

Functional relevance of epigenetic modifications in inflammatory bowel disease

Dissertation

in fulfilment of the requirements for the degree “Dr. rer. nat.”

of the Faculty of Mathematics and Natural Sciences

at Kiel University

submitted by

Zhe Feng

Kiel, 2013

First referee: Prof. Dr. Dr. Thomas C.G. Bosch.....

Second referee: Prof. Dr. Philip Rosenstiel.....

Date of the oral examination: 17.06.2013.....

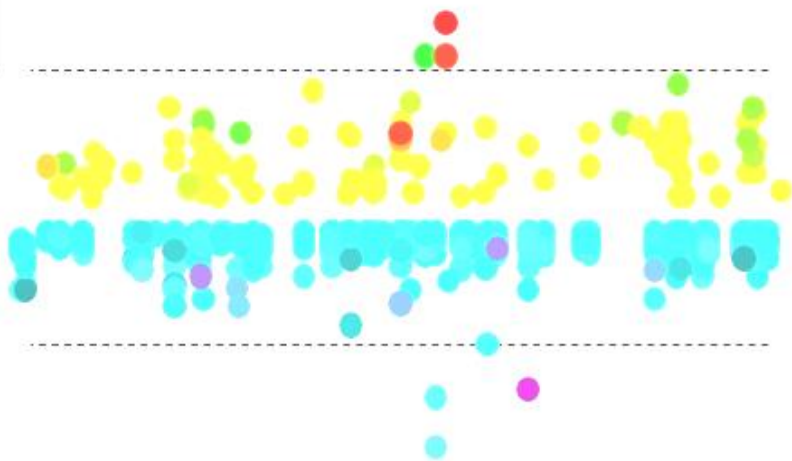
Approved for publication:

Signed: Prof. Dr. Wolfgang J. Duschl, Dean



"I only know that I know nothing."

(Socrates)



To my parents

Parts of this thesis have been published in Genome Research.

Häsler R, **Feng Z**, Bäckdahl L, Spehlmann ME, Franke A, Teschendorff A, Rakyán VK, Down TA, Wilson GA, Feber A, Beck S, Schreiber S, Rosenstiel P. A functional methylome map of ulcerative colitis. **Genome Res.** 2012 Nov;22(11):2130-7. doi: 10.1101/gr.138347.112. Epub 2012 Jul 23.

The final version of this publication can be found in the appendix (Page 120-127).

Table of Contents

1	INTRODUCTION	4
1.1	Ulcerative colitis: a multifactorial induced complex disorder	4
1.1.1	Epidemiology of ulcerative colitis	5
1.1.2	Clinical phenotype of ulcerative colitis	6
1.1.3	Epithelial barrier dysfunction and autoimmunity	7
1.1.4	Immunological aspects of UC	8
1.1.5	Genetic background of UC	9
1.1.6	Environmental factors	10
1.2	Epigenetic regulation in UC	11
1.2.1	Epigenetic mechanisms	13
1.2.2	DNA methylation patterns	14
1.2.3	Non-coding RNA and other epigenetic mechanisms	16
1.2.4	Epigenetics and environment	17
1.2.5	Twin studies	18
1.3	Hypothesis and Aims of the study	19
2	MATERIALS AND METHODS	20
2.1	Patient group composition and sample preparation	20
2.1.1	Patient Recruitment	20
2.1.2	Samples storage and preparation	21
2.1.3	RNA extraction from biopsy	22
2.1.4	DNA extraction from biopsies	22
2.2	Genome-wide scanning of the transcriptional and methylational profiling	24
2.2.1	Genome-wide Transcriptome Profiling (Layer I)	24
2.2.2	Genome-wide Quantification of Differentially Methylated Regions (Layer II)	25

2.2.3	Genome-wide Quantification of individual Methylation Variable Positions (Layer III)	27
2.3	Merging of the transcriptional and methylational data (from Layer I,II & III).....	29
2.4	Validation of the transcriptional and methylational candidates	29
2.4.1	Validation of differential mRNA expression via Real-Time PCR.....	30
2.4.2	Validation of differential DNAm via Pyrosequencing.....	32
2.5	Gene Ontology analysis	34
2.6	Chromatin- immunoprecipitation following quantitative real-time PCR	34
2.6.1	Preparation of samples and control DNA	34
2.6.2	Immunoprecipitation of 5-hydroxymethylcytidine enriched DNA fragments	35
2.6.3	Examination of 5-hydroxymethylcytidine enrichment via qRT PCR.....	35
3	RESULTS.....	37
3.1	Genome-wide profiling of transcription and DNA methylation in ulcerative colitis.....	37
3.1.1	Genome-wide transcriptome profiling (Layer I)	37
3.1.2	Genome-wide methylome profiling (Layer II and layer III)	37
3.1.3	Genome-wide quantification of individual methylated variable positions (Layer III).....	37
3.1.4	Merging of the transcriptional and methylational data (from Layer I,II & III)	38
3.1.5	Intra-class correlations of the three-layer dataset	40
3.1.6	Overview: genomic location of the different methylation in ulcerative colitis	41
3.1.7	Principle component analysis of the significant candidates	41
3.1.8	Correlation between the expressional and methylational data (in cis)	42
3.1.9	Gene ontology (GO) analysis of the disease relevant processes	43
3.2	Validation of selected candidate transcripts under potential epigenetic control.....	44
3.3	Potential new finding of the DNA hydroxymethylation in the acute inflammation	49
4	DISCUSSION.....	52
4.1	Genetics and epigenetics in the pathogenesis of ulcerative colitis	52

4.2	Environment and epigenetics in the pathogenesis of ulcerative colitis	56
4.3	Ulcerative colitis: a complex epigenetic disorder	57
4.4	Genome-wide patterns of hypomethylation are a prominent motive in UC	59
4.5	DNA demethylation in chronic and acute colitis	60
4.5.1	Passive mechanisms of DNA demethylation are potentially absent in UC	62
4.5.2	Demethylation in mammalian cells: a multi-step process	62
4.5.3	The APOBEC Family: Mediators of 5-mC or 5-hmC Deamination	63
4.5.4	Reduced BER activity may lead to carcinogenesis of UC.....	64
4.6	Concluding remarks and future studies	66
5	SUMMARY	67
6	ZUSAMMENFASSUNG	68
7	REFERENCES	69
8	APPENDIX	87
9	CURRICULUM VITAE	130
10	ACKNOWLEDGEMENTS	132
11	DECLARATION	134

1 Introduction

1.1 Ulcerative colitis: a multifactorial induced complex disorder

Several inflammatory diseases of the human mucosa, which have been unknown before industrialization (Bach, 2002), have significantly increased in the Western countries over the recent 100 years. One example is inflammatory bowel disease (IBD) (Abraham and Cho, 2009b). A similar phenomenon has been observed in Asian countries, where hospitalized IBD patients were reported with an increasing frequency during industrialization (Wang et al., 2007). The transformation from an agrarian into an industrial society induced enormous alteration of life-style and environment. Keeping this in mind, the altered mucosa immunity and the luminal composition in the gut that is due to the drastic changes in sanitation, medication, and nutrition, could be a key trigger of an inflammatory barrier disorder (Rook, 2012), but the precise contribution to the disease pathogenesis of these factors are still unknown.

In contrast to environmental factors, it is known that individual genomes differ in millions of variations, some of which might cause different phenotypes or even pathogenesis (Boomsma et al., 2002; Kaminsky et al., 2009). However, complex diseases, such as IBD, might require many predisposing variants (Schreiber et al., 2005). Prior to the first genome-wide association studies (GWAS) in 2005, various genetic factors had been reported to be associated with IBD, including the IBD5 locus, the nucleotide-binding oligomerization domain containing 2/ caspase recruitment domain family, member 15 (NOD2/CARD15) gene, MHC locus (Brant and Shugart, 2004). Since the advent of GWAS, a large number of additional genetic loci have been found in the context of IBD. So far, genome-wide meta analysis confirmed 71 susceptibility loci for Crohn's disease, and 47 risk loci which associated with ulcerative colitis (UC) (Franke et al., 2010; Anderson et al., 2011).

The current GWAS on IBD were mostly based on cohorts of European descent. Studies targeting eastern populations are still lacking. Interestingly, some IBD studies reported that variants, which were identified in the Europeans, are not found in other Asian IBD patients (Ng et al., 2011). This could be partially explained by the fact that the European and Asian populations display substantial genetic and environmental differences, which is of clinical relevance, as the current understanding of disease pathogenesis assumes a complex interplay between environmental and genetic factors (Xavier and Podolsky, 2007).

1.1.1 Epidemiology of ulcerative colitis

UC represents the major subphenotype (OMIM 191390) of IBD (OMIM 266600). It is characterized by an abnormal inflammatory reaction in the gut mucosa. Considerable variation in the epidemiology of UC has been observed when comparing the data worldwide, yet the current understanding suggests that industrialization might be an element in the etiology of UC (Benchimol et al., 2011; Molodecky et al., 2012). Molodecky and colleagues summarized the incidence and prevalence of UC in 2012, and demonstrated the increase over time and in different regions around the world.

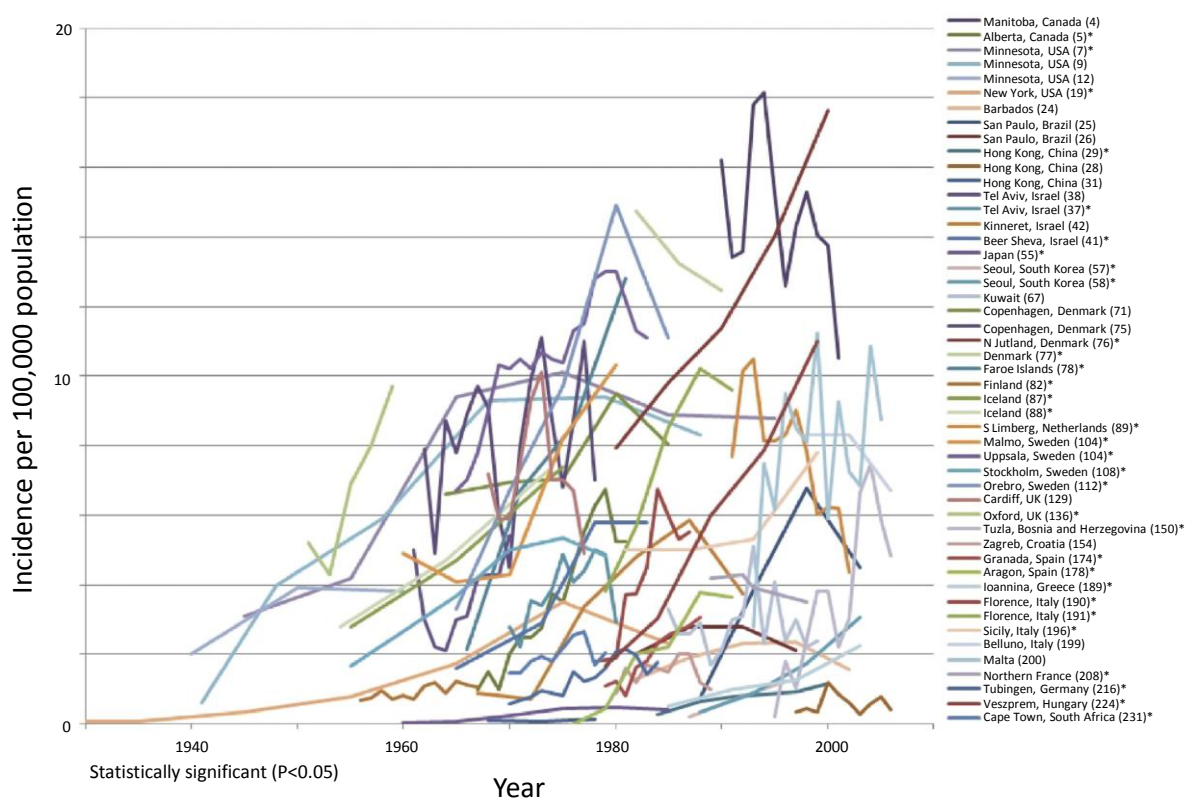


Figure 1. Temporal trends of incidence rates. Data for studies that reported at least 10 years and with at least 3 time points for ulcerative colitis (UC). from Molodecky et al., 2012.

Several complex disorders have been reported that the incidence rates could be stratified by sex. However, no data suggests that the UC has gender special effects. In contrast to that, the age-stratified incidence rates attracted more attention. Some studies estimated the highest incidence rates are found in the range from 20 to 29 years, and that the pediatric incidence is increasing constantly (Benchimol et al., 2011). The results of early onset and chronicity are high treatment costs for the patient, the health care system, and the society.

UC rarely occurs in developing countries, however, the incidence of UC has increased rapidly in these countries in the recent half century. Interestingly, it was reported that individuals from developing regions emigrating to industrialized countries have a higher risk for developing UC, especially among the first-generation children (Barreiro-de Acosta et al., 2011).

1.1.2 Clinical phenotype of ulcerative colitis

Ulcerative colitis is classified as a complex phenotype. Initial symptoms mostly include diarrhea and abdominal pain. However, diarrhea as a lead symptom is unspecific, and stool samples show no pathological findings. UC can show clinical symptoms and an endoscopic picture similar to other infectious colitis diseases. The diagnosis of UC is mainly based on clinical, endoscopic, radiologic, and histological criteria (Nikolaus and Schreiber, 2007).

Ulcerative Colitis	
Symptom	
Abdominal pain	Cramps, mainly left lower quadrant
Diarrhea	Frequent in adults; can alternate with constipation
Hematochezia	Always in active patients; intensity related to disease activity
Abdominal mass	Left lower quadrant, if sigmoid is inflamed in slim individuals
Hyposomia	Rare
Malnutrition	Occasional
Abdominal distention	Only in severe disease
Obstructive symptoms	No
Perianal disease/fistula	No
Laboratory abnormalities	
Acute reactant proteins (eg, CRP)	In extensive or severe disease
Anemia	In severe disease
Macrocytosis	Rare
Hypoalbuminemia	In severe disease
pANCA	in UC ++ (in UC with PSC +++)
ASCA	(+)
Complications/extraintestinal manifestations	
Abscess	Rare
Toxic megacolon	Rare
Ileus	Rare
Primary sclerosing cholangitis	Occasional (5%–15%)
Hepatitis	Occasional
Erythema nodosum	Rare
Pyoderma gangraenosum	Rare
Arthralgia/arthritis	Frequent
Epidemiology	
Current smoker	Rarely
Former smoker	Frequently
Previous appendectomy	Less than expected

Ulcerative Colitis

Endoscopic appearance of disease

Distribution	Continuous spread from the rectum
Small bowel involvement	Rare (back-wash ileitis)
Rectal involvement	Almost always
Uniform, continuous disease	Always
Longitudinal, polycyclic ulcers ("snail track")	No
Cobblestone appearance of ileum	No
Normal mucosa within inflamed areas	No
Strictures	Rare, always suspicious for carcinoma!
Mucosal edema	Frequent
Ulceration	Often flat and extensive
Circumferential inflammation	Frequent

Imaging

Increased bowel wall thickness	Moderate
Mesenteric lymph nodes	Infrequent
Mesenteric fat wrapping	No

Table 1. Diagnosis of ulcerative colitis (UC). Clinical, endoscopical, radiological, and histological characteristics of UC. According to Nikolaus and Schreiber, 2007.

UC patients have a significant higher risk to develop primary sclerosing cholangitis and vice-versa (Cullen and Chapman, 2003) as well as colorectal cancer (Glória et al., 1996; Konishi et al., 2007; March et al., 2011). UC is also associated to other immune-mediated disorders such as arthritis, and multiple sclerosis (Satsangi et al., 1997).

1.1.3 Epithelial barrier dysfunction and autoimmunity

The gut epithelia builds an integrate physical and biochemical barrier and maintains a dynamic balance between the body and the environment. The commensal environmental components in the gut can induce the immune response at the interface, even across the barrier, directly activating immune system (Cho and Blaser, 2012). Unlike in CD patients, inflammation in UC typically occurs in proximity to the epithelium and does not affect the deeper intestinal tissues, therefore the barrier layer should be the main battlefield between the immune system and gut microbiota, and the integrity of the barrier layer might be implicated in the pathogenesis of UC (Kaser et al., 2008). In addition, autoimmunity might play a role in ulcerative colitis. The autoantibody p-ANCA (perinuclear Anti-neutrophil cytoplasmic antibody) has been detected in the blood test of UC patient. It is thought to be against the colonic epithelial antigen. For example tropomyosin 5, a structural protein, might be a putative target of p-ANCA (Geng et al., 1998), but evidence of classical antibody-mediated autoimmunity in ulcerative colitis is still lacking.

1.1.4 Immunological aspects of UC

Individual health depends on a beneficial host–microbe interaction (Duncan and Edberg, 1995). The most intimate contact with the outside world takes place at the surface of the intestine, particularly in the colon, which harbors a greater amount and higher diversity of microorganisms than any other organ (Shanahan, 2002). These potential 'pathogens' are recognized by toll-like receptors (TLRs) and/or caspase recruitment domain (CARDs) containing proteins on/in the epithelial and immune cells, therefore the innate immune response to the microbiota must be controlled correctly and effectively to maintain the intestinal homeostasis (Lee and Kim, 2007). Similarly to Crohn's disease (CD), no specific pathogens were identified in ulcerative colitis, but abnormalities of innate immune such as variants of caspase recruitment domain family, member 9 (CARD9) (Bertin et al., 2000; Glocker et al., 2009; Strasser et al., 2012), increased expression of TLR2 and TRR4 by colonocytes, are involved in microbial recognition and inflammatory signaling. Interestingly, the secretion of proinflammatory cytokines, such as interleukin-1 β , interleukin-6 and tumor necrosis factor α (TNF- α) is not only increased in patients with IBD, but also found in many other inflammatory diseases (Panja et al., 1998; Kaiser et al., 2000).

The humoral and cellular adaptive immunity exhibits various potentially pathological alterations in ulcerative colitis. Immunoglobulins (Ig), such as IgM, IgA, and IgG levels are commonly elevated, but there is a disproportionate increase in IgG1 antibodies in ulcerative colitis (Takahashi and Das, 1985). The differentiation of helper T cells (Th) into Th1 and Th2 has been reported to be altered in IBD. This represents an atypical type 2 helper T cells (Th2) response in the gut mucosa of patients with ulcerative colitis. Interleukin-13 is secreted from the nonclassical natural killer T cells in the colon, which mediates epithelial cell cytotoxicity, apoptosis, and epithelial barrier dysfunction (Fuss et al., 2004; Heller et al., 2005). Interleukin-5–producing Th2-polarized T cells are generally present in affected tissue of ulcerative colitis patients. However, the upregulation of Interleukin-4 (IL-4) levels, which represents a Th2 signature was not observed in UC (Kaser et al., 2010). IL17 which is secreted by recently discovered Th17 cell-lineages is significantly more abundant in UC when compared to controls (Sakuraba et al., 2009). Some studies indicated that although the Th17 response has the potential to be a proinflammatory response, it seemed to be playing mostly a regulatory role rather than an functional role in the inflammation (Waite and Skokos, 2012).

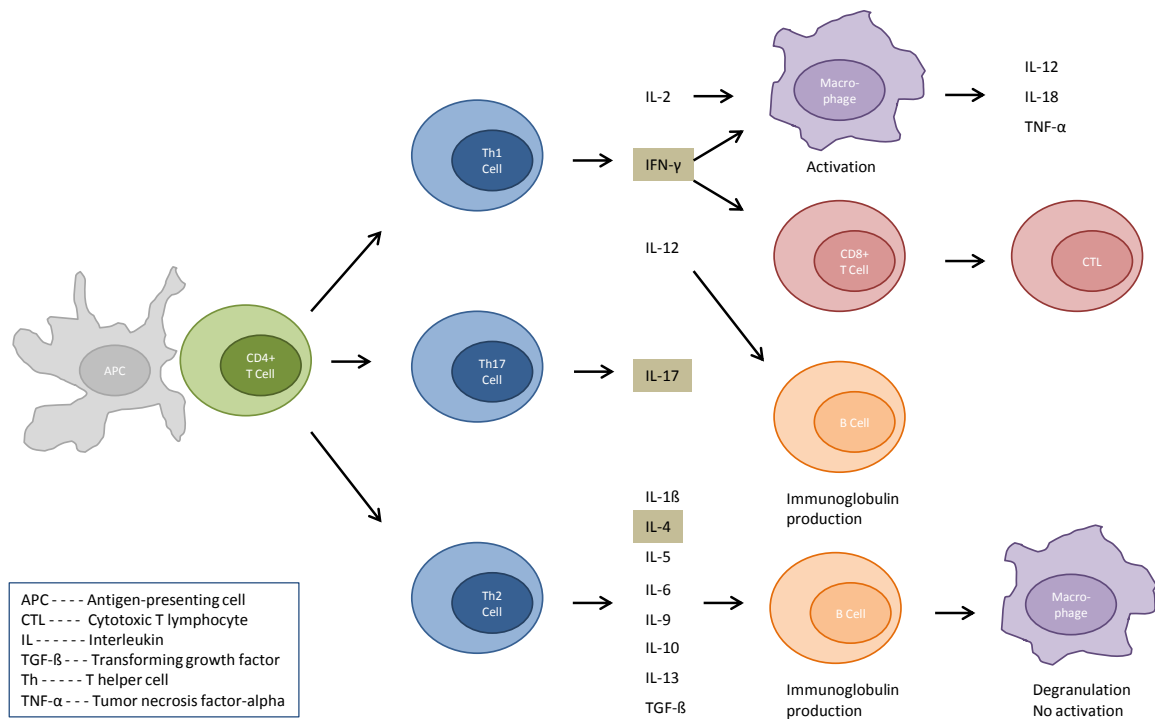


Figure 2. The T helper cell differentiation. The brown marked cytokines are the main signatures of Th1 cells, Th2 cells and Th17 cells respectively. However, in the colonic epithelial layer of ulcerative colitis, it represented an atypical Th2 differentiation, because the upregulation of Interleukin-4 (IL-4) levels has not been observed.

1.1.5 Genetic background of UC

UC is thought to have a prominent heritable component (Duerr, 2007). Prior to the first genome-wide association studies (GWAS) in 2005, genetic factors have been reported to be associated with IBD, including the IBD5 locus, the NOD2/CARD15 gene, and the MHC locus (Brant and Shugart, 2004). Since the advent of GWAS, a large number of genetic variants have been identified to be potentially disease relevant in IBD. A meta-analysis of six studies recently confirmed the presence of 47 loci associated with ulcerative colitis, of which 19 are specific for ulcerative colitis and 28 are shared with Crohn's disease (Franke et al., 2010; Anderson et al., 2011).

Risk loci, which have been identified in the meta-analysis (Anderson et al., 2011) and in individual studies suggested that several pathways are involved in susceptibility, manifestation and progression of UC. Variants of extracellular matrix protein 1 (ECM1) (Merregaert et al., 2010), hepatocyte nuclear factor 4, alpha (HNF4A) (Cattin et al., 2009), cadherin type 1 (CDH1) (Liu et al., 2009), and

laminin beta 1 (LAMB1) (Lee and Gross, 2007) might lead to a dysfunction of the epithelial barrier. The product of DAP (Bialik and Kimchi, 2010) plays very important role in apoptosis and autophagy; and altered PR domain containing 1 with ZNF domain (PRDM1) (Bikoff et al., 2009), interferon regulatory factor 5 (IRF5) (Barnes et al., 2001), and NK2 homeobox 3 (NKX2-3) (Yu et al., 2011) suggest impacts on transcriptional regulation. Interleukin-23 receptor (IL23R), Janus kinase 2 (JAK2), signal transducer and activator of transcription 3 (STAT3), Interleukin-12 beta (IL12B), and protein tyrosine phosphatase non-receptor type 2 (PTPN2) which are shared risk loci with CD are functionally linked to the IL-23 signaling pathway (Abraham and Cho, 2009a; Duvallet et al., 2011), and CARD9 might be involved in immunity to Mycobacterium tuberculosis and Candida albicans (Bertin et al., 2000; Bi et al., 2010; Strasser et al., 2012). particularly several risk loci linked to CD and other autoimmune diseases, such as IL-10, IL7R, IL23R, and Interferon- γ (IFN- γ), which participate helper T-cell types 1 and 17 (Th1 and Th17) differentiation (Atarashi and Honda, 2011).

1.1.6 Environmental factors

Host factors are thought to play a very important role in the pathogenesis of ulcerative colitis. Similarly, environmental factors should not be ignored. The recent worldwide trends in UC epidemiology indicated that the environmental factors might significantly contribute to the pathogenesis of UC, yet quantification of such contributions remains difficult (Bernstein et al., 2010).

Acid, bile and phasic 'housekeeping' motility patterns, which are secreted into the intestinal lumen, can hinder microbial colonization. However, numerous bacteria still can grow in the distal small intestine (Mahida and Rolfe, 2004). The exact number and species cannot be determined with the currently available technologies, moreover, the gut flora is believed to be constantly altered due to substantial environmental changes which are partially fast and irreversible (Benchimol et al., 2011; Molodecky et al., 2012; Rook, 2012). Similar changes could be observed in diet: low-fiber, high-sugar and high animal-fat food composition became widespread (O'Sullivan and O'Morain, 2006). Normally, the mucosal immune system is required to respond to pathogens, while maintaining appropriate tolerance to the commensal bacteria and diet antigens. Current studies indicate that this balance is disrupted in UC (Shanahan, 2002; Mahida and Rolfe, 2004; Danese and Fiocchi, 2011).

Smoking is suggested to be associated with IBD pathogenesis, but interestingly, smoking seems to have a protective effect in UC (Cosnes et al., 2004). A meta-analysis in 1989 revealed a pooled odds ratio of 0.41 (0.34-0.48), indicating that NOT smoking and ulcerative colitis are consistent with a causal relationship (Calkins, 1989). Mahid and colleagues reported a similar analysis result in 2006 and concluded that current smokers have a lower risk for UC (OR 0.58; 95% CI 0.45-0.75), while

former smokers has a higher risk (OR 1.79; 95% CI 1.37-2.34) (Mahid et al., 2006). In a reverse manner, individuals who smoked in their younger years showed a dramatically increased risk for developing UC (Mahid et al., 2007).

Hygiene is thought to be a potential 'pathogenic' factor for UC, given the rising incidence of UC after the industrialization in both, developed and developing countries. The 'IBD hygiene hypothesis' states that the excessively hygienic environment negatively affects immune development of children. Many factors could be involved in this concept, e.g. non-contaminated food, cleaner water, toothpaste, smaller family size, low bacterial load on indoor-surfaces. Many studies revealed that most of these factors are associated with the pathogenesis of UC (Kendler and Baker, 2007). This gradual development has been termed "losing our old friends", such as bacteria, virus and parasites (Rook, 2012).

1.2 Epigenetic regulation in UC

Normally, multicellular organism have identical DNA sequences in each cell (disregarding somatic mutations), however, differences between various cellular phenotypes are maintained by different expressions patterns (Jaenisch and Bird, 2003). This nongenetic program, which records developmental and environmental cues, is the basis of epigenetics. Epigenetics are defined as a heritable mechanism which alters gene expression without altering the DNA sequence (Ptashne, 2007).

In eukaryotic cells, DNA is wrapped around histone proteins and packaged in the nucleus, where repeating units of nucleosomes build a so-called chromatin. Several modifications of amino termini of nucleosomal histones (H) are known, such as methylation and acetylation (Barski et al., 2007; Bartke et al., 2010; Margueron and Reinberg, 2010). Similarly, DNA nucleotides can be methylated on the 5' position of the cytosine (C) pyrimidine ring (Robertson and Wolffe, 2000; Bird, 2002; Bartke et al., 2010). All these modifications potentially influence gene expression (Figure 3). In addition, interactions between the genome and other components in the nucleus, e.g. RNA and proteins, may regulate the gene expression (Bernstein and Allis, 2005; Henikoff, 2008; Nagano et al., 2008; Lee, 2009; Simon and Kingston, 2009).

In a multicellular organisms, the unique gene-expression patterns of each cell determines the identity of its type, its development stage (Takahashi and Yamanaka, 2006; Ieda et al., 2010; Cortázar et al., 2011; Iqbal et al., 2011) and its reaction to internal and external stimuli (Jaenisch and Bird, 2003; Bonfanti et al., 2010; Carone et al., 2010) - this cellular identity is controlled by epigenetic

mechanisms. The molecular basis of epigenetics is complex, and it is known that even single-cell organisms do not have one single epigenome, but instead have multiple epigenomes. These epigenetic mechanisms (Figure 3) work throughout the genome, and interact with each other, to organize the genome in both the one- and the three-dimensional space (Schones and Zhao, 2008).

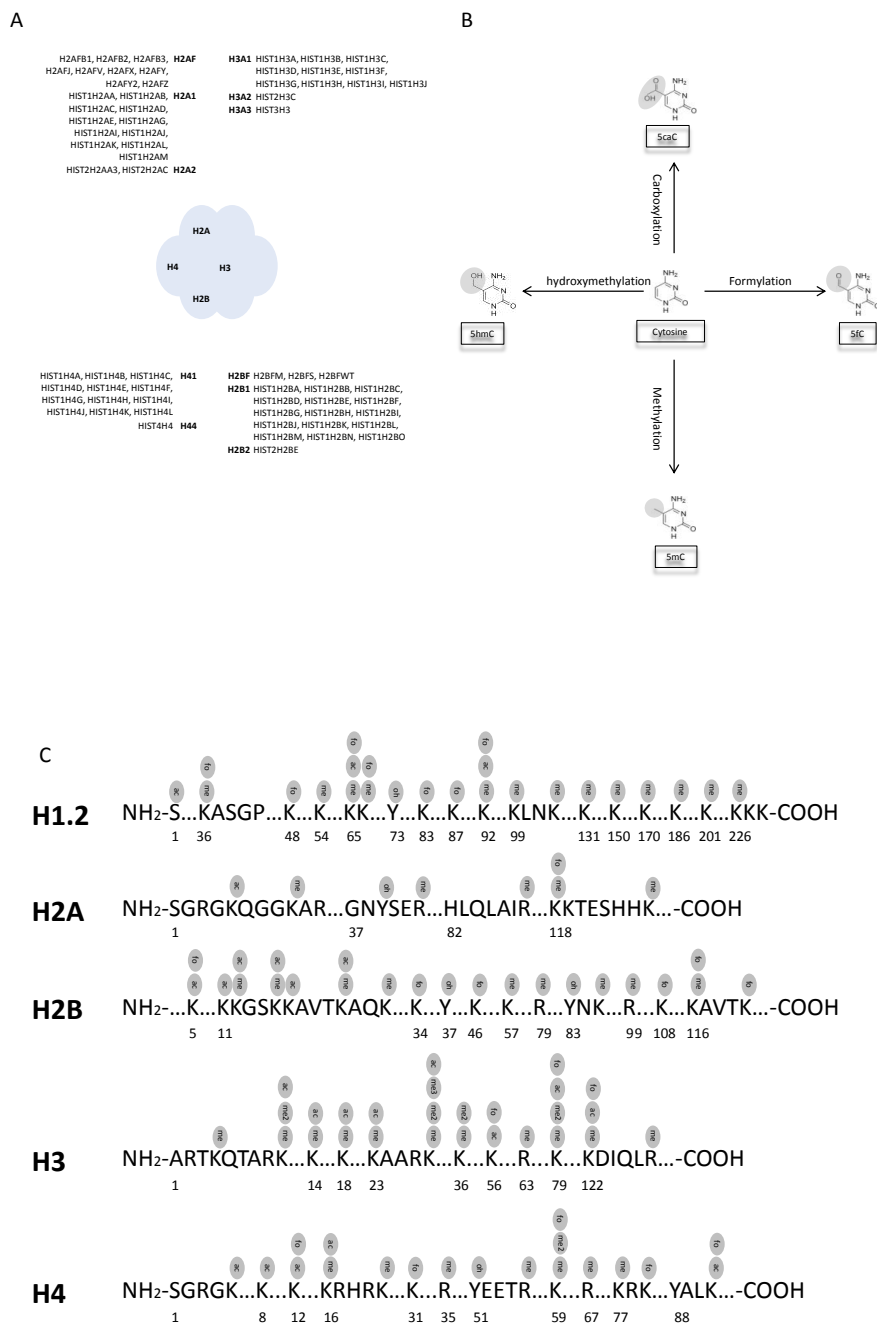


Figure 3. Characteristics of epigenomes. A: DNA methylation, B: histone modification, and C: nucleosome positioning and other factors such as small RNAs contribute to an overall epigenome that regulates transcription and influences the architecture of the genome.

1.2.1 Epigenetic mechanisms

In eukaryotes, nuclear DNA is wrapped around a histone-octamers which consists of two copies of H2A, H2B, H3 and H4. This complex forms nucleosomes which are further compacted into higher-order chromatin structures (Albert et al., 2007; Bartke et al., 2010). Based on the current knowledge, each chromatin is characterized by the presence of histone variants and modifications at the amino-terminal end of the histone proteins, such as acetylation (ac), methylation (me), phosphorylation (p) and ubiquitylation (ub) (Barski et al., 2007; Zeng et al., 2010).

The epigenetic properties of nucleosomes can be differentiated by the incorporation and location of histone variants (Albert et al., 2007; Henikoff, 2008). The H3 is one canonical core histone protein, deposited throughout the chromosome. In the centromere, the H3 is substituted by the centromere-specific histone H3 variant (CenH3), which might be involved in the initiation of chromosome replication (Furuyama et al., 2006). Another important variant of H3 is H3.3, which is thought to be a central player in maintaining epigenetic inheritance at transcriptionally active regions (Tagami et al., 2004). Most H3.3 variants mark the boundaries of active chromatin modifications, such as di- and trimethylation of K4 (H3K4me2 and H3K4me3), dimethylation of K79 (H3K79me2) and acetylation of multiple lysines, while negative modifications, e.g. H3K9me2 are absent. Interestingly, Ahmad and colleagues removed the N-terminal tail of H3.3, but could not document any significant effect on chromosome assembly, which indicates that terminal modifications might have no influences to the localization of H3.3 (Ahmad and Henikoff, 2002).

H2A histone family, member Z (H2A.Z) represents another kind of histone variant. A genome-wide study with budding yeast indicated that it is enriched at tightly packed chromosomes (heterochromatin) (Meneghini et al., 2003). Further studies with human cells revealed that H2A.Z is localized at enhancers and promoters, with a positive correlation between occupancy and transcriptional activity at promoters (Barski et al., 2007). Current evidence indicates that H2A.Z may be functionally linked to transcriptional deactivation. Moreover, H2A.Z might play an important role in the chromosomal stability. In addition, the crystal structure of H2A.Z indicated that it could increase nucleosome mobility by destabilizing nucleosome structure (Suto et al., 2000).

Histone modifications have been suggested to have functional relevance in transcription regulation and genome architecture (Guelen et al., 2008). Numerous genome-wide studies revealed that H3K4me1 and histone acetylation are mostly enriched at transcription start sites (promoter regions and beginning exons) and correlate positively with expression levels. In contrast to that, H3K4me3 seems to silence transcriptions activity. Modifications, which take place at H3K9, H3K27, and H4K20,

might be involved in heterochromatin formation, which is associated with transcriptional repression. The picture is complicated by opposing effects: H3K9me3 and H4K20me3 were detected both silent heterochromatic regions and at active zinc finger (ZNF) genes, respectively (Barski et al., 2007). These modifications are normally co-localized in the same genomic regulatory regions and termed 'bivalent domains'. In embryonic stem cells, it has been suggested that they may provide the potential for both transcriptional activation and repression during differentiation (Bernstein et al., 2006). In addition to that, a similar effect was also found for H2A ubiquitylation which can poise the RNA polymerase II in mouse embryo stem cells in a bidirectional manner (Stock et al., 2007).

1.2.2 DNA methylation patterns

Currently, DNA methylation refers to the addition of a methyl group to the 5' residues of cytosine in the context of cytosine followed by guanine (CpG) dinucleotides. In eukaryotes, DNA methylation has also been found at other bases in diverse sequence contexts (Robertson and Wolffe, 2000; Paulsen and Ferguson-Smith, 2001; Klose and Bird, 2006).

It is known that the CpGs dinucleotides are distributed asymmetrically in the human genome, which features CpG-rich (so-called CpG Island) and CpG-poor regions. Not all CpGs are methylated (Doi et al., 2009; Irizarry et al., 2009). Thirty years ago, it was hypothesized that patterns of DNA methylation exist, and that these patterns vary in different cell types. DNA methylation (DNAm) and its related proteins were suggested to alter expression through interfering the binding affinities of transcription factors (Holliday and Pugh, 1975; Riggs, 1975). In 1980s, Yisraeli and colleagues methylated the promoters of beta-globin gene with purified methylases, and they found that transcription of beta-globin gene was strongly inhibited in mouse fibroblasts (Yisraeli et al., 1988). However, most genes that have CpG-rich promoters lack methylation in this CpG-rich region, yet the gene is silent. One example for this phenomenon are globin genes in non-erythroid cells (Bird et al., 1987). In contrast to that, it was reported that thousands of heavily methylated CpG islands have been identified in oocytes and pre-implantation embryos, and these hypermethylations are preferentially located within the active transcriptional units (Smallwood et al., 2011).

DNA methylation plays essential roles in many key processes, such as genomic imprinting, X-chromosome inactivation and suppression of repetitive elements (Robertson and Wolffe, 2000). In the recent years, DNA methylation was mostly studied in the context of coding regions and potentially regulatory elements, however, current evidence suggests that DNA methylation is conditionally dynamic (Ito et al., 2010; Ficz et al., 2011; Pastor et al., 2011; Wu et al., 2011a), even

outside such regions (Zhang et al., 2006; Alexander et al., 2010; Medvedeva et al., 2010), yet the details of this are still largely unknown.

The variation of DNA methylation in complex diseases has been mainly been investigated in the context of tumor and autoimmune diseases. In oncological scenarios, the general hypothesis is that tumor development is associated with gain or loss of DNA methylation at key elements. In particular, the reprogramming procedure increases the mutation risk and therefore might be involved in the initiation of carcinogenesis (Rai et al., 2010). For non-malignant diseases, studies on discordant monozygotic twins indicated that variations of DNA methylation are associated to disease pathogenesis, such as systemic lupus erythematosus(SLE) (Javierre et al., 2010) and autism-spectrum disorders (Nguyen et al., 2010).

DNA methylation was thought to be stable, however differentiated cellular states can exhibit radically altered epigenetic patterns (Yamanaka and Blau, 2010; Jullien et al., 2011). Recent stem cells studies showed that 5' methylated cytosines (5-mC) in mammalian cells can be hydroxymethylated (5-hydroxymethylcytosine, 5-hmC) during the differentiation (Tahiliani et al., 2009; Ficz et al., 2011; Gu et al., 2011; Williams et al., 2011a; Wu et al., 2011a).

5-hmC was reported originally in the 1950s (Wyatt and Cohen, 1952). It was suggested to be strongly association with methylation, however, detailed knowledge of the required enzymes catalyzing this process in mammalian cells was unknown for decades, until Tahiliani and colleagues discovered the enzymatic activity of the ten-eleven translocation (TET) gene family catalyzing the underling the conversion of 5-mC to 5-hmC. Recent studies indicate that hydroxylation plays an essential role in the epigenetic reprogramming of stem cells, whereas it acts indirectly to mediate active DNA demethylation (Tahiliani et al., 2009).

It has been shown that 5-hmC is abundant in pluripotent embryonic stem cells (ESCs), but it has also been found at lower levels in well differentiated tissues (Branco et al., 2011). 5-hmC seems to activate a transcription which was silenced by methylation (Tahiliani et al., 2009; Ito et al., 2010) and is involved in several other biological processes, however, the role and functional significance of this modification is still largely unknown (Ficz et al., 2011; Pastor et al., 2011; Williams et al., 2011a; Wu et al., 2011a).

Hydroxymethylated Cytosine can be demethylated completely through DNA repair mechanisms. Hydroxyl-groups can be further deaminated by the AID/APOBEC family, finally replaced via DNA repair.

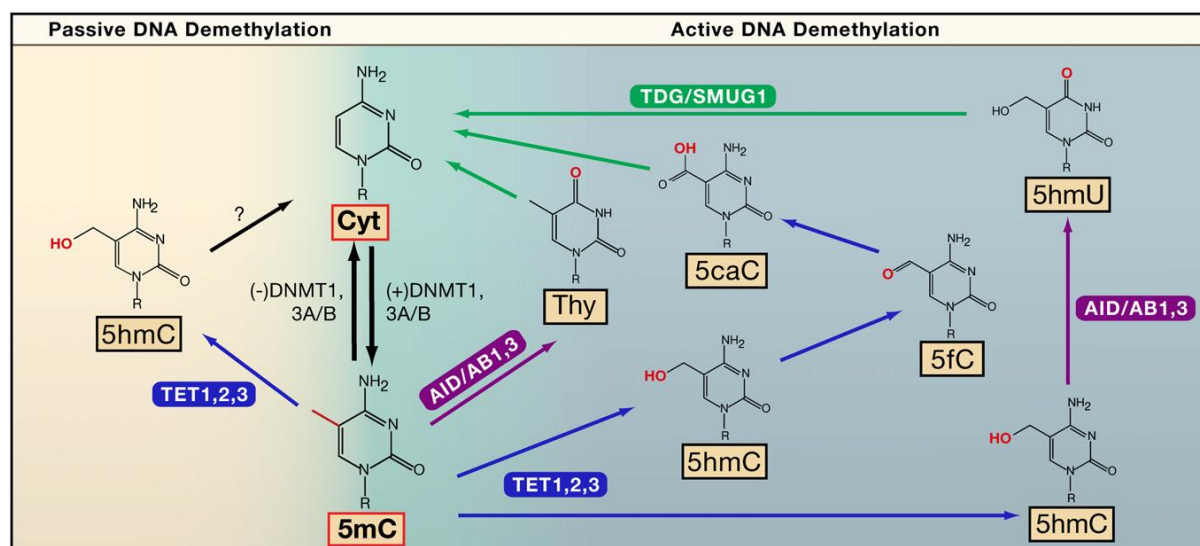


Figure 4. DNA Demethylation Pathways. The absence of DNA methyltransferases (DNMTs) regulates passive DNA demethylation. The ten-eleven translocation (TET) family of enzymes hydroxylates 5-mC to form 5-hydroxymethylcytosine (5-hmC). The AID/APOBEC family members deaminates 5-mC (or 5-hmC) to 5-methyluracil (5-mU) or 5-hydroxymethyluracil (5hmU) which are demethylated completely through the UDG family of base excision repair (BER) glycosylases. Diagram by Bhutani et al., 2011.

1.2.3 Non-coding RNA and other epigenetic mechanisms

In addition to DNA and histone modifications, non-coding RNA (ncRNA) is another important epigenetic element. While 2% of the genome are protein coding regions (Alexander et al., 2010), about 60-70% of the genome is transcribed into non-coding RNA (Washietl et al., 2005; Mattick and Makunin, 2006). Recently, various non-coding transcripts have been described to play crucial functional roles in normal physiology and in the context of disease (Mercer et al., 2009).

Currently, most studies in the field of ncRNAs focus on micro RNAs (miRNAs), which are small ncRNAs of approx. 22 nucleotides that mediate post-transcriptional gene silencing by degrading messenger RNAs (mRNAs) (Pasquinelli, 2012). In human cancer, miRNAs can act as oncogenes or as tumour suppressors and therefore can have key functions in tumorigenesis. For example, the miR-200 family regulates the epithelial-to-mesenchymal transition (Davalos et al., 2011). miR-15 and miR-16 are dysregulated in most B cell chronic lymphocytic leukaemias and result in chromosome 13q14 deletion (Calin et al., 2002).

PIWI-interacting RNAs (piRNAs) are 24–30 nucleotides long. Unlike miRNA, piRNAs are dicer-independent, but bind the PIWI subfamily of argonaute family proteins instead. The piRNA and PIWI protein complex that is formed by piRNAs and PIWI proteins suppresses transposable element expression and mobilization. Interestingly, the piRNA might serve as a navigating element in the

base-pairing recognition (Gunawardane et al., 2007). The mechanisms of piRNAs and PIWI proteins are mostly unknown. It was suggested that they might be involved in potentially disease-relevant processes, such as carcinogenesis (Lee et al., 2006; Taubert et al., 2007; Liu et al., 2010; Sun et al., 2011), stem cell self-renewal (Sharma et al., 2001) and DNA repair (Wang et al., 2011).

lncRNAs (long non-coding RNAs) are normally longer than 200 nucleotides and are involved in many biological processes. Various mechanisms have been proposed by which lncRNAs regulate the expression at transcriptional level. At the human Homeobox (HOX) loci, various lncRNAs are expressed sequentially, temporally or spatially, which regulate genes expression through modulation of chromatin accessibility (Rinn et al., 2007). Similar to the navigational function, X-inactivation specific transcript (XIST) lncRNA (17 kb) recruits the polycomb complex to silence the X chromosome. Its anti-sense lncRNA TSIX is transcribed from the opposite strand to XIST that regulates XIST transcription during X-chromosome inactivation (Navarro et al., 2006).

