the lysate through a QIA shredder column for 2 min at 14000 rpm (16000 xg) in a microfuge. The eluate was centrifuged for 3 min at 14000 rpm (16000 xg) and the cleared lysate was transferred into a new microfuge tube. One volume of 70% ethanol was then added to the lysate, mixed well by vortexing and loaded onto an RNA binding column fitted in a 2 mL collection tube. The solution was centrifuged through the binding column for 15 s at 10000 rpm (8000 xg). The mini-column was washed once by pipetting 350 μ L of RW1 wash buffer on to the column, incubating for 10 min at room temperature and then centrifuging 15s at 10000 rpm (8000 xg).

2.1.2.2 DNase treatment of RNA and final isolation of RNA

DNase treatment of the RNA was carried out while the RNA was still bound on the column. First, DNase stock solution was prepared by reconstituting lyophilized DNase enzyme (Qiagen) by the addition of 550 μ L of RNase-free water. This DNase stock solution (10 μ L, 2.7 Kunitz units/ μ L) was further diluted in 70 μ L RDD buffer (as provided in RNase-free DNase Kit). The diluted DNase (80 μ L) was pipetted directly on to the spin column membrane and allowed to incubate at room temperature for 15 min. Upon completion of the digestion time, the mini-column was washed once with 350 μ L buffer RW1 and twice with 500 μ L buffer RPE. Each wash step was completed by centrifuging the tube 15 s at 10000 rpm (8000 xg). To completely rid the column of wash solution, the column was then centrifuged 1 min at 14000 rpm (16000 xg). Total RNA was eluted from the mini-column into an RNase-free microfuge tube by pipetting 50 μ L of RNase-free water directly

onto the membrane, allowing to completely soak through the membrane at room temperature for 5 min and then centrifuging 1 min at 10000 rpm (8000 xg).

2.1.2.3 Quality control of RNA

RNA concentration was determined by measuring the absorbance of a diluted RNA sample (1:40) in a spectrophotometer at 280 nm and 260 nm. The concentration and was calculated using the following formula:

$$[\text{RNA}]_{\mu\text{g}/\mu\text{L}} = A_{260} \times 40 \mu\text{g RNA /mL} / A_{280} \times 1000 \mu\text{L/mL}$$

Typically, the A260/280 ratio was between 1.8-2.2, indicating a good yield of nucleic acid to protein ratio. The integrity of the RNA was assessed by one of two methods: 1) denaturing RNA gel electrophoresis; or 2) electrophoresis on an RNA 6000 Nano LabChip (Agilent). In the first instance, a 1.5% agarose gel was prepared in 1 X MOPS and 0.9225 M formaldehyde.

Approximately 1 μg of total RNA was heat-denatured in 10 μL of RNA loading solution at 70°C for 10 min and cooled on ice. RNA samples were electrophoresed on the agarose gel at 4-5 V/cm in running buffer of 0.1 X MOPS. After one hour of migration, the gel was photographed under UV light. In the second instance, RNA quality was measured by the Agilent 2100 Bioanalyzer, which used electrophoresis through channels in a microfluidic chip (Nano LabChip) to separate RNA molecules based on size. In both methods, the strong presence of 18S and 28S RNA molecules and the absence of smears at lower molecular weights indicated intact, high quality RNA suitable for further studies.

Genomic DNA contamination of RNA was assessed by amplification of a housekeeper gene (GAPDH) using the following primers: GAPDH_F2, 5'-ACCCACTCCTCCACCTTTGAC-3'; GAPDH_R2, 5'-CTGTTGCTGTAGCCAAATTCGT-3'. Approximately 1 μg of total RNA was used as a template for a PCR reaction containing 1X PCR Buffer, 1.5 mM MgCl_2 , 0.2mM dNTPs, 0.4 μM forward primer, 0.4 μM reverse primer, and 0.1 μL Taq polymerase in a total reaction volume of 50 μL . The cycling program was used as follows: 95°C at 2 min, then 40 cycles of 95°C 30s, 60°C 30 s, 72°C 30 s, and a final extension step of 72°C for 5 min. PCR product was mixed with 10 X DNA loading dye before electrophoresis on a 3% agarose mini-gel in 1X TAE buffer. After 45 min of

electrophoresis, the gel was photographed under UV light. The presence of a 101 bp band indicated that a genomic contamination was present in the RNA.

If the total RNA failed to meet any of the quality criteria, the total RNA was re-isolated from additional biopsies or RNA solution was subject to another round of binding, washing and DNase I digest on a fresh mini-column.

2.1.3 Production of cDNA microarray: Human Unigene Set RZPD1

The production of the cDNA microarrays was carried out by Dr. Christine Costello and Brigitte Mauracher in collaboration with Dr. Holger Eickhoff at Department Lehrach in Max Planck Institute for Molecular Genetics, Berlin (Mah et al., 2004). In brief, amplified PCR products from the Human Unigene Set RZPD 1 clone set (German Resource Center for Genome Research (RZPD), <http://www.rzpd.de/>), which consisted of approximately 34 000 cDNA clones and represented 20 000 UniGene clusters (Build #160), were spotted on 23 X 23 cm Hybond N+ nylon membranes (Amersham, Freiburg, Germany) using a Genetix robot (Genetix, München-Dornach, Germany) attached to a 384 pin gadget-head (250 mm pins). Each cDNA clone was spotted in duplicate within a 6X6 pattern including four blank spots and two *Arabidopsis thaliana* (Genbank Accession U29785) guide spots for gridding orientation. Two hundred and fifty-eight clones representing known genes of interest were additionally spotted to complement the Human Unigene Set 1. A complete description of the clone-based microarray platform is available under GEO accession GPL284.

2.1.4 Hybridization of cDNA microarrays

Using a commercial kit (Qiagen), 100-250 ng poly (A)+ RNA was isolated from 15 µg of total RNA from each patient. Reverse transcribed target cDNA from patients was labeled with ³³P-dCTP as previously described (Eickhoff et al., 2000), with the modification that the labeling reaction was incubated at 42°C for 1 h. RNA was hydrolyzed under alkaline conditions (0.3 M NaOH at 68°C for 20 min) and neutralized. *A. thaliana* cDNA (25 ng) was labeled with 50 mCi ³³P-dCTP (≥2500 Ci/mmol, Amersham) as previously described (Feinberg and Vogelstein, 1983). Unincorporated

radio-nucleotides were removed using MicroSpin G50 columns (Amersham). Prior to the first hybridization, unused filters were washed three times in 1 X SSC; 0.1% SDS at RT for 10 min followed by three washes in 0.1 X SSC; 0.1% SDS for 20 min at 80°C. Filters were pre-hybridized for 2 h at 50°C in hybridization solution (7% SDS, 50% formamide, 5 X SSC, 2% blocking reagent (Roche Diagnostics, Mannheim, Germany), 50 mM sodium phosphate (pH 7.0) and 100 mg/mL denatured salmon sperm DNA). Radiolabeled *A. thaliana* and patient cDNA were heat-denatured and cooled on ice prior to addition to pre-hybridized filter. Following overnight hybridization at 42°C, filters were washed three times in 1 X SSC; 0.1% SDS buffer for 10 min at room temperature followed by two washes for 30 min in 0.2 X SSC; 0.1% SDS at 65°C. Filters were exposed to imaging plates (BAS-MS 2325, Fujifilm, Kanagawa, Japan) for 24 h and scanned at 50 µm resolution on a FLA-3000G imaging system (Fujifilm). There were no saturated probe data points after exposure and scanning. Image gridding was carried out using VisualGrid® software (<http://www.gpc-biotech.com>), and the spot quantitations were imported into a custom-made laboratory information management system (LIMS) database for normalization. A complete description of the sample labeling, hybridization and raw output from spot quantitation software is available under GEO accession numbers GSM5994-GSM5998 (Mah et al., 2004).

2.1.5 Normalization of clone-based microarray data using Zipf's Law

A normalization procedure was developed in-house for the Human Unigene Set 1 RZPD array. Microarray data normalization was conducted in two stages, intra-array normalization and inter-array normalization. Because of the comparatively large size of this membrane-based system, it was observed that some portions of the array displayed higher intensity measures than others (patchiness). Intra-array normalization removed spatial intensity measurement biases by dividing the microarray into 2304 fields, each containing 36 spots (15 duplicated data spots, 4 background controls, and 2 positive controls for grid positioning). The minimum intensity value for the background spots in each field was used to normalize the 30 data points in that field using a simple log transformed global mean method. Data spots for which at least one duplicate was below the filter background median were removed from analysis. Outliers (data spots where duplicate measures were not similar) were identified from a distribution of the difference between duplicates (duplicate 1 minus duplicate 2). This distribution was divided into centiles and the tails of the

distribution where centiles contained less than 10 data points were considered outliers whose value were removed from analysis.

Inter-array normalization was conducted in accordance with the observation that microarray intensity distributions follow Zipf's law (Lu et al., 2005). Genes on the array system were ranked according to median expression intensity over all arrays. Normalization was conducted such that the regression coefficients of the log expression intensity vs. log rank distribution for each microarray was the same as that of the median data distribution of all microarrays. A scaling factor was applied to the pre-normalized data to prevent negative log values from occurring during data transformation, and the same scaling factor was applied to the post-normalized data to return the values to their original magnitude. As an indicator of confidence as to whether a gene could be considered 'present', detection p-values were calculated for each clone probe based on the distribution of background values on each filter. Signals with a detection p-value <0.05 were considered to be 'present'. All relevant raw data values and normalized values are publicly available under GEO accession numbers GSM5994-GSM5998.

2.1.6 Isolation of plasmid DNA

Bacterial clones of interest were cultured in 96-deep-well plates containing 1.4 mL of LB media and 100 µg /mL ampicillin per well. Cultures were grown 20-24 h at 37°C with shaking (220 rpm). Bacterial freezer stocks of the cultures were made by mixing 100 µL of overnight culture with an equal amount of 30% glycerol in LB in a 96-well plate and storing at -80°C. A commercial mini-prep kit (Montage Plasmid Miniprep96 Kit, Millipore) was used to isolate plasmid DNA from the remainder of the bacterial culture. First, the bacteria were pelleted by centrifuging the deep well blocks at 1500 g for 5 min. The culture supernatant was discarded and the bacterial pellets were resuspended in 100 µL solution 1 (cell resuspension solution containing RNase) by vortexing. Cells were then lysed by adding 100 µL of solution 2 (cell lysis solution) to each well, vortexing for one minute and then incubating at room temperature for 2 min to allow lysis to complete. Next, 100 µL of solution 3 (neutralization solution) was added to each well and mixed vigorously on a plate shaker for 2 min. The resulting lysate was then cleared by pipetting the lysate into a MultiScreen

lysate clearing plate, placing this plate on top of the MultiScreen96 plasmid plate, and then applying vacuum (0.27 bar) for 3 minutes or until the lysate was drawn through the clearing plate and into the plasmid plate. The clearing plate was discarded. To isolate the plasmid DNA, the eluate in the plasmid plate was drawn through the DNA-binding membrane by applying vacuum (812.7 millibar) for 5 min or until the all liquid was completely drawn through. Next, the plasmid plate wells were washed by pipetting 200 μL of solution 4 (nuclease-free water) into each well and applying vacuum as before. Plasmid DNA bound to the plasmid plate was then isolated by applying 50 μL of solution 5 (tris buffer for storage) to each well, incubating the plate for 30 min at room temperature and finally pipetting the resuspended plasmid DNA from the plasmid plate into a fresh 96-well plate for storage. Plasmids were stored at -20°C until use.

2.1.7 Sequence verification of bacterial clones

Plasmid DNA was sequenced using Big Dye Terminator chemistry on an ABI 9700 Sequencer in 96-well plate format. Reactions for each well were prepared on ice using the components in Table 2.1. Various sequencing primers were used, including M13 forward/reverse and T7/T3 primers (Section 8.4 Primer and probe sequences).

Table 2.1 Components for sequencing of plasmid DNA

Component	Volume (μL)
Big Dye Terminator enzyme mix	3
Sequencing Primer (10 pmol/ μL)	1
Water (HPLC grade)	1
Plasmid DNA (~ 80 ng/ μL)	5
Total volume	10 μL

Reactions were carried out in 96-well Costar plates sealed with plastic foil. The clone inserts were amplified using the following cycling profile: 96°C for 10 s, then 30 cycles of 95°C for 20 s, 53°C for 30 s, and 60°C for 4 min. PCR products were then allowed to extend at 60°C for 7 min, and finally the reaction was stopped by pausing at 4°C .

Excess fluorescent dye was removed from the reaction by eluting the PCR reaction through a G-50 Sephadex column into a Microarray Optical 96-well reaction plate (Applied Biosystems). The resulting eluate was then subject to capillary sequencing on an ABI Prism 9700 Sequencer.

Chromatograms of the plasmid sequence were analysed using Sequencher 4.0.5 software (GeneCodes).

2.1.8 Hybridization and data analysis of Affymetrix U 95Av2 microarrays

Total RNA was isolated from sigmoid biopsies of five normal controls (Sections 2.1.1 and 2.1.2). For the purpose of the microarray comparison, the RNA from each patient was divided into two aliquots. One aliquot was sent to a collaboration partner (Dr. Anders Thelin, AstraZeneca R&D, Mölndal, Sweden), where the RNA samples were hybridized to Affymetrix U95Av2 microarrays, and the resulting normalized microarray data was shared with the Institute for Clinical Molecular Biology in Kiel. The second aliquot of RNA was hybridized in-house to cDNA microarrays (Sections 2.1.4) and normalized using procedures developed in-house (Section 2.1.5).

2.1.8.1 Hybridization of Affymetrix U 95Av2 microarrays

Affymetrix arrays were hybridized according to standard manufacturer's protocol. Briefly, 15 µg of total RNA was used to generate double-stranded cDNA using Superscript reagents (Life Technologies) and a T7-linked oligo(dT) primer. cRNA were synthesized using the Enzo Bioarray High Yield RNA transcript labeling kit from Affymetrix, resulting in biotinylated cRNA. Labeled cRNA were cleaned using Qiagen RNeasy kit and fragmented into 35-200 bp lengths using fragmentation buffer supplied by Affymetrix. Spike controls B2, bio-B, bio-C, bio-D and Cre-x were added to the hybridization cocktail before overnight hybridization at 45°C for 16 hrs. Arrays were stained and washed using the EukGE-WS2 protocol (a dual staining) before being scanned on an Affymetrix scanner.

2.1.8.2 Microarray suite software (MAS) 5.0

Absolute analysis (Affymetrix Microarray Analysis Suite 5.0) was performed on the arrays using the target intensity value of 150 (global scaling, no baseline) and Tau=0.015 (default parameter which affects discrimination between match/mismatch probe pairs). Target intensity of 150 was chosen based on previous experience with Affymetrix arrays, which showed that this value typically yielded adequate scaling factors (between 0.5 and 10). In contrast to the conventional cutoff of $p < 0.05$ used for the clone-based platform, a signal for an oligonucleotide probe set was determined to be 'present' if the detection p -value of the probe set was less than 0.06, which, for the purposes of this

comparison, included 'present' and 'marginal' calls assigned by MAS5.0 default criteria. The complete datasets are available under GEO accession numbers GSM5999-GSM6003.

2.1.9 Matching of clone-based PCR products to corresponding oligonucleotide probes

For the purpose of the microarray platform comparison, it was necessary to identify a set of genes that were common to both the cDNA and the oligonucleotide microarray platforms. A detailed flowchart of the experimental setup and probe-matching procedure is outlined in Figure 2.1. Two different methods were used to identify probes that theoretically represented the same gene. In a more general approach, cDNA clones or Affymetrix exemplar sequences were matched to human UniGene clusters (build 160) through batch web search tools (<http://genome-www5.stanford.edu/cgi-bin/SMD/source/sourceBatchSearch>). As each UniGene cluster should represent a nonredundant gene (<http://www.ncbi.nlm.nih.gov/UniGene/>), the term "gene" and "UniGene cluster" will be used interchangeably in this thesis. Probes from either system that were grouped to the same UniGene cluster were considered as "matched probes." In cases where a gene was represented by multiple probes, gene detection was only counted once, based on the probe that gave the lowest P value, and was expressed in the highest number of patients. Using this UniGene method of matching, a set of 6645 genes were defined to be present on both platforms. The second approach of gene matching was more sequence specific. Full clone sequences were assembled from sequence-verified data from the 5' and 3' insert ends (up to 700 bp from each end). Public sequence data was used to bridge short gaps (less than 100 bp) between the ends of 5' and 3' sequences in cases where the clone ends did not overlap, giving rise to 272 "reconstructed" full-length clones. The corresponding probes in the Affymetrix platform were identified through BLAST homology searching in the target and probe databases available from NetAffx Analysis Center (<http://www.affymetrix.com/analysis/index.affx>). Probes were considered to be matched if the clone sequence had a BLAST homology score greater than 200 in the target database or if the clone sequence matched more than 10/16 probe pairs with >75% identity. All full-length clone sequences were screened for Alu repeat sequences using the Repeatmasker Web Server (<http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker>).

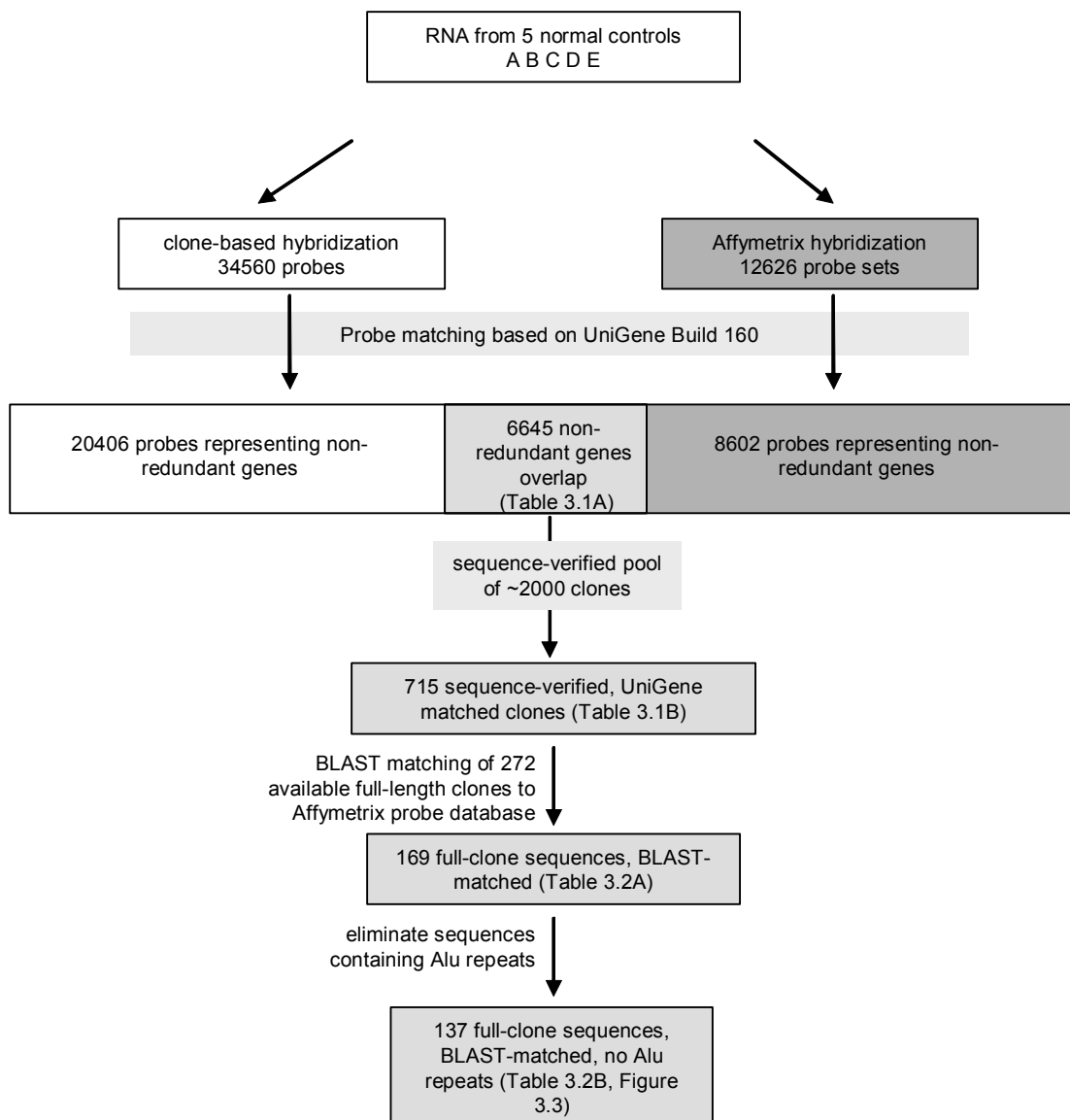


Figure 2.1 Flow-chart of comparison analysis.

Schematic diagram shows the flow of analysis starting from hybridization of five patient samples to both microarray platforms and progresses through probe matching steps with increasing stringency.

2.1.10 Statistical analysis of expression data for microarray comparison

After normalization of microarray data in each respective protocol, the datasets were tested for normality. The distribution of the mean expression data from each expression platform was tested by Kolmogorov-Smirnov test using Analyse-it software in Microsoft Excel. As the mean expression data from each platform were not normally distributed, Spearman rank order coefficients were used to determine the extent of correlation in expression levels between the two systems. To obtain a detection P value for each probe on the array from all five samples, individual P values for each

gene were combined according to the Fisher method for combining probabilities (Sokal and Rohlf, 1995).

2.1.11 Real-time PCR quantitation of selected transcripts

Selected genes (*ACTB*, *CDH11*, *MMP1*, *APOA1*, *PPBP*, *CLDN4*, *CEACAM1*, *MPG*, *PRKCBP1*, *TIMM17B*, *MMP3*, *ALDOB*, and *MUC1*) observed to be present and/or absent on the microarrays were selected for real-time PCR quantification. Primers and probes for real-time PCR were labeled with the fluorescent reporter dye 6-carboxyfluorescein (6-FAM) and the quencher dye 6-carboxytetramethylrhodamine (TAMRA) at the 5' and 3' ends, respectively (Eurogentec, Seraing, Belgium; or Applied Biosystems), and the sequences are supplied in section 8.4. Total RNA (1 µg) from each patient sample was reverse transcribed to cDNA according to manufacturer's instructions (MultiScribe Reverse Transcriptase, Applied Biosystems). Patient cDNA was diluted 1:10, and 5 µL of the cDNA was pipetted in duplicate into a 384-well plate format. Real-time PCR was carried out using the ABI Prism 7900HT Sequence Detection System (Applied Biosystems) in the following 10-µL reaction: 0.9 µL Universal TaqMan Master Mix (Applied Biosystems), 200 nM 6-FAM probe, 300 nM forward and reverse primers, and 5 µL cDNA diluted template. The PCR cycling profile was as follows: 2 min at 50°C, 10 min at 95°C and 40 cycles of 95°C for 15 s, 60°C for 1 min.

All patient samples for each gene assay were analysed in duplicate on one plate, and each gene assay was done in triplicate (except for *MMP1* and *TIMM17B*). Real-time runs were recorded and analyzed using SDS 2.1 software (Applied Biosystems). For an absolute determination of transcript levels, the number of cycles (Ct) required to reach the threshold (set to 0.2) was taken to be a measure of transcript abundance.

2.2 Identification of genes involved in IBD using cDNA microarrays

2.2.1 cDNA microarray expression screening patient cohorts

2.2.1.1 Patient cohort for cDNA microarray expression screening

Group 1 patients included 31 individuals (Table 2.2) for expression profiling on cDNA microarrays. Eleven individuals (5 female (F); mean age: 51.4 y (range: 25-84 y)) were included in the study as the normal population, with endoscopic and histological examination yielding no significant pathological findings in this group. Indications for colonoscopy in this group included colonic cancer surveillance and previous non-specific changes in stool habits. To investigate gene expression in IBD, ten patients with active CD (4F; mean age: 33.6 y (range: 19-66 y)) and ten patients with active UC (2F; mean age: 34.8 y (range: 27-51 y)) were recruited. All endoscopic biopsies were taken from a defined area of the sigmoid colon (at 20-30 cm measured during withdrawal), and immediately snap-frozen in liquid nitrogen. Clinical disease activity was documented using established clinical parameters of the Crohn's disease activity index (CDAI) in CD (Best et al., 1976) and using the colitis activity index (CAI) (Rachmilewitz, 1989; Truelove and Witts, 1955) for UC patients. In order to qualify for inclusion into the study, IBD patients had to be clinically (CDAI>150 or CAI≥4) and endoscopically active at the time of sampling, but free of all medication (other than low dose 5-ASA) for a minimum of 6 weeks prior to endoscopy.

Table 2.2 Clinical characteristics of the IBD patient group for cDNA microarray analysis

	Hospitalized normal controls	CD patients	UC patients
No of patients	11	10	10
Sex (F/M)	(5/6)	(4/6)	(2/6)
Age (y) (mean (range))	51.4 (25-84)	33.6 (18-66)	34.8 (27-51)
Region of biopsy			
Sigmoid colon	11	10	10
Descending colon	0	0	0
Caecum	0	0	0
Transverse colon	0	0	0
Inflammation activity*			
Active	0	10	10
Inactive	11	0	0
Medication			
No treatment	10	6	6
Steroids	0	0	0
5-ASA	0	2	3
NSAID	1	1	1
Antibiotic	0	1	0
Other	0	0	0

2.2.1.2 Patient cohort for real-time PCR verification

In order to verify the cDNA array findings, the initial patient collection described in Table 2.2 was extended to 100 individuals for real-time PCR analysis (Table 2.3). IBD patients used in Group 2 consisted of 27 CD patients (17F; mean age 32.1 y, (range: 17-45 y) and 35 UC patients (20F; mean age 33.8 y, (range: 18-74 y). All patients were selected using similar criteria as outlined above i.e. all patients had active inflammation in the colon and were clinically active (CDAI>150 or CAI \geq 4). Patients were allowed 5-ASA or glucocorticoids, but not immunosuppressants or biologicals. The hospitalized normal control population consisted of 21 individuals (11F; mean age 48.9 y (range 16-84 y) which yielded no significant pathological findings following endoscopic examination for changes in stool habits, abdominal pain, upper gastrointestinal bleeding or cancer surveillance. As disease specificity controls, 17 patients (9F; mean age 56 y (range 17-80 y) with colonic disease but not IBD, were also included (Table 2.3). These disease specificity controls (DC) included 9 non-inflamed and 8 inflamed biopsies samples from patients with infectious diarrhea, other forms of gastrointestinal inflammation or irritable bowel syndrome. No restriction on medications was specified in the inclusion criteria for disease specificity controls and all concomitant medications are listed in Table 2.3. A minimum of 19 normal controls, 16 UC, 19 CD and 14 DC samples were used for each real-time assay, depending on the availability of the patient samples and the plate layout at the time the real-time cDNA plates were produced (Section 2.4.3).

Table 2.3 Characteristics of extended patient group for real-time PCR verification of cDNA microarray results

	Hospitalized normal controls	CD patients	UC patients	DC patients
No of patients	21	27	35	17
Sex (F/M)	(11/10)	(17/10)	(20/15)	(9/8)
Age (y) (mean (range))	48.9 (16-84)	32.1 (17- 45)	33.8 (18-74)	56 (17–80)
Region of biopsy				
Sigmoid colon	13	26	35	16
Descending colon	1	1	0	0
Caecum	7	0	0	0
Transverse colon	0	0	0	1
Inflammation activity*				
Active	0	27	35	8
Inactive	21	0	0	9
Diagnosis (applicable to DC patients)				
Infectious diarrhea	NA	NA	NA	8
Gastrointestinal inflammation	NA	NA	NA	7
Irritable Bowel Syndrome	NA	NA	NA	2
Medication				
No treatment	14	9	11	9
Steroids	0	14	11	1
5-ASA	0	9	18	1
NSAID	2	0	1	3
Antibiotic	1	2	0	4
Other	4	2	4	4

* Inflammation activity of biopsied site, as observed by the endoscopist

Abbreviations used: NA = not applicable, DC = Diseased control patient, 5-ASA: 5-Aminosalicylic acid, NSAID: Non-steroidal anti-inflammatory drug

2.2.2 Microarray data analysis

Total RNA was isolated from biopsies of patients as described in Sections 2.1.2 and 2.2.1.1, and hybridization to Human Unigene I cDNA microarrays was carried out as previously described in Section 2.1.4. The cDNA microarrays were washed and scanned as previously described (Section 2.1.4).

The microarray datasets were normalized according to an in-house method based on Zipf's law (Lu et al., 2005). A background cut off, set to the mean background plus two standard deviations of the background, was used to distinguish detectable, 'present' signals from 'absent' signals, which were set to zero. A fold-change cut off was determined based on the technical variance obtained when the same RNA sample (obtained from a intestinal surgical specimen) was hybridized to five separate arrays of one production batch. Fold changes were calculated between all pairwise

combinations of the five arrays for all 33 792 spotted cDNA probes. For each unique pairwise combination, the fold change observed at the 95th percentile of the distribution was noted. The average of these values (fold change of 1.2) was taken to be the threshold fold change value for differential expression using these cDNA microarrays. Significantly differentially regulated genes were defined to be those genes whose expression was statistically different in two groups (distribution-free Mann-Whitney U-test) and whose fold change was greater than the threshold for technical variance, 1.2. The reported fold changes were calculated based on the median of each experimental group.

To identify broad functional groups of genes that were significantly regulated between normal controls and the IBD subtypes, the frequency of differentially regulated genes associated with a particular functional group was compared to the total number of genes on the array associated with the particular functional group. In this manner, an enrichment of differentially regulated genes belonging to a particular subgroup could be determined using a statistical test based on a hypergeometric distribution (Tavazoie et al., 1999).

2.2.3 Prediction of function in unknown genes

Selected clone probes that were differentially regulated but not associated to known genes were further analysed in detail to decipher possible function. First, cDNA clone sequences, obtained by methods as described in Sections 2.1.6 and 2.1.7, were searched for homology against the GenBank nonredundant database using BLAST alignment. Next, the highest scoring BLAST hit (sequence identity close to 100%) was used to extend the sequences to longer mRNA transcripts. InterPro (<http://www.ebi.ac.uk/InterProScan>), SMART (<http://smart.embl-heidelberg.de>) and Pfam (<http://www.sanger.ac.uk/Software/Pfam>) search methods (with standard parameter sets) were used to detect known protein domains and functional sequence motifs in putative open reading frames (ORFs) longer than 100 amino acids. Additionally, close homologies of the ORFs were found through association to UniGene clusters in humans and other species. In cases of ORFs without a detectable homology to known genes, these transcripts were mapped to genomic locations by BLAST searching and examined in terms of their genomic context.

2.2.4 Verification of cDNA microarray results by fluorogenic real-time PCR

Selected genes from the microarray results were independently quantitated by real-time PCR on an extended patient cohort (Section 2.2.1.2 and Table 2.3). Probes and primers were either designed to non-redundant sequences using Primer Express V2.0 (Applied Biosystems, Foster City, California, United States), or ordered from Applied Biosystems as Assays-on-Demand Gene Expression Assays (Section 8, Appendix, Table 8.3). Total RNA (1 µg) was reverse-transcribed to cDNA according to the manufacturer's instructions (Multi-Scribe Reverse Transcriptase, Applied Biosystems). Reactions were carried out on the ABI PRISM Sequence 7700 Detection System (Applied Biosystems) and relative transcript levels were determined using β -actin as the endogenous control gene. A detailed description of the cDNA preparation and real-time PCR procedure is described in Section 2.1.11. Statistically significant differences between control and IBD samples were determined using a Mann-Whitney U test. A p-value of less than 0.05 was considered significant.

2.2.5 Immunohistochemistry

Paraffin-embedded biopsies from normal controls (n=5) and from patients with CD (n=5) and UC (n=6), which were obtained in parallel from the same sites as the biopsies used for the expression analysis studies, were analysed by Drs. Philip Rosenstiel and Ralph Lucius. Immunohistochemical staining was completed according to a protocol by Waetzig and co-workers (Waetzig et al., 2002). Briefly, 7-µm sections were subjected to heat-induced antigen retrieval in 0.01 M EDTA solution, pH 8 for 10 min. After blocking in 0.75% BSA solution in PBS for 20 min binding, the sections were washed three times in PBS, and incubated for 1 hr with the respective primary antibodies (monoclonal carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) 4D1/C2, protein kinase C beta 1 (PRKCB1; Pharmingen, San Diego, CA, USA), and casein kinase 1, delta (CSNK1D; Santa Cruz, Santa Cruz, CA, USA). The sections were washed in PBS (3 times for 10 min), incubated with peroxidase-conjugated rabbit anti-mouse secondary antibody (Sigma, Deisenhofen, Germany; 1:100) for 30 min, washed, and stained with goat anti-rabbit IgG (Sigma, 1:200, 30 min), before processing with diaminobenzidine and embedding in Aquatek (Merck, Inc., Hawthorne, NY). For specificity controls, (i) the primary antibodies were omitted and (ii) normal

sera were used from the species in which the primary antibodies were raised. No specific staining could be detected using either of these combinations (data not shown).

2.3 Identification of genes involved in inflamed/non-inflamed IBD using Affymetrix HG-U133 arrays

2.3.1 Patient cohorts for microarray analysis on HG-U133 Affymetrix arrays

2.3.1.1 Inflamed and non-inflamed UC cohort

In order to determine differential gene expression in UC patients, 10 UC patients were recruited for this section of the study. Biopsies were taken from both a non-inflamed and an inflamed region from each patient. Additionally, 10 healthy normal controls were also recruited (Table 2.4). All biopsies were taken the lower bowel, with the majority of biopsies taken from sigmoid colon or caecum. Total RNA was prepared from each of the biopsies (Section 2.1.2) and sent to a collaboration partner (Dr. Jun Li, R&D Center, Boehringer Ingelheim Pharmaceuticals, Ridgefield, CT, USA), for hybridization on HG-U133A Affymetrix arrays.

Table 2.4 Patient cohort for UC expression study

	Healthy normal control	UC
No. of patients	10	10
Sex (F/M)	5/5	5/5
Age (y) (mean (range))	30 (20-42)	37 (20-66)
Biopsied region		
Rectum	0	1
Sigmoid	10	7
Ascending colon	0	1
Descending colon	0	2
Caecum	0	9
Inflammation activity		
Active	0	10
Inactive	10	10
Medication		
No treatment	9	6
Steroids	0	0
5-ASA	0	3
NSAID	1	1
Antibiotic	0	0
Other	0	0

2.3.1.2 Inflamed and non-inflamed CD cohort

In order to determine differential gene expression in CD patients, 25 healthy normal controls and 15 CD patients were recruited for this section of the study (Table 2.5). Biopsies were obtained from inflamed (n=8) or non-inflamed (n=7) regions of 11 untreated CD patients. Of these 15 biopsies from the untreated CD patients, four pairs of inflamed/non-inflamed biopsies were obtained from the same patient (matched samples). In the remaining four CD patients, non-inflamed biopsies (n=4) were obtained from immunosuppressive-treated CD patients. Total RNA from each of the biopsies was isolated from each of the biopsies (Section 2.1.2) and sent to a collaboration partner (Dr. Stefan Pierrou, AstraZeneca R&D, Mölndal, Sweden), for hybridization on HG-U133A and HG-U133B Affymetrix arrays.

Table 2.5 Patient cohort for CD expression study

	Healthy normal control	CD
No. of patients	25	15
Sex (F/M)	11/14	10/5
Age (y) (mean (range))	28 (19-50)	35 (22-56)
Biopsied region		
Rectum	0	0
Sigmoid	25	12
Ascending colon	0	2
Transverse colon	0	2
Descending colon	0	0
Caecum	0	3
Inflammation activity		
Active	0	8
Inactive	25	11
Medication		
No treatment	22	9
Steroids	0	2
Azathioprine	0	3
5-ASA	0	6
NSAID	2	0
Antibiotic	0	1
Other	1 [†]	0

[†]topical treatment for nasal congestion

2.3.2 Microarray hybridization and image processing

Using methods described in Section 2.1.2, total RNA was isolated from biopsies of patients described above (Sections 2.3.1.1 and 2.3.1.2) and sent to collaboration partners for hybridization on Affymetrix HG-U133A arrays, which contain 22 000 probe sets and represent 13 239 unique

genes. Hybridization and scanning of the arrays was carried out according to manufacturer's protocols (Section 2.1.8.1).

2.3.3 Affymetrix HG-U133 microarray data analysis

All microarray analysis, including normalization and data exploration (determination of differentially expressed genes, Gene Ontology analysis) was done using open-source tools from Bioconductor (version 1.7; <http://www.bioconductor.org/>) (Gentleman et al., 2004) and the statistical language R (version 2.2.0; <http://www.r-project.org/>). Additionally, scripts in Perl (version 5.8.7; <http://www.cpan.org/>) and MySQL (version 4.1.15; <http://www.mysql.com/>) databases were used for data processing, storage and retrieval. Gene Ontology database (September 2005 version) and Affymetrix annotations (June 2005 version) were used for all analyses.

An overview of the steps taken in Affymetrix microarray data analysis is shown in Figure 2.1.

Starting out with the raw Affymetrix datafiles (*.CEL files), the files were read into an R object using the Bioconductor package *affy* (Gautier et al., 2004). Next, the false colour image of each array was inspected to identify any gross image defects caused by incomplete hybridization or washing, dust particles, saturated or low signals. The raw signals were plotted in a histogram to check for over-saturation.

In contrast to the spotted cDNA microarrays that contain a single clone insert (spotted in duplicate) to represent a gene transcript, the Affymetrix U133 microarrays use 11 pairs of matched/mismatched oligonucleotides to make a probeset which represents the gene transcript. This technical difference necessitates the use of statistical methods which summarize the match and mismatch oligonucleotide pairs into a single expression value. The normalization methods available from the Bioconductor package *affy* contain different methods to first summarize the probe level data and then normalize the summarized probe data. To determine the best normalization method, four different normalization methods were used, including the method provided by Affymetrix (MAS5.0) and three other methods from Bioconductor (GC-RMA, *loess*, *vsn*) (Bolstad et al., 2003; Huber et al., 2002; Irizarry et al., 2003; Wu et al., 2004). The performance of each normalization was determined by the following criteria: 1) ability of normalization to reduce

variance without increasing bias; 2) ability of the normalization procedure to be carried out within a reasonable amount of computing time.

The normalized datasets were subject to filtering by background and variability cutoffs. A background cutoff was calculated for each array, based on a group of bacterial probesets present on the U133 arrays. Since these probesets were bacteria-specific and did not hybridize to human target sequence, signals arising from hybridization of the human targets to these bacterial probesets could be taken as a measure of random hybridization. The background cutoff was set to two standard deviations above the mean background. Based on these cutoffs, probesets were considered 'present' if the signal was higher than the background cutoff. Probesets that were at least 70% present in all arrays and were in the top 25% of the most variable genes (as determined by interquartile range) were considered for further analysis.

A key step in the analysis of microarray data is the determination of genes which are differentially regulated between two defined groups. For pairwise comparisons (between two groups), distribution-free methods, such as permutation tests and rank-based tests, and distribution-dependent methods, such as t-test, were available. In cases where matched samples were available, paired statistical tests were applied. Finally, a linear modeling method (R package: limma (Smyth, 2004)) was used to make multiple comparisons within a dataset. Since the process of testing multiple hypotheses increases the chance of committing false positives, a multiple testing correction was applied to reduce the possibility of false positives. This was accomplished by controlling the False Discovery Rate (FDR) (Benjamini and Hochberg, 1995) as an additional measure of statistical stringency.

Though lists of differentially regulated genes reveal many individual gene changes between two groups, an overall biological context for the differential regulation may be obscured from view. Gene Ontology (GO) is a set of controlled vocabulary intended to standardize the naming of biological processes, molecular functions and cellular components associated with all genes. Though GO contains many associations to genes, many associations between GO terms and genes may not yet be annotated in GO, either because the associations are not yet manually/electronically curated in the GO database, or associations may be misleading because

they are specific to an organism or set of experimental conditions or they are predicted by electronic annotation. Despite these disadvantages, GO can still be used in an exploratory manner to try to identify broader functional groups from lists of differentially regulated genes. Clues to the functional basis of differentially regulated genes may be gleaned by examining functionally related groups of genes (as classified by Gene Ontology) and their enrichment in the subset of differentially regulated genes.

In order to place the differentially regulated probesets into a biological context, the Fisher test for a 2X2 contingency table was used to calculate p -values for the likelihood that differentially regulated genes were enriched (better than chance) within a given Gene Ontology category. First, the primary probeset associations to GO categories were extracted from the HG-U133 annotation file as supplied by Affymetrix. These primary associations only represented the most specific GO term. In order to obtain all higher level associations in the GO tree, Perl scripts were used to follow all branch points in the GO tree from the primary association to the root GO term. In this way, all possible GO associations to a gene were obtained. Probesets without any associations to GO terms were omitted from the analysis, and in cases where a single probeset had multiple associations to GO terms, all associations were included in the analysis. Next, a 2X2 contingency table was created for each of the 18 824 GO terms existing as of September 2005. Using R scripts, all probesets deemed 'present' were classified for differential expression (true/false) and association to GO term (true/false). The Fisher Exact test was then used to calculate the probability that the frequencies varied from those expected by chance. A p -value of less than 0.05 was considered to be significant.

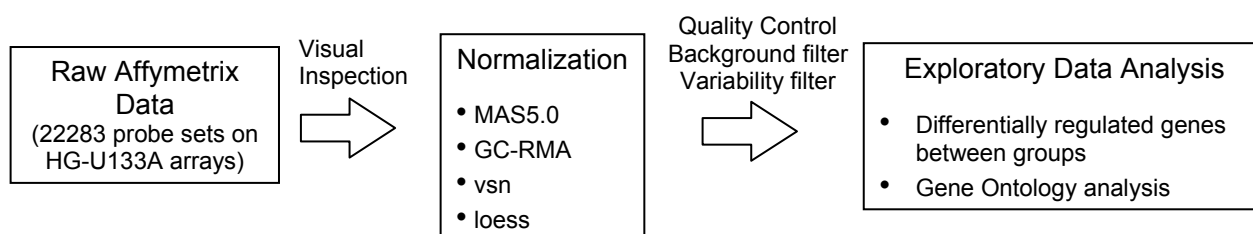


Figure 2.2 Flow diagram of HG-U133A microarray analysis in R statistical language

2.4 Real-time PCR analysis

To assay the gene transcript levels in an extended cohort of a total of 312 patients, cDNA from healthy normal controls, IBD patients and disease-specificity controls was arrayed on three 384-well plates for real-time PCR quantitation.

2.4.1 Patients for real-time PCR analysis

The full real-time PCR cohort consisted of Section 2.2.1.2, and included an additional group of healthy normal controls (Table 2.6).

Table 2.6. Patient and Sample Data for Complete Real-time PCR cohort

	Healthy normal control	Hospitalized normal control	CD	UC	DC
No. of patients	25	34	100	96	57
Sex (F/M)	13/12	18/16	58/42	50/46	31/28
Age (y) (mean (range))	28 (19-50)	55 (27-84)	37 (17-69)	37 (16-74)	47 (16-80)
Biopsied region					
Large bowel	25	34	100	96	57
Small bowel	0	0	0	0	0
Inflammation activity					
Active	0	0	42	57	15
Inactive	25	34	51	39	44
Medication					
No treatment	19	10	19	16	20
Steroids	0	1	36	26	2
5-ASA	0	0	20	44	2
NSAID	1	2	0	2	4
Antibiotic	0	2	4	3	5
Other	5	4	29	22	11

Due to the availability of the samples at the time the cDNA plates were produced (six plate batches were produced in the course of the present study), the plated number of samples per group was not constant between plate batch productions. Ranges of sample numbers used for the assays are listed in the relevant results tables.

2.4.2 cDNA synthesis

Total RNA was isolated from patient biopsies from patients as described in Section 2.1.2. For each patient sample, cDNA was reverse-transcribed from 1 µg of total RNA in a total volume of 100 µL

using the MultiScribe™ Reverse Transcriptase reagents (Applied Biosystems). Components of the reverse transcriptase reaction were mixed on ice (Table 2.7).

Table 2.7 Components for first-strand cDNA synthesis

Component	Volume for one reaction (μL)	Concentration in Final Reaction
10 X TaqMan RT buffer	10	1 X
25 mM MgCl_2	22	5.5 mM
dNTPs (2.5 mM)	20	500 μM
50 μM random hexamer primers	5	2.5 μM
Rnase inhibitor (20 U/ μL)	2	0.4 U/ μL
Multiscribe reverse transcriptase (50 U/ μL)	2.5	1.25 U/ μL
RNase-free water	variable	
1 μg total RNA	variable	10 ng/ μL
Total volume	100	

As a negative control for the reverse transcriptase reaction, the above amounts were halved and the reverse transcriptase was omitted from the reaction. All reactions were prepared on ice and then incubated at 25°C for 10 min, 48°C for 30 min and 95°C for 5 min. The reaction was stopped by addition of 2 μL of 0.5 M EDTA. The success of the first strand cDNA synthesis was then checked by using 2.5 μL of the first strand synthesis as a template for PCR amplification with GAPDH as previously described in section 2.1.2.3. The amplification of the 101 bp GAPDH PCR product in the positive reverse transcriptase reaction, but not in the negative reverse transcriptase reaction, indicated successful first strand synthesis without contamination from genomic DNA. The final theoretical concentration of cDNA, assuming 100% efficiency of reverse transcription, was 10 ng cDNA/ μL .

2.4.3 Plate production

cDNA prepared in section 2.4.2 was diluted 1:5 to a theoretical concentration of 2 ng/ μL before being added to 384-deep well plates. For relative quantitation, standard curve cDNA was prepared from reverse-transcription of a mixture of total RNA obtained from inflamed IBD and normal colonic mucosa to provide the greatest range of gene transcript expression. The standard curve was serially diluted in ten dilutions from 1:1 to 1:200 or 1:1000, in a total volume of 250 μL per deep well. For the real-time PCR reaction, 5 μL of the cDNA solution was added to the 384 well plate in quadruplicate: two wells for β -actin quantitation and two wells for target gene quantitation.

2.4.4 Gene expression assays on ABI prism 7900HT

Real time PCR assays were carried out using either commercial (Table 2.8) or self-designed primers and fluorescent probes (Table 2.9). Sufficient volumes of reaction mix were prepared in bulk and then dispensed by multipipet to 384-well PCR plates containing 5 μL of cDNA (2 ng/ μL). All reagents were kept cooled on ice during the pipetting procedure. Plates were sealed with PCR film and stored at 4 °C or -20°C if they were not immediately run.

Table 2.8 Components for Assay-on-Demand products

Component	Volume per well (μL)	Concentration in Final Reaction
cDNA template (2 ng/ μL)	5	1 ng/ μL
20 X Assay-on-Demand	0.5	1 X
2 X TaqMan® Universal PCR Master Mix	4.5	0.9 X
Total volume	10	

Table 2.9 Components for self-designed primers and probes

Component	Volume per well (μL)	Concentration in Final Reaction
cDNA template (2 ng/ μL)	5	1 ng/ μL
20 μM forward primer	0.15	300 nM
20 μM reverse primer	0.15	300 nM
10 μM FAM-labeled probe	0.2	200 nM
2 X TaqMan® Universal PCR Master Mix	4.5	0.9 X
Total volume	10	

Plates were briefly centrifuged and then run on an ABI 7900HT Sequence Detection System using the following thermocycling profile: 50°C for 2 min; 95°C for 10 min; and 40 cycles of: 95°C for 15 s; 60°C for 1 min.

Upon completion of the run, the data was analysed by the Sequence Detection System (SDS) 2.0 software using the following parameters: 1) threshold line set as low as possible such that the line intersects the amplification curves in the exponential phase yet any amplification in the non-template controls is reduced to negligible levels; 2) baseline set separately for housekeeper and target such that the baseline starts at three cycles and ends just before the first amplification curves rise above threshold; 3) extreme outliers in the standard curve eliminated from the calculation of the standard curve. Upon adjusting all of these parameters, the SDS software automatically calculated the relative quantities, which were then exported as a tab-delimited text file for further data processing.

2.4.5 Real-time PCR data processing

To facilitate systematic analysis and storage of the real-time PCR data, the SDS text files were imported into a self-designed MySQL database and analysed by scripts in Perl and R languages. Gene transcript levels were measured using the standard curve method (Livak, 1997). A flow chart of the data storage and analysis is shown in Figure 2.3.

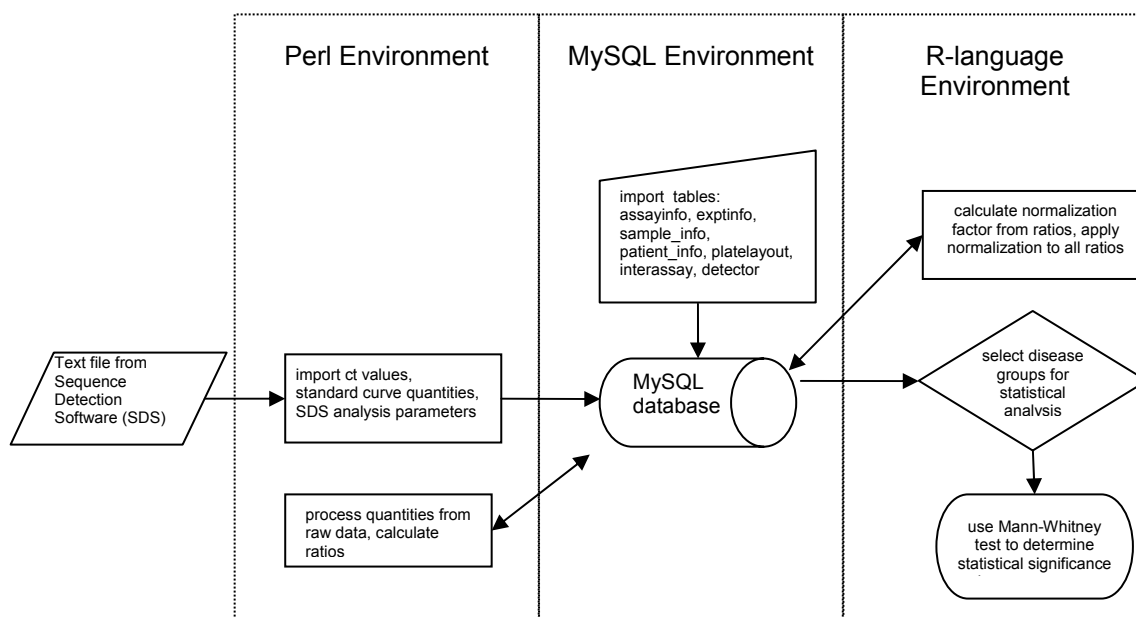


Figure 2.3 Flow diagram showing the data storage and processing of real-time PCR data.

3 Results

3.1 Microarray comparison

To compare microarray results from different microarray platforms, RNA was extracted from colonic mucosa biopsies of five normal controls. Exactly the same RNA was used for subsequent hybridization on both the cDNA and Affymetrix oligonucleotide arrays (HG-U95Av arrays).

3.1.1 Variation in expression signals

The overall distribution of variation in detected signals among the five patient samples was determined by calculating a scaled measure of variation (coefficient of variation) for all “present” signals in both platforms (Figure 3.1). Both systems had a similar range of variation in the distribution of “present” in samples of normal individuals, as demonstrated by median and overlapping semiquartile ranges.

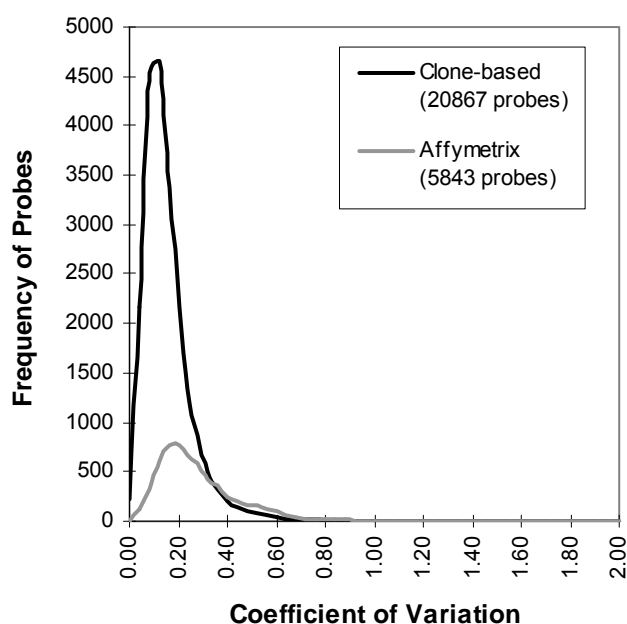


Figure 3.1 Variation in probe detection on both platforms.

Coefficients of variation were calculated for each “present” signal and were found not to be normally distributed by Kolmogorov-Smirnov test. The median coefficients of variation for the clone-based platform (black line) and the Affymetrix platform (gray line) were 0.15 ± 0.11 and 0.26 ± 0.18 , respectively, and were not statistically different.

To facilitate comparison between the microarray platforms, a single call (“present” or “absent”) was assigned to each probe, based on the probability of the probe’s expression signal in each of the five biological replicates. For each probe, p-values were combined from each patient sample using the Fisher method, resulting in a combined detection p-value. Gene probes with a combined

detection p-value of $p < 0.05$ or $p < 0.06$ in the clone-based or Affymetrix platforms, respectively, were defined as “present”. A plot of the coefficient of variation against the \log_{10} mean expression signal illustrated the distribution of overlapping “clouds” formed by groups of “present” signals stratified by three equal groups of combined detection p-values (Figure 3.2). As expected, a higher range of variation was predominantly seen in samples with low expression levels within each platform. Genes called “present” in the top third percentile were expressed at higher levels than those that could be detected in the middle or bottom third percentile of combined p-values.

To obtain an estimate of the actual numbers of genes detected by the platforms, each probe was assigned to a gene using UniGene cluster identifications (Section 2.1.9). A total of 13 639 and 4 660 nonredundant genes were detected (based on combined detection p-values from all samples), representing about two-thirds of the clone-based platform and half of the Affymetrix platform, respectively.

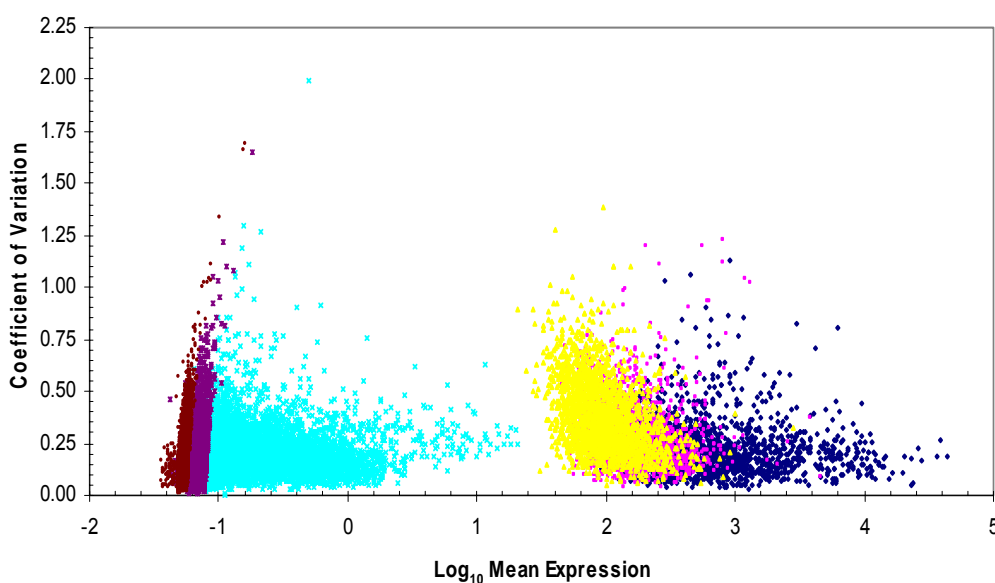


Figure 3.2 Distribution of coefficient of variation and mean expression in present signals.

Coefficient of variation and \log_{10} mean expression of present signals are shown for the clone-based and Affymetrix platforms. Varying p-values of probe detection in the clone-based platform are colored according to top (cyan), middle (purple), or bottom (brown) third percentiles of the detection p-values. Similarly, the Affymetrix platform values are separated by top (dark blue), middle (pink), and bottom (yellow) third percentiles of the detection p-values. The dynamic range of the samples measured extended 3 logs in the clone-based system (1:564 arbitrary units) and 4 logs in the Affymetrix system (1:2092 arbitrary units).

3.1.2 Overlap between gene detection on both platforms

Overlap between the two microarray platforms was defined to be the proportion of calls that were concordantly called “present” or “absent” in genes represented by both platforms. From the 20 406 known and putative genes represented by the clone-based array and the 8 602 known genes represented on the Affymetrix platform, probe matching via UniGene (Section 2.1.9) identified 6 645 genes which were represented on both systems. The 6 645 overlapping genes were classified according to the present/absent calls assigned by combined detection p-values (Table 3.1A and Figure 2.1). Of the 6 645 overlapping genes on both arrays, 4 805 and 4 014 genes were called “present” in the clone-based and Affymetrix platforms, respectively, from which 3 088 genes (46%) were called “present” in both platforms. Overall, 4 002 genes (60%) were in call agreement (overlap) between the two systems. In the remaining discordant fraction, almost twice as many present calls (1 717) were made by the clone-based system compared with the Affymetrix system (926).

