A functional methylome map of ulcerative colitis

Robert Häsler, Zhe Feng, Liselotte Bäckdahl, et al.

Genome Res. 2012 22: 2130-2137 originally published online July 23, 2012
Access the most recent version at doi:10.1101/gr.138347.112

A functional methylome map of ulcerative colitis

Robert Häsler,[1,5] Zhe Feng,[1,5] Liselotte Bäckdahl,[2] Martina E. Spehlmann,[1] Andre Franke,[1] Andrew Teschendorff,[2] Vardhman K. Rakyan,[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]

[1]*Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, 24105 Germany;* [2]*Medical Genomics, UCL Cancer Institute, University College London, London WC1E 6BT, United Kingdom;* [3]*Blizard Institute, Barts and The London, Queen Mary University of London, London E1 2AT, United Kingdom;* [4]*Wellcome 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 UG 133 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 *cis* 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 *CFI*, 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, OMIM 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 (Rakyan 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 (25%–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 (Fraga et al. 2005; Kaminsky et al. 2009; Javierre et al. 2010; Rakyan et al. 2011a).

[5]These authors shared first authorship.
[6]These authors shared senior authorship.
[7]Corresponding authors
E-mail s.schreiber@mucosa.de
E-mail p.rosenstiel@mucosa.de
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.138347.112. Freely available online through the *Genome Research* Open Access option.

22:2130–2137 © 2012, Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/12; www.genome.org

Starting with a collection of mucosal biopsies from 20 MZ twins, discordant for UC, we aimed to generate a genome-wide functional epigenetic map independent 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] ≤ 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 ≤ 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

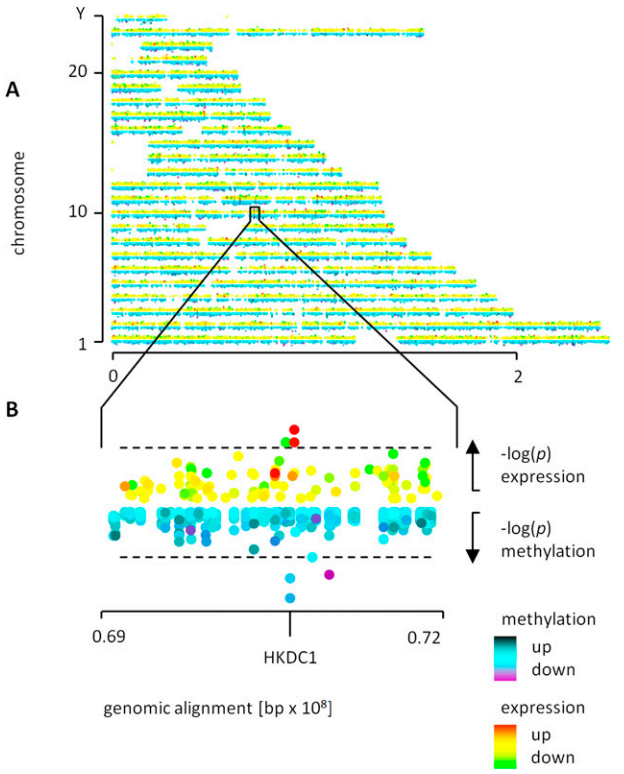

Figure 1. (*A*) Genome-wide profiles of DNAm and gene expression. The *x*-axis represents the genomic location with chromosomes represented as rows. The *y*-axis represents the significance of the differences between UC patients and healthy individuals and is displayed as –log(*p*) for the mRNA and as 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).