Another class of lncRNAs are lincRNAs (long intergenic non-coding RNAs), which are transcribed from intergenic regions. One well-known example of lncRNAs is HOX transcript antisense RNA (HOTAIR), which is expressed in the anti-sense region of HOX genes. It is involved in polycomb protein retargeting throughout the genome, and associated with cancer metastasis (Gupta et al., 2010).

Ultraconserved regions (UCRs) transcribe numerous ncRNAs. So far, 481 ncRNAs were described to origin from UCRs, some of which extend to coding regions. However, more than half of them do not translate into any protein. Transcribed ultraconserved regions (T-UCRs) are expressed in various tissues, while some are specific to cell lines, whereas the function of T-UCRs is still mostly unknown. Interestingly, it was reported that some T-UCRs can bind to miRNAs and might be involved in the carcinogenesis (Calin et al., 2007).

1.2.4 Epigenetics and environment

Epigenetic patterns do not persist throughout life, although the epigenomes, inducing DNA methylation and histone modification, are believed to be stable (Reik, 2007; Kota and Feil, 2010; Law and Jacobsen, 2010). Epigenetic alterations in the context of stem cell reprogramming during differentiation are well studied. However, it is thought that epigenetic changes are regulated by both - environmental factors and intrinsic factors (Feil and Fraga, 2011).

The significant impact of environmental factors on epigenetic patterns is exemplified in honeybees, in which differential nutrition determines whether the bee will be a worker or a queen. Evidence indicates that nutrition-determined phenotype in honey bees is highly associated with DNA methylation (Kucharski et al., 2008). A recent study with *Arabidopsis thaliana* revealed that different ambient temperatures alter the enrichment of a variant of histone H2A.Z incorporation, and nucleosomes containing the alternative histone H2A.Z are essential to perceiving ambient temperature correctly. Nucleosomes containing variants of H2A.Z- therefore result in altering the transcriptome in response to the ambient temperature (Kumar and Wigge, 2010). In mammals, the available S-adenosylmethionine (SAM) is the methyl donor for DNA and histone methylation. The amount of SAM is determined by the one-carbon-metabolism pathway. Numerous nutrients can affect this pathway, such as folate, vitamin B6, vitamin B12, betaine, methionine and choline. It was reported that histone methylation might be positively correlated with the availability of methyl donors (Zhou et al., 2008). However, dietary components do not only provided reagents, but also influence the activity of enzymes linked to epigenetic mechanisms, such as butyrate, which potentially inhibits the histone deacetylases (HDACs) (Dashwood and Ho, 2007). Another example in mammals is sirtuins, a class of HDACs that target acetylated histones which simultaneously require nicotinamide adenine dinucleotide (NAD⁺) (Imai et al., 2000; Vaquero et al., 2004). It is therefore supposed that the enzymatic activity of sirtuins might depend on nutritional and metabolic factors (Kaeberlein et al., 2005; Beher et al., 2009). Moreover, a recent study showed a strong association between DNA methylation and histone methylation, while sirtuin 1 (SIRT1) can deacetylate DNMT1 and alter its activity (Peng et al., 2011).

1.2.5 Twin studies

For multifactorial traits or complex diseases, genome-wide association studies have shown that genetics contribute significantly to the variation observed on population level. Various of these phenotypes significantly increased over the last decades along with sudden social and environmental changes (Willett, 2002). Twin studies provide considerable evidence that risk factors aggregate in siblings attributed to shared genes and shared environment (Bouchard et al., 1999; Eaves et al., 1999; Koopmans et al., 1999; van den Bree et al., 1999).

Naturally, twins have a higher phenotypic resemblance than unrelated individuals. Particularly, monozygotic (MZ) twins have always attracted the curiosity of researchers, as MZ twins derive from a single fertilized egg and therefore are traditionally regarded as genetically identical (Boomsma et al., 2002). Despite the fact that disease-discordant MZ twins have completely identical genetic background, age, sex, similar maternal influences, and even other environmental factors,

interestingly, rates of disease discordance in MZ twins are usually higher than explainable by heritability (Silman et al., 1996; Kendler and Prescott, 1999; MacGregor et al., 2000). This indicates that mechanisms like epigenetics might be involved in the bivalent pathogenesis (Healthy<--->Disease) of discordant MZ twins (Petronis, 2010).

One of the earliest epigenetic studies of MZ discordant twins was performed to interrogate different methylation in the dopamine D2 receptor gene (DRD2) in two schizophrenic discordant twins. In this study, greater methylation differences were observed between discordant MZ twins than between unrelated individuals (Petronis et al., 2003). Further studies identified various other phenotype-associated methylation changes in discordant MZ twins (Mill et al., 2006; Oates et al., 2006; Kuratomi et al., 2008). These evidences showed that MZ twin studies are a powerful approach in detecting disease-related epigenetic changes and probably can detect effects, which cannot be identified by employing traditional studies with unrelated individuals.

1.3 Hypothesis and Aims of the study

Current understanding assumes that both, genetic and environmental factors, are involved in the pathogenesis of ulcerative colitis, yet the etiology of ulcerative colitis is largely unclear. The disease complexity poses a major challenge when identifying mechanisms of pathological relevance. In this context, epigenetic modifications might represent a major interface between internal and external factors.

➤ **Hypothesis:**

The underlying hypothesis of the study is therefore, that epigenetic modification of disease relevant genes results in altered gene expression with pathological consequences, contributing to disease mechanisms.

➤ **Aims:**

- The main aim of the study is to present a high-resolution map of epigenetic modifications with potentially disease relevant effects on the transcriptome of ulcerative colitis. Integrating these results into the current picture of ulcerative colitis disease pathophysiology may help to close the gap between genetic susceptibility, missing heritability and disease manifestation.
- Subsequently, the alteration of DNA methylome in ulcerative colitis has been observed and investigated. DNA demethylation may play a important role in the pathogenesis of ulcerative colitis, and need further investigation.

2 Materials and Methods

The initial part of this project was completed as a collaborative effort between Institute of Clinical Molecular Biology, University of Kiel and the Sanger Institute, University of Cambridge, while the downstream analysis and validation was carried out exclusively at the Institute of Clinical Molecular Biology as part of this thesis.

2.1 Patient group composition and sample preparation

2.1.1 Patient Recruitment

The twenty monozygotic twins, discordant for ulcerative colitis (screening panel; median age: 25, range 18-70) recruited for this study were tested for mono/dizygosity as previously published (Barbaro et al., 2004; von Wurmb-Schwark et al., 2005).

The panel for real time PCR validation of transcript levels consisted of 135 unrelated individuals (validation panel I; n=30 ulcerative colitis, inflamed; n=30 ulcerative colitis, non-inflamed, n=30 healthy individuals; n=15 disease controls, inflamed; n=30 disease controls, non-inflamed; median age: 41 ; age range 18-76), a subgroup of which was used for pyrosequencing validation of methylation levels (validation panel II; n=20 ulcerative colitis, inflamed; n=20 healthy individuals, n=10 disease controls; inflamed, median age: 41, age range 18-68). Another subgroup of which was used for Chromatin immunoprecipitation following quantitative real time PCR (ChIP - qRT PCR) examination of hydroxymethylation levels (Examination panel; n=5 disease controls, inflamed; n=5 ulcerative colitis, inflamed and not inflamed from same patient respectively; median age: 52, age range 25-64).

Healthy individuals included in the study were undergoing colonic cancer surveillance with no previous unspecific changes in stool habits, where endoscopic and histological examination yielded no significant pathological findings. Ulcerative colitis patients were selected to display an endoscopically active disease in the sigmoid colon at the time of sampling. More than 2000 patients were screened to recruit the study population. Disease-specificity controls included individuals with infectious diarrhea, other forms of gastrointestinal inflammation or irritable bowel syndrome. The study setup was approved by the Bioethical Committee of the University of Kiel, where the patients were recruited. All patients gave written informed consent before data and biomaterials were collected. Patient characteristics are summarized in table.

Panel type	Application	Relationship of Individuals	Gender	Age: median, range	Disease representation
Screening panel	Genome-wide screening of the transcriptome and methylome	discordant twins (UC, HN)	10 f	25, 18-70	10 UC
			10 m		10 NC
Validation panel I	validation of the findings in the transcriptome	unrelated individuals	69 f 66 m	41, 18-76	30 UC _i ; 30 UC _{ni} 15 DC _i ; 30 DC _{ni} 30 NC
Validation panel II	validation of the findings in the methylome	unrelated individuals	25 f 25 m	41, 18-68	20 UC 10 DC 20 NC
(€ panel I)					
Examination panel III	Examination of the predictions in the hydroxymethylome	unrelated individuals	5 f 5 m	52, 25-64	5 UC _i ; 5 UC _{ni} 5 DC

Table 2. Study panels I, used for genome-wide assessment of differential expression and DNA methylation as well as for validation and examination of initial findings. f = female, m = male, UC = ulcerative colitis, NC = normal controls, DC = disease specificity controls, suffix: i= inflamed, ni= not inflamed.

2.1.2 Samples storage and preparation

All reagents, glassware and laboratory utensils were specially treated in order to minimize RNA and DNA degradation: Solutions were prepared with 0.1% Diethylpyrocarbonate (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 Grad Celsius (°C) for 12-16 hours (h) before use, in order to inactivate any contaminating nucleases. All plastic ware were purchased as ultraviolet (UV) -sterilized consumables or nuclease-free consumables.

All biopsies used in this study are primary tissue from the intestinal mucosa. Biopsies were taken under endoscopy from a defined area of the sigmoid colon, and immediately snap-frozen in liquid nitrogen. To minimize premature thawing of the biopsies during RNA and DNA isolation, homogenization equipments (mortar, pestle, polyethylene tubes) were cooled with liquid nitrogen prior to the begin of the isolation.

2.1.3 RNA extraction from biopsy

A commercial RNA isolation kit was used to extract total RNA from biopsies (RNeasy mini-kit, Qiagen), according to the manufacturer's protocol. Briefly, prior to the isolation, DNase stock solution was prepared by reconstituting lyophilized DNase enzyme (Qiagen) by the addition of 550 microliter (μL) of RNase-free water. This DNase stock solution ($10 \mu\text{L}$, 2.7 Kunitz units/ μL) was further diluted in $70 \mu\text{L}$ RDD buffer (as provided in RNase-free DNase Kit). One mucosal biopsy was homogenized into a fine powder using a cooled Teflon mortar in a 1.5 milliliter (mL) microfuge tube. The powder was then mixed with $650 \mu\text{L}$ of RLT lysis solution, incubated at room temperature for 10 minutes (min) and vortexed for 30 seconds (s). The lysate was then further centrifuged through a QIA shredder column for 2 min at 14000 rpm in a microfuge. The eluate was centrifuged for 3 min at 14000 rounds per minute (rpm) 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. The mini-column was washed once with $350 \mu\text{L}$ of RW1 buffer on to the column, incubated for 10 min at room temperature and then centrifuged 15 s at 10000 rpm. DNase treatment of the RNA was carried out while the RNA was still bound on the column. The diluted DNase ($80 \mu\text{L}$) was pipetted directly on to the spin column membrane and allowed to incubate at room temperature for 15 min. Upon completion of the digestion step, the mini-column was washed once with $350 \mu\text{L}$ buffer RW1 and twice with $500 \mu\text{L}$ buffer RPE. Each wash step was completed by centrifuging the tube 15 s at 10000 rpm. To completely rid the column of wash solution, the column was then centrifuged 1 min at 14000 rpm. Total RNA was eluted from the mini-column into an RNase-free microfuge tube by pipetting $50 \mu\text{L}$ 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.

2.1.4 DNA extraction from biopsies

A commercial kit was used to isolate total DNA from biopsy material (QIAamp Tissue DNA preparation kit, Qiagen), according to the manufacturer's protocol. Briefly, one mucosal biopsy was homogenized into a fine powder using a cooled Teflon mortar in a 1.5 mL microfuge tube. $100 \mu\text{L}$ Buffer ATL was added, and homogenized with rotor-stator homogenizer. $20 \mu\text{L}$ proteinase K was added, and mixed by vortexing, then incubated on a rocking platform at 56°C until the tissue is completely lysed. The 1.5 ml microcentrifuge tube then centrifuged briefly in order to remove drops from the inside of the lid. As RNA free genomic DNA samples are required for this project, for the RNase-treatment, $4 \mu\text{L}$ RNase A (100 mg/ml) was added at first, then mixed by pulse-vortexing for 15

s, and incubated for 2 min at room temperature. The tube was centrifuged to remove drops from inside the lid before adding 200 μ l Buffer AL to the sample. The sample mixed again by pulse-vortexing for 15 s, and incubated at 70°C for 10 min. 200 μ l ethanol (96–100%) was added to the sample, and mixed by pulse-vortexing for 15 s, the tube was then centrifuged to remove drops from inside the lid. The mixture was carefully applied to the QIAamp Mini spin column (in a 2 ml collection tube) without wetting the rim. Closing the cap, and centrifuged at 8000 rpm for 1 min. The QIAamp Mini spin column then placed in a new 2 ml collection tube, and the tube containing the filtrate was discarded. The QIAamp Mini spin column was washed with 500 μ l Buffer AW1 without wetting the rim and centrifuged at 8000 rpm for 1 min, then placed in a new 2 ml collection tube, and the collection tube containing the filtrate was discarded. Similarly, the QIAamp Mini spin column was washed with AW2 at 14,000 rpm for 3 min. The QIAamp Mini spin column was placed in a new 2 ml collection tube and the old collection tube with the filtrate was discarded. After centrifuging at full speed for 1 min, the QIAamp Mini spin column was placed in a clean 1.5 ml microcentrifuge tube, 50 μ l distilled water was added and incubated at room temperature for 1 min, then centrifuged at 8000 rpm for 1 min.

2.1.4.1 RNA and DNA quantity and quality control

There are several ways to quantitate pure solutions of RNA or DNA samples, however, the most commonly used technique is the determination of absorbency in a UV spectrometer at 260/280 nm or at 260 nanometer (nm) respectively. Briefly, the distilled H₂O was used as a solvent to suspend the nucleic acids, and place each sample in a quartz cuvette. The spectrophotometer must be blanked with a sample of solvent before use. For more accurate readings of the nucleic acid sample, it could be diluted to give readings between 0.1 and 1.0.

As all of the downstream applications in this project were performed with fragmented genomic DNA, the integrity of DNA samples was not examined specially. However, the isolation of intact RNA is essential for the ongoing analysis. The quality control of RNA samples was performed with a microcapillary electrophoresis system (Agilent 2100 Bioanalyzer). Briefly, the gel was prepared at first. 550 μ l of RNA 6000 Nano gel matrix was pipetted into a spin filter, and centrifuged at 14,000 rpm for 10 minutes at room temperature, then 65 μ l filtered gel pipetted into 0.5 ml RNase-free microfuge tubes. Second, the gel-dye mix was prepared. The RNA 6000 Nano dye concentrate was equilibrated to room temperature for 30 min, then RNA 6000 Nano dye concentrate was vortexed for 10 seconds and spun down. 1 μ l of dye was added into a 65 μ l aliquot of filtered gel and vortexed, centrifuged at 12,000 rpm for 10 min at room temperature. Third, the Gel-Dye Mix was loaded on the RNA chip. A new RNA 6000 Nano chip was fixed on the chip priming station. 9.0 μ l of

gel-dye mix was pipetted in the corresponding well, then the chip priming station was closed and the plunger was pressed until it was held by the clip. The clip was released after exactly 30 seconds, then the plunger was slowly pulled back to 1ml position. The chip priming station was then opened and 9.0 μ l of gel-dye mix pipetted in the corresponding wells. The remaining gel-dye mix was discarded. Forth, the Agilent RNA 6000 Nano Marker was loaded. 5 μ l of RNA 6000 Nano was pipetted in all wells, except Gel-Dye Mix wells. At the end, the ladder and samples were loaded on the chip. 1 μ l ladder was pipetted in the ladder well, and 1 μ l of sample was pipetted in each of the 12 sample wells. The chip was horizontally placed in the adapter of the IKA vortexer and vortexed for 1 min at 2400 rpm. Finally, the Agilent 2100 bioanalyzer was started to run the analysis, resulting in one electropherogram per sample.

2.2 Genome-wide scanning of the transcriptional and methylational profiling

All of the genome-wide scanning were performed using the samples from ten pairs monozygotic discordant twins, see table 3 for detail.

Panel type	Application	Relationship of Individuals	Gender	Age: median, range	Disease representation
Screening panel	Genome-wide screening of the transcriptome and methylome	discordant twins (UC, HN)	10 f	25, 18-70	10 UC
			10 m		10 NC

Table 3. Study panels II, used for genome-wide assessment of differential expression and DNA methylation. f = female, m = male, UC = ulcerative colitis, NC = normal controls, suffix: i= inflamed, ni= not inflamed.

2.2.1 Genome-wide Transcriptome Profiling (Layer I)

Total RNA was prepared and hybridized to an Affymetrix UG 133 Plus 2.0 (Affymetrix, Santa Clara, California, US) according to the manufacturer's protocol. Experimental and analytical part of the microarray analysis was performed following the MIAME standards.

2.2.1.1 RNA samples labeling and target hybridization

GeneChip[®] reagent kits was used to prepare biotinylated target from purified RNA samples, which is suitable for hybridization to GeneChip expression probe arrays. Briefly, double-stranded cDNA was synthesized from total RNA isolated from tissue according to the protocol provided in the kit. An in vitro transcription (IVT) reaction was then performed to produce biotin-labeled cRNA from the cDNA. The cRNA was fragmented before hybridization according to the manufacturer's instructions.

A hybridization cocktail was prepared, including 15 micrograms (μg) fragmented and labeled cRNA, 5 μL control oligonucleotide B2 (3 nM), 150 μL 2X hybridization Mix, 30 μL DMSO, and nuclease-free Water to final volume of 300 μL . Probe array was equilibrated to room temperature immediately before use. Then, the hybridization cocktail was heated to 99°C for 5 minutes in a heat block, and at the same time, the array was wet with an 300 μL of Pre- Hybridization Mix by filling it through one of the septa. The probe array filled with pre-hybridization mix was incubated at 45°C for 10 minutes with rotation. The hybridization cocktail was transferred that has been heated at 99°C, to a 45°C heat block for 5 minutes. The hybridization cocktail was centrifuged at maximum speed in a microcentrifuge for 5 minutes to collect any insoluble material from the hybridization mixture. The array was vented with a clean pipette tip and the Pre-Hybridization Mix was extracted from the array with a micropipettor. Then the array was refilled with the appropriate volume of the clarified hybridization cocktail. Probe array was then placed into the hybridization oven rotated at 60 rpm and incubated at 45°C for 16 hours.

2.2.1.2 Bioinformatic analyses

Data was normalized using GCRMA (R, Bioconductor) and signals that were not present in at least 80% of the samples (cutoff: detection p-value ≤ 0.05) were excluded from further analysis. Experimental and analytical part of the microarray analysis was performed following the MIAME standards. Differentially expressed genes were determined using the Mann-Whitney U-test, multiple testing correction was performed using the Benjamini-Hochberg method (Benjamini and Hochberg 1995) and a false discovery rate (FDR) for the signed fold changes (which were based on the ratios of the medians of each group compared) was estimated based on a Westfall and Young permutation, using K=5000 permutations (Westfall and Young 1993). Criteria for transcripts to be categorized as differentially expressed were set to: i) corrected p-value ≤ 0.05 and ii) FDR $\leq 5\%$.

2.2.2 Genome-wide Quantification of Differentially Methylated Regions (Layer II)

DNA was prepared and hybridized to a custom tiling array (Nimblegen, custom 385k array) as previously described (Rakyan et al., 2008). The array was designed to cover known autoimmune/inflammatory linked Loci as well as specific genes with immune-regulatory function and encompassed all known promoters and CpG islands (both promoter- and non-promoter-CpG islands). Data was normalized applying the interquantile normalization using Spotfire for functional genomics (TIBCO, Palo Alto, CA, USA). Differences between ulcerative colitis patients, diseased controls and healthy individuals were determined using the Mann-Whitney U-test, while p-values

were corrected according to Benjamini and Hochberg (Benjamini and Hochberg 1995). DMRs with a corrected p-value ≤ 0.05 were considered significantly differentially methylated.

2.2.2.1 Immunoprecipitation of methylated DNA

DNA concentration in the cell is relatively low. A ligation-mediated PCR (LM-PCR) step was induced to amplify the samples. Comparison of array hybridization performance before and after LM-PCR indicated that the LM-PCR did not introduce significant amplification bias (Down et al., 2008). Briefly, 2.5 μg of genomic DNA was fragmented to a size range of from 300 to 800 bp, then blunt-ended by incubation for 20 min at 12°C in a 120- μL reaction tube containing the DNA sample, 1 \times Buffer 2 (NEB), 10 \times BSA (NEB), 100 μM dNTP mix, and T4 DNA polymerase (NEB). The reaction products were purified with a Zymo-5 kit (Genetix) according to the manufacturer's instructions, while the final elution was carried out in 30 μL of TE buffer (pH 8.5). Ligation of the adaptors was performed by incubating overnight at 16°C in a final volume of 100 μL containing the DNA sample, 40 μL adaptors, T4 DNA ligase 10 \times buffer, 5 μL of T4 DNA ligase (NEB). The reactions were purified with the Zymo-5 kit as described above. In order to fill in the DNA-overhangs, the sample DNA was incubated at 72°C for 10 min in a reaction containing the DNA, 100 micromol (μM) dNTPs, 1 \times AmpliTaq Gold PCR buffer (Applied Biosystems), 1.5 millimol (mM) MgCl_2 , 5 U AmpliTaq Polymerase. The DNA was purified with the Zymo-5 kit as described above. A total of 50 ng of the ligated sample was set aside as the input fraction; 1.2 μg of the ligated DNA sample was subjected to a methyl-DNA immunoprecipitation (MeDIP) assay (Nimblegen) that uses antibodies specific for 5-methyl-cytosine residues according to the manufacturer's instructions. The immunoprecipitated (IP) sample was also purified with Zymo-5 kit (using 700 μL of binding buffer) according to the manufacturer's instructions. 10ng of each IP and input fraction for each sample were amplified with PCR (20 cycles) using the Advantage-GC genomic PCR kit (Clontech). After the LM-PCR, the duplicate reactions were combined, purified with a Qiagen PCR-clean up kit (Qiagen), and eluted in 50 μL H_2O . The MeDIP and input fractions were sent to Nimblegen for hybridization.

2.2.2.2 Array design

The microarray consists of 382,178 50-bp anti-sense oligonucleotides. Although the aim was to target all annotated TSSs and nonpromoter CpG islands (CGIs), it is impossible to design enough suitable unique probes for 18% of the Transcriptions Start Sites (TSSs) and 28% of nonpromoter CGIs at that time, largely due to the presence of repeat elements. However, the array contained 50-bp probes tiled at ~ 100 bp density throughout the entire human Major Histocompatibility Complex, and promoters and nonpromoter CGIs on the X and Y chromosomes. In addition, the array was

originally designed with NCBI build 35 version of the human genome assembly, but then mapped to NCBI build 36 using Exonerate (Slater and Birney, 2005). To be mapped, probes were normally required to align full-length and without gaps or mismatches so that the probes which aligned more than once to the NCBI36 sequence were removed for further analysis. Tiled regions were defined by clustering uniquely mapped probes within 200 bp of one another. Singleton probes were discarded. The tiled regions were then divided into 500-bp ROIs.

2.2.2.3 Bioinformatic analysis

Data was normalized applying the interquartile normalization with Spotfire (TIBCO, Palo Alto, CA, USA). Different methylation between ulcerative colitis patients, diseased controls and healthy individuals were determined using the Mann-Whitney U-test, and p-values were further corrected according to Benjamini and Hochberg (Benjamini and Hochberg 1995). DMRs with a corrected p-value ≤ 0.05 were considered significantly differentially methylated.

2.2.3 Genome-wide Quantification of individual Methylation Variable Positions (Layer III)

DNA Methylation analysis was performed using the Illumina Infinium Human Methylation27 BeadChip (iScan system, Illumina). while the dataset was subject to intra-array and inter-array normalization as published earlier (Teschendorff et al., 2010). Differences between ulcerative colitis patients, diseased controls and healthy individuals were determined using the Mann-Whitney U-test, while p-values were corrected according to Benjamini and Hochberg (Benjamini and Hochberg 1995). MVPs with a corrected p-value ≤ 0.05 were considered significantly differentially methylated.

2.2.3.1 Bisulfite conversion

Genomic DNA samples were treated with bisulfite (so called bisulfite conversion) using the EZ DNA methylation Kit (Zymo Research, D5002) according to the manufacturer's protocol with modifications for the Illumina Infinium Methylation Assay. Briefly, 1mg genomic DNA was mixed and incubated with 5 μ L M-Dilution at 37°C for 15 minutes, then mixed with 100 μ L of CT conversion reagent prepared as mentioned in the protocol and incubated in a thermocycler for 16 cycles at 95°C for 30 seconds and 50°C for 60 minutes. Bisulfite-converted DNA samples were then loaded onto 96-column plates which has provided in the kit for desulphonation and purification according to the protocol. Concentration of eluted DNA was determined with a Nanodrop-1000.

2.2.3.2 Illumina Infinium Human Methylation27 BeadChip

Bisulfite-converted DNA samples were loaded on Illumina's Infinium Human Methylation27 Beadchip Kit. 27,578 CpG loci were designed on the beadchip which located within the proximal promoter regions of transcription start sites covering more than 14,000 human reference sequences (RefSeq) genes at single-nucleotide resolution. Chip processing and data generation were performed according to the manufacturer's manual. Briefly, 4 μ L of bisulfite-converted genomic DNA samples were denatured in 0.014N sodium hydroxide, neutralized and amplified with kit-provided reagents and buffer for 20–24 hours at 37°C. Genomic DNA samples were fragmented. 12 μ L of each sample was loaded onto a 12-sample chip, then placed into a hybridization chamber according to the manual. Chips were incubated at 48°C for 16–20 hours, then washed with wash buffers provided in the kit and assembled into a fluid flow-through station for primer-extension reaction and staining according to the protocol provided in the kit. A single base extension was then performed using labeled DNP- and biotin labeled dNTPs. Polymer-coated chips were image-processed in Illumina's iScan scanner. Data were generated using BeadStudio v3.0 software. Here, each interrogated locus is represented by specific oligomers linked to two bead types: one representing the sequence for methylated DNA (M) and the other for unmethylated DNA (U). The methylation status of a specific CpG site is then calculated from the intensity values of the M and U alleles, as the ratio of fluorescent signals (β -value). This value is based on following definition and calculation:

$$\beta\text{-value} = \frac{\text{Max}(M,0)}{[\text{Max}(M,0) + \text{Max}(U,0) + 100]}.$$

Human HCT116 methylated DNA (Cat# D5014-2, Zymo Research) was used as a positive control for methylation analysis and Human HCT116 DKO DNA (DNA methyltransferase double knock-out cells (DNMT1 and DNMT3b) was used as a negative control for methylation analysis, prepared by Core Facility). Replicate samples (N = 3) were included to assess inter-array reproducibility.

2.2.3.3 Bioinformatic analyses of differential methylation

β -values were normalized applying quantile normalization, the dataset quality was examined based on the detection p-values. Datapoints containing more than two defect (detection p-value >0.05) measurements were eliminated. Differences between ulcerative colitis patients, diseased controls and healthy individuals were determined using the Mann-Whitney U-test, while p-values were corrected according to Benjamini and Hochberg (Benjamini and Hochberg 1995). MVPs with a corrected p-value ≤ 0.05 were considered significantly differentially methylated.

2.3 Merging of the transcriptional and methylational data (from Layer I,II & III)

To identify disease associated transcripts under potential epigenetic control, differentially expressed transcripts from layer I were selected (corrected $p \leq 0.05$, FDR $\leq 5\%$, regulated between healthy and diseased individuals). The genomic transcript locations were used to generate interaction windows of 50 kb up and downstream of the transcription start site (TSS). These windows were examined whether they contain either a DMR (layer II, corrected $p \leq 0.05$, between healthy and diseased individuals) or a MVP (layer III, corrected $p \leq 0.05$, between healthy and diseased individuals). Transcripts, significantly regulated between healthy and diseased individuals with a significantly regulated DMR or MVP within the 50kb window of the TSS were considered candidates for disease associated transcripts under potential epigenetic control.

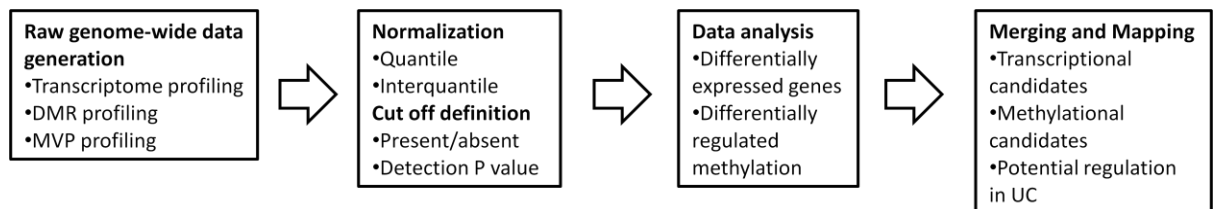


Figure 5. Flow diagram of genome-wide analysis.

2.4 Validation of the transcriptional and methylational candidates

The validation steps were performed using current gold-standard methods, such as quantitative real time PCR for the expression candidates and the bisulfite pyrosequencing for the methylation ones. The individuals in the validation group were selected randomly from the group of unrelated individuals, and all candidates genes with potential epigenetic influence on disease associated mRNA-expression in cis were subjected to validation.

Panel type	Application	Relationship of Individuals	Gender	Age: median, range	Disease representation
Validation panel I	validation of the findings in the transcriptome	unrelated individuals	69 f 66 m	41, 18-76	30 UC _i ; 30 UC _{ni} 15 DC _i ; 30 DC _{ni} ; 30 NC
Validation panel II (⊆ panel I)	validation of the findings in the methylome	unrelated individuals	25 f 25 m	41, 18-68	20 UC 10 DC 20 NC

Table 4. Study panels III, used for validation of initial findings. f = female, m = male, UC = ulcerative colitis, NC = normal controls, DC = disease specificity controls, suffix: i= inflamed, ni= not inflamed.

2.4.1 Validation of differential mRNA expression via Real-Time PCR

Real time PCR (TaqMan) was performed according to the manufacturers guidelines (Applied Biosystems, Foster City, California, USA) using a 7900HT Real-Time PCR System (Applied Biosystems, Foster City, California, USA). Expression levels were calculated relative to beta-actin using the standard-curve method (Livak and Schmittgen 2001). Differences between ulcerative colitis patients, diseased controls and healthy individuals were determined using the Mann-Whitney U-test, while p-values were corrected according to Benjamini and Hochberg (Benjamini and Hochberg 1995).

2.4.1.1 Prepare the cDNA sample

Before running the TaqMan Gene Expression Assays, total RNA from 135 unrelated individuals was isolated as a template for the synthesis of single-stranded cDNA, as described in section 2.1.3. The reverse transcription was performed using High Capacity RNA-to-cDNA Kit (PN 4387406). The quantity of the cDNA was measured using the UV absorbency (A260/A280) method.

2.4.1.2 Prepare the reaction mix and load the plate

TaqMan Gene Expression Assays (20X) and cDNA samples were thawed on ice, completely resuspended by gently vortexing, then briefly centrifuged to bring the liquid to the bottom of the tube. The master mix reagent was mixed by gently swirling the bottle. Each assay run was performed in technical duplicates for each sample, supplemented by endogenous control assays (ACTB) and non template controls (NTCs, containing H₂O).

Symbol	Gene name	TaqMan ID
FLNA	filamin A, alpha (actin binding protein 280)	Hs00155065_m1
MT1H	metallothionein 1H	Hs00823168_g1
SPINK4	serine protease inhibitor, Kazal type 4	Hs00205508_m1
THY1	Thy-1 cell surface antigen	Hs00174816_m1
CFI	complement factor I	Hs00173409_m1
IGH@	immunoglobulin heavy constant gamma 1 (G1m marker);immunoglobulin heavy constant gamma 4 (G4m marker);immunoglobulin heavy locus;immunoglobulin heavy constant gamma 3 (G3m marker);immunoglobulin heavy variable 4-31;immunoglobulin heavy constant mu	Hs00378230_g1
HKDC1	hexokinase domain containing 1	Hs00228405_m1
VIPR1	vasoactive intestinal peptide receptor 1	Hs00270351_m1
PTN	pleiotrophin	Hs01085691_m1
IRS2	insulin receptor substrate 2	Hs00275843_s1
TK1	thymidine kinase 1, soluble	Hs00177406_m1
SLC7A7	solute carrier family 7 (cationic amino acid transporter, y+ system), member 7	Hs00374418_m1
ACTB	actin, beta	Hs99999903_m1

Table 5. TaqMan assays for validation of findings from genome-wide data.

PCR reaction mix component	Volume per 20- μ L reaction (μ L)
20X TaqMan [®] Gene Expression Assay	1
2X TaqMan [®] Gene Expression Master Mix	10
cDNA template (1 to 100 ng)	4
RNase-free water	5

Table 6. Component of the PCR reaction mix.

The mixture was loaded on the 384-well reaction plate, and then the plate was sealed with the appropriate cover and centrifuged briefly. The plate was finally loaded into the 7900HT Fast system (Applied Biosystems) 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.

2.4.1.3 Bioinformatic analyses

Upon completion of the run, the data was generated by the Sequence Detection System (SDS) 2.0 software. Differences between ulcerative colitis patients, diseased controls and healthy individuals were determined using the Mann-Whitney U-test, while p-values were corrected according to Benjamini and Hochberg (Benjamini and Hochberg 1995).

2.4.2 Validation of differential DNAm via Pyrosequencing

Validation of initial findings was performed via bisulfite conversion followed by pyrosequencing (Roche, 454) as previously described (Bollati et al., 2007). Differences between ulcerative colitis patients, diseased controls and healthy individuals were determined using the Mann-Whitney U-test, while p-values were corrected according to Benjamini and Hochberg (Benjamini and Hochberg 1995).

Gene Symbol	Forward primer	Reverse primer	Seqence primer	Target sequence
THY1	agtttaggggtga agaa	cctatccctcccta at	taggggtgaagaa g	AYGTAATGTTTAGAGTAAAAAGY GTTTGTAGTTTTTTGAAGGGTTGG GTGTTTTGGAATAGATGAGGGGG YGAAATGGGGTTGGGGATTAGG GAYGGA
IGH@	gtttgataaagtaa gggtatt	cttctaacaact caatctatat	agataaaggatggt ttg	YGTGGGGAAATGTGGTGATA
SPINK4	ggatttgggttatttt tgt	ctaccaaattttt ttccttaact	gattagtatttaa tagaaa	AAGTAYGGTTTTGATGGAGTATA AATTTAATGGGGTAGAGGAAAAG GGGAAAGGTAATAAATAAGATG GTTATAATGTATGTTAAAYGGT
SLC7A7	gggttttatagtgtt agttaggat	cacctacctttca ttcttaac	taaagtgtgggatt at	AGGYGTGAGTTATYGYGTTTGGTT TGATTTTTTATATTTTTTAGGAAA GTGGAGTTAATTTATATTTGAGG AAGGYG
CFI	ggttgtatttattgt aattgtag	tacttattttcc cctttctat	tttttgtgtgaagg t	ATAGTGTTAYGAAGATAGTGAGA TTGATTTTTTTTTTAATAGTTTTT TAGTAYG
HKDC1	aggggttatagag ttatgtta	accaccaaattctc attacta	tggaaattttaga gtag	TTTTATYGGTTGAAAGTTGGTTTT GTTTAGTGTAAATTTYGTAGAAG ATTTTTTYGGTTGAGTGGGTTTG GTAGAGTTYGGT
HKDC1	agagggaggatt aaag	acattttcccaa att	aggttgaggatttt a	GTTYGTAATATTTYGTTTTTAGG AGGTTTGTAGTTTGGATTGGAA GYGTGTAATATTTTAGAGTYG

Gene Symbol	Forward primer	Reverse primer	Sequence primer	Target sequence
FLNA	ggggatagggatta gaa	cccctctaccaac ctataa	gtgtttttagtaggtg t	TTYGTGAGGTTATTATTGAGTTT AGTGTGGAYGTTYGGGTTTTGAT ATAGATYGGAGGGTYGTAYGTTA AGGTTYG
FLNA	ggggatagggatta gaa	cccctctaccaac ctataa	tggttaatttttagg taa	TTTGAYGGAGATTTAYGTTTAGGA TYGTGGYGATGGTATGTATAAAG TGGAGTATAYGTTTTAYGAGGAG GGTGYG
PTN	ttgggtgaatttgta taa	aaaaccctctcta aacac	tgtattgaatgaaat gat	AATTTTAGAGAYGTAAAGTTTAGT TTTTAAGTTTTATAAATTTTAGAG TTTGAGTTATTTTGAYG
PTN	ttgggtgaatttgta taa	aaaaccctctcta aacac	ggtaagtagtattta attgg	GAATTTTGAGAYG
TK1	gtgtttaagtattgg ggattt	cccaaacaaaat tcctta	agtattggggatttg	GTAYGTATTAGGYGTTTTGTATGT TTATAGGAGTGTTTTAGAYGGTTT TTAAGTTTTTTTTTGGAAATTTTAA TTTAATAAATGAGTTAATTTYG
MT1H	tgtggggagtaggt ta	acaactccttttct aatca	tggggagtaggtta g	GATTTYGGGGTAYGYGGAAYGTT AAAGGGYGGGAGTAGTAGGTAA TTTTAGGGAATTGGGTAGGTTYG G
MT1H	tgtggggagtaggt ta	acaactccttttct aatca	gtagggatgaaata ag	TAGTTTATTYGGGTGTATAGTTTT TTYGYGGGAGTYGTTGYGAAAG GGTYGTTTTYGYGGTGTGTATYGA TTTGYG
MT1H	tgtggggagtaggt ta	acaactccttttct aatca	ttgagtttagggat tttg	YGTTYGATTYGTTTGTGTYGTATA GTTTAGTTTAGGATYGYGGGYGG TTYGGATTTAGTGGGTYG

Table 7. The methylation candidates from both DMR data set and MVP data set, were validated using pyrosequencing.

2.4.2.1 Bisulfite Treatment

1 μ g DNA was denatured in 50 μ L NaOH (2 mol/L) for 20 min at 37°C. Then, the denatured DNA sample was mixed with 30 μ L freshly prepared hydroquinone (10 mmol/L) and 520 μ L sodium bisulfate (3 mol/L) (Sigma-Aldrich, St. Louis, MO) at pH 5.0. The mixture was then incubated at 50°C for 16 hours. The bisulfite-converted DNA was purified using the Wizard DNA Clean-Up System (Promega, Madison, WI). The DNA was eluted in 50 μ L of warm water, and 5.5 μ L NaOH (3 mol/L) were added for 5 min. The DNA was ethanol precipitated with glycogen as a carrier and resuspended in 20 μ L water. Bisulfite-treated DNA was stored at -20°C until further processing.

2.4.2.2 PCR and Pyrosequencing

DNA methylation was quantified using bisulfite pyrosequencing. Briefly, the samples were bisulfite converted and PCR amplified using biotin-labeled primer and the final PCR product was purified

using Sepharose beads. The PCR product was then bound to Streptavidin Sepharose High Performance (Amersham Biosciences, Uppsala, Sweden), and the immobilized PCR product binding the Sepharose beads was purified, washed, denatured using NaOH solution (0.2 mol/L), and washed again using the Pyrosequencing Vacuum Prep Tool (Pyrosequencing, Inc., Westborough, MA) according to the manufacturer's recommendations. Pyrosequencing primer (0.3 μ mol/L) was then annealed to the purified single-stranded PCR product and pyrosequencing was performed using PSQ HS96 Pyrosequencing System (Pyrosequencing). DNA methylation was expressed for each DNA locus as percentage (methylated cytosines / methylated cytosines + unmethylated cytosines). The non-CpG DNA fragments were used as built-in controls to verify bisulfite conversion. Each marker was tested in three replicates and their average was used in the further statistical analysis.

2.4.2.3 Statistical analysis

Differences between ulcerative colitis patients, diseased controls and healthy individuals were determined using the Mann-Whitney U-test, while p-values were corrected according to Benjamini and Hochberg (Benjamini and Hochberg 1995).

2.5 Gene Ontology analysis

Gene Ontology analysis was performed as previously published (Tavazoie et al., 1999). Biological processes associated to the transcripts and candidate genes were retrieved from the Gene Ontology Consortium (www.geneontology.org).

2.6 Chromatin- immunoprecipitation following quantitative real-time PCR

The validated data indicated that the inflammatory response in ulcerative colitis and acute colitis might be mediated on the similar regulatory pattern and through distinct DNA modifications. The following techniques were performed as a pilot project to examine the differences of DNA hydroxylation between ulcerative colitis and acute colitis.

2.6.1 Preparation of samples and control DNA

The genomic DNA samples were fragmented according to the mechanical fragmentation protocol. briefly, 20 μ g genomic DNA was pipetted into a 1,5 ml microcentrifuge and the final volume was adjust to 300 μ l by addition of 10mM Tris-HCl pH 8,5. Then, using a tip probe sonicator, the mixture was sonicated with 15 pulses of 20 seconds, with a 20 second pause on ice between each pulse. The sheared DNA was quantified by ethidium staining after electrophoresis on a 3% agarose gel. Finally, the fragmented DNA was purified using Chromatin IP DNA Purification Kit (Active Motif, Nr. 58002)

Two kinds of control DNA fragments were amplified using PCR for the further experiments to verify the immunoprecipitation efficiency. Two DNA fragments of house keeping genes were amplified with normal primers (GAPDH forward primer: TTAGCGCTGGGGCTGGCATT, reverse primer: CTCCTAGGCCCTCCCCTCTT; ACTB forward primer: CAAGAGATGGCCACGGCTGCT, reverse primer: GGATGCTCGCTCCAACCGACT.), respectively, and in the amplification of GAPDH, special hydroxymethylated cytosine (Zymo Research Corporation) was used and fully hydroxymethylated fragment was generated.

2.6.2 Immunoprecipitation of 5-hydroxymethylcytidine enriched DNA fragments

Immunoprecipitation was performed using a hMeDIP Kit (Active Motif, Nr.55010). The hMeDIP contains a highly specific purified 5-hydroxymethylcytidine antibody and the necessary buffers. Briefly, 200 μ l PCR tube was set up for each immunoprecipitation reaction. 500ng DNA sample and controls DNA were added in the tube and mixed with 10 μ l buffer C, and 1 μ l PIC provided in the Kit, and sterile water to a final volume of 96 μ l respectively. Then 4 μ l 5-hydroxymethylcytidine antibody was added, then incubated overnight with end to end rotation at 4°C. The next day, the PCR tube was centrifuged quickly to collect the contents at the bottom, and the Protein G magnetic beads were resuspended. 25 μ l Protein G beads were added to each tube, and incubated for 2 hours at 4°C with end to end rotation. At this stage, the target DNA was captured on the beads. The tubes were then placed on a magnetic stand to pellet beads to the side of the tube. The supernatant was discarded, and the beads were washed three times with buffer C and D provided in the Kit, according to the protocol. The beads were then resuspended with 50 μ l elution buffer AM2 by pipetting three times, and incubated for 15 minutes at room temperature with end to end rotation to keep the beads in suspension. The tubes were centrifuged to collect liquid from the cap, then, 50 μ l neutralization buffer was added and mixed with pipetting up and down. The tubes were placed on the magnetic stand again, in order to transfer the supernatant, which contains the target DNA, to the fresh tubes.