Table 3.1 Distribution of UniGene cluster detection

		Frequency of Call in Affymetrix Array			Overlap (calls in agreement)	Spearman Rank Order Coefficient
		Present	Absent	Total		
A: Platform-overlapping, nonredundant						
Frequency of call in clone-based array	Present	3 088	1 717	4 802	60%	0.151*
	Absent	926	914	1 840		
	Total	4014	2 631	6 645		
B: Platform-overlapping, sequence-verified, nonredundant						
Frequency of call in clone-based array	Present	384	114	498	64%‡	0.131†
	Absent	144	73	217		
	Total	528	187	715		

Assignment of UniGene cluster is based on non-sequence-verified clone annotation. *Based on median expression signal of 3 088 present calls in both platforms. †Based on median expression signal of 384 present calls in both platforms. ‡ $p < 10^{-14}$, assuming a binomial distribution of probability.

Since discrepancies in the detection of the same genes between two different platforms could be caused by the 20– 30% error rate in cDNA clone annotation (Halgren et al., 2001; Taylor et al., 2001), further analysis was completed using data from only sequence-verified clones.

Approximately 2 000 cDNA clones were sequenced in-house. Sequences were matched to known

Genbank sequences using BLAST homology search. Subsequently, 715 cDNA probes were matched to the corresponding probes on the Affymetrix array through association to UniGene clusters (Table 3.1B and Figure 2.1). Using sequence-verified clones, the fraction of call agreement increased slightly to 64% (457 genes). Additionally, the discordant calls in the remaining 36% of genes appeared to be more evenly split between the two platforms (114 and 144 present calls in the clone-based and Affymetrix platforms, respectively) than previously observed with the non-sequence-verified genes.

As verified sequence identity only appeared to play a small role in overlap of concordant calls between the two systems, the role of probe sequence in overlap was investigated. Available fully sequenced cDNA clones (272) were matched to Affymetrix probe sets by BLAST searching. Over one-third of the probes (103) that were formerly matched to UniGene clusters no longer matched under stringent BLAST matching, due to the representation of completely different parts of the gene (i.e., probe matches did not meet cutoff criteria because probe locations were not sufficiently overlapping). About one-half of the full-length cDNA sequences (142) matched the same probe, and 16% (27) of the fully sequenced clones matched to different probes, for a total of 169 BLAST-matched probes (Table 3.2A, Figure 2.1). Detailed analysis of the differences in detection of the same genes represented between platforms indicated that genes which were present in the cDNA platform were represented by cDNA clones containing Alu repeats in 33 cases (24%) of clone present calls, whereas none of the absent clone calls contained Alu sequences. When Alu repeat-containing clone sequences were excluded from the analysis, 67% of the calls for 137 genes were in agreement (Table 3.2B). This would suggest that although overlap may be slightly improved by increased matching stringency, the same genes on the two systems are still not consistently detected in the remaining 33% of genes, even if the same region of the gene is represented on both arrays.

Table 3.2 Distribution of BLAST-matched full-length clones with and without Alu repeats

		Frequency of Call in Affymetrix Array				
		Present	Absent	Total	Overlap (calls in agreement)	Spearman Rank Order Coefficient
A: BLAST-matched, full-length clones with Alu repeats						
Frequency of call in clone-based array	Present	102	31	133		
	Absent	23	13	36	68%‡	0.015*
	Total	125	44	169		
B: BLAST-matched, full length clones without Alu repeats						
Frequency of call in clone-based array	Present	79	22	101		
	Absent	23	13	36	67%**	0.289†
	Total	102	35	137		

*Based on median expression signal of 102 present calls in both platforms. †Based on median expression signal of 79 present calls in both platforms. ‡ $p < 1.6 \times 10^{-6}$, assuming a binomial distribution of probability. ** $p < 3.7 \times 10^{-5}$, assuming a binomial distribution of probability.

To further investigate the discordant calls from both platforms, the role of combined detection p-value cutoffs was examined in the subset of probes matched by BLAST searching (137 probes, Table 3.2B). The distribution of the discordant calls shows that decreasing the combined p-value cutoff to more stringent values did not increase the level of agreement between the platforms (Figure 3.3). For example, increasing the \log_{10} p-value cutoff to 2 increased the number of genes called absent in both platforms, but also increased the number of discordant calls. At this cutoff, the net effect was an overlap of 65%. In the MAS 5.0 analysis of Affymetrix data, it was possible to influence the stringency of the calls by changing the Tau parameter. Lowering the Tau value increases the number of present calls, at the risk of creating false positives, whereas raising the Tau value decreases the number of present calls, at the risk of calling more false negatives. Decreasing the Tau value to 0.0 did increase the overlap to 72%, but increasing Tau to 0.06 decreased the overlap to 58%. These results were primarily achieved by shifting the Affymetrix p-values to values above or below the cutoff (Figure 3.4 and Figure 3.5). Looking at the distribution of the data points, it was clear that most of the discordant calls were not borderline present/ absent calls, but rather, genes whose p-values from either platform were at the opposite ends of the detection spectrum.

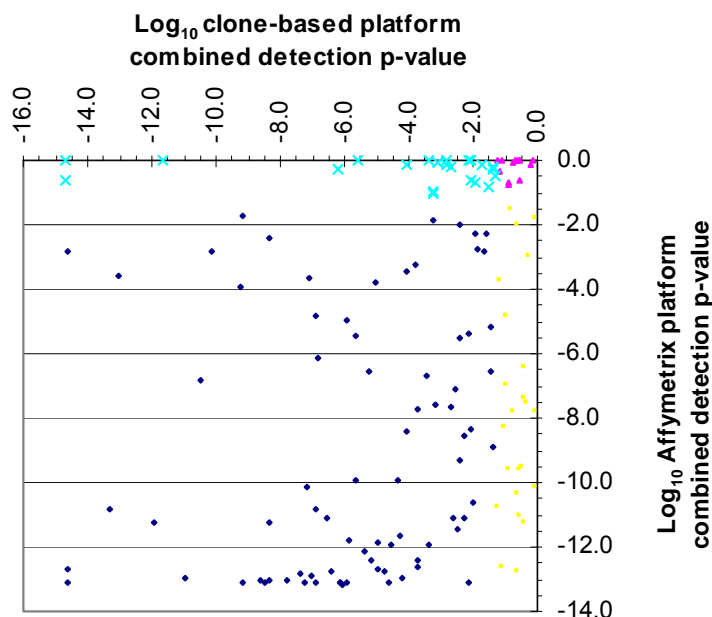


Figure 3.3 One-hundred thirty-seven BLAST-matched genes, without Alu repeats, analyzed in MAS5.0 with Tau=0.015.

Data were generated using combined detection p-value cutoff of 0.05 for clone-based system and p-value cutoff of 0.06 and Tau = 0.015 for the Affymetrix system. Genes present on both platforms (blue diamonds), absent in both platforms (pink triangles), absent in clone-based but present in Affymetrix platform (yellow squares), and present in clone-based but absent in Affymetrix platform (cyan crosses) were plotted according to \log_{10} of the combined detection p-values.

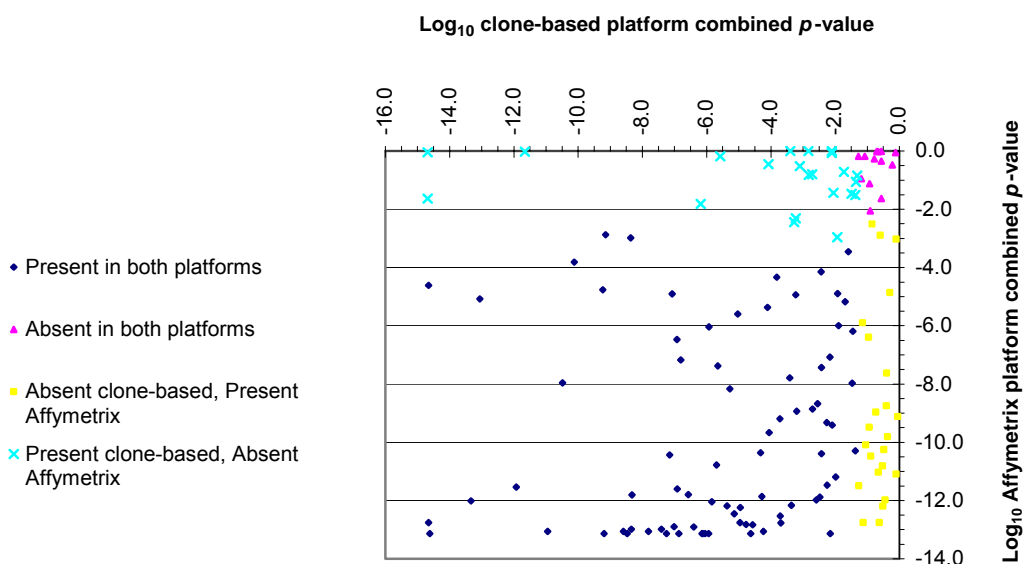


Figure 3.4 One-hundred thirty-seven BLAST-matched clones without Alu repeats, analyzed in MAS5.0 with Tau=0.0.

Data point colors represent calls based on default cutoffs and parameters (clone-based: $p < 0.05$; Affymetrix: Tau=0.015, $p < 0.06$). Decreasing the Tau parameter to 0.0 in MAS5.0 analysis results in less stringent calls by decreasing the combined detection p-values of the genes.

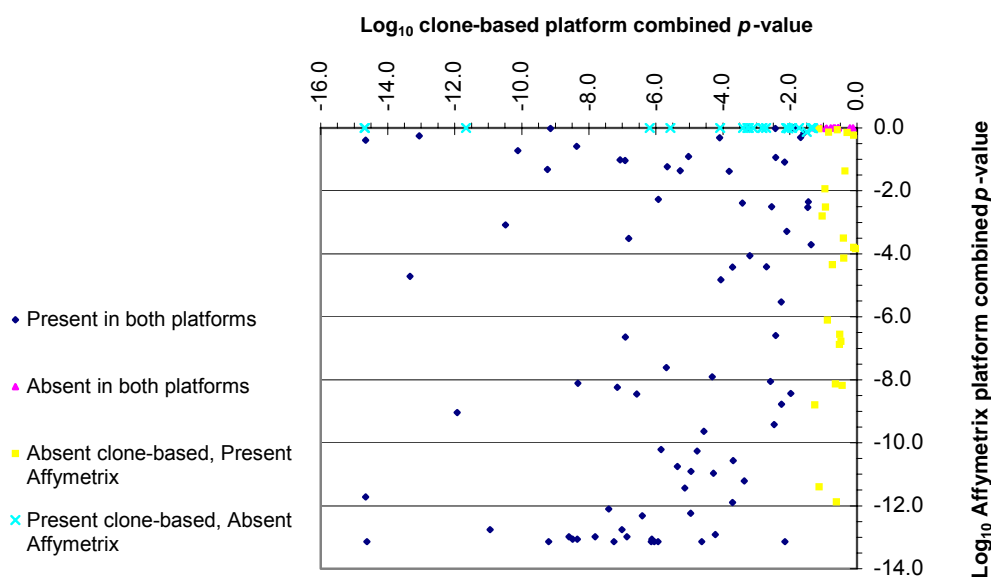


Figure 3.5 One-hundred thirty-seven BLAST-matched clones without Alu repeats, analyzed in MAS5.0 with Tau=0.060.

Data point colors represent calls based on default cutoffs and parameters (clone-based: $p < 0.05$; Affymetrix: Tau=0.015, $p < 0.06$). Increasing the Tau parameter to 0.060 in MAS5.0 analysis results in more stringent calls by increasing the combined p -values of the genes.

Closer inspection of the probe sequences used to detect the discordant genes showed that some of these probes showed some characteristics that could have caused the discordant calls. In at least one-fourth of discordant calls, the probes from either array were overlapping, but often represented exons not found on the alternative array. In the case of two genes expressed in clone-based but not Affymetrix platform, the cDNA probe sequences were chimeric with 28S rRNA. Cross-hybridization was also a potential problem for two genes that were highly polymorphic or belonged to a family of highly similar genes. However, in the remaining discordant cases, no irregularities in probe representation were observed.

3.1.3 Comparison of expression levels between platforms

Spearman rank order correlation coefficients were used to determine the relation between gene expression levels in the two platforms. Probes matched by UniGene cluster were ranked according to their median expression values in the sequence-verified subsets of probes (Figure 2.1). In the first instance, the calculation was restricted to sequence-verified, UniGene-matched probes that were called “present” in both systems. The Spearman rank order coefficient (r_s) was 0.131, based

on 384 genes matched by UniGene cluster (Figure 3.6, Table 3.1B). The correlation did not improve in the BLAST-matched subset ($r_s = -0.015$), when it was further limited to 102 genes that were called present in both systems (Table 3.2A). In the last subset, rank order coefficients were calculated for BLAST-matched probes without Alu repeat sequences and called present in both platforms (79 genes, Table 3.2B). Despite elimination of cDNA probes with potential for Alu cross-hybridization and limitation to concordantly called present genes, the rank orders of the mean expression values obtained in both platforms displayed weak positive correlation ($r_s = 0.289$). As a final test, rank order coefficients were calculated based on the rank order of individual patient expression levels within each platform for 79 genes called present in both platforms (Table 3.2B). In both platforms, patients were ranked by expression level for each gene, such that each of the 79 genes had its own rank order correlation. The rank order correlations ranged from strongly negative ($r_s = -1.00$) to strongly positive ($r_s = 0.80$), with a median and semi-quartile range of $r_s = -0.10 \pm 0.45$. These results imply that the expression values from the two technologies did not correlate with regard to median expression values or with respect to rank order of individual patient expression values within a platform.

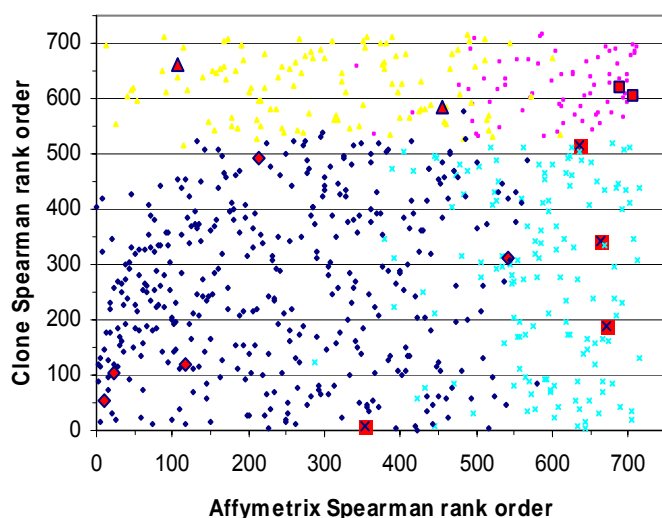


Figure 3.6 Spearman rank order based on mean expression signals for 715 sequence-verified, UniGene-matched genes.

Plot of Spearman rank orders from both array platforms is shown for concordant present calls (blue diamonds), concordant absent calls (pink squares), and discordant calls [absent in Affymetrix but present in clone-based platform (cyan crosses); present in clone-based but absent in Affymetrix (yellow triangles)]. Rank order correlation coefficients display only weak to moderate positive correlation between the two platforms: for the 384 concordant present calls, $r_s = 0.131$; for all 715 genes, $r_s = 0.191$; for the 457 concordant present and absent calls, $r_s = 0.465$. Genes selected for independent real-time PCR quantitation are marked in red.

Independent verification of transcript levels in the five patient samples was carried out by quantitative real-time PCR on 13 selected genes with varying calls between the two platforms (Figure 3.7). Results indicate that for this small subset of genes, the rank order correlation of the median expression values amongst patient samples was strongly positive between the Affymetrix and real-time PCR methods ($r_s = 0.884$) and only moderately to weakly positive between clone-based array and real-time PCR ($r_s = 0.593$) or clone-based and Affymetrix array ($r_s = 0.384$). When the rank order correlation was calculated only for genes called present in either Affymetrix or clone-based arrays, the correlation improved between the clone-based and Affymetrix platforms ($r_s = 0.900$) and between the clone-based and real-time PCR platforms ($r_s = 0.733$) but slightly decreased between the real-time PCR and Affymetrix platforms ($r_s = 0.857$). Expression levels from real-time PCR and overall detection within the array platforms were somewhat consistent. In general, highly expressed genes tended to be present on both array platforms, and low expressed genes were absent in both platforms. However, in two cases, an elevated expression signal from the clone-based platform could have been caused by a cDNA chimeric with 28S rRNA (*MPG*) or a spliced-in Alu repeat (*PRKCBP1*). A diminished expression signal from the Affymetrix platform could have been caused by a probe set which, unlike all the other 12 probe sets, was located in the middle of the gene (*PRKCBP1*), rather than in the last 3' exon, which incurs a slight labeling bias over the 5' end of the gene. Comparison of the gene sequences detected by each system showed that for each of the 13 genes, all three methods were capable of detecting the same splice variant. For the most part, the sequences used for clone-based array probes and real-time PCR probes or clone-based probes and Affymetrix probes were overlapping, but Affymetrix and real-time PCR probes were not necessarily overlapping, although they often detected parts of the same exon in over half of the genes tested.

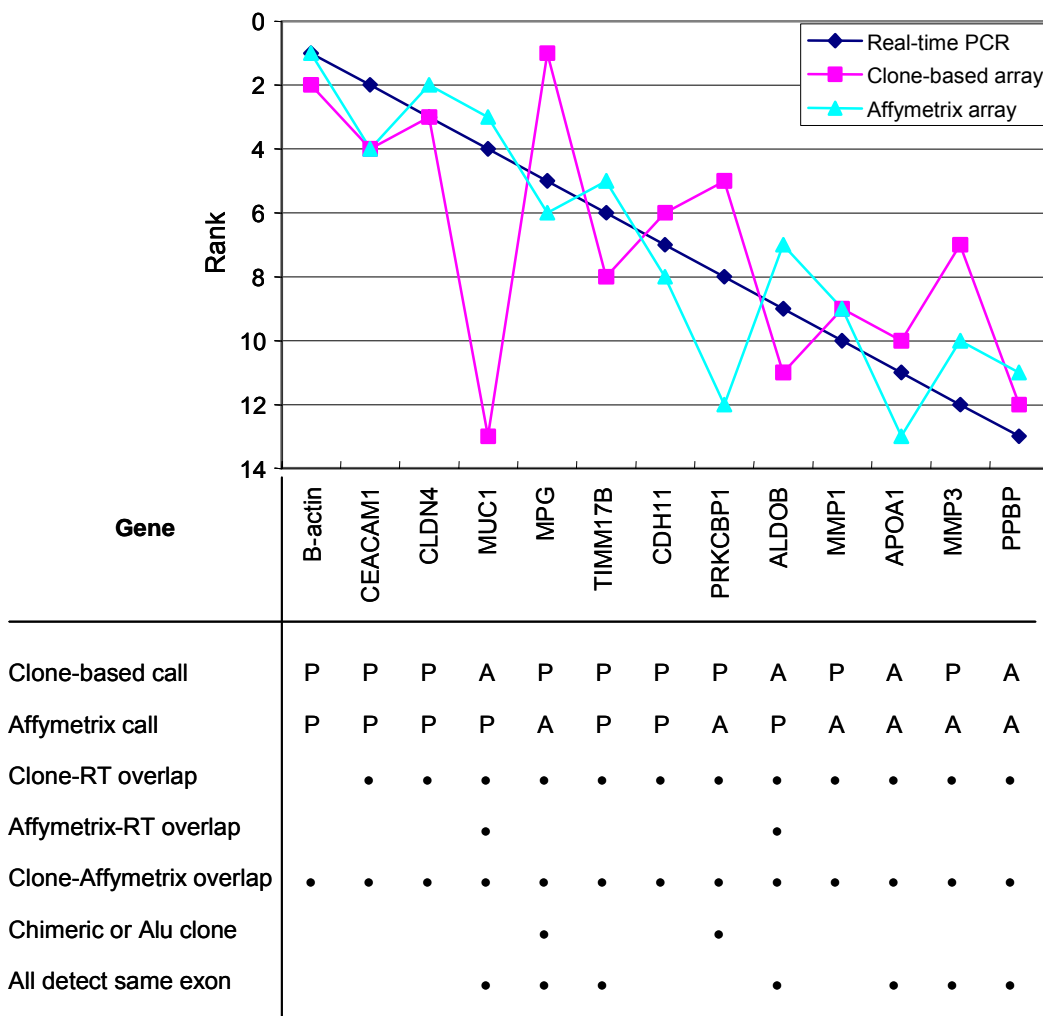


Figure 3.7 Ranking of the 13 genes quantified by real-time PCR, clone-based array, and Affymetrix array.

Genes were ranked according to the median expression signals obtained from quantitative real-time PCR (blue diamonds), the clone-based array (pink squares), and the Affymetrix array (cyan triangles). The table under the graph indicates the clone-based and Affymetrix platform calls and the extent of probe sequence overlap between all three detection methods. For two of the clone-based probes, cross-hybridization likely influenced the rank due to chimerism with 28S rRNA (*MPG*) or Alu repeats (*PRKCBP1*). In over half of the genes tested, all methods detected parts of the same exon. When considering the all 13 genes, regardless of present or absent calls, the rank order correlations were weak between clone-based and Affymetrix arrays ($r_s = 0.385$) and moderate between clone-based arrays and real-time PCR ($r_s = 0.593$), but correlation was high between Affymetrix platform and real-time PCR ($r_s = 0.884$). In contrast, Spearman rank order coefficients showed high correlation when only present genes were considered between clone-based and Affymetrix platforms ($r_s = 0.900$, 5 genes), clone-based array and real-time PCR ($r_s = 0.733$, 9 genes), and Affymetrix array and real-time PCR ($r_s = 0.857$, 7 genes).

3.2 Molecular profiling of IBD using cDNA microarrays

3.2.1 Expression profiles of the IBD colonic mucosa

Differences between gene expression in sigmoid biopsies taken from normal controls (n=11) and inflamed IBD patients (n=10 for each of CD and UC) were determined by hybridizing the samples to the Human Unigene Set RZPD1 cDNA arrays. Statistically significant differential gene expression was calculated by applying a pairwise statistical test (Mann Whitney U test). Genes which had a p-value ≤ 0.0015 and a fold change ≤ 1.2 were considered to be differentially regulated. Based on these cutoff criteria, 650 genes were identified to be differentially regulated between normal controls and at least one of the IBD subtypes. Of these genes, 122 were differentially regulated in both subtypes compared to normal controls. Interestingly, none of the 122 genes were regulated in opposite directions within the two IBD subtypes. Comparing CD to normal controls yielded 500 differentially regulated genes (81 up-regulated), whereas comparing UC to normal controls yielded 270 differentially regulated genes (155 up-regulated). Unknown genes represented 43% (277) of all differentially regulated genes. Overviews of the top 40 up-/down-regulated genes in CD and UC compared to normal controls are shown in Figure 3.8 and Figure 3.9, respectively.

To identify broad functional processes represented by the differentially regulated genes in each IBD subtype, the genes were assigned to functional groups using Gene Ontology classifications obtained through NCBI Gene database. To complement the database information, genes were also manually assigned to groups based on references in the literature (Table 3.3). Using methods described in Section 2.2.2, the significance of differentially regulated genes representing certain functional groups was tested. The functional group "immune and inflammatory responses" was found to be enriched by mostly up-regulated genes in both CD and UC ($p=0.0006$). In the group "cell proliferation and growth" ($p=0.0404$), enrichment was driven by genes up-regulated in UC. Finally, structure and permeability ($p=0.0309$) was largely represented by genes up-regulated in CD.

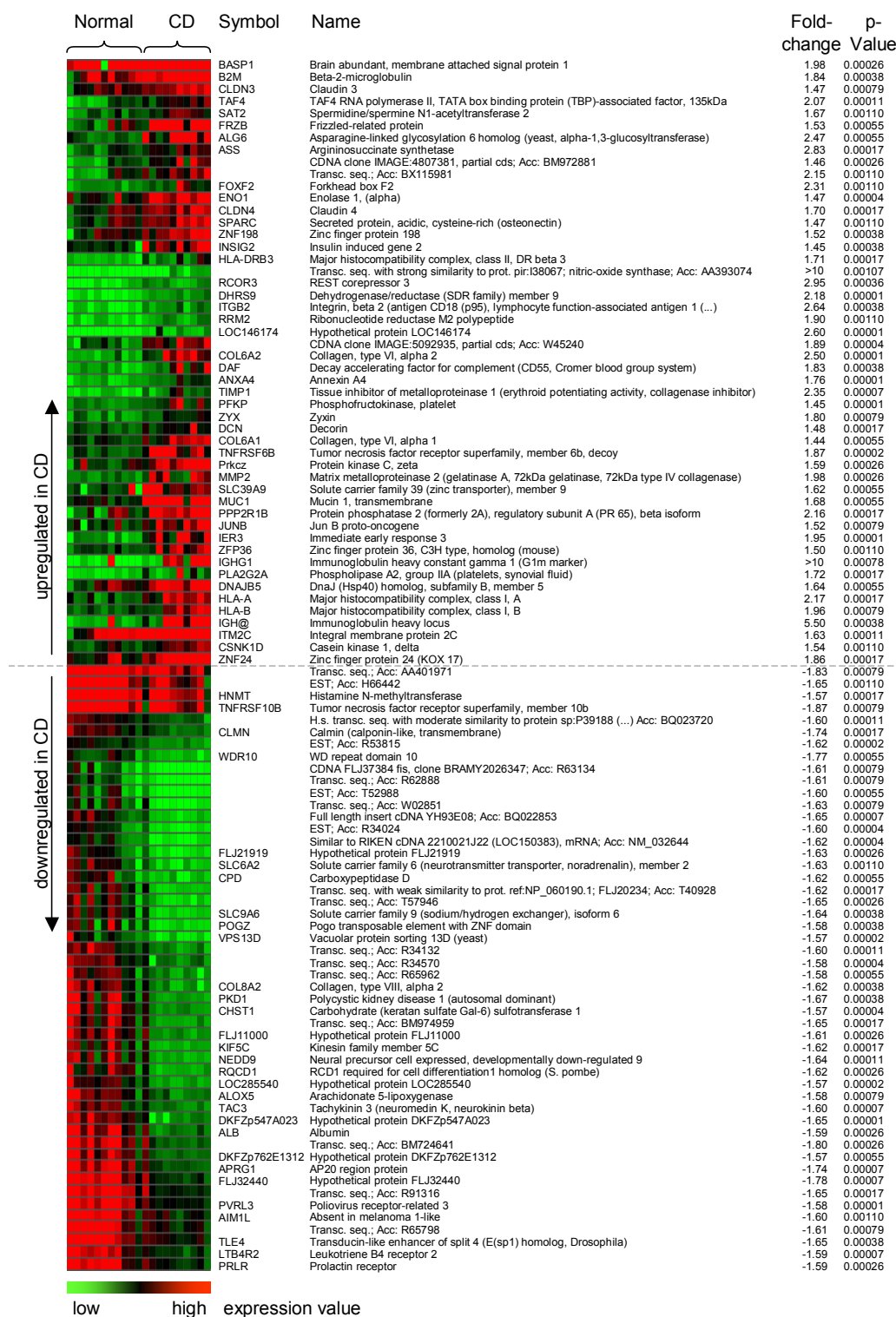


Figure 3.8 Heatmap of differentially regulated genes between normal controls and Crohn's disease.

The expression values of the top 40 up-/down-regulated genes are graphically depicted for each of 11 normal controls and 10 inflamed CD samples. Low to high expression values are colored from green to red, respectively.

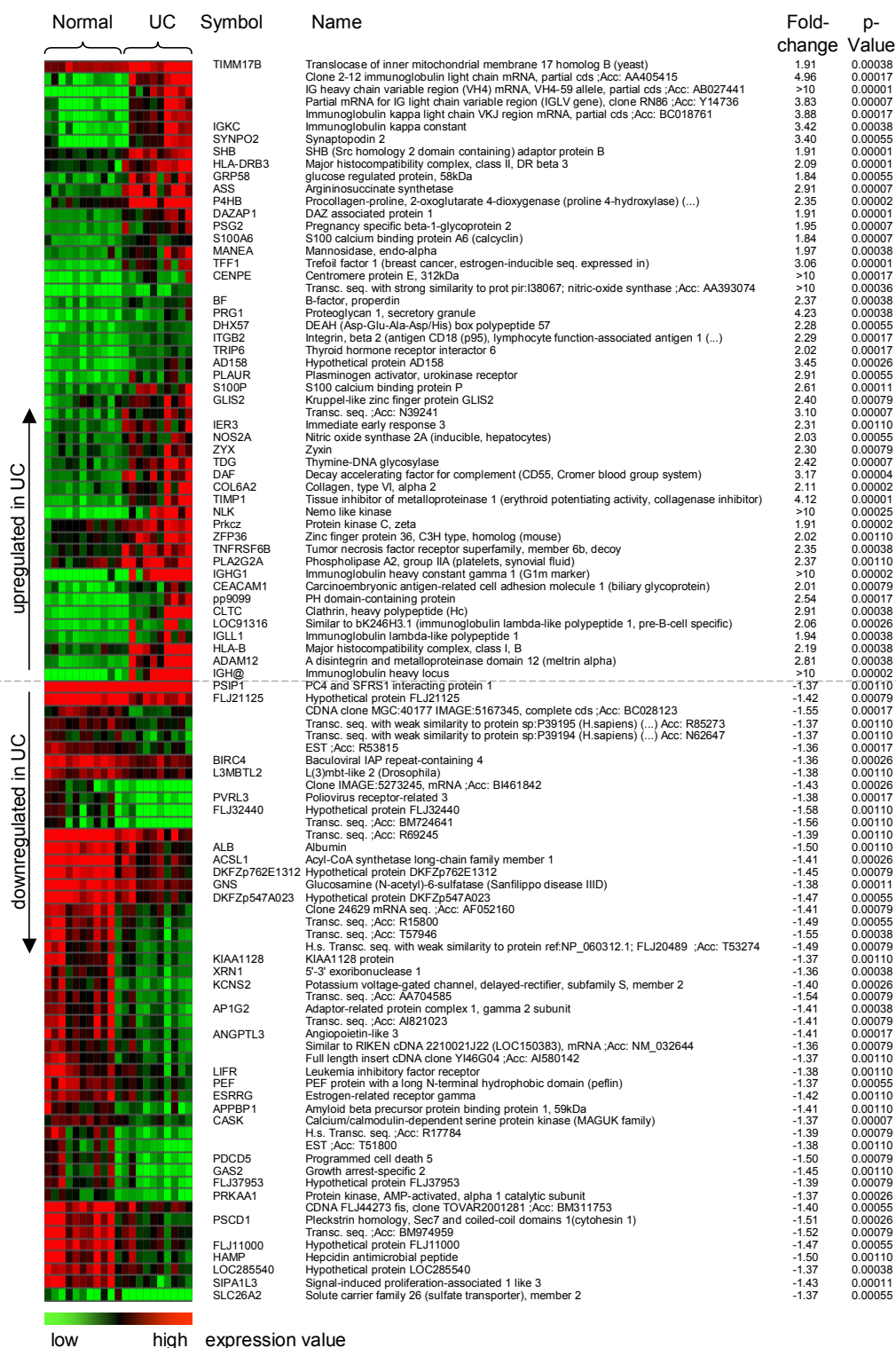


Figure 3.9 Heatmap of differentially regulated genes between normal controls and ulcerative colitis.

The expression values of the top 40 up-/down-regulated genes are graphically depicted for each of 11 normal controls and 10 inflamed UC samples. Low to high expression values are colored from green to red, respectively.

Table 3.3 Functional groups of differentially regulated genes.

Functional Category	Symbol	Name	NC vs. CD		NC vs. UC	
			Fold-change	p-value	Fold-change	p-value
Immune and Inflammatory Response	-	Transcribed sequence with strong similarity to protein pir:I38067 (H. sapiens) I38067 nitric-oxide synthase	>10	0.001066	>10	0.000363
	IGHG1	Immunoglobulin heavy constant gamma 1 (G1m marker)	>10	0.000783	>10	0.000023
	IGH@	Immunoglobulin heavy locus	5.5	0.000380	>10	0.000023
	ASS	Argininosuccinate synthetase	2.83	0.000170	2.91	0.000068
	ITGB2*	Integrin, beta 2 (antigen CD18 (p95), lymphocyte function-associated antigen 1; macrophage antigen 1 (mac-1) beta subunit)	2.64	0.000380	2.29	0.000170
	DAF	Decay accelerating factor for complement (CD55, Cromer blood group system)	1.83	0.000380	3.17	0.000040
	IER3	Immediate early response 3	1.95	0.000006	2.31	0.001100
	HLA-B	Major histocompatibility complex, class I, B	1.96	0.000788	2.19	0.000380
	TNFRSF6B	Tumor necrosis factor receptor superfamily, member 6b, decoy	1.87	0.000023	2.35	0.000380
	PLA2G2A*	Phospholipase A2, group IIA (platelets, synovial fluid)	1.72	0.000170	2.37	0.001100
	HLA-DRB3	Major histocompatibility complex, class II, DR beta 3	1.71	0.000170	2.09	0.000006
	HAMP	Hepcidin antimicrobial peptide	-1.51	0.000788	-1.5	0.001100
	C5	Complement component 5	-1.42	0.000108	-1.28	0.000788
	BIRC4	Baculoviral IAP repeat-containing 4	-1.41	0.000006	-1.36	0.000255
	HLA-A	Major histocompatibility complex, class I, A	2.17	0.000170		
	IL1R1	Interleukin 1 receptor, type I	1.32	0.000380		
	LTB4R2	Leukotriene B4 receptor 2	-1.59	0.000068		
	ALOX5	Arachidonate 5-lipoxygenase	-1.58	0.000788		
	CYLD	Cylindromatosis (turban tumor syndrome)	-1.53	0.001111		
	ALOX5AP	Arachidonate 5-lipoxygenase-activating protein	-1.52	0.000550		
	ADORA3	Adenosine A3 receptor	-1.48	0.000068		
	HLA-DQA1	Major histocompatibility complex, class II, DQ alpha 1	-1.47	0.000108		
	CDKN1A	Cyclin-dependent kinase inhibitor 1A (p21, Cip1)	-1.46	0.000788		
	CCL28	Chemokine (C-C motif) ligand 28	-1.45	0.001100		
	MICA	MHC class I polypeptide-related sequence A	-1.4	0.001100		
	RNPEP	Arginyl aminopeptidase (aminopeptidase B)	-1.37	0.000170		
	MASP1	Mannan-binding lectin serine protease 1	-1.36	0.001100		
	IGKC*	Immunoglobulin kappa constant			3.42	0.000380
	P4HB	Proline 4-hydroxylase			2.37	0.000023
	NOS2A	Nitric oxide synthase 2A (inducible, hepatocytes)			2.03	0.000550
CEACAM1*	Carcinoembryonic antigen-related cell adhesion molecule 1			2.01	0.000788	
IGLL1	Immunoglobulin lambda-like polypeptide 1			1.94	0.000380	
IFITM1	Interferon induced transmembrane protein 1 (9-27)			1.65	0.000068	
CHUK	Conserved helix-loop-helix ubiquitous kinase			1.51	0.001100	
EGLN3	Egl nine homolog 3 (C. Elegans)			1.42	0.000170	
MYBL2	V-myb myeloblastosis viral oncogene homolog (avian)-like 2			1.35	0.000108	
ALDH2	Aldehyde dehydrogenase 2 family (mitochondrial)			1.28	0.00110	
BST1	Bone marrow stromal cell antigen 1			-1.34	0.000550	
Oncogenesis, Cell Proliferation and Growth	TIMP1	Tissue inhibitor of metalloproteinase 1	2.35	0.000068	4.12	0.000006
	PLA2G2A*	Phospholipase A2, group IIA (platelets, synovial fluid)	1.72	0.00017	2.37	0.0011
	CAPNS1	Calpain, small subunit 1	1.24	0.000255	1.36	0.000068
	GAS2	Growth arrest-specific 2	-1.54	0.000055	-1.45	0.0011
	STAG1	Stromal antigen 1	-1.46	0.000011	-1.31	0.000038
	DAB2	Disabled homolog 2, mitogen-responsive phosphoprotein (Drosophila)	-1.33	0.00004	-1.21	0.0011

	DNAJB5	Dnaj (Hsp40) homolog, subfamily B, member 5	1.64	0.00055		
	CSNK1D	Casein kinas 1, delta	1.54	0.0011		
	NME2	Non-metastatic cells 2, protein (NM23B) expressed in	1.28	0.0011		
	NEDD9	Neural precursor cell expressed, developmentally down-regulated 9	-1.64	0.000108		
	CYLD	Cylindromatosis (turban tumor syndrome)	-1.53	0.001111		
	TSSC1	Tumor suppressing subtransferable candidate 1	-1.51	0.000068		
	CISH	Cytokine inducible SH2-containing protein	-1.49	0.00055		
	Rock1	Rho-associated, coiled-coil containing protein kinase 1	-1.48	0.00017		
	PPP2CA	Protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform	-1.48	0.00004		
	PICALM	Phosphatidylinositol binding clathrin assembly protein	-1.46	0.00004		
	WIT-1	Wilms tumor associated protein	-1.45	0.00055		
	FLJ13352	Hypothetical protein FLJ13352	-1.4	0.00038		
	RASA1	RAS p21 protein activator (gtpase activating protein) 1	-1.39	0.0011		
	CDK5RAP1	CDK5 regulatory subunit associated protein 1	-1.37	0.000788		
	RAD1	RAD1 homolog (S. Pombe)	-1.33	0.00055		
	CENPE	Centromere protein E, 312kda			>10	0.00017
	TFF1*	Trefoil factor 1 (breast cancer)			3.06	0.000006
	S100P*	S100 calcium binding protein P			2.61	0.000108
	S100A6	S100 calcium binding protein A6 (calcyclin)			1.84	0.000068
	PRKCN	Protein kinase C, nu			1.68	0.0011
	S100A11*	S100 calcium binding protein A11 (calgizzarin)			1.56	0.000011
	CAPN1	Calpain 1, (mu/l) large subunit			1.37	0.000108
	MYBL2	V-myb myeloblastosis viral oncogene homolog (avian)-like 2			1.35	0.000108
	NDRG3	NDRG family member 3			-1.24	0.0011
Structure and Permeability	TIMP1	Tissue inhibitor of metalloproteinase 1	2.35	0.000068	4.12	0.000006
	COL6A2	Collagen, type VI, alpha 2	2.50	0.000011	2.11	0.000023
	MMP2	Matrix metalloproteinase 2	1.98	0.000255	1.80	0.000108
	MUC1*	Mucin 1, transmembrane	1.68	0.00055	1.74	0.000011
	MUC13	Mucin 13, epithelial transmembrane	1.35	0.00017	1.64	0.000023
	CDH11	Cadherin 11, type 2, OB-cadherin (osteoblast)	1.59	0.00004	1.48	0.00017
	DKFZp547A023	Hypothetical protein dkfzp547a023	-1.65	0.000011	-1.47	0.00055
	FLJ36812	Hypothetical protein FLJ36812	-1.42	0.000023	-1.31	0.000023
	CLDN4	Claudin 4	1.70	0.00017		
	CLDN3	Claudin 3	1.47	0.000788		
	SPARC*	Secreted protein, acidic, cysteine-rich (osteonectin)	1.47	0.0011		
	COL8A2	Collagen, type VIII, alpha 2	-1.62	0.00038		
	COL15A1	Collagen, type XV, alpha 1	-1.55	0.00055		
	PRKCB1*	Protein kinase C, beta 1	-1.5	0.000788		
	Rock1	Rho-associated kinase beta	-1.48	0.00017		
	PDLIM5	PDZ and LIM domain 5	-1.48	0.000788		
	-	Transcribed sequences [N48794]	-1.46	0.000006		
	RDX	Radixin	-1.42	0.00055		
	ANK3	Ankyrin 3, node of Ranvier (ankyrin G)	-1.35	0.00038		
	OCLN	Occludin	-1.34	0.000788		
	FLJ25778	Hypothetical protein FLJ25778	-1.33	0.0011		
	CLTC	Clathrin, heavy polypeptide (Hc)			2.91	0.00038
	P4HB	Procollagen-proline, 2-oxoglutarate 4-dioxygenase			2.35	0.000023
	COL21A1	Collagen, type XXI, alpha 1			1.61	0.000788

Minus sign in the fold-change column indicates down-regulation between normal controls (NC) and IBD individuals. Asterisk (*) indicates a gene previously reported to be differentially expressed in IBD (Dieckgraefe et al., 2000; Dooley et al., 2004; Lawrance et al., 2001). All data based on microarray results.

3.2.2 Functional prediction of unknown genes

In order to decipher potential roles for unknown genes, some of the top differentially expressed unknown genes between normal controls and IBD were further analysed as described in Section 2.2.3 (Table 3.4). All but one of the unknown genes in Table 3.4 yielded protein domains or were homologous to known proteins. The genes represented by N48794 and AF087994 contained protein domains (cadherin and leucine-rich repeats, respectively) commonly involved in protein-protein interactions and may be involved in cell adhesion. Two genes represented by N39296 and AW953679 contain zinc fingers motifs and are therefore predicted to be transcription factors. Another two genes represented by AK056932 and AB067499 are predicted to be involved in enzymatic function (acyltransferase or hydrolase). BC008744 may represent a gene involved in intracellular signaling, as many proteins with pleckstrin homology domains are involved in intracellular signaling. AK022544 and AL117511 share protein homology with other known proteins and may be involved in structure and protein sorting, respectively. BC006384 was the only transcript without protein domains or homology to other proteins. This transcript was located in an intron of a heparin sulfate sulfotransferase, and it may represent a poorly characterized splice variant of this gene. All other genes have homologues in other species, which lends credibility to the hypothesis that these uncharacterized genes must have important functions, since they are conserved in the evolutionary process.

3.2.3 Independent quantitation by fluorescent real-time PCR

Microarray results were independently verified in a larger patient cohort by a more sensitive technique, real-time PCR. The initial cohort of 31 patients (Group 1) used in the cDNA microarray studies was extended to 100 patients (Group 2) for the real-time PCR experiments. No significant differences between the initial and verification cohort were found. As an additional control for the suitability of the samples and the stability of the overall approach, two hallmark genes of IBD pathophysiology (IL8 and TNF) were quantitated by real-time PCR. As expected, IL8 was overexpressed in CD and UC samples, and TNF was only overexpressed in CD samples. Next, 15 additional genes were selected for real-time PCR verification in the extended cohort (Group 2), based on their differential regulation in IBD from the microarrays and their representation in the

three main functional groups described in Section 3.2.1. Uncharacterized genes were also included in the analysis. The real-time PCR experiments are summarized in Figure 3.10 and Table 3.5.

Table 3.4 In silico (InterPro, SMART and Pfam search methods) analysis of top differentially expressed unknown genes between normals and CD or UC, respectively

<i>ProbelD (Symbol)</i>	<i>Name / Homologues in other species</i>	<i>Accession number</i>	<i>Cytoband</i>	<i>Comment</i>	<i>Putative function</i>	<i>NC vs. CD Fold change</i>	<i>NC vs. UC Fold change</i>
54C04 (MGC12916)	Hypothetical protein MGC12916/ Pt	BC006384	17p11.2	ORF has no known protein domains; transcript maps to an intron of heparan sulfate (glucosamine) 3-O-sulfotransferase 3B1 (HS3ST3B1), but may represent an alternate transcript of HS3ST3B1	Sulfotransferase	1.41	n.s.
18C05 (DKFZp547A023)	Hypothetical protein DKFZp547A023 / Mm, Rn	AK022544	1p12	Homology to collagen, myosin tail domain, ATPase domain, high expression similarites to endocytosis genes	Structure, endocytosis	-1.65 †	-1.47 †
44L11 (C14orf125)	Chromosome 14 open reading frame 125 / Mm, Rn	AL117511	14q12	Homology to vacuolar sorting-associated protein VPS13 (SOI1), high expression similarities to genes regulating apoptosis	Protein sorting, apoptosis	-1.33	-1.25
45F21 (TMEM68)	Transmembrane protein 68 / Mm, Rn, Dm, Ce	AK056932	8q12.1	Phosphate acyltransferase domain, high expression similarities to genes involved in lipid metabolism	Phospholipid metabolism	-1.54	n.s.
45H19 (none)	Similar to hypothetical protein FLJ23834 / Mm, Rn	N48794	3p21	Cadherin repeats	Cell adhesion	-1.46	n.s.
41K20 (LRRC57)	Leucine rich repeat containing 57 / Mm, Rn, Dm	AF087994	15q14	Leucine-rich repeats	Cell adhesion	-1.42	-1.31
14D14 (KIAA1912)	KIAA1912 protein / Mm, Rn	AB067499	2p16.2	ORF shows similarities to nucleoside triphosphate hydrolases, expression similarities to genes with guanylate kinase activity	Guanylate kinase	n.s.	-1.25
83J01 (PLEKHQ1)	pleckstrin homology domain containing, family Q member 1 / Mm, Rn	BC008744	15q22.1	Pleckstrin homology domain	Intracellular signaling	n.s.	2.54 †
37B10 (ZCCHC4)	zinc finger, CCHC domain containing 4/ Mm, Rn, Dm, Ce	N39296	4p15.31	GRF zinc finger, shows expression similarities to transcription factors	Transcription factor	-1.50 †	n.s.
32J19 (C20orf17)	Chromosome 20 open reading frame 17/ Mm	AW953679	20q13.2	Contains 5 zinc fingers and hox domain	Transcription factor	-1.22	n.s.

Fold changes are derived from microarray analysis. All reported fold changes are significant unless otherwise indicated. A minus sign in the fold-change columns indicates down-regulation between normal controls (NC) and disease. Single dagger (†) indicates direction of fold change verified by quantitative real-time PCR in Group 2 samples. The fold changes as reported by quantitative PCR for AK022544 (DKFZp547A023) were -1.96 for CD and -1.80 for UC, for BC008744 (PLEKHQ1) were 1.37 for CD and 1.60 for UC and for N39296 (ZCCHC4) were -1.61 and -1.46, respectively. The prediction of the function was based on two criteria: sequence homologies and expression pattern similarities. Expression pattern similarities were calculated based on cosine correlation, accepted similarity had to be > 97% and enrichment significance for the putative function based on Gene Ontology was set to a $p \leq 0.05$.

Abbreviations used: Hs: Homo sapiens; Pt: Pan troglodytes; Mm: Mus musculus; Rn: Rattus norvegicus; Dm: Drosophila melanogaster; Ce: Caenorhabditis elegans; ORF: Open reading frame; NC = normal control; CD = Crohn's disease; UC = ulcerative colitis; n.s. = non-significant.

Table 3.5 Summary of array and real-time RT results.

Symbol	Gene Name	Functional group	NC vs. CD		NC vs. UC		NC vs. Inflamed DC
			Fold-change on array (Real-time PCR)	Array result confirmed by real-time PCR*	Fold-change on array (Real-time PCR)	Array result confirmed by real-time PCR*	Significant by real-time PCR (fold change)
CDH11	Cadherin 11, type 2, OB-cadherin	SP	1.59† (1.71*)	Yes	1.48† (2.46*)	Yes	No (1.18)
CYLD	Cylindromatosis (turban tumor syndrome)	IIR, OCPG	-1.53† (-1.59*)	Yes	-1.21 (-1.92*)	No	Yes (-1.72)
DAF	Decay accelerating factor for complement	IIR	1.83† (2.05*)	Yes	3.17† (3.63*)	Yes	Yes (2.27)
DKFZp547A023	Hypothetical protein DKFZp547A023 / Mm, Rn	Unknown	-1.65† (-1.96*)	Yes	-1.47† (-1.80*)	Yes	Yes (-1.73)
IGHG1	Immunoglobulin heavy constant gamma 1 (G1m marker)	IIR	>10† (10.45*)	Yes	>10† (15.41*)	Yes	No (1.05)
PDLIM5	PDZ and LIM domain 5	SP	-1.48† (-1.49*)	Yes	-1.30 (-1.64*)	No	No (-1.30)
MUC1	Mucin 1, transmembrane	SP	1.68† (2.27*)	Yes	1.74† (2.68*)	Yes	Yes (2.19)
OCLN	Occludin	SP	-1.34† (-1.57*)	Yes	-1.22 (-1.98*)	No	Yes (-1.39)
PLA2G2A	Phospholipase A2, group IIA (platelets, synovial fluid)	IIR, OCPG	1.72† (3.94*)	Yes	2.37† (5.67*)	Yes	Yes (11.40)
PLEKHQ1	pleckstrin homology domain containing, family Q member 1	Unknown	2.23 (1.37*)	No	2.54† (1.60*)	Yes	No (1.34)
RCP9	Calcitonin gene-related peptide-receptor component protein	Unknown	-1.44† (-1.54*)	Yes	-1.31 (-1.88*)	No	No (-1.16)
Rock1	Rho-associated kinase beta	SP, OCPG	-1.48† (-1.65*)	Yes	-1.34 (-1.63*)	No	No (-1.23)
TFF1	Trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in)	OCPG	1.22 (1.05)	Yes	3.06† (2.71*)	Yes	No (3.56)
TIMP1	Tissue inhibitor of metalloproteinase 1	SP, OCPG	2.35† (1.90*)	Yes	4.12† (5.57*)	Yes	Yes (2.04)
ZCCHC4	Zinc finger, CCHC domain containing 4	Unknown	-1.50† (-1.61*)	Yes	-1.22 (-1.46*)	No	Yes (-1.50)

Minus sign in the fold-change column indicates down-regulation between normal controls (NC) and IBD individuals. The direction of the dysregulation seen in the comparison between inflamed DC and NC was the same as that observed in the comparison between IBD and NC. No influence of sex, disease characteristics, age, and pre-treatment could be detected by exploratory statistical groupwise comparison.

A single dagger (†) indicates significant differential expression on microarrays (p-value ≤ 0.0015, fold change ≥ 1.2); an asterisk (*) indicates significant differential expression in real-time PCR (p<0.05).

SP: structure and permeability; IIR: immune and inflammatory response; OCPG: oncogenesis, cell proliferation and growth

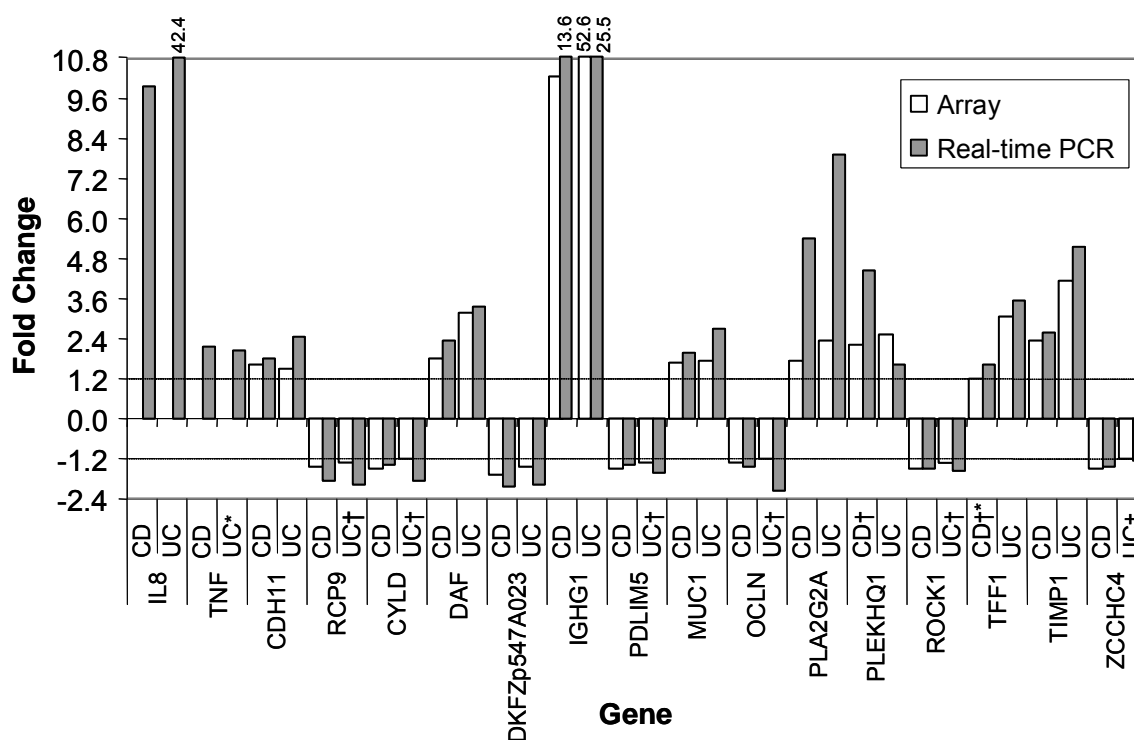


Figure 3.10 Microarray results and corresponding quantitative real-time PCR for differentially regulated genes in CD or UC compared to normal controls.

Genes were chosen based upon their dysregulation in IBD and represent both known genes from the functional groups discussed and genes of unknown function. Quantitative real-time PCR was carried out on individual samples from Group 2 patients (14-18 normal controls, 19-33 UC and 17-22 CD samples, depending on the availability of the patient samples at the time the plates were produced), except for IL8 and TNF (not on array), which were tested in Group 1 patients (11 normal controls, 10 UC and 10 CD patient samples) as a proof-of-principle measure. The extended cohort of Group 2 patients includes those with active disease and using anti-inflammatory drugs (but not immunosuppressants or biologicals), whereas Group 1 patients had active disease and were medication-free for 6 weeks. Results are summarized by a ratio of medians (CD:normal or UC:normal). All results were significantly differentially regulated except for marked results; single dagger (†) indicates that array result was not significant (P -value > 0.0015 or fold change < 1.2); asterisk (*) indicates that real-time PCR result was not significant (P -value > 0.05). A dashed line represents the fold change level of 1.2.

Genes that were significantly up-regulated in both microarray and real-time PCR methods in both IBD subtypes compared to normal controls include *cadherin-11* (*CDH11*), *decay accelerating factor for complement* (*DAF*), *immunoglobulin heavy constant gamma 1* (*IGHG1*), *mucin 1* (*MUC1*), *phospholipase A2, group IIA* (*PLA2G2A*), and *tissue inhibitor of metalloproteinase 1* (*TIMP1*). *Trefoil factor 1* (*TFF1*), was found to be significantly up-regulated in UC on the microarray but not in CD, a result that was confirmed by real-time PCR in the extended group 2 samples. The only down-regulated gene in both subtypes and in both microarray and real-time experiments was the uncharacterized gene *DKFZp547A023* (Genbank accession number AK022544, Table 3.4). The remaining genes showed contrasting results between the microarray and real-time PCR

experiments. In the case of *cylindromatosis (CYLD)*, *calcitonin gene-related peptide receptor component protein (RCP9)*, *PDZ and LIM domain 5 (PDLIM5)*, *occludin (OCLN)*, *Rho-associated, coiled-coil containing protein kinase 1 (ROCK1)*, and *zinc finger, CCHC domain containing 4 (ZCCHC4)*, the microarray results indicated that these genes were significantly down-regulated in CD, but did not reach significance in the UC cohort. However, results from the real-time analysis in the extended cohort showed that these six genes were significantly down-regulated in both IBD subtypes. *Pleckstrin homology domain containing, family Q member 1 (PLEKHQ1, BC008744; see Table 3.4)* was up-regulated in both subtypes on the microarray, but the microarray cutoff criteria were not reached in the CD population. Real-time analysis on the extended IBD population demonstrated this gene was significantly up-regulated in both IBD subtypes. These results illustrate the importance of using an independent technique in larger patient numbers to confirm gene expression identified by microarray experiments.

3.2.4 Expression profiles in non-IBD disease samples

As inflammation may result from different initiating processes, the extent of disease-related inflammation is an important issue in the detection of differential gene expression between normal controls and IBD. Real-time PCR was carried out on patients that had colonic disease (so-called disease controls (DC), n = 17), but not IBD (Figure 3.11 and Table 3.5). These DC samples were further stratified into non-inflamed and inflamed samples. No significant difference was observed between the expression of any of the 15 genes tested between normal controls and non-inflamed disease specificity controls. Eight genes (*CYLD*, *DAF*, *DKFZp547A023*, *MUC1*, *OCLN*, *PLA2G2A*, *TIMP1*, and *ZCCHC4*) were significantly differentially regulated between normal controls and inflamed DC, and the direction of change was the same as that observed in IBD. This observation suggests that these eight gene findings probably reflect general inflammation pathophysiology rather than events specific for IBD. In contrast, IBD-specific gene expression is presented in the expression of the remaining seven genes (*CDH11*, *IGHG1*, *PLEKHQ1*, *TFF1*, *PDLIM5*, *ROCK*, and *RCP9*). In these cases, gene transcripts were not significantly different between normal controls and the inflamed DC group.

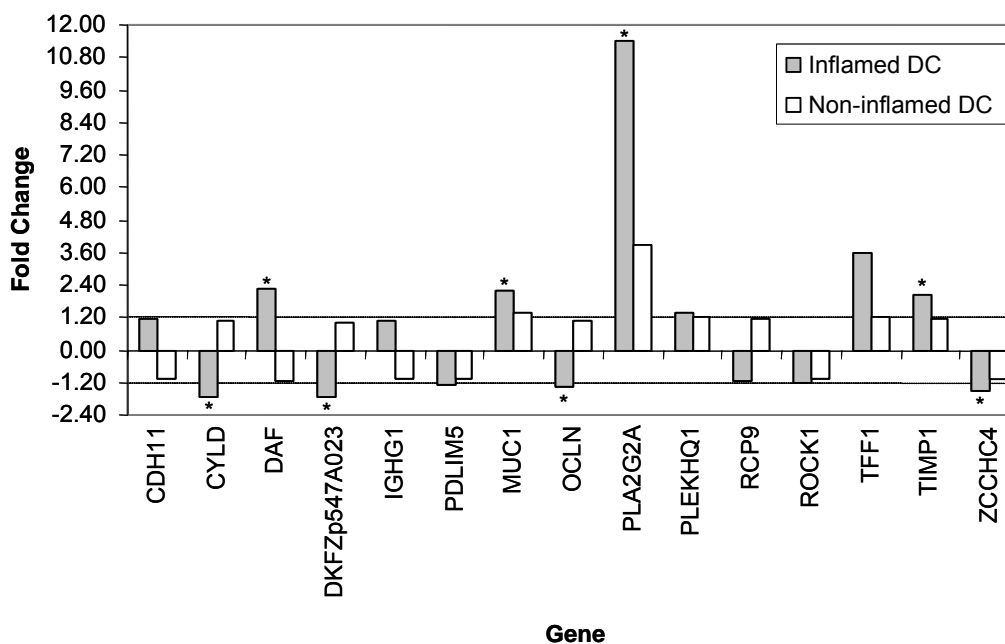


Figure 3.11 Quantitative real-time PCR results between normal controls and both non-inflamed and inflamed disease specificity controls.

