

RESEARCH

Open Access



Peripheral blood methylation profiling of female Crohn's disease patients

Andrew Y. F. Li Yim^{1,4†}, Nicolette W. Duijvis^{2†}, Jing Zhao^{2†}, Wouter J. de Jonge², Geert R. A. M. D'Haens³, Marcel M. A. M. Mannens¹, Adri N. P. M. Mul¹, Anje A. te Velde² and Peter Henneman^{1*}

Abstract

Background: Crohn's disease (CD) is a chronic inflammatory disorder belonging to the inflammatory bowel diseases (IBD). CD affects distinct parts of the gastrointestinal tract, leading to symptoms including diarrhea, fever, abdominal pain, weight loss, and anemia. The aim of this study was to assess whether the DNA methylome of peripheral blood cells can be associated with CD in women.

Methods: Samples were obtained from 18 female patients with histologically confirmed ileal or ileocolic CD and 25 healthy age- and gender-matched controls (mean age and standard deviation: 30.5 ± 6.5 years for both groups). Genome-wide DNA methylation was determined using the Illumina HumanMethylation 450k BeadChip.

Results: Our analysis implicated 4287 differentially methylated positions (DMPs; corrected $p < 0.05$) that are associated to 2715 unique genes. Gene ontology enrichment analysis revealed significant enrichment of our DMPs in immune response processes and inflammatory pathways. Of the 4287 DMPs, 32 DMPs were located on chromosome X with several hits for MIR223 and PABPC5. Comparison with previously performed (epi) genome-wide studies revealed that we replicated 33 IBD-associated genes. In addition to DMPs, we found eight differentially methylated regions (DMRs).

Conclusions: CD patients display a characteristic DNA methylation landscape, with the differentially methylated genes being implicated in immune response. Additionally, DMPs were found on chromosome X suggesting X-linked manifestations of CD that could be associated with female-specific symptoms.

Keywords: Crohn's disease, Inflammatory bowel diseases, Females, DNA methylation, Peripheral blood, Epigenome-wide association study

Background

Crohn's disease (CD) is an inflammatory bowel disease (IBD) characterized by a chronic inflammatory condition of the gastrointestinal tract. On a world-wide basis, CD has a prevalence of 0.5 % with an annual incidence of 12.7 per 100,000 person-years [1]. The inflammation associated with CD can reach through all layers of the intestinal wall, causing complications such as strictures and fistula. The terminal ileum is the most prevalent site for inflammation and strictures, often requiring surgical ileocecal resection.

Different immunosuppressive therapies are commonly applied, such as thiopurines, corticosteroids, and anti-tumor necrosis factor (aTNF) agents, all of which have variable success rates. Aside from complications within the gastrointestinal tract, CD occasionally manifests itself in an extra-intestinal fashion. Certain CD-associated symptoms appear to be gender-specific [2], with female-specific symptoms including irregular menstruation [3, 4] as well as an increased risk of complications during pregnancy [5].

Despite the extensive research performed on CD, the etiology is unknown. Numerous studies have sought to associate genetic changes to the pathogenesis of CD with genome-wide association studies (GWAS) finding many loci that are associated with pathways that have been well established in IBDs, such as pattern recognition

* Correspondence: p.henneman@amc.uva.nl

†Equal contributors

¹Department of Clinical Genetics, Genome Diagnostics Laboratory, Academic Medical Center Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Full list of author information is available at the end of the article

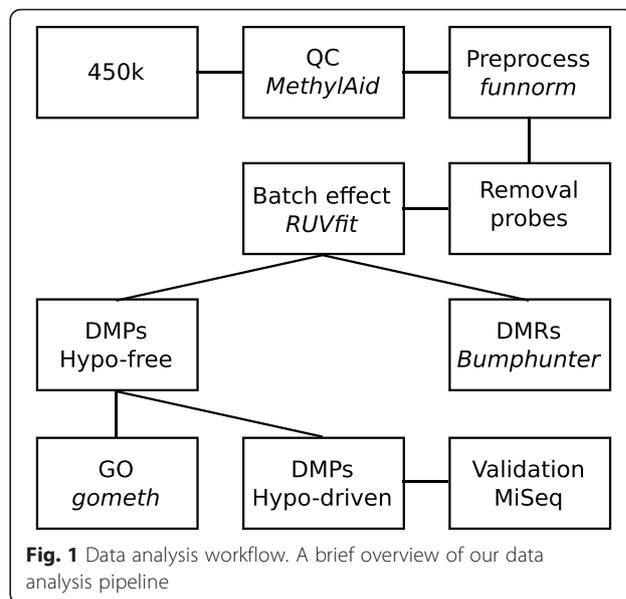
signaling, cytokine production, and autophagy [6, 7]. However, only 20 % of the estimated heritability (30–50 %) of CD can be explained by common genetic variants [6, 8, 9]. A growing body of literature suggests that additional factors such as diet [10], the gut-microbiome [11] and the epigenome [12–15] add to the development and progression of CD.

While the genome remains static for one organism over time and across different cell types, the epigenome can vary considerably. One of the well-described epigenetic modifications is cytosine methylation, which involves the attachment of a methyl group to a cytosine followed by a guanine (CpG site). Aberrant methylation patterns have been implicated in many complex disorders, such as cancers [16], diabetes [17], and juvenile stress [18]. In this study, the aim was to explore how the methylome of peripheral blood is affected in female CD patients. To this end, the HumanMethylation450 BeadChip array (450k) was used to find differentially methylated positions (DMPs) and regions (DMRs) in DNA isolated from peripheral blood. We specifically looked at blood due to the relative ease and non-invasive nature in obtaining the samples. First, we sought to find differentially methylated loci through a hypothesis-free approach. Here, we specifically chose to assess the methylome of female CD patients to see whether CD manifests in the methylome of chromosome X, the results of which could help understand female-specific CD symptoms. Second, we aimed at replicating previously reported genes through a hypothesis-driven approach, whereby we assessed the methylation patterns of CD-associated genes retrieved from GWAS [6, 8, 9] and epigenome-wide association studies (EWAS) [12–14, 19, 20].

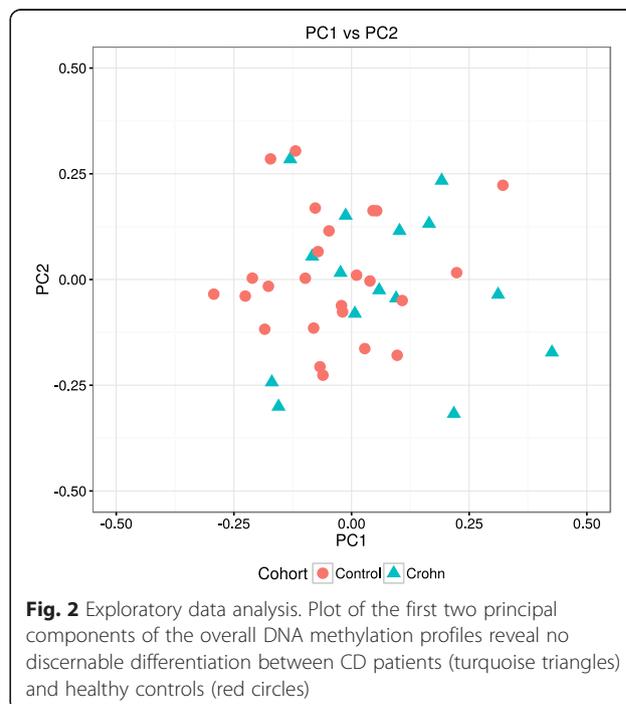
Results

Quality control and exploratory data analysis

Samples were processed according to the flowchart in Fig. 1. Initial quality control using MethylAid [21] indicated that three CD patients failed the bisulfite conversion, hybridization, and overall methylation threshold, resulting in their exclusion from downstream analyses. Subsequent principal component analysis did not reveal any discernable separation of the CD patients from the healthy controls (Fig. 2). Moreover, the first principal component explained only 12.5 % of the variance, suggesting that the DNA methylome does not differ considerably among samples (Additional file 1: Figure S1a). As the DNA samples were obtained from peripheral blood, the concern existed that the heterogeneity of the blood cell composition confounded our data [22, 23]. We therefore estimated the cellular composition per sample using the algorithm described in Houseman