2.6.3 Examination of 5-hydroxymethylcytidine enrichment via qRT PCR

In order to determine the enrichment of the candidate sequences, the following quantitative real time PCR was performed, which the TaqMan assays were designed using primer express 3.0 (Applied Biosystem). Similarly, this quantitative real time PCR was performed according to the manufacturer's guidelines (Applied Biosystems) using a 7900HT Real-Time PCR System, but without cDNA synthesis. Enrichment levels were calculated relative to control DNA samples using the standard-curve method (Livak and Schmittgen 2001). Differences between inflamed mucosa of ulcerative colitis patients and

not inflamed ones, and the inflamed mucosa of diseased controls were determined using the Mann-Whitney U-test, while p-values were corrected according to Benjamini and Hochberg (Benjamini and Hochberg 1995).

Panel type	Application	Relationship of Individuals	Gender	Age: median, range	Disease representation
Examination panel III	Examination of the prediction in the hydroxymethylome	unrelated individuals	5 f 5 m	52, 25-64	5 UCi; 5 Ucn 5 DC

Table 8. Study panels IV, used for examination of initial findings. f = female, m = male, UC = ulcerative colitis, DC = disease specificity controls, suffix: i= inflamed, ni= not inflamed.

Design_ID	Forward Primer	Reverse Primer	Probe
HKDC1_Promoter	GGTGGCACCCAATTCTCAA G	CTGGATGGTAGCTGAGGT TTAGTG	AACCCATCAATACCCTGCT TCCCC
MT1H_Promoter	ACTAGAAAAGGAGCTGCC AACTGT	TTCCCAAGCGAGAAGAGA AGAG	CTCCACCACGCCCTCCAC GTGT
PTN_Promoter	CTTAGCGTCTTTCCTGTATT GAATGA	ATTCGTCAAAGTAGCTCA AACTCTGA	AGACGCAAAGCCCAGCCT CCAAGTT
SPINK4_Promoter	GGGCCATCCCTGCAATCT	GCTTTTTCTGTTTGGGTCA CTAATC	TACCATGTCAGGCACTGT TCTGAGCCC
THY1_Promoter	AAGACGTAATGCCAGAG CAA	CCATTTGCCCCCTCATC	CCTGCAGCCCCCTGAAGG GC
TK1_Promoter	CCACCTACCTCTGCAGACA TCTT	GTAGACGCCATGATGGAT GTGT	AAGGAACCTTGCTTGGGA AACCCACA

Table 9. Designed TaqMan assays.

3 Results

3.1 Genome-wide profiling of transcription and DNA methylation in ulcerative colitis

A 3-layer genome-wide scanning of the transcriptome, differentially methylated regions and differentially methylated variable positions in twins, discordant for ulcerative colitis.

3.1.1 Genome-wide transcriptome profiling (Layer I)

The genome-wide transcriptome analysis (layer #1) of the mucosa of monozygotic twins, discordant for ulcerative colitis, identified 18097 out of 54675 transcripts analyzed as present. Within the present transcripts, 361 are differentially expressed in UC significantly (Benjamini-Hochberg-corrected p -value ≤ 0.05 ; FDR (false discovery rate) $\leq 5\%$), 247 transcripts are downregulated and 114 are upregulated respectively. Most significant candidates associate with inflammation, such as immune response, macrophage activation, complement activation, cell adhesion, cytokines signaling and metabolism (Figure 11). All the significant candidates are listed in table S 1.

3.1.2 Genome-wide methylome profiling (Layer II and layer III)

It is still unknown, whether the single CpG or the regional DNA methylation has the potential regulatory function on transcription. The analysis of differentially methylated regions (DMRs, layer II) examined 392750 regions, of which 389359 were detectable as present in the mucosal tissue samples. 345 DMRs showed differential methylation between healthy and diseased individuals significantly (Benjamini-Hochberg-corrected p -value ≤ 0.05). Within the significant candidates, 240 candidates are hypomethylated and 105 are hypermethylated in UC respectively. Most of them are detected in the promoter, gene body, strongly indicating the potentially regulatory function for the associated transcription in cis. The significant candidates from layer II are listed in table S 2.

3.1.3 Genome-wide quantification of individual methylated variable positions (Layer III)

As mentioned above, the regional methylation and single CpGs were interrogated separately. The analysis of individual methylated variable positions (MVPs, layer III) examined 27578 positions were analyzed, of which 23085 were detected as present in the mucosal tissue samples. 703 MVPs showed differential methylation between healthy and diseased individuals significantly (Benjamini-Hochberg-corrected p -value ≤ 0.05). Surprisingly we found that about 95% (666/703) of the

significant candidates reflect hypomethylation in UC. The significant candidates from layer III are listed in table S 3.

The genome-wide 3-layer screen identified 61 disease-associated transcripts that were also associated with at least one disease-associated MVP or DMR when applying a cis-interaction window of 50 kb from the transcription start site in silico. By definition, those 61 transcript-methylation pairs were called risk loci. Disease-associated transcripts were defined by their significantly differential expression when comparing healthy individuals to UC patients (Benjamini-Hochberg-corrected p-value ≤ 0.05 ; false discovery rate $\leq 5\%$), and disease-associated MVPs and DMRs were defined by their significantly differential methylation when comparing healthy individuals to UC patients (Benjamini-Hochberg-corrected p-value ≤ 0.05 ; false discovery rate $\leq 5\%$). In addition, the 61 candidates do not represent 61 gene-expressions and -methylations (which take place on the gene-body or the regulatory region). While, 1) the 50 kb window is really long, some methylation candidates locate far away from the transcription stop-codon; 2) Some candidates represent the same gene. From the 61 disease- and MVP/DMR-associated transcripts, 12 candidates with the closest proximity between MVP/DMRs and transcript location were subjected to further validation and replication in a larger collection of primary tissue from the intestinal mucosa. Two of those 12 represent controls: altered mRNA expression with no altered methylation (IGHG1) and altered methylation with no altered mRNA expression (SLC7A7).

3.1.4 Merging of the transcriptional and methylational data (from Layer I,II & III)

The combination of the three layers illustrates, that epigenetic differences as well as their potential transcriptional consequences can be monitored in complex primary tissue on a genome-wide scale (genome-wide overview: figure 6).

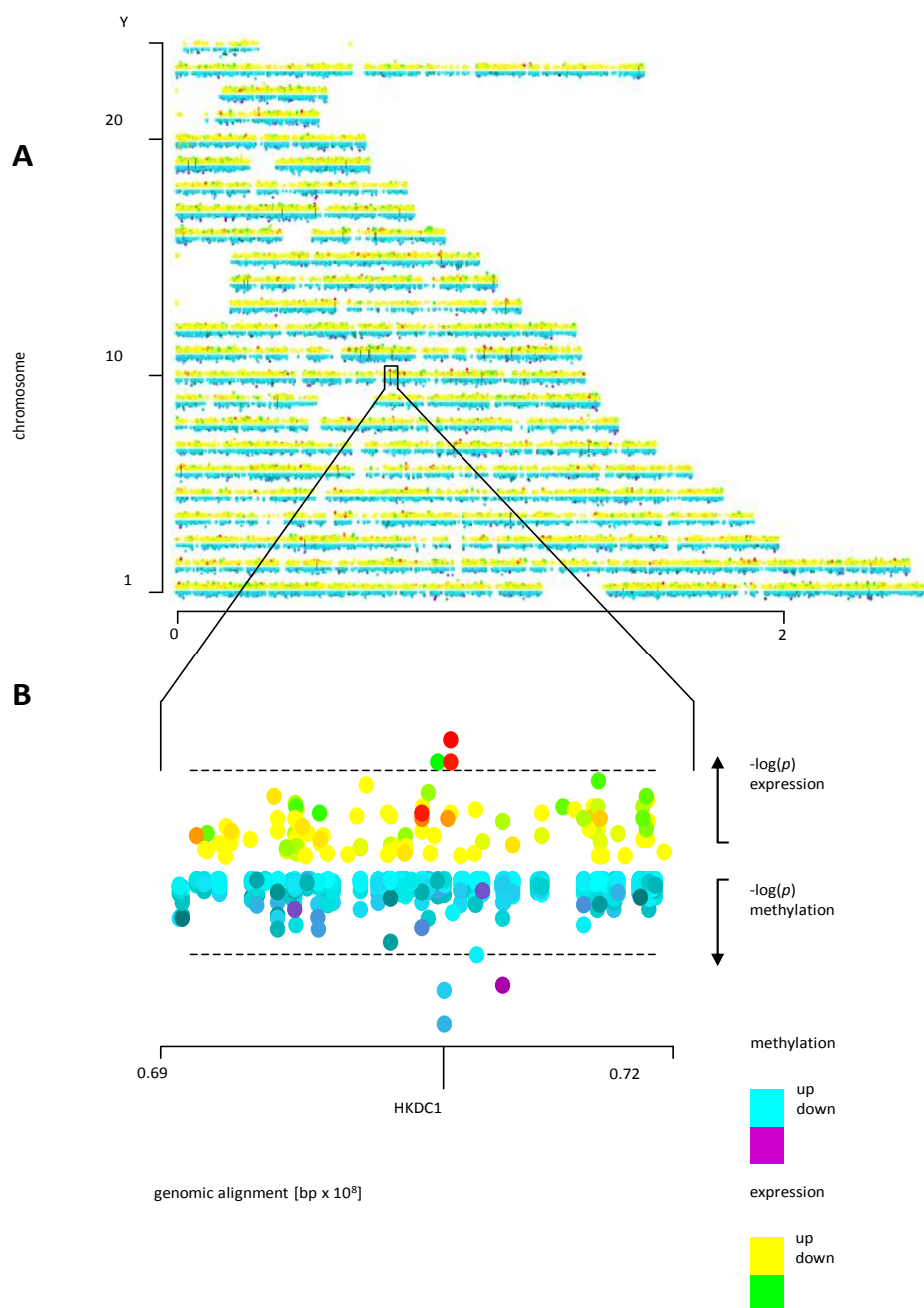


Figure 6. A: Genome-wide profiles of DNAm and gene expression. The x-axis represents the genomic location with chromosomes represented as rows. The significance of the differences between UC patients and healthy individuals is displayed as $-\log(p)$ for the mRNA and as $\log(p)$ for the epigenetic modifications. Effect size (induction/fold change) is encoded by colors (methylation: upregulated = black, not regulated = blue, downregulated = purple; mRNA-expression: upregulated = red, no regulation = yellow, downregulated = green); B: selected candidate HKDC1 (the dotted line indicates the significance threshold).

3.1.5 Intra-class correlations of the three-layer dataset

Similar degrees of heritability have been observed on all 3 levels, as documented by the intra-class correlations (figure 7)

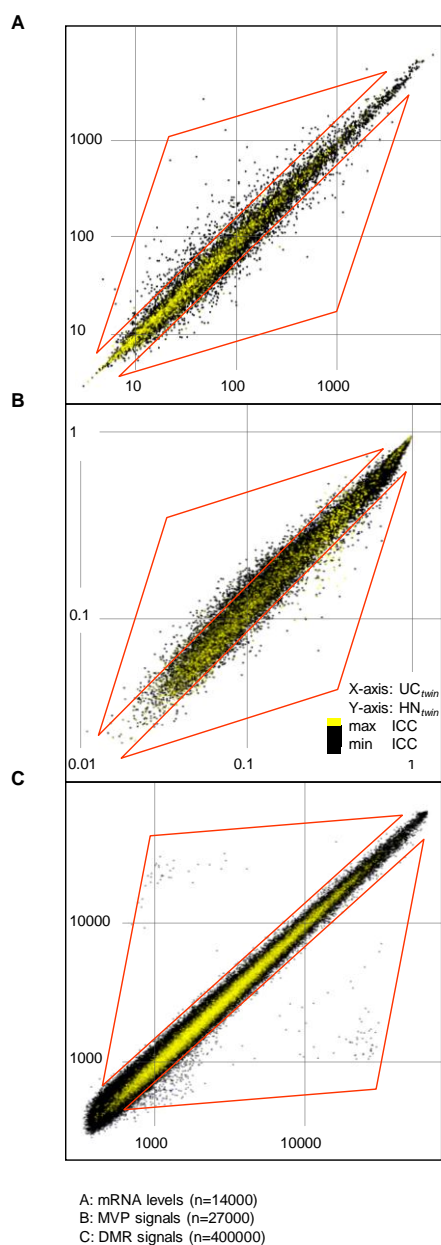
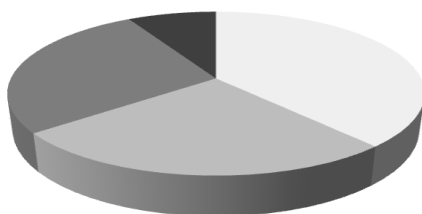


Figure 7. Intra Class Correlation for all three levels of the genome-wide functional epigenetic map. A: mRNA levels (n=1400 transcripts); B: MVPs (n=27.000 signals); C: DMRs (n= 400.000 signals). Red areas indicate where differentially regulated items are found. Intra class correlation coefficient was calculated based on the variance within twins compared to the variance between unrelated individuals.

3.1.6 Overview: genomic location of the different methylation in ulcerative colitis

Hypomethylation is a prominent finding in promoter regions (figure 8A), while hypermethylation is frequent in gene-introns (figure 8B)

A: methylation downregulated



B: methylation upregulated



- Promoter
- Exon
- Intron
- Other

Figure 8. Distribution of significant MVPs/DMRs in different genomic regions (promoters, exons, introns and others). A: Distribution of significantly hypomethylated CpGs when comparing ulcerative colitis patients to healthy controls, B: Distribution of significantly hypermethylated CpGs when comparing ulcerative colitis patients to healthy controls.

3.1.7 Principle component analysis of the significant candidates

The principle component analysis displayed that the disease is the strongest component (PC1), yet mRNA and DNAm do not contribute equally to the separation between healthy and diseased individuals (plots for significant expression changes and significant epigenetic modifications: supplemental figure 10).

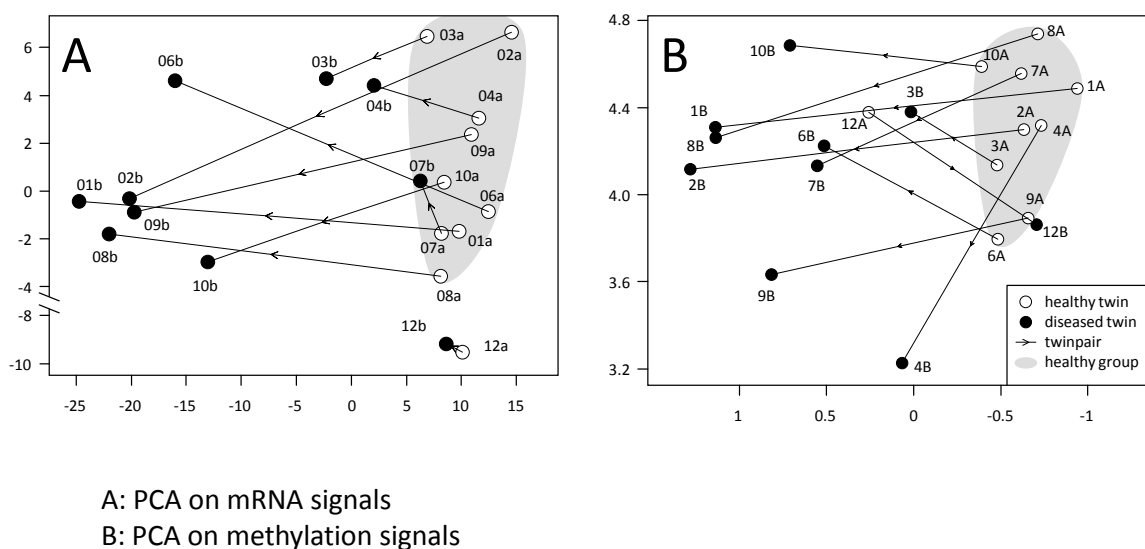


Figure 9. Principal component analysis of differential transcription and DNAm in UC. A: the two strongest components separating the samples based on all significantly differentially expressed mRNA transcripts ($n=361$) are displayed. B: the two strongest components separating the samples based on all significantly differentially methylated CpGs ($n=731$) are displayed. Arrows indicate the connection from the healthy twin to its diseased sibling. Grey areas represent the healthy group, while one outlier (pair 12) was not located within the healthy group in either of the two approaches.

3.1.8 Correlation between the expressional and methylational data (in cis)

Correlation analysis of quantitative expression values and MVP or DMR data revealed a correlation in selected candidate transcripts (median spearman-rho $r=0.58$), correlation in identified candidate loci ($r=0.49$), no correlation in the group of disease-associated transcripts ($r=0.15$) and no correlation when analyzing all transcripts ($r=0.16$). The correlation within the selected candidates was significantly different from all other gene-sets.

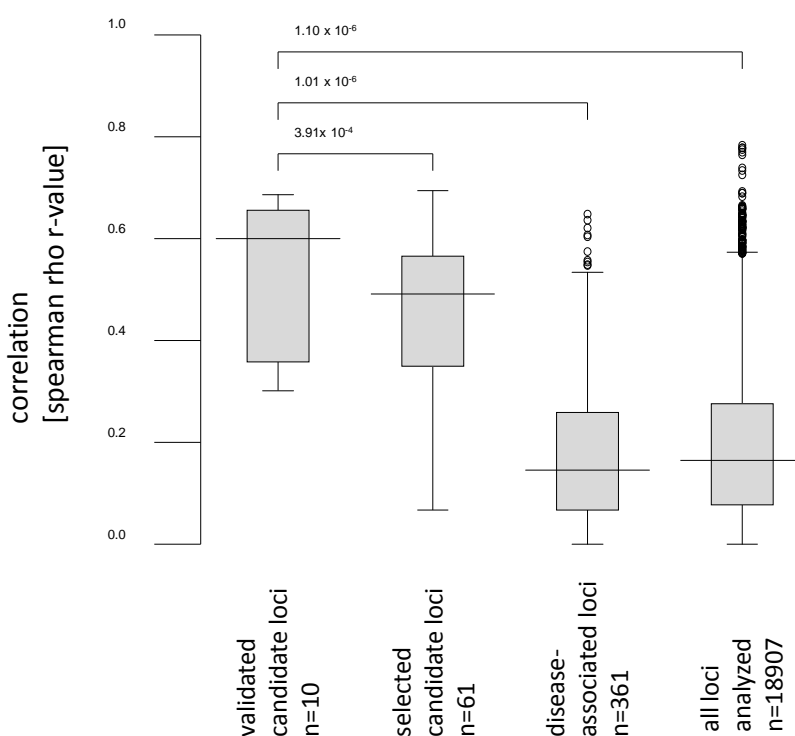


Figure 10. Correlation analysis of mRNA expression and epigenetic modification. Each box-plot represents an individual loci/transcript correlation set based on a Spearman-Rho correlation, displaying the 25th, 50th and 75th percentile. Whiskers correspond to the 5th and 95th percentile while outliers are plotted separately. P-values describe the differences between the individual sets of r-values, calculated by a Mann-Whitney U-test. For better readability r-values were transformed to absolute values.

3.1.9 Gene ontology (GO) analysis of the disease relevant processes

A large number of differentially regulated transcripts (independent of MVPs and DMRs) were associated with inflammation, immune response and other disease relevant processes when performing a Gene Ontology (GO) analysis. Similarly, regulated transcripts in close proximity to MVPs and/or DMRs were also found to be associated to inflammation and immune response (figure 11).

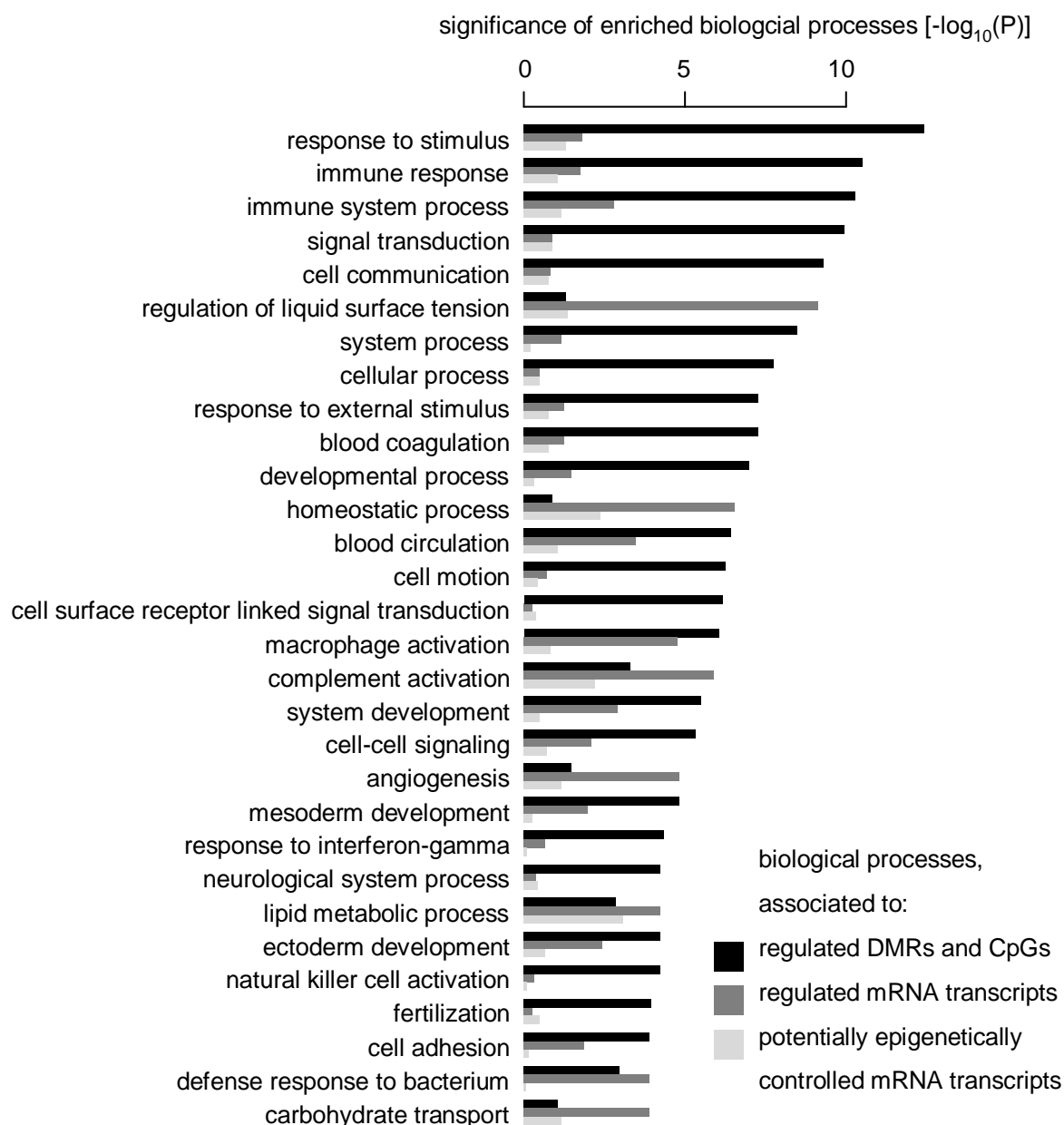


Figure 11. Biological processes affected by differential DNAm or gene expression in UC patients. The 30 most significantly enriched biological processes are displayed for three different categories: i) all MVPs and DMRs, ii) all mRNA transcripts and iii) all mRNA transcripts associated with at least one MVP and/or DMR. Significance is displayed as $-\log_{10}(p)$.

3.2 Validation of selected candidate transcripts under potential epigenetic control

Of the 61 disease- and MVP/DMR-associated transcripts, 10 candidates with the closest proximity between MVP/DMRs and transcript location were subjected to further validation and replication in a larger collection of primary tissue from the intestinal mucosa (validation panel I and II, table 3). This

was performed by pyrosequencing (bisulfite modified DNA, n=50 individuals) for methylation data and TaqMan based real-time PCR (n=135 individuals) for differential mRNA expression. Significance criteria were defined in concordance to the initial screening step (Benjamini-Hochberg-corrected p-value ≤ 0.05). Distinct patterns of differential DNAm/expression pairs as paradigms for different scenarios of epigenetic modulation were selected while focusing on the commonly accepted pattern of hypermethylation with transcriptional repression and hypomethylation with increased mRNA expression: i) Pairs with negative correlation of DNAm and mRNA expression, represented by 40 MVPs and 5 transcripts: increased DNAm accompanied by decreased mRNA expression (MT1H) and, ii) decreased DNAm and with increased mRNA expression (CFI, HKDC1, SPINK4, THY1), iii) pairs with positive correlation of DNAm and mRNA expression, represented by 21 MVPs and 3 transcripts: increased DNAm accompanied by increased mRNA expression (TK1), and iv) decreased DNAm and with decreased mRNA expression (FLNA, PTN). Control transcripts loci were chosen to replicate MVPs/DMRs without a cis-modulation of mRNA expression (SLC7A7, 4 MVPs) or differential mRNA expression without the presence of any MVP or DMR (IGHG1, 1 MVP). The MVPs and DMRs which potentially associated with alterations in expression levels occurred both at the start of the CpG island (HKDC1, TK1) as well as at the end of a CpG island (all other candidates), generally referred to as CpG island shores (Doi et al., 2009; Irizarry et al., 2009). Results for one selected example, representing a validated disease associated transcript, are shown in figure 12 (HKDC1).

Gene symbol; name	prominent gene ontology terms (selected)	Cyto- Genetic band	CpG- position relative to transcript	CpG methylation p-value (observations per gene)	mRNA regulation fold- change; p-value
<i>CFI</i> ; complement factor I	complement activation; innate immune response	4q25	post TSS, SGB	$\downarrow 1.51 \cdot 10^{-05}$ (2)	$\uparrow 5.71$; $4.45 \cdot 10^{-09}$
<i>FLNA</i> ; Filamin A, alpha	positive regulation of I- kappaB kinase/NF-kappaB cascade	Xq28	post TSS, SGB	$\downarrow 1.63 \cdot 10^{-02}$ (2)	$\downarrow -2.85$; $7.72 \cdot 10^{-06}$
<i>HKDC1</i> ; hexokinase domain containing 1	Glycolysis	10q22.1	pre and post TSS	$\downarrow 7.94 \cdot 10^{-06}$ (6)	$\uparrow 2.01$; $5.16 \cdot 10^{-06}$
<i>IGHG1</i> ¹⁾ ; immunoglobulin heavy constant gamma 1 (G1m marker)	immune response; antigen binding	14q32.33	post TSS, SGB	(-) $1.56 \cdot 10^{-01}$ (0)	$\uparrow 8.25$; $8.90 \cdot 10^{-09}$
<i>MT1H</i> ; metallothionein 1H	metal ion binding; protein binding	16q13	post TSS	$\uparrow 1.50 \cdot 10^{-02}$ (2)	$\downarrow -1.66$; $3.04 \cdot 10^{-05}$
<i>PTN</i> ; Pleiotrophin	positive regulation of cell proliferation	7q33	post TSS	$\downarrow 2.42 \cdot 10^{-02}$ (1)	$\downarrow -4.16$; $2.76 \cdot 10^{-08}$
<i>SLC7A7</i> ²⁾ ;solute	blood coagulation,	14q11.2	postTSS	$\downarrow 5.21 \cdot 10^{-06}$ (4)	(-) 1.16;

Gene symbol; name	prominent gene ontology terms (selected)	Cyto- Genetic band	CpG- position relative to transcript	CpG methylation p-value (observations per gene)	mRNA regulation fold- change; p-value
carrier family 7 (cationic amino acid transporter, y+ system), member 7	leukocyte migration				6.73×10^{-01}
<i>SPINK4</i> ; serine peptidase inhibitor, Kazal type 4	Serine-type endopeptidase inhibitor activity	9p13.3	post TSS, SGB	$\downarrow 3.03 \times 10^{-06}$ (2)	$\uparrow 13.05$; 4.45×10^{-09}
<i>THY1</i> ; Thy-1 cell surface antigen	positive regulation of t cell activation	11q23.3	post TSS	$\downarrow 2.28 \times 10^{-03}$ (2)	$\uparrow 2.99$; 4.75×10^{-07}
<i>TK1</i> ; thymidine kinase 1, soluble	DNA replication	17q23.2- q25.3	pre TSS	$\uparrow 3.08 \times 10^{-03}$ (2)	$\uparrow 1.32$; 2.30×10^{-02}

Table 10. Validation of differential DNAm linked to disease-associated transcripts by TaqMan real-time PCR (mRNA expression in n=135 individuals, validation panel I) and pyrosequencing (CpG methylation in n=50 individuals, validation panel II). 1) Negative control candidate for differential expression without epigenetic modification; 2) Negative control candidate for epigenetic modification without differential mRNA expression. TSS = Transcription start site; SGB = spanning to gene body; in case of multiple CpGs, all relative positions are described. Upregulation is indicated by arrow up (\uparrow); downregulation is indicated by arrow down (\downarrow); no regulation is indicated by a dash (-) as measured in the validation panel. P-values presented were corrected for multiple testing using the Benjamini-Hochberg correction. In case of multiple observations, the lowest p-values were presented.

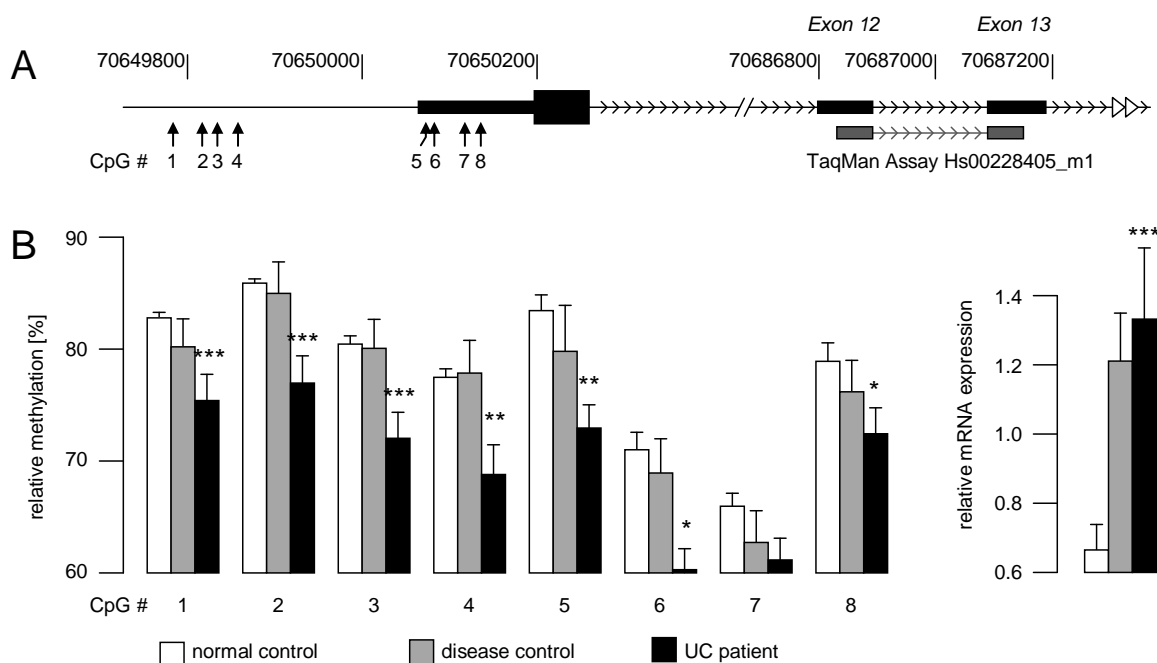


Figure 12. Validation of candidate transcripts, example: HKDC1. A: Relative position of CpGs (arrows up, continuously numbered 1-8) and real-time PCR probe (dark grey); B: quantitative results of the validation: i) methylation (via pyrosequencing, CpGs continuously numbered 1-8 in concordance to

Fig13A) in n=50 individuals (validation panel II), the order of assays displayed corresponds to the order in Fig 13A, ii) mRNA (via real-time PCR, right) in n= 135 individuals (validation panel I). Significant differences between UC patients and normal controls are indicated by a star (* $p \leq 5 \cdot 10^{-3}$; ** $p \leq 5 \cdot 10^{-4}$; *** $p \leq 5 \cdot 10^{-6}$).

The regulatory patterns observed in the validation panel displayed a high robustness as documented by the frequencies of their occurrences: A random individual from the validation panel has a median chance of over 80% to exhibit the regulatory pattern described for the 10 candidate loci (Figure 13).

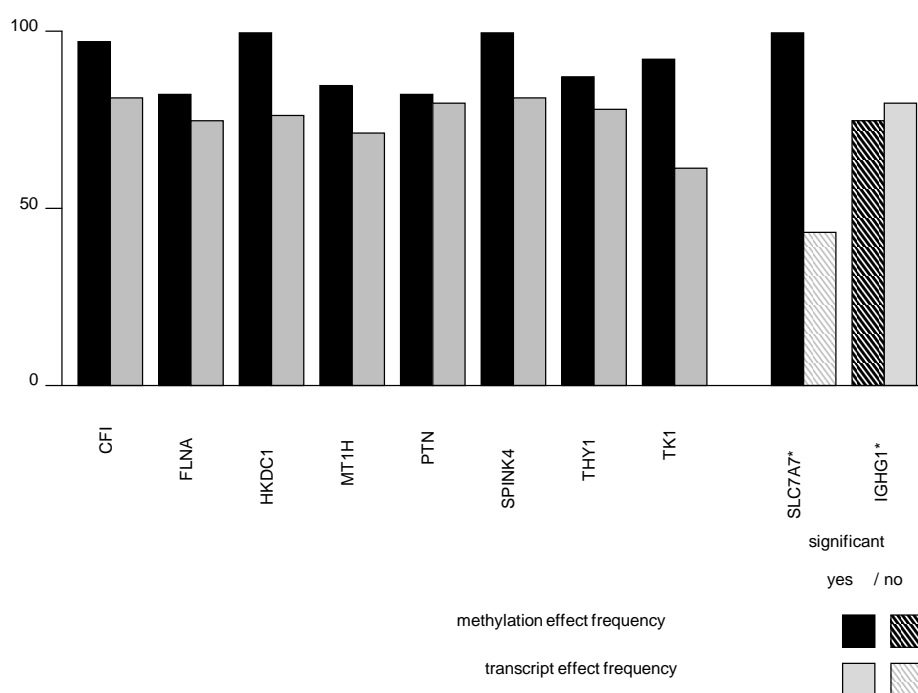


Figure 13. Frequencies of validated effects. Frequencies are presented as percentages at which a significant effect (either epigenetic modification or mRNA regulation at $p \leq 0.05$) is observed in concordance with the initial finding (screening cohort). IGHG1 and TH1 represent control transcripts (marked with a *), where transcript changes were not cis-linked to epigenetic modifications. The 75th percentile is highlighted to emphasize the high frequencies.

A substratification analysis performed on validation panel I and II revealed no influence of the potential confounding factors age, gender, biopsy location and medication on the main findings (mRNA: figure 14 and DNAm: figure 15).

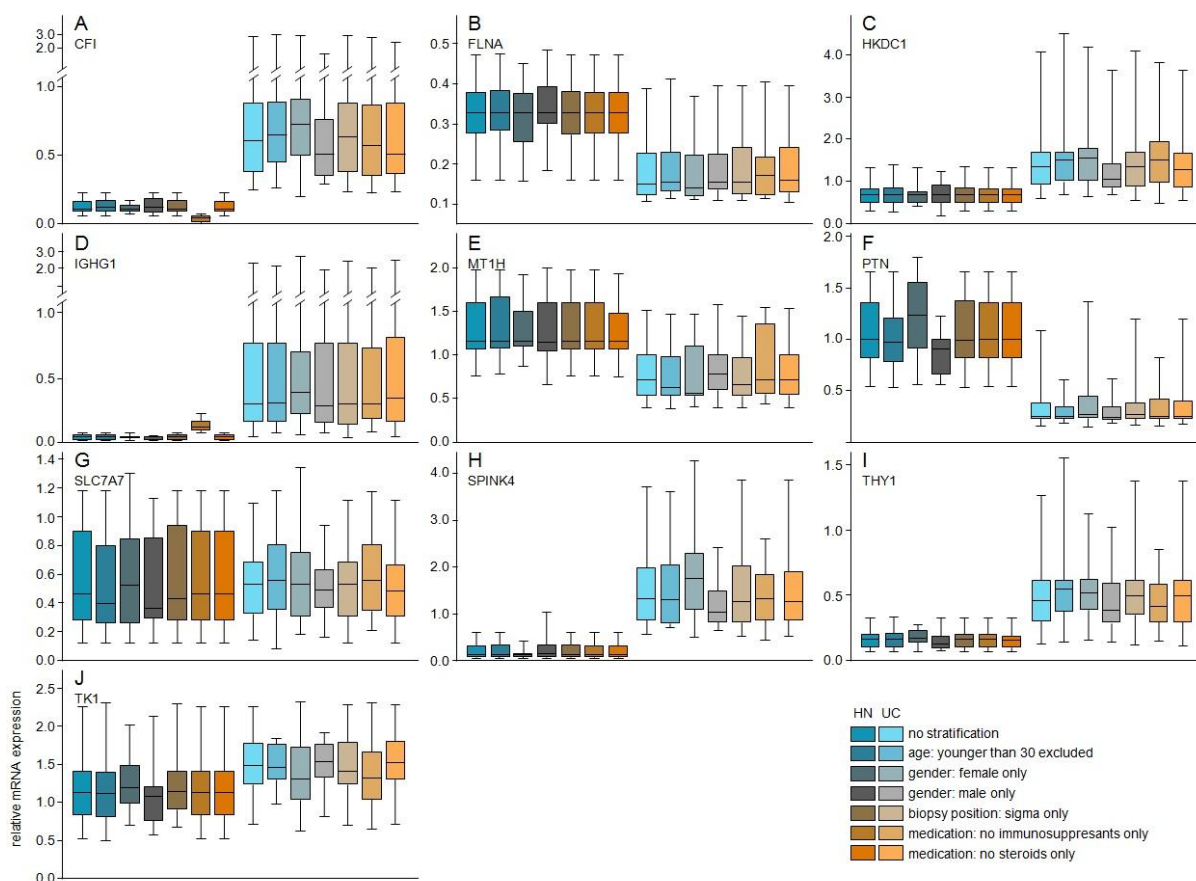


Figure 14. Impact of potential confounding factors on the differential mRNA expression. Each originally selected candidate transcript is represented in a separate panel (A-J). Each panel shows 7 color coded sets of analysis from validation panel I, split into healthy individuals (HN) and ulcerative colitis patients (UC): no substratification (all samples included); substratified by age (all individuals younger than 30 years were excluded); substratified by gender (females only); substratified by gender (males only); substratified by biopsy position (all non-sigmoidal biopsies excluded); substratified by medication (all individuals with immunosuppressant therapy removed) and substratified by medication (all individuals with steroid therapy removed). Box plots represent the 20th, 50th and 75th percentile, while whiskers illustrate the 5th and 95th percentile: 1) negative control candidate for differential expression without epigenetic modification; 2) negative control candidate for epigenetic modification without differential mRNA expression.

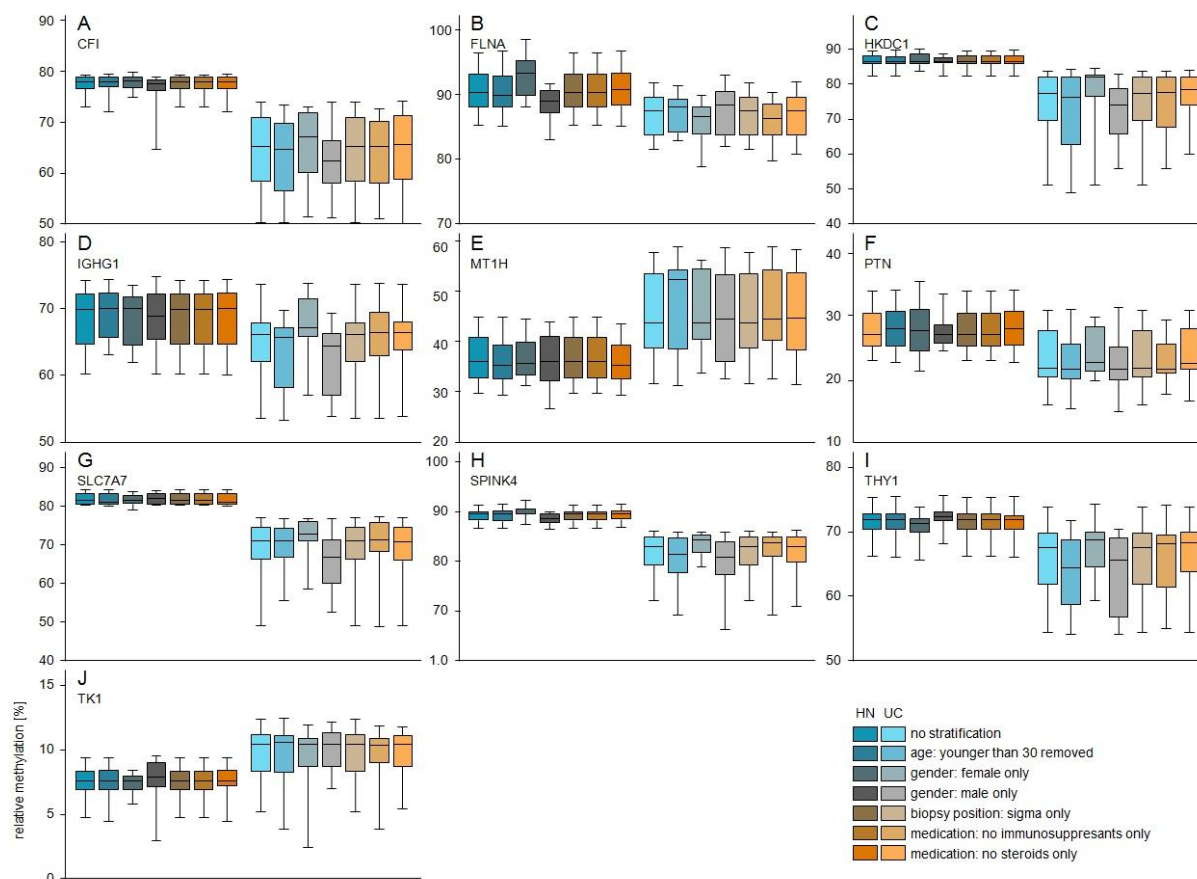


Figure 15. Impact of potential confounding factors on the differential DNA methylation. Each originally selected candidate loci is represented in a separate panel (A-J). Each panel shows 7 color coded sets of analysis from validation panel II, split into healthy individuals (HN) and ulcerative colitis patients (UC): no substratification (all samples included); substratified by age (all individuals younger than 30 years were excluded); substratified by gender (females only); substratified by gender (males only); substratified by biopsy position (all non-sigmoidal biopsies excluded); substratified by medication (all individuals with immunosuppressant therapy removed) and substratified by medication (all individuals with steroid therapy removed). Box plots represent the 20th, 50th and 75th percentile, while whiskers illustrate the 5th and 95th percentile. 1) negative control candidate for differential expression without epigenetic modification; 2) negative control candidate for epigenetic modification without differential mRNA expression.

Combining the the three screening layers and the validation shows 148-fold more differentially methylated MVPs in proximity to the 10 validated loci than expected by chance, corresponding to a p-value of $p=9.11 \times 10^{-44}$ (Fischer's exact test).

3.3 Potential new finding of the DNA hydroxymethylation in the acute inflammation

The altered DNA methylation in ulcerative colitis and disease control individuals (acute interstitial inflammation) indicated that the RNA expression pattern might be associated to with different types of DNA modification (see figure 3). 10 of TaqMan assays (Primer Express 3.0, Applied Biosystems) were designed to examine the DNA hydroxymethylation status of similar regulatory elements (500

bp around TSS) of the 10 candidates in a smaller collection of primary tissue from the intestinal mucosa (validation panel III, table 3). 4 assays did not lead to quantifiable results, the remaining 6 results are shown in figure 16 and 17. It appears that in disease control individuals, DNA hydroxymethylation is higher than in ulcerative colitis when analyzing identical genomic regions, however, the observed trends did not show significance in the individual candidate loci. In contrast to that, a combined analysis taking all loci into account shows a significantly increased hydroxymethylation in disease control individuals when comparing them to ulcerative colitis patients.