Quantitative real-time PCR was carried out on individual samples from group 2 patients (including seven or eight inflamed and seven to nine non-inflamed DCs, depending on the availability of cDNA at the time the plates were produced). The DCs include patients with infectious diarrhea, gastrointestinal inflammation, or irritable bowel syndrome. Patients in this group were not on immunosuppressants or biologicals, but the use of anti-inflammatory drugs was allowed. Results are summarized by a ratio of medians (inflamed DC:normal or non-inflamed DC:normal). Results marked with an asterisk (*) indicate that the real-time PCR result was significant ($p < 0.05$).

3.2.5 Localization by immunohistochemistry

As expression analysis in complex tissues does not allow an identification of the cells responsible for the signal, immunohistochemistry was used to exemplify the approach to genes of interest.

Paraffin-embedded sections of samples from the group 1 patients were examined for three genes, which were selected because they were not previously associated with IBD etiopathogenesis and represented a gene from each of the functional groups presented in Section 3.2.1. Consistent with increased transcript levels measured by microarrays, a marked up-regulation of CEACAM1 protein was seen in IBD when compared to normal controls (Figure 3.12A). Immunoreactivity was found in the apical epithelial lining in the normal mucosa, whereas it extended down into the epithelial cells of the crypts in the inflamed tissue of CD and UC patients. A staining of vascular structures and mononuclear cells, most likely with a lymphocytic phenotype, was only detected in the lamina propria of inflamed biopsies. CSNK1D immunohistochemistry staining (Figure 3.12B) demonstrated

expression in the colon, but the modest up-regulation seen in CD in the microarray was not detected by immunohistochemistry. In the biopsies from CD patients, staining of the apex of the crypts could be observed. Immunoreactivity revealed a strong granular staining pattern of intestinal epithelial cells in the normal and UC group, which was located basolaterally. Staining of PRKCB1 (Figure 3.12C), that appeared to be down-regulated in the microarray results showed only a weak staining in the apical epithelial layer in the normal and CD mucosa, whereas a stronger staining was present in the UC group. Furthermore, scattered lamina propria mononuclear cells underlying the epithelial layer were also positive. Interestingly, in the CD group, immunoreactivity was found nearly exclusively in the marginal zone of small lymph follicles.

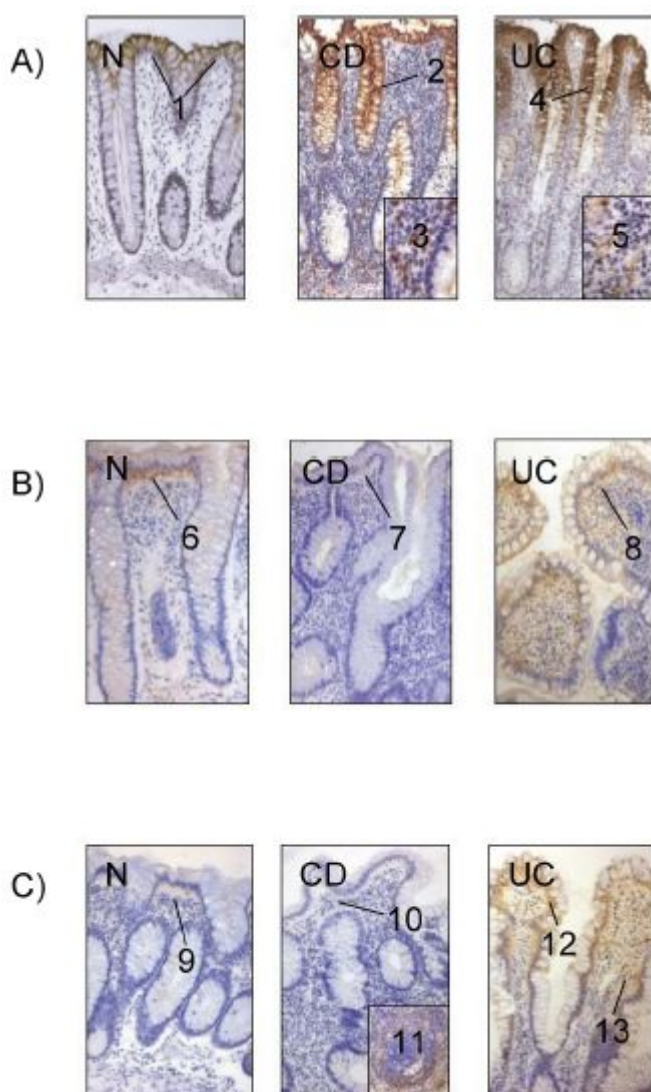


Figure 3.12 Immunohistochemical localization of CEACAM1, CSNK1D, and PRKCB1 in colonic mucosa

Staining of a representative mucosal tissue samples from five normal controls (N), five Crohn disease (CD), and six ulcerative colitis (UC) patients using antibodies against (A) CEACAM1, (B) CSNK1D, and (C) PRKCB1. (A) CEACAM1 immunoreactivity was found in the apical epithelial lining (1), crypts of inflamed tissue (2, 4). Additional staining was detected in immune cells (3) and blood vessels (5). (B) CSNK1D immunoreactivity showed a strong granular staining pattern in normal (6) and the UC group (8), which was located basolaterally. Weaker staining of the apex of the crypts could be detected in CD (7). (C) For PRKCB1, weak staining could be observed in the apical epithelial layer in the normal (9) and CD (10) mucosa and, interestingly, immunoreactivity was found nearly exclusively in the marginal zone of small lymph follicles (11), whereas the lamina propria was immunonegative. In contrast, strong staining was detected in the apical epithelial layer of UC mucosa (12). Furthermore, lamina propria mononuclear cells underlying the epithelial layer were also positive in the UC mucosa (13).

3.3 Molecular profiling of IBD samples using Affymetrix microarrays

3.3.1 Microarray data pre-processing

Four normalization methods (MAS5.0, GC-RMA, vsn, and loess; Section 2.2.2) were evaluated for use. Normalization was carried out separately for the UC and CD microarray datasets. The performance was determined by the following criteria: 1) ability of the normalization procedure to reduce variance without increasing bias; 2) ability of the normalization procedure to be carried out within a reasonable amount of computing time. To assess the variability of signals as a function of signal intensity, the log-intensity ratio (M) and the log-intensity average (A) was computed for each probe set between a pair of arrays. The resulting MA plot showed how the log-intensity ratios were distributed over signal intensity. In unnormalized data, the signals form a thick cloud centred along the x-axis, which tended to have many outlying signals (Figure 3.13). A good normalization method reduces the cloud to a more uniform distribution over all intensities (Figure 3.14).

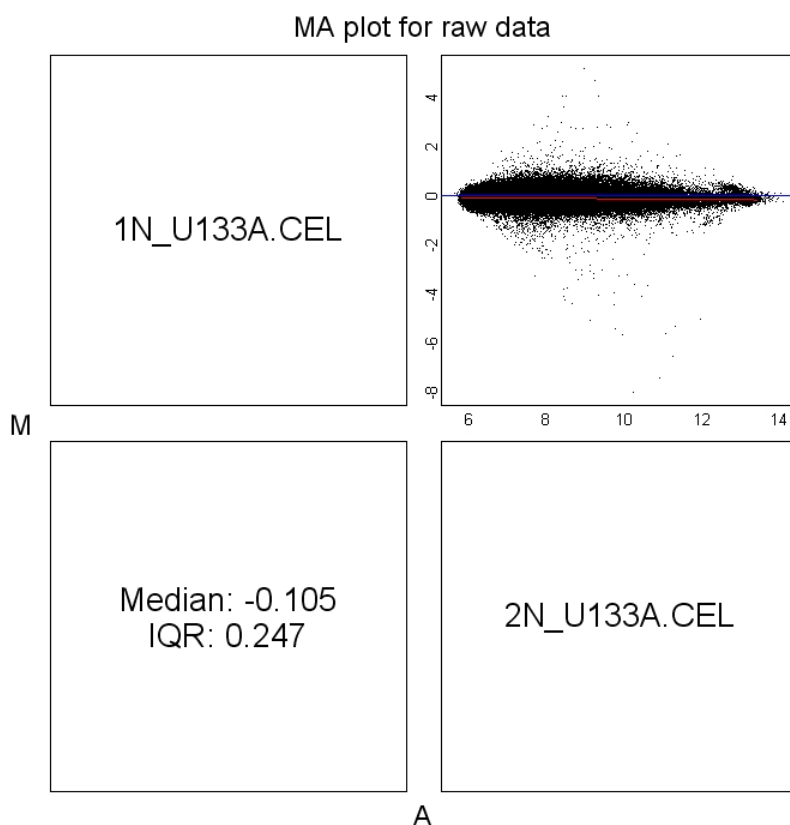


Figure 3.13 MA plot of raw data from one pair of arrays in the UC dataset.

Raw log-intensity ratio (M) is plotted against raw log-intensity average signal (A) for the first two arrays (1N_U133A.CEL and 2N_U133A.CEL) of the UC dataset. The median value of M and the interquartile range (IQR) of M are displayed in the lower left corner.

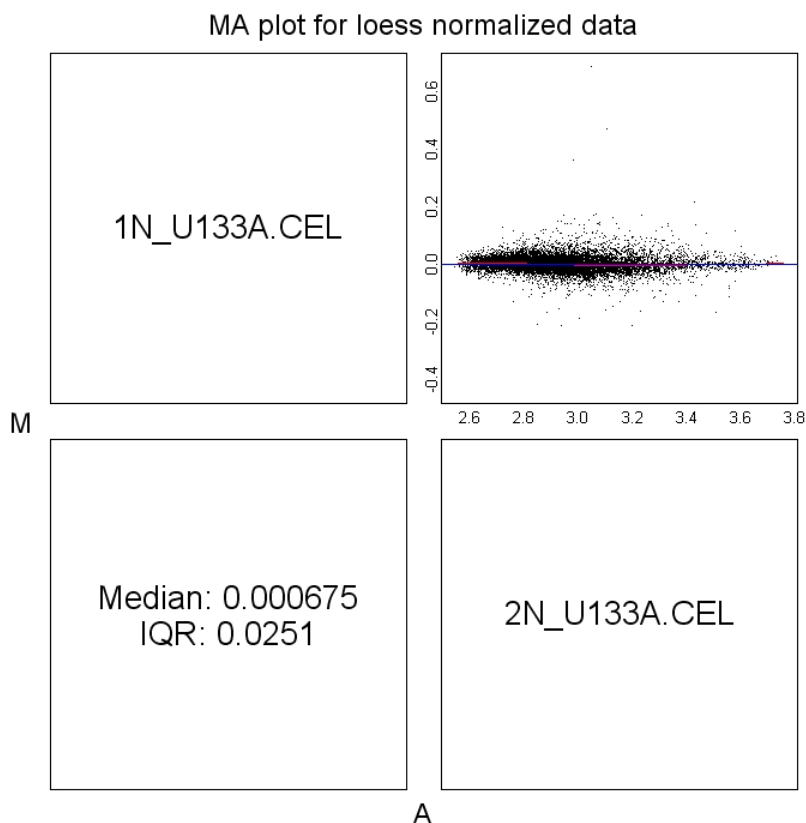


Figure 3.14 MA plot of normalized data from one pair of arrays in the UC dataset.

Normalized log-intensity ratio (M) is plotted against normalized log-intensity average signal (A) for the first two arrays (1N_U133A.CEL and 2N_U133A.CEL) of the UC dataset. The median value of M and the interquartile range (IQR) of M are displayed in the lower left corner.

Rather than plotting many MA plot pairwise comparisons (e.g. 30 x 30 arrays) and examining the variability of each pairwise comparison individually, the MA plots were summarized by taking the median and interquartile range of M for each pairwise comparison. These values were then plotted for each normalization method (Figure 3.15). Ideally, the median and interquartile range of M should be as close to zero as possible, which would indicate the lowest possible overall variability between two arrays. Considering computing time, all procedures except loess could be performed on a desktop computer within one hour. The loess normalization required overnight calculation, but was still feasible for the desktop computer. Of all normalization methods attempted (MAS5.0, GC-RMA, vsn and loess), the loess normalization showed the most compact distribution of points around zero. Vsn and GC-RMA also performed satisfactorily. MAS5.0 yielded the greatest variation between arrays. The normalization procedures yielded similar results for both the UC and CD datasets (CD not shown), and as a result, the loess normalization method was chosen for further data analysis for both datasets.

Density plots of the normalized data identified outlier arrays. In the UC dataset, two samples from one patient were removed from analysis because the overall signals were degraded compared to rest of the arrays (not shown). No samples were removed from the CD dataset.

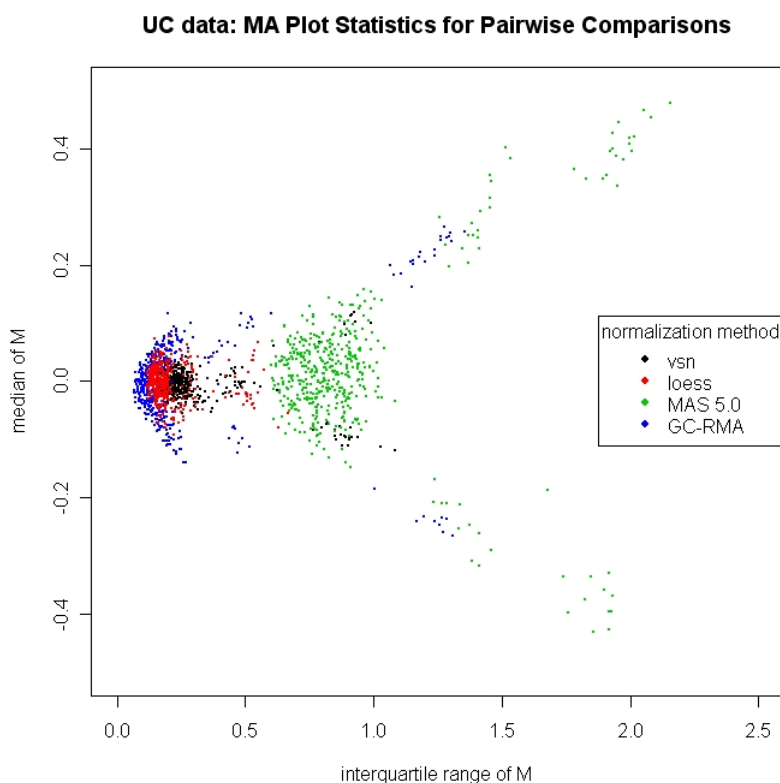


Figure 3.15 Median and interquartile ranges for all pairwise comparisons between arrays in the UC dataset.

Each point represents statistics (interquartile range and median of the log intensity ratio) for one pairwise comparison. Points are colored according to normalization method: vsn (black), loess (red), MAS 5.0 (green), and GC-RMA (blue).

3.3.2 Expression profiling of the non-/inflamed UC mucosa

After normalization and quality control, the UC dataset consisted of 28 arrays, including 10 healthy normal controls, and 18 UC samples (9 paired non-inflamed and inflamed samples from the same patient). Comparisons were made between three groups: normal controls (NC), non-inflamed UC (UC_NI) and inflamed UC (UC_I), as depicted in Figure 3.16. The normalized dataset was further processed by defining which probesets were 'present' on the arrays and eliminating probesets with low variability. A background cutoff was applied to eliminate signals that could not be distinguished from noise. To calculate the background cutoff, a background subset of 27 Affymetrix probesets representing bacterial sequences was extracted from the dataset. The background cutoff was set to the mean of the background subset plus two standard deviations. Next, all probesets were defined to be 'present' if signals were above the background cutoff in at least 70% of arrays in any one group (NC, UC_NI, UC_I). As a final step, only probesets which had an interquartile range in the top 25th percentile of interquartile ranges were selected for differential gene determination. At the end of pre-processing, 5616 probesets remained for further analysis.

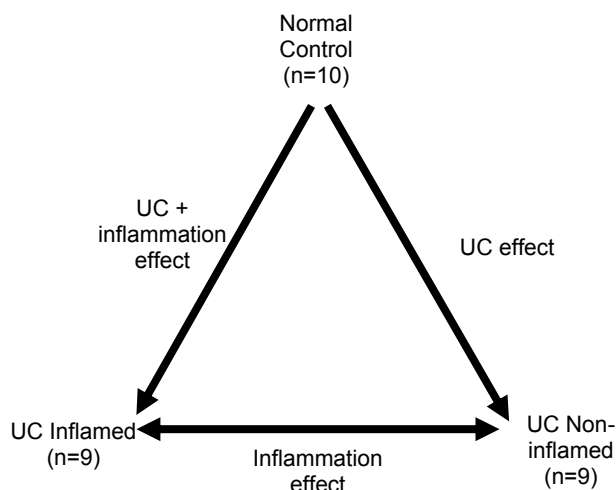


Figure 3.16 Comparisons made between the three groups in the UC dataset.

Comparisons between the three groups (normal controls, non-inflamed UC and inflamed UC) reveal UC effects, inflammation effects, and cumulative UC + inflammation effects. All 18 UC samples were matched samples from the same patient.

3.3.2.1 Differentially regulated genes in the UC dataset

The R package limma was used to apply a linear model to the UC dataset, using a three-factor, one level design. P-values were adjusted for multiple testing using the false discovery rate (Benjamini & Hochberg method). The frequencies of differentially regulated probesets from limma analysis are shown in Table 3.6.

Table 3.6 Number of significantly differentially regulated probesets in the UC dataset, from limma analysis*

Regulation	UC_NI-NC	UC_I-UC_NI	UC_I-NC
Down-regulated	2315 (172)	1435 (144)	2352 (220)
Not regulated	2643 (202)	2205 (142)	1908 (105)
Up-regulated	658 (44)	1976 (132)	1356 (93)

* $p_{adj} < 0.05$, false discovery rate adjustment applied; unknown gene probesets in parentheses

The paired inflamed/non-inflamed samples (from the same patient, but different regions) provided the opportunity to identify differentially regulated genes due to inflammation without the confounding effects of inter-individual differences due to factors such as genotype or local environment. Paired sample significance testing, in the form of t-test and Wilcoxon test, were used to identify differentially expressed genes between pairs of samples (Table 3.7). Paired significance testing yielded approximately 4-7% more differentially regulated probesets than the linear modeling method for UC_I-UC-NI comparison (Figure 3.17).

Table 3.7 Number of significantly differentially regulated genes* in paired inflamed/non-inflamed UC samples

Regulation	UC_I-UC_NI	
	Paired t-test	Paired Wilcoxon test
Down-regulated	1503	1465
Not regulated	1953	2064
Up-regulated	2160	2087

* $p_{adj} < 0.05$, false discovery rate adjustment applied

The top 50 up and down-regulated genes ($p_{adj} < 0.05$ and fold change > 1.2) for each factor comparison are listed in Table 3.8 (UC_NI-NC), Table 3.9 (UC_I-UC_NI), Table 3.10 (UC_I-NC), and Table 3.11 (matched pairs of UC_I-UC_NI). Overall, mRNA expression was observed for known genes associated with intestinal mucosa barrier and immune function. Many genes representing adaptive or innate immune response, such as immunoglobulins (*IGH@*, *IGHm*, *IGLC1*, *IGLC2*, *IGLJ*), MHC genes (*HLA-DRB4*, *HLA-DRA*, *HLA-DQA1*), pro-inflammatory markers (*IL8*, *S100A8*) or anti-microbial peptides (*ADM*, *CXCL9*, *CCL20*, *DEFA5*, *DEFA6*), were up-regulated in inflamed tissue compared to non-inflamed samples (UC or healthy). Additionally, genes involved in colonic mucosal functions such as epithelial repair (*REG1A*; up-regulated in inflamed compared to non-inflamed), pH balance (*CA1*; down-regulated in UC compared to normal), fatty acid transport (*SLC16A1*, *SLC27A2*; up-regulated in non-inflamed UC compared to inflamed UC and normals) and detoxification (*ABCB1*, *CES2*, *CYP2B6*; all down-regulated in inflamed UC compared to normal or non-inflamed UC) were observed in the top 50 differentially regulated gene lists. The differential expression of these intestinal or immunity-related genes validates the fact that the microarrays are detecting expected gene expression differences between inflamed and healthy colon.

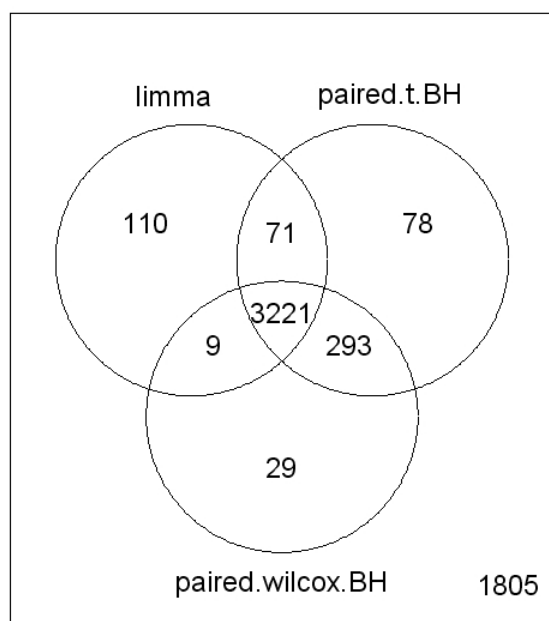


Figure 3.17 Venn diagram showing differentially regulated probesets in the non-/inflamed UC datasets from three analysis methods.

Differentially regulated probesets between inflamed and non-inflamed UC samples were found using different statistical methods. The limma analysis is based on inflammation only, while the paired tests (paired t-test and paired Wilcoxon test) consider that the inflamed/non-inflamed samples are matched pairs. The two paired statistical tests together detected 293 probesets that were not found using the limma method.

Table 3.8 Non-inflamed UC vs. healthy normal controls: top 50 differentially regulated genes from limma analysis

A. Up-regulated probesets from comparison of non-inflamed UC and healthy normal controls

Probe id	Log2Ratio	Gene symbol	Gene name
209600_s_at	0.837	ACOX1	acyl-Coenzyme A oxidase 1, palmitoyl
209869_at	1.34	ADRA2A	adrenergic, alpha-2A-, receptor
206561_s_at	1.113	AKR1B10	aldo-keto reductase family 1, member B10 (aldose reductase)
211357_s_at	0.952	ALDOB	aldolase B, fructose-bisphosphate
204705_x_at	0.852	ALDOB	aldolase B, fructose-bisphosphate
202888_s_at	2.337	ANPEP	alanyl (membrane) aminopeptidase (aminopeptidase N, aminopeptidase M, microsomal aminopeptidase, CD13, p150)
206784_at	1.164	AQP8	aquaporin 8
215100_at	0.894	C6orf105	chromosome 6 open reading frame 105
210133_at	1.158	CCL11	chemokine (C-C motif) ligand 11
214038_at	0.862	CCL8	chemokine (C-C motif) ligand 8
211848_s_at	0.957	CEACAM7	carcinoembryonic antigen-related cell adhesion molecule 7
206198_s_at	0.96	CEACAM7	carcinoembryonic antigen-related cell adhesion molecule 7
206207_at	1.006	CLC	Charcot-Leyden crystal protein
206755_at	0.97	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6
206754_s_at	1.425	CYP2B7P1	cytochrome P450, family 2, subfamily B, polypeptide 7 pseudogene 1
207529_at	1.912	DEFA5	defensin, alpha 5, Paneth cell-specific
207814_at	1.63	DEFA6	defensin, alpha 6, Paneth cell-specific
208250_s_at	1.687	DMBT1	deleted in malignant brain tumors 1
219955_at	1.054	ECAT11	hypothetical protein FLJ10884
222262_s_at	2.244	ETNK1	ethanolamine kinase 1
219017_at	2.734	ETNK1	ethanolamine kinase 1
208228_s_at	0.847	FGFR2	fibroblast growth factor receptor 2
203638_s_at	1.017	FGFR2	fibroblast growth factor receptor 2
219954_s_at	0.864	GBA3	glucosidase, beta, acid 3 (cytosolic)
219722_s_at	0.919	GDPD3	glycerophosphodiester phosphodiesterase domain containing 3
204875_s_at	0.838	GMDS	GDP-mannose 4,6-dehydratase
204607_at	1.601	HMGCS2	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)
205366_s_at	0.924	HOXB6	homeo box B6
206294_at	1.659	HSD3B2	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2
214022_s_at	0.996	IFITM1	interferon induced transmembrane protein 1 (9-27)
201601_x_at	0.95	IFITM1	interferon induced transmembrane protein 1 (9-27)
202718_at	1.82	IGFBP2	insulin-like growth factor binding protein 2, 36kDa
209374_s_at	1.21	IGHM	immunoglobulin heavy constant mu
219543_at	0.83	MAWBP	MAWD binding protein
207251_at	1.014	MEP1B	mepirin A, beta
200632_s_at	0.902	NDRG1	N-myc downstream regulated gene 1
218888_s_at	0.975	NETO2	neuropilin (NRP) and tolloid (TLL)-like 2
210519_s_at	1.029	NQO1	NAD(P)H dehydrogenase, quinone 1
201468_s_at	1.107	NQO1	NAD(P)H dehydrogenase, quinone 1
206340_at	0.942	NR1H4	nuclear receptor subfamily 1, group H, member 4
205552_s_at	0.89	OAS1	2',5'-oligoadenylate synthetase 1, 40/46kDa
205660_at	0.971	OASL	2'-5'-oligoadenylate synthetase-like
212768_s_at	2.662	OLFM4	olfactomedin 4
207558_s_at	1.839	PITX2	paired-like homeodomain transcription factor 2
204304_s_at	1.472	PROM1	prominin 1
209752_at	3.628	REG1A	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)
205815_at	1.705	REG3A	regenerating islet-derived 3 alpha
203225_s_at	0.995	RFK	riboflavin kinase
203224_at	0.891	RFK	riboflavin kinase
217085_at	1.694	SLC14A2	Solute carrier family 14 (urea transporter), member 2
209900_s_at	1.097	SLC16A1	solute carrier family 16 (monocarboxylic acid transporters), member 1
202236_s_at	1.369	SLC16A1	solute carrier family 16 (monocarboxylic acid transporters), member 1
202234_s_at	0.981	SLC16A1	solute carrier family 16 (monocarboxylic acid transporters), member 1

201920_at	1.086	SLC20A1	solute carrier family 20 (phosphate transporter), member 1
205769_at	1.007	SLC27A2	solute carrier family 27 (fatty acid transporter), member 2
205768_s_at	0.831	SLC27A2	solute carrier family 27 (fatty acid transporter), member 2
205799_s_at	0.893	SLC3A1	solute carrier family 3 (cystine, dibasic and neutral amino acid transporters, activator of cystine, dibasic and neutral amino acid transport), member 1
207212_at	1.019	SLC9A3	solute carrier family 9 (sodium/hydrogen exchanger), isoform 3
215125_s_at	0.963	UGT1A10	UDP glucuronosyltransferase 1 family, polypeptide A10
208596_s_at	0.993	UGT1A10	UDP glucuronosyltransferase 1 family, polypeptide A10
207126_x_at	0.939	UGT1A10	UDP glucuronosyltransferase 1 family, polypeptide A10
204532_x_at	0.845	UGT1A10	UDP glucuronosyltransferase 1 family, polypeptide A10
206094_x_at	0.911	UGT1A6	UDP glucuronosyltransferase 1 family, polypeptide A6

B. Down-regulated probesets from comparison of non-inflamed UC and healthy normal controls

213921_at	-2.544	---	represented by NM_001048
202274_at	-0.986	ACTG2	actin, gamma 2, smooth muscle, enteric
220468_at	-1.182	ARF7	ADP-ribosylation factor 7
221530_s_at	-1.012	BHLHB3	basic helix-loop-helix domain containing, class B, 3
213134_x_at	-0.952	BTG3	BTG family, member 3
205950_s_at	-1.046	CA1	carbonic anhydrase I
32128_at	-1.19	CCL18	chemokine (C-C motif) ligand 18 (pulmonary and activation-regulated)
209924_at	-1.045	CCL18	chemokine (C-C motif) ligand 18 (pulmonary and activation-regulated)
221164_x_at	-1.295	CHST5	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 5
200884_at	-0.928	CKB	creatine kinase, brain
214598_at	-2.559	CLDN8	claudin 8
205081_at	-1.759	CRIP1	cysteine-rich protein 1 (intestinal)
201842_s_at	-1.015	EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1
201889_at	-1.007	FAM3C	family with sequence similarity 3, member C
216442_x_at	-2.274	FN1	fibronectin 1
212464_s_at	-2.18	FN1	fibronectin 1
211719_x_at	-2.255	FN1	fibronectin 1
210495_x_at	-2.176	FN1	fibronectin 1
203697_at	-0.899	FRZB	frizzled-related protein
206422_at	-2.604	GCG	glucagon
205498_at	-0.958	GHR	growth hormone receptor
205490_x_at	-0.904	GJB3	gap junction protein, beta 3, 31kDa (connexin 31)
205042_at	-1.057	GNE	glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase
209844_at	-2.044	HOXB13	homeo box B13
210095_s_at	-1.059	IGFBP3	insulin-like growth factor binding protein 3
211372_s_at	-0.853	IL1R2	interleukin 1 receptor, type II
205403_at	-0.938	IL1R2	interleukin 1 receptor, type II
221091_at	-2.636	INSL5	insulin-like 5
212192_at	-1.126	KCTD12	potassium channel tetramerisation domain containing 12
212188_at	-0.914	KCTD12	potassium channel tetramerisation domain containing 12
212573_at	-0.844	KIAA0830	KIAA0830 protein
203726_s_at	-0.935	LAMA3	laminin, alpha 3
202068_s_at	-1.239	LDLR	low density lipoprotein receptor (familial hypercholesterolemia)
206268_at	-1.188	LEFTY1	left-right determination factor 1
208450_at	-1.508	LGALS2	lectin, galactoside-binding, soluble, 2 (galectin 2)
36711_at	-0.847	MAFF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog F (avian)
204673_at	-1.13	MUC2	mucin 2, intestinal/tracheal
207217_s_at	-0.865	NOX1	NADPH oxidase 1
206418_at	-0.881	NOX1	NADPH oxidase 1
209803_s_at	-0.947	PHLDA2	pleckstrin homology-like domain, family A, member 2
208121_s_at	-0.869	PTPRO	protein tyrosine phosphatase, receptor type, O
211253_x_at	-1.287	PYY	peptide YY
207080_s_at	-2.959	PYY	peptide YY
209496_at	-1.155	RARRES2	retinoic acid receptor responder (tazarotene induced) 2
202388_at	-0.919	RGS2	regulator of G-protein signalling 2, 24kDa
204351_at	-1.662	S100P	S100 calcium binding protein P
205979_at	-0.852	SCGB2A1	secretoglobulin, family 2A, member 1
205464_at	-0.883	SCNN1B	sodium channel, nonvoltage-gated 1, beta (Liddle syndrome)
206664_at	-1.073	SI	sucrase-isomaltase (alpha-glucosidase)
207249_s_at	-0.883	SLC28A2	solute carrier family 28 (sodium-coupled nucleoside transporter), member 2
205185_at	-1.338	SPINK5	serine protease inhibitor, Kazal type 5
213994_s_at	-1.086	SPON1	spondin 1, extracellular matrix protein
209436_at	-1.215	SPON1	spondin 1, extracellular matrix protein
205547_s_at	-1.042	TAGLN	transgelin
205009_at	-1.102	TFF1	trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in)
201110_s_at	-0.912	THBS1	thrombospondin 1
215034_s_at	-0.883	TM4SF1	transmembrane 4 L six family member 1
209386_at	-1.011	TM4SF1	transmembrane 4 L six family member 1
214601_at	-0.916	TPH1	tryptophan hydroxylase 1 (tryptophan 5-monoxygenase)
203892_at	-2.256	WFDC2	WAP four-disulfide core domain 2

Table 3.9 Inflamed and non-inflamed UC: top 50 differentially regulated genes from limma analysis

A. Up-regulated probesets from comparison of inflamed and non-inflamed UC			
Probe id	Log2Ratio	Gene symbol	Gene name
202912_at	1.581	ADM	adrenomedullin
217238_s_at	1.256	ALDOB	aldolase B, fructose-bisphosphate
211357_s_at	1.351	ALDOB	aldolase B, fructose-bisphosphate
208747_s_at	1.247	C1S	complement component 1, s subcomponent
217767_at	1.554	C3	complement component 3
210133_at	1.346	CCL11	chemokine (C-C motif) ligand 11
205476_at	1.503	CCL20	chemokine (C-C motif) ligand 20
209396_s_at	2.302	CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)
209395_at	1.997	CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)
202310_s_at	1.775	COL1A1	collagen, type I, alpha 1
202404_s_at	1.769	COL1A2	collagen, type I, alpha 2
202403_s_at	1.782	COL1A2	collagen, type I, alpha 2
215076_s_at	1.349	COL3A1	collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)
201852_x_at	1.387	COL3A1	collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)
201438_at	1.973	COL6A3	collagen, type VI, alpha 3
205927_s_at	1.279	CTSE	cathepsin E
204470_at	1.728	CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)
205242_at	1.897	CXCL13	chemokine (C-X-C motif) ligand 13 (B-cell chemoattractant)
209774_x_at	1.43	CXCL2	chemokine (C-X-C motif) ligand 2
207850_at	1.279	CXCL3	chemokine (C-X-C motif) ligand 3
201289_at	1.366	CYR61	cysteine-rich, angiogenic inducer, 61
201926_s_at	1.713	DAF	decay accelerating factor for complement (CD55, Cromer blood group system)
201925_s_at	1.556	DAF	decay accelerating factor for complement (CD55, Cromer blood group system)
211896_s_at	1.275	DCN	decorin
208250_s_at	1.29	DMBT1	deleted in malignant brain tumors 1
219727_at	2.804	DUOX2	dual oxidase 2
218469_at	2.393	GREM1	gremlin 1, cysteine knot superfamily, homolog (Xenopus laevis)
218468_s_at	2.62	GREM1	gremlin 1, cysteine knot superfamily, homolog (Xenopus laevis)
212671_s_at	1.242	HLA-DQA1	major histocompatibility complex, class II, DQ alpha
210982_s_at	1.341	HLA-DRA	major histocompatibility complex, class II, DR alpha
211959_at	1.368	IGFBP5	insulin-like growth factor binding protein 5
211430_s_at	1.688	IGH@	immunoglobulin heavy locus
211634_x_at	1.455	IGHM	immunoglobulin heavy constant mu
216560_x_at	1.61	IGLC1	Immunoglobulin lambda constant 1 (Mcg marker)
215214_at	2.223	IGLC2	Immunoglobulin lambda variable 3-21
216853_x_at	1.598	IGLJ3	Immunoglobulin lambda joining 3
202859_x_at	1.477	IL8	interleukin 8
212531_at	2.861	LCN2	lipocalin 2 (oncogene 24p3)
204475_at	1.919	MMP1	matrix metalloproteinase 1 (interstitial collagenase)
204580_at	1.452	MMP12	matrix metalloproteinase 12 (macrophage elastase)
205828_at	2.095	MMP3	matrix metalloproteinase 3 (stromelysin 1, progelatinase)
219630_at	1.86	PDZK1IP1	PDZK1 interacting protein 1
41469_at	3.322	PI3	protease inhibitor 3, skin-derived (SKALP)
203691_at	3.256	PI3	protease inhibitor 3, skin-derived (SKALP)
203649_s_at	1.605	PLA2G2A	phospholipase A2, group IIA (platelets, synovial fluid)
205886_at	1.867	REG1B	regenerating islet-derived 1 beta (pancreatic stone protein, pancreatic thread protein)
202917_s_at	1.883	S100A8	S100 calcium binding protein A8 (calgranulin A)
204351_at	2.694	S100P	S100 calcium binding protein P
204855_at	1.702	SERPINB5	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 5
219795_at	2.037	SLC6A14	solute carrier family 6 (amino acid transporter), member 14
200665_s_at	1.256	SPARC	secreted protein, acidic, cysteine-rich (osteonectin)
207214_at	2.037	SPINK4	serine protease inhibitor, Kazal type 4
205185_at	1.569	SPINK5	serine protease inhibitor, Kazal type 5
205009_at	1.767	TFF1	trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in)
201666_at	2.017	TIMP1	tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor)
205890_s_at	2.373	UBD	ubiquitin D
205844_at	1.675	VNN1	vanin 1 // vanin 1

B. Down-regulated probesets from comparison of inflamed and non-inflamed UC			
209735_at	-1.676	ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2
209600_s_at	-0.988	ACOX1	acyl-Coenzyme A oxidase 1, palmitoyl
206262_at	-1.74	ADH1C	alcohol dehydrogenase 1C (class I), gamma polypeptide
209869_at	-0.97	ADRA2A	adrenergic, alpha-2A-, receptor // adrenergic, alpha-2A-, receptor
206561_s_at	-1.077	AKR1B10	aldo-keto reductase family 1, member B10 (aldose reductase)
202888_s_at	-1.929	ANPEP	alanyl (membrane) aminopeptidase (aminopeptidase N, aminopeptidase M, microsomal aminopeptidase, CD13, p150)
206632_s_at	-1.05	APOBEC3B	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B
206784_at	-3.642	AQP8	aquaporin 8
208677_s_at	-0.954	BSG	basigin (OK blood group)
203571_s_at	-1.098	C10orf116	chromosome 10 open reading frame 116
209301_at	-0.949	CA2	carbonic anhydrase II
209668_x_at	-0.969	CES2	carboxylesterase 2 (intestine, liver)
206755_at	-0.993	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6
206754_s_at	-1.348	CYP2B7P1	cytochrome P450, family 2, subfamily B, polypeptide 7 pseudogene 1
222262_s_at	-1.94	ETNK1	ethanolamine kinase 1
219017_at	-2.094	ETNK1	ethanolamine kinase 1
208228_s_at	-1.03	FGFR2	fibroblast growth factor receptor 2
203638_s_at	-1.284	FGFR2	fibroblast growth factor receptor 2
207003_at	-1.256	GUCA2A	guanylate cyclase activator 2A (guanylin)
204607_at	-2.907	HMGCS2	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)
204130_at	-1.061	HSD11B2	hydroxysteroid (11-beta) dehydrogenase 2
204818_at	-1.568	HSD17B2	hydroxysteroid (17-beta) dehydrogenase 2
206294_at	-1.681	HSD3B2	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2
202718_at	-0.957	IGFBP2	insulin-like growth factor binding protein 2, 36kDa
206149_at	-1.761	LOC63928	hepatocellular carcinoma antigen gene 520
212741_at	-1.107	MAOA	monoamine oxidase A
219543_at	-1.125	MAWBP	MAWD binding protein

207251_at	-0.962	MEP1B	meprin A, beta
218756_s_at	-1.243	MGC4172	short-chain dehydrogenase/reductase
217165_x_at	-1.051	MT1F	metallothionein 1F (functional)
213629_x_at	-1.244	MT1F	metallothionein 1F (functional)
204745_x_at	-1.253	MT1G	metallothionein 1G
206461_x_at	-1.04	MT1H	metallothionein 1H
217546_at	-1.008	MT1K	metallothionein 1K
208581_x_at	-1.019	MT1X	metallothionein 1X
212338_at	-1.011	MYO1D	myosin ID
205660_at	-1.022	OASL	2'-5'-oligoadenylate synthetase-like
209791_at	-1.206	PADI2	peptidyl arginine deiminase, type II
208383_s_at	-2.064	PCK1	phosphoenolpyruvate carboxykinase 1 (soluble)
213407_at	-1.174	PHLPPL	PH domain and leucine rich repeat protein phosphatase-like
203335_at	-0.975	PHYH	phytanoyl-CoA hydroxylase (Refsum disease)
207558_s_at	-1.806	PITX2	paired-like homeodomain transcription factor 2
204304_s_at	-0.99	PROM1	prominin 1
217085_at	-1.662	SLC14A2	Solute carrier family 14 (urea transporter), member 2
209900_s_at	-1.107	SLC16A1	solute carrier family 16 (monocarboxylic acid transporters), member 1
202236_s_at	-1.42	SLC16A1	solute carrier family 16 (monocarboxylic acid transporters), member 1
202234_s_at	-0.994	SLC16A1	solute carrier family 16 (monocarboxylic acid transporters), member 1
201920_at	-1.66	SLC20A1	solute carrier family 20 (phosphate transporter), member 1
205074_at	-0.991	SLC22A5	solute carrier family 22 (organic cation transporter), member 5
205097_at	-2.085	SLC26A2	solute carrier family 26 (sulfate transporter), member 2
220435_at	-1.103	SLC30A10	solute carrier family 30 (zinc transporter), member 10
205799_s_at	-1.02	SLC3A1	solute carrier family 3 (cystine, dibasic and neutral amino acid transporters, activator of cystine, dibasic and neutral amino acid transport), member 1
215125_s_at	-1.298	UGT1A10	UDP glucuronosyltransferase 1 family, polypeptide A10
208596_s_at	-1.145	UGT1A10	UDP glucuronosyltransferase 1 family, polypeptide A10
207126_x_at	-1.314	UGT1A10	UDP glucuronosyltransferase 1 family, polypeptide A10
204532_x_at	-1.203	UGT1A10	UDP glucuronosyltransferase 1 family, polypeptide A10
206094_x_at	-1.303	UGT1A6	UDP glucuronosyltransferase 1 family, polypeptide A6
219948_x_at	-1.541	UGT2A3	UDP glucuronosyltransferase 2 family, polypeptide A3

Table 3.10 Inflamed UC and healthy normal controls: Top 50 differentially regulated genes from limma analysis

A. Up-regulated probesets from comparison of inflamed UC and healthy normal controls

Probe id	Log2Ratio	Gene symbol	Gene name
211637_x_at	1.372	---	Rearranged Ig mu-chain (V4-59/DIR1-D5'-D21-9/JH4b)
217238_s_at	2.033	ALDOB	aldolase B, fructose-bisphosphate
211357_s_at	2.303	ALDOB	aldolase B, fructose-bisphosphate
204705_x_at	2.01	ALDOB	aldolase B, fructose-bisphosphate
202357_s_at	1.391	BF	B-factor, properdin
211368_s_at	1.452	CASP1	caspase 1, apoptosis-related cysteine protease (interleukin 1, beta, convertase)
211366_x_at	1.354	CASP1	caspase 1, apoptosis-related cysteine protease (interleukin 1, beta, convertase)
209970_x_at	1.379	CASP1	caspase 1, apoptosis-related cysteine protease (interleukin 1, beta, convertase)
210133_at	2.505	CCL11	chemokine (C-C motif) ligand 11
209396_s_at	2.443	CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)
209395_at	2.247	CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)
202404_s_at	1.593	COL1A2	collagen, type I, alpha 2
201438_at	1.591	COL6A3	collagen, type VI, alpha 3
204470_at	1.757	CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)
205242_at	1.68	CXCL13	chemokine (C-X-C motif) ligand 13 (B-cell chemoattractant)
209774_x_at	1.381	CXCL2	chemokine (C-X-C motif) ligand 2
207850_at	1.32	CXCL3	chemokine (C-X-C motif) ligand 3
203915_at	1.523	CXCL9	chemokine (C-X-C motif) ligand 9
201289_at	1.459	CYR61	cysteine-rich, angiogenic inducer, 61
201926_s_at	1.958	DAF	decay accelerating factor for complement (CD55, Cromer blood group system)
201925_s_at	1.602	DAF	decay accelerating factor for complement (CD55, Cromer blood group system)
207529_at	2.672	DEFA5	defensin, alpha 5, Paneth cell-specific
208250_s_at	2.977	DMBT1	deleted in malignant brain tumors 1
219727_at	3.005	DUOX2	dual oxidase 2
219117_s_at	1.325	FKBP11	FK506 binding protein 11, 19 kDa
202269_x_at	1.572	GBP1	guanylate binding protein 1, interferon-inducible, 67kDa
218469_at	1.831	GREM1	gremlin 1, cysteine knot superfamily, homolog (Xenopus laevis)
218468_s_at	1.997	GREM1	gremlin 1, cysteine knot superfamily, homolog (Xenopus laevis)
209728_at	1.531	HLA-DRB4	major histocompatibility complex, class II, DR beta 4
201601_x_at	1.348	IFITM1	interferon induced transmembrane protein 1 (9-27)
211959_at	1.36	IGFBP5	insulin-like growth factor binding protein 5
211430_s_at	2.153	IGH@	immunoglobulin heavy locus
216491_x_at	1.51	IGHM	immunoglobulin heavy constant mu
211634_x_at	1.342	IGHM	immunoglobulin heavy constant mu
216560_x_at	1.589	IGLC1	Immunoglobulin lambda constant 1 (Mcg marker)
215214_at	1.9	IGLC2	Immunoglobulin lambda variable 3-21
216853_x_at	1.482	IGLJ3	Immunoglobulin lambda joining 3
202859_x_at	1.438	IL8	interleukin 8
212531_at	3.261	LCN2	lipocalin 2 (oncogene 24p3)
204475_at	2.038	MMP1	matrix metalloproteinase 1 (interstitial collagenase)
204580_at	1.945	MMP12	matrix metalloproteinase 12 (macrophage elastase)
205828_at	2.049	MMP3	matrix metalloproteinase 3 (stromelysin 1, progelatinase)
212768_s_at	3.056	OLFM4	olfactomedin 4
219630_at	1.682	PDZK1IP1	PDZK1 interacting protein 1
41469_at	2.564	PI3	protease inhibitor 3, skin-derived (SKALP)
203691_at	2.239	PI3	protease inhibitor 3, skin-derived (SKALP)
205267_at	1.327	POU2AF1	POU domain, class 2, associating factor 1
204279_at	1.506	PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional protease 2)
212187_x_at	1.374	PTGDS	prostaglandin D2 synthase 21kDa (brain)
211748_x_at	1.422	PTGDS	prostaglandin D2 synthase 21kDa (brain)
209752_at	4.783	REG1A	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)
205886_at	2.37	REG1B	regenerating islet-derived 1 beta (pancreatic stone protein, pancreatic thread protein)
205815_at	2.775	REG3A	regenerating islet-derived 3 alpha
209071_s_at	1.328	RGS5	regulator of G-protein signalling 5
202917_s_at	2.21	S100A8	S100 calcium binding protein A8 (calgranulin A)
219795_at	1.842	SLC6A14	solute carrier family 6 (amino acid transporter), member 14
207214_at	2.684	SPINK4	serine protease inhibitor, Kazal type 4

201666_at	1.495	TIMP1	tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor)
205890_s_at	2.296	UBD	ubiquitin D
205844_at	1.604	VNN1	vanin 1

B. Down-regulated probesets from comparison of inflamed UC and healthy normal controls

213921_at	-2.405	---	transcript represented by NM_001048
210524_x_at	-1.106	---	transcript represented by AF078844
204719_at	-1.188	ABCA8	ATP-binding cassette, sub-family A (ABC1), member 8
209993_at	-1.134	ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 8
209994_s_at	-1.128	ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1
206262_at	-1.214	ADH1C	alcohol dehydrogenase 1C (class I), gamma polypeptide
206784_at	-2.477	AQP8	aquaporin 8
220468_at	-1.504	ARF7	ADP-ribosylation factor 7
205950_s_at	-1.896	CA1	carbonic anhydrase I
218309_at	-0.977	CAMK2N1	calcium/calmodulin-dependent protein kinase II inhibitor 1
213509_x_at	-0.989	CES2	carboxylesterase 2 (intestine, liver)
209668_x_at	-1.084	CES2	carboxylesterase 2 (intestine, liver)
204697_s_at	-1.139	CHGA	chromogranin A (parathyroid secretory protein 1)
200884_at	-1.563	CKB	creatine kinase, brain
214598_at	-2.667	CLDN8	claudin 8
205081_at	-1.084	CRIP1	cysteine-rich protein 1 (intestinal)
220161_s_at	-0.97	EPB41L4B	erythrocyte membrane protein band 4.1 like 4B
216442_x_at	-1.176	FN1	fibronectin 1
212464_s_at	-1.151	FN1	fibronectin 1
211719_x_at	-1.059	FN1	fibronectin 1
210495_x_at	-1.116	FN1	fibronectin 1
206422_at	-1.734	GCG	glucagon
207003_at	-1.961	GUCA2A	guanylate cyclase activator 2A (guanylin)
207502_at	-1.426	GUCA2B	guanylate cyclase activator 2B (uroguanylin)
204607_at	-1.306	HMGCS2	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)
209844_at	-1.657	HOXB13	homeo box B13
211372_s_at	-1.518	IL1R2	interleukin 1 receptor, type II
205403_at	-1.633	IL1R2	interleukin 1 receptor, type II
221091_at	-2.484	INSL5	insulin-like 5
212573_at	-1.021	KIAA0830	KIAA0830 protein
208450_at	-1.33	LGALS2	lectin, galactoside-binding, soluble, 2 (galectin 2)
206149_at	-1.608	LOC63928	hepatocellular carcinoma antigen gene 520
212741_at	-1.204	MAOA	monoamine oxidase A
218756_s_at	-1.124	MGC4172	short-chain dehydrogenase/reductase
212859_x_at	-1.208	MT1E	metallothionein 1E (functional)
217165_x_at	-1.649	MT1F	metallothionein 1F (functional)
213629_x_at	-1.73	MT1F	metallothionein 1F (functional)
204745_x_at	-1.712	MT1G	metallothionein 1G
206461_x_at	-1.573	MT1H	metallothionein 1H
217546_at	-1.734	MT1K	metallothionein 1K
208581_x_at	-1.757	MT1X	metallothionein 1X
204326_x_at	-1.495	MT1X	metallothionein 1X
212185_x_at	-1.139	MT2A	metallothionein 2A
209791_at	-1.442	PADI2	peptidyl arginine deiminase, type II
208383_s_at	-1.944	PCK1	phosphoenolpyruvate carboxykinase 1 (soluble)
211253_x_at	-1.158	PYY	peptide YY
207080_s_at	-2.499	PYY	peptide YY
201785_at	-1.125	RNASE1	ribonuclease, RNase A family, 1 (pancreatic)
205464_at	-1.128	SCNN1B	sodium channel, nonvoltage-gated 1, beta (Liddle syndrome)
214433_s_at	-1.406	SELENBP1	selenium binding protein 1
33322_i_at	-0.999	SFN	stratifin
205097_at	-2.019	SLC26A2	solute carrier family 26 (sulfate transporter), member 2
203908_at	-1.305	SLC4A4	solute carrier family 4, sodium bicarbonate cotransporter, member 4
214601_at	-1.308	TPH1	tryptophan hydroxylase 1 (tryptophan 5-monooxygenase)
219948_x_at	-1.58	UGT2A3	UDP glucuronosyltransferase 2 family, polypeptide A3
207432_at	-1.033	VMD2L1	vitelliform macular dystrophy 2-like 1
203892_at	-1.849	WFDC2	WAP four-disulfide core domain 2
214142_at	-0.988	ZG16	zymogen granule protein 16
206059_at	-1.138	ZNF91	zinc finger protein 91 (HPF7, HTF10)

Table 3.11 Table of differentially regulated genes from paired statistical analysis of inflamed and non-inflamed UC

A. Up-regulated probesets from paired statistical analysis, not detected by limma, 17 genes			
Probe id	Log2Ratio	Gene symbol	Gene name
209001_s_at	0.263	ANAPC13	anaphase promoting complex subunit 13
210427_x_at	0.381	ANXA2	annexin A2
203951_at	0.338	CNN1	calponin 1, basic, smooth muscle
213831_at	0.581	HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1
212999_x_at	0.324	HLA-DQB1	Major histocompatibility complex, class II, DQ beta 1
209480_at	0.351	HLA-DQB1	Major histocompatibility complex, class II, DQ beta 1
213674_x_at	0.321	IGHD	immunoglobulin heavy constant delta
201627_s_at	0.357	INSIG1	insulin induced gene 1
201626_at	0.411	INSIG1	insulin induced gene 1
201088_at	0.337	KPNA2	karyopherin alpha 2 (RAG cohort 1, importin alpha 1)
218717_s_at	0.268	LEPREL1	leprecan-like 1
212768_s_at	0.394	OLFM4	olfactomedin 4
206214_at	0.292	PLA2G7	phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)
211924_s_at	0.479	PLAUR	plasminogen activator, urokinase receptor
210845_s_at	0.577	PLAUR	plasminogen activator, urokinase receptor
214288_s_at	0.277	PSMB1	proteasome (prosome, macropain) subunit, beta type, 1
209146_at	0.3	SC4MOL	sterol-C4-methyl oxidase-like
207249_s_at	0.433	SLC28A2	solute carrier family 28 (sodium-coupled nucleoside transporter), member 2
218368_s_at	0.269	TNFRSF12A	tumor necrosis factor receptor superfamily, member 12A
204141_at	0.452	TUBB2	tubulin, beta 2

B. Down-regulated probesets from paired statistical analysis, not detected by limma, 12 genes			
221484_at	-0.265	B4GALT5	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 5
220421_at	-0.321	BTNL8	butyrophilin-like 8
211549_s_at	-0.277	HPGD	hydroxyprostaglandin dehydrogenase 15-(NAD)
220834_at	-1.164	MS4A12	membrane-spanning 4-domains, subfamily A, member 12
220186_s_at	-0.265	PC-LKC	protocadherin LKC
205593_s_at	-0.439	PDE9A	phosphodiesterase 9A
219669_at	-0.786	PRV1	polycythemia rubra vera 1
213716_s_at	-0.319	SECTM1	secreted and transmembrane 1
204019_s_at	-0.512	SH3YL1	SH3 domain containing, Ysc84-like 1 (S. cerevisiae)
204981_at	-0.342	SLC22A18	solute carrier family 22 (organic cation transporter), member 18
207392_x_at	-0.616	UGT2B15	UDP glucuronosyltransferase 2 family, polypeptide B15
207245_at	-1.089	UGT2B17	UDP glucuronosyltransferase 2 family, polypeptide B17

3.3.2.2 Functional grouping of differentially regulated transcripts for UC dataset

In order to place the differentially regulated probesets into a biological context, the Fisher test for a 2X2 contingency table was used to calculate p-values for the likelihood that differentially regulated genes were enriched (better than chance) within a given Gene Ontology category (Table 3.12). A total of 5616 probesets and 18824 GO terms were inputted into Fisher test analysis, resulting in 4281 GO terms being matched to the probesets.

Table 3.12 Frequency of significantly enriched GO terms* in the UC dataset comparisons

Root GO term	UC_NI-NC	UC_I-UC_NI	UC_I-NC
Biological process	313 (139) [†]	360 (80)	377 (83)
Molecular function	225 (96)	274 (66)	215 (45)
Cellular component	85 (41)	69 (12)	94 (15)

* p<0.05, no multiple testing correction applied; [†] numbers in parentheses indicate number of significant GO terms unique for each comparison

Significantly enriched GO terms that were present in all three UC comparisons included: response to stimulus (including defense response, immune response, humoral immune response), apoptosis (including anti-apoptosis, negative regulation of apoptosis, negative regulation of programmed cell

death), cell-cell signaling, caspase activity, MHC class II receptor activity, pattern binding, protein binding, extracellular matrix, membrane fraction, and basement membrane.

More interesting were the significantly enriched GO terms that were specific for a particular comparison. For example, UC-specific effects that were independent of inflammation could be isolated from the comparison between non-inflamed UC and normal controls. These effects included various GO terms involved in transport (monocarboxylic acid transporter activity, channel or pore class transporter activity, protein transporter activity, water channel activity, solute:sodium symporter activity), functions associated with RNA (RNA processing, RNA splicing, mRNA catabolism, nonsense-mediated decay), and nuclear transport (nucleocytoplasmic transport, protein-nucleus import, protein-nucleus import, docking). Inflammation effects within the context of UC (UC_I-UC_NI comparison) included GO terms for fatty acid biosynthesis/metabolism (butyrate metabolism, gamma-aminobutyric acid metabolism), innate immunity (immune cell mediated cytotoxicity, natural killer cell mediated cytotoxicity), and signaling (epidermal growth factor receptor signaling pathway, regulation of G-protein coupled receptor protein signaling pathway, nitric oxide mediated signal transduction). The cumulative effects of UC and inflammation compared to normal controls (UC_I-NC comparison) yielded a significant enrichment in signaling (intracellular signaling cascade, enzyme linked receptor protein signaling pathway, Wnt receptor signaling pathway, transmembrane receptor protein tyrosine phosphatase signaling pathway), immune response (cell-mediated immune response, innate immune response, positive regulation of cytokine biosynthesis, T-helper 1 type immune response), and cell junction.

3.3.3 Expression profiling of the non-/inflamed CD mucosa

After normalization and quality control, the CD dataset consisted of 44 arrays, including 25 healthy normal controls, and 19 CD samples (8 inflamed, 7 non-inflamed and 4 inflamed and treated with immunosuppressives (steroids or azathioprine)). Four of the non-inflamed/inflamed samples were paired samples from the same patient, but taken from different regions. Comparisons were made between four groups: normal controls (NC), non-inflamed CD (CD_NI) and inflamed CD (CD_I), and inflamed CD patients treated with immunosuppressives (CD_IS) as depicted in Figure 3.18.