et al. [23] (Additional file 1: Figure S1b). When comparing CD versus healthy controls, a difference in blood cell composition was observed, which was nominally statistically significant at an α (p value threshold) of 0.05. However, after correcting for multiple testing using the Bonferroni method, the associations were almost statistically significant suggesting that CD is potentially associated with changes in the cellular composition. Calculating the Pearson correlation coefficient for the blood cell



distribution with each principal component revealed a strong correlation of the blood cell distribution with the first principal component. This correlation was statistically significant for CD8T cells, CD4T cells, natural killer cells, and granulocytes after correcting for multiple testing using the Benjamini-Hochberg (BH) procedure (Additional file 1: Figure S1c). We surmised that additional biological confounders included age [24] and the usage of aTNF medication at the time of phlebotomy. To prevent age from confounding our data, we had age-matched our cohort prior to sampling. Correlating age and aTNF usage with the principal components revealed no significant correlation ($R^2 > 0.10$), suggesting that neither affect the methylome significantly (Additional file 1: Figure S1d, e). To correct for the most prominent (hidden) biological confounders, such as the cellular composition, we performed factor analysis using the RUVfit function [22, 25–27]. RUVfit is a wrapper function for the “remove unwanted variation” (RUV) methods [25–27]. While it would have been possible to include the estimated blood cell composition obtained from the Houseman algorithm as covariates in the linear model, as described in Guintivano et al. [28] and Hannum et al. [24], this method was discouraged in Montañó et al. [29] and Jaffe and Irizarry [22] as the estimated blood cell composition was found to yield biased results. Instead, Jaffe and Irizarry suggested the usage of RUV as a way for correcting for composition-based confounding [22]. The advantage of RUV over other conventional methods is its ability to discover (hidden) biological confounders aside from blood cell composition. For more information about our implementation of the RUVfit function, see Section 5.

Differentially methylated positions in Crohn’s disease patients

After normalizing the data and correcting for confounders, we observed 4287 significant DMPs (BH-adjusted $p < 0.05$) that were associated to 2715 unique genes. Of the 4287 significant DMPs, 949 were hypomethylated with the remaining 3338 being hypermethylated (Additional file 2: Table S1). The two most significant DMPs were found within the protein tyrosine phosphatase *PTPRN2* [Ensembl: ENSG00000155093] and the zinc-finger protein *BCL11A* [Ensembl: ENSG00000119866], which were moderately hypermethylated in CD patients versus healthy controls (see dot-boxplots on the right of Fig. 3a).

Differentially methylated position distribution analysis

The precise fashion through which methylation affects transcription remains unknown with the current dogma being that hypermethylated regions within the transcription

start site (TSS) silence the respective gene [30, 31]. To this end, we investigated the DMP distribution per genetic feature. Here, we used a Fisher exact test to calculate whether the ratio of DMPs versus the total number of 450k probes per genetic feature was significantly different from a DMP distribution originating by chance. We observed a statistically significant difference in the DMP distribution for the transcription start sites (TSS1500 and TSS200), the gene body, the first exon, the 3’ untranslated region (3’ UTR) and the intergenic region (Additional file 3: Figure S2a and Additional file 4: Table S2). Only the 5’UTR was not statistically significant, suggesting that the DMPs are not randomly distributed. Next, we sought to test whether the direction of methylation was significantly different for any of the genetic features using a second Fisher exact test. Here, we found no statistically significant differences in the distribution of hypo- and hypermethylated DMPs for any of the genetic features (Additional file 3: Figure S2b, c and Additional file 4: Table S2).

A similar approach was used to assess the DMP distribution per chromosome. Here, we found a significantly different DMP distribution for chromosomes 1, 19, and X (Fig. 3b). Furthermore, analysis of the hypo- and hypermethylated DMP distribution revealed that while the autosomal chromosomes contained more hypermethylated DMPs than hypomethylated DMPs, the inverse was true for chromosome X (Fig. 3c and Additional file 5: Table S3). As we had a female-only cohort, we investigated chromosome X in further detail. Analysis of the X-associated DMPs yielded 32 DMPs of the 10,246 probes on chromosome X (Additional file 11: Table S4). Analysis of the genes associated to the X-linked DMPs revealed an enrichment of only two genes: *MIR223* [Ensembl: ENSG00000207939] (Fig. 4a) and *PABPC5* [Ensembl: ENSG00000174740] (Fig. 4b), which were represented by two and four DMPs, respectively.

Differentially methylated regions in Crohn’s disease patients

Using the *bumphunter* function [32], we found eight DMRs, which we associated to *HLA-J* [Ensembl: ENSG00000204622], *BOLA3* [Ensembl: ENSG00000163170], *TACSTD2* [Ensembl: ENSG00000184292], *APOBEC1* [Ensembl: ENSG00000111701], *MOVIOL1* [Ensembl: ENSG00000073146], *OR2L13* [Ensembl: ENSG00000196071], *LINC00612* [Ensembl: ENSG00000214851], and *SHANK2* [Ensembl: ENSG00000162105] (Table 1). While the individual CpGs comprising the DMRs were not significantly differentially methylated, the mean difference across the entire region was moderate but noticeable (Additional file 6: Figure S3).