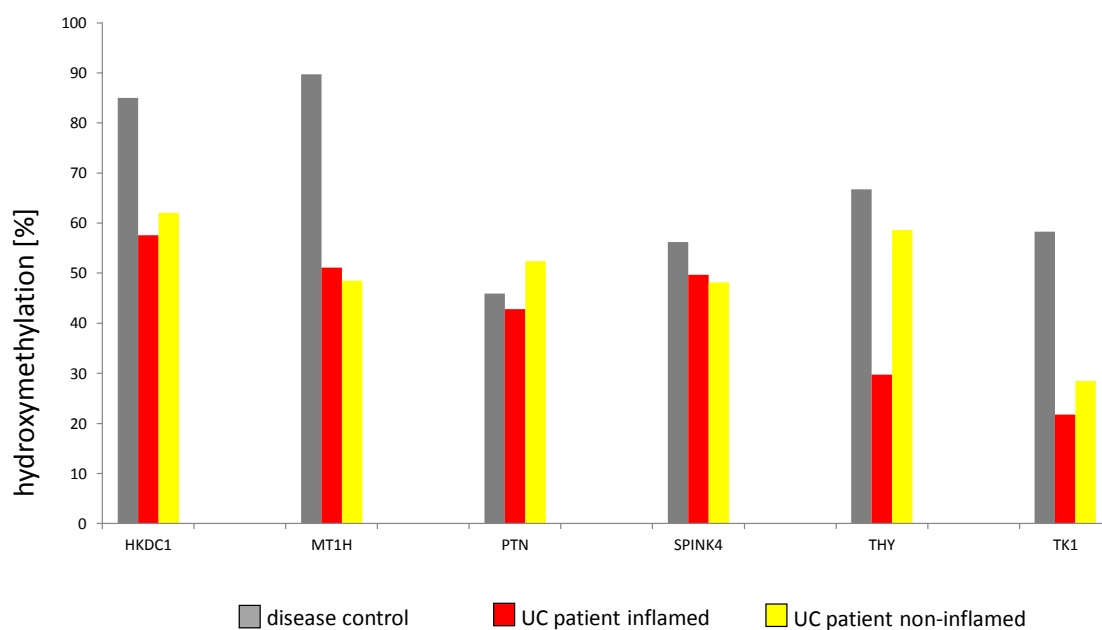


Figure 16. Hydroxymethylation levels in promoters of candidate loci: HKDC1, MT1H, PTN, SPINK4, THY and TK1. The detected hydroxymethylation levels of each promoter is displayed as relative methylation (%).

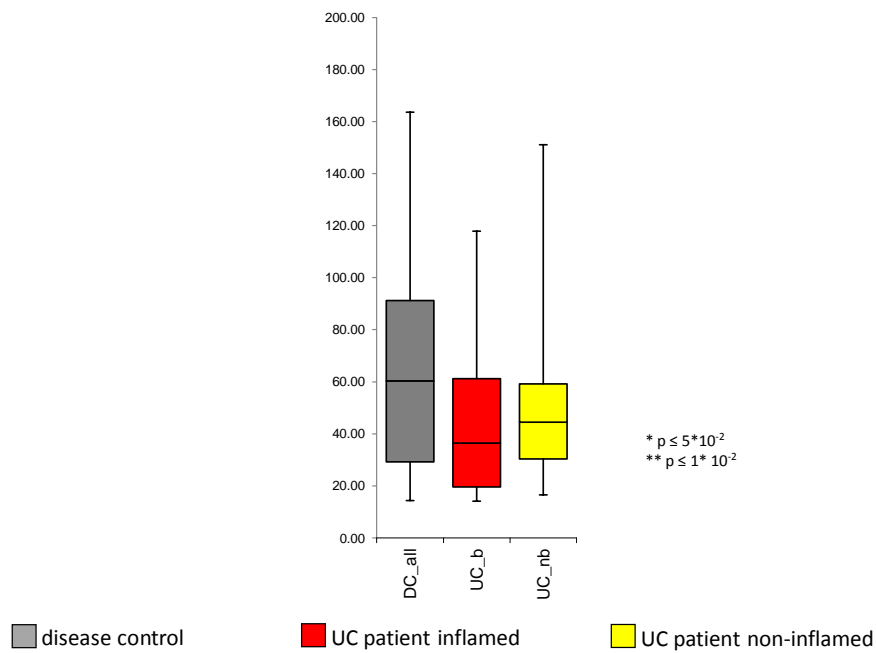


Figure 17. Hydroxymethylation detection of the candidate promoters: HKDC1, MT1H, PTN, SPINK4, THY and TK1. The detected hydroxymethylational level of all promoters in disease controls (5 individuals, biopsy from inflamed region) is significantly higher than in biopsies from ulcerative colitis patients (5 individuals, each provided two biopsies, one from an inflamed region (UC_b), the other from not inflamed region(UC_nb)).

4 Discussion

This study aims to create a functional epigenetic map by combining genome-wide data of differentially methylated regions (DMRs), methylation variable positions (MVPs) and transcriptome data, reflecting potential effects of the epigenetic variations. A group of 20 monozygotic twins, discordant for ulcerative colitis (UC) were selected as an entry level to perform this three-layer genome-wide scan to target mechanisms that show significant differences between healthy and diseased individuals. From this genome-wide map, selected candidate loci were subjected to further validation. Finally, initial results on the role of DNA demethylation inflammatory disorders, representing a key element in the dynamics of epigenetics, were collected as a starting point for follow-up studies. Genetics and epigenetics in the pathogenesis of ulcerative colitis.

4.1 Genetics and epigenetics in the pathogenesis of ulcerative colitis

Recently, substantial advances have been made in the understanding of the molecular pathogenesis of ulcerative colitis, a large number of genetic and environmental factors were suggested to be associated with the susceptibility, manifestation and progression of UC (Jaenisch and Bird, 2003; Feil and Fraga, 2011).

In the last few years, genome-wide studies and meta-analysis focusing on UC susceptibility loci have been highly successful in identifying candidates that contribute to disease susceptibility: Up-to-date, 47 risk loci have been identified (Anderson et al., 2011). The genetic variants are thought to be strongly associated with innate and adaptive immune response, autophagy, apoptosis, cellular metabolism and phagocytosis (Anderson et al., 2011). This is in concordance with these observations on the transcriptome side, where the identified candidates are either directly involved in inflammatory processes, or modulate them. However, while genetic variation represents a rather static element, the transcriptome displays many dynamic features. Generally, genetic variants have at least two possibilities to participate the initiation and manifestation of disease:, either by changing the structure of the mRNA transcript, or by altering the transcriptional status. For example rs11209026, which located in the transcribed region and encodes the stop codon, results in a dysfunctional stop-codon, leading to altered amino-acid sequences. In addition, other variants which are localized within regulatory elements, such as rs6983267, has been mapped to conserved sequences within a 500-kilobases region upstream of the MYC oncogene, which may affect cancer development through altered regulation of MYC expression (Sur et al., 2012). Recently, Cowper-Sal-Lari and colleagues reported that the disease risk-associated Single-nucleotide polymorphism (SNP) can modulate the affinity of chromatin for transcription factor and alter gene expression

(Cowper-Sal Lari et al., 2012). Based on the current knowledge, none of the observed disease-associated alterations on transcriptional level are associated with the 47 previously published UC risk loci. Consequently, the 61 risk loci identified in this study are novel. Interestingly, interleukin-1 receptor type II (IL1R2) was detected to be significantly downregulated in this genome-wide transcriptome analysis. This gene is a confirmed risk loci (rs2310173) for UC and therefore a genes of interest, yet the question whether rs2310173 is associated with altered IL1R2 mRNA expression remains unanswered. In fact, the variant is localized downstream of the IL1R2 coding region in the transcriptional direction, and it is well known that IL1R2 belongs to the family of interleukin-1 receptor-like genes (Dale and Nicklin, 1999). In this gene family, the type I IL-1 receptor (IL1R1) is the high-affinity component of the receptor for IL-1. The type II IL-1 receptor (IL1R2) has a three-Ig-domain extracellular portion, like IL1R1, and binds IL-1, but lacks a Toll-like cytoplasmic domain and gives no detectable cytoplasmic signal. IL1R1 and IL1R2 are mapped to the pericentromeric region of chromosome 2q, and are both transcribed in the same direction. It was thought that both genes are in the same transcriptional cluster (Dale and Nicklin, 1999), and surprisingly, the risk loci rs2310173 localized between them, and near to IL1R2. The genome-wide data showed that IL1R2 was downregulated in UC significantly, but IL1R1 not. This may indicate indirectly that the risk loci could play a role in the transcription, and contribute to the initiation or manifestation of UC.

dbSNP	Chr	MVP in proximity	MVP distance [bp]	MVP p-value	MVP fold change	DMR in proximity	DMR distance [bp]	DMR p-value	DMR fold change	Associated genes/transcripts (p-value / fold change)
rs254560	5	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	²⁾
rs267939	5	cg21046940	8544	0.789	1.24	4_750	3442	0.974	1.08	<i>DAP</i> (0.517 / 1.08)
rs678170	11	cg27227156	1670	¹⁾	¹⁾	97_557	1049	0.955	1.08	²⁾
rs734999	1	cg05610148	4210	0.525	1.23	388_658	4183	1.031	-1.08	<i>TNFRSF14</i> (0.971 / -1.0 ²⁾ ; <i>MMEL1</i> (0.390 / -1.1 ³⁾ ; <i>PLCH2</i> ³⁾
rs798502	7	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	<i>GNA12</i> (0.199 / 1.05)
rs907611	11	cg26158194	249	0.496	-1.16	584_700	13	1.031	1.02	<i>LSP1</i> (0.937 / 1.16)
rs941823	13	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	²⁾
rs943072	6	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	²⁾
rs1297265	21	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	²⁾
rs1801274	1	cg27470554	4504	0.315	-1.12	358_172	3917	0.974	-1.01	<i>FCGR2B</i> (0.679 / 1.15)
rs2155219	11	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	²⁾
rs2297441	20	cg18611245	10654	0.853	-1.26	159_97	433	0.955	1.07	<i>SLC2A4RG</i> ³⁾ ; <i>STMN3</i> ³⁾ ; <i>ZBTB46</i> ³⁾ ; <i>ZGPAT</i> (0.945 / -1.00); <i>RTEL1</i> ³⁾
rs2310173	2	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	<i>IL1R2</i> (0.065 / -2.31)
rs2836878	21	¹⁾	¹⁾	¹⁾	¹⁾	741_343	33734	0.955	-1.01	²⁾
rs2838519	21	¹⁾	¹⁾	¹⁾	¹⁾	517_389	20907	0.974	-1.06	<i>ICOSLG</i> ³⁾
rs2872507	17	cg14914852	15965	0.525	1.08	600_252	15642	0.955	-1.06	<i>IKZF3</i> ³⁾ ; <i>ORMDL3</i> (0.937 / -1.0 ³⁾ ; <i>IKZF3</i> ³⁾ ; <i>PNMT</i> ³⁾ ; <i>ZBP2</i> ³⁾
rs3024505	1	cg17067005	5443	0.315	-1.08	143_227	22	0.974	-1.09	<i>IL10</i> ³⁾ ; <i>IL19</i> ³⁾
rs3194051	5	cg04312209	19143	¹⁾	¹⁾	105_1001	18996	0.974	1.27	<i>IL7R</i> ³⁾
rs4246905	9	cg11809085	14508	¹⁾	¹⁾	643_137	14915	1.031	1.03	<i>TNFSF8</i> ³⁾ ; <i>TNFSF15</i> ³⁾
rs4510766	7	cg01025762	38575	¹⁾	¹⁾	207_929	38617	0.955	1.19	²⁾
rs4676406	2	cg07432969	9550	0.525	-1.02	218_140	11648	0.955	-1.07	<i>GPR35</i> (0.846 / 1.0 ²⁾
rs4728142	7	cg26616347	3786	0.912	1.25	168_892	3779	1.031	-1.02	<i>IRF5</i> ³⁾ ; <i>TNPO3</i> (0.954 / -1.00)
rs5771069	22	¹⁾	¹⁾	¹⁾	¹⁾	195_653	2214	1.031	-1	<i>PIM3</i> (0.837 / 1.70); <i>IL17REL</i> ³⁾
rs6017342	20	cg00431114	39000	¹⁾	¹⁾	645_453	28952	1.031	-1.02	<i>SERINC3</i> (0.199 / -1.16)
rs6426833	1	cg09195271	29362	0.525	1.08	53_167	29661	0.955	1.14	²⁾
rs6451493	5	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	<i>PTGER4</i> (0.670 / -1.05)
rs6499188	16	cg19034028	2577	¹⁾	¹⁾	693_223	3875	0.974	-1.11	<i>ZFP90</i> (0.945 / -1.00)
rs6584283	10	¹⁾	¹⁾	¹⁾	¹⁾	39_119	326	1.031	-1.08	²⁾
rs6871626	5	¹⁾	¹⁾	¹⁾	¹⁾	18_750	46859	0.974	-1.01	<i>IL12B</i> ³⁾
rs6911490	6	cg19464016	11932	0.789	1.38	736_824	11920	1	1.03	<i>PRDM1</i> ³⁾
rs6920220	6	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	³⁾
rs7134599	12	¹⁾	¹⁾	¹⁾	¹⁾	521_873	43646	1.031	1.15	<i>IFNG</i> ³⁾ ; <i>IL26</i> ³⁾
rs7524102	1	¹⁾	¹⁾	¹⁾	¹⁾	440_344	47983	1.031	-1.06	²⁾

dbSNP	Chr	MVP in proximity	MVP distance [bp]	MVP p-value	MVP fold change	DMR in proximity	DMR distance [bp]	DMR p-value	DMR fold change	Associated genes/transcripts (p-value / fold change)
rs7554511	1	cg15250507	17196	0.525	1.14	248_116	3033	0.955	-1.17	<i>C1orf106</i> (0.945 / -1.0 ²)
rs7608910	2	cg09469394	39133	0.853	1.17	200_246	39766	1.031	1.03	<i>PUS10</i> ³)
rs9268853	6	cg19248557	21804	0.912	1.09	328_1008	450	0.955	1.04	<i>HLA-DRB5</i> ³); <i>HLA-DQA1</i> (0.856 / 1.0 ²); <i>HLA-DRB1</i> (0.517 / 1.61); <i>HLA-DRA</i> (0.517 / 1.68)
rs9822268	3	cg21958034	6112	0.315	-1.07	374_156	3416	0.955	1.19	<i>UBA7</i> ³); <i>MST1</i> ³); <i>APEH</i> (0.665 / -1.30); <i>AMIGO3</i> ³); <i>GMPPB</i> (0.937 / 1.19)
rs10758669	9	cg20394284	2605	0.853	1.02	20_920	2943	1.031	-1.05	<i>JAK2</i> (0.517 / 1.06)
rs10781499	9	cg02516189	1688	0.525	-1.11	480_724	672	1.031	-1.02	<i>CARD9</i> ³); <i>INPP5E</i> (0.962 / -1.00); <i>SDCCAG3</i> (0.665 / -1.1 ²); <i>SEC16A</i> ³)
rs11209026	1	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	<i>IL23R</i> ³)
rs11676348	2	cg14150666	18835	0.496	-1.07	188_388	19018	1.031	1.04	<i>IL8RA</i> ³); <i>SLC11A1</i> ³); <i>IL8RB</i> ³); <i>AAMP</i> (0.856 / -1.07)
rs11739663	5	cg19108785	17929	0.853	1.42	467_749	8145	0.974	-1.08	<i>EXOC3</i> (0.040 / -1.36)
rs12261843	10	cg23218877	18043	0.912	1.02	¹⁾	¹⁾	¹⁾	¹⁾	<i>CCNY</i> ³)
rs16940202	16	¹⁾	¹⁾	¹⁾	¹⁾	700_466	23	0.955	-1.16	²⁾
rs17085007	13	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	¹⁾	²⁾
rs17388568	4	cg17850932	48539	0.789	-1.08	3_485	28543	1.031	-1.04	<i>IL21</i> ³)
rs35675666	1	cg25007680	151	0.496	1.16	509_517	40	0.955	-1.08	<i>TNFRSF9</i> ³); <i>ERFF1</i> ³); <i>UTS2</i> ³)

Table 11. dbSNP: previously published loci, associated to ulcerative colitis (Anderson et al., 2011); MVP: methylation variable positions (listed with its ID); DMR: differentially methylated region; distance describes the distance between the SNP and MVP/DMR position in bp; associated genes/transcripts are listed with their gene symbol, p-value and fold change; 1) the SNP is not in proximity of a detectable MVP/DMR; 2) the SNP is not in proximity of a detectable gene/transcript; 3) the associated gene/transcript was categorized as not expressed in the tissue.

Interestingly, some risk alleles are cytosine or represent a CG element, but alteration of DNA methylation between UC and healthy individuals could not be detected in the genome-wide data, since the monozygotic twins have theoretically the same genotype: Consequently, the presented genome-wide study setup employing monozygotic twins does not favor detecting such effects, yet they may play a disease relevant role.

4.2 Environment and epigenetics in the pathogenesis of ulcerative colitis

Environmental factors have been considered to play a role in the pathogenesis of numerous complex disorders, such as UC (Jaenisch and Bird, 2003; Feil and Fraga, 2011). Epidemiological studies investigating risk factors are based on tangible evidence, for example, food, air pollution, as well as many other elements of the biological and psychological environment, all of which may contribute to disease susceptibility. This is documented by studies employing adopted individuals, families and monozygotic twins, which are carried out with the implicit assumption that any variation that is not attributable to genetic factors must origin from the environment (Hemminki et al., 2006). This assumption, however, could not be validated in this study. Genetic and non-genetic factors such as environment are not always easily distinguishable, as the microbiota, which is often considered an environmental factor, is also strongly dependent on the host genotype: In an experimental colitis model in mice Nod2 deficient mice could transmit their disease risk to wildtype animals via their microbiota (Couturier et al., JCI 2012).

Epidemiological studies targeting environmental influence on complex diseases are vulnerable to variations due to the large variety of multidirectional environmental influences which oftentimes cannot be estimated objectively (Taubes, 1995). These difficulties indicate that data from epidemiological studies require to be supported other experiments or data.

Another uncertainty is introduced by the observation that environmental risk factors seem to have 'heritable' partners which are embedded in the human genome (Kendler and Baker, 2007). For example, a Swedish twin study demonstrated that the tendency to smoke has a heritable component, indicating that Genetic factors play an important etiologic role in regular tobacco use (Kendler et al., 2000). Based on studies targeting depression, Kendler and colleagues suggested that genetic and environmental risk factors both contribute to the aetiology of complex disorders, partially explained by the finding that individuals with a defined genetic risk have the tendency to select high-risk environments (Kendler and Karkowski-Shuman, 1997).

This study setup is based on ten pairs monozygotic discordant twins, following the concept of a 'non-shared environment' as suggested earlier (Lyons, 1996; Hebebrand et al., 2001). Briefly, this concept suggested that such setups have the possibility to find variations that cannot be attributed to heritable factors and that environmental influences which result in similarities in family members are not considered. In this context, a meta-analysis revealed that the proportion of the variation of individual traits that is attributable to non-shared environment among individuals is 45–60%, whereas the variation attributable to shared environment is nearly zero (Bouchard and McGue, 2003). Consequently, non-shared environment could make an important contribution to the risk of developing complex diseases, such as UC.

Taken together, investigating of the role of environmental factors in complex diseases generally needs to be broken down into specific questions. Currently, it is not possible to design a study that could monitor all environmental influences. For this reason, monozygotic discordant twins were selected to generate an in-depth picture of the UC epigenome aiming to minimize genetic and environmental influences.

4.3 Ulcerative colitis: a complex epigenetic disorder

In total, 361 differently expressed transcripts and 1048 differentially methylated DNA elements were identified when comparing UC patients to healthy individuals. A subset of 10 cis linked expression-methylation pairs and 2 negative controls were further verified in a larger patient group. It is unclear to which extend epigenetic modifications might contribute to disease susceptibility, manifestation and progression, however, epigenetic mechanism are currently being discussed to be involved in establishing specific transcription patterns (Feil and Fraga, 2011). DNA methylation represents only one of many epigenetic modifications. As described in the introduction, certain forms of histone modification cause local formation of heterochromatin, which is readily reversible, whereas DNA methylation leads to a more stable long-term repression. Interestingly, recent studies reported that DNA methylation and histone modification do not act in an isolated manner, but can be dependent on one another, which was demonstrated in the context of tumorigenesis (Cameron et al., 1999). It is important to note that DNA methylation and histone modification are carried out by different chemical reactions and require different sets of enzymes. The relationship between DNA methylation and histone modification might be partially mediated through methylcytosine-binding proteins, such as methyl-CpG binding domain protein 2 (MeCP2 or MBD2), which are capable of recruiting histone deacetylases to the methylated region (Cedar and Bergman, 2009). However, the exact nature of the relationship between these epigenetic mechanisms is still not clear. DNA hypermethylation is not in all cases the cause of transcriptional silencing. Early experiments by Lock

and colleagues showed that the hypoxanthine phosphoribosyltransferase (Hprt) gene on the inactive X chromosome is methylated only after inactivation of the whole chromosome (Lock et al., 1987). Other genome-wide studies in cancer cells showed that genes with CpG island promoters that are already silenced by Polycomb complexes are being methylated earlier than other genes in cancer scenarios (Ohm et al., 2007; Schlesinger et al., 2007; Widschwendter et al., 2007; Gal-Yam et al., 2008). In addition to epigenetic modifications, transcriptional patterns sometimes depend on specific transcription factors. In 2010, Chi and colleagues reported that ets variant 1 (ETV1) is universally highly expressed in gastrointestinal stromal tumor (GIST) cell lines and transcriptome profiling and analysis of ETV1-binding sites indicated that ETV1 is a master regulator of an ICC-GIST-specific transcription network (Chi et al., 2010). Similarly it was recently reported that loss of epigenetic modification driven by the foxp3 transcription factor leads to regulatory T cell insufficiency (Bettini et al., 2012).

The majority of identified loci in the presented study are associated to biological processes which are either directly or indirectly linked to immune processes, which is consistent with previous findings on functional genomics of UC (Dieckgraefe et al., 2000; Lawrance et al., 2001; Costello et al., 2005) as well as with findings on UC genetics (Anderson et al., 2011). Interestingly, a substantially smaller proportion of biological processes identified are distinct from bona fide immune processes: Structural, developmental and metabolic processes are prominent findings. This suggests additional levels of transcriptional control and strongly indicates that altered immune functions in cells of the intestinal mucosal might be involved in the aetiology of UC. In this context, it is widely accepted that the final target of the inflammatory processes appears to be the colonic epithelium. Numerous proinflammatory cytokines, such as IL-1, IL-6 and TNF- α , have been reported to be increased in serum (Gross et al., 1991) and in primary tissues (Isaacs et al., 1992; Stevens et al., 1992; Breese et al., 1993) of UC patients. On the other hand, most cytokine receptors are not only detected on the antigen presenting cells (APCs), but also constitutively expressed on intestinal epithelial cells (IECs) (Panja et al., 1998). There is increasing evidence that IECs may serve as targets for these locally produced cytokines. For example, IL-1 β up-regulates IL-6 synthesis (a proinflammatory cytokine) in epithelial cells (Panja et al., 1995). Increasing evidence exists indicating that epithelial cell growth and differentiation ensures maintenance of homeostasis and plays an indispensable role in the pathogenesis of inflammatory disorders.

Previous studies have addressed the epigenetic modulation in the context of IBD mostly on a candidate gene approach, focused on IBD-associated colorectal cancer (Moriyama et al., 2007; Dhir et al., 2008; Edwards et al., 2009; Gonsky et al., 2011). Together with the high tissue specificity

(Rakyan et al., 2008) and the fundamental differences in detection, these studies are only of limited use for comparison to the systematic genome-wide approach. However, The results support one recent report stating that epigenetic dysregulation of the IRF5 promoter is not associated with IBD (Balasa et al., 2010). In contrast to that, many of these findings on the mRNA level are in concordance with previously published studies (Dieckgraefe et al., 2000; Lawrance et al., 2001; Anderson et al., 2011), potentially attributed to the lower technical and/or biological variance in inflammation associated mRNA patterns. Among the replicated findings, several genes have been directly associated with chronic intestinal inflammation: pleiotrophin (PTN) has been shown to be functionally linked to inflammation and cancer (Kadomatsu, 2005), while Thy-1 cell surface antigen (THY1) mediates cell adhesion during inflammation (Jurisic et al., 2010). In the same context, the serine protease inhibitor 4 (SPINK4) has been shown to be differentially expressed in chronic autoimmune intestinal inflammation (celiac disease), likely derived from altered goblet cell activity, while no causative genetic variant was identified (Wapenaar et al., 2007). Although these candidate approaches shed light on individual genes, the disease relevance for UC and other chronic or autoimmune disease was not demonstrated. In fact, all the those results are linked to non-chronic inflammation, leading to the fundamental question which mechanisms leads a stimulation of these inflammatory transcripts persistently or reiteratively in the intestinal mucosa of UC.

4.4 Genome-wide patterns of hypomethylation are a prominent motive in UC

The presented study reported cis-linked DNA methylation transcription partners in UC. Interestingly, most significant methylation candidates are hypomethylated when comparing UC patients to healthy individuals (Table S 1. &2.).

Loss of 5-mC is a broadly discussed topic in the field of epigenetic research, numerous studies indicate that DNA demethylation plays significant roles in various malignant events (Ehrlich, 2009; Rai et al., 2010; Watanabe and Maekawa, 2010; Wild and Flanagan, 2010). Thus, the degree of demethylation may serve as a putative molecular biomarker with predictive and prognostic value. This study provides insight into the genome-wide 5-mC demethylation in primary human tissue from the intestinal mucosa of UC patients and healthy individuals while controlling for genetic interference. It is tempting to speculate that the high prevalence of tumorigenesis in UC patients might be partially attributed to epigenetic modifications like the observed decreased genome-wide demethylation (Figure 18). In relevant cell culture models, an decrease in 5-mC levels via alteration of DNMT1 activity was shown to silence tumor suppressor genes (Peng et al., 2011). Taken together, this study provides multiple layers of support that genome-wide loss of 5-mC is a epigenetic motive

in ulcerative colitis, with potential diagnostic and prognostic advantages for the future when compared to global DNA hypomethylation, a recognized epigenetic motive of carcinogenesis.

The potential clinical application of 5-mC could be supported by the possibility to detect 5-mC immunohistochemically. DNA demethylation has been extensively observed in numerous malignancies such as breast, ovarian, and colorectal carcinomas, the latter being closely associated to UC (Ehrlich, 2009; Watanabe and Maekawa, 2010; Wild and Flanagan, 2010). As the anti-5-mC antibody for immunohistochemical staining is commercially available, this approach has been well established in several diagnostic laboratories. In this context, These findings could be one of the very first steps, leading to the development of a new, simple, sensitive and practical adjuvant diagnostic assay, following the example of Lian and colleagues (Lian et al., 2012).

DNA demethylation in stem cells and tumor cells raises the intriguing question about the nature of the biological role of 5-mC in cell differentiation, self-renewal and malignant transformation. Studies of genomic reprogramming indicate that DNA demethylation is involved in the regulation of cell differentiation and carcinogenesis (Cameron et al., 1999; Ito et al., 2010; Rai et al., 2010; Ficiz et al., 2011; Williams et al., 2011b; Wu et al., 2011b; Doege et al., 2012; Lian et al., 2012). Until now, tissue-specific 5-mC distribution and genome-wide mapping of 5-mC in a non-malignant disease have not been well studied. The findings presented here on genome-wide hypomethylation in the intestinal mucosa of UC patients provide new insights on the supporting role of DNA demethylation in pathways that are fundamental to cellular differentiation and dedifferentiation.

4.5 DNA demethylation in chronic and acute colitis

In general, the term DNA methylation refers to the enzymatic modification at the 5-position of cytosine in a CpG dinucleotide (Wigler et al., 1981; Robertson and Wolffe, 2000; Zhang et al., 2006; Heyn and Esteller, 2012; Jones, 2012) sequence context. This DNA modification was thought inheritable and 'stable' without primary DNA base sequence changes resulting in possible epigenetic regulation of phenotype and gene expression (Robertson and Wolffe, 2000; Weaver et al., 2004; Ma et al., 2009; Urduinguio et al., 2009; Wu et al., 2010). In the presented study, a genome-wide demethylation was found in the colonic mucosa. Following this pattern, the activation of various inflammatory genes associated a downregulation of DNA methylation in their regulatory regions, such as promoter regions or the gene body was observed. Especially, the correlation analysis of quantitative expression and methylation data revealed a significant negative correlation in all verified candidate transcripts (median spearman-rho $r = -0.58$). This finding is in concordance with the former hypothesis, that the DNA methylation regulates transcription negatively (Holliday and

Pugh, 1975; Wigler et al., 1981; Clark et al., 1997). The presented observations indicate that DNA demethylation might play a very important role in the pathogenesis of UC. On the other hand, it could be demonstrated that DNA demethylation in UC might be entirely different from DNA demethylation found in acute colitis. As showed in the example (figure 13), the upregulated transcription of HKDC1 accompanied by a significant downregulated DNA methylation of its regulatory region in UC, but not in disease controls. In the pilot study, I found that the hydroxymethylation levels of selected transcripts in biopsies from disease controls are significant higher than in biopsies from ulcerative colitis patients.

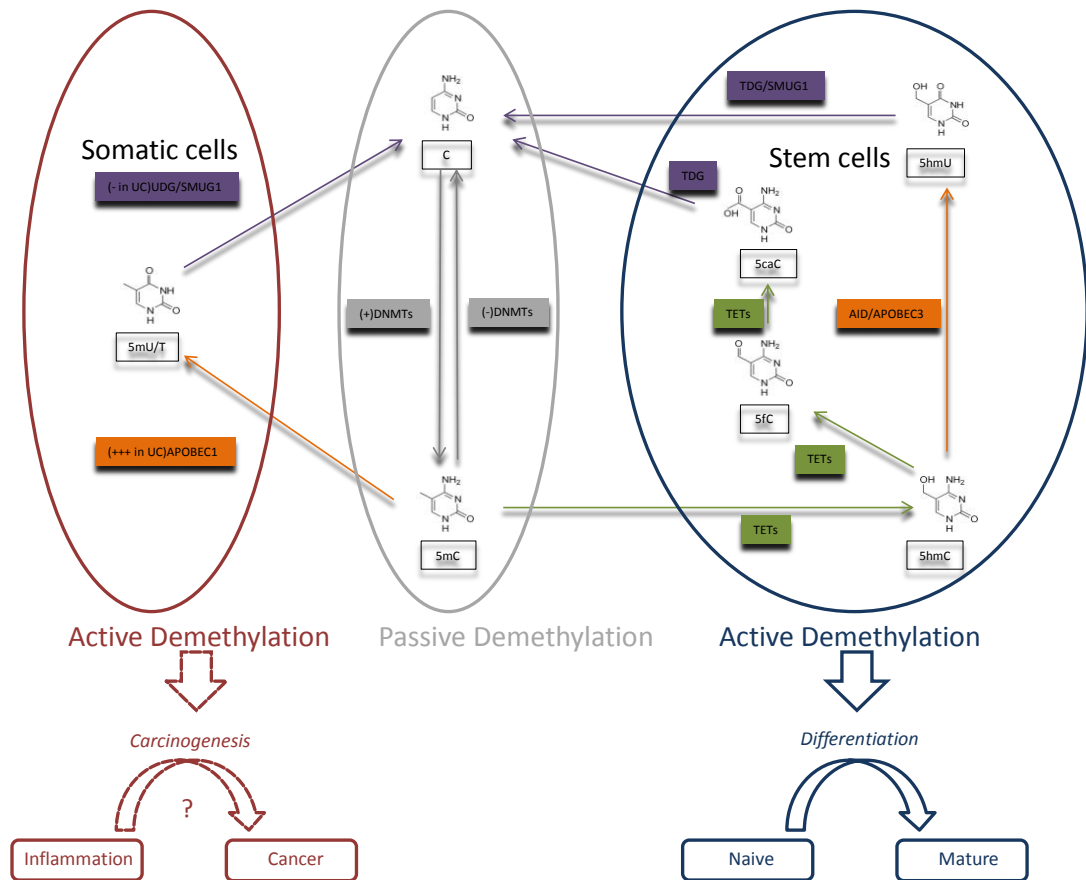


Figure 18. Potential DNA demethylation pathways. Passive DNA demethylation might be due to a reduction in activity or absence of DNA methyl transferases (DNMTs) (grey). Three enzyme families have been identified currently in active DNA demethylation. First, 5-methylcytosine (5-mC) can be hydroxylated by the ten-eleven translocation (TET) family of enzymes (green), to form 5-hydroxymethylcytosine (5-hmC). The 5-hmC could be further oxidized to 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC). Second, 5-mC and 5-hmC could be further deaminated by the APOBEC family members (orange) to form 5-methyluracil (5mU/Thymine) and 5-hydroxymethyluracil (5hmU) respectively. Third, 5mU, 5hmU and 5-caC could be replaced with unmodified cytosine by DNA repair of base excision repair (BER) glycosylases (violet), such as uracil-DNA glycosylase/ single-strand-selective monofunctional uracil-DNA glycosylase 1 (UDG/SMUG1) and thymine-DNA glycosylase (TDG)/SMUG1 in the somatic cells and stem cells respectively, culminating in DNA demethylation.

4.5.1 Passive mechanisms of DNA demethylation are potentially absent in UC

DNA methyl transferases (DNMTs) are responsible for the establishment and maintenance of DNA methylation in mammalian cells. It has been thought that downregulation of the DNMT activity leads a loss of DNA methylation, such as in early mammalian development (Monk et al., 1991; Rougier et al., 1998). More recent studies have reported that *de novo* methylation in early development is established by DNA methyltransferases 3A (DNMT3A) and 3B (DNMT3B) (Aoki et al., 2001; Zhao et al., 2009; Wu et al., 2010), and the methylation patterns are inheritably maintained through cell divisions by DNMT1 (Klein et al., 2011; Shock et al., 2011; Song et al., 2012). The genome-wide data presented in this study reflected no significant transcriptional alteration of DNMT family members between UC patients and healthy twins, and it could not be shown that the activity of DNMTs is downregulated in UC, suggesting that the passive DNA demethylation might be not responsible for the demethylation in UC. Consequently, the loss of methylation in mucosal tissue of UC patients could be mediated by an active mechanism. Previous studies suggested that DNA demethylation could be catalyzed by a rapid and active mechanism, independent of cell division (Paroush et al., 1990; Hsieh et al., 2009; Wu and Zhang, 2010). Until now, an enzyme which can directly 'cut off' the methyl group was not found, suggesting that such a biochemical reaction might not be as simple as originally proposed (Bird, 2002). The presented data indicates that genome-wide hypomethylation in the colonic epithelial tissue might be reprogrammed enzymatically (so-called active DNA demethylation), thus DNA demethylation could play an essential role in the initiation and manifestation of UC.

4.5.2 Demethylation in mammalian cells: a multi-step process

In contrast to histone modifications, epigenetic variations which take place directly on the DNA molecular are not too much. For example, the recently rediscovered 5-hmC, which has been reported half century ago (Wyatt and Cohen, 1952), but its biological significance was ignored until the identification of the ten-eleven translocation (TETs) enzyme family, which hydroxylated 5-mC to 5-hmC. A series of studies (Ficz et al., 2011; Gu et al., 2011; He et al., 2011; Williams et al., 2011b; Wu et al., 2011b; Doege et al., 2012) has identified key players in this process, for example, the 5-hmC may serve as the intermediates in the DNA demethylation. Initially, the TET family members in mammalian cells were characterized as fusion partners of the MLL protein in acute myeloid leukemia and named leukemia-associated proteins (LCX)(Ono et al., 2002a)(Ono et al., 2002b). However, recently, Tahiliani and colleagues reported that TETs could potentially modify the 5-mC, and convert 5-mC to 5-hmC (Tahiliani et al., 2009). This finding rekindled the interest on 5-hmC, as the hydroxymethylated cytosine can be deaminated by activation-induced cytidine deaminase/

apolipoprotein B mRNA editing enzyme, catalytic polypeptide (AID/APOBEC) family members, leading to the possibility of complete demethylation through DNA repair. Current studies targeting 5-hmC were mostly focused on the epigenetic reprogramming of embryonic stem cells. It is important to note, that 5-hmC is abundant in proliferative tissues such as pluripotent embryonic stem cells (ESCs), but it is also present at lower levels in somatic tissues and organs, such as blood, lung, kidney, and muscle (Tahiliani et al., 2009; Globisch et al., 2010; Khare et al., 2012). This is in concordance with the presented study, where hydroxymethylation was altered in the colonic mucosa, suggesting an active mechanism.

Based on the current knowledge, three enzymatic families might be involved in the active DNA demethylation in mammalian cells (Gehring et al., 2009; Tahiliani et al., 2009; Wu and Zhang, 2010; Cortellino et al., 2011): The ten-eleven translocation (TET) family, which can further modify 5-mC and hydroxylate it; the AID/APOBEC family, which can deaminate the modified cytosine and the family of base excision repair (BER) glycosylases, which can replace the Anomalous base with an unmodified cytosine, and finalize the demethylation (Figure 18). The genome-wide transcriptional data presented here revealed that APOBEC1, one member of the AID/APOBEC family, is significantly upregulated in UC than in healthy Co-twins (P value 0.04994; Foldchange 8.1304), indicating an enhanced enzymatic activity of deamination in the inflamed tissue of UC patients, suggesting a potentially causal role of this enzyme in maintaining the hypomethylated state in tissue of UC patients.

4.5.3 The APOBEC Family: Mediators of 5-mC or 5-hmC Deamination

APOBEC1, which was shown to be functionally associated to UC, belongs to a family of enzymes, which were originally identified as RNA editors, so called “apolipoprotein B mRNA-editing catalytic polypeptides,” or APOBECs (Conticello et al., 2007). Recent studies indicated that members of the APOBEC family could also mediate deamination of cytosine residues to uracil, followed by the deamination of the “mismatched cytosine” which then is repaired by base excision repair (BER) proteins (Nabel et al., 2012; Roberts et al., 2012). Like most DNA repair mechanism, BER is error prone, leading to mutations in mammals (Liu and Schatz, 2009; Maul and Gearhart, 2010). The upregulated APOBEC1 could potentially stimulate the genomic demethylation through deaminating methylated cytosines, while the deamination would also provide the probability of introducing mutations (Berger et al., 2011; Burns et al., 2013; Greenman et al., 2007; Jones et al., 2010; Parsons et al., 2011; Sjöblom et al., 2006; Stransky et al., 2011) due to the subsequently downregulated DNA repair mechanisms. This might partially explain why UC patients have a significantly higher risk to

develop colorectal cancer than CD patients, as the differential expression of APOBEC1 was observed only in UC (P value, 0.00005), but not in CD.

The presented study suggested that APOBEC1 might be a key player in active DNA demethylation in UC. This is in concordance with previous studies targeting nuclear reprogramming which provided support for the concept that AID (another APOBEC family member) plays an essential role in active DNA demethylation in mammalian cells (Conticello et al., 2007; Nabel et al., 2012). In a fusion-cell study, rapid demethylation at the promoters of human pluripotency genes (OCT4 and NANOG) was detected in human somatic cells (fibroblasts) while a fusion of an excess of mouse ESCs was observed (Tat et al., 2011). The knock-down of AID with RNA interference blocked pluripotency gene promoter demethylation and gene expression. Unlike TET, which has a DNA-binding motif (Ono et al., 2002b), AID and other APOBEC family members do not possess such motives. Cairns and colleagues showed the role of APOBECs in genome-wide demethylation in zebrafish embryos for the first time in 2008 (Rai et al., 2008): Upon overexpression of AID or zebrafish APOBEC deaminases and the glycosylase MBD4, active DNA demethylation was observed in zebrafish embryos injected with a methylated linearized nonreplicating DNA. Popp and colleagues reported a similar effect of AID in global DNA demethylation at a later stage of embryogenesis in mice (Popp et al., 2010). AID knock-out mice (AID^{-/-}) is viable and fertile, although with somewhat smaller litter sizes. This mouse however, exhibited a genome-wide hypermethylation in its primordial germ cells (PGCs) when compared to wild type animals, suggesting that AID is involved in DNA demethylation (Muramatsu et al., 2000). In this study, I found a higher expression of APOBEC family member (APOBEC1), and a global hypomethylation in the colonic biopsies of UC patients. All these findings raise the possibility that the APOBEC family members may play important roles in compensating high DNA demethylation levels.

4.5.4 Reduced BER activity may lead to carcinogenesis of UC

DNA methylation was sometimes suggested to be associated with tumorigenesis. Three main molecular mechanisms have been proposed: (1) a global decrease in DNA methylation, (2) increased uracil misincorporation during DNA replication, and (3) increased cytosine deamination at sites of DNA methylation (Rai et al., 2010; Loenarz and Schofield, 2011; Moréra et al., 2012). In the presented study, a genome-wide hypomethylation in the intestinal tissue of UC patients could be shown, and it is well known that UC has a relative higher risk of carcinogenesis. As discussed above, DNMTs were transcribed normally, and no evidence showed a reduced activity in UC, suggesting that a high frequency of demethylation and remethylation might contribute to the increased risk of carcinogenesis in UC patients.

Transitions of cytosine to thymine or uracil are the most common mutation in human cells and it are also found very frequently in human tumors (Berger et al., 2011; Burns et al., 2013; Greenman et al., 2007; Jones et al., 2010; Parsons et al., 2011; Sjöblom et al., 2006; Stransky et al., 2011). These transitions take place usually in a CpG context, and are strongly associated to deamination of 5-methylcytosine (5-mC) to thymine, or cytosine to uracil (Waters and Swann, 2000). Mismatches caused by deamination (T to G or U to G) are generally repaired by the base excision repair (BER) pathway. Recently, deamination has been suggested to precede BER in mammalian DNA demethylation (Cortellino et al., 2011; Gu et al., 2011). Based on the current knowledge, at least four different human DNA glycosylases may remove uracil and mismatched thymine in the genome, which is itself cytotoxic and potentially mutagenic. These enzymes are UNG, SMUG1, TDG and MBD4 (Krokan et al., 2002).

The glycosylases which are part of the BER pathway can be further classified based on their function: the involvement in the embryogenetic reprogramming, such as thymine-DNA glycosylase (TDG) and single-strand-selective monofunctional uracil-DNA glycosylase 1 (SMUG1) (Cortellino et al., 2011; Guo et al., 2011). Both glycosylases can convert diaminated or oxidated 5-mC to cytosine, and TDG knock-out mice are embryonic lethal. The direct interaction between AID and TDG has been demonstrated by coimmunoprecipitation experiments (Cortellino et al., 2011), suggesting that they might act in a partnership with TET and AID/APOBEC in the embryogenetic reprogramming. The other group seems to be responsible for the routine conversion of U to C or T to C in the somatic cells, including SMUG1 and UDG (Linhart et al., 2009; Skjeldam et al., 2010; Kemmerich et al., 2012). Indeed, SMUG1-knockout and SMUG1/UNG-double knockout mice could be bred normally and healthy beyond 12 months of age (Kemmerich et al., 2012), indicating that DNA glycosylases might be not only involved in the embryogenetic reprogramming, but also in other pathophysiological events, such as UC. In the presented study, apart from APOBEC1, which belongs to APOBEC deaminase family and is upregulated in UC when compared to healthy controls, it was also shown that UDG, which is one glycosylase of BER, is significantly downregulated in UC (P value, 0.0002), suggesting a stronger tendency of mutationogenesis. Consequently, the balance between demethylation and remethylation might be disturbed, increasing the probability of introducing mutations.

In addition, another potential DNA demethylation intermediates - 5-formylcytosine and 5-carboxylcytosine, should be mentioned here in short. Recent reports suggested that 5-hmC can be further oxidized by TET proteins to form 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) (He et al., 2011). Interestingly, 5-caC could not only be repaired by TDG, another study also raises the

possibility that a putative decarboxylase could directly convert 5-caC to cytosine independent of DNA repair (Ito et al., 2011). However, 5-fC and 5-caC are much less abundant than 5-hmC in the mammalian cells, and they have been detected only in ES cells recently, therefore, I did not perform any experiment to detect 5-fC or 5-caC in this study.

4.6 Concluding remarks and future studies

To further support the clinical relevance of the presented findings, especially in the context of genome-wide demethylation and the cis-link between DNA methylation and gene expression, further studies validating these results on a functional level are required. In a first step, which is currently on-going, differences in demethylation in the gut mucosa between ulcerative colitis and other non-IBD colitis are being assessed. This is carried out by examining the DNA hydroxymethylation and DNA methylation status in the gut mucosa between UC in comparison to non-IBD colitis. Bisulfite sequencing represents the gold standard for validating DNA methylation, but it cannot distinguish DNA methylation and hydroxymethylation (Jin et al., 2010). The recent development of a method for the quantitative validation of DNA hydroxymethylation, so-called oxidative bisulfite sequencing (Booth et al., 2012) will help to close this methodological gap.