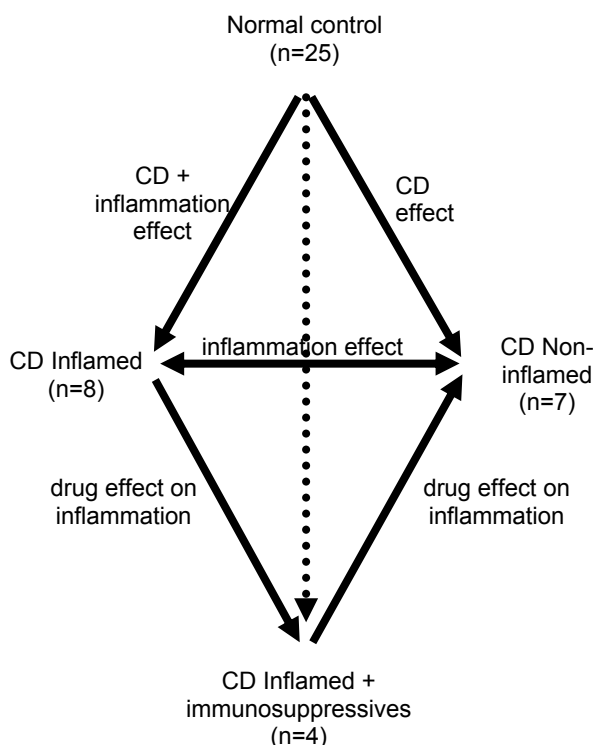


Figure 3.18 Comparison groups for inflamed/non-inflamed CD analysis

Comparisons made between each of the four groups (normal controls, non-inflamed CD, inflamed CD, and inflamed CD treated with immunosuppressives, respectively) reveal CD effects, inflammation effects, cumulative CD + inflammation effects, and effects of immunosuppressives on inflamed samples.

As previously described for the UC dataset, the normalized CD dataset was further processed by defining which probesets were 'present' on the arrays and eliminating probesets with low variability. The background cutoff was set to the mean of the background subset plus two standard deviations. Next, all probesets were defined to be 'present' if signals were above the background cutoff in at least 70% of arrays in any one group (NC, CD_NI, CD_I, CD_IS). As a final step, only probesets which had an interquartile range in the top 25th percentile of interquartile ranges were selected for differential gene determination. At the end of pre-processing, 5985 probesets from U133 chip A and 5490 probesets from U133 chip B remained for further analysis.

3.3.3.1 Differentially regulated genes in the CD dataset

Differentially regulated genes in the CD dataset were determined using the limma R package with a four-factor, one level design. Limma analysis was carried out separately for chips A and B. P-values were adjusted for multiple testing using the false discovery rate (Benjamini & Hochberg method). Six possible unique comparisons (Contrasts 1 to 6) were possible using four factors (Table 3.13). In contrast to the UC dataset, much fewer differentially regulated genes were detected between inflamed and non-inflamed disease samples.

Table 3.13 Number of significantly differentially regulated genes in the CD dataset including results combined from chip A and B analysis*

Regulation	Contrast					
	1	2	3	4	5	6
	CD_NI-NC	CD_I- CD_NI	CD_I-NC	CD_IS- CD_NI	CD_IS- NC	CD_I- CD_IS
Down-regulated	3575 (563)	96 (25)	3924 (719)	30 (13)	675 (146)	1065 (175)
Not regulated	7580 (1547)	11247 (2142)	6979 (1364)	10621 (2090)	10707 (2017)	10194 (1956)
Up-regulated	320 (73)	132 (16)	572 (100)	824 (80)	93 (20)	216 (52)

* $p_{\text{adj}} < 0.05$, false discovery rate adjustment applied; unknown gene probesets in parentheses

Differential expression between matched samples of inflamed/non-inflamed regions of the same patient were analysed using paired significance testing, as previously done for the UC patient samples. In stark contrast to the paired UC dataset, the paired CD dataset revealed no significantly differentially expressed genes after a multiple testing correction was applied. Upon omitting the multiple testing correction, about 300 probesets were significantly regulated from raw paired t-test p-values, and none were significantly regulated from the raw paired Wilcoxon test p-values (Table 3.14). The overlap between significantly regulated probesets derived from limma and raw paired t-test was also much lower in the matched CD dataset compared to the matched UC dataset (compare Figure 3.17 and Figure 3.19). The dramatic reduction in significantly regulated probesets of the matched CD samples is likely influenced by the reduced power of the small sample size between the CD and UC datasets (4 matched CD pairs compared to 9 matched UC samples). Additionally, the matched CD pairs involved more comparisons being taken from different regions (Figure 3.20), which may result in additional variability.

Table 3.14 Number of significantly differentially regulated genes in paired inflamed/non-inflamed CD samples (combined results from chip A and B)

Regulation	CD_I-CD_NI	
	Paired t-test*	Paired Wilcoxon test*
Down-regulated	204	0
Not regulated	10957	11475
Up-regulated	314	0

* $p_{\text{raw}} < 0.05$

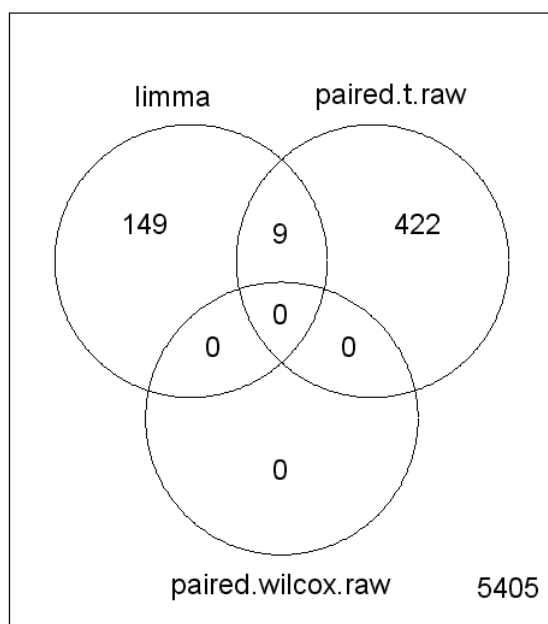


Figure 3.19 Venn diagram of differentially regulated probesets found using different statistical methods (limma, paired t-test, paired Wilcoxon test) for the CD dataset.

Paired statistical tests identify additional probesets not detected by inter-individual statistical analysis (limma), but the overlap between limma and the paired t-test is much lower than in the matched UC dataset.

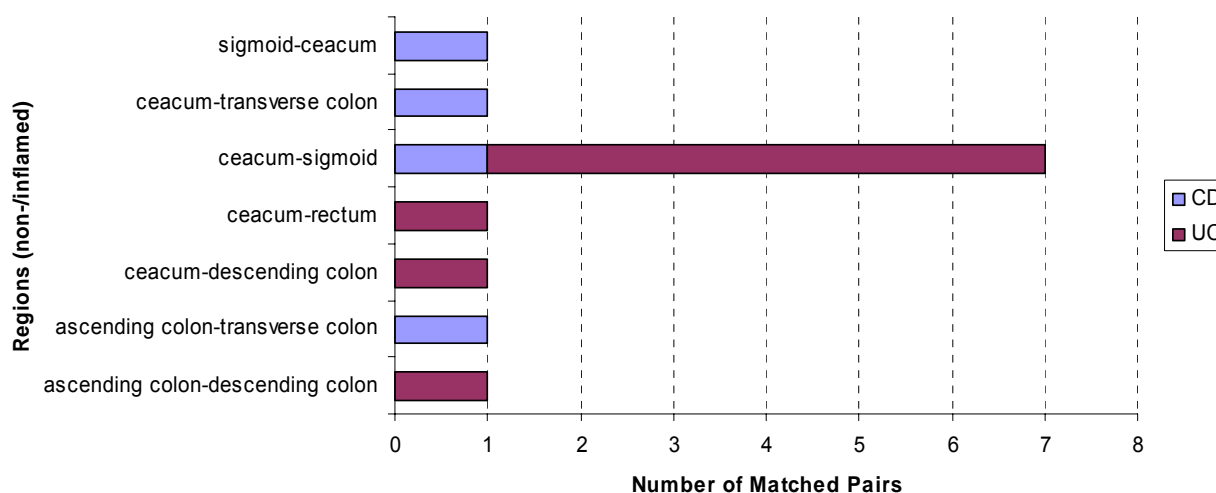


Figure 3.20 Regions of lower bowel that were involved in the matched samples of CD and UC datasets.

The paired statistical tests were carried out between non-inflamed and inflamed regions of the same patient. The CD dataset involved 4 matched pairs, while 9 matched pairs were used in the UC dataset.

Top differentially expressed genes are shown in Table 3.15 (CD_NI-NC), Table 3.16 (CD_I-CD_NI), Table 3.17 (CD_I-NC), Table 3.18 (CD_IS-CD_NI), Table 3.19 (CD_IS-NC) and Table 3.20 (CD_I-CD_IS). Inspection of the top genes in these tables showed that microarrays detected mRNA expression of genes involved in intestinal function, inflammation and immune response. For example, differential expression of gut endocrine hormones (*PYY*, *GCG*) and intestinal enzymes (*ALDOB*, *CA1*, *CES2*) was observed in various conditions. In comparisons involving inflamed CD samples and non-inflamed samples, augmented immune response was evidenced by the down-regulation of the pro-inflammatory antagonist *IL1R2* and up-regulation of

various immunoglobulin genes (*IGHM*, *IGKC*, *IGLJ3*, *IGKV1-5*, *IGH@*, *IGHD*, *IGHA1*, *IGLC2*, *IGLV2-14*), pro-inflammatory cytokines (*CASP1*), antimicrobial peptides (*DEFA5*, *DEFA6*) and genes previously shown in the literature to be expressed in inflamed colon (*S100A8*, *OLFM4*, *REG1A*, *REG3A*) (Beaven and Abreu, 2004; Ogawa et al., 2003; Shinozaki et al., 2001).

Table 3.15 Non-inflamed CD and healthy normal controls (Contrast 1): Top 50 differentially regulated genes from limma analysis

A. Up-regulated probesets from comparison of CD_NI...NC			
Probe id	Log2Ratio	Gene symbol	Gene name
214777_at	0.755	---	Anti-HIV-1 gp120 V3 loop antibody DO142-10 light chain variable region
211643_x_at	0.645	---	Anti-rabies virus immunoglobulin rearranged kappa chain V-region
211645_x_at	0.595	---	Cationic anti-DNA autoantibody
211650_x_at	0.510	---	IgM VDJ-region
217281_x_at	0.498	---	IgM rheumatoid factor RF-TT9, variable heavy chain
211908_x_at	0.491	---	IgM VDJ-region
215176_x_at	0.487	---	HRV Fab 027-VL
211637_x_at	0.474	---	Rearranged Ig mu-chain (V4-59/DIR1-D5'-D21-9/JH4b)
202003_s_at	0.607	ACAA2	acetyl-Coenzyme A acyltransferase 2 (mitochondrial 3-oxoacyl-Coenzyme A thiolase)
217238_s_at	0.844	ALDOB	aldolase B, fructose-bisphosphate
202888_s_at	0.952	ANPEP	alanyl (membrane) aminopeptidase (aminopeptidase N, aminopeptidase M, microsomal aminopeptidase, CD13, p150)
202357_s_at	0.475	BF	B-factor, properdin
229070_at	0.482	C6orf105	chromosome 6 open reading frame 105
211366_x_at	0.548	CASP1	caspase 1, apoptosis-related cysteine protease (interleukin 1, beta, convertase)
210133_at	1.127	CCL11	chemokine (C-C motif) ligand 11
206407_s_at	0.626	CCL13	chemokine (C-C motif) ligand 13
214038_at	0.608	CCL8	chemokine (C-C motif) ligand 8
206207_at	0.880	CLC	Charcot-Leyden crystal protein
210107_at	0.506	CLCA1	chloride channel, calcium activated, family member 1
207529_at	1.807	DEFA5	defensin, alpha 5, Paneth cell-specific
207814_at	1.005	DEFA6	defensin, alpha 6, Paneth cell-specific
225458_at	1.989	DKFZP564I1171	DKFZP564I1171 protein
208250_s_at	1.229	DMBT1	deleted in malignant brain tumors 1
219017_at	0.679	ETNK1	ethanolamine kinase 1
227194_at	0.540	FAM3B	family with sequence similarity 3, member B
202269_x_at	0.481	GBP1	guanylate binding protein 1, interferon-inducible, 67kDa
227614_at	0.708	HKDC1	hexokinase domain containing 1
202411_at	0.510	IFIT2	interferon, alpha-inducible protein 27
201601_x_at	0.549	IFITM1	interferon induced transmembrane protein 1 (9-27)
202718_at	0.842	IGFBP2	insulin-like growth factor binding protein 2, 36kDa
214916_x_at	0.639	IGH@	immunoglobulin heavy locus
216557_x_at	0.621	IGHA1	immunoglobulin heavy constant alpha 1
211644_x_at	0.593	IGKC	immunoglobulin kappa constant
214768_x_at	0.583	IGKV1-5	immunoglobulin kappa variable 1-5
217235_x_at	0.569	IGLC2	Immunoglobulin lambda variable 3-21
217148_x_at	0.532	IGLV2-14	immunoglobulin lambda variable 2-14
212531_at	1.022	LCN2	lipocalin 2 (oncogene 24p3)
229860_x_at	0.807	LOC401115	hypothetical gene supported by BC038466; BC062790
200632_s_at	0.577	NDRG1	N-myc downstream regulated gene 1
223217_s_at	0.639	NFKBIZ	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, zeta
210519_s_at	0.660	NQO1	NAD(P)H dehydrogenase, quinone 1
212768_s_at	1.714	OLFM4	olfactomedin 4
229230_at	0.858	OSTalpha	organic solute transporter alpha
230830_at	0.640	OSTbeta	organic solute transporter beta
226147_s_at	0.524	PIGR	polymeric immunoglobulin receptor
243669_s_at	0.946	PRAP1	proline-rich acidic protein 1
204279_at	0.552	PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional protease 2)
209752_at	1.719	REG1A	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)
205815_at	1.328	REG3A	regenerating islet-derived 3 alpha
202917_s_at	0.911	S100A8	S100 calcium binding protein A8 (calgranulin A)
201061_s_at	0.604	STOM	stomatin

B. Down-regulated probesets from comparison of CD_NI...NC

200729_s_at	-0.347	ACTR2	ARP2 actin-related protein 2 homolog (yeast)
200612_s_at	-0.306	AP2B1	adaptor-related protein complex 2, beta 1 subunit
200602_at	-0.307	APP	amyloid beta (A4) precursor protein (protease nexin-II, Alzheimer disease)
200779_at	-0.370	ATF4	activating transcription factor 4 (tax-responsive enhancer element B67)
200078_s_at	-0.401	ATP6V0B	ATPase, H+ transporting, lysosomal 21kDa, V0 subunit c'
200041_s_at	-0.269	BAT1	HLA-B associated transcript 1
200777_s_at	-0.362	BZW1	basic leucine zipper and W2 domains 1
200614_at	-0.393	CLTC	clathrin, heavy polypeptide (Hc)
200621_at	-0.409	CSRP1	cysteine and glycine-rich protein 1
200694_s_at	-0.320	DDX24	DEAD (Asp-Glu-Ala-Asp) box polypeptide 24
200033_at	-0.555	DDX5	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5
200664_s_at	-0.643	DNAJB1	DnaJ (Hsp40) homolog, subfamily B, member 1
200762_at	-0.277	DPYSL2	dihydropyrimidinase-like 2
200606_at	-0.419	DSP	desmoplakin
200004_at	-0.396	EIF4G2	eukaryotic translation initiation factor 4 gamma, 2
1438_at	-0.290	EPHB3	EPH receptor B3
200709_at	-0.313	FKBP1A	FK506 binding protein 1A, 12kDa
200648_s_at	-0.617	GLUL	glutamate-ammonia ligase (glutamine synthase)
200708_at	-0.346	GOT2	glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)
200696_s_at	-0.380	GSN	gelsolin (amyloidosis, Finnish type)
200073_s_at	-0.299	HNRPD	heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA binding protein 1, 37kDa)
200775_s_at	-0.331	HNRPK	heterogeneous nuclear ribonucleoprotein K
200052_s_at	-0.302	ILF2	interleukin enhancer binding factor 2, 45kDa
200791_s_at	-0.398	IQGAP1	IQ motif containing GTPase activating protein 1
200048_s_at	-0.289	JTB	jumping translocation breakpoint
200698_at	-0.444	KDELR2	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 2
200673_at	-0.291	LAPTM4A	lysosomal-associated protein transmembrane 4 alpha
200713_s_at	-0.313	MAPRE1	microtubule-associated protein, RP/EB family, member 1
200644_at	-0.288	MARCKSL1	MARCKS-like 1
200768_s_at	-0.448	MAT2A	methionine adenosyltransferase II, alpha
200624_s_at	-0.277	MATR3	matrin 3
200797_s_at	-0.439	MCL1	myeloid cell leukemia sequence 1 (BCL2-related)
200027_at	-0.291	NARS	asparaginyl-tRNA synthetase
200790_at	-0.593	ODC1	ornithine decarboxylase 1
200604_s_at	-0.394	PRKAR1A	protein kinase, cAMP-dependent, regulatory, type I, alpha (tissue specific extinguisher 1)
200730_s_at	-0.580	PTP4A1	protein tyrosine phosphatase type IVA, member 1
200677_at	-0.419	PTTG1IP	pituitary tumor-transforming 1 interacting protein
200608_s_at	-0.273	RAD21	RAD21 homolog (S. pombe)

200749_at	-0.332	RAN	RAN, member RAS oncogene family
200060_s_at	-0.296	RNPS1	RNA binding protein S1, serine-rich domain
200631_s_at	-0.333	SET	SET translocation (myeloid leukemia-associated)
200754_x_at	-0.332	SFRS2	splicing factor, arginine/serine-rich 2
200718_s_at	-0.282	SKP1A	S-phase kinase-associated protein 1A (p19A)
200672_x_at	-0.286	SPTBN1	spectrin, beta, non-erythrocytic 1
200620_at	-0.301	TMEM59	transmembrane protein 59
200662_s_at	-0.373	TOMM20	translocase of outer mitochondrial membrane 20 homolog (yeast)
200668_s_at	-0.286	UBE2D3	ubiquitin-conjugating enzyme E2D 3 (UBC4/5 homolog, yeast)
200083_at	-0.297	USP22	ubiquitin specific protease 22
200611_s_at	-0.281	WDR1	WD repeat domain 1
200641_s_at	-0.575	YWHAZ	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide

Table 3.16 Inflamed CD and Non-inflamed CD (Contrast 2): Top 50 differentially regulated genes from limma analysis

A. Up-regulated probesets from comparison of CD _I ...CD _{NI}			
Probe id	Log2Ratio	Gene symbol	Gene name
214777_at	0.699	---	Anti-HIV-1 gp120 V3 loop antibody DO142-10 light chain variable region
211641_x_at	0.623	---	Immunoglobulin heavy chain VH3 (H11)
211637_x_at	0.615	---	Rearranged Ig mu-chain (V4-59/DIR1-D5'-D21-9/JH4b)
217227_x_at	0.541	---	Hepatitis B surface antigen antibody variable domain
217179_x_at	0.535	---	Ig rearranged lambda-chain gene V-JI2/I3-region
211645_x_at	0.494	---	Cationic anti-DNA autoantibody
217281_x_at	0.460	---	IgM rheumatoid factor RF-TT9, variable heavy chain
217258_x_at	0.448	---	IgG to Puumala virus G2, light chain variable region
202357_s_at	0.378	BF	B-factor, properdin
201641_at	0.414	BST2	bone marrow stromal cell antigen 2
209970_x_at	0.473	CASP1	caspase 1, apoptosis-related cysteine protease (interleukin 1, beta, convertase)
211980_at	0.373	COL4A1	collagen, type IV, alpha 1
203915_at	0.808	CXCL9	chemokine (C-X-C motif) ligand 9
207529_at	1.467	DEFA5	defensin, alpha 5, Paneth cell-specific
207814_at	1.027	DEFA6	defensin, alpha 6, Paneth cell-specific
219955_at	0.448	ECAT11	hypothetical protein FLJ10884
217294_s_at	0.391	ENO1	enolase 1, (alpha)
227194_at	0.640	FAM3B	family with sequence similarity 3, member B
202269_x_at	0.528	GBP1	guanylate binding protein 1, interferon-inducible, 67kDa
209276_s_at	0.430	GLRX	glutaredoxin (thioltransferase)
210982_s_at	0.651	HLA-DRA	major histocompatibility complex, class II, DR alpha
201601_x_at	0.687	IFITM1	interferon induced transmembrane protein 1 (9-27)
201315_x_at	0.501	IFITM2	interferon induced transmembrane protein 2 (1-8D)
212203_x_at	0.492	IFITM3	interferon induced transmembrane protein 3 (1-8U)
214916_x_at	0.470	IGH@	immunoglobulin heavy locus
216557_x_at	0.435	IGHA1	immunoglobulin heavy constant alpha 1
214973_x_at	0.750	IGHD	immunoglobulin heavy constant delta
211633_x_at	0.498	IGHG1	Immunoglobulin heavy constant gamma 1 (G1m marker)
209374_s_at	0.959	IGHM	immunoglobulin heavy constant mu
211644_x_at	0.492	IGKC	immunoglobulin kappa constant
214768_x_at	0.830	IGKV1-5	immunoglobulin kappa variable 1-5
217235_x_at	0.607	IGLC2	Immunoglobulin lambda variable 3-21
211798_x_at	0.422	IGLJ3	immunoglobulin lambda joining 3
217148_x_at	0.665	IGLV2-14	immunoglobulin lambda variable 2-14
210029_at	0.816	INDO	indoleamine-pyrrole 2,3 dioxygenase
217933_s_at	0.397	LAP3	leucine aminopeptidase 3
204580_at	0.938	MMP12	matrix metalloproteinase 12 (macrophage elastase)
218698_at	0.413	MMRP19	likely ortholog of mouse monocyte macrophage 19
210037_s_at	0.431	NOS2A	nitric oxide synthase 2A (inducible, hepatocytes)
204279_at	0.495	PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional protease 2)
201762_s_at	0.415	PSME2	proteasome (prosome, macropain) activator subunit 2 (PA28 beta)
212187_x_at	0.647	PTGDS	prostaglandin D2 synthase 21kDa (brain)
209752_at	1.584	REG1A	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)
205815_at	1.516	REG3A	regenerating islet-derived 3 alpha
201060_x_at	0.431	STOM	stomatin
211796_s_at	0.427	TRBV21-1	T cell receptor beta variable 21-1
205890_s_at	0.673	UBD	ubiquitin D
201649_at	0.498	UBE2L6	ubiquitin-conjugating enzyme E2L 6
200629_at	0.663	WARS	tryptophanyl-tRNA synthetase
B. Down-regulated probesets from comparison of CD _I ...CD _{NI}			
230269_at	-0.617	---	Transcribed locus, strongly similar to NP_001186.1 filensin; cytoskeletal protein, 115 KD [Homo sapiens]
220645_at	-0.550	C11orf33	chromosome 11 open reading frame 33
223194_s_at	-0.338	C6orf85	chromosome 6 open reading frame 85
229964_at	-0.268	C9orf152	chromosome 9 open reading frame 152
209668_x_at	-0.474	CES2	carboxylesterase 2 (intestine, liver)
223942_x_at	-0.517	CHST5	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 5
200884_at	-0.887	CKB	creatine kinase, brain
224998_at	-0.272	CKLFSF4	chemokine-like factor super family 4
225809_at	-0.429	DKFZP564O0823	DKFZP564O0823 protein
229254_at	-0.418	DKFZp761N1114	hypothetical protein DKFZp761N1114
227676_at	-0.327	FAM3D	family with sequence similarity 3, member D
225667_s_at	-0.332	FAM84A	family with sequence similarity 84, member A
203697_at	-0.300	FRZB	frizzled-related protein
207003_at	-0.831	GUCA2A	guanylate cyclase activator 2A (guanylin)
209844_at	-0.582	HOXB13	homeo box B13
211372_s_at	-0.466	IL1R2	interleukin 1 receptor, type II
212573_at	-0.405	KIAA0830	KIAA0830 protein
208450_at	-0.576	LGALS2	lectin, galactoside-binding, soluble, 2 (galectin 2)
225270_at	-0.293	NEO1	neogenin homolog 1 (chicken)
213369_at	-0.430	PCDH21	protocadherin 21
228507_at	-0.264	PDE3A	Phosphodiesterase 3A, cGMP-inhibited
230784_at	-0.864	PRAC	small nuclear protein PRAC
200844_s_at	-0.595	PRDX6	peroxiredoxin 6
201481_s_at	-0.323	PYGB	phosphorylase, glycogen; brain
232707_at	-0.546	RAXLX	RAX-like homeobox
213994_s_at	-0.294	SPON1	spondin 1, extracellular matrix protein
222571_at	-0.285	ST6GALNAC6	ST6 (alpha-N-acetylneuraminy-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 6
223103_at	-0.332	STARD10	START domain containing 10
232914_s_at	-0.336	SYTL2	synaptotagmin-like 2
227642_at	-0.425	TFCP2L1	Transcription factor CP2-like 1
228716_at	-0.294	THRB	thyroid hormone receptor, beta (erythroblastic leukemia viral (v-erb-a) oncogene homolog 2, avian)
224412_s_at	-0.935	TRPM6	transient receptor potential cation channel, subfamily M, member 6
228232_s_at	-0.392	VSIG2	V-set and immunoglobulin domain containing 2
203892_at	-0.488	WFDC2	WAP four-disulfide core domain 2

Table 3.17 Inflamed CD and healthy normal controls (Contrast 3): Top 50 differentially regulated genes from limma analysis

A. Up-regulated probesets from comparison of CD_I...NC			
Probe id	Log2Ratio	Gene symbol	Gene name
214777_at	1.454	---	Anti-HIV-1 gp120 V3 loop antibody DO142-10 light chain variable region
211641_x_at	1.093	---	Immunoglobulin heavy chain VH3 (H11)
211637_x_at	1.089	---	Rearranged Ig mu-chain (V4-59/DIR1-D5'-D21-9/JH4b)
211645_x_at	1.089	---	Cationic anti-DNA autoantibody
211643_x_at	1.035	---	Anti-rabies virus immunoglobulin rearranged kappa chain V-region
224342_x_at	0.959	---	Ig rearranged lambda-chain gene V-JI2/I3-region
217281_x_at	0.958	---	IgM rheumatoid factor RF-TT9, variable heavy chain
217179_x_at	0.949	---	Ig rearranged lambda-chain gene V-JI2/I3-region
217227_x_at	0.862	---	Hepatitis B surface antigen antibody variable domain
211650_x_at	0.859	---	IgM VDJ-region
204705_x_at	0.879	ALDOB	aldolase B, fructose-bisphosphate
209970_x_at	1.016	CASP1	caspase 1, apoptosis-related cysteine protease (interleukin 1, beta, convertase)
210133_at	1.634	CCL11	chemokine (C-C motif) ligand 11
206207_at	1.134	CLC	Charcot-Leyden crystal protein
203915_at	1.094	CXCL9	chemokine (C-X-C motif) ligand 9
207529_at	3.275	DEFA5	defensin, alpha 5, Paneth cell-specific
207814_at	2.033	DEFA6	defensin, alpha 6, Paneth cell-specific
225458_at	2.186	DKFZP5641171	DKFZP5641171 protein
208250_s_at	1.817	DMBT1	deleted in malignant brain tumors 1
219727_at	1.180	DUOX2	dual oxidase 2
222262_s_at	1.052	ETNK1	ethanolamine kinase 1
227194_at	1.180	FAM3B	family with sequence similarity 3, member B
202269_x_at	1.008	GBP1	guanylate binding protein 1, interferon-inducible, 67kDa
227614_at	0.894	HKDC1	hexokinase domain containing 1
201601_x_at	1.235	IFITM1	interferon induced transmembrane protein 1 (9-27)
202718_at	1.117	IGFBP2	insulin-like growth factor binding protein 2, 36kDa
211430_s_at	1.194	IGH@	immunoglobulin heavy locus
216557_x_at	1.056	IGHA1	immunoglobulin heavy constant alpha 1
214973_x_at	1.211	IGHD	immunoglobulin heavy constant delta
216542_x_at	0.860	IGHG1	immunoglobulin heavy constant gamma 1 (G1m marker)
209374_s_at	1.613	IGHM	immunoglobulin heavy constant mu
211644_x_at	1.084	IGKC	immunoglobulin kappa constant
214768_x_at	1.413	IGKV1-5	immunoglobulin kappa variable 1-5
217235_x_at	1.177	IGLC2	Immunoglobulin lambda variable 3-21
211798_x_at	0.874	IGLJ3	immunoglobulin lambda joining 3
217148_x_at	1.198	IGLV2-14	immunoglobulin lambda variable 2-14
210029_at	1.104	INDO	indoleamine-pyrrole 2,3 dioxygenase
212531_at	1.834	LCN2	lipocalin 2 (oncogene 24p3)
204580_at	1.143	MMP12	matrix metalloproteinase 12 (macrophage elastase)
223217_s_at	0.877	NFKBIZ	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, zeta
210519_s_at	0.907	NQO1	NAD(P)H dehydrogenase, quinone 1
212768_s_at	2.474	OLFM4	olfactomedin 4
229230_at	1.026	OStalpha	organic solute transporter alpha
204279_at	1.047	PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional protease 2)
209752_at	3.302	REG1A	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)
205815_at	2.844	REG3A	regenerating islet-derived 3 alpha
202917_s_at	1.491	S100A8	S100 calcium binding protein A8 (calgranulin A)
201061_s_at	1.019	STOM	stomatin
205890_s_at	0.991	UBD	ubiquitin D
200629_at	0.857	WARS	tryptophanyl-tRNA synthetase

B. Down-regulated probesets from comparison of CD_I...NC			
200965_s_at	-0.281	ABLIM1	actin binding LIM protein 1
200974_at	-0.503	ACTA2	actin, alpha 2, smooth muscle, aorta
200612_s_at	-0.429	AP2B1	adaptor-related protein complex 2, beta 1 subunit
200602_at	-0.322	APP	amyloid beta (A4) precursor protein (protease nexin-II, Alzheimer disease)
200950_at	-0.307	ARPC1A	actin related protein 2/3 complex, subunit 1A, 41kDa
200779_at	-0.275	ATF4	activating transcription factor 4 (tax-responsive enhancer element B67)
200078_s_at	-0.373	ATP6V0B	ATPase, H+ transporting, lysosomal 21kDa, V0 subunit c'
200921_s_at	-0.420	BTG1	B-cell translocation gene 1, anti-proliferative
200953_s_at	-0.494	CCND2	cyclin D2
200884_at	-1.690	CKB	creatine kinase, brain
200614_at	-0.422	CLTC	clathrin, heavy polypeptide (Hc)
200086_s_at	-0.282	COX4I1	cytochrome c oxidase subunit IV isoform 1
200621_at	-0.380	CSRP1	cysteine and glycine-rich protein 1
1007_s_at	-0.337	DDR1	discoidin domain receptor family, member 1
200694_s_at	-0.301	DDX24	DEAD (Asp-Glu-Ala-Asp) box polypeptide 24
200033_at	-0.371	DDX5	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5
200880_at	-0.389	DNAJA1	DnaJ (Hsp40) homolog, subfamily A, member 1
200666_s_at	-0.543	DNAJB1	DnaJ (Hsp40) homolog, subfamily B, member 1
200606_at	-0.477	DSP	desmoplakin
200004_at	-0.367	EIF4G2	eukaryotic translation initiation factor 4 gamma, 2
1438_at	-0.298	EPHB3	EPH receptor B3
200709_at	-0.352	FKBP1A	FK506 binding protein 1A, 12kDa
200648_s_at	-0.413	GLUL	glutamate-ammonia ligase (glutamine synthase)
200708_at	-0.364	GOT2	glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)
200678_x_at	-0.269	GRN	granulin
200696_s_at	-0.512	GSN	gelsolin (amyloidosis, Finnish type)
200791_s_at	-0.372	IQGAP1	IQ motif containing GTPase activating protein 1
200699_at	-0.415	KDELR2	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 2
200851_s_at	-0.294	KIAA0174	KIAA0174
200915_x_at	-0.376	KTN1	kinesin 1 (kinesin receptor)
200644_at	-0.380	MARCKSL1	MARCKS-like 1
200797_s_at	-0.355	MCL1	myeloid cell leukemia sequence 1 (BCL2-related)
200816_s_at	-0.286	PAFAH1B1	platelet-activating factor acetylhydrolase, isoform 1b, alpha subunit 45kDa
200845_s_at	-0.980	PRDX6	peroxiredoxin 6
200732_s_at	-0.725	PTP4A1	protein tyrosine phosphatase type IVA, member 1
200677_at	-0.569	PTTG1IP	pituitary tumor-transforming 1 interacting protein
200749_at	-0.283	RAN	RAN, member RAS oncogene family
200060_s_at	-0.269	RNPS1	RNA binding protein S1, serine-rich domain
200872_at	-0.580	S100A10	S100 calcium binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11))
200893_at	-0.390	SFRS10	splicing factor, arginine/serine-rich 10 (transformer 2 homolog, Drosophila)
200718_s_at	-0.305	SKP1A	S-phase kinase-associated protein 1A (p19A)

200672_x_at	-0.345	SPTBN1	spectrin, beta, non-erythrocytic 1
200911_s_at	-0.643	TACC1	transforming, acidic coiled-coil containing protein 1
200804_at	-0.336	TEGT	testis enhanced gene transcript (BAX inhibitor 1)
200620_at	-0.526	TMEM59	transmembrane protein 59
200662_s_at	-0.314	TOMM20	translocase of outer mitochondrial membrane 20 homolog (yeast)
200973_s_at	-0.525	TSPAN3	tetraspanin 3
200083_at	-0.326	USP22	ubiquitin specific protease 22
200931_s_at	-0.389	VCL	vinculin
200867_at	-0.304	ZNF313	zinc finger protein 313

Table 3.18 Inflamed CD treated with immunosuppressives and non-inflamed CD (Contrast 4): Top 50 differentially regulated genes from limma analysis

A. Up-regulated probesets from comparison of CD_IS...CD_NI			
Probe id	Log2Ratio	Gene symbol	Gene name
213921_at	1.625	---	transcript represented by NM_001048
200974_at	0.673	ACTA2	actin, alpha 2, smooth muscle, aorta
202274_at	0.634	ACTG2	actin, gamma 2, smooth muscle, enteric
228969_at	0.703	AGR2	anterior gradient 2 homolog (Xenopus laevis)
214953_s_at	0.960	APP	amyloid beta (A4) precursor protein (protease nexin-II, Alzheimer disease)
228241_at	0.537	BCMP11	breast cancer membrane protein 11
205950_s_at	0.883	CA1	carbonic anhydrase I
201946_s_at	0.593	CCT2	chaperonin containing TCP1, subunit 2 (beta)
201735_s_at	0.566	CLCN3	chloride channel 3
214598_at	0.846	CLDN8	claudin 8
230360_at	0.911	COLM	collomin
216607_s_at	0.770	CYP51A1	cytochrome P450, family 51, subfamily A, polypeptide 1
242372_s_at	0.582	DKFZp761N1114	hypothetical protein DKFZp761N1114
202345_s_at	0.548	FABP5	fatty acid binding protein 5 (psoriasis-associated)
201889_at	0.547	FAM3C	family with sequence similarity 3, member C
201798_s_at	0.655	FER1L3	fer-1-like 3, myoferlin (C. elegans)
201540_at	0.550	FHL1	four and a half LIM domains 1
211719_x_at	1.038	FN1	fibronectin 1
206422_at	1.211	GCG	glucagon
200648_s_at	0.528	GLUL	glutamate-ammonia ligase (glutamine synthase)
205042_at	0.688	GNE	glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase
202539_s_at	0.553	HMGCR	3-hydroxy-3-methylglutaryl-Coenzyme A reductase
209844_at	0.639	HOXB13	homeo box B13
236681_at	0.581	HOXD13	Homeo box D13
200806_s_at	0.561	HSPD1	heat shock 60kDa protein 1 (chaperonin)
205403_at	0.713	IL1R2	interleukin 1 receptor, type II
201627_s_at	0.841	INSIG1	insulin induced gene 1
221091_at	1.435	INSL5	insulin-like 5
212192_at	0.619	KCTD12	potassium channel tetramerisation domain containing 12
206043_s_at	0.546	KIAA0703	KIAA0703 gene product
213564_x_at	0.582	LDHB	lactate dehydrogenase B
231814_at	0.588	LOC219612	hypothetical gene supported by AK025404
226654_at	1.145	MUC12	mucin 12
201497_x_at	0.702	MYH11	myosin, heavy polypeptide 11, smooth muscle
218189_s_at	0.585	NANS	N-acetylneuraminic acid synthase (sialic acid synthase)
207217_s_at	0.640	NOX1	NADPH oxidase 1
200845_s_at	0.559	PRDX6	peroxiredoxin 6
207080_s_at	1.578	PYY	peptide YY
205158_at	0.569	RNASE4	ribonuclease, RNase A family, 4
209146_at	0.848	SC4MOL	sterol-C4-methyl oxidase-like
205464_at	0.898	SCNN1B	sodium channel, nonvoltage-gated 1, beta (Liddle syndrome)
201742_x_at	0.532	SFRS1	splicing factor, arginine/serine-rich 1 (splicing factor 2, alternate splicing factor)
205185_at	1.057	SPINK5	serine protease inhibitor, Kazal type 5
209436_at	0.754	SPON1	spondin 1, extracellular matrix protein
209218_at	0.783	SQLE	squalene epoxidase
225721_at	0.639	SYNP02	Synaptopodin 2
205547_s_at	0.724	TAGLN	transgelin
217979_at	0.558	TM4SF13	Tetraspanin 13
238846_at	0.661	TNFRSF11A	tumor necrosis factor receptor superfamily, member 11a, NFkB activator
201689_s_at	0.537	TPD52	tumor protein D52
B. Down-regulated probesets from comparison of CD_IS...CD_NI			
211908_x_at	-0.376	---	IgM VDJ-region
214836_x_at	-0.390	---	HRV Fab N8-VL
215176_x_at	-0.498	---	HRV Fab 027-VL
216365_x_at	-0.328	---	Clone bsmneg3-t5 nonfunctional immunoglobulin light chain (IGL) mRNA, partial sequence
234884_x_at	-0.522	---	Transcript represented by L21961
206407_s_at	-0.512	CCL13	chemokine (C-C motif) ligand 13
214038_at	-0.486	CCL8	chemokine (C-C motif) ligand 8
225457_s_at	-1.171	DKFZP56411171	DKFZP56411171 protein
212788_x_at	-0.344	FTL	ferritin, light polypeptide
217022_s_at	-0.394	IGHA1	immunoglobulin heavy constant alpha 1
211649_x_at	-0.326	IGHM	Immunoglobulin heavy constant mu
214677_x_at	-0.420	IGL@	immunoglobulin lambda locus
209138_x_at	-0.354	IGLC2	Immunoglobulin lambda variable 3-21
211798_x_at	-0.508	IGLJ3	immunoglobulin lambda joining 3
229860_x_at	-0.428	LOC401115	hypothetical gene supported by BC038466; BC062790
215946_x_at	-0.294	LOC91353	similar to omega protein
212226_s_at	-0.411	PPAP2B	phosphatidic acid phosphatase type 2B
243669_s_at	-0.683	PRAP1	proline-rich acidic protein 1

Table 3.19 Inflamed CD treated with immunosuppressives and healthy normal controls (Contrast 5): Top 50 differentially regulated genes from limma analysis

A. Up-regulated probesets from comparison of CD_IS...NC			
Probe id	Log2Ratio	Gene symbol	Gene name
235496_at	0.523	---	Clone DNA49141 LGLL338 (UNQ338) mRNA, complete cds
227452_at	0.313	---	Hypothetical gene supported by AK000477
206469_x_at	0.315	AKR7A3	aldo-keto reductase family 7, member A3 (aflatoxin aldehyde reductase)
222446_s_at	0.457	BACE2	beta-site APP-cleaving enzyme 2
229070_at	0.586	C6orf105	chromosome 6 open reading frame 105
210133_at	1.021	CCL11	chemokine (C-C motif) ligand 11
206576_s_at	0.594	CEACAM1	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)
203757_s_at	0.561	CEACAM6	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)
206207_at	0.578	CLC	Charcot-Leyden crystal protein
210107_at	0.630	CLCA1	chloride channel, calcium activated, family member 1
207529_at	1.107	DEFA5	defensin, alpha 5, Paneth cell-specific
222847_s_at	0.564	EGLN3	egl nine homolog 3 (C. elegans)
210950_s_at	0.467	FDFT1	farnesyl-diphosphate farnesyltransferase 1
218986_s_at	0.529	FLJ20035	hypothetical protein FLJ20035
233604_at	0.288	FLJ22763	hypothetical gene supported by AK026416
204875_s_at	0.454	GMSD	GDP-mannose 4,6-dehydratase
202934_at	0.362	HK2	hexokinase 2
227614_at	0.486	HKDC1	hexokinase domain containing 1
210619_s_at	0.335	HYAL1	hyaluronoglucosaminidase 1
202718_at	0.475	IGFBP2	insulin-like growth factor binding protein 2, 36kDa
209374_s_at	1.161	IGHM	immunoglobulin heavy constant mu
213564_x_at	0.332	LDHB	lactate dehydrogenase B
226702_at	0.333	LOC129607	hypothetical protein LOC129607
229860_x_at	0.379	LOC401115	hypothetical gene supported by BC038466; BC062790
223940_x_at	0.556	MALAT1	metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA)
231736_x_at	0.409	MGST1	microsomal glutathione S-transferase 1
218698_at	0.480	MMRP19	likely ortholog of mouse monocyte macrophage 19
211695_x_at	0.322	MUC1	mucin 1, transmembrane
200632_s_at	0.528	NDRG1	N-myc downstream regulated gene 1
218888_s_at	0.575	NETO2	neuropilin (NRP) and tolloid (TLL)-like 2
210519_s_at	0.732	NQO1	NAD(P)H dehydrogenase, quinone 1
205552_s_at	0.375	OAS1	2',5'-oligoadenylate synthetase 1, 40/46kDa
212768_s_at	2.225	OLFM4	olfactomedin 4
230830_at	0.592	OSTbeta	organic solute transporter beta
202619_s_at	0.343	PLOD2	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2
202483_s_at	0.335	RANBP1	RAN binding protein 1
218353_at	0.267	RGS5	regulator of G-protein signalling 5
218424_s_at	0.387	STEAP3	STEAP family member 3
205890_s_at	0.861	UBD	ubiquitin D
217717_s_at	0.514	YWHAH	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide
210996_s_at	0.295	YWHAE	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypeptide
B. Down-regulated probesets from comparison of CD_IS...NC			
201752_s_at	-0.475	ADD3	adducin 3 (gamma)
204288_s_at	-0.309	ARGBP2	Arg/Abi-interacting protein ArgBP2
200779_at	-0.349	ATF4	activating transcription factor 4 (tax-responsive enhancer element B67)
202985_s_at	-0.296	BAG5	BCL2-associated athanogene 5
200920_s_at	-0.322	BTG1	B-cell translocation gene 1, anti-proliferative
200953_s_at	-0.440	CCND2	cyclin D2
201743_at	-0.371	CD14	CD14 antigen
202910_s_at	-0.314	CD97	CD97 antigen
203953_s_at	-0.323	CLDN3	claudin 3
201428_at	-0.293	CLDN4	claudin 4
202224_at	-0.402	CRK	v-crk sarcoma virus CT10 oncogene homolog (avian)
201220_x_at	-0.343	CTBP2	C-terminal binding protein 2
202157_s_at	-0.311	CUGBP2	CUG triplet repeat, RNA binding protein 2
203917_at	-0.552	CXADR	coxsackie virus and adenovirus receptor
200666_s_at	-0.357	DNAJB1	DnaJ (Hsp40) homolog, subfamily B, member 1
201842_s_at	-0.348	EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1
202668_at	-0.498	EFNB2	ephrin-B2
203499_at	-0.452	EPHA2	EPH receptor A2
204131_s_at	-0.663	FOXO3A	forkhead box O3A
201738_at	-0.339	GC20	translation factor sui1 homolog
202794_at	-0.364	INPP1	inositol polyphosphate-1-phosphatase
202597_at	-0.282	IRF6	interferon regulatory factor 6
203752_s_at	-0.417	JUND	jun D proto-oncogene
203130_s_at	-0.317	KIF5C	kinesin family member 5C
203726_s_at	-0.540	LAMA3	laminin, alpha 3
202267_at	-0.314	LAMC2	laminin, gamma 2
201212_at	-0.307	LGMN	legumain
202822_at	-0.292	LPP	LIM domain containing preferred translocation partner in lipoma
202431_s_at	-0.351	MYC	v-myc myelocytomatosis viral oncogene homolog (avian)
202555_s_at	-0.349	MYLK	myosin, light polypeptide kinase
201865_x_at	-0.300	NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)
202425_x_at	-0.270	PPP3CA	protein phosphatase 3 (formerly 2B), catalytic subunit, alpha isoform (calcineurin A alpha)
200732_s_at	-0.392	PTP4A1	protein tyrosine phosphatase type IVA, member 1
201164_s_at	-0.340	PUM1	pumilio homolog 1 (Drosophila)
204020_at	-0.275	PURA	purine-rich element binding protein A
201785_at	-0.463	RNASE1	ribonuclease, RNase A family, 1 (pancreatic)
203843_at	-0.280	RPS6KA3	ribosomal protein S6 kinase, 90kDa, polypeptide 3
203408_s_at	-0.322	SATB1	special AT-rich sequence binding protein 1 (binds to nuclear matrix/scaffold-associating DNA's)
202071_at	-0.308	SDC4	syndecan 4 (amphiglycan, ryudocan)
202656_s_at	-0.295	SERTAD2	SERTA domain containing 2
203509_at	-0.454	SORL1	sorilin-related receptor, L(DLR class) A repeats-containing
204011_at	-0.388	SPRY2	sprouty homolog 2 (Drosophila)
202021_x_at	-0.496	SUI1	putative translation initiation factor
200911_s_at	-0.456	TACC1	transforming, acidic coiled-coil containing protein 1
201108_s_at	-0.662	THBS1	thrombospondin 1
201986_at	-0.351	THRAP1	thyroid hormone receptor associated protein 1
204094_s_at	-0.369	TSC2D2	TSC2 domain family, member 2
203892_at	-0.400	WFDC2	WAP four-disulfide core domain 2
201367_s_at	-0.357	ZFP36L2	zinc finger protein 36, C3H type-like 2

200868 s at

-0.301

ZNF313

zinc finger protein 313

**Table 3.20 Inflamed CD and inflamed CD treated with immunosuppressives (Contrast 6):
Top 50 differentially regulated genes from limma analysis**

A. Up-regulated probesets from comparison of CD_I...CD_IS			
Probe id	Log2Ratio	Gene symbol	Gene name
234884_x_at	0.690	---	Transcript represented by L21961
214777_at	1.064	---	Anti-HIV-1 gp120 V3 loop antibody DO142-10 light chain variable region
211643_x_at	0.878	---	Anti-rabies virus immunoglobulin rearranged kappa chain V-region
211645_x_at	0.823	---	Cationic anti-DNA autoantibody
216401_x_at	0.699	---	Clone plaque 3.4-112 anti-oxidized LDL immunoglobulin light chain variable region mRNA, partial cds
217227_x_at	0.913	---	Hepatitis B surface antigen antibody variable domain
215176_x_at	0.801	---	HRV Fab 027-VL
224342_x_at	0.870	---	Ig rearranged lambda-chain gene V-JI2/I3-region
217258_x_at	0.901	---	IgG to Puumala virus G2, light chain variable region
211635_x_at	0.749	---	IgM rheumatoid factor RF-TT1, variable heavy chain
217281_x_at	0.731	---	IgM rheumatoid factor RF-TT9, variable heavy chain
211650_x_at	0.792	---	IgM VDJ-region
211641_x_at	0.769	---	Immunoglobulin heavy chain VH3 (H11)
211637_x_at	1.000	---	Rearranged Ig mu-chain (V4-59/DIR1-D5'-D21-9/JH4b)
204705_x_at	0.717	ALDOB	aldolase B, fructose-bisphosphate
209970_x_at	0.706	CASP1	caspase 1, apoptosis-related cysteine protease (interleukin 1, beta, convertase)
201926_s_at	0.931	DAF	decay accelerating factor for complement (CD55, Cromer blood group system)
207529_at	2.167	DEFA5	defensin, alpha 5, Paneth cell-specific
207814_at	1.612	DEFA6	defensin, alpha 6, Paneth cell-specific
225457_s_at	1.242	DKFZP56411171	DKFZP56411171 protein
208250_s_at	1.626	DMBT1	deleted in malignant brain tumors 1
219017_at	0.789	ETNK1	ethanolamine kinase 1
227194_at	1.116	FAM3B	family with sequence similarity 3, member B
202269_x_at	0.705	GBP1	guanylate binding protein 1, interferon-inducible, 67kDa
201601_x_at	0.860	IFITM1	interferon induced transmembrane protein 1 (9-27)
214916_x_at	0.700	IGH@	immunoglobulin heavy locus
216557_x_at	0.889	IGHA1	immunoglobulin heavy constant alpha 1
214973_x_at	1.154	IGHD	immunoglobulin heavy constant delta
211633_x_at	0.811	IGHG1	Immunoglobulin heavy constant gamma 1 (G1m marker)
216491_x_at	1.030	IGHM	immunoglobulin heavy constant mu
211644_x_at	0.895	IGKC	immunoglobulin kappa constant
214768_x_at	0.898	IGKV1-5	immunoglobulin kappa variable 1-5
216207_x_at	0.699	IGKV1D-13	immunoglobulin kappa variable 1D-13
216560_x_at	0.789	IGLC1	Immunoglobulin lambda constant 1 (Mcg marker)
234764_x_at	1.101	IGLC2	Ig lambda chain V-region (VL-AIG)
211798_x_at	0.930	IGLJ3	immunoglobulin lambda joining 3
217148_x_at	0.834	IGLV2-14	immunoglobulin lambda variable 2-14
217480_x_at	0.713	LOC339562	similar to Ig kappa chain
217378_x_at	0.688	LOC391427	similar to Ig kappa chain precursor V region (orphan V108) - human (fragment)
204580_at	0.991	MMP12	matrix metalloproteinase 12 (macrophage elastase)
229230_at	0.761	OSTalpha	organic solute transporter alpha
204279_at	0.674	PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional protease 2)
209752_at	2.963	REG1A	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)
205815_at	2.075	REG3A	regenerating islet-derived 3 alpha
201061_s_at	0.736	STOM	stomatin
221253_s_at	0.664	TXNDC5	thioredoxin domain containing 5

B. Down-regulated probesets from comparison of CD_I...CD_IS

201128_s_at	-0.309	ACLY	ATP citrate lyase
201034_at	-0.502	ADD3	adducin 3 (gamma)
200612_s_at	-0.615	AP2B1	adaptor-related protein complex 2, beta 1 subunit
200602_at	-0.472	APP	amyloid beta (A4) precursor protein (protease nexin-II, Alzheimer disease)
201176_s_at	-0.320	ARCN1	archain 1
201097_s_at	-0.477	ARF4	ADP-ribosylation factor 4
200950_at	-0.404	ARPC1A	actin related protein 2/3 complex, subunit 1A, 41kDa
200818_at	-0.311	ATP5O	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit (oligomycin sensitivity conferring protein)
200041_s_at	-0.293	BAT1	HLA-B associated transcript 1
201032_at	-0.489	BLCAP	bladder cancer associated protein
201005_at	-0.356	CD9	CD9 antigen (p24)
200999_s_at	-0.484	CKAP4	cytoskeleton-associated protein 4
200884_at	-1.497	CKB	creatine kinase, brain
200614_at	-0.367	CLTC	clathrin, heavy polypeptide (Hc)
201116_s_at	-0.308	CPE	carboxypeptidase E
200621_at	-0.283	CSRP1	cysteine and glycine-rich protein 1
200606_at	-0.505	DSP	desmoplakin
201016_at	-0.350	EIF1AX	eukaryotic translation initiation factor 1A, X-linked
200004_at	-0.341	EIF4G2	eukaryotic translation initiation factor 4 gamma, 2
200043_at	-0.292	ERH	enhancer of rudimentary homolog (Drosophila)
200709_at	-0.361	FKBP1A	FK506 binding protein 1A, 12kDa
200979_at	-0.343	FLJ16518	Mitogen-activated protein kinase kinase 15
201103_x_at	-0.284	FLJ20719	hypothetical protein FLJ20719
200648_s_at	-0.324	GLUL	glutamate-ammonia ligase (glutamine synthase)
200708_at	-0.388	GOT2	glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)
200696_s_at	-0.470	GSN	gelsolin (amyloidosis, Finnish type)
201209_at	-0.281	HDAC1	histone deacetylase 1
200955_at	-0.299	IMMT	inner membrane protein, mitochondrial (mitofilin)
200698_at	-0.388	KDELRL2	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 2
200915_x_at	-0.394	KTN1	kinesin 1 (kinesin receptor)
201030_x_at	-0.296	LDHB	lactate dehydrogenase B
200644_at	-0.321	MARCKSL1	MARCKS-like 1
201153_s_at	-0.273	MBNL1	muscleblind-like (Drosophila)
200978_at	-0.310	MDH1	malate dehydrogenase 1, NAD (soluble)
201227_s_at	-0.292	NDUFB8	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 8, 19kDa
200658_s_at	-0.265	PHB	prohibitin
200845_s_at	-1.039	PRDX6	peroxiredoxin 6
200677_at	-0.411	PTTG1IP	pituitary tumor-transforming 1 interacting protein
201047_x_at	-0.273	RAB6A	RAB6A, member RAS oncogene family
200872_at	-0.379	S100A10	S100 calcium binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11))
200893_at	-0.322	SFRS10	splicing factor, arginine/serine-rich 10 (transformer 2 homolog, Drosophila)
200718_s_at	-0.266	SKP1A	S-phase kinase-associated protein 1A (p19A)
200030_s_at	-0.290	SLC25A3	solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 3

200672_x_at	-0.293	SPTBN1	spectrin, beta, non-erythrocytic 1
201023_at	-0.269	TAF7	TAF7 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 55kDa
200620_at	-0.437	TMEM59	transmembrane protein 59
201175_at	-0.266	TMX2	thioredoxin-related transmembrane protein 2
201002_s_at	-0.277	UBE2V1	ubiquitin-conjugating enzyme E2 variant 1
200931_s_at	-0.466	VCL	vinculin
200609_s_at	-0.297	WDR1	WD repeat domain 1

3.3.3.2 Functional grouping of differentially regulated transcripts for CD dataset

As previously described for the UC dataset, the Fisher test for a 2X2 contingency table was used to check for enrichment of differentially regulated genes in GO classification terms (Table 3.21). A total of 11 268 probesets (from chips A and B) and 18 824 GO terms were input into Fisher test analysis, resulting in 4614 GO terms being matched to the probesets.

Table 3.21 Frequency of significantly enriched GO terms* in the CD dataset (chips A and B)

Contrast	Without Immunosuppressives			Involving Immunosuppressives		
	1	2	3	4	5	6
Root GO term	CD_NI-NC	CD_I- CD_NI	CD_I-NC	CD_IS- CD_NI	CD_IS- NC	CD_I- CD_IS
Biological process	239 (47) [†]	135 (55)	248 (51)	199 (79)	168 (70)	198 (34)
Molecular function	153 (27)	65 (29)	178 (36)	158 (48)	69 (29)	139 (12)
Cellular component	76 (10)	17 (8)	57 (7)	73 (22)	24 (4)	54 (7)

* $p < 0.05$, no correction applied; [†] numbers in parentheses indicate number of GO terms unique for that comparison

A sampling of the overall GO profile of the three comparisons without immunosuppressive-treated samples (first three contrasts in Table 3.21) indicated that immune-related processes (defense response, immune response, antigen binding, apoptotic protease activator activity, caspase activator activity), signaling (signal transducer activity, cell-cell signaling), detoxification (xenobiotic metabolism), and transport (calcium ion transport) were represented by differentially regulated genes in all three comparisons.

Considering all GO terms significant in both CD to normal control comparisons (CD_NI-NC and CD_I-NC), cell junction was extremely well represented, including 75% of all 'cellular component' GO terms relating to cell junction from all present genes: intercellular junction, adherens junction, cell-cell adherens junction, tight junction, cell-substrate adherens junction, cell-matrix junction, apical junction complex, zonula adherens. To a lesser extent, terms relating to apoptosis (regulation of apoptosis, anti-apoptosis, negative regulation of apoptosis) and transport (calcium-transporting ATPase activity, ion transport, phosphatidylinositol transporter activity, channel or pore class transporter activity) were also identified.

Analysis of GO terms that were significant between non-inflamed CD and normal controls only (ie. the effect of disease without inflammation effects) revealed a significant number of differentially regulated genes related to RNA (RNA binding, helicase activity, processing, metabolism and splicing), localization to diverse cellular locations (for example, nuclear membrane, mitochondrial outer membrane, endosome, coated vesicle), and transport (intracellular, protein, hexose, and cation transport), to name but a few select GO groups. Among the more specific terms revealed by GO analysis, Notch signaling pathway, low-density lipoprotein receptor activity, peptide hormone processing, regulation of cell migration, and regulation of neurotransmitter levels were identified.

The comparison of inflamed CD to non-inflamed CD has the potential to reveal inflammation effects specifically within the context of CD. GO analysis yielded terms with implications in the inflammatory/immune response (prostaglandin-D synthase activity, cytokine activity), lipid metabolism (fatty acid biosynthesis, lipid biosynthesis), protein degradation (proteasome complex (sensu Eukaryota), proteolysis and peptidolysis), apoptosis (regulation of I-kappaB kinase/NF-kappaB cascade, caspase activity) and reactive oxygen species (response to oxidative stress).