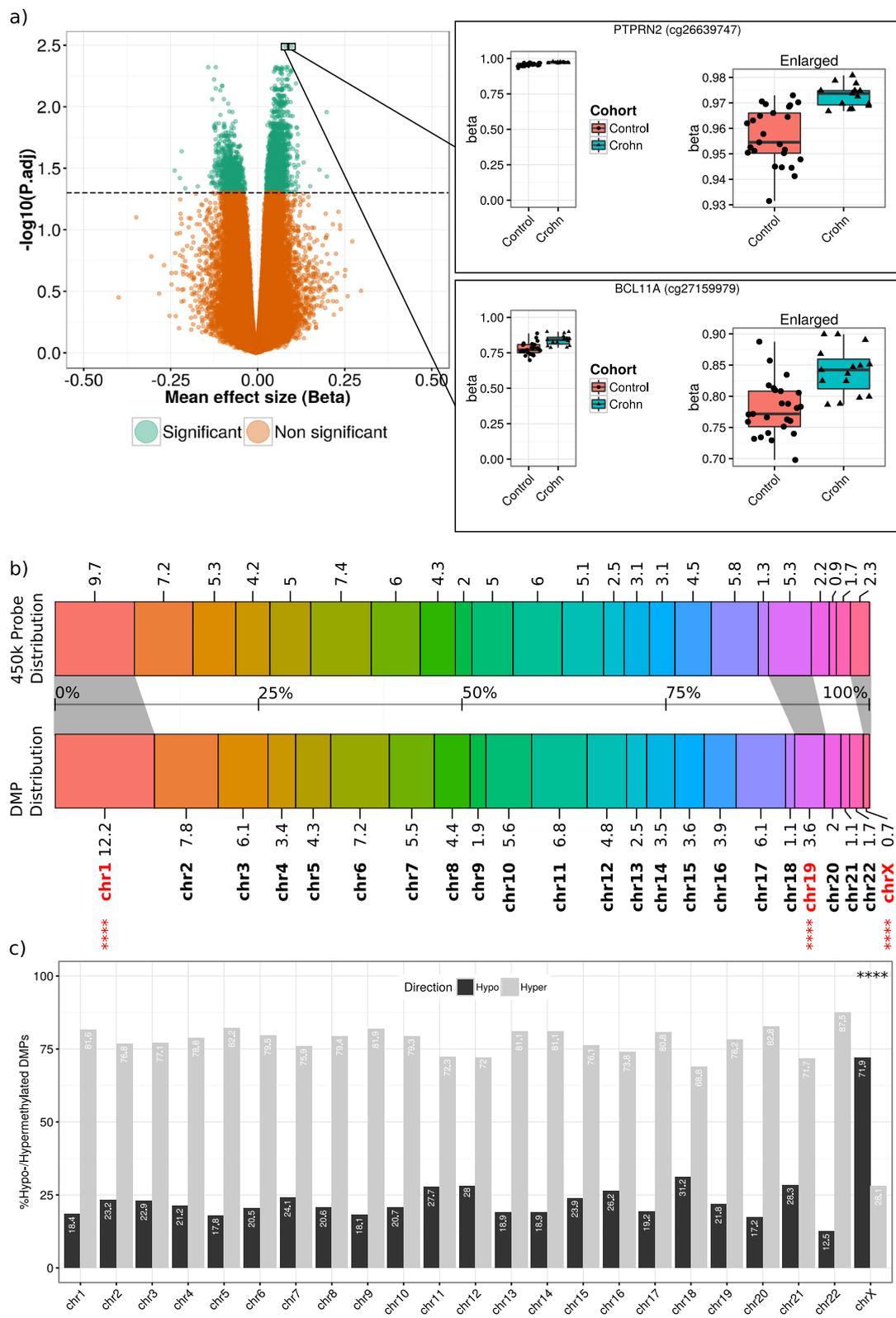


Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Differentially methylated positions. **a** Left: Volcano plot of the $-\log_{10}$ transformed BH-adjusted p on the Y-axis versus the mean effect size in methylation (beta) on the X-axis. DMPs are indicated in green. Right: Dot-boxplots of the two most significant DMPs: cg26639747 (*PTPRN2*) and cg27159979 (*BLC11A*). **b** Comparison of the probe distribution on the 450k versus the DMP distribution per chromosome where the different colors represent the different chromosomes. The numbers along the barplot represent the percentages of the 450k probes (top) or DMPs (bottom) per chromosome. Significantly different DMP distributions are indicated in bold red with the asterisks indicating the level of significance as found in Additional file 5: Table S3 ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$). **c** For each chromosome, the percentage hypo- and hypermethylated DMPs is indicated with barplots in black and gray, respectively

Pathway enrichment analysis of the differentially methylated positions

To understand the functional relevance of our reported DMPs, gene ontology (GO) enrichment analysis was performed. GO enrichment yielded 32 significantly enriched (BH-adjusted $p < 0.05$) processes, with notable hits for immune response (GO:0006955 and GO:0002376) and leukocyte activation (GO:0045321) as well as neutrophil chemotaxis (GO:0030593) (Table 2).

Overlap with previous studies

Next, we compared the genes associated to our DMPs with genes associated to CD and IBD from previous

GWAS [6, 8, 9] and EWAS [12–14, 19, 20] data. The GWAS-derived list contained 275 genes whereas the EWAS-derived list contained 4388 genes. When comparing the GWAS, the EWAS and our own data, we found 33 genes that were present in all three datasets. Analysis of the CpGs associated to the 33 overlapping genes yielded 136 statistically significant hypothesis-driven DMPs (BH-adjusted $p < 0.05$) (Additional file 7: Table S5). Of the ten most significant hypothesis-driven DMPs, five DMPs were associated to *TNF* [Ensembl: ENSG00000232810] (Fig. 5c) and two were associated to *SP140* [Ensembl: ENSG00000079263] (Fig. 5b). Interestingly, while the hypothesis-driven DMPs found in *TNF*

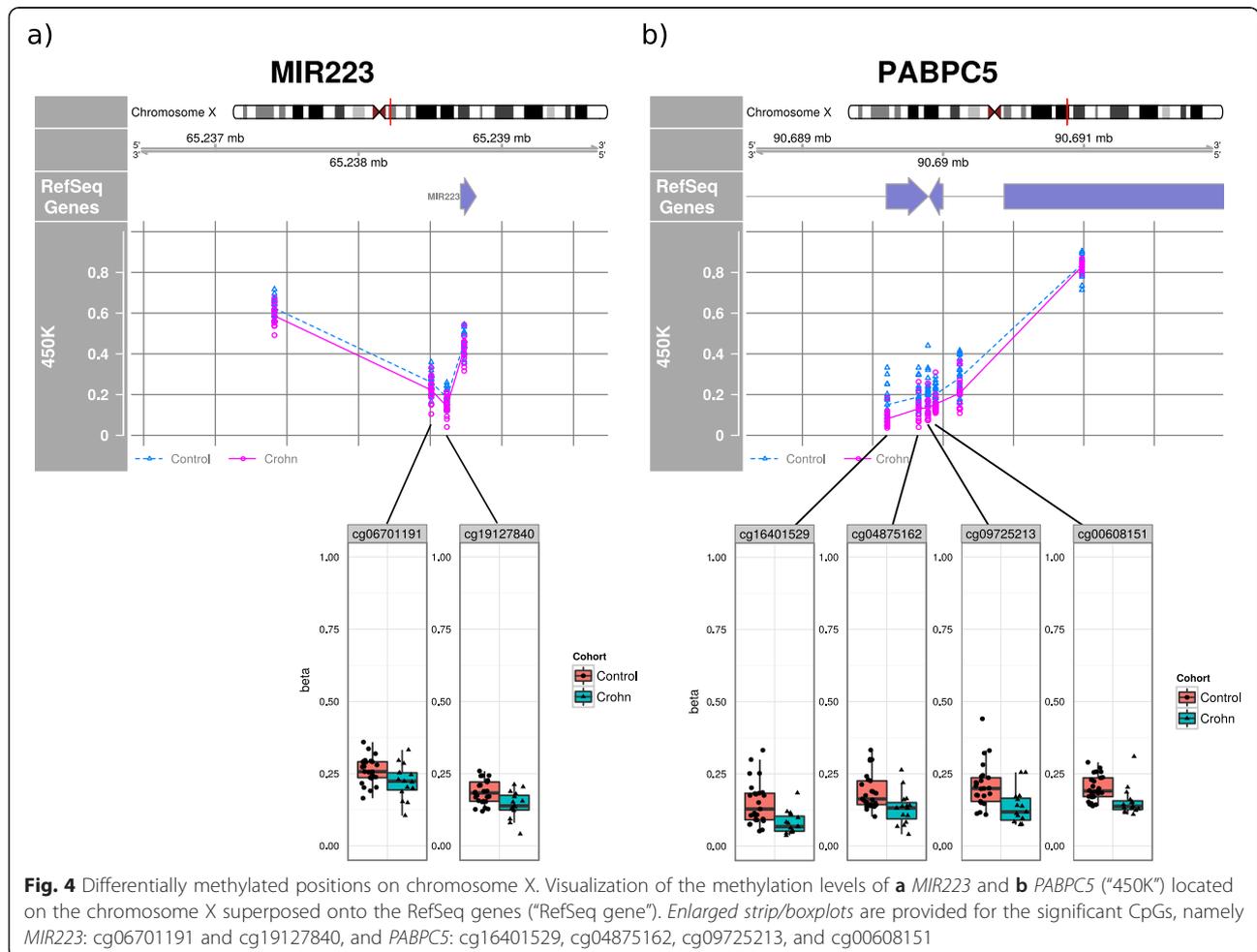


Table 1 DMRs as predicted by the bump hunter function, containing four or more consecutive DMPs

DMR location (hg19)	Mean effect size	Area DMR	DMPs	Nearest gene
chr6: 29895175-29895260	-0.172	6.89E-01	4	HLA-J
chr2: 74357527-74357872	0.127	5.10E-01	4	BOLA3
chr1: 59043199-59043280	0.121	4.83E-01	4	TACSTD2
chr12: 7781004-7781288	-0.118	4.72E-01	4	APOBEC1
chr22: 50528213-50528312	-0.0992	3.97E-01	4	MOV10L1
chr1: 248100345-248100585	0.0935	3.74E-01	4	OR2L13
chr12: 9217510-9217669	-0.0918	3.67E-01	4	LINC00612
chr11: 70672841-70672878	0.0911	3.65E-01	4	SHANK2

appear to occur consecutively, our previous DMR analysis did not yield any hits for *TNF*, which might be due to the limited mean difference observed across the *TNF*-associated DMPs. To validate our findings for *SP140* and *TNF*, we performed MiSeq amplicon sequencing and correlated the results with our findings obtained from the 450k data. The methylation levels obtained from the MiSeq sequencing were found to be concordant with the 450k results for *SP140* (see MiSEQ track in Fig. 5b). Unfortunately, we were unable to obtain sufficient reads with the primers designed for our region of interest for *TNF*. We therefore sequenced downstream of our region of interest, which yielded adequate reads and revealed methylation levels similar to what was found using the 450k (see MiSEQ track in Fig. 5c). In addition to *SP140* and *TNF*, specific regions within *TNFSF4* [Ensembl: ENSG00000117586] (Additional file 8: Figure S4b), *IL10/IL19* [Ensembl: ENSG00000136634] (Additional file 8: Figure S4c), and *ORMDL3* [Ensembl: ENSG00000172057] (Additional file 8: Figure S4d) were also sequenced, as they had been associated with CD previously [6]. Overall, the methylation levels obtained through MiSeq sequencing were found to be concordant with the methylation levels obtained from the 450k array (Additional file 8: Figure S4a), but the differences between CD patients and healthy controls were not statistically significant.