DNA demethylation in the gut mucosa of UC is a hallmark of these findings. The enzymes APOBEC1 and UDG display a significantly altered expression as shown in the genome-wide data. One of the currently on-going therefore is to validate important enzymes that are potentially involved in the DNA demethylation pathways, as showed in figure 20. This will help to better understand DNA demethylation in the pathogenesis of UC.

Many studies indicate that dynamics of DNA demethylation might play a significant role in the pathogenesis of numerous malignant diseases. This suggests to monitor the DNA methylation and hydroxymethylation of colonic biopsies for classification and prediction of primary colorectal cancer.

DNA methylation and hydroxymethylation are not the only mechanisms regulating transcription. There are numerous other epigenetic mechanisms participating in this network. The presented results emphasize the necessity to characterize the dynamics of these epigenetic modifications and their cis-linked transcription to unravel the nature of this regulatory principle. Since ulcerative colitis represents an exemplary scenario for inflammatory disorders at interfaces, the strong disease associated epigenetic signatures presented here might encourage future studies in related diseases.

5 Summary

The etiology of ulcerative colitis (UC), a chronic inflammatory bowel disease, is still not fully understood. One of the current hypotheses is, that genetic and environmental factors modulate the epigenetic landscape and thus contribute to susceptibility, manifestation and progression of the disease. In the presented thesis, a three-layer epigenome-wide association study (EWAS) is reported, using intestinal biopsies from ten monozygotic twin pairs discordant for the manifestation of UC (n=20). The primary result is a map of links between disease-associated transcripts and epigenetic modifications. These candidates of cis-linked gene expression and DNA methylation (DNAm) were validated in two larger independent patient populations (n=185) by quantitative real-time PCR and bisulfite-pyrosequencing, resulting high DNAm/RNA correlations, indicating a potential impact of DNAm on transcription. Many of the identified candidate genes have been functionally implicated in inflammatory processes. In contrast to the observed hypomethylation of several candidate loci in UC, patients with acute inflammation showed no altered methylation. To validate an alternative epigenetic modification in these patients, hydroxymethylation was quantified at selected loci, indicating that this mechanism might play a role in acute inflammation when compared to chronic inflammation. In conclusion, this study represents the first replicated EWAS in affected tissue of UC patients, integrated with transcriptional signatures. The results indicate a potential role of epigenetic modification in disease manifestation and progression of UC.

6 Zusammenfassung

Die Ätiologie von colitis Ulcerosa (UC), einer chronisch entzündlichen Darmerkrankung, ist bisher noch nicht vollständig aufgeklärt. Eine der gängigen Hypothesen geht davon aus, daß genetische Faktoren und Umwelteinflüsse das epigenetische Profil verändern und somit zu Suszeptibilität, Manifestation und Progression der Erkrankung beitragen. In der vorliegenden Dissertation wird eine epigenomweite Assoziationsstudie (EWAS) vorgestellt, die auf der Verwendung von intestinalen Biopsien von 10 eineiigen Zwillingspaaren basiert, die diskordant für colitis Ulcerosa (UC) sind (n=20). Als primäres Ergebnis wurden krankheitsassoziierte Transkripte identifiziert, die von epigenetische Modifikationen begleitet werden. Diese Kandidaten von Genexpression und DNA-Methylierung (DNAm) in direkter Nachbarschaft („in cis“) wurden in zwei größeren Patientengruppen (n=185) mittels quantitativer Real-Time PCR und Bisulfid-Pyrosequenzierung validiert. Die resultierende hohe Korrelation zwischen DNAm und mRNA deutet auf den potentiellen Einfluss der DNAm auf das Transkriptom in UC hin. Viele der identifizierten Kandidatengene stehen in funktionellem Zusammenhang mit Entzündungsprozessen. Im Gegensatz zur beobachteten Hypomethylierung vieler Kandidatengene in UC zeigten Patienten mit akuten Darmentzündungen keine solchen Veränderungen. Um alternative epigenetische Modifikationen zu prüfen, wurde in diesen Patienten die Hydroxymethylierung quantifiziert. Die Ergebnisse hieraus sprechen für eine mögliche Rolle der Hydroxymethylierung bei akuter Entzündung im Vergleich zur chronischen Entzündung. Zusammenfassend stellt die vorgestellte Arbeit die erste replizierte EWAS in entzündetem Gewebe von UC Patienten dar, die mit dem Transkriptom verknüpft wurde. Die Ergebnisse deuten auf eine mögliche Rolle der epigenetischen Modifikationen bei Manifestation und Progression von UC hin.

7 References

- Abraham, C., and Cho, J. (2009a). Interleukin-23/Th17 pathways and inflammatory bowel disease. *Inflamm. Bowel Dis.* *15*, 1090–1100.
- Abraham, C., and Cho, J.H. (2009b). Inflammatory bowel disease. *N. Engl. J. Med.* *361*, 2066–2078.
- Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* *9*, 1191–1200.
- Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* *446*, 572–576.
- Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M., and Gerstein, M.B. (2010). Annotating non-coding regions of the genome. *Nat. Rev. Genet* *11*, 559–571.
- Anderson, C.A., Boucher, G., Lees, C.W., Franke, A., D’Amato, M., Taylor, K.D., Lee, J.C., Goyette, P., Imielinski, M., Latiano, A., et al., (2011). Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet.*
- Aoki, A., Suetake, I., Miyagawa, J., Fujio, T., Chijiwa, T., Sasaki, H., and Tajima, S. (2001). Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. *Nucleic Acids Res* *29*, 3506–3512.
- Atarashi, K., and Honda, K. (2011). Microbiota in autoimmunity and tolerance. *Curr. Opin. Immunol.* *23*, 761–768.
- Bach, J.-F. (2002). The effect of infections on susceptibility to autoimmune and allergic diseases. *N. Engl. J. Med.* *347*, 911–920.
- Balasa, A., Gathungu, G., Kisfali, P., Smith, E.O., Cho, J.H., Melegh, B., and Kellermayer, R. (2010). Assessment of DNA methylation at the interferon regulatory factor 5 (IRF5) promoter region in inflammatory bowel diseases. *Int J Colorectal Dis* *25*, 553–556.
- Barbaro, A., Cormaci, P., and Barbaro, A. (2004). DNA analysis from mixed biological materials. *Forensic Sci. Int.* *146 Suppl*, S123–125.
- Barnes, B.J., Moore, P.A., and Pitha, P.M. (2001). Virus-specific activation of a novel interferon regulatory factor, IRF-5, results in the induction of distinct interferon alpha genes. *J. Biol. Chem.* *276*, 23382–23390.
- Barreiro-de Acosta, M., Alvarez Castro, A., Souto, R., Iglesias, M., Lorenzo, A., and Dominguez-Muñoz, J.E. (2011). Emigration to western industrialized countries: A risk factor for developing inflammatory bowel disease. *J Crohns Colitis* *5*, 566–569.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* *129*, 823–837.
- Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* *143*, 470–484.

- Beher, D., Wu, J., Cumine, S., Kim, K.W., Lu, S.-C., Atangan, L., and Wang, M. (2009). Resveratrol is not a direct activator of SIRT1 enzyme activity. *Chem Biol Drug Des* 74, 619–624.
- Benchimol, E.I., Fortinsky, K.J., Gozdyra, P., Van den Heuvel, M., Van Limbergen, J., and Griffiths, A.M. (2011). Epidemiology of pediatric inflammatory bowel disease: a systematic review of international trends. *Inflamm. Bowel Dis.* 17, 423–439.
- Berger, M.F., Lawrence, M.S., Demichelis, F., Drier, Y., Cibulskis, K., Sivachenko, A.Y., Sboner, A., Esgueva, R., Pflueger, D., Sougnez, C., et al., (2011). The genomic complexity of primary human prostate cancer. *Nature* 470, 214–220.
- Bernstein, E., and Allis, C.D. (2005). RNA meets chromatin. *Genes Dev* 19, 1635–1655.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al., (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326.
- Bernstein, C.N., Fried, M., Krabshuis, J.H., Cohen, H., Eliakim, R., Fedail, S., Gearry, R., Goh, K.L., Hamid, S., Khan, A.G., et al., (2010). World Gastroenterology Organization Practice Guidelines for the diagnosis and management of IBD in 2010. *Inflamm. Bowel Dis* 16, 112–124.
- Bertin, J., Guo, Y., Wang, L., Srinivasula, S.M., Jacobson, M.D., Poyet, J.L., Merriam, S., Du, M.Q., Dyer, M.J., Robison, K.E., et al., (2000). CARD9 is a novel caspase recruitment domain-containing protein that interacts with BCL10/CLAP and activates NF-kappa B. *J. Biol. Chem.* 275, 41082–41086.
- Bettini, M.L., Pan, F., Bettini, M., Finkelstein, D., Rehg, J.E., Floess, S., Bell, B.D., Ziegler, S.F., Huehn, J., Pardoll, D.M., et al., (2012). Loss of Epigenetic Modification Driven by the Foxp3 Transcription Factor Leads to Regulatory T Cell Insufficiency. *Immunity*.
- Bhutani, N., Burns, D.M., and Blau, H.M. (2011). DNA demethylation dynamics. *Cell* 146, 866–872.
- Bi, L., Gojestani, S., Wu, W., Hsu, Y.-M.S., Zhu, J., Ariizumi, K., and Lin, X. (2010). CARD9 mediates dectin-2-induced I kappa B alpha kinase ubiquitination leading to activation of NF-kappa B in response to stimulation by the hyphal form of *Candida albicans*. *J. Biol. Chem.* 285, 25969–25977.
- Bialik, S., and Kimchi, A. (2010). Lethal weapons: DAP-kinase, autophagy and cell death: DAP-kinase regulates autophagy. *Curr. Opin. Cell Biol.* 22, 199–205.
- Bikoff, E.K., Morgan, M.A., and Robertson, E.J. (2009). An expanding job description for Blimp-1/PRDM1. *Curr. Opin. Genet. Dev.* 19, 379–385.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev* 16, 6–21.
- Bird, A.P., Taggart, M.H., Nicholls, R.D., and Higgs, D.R. (1987). Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene. *EMBO J.* 6, 999–1004.
- Bonfanti, P., Claudinot, S., Amici, A.W., Farley, A., Blackburn, C.C., and Barrandon, Y. (2010). Microenvironmental reprogramming of thymic epithelial cells to skin multipotent stem cells. *Nature* 466, 978–982.
- Boomsma, D., Busjahn, A., and Peltonen, L. (2002). Classical twin studies and beyond. *Nat. Rev. Genet* 3, 872–882.

- Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336, 934–937.
- Bouchard, T.J., Jr, and McGue, M. (2003). Genetic and environmental influences on human psychological differences. *J. Neurobiol.* 54, 4–45.
- Bouchard, T.J., Jr, McGue, M., Lykken, D., and Tellegen, A. (1999). Intrinsic and extrinsic religiousness: genetic and environmental influences and personality correlates. *Twin Res* 2, 88–98.
- Branco, M.R., Ficz, G., and Reik, W. (2011). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nature Reviews. Genetics*.
- Brant, S.R., and Shugart, Y.Y. (2004). Inflammatory bowel disease gene hunting by linkage analysis: rationale, methodology, and present status of the field. *Inflamm. Bowel Dis.* 10, 300–311.
- Van den Bree, M.B., Eaves, L.J., and Dwyer, J.T. (1999). Genetic and environmental influences on eating patterns of twins aged ≥ 50 y. *Am. J. Clin. Nutr.* 70, 456–465.
- Breese, E., Braegger, C.P., Corrigan, C.J., Walker-Smith, J.A., and MacDonald, T.T. (1993). Interleukin-2- and interferon-gamma-secreting T cells in normal and diseased human intestinal mucosa. *Immunology* 78, 127–131.
- Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., et al., (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*.
- Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., et al., (2002). Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15524–15529.
- Calin, G.A., Liu, C., Ferracin, M., Hyslop, T., Spizzo, R., Sevignani, C., Fabbri, M., Cimmino, A., Lee, E.J., Wojcik, S.E., et al., (2007). Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* 12, 215–229.
- Calkins, B.M. (1989). A meta-analysis of the role of smoking in inflammatory bowel disease. *Dig. Dis. Sci.* 34, 1841–1854.
- Cameron, E.E., Bachman, K.E., Myöhänen, S., Herman, J.G., and Baylin, S.B. (1999). Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nat. Genet* 21, 103–107.
- Carone, B.R., Fauquier, L., Habib, N., Shea, J.M., Hart, C.E., Li, R., Bock, C., Li, C., Gu, H., Zamore, P.D., et al., (2010). Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* 143, 1084–1096.
- Cattin, A.-L., Le Beyec, J., Barreau, F., Saint-Just, S., Houllier, A., Gonzalez, F.J., Robine, S., Pinçon-Raymond, M., Cardot, P., Lacasa, M., et al., (2009). Hepatocyte nuclear factor 4alpha, a key factor for homeostasis, cell architecture, and barrier function of the adult intestinal epithelium. *Mol. Cell. Biol.* 29, 6294–6308.
- Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* 10, 295–304.

- Chi, P., Chen, Y., Zhang, L., Guo, X., Wongvipat, J., Shamu, T., Fletcher, J.A., Dewell, S., Maki, R.G., Zheng, D., et al., (2010). ETV1 is a lineage survival factor that cooperates with KIT in gastrointestinal stromal tumours. *Nature* 467, 849–853.
- Cho, I., and Blaser, M.J. (2012). The human microbiome: at the interface of health and disease. *Nature Reviews. Genetics*.
- Clark, S.J., Harrison, J., and Molloy, P.L. (1997). Sp1 binding is inhibited by (m)Cp(m)CpG methylation. *Gene* 195, 67–71.
- Conticello, S.G., Langlois, M.-A., Yang, Z., and Neuberger, M.S. (2007). DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv. Immunol.* 94, 37–73.
- Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., Macdougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., et al., (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature*.
- Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al., (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 146, 67–79.
- Cosnes, J., Nion-Larmurier, I., Afchain, P., Beaugerie, L., and Gendre, J.-P. (2004). Gender differences in the response of colitis to smoking. *Clin. Gastroenterol. Hepatol.* 2, 41–48.
- Costello, C.M., Mah, N., Häslér, 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.
- Cowper-Sal Lari, R., Zhang, X., Wright, J.B., Bailey, S.D., Cole, M.D., Eeckhoutte, J., Moore, J.H., and Lupien, M. (2012). Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat. Genet.*
- Cullen, S., and Chapman, R. (2003). Primary sclerosing cholangitis. *Autoimmun Rev* 2, 305–312.
- Dale, M., and Nicklin, M.J. (1999). Interleukin-1 receptor cluster: gene organization of IL1R2, IL1R1, IL1RL2 (IL-1Rrp2), IL1RL1 (T1/ST2), and IL18R1 (IL-1Rrp) on human chromosome 2q. *Genomics* 57, 177–179.
- Danese, S., and Fiocchi, C. (2011). Ulcerative colitis. *N. Engl. J. Med.* 365, 1713–1725.
- Dashwood, R.H., and Ho, E. (2007). Dietary histone deacetylase inhibitors: from cells to mice to man. *Semin. Cancer Biol.* 17, 363–369.
- Davalos, V., Moutinho, C., Villanueva, A., Boque, R., Silva, P., Carneiro, F., and Esteller, M. (2011). Dynamic epigenetic regulation of the microRNA-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis. *Oncogene*.
- Dhir, M., Montgomery, E.A., Glöckner, S.C., Schuebel, K.E., Hooker, C.M., Herman, J.G., Baylin, S.B., Gearhart, S.L., and Ahuja, N. (2008). Epigenetic regulation of WNT signaling pathway genes in inflammatory bowel disease (IBD) associated neoplasia. *J. Gastrointest. Surg.* 12, 1745–1753.

- 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.
- Doerge, C.A., Inoue, K., Yamashita, T., Rhee, D.B., Travis, S., Fujita, R., Guarnieri, P., Bhagat, G., Vanti, W.B., Shih, A., et al., (2012). Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature*.
- Doi, A., Park, I.-H., Wen, B., Murakami, P., Aryee, M.J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., et al., (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet* 41, 1350–1353.
- Down, T.A., Rakyanc, V.K., Turner, D.J., Flicek, P., Li, H., Kulesha, E., Gräf, S., Johnson, N., Herrero, J., Tomazou, E.M., et al., (2008). A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol* 26, 779–785.
- Duerr, R.H. (2007). Genome-wide association studies herald a new era of rapid discoveries in inflammatory bowel disease research. *Gastroenterology* 132, 2045–2049.
- Duncan, H.E., and Edberg, S.C. (1995). Host-microbe interaction in the gastrointestinal tract. *Crit. Rev. Microbiol.* 21, 85–100.
- Duvallet, E., Semerano, L., Assier, E., Falgarone, G., and Boissier, M.-C. (2011). Interleukin-23: a key cytokine in inflammatory diseases. *Ann. Med.* 43, 503–511.
- Eaves, L., Heath, A., Martin, N., Maes, H., Neale, M., Kendler, K., Kirk, K., and Corey, L. (1999). Comparing the biological and cultural inheritance of personality and social attitudes in the Virginia 30,000 study of twins and their relatives. *Twin Res* 2, 62–80.
- Edwards, R.A., Witherspoon, M., Wang, K., Afrasiabi, K., Pham, T., Birnbaumer, L., and Lipkin, S.M. (2009). Epigenetic repression of DNA mismatch repair by inflammation and hypoxia in inflammatory bowel disease-associated colorectal cancer. *Cancer Res.* 69, 6423–6429.
- Ehrlich, M. (2009). DNA hypomethylation in cancer cells. *Epigenomics* 1, 239–259.
- Feil, R., and Fraga, M.F. (2011). Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* 13, 97–109.
- Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*.
- Franke, A., McGovern, D.P.B., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R., et al., (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn’s disease susceptibility loci. *Nat. Genet* 42, 1118–1125.
- Furuyama, T., Dalal, Y., and Henikoff, S. (2006). Chaperone-mediated assembly of centromeric chromatin in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6172–6177.
- Fuss, I.J., Heller, F., Boirivant, M., Leon, F., Yoshida, M., Fichtner-Feigl, S., Yang, Z., Exley, M., Kitani, A., Blumberg, R.S., et al., (2004). Nonclassical CD1d-restricted NK T cells that produce IL-13 characterize an atypical Th2 response in ulcerative colitis. *J. Clin. Invest.* 113, 1490–1497.

- Gal-Yam, E.N., Egger, G., Iniguez, L., Holster, H., Einarsson, S., Zhang, X., Lin, J.C., Liang, G., Jones, P.A., and Tanay, A. (2008). Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proc. Natl. Acad. Sci. U.S.A.* *105*, 12979–12984.
- Gehring, M., Reik, W., and Henikoff, S. (2009). DNA demethylation by DNA repair. *Trends Genet.* *25*, 82–90.
- Geng, X., Biancone, L., Dai, H.H., Lin, J.J., Yoshizaki, N., Dasgupta, A., Pallone, F., and Das, K.M. (1998). Tropomyosin isoforms in intestinal mucosa: production of autoantibodies to tropomyosin isoforms in ulcerative colitis. *Gastroenterology* *114*, 912–922.
- Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* *5*, e15367.
- Glocker, E.-O., Hennigs, A., Nabavi, M., Schäffer, A.A., Woellner, C., Salzer, U., Pfeifer, D., Veelken, H., Warnatz, K., Tahami, F., et al., (2009). A homozygous CARD9 mutation in a family with susceptibility to fungal infections. *N. Engl. J. Med.* *361*, 1727–1735.
- Glória, L., Cravo, M., Pinto, A., de Sousa, L.S., Chaves, P., Leitão, C.N., Quina, M., Mira, F.C., and Soares, J. (1996). DNA hypomethylation and proliferative activity are increased in the rectal mucosa of patients with long-standing ulcerative colitis. *Cancer* *78*, 2300–2306.
- Gonsky, R., Deem, R.L., Landers, C.J., Derkowski, C.A., Berel, D., McGovern, D.P.B., and Targan, S.R. (2011). Distinct IFNG methylation in a subset of ulcerative colitis patients based on reactivity to microbial antigens. *Inflamm. Bowel Dis.* *17*, 171–178.
- Greenman, C., Stephens, P., Smith, R., Dalgliesh, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., et al., (2007). Patterns of somatic mutation in human cancer genomes. *Nature* *446*, 153–158.
- Gross, V., Andus, T., Leser, H.G., Roth, M., and Schölmerich, J. (1991). Inflammatory mediators in chronic inflammatory bowel diseases. *Klin. Wochenschr.* *69*, 981–987.
- Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S.-G., et al., (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*.
- Guelen, L., Pagie, L., Brassat, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al., (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* *453*, 948–951.
- Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* *315*, 1587–1590.
- Guo, J.U., Su, Y., Zhong, C., Ming, G.-L., and Song, H. (2011). Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. *Cell*.
- Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J.L., et al., (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* *464*, 1071–1076.

- He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al., (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307.
- Hebebrand, J., Sommerlad, C., Geller, F., Görg, T., and Hinney, A. (2001). The genetics of obesity: practical implications. *Int. J. Obes. Relat. Metab. Disord.* **25 Suppl 1**, S10–18.
- Heller, F., Florian, P., Bojarski, C., Richter, J., Christ, M., Hillenbrand, B., Mankertz, J., Gitter, A.H., Bürgel, N., Fromm, M., et al., (2005). Interleukin-13 is the key effector Th2 cytokine in ulcerative colitis that affects epithelial tight junctions, apoptosis, and cell restitution. *Gastroenterology* **129**, 550–564.
- Hemminki, K., Lorenzo Bermejo, J., and Försti, A. (2006). The balance between heritable and environmental aetiology of human disease. *Nat. Rev. Genet.* **7**, 958–965.
- Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat. Rev. Genet* **9**, 15–26.
- Heyn, H., and Esteller, M. (2012). DNA methylation profiling in the clinic: applications and challenges. *Nat. Rev. Genet.*
- Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. *Science* **187**, 226–232.
- Hsieh, T.-F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., and Zilberman, D. (2009). Genome-wide demethylation of Arabidopsis endosperm. *Science* **324**, 1451–1454.
- Ieda, M., Fu, J.-D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386.
- Imai, S., Armstrong, C.M., Kaeberlein, M., and Guarente, L. (2000). Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–800.
- Iqbal, K., Jin, S.-G., Pfeifer, G.P., and Szabó, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl. Acad. Sci. U.S.A* **108**, 3642–3647.
- Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al., (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet* **41**, 178–186.
- Isaacs, K.L., Sartor, R.B., and Haskill, S. (1992). Cytokine messenger RNA profiles in inflammatory bowel disease mucosa detected by polymerase chain reaction amplification. *Gastroenterology* **103**, 1587–1595.
- Ito, S., D’Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303.

- Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet* *33 Suppl*, 245–254.
- Javierre, B.M., Fernandez, A.F., Richter, J., Al-Shahrour, F., Martin-Subero, J.I., Rodriguez-Ubreva, J., Berdasco, M., Fraga, M.F., O’Hanlon, T.P., Rider, L.G., et al., (2010). Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res.* *20*, 170–179.
- Jin, S.-G., Kadam, S., and Pfeifer, G.P. (2010). Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res* *38*, e125.
- Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews. Genetics*.
- Jones, S., Wang, T.-L., Shih, I.-M., Mao, T.-L., Nakayama, K., Roden, R., Glas, R., Slamon, D., Diaz, L.A., Jr, Vogelstein, B., et al., (2010). Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* *330*, 228–231.
- Jullien, J., Pasque, V., Halley-Stott, R.P., Miyamoto, K., and Gurdon, J.B. (2011). Mechanisms of nuclear reprogramming by eggs and oocytes: a deterministic process? *Nat. Rev. Mol. Cell Biol.* *12*, 453–459.
- Juriscic, G., Iolyeva, M., Proulx, S.T., Halin, C., and Detmar, M. (2010). Thymus cell antigen 1 (Thy1, CD90) is expressed by lymphatic vessels and mediates cell adhesion to lymphatic endothelium. *Exp. Cell Res.* *316*, 2982–2992.
- Kadomatsu, K. (2005). The midkine family in cancer, inflammation and neural development. *Nagoya J Med Sci* *67*, 71–82.
- Kaeberlein, M., McDonagh, T., Heltweg, B., Hixon, J., Westman, E.A., Caldwell, S.D., Napper, A., Curtis, R., DiStefano, P.S., Fields, S., et al., (2005). Substrate-specific activation of sirtuins by resveratrol. *J. Biol. Chem.* *280*, 17038–17045.
- Kaiser, P., Rothwell, L., Galyov, E.E., Barrow, P.A., Burnside, J., and Wigley, P. (2000). Differential cytokine expression in avian cells in response to invasion by *Salmonella typhimurium*, *Salmonella enteritidis* and *Salmonella gallinarum*. *Microbiology (Reading, Engl.)* *146 Pt 12*, 3217–3226.
- Kaminsky, Z.A., Tang, T., Wang, S.-C., Ptak, C., Oh, G.H.T., Wong, A.H.C., Feldcamp, L.A., Virtanen, C., Halfvarson, J., Tysk, C., et al., (2009). DNA methylation profiles in monozygotic and dizygotic twins. *Nat. Genet* *41*, 240–245.
- Kaser, A., Lee, A.-H., Franke, A., Glickman, J.N., Zeissig, S., Tilg, H., Nieuwenhuis, E.E.S., Higgins, D.E., Schreiber, S., Glimcher, L.H., et al., (2008). XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell* *134*, 743–756.
- Kaser, A., Zeissig, S., and Blumberg, R.S. (2010). Inflammatory bowel disease. *Annu. Rev. Immunol.* *28*, 573–621.
- Kemmerich, K., Dingler, F.A., Rada, C., and Neuberger, M.S. (2012). Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of *Ung*^{-/-}*Msh2*^{-/-} mice. *Nucleic Acids Res.* *40*, 6016–6025.

- Kendler, K.S., and Baker, J.H. (2007). Genetic influences on measures of the environment: a systematic review. *Psychol Med* 37, 615–626.
- Kendler, K.S., and Karkowski-Shuman, L. (1997). Stressful life events and genetic liability to major depression: genetic control of exposure to the environment? *Psychol Med* 27, 539–547.
- Kendler, K.S., and Prescott, C.A. (1999). A population-based twin study of lifetime major depression in men and women. *Arch. Gen. Psychiatry* 56, 39–44.
- Kendler, K.S., Thornton, L.M., and Pedersen, N.L. (2000). Tobacco consumption in Swedish twins reared apart and reared together. *Arch. Gen. Psychiatry* 57, 886–892.
- Khare, T., Pai, S., Koncivicius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., Irimia, M., Jia, P., Ptak, C., Xia, M., et al., (2012). 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat. Struct. Mol. Biol.*
- Klein, C.J., Botuyan, M.-V., Wu, Y., Ward, C.J., Nicholson, G.A., Hammans, S., Hojo, K., Yamanishi, H., Karpf, A.R., Wallace, D.C., et al., (2011). Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nat. Genet* 43, 595–600.
- Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci* 31, 89–97.
- Konishi, K., Shen, L., Wang, S., Meltzer, S.J., Harpaz, N., and Issa, J.-P.J. (2007). Rare CpG island methylator phenotype in ulcerative colitis-associated neoplasias. *Gastroenterology* 132, 1254–1260.
- Koopmans, J.R., Slutske, W.S., Heath, A.C., Neale, M.C., and Boomsma, D.I. (1999). The genetics of smoking initiation and quantity smoked in Dutch adolescent and young adult twins. *Behav. Genet.* 29, 383–393.
- Kota, S.K., and Feil, R. (2010). Epigenetic transitions in germ cell development and meiosis. *Dev. Cell* 19, 675–686.
- Krokan, H.E., Drabløs, F., and Slupphaug, G. (2002). Uracil in DNA--occurrence, consequences and repair. *Oncogene* 21, 8935–8948.
- Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319, 1827–1830.
- Kumar, S.V., and Wigge, P.A. (2010). H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. *Cell* 140, 136–147.
- Kuratomi, G., Iwamoto, K., Bundo, M., Kusumi, I., Kato, N., Iwata, N., Ozaki, N., and Kato, T. (2008). Aberrant DNA methylation associated with bipolar disorder identified from discordant monozygotic twins. *Mol. Psychiatry* 13, 429–441.
- Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220.
- 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.

- Lee, J.T. (2009). Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev* 23, 1831–1842.
- Lee, J., and Gross, J.M. (2007). Laminin beta1 and gamma1 containing laminins are essential for basement membrane integrity in the zebrafish eye. *Invest. Ophthalmol. Vis. Sci.* 48, 2483–2490.
- Lee, M.S., and Kim, Y.-J. (2007). Signaling pathways downstream of pattern-recognition receptors and their cross talk. *Annu. Rev. Biochem.* 76, 447–480.
- Lee, J.H., Schütte, D., Wulf, G., Füzesi, L., Radzun, H.-J., Schweyer, S., Engel, W., and Nayernia, K. (2006). Stem-cell protein Piwil2 is widely expressed in tumors and inhibits apoptosis through activation of Stat3/Bcl-XL pathway. *Hum. Mol. Genet.* 15, 201–211.
- Lian, C.G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q., et al., (2012). Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell* 150, 1135–1146.
- Linhart, H.G., Troen, A., Bell, G.W., Cantu, E., Chao, W.-H., Moran, E., Steine, E., He, T., and Jaenisch, R. (2009). Folate deficiency induces genomic uracil misincorporation and hypomethylation but does not increase DNA point mutations. *Gastroenterology* 136, 227–235.e3.
- Liu, M., and Schatz, D.G. (2009). Balancing AID and DNA repair during somatic hypermutation. *Trends Immunol.* 30, 173–181.
- Liu, J.J., Shen, R., Chen, L., Ye, Y., He, G., Hua, K., Jarjoura, D., Nakano, T., Ramesh, G.K., Shapiro, C.L., et al., (2010). Piwil2 is expressed in various stages of breast cancers and has the potential to be used as a novel biomarker. *Int J Clin Exp Pathol* 3, 328–337.
- Liu, L., Guo, X., Rao, J.N., Zou, T., Xiao, L., Yu, T., Timmons, J.A., Turner, D.J., and Wang, J.-Y. (2009). Polyamines regulate E-cadherin transcription through c-Myc modulating intestinal epithelial barrier function. *Am. J. Physiol., Cell Physiol.* 296, C801–810.
- Lock, L.F., Takagi, N., and Martin, G.R. (1987). Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell* 48, 39–46.
- Loenarz, C., and Schofield, C.J. (2011). Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem. Sci.* 36, 7–18.
- Lyons, M.J. (1996). A twin study of self-reported criminal behaviour. *Ciba Found. Symp.* 194, 61–70; discussion 70–75.
- Ma, D.K., Jang, M.-H., Guo, J.U., Kitabatake, Y., Chang, M.-L., Pow-Anpongkul, N., Flavell, R.A., Lu, B., Ming, G.-L., and Song, H. (2009). Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science* 323, 1074–1077.
- MacGregor, A.J., Snieder, H., Rigby, A.S., Koskenvuo, M., Kaprio, J., Aho, K., and Silman, A.J. (2000). Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.* 43, 30–37.
- Mahid, S.S., Minor, K.S., Soto, R.E., Hornung, C.A., and Galandiuk, S. (2006). Smoking and inflammatory bowel disease: a meta-analysis. *Mayo Clin. Proc.* 81, 1462–1471.

- Mahid, S.S., Minor, K.S., Stromberg, A.J., and Galandiuk, S. (2007). Active and passive smoking in childhood is related to the development of inflammatory bowel disease. *Inflamm. Bowel Dis.* *13*, 431–438.
- Mahida, Y.R., and Rolfe, V.E. (2004). Host-bacterial interactions in inflammatory bowel disease. *Clin. Sci.* *107*, 331–341.
- March, H.N., Rust, A.G., Wright, N.A., Ten Hoeve, J., de Ridder, J., Eldridge, M., van der Weyden, L., Berns, A., Gadiot, J., Uren, A., et al., (2011). Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. *Nat. Genet.* *43*, 1202–1209.
- Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. *Nat. Rev. Genet.* *11*, 285–296.
- Mattick, J.S., and Makunin, I.V. (2006). Non-coding RNA. *Hum. Mol. Genet.* *15 Spec No 1*, R17–29.
- Maul, R.W., and Gearhart, P.J. (2010). AID and somatic hypermutation. *Adv. Immunol.* *105*, 159–191.
- Medvedeva, Y.A., Fridman, M.V., Oparina, N.J., Malko, D.B., Ermakova, E.O., Kulakovskiy, I.V., Heinzl, A., and Makeev, V.J. (2010). Intergenic, gene terminal, and intragenic CpG islands in the human genome. *BMC Genomics* *11*, 48.
- Meneghini, M.D., Wu, M., and Madhani, H.D. (2003). Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* *112*, 725–736.
- Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* *10*, 155–159.
- Merregaert, J., Van Langen, J., Hansen, U., Ponsaerts, P., El Ghalbzouri, A., Steenackers, E., Van Ostade, X., and Sercu, S. (2010). Phospholipid scramblase 1 is secreted by a lipid raft-dependent pathway and interacts with the extracellular matrix protein 1 in the dermal epidermal junction zone of human skin. *J. Biol. Chem.* *285*, 37823–37837.
- Mill, J., Dempster, E., Caspi, A., Williams, B., Moffitt, T., and Craig, I. (2006). Evidence for monozygotic twin (MZ) discordance in methylation level at two CpG sites in the promoter region of the catechol-O-methyltransferase (COMT) gene. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* *141B*, 421–425.
- Molodecky, N.A., Soon, I.S., Rabi, D.M., Ghali, W.A., Ferris, M., Chernoff, G., Benchimol, E.I., Panaccione, R., Ghosh, S., Barkema, H.W., et al., (2012). Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* *142*, 46–54.e42; quiz e30.
- Monk, M., Adams, R.L., and Rinaldi, A. (1991). Decrease in DNA methylase activity during preimplantation development in the mouse. *Development* *112*, 189–192.
- Moréra, S., Grin, I., Vigouroux, A., Couvé, S., Henriot, V., Saparbaev, M., and Ishchenko, A.A. (2012). Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyluracil-containing DNA. *Nucleic Acids Research*.
- Moriyama, T., Matsumoto, T., Nakamura, S., Jo, Y., Mibu, R., Yao, T., and Iida, M. (2007). Hypermethylation of p14 (ARF) may be predictive of colitic cancer in patients with ulcerative colitis. *Dis. Colon Rectum* *50*, 1384–1392.

- Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* *102*, 553–563.
- Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* *8*, 751–758.
- Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* *322*, 1717–1720.
- Navarro, P., Page, D.R., Avner, P., and Rougeulle, C. (2006). Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program. *Genes Dev.* *20*, 2787–2792.
- Ng, S.C., Tsoi, K.K.F., Kamm, M.A., Xia, B., Wu, J., Chan, F.K.L., and Sung, J.J.Y. (2011). Genetics of inflammatory bowel disease in Asia: Systematic review and meta-analysis. *Inflamm. Bowel Dis.*
- Nguyen, A., Rauch, T.A., Pfeifer, G.P., and Hu, V.W. (2010). Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB J.* *24*, 3036–3051.
- Nikolaus, S., and Schreiber, S. (2007). Diagnostics of inflammatory bowel disease. *Gastroenterology* *133*, 1670–1689.
- O’Sullivan, M., and O’Morain, C. (2006). Nutrition in inflammatory bowel disease. *Best Pract Res Clin Gastroenterol* *20*, 561–573.
- Oates, N.A., van Vliet, J., Duffy, D.L., Kroes, H.Y., Martin, N.G., Boomsma, D.I., Campbell, M., Coulthard, M.G., Whitelaw, E., and Chong, S. (2006). Increased DNA methylation at the AXIN1 gene in a monozygotic twin from a pair discordant for a caudal duplication anomaly. *Am. J. Hum. Genet.* *79*, 155–162.
- Ohm, J.E., McGarvey, K.M., Yu, X., Cheng, L., Schuebel, K.E., Cope, L., Mohammad, H.P., Chen, W., Daniel, V.C., Yu, W., et al., (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat. Genet.* *39*, 237–242.
- Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., and Hayashi, Y. (2002a). LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res.* *62*, 4075–4080.
- Ono, S., Umezaki, M., Nabari, H., Nakayama, R., Hasegawa, K., Sako, S., Fujii, T., Yamazaki, I., and Yoshimura, T. (2002b). Selective removal of chymotrypsin using diphenyl alpha-aminoalkylphosphonate immobilized on sepharose gel. *Biosci. Biotechnol. Biochem.* *66*, 1111–1115.
- Panja, A., Siden, E., and Mayer, L. (1995). Synthesis and regulation of accessory/proinflammatory cytokines by intestinal epithelial cells. *Clin. Exp. Immunol.* *100*, 298–305.
- Panja, A., Goldberg, S., Eckmann, L., Krishen, P., and Mayer, L. (1998). The regulation and functional consequence of proinflammatory cytokine binding on human intestinal epithelial cells. *J. Immunol.* *161*, 3675–3684.

- Paroush, Z., Keshet, I., Yisraeli, J., and Cedar, H. (1990). Dynamics of demethylation and activation of the alpha-actin gene in myoblasts. *Cell* 63, 1229–1237.
- Parsons, D.W., Li, M., Zhang, X., Jones, S., Leary, R.J., Lin, J.C.-H., Boca, S.M., Carter, H., Samayoa, J., Bettegowda, C., et al., (2011). The genetic landscape of the childhood cancer medulloblastoma. *Science* 331, 435–439.
- Pasquinelli, A.E. (2012). MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nature Reviews. Genetics*.
- Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P., et al., (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*.
- Paulsen, M., and Ferguson-Smith, A.C. (2001). DNA methylation in genomic imprinting, development, and disease. *J. Pathol* 195, 97–110.
- Peng, L., Yuan, Z., Ling, H., Fukasawa, K., Robertson, K., Olashaw, N., Koomen, J., Chen, J., Lane, W.S., and Seto, E. (2011). SIRT1 deacetylates the DNA methyltransferase 1 (DNMT1) protein and alters its activities. *Mol. Cell. Biol.* 31, 4720–4734.
- Petronis, A. (2010). Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* 465, 721–727.
- Petronis, A., Gottesman, I.I., Kan, P., Kennedy, J.L., Basile, V.S., Paterson, A.D., and Pependikyte, V. (2003). Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? *Schizophr Bull* 29, 169–178.
- Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463, 1101–1105.
- Ptashne, M. (2007). On the use of the word “epigenetic.” *Curr. Biol.* 17, R233–236.
- Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* 135, 1201–1212.
- Rai, K., Sarkar, S., Broadbent, T.J., Voas, M., Grossmann, K.F., Nadauld, L.D., Dehghanizadeh, S., Hagos, F.T., Li, Y., Toth, R.K., et al., (2010). DNA Demethylase Activity Maintains Intestinal Cells in an Undifferentiated State Following Loss of APC. *Cell* 142, 930–942.
- Rakyan, V.K., Down, T.A., Thorne, N.P., Flicek, P., Kulesha, E., Gräf, S., Tomazou, E.M., Bäckdahl, L., Johnson, N., Herberth, M., et al., (2008). An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res* 18, 1518–1529.
- Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447, 425–432.
- Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* 14, 9–25.

- Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Bruggmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al., (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* *129*, 1311–1323.
- Roberts, V.A., Pique, M.E., Hsu, S., Li, S., Slupphaug, G., Rambo, R.P., Jamison, J.W., Liu, T., Lee, J.H., Tainer, J.A., et al., (2012). Combining H/D exchange mass spectroscopy and computational docking reveals extended DNA-binding surface on uracil-DNA glycosylase. *Nucleic Acids Research* *40*, 6070–6081.
- Robertson, K.D., and Wolffe, A.P. (2000). DNA methylation in health and disease. *Nat. Rev. Genet* *1*, 11–19.
- Rook, G.A.W. (2012). Hygiene hypothesis and autoimmune diseases. *Clin Rev Allergy Immunol* *42*, 5–15.
- Rougier, N., Bourc'his, D., Gomes, D.M., Niveleau, A., Plachot, M., Pàldi, A., and Viegas-Péquignot, E. (1998). Chromosome methylation patterns during mammalian preimplantation development. *Genes Dev.* *12*, 2108–2113.
- Sakuraba, A., Sato, T., Kamada, N., Kitazume, M., Sugita, A., and Hibi, T. (2009). Th1/Th17 immune response is induced by mesenteric lymph node dendritic cells in Crohn's disease. *Gastroenterology* *137*, 1736–1745.
- Satsangi, J., Jewell, D.P., and Bell, J.I. (1997). The genetics of inflammatory bowel disease. *Gut* *40*, 572–574.
- Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., et al., (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat. Genet.* *39*, 232–236.
- Schones, D.E., and Zhao, K. (2008). Genome-wide approaches to studying chromatin modifications. *Nat. Rev. Genet.* *9*, 179–191.
- Schreiber, S., Rosenstiel, P., Albrecht, M., Hampe, J., and Krawczak, M. (2005). Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nat. Rev. Genet* *6*, 376–388.
- Shanahan, F. (2002). Gut flora in gastrointestinal disease. *Eur J Surg Suppl* 47–52.
- Sharma, A.K., Nelson, M.C., Brandt, J.E., Wessman, M., Mahmud, N., Weller, K.P., and Hoffman, R. (2001). Human CD34(+) stem cells express the hiwi gene, a human homologue of the *Drosophila* gene piwi. *Blood* *97*, 426–434.
- Shock, L.S., Thakkar, P.V., Peterson, E.J., Moran, R.G., and Taylor, S.M. (2011). DNA methyltransferase 1, cytosine methylation, and cytosine hydroxymethylation in mammalian mitochondria. *Proc. Natl. Acad. Sci. U.S.A.* *108*, 3630–3635.
- Silman, A.J., Newman, J., and MacGregor, A.J. (1996). Cigarette smoking increases the risk of rheumatoid arthritis. Results from a nationwide study of disease-discordant twins. *Arthritis Rheum.* *39*, 732–735.
- Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol* *10*, 697–708.

- Sjöblom, T., Jones, S., Wood, L.D., Parsons, D.W., Lin, J., Barber, T.D., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N., et al., (2006). The consensus coding sequences of human breast and colorectal cancers. *Science* 314, 268–274.
- Skjeldam, H.K., Kassahun, H., Fensgård, O., SenGupta, T., Babaie, E., Lindvall, J.M., Arczewska, K., and Nilsen, H. (2010). Loss of *Caenorhabditis elegans* UNG-1 uracil-DNA glycosylase affects apoptosis in response to DNA damaging agents. *DNA Repair (Amst.)* 9, 861–870.
- Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6, 31.
- Smallwood, S.A., Tomizawa, S.-I., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat. Genet.* 43, 811–814.
- Song, J., Teplova, M., Ishibe-Murakami, S., and Patel, D.J. (2012). Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science* 335, 709–712.
- Stevens, C., Walz, G., Singaram, C., Lipman, M.L., Zanker, B., Muggia, A., Antonioli, D., Peppercorn, M.A., and Strom, T.B. (1992). Tumor necrosis factor-alpha, interleukin-1 beta, and interleukin-6 expression in inflammatory bowel disease. *Dig. Dis. Sci.* 37, 818–826.
- Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat. Cell Biol.* 9, 1428–1435.
- Stransky, N., Egloff, A.M., Tward, A.D., Kostic, A.D., Cibulskis, K., Sivachenko, A., Kryukov, G.V., Lawrence, M.S., Sougnez, C., McKenna, A., et al., (2011). The mutational landscape of head and neck squamous cell carcinoma. *Science* 333, 1157–1160.
- Strasser, D., Neumann, K., Bergmann, H., Marakalala, M.J., Guler, R., Rojowska, A., Hopfner, K.-P., Brombacher, F., Urlaub, H., Baier, G., et al., (2012). Syk Kinase-Coupled C-type Lectin Receptors Engage Protein Kinase C- δ to Elicit Card9 Adaptor-Mediated Innate Immunity. *Immunity*.
- Sun, G., Wang, Y., Sun, L., Luo, H., Liu, N., Fu, Z., and You, Y. (2011). Clinical significance of Hiwi gene expression in gliomas. *Brain Res.* 1373, 183–188.
- Sur, I.K., Hallikas, O., Vähärautio, A., Yan, J., Turunen, M., Enge, M., Taipale, M., Karhu, A., Aaltonen, L.A., and Taipale, J. (2012). Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* 338, 1360–1363.
- Suto, R.K., Clarkson, M.J., Tremethick, D.J., and Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nat. Struct. Biol.* 7, 1121–1124.
- Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* 116, 51–61.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al., (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935.