A wide variety of significant GO terms were identified from the comparison between normal controls and inflamed CD. As expected, many inflammatory or immune-related terms were identified, for example (coagulation, immune cell activation, MHC class I receptor activity). Other significant terms, which were not directly related to inflammation or immune response, included terms related to wound healing (vasculature development, growth factor activity, platelet activation, homophilic cell adhesion), binding (cation binding, growth factor binding, ligand-dependent nuclear receptor activity, enzyme linked receptor protein signaling pathway), and other terms, such as reproduction, transcription regulator activity, and regulation of cell activation, to name but a few.

It is possible to observe the effect of immunosuppressive treatment in the comparisons only involving CD samples treated with immunosuppressives (Table 3.21). Uniquely significant GO terms in comparisons between CD_IS and CD_NI samples include steroid metabolism, ribonucleotide metabolism, purine nucleotide metabolism, purine ribonucleotide metabolism, ribonucleoside monophosphate metabolism, and purine ribonucleoside monophosphate metabolism—all of these terms play a role in the metabolism of steroids or the response to

azathioprine treatment. Notably absent from the above contrast involving CD_IS and CD_NI was an abundance of terms related to immune response or inflammation. The comparison between CD_IS and normal controls shows that GO terms related to nucleotide metabolism (regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism), apoptosis (induction of apoptosis, positive regulation of apoptosis), immune response (chemokine receptor binding, leukocyte adhesion) and cell cycle (regulation of cell cycle, G0 to G1 transition) were noted. In the last of the three contrasts involving samples treated with immunosuppressives (CD_I vs. CD_IS), unique GO terms for this contrast included terms relating to immune response (regulation of T-cell differentiation, cell-mediated immune response, antibacterial humoral response, natural killer cell activation), detoxification (carboxypeptidase activity, metalloproteinase activity, aromatic compound metabolism), and other groups (negative regulation of Wnt receptor signaling pathway, hormone binding, intercellular junction assembly and/or maintenance), just to name a few.

Analysis of the GO terms represented by differentially regulated genes between inflamed/non-inflamed CD and UC or normal controls shows that biological processes and molecular functions (such as inflammation/immune response and apoptosis) were observed, as expected. Additionally, known processes related to immunosuppressive treatment in CD were noted. Though many of the identified GO groups (and genes) could have been the basis of further investigation, only two subject areas were chosen for further study, based on the fact that they were not directly related to immune function. The themes chosen for further real-time analysis include WNT signaling pathway and epithelial barrier function.

3.3.4 Functional themes supported by microarray analysis in IBD

The results from cDNA (Costello et al., 2005) and Affymetrix microarrays used in this thesis cannot be directly combined in terms of expression values; however, the results should be complementary in terms of the broad functional groups represented by the differentially expressed genes from both platforms. The findings of all 11 group comparisons from all microarray datasets are summarized in Table 3.22. The differentially regulated genes in each of the group comparisons were categorized into three major functional groups: immune and inflammatory response; oncogenesis, cell proliferation and growth; and structure and permeability.

Table 3.22 Log₂ ratios of differentially regulated genes from comparisons in the cDNA and Affymetrix microarray cohorts*.

Symbol		Name		cDNA arrays		Affymetrix arrays								
						CD						UC		
						1	2	3	4	5	6	1	2	3
CD_-I-NC	UC_-I-NC	CD_-NI-NC	CD_-I-CD_-NI	CD_-I-NC	CD_-IS-CD_-NI	CD_-IS-NC	CD_-I-CD_-IS	UC_-NI-NC	UC_-I-UC_-NI	UC_-I-NC				

Immune and Inflammatory Response

---	Transcribed sequence with strong similarity to protein pir:I38067 (H.sapiens) I38067 nitric-oxide synthase	6.34	6.89											
ADORA3	adenosine A3 receptor	-0.57												
ALDH2	aldehyde dehydrogenase 2 family (mitochondrial)		0.36								0.33	-0.31		
ALOX5	arachidonate 5-lipoxygenase	-0.66										0.54	0.31	
ALOX5AP	arachidonate 5-lipoxygenase-activating protein	-0.60												
APOL2	apolipoprotein L, 2				0.36								0.32	0.45
ASS	argininosuccinate synthetase	1.50	1.54		0.67						0.33	0.64	0.97	
BIRC4	baculoviral IAP repeat-containing 4	-0.50	-0.44	-0.31		-0.34			-0.34					
BST1	bone marrow stromal cell antigen 1		-0.42											
BST2	bone marrow stromal cell antigen 2				0.32	0.41	0.73			0.56		0.59	0.61	
C2	complement component 2				0.28	0.29	0.57			0.39		0.55	0.59	
C5	complement component 5	-0.51	-0.36											
CASP1	caspase 1, apoptosis-related cysteine protease (interleukin 1, beta, convertase)				0.54	0.47	1.02			0.71	0.53	0.78	1.32	
CCL11	chemokine (C-C motif) ligand 11				1.13		1.63		1.02		1.17	1.17	2.33	
CCL13	chemokine (C-C motif) ligand 13				0.63		0.65	-0.51		0.54	0.46		0.49	
CCL28	chemokine (C-C motif) ligand 28	-0.54												
CCL8	chemokine (C-C motif) ligand 8				0.61		0.43	-0.49			0.67		0.56	
CD164	CD164 antigen, sialomucin				-0.50		-0.54			-0.38			-0.35	
CD58	CD58 antigen, (lymphocyte function-associated antigen 3)				-0.42		-0.42	0.30		-0.30	-0.40		-0.28	
CD79A	CD79A antigen (immunoglobulin-associated alpha)						0.34			0.32		0.98	0.88	
CD9	CD9 antigen (p24)				-0.34		-0.37			-0.36		-0.50	-0.49	
CD97	CD97 antigen				-0.42		-0.41		-0.31		-0.48	0.30		
CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	-0.55												
CEACAM1§	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein) (Affymetrix probe: 211883_x_at)		1.01				0.43			0.47	0.66		0.60	
CEACAM1§	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein) (Affymetrix probe: 209498_at)		1.01								0.77	-0.80		
CFB	complement factor B				0.48	0.38	0.85			0.50	0.48	0.95	1.44	
CFI	complement factor I				0.41		0.68			0.59	0.44	0.50	0.94	
CHUK	conserved helix-loop-helix ubiquitous kinase	0.59												
CLC	Charcot-Leyden crystal protein				0.88		1.13			0.58	0.56	0.96	0.95	

Symbol		Name		cDNA arrays		Affymetrix arrays								
						CD						UC		
						1	2	3	4	5	6	1	2	3
CD_I-NC	UC_I-NC	CD_NI-NC	CD_I-CD_NI	CD_I-NC	CD_IS-CD_NI	CD_IS-NC	CD_I-CD_IS	UC_NI-NC	UC_I-UC_NI	UC_I-NC				
CRIP1	cysteine-rich protein 1 (intestinal)			-1.31		-1.56			-1.37		-1.65	0.57	-1.09	
CTSC	cathepsin C				0.30	0.54					0.32	0.34	0.67	
CTSE	cathepsin E			-0.78		-0.77	0.51	-0.27	-0.50		-0.74	0.94		
CXCL12	chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)			-0.36		-0.31					-0.64		-0.38	
CXCL2	chemokine (C-X-C motif) ligand 2					0.47				0.43		1.64	1.84	
CYLD††	cylindromatosis (turban tumor syndrome)	-0.61										0.39		
DAF	decay accelerating factor for complement (CD55, Cromer blood group system)	0.87	1.66			0.61				0.93		1.51	1.78	
DEFA5	defensin, alpha 5, Paneth cell-specific			1.81	1.47	3.27		1.11	2.17	2.22			2.42	
DEFA6	defensin, alpha 6, Paneth cell-specific			1.01	1.03	2.03			1.61	1.34			1.51	
DMBT1†	deleted in malignant brain tumors 1			1.23		1.82			1.63	1.64	1.35		2.99	
ECGF1	endothelial cell growth factor 1 (platelet-derived)				0.28	0.51			0.37		0.39	0.56		
EGLN3	egl nine homolog 3 (C. elegans) (Affymetrix probe: 222847_s_at)		0.51	0.41		0.59		0.56						
EGLN3	egl nine homolog 3 (C. elegans) (Affymetrix probe: 219232_s_at)		0.51							0.29			0.45	
ELF4	E74-like factor 4 (ets domain transcription factor)			-0.36		-0.45						-0.28	-0.53	
F2RL1	coagulation factor II (thrombin) receptor-like 1			-0.65		-0.65						-0.54	-0.80	
FAM3C	family with sequence similarity 3, member C			-0.70		-0.77	0.55		-0.61	-1.07	0.74		-0.34	
FAS	Fas (TNF receptor superfamily, member 6)			-0.50		-0.40	0.34			-0.51			-0.29	
FCGR3A	Fc fragment of IgG, low affinity IIIa, receptor (CD16a)			0.29		0.45			0.38	0.30	0.64		0.94	
FCGR3B	Fc fragment of IgG, low affinity IIIb, receptor (CD16b)			0.38		0.43					0.78		1.02	
FN1	fibronectin 1			-1.57		-1.58	1.04	-0.53	-1.05	-2.13	1.03		-1.10	
FTH1	ferritin, heavy polypeptide 1					-0.59			-0.73	-0.36	-0.29		-0.65	
GBP1	guanylate binding protein 1, interferon-inducible, 67kDa			0.48	0.53	1.01			0.70	0.47	0.93		1.40	
GBP2	guanylate binding protein 2, interferon-inducible				0.33	0.56			0.36		0.61		0.75	
GMDS	GDP-mannose 4,6-dehydratase			0.39	0.32	0.72		0.45		0.82			0.78	
GNAQ	Guanine nucleotide binding protein (G protein), q polypeptide			-0.46		-0.41							-0.30	
HAMP	hepcidin antimicrobial peptide	-0.59	-0.58											
HLA-A	major histocompatibility complex, class I, A	1.12		0.29		0.36								
HLA-B	major histocompatibility complex, class I, B	0.97	1.13			0.43			0.36		0.27		0.36	
HLA-DMB	major histocompatibility complex, class II, DM beta				0.35	0.51			0.41		0.53		0.49	
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	-0.56								-0.43	0.85		0.41	
HLA-DRA	major histocompatibility complex, class II, DR alpha				0.65	0.80				-0.37	1.14		0.78	
HLA-E	major histocompatibility complex, class I, E			0.32	0.33	0.65			0.48		0.51		0.77	
IER3	immediate early response 3	0.96	1.21											
IFITM1	interferon induced transmembrane protein 1 (9-27)		0.72	0.55	0.69	1.24			0.86	0.79	0.49		1.28	
IFITM2	interferon induced transmembrane protein 2 (1-8D)			0.33	0.50	0.83			0.59		0.87		1.11	
IFITM3	interferon induced transmembrane protein 3 (1-8U)			0.33	0.49	0.82			0.56		0.84		1.06	
IGH@	immunoglobulin heavy locus	2.46	3.32	0.64	0.47	1.11			0.70	0.47	0.60		1.07	
IGHA1	immunoglobulin heavy constant alpha 1			0.62	0.43	1.06			0.89		0.71		1.04	

Symbol		Name		cDNA arrays		Affymetrix arrays								
						CD						UC		
						1	2	3	4	5	6	1	2	3
CD_I-NC	UC_I-NC	CD_NI-NC	CD_I-CD_NI	CD_I-NC	CD_IS-CD_NI	CD_IS-NC	CD_I-CD_IS	UC_NI-NC	UC_I-UC_NI	UC_I-NC				
IGHD	immunoglobulin heavy constant delta			0.46	0.75	1.21			1.15		1.01	1.35		
IGHG1	Immunoglobulin heavy constant gamma 1 (G1m marker)	3.32	3.32		0.50	0.73			0.81		1.15	1.37		
IGHM	immunoglobulin heavy constant mu				0.96	1.61		1.16		0.90		1.12		
IGKC	immunoglobulin kappa constant		1.77	0.59	0.49	1.08			0.89		0.91	0.99		
IGKV1-5	immunoglobulin kappa variable 1-5			0.58	0.83	1.41			0.90	0.44	0.86	1.29		
IGL@	immunoglobulin lambda locus			0.34	0.54	-0.42			0.62		0.36	0.33		
IGLC2	Immunoglobulin lambda variable 3-21			0.57	0.61	1.18			0.88	0.39	0.65	1.04		
IGLL1	immunoglobulin lambda-like polypeptide 1		0.96											
IL1R1	interleukin 1 receptor, type I	0.40												
IL1R2	interleukin 1 receptor, type II			-0.96		-1.41	0.71		-1.16	-0.86	-0.70	-1.56		
IL2RG	interleukin 2 receptor, gamma (severe combined immunodeficiency)			-0.43		-0.66	0.39		-0.61	-0.34		-0.30		
ITGB2	integrin, beta 2 (antigen CD18 (p95), lymphocyte function-associated antigen 1; macrophage antigen 1 (mac-1) beta subunit)	1.40	1.20			0.47					0.58	0.55		
LEFTY1	left-right determination factor 1			-0.89		-1.20			-0.79	-1.04		-0.79		
LTB4R2	leukotriene B4 receptor 2	-0.67												
MASP1	mannan-binding lectin serine protease 1 (C4/C2 activating component of Ra-reactive factor)	-0.44												
MICA	MHC class I polypeptide-related sequence A	-0.49												
MYBL2	v-myb myeloblastosis viral oncogene homolog (avian)-like 2		0.43											
NMI	N-myc (and STAT) interactor				0.29	0.46			0.27		0.31	0.55		
NOS2A	nitric oxide synthase 2A (inducible, hepatocytes)		1.02		0.43	0.71			0.52		0.95	1.15		
OAS1	2',5'-oligoadenylate synthetase 1, 40/46kDa					0.29		0.38		0.68	-0.34	0.35		
P4HB	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide (protein disulfide isomerase-associated 1)		1.24											
PDGFA	platelet-derived growth factor alpha polypeptide			-0.61		-0.67		-0.74		-0.29		-0.44		
PLA2G2A	phospholipase A2, group IIA (platelets, synovial fluid)	0.78	1.24							-1.21	1.88			
PROS1	protein S (alpha)			-0.32		-0.50	0.36		-0.53	-0.58		-0.33		
PSMB10	proteasome (prosome, macropain) subunit, beta type, 10				0.31	0.47					0.31	0.50		
PSMB8	proteasome (prosome, macropain) subunit, beta type, 8 (large multifunctional protease 7)				0.36	0.59				0.29	0.51	0.80		
PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional protease 2)			0.55	0.50	1.05			0.67	0.54	0.92	1.45		
PTGDS	prostaglandin D2 synthase 21kDa (brain)				0.65	0.77				0.32	1.09	1.41		
PTGER4	prostaglandin E receptor 4 (subtype EP4)			-0.40		-0.47		-0.36		-0.49		-0.28		
PTN	pleiotrophin (heparin binding growth factor 8, neurite growth-promoting factor 1)					-0.28				-0.68		-0.61		
REG3A	regenerating islet-derived 3 alpha			1.33	1.52	2.84			2.07	1.67	0.98	2.65		
RNPEP	arginyl aminopeptidase (aminopeptidase B)	-0.45												
S100A8	S100 calcium binding protein A8 (calgranulin A)			0.91		1.49				0.78	1.92	2.70		
S100A9	S100 calcium binding protein A9 (calgranulin B)			0.34		0.56			0.44		0.80	1.07		

Symbol		Name		cDNA arrays		Affymetrix arrays								
						CD						UC		
						1	2	3	4	5	6	1	2	3
CD_I-NC	UC_I-NC	CD_NI-NC	CD_I-CD_NI	CD_I-NC	CD_IS-CD_NI	CD_IS-NC	CD_I-CD_IS	UC_NI-NC	UC_I-UC_NI	UC_I-NC				
SPINK5†	serine protease inhibitor, Kazal type 5			-0.99		-1.49	1.06		-1.55	-1.59	1.43			
SPRED2	sprouty-related, EVH1 domain containing 2			-0.47		-0.51		-0.28		-0.41	-0.30	-0.71		
ST6GAL1	ST6 beta-galactosamide alpha-2,6-sialyltransferase 1			-0.37		-0.42			-0.27	-0.70	0.75			
TFF1	trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in)		1.61	-0.80		-0.89				-1.07	1.71	0.65		
THBS1	thrombospondin 1			-0.75		-1.02		-0.66		-0.79		-0.70		
TNFRSF11A	tumor necrosis factor receptor superfamily, member 11a, NFKB activator			-0.33		-0.43	0.40		-0.50	-0.50		-0.32		
TNFRSF6B	tumor necrosis factor receptor superfamily, member 6b, decoy	0.90	1.23								0.98	1.09		
TRBV21-1	T cell receptor beta variable 21-1				0.43	0.41			0.52		0.80	0.75		
UBD	ubiquitin D				0.67	0.99		0.86			2.18	2.23		
VWF	von Willebrand factor					0.40					0.67	0.90		

Oncogenesis, Cell Proliferation and Growth

ADRA2A	adrenergic, alpha-2A-, receptor					0.31		0.55		0.51	1.36	-0.98	0.38
BHLHB3	basic helix-loop-helix domain containing, class B, 3			-0.49		-0.61			-0.35	-0.89	0.41	-0.48	
BST2	bone marrow stromal cell antigen 2			0.32	0.41	0.73				0.56	0.59	0.61	
BTG3	BTG family, member 3			-0.57		-0.70	0.36		-0.49	-0.88	0.35	-0.53	
CAPN1	calpain 1, (mu/l) large subunit		0.45										
CAPNS1	calpain, small subunit 1	0.31	0.44										
CCND2	cyclin D2			-0.56		-0.49		-0.44		-0.76		-0.55	
CCT2	chaperonin containing TCP1, subunit 2 (beta)			-0.38			0.43			-0.40		-0.31	
CD164	CD164 antigen, sialomucin			-0.50		-0.54			-0.38			-0.35	
CDK2AP1	CDK2-associated protein 1			-0.31			0.37		-0.27	-0.29	0.35		
CDK5RAP1	CDK5 regulatory subunit associated protein 1	-0.45											
CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	-0.55											
CENPE	centromere protein E, 312kDa		3.32										
CISH	cytokine inducible SH2-containing protein	-0.58											
CRIP1	cysteine-rich protein 1 (intestinal)			-1.31		-1.56			-1.37	-1.65	0.57	-1.09	
CSNK1D	casein kinase 1, delta	0.62		-0.28		-0.29						-0.39	
CYLD	cylindromatosis (turban tumor syndrome)	-0.61									0.39		
DAB2	disabled homolog 2, mitogen-responsive phosphoprotein (Drosophila)	-0.41	-0.28	-0.27		-0.36	0.44		-0.53	-0.51		-0.41	
DKC1	dyskeratosis congenita 1, dyskerin			-0.38		-0.28	0.37			-0.40		-0.31	
DMBT1†	deleted in malignant brain tumors 1			1.23		1.82				1.63	1.64	1.35	2.99
DNAJB5	DnaJ (Hsp40) homolog, subfamily B, member 5	0.71											
DST	dystonin			-0.41		-0.49	0.34		-0.42	-0.41		-0.38	
EIF4G2	eukaryotic translation initiation factor 4 gamma, 2			-0.40		-0.37	0.37		-0.34				
ELF4	E74-like factor 4 (ets domain transcription factor)			-0.36		-0.45					-0.28	-0.53	
FGFBP1	fibroblast growth factor binding protein 1			-0.34			0.49		-0.40				

Symbol		Name		cDNA arrays		Affymetrix arrays								
						CD						UC		
						1	2	3	4	5	6	1	2	3
CD_I-NC	UC_I-NC	CD_NI-NC	CD_I-CD_NI	CD_I-NC	CD_IS-CD_NI	CD_IS-NC	CD_I-CD_IS	UC_NI-NC	UC_I-UC_NI	UC_I-NC				
FGFR2	fibroblast growth factor receptor 2 (bacteria-expressed kinase, keratinocyte growth factor receptor, craniofacial dysostosis 1, Crouzon syndrome, Pfeiffer syndrome, Jackson-Weiss syndrome)			0.35		0.44					1.01	-1.23		
FGFR3	fibroblast growth factor receptor 3 (achondroplasia, thanatophoric dwarfism)			-0.27		-0.47			-0.27			-0.64	-0.70	
FHL1	four and a half LIM domains 1			-0.72		-0.58	0.55			-0.41	-0.57			
FTH1	ferritin, heavy polypeptide 1					-0.59				-0.73	-0.36	-0.29	-0.65	
GAS2	growth arrest-specific 2	-0.62	-0.54											
GCG	glucagon			-1.64		-2.06	1.21			-1.63	-2.63	0.74	-1.89	
GMNN	geminin, DNA replication inhibitor			-0.30		-0.32	0.34			-0.35		-0.37	-0.59	
GSPT1	G1 to S phase transition 1			-0.28			0.36				-0.41		-0.27	
HRASLS3	HRAS-like suppressor 3					0.37							0.44	
HSPA2	heat shock 70kDa protein 2			-0.33		-0.46				-0.47	-0.48		-0.35	
IFITM1	interferon induced transmembrane protein 1 (9-27)		0.72	0.55	0.69	1.24				0.86	0.79	0.49	1.28	
IGFBP2	insulin-like growth factor binding protein 2, 36kDa			0.84		1.12		0.47		0.64	1.82	-0.88	0.93	
IGFBP3	insulin-like growth factor binding protein 3			-0.66		-0.85				-0.54	-1.17		-0.95	
IL2RG	interleukin 2 receptor, gamma (severe combined immunodeficiency)			-0.43		-0.66	0.39			-0.61	-0.34		-0.30	
INSIG1	insulin induced gene 1			-0.51		-0.39	0.84			-0.72	-0.43			
KLF4	Kruppel-like factor 4 (gut)			-0.55		-0.42	0.38					-0.32	-0.51	
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog			-0.46		-0.55				-0.41	-0.46		-0.55	
LEFTY1	left-right determination factor 1			-0.89		-1.20				-0.79	-1.04		-0.79	
MACF1	microtubule-actin crosslinking factor 1			-0.32		-0.30					-0.42	0.34		
MYBL2	v-myb myeloblastosis viral oncogene homolog (avian)-like 2		0.43											
NDRG3	NDRG family member 3		-0.31											
NEDD9	neural precursor cell expressed, developmentally down-regulated 9	-0.71				-0.30				-0.40				
NME2	non-metastatic cells 2, protein (NM23B) expressed in	0.36												
PDGFA	platelet-derived growth factor alpha polypeptide			-0.61		-0.67				-0.74		-0.29	-0.44	
PICALM	phosphatidylinositol binding clathrin assembly protein	-0.55												
PIM2	pim-2 oncogene					0.45				0.49		0.68	0.86	
PLA2G2A	phospholipase A2, group IIA (platelets, synovial fluid)	0.78	1.24								-1.21	1.88		
PMP22	peripheral myelin protein 22			-0.58		-0.43			-0.54		-0.75	0.61		
PPP2CA	protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform	-0.57		-0.39		-0.28	0.28				-0.36			
PPP6C	protein phosphatase 6, catalytic subunit					-0.35				-0.33			-0.40	
PRDX1	peroxiredoxin 1					0.45					0.43		0.38	
PRKD3	protein kinase D3		0.75											
PTN	pleiotrophin (heparin binding growth factor 8, neurite growth-promoting factor 1)					-0.28					-0.68		-0.61	
PTP4A1	protein tyrosine phosphatase type IVA, member 1			-0.66		-0.73			-0.39		-0.33	-0.62	-0.95	
PYY	peptide YY			-1.72		-2.49	1.58			-2.35	-3.09	0.54	-2.56	
QSCN6	quiescin Q6			-0.28		-0.47				-0.38	-0.50	0.65		
RABGAP1	RAB GTPase activating protein 1			-0.46		-0.59	0.33			-0.45	-0.55		-0.45	

Symbol		Name		cDNA arrays		Affymetrix arrays								
						CD						UC		
						1	2	3	4	5	6	1	2	3
CD_I-NC	UC_I-NC	CD_NI-NC	CD_I-CD_NI	CD_I-NC	CD_IS-CD_NI	CD_IS-NC	CD_I-CD_IS	UC_NI-NC	UC_I-UC_NI	UC_I-NC				
RAD1	RAD1 homolog (<i>S. pombe</i>)	-0.41												
RASA1	RAS p21 protein activator (GTPase activating protein) 1	-0.48		-0.27		-0.28								
REG1A	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)			1.72	1.58	3.30			2.96	3.62		4.35		
REG3A	regenerating islet-derived 3 alpha			1.33	1.52	2.84			2.07	1.67	0.98	2.65		
ROCK1	Rho-associated, coiled-coil containing protein kinase 1	-0.57												
S100A11	S100 calcium binding protein A11 (calgizzarin)		0.64								0.75	0.48		
S100A6	S100 calcium binding protein A6 (calcyclin)		0.88											
S100P	S100 calcium binding protein P		1.38							-1.05	2.07	1.02		
SFN	stratifin			-0.72		-0.89		-0.51		-0.50	-0.37	-0.87		
SH3BP4	SH3-domain binding protein 4			-0.41		-0.42		-0.28		-0.39		-0.36		
STAG1	stromal antigen 1	-0.55	-0.39											
STAT1	signal transducer and activator of transcription 1, 91kDa				0.32	0.53			0.38		0.88	1.07		
TACC1	transforming, acidic coiled-coil containing protein 1			-0.61		-0.64		-0.46		-0.67		-0.46		
TFF1	trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in)		1.61	-0.80		-0.89				-1.07	1.71	0.65		
TIMP1	tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor)	1.23	2.04							-0.40	1.85	1.45		
TNFRSF11A	tumor necrosis factor receptor superfamily, member 11a, NFkB activator			-0.33		-0.43	0.40		-0.50	-0.50		-0.32		
TSSC1	tumor suppressing subtransferable candidate 1	-0.59												
TXN	thioredoxin			-0.37		-0.50			-0.48	-0.59		-0.40		
WARS	tryptophanyl-tRNA synthetase				0.65	0.79			0.54		1.03	1.09		
WIT-1	Wilms tumor associated protein	-0.54												

Structure and Permeability

---	Transcribed sequences (N48794)	-0.55												
ACTA2	actin, alpha 2, smooth muscle, aorta			-0.82		-0.50	0.67			-0.68	0.70			
ACTG2	actin, gamma 2, smooth muscle, enteric			-0.87		-0.64	0.63			-0.89	0.61			
ADD3	adducin 3 (gamma)			-0.66		-0.67		-0.48		-0.29	-0.40	-0.69		
ASPH	aspartate beta-hydroxylase					-0.27	0.34		-0.38	-0.32	0.47			
C5	complement component 5	-0.51	-0.36											
CD58	CD58 antigen, (lymphocyte function-associated antigen 3)			-0.42		-0.42	0.30		-0.30	-0.40		-0.28		
CDH1	cadherin 1, type 1, E-cadherin (epithelial)			-0.37		-0.42						-0.38		
CEACAM1§	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein) (Affymetrix probe: 211883_x_at)		1.01			0.43		0.47		0.66		0.60		
CEACAM1§	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein) (Affymetrix probe: 209498_at)		1.01							0.77	-0.80			
CRYBA2	crystallin, beta A2			-0.28		-0.33				-0.45		-0.46		
DES	desmin			-0.42		-0.33	0.32			-0.54	0.39			
DKFZp547A023	hypothetical protein DKFZp547A023	-0.72	-0.56											

Symbol		Name		cDNA arrays		Affymetrix arrays								
						CD						UC		
						1	2	3	4	5	6	1	2	3
CD_I-NC	UC_I-NC	CD_NI-NC	CD_I-CD_NI	CD_I-NC	CD_IS-CD_NI	CD_IS-NC	CD_I-CD_IS	UC_NI-NC	UC_I-UC_NI	UC_I-NC				
EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1			-0.70		-0.63	0.35	-0.35	-0.28	-1.02	0.72	-0.31		
EPB41L4B	erythrocyte membrane protein band 4.1 like 4B			-0.41		-0.69			-0.60		-0.71	-0.89		
EVA1	epithelial V-like antigen 1			-0.37		-0.28	0.27			-0.31		-0.37		
FAT	FAT tumor suppressor homolog 1 (Drosophila)			-0.33		-0.40	0.35		-0.42	-0.29				
FLJ25778	hypothetical protein FLJ25778	-0.41												
FLJ36812	hypothetical protein FLJ36812	-0.51	-0.39											
FMOD	fibromodulin			-0.31		-0.35	0.31		-0.35	-0.70	0.41	-0.29		
GJB3	gap junction protein, beta 3, 31kDa (connexin 31)			-0.53		-0.64		-0.40		-0.84	0.35	-0.49		
GMDS	GDP-mannose 4,6-dehydratase			0.39	0.32	0.72		0.45		0.82		0.78		
ITGB2	integrin, beta 2 (antigen CD18 (p95), lymphocyte function-associated antigen 1; macrophage antigen 1 (mac-1) beta subunit)	1.40	1.20			0.47					0.58	0.55		
KRT19	keratin 19			-0.60		-0.78			-0.51	-0.73	0.41	-0.32		
MAP7	microtubule-associated protein 7			-0.30		-0.30	0.40		-0.40	-0.31		-0.42		
MMP12	matrix metalloproteinase 12 (macrophage elastase)				0.94	1.14			0.99	0.78	1.36	2.14		
MMP2	matrix metalloproteinase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)	0.99	0.85			0.33					0.71	0.80		
MRPS18A	mitochondrial ribosomal protein S18A			-0.36		-0.41			-0.29	-0.43		-0.40		
MUC1	mucin 1, transmembrane	0.75	0.80					0.32		-0.29	0.57	0.28		
MUC13	mucin 13, epithelial transmembrane	0.43	0.71	-0.42		-0.29	0.37			-0.48	0.32			
MUC2	mucin 2, intestinal/tracheal			-0.50		-0.69				-1.08	0.74	-0.34		
MUC5B	mucin 5, subtype B, tracheobronchial					-0.37	0.48		-0.58	-0.47				
MYL9	myosin, light polypeptide 9, regulatory			-0.48		-0.34	0.44			-0.62	0.61			
NEBL	nebulette			-0.33		-0.36				-0.55	0.44			
OCLN	occludin	-0.42		-0.30		-0.33			-0.27	-0.27		-0.46		
PCDH21	protocadherin 21			-0.36	-0.43	-0.79			-0.60	-0.73		-0.79		
PDLIM5	PDZ and LIM domain 5	-0.57		-0.29										
PRKCB1	protein kinase C, beta 1	-0.58										0.35		
REG3A	regenerating islet-derived 3 alpha			1.33	1.52	2.84			2.07	1.67	0.98	2.65		
RPL22	ribosomal protein L22					-0.32			-0.34			-0.45		
SPINK5†	serine protease inhibitor, Kazal type 5			-0.99		-1.49	1.06		-1.55	-1.59	1.43			
SPON1	spondin 1, extracellular matrix protein			-0.88		-1.23	0.75		-1.10	-1.25	0.39	-0.86		
SPTBN1	spectrin, beta, non-erythrocytic 1			-0.27		-0.39			-0.33	-0.34		-0.33		
SYNE2	spectrin repeat containing, nuclear envelope 2			-0.40		-0.55			-0.38		-0.43	-0.67		
THBS1	thrombospondin 1			-0.75		-1.02		-0.66		-0.79		-0.70		
TIMP1	tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor)	1.23	2.04							-0.40	1.85	1.45		
TPM1	tropomyosin 1 (alpha)			-0.33		-0.34	0.36		-0.36	-0.38		-0.42		
TPM2	tropomyosin 2 (beta)			-0.43			0.36			-0.68	0.68			
TRAPPC4	trafficking protein particle complex 4			-0.32		-0.35			-0.27	-0.44		-0.37		
TUBB6	tubulin, beta 6			-0.29						-0.52	0.66			

Symbol		Name		cDNA arrays		Affymetrix arrays												
						CD						UC						
						1	2	3	4	5	6	1	2	3				
VIL2	villin 2 (ezrin)	CD_I-NC	UC_I-NC	CD_NI-NC	CD_I-CD_NI	CD_I-NC	CD_IS-CD_NI	CD_IS-NC	CD_I-CD_IS	UC_NI-NC	UC_I-UC_NI	UC_I-NC	-0.66	-0.53	-0.52	-0.42	0.67	-0.50
VWF	von Willebrand factor													0.40				0.90

* Shown are log₂ ratios for all microarray comparisons; all displayed values are log₂ ratio > 0.268 (greater than 1.2 fold change). Baseline group is always the second group listed in the column heading. For example, values in the column "CD_I-NC" are expressed as the log₂ ratio value of CD_I with respect to NC group. A minus sign before the log₂ ratio value indicates down-regulation with respect to baseline group.

[§] Log₂ ratios for CEACAM1 from real-time PCR on extended cohort: CD_NI-NC (0.68); CD_I-CD_NI (no regulation); CD_I-NC (0.59); UC_NI-NC (0.62); UC_I-UC_NI (-0.74); UC_I-NC (no regulation); DC_NI-NC (no regulation); DC_I-DC_NI (no regulation); DC_I-NC (no regulation).

^{††} Log₂ ratios for CYLD from real-time PCR on extended cohort: CD_NI-NC (-0.66); CD_I-CD_NI (no regulation); CD_I-NC (-0.71); UC_NI-NC (-1.06); UC_I-UC_NI (no regulation); UC_I-NC (-1.07); DC_NI-NC (no regulation); DC_I-DC_NI (-0.76); DC_I-NC (-0.92).

[†] Log₂ ratios for DMBT1 from real-time PCR on extended cohort: CD_NI-NC (0.78); CD_I-CD_NI (2.88); CD_I-NC (3.66); UC_NI-NC (0.65); UC_I-UC_NI (2.76); UC_I-NC (3.41); DC_NI-NC (0.79); DC_I-DC_NI (2.94); DC_I-NC (3.73).

[‡] Log₂ ratios for SPINK5 from real-time PCR on extended cohort: CD_NI-NC (no regulation); CD_I-CD_NI (no regulation); CD_I-NC (no regulation); UC_NI-NC (-1.01); UC_I-UC_NI (1.25); UC_I-NC (no regulation); DC_NI-NC (no regulation); DC_I-DC_NI (no regulation); DC_I-NC (no regulation).

3.3.5 mRNA expression of genes involved in Wnt signaling in IBD

To further investigate the role of the Wnt pathway in IBD, a comprehensive list of genes directly or indirectly involved in the Wnt signaling pathway was compiled using the Affymetrix pathway annotations and complemented by literature searching. Genes involved in the Wnt pathway may be divided into the following groups: A) upstream of Wnt (processing of wnt ligands or co-receptors); B) initiation of Wnt signaling (extracellular and integral membrane Wnt components); C) cytoplasmic Wnt signaling transduction; D) nuclear transcription of TCF/LEF; E) transcriptional targets of Wnt. The relative expression values of these genes on both the cDNA and Affymetrix microarrays are summarized in Table 3.23.

Table 3.23 Microarray expression of genes associated with Wnt signaling*

Gene symbol	CD Probe id	CD log2ratio			UC Probe id	UC log2ratio		
		CD_NI- NC	CD_I- CD_NI	CD_I- NC		UC_NI- NC	UC_I- UC_NI	UC_I- NC
A. upstream of Wnt (processing of wnt ligands or co-receptors)								
ADAM12	204943_at	-	-	-	204943_at	-	-	-
	28B08†	NA	NA	-	28B08	NA	NA	1.49
FOXL1	216572_at	-	-	-	216572_at	-	-	-
GPR177	221958_s_at	-0.40	-	-0.47	221958_s_at	-0.32	-0.18	-0.50
HS3ST1	205466_s_at	-	-	-	205466_s_at	-	0.17	-
HS3ST4	228206_at	-	-	-	228206_at	-	-	-
XYLT2	219401_at	-0.16	-	-0.15	219401_at	-0.23	0.11	-0.12
	231550_at	-0.21	-	-0.21	231550_at	-	-	-
B. initiation of Wnt signaling (extracellular and integral membrane Wnt components)								
CDH1	201130_s_at	-0.49	-	-0.37	201130_s_at	-	-	-0.38
CDH11	207173_x_at	-	-	-	207173_x_at	-	1.05	0.86
	54E05†	NA	NA	0.67	54E05	NA	NA	0.57
DKK3	214247_s_at	-	-	-	214247_s_at	-	0.52	0.32
FRZB	203697_at	-0.24	-0.30	-0.54	203697_at	-0.75	-	-0.81
	62O09†	NA	NA	0.61	62O09	NA	NA	0.85
FZD1	204451_at	-	-	-	204451_at	-	-	-
FZD2	210220_at	-	-	-	210220_at	-	-	-
FZD4	218665_at	-	-	-	218665_at	-	-	-
	92H05	NA	NA	-0.45	92H05	NA	NA	-
FZD6	203987_at	-	-	-	203987_at	-	-	-
	18M23†	NA	NA	-0.53	18M23	NA	NA	-0.40
FZD7	203706_s_at	-	-	-	203706_s_at	-0.37	-	-0.28
FZD8	227405_s_at	-0.26	-	-0.32	216587_s_at	-	-	-
GPC3	209220_at	-	-	-	209220_at	0.52	-0.40	-
LDLR	202068_s_at	-0.92	-	-0.79	202068_s_at	-1.07	0.61	-0.46
LRP6	225745_at	-0.14	-	-0.23	205606_at	-	-	-
SDC1	201287_s_at	-	-	-	201287_s_at	-	-	-
SFRP1	202037_s_at	-	-	-	202037_s_at	-0.40	-	-0.27
WNT4	230751_at	-	-	-	208606_s_at	-	0.20	0.20
WNT5A	205990_s_at	-	-	-	205990_s_at	-	0.44	0.61
C. cytoplasmic WNT signaling transduction								
APC	203527_s_at	-	-	-	203527_s_at	-	-	-
ARHGEF10	216620_s_at	-	-	-	216620_s_at	-	-	-
	47I05†	NA	NA	0.38	47I05	NA	NA	-
AXIN1	212849_at	-	-	-	212849_at	-	-	-
C2orf31	221245_s_at	-	-	-0.47	221245_s_at	-	-0.74	-0.79
CEACAM1	206576_s_at	-	-	0.48	206576_s_at	0.79	-	-

Gene symbol	CD Probe id	CD log2ratio			UC Probe id	UC log2ratio		
		CD_NI- NC	CD_I- CD_NI	CD_I- NC		UC_NI- NC	UC_I- UC_NI	UC_I- NC
	209498_at	-	-	-	209498_at	0.77	-0.80	-
	211883_x_at	-	-	0.43	211883_x_at	0.66	-	0.60
	211889_x_at	-	-	0.44	211889_x_at	0.80	-	-
	71D13†	NA	NA	-	71D13	NA	NA	1.01
CSNK1D	207945_s_at	-	-	-	207945_s_at	-	-	-
CSNK1D	208774_at	-0.28	-	-0.29	208774_at	-0.21	-0.18	-0.39
	59L04†	NA	NA	0.62	59L04†	NA	NA	-
CSNK1E	226858_at	-	-	-	222015_at	-	-	-
CTNNA1	200764_s_at	-	-	-	200764_s_at	-	-	-
	200765_x_at	-	-	-	200765_x_at	-	-	-
	210844_x_at	-0.27	-	-0.22	210844_x_at	-	-	-
CTNNB1	201533_at	-	-	-	201533_at	-	-	-
	223679_at	-0.19	-	-0.17				
DAAM1	216060_s_at	-0.33	-	-0.34	216060_s_at	-0.26	-	-0.35
DVL1	203230_at	-0.19	-	-0.21	203230_at	-	-	-
DVL2	57532_at	-	-	-	57532_at	-	-	-
	18E06†	NA	NA	-0.60	18E06†	NA	NA	-
DVL3	201908_at	-0.25	-	-0.30	201908_at	-0.21	-	-0.27
FBXW2	209630_s_at	-0.34	-	-0.29	209630_s_at	-0.30	-	-0.28
	91B05†	NA	NA	-	91B05†	NA	NA	0.57
FRAT2	209864_at	-	-	-	209864_at	-	-0.27	-0.28
GSK3B	226183_at	-	-	-0.12	209945_s_at	-	-	-
	226191_at	-0.19	-	-0.23				
MAPK9	225781_at	-	-	-0.21				
	203218_at	-0.21	-	-0.21	203218_at	-	-	-
MUC1	207847_s_at	-	-	-	207847_s_at	-0.56	0.88	-
	211695_x_at	-	-	-	211695_x_at	-0.29	0.57	0.28
	213693_s_at	-	-	-	213693_s_at	-0.52	0.92	-
	22E24†	NA	NA	0.75	22E24†	NA	NA	0.80
PAFAH1B1	200816_s_at	-0.30	-	-0.29	200816_s_at	-0.15	-0.15	-0.30
PLAU	205479_s_at	-	-	-	205479_s_at	-	0.73	0.64
PPP2R5C	201877_s_at	-0.27	-	-	201877_s_at	-0.20	-	-
PPP2R5E	203338_at	-0.35	-	-0.37	203338_at	-0.49	-	-0.38
PRKCA	213093_at	-	-	-	213093_at	-	-0.29	-0.20
PRKCB1	209685_s_at	0.17	-	0.17	209685_s_at	-	0.24	0.35
	01M08†	NA	NA	-0.59	01M08†	NA	NA	-
PRKCD	202545_at	-	-	-	202545_at	-	-0.20	-0.17
PRKCE	226101_at	-	-	-0.10				
PRKCH	218764_at	-	-	-	218764_at	-	0.33	0.28
	18L05†	NA	NA	-0.39	18L05†	NA	NA	-
PRKCI	209678_s_at	-	-	-	209678_s_at	-0.21	-	-0.32
PRKCZ	202178_at	-	-	-	202178_at	-	-0.31	-0.32
RHOA	200059_s_at	-0.24	-	-	200059_s_at	-	-	-
VCL	200931_s_at	-0.38	-	-0.39	200931_s_at	-0.38	-	-0.27

D. nuclear transcription of TCF/LEF

CREBBP	202160_at	-	-	-	202160_at	-0.26	-	-0.25
CTBP1	203392_s_at	-0.29	-	-0.28	203392_s_at	-0.42	-	-0.47
CTBP2	210835_s_at	-0.36	-	-0.38	210835_s_at	-0.30	-	-0.37
HDAC1	201209_at	-0.25	-	-	201209_at	-	-	-
HDAC2	201833_at	-0.19	-	-0.14	201833_at	-	-	-
HDAC7A	217937_s_at	-	-	0.18	217937_s_at	-	-	-
LEF1	221558_s_at	-	-	-	221558_s_at	-	0.54	0.54
TCF7L2	212759_s_at	-	-	-	212759_s_at	-	-0.20	-0.23
	212762_s_at	-	-	-	212762_s_at	-	-0.28	-0.34
	216035_x_at	-	-	-	216035_x_at	-	-0.20	-0.23
	216037_x_at	-	-	-	216037_x_at	-	-0.19	-0.26
	216511_s_at	-	-	-	216511_s_at	-	-	-0.27
TLE1	203221_at	-0.39	-	-0.28	203221_at	-0.34	-	-0.23
TLE2	40837_at	-0.17	-	-0.30	40837_at	-0.26	-0.18	-0.44
TLE3	212770_at	-	-	-0.16	212770_at	-	-	-

Gene symbol	CD Probe id	CD log ₂ ratio			UC Probe id	UC log ₂ ratio		
		CD_NI-NC	CD_I-CD_NI	CD_I-NC		UC_NI-NC	UC_I-UC_NI	UC_I-NC
E. transcriptional targets of WNT								
CCND1	208712_at	-0.30	-	-0.26	208712_at	-0.42	-	-0.35
CCND2	200951_s_at	-	-	-	200951_s_at	-0.28	0.34	-
	200952_s_at	-0.33	-	-0.28	200952_s_at	-0.23	-	-0.15
	200953_s_at	-0.56	-	-0.49	200953_s_at	-0.76	0.21	-0.55
CCND3	201700_at	-	-	-0.20	201700_at	-	-	-
EMP3	203729_at	-	-	-	203729_at	-0.30	0.62	0.32
EPHB2	209589_s_at	-0.19	-	-0.25	209589_s_at	-0.23	-	-0.23
EPHB3	1438_at	-0.29	-	-0.30	1438_at	-0.45	0.15	-0.30
FOSL1	204420_at	-	-	-	204420_at	-0.21	0.19	-
JUN	201465_s_at	-	-	-	201465_s_at	0.16	-	0.14
JUNB	201473_at	-	-	-	201473_at	-	0.37	-
	22G16†	NA	NA	0.60	22G16†	NA	NA	0.69
MYC	202431_s_at	-0.39	-	-	202431_s_at	-0.35	-	-0.22
PPARD	37152_at	-	-	-	37152_at	-	-0.27	-0.28
VANGL2	226029_at	-0.20	-	-0.34				

* Significant log₂ ratio values are shown from both the cDNA and Affymetrix microarrays. Baseline group is always the second group listed in the column heading. For example, values in the column "CD_NI-NC" are expressed as the log₂ ratio value of CD_NI with respect to NC group. A minus sign before the log₂ ratio value indicates down-regulation with respect to baseline group. All probes listed are from the Affymetrix arrays, unless indicated with a "†"—these probes originated from the cDNA arrays. A lone minus sign within a cell indicates that the probe was not significantly regulated. "NA" indicates that the comparison was not applicable for that probe id.

Genes involved in processes preceding the initiation of Wnt signaling include those involved in ligand processing (proteases), heparin sulfate proteoglycan (HSPG) synthesis (sulfotransferase, xylosyltransferase), and transcription factors involved in modulating heparin sulfate proteoglycan levels (*FOX L1*). A recently characterized gene involved in intracellular processing of Wnt ligands, *G protein-coupled receptor 177 (GPR177)*, was suggested to promote secretion of Wnt ligands by either directing proper secretion of Wnt or by post-translational modification of the Wnt ligand (Banziger et al., 2006; Bartscherer et al., 2006). Microarray analysis indicated that this gene was down-regulated in IBD (inflamed or non-inflamed) compared to normal controls, possibly indicating that Wnt ligand processing, and thus Wnt pathway activation, is reduced in CD and UC. Differential regulation of *HS3ST1* and *XYLT2*, genes involved in HSPG synthesis, may indicate that different HSPG modifications are made in IBD mucosa, which could modulate Wnt signaling in a different manner compared to normal or non-inflamed mucosa.

In contrast to events preceding Wnt signaling, many genes involved in signal transduction process have been well-characterized in the literature. Wnt signaling is initiated by binding of Wnt ligands to the Frizzled or LRP receptors. Wnt ligand binding can be inhibited by specific proteins (secreted frizzled-related proteins, Wnt inhibitory factor, dickkopf) and the presence of co-receptors (HSPG, KREMEN) may promote/inhibit Wnt ligand binding. The number and variety of Wnt ligands and

receptors and co-receptors for Wnt ligands is extensive. Over half of the 20 Wnt family ligand genes were detected as being 'present' on the microarrays, and interestingly, two Wnt genes involved in the non-canonical Wnt pathway, *WNT4* and *WNT5A* were only differentially regulated in UC. In terms of colon maintenance, *WNT4* and *5A* inhibit proliferation and promote differentiation (Gregorieff et al., 2005). *WNT5A* up-regulation may be particularly relevant to IBD in light of the recent finding that *WNT5A/FZD5* signaling is able to regulate anti-microbial response in human antigen presenting cells and macrophages (Blumenthal et al., 2006).

At least 70% of the ten frizzled family of receptors (*FRZ1-10*) were detected on the arrays, and two genes were differentially regulated, *FZD7* in UC and *FZD8* in CD. Of the frizzled-related genes, which include 5 family members (*SFRP1,2,4,5* and *FRZB*), 60% were detected and differential expression was monitored in *FRZB* (both CD and UC) and *SFRP1* (only UC). The dickkopf (*DKK1-4*) and dickkopf-like (*DKKL1-2*) family of genes, of which there are six members, is represented by the expression of three members on the arrays (*DKK3*, *DKK4* and *DKKL1*), and *DKK3* showed differential expression in UC. Known co-receptors of the Wnt pathway (*KREMEN1* and *2*, *LRP5* and *6*) and potential modulators, including the heparin sulfate proteoglycan proteins (syndecan, perlecan and glypican), were detected. Of these, *GPC3* was differentially regulated in UC.

Nuclear members of the canonical Wnt signaling pathway (Groucho, CtBP, HDAC, CBP, TCF) were also generally down-regulated with respect to non-inflamed tissue (normal or IBD). In the inactive state, Groucho (official name: *TLE1-4*; transducin-like enhancer of split), CtBP (official name: *CTBP1-2*; C-terminal binding protein) and histone deacetylases (HDAC) repress transcription activation by TCF (Chinnadurai, 2002). Upon Wnt pathway activation, β -catenin translocates to the nucleus, and along with the co-activator CBP (official name: *CREBBP*; CREB-binding protein), β -catenin displaces the repressor Groucho, and converts the repressor complex into a transcription activation complex (Logan and Nusse, 2004). As observed on the microarrays, down-regulation of the repressors *CTBP1-2*, *TLE1-3*, *HDAC1,2,7A* in IBD compared to normal controls favored activation of the Wnt pathway in IBD.

In the comparisons between inflamed IBD and normal controls, it was also possible to incorporate microarray data from the cDNA microarray dataset, which consisted of a similar comparison

between inflamed IBD and hospitalized normal controls. Results from both array platforms were available for 14 of the Wnt signaling genes. For six genes (*ADAM12*, *ARHGEF10*, *DVL2*, *JUNB*, *FZD6*, *FZD4*), differential regulation was detected in the cDNA microarray only. In five cases, (*PRKCH*, *FBXW2*, *CSNK1D*, *FRZB*, *PRKCB1*), the regulation from the cDNA or the Affymetrix platform was in the opposite direction in one or both IBD subtypes. In the remaining three cases (*CEACAM1*, *MUC1*, *CDH11*) either or both IBD subtypes showed differential regulation in the same direction (up- or down-regulation) as the other microarray platform.

The mRNA expression of Wnt pathway associated genes in an expanded real-time PCR cohort was further investigated in 27 genes from the microarray results in Table 3.24. In general, differentially regulated Wnt genes were heavily influenced by inflammation state—about 80% of Wnt genes quantitated in the extended cohort were differentially expressed in inflamed CD or UC compared to normal controls. As for comparisons between inflammation states of a particular disease group, most regulated genes were down-regulated in all diseases, and the UC group contained the most up-regulated genes (seven genes in total). Eight genes (*ADAM12*, *CDH11*, *SFRP2*, *LEF1*, *TCF7L2*, *EMP3*, *JUNB* and *TCF4*), representing 30% of the transcripts quantitated by real-time PCR, were specifically regulated in IBD, as no differential regulation was detected in the disease control group.

Table 3.24 Relative mRNA transcript levels of Wnt pathway associated genes as determined by real-time PCR[†]

Gene symbol	Assay ID	CD log2ratio			UC log2ratio			DC log2ratio		
		CD_NI -NC	CD_I- CD_NI	CD_I- NC	UC_NI -NC	UC_I- UC_NI	UC_I- NC	DC_N I-NC	DC_I- DC_NI	DC_I- NC
A. upstream of WNT (processing of wnt ligands or co-receptors)										
ADAM12	Hs00291596_s1	-	0.28	0.39*	-	1.16*	0.99*	-	-	-
FOXL1	Hs00534264_s1	-0.62*	-	-0.71*	-0.78*	-	-0.74*	-	-	-0.53
B. initiation of WNT signaling (extracellular and integral membrane WNT components)										
CDH1	Hs00170423_m1	-	-0.80*	-0.89*	-0.21	-0.72	-0.93*	-	-0.54	-0.63*
CDH11	92E24_CDH11	-	-	0.48	-	0.72	1.13*	-	-	-
FRZB	AAX00559008s1	-	-0.52*	-0.33	-	-0.49	-0.55*	-	-0.62*	-0.59
FRZB	Hs00173503_m1	-1.11*	-0.55	-1.66	-0.47	-1.35*	-1.82*	-1.44*	-	-2.08*
FZD1	Hs00268943_s1	-0.27	-0.45	-0.72*	-0.45	-0.79	-1.23*	-0.35	-	-1.17
FZD2	Hs00361432_s1	-	-	-	-0.23	0.42	-	-	-	-0.81
FZD7	Hs00275833_s1	-	-	-0.78	-	-0.39	-0.98*	-0.74	-	-1.25
SDC1	Hs00174579_m1	-	-0.36	-0.38	-0.25	-	-0.20	-	-0.44	-0.52
SFRP2	SERP2.1M1	-	-	1.41	-	2.32	2.68*	-	-	-
C. cytoplasmic WNT signaling transduction										
APC	Hs00181051_m1	-	-	-0.32	-0.27	-0.23	-0.50*	-0.20	-	-0.51
ARHGEF10	Hs00206020_m1	-0.67*	0.32	-0.35	-0.78*	0.50*	-0.28	-0.53*	-	-
AXIN1	Hs00394718_m1	-	-0.60*	-0.76*	-	-0.57*	-0.77*	-	-0.64	-0.83
CSNK1D	AAX00503624m1	-	-0.15	-0.16*	-0.23	-0.14	-0.36*	-	-0.56	-0.43

Gene symbol	Assay ID	CD log2ratio			UC log2ratio			DC log2ratio		
		CD_NI -NC	CD_I- CD_NI	CD_I- NC	UC_NI -NC	UC_I- UC_NI	UC_I- NC	DC_N I-NC	DC_I- DC_NI	DC_I- NC
CSNK1E	Hs00266431_m1	-0.77*	-	-0.81*	-0.72*	-	-0.51	-0.75*	-	-1.00
CTNNA1	Hs00426996_m1	-0.43	-	-0.59*	-0.23	-0.51*	-0.74*	-0.24	-	-0.39
CTNNB1	Hs00170025_m1	-0.37*	-	-0.44*	-0.42*	-	-0.41*	-	-	-0.44
DVL1	Hs00182896_m1	-0.60*	-	-0.62*	-0.37	-	-0.48*	-0.40*	-	-0.59
GSK3B	Hs00275656_m1	-0.85*	-0.21	-1.06*	-0.55*	-0.48*	-1.03*	-0.52*	-0.45	-0.97*
VCL	Hs00243320_m1	-0.73*	-0.31	-1.03*	-0.71*	-	-0.85*	-0.37	-	-
D. nuclear transcription of TCF/LEF										
LEF1	Hs00212390_m1	-	0.95	-	-	0.88*	0.43	-	-	-
TCF7	Hs00175273_m1	-	-	0.57	-	-	-	0.58	-	-
TCF7L2	Hs00181036_m1	0.26	-0.89*	-0.64*	-	-	-0.38	-	-	-
E. transcriptional targets of WNT										
EMP3	Hs00171319_m1	-	-	-	-	0.40	0.56	-	-	-
IL8	IL-8_	-	3.25	2.66	-	3.57	3.48	NA	NA	NA
JUNB	Hs00357891_s1	-	-	-	-	0.72	-	-	-	-
TCF4	Hs00162613_m1	-	-	0.49	-	0.47	0.78	-	-	-

[†] Significant log₂ ratio values (p<0.05) are shown from both the cDNA and Affymetrix microarrays. Baseline group is always the second group listed in the column heading. For example, values in the column "CD_NI-NC" are expressed as the log₂ ratio value of CD_NI with respect to NC group. A minus sign before the log₂ ratio value indicates down-regulation with respect to baseline group. A lone minus sign within a cell indicates that the probe was not significantly regulated. An asterisk indicates p<0.001 for that ratio. The number of samples in each group range as follows: NC (12-25), inflamed CD (17-36), non-inflamed CD (16-32), inflamed UC (27-47), non-inflamed UC (8-30), inflamed DC (6-13), non-inflamed DC (17-27).

3.3.6 mRNA expression of genes involved in barrier function

Genes related to cellular junction were extracted from the GO annotations of the Affymetrix probesets and additional genes associated to known cell junction proteins were complemented by literature search. The genes were categorized into five categories: A) intercellular junction; B) tight junction; C) cell-cell adherens junction; D) cell-matrix junction; and E) other genes associated with barrier function. In cases where a gene was associated to multiple categories, the gene was classified by its most specific description, i.e. GO term furthest from the root term. The mRNA expression, as detected by cDNA and Affymetrix arrays, is depicted in Table 3.25.