In particular, we assessed the methylation levels of the top DMPs as reported by McDermott et al. due to the similarity in design and goals with our current study [13]. While our results mostly correspond with respect to the direction of methylation, our reported effect sizes differ (Additional file 9: Table S6). Visualization of the DMPs found in *TIFAB* [Ensembl: ENSG00000255833] (cg16176675) and *TRAF6* [Ensembl: ENSG00000175104] (cg01476222), which represent the top DMP and the validated DMP reported by McDermott et al., displayed a minor difference that was not statistically significant in our data (Fig. 5d, e). For certain DMPs, we appear to observe opposite effects. Analysis of the contentious

Table 2 Statistically significant gene ontology enrichment on our significant DMPs

GO	Term	p value	BH-adjusted p value
GO:0002376	Immune system process	1.58E-07	1.60E-03
GO:0006955	Immune response	9.61E-08	1.60E-03
GO:0007166	Cell surface receptor signaling pathway	7.25E-07	4.86E-03
GO:0060326	Cell chemotaxis	9.63E-07	4.86E-03
GO:0006909	Phagocytosis	3.92E-06	1.55E-02
GO:0030593	Neutrophil chemotaxis	4.60E-06	1.55E-02
GO:0098602	Single organism cell adhesion	5.62E-06	1.62E-02
GO:0006952	Defense response	1.97E-05	2.79E-02
GO:0048583	Regulation of response to stimulus	1.44E-05	2.79E-02
GO:0016337	Single organismal cell-cell adhesion	2.07E-05	2.79E-02
GO:0045321	Leukocyte activation	1.54E-05	2.79E-02
GO:0050900	Leukocyte migration	1.84E-05	2.79E-02
GO:0030595	Leukocyte chemotaxis	2.06E-05	2.79E-02
GO:1990266	Neutrophil migration	1.84E-05	2.79E-02
GO:0071621	Granulocyte chemotaxis	1.80E-05	2.79E-02
GO:0071944	Cell periphery	2.28E-05	2.88E-02
GO:0001775	Cell activation	2.66E-05	3.10E-02
GO:0034109	Homotypic cell-cell adhesion	2.76E-05	3.10E-02
GO:0016477	Cell migration	3.42E-05	3.63E-02
GO:0006954	Inflammatory response	3.94E-05	3.76E-02
GO:0007165	Signal transduction	4.36E-05	3.76E-02
GO:0048870	Cell motility	4.61E-05	3.76E-02
GO:0051674	Localization of cell	4.61E-05	3.76E-02
GO:0070486	Leukocyte aggregation	4.21E-05	3.76E-02
GO:0002696	Positive regulation of leukocyte activation	4.98E-05	3.76E-02
GO:0071800	Podosome assembly	4.69E-05	3.76E-02
GO:0009897	External side of plasma membrane	5.02E-05	3.76E-02
GO:0007159	Leukocyte cell-cell adhesion	5.70E-05	4.11E-02
GO:0098552	Side of membrane	6.54E-05	4.56E-02
GO:0044700	Single organism signaling	7.07E-05	4.72E-02
GO:0050867	Positive regulation of cell activation	7.25E-05	4.72E-02
GO:0009611	Response to wounding	7.84E-05	4.95E-02

DMPs reveals an association with UC in the dataset of McDermott et al. suggesting CD-specific methylation.

Discussion

Quality control and exploratory data analysis

In this study, we studied the methylation differences between female CD patients versus healthy controls in peripheral blood. To our knowledge, we are the first to perform methylation analysis in peripheral blood using a