- Takahashi, F., and Das, K.M. (1985). Isolation and characterization of a colonic autoantigen specifically recognized by colon tissue-bound immunoglobulin G from idiopathic ulcerative colitis. *J. Clin. Invest.* *76*, 311–318.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* *126*, 663–676.
- Tat, P.A., Sumer, H., Pralong, D., and Verma, P.J. (2011). The efficiency of cell fusion-based reprogramming is affected by the somatic cell type and the in vitro age of somatic cells. *Cell Reprogram* *13*, 331–344.
- Taubert, H., Greither, T., Kaushal, D., Würfl, P., Bache, M., Bartel, F., Kehlen, A., Lautenschläger, C., Harris, L., Kraemer, K., et al., (2007). Expression of the stem cell self-renewal gene *Hiwi* and risk of tumour-related death in patients with soft-tissue sarcoma. *Oncogene* *26*, 1098–1100.
- Taubes, G. (1995). Epidemiology faces its limits. *Science* *269*, 164–169.
- Teschendorff, A.E., Menon, U., Gentry-Maharaj, A., Ramus, S.J., Weisenberger, D.J., Shen, H., Campan, M., Noushmehr, H., Bell, C.G., Maxwell, A.P., et al., (2010). Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* *20*, 440–446.
- Urduingio, R.G., Sanchez-Mut, J.V., and Esteller, M. (2009). Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurol* *8*, 1056–1072.
- Vaquero, A., Scher, M., Lee, D., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2004). Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Mol. Cell* *16*, 93–105.
- Waite, J.C., and Skokos, D. (2012). Th17 response and inflammatory autoimmune diseases. *Int J Inflam* *2012*, 819467.
- Wang, Q.-E., Han, C., Milum, K., and Wani, A.A. (2011). Stem cell protein *Piwil2* modulates chromatin modifications upon cisplatin treatment. *Mutat. Res.* *708*, 59–68.
- Wang, Y.F., Zhang, H., and Ouyang, Q. (2007). Clinical manifestations of inflammatory bowel disease: East and West differences. *J Dig Dis* *8*, 121–127.
- Wapenaar, M.C., Monsuur, A.J., Poell, J., van 't Slot, R., Meijer, J.W.R., Meijer, G.A., Mulder, C.J., Mearin, M.L., and Wijmenga, C. (2007). The *SPINK* gene family and celiac disease susceptibility. *Immunogenetics* *59*, 349–357.
- Washietl, S., Hofacker, I.L., Lukasser, M., Hüttenhofer, A., and Stadler, P.F. (2005). Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat. Biotechnol.* *23*, 1383–1390.
- Watanabe, Y., and Maekawa, M. (2010). Methylation of DNA in cancer. *Adv Clin Chem* *52*, 145–167.
- Waters, T.R., and Swann, P.F. (2000). Thymine-DNA glycosylase and G to A transition mutations at CpG sites. *Mutat. Res.* *462*, 137–147.

- Weaver, I.C.G., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M., and Meaney, M.J. (2004). Epigenetic programming by maternal behavior. *Nat. Neurosci* 7, 847–854.
- Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I., et al., (2007). Epigenetic stem cell signature in cancer. *Nat. Genet.* 39, 157–158.
- Wigler, M., Levy, D., and Perucho, M. (1981). The somatic replication of DNA methylation. *Cell* 24, 33–40.
- Wild, L., and Flanagan, J.M. (2010). Genome-wide hypomethylation in cancer may be a passive consequence of transformation. *Biochim. Biophys. Acta* 1806, 50–57.
- Willett, W.C. (2002). Balancing life-style and genomics research for disease prevention. *Science* 296, 695–698.
- Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappsilber, J., and Helin, K. (2011a). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*.
- Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappsilber, J., and Helin, K. (2011b). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*.
- Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol* 11, 607–620.
- Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* 329, 444–448.
- Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Eve Sun, Y., and Zhang, Y. (2011a). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature*.
- Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Eve Sun, Y., and Zhang, Y. (2011b). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature*.
- Von Wurmb-Schwark, N., Ringleb, A., Gebühr, M., and Simeoni, E. (2005). Genetic analysis of modern and historical burned human remains. *Anthropol Anz* 63, 1–12.
- Wyatt, G.R., and Cohen, S.S. (1952). A new pyrimidine base from bacteriophage nucleic acids. *Nature* 170, 1072–1073.
- Xavier, R.J., and Podolsky, D.K. (2007). Unravelling the pathogenesis of inflammatory bowel disease. *Nature* 448, 427–434.
- Yamanaka, S., and Blau, H.M. (2010). Nuclear reprogramming to a pluripotent state by three approaches. *Nature* 465, 704–712.
- Yisraeli, J., Frank, D., Razin, A., and Cedar, H. (1988). Effect of in vitro DNA methylation on beta-globin gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 85, 4638–4642.

Yu, W., Hegarty, J.P., Berg, A., Chen, X., West, G., Kelly, A.A., Wang, Y., Poritz, L.S., Koltun, W.A., and Lin, Z. (2011). NKX2-3 transcriptional regulation of endothelin-1 and VEGF signaling in human intestinal microvascular endothelial cells. *PLoS ONE* 6, e20454.

Zeng, L., Zhang, Q., Li, S., Plotnikov, A.N., Walsh, M.J., and Zhou, M.-M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* 466, 258–262.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W.-L., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., et al., (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* 126, 1189–1201.

Zhao, Q., Rank, G., Tan, Y.T., Li, H., Moritz, R.L., Simpson, R.J., Cerruti, L., Curtis, D.J., Patel, D.J., Allis, C.D., et al., (2009). PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat. Struct. Mol. Biol* 16, 304–311.

Zhou, W., Alonso, S., Takai, D., Lu, S.C., Yamamoto, F., Perucho, M., and Huang, S. (2008). Requirement of RIZ1 for cancer prevention by methyl-balanced diet. *PLoS ONE* 3, e3390.

8 Appendix

Table S 1 Genome-wide significant expressional candidates (Layer I).

Experimental ID	Gene Symbol	Chromosome	Transcription Start Sites	P value
1554436_a_at	REG4	1	120138164	0.041144451
200662_s_at	TOMM20	1	233339282	0.018875861
200771_at	LAMC1	1	181259217	0.029076187
201969_at	NASP	1	45822303	0.035722791
202651_at	LPGAT1	1	209983422	0.006164888
202743_at	PIK3R3	1	46278398	0.035722791
202834_at	AGT	1	228904891	0.007933587
203073_at	COG2	1	228844824	0.018875861
203508_at	TNFRSF1B	1	12149646	0.049944792
203801_at	MRPS14	1	173249750	0.049944792
204607_at	HMGCS2	1	120092525	0.035722791
205309_at	SMPDL3B	1	28134090	0.029076187
205483_s_at	ISG15	1	938709	0.010231873
207826_s_at	ID3	1	23756995	0.049944792
209071_s_at	RGS5	1	161378720	0.011839038
209589_s_at	EPHB2	1	22909917	0.029076187
210651_s_at	EPHB2	1	22909917	0.015061861
212348_s_at	KDM1	1	23218527	0.010231873
	HIST2H2AA3	///		
214290_s_at	HIST2H2AA4	1	148089280	0.049944792
214433_s_at	SELENBP1	1	149603403	0.035722791
	MT1E /// MT1H ///			
216336_x_at	MT1M /// MT1P2	1	33189329	0.029076187
218168_s_at	CABC1	1	225194560	0.018875861
219134_at	ELTD1	1	79128036	0.041144451
219481_at	TTC13	1	229108611	0.023555392
221486_at	ENSA	1	148861222	0.015061861
221656_s_at	ARHGEF10L	1	17738916	0.010231873
221726_at	RPL22	1	6167666	0.023555392
223223_at	ARV1	1	229181445	0.049944792
223272_s_at	C1orf57	1	231152992	0.007933587
223339_at	ATPIF1	1	28435197	0.035722791
223402_at	DUSP23	1	158017382	0.015061861
223447_at	REG4	1	120138164	0.041144451
226020_s_at	DAB1 /// OMA1	1	58718725	0.041144451
226388_at	TCEA3	1	23580141	0.011839038
227582_at	KLHDC9	1	159334777	0.018875861
229254_at	MFSD4	1	203804734	0.035722791
232219_x_at	USP21	1	159395877	0.049944792

Experimental ID	Gene Symbol	Chromosome	Transcription Start Sites	P value
240110_at	HMGCS2	1	120092525	0.049944792
200887_s_at	STAT1	2	191542006	0.007933587
201302_at	ANXA4	2	69822630	0.007933587
201438_at	COL6A3	2	237897393	0.010231873
202274_at	ACTG2	2	73973600	0.035722791
202351_at	ITGAV	2	187163034	0.006164888
203697_at	FRZB	2	183406981	0.006164888
203698_s_at	FRZB	2	183406981	0.010231873
205403_at	IL1R2	2	101974737	0.029076187
209969_s_at	STAT1	2	191542006	0.018875861
211161_s_at	COL3A1	2	189547343	0.035722791
211372_s_at	IL1R2	2	101974737	0.018875861
211959_at	IGFBP5	2	217245072	0.041144451
212012_at	PXDN	2	1614665	0.001717581
212013_at	PXDN	2	1614665	0.011839038
212017_at	FAM168B	2	131521918	0.035722791
212119_at	RHOQ	2	46623370	0.049944792
	LOC732160	///		
217860_at	NDUFA10	2	240548830	0.015061861
219093_at	PID1	2	229596932	0.035722791
219735_s_at	TFCP2L1	2	121690635	0.035722791
221729_at	COL5A2	2	189604885	0.035722791
223297_at	AMMECR1L	2	128335675	0.035722791
223312_at	C2orf7	2	73308641	0.015061861
225174_at	DNAJC10	2	183289243	0.006164888
226749_at	MRPS9	2	105020914	0.049944792
227642_at	TFCP2L1	2	121690635	0.010231873
227867_at	LOC129293	2	84902306	0.023555392
229211_at	DUSP28	2	241148143	0.041144451
AFFX-				
HUMISGF3A/M97935_3_at	STAT1	2	191542006	0.006164888
1553296_at	GPR128	3	101811122	0.010231873
201904_s_at	CTDSPL	3	37878672	0.00368053
201906_s_at	CTDSPL	3	37878672	0.049944792
203140_at	BCL6	3	188921858	0.023555392
204245_s_at	RPP14	3	58267011	0.006164888
204294_at	AMT	3	49429214	0.023555392
205019_s_at	VIPR1	3	42519107	0.041144451
206286_s_at	TDGF1 /// TDGF3	3	46594216	0.049944792
207808_s_at	PROS1	3	95074571	0.041144451
208782_at	FSTL1	3	121595751	0.023555392
208911_s_at	PDHB	3	58388393	0.023555392
218440_at	MCCC1	3	184215699	0.049944792
221872_at	RARRES1	3	159897590	0.023555392
223220_s_at	PARP9	3	123729451	0.041144451

Experimental ID	Gene Symbol	Chromosome	Transcription Start Sites	P value
223969_s_at	RETNLB	3	109957175	0.041144451
224701_at	PARP14	3	123882361	0.035722791
226850_at	SUMF1	3	4377829	0.029076187
226987_at	RBM15B	3	51403770	0.023555392
227676_at	FAM3D	3	58594709	0.002208318
227679_at	HDAC11	3	13496714	0.041144451
228754_at	SLC6A6	3	14419109	0.007933587
229544_at	---	3	40483031	0.029076187
1568634_a_at	LRRC66	4	52554622	0.018875861
203854_at	CFI	4	110881296	0.035722791
204351_at	S100P	4	6746466	0.023555392
204379_s_at	FGFR3	4	1764836	0.023555392
223157_at	C4orf14	4	57524272	0.023555392
225179_at	UBE2K	4	39376058	0.049944792
225391_at	LOC93622	4	6726583	0.029076187
226001_at	KLHL5	4	38723053	0.011839038
228992_at	MED28	4	17225370	0.007933587
230329_s_at	NUDT6	4	124033248	0.041144451
200665_s_at	SPARC	5	151021201	0.015061861
201278_at	DAB2	5	39407536	0.010231873
201506_at	TGFBI	5	135392482	0.029076187
203286_at	RNF44	5	175886305	0.029076187
203989_x_at	F2R	5	76047623	0.029076187
204020_at	PURA	5	139473891	0.015061861
206430_at	CDX1	5	149526536	0.041144451
	ANKHD1-EIF4EBP3 ///			
214919_s_at	EIF4EBP3	5	139888594	0.015061861
217840_at	DDX41	5	176871183	0.029076187
218398_at	MRPS30	5	44844783	0.049944792
224782_at	ZMAT2	5	140060215	0.041144451
225275_at	EDIL3	5	83273881	0.010231873
226826_at	---	5	157117434	0.049944792
227314_at	ITGA2	5	52320912	0.018875861
227339_at	RGMB	5	98132898	0.035722791
227628_at	GPX8	5	54491740	0.049944792
204214_s_at	RAB32	6	146906520	0.041144451
204279_at	PSMB9	6	32929915	0.041144451
204565_at	ACOT13	6	24775241	0.049944792
205771_s_at	AKAP7	6	131508153	0.023555392
208788_at	ELOVL5	6	53240154	0.049944792
209193_at	PIM1	6	37245899	0.015061861
212265_at	QKI	6	163755664	0.035722791
213540_at	HSD17B8	6	33280396	0.015061861
214106_s_at	GMDS	6	1569039	0.035722791

Experimental ID	Gene Symbol	Chromosome	Transcription Start Sites	P value
219006_at	NDUFAF4	6	97444860	0.018875861
224430_s_at	MTO1	6	74228174	0.011839038
225664_at	COL12A1	6	75850761	0.002208318
225973_at	TAP2	6	32897587	0.029076187
227226_at	MRAP2	6	84800138	0.023555392
202404_s_at	COL1A2	7	93861808	0.002797203
203324_s_at	CAV2	7	115926679	0.010231873
209465_x_at	PTN	7	136562634	0.002797203
209466_x_at	PTN	7	136562634	0.029076187
211737_x_at	PTN	7	136562634	0.029076187
212345_s_at	CREB3L2	7	137210264	0.011839038
212792_at	DPY19L1	7	34935017	0.018875861
213333_at	MDH2	7	75515328	0.010231873
213546_at	DKFZP58611420	7	30376190	0.041144451
214435_x_at	RALA	7	39629686	0.015061861
218543_s_at	PARP12	7	139370017	0.049944792
218654_s_at	MRPS33	7	140352429	0.035722791
219041_s_at	REPIN1	7	149696811	0.035722791
224848_at	CDK6	7	92072172	0.049944792
224890_s_at	C7orf59	7	99584465	0.041144451
227753_at	TMEM139	7	142692184	0.041144451
230036_at	SAMD9L	7	92597303	0.041144451
201066_at	CYC1	8	145221947	0.035722791
202625_at	LYN	8	56954939	0.015061861
202626_s_at	LYN	8	56954939	0.015061861
203000_at	STMN2	8	80685934	0.011839038
203001_s_at	STMN2	8	80685934	0.049944792
203560_at	GGH	8	64090196	0.015061861
204505_s_at	EPB49	8	21967026	0.049944792
210754_s_at	LYN	8	56954939	0.035722791
213423_x_at	TUSC3	8	15442100	0.049944792
225478_at	MFHAS1	8	8679408	0.035722791
225591_at	FBXO25	8	346807	0.015061861
225681_at	CTHRC1	8	104452961	0.006164888
231270_at	CA13	8	86344967	0.029076187
235849_at	SCARA5	8	27783668	0.00473211
201060_x_at	STOM	9	123141173	0.041144451
201061_s_at	STOM	9	123141173	0.015061861
201645_at	TNC	9	116822625	0.023555392
202151_s_at	UBAC1	9	137964635	0.006164888
202893_at	UNC13B	9	35151988	0.023555392
204083_s_at	TPM2	9	35671989	0.041144451
204772_s_at	TTF1	9	134240757	0.041144451
207214_at	SPINK4	9	33230195	0.041144451

Experimental ID	Gene Symbol	Chromosome	Transcription Start Sites	P value
212068_s_at	BAT2L	9	133295297	0.049944792
212488_at	COL5A1	9	136673472	0.023555392
218189_s_at	NANS	9	99858779	0.006164888
225568_at	TMEM141	9	138805597	0.041144451
226499_at	NRARP	9	139313903	0.011839038
227186_s_at	MRPL41	9	139566129	0.041144451
228885_at	MAMDC2	9	71848316	0.029076187
229720_at	BAG1	9	33242469	0.018875861
208813_at	GOT1	10	101146617	0.041144451
212894_at	SUPV3L1	10	70609998	0.035722791
218174_s_at	C10orf57	10	81828405	0.010231873
225143_at	SFXN4	10	120890414	0.041144451
227163_at	GSTO2	10	106018620	0.035722791
227405_s_at	FZD8	10	35967182	0.035722791
227614_at	HKDC1	10	70650064	0.007933587
229450_at	IFIT3	10	91077581	0.007933587
1552703_s_at	CARD16 /// CASP1	11	104417331	0.049944792
1553828_at	FAM55A	11	113897646	0.015061861
1560587_s_at	PRDX5	11	63842144	0.002797203
200747_s_at	NUMA1	11	71391558	0.035722791
204070_at	RARRES3	11	63060848	0.029076187
204580_at	MMP12	11	102238673	0.035722791
205412_at	ACAT1	11	107497467	0.035722791
205482_x_at	SNX15	11	64551485	0.015061861
205547_s_at	TAGLN	11	116575249	0.011839038
208714_at	NDUFV1	11	67130898	0.010231873
208745_at	ATP5L	11	117777313	0.035722791
209436_at	SPON1	11	13940489	0.035722791
211368_s_at	CASP1	11	104401446	0.035722791
212573_at	ENDOD1	11	94462664	0.041144451
213869_x_at	THY1	11	118793865	0.015061861
217969_at	C11orf2	11	64620258	0.011839038
219188_s_at	MACROD1	11	63522606	0.010231873
219549_s_at	RTN3	11	63205497	0.015061861
222994_at	PRDX5	11	63842144	0.00473211
223103_at	STARD10	11	72143421	0.015061861
223305_at	TMEM216	11	60916440	0.029076187
225510_at	OAF	11	119586956	0.041144451
229369_at	VSIG2	11	124122579	0.006164888
229715_at	DKFZp686O24166	11	17329892	0.010231873
231530_s_at	C11orf1	11	111255157	0.029076187
232322_x_at	STARD10	11	72143421	0.006164888
1554599_x_at	---	12	122684337	0.049944792
202330_s_at	UNG	12	108019797	0.010231873

Experimental ID	Gene Symbol	Chromosome	Transcription Start Sites	P value
207158_at	APOBEC1	12	7693262	0.049944792
207170_s_at	LETMD1	12	49728350	0.029076187
208121_s_at	PTPRO	12	15366753	0.010231873
209195_s_at	ADCY6	12	47446241	0.011839038
210976_s_at	PFKM	12	46799294	0.035722791
212959_s_at	GNPTAB	12	100663407	0.018875861
218434_s_at	AACS	12	124115877	0.049944792
219076_s_at	PXMP2	12	131774264	0.011839038
224974_at	SUDS3	12	117298740	0.029076187
230252_at	LPAR5	12	6598261	0.041144451
233841_s_at	SUDS3	12	117298740	0.011839038
	GAS6	///		
1598_g_at	LOC100133684	13	113546912	0.023555392
	GAS6	///		
202177_at	LOC100133684	13	113546912	0.015061861
205677_s_at	DLEU1	13	49554414	0.041144451
209185_s_at	IRS2	13	109204184	0.035722791
211964_at	COL4A2	13	109757631	0.041144451
211980_at	COL4A1	13	109599310	0.015061861
211981_at	COL4A1	13	109599310	0.041144451
212768_s_at	OLFM4	13	52500972	0.00473211
218656_s_at	LHFP	13	38815028	0.035722791
221816_s_at	PHF11	13	48967801	0.029076187
227609_at	EPST11	13	42360121	0.049944792
200884_at	CKB	14	103055748	0.000736106
200989_at	HIF1A	14	61231871	0.00473211
201762_s_at	PSME2	14	23682413	0.023555392
202411_at	IFI27	14	93646831	0.018875861
204588_s_at	SLC7A7	14	22312271	0.049944792
206453_s_at	NDRG2	14	20554761	0.007933587
209210_s_at	FERMT2	14	52393739	0.023555392
209531_at	GSTZ1	14	76856982	0.049944792
	IGH@ /// IGHG1 ///			
	IGHG2 /// IGHM ///			
211430_s_at	IGHV4-31	14	105278720	0.035722791
215118_s_at	IGHG1	14	105241338	0.007933587
218298_s_at	C14orf159	14	90650109	0.035722791
226747_at	TXNDC16	14	51967058	0.049944792
227274_at	SYNJ2BP	14	69902965	0.049944792
228569_at	PAPOLA	14	96038472	0.041144451
237400_at	---	14	49862810	0.035722791
202766_s_at	FBN1	15	46487796	0.015061861
204321_at	NEO1	15	71131927	0.029076187
204698_at	ISG20	15	86983042	0.029076187
218791_s_at	C15orf29	15	32220166	0.049944792

Experimental ID	Gene Symbol	Chromosome	Transcription Start Sites	P value
225311_at	IVD	15	38484977	0.029076187
225330_at	IGF1R	15	97010283	0.049944792
225337_at	ABHD2	15	87432384	0.010231873
1553715_s_at	C16orf14	16	631849	0.029076187
1554078_s_at	DNAJA3	16	4415858	0.049944792
200708_at	GOT2	16	57298535	0.049944792
201468_s_at	NQO1	16	68300804	0.007933587
202183_s_at	KIF22	16	89374	0.049944792
204130_at	HSD11B2	16	66022536	0.007933587
204326_x_at	MT1X	16	55273882	0.015061861
204745_x_at	MT1G	16	55258153	0.00368053
206043_s_at	ATP2C2	16	82959633	0.029076187
206461_x_at	MT1H	16	55261226	0.007933587
207173_x_at	CDH11	16	63538183	0.015061861
208581_x_at	MT1X	16	55273882	0.010231873
208760_at	UBE2I	16	1299180	0.041144451
209230_s_at	NUPR1	16	28456162	0.041144451
209759_s_at	DCI	16	2229898	0.049944792
210519_s_at	NQO1	16	68300804	0.015061861
211456_x_at	MT1P2	16	55261251	0.010231873
212056_at	KIAA0182	16	84202529	0.007933587
212228_s_at	COQ9	16	56038902	0.029076187
212859_x_at	MT1E	16	55217085	0.007933587
213237_at	C16orf88	16	19625174	0.007933587
213629_x_at	MT1F	16	55249355	0.007933587
214035_x_at	LOC399491	16	15105673	0.041144451
214798_at	ATP2C2	16	82959633	0.007933587
217165_x_at	MT1F	16	55249355	0.00368053
217546_at	MT1M	16	55224034	0.007933587
218606_at	ZDHHC7	16	83565567	0.006164888
220335_x_at	CES3	16	65552638	0.029076187
223112_s_at	NDUFB10	16	1949517	0.001717581
224998_at	CMTM4	16	65206153	0.010231873
225129_at	CPNE2	16	55684010	0.035722791
225276_at	GSPT1	16	11869485	0.041144451
228049_x_at	---	16	2144803	0.015061861
240303_at	TMC5	16	19329557	0.049944792
242414_at	QPRT	16	29597941	0.018875861
200659_s_at	PHB	17	44836418	0.049944792
202311_s_at	COL1A1	17	45616455	0.002797203
203858_s_at	COX10	17	13913443	0.015061861
204690_at	STX8	17	9094512	0.049944792
204719_at	ABCA8	17	64375025	0.007933587
205601_s_at	HOXB5	17	44023617	0.029076187

Experimental ID	Gene Symbol	Chromosome	Transcription Start Sites	P value
210524_x_at	---	17	59248850	0.00473211
212338_at	MYO1D	17	27843740	0.011839038
218307_at	RSAD1	17	45911188	0.023555392
219909_at	MMP28	17	31116988	0.015061861
223170_at	TMEM98	17	28279040	0.018875861
224784_at	MLLT6	17	34115398	0.00368053
225165_at	PPP1R1B	17	35036704	0.007933587
225947_at	MYO19	17	31925711	0.00473211
226905_at	FAM101B	17	289998	0.015061861
228617_at	XAF1	17	6599879	0.018875861
230784_at	PRAC	17	44154080	0.011839038
231250_at	---	17	44068914	0.041144451
203126_at	IMPA2	18	11971454	0.035722791
212386_at	TCF4	18	51040559	0.029076187
219048_at	PIGN	18	57862437	0.035722791
227049_at	ZADH2	18	71038266	0.018875861
203576_at	BCAT2	19	53990130	0.035722791
205382_s_at	CFD	19	810664	0.011839038
205861_at	SPIB	19	55614006	0.035722791
216699_s_at	KLK1	19	56014215	0.035722791
217780_at	C19orf56	19	12639880	0.018875861
218312_s_at	ZSCAN18	19	63287021	0.035722791
218429_s_at	C19orf66	19	10057805	0.041144451
219360_s_at	TRPM4	19	54352863	0.035722791
219613_s_at	SIRT6	19	4125105	0.029076187
221692_s_at	MRPL34	19	17277476	0.041144451
226010_at	SLC25A23	19	6391074	0.023555392
227878_s_at	ALKBH7	19	6323443	0.049944792
234339_s_at	GLTSCR2	19	52940604	0.023555392
40837_at	TLE2	19	2948636	0.041144451
53720_at	C19orf66	19	10057805	0.007933587
202878_s_at	CD93	20	23007992	0.018875861
203892_at	WFDC2	20	43531807	0.015061861
217792_at	SNX5	20	17870243	0.010231873
225903_at	PIGU	20	32612006	0.029076187
226805_at	---	20	42364894	0.035722791
227160_s_at	C20orf7	20	13713681	0.023555392
202217_at	C21orf33	21	44377921	0.041144451
205548_s_at	BTG3	21	17887840	0.010231873
213134_x_at	BTG3	21	17887840	0.011839038
217867_x_at	BACE2	21	41461597	0.035722791
222446_s_at	BACE2	21	41461597	0.023555392
227125_at	---	21	33558910	0.035722791
1553974_at	C22orf39	22	17810894	0.007933587

Experimental ID	Gene Symbol	Chromosome	Transcription Start Sites	P value
201247_at	SREBF2	22	40559051	0.006164888
212100_s_at	POLDIP3	22	41309670	0.035722791
217719_at	EIF3L	22	36575315	0.011777696
219716_at	APOL6	22	34374369	0.00473211
1555864_s_at	PDHA1	X	19271931	0.015061861
200980_s_at	PDHA1	X	19271931	0.029076187
201215_at	PLS3	X	114701461	0.049944792
202242_at	TSPAN7	X	38305674	0.029076187
203446_s_at	OCRL	X	128501932	0.029076187
212741_at	MAOA	X	43400352	0.041144451
213746_s_at	FLNA	X	153230093	0.041144451
214752_x_at	FLNA	X	153230093	0.035722791
215440_s_at	BEX4	X	102356675	0.035722791
217963_s_at	NGFRAP1	X	102517923	0.015061861
219297_at	WDR44	X	117364069	0.010231873
221196_x_at	BRCC3	X	153952903	0.041144451
221620_s_at	APOO	X	23761397	0.041144451
222257_s_at	ACE2	X	15489076	0.041144451
224367_at	BEX2	X	102450929	0.010231873
227425_at	REPS2	X	16874734	0.035722791
228752_at	BCORL1	X	128944349	0.041144451

Genome-wide data set (layer I) of all individuals included in the study have been submitted to the NCBI Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) under series accession numbers GSE22619 (samples: GSM560961–GSM560976).

Table S 2 Genome-wide significant methylational candidates (layerII)

Experimental ID	Chromosome	Region Start	Region End	P value
3800_0606_0724	1	89436901	89437900	0.006256901
3800_0755_0573	1	228904908	230211618	0.006256901
3800_0681_0089	1	85438407	85440238	0.006256901
3800_0713_0797	1	115101847	115102877	0.008342535
3800_0023_0719	1	195123584	195124583	0.008342535
3800_0063_0819	1	222584197	222585257	0.008342535
3800_0146_0154	1	203556856	203558256	0.008342535
3800_0632_0866	1	2333246	2334620	0.015016562
3800_0055_0193	1	228904908	230211618	0.015016562
3800_0143_0457	1	228904908	230211618	0.015016562
3800_0081_0915	1	228904908	230211618	0.015016562
3800_0213_0229	1	228904908	230211618	0.015016562
3800_0243_0005	1	228904908	230211618	0.015016562
3800_0285_0661	1	148088526	148091815	0.01981352
3800_0476_0484	1	36544418	36546055	0.02681529
3800_0050_0050	1	85814818	85815523	0.031284505
3800_0659_0845	1	224031905	224032909	0.031284505
3800_0638_0940	1	209498729	209500757	0.031284505
3800_0049_0567	1	226394037	226395364	0.031284505
3800_0615_0623	1	233878465	233881427	0.031284505
3800_0560_0390	1	47552125	47553124	0.031284505
3800_0295_0219	1	9979593	9980799	0.031284505
3800_0539_0495	1	78017648	78018922	0.031284505
3800_0111_0929	1	228904908	230211618	0.031284505
3800_0021_0625	1	159767053	159767736	0.041921237
3800_0344_0398	1	54726482	54727875	0.041921237
3800_0583_0461	1	226403927	226404926	0.041921237
3800_0731_1023	1	11041618	11043394	0.041921237
3800_0467_0019	1	1299004	1301193	0.041921237
3800_0146_0572	1	25536194	25538146	0.041921237
3800_0068_0956	1	26478005	26479500	0.041921237
3800_0142_0932	1	32478112	32479728	0.041921237
3800_0530_0460	1	32573447	32575171	0.041921237
3800_0748_0138	1	36046087	36047165	0.041921237
3800_0043_0731	1	228904908	230211618	0.041921237
3800_0465_0141	2	236736534	236743501	0.006256901
3800_0540_0720	2	24436125	24437115	0.008342535
3800_0558_0290	2	25207788	25208506	0.010949577
3800_0729_0331	2	89776352	89777351	0.010949577
3800_0326_0102	2	91300424	91301423	0.010949577
3800_0335_0975	2	136589730	136592494	0.010949577
3800_0681_0569	2	66656395	66658335	0.015016562

Experimental ID	Chromosome	Region Start	Region End	P value
3800_0434_0878	2	241690148	241691147	0.015016562
3800_0041_0685	2	203901401	204925184	0.015016562
3800_0146_0050	2	241618304	241619303	0.01981352
3800_0441_0435	2	203901401	204925184	0.01981352
3800_0143_0599	2	160181077	160182076	0.02681529
3800_0577_0787	2	47030815	47031814	0.02681529
3800_0383_0259	2	12774077	12776470	0.02681529
3800_0653_0815	2	173000314	173001429	0.02681529
3800_0720_0674	2	187764982	188209235	0.02681529
3800_0052_1016	2	45093877	45095083	0.031284505
3800_0516_0902	2	160626711	160628117	0.031284505
3800_0754_0494	2	177965416	177966499	0.031284505
3800_0203_0451	2	203901401	204925184	0.031284505
3800_0484_0618	2	219505705	220755705	0.031284505
3800_0623_0873	2	219505705	220755705	0.031284505
3800_0623_0353	2	156884498	156885592	0.041921237
3800_0143_0049	2	232959458	232961658	0.041921237
3800_0067_0899	2	53848207	53849206	0.041921237
3800_0760_0782	2	230286307	230288268	0.041921237
3800_0362_0276	2	241156031	241157721	0.041921237
3800_0639_0875	2	70847956	70849587	0.041921237
3800_0208_1010	2	187764982	188209235	0.041921237
3800_0608_0248	2	203901401	204925184	0.041921237
3800_0697_0449	2	203901401	204925184	0.041921237
3800_0047_0353	3	174595965	174596624	0.006256901
3800_0414_0038	3	106854807	107075518	0.006256901
3800_0667_0713	3	42517825	42519871	0.006256901
3800_0556_0512	3	198153093	198154692	0.008342535
3800_0056_0146	3	143824785	143825784	0.015016562
3800_0331_0061	3	79721502	79722501	0.015016562
3800_0001_0233	3	139530646	139531645	0.015016562
3800_0632_0446	3	48673340	48676671	0.01981352
3800_0125_0565	3	106854807	107075518	0.02681529
3800_0153_1005	3	186752801	186753813	0.02681529
3800_0439_0793	3	39123382	39125608	0.02681529
3800_0044_0878	3	106854807	107075518	0.031284505
3800_0053_0693	3	158193315	158194314	0.031284505
3800_0514_0324	3	40403440	40404439	0.031284505
3800_0487_0879	3	72578215	72579542	0.031284505
3800_0143_0549	3	106854807	107075518	0.041921237
3800_0704_0726	3	9152626	9153189	0.041921237
3800_0597_0677	3	115149050	115150188	0.041921237
3800_0600_0274	3	49178290	49179522	0.041921237
3800_0071_0785	3	5203959	5205732	0.041921237
3800_0114_0722	4	110968335	110969334	0.010949577

Experimental ID	Chromosome	Region Start	Region End	P value
3800_0434_0672	4	1386292	1391730	0.010949577
3800_0174_0134	4	69464892	69465891	0.01981352
3800_0033_0053	4	114901519	114902927	0.01981352
3800_0067_0559	4	95591825	95592884	0.01981352
3800_0080_0806	4	143571613	143572612	0.031284505
3800_0027_0023	4	114901519	114902927	0.031284505
3800_0434_0050	4	41057010	41058396	0.031284505
3800_0556_0510	4	75122818	75124020	0.031284505
3800_0764_0066	4	42093910	42095559	0.041921237
3800_0336_0154	4	104216658	104218129	0.041921237
3800_0540_0122	4	140695832	140698078	0.041921237
3800_0526_1014	5	139818777	139819776	0.006256901
3800_0161_0689	5	148864177	148865176	0.010949577
3800_0145_0669	5	153805517	153806719	0.015016562
3800_0208_0420	5	158584694	158713204	0.02681529
3800_0757_0499	5	158584694	158713204	0.02681529
3800_0422_0930	5	159671360	159672491	0.031284505
3800_0735_0905	5	72778365	72780693	0.031284505
3800_0057_0307	5	158584694	158713204	0.031284505
3800_0198_0708	5	39238570	39239569	0.041921237
3800_0623_0049	5	146812754	146814203	0.041921237
3800_0722_0278	5	1576026	1577826	0.041921237
3800_0146_0774	5	176169343	176170916	0.041921237
3800_0022_0824	5	39109639	39111012	0.041921237
3800_0674_0012	6	1549606	1560865	0.006256901
3800_0110_0028	6	110974237	112440374	0.008342535
3800_0386_0350	6	110974237	112440374	0.010949577
3800_0375_0011	6	116971007	116972006	0.015016562
3800_0600_0116	6	16868158	16870964	0.015016562
3800_0542_0694	6	29797000	33606563	0.015016562
3800_0130_0714	6	29797000	33606563	0.015016562
3800_0723_0969	6	29797000	33606563	0.015016562
3800_0050_0458	6	29797000	33606563	0.015016562
3800_0026_0774	6	29797000	33606563	0.01981352
3800_0207_0831	6	158572694	158574058	0.02681529
3800_0072_0648	6	116681615	116682614	0.02681529
3800_0303_0069	6	139391286	139392433	0.02681529
3800_0623_0707	6	29797000	33606563	0.02681529
3800_0527_0671	6	152160379	152502429	0.02681529
3800_0254_0374	6	110974237	112440374	0.02681529
3800_0649_0125	6	110974237	112440374	0.02681529
3800_0130_0902	6	27968657	27969932	0.031284505
3800_0673_0897	6	29797000	33606563	0.031284505
3800_0027_0065	6	110974237	112440374	0.031284505
3800_0242_0384	6	29797000	33606563	0.04152307

Experimental ID	Chromosome	Region Start	Region End	P value
3800_0239_0047	6	29797000	33606563	0.041921237
3800_0051_0183	6	29797000	33606563	0.041921237
3800_0101_0177	6	152160379	152502429	0.041921237
3800_0527_0977	6	152160379	152502429	0.041921237
3800_0678_0634	6	110974237	112440374	0.041921237
3800_0008_0484	6	110974237	112440374	0.041921237
3800_0237_0193	6	110974237	112440374	0.041921237
3800_0043_0007	7	86965883	87185500	0.015016562
3800_0443_0775	7	102576500	102577499	0.015016562
3800_0576_0808	7	92056997	92058390	0.015016562
3800_0539_0941	7	98083742	98085695	0.025809717
3800_0114_0594	7	86965883	87185500	0.02681529
3800_0613_0745	7	71987711	71988701	0.02681529
3800_0455_0711	7	136678837	136679836	0.031284505
3800_0616_0718	7	107318598	107319597	0.031284505
3800_0653_0759	7	138370585	138372065	0.041921237
3800_0689_0467	7	149706723	149708133	0.041921237
3800_0434_0152	8	145390882	145392024	0.006256901
3800_0596_0152	8	65655119	65657006	0.010949577
3800_0335_0151	8	96350021	96351344	0.015016562
3800_0609_0887	8	124849468	124850634	0.01981352
3800_0500_0654	8	20098735	20099734	0.02681529
3800_0503_0997	8	143691433	143693585	0.02681529
3800_0097_0807	8	116749800	116751179	0.041921237
3800_0075_0967	8	33543132	33544935	0.041921237
3800_0239_0049	8	61591729	61592863	0.041921237
3800_0626_0776	9	139265302	139266301	0.006256901
3800_0685_0577	9	139083977	139085599	0.010949577
3800_0369_0517	9	2611309	2613685	0.010949577
3800_0153_0391	9	8877887	8878886	0.010949577
3800_0735_0181	9	101621613	101627382	0.015016562
3800_0723_0709	9	93225473	93227117	0.015016562
3800_0748_0384	9	99784804	99787589	0.015016562
3800_0496_0218	9	71476625	71477901	0.01981352
3800_0669_0225	9	130077413	130079072	0.02681529
3800_0653_0127	9	139006077	139010164	0.02681529
3800_0483_0803	9	97307956	97313729	0.031284505
3800_0530_0052	9	136949289	136950380	0.041921237
3800_0023_0171	9	101621613	101627382	0.041921237
3800_0523_0201	9	204587	206146	0.041921237
3800_0166_0348	9	37312007	37313006	0.041921237
3800_0626_0152	10	26540600	26638493	0.006256901
3800_0614_0804	10	99521016	99522477	0.006256901
3800_0434_0154	10	124729671	124730670	0.008342535
3800_0753_0517	10	122698501	122699322	0.010949577

Experimental ID	Chromosome	Region Start	Region End	P value
3800_0527_0049	10	44180619	44205548	0.015016562
3800_0089_0561	10	4995204	4996203	0.01981352
3800_0693_0657	10	43421612	43422611	0.01981352
3800_0606_0716	10	75303443	75305099	0.01981352
3800_0321_0723	10	31642430	31863133	0.01981352
3800_0684_0182	10	5126319	5127318	0.02681529
3800_0527_0877	10	50639665	50641125	0.02681529
3800_0273_0289	10	79215557	79303959	0.031284505
3800_0582_0638	10	38685045	38686115	0.031284505
3800_0533_0245	10	126595111	126596202	0.041921237
3800_0010_0752	10	101180087	101181086	0.041921237
3800_0575_0107	10	13429212	13431036	0.041921237
3800_0208_0896	10	63092476	63093475	0.041921237
3800_0555_0773	10	72826171	72827960	0.041921237
3800_0719_0353	11	117252313	117253327	0.008342535
3800_0764_0068	11	13440850	13442314	0.015016562
3800_0581_0053	11	66790252	66791257	0.01981352
3800_0680_0310	11	62449950	62451348	0.02681529
3800_0053_0491	11	30909778	30910777	0.02681529
3800_0494_0568	11	45899523	45902108	0.02681529
3800_0361_0657	11	47163765	47165752	0.02681529
3800_0639_0701	11	60885610	60887182	0.02681529
3800_0582_0106	11	65535789	65536788	0.02681529
3800_0143_0047	11	109468451	109469887	0.031284505
3800_0645_0677	11	45895873	45897055	0.031284505
3800_0679_0245	11	808653	810271	0.031284505
3800_0050_0154	11	87710150	87711336	0.031284505
3800_0458_0656	11	61032402	61033905	0.041921237
3800_0239_0337	11	55776057	55777056	0.041921237
3800_0738_0968	11	109671775	109673297	0.041921237
3800_0107_0715	11	119103593	119105710	0.041921237
3800_0690_0330	11	120668455	120669458	0.041921237
3800_0600_0270	11	47192840	47193839	0.041921237
3800_0160_0742	11	63815385	63816694	0.041921237
3800_0653_0613	11	67831677	67978301	0.041921237
3800_0726_0768	12	55767065	55770072	0.008342535
3800_0050_0356	12	108755332	108756345	0.010949577
3800_0722_0978	12	124236071	124238065	0.015016562
3800_0143_0671	12	122964710	122965709	0.01981352
3800_0239_0671	12	52696691	52697690	0.01981352
3800_0528_0460	12	73889529	73890528	0.01981352
3800_0269_0247	12	61225477	61226476	0.01981352
3800_0599_0021	12	122016653	122017746	0.02681529
3800_0434_0776	12	51394128	51394738	0.031284505
3800_0678_1014	12	31634028	31635969	0.031284505

Experimental ID	Chromosome	Region Start	Region End	P value
3800_0014_0932	12	54610964	54612829	0.031284505
3800_0542_0680	12	67012098	67013129	0.031284505
3800_0473_0727	12	59165655	59179739	0.031284505
3800_0528_0154	13	33014558	33015140	0.010949577
3800_0242_0254	13	92677247	92678878	0.015016562
3800_0012_0446	13	109232468	109238181	0.041921237
3800_0541_0071	13	114018227	114019408	0.041921237
3800_0434_0354	13	78073612	78076446	0.041921237
3800_0146_0254	14	77034325	77035696	0.010949577
3800_0713_0805	14	35347530	35348371	0.015016562
3800_0757_0343	14	64001633	64002295	0.015016562
3800_0719_0773	14	105138861	105139860	0.015016562
3800_0040_0298	14	19800342	19999532	0.015016562
3800_0681_0925	14	56116015	56117046	0.02681529
3800_0242_0980	14	36056114	36060327	0.02681529
3800_0343_0375	14	23674923	23675968	0.031284505
3800_0140_0458	14	92720499	92721832	0.031284505
3800_0044_0060	14	100510647	100511646	0.041921237
3800_0001_0395	14	100514843	100516832	0.041921237
3800_0239_0669	15	68174857	68180286	0.006256901
3800_0765_0391	15	92561236	94081236	0.006256901
3800_0432_0010	15	92561236	94081236	0.010949577
3800_0061_0931	15	32446268	32447521	0.01981352
3800_0599_0369	15	92561236	94081236	0.01981352
3800_0370_0526	15	78331070	78331717	0.02681529
3800_0457_0525	15	88446003	88447462	0.02681529
3800_0004_0784	15	92561236	94081236	0.02681529
3800_0335_0773	15	92561236	94081236	0.02681529
3800_0087_0071	15	87541220	87542219	0.031284505
3800_0296_0348	15	92561236	94081236	0.031284505
3800_0012_0688	15	92561236	94081236	0.031284505
3800_0115_0801	15	63870894	63872390	0.041921237
3800_0562_0316	15	81471012	81472032	0.041921237
3800_0088_0756	15	92561236	94081236	0.041921237
3800_0146_0458	15	92561236	94081236	0.041921237
3800_0272_0806	15	92561236	94081236	0.041921237
3800_0152_0758	15	43052799	43053798	0.041921237
3800_0431_0047	16	3096251	3096890	0.008342535
3800_0718_0656	16	74024352	74025638	0.010949577
3800_0673_0187	16	18719889	18721712	0.015016562
3800_0573_0999	16	66119972	66121833	0.015016562
3800_0662_0998	16	87280022	87281040	0.015016562
3800_0046_0884	16	85156857	85160150	0.01981352
3800_0617_0235	16	78190121	78193038	0.02681529
3800_0606_0650	16	1410504	1411283	0.031284505

Experimental ID	Chromosome	Region Start	Region End	P value
3800_0613_0959	16	2226220	2228074	0.041921237
3800_0338_0570	16	30528502	30529933	0.041921237
3800_0510_1018	17	43455797	43457903	0.006256901
3800_0623_0355	17	1873749	1876158	0.008342535
3800_0623_0975	17	30593255	30594953	0.008342535
3800_0419_0861	17	33971254	33972119	0.010949577
3800_0527_0771	17	7965484	7967587	0.010949577
3800_0478_0482	17	35549861	35550860	0.010949577
3800_0651_0679	17	73694477	73695788	0.015016562
3800_0527_0669	17	37422315	37423947	0.015016562
3800_0259_0699	17	16282823	16284825	0.015016562
3800_0074_0362	17	18624891	18625890	0.01981352
3800_0367_0523	17	7995276	7997177	0.01981352
3800_0722_0280	17	77279977	77281673	0.031284505
3800_0626_0358	17	59349681	59350680	0.041921237
3800_0335_0251	17	77128351	77130594	0.041921237
3800_0242_0878	18	45973865	45975926	0.006256901
3800_0661_0691	18	12647045	12649479	0.006256901
3800_0767_0969	18	32020390	32022230	0.008342535
3800_0434_0052	18	46339837	46341587	0.015016562
3800_0128_0326	18	17662798	17663797	0.031284505
3800_0570_0548	18	12981062	12981696	0.041921237
3800_0573_0187	18	44755333	44756974	0.041921237
3800_0011_0049	18	42936233	42937232	0.041921237
3800_0477_0983	18	64654707	64656295	0.041921237
3800_0199_0211	19	38978025	38981323	0.006256901
3800_0618_0436	19	218086	218762	0.015016562
3800_0335_0049	19	8814083	8815082	0.01981352
3800_0235_0323	19	4074522	4075876	0.01981352
3800_0143_0149	19	55012577	55013574	0.031284505
3800_0329_0681	19	63720165	63720751	0.031284505
3800_0688_0250	19	12941183	12942182	0.031284505
3800_0490_0500	19	59384842	59387309	0.031284505
3800_0634_0214	19	2439288	2440641	0.041921237
3800_0650_0580	19	44689431	44690131	0.041921237
3800_0529_0841	20	60557856	60559928	0.006256901
3800_0667_0063	20	48844689	48845765	0.006256901
3800_0157_0483	20	29656504	29657686	0.015016562
3800_0369_0611	20	34005017	34006592	0.02681529
3800_0698_0798	20	47986072	47987071	0.02681529
3800_0662_0992	20	47965011	47966225	0.041921237
3800_0719_0357	20	57947467	57949701	0.041921237
3800_0549_0761	21	46911629	46912534	0.010949577
3800_0755_0651	21	42512018	42513827	0.010949577
3800_0516_0214	21	43400233	43401453	0.015016562

Experimental ID	Chromosome	Region Start	Region End	P value
3800_0588_0370	21	37366576	37368367	0.031284505
3800_0506_1002	21	16833031	16834764	0.041921237
3800_0550_0472	22	30016637	30017371	0.008342535
3800_0241_0975	22	21742225	21743597	0.015016562
3800_0646_0848	22	48980775	48982516	0.015016562
3800_0146_0456	22	19386259	19388793	0.02681529
3800_0494_0448	22	37231500	37233012	0.02681529
3800_0735_0509	22	27467770	27468769	0.031284505
3800_0456_0290	22	45307982	45313525	0.031284505
3800_0316_0228	X	266967	268377	0.01981352
3800_0527_0875	X	153241104	153242142	0.02681529
3800_0611_0231	X	48803211	48804210	0.031284505
3800_0013_1017	X	13957665	13958682	0.031284505
3800_0431_0787	X	152699463	152700462	0.041921237
3800_0632_0766	X	133757705	133759651	0.041921237
3800_0177_0839	Y	19477494	19478493	0.015016562
3800_0211_0071	Y	19264407	19265406	0.02681529
3800_0083_0777	Y	16414655	16415654	0.02681529
3800_0291_0669	Y	15698041	15699040	0.031284505
3800_0063_0693	*	*	*	0.006256901
3800_0017_0957	*	*	*	0.010949577
3800_0532_0440	*	*	*	0.02681529
3800_0060_0448	*	*	*	0.02681529
3800_0070_0936	*	*	*	0.031284505
3800_0057_0603	*	*	*	0.031284505
3800_0656_0846	*	*	*	0.041921237
3800_0737_0587	*	*	*	0.041921237
3800_0029_0879	*	*	*	0.041921237
3800_0066_0450	*	*	*	0.041921237
3800_0530_0570	*	*	*	0.041921237
3800_0695_0985	*	*	*	0.041921237
3800_0039_0401	*	*	*	0.041921237
3800_0750_0244	*	*	*	0.041921237

* indicates that for the artificial control does not map to the human genome, and genome-wide data sets (layer II) of all individuals included in the study have been submitted to the NCBI Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) under series accession numbers GSE 27899 (samples GSM688907–GSM688926).