Table 3.25 Microarray expression of genes associated with barrier function*

Gene sym.bol	CD Probe id	CD log2ratio			UC Probe id	UC log2ratio		
		CD_NI- NC	CD_I- CD_NI	CD_I- NC		UC_NI- NC	UC_I- UC_NI	UC_I- NC
A. intercellular junction								
CNN2	201605_x_at	-	-	-	201605_x_at	-0.10	0.28	0.17
DLG1	202514_at	-0.23	-	-0.26	202514_at	-0.21	-	-0.23
DLG5	201681_s_at	-0.14	-	-0.15	201681_s_at	-	-	-
DSC2	204750_s_at	-	-	-	204750_s_at	-0.23	-	-0.40
	204751_x_at	-0.40	-	-0.32	204751_x_at	-0.23	-0.20	-0.43
DSG2	217901_at	-0.26	-	-0.26	217901_at	-	-	-0.25
GJA1	201667_at	-	-	-	201667_at	-0.70	0.89	-
GJB1	204973_at	-	-	-	204973_at	0.25	-0.50	-0.25
GJB3	205490_x_at	-0.53	-	-0.64	205490_x_at	-0.84	0.35	-0.49
	215243_s_at	-0.40	-	-0.44	215243_s_at	-0.69	0.32	-0.37
PDZK2	220303_at	-	-	-	220303_at	-	-0.25	-0.15
PECAM1	208982_at	-	-	-	208982_at	-	0.87	0.93
	208983_s_at	-	-	0.37	208983_s_at	-	0.77	0.90
PKP2	207717_s_at	-0.28	-	-	207717_s_at	-	-0.38	-0.51
PNN	212036_s_at	-0.28	-	-0.18	212036_s_at	-0.21	-	-
	212037_at	-0.32	-	-0.29	212037_at	-0.35	-	-0.35
STEAP1	205542_at	-	-	-	205542_at	0.50	-	0.34
VCL	200931_s_at	-0.38	-	-0.39	200931_s_at	-0.38	-	-0.27
B. tight junction								
ALS2CR19	228411_at	-0.16	-	-0.23				
ARHGEF10	216620_s_at	-	-	-	216620_s_at	-	-	-
ASH1L	226447_at	-	-	-0.22				
C3orf4	208925_at	-0.25	-	-0.20	208925_at	-0.23	0.19	-
CASK	207620_s_at	-0.25	-	-0.21	207620_s_at	-0.21	-	-0.17
CASK	211208_s_at	-0.27	-	-0.25	211208_s_at	-0.20	-	-0.27
	60L19†	NA	NA	-0.37	60L19†	NA	NA	-0.46
CDC42	208727_s_at	-	-	-	208727_s_at	-	-	-
CLDN1	218182_s_at	-	-	-	218182_s_at	-	-	-
CLDN15	219640_at	-	-	-	219640_at	-	-0.32	-0.30
CLDN3	203953_s_at	-0.75	-	-0.82	203953_s_at	-0.47	-0.36	-0.83
CLDN3	203954_x_at	-0.52	-	-0.67	203954_x_at	-0.57	-0.28	-0.85
	53N07†	NA	NA	0.56	53N07†	NA	NA	-
CLDN4	201428_at	-0.55	-	-0.72	201428_at	-0.49	-	-0.63
	59O01†	NA	NA	0.77	59O01†	NA	NA	-
CLDN5	204482_at	-	-	-	204482_at	-	0.20	0.21
CLDN7	202790_at	-0.29	-	-0.30	202790_at	-	-0.40	-0.35
CLDN8	214598_at	-1.57	-	-2.09	214598_at	-2.62	-	-2.62
CRB3	236279_at	-0.24	-	-0.36				
CSDA	201160_s_at	-0.32	-	-0.23	201160_s_at	-0.33	0.26	-
	201161_s_at	-	-	-	201161_s_at	-0.32	0.31	-
CXADR	203917_at	-0.79	-	-0.85	203917_at	-0.38	-0.17	-0.55
	239155_at	-0.22	-	-0.29				
F11R	221664_s_at	-0.33	-	-0.19	221664_s_at	-	-	-
INADL	223681_s_at	-0.23	-	-0.30				
LIN7C	219399_at	-0.22	-	-0.17	219399_at	-	-	-
	221568_s_at	-0.16	-	-0.15	221568_s_at	-0.13	-0.15	-0.28
MAG11	225465_at	-	-	-0.22				
OCLN	209925_at	-0.30	-	-0.33	209925_at	-0.27	-0.19	-0.46
	38K23†	NA	NA	-0.42	38K23	NA	NA	-
PARD3	221526_x_at	-0.27	-	-0.30	221526_x_at	-0.23	-	-0.29
PPP2CA	208652_at	-0.39	-	-0.28	208652_at	-0.36	-	-0.24
	238719_at	-0.11	-	-0.11				
	33E15†	NA	NA	-0.56	33E15†	NA	NA	-
PPP2CB	201375_s_at	-0.39	-	-0.35	201375_s_at	-0.19	-	-0.30
PPP2R1B	202883_s_at	-0.28	-	-0.29	202883_s_at	-0.15	-0.17	-0.32
	202886_s_at	-0.26	-	-0.17	202886_s_at	-	-	-
	47O03†	NA	NA	1.11	47O03	NA	NA	-
PPP2R2A	202313_at	-0.24	-	-	202313_at	-	-	-
	228013_at	-0.16	-	-0.13				
PPP2R3A	209633_at	-	-	-	209633_at	-	-0.36	-0.28

Gene sym.bol	CD Probe id	CD log2ratio			UC Probe id	UC log2ratio		
		CD_NI-NC	CD_I-CD_NI	CD_I-NC		UC_NI-NC	UC_I-UC_NI	UC_I-NC
PRKCA	213093_at	-	-	-	213093_at	-	-0.29	-0.20
PRKCD	202545_at	-	-	-	202545_at	-	-0.20	-0.17
PRKCH	218764_at	-	-	-	218764_at	-	0.33	0.28
	18L05†	NA	NA	-0.39	18L05	NA	NA	-
PRKCZ	202178_at	-	-	-	202178_at	-	-0.31	-0.32
PTEN	228006_at	-	-	-0.21				
RAB13	202252_at	-0.25	-	-0.22	202252_at	-0.19	0.32	0.13
ROCK1	230239_at	-	-	-0.10				
	18N21†	NA	NA	-0.57	18N21	NA	NA	-
ROCK2	202762_at	-	-	-0.22	202762_at	-0.18	-0.14	-0.31
TJP2	202085_at	-0.36	-	-0.27	202085_at	-	-	-0.28
TJP3	213412_at	-	-	-	213412_at	0.31	-0.42	-
	35148_at	-	-	-	35148_at	0.20	-0.47	-0.27
VAPA	208780_x_at	-0.18	-	-0.19	208780_x_at	-	-	-
WNK4	229158_at	-0.25	-	-0.37				
C. cell-cell adherens junction								
CDH1	201130_s_at	-0.49	-	-0.37	201130_s_at	-	-	-0.38
	201131_s_at	-0.37	-	-0.42	201131_s_at	-0.14	-0.24	-0.38
CDH11	207173_x_at	-	-	-	207173_x_at	-	1.05	0.86
	54E05†	NA	NA	0.67	54E05	NA	NA	0.57
CDH17	209847_at	-	-	-	209847_at	-	-0.30	-0.21
CDH3	203256_at	-	-	-	203256_at	-	0.65	0.66
CDH5	204677_at	-	-	-	204677_at	-	0.36	0.43
CTNNB1	223679_at	-0.19	-	-0.17				
	201533_at	-	-	-	201533_at	-	-	-
CTNND1	208407_s_at	-0.38	-	-0.48	208407_s_at	-0.22	-0.40	-0.62
DSP	200606_at	-0.42	-	-0.48	200606_at	-0.27	-0.33	-0.60
MLLT4	214939_x_at	-0.20	-	-0.13	214939_x_at	-	-	-
	224685_at	-0.23	-0.24	-0.47				
PGM5	226303_at	-0.21	-	-0.24				
SDCBP	200958_s_at	-	-	-	200958_s_at	-	0.30	-
TJP1	202011_at	-0.47	-	-0.52	202011_at	-0.47	-	-0.59
	214168_s_at	-0.23	-	-0.15	214168_s_at	-	-	-
D. cell-matrix junction								
DST	212254_s_at	-0.41	-	-0.49	212254_s_at	-0.41	-	-0.38
ERBB2IP	217941_s_at	-	-	-	217941_s_at	0.35	-0.46	-
SORBS1	218087_s_at	-0.53	-	-0.60	218087_s_at	-0.80	-	-0.63
E. Other genes associated with barrier function								
CLTA	216295_s_at	-0.18	-	-0.15	216295_s_at	-	-	-
CLTC	200614_at	-0.39	-	-0.42	200614_at	-0.26	-	-0.35
	22P10†	NA	NA	-		NA	NA	1.54
FLNA	200859_x_at	-0.27	-	-	200859_x_at	-0.45	0.40	-
	213746_s_at	-	-	-	213746_s_at	-0.36	0.38	-
	214752_x_at	-	-	-	214752_x_at	-0.49	0.43	-
	67G22†	NA	NA	-		NA	NA	0.51
ITM2C	221004_s_at	-0.26	-	-0.51	221004_s_at	-0.54	-	-0.59
	55I12†	NA	NA	0.71	55I12	NA	NA	-
MUC1	207847_s_at	-	-	-	207847_s_at	-0.56	0.88	-
	211695_x_at	-	-	-	211695_x_at	-0.29	0.57	0.28
PICALM	212506_at	-0.23	-	-	212506_at	-	-	-
	01A04†	NA	NA	-0.55	01A04	NA	NA	-

* Significant log₂ ratio values are shown from both the cDNA and Affymetrix microarrays. Baseline group is always the second group listed in the column heading. For example, values in the column "CD_NI-NC" are expressed as the log₂ ratio value of CD_NI with respect to NC group. A minus sign before the log₂ ratio value indicates down-regulation with respect to baseline group. All probes listed are from the Affymetrix arrays, unless indicated with a "†"—these probes originated from the cDNA arrays. A lone minus sign within a cell indicates that the probe was not significantly regulated. "NA" indicates that the comparison was not applicable for that probe id.

The GO term intercellular junction (Table 3.25) is a broad category that includes genes involved in specialized regions that connect two cells. The gap junction family of genes form hemichannels, or connexons, which allow cell-cell communication in adjacent cells. Of 17 gap junction family members represented on the array, 14 were detected (*CX36*, *CX40.1*, *GJA1*, *GJA3*, *GJA4*, *GJA5*, *GJA7*, *GJA8*, *GJB1*, *GJB2*, *GJB3*, *GJB4*, *GJB5*, *GJC1*), and three genes (*GJA1*, *GJB1*, *GJB3*) were differentially regulated, mostly down-regulated in inflamed CD or UC samples compared to healthy controls. The desmoglein subfamily (*DSG1-3*) and plakophilins (*PKP1* and *2*) are involved in cell-cell junctions (desmosomes). *DSG2* is represented by three probes on the Affymetrix microarray, each of which indicates different pattern of regulation in the six comparisons, probably due to different gene transcripts; however, in general, *DSG2* was down-regulated in inflamed IBD compared to normal controls. Both plakophilins were detected by the arrays, but only *PKP2* was differentially down-regulated in disease compared to normal controls. Other genes associated with cell adhesion and/or cell matrix interactions (Table 3.25) were also generally down-regulated in inflamed samples compared to normal controls. These include MAGUK family members *DLG1* and *DLG5*, regulator of E-cadherin *PNN*, and cell matrix components *VCL*, *DST* and *SORBS1*. Within the intercellular junction group, the few genes that displayed a general up-regulation in inflamed CD and/or UC compared to non-inflamed samples include cell-adhesion receptor *PECAM1*, cell adhesion associated *CNN2*, and the cell-cell junction gene *STEAP1*.

The tight junction complex is the most apical cell-cell barrier in the epithelial layer. Transmembrane proteins, such as the junctional adhesion molecule (JAM) and claudin (CLDN) families and occludin (OCLN), form contacts to adjacent cells. On the microarrays, 16/18 claudin family members were detected as being present on the array, and six claudin genes were differentially expressed in at least one IBD subtype. *CLDN3,4,7,8* were down-regulated in inflamed compared to non-inflamed samples on the Affymetrix arrays, while *CLDN5* was up-regulated. In contrast, the cDNA microarrays showed that *CLDN 3* and *4* were up-regulated in inflamed CD compared to hospitalized normal controls. All three JAM genes were detected (*F11R*, *JAM2*, *JAM3*), and only *F11R* was differentially down-regulated in CD compared to normal controls. *OCLN* was down-regulated in inflamed IBD compared to normal controls, a result which was replicated in the cDNA microarrays. Adaptor proteins, such as the tight junction (TJP) and membrane associated guanylate kinase (MAGI) families, interact between the transmembrane and cytoplasmic proteins

and serve to transmit signals to and from the tight junction. Tight junction genes *TJP1-3* were all down-regulated in inflamed IBD compared to normals, while *TJP3* was up-regulated in non-inflamed UC compared to normal controls. Membrane associated guanylate kinases *MAGI1-3* were only detected in CD, and *MAGI1* was down-regulated in inflamed CD compared to normals. Finally, genes regulating tight junction assembly or paracellular permeability, including the protein phosphatase 2 (*PPP2C* and *PPP2R*), protein kinase C (*PRKCD* and *PRKCZ*) and the Rho-associated, coiled-coil containing protein kinase genes (*ROCK1* and *2*), were down-regulated in at least one inflamed IBD subtype compared to normal controls or non-inflamed samples. RAS oncogene family member 13 (*RAB13*) and *PRKCH* were the only two genes with up-regulation in inflamed UC compared to non-inflamed samples.

Many of the genes involved in cell-cell adhesion are represented by the cadherin and catenin families. From 17 cadherin family members on the Affymetrix arrays, 15 were present. *CDH1* and *17* were down-regulated in inflamed IBD compared to normal or non-inflamed samples. *CDH 3, 5* and *11* were up-regulated only in inflamed UC compared to non-inflamed UC or normal controls. The up-regulation of *CDH11* in inflamed UC compared to hospitalized normal controls was replicated in the cDNA array, but additional gene regulation was observed for CD as well. Two forms of the catenin (cadherin-associated protein) family (*CTNNB1*, *CTNND1*) were down-regulated (inflamed or diseased compared to non-inflamed or normal control) in the Affymetrix microarrays.

Other genes associated with barrier function, either by endocytosis or association with known components of cell junction, include the clathrins or clathrin binding proteins (*CLTA*, *CLTC*, *PICALM*; generally down-regulated in disease compared to normal controls) and *MUC1* and *FLNA*, which were both down-regulated in non-inflamed UC compared to normal controls but up-regulated in inflamed UC compared to non-inflamed UC.

A select number of genes (23 genes) involved in barrier function were further investigated using real-time PCR in an expanded number of IBD samples, as well as in a cohort of inflamed and non-inflamed disease controls (DC). Overall, 18 genes were down-regulated in either or both inflamed IBD subtypes compared to normal controls (Table 3.26). In four genes, up-regulation was observed

in either or both inflamed IBD subtypes compared to normal controls: *CLDN1*, *CLDN2*, *CDH11* and *MUC1*. For *TJP2* and *CDH11*, the observed up-regulation appeared to be specific to IBD, since no regulation was observed in any of the comparisons involving disease controls. Furthermore, the down-regulation in *VCL*, *ARHGEF10*, and *ITM2C* were observed in inflamed IBD but not in inflamed disease controls, possibly indicating that regulation of these genes is influenced by disease rather than general inflammation. On the other hand, all other genes (*DLG5*, *CLDN1*, *CLDN2*, *CLDN3*, *CLDN4*, *CLDN7*, *OCLN*, *TJP1*, *TJP3*, *CHD1*, *CTNNA1*, *CTNNB1*, *ERBB2IP*, *CLTC*, *FNLA*, *MUC1*, *NRG1*, *SORBS3*) were regulated in all inflamed disease (CD, UC, DC) samples compared to normal controls.

Table 3.26 Relative mRNA transcript levels of gene associated with barrier function, as determined by real-time PCR[†]

Gene symbol	Assay ID	CD log2ratio			UC log2ratio			DC log2ratio		
		CD_NI-NC	CD_I-CD_NI	CD_I-NC	UC_NI-NC	UC_I-UC_NI	UC_I-NC	DC_NI-NC	DC_I-DC_NI	DC_I-NC
A. Intercellular junction										
DLG5	DLG5_A	-0.50*	-	-0.53*	-0.58*	-	-0.41*	-0.68*	-	-0.75*
VCL	Hs00243320_m1	-0.73*	-0.31	-1.03*	-0.71*	-	-0.85*	-0.37	-	-
B. Tight junction										
ARHGEF10	Hs00206020_m1	-0.67*	0.32	-0.35	-0.78*	0.5 *	-0.28	-0.53*	-	-
CLDN1	Hs00221623_m1	-	1.47*	1.27*	-	2.00*	1.85*	0.62	-	1.62
CLDN2	Claudin2.2s1	-	3.1*	3.52*	-	4.86*	5.30*	-	2.94	4.01*
CLDN3	Hs00265816_s1	-0.51	-0.43	-0.94*	-0.47*	-0.64	-1.11*	-0.56*	-	-0.54*
CLDN4	Hs00533616_s1	-1.03*	-	-1.23*	-1.15*	-0.63	-1.78*	-0.59	-	-0.82*
CLDN7	Hs00600772_m1	-	-0.46*	-0.63*	-	-0.85*	-0.73*	-	-0.47	-0.46
OCLN	Hs00170162_m1	-0.48*	-0.47	-0.95*	-0.68*	-0.72	-1.40*	-0.54*	-0.36	-0.90*
TJP1	Hs00268480_m1	-0.91*	-0.36	-1.28*	-0.94*	-0.46	-1.40*	-0.70*	-	-0.96*
TJP2	Hs00178081_m1	-	0.48	-	-	0.58	0.56	-	-	-
TJP3	Hs00274276_m1	-	-0.21	-0.44	-	-0.81*	-0.82*	-	-0.56	-0.43
C. Cell-cell adherens junction										
CDH1	Hs00170423_m1	-	-0.80*	-0.89*	-0.21	-0.72	-0.93*	-	-0.54	-0.63*
CDH11	92E24_CDH11	-	-	0.48	-	0.88*	1.13*	-	-	-
CTNNA1	Hs00426996_m1	-0.43	-	-0.59*	-0.23	-0.51*	-0.74*	-0.24	-	-0.39
CTNNB1	Hs00170025_m1	-0.37*	-	-0.44*	-0.42*	-	-0.41*	-	-	-0.44
D. Cell-matrix junction										
ERBB2IP	Hs00180965_m1	-	-0.34	-0.52	-0.38	-0.63	-1.01*	-	-0.76	-0.73
E. Other genes associated with barrier function										
CLTC	Hs00191535_m1	-0.52*	-0.34	-0.86*	-0.71*	-0.37	-1.08*	-0.38	-	-0.41
FNLA	Hs00155065_m1	-1.64*	-	-1.62*	-1.32*	-	-1.50*	-1.31*	-	-1.33
ITM2C	ITM2C_1s_	-0.66	-	-0.79*	-0.61*	-0.47	-1.08*	-0.79*	-	-
MUC1	Mucin(22E24)	-	0.82	0.77	-	1.40*	0.94*	-	1.11	0.65
NRG1	Hs00247620_m1	-0.65	-	-0.78*	-1.04	-	-0.49	-0.57	-	-1.14
SORBS3	Hs00195059_m1	-	-0.94*	-1.10*	-0.39*	-1.22	-1.61*	-0.35	-0.81	-1.16*

[†] Significant log₂ ratio values (p<0.05) are shown from both the cDNA and Affymetrix microarrays. Baseline group is always the second group listed in the column heading. For example, values in the column "CD_NI-NC" are expressed as the log₂ ratio value of CD_NI with respect to NC group. A minus sign before the log₂ ratio value indicates down-regulation with respect to baseline group. A lone minus sign within a cell indicates that the probe was not significantly regulated. An asterisk indicates p<0.001 for that ratio. The number of samples in each group range as follows: NC (16-25), inflamed CD (29-36), non-inflamed CD (15-21), inflamed UC (16-25), non-inflamed UC (10-16), inflamed DC (4-9), non-inflamed DC (15-24).

4 Discussion

The objective of this study was to identify novel genes involved in the pathogenesis of IBD. To this end, genome-wide microarray expression screening was employed to identify mRNA transcript differences between normal controls and different IBD disease states. In a first step, the two major types of microarray platforms, namely cDNA clone-based and oligonucleotide-based arrays, were compared using the same starting material. This initial experiment determined the extent of compatibility between results received from different platforms. Having established that quantitative results from different microarray platforms could not be directly combined, results from different experimental approaches were analysed separately. Comparing expression profiles from normal controls and inflamed/non-inflamed CD or UC yielded thousands of differentially expressed genes. To place these differentially regulated genes in a tangible context, Gene Ontology (GO) categories that were over-represented by the transcripts were elucidated, leading to the identification of the following three broad functional themes in IBD: immune and inflammatory response; oncogenesis, cell proliferation and growth; and structure and permeability. Additionally, two specific themes, namely Wnt signaling pathway and cell-cell adhesion, were also highlighted by GO analysis. Over 50 transcripts involved in the specific themes and 15 transcripts involved in the broad themes were subject to real-time PCR quantitation in an extended cohort of IBD, disease controls and normal controls.

4.1 Microarray comparison

Microarray technology has rapidly evolved from high-density cDNA clone-based arrays to oligonucleotide-based arrays. In this study, these two main microarray platforms were compared using the same starting target RNA from human colonic mucosa biopsies. This study is the first to compare fluorescent and radioactively labeled platforms, probed with complex tissue, rather than homogenous cell lines (Mah et al., 2004).

Using full-length clone sequences and publicly released Affymetrix probe sequences, an in-depth analysis of the probe sequences and their corresponding expression levels was carried out. The results clearly showed that there was no rank order correlation of either 1) mean expression values from matched probes between both arrays, irrespective of the probe-matching methods (through

UniGene or BLAST search of the Affymetrix probe database); or 2) individual expression values for each patient within a platform. Moreover, the overlap in detection of the same genes on both platforms, in terms of concordant present/absent calls, was only 60-68%. Less than 8% of the discordant calls could be attributed to clone sequencing errors or differences in probe design for the same gene, suggesting that other factors influenced the discordant call rate.

All stages of the microarray experiment, from platform design, experimental conditions, spot quantitation, and further data processing, to the final gene expression values, may contribute to differences in expression results. The Affymetrix design bases its quantitation value on the difference in fluorescence between the 16 match and mismatch probe sets, whereas the clone-based system uses the hybridization signal from a single clone spotted in duplicate. Each method has its disadvantages. The use of a difference between match and mismatch as a quantitative measure is potentially vulnerable to single base changes due to polymorphism or sequencing errors in the original sequence used for oligonucleotide design. As it is possible for a mismatched oligonucleotide to produce 5–54% signal of a perfect match (Okamoto et al., 2000), cross-hybridization to mismatch oligonucleotides may result in an underestimation of gene expression. On the other hand, cDNA arrays are less sensitive to single base pair changes in the probe sequence because the spotted probe is longer (200–1500 bp), but they are more open to cross-hybridization (i.e. gene families) and may contain latent non-specific sequence (i.e., repetitive elements) since the cDNA clones are not completely sequenced. Compared with the Affymetrix arrays, which are directly synthesized on glass under controlled reaction conditions (McGall et al., 1996), cDNA probes may be subject to concentration variations due to printing effects and PCR amplification efficiency. The region of the gene represented by the probe may affect detected expression levels if the probes matched solely by UniGene cluster association detect different parts of the gene that are absent or poorly represented in target DNA, possibly due to alternative splicing or 3' bias in target labeling. Taken together, there are many opportunities for platform-specific biases to render cross-platform data incomparable.

The microarray comparison was limited to the analysis of baseline quantitation of biological replicates and did not attempt to compare the arrays' ability to detect changes in gene transcript levels. However, comprehensive analysis of the basic technique of hybridization alone shows that

there are many factors influencing microarray compatibility at a basal level, some of which may also influence differential gene analysis. Typically, microarrays are used to identify differentially expressed genes in an experimental setup that includes a group of arrays from two conditions (control and test condition), with the test and control samples being processed on the same array (in the case of two-color hybridization) or copies of the same array (as in the case of single-channel hybridizations). Usually, the differences are expressed as fold changes or log ratios for a single gene, rather than differences between different genes. Differences in a gene's expression, not explicit levels of gene expression, are the final values reported. In theory, ratios should be a more reliable output than baseline expression signals, because the ratio of test to control samples reduces the effect of systematic platform variables such as spotted probe concentration, printing effects, and probe sequence and length.

It should be noted that the aim of the microarray comparison was not to conclusively determine which platform outperformed the other. Such an endeavour would only be possible if microarray results could be compared to a "gold standard" on a large scale. As a cursory inquiry, absolute signal intensities from each array were compared to independent real-time PCR quantitation of 13 selected genes. The rank order of the expression levels from the cDNA and oligonucleotide platforms showed a strong positive correlation ($r_s = 0.75 - 0.90$) with the real-time PCR quantities in genes that were called "present" on the arrays. However, one cannot extrapolate these results, which were based on a small number of real-time assays, to conclude that this correlation is true for larger data sets. Contrary to other published comparisons, which matched cross-platform probes by gene name or homology to expressed sequence tags (ESTs), this study eliminated such variables as sequence ambiguity by basing the analysis on sequence-verified clones and fully sequenced clones, an important detail that was not adequately addressed in previous studies. No other microarray comparison has matched cross-platform microarray probes using the complete and exact probe sequence printed on the array. Additionally, this study did not restrict the analysis to a few selected genes of interest (i.e., highly up-regulated genes or those known to be involved in a specific process), but rather chose arbitrary genes in a non-hypothesis-driven fashion. Since large-scale analysis may involve whole gene sets, it is imperative to know whether cross-platform microarray data correlates as a whole, not just on a few select genes.

The results of the microarray comparison indicate that oligonucleotide-based arrays, such as those produced by Affymetrix, and full-length clone-based arrays may be too different in experimental design to give directly correlating results for the same genes. This suggests that microarray technologies should not be used as an absolute quantitation method and that pooling of global expression profiles from different microarray platforms for the purposes of large-scale data mining should be undertaken with caution. The observation that there is only moderate overlap in genes that are called "present" and no correlation in the expression levels from each platform necessitates complementary approaches to confirm findings solely based on microarray data (Mah et al., 2004).

4.2 Expression profiling in inflammatory bowel disease

The pathology of IBD is influenced by various factors, including genetic and environmental factors, which contribute to an inappropriate immune response in the lower gastrointestinal tract. To date, genetic linkage and association studies have identified sequence variants that confer disease susceptibility, with the strongest association being found in *CARD15* for CD and lesser associations in *SLC22A4* and *SLC22A5* for CD, *AGR2* and *MUC3* for UC, and *DLG5* and *NR1I2* for IBD (Dring et al., 2006; Hugot et al., 2001; Kyo et al., 1999; Peltekova et al., 2004; Stoll et al., 2004; Waller et al., 2006; Zheng et al., 2006). At the mRNA level, genome-wide expression screening has the potential to reveal disease-relevant genes by comparing expression profile snapshots of different disease states. Previous microarray expression studies in IBD have highlighted the dysregulation of the REG gene family (Dieckgraefe et al., 2000; Lawrance et al., 2001) and detoxification genes (Langmann et al., 2004).

In this thesis, two experimental approaches were used to investigate differential transcription profiles in IBD. In the first setup, cDNA microarrays were used to differentiate mRNA transcript expression between patients with active IBD and hospitalized normal controls. A particular advantage of this cohort was that all individuals in the cohort were accustomed to the hospital environment. In contrast, the second experimental setup compared active and non-active IBD hospitalized patients to healthy normal controls who were not currently hospitalized. However, it should be noted that both hospitalized normal controls and healthy normal controls were treated in the same manner throughout the colonoscopy procedure, except that the healthy normal controls

were given an enema if required. The patient cohort of the second experimental setup was advantageous due to the fact that macroscopically non-inflamed IBD samples compared to healthy normal controls could reveal IBD-specific effects independent of inflammation (CD_NI-NC or UC_NI-NC comparisons). Additionally, IBD-specific inflammation effects could be revealed by the comparison between inflamed and non-inflamed IBD (CD_I-CD_NI or UC_I-UC_NI), and cumulative effects between active IBD and healthy normal controls could be observed (CD_I-NC or UC_I-NC). Finally, the second experimental setup contained a number of inflamed and non-inflamed samples from the same patient, thereby eliminating noise due to inter-individual differences such as bacterial flora or genetic background.

The comparison between active IBD and hospitalized normal controls on cDNA microarrays identified 650 differentially regulated genes, with 500 and 272 regulated transcripts from the comparison between hospitalized normal controls and CD or UC, respectively. Interestingly, there was an imbalance in the proportion of genes that were dysregulated—approximately 84% of the differentially regulated genes were down-regulated in CD, compared to 42% down-regulated in UC. However, this result cannot be interpreted to infer that a broad up-regulation in UC and broad down-regulation in CD are distinguishing features of the two subtypes, since the genes found are highly influenced by the types and numbers of genes present on the particular microarray platform that was used. Of the overlapping 122 genes that were dysregulated in both CD and UC, none were found to be regulated in opposite direction.

Using Affymetrix-based arrays, the comparison of inflamed or non-inflamed IBD and healthy normal controls identified thousands of differentially regulated transcripts. The greater number of differentially regulated transcripts in this experimental setup is partly due to the fact that the oligonucleotide probes did not have to be verified by cDNA clone sequencing, which was a necessary and limiting step in the cDNA microarray protocol. There was also a markedly greater number of differentially regulated transcripts for the CD comparison than the UC comparison, since the CD dataset derived from two Affymetrix chips (U133A and B), while results were only available from chip U133A in the UC dataset. In all comparisons where IBD was compared to healthy normal controls, 63%-92% of the regulated transcripts were down-regulated compared to normal controls. However, in the comparisons between inflamed and non-inflamed within an IBD subtype, 58% of

the differentially regulated transcripts were up-regulated (with respect to non-inflamed) in both CD and UC, which suggests that disease-specific inflammation results in a broad up-regulation of genes, at least on these Affymetrix arrays (HG-U133). Of transcripts that were differentially regulated in both CD and UC for the three analogous comparisons (non-inflamed IBD vs. healthy normal controls, inflamed IBD vs. non-inflamed IBD, and inflamed IBD vs. healthy normal controls), only a handful of transcripts were regulated in the opposite direction. Of 1487 transcripts differentially regulated in both the non-inflamed IBD subtypes vs. healthy normal controls (CD_NI-NC and UC_NI-NC), only six genes were oppositely regulated (*C14orf43*, *ATP5F1*, *DLD*, *RAB5C*, *RRAS2*, *SRRM2*), but the fold changes were lower than 1.2-fold in at least one of the comparisons. With respect to disease-specific inflammation, there were 122 overlapping transcripts in the inflamed IBD vs. non-inflamed IBD comparison (CD_I-CD_NI and UC_I-UC_NI), and very interestingly, eight transcripts were oppositely regulated. *CHST5*, *HOXB13*, *LGALS2*, *WFDC2*, and *HOXB6* were differentially regulated with changes greater than 1.2-fold in their respective IBD subtype, while *S100A10*, *SPON1*, and *CDK2AP2* experienced changes lower than 1.2-fold. In the last comparison between inflamed IBD and healthy normal controls (CD_I-NC and UC_I-NC), 11 genes were oppositely regulated (*ASPN*, *GSN*, *PARP8*, *QSCN6*, *RAB13*, *RAB27A*, *RAB5C*, *RDH11*, *TFF1*, *FGFR2*, *PAPSS2*). However, these opposing microarray results should not be taken at face value, as one of these genes, *TFF1*, though it was regulated greater than 1.2 fold, was shown not to be oppositely regulated in CD and UC in real-time PCR experiments. Together with the cDNA microarray results, these results show that although there are genes that are differentially regulated in one subtype but not the other, diametrically opposite regulation of the same genes in CD and UC was quite rare, and in one case, not verifiable by real-time PCR in an extended cohort of samples. This observation supports the notion of a shared general inflammation profile underlying two clinically divergent forms of IBD.

4.2.1 Broad functional themes in the pathology of IBD

Interpreting the functional consequences of changes in gene expression observed in microarray expression screening is one of the major goals of exploratory microarray data analysis. One method used to attempt functional interpretation of microarray data is the use of annotation-based pathway databases such as Gene Ontology (GO). The use of GO does have its limitations, in particular the amount and quality of annotation (Fraser and Marcotte, 2004), but it does allow a

broad overview of terms to decipher gene pathways. In the present study, GO terms were supplemented with literature references to yield the classification of differential expression of genes from all microarray results. Differentially regulated genes were classified into three major groups: immune and inflammatory response; oncogenesis, cell proliferation and growth; and structure and permeability.

4.2.1.1 Immune and inflammatory response

The intestinal mucosa represents a key barrier between the hostile environment of pathogenic luminal antigens and internal tissues. It is no surprise that the IBD mucosa, compared to normal controls (healthy or hospitalized normal controls), exhibits differential regulation of many immune-related genes. In both subtypes, many genes which code for proteins with immunoglobulin receptor function, including (*BST2*, *IGH@*, *IGHG1*, *IGKC*, *IGLL1*, *IGKV1-5*, *IGLC2*, *IGHA1*, *IGHD*, *IGL@*, *IGHM*, *FCGR3B* and *CD79A*), were up-regulated in inflamed IBD compared to normal controls or non-inflamed IBD. Similarly, genes involved in antigen presentation (*HLA-A*, *HLA-B*, *HLA-E*, *HLA-DMB*, *HLA-DRA*, *PSMB10*) or anti-bacterial peptides (*DEFA5*, *DEFA6*) were largely up-regulated in inflamed IBD compared to non-inflamed IBD or normal controls. The acute phase response genes *C5* and *FTH1* were down-regulated, while *CFB* and *C2* were up-regulated in inflamed compared to non-inflamed samples. However, some of the aforementioned genes (*IGLC2*, *IGH@*, *IGHG1*, *DEFA5*, *DEFA6*, *HLA-A*, *HLA-E*, *HLA-DRA*, *CFB*, *C2*) did demonstrate up-regulation between non-inflamed IBD and normal controls, which may indicate that macroscopically non-inflamed mucosa maintained residual molecular traces of an immune response.

Of particular interest in this group, *cylindromatosis (turban tumor syndrome)* (*CYLD*) is a tumour suppressor and negative regulator of nuclear factor-kappa B (NF- κ B) (Neurath et al., 1996; Rogler et al., 1998; Schreiber et al., 1998), a key transcription factor which regulates adaptive and innate immune responses. *In vitro* investigations into the regulation of *CYLD* have revealed that *CYLD* is part of an autoregulatory control mechanism for TNFR-induced NF κ B activation. In this pathway, TNF up-regulates *CYLD*, which in turn leads to down-regulation of NF- κ B signaling (Jono et al., 2004; Yoshida et al., 2005). Recent *in vivo* studies have demonstrated that *CYLD* is a deubiquitination enzyme with different effects. In one study, *CYLD* regulates cell growth and proliferation by deubiquitinating Bcl-3 in keratinocytes, leading to degradation of Bcl-3 and preventing its co-

activation with NFkB (Massoumi et al., 2006). Another study showed that CYLD participates in T-cell development by regulating T-cell receptor signaling in T-cell clonal selection (Reiley et al., 2006). Despite its involvement with NFkB regulation, CYLD has not yet been examined in IBD. In the present study, microarray analysis showed that the tumour suppressor CYLD was down-regulated in CD compared to hospitalized normal controls and up-regulated in inflamed UC compared to non-inflamed UC. Real-time PCR analysis showed that CYLD was in fact down-regulated in both inflamed CD and UC compared to normal controls. Down-regulation was also observed in non-inflamed CD and UC compared to normal controls, showing that its regulation was dependent on disease but not on inflammation. Potentially, deficient amounts of CYLD transcript precede active inflammation and perpetuate inflammation by insufficient down-regulation of NFkB in IBD.

Another gene with a dual immune/tumour function is *deleted in malignant brain tumors 1 (DMBT1)*. DMBT1 was originally identified as a scavenger receptor with tumour suppressor properties, since its absence or reduction in expression was observed in different cancers (Braidotti et al., 2004; Wu et al., 1999). Two main functions for DMBT1 have been suggested, including mucosal defence and epithelial differentiation (Kang and Reid, 2003); however, to date, no role for DMBT1 has been suggested in the pathogenesis of IBD. Real-time PCR shows that DMBT1 transcript is strongly up-regulated (\log_2 ratio ≥ 2.76) in inflamed mucosa of any disease state (CD, UC or DC) compared to any non-inflamed mucosa (IBD or normal control). DMBT1 is also up-regulated to a lesser extent (\log_2 ratio ranges from 0.65 to 0.79) in non-inflamed IBD or disease controls compared to normal controls (Table 3.22, footnote), indicating that the up-regulation is evident in all disease states. On the microarrays, immunosuppressive treatment appears to reduce DMBT1 expression to normal levels, as no significant regulation is detected between immunosuppressive treated CD patients and non-inflamed CD or normal controls. Though there is no clear function for DMBT1, a peptide from a splice variant of DMBT1 was found to have the ability to bind various bacterial species (Bikker et al., 2002), suggesting its function as an innate immunity receptor. Alternatively, DMBT1 up-regulation in IBD may reflect an increase in epithelial restitution in response to damage, since a porcine splice variant of DMBT1, CRP-ductin, has been suggested as a receptor for trefoil factor (Thim and Mortz, 2000), which is highly expressed in the intestine and promotes healing (Taupin and Podolsky, 2003).

Carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) has roles including negative regulation of epithelial cell growth, differentiation, neutrophil activation and most interestingly, negative regulation of Th-1 mediated inflammation in a mouse colitis model (Chen et al., 2004; Iijima et al., 2004; Izzi et al., 1999; Rubel et al., 2001). The microarrays detected up-regulation of CEACAM1 in inflamed CD or UC compared to normal controls and in non-inflamed UC compared to normal controls (Table 3.22). Real-time PCR detected significant up-regulation of CEACAM1 transcript in inflamed CD, but interestingly, not in inflamed UC compared to normal controls (Table 3.22, footnote). Up-regulation of CEACAM1 was also detected in non-inflamed CD/UC compared to healthy controls (Table 3.22, footnote). Conversely, one microarray probeset (209498_at) and real-time PCR showed down-regulation in inflamed UC compared to non-inflamed UC. Consistent detection of CEACAM1 transcript was complicated by the fact that multiple splice variants of CEACAM1 were detected by the microarray and real-time probes; however, there was consistent up-regulation in CD (inflamed or non-inflamed) compared to normal controls, thus pointing to unique up-regulation due to the CD condition. In addition, immunohistochemical staining of mucosal sections (Figure 3.12, Panel A) showed that infiltrating immune cells in the lamina propria were stained by anti-CEACAM antibody in CD, but not in UC, further suggesting the importance of CEACAM1 in the pathology of CD. As the mechanism by which CEACAM1 regulates T-cells does not occur at the transcript level, but rather involves phosphorylation of the protein and/or interaction with other CEACAM molecules (Chen et al., 2004; Iijima et al., 2004), increased transcript levels could result in an over-expression of the short form of CEACAM1 protein, which is a co-stimulatory receptor in T-cell regulation, over the long form of CEACAM1, which is an inhibitory receptor. As a consequence, the short form could dominate over the long form of CEACAM1, resulting in a Th-1 cell driven response in CD.

4.2.1.2 Oncogenesis, cell proliferation and growth

The intestinal epithelium is one of the most rapidly self-renewing tissues in the body, making it more susceptible to malignant transformation. Inflammation is characterized by expression of chemotactic factors, growth factors, and proteases which facilitate leukocyte migration and healing at the site of injury. The normal intestinal epithelium is under a state of controlled inflammation, such that an immune response is invoked by pathogenic antigens, yet commensal bacteria are

tolerated. Within the context of IBD, abnormal processes in tissue renewal, either through normal epithelial maintenance or tissue injury, and chronic inflammation, increase the potential for cancer to develop. Dysregulation of genes, which either promote tumour progression or ultimately fail to suppress tumours, is of particular interest in IBD, as there is an increased risk of developing colorectal carcinoma in UC (Brostrom et al., 1987; Ekblom et al., 1990). Tumour progression genes in this group include *TIMP1* and members of the S100 calcium binding protein family (*S100A6*, *S100A11* and *S100P*), whose expression promotes cell proliferation and survival and are associated with tumor progression in pancreatic and colon cancers (Arumugam et al., 2004; Bronckart et al., 2001; Holten-Andersen et al., 2004; Komatsu et al., 2002; Melle et al., 2005). These genes show an up-regulation on the microarrays, primarily in inflamed UC compared to non-inflamed controls. The up-regulation of *TIMP1* expression in inflamed IBD compared to normal controls was confirmed by real-time PCR (Figure 3.10).

Trefoil factor 1 (*TFF1*) has both tumour suppressor and tumour progression properties (Emami et al., 2004), in addition to a regenerative role in the gut, which specifically promotes epithelial cell migration without apoptosis in response to injury (Vieten et al., 2005). Microarray expression results from both cDNA and Affymetrix arrays show up-regulation in *TFF1* in inflamed UC compared to the respective control group, while in CD, significant down-regulation was detected in CD compared to healthy normal controls on the Affymetrix arrays but not on the cDNA arrays (Table 3.22). In the real-time PCR cohort comparing active IBD to hospitalized normal controls, the up-regulation in active UC was confirmed (Figure 3.10), while no regulation was observed in the CD or disease controls. This may suggest that *TFF1* expression in inflamed UC mucosa may play a specific role in *TFF1* tumour progression, in accordance to a previous report showing that *TFF1* protein expression is significantly over expressed in UC-associated colorectal adenocarcinoma compared to normal mucosa (Hirota et al., 2000).

Finally, within this category, the Wnt pathway is of interest since inappropriate activation is a common signaling abnormality in human cancers (Ergun et al., 2000; Hauck, 2002). As members of the casein kinase I gene family (*CSNK1*) have been implicated in the regulation of Wnt-targeted gene expression (Amit et al., 2002; Liu et al., 2002), both transcript levels and immunohistochemical protein localization were examined in IBD biopsies. Real-time PCR results

indicate that CSNK1D was down-regulated in all UC comparisons, while in CD, down-regulation was only observed for comparisons involving inflamed CD (Table 3.24). Immunohistochemistry staining of the mucosa (Figure 3.12, Panel C) showing basolateral staining of CSNK1D in normal and active UC epithelial cells, but only faint staining in the crypts of CD, may imply a role for Wnt signaling, especially in UC pathogenesis.

4.2.1.3 Structure and permeability

The intestinal epithelial barrier, in terms of a physical barrier of surface glycocalyx and a layer of epithelial cells, prevents luminal contents, such as bacteria and viruses from transversing the epithelial layer. Impaired barrier function makes the gut vulnerable to microbial invasion and invokes an immune response. In some respects, IBD can be thought of as being caused by chronic inflammation produced by a vicious circle of barrier breach and immune response, perpetuated by inefficient healing and re-infection by luminal contents (Clayburgh et al., 2004; Korzenik, 2005). Several genes in this category were ubiquitously regulated in both IBD and non-IBD samples, reflecting known gene dysregulation in disease where inflammation and wound healing are recurrent events. These include paracellular permeability (down-regulation of OCLN, Figure 3.10) (Kucharzik et al., 2001), degradation of extracellular matrix (up-regulation of TIMP1 and MMP2, Figure 3.10 and Table 3.3, respectively) (Massova et al., 1998; von Lampe et al., 2000) and barrier protection against bacterial invasion of the epithelial surface (up-regulation of MUC1, Figure 3.10) (Carraway et al., 2003).

Furthermore, the present study identified several genes (*CDH11*, transcripts represented by N48794 and AF087994, *SPINK5*, *PDLIM5* and *ROCK1*) that were novel within the context of IBD. Members of the cadherin superfamily, integral membrane proteins that mediate calcium-dependent cell-cell adhesion, have been shown to be involved in epithelial cell migration and resealing the site of tissue damage in the intestinal mucosa (Wilson and Gibson, 1997). Up-regulation of cadherin 11 (*CDH11*) in both IBD subtypes, but not in inflamed non-IBD tissue (Table 3.26), furthers speculation that this member of the cadherin family could also be involved in IBD-specific restructuring processes in the intestinal mucosa. Continuing with the structure theme, two uncharacterized transcripts (GenBank accession numbers N48794 and AF087994), which were assigned a role in

cell-cell adhesion (one of them containing cadherin repeats), were also found to be differentially regulated (Table 3.4).

Genes which are involved in extracellular matrix maintenance or cell-cell adhesion have the potential to impair barrier function when dysregulated. One such gene is *serine peptidase inhibitor, Kazal type 5 (SPINK5)*, which borders the IBD5 locus on 5q31. IBD5 contains many susceptibility genes, including clusters of cytokines and transporters, and an association to CD was shown (Rioux et al., 2001), though no one causal gene has been identified. SPINK5 has been postulated to provide anti-microbial protection of the mucosal epithelia (Magert et al., 1999), regulate differentiation of T-lymphocytes, or regulate extracellular matrix remodeling (Chavanas et al., 2000). Already implicated in Netherton Syndrome (NS), the lack of SPINK5 allows unrestricted serine protease activation and eventual degradation of the stratum corneum, resulting in impaired barrier function (Hachem et al., 2006). In the present study, the microarrays showed that transcription of this gene was down-regulated in both non-inflamed IBD subtypes compared to normal controls; however, real-time PCR on an extended patient cohort confirmed differential regulation in UC, but not in CD or disease controls (Table 3.22, footnote). There was no difference detected between inflamed UC and normal controls, but SPINK5 transcript was down-regulated in non-inflamed UC compared to normal controls. A down-regulation in SPINK5 could possibly allow unlimited digestion of the extracellular matrix, thereby compromising the epithelial barrier and allowing bacteria to induce prolonged inflammation responses.

Cell migration in response to wound healing is regulated in part by cell adhesion processes, such as Rho-ROCK-mediated cytoskeletal reorganization (Kuroda et al., 1996; Nakagawa et al., 2000). Interestingly, members of this pathway, ROCK1 and PDLIM5, were down-regulated in both IBD subtypes (but not in inflamed DC) in the real-time PCR experiments using hospitalized normal controls, which might indicate a potential decline in cell migration and an impaired ability to maintain epithelial integrity in IBD. Furthermore, immunohistochemistry staining of PRKCB1, which has been shown to interact with PDLIM5 (Kuroda et al., 1996), showed that lamina propria cells were preferentially stained in UC, but not CD or normal controls, possibly indicating a unique role for this protein in UC. Interestingly, a genetic variant in *DLG5*, which was recently associated with IBD (Stoll et al., 2004), is a PDZ-containing scaffolding protein that associates with beta-catenin

and is a binding partner of vinexin at cell-cell contacts (Wakabayashi et al., 2003). PDLIM5, in addition to its signaling function by interacting with PRKCB1 through its LIM domain (Kuroda et al., 1996), also contains PDZ domains, which may be used as sites for interaction of scaffolding proteins. These findings might represent interesting targets for further functional characterization in the context of wound healing and regeneration.

4.2.2 Specific themes in the pathology of IBD

From exploratory data analysis of GO terms represented by differentially regulated genes between inflamed/ non-inflamed CD and UC or normal controls, many GO terms were found to be over-represented by the differentially regulated genes. Of particular interest were GO terms that were not directly related to immune function. Two such specific themes, namely Wnt pathway and cell-cell adhesion, were chosen for further analysis by real-time PCR.

4.2.2.1 Wnt pathway

The role of the Wnt pathway in the context of IBD is interesting in two respects. Firstly, the Wnt pathway controls intestinal epithelial cell proliferation, which is essential for development and renewal processes in the colon (Batlle et al., 2002; van de Wetering et al., 2002). Secondly, aberrations in the Wnt pathway have been shown to cause cancer (Moon et al., 2004), and as there is an elevated risk for colorectal cancer in UC, there is the possibility that the Wnt pathway could be involved. To date, one publication has used cDNA macroarrays to identify up-regulation of five Wnt transcripts (FRZB, FRZ5, FZD3, DVL3, SFRP2) in a single UC sample compared to one CD sample (Uthoff et al., 2001). In contrast, the present study is the first to report differential expression of 28 Wnt pathway transcripts, as measured by real-time PCR, in a larger number of CD and UC samples, as well as disease and normal controls. Since the anomalous expression of gene transcripts at any stage of Wnt signaling have the potential to influence Wnt activation, the potential contributions of each gene are discussed in the following groups: 1) Wnt ligand modifiers or co-receptors; 2) Wnt receptors and ligands; 3) cytoplasmic components of Wnt signaling; 4) nuclear components of Wnt signaling; and 5) Wnt target genes.

Genes involved in processes preceding the initiation of Wnt signaling include those involved in ligand processing of Wnt ligands, heparin sulfate proteoglycan (HSPG) synthesis, and transcription factors involved in modulating heparin sulfate proteoglycan levels. Heparin sulfate proteoglycans may participate in Wnt signaling by acting as low-affinity co-receptors for Wnt ligands, facilitating transport or stabilizing Wnt ligands (Hacker et al., 2005; Logan and Nusse, 2004; Wodarz and Nusse, 1998). Wnt activation can be promoted by the heparin sulfate proteoglycan protein syndecan (Alexander et al., 2000; Munoz et al., 2006), which binds ADAM12. Although ADAM12 has primarily been characterized in its role for ADAM12/syndecan-4-mediated cell adhesion in mesenchymal cell lines (Iba et al., 2000; Thodeti et al., 2003), its specific up-regulation in IBD (but not disease controls) compared to non-inflamed groups (normal control or non-inflamed IBD) in real-time PCR (Table 3.24) has the potential to affect Wnt pathway activation by engaging more syndecan molecules for ADAM12/syndecan signaling and leaving less syndecan available to act as a co-receptor for Wnt activation. The transcription factor FOXL1 has been shown to be involved in gastrointestinal cell proliferation and negative regulation of the Wnt pathway by suppressing the expression of HSPGs, which act as co-receptors for Wnt ligand binding (Perreault et al., 2001). Real-time PCR noted a down-regulation of FOXL1 in IBD (inflamed or non-inflamed) or inflamed DC compared to normal controls (Table 3.24). This may suggest that in IBD, regardless of inflammation status, FOXL1 fails to negatively regulate HSPG expression, which may lead to an increased HSPG expression and higher efficiency for Wnt activation. However, the expression of the HSPG transcript syndecan-1 (SDC1) was confirmed by real-time PCR to be down-regulated in inflamed IBD or DC compared to normal controls (Table 3.24). Considering the fact that the mechanism of FOXL1 regulation of proteoglycan expression has not been fully elucidated but is thought to occur at the post-transcriptional level (Perreault et al., 2001), it is possible that the observed down-regulation of SDC1 cannot directly be associated with diminished FOXL1 transcript levels.

Initiation of Wnt signaling, be it in the canonical or non-canonical Wnt pathway, requires binding of Wnt ligands to the transmembrane receptor complex (LRP5/6 and Frizzled family members). Once secreted from the initiating cell into the extracellular space, Wnt ligands can be modulated by the secreted frizzled related protein family (sFRPs), which have been shown to either promote or inhibit Wnt activation, depending on the Wnt ligands present, the tissue specificity, and the concentration

of sFRPs (Kawano and Kypta, 2003). Down-regulation of SFRP1 or deletion of chromosomal regions (2q) containing FRZB have been observed in many cancers (Ko et al., 2002; Schmitt et al., 2002; Ugolini et al., 2001; Zhou et al., 1998), potentially because SFRP proteins can promote apoptosis (Kawano and Kypta, 2003). Frizzled related protein (FRZB) was shown to be down-regulated in inflamed IBD compared to normal or non-inflamed IBD by real-time PCR (Table 3.24), while on the microarrays, SFRP1 was down-regulated in UC (inflamed or non-inflamed) but not in CD compared to normal controls (Table 3.23). In contrast, real-time PCR indicated that SFRP2 was specifically up-regulated in inflamed IBD, but not inflamed DC compared to normal controls (Table 3.24). This result is compatible with the observation that SFRP2 has been shown to inhibit Wnt (Kawano and Kypta, 2003), as well as SFRP1 (Yoshino et al., 2001).

The diversity of Wnt functions in various tissues is achieved through the binding various combinations of Wnt ligands to Frizzled-LRP4/5 transmembrane receptor complex (Miller, 2002). Though many studies have examined ligand-receptor associations in developing *Drosophila* and *C. elegans* cells, there is a paucity of information on the function and specificity of the intact Wnt-receptor interactions in mammalian cells (Dale, 1998; Takada et al., 2005) and especially in adult colon, partly due to the functional redundancy of the ligand-receptor interactions and the requirement for conditional activation methods to single out specific interactions (Kuhnert et al., 2004). Nevertheless, real-time PCR detected a down-regulation of transcripts for FZD1, FZD2, and FZD7 in normal controls compared to inflamed IBD or disease controls (Table 3.24), suggesting that down-regulation of these receptor transcripts is likely influenced by inflammation and not by IBD alone.

Cytoplasmic members of the canonical Wnt signaling pathway were primarily down-regulated in inflamed IBD compared to normal controls or even non-inflamed IBD, as shown by real-time PCR experiments. However, the overwhelming down-regulation does not unequivocally support a decrease or increase in Wnt activation, since many of the involved genes affect the pathway by different means. For example, members of the destruction complex, casein kinase I and GSK3, phosphorylate β -catenin, thereby targeting it for protein degradation. The observed down-regulation of two casein kinase I family members (CSNK1D, CSNK1E), in inflamed IBD compared to normal controls (Table 3.24), may indicate that inflammation has the potential to transduce (i.e., positively

regulate) Wnt signaling by reducing the amount of free casein kinase I and decreasing the phosphorylation of β -catenin that targets it for degradation. The observed down-regulation of AXIN1 in inflamed tissue only (both IBD and disease controls) in real-time PCR (Table 3.24) may be interpreted as positive regulation of the Wnt pathway, since axin forms the scaffolding structure for the destruction complex. As a reduction in axin transcript could reduce its availability to the destruction complex, the functionality of the destruction complex could be impaired, leading to an increase of β -catenin, translocation of β -catenin to the nucleus and transcription of Wnt target genes. APC, along with axin, forms the protein scaffolding to which other elements of the destruction complex bind, and also functions to chaperone nuclear β -catenin to the destruction complex (Giles et al., 2003). Down-regulation of APC in inflamed IBD and DC compared to normal controls (Table 3.24) could be interpreted as a positive effect on the Wnt pathway, since a reduction in APC could impair effective degradation of β -catenin. On the other hand, a decrease in disheveled 1 (DVL1) in all disease states compared to non-inflamed tissue (normal, IBD, or DC; Table 3.24) may lead to negative regulation of Wnt signaling, since dishevelled is required to transduce Wnt ligand activation between the Frizzled receptor and the destruction complex through axin (Moon et al., 2004). Moreover, the observed decrease in β -catenin in all IBD states compared to normals in real-time PCR (Table 3.24) could suggest that the concentration of the co-transcriptional activator is reduced, thus limiting transcription of Wnt targets.

The expression of the TCF/LEF transcription factors, as measured by real-time PCR, did not reach a clear consensus on Wnt pathway activation or repression. Transcription factor TCF7L2 (commonly known by unofficial name Tcf-4), the most prominently expressed TCF transcription family member in colon epithelium (Korinek et al., 1997), was seen to be down-regulated in inflamed IBD compared to normals, while its co-transcription factor LEF1 was up-regulated in inflamed IBD compared to non-inflamed IBD tissue (Table 3.24). Both TCF7L2 and LEF1 were specifically regulated in IBD, as no regulation was observed in the disease control group. TCF7, which can act as an antagonist of TCF7L2 (Giles et al., 2003), was up-regulated in inflamed CD compared to normals (Table 3.24), thus supporting repression of Wnt activation. Interestingly, the differential expression of the antagonist TCF7 (commonly known by unofficial name Tcf-1) in CD but not in UC, may protect the CD colon by inhibiting Wnt target gene transcription to a greater extent than in the UC colon, where TCF7 expression is not elevated. Overall, the pattern of TCF

family expression supports repression of Wnt target genes, while LEF1 expression supports transcription of Wnt target genes.

All of the previously described gene transcript regulations, from upstream of Wnt processing to nuclear transcription of TCF/LEF, culminate in the transcription initiation of Wnt target genes. Many target genes of canonical Wnt pathway transcription have been identified from studies investigating β -catenin/TCF signaling in colorectal and other cancers (Giles et al., 2003). Some of these target genes have clear roles in intestinal epithelial homeostasis and renewal, such as cell proliferation mediated by *MYC* and *PPARD* (He et al., 1999; van de Wetering et al., 2002) and cell migration and proliferation mediated by ephrin receptors *EPHB2* and *EPHB3* (Holmberg et al., 2006). Microarray results showed that transcripts of these aforementioned genes were down-regulated in inflamed IBD compared to normals (Table 3.23), which could imply that the inflamed IBD colon has impaired ability to regenerate, resulting in barrier breach by bacteria and subsequent chronic inflammation.

Interestingly, a recent study showed that bone marrow transplant lead to a reduction in disease severity in an IL10 knockout colitic mouse model (Bamba et al., 2006). The improvement was attributed to the localization and differentiation of bone-marrow derived cells to colonic subepithelial myoblasts (SEMFs), which are located in the mesenchyme surrounding inflamed colon crypts. The authors suggested that the improvement was due to the cytokine-mediated effector function of the colonic SEMFs. In the same vein, it is also possible that the SEMFs could improve colon regeneration by providing proper Wnt signals to the proliferating colon crypts, as Wnt signaling to the colon crypts originate from the mesenchyme (Gregorieff et al., 2005).

However, it is difficult to discern whether or not other genes involved in Wnt-related cancer progression also have a specific role in IBD, since many of the differentially regulated Wnt targets are also transcription regulators (e.g. *JUN*, *MYC*, *TCF4*, *TLE*, *VANGL2*, *FOSL1*, *PPARD*) and affect genes further downstream of Wnt, or are cross-regulated by other pathways. For example, *MYC* induces the expression of *EMP3*, which is thought to be involved in cell proliferation and cell-cell interactions (Ben-Porath et al., 1998), and *JUN* may be cross-regulated by its interaction with *TJP2* (Betanzos et al., 2004). In at least two instances, immune responses are likely to represent

dominant control of two Wnt targets. IL8 was shown to be inducible by Wnt/ β -catenin pathway and is up-regulated in many cancers, including colorectal cancer (Levy et al., 2002). As an inflammation marker, IL8 is also regulated by inflammatory response mediated by Toll-like receptor signaling (Yoshida et al., 2005). In another example, TCF4 (also known as immunoglobulin transcription factor-2) was also shown to be induced by the β -catenin /TCF to promote cell growth in neoplastic transformation (Kolligs et al., 2002), but since TCF4 binds to an immunoglobulin enhancer (Gstaiger et al., 1996; Henthorn et al., 1990) it is likely to be strongly involved in antibody expression. Therefore, the observed up-regulation of IL8 and TCF4 in inflamed IBD compared to normal controls from real-time PCR (Table 3.24) may be primarily due to immune response rather than Wnt activation. Interestingly, three of the downstream targets of Wnt were shown to be exclusively regulated in IBD and not in disease controls (real-time PCR result): TCF4 was specific for IBD; EMP3 and JUNB were specific for UC only (Table 3.24). These results may implicate a specific role for these transcripts in IBD pathogenesis. However, within the context of IBD, the complexity of pathway interactions indicates that more research is needed to elucidate the downstream roles of other Wnt targets, many of which are transcription regulators themselves.