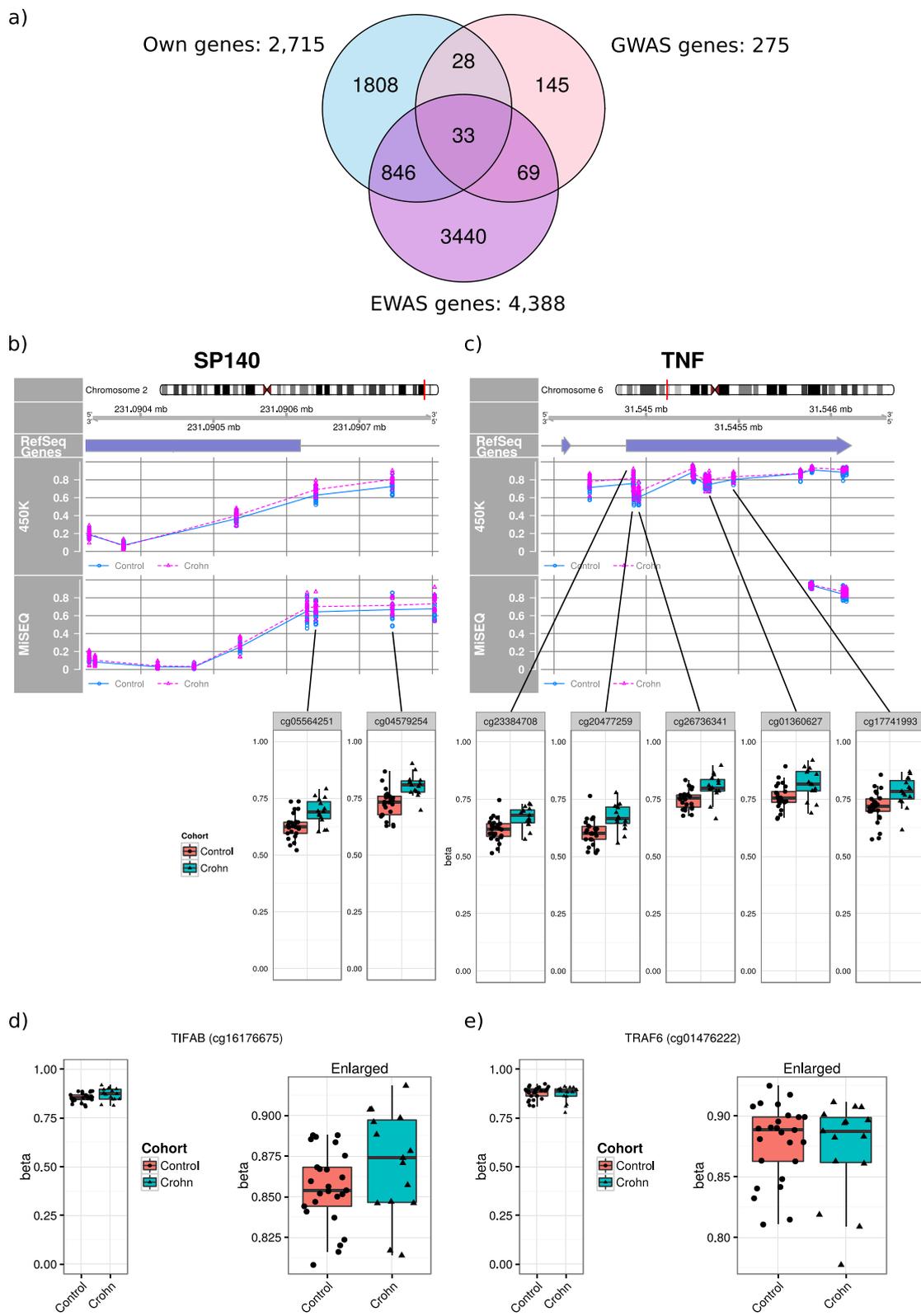


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Hypothesis-driven differentially methylated positions. **a** Venn diagram representing the overlap between CD-associated genes from our data (2715 genes), GWAS data (275 genes), and EWAS data (4388 genes). Genomic plots of the methylation levels of the DMPs obtained from the 450k ("450K") compared to the methylation levels obtained from MiSeq sequencing ("MiSEQ") superposed on known RefSeq genes ("RefSeq gene") for **b** *SP140* and **c** *TNF*. Note that the MiSeq sequencing of *SP140* missed one CpG covered by the 450k, which was specifically removed due to low read count (<100; see Section 5 "Illumina MiSeq Sequencing"). *Enlarged dot-boxplots* are provided for the significant CpGs associated to *TNF*: cg23384708, cg20477259, cg26736341, cg1360627, and cg17741993, and *SP140*: cg05564251 and cg04579254. Dot-boxplots of **d** cg16176675 (*TIFAB*) and **e** cg01476222 (*TRAF6*), as reported from McDermott et al.

female-only cohort, which provided us with the opportunity to investigate CD-associated methylome manifestations on chromosome X. We used peripheral blood as our sample of interest, with the intention of discovering epigenetic loci that could be of use in the clinical setting. As peripheral blood is a heterogeneous population, our results were confounded by the change in blood cell distribution in the presence of CD. To correct for the blood cell distribution, we implemented the RUVfit function [22, 25].

Differentially methylated positions in Crohn's disease patients

Our analysis yielded 4287 sites that displayed statistically significant differences in methylation for CD patients versus healthy controls. Despite finding many DMPs, the effect sizes were limited, which reflects the results obtained by Harris et al. [12] and McDermott et al. [13]. The two most significant DMPs were found in *PTPRN2* and *BCL11A*, with the former being associated to type 1 diabetes in mice [33–35] and the latter to type 2 diabetes in human males [36]. Previous research showed that *PTPRN2* in rats displays phosphatase activity towards inositol phospholipids [37], whereas *BCL11A* in mice acts as a negative regulator of p53 [38]. From the literature, it appears as though *PTPRN2* and *BCL11A* are involved in generic pathways suggesting that deregulation of generic pathways underlie complex disorders such as CD.

Differentially methylated position distribution analysis

Analysis of the distribution of DMPs across genetic features revealed that the DMPs are not randomly distributed. However, no particular enrichment of either the hyper- or hypomethylated DMPs was observed. A similar DMP distribution analysis for the chromosomes revealed significant differences in DMP distributions for chromosomes 1, 19, and X, where chromosome X displayed a significant depletion in DMPs versus the other chromosomes. The limited number of DMPs on chromosome X corroborates the overall nongender-specific nature and incidence of CD [1]. Of the limited number of X-linked DMPs, we found an enrichment of DMPs associated to the microRNA *MIR-223* and *PABPC5*. *MIR-223* plays an important role in promoting

granulocyte differentiation [39] whose deregulation is associated with various cancers [40–42] as well as endothelial cell apoptosis [43], implicating a putative role in the formation of ulcers in CD patients. Additionally, *MIR-223* expression was found to be elevated in the inflamed ileum of CD patients [44], with the expression of *MIR-223* being tightly regulated through histone acetylation and DNA methylation by the AML1/ETO fusion protein, making it an interesting target for future research [45]. The available literature on *PABPC5* describes its discovery based on similarity towards poly(A) binding proteins, suggesting a role in transcriptional regulation. Similar to *PTPRN2* and *BCL11A*, it appears as though *PABPC5* is involved in a generic pathway. Nonetheless, the fact that four DMPs were associated to *PABPC5* makes it an interesting candidate for future research on CD.

Differentially methylated regions in Crohn's disease patients

In addition to DMPs, we found eight DMRs. One of the DMRs was located upstream of the major histocompatibility complex *HLA-J* (Additional file 6: Figure S3). HLA genes are involved in immunoregulation and have been implicated in the pathogenesis of CD previously [6]. Another DMR was associated to *MOV10L1*, which has been described as an RNA helicase involved in piRNA processing [46, 47]. Using the ENCODE data in the UCSC Genome Browser, we observed that the *MOV10L1*-DMR associates to a region that contains transcription factor binding sites (TFBS) [48] for two genes, namely *EGR1* [Ensembl: ENSG00000120738] and *ZBTB33* [Ensembl: ENSG00000177485]. *EGR1* is involved in inflammation through its regulation of downstream targets such as *TNF* [49, 50]. Inflamed intestinal tissue was found to display increased levels of *EGR1* expression in CD patients [51]. The *ZBTB33* protein is a zinc-finger transcriptional regulator, which binds methylated CpG sites conferring transcriptional repression in an in vitro setting [52]. Unlike *EGR1*, no studies have associated *ZBTB33* to IBD. While it is enticing for us to suggest a link between our DMRs and the transcription factor binding sites obtained from UCSC Genome Browser, no proper conclusions can be drawn given that our samples are not the same and that TFBS are often cell-

type specific [53]. Further research is necessary to elucidate a putative interplay between our MOV10L1-DMR and *EGR1*.

Pathway enrichment analysis of the differentially methylated positions

Our GO-enrichment analysis revealed that our DMPs were enriched in pathways involved in inflammation and cell activation. Comparable results were reported by McDermott et al. where differential methylation in peripheral blood mononuclear cells from IBD patients was associated to genes involved in immune response and T cell activation [13]. Our data suggests that the DNA methylome is affected in genes that are involved in pathways associated to inflammation and immune response.

Overlap with previous studies

By comparing our results with previous CD studies, we managed to replicate 33 genes. We confirmed the methylation status of *TNF*, *SP140*, *TNFSF4*, *IL10/IL19*, and *ORMDL3* through MiSeq amplicon sequencing. Our results therefore suggest that deregulation of the previously mentioned genes could occur at an epigenetic and genetic level, thereby contributing to the observed inflammatory phenotype.

Limitations of the current study

It is important to realize that the results obtained in the present study cannot be used as biomarkers. The limited sample size and the minor effect sizes observed obscure the number of true positives and negatives due to the lack of power. Increasing the power could be achieved through a meta-analysis whereby various studies of similar in design are combined. While we have provided a brief comparison of our results with other studies of similar design, a systematic meta-analysis is necessary to ascertain the limited effect sizes observed. As such, our results merely provide CpGs that are found to be associated to CD in our cohort, which nonetheless provide a platform for future studies to elucidate the role of methylation in CD.

Conclusions

This study has shown that CD affects the DNA methylome of peripheral blood in female CD patients versus healthy controls, with the affected genes being enriched in inflammatory pathways. While we report differentially methylated loci in peripheral blood, the effect sizes are limited which was expected given the multifactorial nature of CD. By elucidating the methylome-associated changes in CD, we sought to gain a better understanding of the role of epigenetics in the pathogenesis of CD, thereby opening up new windows of opportunities for research in the diagnosis or treatment of CD.

Methods

Patient inclusion

Our CD samples consisted of 18 female CD patients with histologically confirmed intestinal CD (age range: 22 to 43) that visited the outpatient clinic at the Academic Medical Centre (AMC) IBD department in Amsterdam, the Netherlands. Of the 18 CD cases, only 15 remained after quality control using the *MethylAid* (version 1.4.0) package [21]. The healthy control samples were obtained from 25 anonymous healthy women (age range: 21 to 43) from the biobank located at the AMC Department of Clinical Genetics, DNA Diagnostics laboratory. Healthy female controls were defined as patients that tested screen-negative for specific DNA-mutations as part of genetic family studies. The assembly of this cohort was approved by the medical ethics committee of the Academic Medical Hospital (METC 08/330 # 09.17.0268), and written informed consent was obtained from both the CD patients and control subjects.