Table S 3 Genome-wide significant methylational candidates (Layer III)

Experimental ID	Chromosome	MVP	P value
cg02813121	1	151615535	0.001584793
cg01053621	1	159460615	0.001584793
cg25856811	1	151240581	0.002091046
cg15756091	1	180686278	0.003714358
cg00625425	1	62558100	0.004915793
cg10904672	1	148019991	0.004915793
cg04567009	1	159867393	0.008344175
cg25020850	1	180908809	0.008344175
cg22077553	1	169888355	0.008344175
cg01377755	1	145609882	0.010620314
cg01426743	1	177260866	0.016932758
cg14467840	1	151867596	0.016932758
cg19685066	1	17954590	0.016932758
cg15361231	1	191341814	0.016932758
cg12234947	1	109958343	0.016932758
cg01459162	1	17446917	0.016932758
cg05190718	1	116112235	0.016932758
cg16776350	1	158815782	0.016932758
cg01214209	1	38031582	0.021130572
cg06589885	1	144124550	0.021130572
cg00974864	1	159867677	0.021130572
cg08763351	1	151209834	0.021130572
cg08477744	1	17180619	0.021130572
cg21919219	1	103999422	0.021130572
cg26073060	1	32459323	0.021130572
cg11750883	1	150753574	0.021130572
cg24987706	1	165327412	0.024548753
cg03329572	1	155789877	0.024548753
cg04743063	1	100504122	0.024548753
cg01459453	1	167865836	0.024548753
cg02327719	1	26057973	0.024548753
cg27341860	1	246166503	0.024548753
cg09118625	1	68285559	0.024548753
cg12467090	1	202725762	0.024548753
cg14615807	1	55236761	0.024548753
cg18354594	1	245343270	0.024548753
cg16456919	1	202922001	0.024548753
cg00095526	1	199348309	0.030265142
cg06101324	1	151222167	0.030265142
cg17241310	1	90955444	0.030265142
cg05606799	1	202431930	0.030265142
cg10331038	1	173642651	0.030265142

Experimental ID	Chromosome	MVP	P value
cg18354264	1	224477631	0.030265142
cg19370451	1	150653185	0.030265142
cg07689731	1	31166502	0.030265142
cg12869058	1	94360460	0.030265142
cg03941108	1	22858524	0.030265142
cg01770400	1	172153108	0.036908993
cg13626881	1	201325801	0.036908993
cg09448880	1	151550465	0.036908993
cg16501235	1	148512612	0.036908993
cg27210447	1	16221436	0.036908993
cg08555657	1	151334855	0.036908993
cg12114524	1	199657417	0.036908993
cg24051818	1	54880506	0.036908993
cg05248781	1	150749823	0.036908993
cg20839149	1	155053225	0.036908993
cg23909633	1	205137162	0.036908993
cg23282674	1	205105925	0.044803416
cg04629204	1	26220912	0.044803416
cg14448116	1	159436734	0.044803416
cg00532890	1	162795306	0.044803416
cg19728577	1	42391750	0.044803416
cg22165507	1	163867631	0.044803416
cg13447818	1	150565508	0.044803416
cg11832543	1	176748441	0.044803416
cg27336379	1	75850818	0.044803416
cg04138756	1	151240473	0.044803416
cg24641737	1	111544794	0.044803416
cg03641225	1	68285127	0.044803416
cg26112639	1	245646729	0.044803416
cg18854666	2	218955299	0.001584793
cg22415472	2	107968793	0.001584793
cg27485921	2	46600883	0.003714358
cg24697329	2	131390846	0.004915793
cg17518825	2	46376965	0.006405205
cg11346450	2	234302782	0.006405205
cg11827101	2	8385873	0.008344175
cg18638581	2	74913110	0.010620314
cg10062065	2	220016991	0.013511376
cg13877895	2	69054513	0.013511376
cg18374517	2	232979819	0.013511376
cg19132372	2	206787397	0.016932758
cg23667432	2	232952683	0.016932758
cg25066857	2	85774949	0.016932758
cg10807560	2	40510169	0.021130572
cg11078738	2	69093957	0.021130572

Experimental ID	Chromosome	MVP	P value
cg21004129	2	218740269	0.021130572
cg21364150	2	183651796	0.021130572
cg00573606	2	237987419	0.021130572
cg03380645	2	85748559	0.021130572
cg17327492	2	49235663	0.024548753
cg11024597	2	106047843	0.024548753
cg06627364	2	87536203	0.024548753
cg09425228	2	228386249	0.030265142
cg06971096	2	219881835	0.030265142
cg11811840	2	234333905	0.030265142
cg00918005	2	79106718	0.030265142
cg13531460	2	231499057	0.030265142
cg24560809	2	95303066	0.030265142
cg27342801	2	79240781	0.030265142
cg02799466	2	229844655	0.030265142
cg26385743	2	130817933	0.036908993
cg07412254	2	120940004	0.036908993
cg11206634	2	128174869	0.036908993
cg26154999	2	224974590	0.036908993
cg25651505	2	85665534	0.036908993
cg19394737	2	219600605	0.036908993
cg27003571	2	144768299	0.044803416
cg15746445	2	234489672	0.044803416
cg25928444	2	220145681	0.044803416
cg24579667	2	156988657	0.044803416
cg00795812	2	242450682	0.044803416
cg05020203	2	242617502	0.044803416
cg08861115	2	113451848	0.044803416
cg09947274	2	27293419	0.044803416
cg23759710	2	42844461	0.044803416
cg10159529	3	3127530	0.001584793
cg04569233	3	39296453	0.002091046
cg15051063	3	144376637	0.003714358
cg24287460	3	130230716	0.003714358
cg14483391	3	185231921	0.004915793
cg04387658	3	123257770	0.004915793
cg17142134	3	172227335	0.008344175
cg14717170	3	166397165	0.008344175
cg00400028	3	142432606	0.008344175
cg24315815	3	147452269	0.010620314
cg22083047	3	64186504	0.010620314
cg13742532	3	42825852	0.010620314
cg00037940	3	139961975	0.010620314
cg13550608	3	50289658	0.010620314
cg26838900	3	195571421	0.010620314

Experimental ID	Chromosome	MVP	P value
cg21706946	3	12775469	0.010620314
cg14375111	3	14140187	0.013511376
cg15113803	3	130729809	0.013511376
cg04347447	3	50516471	0.013511376
cg24389347	3	127760383	0.013511376
cg05564657	3	153014065	0.016932758
cg13144783	3	46224799	0.016932758
cg04739149	3	101315983	0.021130572
cg12894629	3	197427114	0.021130572
cg03109316	3	115440593	0.021130572
cg05445326	3	197549966	0.021130572
cg15439078	3	46879647	0.024548753
cg03391568	3	46893270	0.024548753
cg07361385	3	187813720	0.024548753
cg15946807	3	30868233	0.024548753
cg04848046	3	173312843	0.024548753
cg10883352	3	144378065	0.024548753
cg21044104	3	42426880	0.024548753
cg14458615	3	120347642	0.030265142
cg12242338	3	141879779	0.030265142
cg16396948	3	113086084	0.030265142
cg22356117	3	184364121	0.036908993
cg07084746	3	129777337	0.036908993
cg17386185	3	52296306	0.036908993
cg15423862	3	157320803	0.036908993
cg21572897	3	120760511	0.036908993
cg03019000	3	51679391	0.036908993
cg25982743	3	12175236	0.036908993
cg25298754	3	112796792	0.036908993
cg11761535	3	150533721	0.036908993
cg25781123	3	9379598	0.044803416
cg12069042	3	48441704	0.044803416
cg21951612	3	139574779	0.044803416
cg13099330	3	140740489	0.044803416
cg18669588	3	52249091	0.044803416
cg02912041	3	191650257	0.044803416
cg09841009	4	145281410	0.001584793
cg08649013	4	9651296	0.003714358
cg26701826	4	109034053	0.006405205
cg07364841	4	3413945	0.008344175
cg16998872	4	145046282	0.010620314
cg16418329	4	71235340	0.013511376
cg19000186	4	47650768	0.013511376
cg17725968	4	96980049	0.016932758
cg07625840	4	119493789	0.016932758

Experimental ID	Chromosome	MVP	P value
cg02255004	4	80967684	0.016932758
cg22252999	4	77215438	0.016932758
cg26125600	4	74937842	0.016932758
cg27108154	4	96980400	0.016932758
cg13255629	4	88971684	0.016932758
cg10375110	4	40032200	0.021130572
cg10307548	4	24404928	0.021130572
cg20357806	4	75072643	0.021130572
cg12243271	4	110942151	0.024548753
cg07360692	4	106286401	0.024548753
cg22670329	4	74920939	0.024548753
cg09816180	4	72868952	0.030265142
cg08493463	4	185376072	0.030265142
cg24829483	4	14950739	0.030265142
cg14078518	4	80548308	0.030265142
cg02992596	4	155921670	0.036908993
cg22461018	4	174774586	0.036908993
cg22960952	4	76700193	0.036908993
cg02936740	4	68678590	0.036908993
cg02029926	4	74953738	0.036908993
cg17738194	4	80548412	0.044803416
cg25421002	4	146778997	0.044803416
cg10409560	4	76700323	0.044803416
cg20950277	4	122304691	0.044803416
cg02955988	4	129104744	0.044803416
cg17854440	4	111617030	0.044803416
cg12845808	5	141318788	0.001155578
cg20131596	5	102229027	0.001155578
cg24585690	5	135259413	0.004915793
cg03503295	5	13997491	0.008344175
cg02196805	5	131437536	0.008344175
cg19421752	5	1278224	0.013511376
cg19530885	5	147238325	0.016932758
cg10345936	5	150708005	0.016932758
cg16434306	5	32823885	0.021130572
cg08818984	5	142795020	0.021130572
cg24363955	5	32824224	0.024548753
cg16360372	5	147192921	0.030265142
cg21652012	5	170143587	0.030265142
cg27606341	5	39255389	0.030265142
cg06501790	5	176743863	0.030265142
cg07233761	5	54317194	0.036908993
cg03504701	5	20016614	0.036908993
cg13590277	5	149999689	0.044803416
cg15512851	6	37081496	0.002091046

Experimental ID	Chromosome	MVP	P value
cg24063382	6	29565040	0.003714358
cg14023451	6	24597866	0.006405205
cg06196379	6	41362863	0.008344175
cg17407908	6	32261093	0.008344175
cg01669948	6	39398855	0.008344175
cg01980222	6	41238895	0.008344175
cg25488021	6	131319140	0.008344175
cg09299388	6	49863149	0.008344175
cg00333528	6	89984204	0.010620314
cg20790540	6	42992328	0.010620314
cg09508556	6	31214961	0.010620314
cg26424956	6	34209504	0.013511376
cg07434278	6	32114223	0.013511376
cg18201198	6	6568577	0.016932758
cg17741572	6	32022284	0.021130572
cg21789545	6	71069112	0.021130572
cg20095587	6	41238696	0.030265142
cg03835296	6	25938901	0.030265142
cg02611419	6	39390510	0.030265142
cg00024396	6	53321967	0.036908993
cg11847949	6	32114000	0.044803416
cg20856834	6	29450501	0.044803416
cg16268563	6	31621659	0.044803416
cg04645843	6	31026632	0.044803416
cg12629515	6	27967790	0.044803416
cg21440587	6	31691437	0.044803416
cg10681065	7	100077108	0.002091046
cg11693019	7	133862160	0.003714358
cg15001372	7	44196417	0.004915793
cg14898779	7	23716132	0.006405205
cg06506864	7	34664713	0.013511376
cg25650811	7	31523600	0.013511376
cg12150401	7	122422758	0.016932758
cg26756862	7	142723082	0.016932758
cg20748065	7	75421357	0.021130572
cg11808544	7	55724354	0.021130572
cg03843951	7	50597128	0.024548753
cg17754680	7	44548356	0.024548753
cg09966445	7	38638184	0.024548753
cg18463686	7	141293159	0.024548753
cg06943865	7	94131716	0.030265142
cg19465374	7	99411214	0.030265142
cg06762858	7	99809192	0.030265142
cg00644033	7	100382640	0.030265142
cg17628717	7	43317733	0.030265142

Experimental ID	Chromosome	MVP	P value
cg11091262	7	151284387	0.030265142
cg11828089	7	71516057	0.036908993
cg05279864	7	86947255	0.036908993
cg12643449	7	111849574	0.036908993
cg06942110	7	1511911	0.036908993
cg18117847	7	50411395	0.036908993
cg25612145	7	23716164	0.044803416
cg00245878	7	64088310	0.044803416
cg20080624	7	96489047	0.044803416
cg06981910	7	129634282	0.044803416
cg22066521	7	20653748	0.044803416
cg03716937	8	39815019	0.010620314
cg23704362	8	67568296	0.010620314
cg22218909	8	6863796	0.013511376
cg02658251	8	7788776	0.013511376
cg16303562	8	16094704	0.013511376
cg06276653	8	139233817	0.013511376
cg17191715	8	86478130	0.016932758
cg23566335	8	39561504	0.016932758
cg00319692	8	87180235	0.021130572
cg19391527	8	28229651	0.021130572
cg09314766	8	39814854	0.021130572
cg11695358	8	144869951	0.024548753
cg18939260	8	125810527	0.024548753
cg07005767	8	107851714	0.030265142
cg04017769	8	38764096	0.036908993
cg24243265	8	130869033	0.036908993
cg23067535	8	124264314	0.036908993
cg14120879	8	7332048	0.036908993
cg27212977	8	6770817	0.036908993
cg07459489	8	118216344	0.036908993
cg19982860	9	21156581	0.00066033
cg01074640	9	21218142	0.002091046
cg22161874	9	76692862	0.003714358
cg17357062	9	136950169	0.006405205
cg04103317	9	33230960	0.008344175
cg17683775	9	131555630	0.010620314
cg02602411	9	103396998	0.010620314
cg14611112	9	138763172	0.013511376
cg26164184	9	136911327	0.016932758
cg21122774	9	135594817	0.021130572
cg23815000	9	137552833	0.021130572
cg10726357	9	114693364	0.021130572
cg17301902	9	79453039	0.030265142
cg26776077	9	65233711	0.030265142

Experimental ID	Chromosome	MVP	P value
cg21023114	9	21231233	0.036908993
cg09547190	9	94897398	0.036908993
cg11873854	9	138762696	0.036908993
cg18506679	9	126603681	0.036908993
cg09427311	9	128925551	0.036908993
cg15248035	9	36159949	0.036908993
cg16112945	9	135275627	0.036908993
cg15582789	9	98185346	0.036908993
cg00563932	9	138990870	0.044803416
cg04995717	9	27099293	0.044803416
cg15526708	9	100906254	0.044803416
cg23693510	9	23815759	0.044803416
cg00691625	9	126309628	0.044803416
cg11959435	9	21430836	0.044803416
cg24392274	9	129961663	0.044803416
cg05050341	9	129657355	0.044803416
cg08458487	10	81699171	0.001155578
cg05441133	10	48036897	0.003714358
cg11762346	10	70650118	0.004915793
cg18833140	10	115302543	0.006405205
cg00532335	10	95506851	0.008344175
cg10548978	10	72215059	0.010620314
cg00623593	10	128657600	0.013511376
cg13882988	10	54201351	0.013511376
cg02876062	10	14856660	0.016932758
cg11639651	10	70649817	0.016932758
cg20340596	10	100196915	0.016932758
cg00899659	10	44815977	0.021130572
cg06692050	10	99522388	0.021130572
cg02803836	10	131302497	0.024548753
cg00025991	10	726625	0.024548753
cg24949488	10	98054352	0.024548753
cg04189838	10	96513337	0.030265142
cg15777781	10	11087277	0.030265142
cg13434852	10	128658171	0.030265142
cg21194776	10	74984134	0.030265142
cg05322019	10	102722313	0.030265142
cg13181019	10	28611588	0.030265142
cg05064181	10	116435388	0.036908993
cg06825166	10	98109069	0.036908993
cg05046097	10	88707906	0.036908993
cg05835416	10	96432664	0.036908993
cg25833031	10	30679302	0.036908993
cg04444771	10	44793718	0.044803416
cg22545356	10	88706639	0.044803416

Experimental ID	Chromosome	MVP	P value
cg21033494	10	51220120	0.044803416
cg06206628	10	49152832	0.044803416
cg07882535	11	2401968	0.000880441
cg25943276	11	131038494	0.000880441
cg13916742	11	61714364	0.001155578
cg18484189	11	7941549	0.001155578
cg09632136	11	113670871	0.002091046
cg20910746	11	27018992	0.002829987
cg25406518	11	60858741	0.002829987
cg17092637	11	104345246	0.003714358
cg21111471	11	63627576	0.003714358
cg24252809	11	18914027	0.004915793
cg24776407	11	68818210	0.004915793
cg19906550	11	2877375	0.006405205
cg05921324	11	116199324	0.008344175
cg19560971	11	118723448	0.008344175
cg13412615	11	63814383	0.008344175
cg20029201	11	118287023	0.010620314
cg25072962	11	60281151	0.010620314
cg13765961	11	59979814	0.010620314
cg18589858	11	74538633	0.013511376
cg16873863	11	2877752	0.013511376
cg04048249	11	116205766	0.016932758
cg15227610	11	111287226	0.016932758
cg16466334	11	102219568	0.016932758
cg02347487	11	7015679	0.016932758
cg21407055	11	3623059	0.021130572
cg21633698	11	118799744	0.021130572
cg10148841	11	124272665	0.021130572
cg21019522	11	2877365	0.021130572
cg22658979	11	102331679	0.021130572
cg24411312	11	129534054	0.021130572
cg03087937	11	26550132	0.024548753
cg03856801	11	116214348	0.024548753
cg11802013	11	69178356	0.024548753
cg22994720	11	74119636	0.024548753
cg22862656	11	61794235	0.024548753
cg20311730	11	7941759	0.024548753
cg06917325	11	62539699	0.024548753
cg25101936	11	113434374	0.030265142
cg11173246	11	831376	0.030265142
cg11494699	11	36544793	0.036908993
cg01697865	11	117361217	0.036908993
cg04962134	11	55409886	0.036908993
cg26202340	11	125657567	0.036908993

Experimental ID	Chromosome	MVP	P value
cg26149678	11	71387565	0.036908993
cg01987509	11	100507255	0.036908993
cg13179915	11	65119525	0.036908993
cg12493906	11	4965887	0.044803416
cg00041575	11	2279317	0.044803416
cg02477931	11	131037931	0.044803416
cg12949760	11	2499438	0.044803416
cg24506604	12	50865769	0.002091046
cg09447105	12	15017287	0.008344175
cg25463779	12	123339189	0.008344175
cg23256150	12	99391889	0.008344175
cg08640498	12	56114782	0.008344175
cg25697314	12	66905823	0.008344175
cg11465372	12	51457458	0.010620314
cg21922731	12	55443397	0.010620314
cg14162076	12	8556359	0.010620314
cg03986640	12	55134622	0.010620314
cg11051139	12	50866695	0.013511376
cg03014680	12	10013789	0.013511376
cg02554564	12	5473392	0.013511376
cg23696712	12	51514408	0.016932758
cg13446622	12	55614673	0.016932758
cg25372195	12	53328560	0.021130572
cg00495491	12	26714080	0.021130572
cg07017706	12	51299598	0.021130572
cg20200335	12	51253821	0.021130572
cg12572827	12	15016776	0.024548753
cg09418321	12	4569879	0.024548753
cg12650635	12	48630730	0.024548753
cg15711744	12	47152405	0.030265142
cg13053396	12	7038806	0.030265142
cg03490200	12	8866463	0.030265142
cg23767977	12	51233732	0.030265142
cg01007201	12	8166472	0.030265142
cg24664957	12	121946956	0.036908993
cg13928961	12	51298276	0.036908993
cg19824441	12	55674579	0.036908993
cg15569340	12	7173757	0.036908993
cg03064067	12	83831047	0.036908993
cg06270401	12	4569346	0.036908993
cg03538436	12	116283753	0.044803416
cg04303901	12	96410443	0.044803416
cg04375036	12	109666202	0.044803416
cg14338062	12	56169546	0.044803416
cg20870559	12	111900901	0.044803416

Experimental ID	Chromosome	MVP	P value
cg12682367	13	23421698	0.006405205
cg27065979	13	51632135	0.008344175
cg11635563	13	43139849	0.008344175
cg12071073	13	51334282	0.013511376
cg05823029	13	113360759	0.021130572
cg15648315	13	112349637	0.021130572
cg24821554	13	50537954	0.021130572
cg16194715	13	36351926	0.021130572
cg07744166	13	26029484	0.030265142
cg09878269	13	112860525	0.030265142
cg10246520	13	52499945	0.036908993
cg24222324	13	42043414	0.036908993
cg03721967	13	113057721	0.036908993
cg22539738	13	44460280	0.044803416
cg18960218	14	22355205	0.000880441
cg08495878	14	94097612	0.001584793
cg26191951	14	20493306	0.004915793
cg03770548	14	22594880	0.004915793
cg24812523	14	31868513	0.006405205
cg24526899	14	53493899	0.010620314
cg02507952	14	105461686	0.013511376
cg19042947	14	94097194	0.016932758
cg14541311	14	22946970	0.021130572
cg05788638	14	93828773	0.024548753
cg26514492	14	104602938	0.030265142
cg02681442	14	28305759	0.030265142
cg05485062	14	94053616	0.036908993
cg27504299	14	95221862	0.036908993
cg16937611	14	94117855	0.036908993
cg09721659	14	80491864	0.044803416
cg22090592	14	104245687	0.044803416
cg00174500	14	22916319	0.044803416
cg24355048	14	24114868	0.044803416
cg01429391	14	22974770	0.044803416
cg06319346	15	28474184	0.001155578
cg12703269	15	75073287	0.004915793
cg22563815	15	76644004	0.006405205
cg27257408	15	76313891	0.010620314
cg14694011	15	49418316	0.013511376
cg10173075	15	87565840	0.013511376
cg08433095	15	53588361	0.016932758
cg25003924	15	50649046	0.021130572
cg20225681	15	63290500	0.024548753
cg02982734	15	21442141	0.030265142
cg11540997	15	43195110	0.030265142

Experimental ID	Chromosome	MVP	P value
cg26704579	15	72805743	0.044803416
cg09207718	15	72828439	0.044803416
cg07888040	15	49456818	0.044803416
cg21917349	15	27001152	0.044803416
cg04731384	16	88604267	0.003714358
cg07842062	16	3246525	0.003714358
cg27298262	16	88603817	0.003714358
cg13897627	16	47935998	0.004915793
cg06810461	16	66075645	0.006405205
cg12564453	16	55553341	0.006405205
cg12125117	16	56258962	0.008344175
cg21686987	16	73810474	0.008344175
cg25101056	16	82831498	0.010620314
cg03297731	16	30031794	0.010620314
cg15206445	16	88100813	0.010620314
cg01351032	16	10878692	0.010620314
cg23444894	16	19205244	0.013511376
cg21085768	16	2807089	0.013511376
cg06793062	16	74901369	0.013511376
cg14221831	16	65510215	0.016932758
cg02164442	16	31312602	0.016932758
cg06172871	16	70645745	0.016932758
cg26916607	16	55552271	0.021130572
cg24110063	16	31347253	0.024548753
cg14296374	16	82398640	0.030265142
cg16480209	16	56561961	0.030265142
cg11667117	16	86527463	0.036908993
cg07595943	16	82782402	0.036908993
cg23514672	16	10696145	0.036908993
cg07434382	16	68541873	0.036908993
cg24217877	16	47936080	0.036908993
cg14792480	16	65511758	0.036908993
cg24652919	16	3013192	0.044803416
cg12343082	16	86527646	0.044803416
cg23487586	16	83873424	0.044803416
cg16799087	16	88655862	0.044803416
cg02552572	16	55155247	0.044803416
cg11258532	16	65435374	0.044803416
cg06100324	16	751348	0.044803416
cg14893129	17	75766646	0.002091046
cg22264436	17	39191949	0.003714358
cg16826286	17	31290403	0.004915793
cg02580606	17	36780252	0.004915793
cg00597076	17	1342630	0.004915793
cg22392708	17	36347296	0.010620314

Experimental ID	Chromosome	MVP	P value
cg15005385	17	31650167	0.013511376
cg10315334	17	31231445	0.013511376
cg10057218	17	35327675	0.013511376
cg16806597	17	18891060	0.013511376
cg27622610	17	2976935	0.016932758
cg02784874	17	36982488	0.016932758
cg12954718	17	4972090	0.021130572
cg22377428	17	69875477	0.021130572
cg10213812	17	23874449	0.021130572
cg06144905	17	24393906	0.021130572
cg23674788	17	36876839	0.021130572
cg23276115	17	18891062	0.021130572
cg13337865	17	6675715	0.024548753
cg03330678	17	72827828	0.024548753
cg17839611	17	44641801	0.024548753
cg21230133	17	36807046	0.024548753
cg23355492	17	43288592	0.024548753
cg12380854	17	43262759	0.030265142
cg02593766	17	45963676	0.030265142
cg03742272	17	7931430	0.030265142
cg13098960	17	7434875	0.030265142
cg22467071	17	59817613	0.030265142
cg02756698	17	36276534	0.030265142
cg21636577	17	7062605	0.030265142
cg00689014	17	37122664	0.030265142
cg05766474	17	31332781	0.036908993
cg12894984	17	35077157	0.036908993
cg13053608	17	37599199	0.036908993
cg05187322	17	75766923	0.036908993
cg08696192	17	35910836	0.036908993
cg19056418	17	73638812	0.036908993
cg24747122	17	60484391	0.044803416
cg06799664	17	74078716	0.044803416
cg16678925	17	3047722	0.044803416
cg22182666	17	53760154	0.044803416
cg04882759	17	12510626	0.044803416
cg10539808	18	22474735	0.003714358
cg11435943	18	59373667	0.013511376
cg18121684	18	59405541	0.016932758
cg24691255	18	59704946	0.030265142
cg18382305	18	29275063	0.030265142
cg13445249	18	27211109	0.044803416
cg10533434	18	59480206	0.044803416
cg26026726	19	40554254	0.00066033
cg13407883	19	56319655	0.00066033

Experimental ID	Chromosome	MVP	P value
cg10126923	19	56567263	0.002829987
cg02903525	19	52505329	0.002829987
cg10305797	19	40673517	0.004915793
cg09018862	19	47574484	0.006405205
cg26136776	19	12859426	0.008344175
cg05348870	19	6622045	0.008344175
cg14366490	19	17432825	0.008344175
cg13434203	19	47936342	0.013511376
cg24654350	19	60020679	0.013511376
cg23264413	19	48402117	0.013511376
cg24734575	19	38052466	0.016932758
cg12387247	19	7672974	0.016932758
cg08475088	19	60941103	0.016932758
cg08539991	19	40895672	0.016932758
cg16069910	19	17432528	0.016932758
cg10280342	19	6327576	0.021130572
cg07425555	19	13910018	0.021130572
cg18967533	19	56164537	0.021130572
cg10857774	19	50978523	0.024548753
cg15586352	19	48382983	0.024548753
cg14757492	19	18890194	0.024548753
cg04000821	19	59705878	0.024548753
cg02981853	19	7351850	0.030265142
cg18493147	19	425445	0.030265142
cg25221254	19	6284547	0.030265142
cg27519373	19	62042104	0.030265142
cg02686662	19	243245	0.036908993
cg25187533	19	48278627	0.036908993
cg21930712	19	56452928	0.036908993
cg11170796	19	1601224	0.036908993
cg04349727	19	56149201	0.036908993
cg19776453	19	47725641	0.036908993
cg25033144	19	59740102	0.036908993
cg22039846	19	59973065	0.044803416
cg09395833	19	15097755	0.044803416
cg05241571	19	40673064	0.044803416
cg14333565	19	5775065	0.044803416
cg18971671	19	54093533	0.044803416
cg11159299	19	494432	0.044803416
cg06123346	19	40746359	0.044803416
cg26149550	19	56026308	0.044803416
cg25843439	19	15713574	0.044803416
cg16101800	20	55569975	0.001584793
cg09868035	20	61962518	0.002829987
cg26206598	20	46878839	0.002829987

Experimental ID	Chromosome	MVP	P value
cg09076077	20	58063710	0.002829987
cg23713742	20	33666135	0.003714358
cg02882813	20	23808078	0.004915793
cg11061975	20	1587325	0.006405205
cg24513045	20	62181963	0.010620314
cg20485165	20	43186463	0.013511376
cg21835643	20	43368775	0.016932758
cg20732137	20	61721671	0.016932758
cg26540515	20	844349	0.021130572
cg03658707	20	32312087	0.024548753
cg24122922	20	24397392	0.024548753
cg00138126	20	55720103	0.030265142
cg05932408	20	47618046	0.030265142
cg14271690	20	36643413	0.030265142
cg19241311	20	29492554	0.030265142
cg09458237	20	3660191	0.030265142
cg25718438	20	43177335	0.030265142
cg18462653	20	29441837	0.036908993
cg10938286	20	23754999	0.036908993
cg18223379	20	31083308	0.036908993
cg09276978	20	34428931	0.036908993
cg26873164	20	48032382	0.036908993
cg10099900	20	1041733	0.044803416
cg18034859	20	29870349	0.044803416
cg13265003	21	42792092	0.004915793
cg00895324	21	40160387	0.008344175
cg03801286	21	34806378	0.008344175
cg27456885	21	30894784	0.010620314
cg07658590	21	45788015	0.013511376
cg24423088	21	31107236	0.021130572
cg25608949	21	42689424	0.021130572
cg10569414	21	42315323	0.024548753
cg10719920	21	45549271	0.025629073
cg02066681	21	45648228	0.030265142
cg09964921	21	34807637	0.036908993
cg24523456	21	37285030	0.036908993
cg00516481	21	42946271	0.036908993
cg15536230	21	43809520	0.036908993
cg07374637	21	30797319	0.044803416
cg17398613	21	42792922	0.044803416
cg05615487	22	49523273	0.002091046
cg10073042	22	21303200	0.002829987
cg00556408	22	35830534	0.008344175
cg17505463	22	17160198	0.008344175
cg07986773	22	43937273	0.016932758

Experimental ID	Chromosome	MVP	P value
cg15303841	22	28163804	0.016932758
cg08088989	22	21204447	0.021130572
cg18365380	22	22971011	0.021130572
cg18294257	22	29197673	0.024548753
cg04541607	22	25343978	0.030265142
cg02763671	22	18483541	0.030265142
cg21271753	22	35829397	0.030265142
cg10490064	22	23947422	0.036908993
cg23173455	22	21318684	0.044803416
cg17568996	22	41158069	0.044803416
cg01474260	22	15453884	0.044803416
cg10127415	X	26120444	0.002091046
cg18787449	X	151670926	0.004915793
cg15408454	X	151617883	0.008344175
cg12391921	X	70438375	0.010620314
cg05772663	X	113724668	0.016932758
cg21489722	X	8961881	0.021130572
cg14942312	X	129346742	0.021130572
cg00596686	X	7147332	0.024548753
cg22762309	X	69198828	0.024548753
cg22495120	X	146870786	0.030265142
cg03791917	X	100527943	0.030265142
cg16022279	X	151837379	0.036908993
cg03291145	X	2994515	0.036908993
cg06220755	X	17730979	0.036908993
cg12728629	X	151673029	0.036908993
cg15741706	X	134134072	0.044803416
cg19481953	X	99785577	0.044803416
cg11110686	X	140819380	0.044803416
cg24489034	X	148485237	0.044803416
cg21825364	Y	14677489	0.00066033

Genome-wide data set (layer III) of all individuals included in the study have been submitted to the NCBI Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) under series accession numbers GSE 27899 (samples GSM688887–GSM688906).

Research

A functional methylome map of ulcerative colitis

Robert Häslér,^{1,5} Zhe Feng,^{1,5} Liselotte Bäckdahl,² Martina E. Spehlmann,¹ Andre Franke,¹ Andrew Teschendorff,² Vardhman K. Ramanan,³ Thomas A. Down,⁴ Gareth A. Wilson,² Andrew Feber,² Stephan Beck,^{2,6} Stefan Schreiber,^{1,6,7} and Philip Rosenstiel^{1,6,7}

¹Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, 24105 Germany; ²Medical Genomics, UCL Cancer Institute, University College London, London WC1E 6BT, United Kingdom; ³Blizard Institute, Barts and The London, Queen Mary University of London, London E1 2AT, United Kingdom; ⁴Wellcome Trust Cancer Research UK Gurdon Institute and Department of Genetics, University of Cambridge, Cambridge CB2 1QN, United Kingdom

The etiology of inflammatory bowel diseases is only partially explained by the current genetic risk map. It is hypothesized that environmental factors modulate the epigenetic landscape and thus contribute to disease susceptibility, manifestation, and progression. To test this, we analyzed DNA methylation (DNAm), a fundamental mechanism of epigenetic long-term modulation of gene expression. We report a three-layer epigenome-wide association study (EWAS) using intestinal biopsies from 10 monozygotic twin pairs ($n = 20$ individuals) discordant for manifestation of ulcerative colitis (UC). Genome-wide expression scans were generated using Affymetrix UG133 Plus 2.0 arrays (layer 1). Genome-wide DNAm scans were carried out using Illumina 27k Infinium Bead Arrays to identify methylation variable positions (MVPs, layer 2), and MeDIP-chip on Nimblegen custom 385k Tiling Arrays to identify differentially methylated regions (DMRs, layer 3). Identified MVPs and DMRs were validated in two independent patient populations by quantitative real-time PCR and bisulfite-pyrosequencing ($n = 185$). The EWAS identified 61 disease-associated loci harboring differential DNAm in *de* of a differentially expressed transcript. All constitute novel candidate risk loci for UC not previously identified by GWAS. Among them are several that have been functionally implicated in inflammatory processes, e.g., complement factor *CF1*, the serine protease inhibitor *SPINK4*, and the adhesion molecule *THY1* (also known as *CD90*). Our study design excludes nondisease inflammation as a cause of the identified changes in DNAm. This study represents the first replicated EWAS of UC integrated with transcriptional signatures in the affected tissue and demonstrates the power of EWAS to uncover unexplained disease risk and molecular events of disease manifestation.

[Supplemental material is available for this article.]

Ulcerative colitis (UC) represents one major subphenotype (OMIM 191390) of human inflammatory bowel disease (IBD, OMM 266600) and is characterized by chronic inflammation of the intestinal mucosa, exhibiting a continuous pattern in the affected tissue. In the past decades, the disease displayed a remarkably steep rise in incidence, which cannot be explained by genetic variants exclusively. Current estimations attribute ~16% of the disease heritability to identified genetic variants (Anderson et al. 2011), while environmental changes interacting with genetic predisposition (Hampe et al. 1999; Stoll et al. 2004; Franke et al. 2008, 2010) are discussed as major determinants for disease manifestation (Rosenstiel et al. 2009). This is in concordance with observations for other complex diseases (Manolio et al. 2009). However, neither the relapsing/remitting nature of the inflammatory process nor the delayed onset after several decades of health is understood. To address this gap, several studies have focused on the functional genomics of UC in order to obtain a high-resolution transcriptional picture of disease processes in the inflamed mucosa

(Dieckgraefe et al. 2000; Lawrance et al. 2001; Costello et al. 2005). Beyond germline DNA variants, epigenetic variants, e.g., DNA methylation (DNAm) and histone modifications, could modulate disease-relevant gene function (Petronis 2010). Epigenetic patterns are highly tissue-specific (Ramanan et al. 2008) and can be influenced by environmental factors (Mann et al. 2004; Heijmans et al. 2008). Recent technological advances have assessed the degree of methylation of specific CpG sites as regulators of disease-associated transcripts (van Overveld et al. 2003; Petronis 2010). Consequently, epigenetic modifications represent promising candidates for elucidating processes of disease manifestation beyond the identified risk loci. A genome-wide study assessing the role of DNAm in the primary diseased tissue of a chronic inflammatory disorder, such as UC, could therefore lead to signatures that capture the environmental influence on patients. A particularly powerful approach for addressing epigenetic mechanisms is the study of discordant monozygotic (MZ) twins (Bell and Spector 2011). Except for highly penetrant Mendelian disorders, MZ twins exhibit a broad range (2.9%–95%) of disease discordance for most complex diseases that cannot be explained by classical genetics (Nance 1977). A recent report showed a discordance rate of 84% in MZ twins for UC (Spehlmann et al. 2008). While, historically, twin-based studies have been employed for the identification of disease genes, recent interest in MZ twin studies has moved to mapping epigenetic changes in identical genomes that have precipitated into different phenotypes (Praga et al. 2005; Kaminsky et al. 2009; Javierre et al. 2010; Ramanan et al. 2011a).

⁵These authors shared first authorship.

⁶These authors shared senior authorship.

⁷Corresponding authors

E-mail: s.schreiber@ucl.ac.uk

E-mail: p.rosenstiel@ucl.ac.uk

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/g.138847.112>. Freely available online through the Genome Research Open Access option.

A functional methylome map of ulcerative colitis

Starting with a collection of mucosal biopsies from 20 MZ twins, discordant for UC, we aimed to generate a genome-wide functional epigenetic map in dependent of the genetic background by combining three essential layers of information: (1) Genome-wide mRNA expression profiling; (2) genome-wide quantification of methylation variable positions (MVPs) representing individual methylation events within the proximal promoter regions of transcription start sites; and (3) genome-wide assessment of differentially methylated regions (DMRs) representing group effects of linked MVPs. Combining these three layers and validation of selected findings in an independent cohort of unrelated individuals illustrates the potential impact of functional epigenetics on disease mechanisms in UC.

Results

Three-layer genome-wide scans for differential gene expression, MVPs, and DMRs in twins, discordant for UC

The transcriptome analysis (layer #1) of the mucosa of 20 MZ twins, discordant for UC, identified 18,097 out of 54,675 transcripts analyzed as expressed. Of these, 361 were significantly differentially expressed (Benjamini-Hochberg-corrected P -value ≤ 0.05 ; FDR [false discovery rate] $\leq 5\%$) and 356 were in close proximity (50 kb) of at least one MVP or DMR. The MVP analysis (layer #2) involved interrogation of 27,578 CpG sites, of which 23,085 were informative, resulting in the identification of 703 MVPs between healthy and diseased individuals (Benjamini-Hochberg-corrected P -value ≤ 0.05). The DMR analysis (layer #3) involved the examination of 392,750 positions, of which 389,359 were informative in the mucosal tissue sample, resulting in the identification of 345 DMRs between healthy and diseased individuals (Benjamini-Hochberg-corrected P -value ≤ 0.05).