To date, it is not conclusive whether disruption of the Wnt pathway can lead to a higher rate of colorectal cancer in IBD patients, and in particular UC patients, since *APC* and *CTNNB1* mutation detection studies have not conclusively demonstrated the involvement of the Wnt pathway in UC-related colorectal carcinomas (Aust et al., 2002). Overall, of the 28 Wnt pathway transcripts monitored in real-time PCR analysis (Table 3.24), 64% of the genes were differentially regulated in inflamed UC compared to non-inflamed tissue (UC or normal control). In contrast to UC, only 46% of the Wnt transcripts were differentially regulated in inflamed CD compared to non-inflamed tissue. Despite the lack of mutation detection results in two key genes of the Wnt pathway, differential transcript regulation of many different Wnt pathway genes may indicate that overall, the UC condition has a predisposition towards cancer, since more Wnt transcripts are dysregulated in UC than in the CD condition.

4.2.2.2 Cell-cell adhesion

An increase in intestinal permeability—be it the result of stress, genetic predisposition, or bacterial environment—has been suggested as an exacerbating factor to the pathology of IBD (Nazli et al., 2004; Soderholm et al., 1999; Soderholm et al., 2002; Yang et al., 2006). Broadly speaking, permeability in the intestinal epithelium may refer to surface barrier protection (glycocalyx), permeability between cells (paracellular) or permeability through cells (transcellular). Here, discussions relating to barrier function will be focused on junctions between cells, particularly on genes involved in cell-cell adhesion interactions.

The coordinated interaction of proteins involved in cell-cell junctional complexes, which include the tight junction, adherens junction and desmosome, maintain the integrity of the intestinal epithelium, while selectively regulating paracellular traffic (Sawada et al., 2003). The tight junction (TJ) is located most apically in the polarized intestinal epithelial cell and forms a semi-permeable ring that seals the gap between neighbouring cells and further serves to separate basolateral from apical membrane components. Three transmembrane proteins, including claudin (CLDN), occludin (OCLN) and junctional adhesion molecule (JAM), form the tight junction. Within the claudin family, different dysregulation patterns were observed, which would suggest that claudin members are regulated by different mechanisms. CLDN3, 4, and 7 were confirmed in real-time PCR to be consistently down-regulated in inflamed IBD compared to normal controls, while CLDN1 and 2 were strikingly up-regulated (ranging from a \log_2 ratio of 1.27 to 5.30) in the same comparison (Table 3.26). The differential expression of the claudins was dependent on inflammation, as inflamed disease controls mirrored the results seen in inflamed IBD. These results are consistent with a recent study, which showed that CLDN2 protein expression was significantly increased in the colonic epithelium of active UC and CD, whereas CLDN3 and 4 were sporadically reduced or redistributed compared to normal controls, a difference which the authors suggest may be due to differential responses to IL13 and TNF/IFNG, respectively (Prasad et al., 2005). The study further demonstrated that both the up-regulation of CLDN2 and down-regulation of CLDN3 and 4 were associated with increased permeability in a model system (Prasad et al., 2005), indicating that the net effect of CLDN dysregulation lead to barrier dysfunction.

Another tight junction transmembrane component, occludin (OCLN), was down-regulated in real-time PCR in all disease states compared to non-inflamed samples (Table 3.26), suggesting that dysregulation of this gene was wholly dependent on disease condition (either IBD or disease control), rather than inflammation. The effect of inflammation on OCLN expression is supported by a study showing that immunofluorescence labeling of OCLN was reduced in the colon epithelium of active CD and UC patients (Kucharzik et al., 2001); however, the same study reported no differences in immunofluorescence labeling between normal controls and inactive IBD. As OCLN transcript expression can be negatively regulated by TNF or IFNG (Mankertz et al., 2000) and TNF is up-regulated in IBD (Breese et al., 1994), these results may suggest that in the inactive IBD mucosa, underlying events at the transcriptional level have not yet translated into an effect detectable by immunofluorescence microscopy.

The third major component of tight junctions, the JAM proteins, are single-transmembrane proteins which recruit other proteins (claudins, occludin, tight junction protein family and partitioning defective family members) to the tight junction through its PDZ domains (Ebnet et al., 2004; Mandell and Parkos, 2005). One JAM family member, F11R, was observed to be down-regulated on the microarrays in inflamed or non-inflamed CD compared to normal controls, and no regulation was observed in UC (Table 3.25). The CD-specific dysregulation may implicate a role for F11R in the pathogenesis of CD, through TJ disruption and additionally through leukocyte migration events, since F11R is also expressed on leukocytes and participates in transendothelial migration (Ebnet et al., 2004). The specificity of this interaction, however, remains to be confirmed by additional studies.

While the transmembrane claudins, occludin and JAM proteins seal the extracellular space between epithelial cells, a network of cytoplasmic proteins provide scaffolding for the transmembrane proteins and transmit signals between the tight junction and actin cytoskeleton (Matter and Balda, 2003; Mitic et al., 2000). In particular, the tight junction protein family (TJP or commonly known as zonula occludens (ZO)), which is a member of the membrane associated guanylate kinase family (MAGUK), contains PDZ domains and a guanylate kinase-like (GUK) domain, which interact with cytoplasmic domains of claudin and occludin, respectively (Itoh et al., 1999). In real-time PCR, TJP1 and TJP3 exhibited consistent down-regulation in inflamed IBD and

disease controls when compared to normal controls (Table 3.26). In contrast, TJP2 was exclusively up-regulated in IBD. TJP2 was up-regulated in inflamed UC compared to normal controls and up-regulated in inflamed IBD compared to non-inflamed IBD (Table 3.26). The up-regulation of TJP2 in inflamed IBD tissue may represent a functional redundancy to attempt to compensate for the reduction in TJP1, as it has been shown that TJP2 is recruited to tight junctions in TJP1-deficient cells (Umeda et al., 2004). Alternatively, TJP2 may function in inflamed IBD as a regulator of cell proliferation. TJP2 has been shown to act as a messenger molecule between the TJ and the nucleus, since TJP2 has been associated with transcription factors in the nucleus upon loss of cell-cell contact (Betanzos et al., 2004; Islas et al., 2002).

Formation of the tight junction is thought to be preceded by the formation of the adherens junction, which is initiated by CDH1 (E-cadherin) binding to extracellular domains of other CDH1 molecules in adjacent cells in a Ca^{2+} dependent manner (Sawada et al., 2003). Through its cytoplasmic domains, CDH1 interacts directly with members of the catenin family (beta, gamma and delta catenin (p120)), while vinculin (VCL) and α -catenin link the cadherin-catenin complex to the actin cytoskeleton (Gumbiner, 2005; Zbar et al., 2004). Other cytoplasmic proteins, such as TJP1 and JAM, are also recruited to the developing adherens junction, and subsequent recruitment of occludin, claudin, partitioning defective family members (PAR3 and PAR6) and atypical protein kinase C (PRKCI; common name: aPKC) lead to maturation of the intercellular junction and segregation of the apical junction complex into adherens and tight junctions (Sawada et al., 2003). Within the context of IBD, cadherins CDH1 and CDH11 demonstrate opposite behaviours in real-time PCR. CDH1 is down-regulated in all inflamed disease conditions (CD, UC and DC) compared to normal controls, while CDH11 is specifically up-regulated in inflamed IBD but not in inflamed disease controls (Table 3.26). Similar to CDH1, the actin linking proteins α - β -catenin (CTNNA1, CTNNB1) and vinculin (VCL) are down-regulated in inflamed disease conditions compared to normal controls in real-time PCR, an effect which is also evident when non-inflamed IBD is compared to normal controls (Table 3.26). The down-regulation of CDH1, CTNNA1 and CTNNB1 in inflamed mucosa may be indicative of cell adhesion disruption resulting from cell migration, since epithelial restitution is the primary repair mechanism induced in response to ulcerations in the gastrointestinal mucosa (Zbar et al., 2004). Furthermore, the down-regulation of these genes in non-inflamed CD or UC mucosa compared to normal controls may point towards a general defect

in the IBD mucosa to produce sufficient amounts of these proteins in the absence of inflammation. Interestingly, in stark contrast to CDH1, CDH11 is specifically up-regulated in inflamed IBD compared to normal controls. Though there are no studies specifically addressing the role of CDH11 in the gut mucosa, a recent study in a murine fibroblast cell line has shown that CDH11 is a cellular motility factor that regulates remodeling of cell-cell contacts and the actin cytoskeleton (Kiener et al., 2006). It is therefore conceivable that the up-regulation of CDH11 and the down-regulation of CDH1, CTNNB1 and CTNNA1 work together to promote epithelial cell restitution in the inflamed IBD mucosa.

The fact that components of the cell-cell adhesion interaction are also associated with the Wnt pathway suggests that there could be cross-talk between cell-cell adhesion mechanism and the Wnt pathway. β -catenin and APC are both multifunctional proteins that interact with both the Wnt pathway and cell-cell adhesion. In the Wnt pathway, APC forms part of the degradation complex that inactivates the Wnt pathway by targeting cytoplasmic β -catenin for proteomic degradation. In the role of cell-cell adhesion, β -catenin links the adherens junction to the actin cytoskeleton through E-cadherin and α -catenin. APC has been shown act as a shuttling chaperone for β -catenin in and out of the nucleus (Henderson and Fagotto, 2002). APC has also been localized to various parts of the cell (cytoplasm, microtubule tips, plasma membrane, nuclei) (Bienz, 2002). Experiments in a human colon cancer cell line have also shown that restoration of APC expression in a APC-mutated cell line improved cell-cell adhesion. Based on these results, it has been proposed that APC interacts with β -catenin to promote translocation of E-cadherin (CDH1) and β -catenin to the cell membrane (Faux et al., 2004). Though APC and β -catenin transcripts are regulated in the inflamed IBD and DC (Table 3.24 and Table 3.26), it is currently not clear whether inappropriate cross-talk between the Wnt pathway and cell-cell adhesion processes contribute to IBD pathogenesis.

Overall, many genes involved in tight junction and adherens junction were shown to be dysregulated at the transcriptional level in inflamed or diseased colonic mucosa biopsies compared to normal controls. Tight junction genes, including CLDN1,2, and 7, displayed inflammation-dependent dysregulation, whereas occludin showed disease-dependent down-regulation. Cytoplasmic proteins associated with the tight junction, including TJP1, 2, and 3, were dysregulated

in inflamed IBD compared to non-inflamed conditions. Collectively, these results support the notion that inflammation and disease dysregulate the expression of tight junction components, which can disrupt the function of the tight junction. Finally, the differential expression of adherens junction components, CDH1, CDH11, VCL, CTNNA1, and CTNNB1 suggest that adherens junction disruption may be part of the epithelial cell restitution process at work in IBD.

4.3 Conclusions

Genome-wide microarray expression profiling is increasingly being used as a screening tool for hypothesis generation. Examples of the two main types of microarray platforms, mainly the cDNA and oligonucleotide platforms, were compared by hybridizing the same RNA samples on each platform. Contrary to previously published microarray comparisons, the present study was the first to match like-probes from different microarray platforms using sequence-verified probes.

The present study demonstrates for the first time that oligonucleotide-based arrays and full-length clone-based arrays may be too different in experimental design to give directly correlating results for the same genes. Pooling of global expression profiles for different platforms for the purpose of meta-analysis should be undertaken with caution, and results obtained from microarrays should be confirmed by complementary approaches.

A thorough analysis of the microarray comparison data showed that there was no rank order correlation of mean expression values from matched probes between both arrays, and that individual expression values for each patient within a platform showed no rank order correlation. Furthermore, the overlap in detection of the same genes on both platforms, in terms of concordant present/absent calls, was only 60-68%. Detailed analysis moreover showed that less than 8% of the discordant calls were caused by clone sequencing errors or differences in probe design for the same gene. Comparison of microarray expression levels with that of real-time PCR indicated that the data from both microarray platforms did display a strong positive rank order correlation, however, this was based on a small number of genes (13 genes) and was not a reliable indicator for larger datasets.

The present study also provided a whole genome snapshot of molecular processes taking place at the transcriptional level in the active and inactive IBD colonic mucosa. The transcription profiles of colonic mucosal biopsies from CD, UC and control individuals were obtained using two different microarray platforms. Keeping the results of the microarray comparison in mind, the data from each microarray experiment was analysed separately, and many of the key results were subject to real-time PCR quantitation in an extended cohort of over 300 individuals.

The main findings of these microarray analyses point to genes involved in immune and inflammatory response (*CYLD*, *DMBT1*, *CEACAM1*); oncogenesis, cell proliferation and growth (*TIMP1*, *CSNKD1*); and structure and permeability (*PDLIM5*, *ROCK1*, *CDH11*, *SPINK5*, *PRKCB1*). All of these genes have some known functions but have not been previously described in the context of inflammatory bowel disease. Based on microarray results and references in the literature, further verification was carried out in over 50 genes in two more specific subject areas: Wnt pathway and cell-cell adhesion. The differential regulation (detected by real-time PCR) of almost all of the cell-cell adhesion genes and 85% of the Wnt pathway members was influenced by inflammation. Of particular note were members of the TCF family of transcription factors in the Wnt pathway, namely *TCF7* and *TCF7L2*, which exhibited distinctly different patterns of expression between CD and UC, with the end result favoring Wnt transcription factor transcription in UC over CD. Additionally, three Wnt target genes (*TCF4*, *EMP3*, *JUNB*), all of which are involved in cell growth or proliferation, exhibited specific regulation in IBD, with *EMP3* and *JUNB* regulation exclusive for UC. UC-specific regulation of genes in the Wnt pathway could suggest that dysregulation of Wnt pathway may be more commonplace in UC than in CD, consequently leading to a higher predisposition for UC-related cancer via Wnt pathway.

Within the cell-cell adhesion theme, the differential regulation of *CDH11*, *CDH1*, *VCL*, *CTNNB1*, and *CTNNA1* may be working in an concerted effort to promote epithelial cell restitution in the inflamed IBD mucosa. The role of cell migration in restitution process is additionally supported by Rho-ROCK-mediated cytoskeletal reorganization mediated by regulation of *ROCK1* and *PDLIM5*. Also, *TJP2* is of particular interest since its up-regulation is specific to inflamed IBD, and it could be an important signaling messenger between the tight junction and the nucleus, with possible cross-talk with Wnt pathway target genes.

A wealth of transcript information can be obtained by microarray expression profiling, which can be used to identify processes or pathways for further study. In the present study, microarray data analysis provided a number of potential leads. As it was not possible to follow all leads, gene transcripts from five focused functional themes were subject to verification using quantitative real-time PCR. However, the true fruits of microarray analysis will not be realized until further specific investigations are pursued.

5 Summary

The etiology of the two main forms of inflammatory bowel disease (IBD), Crohn's disease (CD) and ulcerative colitis (UC) remains unknown; however, a combination of environmental and genetic factors participate in the etiopathogenesis of IBD, including microbial, dietary, immune, inflammatory, and intestinal permeability factors. The advent of microarray expression screening facilitates the investigation of complex, polygenic diseases by enabling the scientist to monitor the gene transcript levels of the whole genome. As a first step, a comparison was made between two such methods, namely cDNA-based and oligonucleotide-based microarray platforms. For a given set of genes known to be present on both platforms, there was moderate overlap in 'present' or 'absent' gene calls and weak correlation in the expression levels of transcripts detected. Keeping these results in mind, both cDNA- and oligo-based microarrays were used to perform whole-genome expression profiling in colonic mucosa biopsies obtained from IBD patients and normal controls (> 80 individuals), and the resulting microarray datasets were analysed separately. This study identified many functional groups of gene transcripts that were differentially expressed between inflamed/non-inflamed IBD and normal controls. Though many different lines of investigation could have been pursued, specific focus was placed on immune and inflammatory response, oncogenesis, and structure and permeability. Particular attention was further limited to the role of genes involved in Wnt pathway and cell-cell adhesion in IBD. Selected genes of interest were further verified using an independent technique (quantitative real-time PCR) in a larger cohort of IBD patients, disease-specificity controls and normal controls (>300 individuals). Taken together, a comprehensive whole-genome expression analysis revealed novel genes (*CYLD*, *DMBT1*, *CEACAM1*, *CSNKD1*, *PDLIM5*, *ROCK1*, *CDH11*, *SPINK5*, *PRKCB1*, *TJP2*, *VCL*, *ADAM12*, *FOXL1*, *TCF7*, *TCF7L2*, *TCF4*, *EMP3*, *JUNB*) that are associated with the pathogenesis of IBD and serve to provide fresh insight into the mechanisms of this complex disease.

6 Zusammenfassung

Die Ursachen von Morbus Crohn (MC) und Colitis Ulcerosa (CU), den beiden Hauptformen der chronisch entzündlichen Darmerkrankungen (CED), sind noch weitestgehend unbekannt. Derzeit geht man davon aus, dass eine Reihe von Faktoren wie genetischer Hintergrund, pathogene Keime, Ernährung, verschiedene Umweltfaktoren, Immunsystem sowie Entzündung und intestinale Permeabilität einen Einfluss auf CED ausüben. Die systematische Expressionsanalyse mittels Microarrays bietet die Möglichkeit, mRNA-Transkriptmengen des gesamten Genoms zu quantifizieren und stellt somit ein geeignetes Werkzeug dar, komplexe polygene Erkrankungen zu untersuchen.

In einem ersten Schritt wurden cDNA- und Oligonucleotid-basierte Microarrays systematisch verglichen. Dabei zeigten Transkripte, die mit beiden Systemen nachgewiesen werden konnten, eine mäßige Übereinstimmung in der Detektionsrate (Transkript detektiert / nicht detektiert) und eine schwache Korrelation zwischen den quantitativen Expressionswerten beider Systeme. Vor dem Hintergrund dieser Ergebnisse wurden in einem zweiten Schritt genomweite Expressionsprofile aus Darmbiopsien von CED Patienten und gesunden Individuen gewonnen (ca. 80 Probanden). Im Rahmen dieser Untersuchungen wurden verschiedene funktionelle Gruppen von Transkripten identifiziert, deren differentielle Expression zwischen entzündeten / nicht-entzündeten CED Patientengeweben und Normalgeweben deutliche Unterschiede aufwies. Im weiteren Verlauf wurde biologischen Prozessen, die mit Immun- und Entzündungsantworten, der Oncogenese sowie Struktur und Permeabilität assoziiert sind, besondere Aufmerksamkeit gewidmet. Dabei wurde im Speziellen der Einfluss von Genen, welche Funktionen im Wnt-Signaltransduktionsweg ausüben und an der Zell-Zell Adhäsion beteiligt sind, auf die Pathophysiologie von CED untersucht. Die Ergebnisse für ausgewählte Gene wurden mittels einer unabhängigen Technologie (Real-Time PCR) in einer größeren Stichprobe bestehend aus CED-Patienten, erkrankten Kontrollindividuen und gesunden Individuen (>300 Probanden) verifiziert.

Die vorliegende umfassende Studie konnte mittels genomweiter Expressionanalyse eine Reihe von neuen Genen (*CYLD*, *DMBT1*, *CEACAM1*, *CSNKD1*, *PDLIM5*, *ROCK1*, *CDH11*, *SPINK5*, *PRKCB1*, *TJP2*, *VCL*, *ADAM12*, *FOXL1*, *TCF7*, *TCF7L2*, *TCF4*, *EMP3*, *JUNB*) identifizieren, die

mit der Pathophysiologie von CED assoziiert sind. Sie liefert damit neue Erkenntnisse zu molekularen Mechanismen dieser komplexen Erkrankungen.

7 References

- (1998). Ulcerative colitis and colon carcinoma: epidemiology, surveillance, diagnosis, and treatment. The Society for Surgery of the Alimentary Tract, American Gastroenterological Association American Society for Liver Diseases, American Society for Gastrointestinal Endoscopy, American Hepato-Pancreato-Biliary Association. *J Gastrointest Surg* 2, 305-306.
- Ai, X., Do, A. T., Lozynska, O., Kusche-Gullberg, M., Lindahl, U., and Emerson, C. P., Jr. (2003). QSulf1 remodels the 6-O sulfation states of cell surface heparan sulfate proteoglycans to promote Wnt signaling. *J Cell Biol* 162, 341-351. Epub 2003 Jul 14.
- Alexander, C. M., Reichsman, F., Hinkes, M. T., Lincecum, J., Becker, K. A., Cumberland, S., and Bernfield, M. (2000). Syndecan-1 is required for Wnt-1-induced mammary tumorigenesis in mice. *Nat Genet* 25, 329-332.
- Amit, S., Hatzubai, A., Birman, Y., Andersen, J. S., Ben-Shushan, E., Mann, M., Ben-Neriah, Y., and Alkalay, I. (2002). Axin-mediated CKI phosphorylation of beta-catenin at Ser 45: a molecular switch for the Wnt pathway. *Genes Dev* 16, 1066-1076.
- Arumugam, T., Simeone, D. M., Schmidt, A. M., and Logsdon, C. D. (2004). S100P stimulates cell proliferation and survival via receptor for activated glycation end products (RAGE). *J Biol Chem* 279, 5059-5065. Epub 2003 Nov 14.
- Aust, D. E., Terdiman, J. P., Willenbacher, R. F., Chang, C. G., Molinaro-Clark, A., Baretton, G. B., Loehrs, U., and Waldman, F. M. (2002). The APC/beta-catenin pathway in ulcerative colitis-related colorectal carcinomas: a mutational analysis. *Cancer* 94, 1421-1427.
- Bamba, S., Lee, C. Y., Brittan, M., Preston, S. L., Direkze, N. C., Poulson, R., Alison, M. R., Wright, N. A., and Otto, W. R. (2006). Bone marrow transplantation ameliorates pathology in interleukin-10 knockout colitic mice. *J Pathol* 209, 265-273.
- Banziger, C., Soldini, D., Schutt, C., Zipperlen, P., Hausmann, G., and Basler, K. (2006). Wntless, a conserved membrane protein dedicated to the secretion of Wnt proteins from signaling cells. *Cell* 125, 509-522.
- Bartscherer, K., Pelte, N., Ingelfinger, D., and Boutros, M. (2006). Secretion of Wnt ligands requires Evi, a conserved transmembrane protein. *Cell* 125, 523-533.
- Battle, E., Henderson, J. T., Beghtel, H., van den Born, M. M., Sancho, E., Huls, G., Meeldijk, J., Robertson, J., van de Wetering, M., Pawson, T., and Clevers, H. (2002). Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* 111, 251-263.
- Beaven, S. W., and Abreu, M. T. (2004). Biomarkers in inflammatory bowel disease. *Curr Opin Gastroenterol* 20, 318-327.
- Beer, D. G., Kardia, S. L., Huang, C. C., Giordano, T. J., Levin, A. M., Misek, D. E., Lin, L., Chen, G., Ghazizadeh, T. G., Thomas, D. G., *et al.* (2002). Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 8, 816-824.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 57, 289-300.
- Ben-Porath, I., Kozak, C. A., and Benvenisty, N. (1998). Chromosomal mapping of Tmp (Emp1), Xmp (Emp2), and Ymp (Emp3), genes encoding membrane proteins related to Pmp22. *Genomics* 49, 443-447.
- Best, W. R., Bechtel, J. M., Singleton, J. W., and Kern, F., Jr. (1976). Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. *Gastroenterology* 70, 439-444.
- Betanzos, A., Huerta, M., Lopez-Bayghen, E., Azuara, E., Amerena, J., and Gonzalez-Mariscal, L. (2004). The tight junction protein ZO-2 associates with Jun, Fos and C/EBP transcription factors in epithelial cells. *Exp Cell Res* 292, 51-66.
- Bienz, M. (2002). The subcellular destinations of APC proteins. *Nat Rev Mol Cell Biol* 3, 328-338.
- Bikker, F. J., Ligtenberg, A. J., Nazmi, K., Veerman, E. C., van't Hof, W., Bolscher, J. G., Poustka, A., Nieuw Amerongen, A. V., and Mollenhauer, J. (2002). Identification of the bacteria-binding

- peptide domain on salivary agglutinin (gp-340/DMBT1), a member of the scavenger receptor cysteine-rich superfamily. *J Biol Chem* 277, 32109-32115.
- Birrenbach, T., and Bocker, U. (2004). Inflammatory bowel disease and smoking: a review of epidemiology, pathophysiology, and therapeutic implications. *Inflamm Bowel Dis* 10, 848-859.
- Blumenthal, A., Ehlers, S., Lauber, J., Buer, J., Lange, C., Goldmann, T., Heine, H., Brandt, E., and Reiling, N. (2006). The Wingless homologue Wnt5a and its receptor Frizzled-5 regulate inflammatory responses of human mononuclear cells induced by microbial stimulation. *Blood*.
- Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185-193.
- Bouma, G., and Strober, W. (2003). The immunological and genetic basis of inflammatory bowel disease. *Nat Rev Immunol* 3, 521-533.
- Braidotti, P., Nuciforo, P. G., Mollenhauer, J., Poustka, A., Pellegrini, C., Moro, A., Bulfamante, G., Coggi, G., Bosari, S., and Pietra, G. G. (2004). DMBT1 expression is down-regulated in breast cancer. *BMC Cancer* 4, 46.
- Breese, E. J., Michie, C. A., Nicholls, S. W., Murch, S. H., Williams, C. B., Domizio, P., Walker-Smith, J. A., and MacDonald, T. T. (1994). Tumor necrosis factor alpha-producing cells in the intestinal mucosa of children with inflammatory bowel disease. *Gastroenterology* 106, 1455-1466.
- Bronckart, Y., Decaestecker, C., Nagy, N., Harper, L., Schafer, B. W., Salmon, I., Pochet, R., Kiss, R., and Heizman, C. W. (2001). Development and progression of malignancy in human colon tissues are correlated with expression of specific Ca(2+)-binding S100 proteins. *Histol Histopathol* 16, 707-712.
- Brostrom, O., Lofberg, R., Nordenvall, B., Ost, A., and Hellers, G. (1987). The risk of colorectal cancer in ulcerative colitis. An epidemiologic study. *Scand J Gastroenterol* 22, 1193-1199.
- Carraway, K. L., Ramsauer, V. P., Haq, B., and Carothers Carraway, C. A. (2003). Cell signaling through membrane mucins. *Bioessays* 25, 66-71.
- Chavanas, S., Bodemer, C., Rochat, A., Hamel-Teillac, D., Ali, M., Irvine, A. D., Bonafe, J. L., Wilkinson, J., Taieb, A., Barrandon, Y., *et al.* (2000). Mutations in SPINK5, encoding a serine protease inhibitor, cause Netherton syndrome. *Nat Genet* 25, 141-142.
- Chen, D., Iijima, H., Nagaishi, T., Nakajima, A., Russell, S., Raychowdhury, R., Morales, V., Rudd, C. E., Utku, N., and Blumberg, R. S. (2004). Carcinoembryonic antigen-related cellular adhesion molecule 1 isoforms alternatively inhibit and costimulate human T cell function. *J Immunol* 172, 3535-3543.
- Chinnadurai, G. (2002). CtBP, an unconventional transcriptional corepressor in development and oncogenesis. *Mol Cell* 9, 213-224.
- Cho, J. H., Nicolae, D. L., Ramos, R., Fields, C. T., Rabenau, K., Corradino, S., Brant, S. R., Espinosa, R., LeBeau, M., Hanauer, S. B., *et al.* (2000). Linkage and linkage disequilibrium in chromosome band 1p36 in American Chaldeans with inflammatory bowel disease. *Hum Mol Genet* 9, 1425-1432.
- Clayburgh, D. R., Shen, L., and Turner, J. R. (2004). A porous defense: the leaky epithelial barrier in intestinal disease. *Lab Invest* 84, 282-291.
- Costello, C. M., Mah, N., Hasler, R., Rosenstiel, P., Waetzig, G. H., Hahn, A., Lu, T., Gurbuz, Y., Nikolaus, S., Albrecht, M., *et al.* (2005). Dissection of the inflammatory bowel disease transcriptome using genome-wide cDNA microarrays. *PLoS Med* 2, e199. Epub 2005 Aug 2003.
- Cuff, M., Dyer, J., Jones, M., and Shirazi-Beechey, S. (2005). The human colonic monocarboxylate transporter Isoform 1: its potential importance to colonic tissue homeostasis. *Gastroenterology* 128, 676-686.
- Dale, T. C. (1998). Signal transduction by the Wnt family of ligands. *Biochem J* 329 (Pt 2), 209-223.
- Dieckgraefe, B. K., Stenson, W. F., Korzenik, J. R., Swanson, P. E., and Harrington, C. A. (2000). Analysis of mucosal gene expression in inflammatory bowel disease by parallel oligonucleotide arrays. *Physiol Genomics* 4, 1-11.

- Dooley, T. P., Curto, E. V., Reddy, S. P., Davis, R. L., Lambert, G. W., Wilborn, T. W., and Elson, C. O. (2004). Regulation of gene expression in inflammatory bowel disease and correlation with IBD drugs: screening by DNA microarrays. *Inflamm Bowel Dis* 10, 1-14.
- Dring, M. M., Goulding, C. A., Trimble, V. I., Keegan, D., Ryan, A. W., Brophy, K. M., Smyth, C. M., Keeling, P. W., O'Donoghue, D., O'Sullivan, M., *et al.* (2006). The pregnane X receptor locus is associated with susceptibility to inflammatory bowel disease. *Gastroenterology* 130, 341-348; quiz 592.
- Duchmann, R., Kaiser, I., Hermann, E., Mayet, W., Ewe, K., and Meyer zum Buschenfelde, K. H. (1995). Tolerance exists towards resident intestinal flora but is broken in active inflammatory bowel disease (IBD). *Clin Exp Immunol* 102, 448-455.
- Duerr, R. H., Barmada, M. M., Zhang, L., Pfutzer, R., and Weeks, D. E. (2000). High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12. *Am J Hum Genet* 66, 1857-1862.
- Ebnet, K., Suzuki, A., Ohno, S., and Vestweber, D. (2004). Junctional adhesion molecules (JAMs): more molecules with dual functions? *J Cell Sci* 117, 19-29.
- Eickhoff, H., Schuchhardt, J., Ivanov, I., Meier-Ewert, S., O'Brien, J., Malik, A., Tandon, N., Wolski, E. W., Rohlf, E., Nyarsik, L., *et al.* (2000). Tissue gene expression analysis using arrayed normalized cDNA libraries. *Genome Res* 10, 1230-1240.
- Ekbom, A., Helmick, C., Zack, M., and Adami, H. O. (1990). Ulcerative colitis and colorectal cancer. A population-based study. *N Engl J Med* 323, 1228-1233.
- Emami, S., Rodrigues, S., Rodrigue, C. M., Le Floch, N., Rivat, C., Attoub, S., Bruyneel, E., and Gespach, C. (2004). Trefoil factor family (TFF) peptides and cancer progression. *Peptides* 25, 885-898.
- Ergun, S., Kilik, N., Ziegeler, G., Hansen, A., Nollau, P., Gotze, J., Wurmbach, J. H., Horst, A., Weil, J., Fernando, M., and Wagener, C. (2000). CEA-related cell adhesion molecule 1: a potent angiogenic factor and a major effector of vascular endothelial growth factor. *Mol Cell* 5, 311-320.
- Faux, M. C., Ross, J. L., Meeker, C., Johns, T., Ji, H., Simpson, R. J., Layton, M. J., and Burgess, A. W. (2004). Restoration of full-length adenomatous polyposis coli (APC) protein in a colon cancer cell line enhances cell adhesion. *J Cell Sci* 117, 427-439.
- Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 132, 6-13.
- Fraser, A. G., and Marcotte, E. M. (2004). A probabilistic view of gene function. *Nat Genet* 36, 559-564.
- Fuss, I. J., Neurath, M., Boirivant, M., Klein, J. S., de la Motte, C., Strong, S. A., Fiocchi, C., and Strober, W. (1996). Disparate CD4+ lamina propria (LP) lymphokine secretion profiles in inflammatory bowel disease. Crohn's disease LP cells manifest increased secretion of IFN-gamma, whereas ulcerative colitis LP cells manifest increased secretion of IL-5. *J Immunol* 157, 1261-1270.
- Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20, 307-315.
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5, R80. Epub 2004 Sep 2015.
- Geschwind, D. H. (2001). Sharing gene expression data: an array of options. *Nat Rev Neurosci* 2, 435-438.
- Gewirtz, A. T., Navas, T. A., Lyons, S., Godowski, P. J., and Madara, J. L. (2001). Cutting edge: bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. *J Immunol* 167, 1882-1885.
- Giles, R. H., van Es, J. H., and Clevers, H. (2003). Caught up in a Wnt storm: Wnt signaling in cancer. *Biochim Biophys Acta* 1653, 1-24.
- Gregorieff, A., Pinto, D., Begthel, H., Destree, O., Kielman, M., and Clevers, H. (2005). Expression pattern of Wnt signaling components in the adult intestine. *Gastroenterology* 129, 626-638.

- Gress, T. M., Hoheisel, J. D., Lennon, G. G., Zehetner, G., and Lehrach, H. (1992). Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. *Mamm Genome* 3, 609-619.
- Gstaiger, M., Hovens, C., Georgiev, O., Knoepfel, L., and Schaffner, W. (1996). BZLF1 (ZEBRA, Zta) protein of Epstein-Barr virus selected in a yeast one-hybrid system by binding to a consensus site in the IgH intronic enhancer: a role in immunoglobulin expression? *Biol Chem* 377, 669-673.
- Gumbiner, B. M. (2005). Regulation of cadherin-mediated adhesion in morphogenesis. *Nat Rev Mol Cell Biol* 6, 622-634.
- Hachem, J. P., Wagberg, F., Schmuth, M., Crumrine, D., Lissens, W., Jayakumar, A., Houben, E., Mauro, T. M., Leonardsson, G., Brattsand, M., *et al.* (2006). Serine Protease Activity and Residual LEKTI Expression Determine Phenotype in Netherton Syndrome. *J Invest Dermatol* 126, 1609-1621.
- Hacker, U., Nybakken, K., and Perrimon, N. (2005). Heparan sulphate proteoglycans: the sweet side of development. *Nat Rev Mol Cell Biol* 6, 530-541.
- Halgren, R. G., Fielden, M. R., Fong, C. J., and Zacharewski, T. R. (2001). Assessment of clone identity and sequence fidelity for 1189 IMAGE cDNA clones. *Nucleic Acids Res* 29, 582-588.
- Hampe, J., Frenzel, H., Mirza, M. M., Croucher, P. J., Cuthbert, A., Mascheretti, S., Huse, K., Platzer, M., Bridger, S., Meyer, B., *et al.* (2002). Evidence for a NOD2-independent susceptibility locus for inflammatory bowel disease on chromosome 16p. *Proc Natl Acad Sci U S A* 99, 321-326.
- Hampe, J., Lynch, N. J., Daniels, S., Bridger, S., Macpherson, A. J., Stokkers, P., Forbes, A., Lennard-Jones, J. E., Mathew, C. G., Curran, M. E., and Schreiber, S. (2001). Fine mapping of the chromosome 3p susceptibility locus in inflammatory bowel disease. *Gut* 48, 191-197.
- Hampe, J., Schreiber, S., Shaw, S. H., Lau, K. F., Bridger, S., Macpherson, A. J., Cardon, L. R., Sakul, H., Harris, T. J., Buckler, A., *et al.* (1999a). A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am J Hum Genet* 64, 808-816.
- Hampe, J., Shaw, S. H., Saiz, R., Leysens, N., Lantermann, A., Mascheretti, S., Lynch, N. J., MacPherson, A. J., Bridger, S., van Deventer, S., *et al.* (1999b). Linkage of inflammatory bowel disease to human chromosome 6p. *Am J Hum Genet* 65, 1647-1655.
- Harren, M., Schonfelder, G., Paul, M., Horak, I., Riecken, E. O., Wiedenmann, B., and John, M. (1998). High expression of inducible nitric oxide synthase correlates with intestinal inflammation of interleukin-2-deficient mice. *Ann N Y Acad Sci* 859, 210-215.
- Hauck, C. R. (2002). Cell adhesion receptors - signaling capacity and exploitation by bacterial pathogens. *Med Microbiol Immunol (Berl)* 191, 55-62.
- He, T. C., Chan, T. A., Vogelstein, B., and Kinzler, K. W. (1999). PPARdelta is an APC-regulated target of nonsteroidal anti-inflammatory drugs. *Cell* 99, 335-345.
- Henderson, B. R., and Fagotto, F. (2002). The ins and outs of APC and beta-catenin nuclear transport. *EMBO Rep* 3, 834-839.
- Hendrickson, B. A., Gokhale, R., and Cho, J. H. (2002). Clinical aspects and pathophysiology of inflammatory bowel disease. *Clin Microbiol Rev* 15, 79-94.
- Henthorn, P., Kiledjian, M., and Kadesch, T. (1990). Two distinct transcription factors that bind the immunoglobulin enhancer microE5/kappa 2 motif. *Science* 247, 467-470.
- Hermiston, M. L., and Gordon, J. I. (1995). Inflammatory bowel disease and adenomas in mice expressing a dominant negative N-cadherin. *Science* 270, 1203-1207.
- Hirota, Y., Tanaka, S., Haruma, K., Yoshihara, M., Sumii, K., Kajiyama, G., Shimamoto, F., and Kohno, N. (2000). pS2 expression as a possible diagnostic marker of colorectal carcinoma in ulcerative colitis. *Oncol Rep* 7, 233-239.
- Holmberg, J., Genander, M., Halford, M. M., Anneren, C., Sondell, M., Chumley, M. J., Silvano, R. E., Henkemeyer, M., and Frisen, J. (2006). EphB receptors coordinate migration and proliferation in the intestinal stem cell niche. *Cell* 125, 1151-1163.

- Holten-Andersen, M. N., Fenger, C., Nielsen, H. J., Rasmussen, A. S., Christensen, I. J., Brunner, N., and Kronborg, O. (2004). Plasma TIMP-1 in patients with colorectal adenomas: a prospective study. *Eur J Cancer* *40*, 2159-2164.
- Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* *18*, S96-104.
- Hugot, J. P., Alberti, C., Berrebi, D., Bingen, E., and Cezard, J. P. (2003). Crohn's disease: the cold chain hypothesis. *Lancet* *362*, 2012-2015.
- Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J. P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C. A., Gassull, M., *et al.* (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* *411*, 599-603.
- Iba, K., Albrechtsen, R., Gilpin, B., Frohlich, C., Loechel, F., Zolkiewska, A., Ishiguro, K., Kojima, T., Liu, W., Langford, J. K., *et al.* (2000). The cysteine-rich domain of human ADAM 12 supports cell adhesion through syndecans and triggers signaling events that lead to beta1 integrin-dependent cell spreading. *J Cell Biol* *149*, 1143-1156.
- Iijima, H., Neurath, M. F., Nagaishi, T., Glickman, J. N., Nieuwenhuis, E. E., Nakajima, A., Chen, D., Fuss, I. J., Utku, N., Lewicki, D. N., *et al.* (2004). Specific regulation of T helper cell 1-mediated murine colitis by CEACAM1. *J Exp Med* *199*, 471-482.
- Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* *4*, 249-264.
- Islas, S., Vega, J., Ponce, L., and Gonzalez-Mariscal, L. (2002). Nuclear localization of the tight junction protein ZO-2 in epithelial cells. *Exp Cell Res* *274*, 138-148.
- Itoh, M., Furuse, M., Morita, K., Kubota, K., Saitou, M., and Tsukita, S. (1999). Direct binding of three tight junction-associated MAGUKs, ZO-1, ZO-2, and ZO-3, with the COOH termini of claudins. *J Cell Biol* *147*, 1351-1363.
- Izzi, L., Turbide, C., Houde, C., Kunath, T., and Beauchemin, N. (1999). cis-Determinants in the cytoplasmic domain of CEACAM1 responsible for its tumor inhibitory function. *Oncogene* *18*, 5563-5572.
- Jono, H., Lim, J. H., Chen, L. F., Xu, H., Trompouki, E., Pan, Z. K., Mosialos, G., and Li, J. D. (2004). NF-kappaB is essential for induction of CYLD, the negative regulator of NF-kappaB: evidence for a novel inducible autoregulatory feedback pathway. *J Biol Chem* *279*, 36171-36174.
- Jordan, B. (2002). Historical Background and Anticipated Developments. *Ann N Y Acad Sci* *975*, 24-32.
- Kane, M. D., Jatko, T. A., Stumpf, C. R., Lu, J., Thomas, J. D., and Madore, S. J. (2000). Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* *28*, 4552-4557.
- Kang, W., and Reid, K. B. (2003). DMBT1, a regulator of mucosal homeostasis through the linking of mucosal defense and regeneration? *FEBS Lett* *540*, 21-25.
- Kawano, Y., and Kypta, R. (2003). Secreted antagonists of the Wnt signalling pathway. *J Cell Sci* *116*, 2627-2634.
- Kett, K., Rognum, T. O., and Brandtzaeg, P. (1987). Mucosal subclass distribution of immunoglobulin G-producing cells is different in ulcerative colitis and Crohn's disease of the colon. *Gastroenterology* *93*, 919-924.
- Kiener, H. P., Stipp, C. S., Allen, P. G., Higgins, J. M., and Brenner, M. B. (2006). The cadherin-11 cytoplasmic juxtamembrane domain promotes alpha-catenin turnover at adherens junctions and intercellular motility. *Mol Biol Cell* *17*, 2366-2376.
- Klement, E., Cohen, R. V., Boxman, J., Joseph, A., and Reif, S. (2004). Breastfeeding and risk of inflammatory bowel disease: a systematic review with meta-analysis. *Am J Clin Nutr* *80*, 1342-1352.

- Ko, J., Ryu, K. S., Lee, Y. H., Na, D. S., Kim, Y. S., Oh, Y. M., Kim, I. S., and Kim, J. W. (2002). Human secreted frizzled-related protein is down-regulated and induces apoptosis in human cervical cancer. *Exp Cell Res* 280, 280-287.
- Kolligs, F. T., Nieman, M. T., Winer, I., Hu, G., Van Mater, D., Feng, Y., Smith, I. M., Wu, R., Zhai, Y., Cho, K. R., and Fearon, E. R. (2002). ITF-2, a downstream target of the Wnt/TCF pathway, is activated in human cancers with beta-catenin defects and promotes neoplastic transformation. *Cancer Cell* 1, 145-155.
- Komatsu, K., Murata, K., Kameyama, M., Ayaki, M., Mukai, M., Ishiguro, S., Miyoshi, J., Tatsuta, M., Inoue, M., and Nakamura, H. (2002). Expression of S100A6 and S100A4 in matched samples of human colorectal mucosa, primary colorectal adenocarcinomas and liver metastases. *Oncology* 63, 192-200.
- Korinek, V., Barker, N., Morin, P. J., van Wichen, D., de Weger, R., Kinzler, K. W., Vogelstein, B., and Clevers, H. (1997). Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. *Science* 275, 1784-1787.
- Korzenik, J. R. (2005). Past and current theories of etiology of IBD: toothpaste, worms, and refrigerators. *J Clin Gastroenterol* 39, S59-65.
- Kosiewicz, M. M., Nast, C. C., Krishnan, A., Rivera-Nieves, J., Moskaluk, C. A., Matsumoto, S., Kozaiwa, K., and Cominelli, F. (2001). Th1-type responses mediate spontaneous ileitis in a novel murine model of Crohn's disease. *J Clin Invest* 107, 695-702.
- Kozaiwa, K., Sugawara, K., Smith, M. F., Jr., Carl, V., Yamschikov, V., Belyea, B., McEwen, S. B., Moskaluk, C. A., Pizarro, T. T., Cominelli, F., and McDuffie, M. (2003). Identification of a quantitative trait locus for ileitis in a spontaneous mouse model of Crohn's disease: SAMP1/YitFc. *Gastroenterology* 125, 477-490.
- Kraus, T. A., Cheifetz, A., Toy, L., Meddings, J. B., and Mayer, L. (2006). Evidence for a Genetic Defect in Oral Tolerance Induction in Inflammatory Bowel Disease. *Inflamm Bowel Dis* 12, 82-88.
- Kraus, T. A., Toy, L., Chan, L., Childs, J., and Mayer, L. (2004). Failure to induce oral tolerance to a soluble protein in patients with inflammatory bowel disease. *Gastroenterology* 126, 1771-1778.
- Kucharzik, T., Walsh, S. V., Chen, J., Parkos, C. A., and Nusrat, A. (2001). Neutrophil transmigration in inflammatory bowel disease is associated with differential expression of epithelial intercellular junction proteins. *Am J Pathol* 159, 2001-2009.
- Kuhnert, F., Davis, C. R., Wang, H. T., Chu, P., Lee, M., Yuan, J., Nusse, R., and Kuo, C. J. (2004). Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. *Proc Natl Acad Sci U S A* 101, 266-271.
- Kuo, W. P., Jenssen, T. K., Butte, A. J., Ohno-Machado, L., and Kohane, I. S. (2002). Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* 18, 405-412.
- Kuroda, S., Tokunaga, C., Kiyohara, Y., Higuchi, O., Konishi, H., Mizuno, K., Gill, G. N., and Kikkawa, U. (1996). Protein-protein interaction of zinc finger LIM domains with protein kinase C. *J Biol Chem* 271, 31029-31032.
- Kyo, K., Parkes, M., Takei, Y., Nishimori, H., Vyas, P., Satsangi, J., Simmons, J., Nagawa, H., Baba, S., Jewell, D., *et al.* (1999). Association of ulcerative colitis with rare VNTR alleles of the human intestinal mucin gene, MUC3. *Hum Mol Genet* 8, 307-311.
- Langmann, T., Moehle, C., Mauerer, R., Scharl, M., Liebisch, G., Zahn, A., Stremmel, W., and Schmitz, G. (2004). Loss of detoxification in inflammatory bowel disease: dysregulation of pregnane X receptor target genes. *Gastroenterology* 127, 26-40.
- Lawrance, I. C., Fiocchi, C., and Chakravarti, S. (2001). Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes. *Hum Mol Genet* 10, 445-456.
- Lennon, G. G., and Lehrach, H. (1991). Hybridization analyses of arrayed cDNA libraries. *Trends Genet* 7, 314-317.
- Levy, L., Neuveut, C., Renard, C. A., Charneau, P., Branchereau, S., Gauthier, F., Van Nhieu, J. T., Cherqui, D., Petit-Bertron, A. F., Mathieu, D., and Buendia, M. A. (2002). Transcriptional activation of interleukin-8 by beta-catenin-Tcf4. *J Biol Chem* 277, 42386-42393.

- Lindberg, E., Tysk, C., Andersson, K., and Järnerot, G. (1988). Smoking and inflammatory bowel disease. A case control study. *Gut* 29, 352-357.
- Liu, C., Li, Y., Semenov, M., Han, C., Baeg, G. H., Tan, Y., Zhang, Z., Lin, X., and He, X. (2002). Control of beta-catenin phosphorylation/degradation by a dual-kinase mechanism. *Cell* 108, 837-847.
- Livak, K. J. (1997). ABI Prism 7700 Sequence Detection System. User Bulletin no. 2. (Applied Biosystems).
- Logan, C. Y., and Nusse, R. (2004). The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* 20, 781-810.
- Lu, T., Costello, C. M., Croucher, P. J., Hasler, R., Deuschl, G., and Schreiber, S. (2005). Can Zipf's law be adapted to normalize microarrays? *BMC Bioinformatics* 6, 37.
- Ma, Y., Ohmen, J. D., Li, Z., Bentley, L. G., McElree, C., Pressman, S., Targan, S. R., Fischel-Ghodsian, N., Rotter, J. I., and Yang, H. (1999). A genome-wide search identifies potential new susceptibility loci for Crohn's disease. *Inflamm Bowel Dis* 5, 271-278.
- Magert, H. J., Standker, L., Kreutzmann, P., Zucht, H. D., Reinecke, M., Sommerhoff, C. P., Fritz, H., and Forssmann, W. G. (1999). LEKTI, a novel 15-domain type of human serine proteinase inhibitor. *J Biol Chem* 274, 21499-21502.
- Mah, N., Thelin, A., Lu, T., Nikolaus, S., Kuhbacher, T., Gurbuz, Y., Eickhoff, H., Kloppel, G., Lehrach, H., Mellgard, B., *et al.* (2004). A comparison of oligonucleotide and cDNA-based microarray systems. *Physiol Genomics* 16, 361-370. Epub 2003 Nov 2025.
- Mahler, M., Bristol, I. J., Leiter, E. H., Workman, A. E., Birkenmeier, E. H., Elson, C. O., and Sundberg, J. P. (1998). Differential susceptibility of inbred mouse strains to dextran sulfate sodium-induced colitis. *Am J Physiol* 274, G544-551.
- Mahler, M., Most, C., Schmidtke, S., Sundberg, J. P., Li, R., Hedrich, H. J., and Churchill, G. A. (2002). Genetics of colitis susceptibility in IL-10-deficient mice: backcross versus F2 results contrasted by principal component analysis. *Genomics* 80, 274-282.
- Mandell, K. J., and Parkos, C. A. (2005). The JAM family of proteins. *Adv Drug Deliv Rev* 57, 857-867.
- Mankertz, J., Tavalali, S., Schmitz, H., Mankertz, A., Riecken, E. O., Fromm, M., and Schulzke, J. D. (2000). Expression from the human occludin promoter is affected by tumor necrosis factor alpha and interferon gamma. *J Cell Sci* 113 (Pt 11), 2085-2090.
- Mannon, P. J., Fuss, I. J., Mayer, L., Elson, C. O., Sandborn, W. J., Present, D., Dolin, B., Goodman, N., Groden, C., Hornung, R. L., *et al.* (2004). Anti-interleukin-12 antibody for active Crohn's disease. *N Engl J Med* 351, 2069-2079.
- Massoumi, R., Chmielarska, K., Hennecke, K., Pfeifer, A., and Fassler, R. (2006). Cyld inhibits tumor cell proliferation by blocking Bcl-3-dependent NF-kappaB signaling. *Cell* 125, 665-677.
- Massova, I., Kotra, L. P., Fridman, R., and Mobashery, S. (1998). Matrix metalloproteinases: structures, evolution, and diversification. *Faseb J* 12, 1075-1095.
- Matter, K., and Balda, M. S. (2003). Signalling to and from tight junctions. *Nat Rev Mol Cell Biol* 4, 225-236.
- Mayberry, J. F., and Rhodes, J. (1984). Epidemiological aspects of Crohn's disease: a review of the literature. *Gut* 25, 886-899.
- McDonald, M. J., and Rosbash, M. (2001). Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell* 107, 567-578.
- McGall, G., Labadie, J., Brock, P., Wallraff, G., Nguyen, T., and Hinsberg, W. (1996). Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proc Natl Acad Sci U S A* 93, 13555-13560.
- Meijssen, M. A., Brandwein, S. L., Reinecker, H. C., Bhan, A. K., and Podolsky, D. K. (1998). Alteration of gene expression by intestinal epithelial cells precedes colitis in interleukin-2-deficient mice. *Am J Physiol* 274, G472-479.

- Melle, C., Osterloh, D., Ernst, G., Schimmel, B., Bleul, A., and von Eggeling, F. (2005). Identification of proteins from colorectal cancer tissue by two-dimensional gel electrophoresis and SELDI mass spectrometry. *Int J Mol Med* 16, 11-17.
- Middleton, F. A., Mirnics, K., Pierri, J. N., Lewis, D. A., and Levitt, P. (2002). Gene expression profiling reveals alterations of specific metabolic pathways in schizophrenia. *J Neurosci* 22, 2718-2729.
- Miller, J. R. (2002). The Wnts. *Genome Biol* 3, REVIEWS3001.
- Mitic, L. L., Van Itallie, C. M., and Anderson, J. M. (2000). Molecular physiology and pathophysiology of tight junctions I. Tight junction structure and function: lessons from mutant animals and proteins. *Am J Physiol Gastrointest Liver Physiol* 279, G250-254.
- Monteleone, G., Biancone, L., Marasco, R., Morrone, G., Marasco, O., Lizza, F., and Pallone, F. (1997). Interleukin 12 is expressed and actively released by Crohn's disease intestinal lamina propria mononuclear cells. *Gastroenterology* 112, 1169-1178.
- Moon, R. T., Kohn, A. D., De Ferrari, G. V., and Kaykas, A. (2004). WNT and beta-catenin signalling: diseases and therapies. *Nat Rev Genet* 5, 691-701.
- Munoz, R., Moreno, M., Oliva, C., Orbenes, C., and Larrain, J. (2006). Syndecan-4 regulates non-canonical Wnt signalling and is essential for convergent and extension movements in *Xenopus* embryos. *Nat Cell Biol* 8, 492-500.
- Nakagawa, N., Hoshijima, M., Oyasu, M., Saito, N., Tanizawa, K., and Kuroda, S. (2000). ENH, containing PDZ and LIM domains, heart/skeletal muscle-specific protein, associates with cytoskeletal proteins through the PDZ domain. *Biochem Biophys Res Commun* 272, 505-512.
- Nazli, A., Yang, P. C., Jury, J., Howe, K., Watson, J. L., Soderholm, J. D., Sherman, P. M., Perdue, M. H., and McKay, D. M. (2004). Epithelia under metabolic stress perceive commensal bacteria as a threat. *Am J Pathol* 164, 947-957.
- Nelson, W. J., and Nusse, R. (2004). Convergence of Wnt, beta-catenin, and cadherin pathways. *Science* 303, 1483-1487.
- Neurath, M. F., Pettersson, S., Meyer zum Buschenfelde, K. H., and Strober, W. (1996). Local administration of antisense phosphorothioate oligonucleotides to the p65 subunit of NF-kappa B abrogates established experimental colitis in mice. *Nat Med* 2, 998-1004.
- Nielsen, T. O., West, R. B., Linn, S. C., Alter, O., Knowling, M. A., O'Connell, J. X., Zhu, S., Fero, M., Sherlock, G., Pollack, J. R., *et al.* (2002). Molecular characterisation of soft tissue tumours: a gene expression study. *Lancet* 359, 1301-1307.
- Ogawa, H., Fukushima, K., Naito, H., Funayama, Y., Unno, M., Takahashi, K., Kitayama, T., Matsuno, S., Ohtani, H., Takasawa, S., *et al.* (2003). Increased expression of HIP/PAP and regenerating gene III in human inflammatory bowel disease and a murine bacterial reconstitution model. *Inflamm Bowel Dis* 9, 162-170.
- Ohki, R., Yamamoto, K., Mano, H., Lee, R. T., Ikeda, U., and Shimada, K. (2002). Identification of mechanically induced genes in human monocytic cells by DNA microarrays. *J Hypertens* 20, 685-691.
- Okamoto, T., Suzuki, T., and Yamamoto, N. (2000). Microarray fabrication with covalent attachment of DNA using bubble jet technology. *Nat Biotechnol* 18, 438-441.
- Orholm, M., Binder, V., Sorensen, T. I., Rasmussen, L. P., and Kyvik, K. O. (2000). Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. *Scand J Gastroenterol* 35, 1075-1081.
- Parronchi, P., Romagnani, P., Annunziato, F., Sampognaro, S., Becchio, A., Giannarini, L., Maggi, E., Pupilli, C., Tonelli, F., and Romagnani, S. (1997). Type 1 T-helper cell predominance and interleukin-12 expression in the gut of patients with Crohn's disease. *Am J Pathol* 150, 823-832.
- Peifer, M. (2002). Developmental biology: colon construction. *Nature* 420, 274-275, 277.
- Peifer, M., and Polakis, P. (2000). Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. *Science* 287, 1606-1609.