DNA isolation and bisulfite conversion

Peripheral blood was drawn and stored in EDTA to prevent coagulation. Erythrocytes were lysed before proteins were aggregated out of the sample. Genomic DNA was extracted through ethanol precipitation, after which the DNA was dissolved in tris-ethylenediaminetetraacetic buffer (Tris-EDTA) and stored at 4 °C. Subsequent bisulfite conversion of the DNA was performed using the Zymo EZ DNA Methylation™ kit following the manufacturer's protocol.

Methylation analysis

Whole-genome DNA methylation profiles were quantified using the Illumina HumanMethylation450k BeadChip Array, which measures 485,577 CpG sites at ServiceXS in Leiden, the Netherlands. Prior to 450k analysis, quality control of converted DNA was performed by means of high-resolution melting analysis of the *H19* locus [Ensembl: ENSG00000130600] according to the diagnostics workflow as described by Alders et al. [54].

Differentially methylated loci analysis

The methylation data was imported into the R statistical programming environment (version 3.2.2) using the Bioconductor package *minfi* (version 1.16.0) [55]. Initial quality control was performed using the *MethylAid* package, whereby the quality of each sample was assessed using the internal control probes located on the BeadChip array [21]. Subsequently, probes were removed that were known to be promiscuous, located on the Y-chromosome, or associated with CpGs with known SNPs (minor allele frequency >0). The remaining probes were normalized using the functional normalization method [56], after which *M* values ($M = \log_2(M/U)$) were used

for statistical analyses and β -values ($\beta = M/(M + U + 100)$) were used for the visualization of the methylation levels [57]. DMPs were obtained through linear regression using the *limma* package [58, 59].

DMRs were obtained using the DMR-finding function in *minfi* called *bumphunter* [32, 55]. In brief, *bumphunter* searches for DMRs by looking for CpGs with a mean difference above a certain threshold. We set the inclusion threshold to 0.08. To remove single CpGs that exceeded the inclusion threshold from our DMRs, we filtered for at least four consecutive CpGs to minimize the probability of randomly obtaining consecutive CpGs whose mean effect size are above 0.08 by chance. See Fig. 1 for a brief summary of our workflow.

Batch effect correction using factor analysis

We accounted for technical batch effects using the functional normalization method, which estimates technical variation through the internal technical control probes located on the 450k array [56]. Unlike technical batch effects, the technical control probes on the 450k array are unaffected by biological confounders. Finding and correcting for biological confounders was done through factor analysis, using the R function *RUVfit* found within the *missMethyl* package (version 1.4.0) [60]. *RUVfit* implements the RUV (“remove unwanted variation”) functions where negative control probes are used to estimate the effects of unwanted variation [26, 27]. Negative control probes are CpGs that are unaffected by the factor of interest but are affected by the batch effect. Due to the fact that we did not know a priori which CpGs were not differentially methylated, we followed the guidelines posted in the vignette of the *missMethyl* package [25]. In short, a linear regression was performed on the CD status against the uncorrected *M* values yielding statistically non-significant CpGs (BH-adjusted $p > 0.5$). Such statistically non-significant CpGs were deemed unassociated with CD and were therefore used as negative control probes. We then called the *RUVfit* function using the *RUV-inverse* (“*RUVinv*”) function from the *ruv* package (version 0.9.6) to estimate and correct for batch effects [25–27].

Differentially methylated position distribution analysis

The DMPs were stratified per genetic feature/chromosome and compared to the total number of 450k probes associated to the respective genetic feature/chromosome. A Fisher exact test of independence was then used to calculate the probability that the number of DMPs found for a specific genetic feature/chromosome was significantly different from the expected number of DMPs. A second Fisher exact test was then performed on the number of hypermethylated DMPs versus the hypomethylated DMPs to assess whether the distribution was significantly different in any genetic

feature/chromosome. Our threshold for statistical significance was set to a Bonferroni-adjusted α of 0.05.

Gene ontology enrichment analysis

Gene ontology (GO) enrichment analysis was performed on the DMPs using the *gometh* function from the *RmissMethyl* package [60]. The *gometh* function corrects for the number of probes per pathway thereby giving a balanced overview of the enriched pathways.

Hypothesis-driven analysis

To compare our data with previous GWAS and EWAS data, we generated lists of unique genes acquired from GWAS and EWAS. The GWAS genes consisted of genes associated to the significant loci reported in the summary statistics obtained from Franke et al. [9], Jostins et al. [6], and Liu et al. [8] whereas the EWAS genes consisted of genes associated to significant loci reported in the summary statistics obtained from Lin et al. [20], Nimmo et al. [14], Karatzas et al. [19], and McDermott et al. [13]. We then compared and looked for the genes that were present in all three gene lists and extracted the CpGs associated to these genes from our own data after which we adjusted for multiple testing accordingly.

Illumina MiSeq sequencing

Technical validation of several promising 450k CpG sites was performed through targeted amplicon sequence analysis using the Illumina MiSeq platform. Primers were designed with a bisulfite-converted reference sequence, human genome build 19 (hg19), using Primer3 [61, 62]. Primer information is described in Additional file 10: Table S7. Amplicons were amplified through PCR and pooled per subject after which non-specific products were removed using the Agencourt AMPure PCR purification kit (Beckman Coulter). Pooled amplicons were elongated using TruSeq indices and adapter sequences after which they were purified. Quality control of the amplicon length within the pools was performed using Agilent 2100 Bioanalyzer. DNA concentration was measured using Qubit 2.0 Fluorometer (ThermoFisher) and equalized to equimolar concentrations for all subject pools. MiSeq amplicon sequencing was then performed according to the standard protocol (Additional file 11: Table S4). Raw sequence data was mapped, aligned, and analyzed using GATK [63, 64], BWA, and Integrative Genomics viewer (version 2.3.57) [65], respectively, against the bisulfite-converted hg19. A minimum of 100 reads per patient amplicon was deemed successful. While we were capable of correcting for (hidden) technical and biological confounders during the 450k methylation analysis, we were unable to correct for confounding factors during the MiSeq amplicon sequencing experiment.

Visualization of the differentially methylated loci

Individual CpGs were visualized as a strip/boxplot using the *ggplot2* package (version 1.0.1) [66]. Regions of CpGs as well as the CpG islands, the histone 3 single- and triple methylation, the DNase I hypersensitivity sites and the transcription factor-binding sites were retrieved from the UCSC Genome Browser and visualized using the *Gviz* package (version 1.14.0) [67].

Additional files

Additional file 1: Figure S1. Exploratory data analysis of putative biological confounders. a) Variance explained per principal component based on our 450k data (turquoise triangles) versus randomly generated data (red circles). b) Dot-boxplot of the cellular composition as estimated by the Houseman algorithm [22, 23]. c) Pearson correlation coefficient (R^2) of each blood cell proportion with the principal components. The statistically significant correlations are indicated in turquoise, whereas statistically non-significant associations are indicated in red. Similar correlations were calculated for d) age and e) anti-TNF usage. (PDF 119 kb)

Additional file 2: Table S1. Annotated data for the DMPs passing BH correction for the hypothesis-free approach. The location data: Illumina probe ID, chromosome, position, strand, UCSC gene symbol, UCSC genetic feature and regulatory feature, are shown along with the methylation statistics: mean beta difference, the p values and the BH-adjusted p . (XLSX 403 kb)

Additional file 3: Figure S2. DMP-distribution per genetic feature and per chromosome. a) Comparison of the probe distribution on the 450k versus the DMP distribution per genetic feature where the different colors represent the different genetic features. The numbers along the barplot represent the percentages of the 450k probes (top) or DMPs (bottom) per genetic feature. Significantly different DMP-distributions are indicated in bold red with the asterisks indicating the level of significance as found in Additional file 5: Table S3 (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$). b) For each genetic feature the percentage hypo- and hypermethylated DMPs is indicated with barplots in black and gray respectively. (PDF 54 kb)