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 (*MT1H*) 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 ~16% (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*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

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

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
CFI; complement factor I	Complement activation; innate immune response	4q25	Post-TSS, SGB	↓ 1.51×10^{-05} (2)	↑ 5.71; 4.45×10^{-09}
FLNA; Filamin A, alpha	Positive regulation of I-κB kinase/NF-κB cascade	Xq28	Post-TSS, SGB	↓ 1.63×10^{-02} (2)	↓ −2.85; 7.72×10^{-06}
HKDC1; hexokinase domain containing 1	Glycolysis	10q22.1	Pre-TSS and post-TSS	↓ 7.94×10^{-06} (6)	↑ 2.01; 5.16×10^{-06}
IGHG1[a]; immunoglobulin heavy constant gamma 1 (G1m marker)	Immune response; antigen binding	14q32.33	Post-TSS, SGB	(–) 1.56×10^{-01} (0)	↑ 8.25; 8.90×10^{-09}
MT1H; metallothionein 1H	Metal ion binding; protein binding	16q13	Post-TSS	↑ 1.50×10^{-02} (2)	↓ −1.66; 3.04×10^{-05}
PTN; pleiotrophin	Positive regulation of cell proliferation	7q33	Post-TSS	↓ 2.42×10^{-02} (1)	↓ −4.16; 2.76×10^{-08}
SLC7A7[b]; solute carrier family 7 (cationic amino acid transporter, y+ system), member 7	Blood coagulation, leukocyte migration	14q11.2	Post-TSS	↓ 5.21×10^{-06} (4)	(–) 1.16; 6.73×10^{-01}
SPINK4; serine peptidase inhibitor, Kazal type 4	Serine-type endopeptidase inhibitor activity	9p13.3	Post-TSS, SGB	↓ 3.03×10^{-06} (2)	↑ 13.05; 4.45×10^{-09}
THY1; Thy-1 cell surface antigen	Positive regulation of t cell activation	11q23.3	Post-TSS	↓ 2.28×10^{-03} (2)	↑ 2.99; 4.75×10^{-07}
TK1; thymidine kinase 1, soluble	DNA replication	17q23.2–q25.3	Pre-TSS	↑ 3.08×10^{-03} (2)	↑ 1.32; 2.30×10^{-02}

(TSS) Transcription start site; (SGB) 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.
[a]Negative control candidate for differential expression without epigenetic modification.
[b]Negative 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 sigmoid 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