- Pelteková, V. D., Wintle, R. F., Rubin, L. A., Amos, C. I., Huang, Q., Gu, X., Newman, B., Van Oene, M., Cescon, D., Greenberg, G., *et al.* (2004). Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet* 36, 471-475.
- Perreault, N., Katz, J. P., Sackett, S. D., and Kaestner, K. H. (2001). Foxl1 controls the Wnt/beta-catenin pathway by modulating the expression of proteoglycans in the gut. *J Biol Chem* 276, 43328-43333.
- Persson, P. G., Ahlbom, A., and Hellers, G. (1992). Diet and inflammatory bowel disease: a case-control study. *Epidemiology* 3, 47-52.
- Prasad, S., Mingrino, R., Kaukinen, K., Hayes, K. L., Powell, R. M., MacDonald, T. T., and Collins, J. E. (2005). Inflammatory processes have differential effects on claudins 2, 3 and 4 in colonic epithelial cells. *Lab Invest* 85, 1139-1162.
- Rachmilewitz, D. (1989). Coated mesalazine (5-aminosalicylic acid) versus sulphasalazine in the treatment of active ulcerative colitis: a randomised trial. *Bmj* 298, 82-86.
- Regueiro, M., Kip, K. E., Cheung, O., Hegazi, R. A., and Plevy, S. (2005). Cigarette smoking and age at diagnosis of inflammatory bowel disease. *Inflamm Bowel Dis* 11, 42-47.
- Reif, S., Klein, I., Lubin, F., Farbstein, M., Hallak, A., and Gilat, T. (1997). Pre-illness dietary factors in inflammatory bowel disease. *Gut* 40, 754-760.
- Reiley, W. W., Zhang, M., Jin, W., Losiewicz, M., Donohue, K. B., Norbury, C. C., and Sun, S. C. (2006). Regulation of T cell development by the deubiquitinating enzyme CYLD. *Nat Immunol* 7, 411-417.
- Rhodes, D. R., Barrette, T. R., Rubin, M. A., Ghosh, D., and Chinnaiyan, A. M. (2002). Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res* 62, 4427-4433.
- Rioux, J. D., Daly, M. J., Silverberg, M. S., Lindblad, K., Steinhart, H., Cohen, Z., Delmonte, T., Kocher, K., Miller, K., Guschwan, S., *et al.* (2001). Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 29, 223-228.
- Rioux, J. D., Silverberg, M. S., Daly, M. J., Steinhart, A. H., McLeod, R. S., Griffiths, A. M., Green, T., Brettin, T. S., Stone, V., Bull, S. B., *et al.* (2000). Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet* 66, 1863-1870.
- Roediger, W. E., Duncan, A., Kapaniris, O., and Millard, S. (1993). Reducing sulfur compounds of the colon impair colonocyte nutrition: implications for ulcerative colitis. *Gastroenterology* 104, 802-809.
- Rogler, G., Brand, K., Vogl, D., Page, S., Hofmeister, R., Andus, T., Knuechel, R., Baeuerle, P. A., Scholmerich, J., and Gross, V. (1998). Nuclear factor kappaB is activated in macrophages and epithelial cells of inflamed intestinal mucosa. *Gastroenterology* 115, 357-369.
- Rotter, J. I. (1994). Inflammatory bowel disease. *Lancet* 343, 1360.
- Rubel, C., Fernandez, G. C., Dran, G., Bompadre, M. B., Isturiz, M. A., and Palermo, M. S. (2001). Fibrinogen promotes neutrophil activation and delays apoptosis. *J Immunol* 166, 2002-2010.
- Sadlack, B., Merz, H., Schorle, H., Schimpl, A., Feller, A. C., and Horak, I. (1993). Ulcerative colitis-like disease in mice with a disrupted interleukin-2 gene. *Cell* 75, 253-261.
- Satsangi, J., Parkes, M., Louis, E., Hashimoto, L., Kato, N., Welsh, K., Terwilliger, J. D., Lathrop, G. M., Bell, J. I., and Jewell, D. P. (1996). Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* 14, 199-202.
- Sawada, N., Murata, M., Kikuchi, K., Osanai, M., Tobioka, H., Kojima, T., and Chiba, H. (2003). Tight junctions and human diseases. *Med Electron Microsc* 36, 147-156.
- Saxon, A., Shanahan, F., Landers, C., Ganz, T., and Targan, S. (1990). A distinct subset of antineutrophil cytoplasmic antibodies is associated with inflammatory bowel disease. *J Allergy Clin Immunol* 86, 202-210.
- Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467-470.

- Scheppach, W., Christl, S. U., Bartram, H. P., Richter, F., and Kasper, H. (1997). Effects of short-chain fatty acids on the inflamed colonic mucosa. *Scand J Gastroenterol Suppl* 222, 53-57.
- Schmitt, J. F., Millar, D. S., Pedersen, J. S., Clark, S. L., Venter, D. J., Frydenberg, M., Molloy, P. L., and Risbridger, G. P. (2002). Hypermethylation of the inhibin alpha-subunit gene in prostate carcinoma. *Mol Endocrinol* 16, 213-220.
- Schreiber, S., Nikolaus, S., and Hampe, J. (1998). Activation of nuclear factor kappa B inflammatory bowel disease. *Gut* 42, 477-484.
- Schultz, M., Tonkonogy, S. L., Sellon, R. K., Veltkamp, C., Godfrey, V. L., Kwon, J., Grenther, W. B., Balish, E., Horak, I., and Sartor, R. B. (1999). IL-2-deficient mice raised under germfree conditions develop delayed mild focal intestinal inflammation. *Am J Physiol* 276, G1461-1472.
- Sheldahl, L. C., Slusarski, D. C., Pandur, P., Miller, J. R., Kuhl, M., and Moon, R. T. (2003). Dishevelled activates Ca²⁺ flux, PKC, and CamKII in vertebrate embryos. *J Cell Biol* 161, 769-777.
- Shinozaki, S., Nakamura, T., Imura, M., Kato, Y., Iizuka, B., Kobayashi, M., and Hayashi, N. (2001). Upregulation of Reg 1alpha and GW112 in the epithelium of inflamed colonic mucosa. *Gut* 48, 623-629.
- Smyth, G. K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. In *Statistical Applications in Genetics and Molecular Biology*.
- Soderholm, J. D., Olaison, G., Lindberg, E., Hannestad, U., Vindels, A., Tysk, C., Järnerot, G., and Sjö Dahl, R. (1999). Different intestinal permeability patterns in relatives and spouses of patients with Crohn's disease: an inherited defect in mucosal defence? *Gut* 44, 96-100.
- Soderholm, J. D., Olaison, G., Peterson, K. H., Franzen, L. E., Lindmark, T., Wiren, M., Tagesson, C., and Sjö Dahl, R. (2002). Augmented increase in tight junction permeability by luminal stimuli in the non-inflamed ileum of Crohn's disease. *Gut* 50, 307-313.
- Sokal, R. R., and Rohlf, F. J. (1995). *Biometry: the principles and practice of statistics in biological research*, 3rd edn (New York, W. H. Freeman and Company).
- Stillman, B. A., and Tonkinson, J. L. (2001). Expression microarray hybridization kinetics depend on length of the immobilized DNA but are independent of immobilization substrate. *Anal Biochem* 295, 149-157.
- Stoeckert, C. J., Causton, H. C., and Ball, C. A. (2002). Microarray databases: standards and ontologies. *Nat Genet* 32 *Suppl* 2, 469-473.
- Stoll, M., Corneliussen, B., Costello, C. M., Waetzig, G. H., Mellgard, B., Koch, W. A., Rosenstiel, P., Albrecht, M., Croucher, P. J., Seeger, D., *et al.* (2004). Genetic variation in DLG5 is associated with inflammatory bowel disease. *Nat Genet* 36, 476-480.
- Strutt, D. (2003). Frizzled signalling and cell polarisation in *Drosophila* and vertebrates. *Development* 130, 4501-4513.
- Takada, R., Hijikata, H., Kondoh, H., and Takada, S. (2005). Analysis of combinatorial effects of Wnts and Frizzleds on beta-catenin/armadillo stabilization and Dishevelled phosphorylation. *Genes Cells* 10, 919-928.
- Taupin, D., and Podolsky, D. K. (2003). Trefoil factors: initiators of mucosal healing. *Nat Rev Mol Cell Biol* 4, 721-732.
- Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J., and Church, G. M. (1999). Systematic determination of genetic network architecture. *Nat Genet* 22, 281-285.
- Taylor, E., Cogdell, D., Coombes, K., Hu, L., Ramdas, L., Tabor, A., Hamilton, S., and Zhang, W. (2001). Sequence verification as quality-control step for production of cDNA microarrays. *Biotechniques* 31, 62-65.
- Thim, L., and Mortz, E. (2000). Isolation and characterization of putative trefoil peptide receptors. *Regul Pept* 90, 61-68.
- Thodeti, C. K., Albrechtsen, R., Grauslund, M., Asmar, M., Larsson, C., Takada, Y., Mercurio, A. M., Couchman, J. R., and Wewer, U. M. (2003). ADAM12/syndecan-4 signaling promotes beta 1 integrin-dependent cell spreading through protein kinase Calpha and RhoA. *J Biol Chem* 278, 9576-9584.

- Thompson, N. P., Driscoll, R., Pounder, R. E., and Wakefield, A. J. (1996). Genetics versus environment in inflammatory bowel disease: results of a British twin study. *Bmj* *312*, 95-96.
- Truelove, S. C., and Witts, L. J. (1955). Cortisone in ulcerative colitis; final report on a therapeutic trial. *Br Med J*, 1041-1048.
- Tysk, C., Lindberg, E., Järnerot, G., and Floderus-Myrhed, B. (1988). Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* *29*, 990-996.
- Ugolini, F., Charafe-Jauffret, E., Bardou, V. J., Geneix, J., Adelaide, J., Labat-Moleur, F., Penault-Llorca, F., Longy, M., Jacquemier, J., Birnbaum, D., and Pebusque, M. J. (2001). WNT pathway and mammary carcinogenesis: loss of expression of candidate tumor suppressor gene SFRP1 in most invasive carcinomas except of the medullary type. *Oncogene* *20*, 5810-5817.
- Umeda, K., Matsui, T., Nakayama, M., Furuse, K., Sasaki, H., Furuse, M., and Tsukita, S. (2004). Establishment and characterization of cultured epithelial cells lacking expression of ZO-1. *J Biol Chem* *279*, 44785-44794.
- Utech, M., Ivanov, A. I., Samarina, S. N., Bruewer, M., Turner, J. R., Mrsny, R. J., Parkos, C. A., and Nusrat, A. (2005). Mechanism of IFN-gamma-induced endocytosis of tight junction proteins: myosin II-dependent vacuolarization of the apical plasma membrane. *Mol Biol Cell* *16*, 5040-5052.
- Uthoff, S. M., Eichenberger, M. R., Lewis, R. K., Fox, M. P., Hamilton, C. J., McAuliffe, T. L., Grimes, H. L., and Galandiuk, S. (2001). Identification of candidate genes in ulcerative colitis and Crohn's disease using cDNA array technology. *Int J Oncol* *19*, 803-810.
- van de Wetering, M., Sancho, E., Verweij, C., de Lau, W., Oving, I., Hurlstone, A., van der Horn, K., Batlle, E., Coudreuse, D., Haramis, A. P., *et al.* (2002). The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* *111*, 241-250.
- van Hogezaand, R. A., Eichhorn, R. F., Choudry, A., Veenendaal, R. A., and Lamers, C. B. (2002). Malignancies in inflammatory bowel disease: fact or fiction? *Scand J Gastroenterol Suppl*, 48-53.
- Vieten, D., Corfield, A., Carroll, D., Ramani, P., and Spicer, R. (2005). Impaired mucosal regeneration in neonatal necrotising enterocolitis. *Pediatr Surg Int* *21*, 153-160.
- von Lampe, B., Barthel, B., Coupland, S. E., Riecken, E. O., and Rosewicz, S. (2000). Differential expression of matrix metalloproteinases and their tissue inhibitors in colon mucosa of patients with inflammatory bowel disease. *Gut* *47*, 63-73.
- Waetzig, G. H., Seegert, D., Rosenstiel, P., Nikolaus, S., and Schreiber, S. (2002). p38 mitogen-activated protein kinase is activated and linked to TNF-alpha signaling in inflammatory bowel disease. *J Immunol* *168*, 5342-5351.
- Wakabayashi, M., Ito, T., Mitsushima, M., Aizawa, S., Ueda, K., Amachi, T., and Kioka, N. (2003). Interaction of Ip-dlg/KIAA0583, a membrane-associated guanylate kinase family protein, with vinexin and beta-catenin at sites of cell-cell contact. *J Biol Chem* *278*, 21709-21714.
- Waller, S., Tremelling, M., Bredin, F., Godfrey, L., Howson, J., and Parkes, M. (2006). Evidence for association of OCTN genes and IBD5 with ulcerative colitis. *Gut* *55*, 809-814.
- Weinstock, J. V., Blum, A., Metwali, A., Elliott, D., Bunnett, N., and Arsenescu, R. (2003). Substance P regulates Th1-type colitis in IL-10 knockout mice. *J Immunol* *171*, 3762-3767.
- Weinstock, J. V., Summers, R., and Elliott, D. E. (2004). Helminths and harmony. *Gut* *53*, 7-9.
- Wills-Karp, M., Santeliz, J., and Karp, C. L. (2001). The germless theory of allergic disease: revisiting the hygiene hypothesis. *Nat Rev Immunol* *1*, 69-75.
- Wilson, A. J., and Gibson, P. R. (1997). Epithelial migration in the colon: filling in the gaps. *Clin Sci (Lond)* *93*, 97-108.
- Wodarz, A., and Nusse, R. (1998). Mechanisms of Wnt signaling in development. *Annu Rev Cell Dev Biol* *14*, 59-88.
- Wu, W., Kemp, B. L., Proctor, M. L., Gazdar, A. F., Minna, J. D., Hong, W. K., and Mao, L. (1999). Expression of DMBT1, a candidate tumor suppressor gene, is frequently lost in lung cancer. *Cancer Res* *59*, 1846-1851.

- Wu, Z., Irizarry, R. A., Gentleman, R., Murillo, F. M., and Spencer, F. (2004). A Model Based Background Adjustment for Oligonucleotide Expression Arrays. In Johns Hopkins University, Dept. of Biostatistics Working Papers.
- Yang, H., Taylor, K. D., and Rotter, J. I. (2001). Inflammatory bowel disease. I. Genetic epidemiology. *Mol Genet Metab* 74, 1-21.
- Yang, P. C., Jury, J., Soderholm, J. D., Sherman, P. M., McKay, D. M., and Perdue, M. H. (2006). Chronic psychological stress in rats induces intestinal sensitization to luminal antigens. *Am J Pathol* 168, 104-114; quiz 363.
- Ye, D., Ma, I., and Ma, T. Y. (2006). Molecular mechanism of tumor necrosis factor-alpha modulation of intestinal epithelial tight junction barrier. *Am J Physiol Gastrointest Liver Physiol* 290, G496-504.
- Yoshida, H., Jono, H., Kai, H., and Li, J. D. (2005). The tumor suppressor cylindromatosis (CYLD) acts as a negative regulator for toll-like receptor 2 signaling via negative cross-talk with TRAF6 AND TRAF7. *J Biol Chem* 280, 41111-41121.
- Yoshimoto, H., Saltsman, K., Gasch, A. P., Li, H. X., Ogawa, N., Botstein, D., Brown, P. O., and Cyert, M. S. (2002). Genome-wide analysis of gene expression regulated by the calcineurin/Crz1p signaling pathway in *Saccharomyces cerevisiae*. *J Biol Chem* 277, 31079-31088.
- Yoshino, K., Rubin, J. S., Higinbotham, K. G., Uren, A., Anest, V., Plisov, S. Y., and Perantoni, A. O. (2001). Secreted Frizzled-related proteins can regulate metanephric development. *Mech Dev* 102, 45-55.
- Yuen, T., Wurmbach, E., Pfeffer, R. L., Ebersole, B. J., and Sealfon, S. C. (2002). Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. *Nucleic Acids Res* 30, e48.
- Zbar, A. P., Simopoulos, C., and Karayiannakis, A. J. (2004). Cadherins: an integral role in inflammatory bowel disease and mucosal restitution. *J Gastroenterol* 39, 413-421.
- Zheng, W., Rosenstiel, P., Huse, K., Sina, C., Valentonyte, R., Mah, N., Zeitlmann, L., Grosse, J., Ruf, N., Nurnberg, P., *et al.* (2006). Evaluation of AGR2 and AGR3 as candidate genes for inflammatory bowel disease. *Genes Immun* 7, 11-18.
- Zhou, Z., Wang, J., Han, X., Zhou, J., and Linder, S. (1998). Up-regulation of human secreted frizzled homolog in apoptosis and its down-regulation in breast tumors. *Int J Cancer* 78, 95-99.

8 Appendix

8.1 Materials

Table 8.1. Materials and Sources

Material	Manufacturer / Supplier
100 bp DNA ladder [³² P]dCTP (≥2,500 Ci/mmol)	Invitrogen; Karlsruhe, Germany AmershamBiosciences/GE Healthcare UK Ltd, Little Chalfont, UK
10 X MOPS	Sigma; München, Germany
20 X SSC	Sigma; München, Germany
2-Mercaptoethanol	Sigma; München, Germany
37% Formaldehyde	Sigma; München, Germany
384-deep-well storage plate	ABgene, Epsom, UK
ABI PRISM® 96-Well Optical Reaction Plate	Applied Biosystems; Weiterstadt, Germany
Agarose	Eurogentec; Köln, Germany
Ampicillin	Sigma; München, Germany
AmpliTaq® DNA Polymerase	Applied Biosystems; Weiterstadt, Germany
AmpliTaq® Gold DNA Polymerase	Applied Biosystems; Weiterstadt, Germany
Anti-CSNK1D antibody	Santa Cruz Biotechnology, Santa Cruz, CA, USA
Anti-PRKCB1 antibody	Pharmingen, San Diego, CA, USA
Aquatek	Merck, Hawthorne, NY, USA
Bacto-Tryptone	BD Biosciences, Heidelberg
Bacto-Yeast Extract	BD Biosciences, Heidelberg
BigDye Terminator Ready reaction kit	Applied Biosystems; Weiterstadt, Germany
Bovine serum albumin (albumin fraction V)	Merck; Darmstadt, Germany
Bromophenol blue	Sigma; München, Germany
Cryotubes (2ml)	Greiner Bio-One GmbH; Frickenhausen, Germany
Diethyl pyrocarbonate (DEPC)	Sigma; München, Germany
di-Sodium hydrogen phosphate	Merck; Darmstadt, Germany
dNTP set (100mM solutions 100µM each)	Amersham Biosciences; Freiburg, Germany
Easy peel heat seal foil	ABgene, Epsom, UK
EDTA	Sigma; München, Germany
Enzo Bioarray High Yield RNA transcript labeling kit	Affymetrix, USA
Ethanol <i>p. a.</i>	Merck; Darmstadt, Germany
Ethidium bromide solution (10mg/ml)	Invitrogen; Karlsruhe, Germany
Formamide	Sigma; München, Germany
GeneAmp PCR buffer system (10x buffer w/o MgCl ₂ ; 25mM MgCl ₂ solution)	Applied Biosystems; Weiterstadt, Germany
Goat anti-rabbit IgG	Sigma; München, Germany
Glycerol	Sigma; München, Germany
HG-U133A GeneChips	Affymetrix, Santa Clara, USA
HG-U95Av2 GeneChips	Affymetrix, Santa Clara, USA
Hybridization blocking reagent	Roche Diagnostics, Mannheim, Germany
Isopropanol	Merck; Darmstadt, Germany
MgCl ₂	Merck; Darmstadt, Germany
MicroAmp® optical 96 well reaction plate	Applied Biosystems; Weiterstadt, Germany
MicroAmp® single strips	Applied Biosystems; Weiterstadt, Germany
Microtiter 384 well plates	Sarstedt; Nürnberg, Germany
Microtiter 96 well plates	Costar Corning Incorporated; Cambridge, MA, USA
Montage Plasmid Miniprep96 Kit	Millipore, Billerica, MA
Multiscreen column loader	Amersham Biosciences; Freiburg, Germany
MultiScribe Reverse Transcriptase	Applied Biosystems; Weiterstadt, Germany
Oligotex mRNA Mini Kit	Qiagen, Hilden, Germany
Peroxidase-conjugated rabbit anti-mouse secondary antibody	Sigma, Deisenhofen, Germany
Phosphate buffered saline (PBS)	Invitrogen/Gibco; Karlsruhe, Germany
Pipette (serological, sterile with filter 5 / 10 / 25 ml)	Sarstedt; Nümbrecht, Germany
Pipette tips with filter (10 / 200 / 1000 µl)	Sarstedt; Nürnberg, Germany

Material	Manufacturer / Supplier
Primers and Probes	Eurogentec, Seraing, Belgium
QIAshredder homogenizer kit	Qiagen; Hilden, Germany
RNA Ladder	Invitrogen/Gibco; Karlsruhe, Germany
RNase-Free DNase Set	Qiagen, Hilden, Germany
RNaseZap wipes	Ambion; Austin, TX, USA
RNeasy mini RNA extraction kit	Qiagen, Hilden, Germany
Salmon sperm DNA	Sigma; München, Germany
SAP shrimp alkaline phosphatase	Amersham Biosciences; Freiburg, Germany
Sephadex powder (G50 superfine)	Amersham Biosciences; Freiburg, Germany
Sephadex spin column plates MAHVN 4550	Amersham Biosciences; Freiburg, Germany
SmartLadder DNA marker	Eurogentec; Köln, Germany
Sodium acetate	Sigma; München, Germany
Sodium carbonate	Sigma; München, Germany
Sodium chloride	Merck; Darmstadt, Germany
Sodium citrate	Sigma; München, Germany
Sodium dodecyl sulfate	Sigma; München, Germany
Sodium hydrogen carbonate	Sigma; München, Germany
Sodium hydroxide	Merck; Darmstadt, Germany
Sodium phosphate	Sigma; München, Germany
SuperScript reverse transcriptase	Invitrogen/Life Technologies, Carlsbad, USA
TAE Buffer 25 × Ready	Pack Amresco; Solon, OH, USA
TAE Buffer 25x ready pack	Amresco; Solon, OH, USA
TaqMan® Universal PCR Master Mix	Applied Biosystems; Weiterstadt, Germany
Tris	Merck; Darmstadt, Germany
Tubes (0.5 / 1.5 / 2.0 mL)	Eppendorf; Köln, Germany
Tubes, sterile (15 mL)	Sarstedt; Nümbrecht, Germany
Tubes, sterile (50 mL)	BD Biosciences; Heidelberg, Germany
Universal TaqMan Master Mix	Applied Biosystems; Weiterstadt, Germany

8.2 Equipment

Table 8.2. Equipment and Sources

Item	Manufacturer/ Supplier
ABI 9700 Sequencer	Applied Biosystems, Foster City, USA
ABI Prism 7900HT Sequence Detection System	Applied Biosystems, Foster City, USA
Agilent 2100 Bioanalyzer	Agilent Technologies, Palo Alto, USA
GeneAmp PCR system 9700	Applied Biosystems, Foster City, USA
General purpose centrifuge (Labofuge 400R)	Heraeus Instruments, Osterode, Germany
Homogenizer for microfuge tubes, Teflon head	Omnilab, Bremen, Germany
Microfuge (Biofuge fresco)	Heraeus Instruments, Osterode, Germany
Phosphorimaging plate (BAS-MS 2325)	Fujifilm, Kanagawa, Japan
Phosphorimaging scanner (FLA-3000G)	Fujifilm, Kanagawa, Japan
Shaking incubator (GFL 3033)	Gesellschaft für Labortechnik mbH, Burgwedel, Germany
Thermomixer compact	Eppendorf, Hamburg, Germany
Ultrospec 3100 pro spectrophotometer	Amersham Pharmacia Biotech, Uppsala, Sweden
Vortex-genie 2	Scientific Industries, Bohemia, USA

8.3 Solutions and media

Solution	Composition
10 X DNA gel loading buffer	50% v/v glycerol, 0.1% bromophenol blue (w/v)
RNA gel loading buffer	40 % (v/v) formaldehyde, 40% (v/v) formamide, 0.9 X MOPS, 0.3 µg/µL ethidium bromide, 1 mM EDTA (pH 7.5), 1 X DNA gel loading buffer
TE (pH 7.5, 8.0)	10 mM Tris-HCl, 1 mM EDTA
DEPC treated water	1 mL DEPC in 1 L DDW, shake vigorously and autoclave
LB media	10g/L Tryptone, 5g/L Yeast Extract, 5 g/L NaCl

8.4 Primer and probe sequences

Primers and probes are listed according to the section in which they were mentioned. For Assays-On-Demand Gene expression assays from Applied Biosystems, only a context sequence is provided by the manufacturer. All fluorogenic probes are labeled with 5'FAM and 3'TAMRA.

Table 8.3 Primer and probe sequences for real-time PCR or regular PCR

Section	Gene Name or Primer Pair	Gene Symbol	Forward Primer Sequence 5' to 3'	Reverse Primer Sequence 5' to 3'	Probe Sequence/ Context sequence
2.1	M13F/ M13R	n/a	CGTTGTAAACGACGGCCAGT	TTTCACACAGGAAACAGCTATGAC	
2.1	M13for23/ M13rev30	n/a	GACGTTGTAAACGACGGCCAGT	ATAACAATTTACACAGGAAACAGC TATGA	
2.1	T7/ T3	n/a	TAATACGACTCACTATAGGG	AATTAACCCCTACTAAAGGG	
2.1, 2.2	actin, beta	ACTB	AGCCTCGCCTTTGCCGA	CTGGTGCCTGGGGCG	CCGCCGCCGTCCACACCCGCC
2.1	glyceraldehyde-3-phosphate dehydrogenase	GAPDH	ACCCACTCCTCCACCTTTGAC	CTGTTGCTGTAGCCAAATTCGT	N/A
2.1	cadherin 11	CDH11	ACCAACACCCTCACCATCAAA	GGCCGGCGTTCAGAA	TGCGGGTGCGACGTGAACG
2.1	claudin 4	CLDN4	GTGTACCAACTGCCTGGAGGAT	ACCCCGGCACTATCACCATA	CCAAGACCATGATCGTGGCGGTG
2.1	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	CEACAM1	TTCTGCTCACAGCCTCACTTCTA	AACCTCCTTCCCCTCTGCAA	CCTTCTGGAACCCGCCACC
2.1	N-methylpurine-DNA glycosylase	MPG	AGCATCTATTTCTCAAGCCAAA	GGAGTTCTGTGCCATTAGGAAGTC	CCAGCCGGCAGTCCCCTG
2.1	protein kinase C binding protein 1	PRKCBP1	AAGCGCCAGATTCGTAGCA	CTTCCGAGATTTCAATTTATGCCTAA	TTGACAAGACCATAGAGAGTTGCAAAGCAC A
2.1	matrix metalloproteinase 3	MMP3	TGTAGAAGGCACAATATGGGCAC	AGTCACTTGCTGTTGCACACGA	ATGTTTCGTTTTCTCCTGCCTGTGCTGTGA
2.1	aldolase B, fructose-bisphosphate	ALDOB	GGAGGCTTTTATGAAGCGGG	TGAAGAGCGACTGGGTGGAA	TAAGTCCAGGCGGCCAAAGGAC
2.1	mucin 1, transmembrane	MUC1	GTGCCCCCTAGCAGTACCG	GAAGTGGCTGCCACTGCTG	TGAGAAAGTTTCTGCAGGTAATGGTGGCA
2.1	translocase of inner mitochondrial membrane 17 homolog B (yeast)	TIMM17B	CAAGGGTTTTCCGCAATGC	TGCGAAGCTACCTCCAATCTG	CCGTTGAGAGGTAGTGCCAATGCTGT
2.1	matrix metalloproteinase 1	MMP1	ATGCGCACAAATCCCTTCTACC	TCCTCAGAAAGAGCAGCATCGATA	ACGAATTTGCCGACAGAGATGAAGTCCGGT
2.1	pro-platelet basic protein	PPBP	ACACTGAAGGATGGGAGGAAAATC	GCAGATTCATCACCTGCCAATTT	CAGATGCTCCCAGAATC
2.1	apolipoprotein A-I	APOA1	ACAGCGTGACCTCCACCTTC	CTGCCTCAGGCCCTCTGTCT	TCGGCCCTGTGACCCAGGAGTTCTG
2.2	interleukin 8	IL8	GGAATTGAATGGGTTTGCTAGAA	TGTGGATCCTGGCTAGCAGAC	TGGAAATCCTGGATTTTTTCTGTAAATCTG GCAA
2.2	tumor necrosis factor (TNF superfamily, member 2)	TNF	AGGCGGTGCTTGTTCCTCA	GTTTCGAGAAGATGATCTGACTGCC	CCAGAGGGAAGAGTTCCCCAGGGAC
2.2	cadherin 11, type 2, OB-cadherin (osteoblast)	CDH11	ACCAACACCCTCACCATCAAA	GGCCGGCGTTCAGAATG	TGCGGGTGCGACGTGAACG
2.2	decay accelerating factor for complement (CD55, Cromer blood group system)	DAF	CGAGGTGCCAACAAGGCTAA	CCTGGACGGCACTCATATTCC	TTCTGCATCCCTCAAACAGCCTTATATCACT CT
2.2	immunoglobulin heavy constant gamma 1 (G1m marker)	IGHG1	GCCCCAGTTGCTGGC	AGCTGGGTCCCCTGCTG	TTTCAATCCAAGCATAACTCAGTGACGCAT
2.2	mucin 1, transmembrane	MUC1	GTGCCCCCTAGCAGTACCG	GAAGTGGCTGCCACTGCTG	TGAGAAAGTTTCTGCAGGTAATGGTGGCA
2.2	trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in)	TFF1	GACACCTCAGACAGCTTCT	GAGTAGTCAAAGTCAGAGCAGTCAA	CCTCGGCTCACAAAC
2.2	TIMP metalloproteinase inhibitor 1	TIMP1	TTCTGGCATCCTGTTGTTGCT	TGACGAGGTCGGAATTGCA	CCCCCAGCAGGGCCTGCACC
2.2	cylindromatosis (turban tumor syndrome)	CYLD			TCGGGATGGTGGTCAGAAATGGCTTC
2.2	CTTNBP2 N-terminal like	DKFz547A0 23			AGGCTGACTCAACAGTTGGAATTTG
2.2	PDZ and LIM domain 5	PDLIM5			AAAATGGCCCAAGAAAACACAT
2.2	occludin	OCLN			CCACGCCGGTTCTGAAGTGGTTCA
2.2	phospholipase A2, group IIA (platelets,	PLA2G2A			AGCAGAATCACCTGTGCAAAAACAGG

Section	Gene Name or Primer Pair	Gene Symbol	Forward Primer Sequence 5' to 3'	Reverse Primer Sequence 5' to 3'	Probe Sequence/ Context sequence
2.2	synovial fluid) pleckstrin homology domain containing, family Q member 1	PLEKHQ1			GAACAAGGTCAGCGACATCAAATTC
2.2	rho-associated, coiled-coil containing protein kinase 1	ROCK1			AATTGGATGAAGAGGGAAATCAAAG
2.2	calcitonin gene-related peptide-receptor component protein	RCP9			GGAGATCCAGCTGATGGTGAAGAG
2.2	zinc finger, CCHC domain containing 4 (28B08) Homo sapiens cDNA FLJ31066	ZCCHC4			GCACATCCAAGGATGGCAGGAAATG
2.4	ADAM12 fis, clone HSYRA2001153	ADAM12			CCACGGCACACTCAGGCTGGAGATC
2.4	forkhead box l1	FOXL1			CTCGACCCGCATGTCCAGGGCGGGCT
2.4	cadherin 1, type 1, e-cadherin (epithelial)	CDH1			CCCCATCAGGCCTCCGTTTCTGGAA
2.4	cadherin 11, type 2, OB-cadherin (osteoblast)	CDH11	ACCAACACCCTCACCATCAAA	GGCCGGCGTTCAGAATG	TGCGGGTGCGACGTGAACG
2.4	frizzled-related protein	FRZB	CTGGACTAGCAAAGGAAAATTGCA	GCCCCAAACCATTACAAGG	AAGCAGAAACCAAAAGAG
2.4	frizzled-related protein	FRZB			GCAAGCAGTGAACGCTGTAAATGTA
2.4	frizzled homolog 1 (drosophila)	FZD1			CGGCGCCGGCGAGCTGTGCGTGGGC
2.4	frizzled homolog 2 (drosophila)	FZD2			CTCACAGGAGGAGACGCGTTTCGCG
2.4	frizzled homolog 7 (drosophila)	FZD7			CCCGTACTGCTTTCAGGTCTTGTA
2.4	syndecan 1	SDC1			CGGCCCTGCCGCAAAATTGTGGCTAC
2.4	secreted frizzled-related protein 2	SFRP2	CTTGAGTGCACCGTTTCC	CGTTTCCATTATGTCGTTGTCATCA	ACCTTTGGAGCTTCCCTCG
2.4	adenomatosis polyposis coli	APC			GAAAAACGAGCACAGCGAAGAATAG
2.4	Rho guanine nucleotide exchange factor (gef) 10	ARHGEF10			CATGCTGCCAGGGCACCAGCGGCTG
2.4	Homo sapiens axin (axin) mrna, partial cds.	AXIN1			CACGGACAGCAGCGTGGATGGGATC
2.4	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	CEACAM1	CAGCCCCACTTCACAGAGT	GGCGGGTTCAGAAGGTT	CTCACAGCCTCACTTCT
2.4	casein kinase 1, delta	CSNK1D	CGACAAGATACCTCTCGCATGT	GCCGTGGTGTTCGAAAGGA	CACCTCACAGAATAGC
2.4	casein kinase 1, epsilon	CSNK1E			GGCTATCCCTCGAATTCTCAACAT
2.4	catenin (cadherin-associated protein), alpha 1, 102kDa	CTNNA1			AGACAACAGGAATTGAAAGATGTTG
2.4	catenin (cadherin-associated protein), beta 1, 88kda	CTNNB1			TAGAAGCTGGTGAATGCAAGCTTT
2.4	dishevelled, dsh homolog 1 (drosophila)	DVL1			ACGAGCAGGCTCAGCAGCTCCACGG
2.4	glycogen synthase kinase 3 beta	GSK3B			TAGAAATAATCAAGGCTCTGGGAAC
2.4	Rho-associated, coiled-coil containing protein kinase 1	ROCK1			AATTGGATGAAGAGGGAAATCAAAG
2.4	vinculin	VCL			GAGTCTGAGCAGGCCACAGAGATGC
2.4	lymphoid enhancer-binding factor 1	LEF1			AGCAACGGACACGAGGTGCCAGAC
2.4	transcription factor 7 (T-cell specific, HMG-box)	TCF7			ACTGTGAGCTGTTTACCCACCCAT
2.4	transcription factor 7-like 2 (T-cell specific, HMG-box)	TCF7L2			GTGTCCAGTTCCCTCCCATATGG
2.4	epithelial membrane protein 3	EMP3			GTCAGCGAGAATGGCTGGCTGAAGG
2.4	interleukin 8	IL8	GGAATTGAATGGGTTTGCTAGAA	TGTGGATCCTGGCTAGCAGAC	TGGAATCCTGGATTTTTTCTGTTAAATCTG

Section	Gene Name or Primer Pair	Gene Symbol	Forward Primer Sequence 5' to 3'	Reverse Primer Sequence 5' to 3'	Probe Sequence/ Context sequence
2.4	jun b proto-oncogene	JUNB			GCAA
2.4	transcription factor 4	TCF4			GGTGGCGGCAGCTACTTTTCTGGTC
2.4	discs, large (Drosophila) homolog 5	DLG5	GGA GAA GCG GCC ATT TCG	CTC AAT AGC GTG CCG AGC AA	CTTCTTCATGCAAGATGGCCATCAC
2.4	cinculin	VCL			AAA GGA GAT CAC AGA AAA GAA CCG ACA
2.4	Rho guanine nucleotide exchange factor (gef) 10	ARHGEF10			CTG C
2.4	claudin 1	CLDN1			GAGTCTGAGCAGGCCACAGAGATGC
2.4	claudin 2	CLDN2	GAAGGTGCTGCTGAGGATAGAC	CCCATTCTGCCTTTGCTCAAT	CATGCTGCCAGGGCACCAGCGGCTG
2.4	claudin 3	CLDN3			TTTCTTCTGCAAGTCTGGCTATTT
2.4	claudin 4	CLDN4			CCAATGGCCAAAGTCA
2.4	claudin 7	CLDN7			CGGCCACCAAGGTGCTACTCCGC
2.4	occludin	OCLN			CGGGTTTAGAGGGGAGGGGCGAAGG
2.4	tight junction protein 1 (zona occludens 1)	TJP1			TGCTCGCCCTGTCCGCGCCTTGCA
2.4	tight junction protein 2 (zona occludens 2)	TJP2			CCACGCCGTTCTGAAATGGTTCA
2.4	tight junction protein 3 (zona occludens 3)	TJP3			GCCGCCAAGAGCACAGCAATGGAGG
2.4	cadherin 1, type 1, e-cadherin (epithelial)	CDH1			TTGGCTCCGCGCCCAGGCATGGAA
2.4	cadherin 11, type 2, OB-cadherin (osteoblast)	CDH11	ACCAACACCCTCACCATCAAA	GGCCGGCGTTCAGAATG	AGATTCTGCAGGTGAATGACGTGCC
2.4	catenin (cadherin-associated protein), alpha 1, 102kDa	CTNNA1			CCCCATCAGGCCTCCGTTTCTGGAA
2.4	catenin (cadherin-associated protein), beta 1, 88kda	CTNNB1			TGCGGGTGCGACGTGAACG
2.4	ErbB2 interacting protein	ERBB2IP			AGACAACAGGAATTGAAAGATGTTG
2.4	clathrin, heavy polypeptide (Hc)	CLTC			TAGAAGCTGGTGAATGCAAGCTTT
2.4	filamin a, alpha (actin binding protein 280)	FLNA			CCC GCCGAAAGAATGTTGGCTCAAT
2.4	integral membrane protein 3 (Interim)	ITM2C	GGGCCAAGAAGTCAATGC	GAGCGTCTCCACCAGAA	CATCTCCAGCTCCAGAACCTGGGTA
2.4	mucin 1, transmembrane	MUC1	GTGCCCCCTAGCAGTACCG	GAAGTGGCTGCCACTGCTG	GAACAGCAGCGCAACCTCTGCTCCC
2.4	neuregulin 1	NRG1			CCACTTCGAGAACACC
2.4	vinexin beta (sh3-containing adaptor molecule-1) (interim)	SORBS3			TGAGAAGGTTTCTGCAGGTAATGGTGGCA
2.4	deleted in malignant brain tumors 1	DMBT1			ACAAAAAAGCCAGGGAAAGTCAGAA
2.4	serine peptidase inhibitor, Kazal type 5	SPINK5			GATGGCTGGTTTGTGGGTGTCTCCC
2.4					AATCCGGAGTGGATGCGTGAGGGAT
2.4					ATGCCCCCGTCTGACGAATGACAGG

* AoD gene assays are provided with context sequence surrounding the assay location.

8.5 Abbreviations

Table 8.4. List of abbreviations used in thesis text

Abbreviation	Description
5-ASA	5-Aminosalicylate
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1
AGR2	anterior gradient 2 homolog (Xenopus laevis)
APC	adenomatous polyposis coli
ASCA	anti-Saccharomyces cerevisiae antibody
ASPN	asporin (LRR class 1)
ATP5F1	ATP synthase, H ⁺ transporting, mitochondrial F ₀ complex, subunit B1
BLAST	Basic Local Alignment Search Tool
BST2	bone marrow stromal cell antigen 2
C14orf43	chromosome 14 open reading frame 43
C2	complement component 2
C3	complement component 3
C5	complement component 5
CAI	colitis activity index
CARD15	caspase recruitment domain family, member 15
CBP	CREB-binding protein
CD	Crohn's disease
CD_I	inflamed Crohn's disease
CD_IS	inflamed Crohn's disease treated with immunosuppressive
CD_NI	non-inflamed Crohn's disease
CD11	ITGAL; integrin, alpha L (antigen CD11A (p180), lymphocyte function-associated antigen 1
CD14	CD14 molecule
CD19	CD19 molecule
CD79A	CD79a molecule, immunoglobulin-associated alpha
CDAI	Crohn's disease activity index
CDK2AP2	CDK2-associated protein 2
cDNA	complementary DNA
CFB	complement factor B
CHST5	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 5
CKI	casein kinase I
CSF2	colony stimulating factor 2 (granulocyte-macrophage)
CSNK1	casein kinase I gene family
CSNK1D	casein kinase 1, delta
Ct	number of cycles to reach threshold
CTNNB1	beta-catenin
DC	disease control
DEFA5	defensin, alpha 5, Paneth cell-specific
DEFA6	defensin, alpha 6, Paneth cell-specific
DKK	dickkopf
DLD	dihydrolipoamide dehydrogenase
DLG5	discs, large homolog 5 (Drosophila)
DNA	deoxyribonucleic acid
DSH/DVL	dishevelled
DSS	dextran sulfate sodium
EDTA	ethylenediaminetetraacetic acid
EGFR	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)
EPHB2	EPH receptor B2
EPHB3	EPH receptor B3
F	female
F11R	F11 receptor, CD321, JAM, JAM-1, JAM-A
FAM	6-carboxyfluorescein
FCGR3B	Fc fragment of IgG, low affinity IIIb, receptor (CD16b)

Abbreviation	Description
FDR	false discovery rate
FGFR2	fibroblast growth factor receptor 2 (bacteria-expressed kinase, keratinocyte growth factor receptor, craniofacial dysostosis 1, Crouzon syndrome, Pfeiffer syndrome, Jackson-Weiss syndrome)
FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog; c-fos
FTH1	ferritin, heavy polypeptide 1
FZ	Frizzled
g	gram
GNAI2	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2
GO	Gene Ontology
GSK3	glycogen synthase-3
GSN	gelsolin (amyloidosis, Finnish type)
HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
HLA-A	major histocompatibility complex, class I, A
HLA-B	major histocompatibility complex, class I, B
HLA-DMB	major histocompatibility complex, class II, DM beta
HLA-DRA	major histocompatibility complex, class II, DR alpha
HLA-E	major histocompatibility complex, class I, E
HOXB13	homeobox B13
HOXB6	homeobox B6
HSPG	heparan sulphate proteoglycans
IBD	inflammatory bowel disease
ICAM1	intercellular adhesion molecule 1 (CD54), human rhinovirus receptor
ICAM-1	intercellular adhesion molecule 1 (CD54),
IFNG	Interferon, gamma
IgG	immunoglobulin G
IGH@	immunoglobulin heavy locus
IGHA1	immunoglobulin heavy constant alpha 1
IGHD	immunoglobulin heavy constant delta
IGHG1	immunoglobulin heavy constant gamma 1 (G1m marker)
IGHM	immunoglobulin heavy constant mu
IGKC	immunoglobulin kappa constant
IGKV1-5	immunoglobulin kappa variable 1-5
IgL@	immunoglobulin lambda locus
IGLC2	immunoglobulin lambda variable 3-21
IPLL1	immunoglobulin lambda-like polypeptide 1
IL10	interleukin 10
IL12	interleukin 12
IL13	interleukin 13
IL2	interleukin 2
IL4	interleukin 4
IL4R	interleukin 4 receptor
IL5	interleukin 5 (colony-stimulating factor, eosinophil)
IQR	interquartile range
ITGAD	integrin, alpha D
ITGAL	integrin, alpha L (antigen CD11A (p180), lymphocyte function-associated antigen 1; alpha polypeptide)
ITGAM	integrin, alpha M (complement component 3 receptor 3 subunit)
ITGAX	integrin, alpha X (complement component 3 receptor 4 subunit)
JAK3	janus kinase 3 (a protein tyrosine kinase, leukocyte)
JNK	c-Jun N-terminal kinase
JUN	jun oncogene; c-Jun; AP1
L	liter
LEF	lymphoid enhancer-binding transcription factor
LGALS2	lectin, galactoside-binding, soluble, 2 (galectin 2)
LIMS	laboratory information management system
LRP5/6	lipoprotein-receptor-related proteins 5 and 6
m	milli; 10 ⁻³
μ	micro; 10 ⁻⁶
M	molar (mol/L)

Abbreviation	Description
MAS	microarray suite
MHC	major histocompatibility complex
min	minute
MOPS	3-(4-morpholino)propane sulfonic acid
mRNA	messenger RNA
MUC3	mucin 3A, cell surface associated
MYC	v-myc myelocytomatosis viral oncogene homolog (avian); c-myc
NC	normal control
NCBI	National Center For Biotechnology Information
NOD2	nucleotide-binding oligomerization domain 2; CARD15
NOS	nitric oxide synthase
NSAID	non-steroidal anti-inflammatory drug
OCTN	organic cation transporter
pANCA	anti-neutrophil cytoplasmic antibody
PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2
PARP8	soly (ADP-ribose) polymerase family, member 8
PCR	polymerase chain reaction
PKC	protein kinase C
PPARD	peroxisome proliferative activated receptor, delta
PRKCB1	srotein kinase C, beta 1
PSMB10	proteasome (prosome, macropain) subunit, beta type, 10
QSCN6	Quiescin Q6
RAB13	RAB13, member RAS oncogene family
RAB27A	RAB27A, member RAS oncogene family
RAB5C	RAB5C, member RAS oncogene family
RAB5C	RAB5C, member RAS oncogene family
RDH11	Retinol dehydrogenase 11 (all-trans/9-cis/11-cis)
RNA	ribonucleic acid
ROK	Rho kinase
rpm	revolutions per minute
RRAS2	related RAS viral (r-ras) oncogene homolog 2
RZPD	Deutsches Ressourcenzentrum fuer Genomforschung
s	second
S100A10	S100 calcium binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11))
SCFA	short chain fatty acids
SEMFs	subepithelial myoblasts
sFRP	secreted Frizzled-related protein
SLC11A2	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2
SLC22A4	solute carrier family 22 (organic cation transporter), member 4
SLC22A5	solute carrier family 22 (organic cation transporter), member 5
SPN	sialophorin (leukosialin, CD43)
SPON1	spondin 1, extracellular matrix protein
SRRM2	serine/arginine repetitive matrix 2
TBXA2R	thromboxane A2 receptor
TCF	T-cell-specific transcription factor
TET	6-carboxytetramethylrhodamine
Th1	type 1 helper T cell
Th2	type 2 helper T cell
Th3	suppressor T-cells
TLR5	Toll-like receptor 5
TNF	tumor necrosis factor (TNF superfamily, member 2)
TYK2	tyrosine kinase 2
UC	ulcerative colitis
UC_I	inflamed ulcerative colitis
UC_NI	non-inflamed ulcerative colitis
WFDC2	WAP four-disulfide core domain 2
WIF1	Wnt-inhibitory factor-1
xg	acceleration due to the force of gravity
y	year

9 Curriculum vitae

PERSONAL INFORMATION

Name: Nancy Mah
 Date of birth: October 1, 1970
 Place of birth: Regina, Canada
 Marital status: Single
 Nationality: Canadian

EDUCATION

2001-present Ph.D. candidate, Faculty of Mathematics and Natural Sciences, Institute for Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel, Germany: Characterization of novel genes involved in inflammatory bowel disease, Supervisor: Prof. Dr. Stefan Schreiber
 1996-1999 M.Sc. thesis, Faculty of Graduate Studies, University of British Columbia, Vancouver, Canada: Heterologous expression and purification of bovine rod photoreceptor glutamic acid-rich protein, Supervisor: Prof. Dr. Robert Molday
 1988-1993 B.Sc. with Specialization in Biochemistry, Department of Biochemistry, University of Alberta, Edmonton, Canada
 1990 French Summer Language Exchange Program, University of Trois-Rivières, Trois-Rivières, Quebec, Canada
 1988 Advanced High School Diploma with Excellence, Roland Michener Junior/Senior High School, Slave Lake, Alberta, Canada

WORK HISTORY

1999-2000 Research assistant, Institute of Human Genetics, University of Würzburg, Würzburg, Germany: Isolation and characterization of retina-specific transcripts
 1995-1996 Research assistant, Department of Botany, University of British Columbia, Vancouver, Canada: Characterization of cinnamate-4-hydroxylase from poplar trees
 1993-1995 Research assistant, Research and Development Group, Biomira Inc., Edmonton, Canada: Recombinant single chain antibody production in the yeast *Pichia pastoris*

PUBLICATION LIST

Zheng, W., Rosenstiel, P., Huse, K., Sina, C., Valentonyte, R., **Mah, N.**, Zeitlmann, L., Grosse, J., Ruf, N., Nurnberg, P., *et al.* (2006). Evaluation of AGR2 and AGR3 as candidate genes for inflammatory bowel disease. *Genes Immun* 7, 11-18.

Costello, C. M., **Mah, N.**, Hasler, R., Rosenstiel, P., Waetzig, G. H., Hahn, A., Lu, T., Gurbuz, Y., Nikolaus, S., Albrecht, M., *et al.* (2005). Dissection of the inflammatory bowel disease transcriptome using genome-wide cDNA microarrays. *PLoS Med* 2, e199. Epub 2005 Aug 2003.

Mah, N., Thelin, A., Lu, T., Nikolaus, S., Kuhbacher, T., Gurbuz, Y., Eickhoff, H., Kloppel, G., Lehrach, H., Mellgard, B., *et al.* (2004). A comparison of oligonucleotide and cDNA-based microarray systems. *Physiol Genomics* 16, 361-370. Epub 2003 Nov 2005.

Mah, N., Stoehr, H., Schulz, H. L., White, K., and Weber, B. H. (2001). Identification of a novel retina-specific gene located in a subtelomeric region with polymorphic distribution among multiple human chromosomes. *Biochim Biophys Acta* 1522, 167-174.

Ro, D. K., **Mah, N.**, Ellis, B. E., and Douglas, C. J. (2001). Functional characterization and subcellular localization of poplar (*Populus trichocarpa* x *Populus deltoides*) cinnamate 4-hydroxylase. *Plant Physiol* 126, 317-329.

Stohr, H., **Mah, N.**, Schulz, H. L., Gehrig, A., Frohlich, S., and Weber, B. H. (2000). EST mining of the UniGene dataset to identify retina-specific genes. *Cytogenet Cell Genet* 91, 267-277.

Luo, D., **Mah, N.**, Wishart, D., Zhang, Y., Jacobs, F., and Martin, L. (1996). Construction and expression of bi-functional proteins of single-chain Fv with effector domains. *J Biochem (Tokyo)* *120*, 229-232.

Luo, D., **Mah, N.**, Krantz, M., Wilde, K., Wishart, D., Zhang, Y., Jacobs, F., and Martin, L. (1995). VI-linker-Vh orientation-dependent expression of single chain Fv-containing an engineered disulfide-stabilized bond in the framework regions. *J Biochem (Tokyo)* *118*, 825-831.

CONGRESS ABSTRACTS

Mah, N., Costello, C. M., Haesler, R., Waetzig, G. H., Nikolaus, S., and Schreiber, S. (2005). Disruption of epithelial barrier function in inflammatory bowel disease. Poster presented at: National Genome Research Network Symposium (Kiel, Germany).

Mah, N., Costello, C. M., Haesler, R., Rosenstiel, P., Waetzig, G. H., Lu, T., Gurbuz, Y., Nikolaus, S., Kloppel, G., Eickhoff, H., *et al.* (2004). Molecular profiling of inflammatory bowel disease. Talk presented at: National Genome Research Network Symposium (Berlin, Germany).

Mah, N., Thelin, A., Costello, C. M., Lu, T., Gurbuz, Y., Eickhoff, H., Nikolaus, S., Lehrach, H., Mellgard, B., and Schreiber, S. (2002). Oligonucleotide versus cDNA-based microarray systems: a comparison using normal human colonic mucosa. Poster presented at: National Genome Research Network Symposium (Berlin, Germany).

Mah, N., Thelin, A., Costello, C. M., Lu, T., Gurbuz, Y., Mauracher, B., Eickhoff, H., Nikolaus, S., Mellgard, B., and Schreiber, S. (2002). Oligonucleotide versus cDNA-based microarray systems: a comparison using normal human colonic mucosa. Poster presented at: United European Gastroenterology Week (Geneva, Switzerland).

Mah, N., Thelin, A., Costello, C. M., Lu, T., Gurbuz, Y., Eickhoff, H., Nikolaus, S., Lehrach, H., Mellgard, B., and Schreiber, S. (2002). Oligonucleotide versus cDNA-based microarray systems: a comparison using normal human colonic mucosa. Poster presented at: German Human Genome Project Meeting (Leipzig, Germany).

Mah, N., Thelin, A., Costello, C. M., Lu, T., Gurbuz, Y., Eickhoff, H., Nikolaus, S., Mellgard, B., and Schreiber, S. (2002). Gene expression in human intestine—comparative analysis using two different microarray platforms. Poster presented at: National Genome Research Network Symposium (Kiel, Germany).

Mah, N., Thelin, A., Costello, C. M., Lu, T., Gurbuz, Y., Eickhoff, H., Nikolaus, S., Mellgard, B., and Schreiber, S. (2001). Expression profiling of normal human sigmoidal colon using microarray expression screening. Poster presented at: German Human Genome Project Meeting (Braunschweig, Germany).

Mah, N., H. Schulz, Stohr H, Weber, BHF (2000). Characterisation of two complex retinal transcripts with unknown functional properties. Poster presented at: German Human Genome Project Meeting (Heidelberg, Germany).

10 Declaration (Erklärung)

I was the primary scientist for the work done on the microarray comparison and real-time PCR work, substantially contributed to the cDNA IBD microarray study, and performed all of the microarray data analysis for the Affymetrix IBD microarray studies. This project was performed in collaboration with AstraZeneca R&D (Drs. Anders Thelin, Stefan Pierrou) and Boehringer Ingelheim Pharmaceuticals (Dr. Jun Li), all of whom were responsible for running the hybridization and scanning of the Affymetrix arrays. I am extremely grateful to all collaborators for their contributions to the project.

No part of this thesis has been submitted to any other board for another qualification. Most of the results have already been published (see below).

Kiel,(Nancy Mah)

PUBLICATIONS RELATED TO THIS THESIS

Costello CM, **Mah N**, Häslér R, Rosenstiel P, Waetzig GH, Hahn A, Lu T, Gurbuz Y, Nikolaus S, Albrecht M, Hampe J, Lucius R, Klöppel G, Eickhoff H, Lehrach H, Lengauer T and Schreiber S. Dissection of the inflammatory bowel disease transcriptome using genome-wide cDNA microarrays identifies novel candidate disease genes. *PLoS Medicine* 2: 0771-0787., 2005.

Mah N, Thelin A, Lu T, Nikolaus S, Kuhbacher T, Gurbuz Y, Eickhoff H, Kloppel G, Lehrach H, Mellgard B, Costello CM and Schreiber S. A comparison of oligonucleotide and cDNA-based microarray systems. *Physiol Genomics* 16: 361-370., 2004.

CONGRESS ABSTRACTS RELATED TO THIS THESIS

Mah, N., Costello, C. M., Haesler, R., Waetzig, G. H., Nikolaus, S., and Schreiber, S. (2005). Disruption of epithelial barrier function in inflammatory bowel disease. Poster presented at: National Genome Research Network Symposium (Kiel, Germany).

Mah, N., Costello, C. M., Haesler, R., Rosenstiel, P., Waetzig, G. H., Lu, T., Gurbuz, Y., Nikolaus, S., Kloppel, G., Eickhoff, H., *et al.* (2004). Molecular profiling of inflammatory bowel disease. Talk presented at: National Genome Research Network Symposium (Berlin, Germany).

Mah, N., Thelin, A., Costello, C. M., Lu, T., Gurbuz, Y., Eickhoff, H., Nikolaus, S., Lehrach, H., Mellgard, B., and Schreiber, S. (2002). Oligonucleotide versus cDNA-based microarray systems: a comparison using normal human colonic mucosa. Poster presented at: National Genome Research Network Symposium (Berlin, Germany).

Mah, N., Thelin, A., Costello, C. M., Lu, T., Gurbuz, Y., Mauracher, B., Eickhoff, H., Nikolaus, S., Mellgard, B., and Schreiber, S. (2002). Oligonucleotide versus cDNA-based microarray systems: a comparison using normal human colonic mucosa. Poster presented at: United European Gastroenterology Week (Geneva, Switzerland).

Mah, N., Thelin, A., Costello, C. M., Lu, T., Gurbuz, Y., Eickhoff, H., Nikolaus, S., Lehrach, H., Mellgard, B., and Schreiber, S. (2002). Oligonucleotide versus cDNA-based microarray systems: a

comparison using normal human colonic mucosa. Poster presented at: German Human Genome Project Meeting (Leipzig, Germany).

Mah, N., Thelin, A., Costello, C. M., Lu, T., Gurbuz, Y., Eickhoff, H., Nikolaus, S., Mellgard, B., and Schreiber, S. (2002). Gene expression in human intestine—comparative analysis using two different microarray platforms. Poster presented at: National Genome Research Network Symposium (Kiel, Germany).

Mah, N., Thelin, A., Costello, C. M., Lu, T., Gurbuz, Y., Eickhoff, H., Nikolaus, S., Mellgard, B., and Schreiber, S. (2001). Expression profiling of normal human sigmoidal colon using microarray expression screening. Poster presented at: German Human Genome Project Meeting (Braunschweig, Germany).

11 Acknowledgements

I would like to thank Prof. Stefan Schreiber for his support and encouragement and for giving me the opportunity to work in a very well-equipped scientific environment. His vision to increase our technical capacity through external collaboration partners was vital to the success of this project.

Thank you to Prof. Karin Krupinska from the Botanical Institute of the Christian-Albrechts University of Kiel for taking on the role of my official supervisor in the Faculty of Mathematics and Natural Sciences.

I would like to thank the head of the Department of General Internal Medicine of the University Clinic Schleswig-Holstein Campus Kiel, Prof. Dr. Ulrich R Fölsch, for the allowing me to work in his hospital.

I would like to thank Dr. Christine Costello, who produced the cDNA microarrays from scratch and first implemented the high-throughput real-time PCR platform. Her guidance and motivation have been very helpful during the course of my studies.

I am grateful to Dr. Anders Thelin and Dr. Stefan Pierrou from Astra Zeneca R&D for performing hybridizations on the Affymetrix arrays for the microarray comparison and the expression screening in CD patients. I would like to thank Dr. Jun Li from Boehringer Ingelheim Pharmaceuticals for his contributions on the expression profiling using Affymetrix arrays in UC.

The work presented in this Ph.D. thesis could not have been completed without the contributions of patients and the staff of the First Medical Clinic (University Clinic Schleswig-Holstein, Campus Kiel) and the Institute for Clinical Molecular Biology (Christian Albrechts University Kiel). In particular, I would like to thank Drs. Susanna Nikolaus, Tanja Kühbacher and Christian Sina for collecting patient material and data. I would like to thank Drs. Philip Rosenstiel and Ralph Lucius for providing immunohistochemistry support, and Carl Manaster, Marcus Will, Markus Teuber, Martin Kerick and Tim Lu for answers to my computer/bioinformatic questions. I would like to thank Brigitte Mauracher, Katja Tamms, Nicola Dierkes and Dorina Oelsner for excellent technical assistance in expression screening. Thank you to Prof. Monika Stoll, Dr. Georg Wätzig, and Dr. Andreas Fiebig for encouragement and Dr. Almut Nebel for critically reviewing my thesis.

Special thanks go to Tim Lu, Dr. Peter Croucher, Dr. Silvia Mascheretti-Croucher, Martin Kerick, and Dr. Robert Häsler for their interesting and encouraging scientific discussions. Thank you to the former staff of 'Ikea' for their service over the last few years.

Finally, thank you to my family and Michi for their support, food imports, and hockey updates during all the years I have been abroad.