Additional file 4: Table S2. DMP-distribution statistics per genetic feature. Results are ordered by the first Fisher test (left three columns), which tests for differences in DMP-distribution per genetic feature, and the second Fisher test (right three columns), which tests for differences in the distribution of the hypo-/hypermethylated DMPs. Statistics provided are the odds ratios with the 95 % confidence intervals ("OR (CI-95)"), the p values and the Bonferroni-adjusted p values. (DOCX 53 kb)

Additional file 5: Table S3. DMP-distribution statistics per chromosome. Results are ordered by the first Fisher test (left three columns), which tests for differences in DMP-distribution per chromosome, and the second Fisher test (right three columns), which tests for differences in the distribution of the hypo-/hypermethylated DMPs. Statistics provided are the odds ratios with the 95 % confidence intervals ("OR (CI-95)"), the p values and the Bonferroni-adjusted p values. (DOCX 98 kb)

Additional file 6: Figure S3. Differentially methylated regions. Plots of the methylation levels of the DMRs nearest to: a) *HLA-J*, b) *MOV10L1*, c) *LINC00612*, d) *SHANK2*, e) *APOBEC1*, f) *OR2L13* and g) *TACSTD2* from the 450k ("450K") superposed onto the RefSeq gene ("RefSeq gene"), the CpG island ("CGI") and the transcription factor binding sites ("TFBS"), as retrieved from the UCSC Genome Browser. The red transparent rectangle indicates the DMR as reported by bump hunter. (PDF 420 kb)

Additional file 7: Table S5. Annotated data for the DMPs passing BH-correction for the hypothesis-driven approach. The location data: Illumina probe ID, chromosome, position, strand, UCSC gene symbol, UCSC genetic feature and regulatory feature, are shown alongside the methylation statistics: mean difference in beta, the p values and the BH-adjusted p values calculated for the hypothesis-free approach and the hypothesis-driven approach. (XLSX 26 kb)

Additional file 8: Figure S4. MiSeq validation. a) Correlation of the methylation levels obtained from 450k and MiSeq. Each color represents the gene associated to the plotted CpG. Visualization of the methylation levels in beta of the DMPs obtained from the 450k ("450K") compared to the methylation levels obtained from MiSeq sequencing ("MiSeq") superposed onto the RefSeq gene ("RefSeq gene") for b) *TNFSF4*, c) *IL10/IL19*, and d) *ORMDL3*. (PDF 146 kb)

Additional file 9: Table S6. Comparison of the top DMPs reported by McDermott et al. with our own data. The location data: Illumina probe ID, chromosome, position, and associated gene are shown alongside the effect size per study. CpGs where opposite effects were found are indicated in bold. (DOCX 15 kb)

Additional file 10: Table S7. MiSeq primers. The primer data used for MiSeq amplicon sequencing is shown alongside the included 450k probes. (DOCX 16 kb)

Additional file 11: Table S4. DMPs on chromosome X. The location data: Illumina probe ID, chromosome, position, strand, UCSC gene symbol, UCSC genetic feature and regulatory feature, are shown alongside the methylation statistics: mean difference in beta, the p values and the BH-adjusted p values. DMPs associated to either *PABPC5* or *MIR223* are indicated in bold. (XLSX 12 kb)

Abbreviations

450k, Illumina HumanMethylation450 BeadChip array; BH, Benjamini-Hochberg; CD, Crohn's disease; DMP, differentially methylated position; DMR, differentially methylated region; EWAS, epigenome-wide association study; GO, gene ontology; GWAS, genome-wide association study; IBD, inflammatory bowel disease; RUV, remove unwanted variation; TFBS, transcription factor binding sites; UC, ulcerative colitis

Acknowledgements

We are grateful to Dr. Frank Harders from the Department of Infection Biology at the University of Wageningen, the Netherlands, and Martin Haagmans and Dr. Olaf Mook from the Department of Clinical Genetics at the Academic Medical Center in Amsterdam, the Netherlands, for their contribution in generating the data obtained from the MiSeq sequencing experiment.

Availability of supporting data

The dataset supporting the results are available in the GEO repository (GSE81961).

Authors' contributions

NWD analyzed the disease status of the Crohn's disease patients, assembled the patient cohort, and isolated the DNA. JZ designed the primers for MiSeq amplicon sequencing. PH performed the MiSeq data acquisition. ANPM aligned the MiSeq sequences to the reference genome. PH and AYFLY analyzed the results acquired from the HumanMethylation 450k array and the MiSeq sequencing data. AYFLY and NWD drafted the manuscript. AAV and PH conceived the study and participated together with GRAMDH, MMAMM, NWD, and WJJ in the design and coordination of the study. All authors read and approved the final manuscript.

Competing interests

AYFLY was financially supported by GlaxoSmithKline to work on the Marie Skłodowska-Curie EpiMac project (Grant No. SEP-210163258). WJJ was financially supported by GlaxoSmithKline, Maed Johnsson, Schwabe and is a co-owner of Gut Research BV.

Ethics approval and consent to participate

The assembly of this cohort was approved by the medical ethics committee of the Academic Medical Hospital (METC 08/330 # 09.17.0268) and written informed consent was obtained from both the CD patients and control subjects.

Author details

¹Department of Clinical Genetics, Genome Diagnostics Laboratory, Academic Medical Center Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. ²Tytgat Institute for Liver & Intestinal

Research, Academic Medical Center, Amsterdam, The Netherlands.

³Department of Gastroenterology, Academic Medical Center, Amsterdam, The Netherlands. ⁴Epinova Discovery Performance Unit, GlaxoSmithKline, Stevenage, UK.