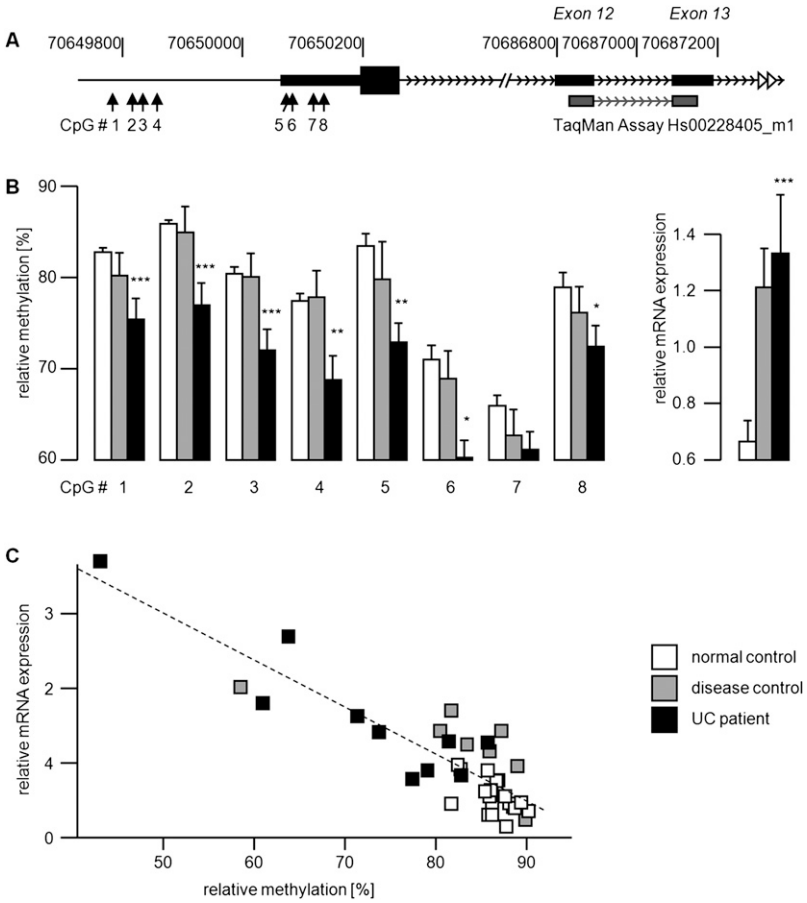

Figure 2. Validation of DNAm and gene expression in selected candidate transcripts; example: *HKDC1*. (*A*) Relative position of CpGs (arrows up, continuously numbered 1–8) and real-time PCR probe (dark gray). (*B*) Quantitative results of the validation: (1) methylation (via pyrosequencing, CpGs continuously numbered 1–8 in concordance with Fig. 2A) in $n = 50$ individuals (validation panel II); 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-ρ r = −7.43). Significant differences between UC patients and normal controls are indicated by asterisks ([*] $P \leq 5 \times 10^{-3}$; [**] $P \leq 5 \times 10^{-4}$; [***] $P \leq 5 \times 10^{-6}$).

are independent of genetic effects. This also reflects fundamental differences between GWAS and EWAS approaches, as outlined recently (Rakyan 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 Cormaci 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).

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 $\leq$ 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 $\leq$ 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 data set 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 $\leq$ 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 $\leq$ 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 $\leq$ 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 GSE 27899 (samples GSM688887–GSM688926).

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.F., L.B., and M.E.S. collected the samples and conducted the experiments; R.H., Z.F., A.F., A.T., V.K.R., T.A.D., G.A.W., and A.F. analyzed the data; R.H., Z.F., 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'Amato 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, Kisfali P, Smith EO, Cho JH, Melegh B, 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. 2004. DNA analysis from mixed biological materials. *Forensic Sci Int* (Suppl) **146:** S123–S125.

Bell JT, 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, Baccarelli A, Hou L, Bonzini M, Fustinoni 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, Waetzig GH, 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. doi: 10.1371/journal.pmed.0020199.

Dhir M, Montgomery EA, Glöckner SC, Schuebel KE, Hooker CM, Herman JG, Baylin 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.

Dieckgraefe BK, Stenson WF, 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 I-H, Wen B, Murakami P, Aryee MJ, 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.

Edwards RA, Witherspoon M, Wang K, Afrasiabi K, Pham T, Birnbaumer L, 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 E, 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* **102:** 10604–10609.

Franke A, Balschun T, Karlsen TH, Sventoraityte J, Nikolaus S, Mayr G, Domingues FS, Albrecht M, Nothnagel 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, Häsler 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, Hammerø 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, Sakul H, Harris TJ, Buckler A, et al. 1999. A genomewide 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, Zvirbliene A, Spehlmann ME, von Wurmb-Schwark N, Kupcinskas L, 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, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE, Lumey LH. 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci* **105:** 17046–17049.

Irizarry RA, 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.

Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, Berdasco 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.

Jurisic G, Iolyeva M, Proulx 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 midkine family in cancer, inflammation and neural development. *Nagoya J Med Sci* **67:** 71–82.

Kaminsky ZA, Tang T, Wang S-C, Ptak C, Oh GHT, Wong AHC, Feldcamp LA, Virtanen C, Halfvarson J, Tysk C, et al. 2009. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet* **41:** 240–245.

Lawrance IC, Fiocchi C, 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 $2^{-\Delta\Delta C_T}$ method. *Methods* **25:** 402–408.

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

Mann MRW, Lee SS, Doherty AS, Verona RI, Nolen LD, Schultz RM, Bartolomei 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, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon Lon R, Chakravarti A, et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461:** 747–753.

Moriyama 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, 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.

Rakyan VK, Beyan H, Down TA, Hawa MI, Maslau S, Aden D, Daunay A, Busato F, Mein CA, Manfras 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. *Semin Immunol* **21:** 334–345.

Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M, Siegmund KD, Koss 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.

Stoll M, Corneliussen B, Costello CM, Waetzig GH, Mellgard B, Koch WA, Rosenstiel P, Albrecht M, Croucher PJP, Seegert 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.

Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Noushmehr 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, Suo L, Van Neste L, van Dam P, Van Marck 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 RJFL, Sandkuijl LA, Enthoven L, Winokur ST, Bakels F, Padberg GW, van Ommen G-JB, Frants 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 E, Lignitz E, Poetsch 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.