A genome-wide functional epigenetic map for UC

Integration of the three layers illustrates that epigenetic differences as well as their potential transcriptional consequences can be monitored in complex primary tissue on a genome-wide scale (Fig. 1; high resolution maps of individual chromosomes, see Supplemental Fig. S1A-X). The observed DNAm follows the expected bimodal distribution (Supplemental Fig. S2). Hypomethylation is prominent in promoter regions (Supplemental Fig. S3A), while hypermethylation is more frequent in gene-introns (Supplemental Fig. S3B). Similar patterns of variability were observed for all three layers, as documented by the intra-class correlations (Supplemental Fig. S4). The principal component analysis identified disease as the strongest component. Notably, mRNA and DNAm did not contribute equally to the separation between healthy and diseased individuals (Supplemental Fig. S5). Collectively, the three-layer analysis identified 61 disease-associated genes that were defined by differential expression and at least one MVP or DMR within a *cis*-interaction window of 50 kb (Supplemental Table S1) from the transcription start site. Disease-associated transcripts were defined by their significantly differential mRNA expression when comparing healthy individuals to UC patients (Benjamini-Hochberg-corrected P -value ≤ 0.05 ; false discovery rate $\leq 5\%$). Unsupervised clustering of these 61 mRNA/DNAm pairs revealed patient similarities of mRNA and DNAm levels (Supplemental Fig. S6). Correlation analysis of quantitative expression values and MVP or DMR data revealed a correlation in selected candidate transcripts (median Spearman- ρ $r = 0.58$), a correlation in identified

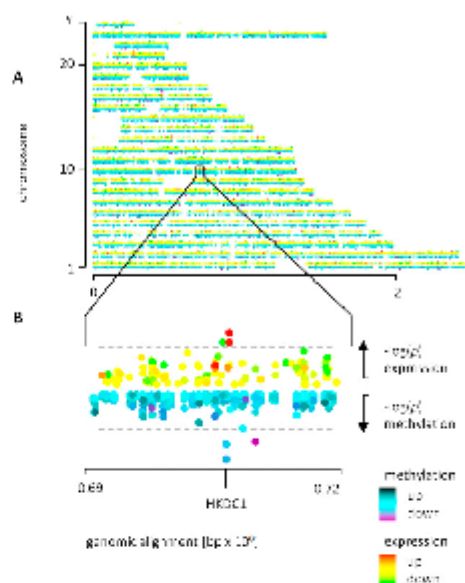


Figure 1. (A) Genome-wide profiles of DNAm and gene expression. The x-axis represents the genomic location with chromosomes represented as rows. The y-axis represents the significance of the differences between UC patients and healthy individuals and is displayed as $-\log(p)$ for the mRNA and as $\log(p)$ for the epigenetic modifications. Effect size (induction/fold change) is encoded by colors. Methylation: (black) up-regulated, (blue) not regulated, (purple) down-regulated; mRNA-expression: (red) up-regulated, (yellow) no regulation, (green) down-regulated. (B) Higher resolution map for selected candidate locus *HKDC1* (display and color coding according to Fig. 1A; the dotted line indicates the significance threshold). A higher resolution of this map is provided in Supplemental Figure S1A-X.

candidate loci ($r = 0.49$), no correlation in the group of disease-associated transcripts ($r = 0.15$), and no correlation when analyzing all transcripts ($r = 0.16$). The correlation within the selected candidates was significantly different from all other gene-sets (Supplemental Fig. S7). A large number of differentially expressed transcripts (independent of MVPs and DMRs) were associated with inflammation, immune response, and other disease-relevant processes when performing a gene ontology (GO) analysis. Similarly, differentially expressed transcripts in close proximity to MVPs and/or DMRs were also found to be associated with inflammation and immune response (Supplemental Fig. S8). The variant patterns observed in the validation panel displayed high robustness, as documented by the frequencies of their respective occurrences: A random individual from the validation panel has a median chance of over 80% of exhibiting the variant pattern described for the 10 candidate loci (Supplemental Fig. S9).

A direct comparison of the results for mRNA, MVP, and DMR levels to 47 previously published UC-associated loci (Anderson et al. 2011) revealed that none of those loci exhibited significant alterations for any of the three layers investigated (Supplemental Table S2).

Häsler et al.

Validation of selected candidate transcripts under potential epigenetic control

Of the 61 disease- and MVP/DMR-associated transcripts, 10 candidates with the closest proximity between MVP/DMRs and transcript location were subjected to further validation and replication in a larger collection of primary tissue from the intestinal mucosa (Table 1, validation panel I and II). This was performed by pyrosequencing (bisulfite-modified DNA, $n = 50$ individuals) for methylation analysis and TaqMan-based real-time PCR ($n = 135$ individuals) for differential mRNA expression. Significance criteria were defined in concordance with the initial scan (Benjamini-Hochberg-corrected P -value ≤ 0.05) (details on the panel of patients and controls are listed in Table 1; for detailed description of the methods, see Supplemental Material). The results are summarized in Table 2. We selected distinct patterns of differential DNAm/expression pairs as paradigms for different scenarios of epigenetic modulation while focusing on the commonly accepted pattern of hypermethylation with transcriptional repression and hypomethylation to be associated with increased mRNA expression. (1) Pairs with negative correlation of DNAm and mRNA expression, represented by 40 MVPs and five transcripts: increased DNAm accompanied by decreased mRNA expression (*MTIH*) and decreased DNAm with increased mRNA expression (*CFI*, *HKDC1*, *SPINK4*, *THY1*); (2) pairs with positive correlation of DNAm and mRNA expression, represented by 21 MVPs and three transcripts: increased DNAm accompanied by increased mRNA expression (*TK1*), and decreased DNAm with decreased mRNA expression (*FLNA*, *PTN*). Control loci were chosen to replicate MVPs/DMRs without a *cis*-modulation of mRNA expression (*SLC7A7*, 4 MVPs) or differential mRNA expression without multiple MVPs or DMRs (*IGHG1*, 1 MVP). We found MVPs and DMRs to be associated with alterations in expression levels occurring both at the start of a CpG island (*HKDC1*, *TK1*) as well as at the end of a CpG island (all other candidates), generally referred to as CpG island shores (Doi et al. 2009; Irizarry et al. 2009). Results for one selected example, representing a validated disease-associated transcript, is shown in Figure 2 (*HKDC1*). A substratification analysis performed on validation panel I and II revealed no influence of the potential confounding factors such as age, gender, biopsy location, and medication on the main findings (Supplemental Fig. S10, mRNA; Supplemental Fig. S11, DNAm).

Combining the results of the three-layer scan and validation showed 148-fold more MVPs in proximity to the 10 validated loci than expected by chance, corresponding to a P -value of 9.11×10^{-44} (Fisher's exact test).

Discussion

In ulcerative colitis, the contribution of genetic variants to the disease risk is estimated at $\sim 10\%$ (Anderson et al. 2011). Together with other factors (Hampe et al. 1999; Stoll et al. 2004; Franke et al. 2008, 2010), epigenetic modifications may explain parts of the remaining disease risk. While disease-relevant epigenetic modifications as well as their interaction with environmental factors has been shown for other diseases (van Overveld et al. 2003; Mann et al. 2004; Heijmans et al. 2008; Rakyan et al. 2008, 2011a; Petronis 2010), such effects have not been shown in ulcerative colitis. Therefore, this study aimed to create a functional epigenetic map by combining genome-wide data of differentially methylated regions, methylation variable positions, and transcriptome data, reflecting potential effects of the epigenetic variations. A group of 20 monozygotic twins, discordant for UC, was selected as an entry level to perform a three-layer genome-wide scan to target mechanisms that show significant differences between healthy and diseased individuals. The validity of such an approach has been recently demonstrated in psoriasis and lung adenocarcinoma (Gervin et al. 2012; Selamat et al. 2012). Obviously, such approaches cannot demonstrate the causality of epigenetic modifications for altered mRNA levels. In addition, the features measured by the systems employed restrict the number of detectable interactions between DNAm and mRNA expression. Keeping these limitations in mind, the presented identification of 61 *cis*-links between epigenetic modifications and disease-associated transcripts supports the hypothesis that pathophysiological events are a reflection of—and potentially controlled by—epigenetic modifications with consequences on transcriptional changes. The identified *cis*-links show both negative and positive correlations between DNAm and RNA expression, which is in concordance with two recent studies applying similar approaches (Gervin et al. 2012; Selamat et al. 2012), yet the nature of positive correlations remains unexplained by the current understanding of epigenetic regulation. The majority of identified loci are associated with biological processes which are either directly or indirectly linked to immune processes, which is consistent with previous findings on functional genomics of UC (Dieckgraefe et al. 2000; Lawrance et al. 2001; Costello et al. 2005), as well as with findings on UC genetics (Anderson et al. 2011). Interestingly, a smaller proportion of biological processes identified are distinct from bona fide immune processes: Structural, developmental, and metabolic processes are prominent findings. This suggests additional levels of transcriptional control and is concordant with a previous study showing that immune response-associated biological processes are not primarily under genetic control in mucosal tissue (Häsler et al. 2009).

Table 1. Study panels used for genome-wide assessment of differential expression and DNAm as well as for validation of initial findings

Panel type	Application	Relationship of individuals	Gender representation	Age: Median, range	Disease representation
Screening panel	Genome-wide screening of the transcriptome and methylome	Discordant twins (UC, HN)	10 f 10 m	25, 18–70	10 UC 10 NC
Validation panel I	Validation of the findings in the transcriptome	Unrelated individuals	69 f 66 m	41, 18–76	30 UC; 30 UCn/ 15 DC; 30 DCn/ 30 NC
Validation panel II (\subseteq panel I)	Validation of the findings in the methylome	Unrelated individuals	25 f 25 m	41, 18–68	20 UC 10 DC 20 NC

(f) Female, (m) male, (UC) ulcerative colitis, (NC) normal controls, (DC) disease specificity controls; suffix: (i) inflamed, (ni) not inflamed.

A functional methylome map of ulcerative colitis

Table 2. Validation of differential DNAm linked to disease-associated transcripts by TaqMan real-time PCR (mRNA expression in $n = 135$ individuals, validation panel I) and pyrosequencing (CpG methylation in $n = 50$ individuals, validation panel II)

Gene symbol; name	Prominent gene ontology terms (selected)	Cyto-genetic band	CpG-position relative to transcript	CpG methylation P-value (observations per gene)	mRNA regulation fold-change; P-value
<i>CFI</i> ; complement factor I	Complement activation; innate immune response	4q25	Post-TSS, SCB	$\downarrow 1.51 \times 10^{-05}$ (2)	$\uparrow 5.71$; 4.45×10^{-09}
<i>FLNA</i> ; Flamin A, alpha	Positive regulation of I- κ B kinase/NF- κ B cascade	Xq28	Post-TSS, SCB	$\downarrow 1.63 \times 10^{-02}$ (2)	$\downarrow -2.85$; 7.72×10^{-04}
<i>HKDC1</i> ; hexokinase domain containing 1	Glycolysis	10q22.1	Pre-TSS and post-TSS	$\downarrow 7.94 \times 10^{-06}$ (6)	$\uparrow 2.01$; 5.16×10^{-04}
<i>IGHG1</i> ^a ; immunoglobulin heavy constant gamma 1 (Cl m marker)	immune response; antigen binding	14q32.33	Post-TSS, SCB	$(-) 1.56 \times 10^{-01}$ (0)	$\uparrow 8.25$; 8.90×10^{-09}
<i>MTTH</i> ; metallothionein 1H	Metal ion binding; protein binding	16q13	Post-TSS	$\uparrow 1.50 \times 10^{-02}$ (2)	$\downarrow -1.66$; 3.04×10^{-05}
<i>PTN</i> ; pleiotrophin	Positive regulation of cell proliferation	7q33	Post-TSS	$\downarrow 2.42 \times 10^{-02}$ (1)	$\downarrow -4.16$; 2.76×10^{-04}
<i>SLC7A7</i> ^b ; solute carrier family 7 (cationic amino acid transporter, y+ system), member 7	Blood coagulation, leukocyte migration	14q11.2	Post-TSS	$\downarrow 5.21 \times 10^{-06}$ (4)	$(-) 1.16$; 6.73×10^{-01}
<i>SPINK4</i> ; serine peptidase inhibitor, Kazal type 4	Serine-type endopeptidase inhibitor activity	9p13.3	Post-TSS, SCB	$\downarrow 3.03 \times 10^{-06}$ (2)	$\uparrow 13.05$; 4.45×10^{-09}
<i>THY1</i> ; Thy-1 cell surface antigen	Positive regulation of t cell activation	11q23.3	Post-TSS	$\downarrow 2.28 \times 10^{-01}$ (2)	$\uparrow 2.99$; 4.75×10^{-07}
<i>TK1</i> ; thymidine kinase 1, soluble	DNA replication	17q23.2-q25.3	Pre-TSS	$\uparrow 3.08 \times 10^{-01}$ (2)	$\uparrow 1.32$; 2.30×10^{-02}

(TSS) Transcription start site; (SCB) spanning to gene body; in case of multiple CpGs, all relative positions are described. (†) Up-regulation; (‡) down-regulation; (-) no regulation as measured in the validation panel. P-values presented were corrected for multiple testing using the Benjamini-Hochberg correction. In case of multiple observations, the lowest P-values were presented.

^aNegative control candidate for differential expression without epigenetic modification.

^bNegative control candidate for epigenetic modification without differential mRNA expression.

The design used here allowed direct comparison of individual MVPs and DMRs. While DMRs may be expected to have greater functional relevance than MVPs, this has not yet been formally demonstrated by any study. Consequently, both should be part of an epigenetic map to ensure the inclusion of all potentially relevant information. Based on the assessment of 23,085 CpG sites in our study, no single MVP was found that could serve as representative for an entire DMR. In addition, highly significant MVPs were found in loci not showing any differential gene expression. Together with our observation that a variable number of MVPs (in the case of our validated MVP/DMR-transcript pairs: 1-6) may potentially influence mRNA expression, these data provide no evidence that individual MVPs are of higher relevance than DMRs or vice versa.

Previous studies have addressed the epigenetic modulation in the context of IBD mostly from a candidate gene approach, focused on IBD-associated colorectal cancer (Moriyama et al. 2007; Dhir et al. 2008; Edwards et al. 2009; Gonsky et al. 2011). Together with the high tissue specificity (Rakyan et al. 2008) and the fundamental differences in detection, these studies are only of limited use for comparison to our systematic genome-wide approach. However, our results support one recent report stating that epigenetic dysregulation of the *IRF5* promoter is not associated with IBD (Balasa et al. 2010). In contrast to that, many of our findings on the mRNA level are in concordance with previously published studies (Dieckgraefe et al. 2000; Lawrance et al. 2001; Anderson et al. 2011), potentially attributed to the lower technical and/or biological variance in inflammation-associated mRNA patterns. Among the replicated findings, several genes have been directly associated with chronic intestinal inflammation; *PTN* (pleiotrophin) has been shown to be functionally linked to inflammation and cancer

(Kadomatsu 2005), while *THY1* (Thy-1 cell surface antigen) mediates cell adhesion during inflammation (Jurisic et al. 2010). The serine protease inhibitor *SPINK4* has been shown to be differentially expressed in chronic autoimmune intestinal inflammation (celiac disease), likely derived from altered goblet cell activity, while no causative genetic variant was identified (Wapenaar et al. 2007).

The ontogenetic stability of DNAm over time cannot be assessed with our study design; however, our findings in the enlarged validation and replication panel of unrelated patients and controls document the relevance of our results for a more heterogeneous population. It is important to note that we examined the disease specificity of our results by including inflammatory disease controls (non-IBD-associated inflammation in the sigmoidal mucosa) in our validation panels. None of the validated candidate genes showed DNAm variation together with altered transcript levels when comparing diseased controls to healthy individuals. A significantly altered DNAm between inflammatory disease controls and healthy individuals was only found for the *SLC7A7* locus; however, this was not accompanied by altered mRNA levels. These results corroborate the potential disease specificity of the identified effects and further support the concept of a distinct impairment of mucosal homeostasis in UC that is not common to intestinal inflammation in general.

Interestingly, all of the identified disease-associated alterations on mRNA, MVP, or DMR levels are novel as different to the 47 previously published UC risk loci identified by GWAS (Anderson et al. 2011). None of the UC risk loci are considered to be methylation quantitative trait loci (Zhang et al. 2010). Furthermore, none of the risk loci are located in a regulatory region potentially affecting mRNA expression. In addition, the use of monozygotic twins for the discovery phase favors identification of candidate variants that

Häsler et al.

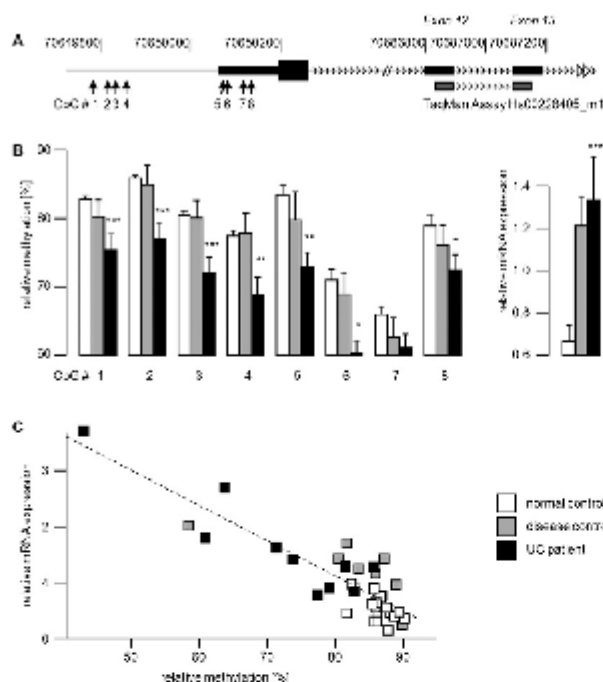


Figure 2. Validation of DNAm and gene expression in selected candidate transcripts example: HKDC1. (A) Relative position of CpGs (arrows up, continuously numbered 1–8) and real-time PCR probe (dark gray). (B) Quantitative results of the validation: (1) methylation (via pyrosequencing, CpGs continuously numbered 1–8 in concordance with Fig. 2A) in $n = 50$ individuals (validation panel I); the order of assays displayed corresponds to the order in Figure 2A; (2) mRNA (via real-time PCR, right) in $n = 135$ individuals (validation panel I). (C) Correlation between mRNA expression (x-axis) and relative methylation (y-axis) in HKDC1 for a selected CpG; the dotted line represents the correlation trend (Spearman- $r = -7.43$). Significant differences between UC patients and normal controls are indicated by asterisks (* $P \leq 5 \times 10^{-2}$; ** $P \leq 5 \times 10^{-4}$; *** $P \leq 5 \times 10^{-6}$).

are independent of genetic effects. This also reflects fundamental differences between GWAS and EWAS approaches, as outlined recently (Radyan et al. 2011b): EWAS, especially when conducted in monozygotic discordant twins and linked to transcriptome analysis, represents a powerful complementary tool to detect variations which cannot be captured by a GWAS.

Finally, the potential clinical relevance of our findings is supported by the high frequencies at which these effects occur: In 93% of all individuals, decreased methylation at these candidate loci is reflected by up-regulation of the corresponding transcript (or vice versa) when comparing UC patients to healthy controls. Moreover, we observed ~150-fold more MVPs in the proximity to the validated candidate loci than expected by chance. This also indicates a high robustness of cis-links between epigenetic modifications and regulation of disease-associated transcripts.

In conclusion, our results indicate that changes in DNAm and their consequences on the transcriptome may represent disease mechanisms for UC, independent of genetic variation. The use of primary tissue from monozygotic disease-discordant twins followed by the validation of the findings in a larger and independent panel

of UC patients supports the potential disease relevance of our observations. In addition, the disease specificity of the observed events is strengthened by our study design, which controlled for confounding non-disease-associated variation, e.g., due to inflammation. While our data are suggestive of a link between DNAm and its effects on mRNA expression, it remains a challenge to formally establish causality. Nevertheless, the integrated three-layer functional map reported here will contribute toward a better understanding of IBD pathophysiology, further closing the gap between unexplained disease susceptibility and disease manifestation.

Methods

Patient recruitment and patient characteristics

The twenty monozygotic twins, discordant for ulcerative colitis (screening panel; median age 25, range 18–70) recruited for this study were tested for mono/dizygosity as previously published (Barbaro and Cornaci 2004; von Wurmb-Schwark et al. 2005). The panel for real-time PCR validation of transcript levels consisted of 135 unrelated individuals (validation panel I; $n = 30$ ulcerative colitis, inflamed; $n = 30$ ulcerative colitis, noninflamed; $n = 30$ healthy individuals; $n = 15$ disease controls, inflamed; $n = 30$ disease controls, noninflamed; median age: 41, age range 18–76), a subgroup of which was used for pyrosequencing validation of methylation levels (validation panel II; $n = 20$ ulcerative colitis, inflamed; $n = 20$ healthy individuals, $n = 10$ disease controls, inflamed; median age: 41, age range 18–68). Healthy individuals included in the study were undergoing colonic cancer surveillance with no previous unspecific changes in stool habits, where endoscopic and histological examination yielded no significant pathological findings. Ulcerative colitis patients were selected to display an endoscopically active disease in the sigmoid colon at the time of sampling. Inflammation was assessed macroscopically during colonoscopy and categorized into (1) no signs of inflammation, (2) low inflammation, and (3) moderate/high inflammation, while only tissue with no or moderate/high inflammation was included in the study (see Supplemental Table S3). More than 2000 patients were screened to recruit the study population. Disease-specificity controls included individuals with infectious diarrhea, other forms of gastrointestinal inflammation or irritable bowel syndrome. The study setup was approved by the Bioethical Committee of the University of Kiel, where the patients were recruited. All patients gave written informed consent before data and biomaterials were collected. Patient characteristics are summarized in Table 1; a more detailed description of patient characteristics is presented in Supplemental Table S3 (A: screening panel; B: validation panel I; C: validation panel II).

A functional methylome map of ulcerative colitis

Biopsy processing and sample preparation

All biopsies used in this study are primary tissue from the intestinal mucosa. Biopsies were taken endoscopically from a defined area of the colon and immediately snap-frozen in liquid nitrogen. Total RNA was extracted and processed as previously described (Mah et al. 2004) and quality controlled using an Agilent Bioanalyzer according to the manufacturer's guidelines. DNA was extracted from biopsies using the QIAamp Tissue DNA preparation kit (Qiagen).

Expression profiling (mRNA, layer I)

Total RNA was prepared and hybridized to an Affymetrix UG 133 Plus 2.0 according to the manufacturer's protocol. Data was normalized using GCRMA (R, Bioconductor), and signals that were not present in at least 80% of the samples (cutoff: detection P -value ≤ 0.05) were excluded from further analysis. The experimental and analytical part of the microarray analysis was performed following the MIAME standards, including the data submission to Gene Expression Omnibus (GEO, URL: <http://www.ncbi.nlm.nih.gov/geo>, series: GSE22619; samples: GSM560961–GSM560976). Differentially expressed genes were determined using the Mann-Whitney U -test, multiple testing correction was performed using the Benjamini-Hochberg method (Benjamini and Hochberg 1995), and a false discovery rate for the signed fold changes (which were based on the ratios of the medians of each group compared) was estimated based on a Westfall and Young permutation, using $K = 5000$ permutations (Westfall and Young 1993). Criteria for transcripts to be categorized as differentially expressed were set to: (1) corrected P -value ≤ 0.05 , and (2) FDR $\leq 5\%$.

Analysis of methylation variable positions (MVPs, layer II)

DNA was prepared and analyzed using HumanMethylation27 BeadChips (iScan system, Illumina) as previously described (Van der Auwera et al. 2010), while the dataset was subject to intra-array and inter-array normalization as published earlier (Teschendorff et al. 2010). Differences between ulcerative colitis patients, diseased controls, and healthy individuals were determined using the Mann-Whitney U -test, while P -values were corrected according to Benjamini and Hochberg (Benjamini and Hochberg 1995). MVPs with a corrected P -value ≤ 0.05 were considered significantly differentially methylated.

Analysis of differentially methylated regions (DMRs, layer III)

DNA was prepared and hybridized to a custom tiling array (Nimblegen, custom 385k array) as previously described (Rakyan et al. 2008). The array was designed to cover known autoimmune/inflammatory-linked loci as well as specific genes with immune-regulatory function and encompassed all known promoters and CpG islands (both promoter- and non-promoter-CpG islands). Data was normalized applying the inter-quantile normalization using Spotfire for functional genomics (TIBCO). Differences between ulcerative colitis patients, diseased controls, and healthy individuals were determined using the Mann-Whitney U -test, while P -values were corrected according to Benjamini and Hochberg (Benjamini and Hochberg 1995). DMRs with a corrected P -value ≤ 0.05 were considered significantly differentially methylated.

Integration of three layers of genome-wide scans, candidate selection

To identify disease associated transcripts under potential epigenetic control, differentially expressed transcripts from layer I were

selected (corrected $P \leq 0.05$, FDR $\leq 5\%$, regulated between healthy and diseased individuals). The genomic transcript locations were used to generate interaction windows of 50 kb upstream of and downstream from the transcription start site (TSS). These windows were examined to see whether they contain either a DMR (layer II, corrected $P \leq 0.05$, between healthy and diseased individuals) or a MVP (layer III, corrected $P \leq 0.05$, between healthy and diseased individuals). Transcripts significantly regulated between healthy and diseased individuals, with a significantly regulated DMR or MVP within the 50-kb window of the TSS, were considered candidates for disease-associated transcripts under potential epigenetic control. Correlating quantitative expression values with DMRs and MVPs within this window was carried out using the Spearman- ρ correlation. Differences between sets of correlation (all genes, disease-associated transcripts, and validated transcripts) were assessed using the Mann-Whitney U -test.

Validation of differential mRNA expression via real-time PCR

Real-time PCR (TaqMan) was performed according to the manufacturer's guidelines (Applied Biosystems) using a 7900HT Real-Time PCR System. Expression levels were calculated relative to beta-actin using the standard-curve method (Livak and Schmittgen 2001). Differences between ulcerative colitis patients, diseased controls, and healthy individuals were determined using the Mann-Whitney U -test, while P -values were corrected according to Benjamini and Hochberg (Benjamini and Hochberg 1995).

Validation of differential DNAm via pyrosequencing

Validation of initial findings was performed via bisulfite conversion followed by pyrosequencing (Roche, 454) as previously described (Bollati et al. 2007). Differences between ulcerative colitis patients, diseased controls, and healthy individuals were determined using the Mann-Whitney U -test, while P -values were corrected according to Benjamini and Hochberg (Benjamini and Hochberg 1995).

Gene ontology analysis

Gene ontology analysis was performed as previously published (Tavazoie et al. 1999). Biological processes associated with the transcripts and candidate genes were retrieved from the Gene Ontology Consortium (www.geneontology.org).

Determination of effect frequencies

Frequencies of validated effects were determined by assessing the number of occurrences, where an effect was following the variation pattern observed when comparing the medians of ulcerative colitis versus healthy individual signals. An effect was considered significant when both differential DNAm and differential mRNA expression showed Benjamini-Hochberg-corrected P -values ≤ 0.05 in the validation experiment.

Data access

Genome-wide data sets (three layers) of all individuals included in the study have been submitted to the NCBI Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) under series accession numbers GSE22619 (samples: GSM560961–GSM560976) and GSE27899 (samples: GSM688887–GSM688926).

Häsler et al.

Acknowledgments

We thank Dorina Oelsner for technical assistance. Additionally, we thank Nicole von Wurmb-Schwark for carrying out the zygosity testing of the twins. This study was supported by the National German Genome Network, the SFB 415 Project Z1, the SFB877 B9 project, and the Excellence Cluster Inflammation at Interfaces. P.R. and S.S. are supported by the German Epigenome Project DEEP (BMBF). The Beck laboratory was supported by Wellcome Trust (084071), Royal Society Wolfson Research Merit Award (WM100023), and EU-FP7 project BLUEPRINT (282510). Finally, we thank all the patients, healthy volunteers, and physicians who took part in this study for their participation. We thank the three anonymous reviewers for excellent suggestions and constructive criticism.

Author contributions: R.H., Z.E., L.B., and M.E.S. collected the samples and conducted the experiments; R.H., Z.E., A.E., A.T., V.K.R., T.A.D., G.A.W., and A.E. analyzed the data; R.H., Z.E., S.B., S.S., and P.R. contributed to writing the manuscript; and S.B., S.S., and P.R. initiated, designed, and supervised the project.

References

- Anderson CA, Boucher G, Lees CW, Franke A, D'Aimato M, Taylor KD, Lee JC, Goyette P, Imielinski M, Latiano A, et al. 2011. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* **43**: 246–252.
- Balasa A, Gathungu G, Kistall P, Smith EO, Cho JH, Melegh B, Kelleymayer R. 2010. Assessment of DNA methylation at the intergenic regulatory factor 5 (RFX5) promoter region in inflammatory bowel diseases. *Int J Colorectal Dis* **25**: 553–556.
- Barbano A, Cormac P. 2004. DNA analysis from mixed biological materials. *Forensic Sci Int (Suppl)* **146**: S123–S125.
- Beilert S, Spector TD. 2011. A twin approach to unraveling epigenetics. *Trends Genet* **27**: 116–125.
- Benjamini Y, Hochberg Y. 1995. A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* **57**: 289–300.
- Bollati V, Baccaelli A, Hou L, Bonzini M, Rustioni S, Cavallo D, Byun H-M, Jiang J, Marinelli B, Pesatori AC, et al. 2007. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. *Cancer Res* **67**: 876–880.
- Costello CM, Mah N, Häsler R, Rosenstiel P, Wetzig GH, Hahn A, Lu T, Gurbuz V, Nikolaus S, Albrecht M, et al. 2005. Dissection of the inflammatory bowel disease transcriptome using genome-wide cDNA microarrays. *PLoS Med* **2**: e199. doi: 10.1371/journal.pmed.0020199.
- Dhir M, Montgomery EA, Götzker SC, Schaefer KE, Hooker CM, Herman JC, Bayliff SB, Gearhart SL, Ahuja N. 2008. Epigenetic regulation of WNT signaling pathway genes in inflammatory bowel disease (IBD) associated neoplasia. *J Gastrointest Surg* **12**: 1745–1753.
- Dieckmeyer BK, Stenson WE, Korzenik JR, Swanson PE, Harrington CA. 2000. Analysis of mucosal gene expression in inflammatory bowel disease by parallel oligonucleotide arrays. *Physiol Genomics* **4**: 1–11.
- Doi A, Park IH, Wen B, Munkami P, Auye MJ, Irizarry R, Herb R, Ladd-Acosta C, Rho J, Loefer S, et al. 2009. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* **41**: 1350–1353.
- Edwards RA, Witherspoon M, Wang K, Afshari K, Pham T, Birnbaumer I, Lipkin SM. 2009. Epigenetic repression of DNA mismatch repair by inflammation and hypoxia in inflammatory bowel disease-associated colorectal cancer. *Cancer Res* **69**: 6423–6429.
- Fraga MF, Ballestar F, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, et al. 2005. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* **102**: 10604–10609.
- Franke A, Balschun T, Karlsen TH, Sventoraitis J, Nikolaus S, Mayr G, Domingues B, Albrecht M, Northrup M, Ellinghaus D, et al. 2008. Sequence variants in *IL10*, *ARPC2* and multiple other loci contribute to ulcerative colitis susceptibility. *Nat Genet* **40**: 1319–1323.
- Franke A, Balschun T, Sina C, Ellinghaus D, Hildner R, Mayr G, Albrecht M, Wittig M, Buchert E, Nikolaus S, et al. 2010. Genome-wide association study for ulcerative colitis identifies risk loci at 7q22 and 22q13 (*IL17REL*). *Nat Genet* **42**: 292–294.
- Gervin K, Vigeland MD, Mattingsdal M, Hammer M, Nygård H, Olsen AO, Brandt I, Harris JR, Undlien DE, Lyle R. 2012. DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: Identification of epigenetically dysregulated genes. *PLoS Genet* **8**: e1002454. doi: 10.1371/journal.pgen.1002454.
- Gonsky R, Deem RL, Landers CJ, Derkowski CA, Berel D, McGovern DPB, Targan SR. 2011. Distinct IFNG methylation in a subset of ulcerative colitis patients based on reactivity to microbial antigens. *Inflamm Bowel Dis* **17**: 171–178.
- Hampe J, Schreiber S, Shaw SH, Lau KF, Bridger S, Macpherson AJ, Cardon LR, Sakai H, Harris TJ, Buckler A, et al. 1999. A genome-wide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am J Hum Genet* **64**: 808–816.
- Häsler R, Begun A, Freitag-Wolf S, Kerick M, Mah N, Zvirbilene A, Spehlmann ME, von Wurmb-Schwark N, Kupcinskas I, Rosenstiel P, et al. 2009. Genetic control of global gene expression levels in the intestinal mucosa: A human twin study. *Physiol Genomics* **38**: 73–79.
- Heijmans BT, Tobin EW, Stein AD, Putter H, Blauw GJ, Sussner ES, Slagboom PE, Lumey LH. 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A* **105**: 17046–17049.
- Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Hongmei M, Webster M, et al. 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* **41**: 178–186.
- Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, Bastasco M, Fraga Mario F, O'Hanlon TP, Rider LG, et al. 2010. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res* **20**: 170–179.
- Juricic G, Ilyeva M, Pasule ST, Halin C, Detmar M. 2010. Thymus cell antigen 1 (Thy1, CD90) is expressed by lymphatic vessels and mediates cell adhesion to lymphatic endothelium. *Exp Cell Res* **316**: 2982–2992.
- Kadomatsu K. 2005. The midline family in cancer, inflammation and neural development. *Nagoya J Med Sci* **67**: 71–82.
- Kamirsky ZA, Tang T, Wang S-C, Prisk C, Oh GHT, Wong AHC, Feldcamp LA, Vitman C, Halverson J, Tyak C, et al. 2009. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet* **41**: 240–245.
- Lawrance IC, Koch HC, Chakravarti S. 2001. Ulcerative colitis and Crohn's disease: Distinctive gene expression profiles and novel susceptibility candidate genes. *Hum Mol Genet* **10**: 445–456.
- Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔC_T} method. *Methods* **25**: 402–408.
- Mah N, Thelin A, Lu T, Nikolaus S, Rühbacher T, Gurbuz V, Eickhoff H, Koppel G, Lehrach H, Mellgard B, et al. 2004. A comparison of oligonucleotide and cDNA-based microarray systems. *Physiol Genomics* **16**: 361–370.
- Mann MRW, Lee SS, Doherty AS, Verona RI, Nolen LD, Schultz RM, Batolometti MS. 2004. Selective loss of imprinting in the placenta following preimplantation development in culture. *Development* **131**: 3727–3735.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Caudon Lon R, Chakravarti A, et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461**: 747–753.
- Motiyama T, Matsumoto T, Nakamura S, Jo Y, Mibu R, Yao T, Iida M. 2007. Hypermethylation of p14 (ARF) may be predictive of colitic cancer in patients with ulcerative colitis. *Dis Colon Rectum* **50**: 1384–1392.
- Nance WE. 1977. The use of twins in clinical research. *Birth Defects Orig Artic Ser* **13**: 19–44.
- Petronis A. 2010. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* **465**: 721–727.
- Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Gräf S, Tomazou EM, Bickelohr L, Johnson N, Heberth M, et al. 2008. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res* **18**: 1518–1529.
- Rakyan VK, Beyan H, Down TA, Hava M, Maitau S, Aden D, Danay A, Busato F, Mein CA, Munias B, et al. 2011a. Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet* **7**: e1002300. doi: 10.1371/journal.pgen.1002300.
- Rakyan VK, Down TA, Balding DJ, Beck S. 2011b. Epigenome-wide association studies for common human diseases. *Nat Rev Genet* **12**: 529–541.
- Rosenstiel P, Sina C, Franke A, Schreiber S. 2009. Towards a molecular risk map—recent advances on the etiology of inflammatory bowel disease. *Sozialwissenschaft* **2**: 334–345.
- Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M, Segmund KD, Kost MN, Hagen JA, Lam WL, et al. 2012. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Res* **22**: 1197–1211.
- Spehlmann ME, Begun AZ, Burghardt J, Lepage P, Raedler A, Schreiber S. 2008. Epidemiology of inflammatory bowel disease in a German twin cohort: Results of a nationwide study. *Inflamm Bowel Dis* **14**: 968–976.

A functional methylome map of ulcerative colitis

- Stoll M, Comelliusen B, Costello CM, Wietzig GH, Melgard B, Koch WA, Rosenstiel P, Albrecht M, Claucher FJP, Seeger D, et al. 2004. Genetic variation in *DLG5* is associated with inflammatory bowel disease. *Nat Genet* **36**: 476–480.
- Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. 1999. Systematic determination of genetic network architecture. *Nat Genet* **22**: 281–285.
- Teichendorf AF, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Nourmehr H, Bell CG, Maxwell AP, et al. 2010. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* **20**: 440–446.
- Van der Auwera I, Yu W, Siao L, Van Neste L, van Dam P, Van Maack EA, Pauwels P, Vermeulen PB, Dirix LY, Van Laere SJ. 2010. Array-based DNA methylation profiling for breast cancer subtype discrimination. *PLoS ONE* **5**: e12616. doi: 10.1371/journal.pone.0012616.
- van Overveld PGM, Lemmers RJJ, Sandkuijl LA, Erthoven L, Winokur ST, Bakels F, Padberg GW, van Ommen G-JB, Franis RR, van der Maarel SM. 2003. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat Genet* **35**: 315–317.
- von Wurmb-Schwark N, Malyusz V, Simeoni F, Lignitz E, Poesch M. 2005. Possible pitfalls in motherless paternity analysis with related putative fathers. *Forensic Sci Int* **159**: 92–97.
- Wapenaar MC, Monsuur AJ, Poell J, van 't Slot R, Meijer JWR, Meijer GA, Mulder CJ, Mearin ML, Wijmenga C. 2007. The SPINK gene family and celiac disease susceptibility. *Immunogenetics* **59**: 349–357.
- Westfall PH, Young S. 1993. *Resampling-based multiple testing*. Wiley-Interscience, New York.
- Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W, Craig DW, Redman M, Gershon ES, Liu C. 2010. Genetic control of individual differences in gene-specific methylation in human brain. *Am J Hum Genet* **86**: 411–419.

Received January 31, 2012; accepted in revised form July 11, 2012.

List of figures

Figure 1.	Temporal trends of incidence rates.....	5
Figure 2.	The T helper cell differentiation.	9
Figure 3.	Characteristics of epigenomes.....	12
Figure 4.	DNA Demethylation Pathways.....	16
Figure 5.	Flow diagram of genome-wide analysis	29
Figure 6.	A: Genome-wide profiles of DNAm and gene expression	39
Figure 7.	Intra Class Correlation.....	40
Figure 8.	Distribution of significant MVPs/DMRs in different genomic regions	41
Figure 9.	Principal component analysis of differential transcription and DNAm in UC	42
Figure 10.	Correlation analysis of mRNA expression and epigenetic modification.....	43
Figure 11.	Biological processes affected by differential DNAm or gene expression	44
Figure 12.	Validation of candidate transcripts, example: HKDC1.....	46
Figure 13.	Frequencies of validated effects.....	47
Figure 14.	Impact of potential confounding factors on the differential mRNA expression.	48
Figure 15.	Impact of potential confounding factors on the differential DNA methylation.....	49
Figure 16.	Hydroxymethylation levels in promoters of candidate loci.....	50
Figure 17.	Hydroxymethylation detection of the candidate promoters.....	51
Figure 18.	Potential DNA demethylation pathways.....	61

List of tables

Table 1.	<i>Diagnosis of ulcerative colitis (UC)</i>	7
Table 2.	<i>Study panels I</i>	21
Table 3.	<i>Study panels II</i>	24
Table 4.	<i>Study panels III</i>	30
Table 6.	<i>Component of the PCR reaction mix</i>	31
Table 7.	<i>The methylation candidates from both DMR data set and MVP data set</i>	33
Table 8.	<i>Study panels IV</i>	36
Table 9.	<i>Designed TaqMan assays</i>	36
Table 10.	<i>Validation of differential DNAm linked to disease-associated transcripts</i>	46
Table 11.	<i>dbSNP: previously published loci, associated to ulcerative colitis</i>	55

9 Curriculum Vitae

Personal Data

Name	Zhe Feng
Date of birth	02 nd .Oct.1978
Place of birth	Liaoyuan, China
Citizenship	Chinese
Telephone	0049 (0)431 21 91 965
Email	fengzhe1002@hotmail.com

Education

09.1985 - 07.1990	The second primary school, Liaoyuan, China
09.1990 - 07.1993	The third junior middle school, Liaoyuan, China
09.1993 - 07.1996	The fifth senior middle school, Liaoyuan, China

Study

09.1996 - 07.2001	Study of Medicine at Beihua University, Jilin, China
09.2002 - 07.2006	Study of Molecular Medicine at Friedrich-Alexander-University of Erlangen-Nuremberg, Germany
10.2006 - 01.2009	Study of Biomedicine at Medical High School Hanover, Germany
08.2009 - Today	PhD work at the Institute of Clinical Molecular Biology, Christian-Albrechts University of Kiel, Germany

Publications

1. Häslér R*, Feng Z*, Bäckdahl L, Spehlmann ME, Franke A, Teschendorff A, Rakyán VK, Down TA, Wilson GA, Feber A, Beck S, Schreiber S, Rosenstiel P. A functional methylome map of ulcerative colitis. *Genome Res.* 2012 Nov;22 (11):2130-7. doi: 10.1101/gr.138347.112. Epub 2012 Jul 23.
- 2[#]. Song P, Groos S, Riederer B, Feng Z, Krabbenhöft A, Manns MP, Smolka A, Hagen SJ, Neusch C, Seidler U. Kir4.1 channel expression is essential for parietal cell control of acid secretion. *J Biol Chem.* 2011 Apr 22;286 (16):14120-8. Epub 2011 Mar 2.

3[#]. Song P, Groos S, Riederer B, Feng Z, Krabbenhöft A, Smolka A, Seidler U. KCNQ1 is the luminal K⁺ recycling channel during stimulation of gastric acid secretion. J Physiol. 2009 Aug 1;587 (Pt 15):3955-65. Epub 2009 Jun 2.

Awards and honors

Chinese government award for outstanding self-financed students abroad 2012

*. Shared the first authorship.

#. not related to this thesis.

10 Acknowledgements

This study would have been impossible without the help of many competent and kind contributors. Therefore, I am very grateful to...

Prof. Dr. Stefan Schreiber & Prof.
Dr. Philip Rosenstiel for providing me with an attractive research project and excellent working conditions. Especially, I am grateful for their trust and encouragement.

Prof. Dr. Dr. Thomas C.G. Bosch for his support and supervision and the possibilities to present my work during the laboratory meetings of his workgroup.

Dr. Robert Häslar for his supervision throughout the three years and his introduction into epigenetics and genetics. His scientific work provided the basis of this project. Furthermore, his time-intensive investments into this thesis are gratefully acknowledged.

Dorina Oelsner for her expert help with real-time gene expression studies, and many other kinds of laboratory works.

Geetha Venkatesh for her kindly help with some special statistic analysis.

Prof. Dr. Stephan Beck, Dr. Liselotte Bäckdahl, Dr. Andrew Teschendorff, Dr. Vardhman K. Rakyan, Dr. Thomas A. Down, Dr. Gareth A. Wilson & Dr. Andrew Feber for their generating genome-wide scanning with Nimblegen and Illumina arrays as well as for the pre-analysis.

Manuela Kramp, Sanaz Sedghpour Sabet	for their expert help with the DNA & RNA technique.
Nicole von Wurmb-Schwark	for carrying out the zygosity testing of the twins.
Tanja Kaacksteen, Melanie Friskovec, Tanja Wesse, Catharina von der Lancken	for their extremely competent help with performing ChIP-qRT PCR.
Dr. Maren Paulsen, Dr. Simone Lipinski	for sharing their experience of culturing primary cell lines.
Dr. Markus Schilhabel	for his help with the DNA preparing.
Dr. Martina E. Spehlmann	for the clinic samples collection.
Marcus Will	for his IT expertise and maintenance of the server and the network.
My parents Shoufang Wang and Huashan Feng	for their constant motivation, and support on all levels since the beginning of my studies.

I thank the entire staff of the Institute of Clinical Molecular Biology for constructing the nice atmosphere and providing their friendly support. In addition, I apologize to those, who I forgot to mention in this acknowledgement. Finally, I would like to express my appreciation for the expert help of the entire team of the 1st medical department, University hospital Schleswig-Holstein, especially for the cooperation of patients, who agreed on the use of their biopsies and blood samples for academic research.

11 Declaration

Declaration

Apart from the advice of my supervisors Dr. Robert Häsler, Prof. Dr. Philip Rosenstiel and Prof. Dr. Dr. Thomas Bosch, this thesis is completely the result of my own work. No part of it has been submitted to any other board for another qualification. Most of the results have been or are about to be published (see "Appendix", Page 120-127).

Kiel, April 30th, 2013 (Zhe Feng)

Erklärung

Hiermit erkläre ich, daß diese Dissertation, abgesehen von der Beratung durch meine akademischen Lehrer Dr. Robert Häsler, Prof. Dr. Philip Rosenstiel and Prof. Dr. Dr. Thomas Bosch, nach Inhalt und Form meine eigene Arbeit ist. Sie hat weder im Ganzen noch zum Teil an anderer Stelle im Rahmen eines Promotionsverfahrens vorgelegen. Die meisten Ergebnisse dieser Arbeit wurden zur Veröffentlichung eingereicht (siehe "Appendix", Seite 120-127).

Kiel, den 30.Apr.2013 (Zhe Feng)