Received: 22 January 2016 Accepted: 22 May 2016

Published online: 08 June 2016

References

- Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*. 2012;142:46–54. e42.
- Wagtmans M, Verspaget H, Lamers C, Hogezaand R. Gender-related differences in the clinical course of Crohn's disease. *Am J Gastroenterol*. 2001;96:1541–6.
- Mountfield R, Prosser R, Bampton P, Muller K, Andrews JM. Pregnancy and IBD treatment: this challenging interplay from a patients' perspective. *J Crohns Colitis*. 2010;4:176–82.
- Saha S, Zhao Y-Q, Shah SA, Esposti SD, Lidofsky S, Salih S, et al. Menstrual cycle changes in women with inflammatory bowel disease: a study from the ocean state Crohn's and colitis area registry. *Inflamm Bowel Dis*. 2014; 20:534–40.
- Plavšić I, Štimac T, Hauser G. Crohn's disease in women. *Int J Womens Health*. 2013;5:681–8.
- Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491:119–24.
- McGovern DPB, Kugathasan S, Cho JH. Genetics of inflammatory bowel diseases. *Gastroenterology*. 2015;149:1163–76. e2.
- Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*. 2015;47:979.
- Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*. 2010;42:1118–25.
- Pituch-Zdanowska A, Banaszkiwicz A, Albrecht P. The role of dietary fibre in inflammatory bowel disease. *Prz Gastroenterol*. 2015;10:135–41.
- Xavier RJ, Podolsky DK. Unravelling the pathogenesis of inflammatory bowel disease. *Nature*. 2007;448:427–34.
- Harris RA, Nagy-Szakal D, Pedersen N, Opekun A, Bronsky J, Munkholm P, et al. Genome-wide peripheral blood leukocyte DNA methylation microarrays identified a single association with inflammatory bowel diseases. *Inflamm Bowel Dis*. 2012;18:2334–41.
- McDermott E, Ryan EJ, Tosetto M, Gibson D, Burrage J, Keegan D, et al. DNA methylation profiling in inflammatory bowel disease provides new insights into disease pathogenesis. *J Crohns Colitis*. 2015;10:77.
- Nimmo ER, Prendergast JG, Aldhous MC, Kennedy NA, Henderson P, Drummond HE, et al. Genome-wide methylation profiling in Crohn's disease identifies altered epigenetic regulation of key host defense mechanisms including the Th17 pathway. *Inflamm Bowel Dis*. 2012;18:889–99.
- Cooke J, Zhang H, Greger L, Silva A-L, Massey D, Dawson C, et al. Mucosal genome-wide methylation changes in inflammatory bowel disease. *Inflamm Bowel Dis*. 2012;18:2128–37.
- Daura-Oller E, Cabre M, Montero MA, Paternain JL, Romeu A. Specific gene hypomethylation and cancer: new insights into coding region feature trends. *Bioinformation*. 2009;3:340–3.
- Majumder S, Advani A. The epigenetic regulation of podocyte function in diabetes. *J Diabetes Complications*. 2015;29:1337.
- Szyf M. DNA methylation, behavior and early life adversity. *J Genet Genomics*. 2013;40:331–8.
- Karatzas PS, Gazouli M, Safioleas M, Mantzaris GJ. DNA methylation changes in inflammatory bowel disease. *Ann Gastroenterol Q Publ Hell Soc Gastroenterol*. 2014;27:125–32.
- Lin Z, Hegarty JP, Yu W, Cappel JA, Chen X, Faber PW, et al. Identification of disease-associated DNA methylation in B cells from Crohn's disease and ulcerative colitis patients. *Dig Dis Sci*. 2012;57:3145–53.
- van Iterson M, Tobi EW, Sliker RC, den Hollander W, Luijk R, Slagboom PE, et al. MethylAid: visual and interactive quality control of large Illumina 450 k data sets. *Bioinformatics*. 2014;30:3435–7.
- Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2014;15:R31.
- Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86.
- Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013;49:359–67.
- Maksimovic J, Gagnon-Bartsch JA, Speed TP, Oshlack A. Removing unwanted variation in a differential methylation analysis of Illumina HumanMethylation450 array data. *Nucleic Acids Res*. 2015;43:e106.
- Gagnon-Bartsch JA, Jacob L, Speed TP. Removing unwanted variation from high dimensional data with negative controls. Berkeley: Tech Reports from Dep Stat Univ California; 2013. p. 1–112.
- Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. *Biostatistics*. 2012;13:539–52.
- Guintivano J, Aryee MJ, Kaminsky ZA. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics*. 2013;8:290–302.
- Montaño CM, Irizarry RA, Kaufmann WE, Talbot K, Gur RE, Feinberg AP, et al. Measuring cell-type specific differential methylation in human brain tissue. *Genome Biol*. 2013;14:R94.
- Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev*. 2011;25:1010–22.
- Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*. 2012;13:484–92.
- Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol*. 2012;41:200–9.
- Hendriks WJ, Pulido R. Protein tyrosine phosphatase variants in human hereditary disorders and disease susceptibilities. *Biochim Biophys Acta*. 1832; 2013:1673–96.
- Cui L, Yu WP, DeAizpurua H, Schmidli R, Pallen C. Cloning and characterization of islet cell antigen-related protein-tyrosine phosphatase (PTP), a novel receptor-like PTP and autoantigen in insulin-dependent diabetes. *J Biol Chem*. 1996;271:24817–23.
- Lu J, Li Q, Xie H, Chen Z, Borovitskaya A, Maclaren N, et al. Identification of a second transmembrane protein tyrosine phosphatase, IA-2beta, as an autoantigen in insulin-dependent diabetes mellitus: precursor of the 37-kDa tryptic fragment. *Proc Natl Acad Sci U S A*. 1996;93:2307–11.
- Tang L, Wang L, Ye H, Xu X, Hong Q, Wang H, et al. BCL11A gene DNA methylation contributes to the risk of type 2 diabetes in males. *Experimental and Therapeutic Medicine*. 2014;8:459–63.
- Caromile LA, Oganessian A, Coats SA, Seifert RA, Bowen-Pope DF. The neurosecretory vesicle protein phogrin functions as a phosphatidylinositol phosphatase to regulate insulin secretion. *J Biol Chem*. 2010;285:10487–96.
- Yu Y, Wang J, Khaled W, Burke S, Li P, Chen X, et al. Bcl11a is essential for lymphoid development and negatively regulates p53. *J Exp Med*. 2012; 209:2467–83.
- Johnnidis JB, Harris MH, Wheeler RT, Stehling-Sun S, Lam MH, Kirak O, et al. Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. *Nature*. 2008;451:1125–9.
- Laios A, O'Toole S, Flavin R, Martin C, Kelly L, Ring M, et al. Potential role of miR-9 and miR-223 in recurrent ovarian cancer. *Mol Cancer*. 2008;7:35.
- Wong QW, Lung RW, Law PT, Lai PB, Chan KY, To K, et al. MicroRNA-223 is commonly repressed in hepatocellular carcinoma and potentiates expression of Stathmin1. *Gastroenterology*. 2008;135:257–69.
- Liu T-Y, Chen S-U, Kuo S-H, Cheng A-L, Lin C-W. E2A-positive gastric MALT lymphoma has weaker plasmacytoid infiltrates and stronger expression of the memory B-cell-associated miR-223: possible correlation with stage and treatment response. *Mod Pathol an Off J United States Can Acad Pathol Inc*. 2010;23:1507–17.
- Pan Y, Liang H, Liu H, Li D, Chen X, Li L, et al. Platelet-secreted microRNA-223 promotes endothelial cell apoptosis induced by advanced glycation end products via targeting the insulin-like growth factor 1 receptor. *J Immunol*. 2014;192:437–46.
- Wu F, Zhang S, Dassopoulos T, Harris ML, Bayless TM, Meltzer SJ, et al. Identification of microRNAs associated with ileal and colonic Crohn's disease. *Inflamm Bowel Dis*. 2010;16:1729–38.
- Fazi F, Raccanichi S, Zardo G, Starnes LM, Mancini M, Travaglini L, et al. Epigenetic silencing of the myelopoiesis regulator microRNA-223 by the AML1/E2O oncoprotein. *Cancer Cell*. 2007;12:457–66.

46. Vourekas A, Zheng K, Fu Q, Maragkakis M, Alexiou P, Ma J, et al. The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. *Genes Dev.* 2015;29:617–29.
47. Zhu X, Zhi E, Li Z. MOV10L1 in piRNA processing and gene silencing of retrotransposons during spermatogenesis. *Reproduction.* 2015;149:R229–35.
48. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods.* 2012;9:215–6.
49. Pritchard MT, Nagy LE. Ethanol-induced liver injury: potential roles for egr-1. *Alcohol Clin Exp Res.* 2005;29(11 Suppl):146S–50.
50. McMahon SB, Monroe JG. The role of early growth response gene 1 (egr-1) in regulation of the immune response. *J Leukoc Biol.* 1996;60:159–66.
51. Yu W, Lin Z, Hegarty JP, Chen X, Kelly AA, Wang Y, et al. Genes differentially regulated by NKX2-3 in B cells between ulcerative colitis and Crohn's disease patients and possible involvement of EGR1. *Inflammation.* 2012;35:889–99.
52. Prokhortchouk A, Hendrich B, Jørgensen H, Ruzov A, Wilm M, Georgiev G, et al. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev.* 2001;15:1613–18.
53. Gertz J, Savic D, Varley KE, Partridge EC, Safi A, Jain P, et al. Distinct properties of cell-type-specific and shared transcription factor binding sites. *Mol Cell.* 2013;52:25–36.
54. Alders M, Bliiek J, vd Lip K, vd Bogaard R, Mannens M. Determination of KCNQ1OT1 and H19 methylation levels in BWS and SRS patients using methylation-sensitive high-resolution melting analysis. *Eur J Hum Genet.* 2009;17:467–73.
55. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 2014;30:1363–9.
56. Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450 k methylation array data improves replication in large cancer studies. *Genome Biol.* 2014;15:503.
57. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics.* 2010;11:587.
58. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNAsequencing and microarray studies. *Nucleic Acids Res* 2015, 43:1–13.
59. Smyth GK: Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat Appl Genet Mol Biol* 2004, 3:1–25.
60. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics.* 2015;32:286–8.
61. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 2012;40:e115.
62. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics.* 2007;23:1289–91.
63. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky AM, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297–303.
64. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43:491–8.
65. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotech.* 2011;29:24–6.
66. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer; 2009.
67. Hahne F, Ivanek R: Statistical Genomics: Methods and Protocols. Edited by Mathé E, Davis S. New York, NY: Springer New York; 2016:335–51.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

