Epigenome-wide profiling identifies significant differences in DNA methylation between matched-pairs of T- and B-lymphocytes from healthy individuals.

Dr. John R. Glossop[1,2]*, Mrs. Nicola B. Nixon[2], Dr. Richard D. Emes[3], Dr. Kim E. Haworth[1], Dr. Jon C. Packham[2], Dr. Peter T. Dawes[2], Prof. Anthony A. Fryer[1], Dr. Derek L. Mattey[1,2], Prof. William E. Farrell[1]

[1]Institute for Science and Technology in Medicine, Keele University, Guy Hilton Research Centre, Thornburrow Drive, Hartshill, Stoke-on-Trent, Staffordshire, ST4 7QB, UK.

[2]Haywood Rheumatology Centre, Haywood Hospital, High Lane, Burslem, Stoke-on-Trent, Staffordshire, ST6 7AG, UK.

[3]School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington Campus, Sutton Bonington, Leicestershire, LE12 5RD, UK.

***Corresponding author:** Dr. John. R. Glossop, Institute for Science and Technology in Medicine, Keele University, Haywood Rheumatology Centre, Haywood Hospital, High Lane, Burslem, Stoke-on-Trent, Staffordshire, ST6 7AG, UK.

Tel: +44 (0) 1782 673812, Fax: +44 (0) 1782 673509

E-mail: j.r.glossop@keele.ac.uk

Conflict-of-interest and financial disclosure statement: All listed authors (JRG, NBN, RDE, KEH, JCP, PTD, AAF, DLM and WEF) have no conflicts-of-interest to declare, and have no relevant financial relationships to disclose.

Running title: Genome-wide DNA methylation profiling in healthy T- and B-lymphocytes

Key words: DNA methylation, CpG, genome-wide, Illumina 450K array, T-lymphocyte, B-lymphocyte, rheumatoid arthritis.

Abbreviations: CpG, cytosine-guanine dinucleotide; RA, rheumatoid arthritis; PCR, polymerase chain reaction; 450K array/array, HumanMethylation450 BeadChip; NIMBL, numerical identification of methylation biomarker lists; DAVID, database for annotation, visualization and integrated discovery.

Abstract

Multiple reports now describe changes to the DNA methylome in rheumatoid arthritis and in many cases have analyzed methylation in mixed cell populations from whole blood. However, these approaches may preclude the identification of cell type-specific methylation, which may subsequently bias identification of disease-specific changes. To address this possibility, we conducted genome-wide DNA methylation profiling using HumanMethylation450 BeadChips to identify differences within matched pairs of T-lymphocytes and B-lymphocytes isolated from the peripheral blood of 10 healthy females. Array data were processed and differential methylation identified using NIMBL software. Validation of array data was performed by bisulfite Pyrosequencing. Genome-wide DNA methylation was initially determined by analysis of LINE-1 sequences and was higher in B-lymphocytes than matched T-lymphocytes (69.8 vs. 65.2%, $p \leq 0.01$). Pairwise analysis identified 679 CpGs, representing 250 genes, which were differentially methylated between T-lymphocytes and B-lymphocytes. The majority of sites (76.6%) were hypermethylated in B-lymphocytes. Pyrosequencing of selected candidates confirmed the array data in all cases. Hierarchical clustering revealed perfect segregation of samples into two distinct clusters based on cell type. Differentially methylated genes showed enrichment for biological functions/pathways associated with leukocytes and T-lymphocytes. Our work for the first time shows that T-lymphocytes and B-lymphocytes possess intrinsic differences in DNA methylation within a restricted set of functionally-related genes. These data provide a foundation for investigating DNA methylation in diseases in which these cell types play important and distinct roles.

Introduction

DNA methylation, one of several prototypical epigenetic mechanisms, regulates gene expression without imposing alterations to the underlying nucleotide sequence. It is important in a variety of fundamental biological processes[1-3] and in humans occurs almost exclusively on cytosines in the context of cytosine-guanine dinucleotides (CpGs). These dinucleotides are found to cluster at high density in regions of DNA termed CpG islands, and are often associated with gene promoters.[4,5] With particular exceptions, they are typically not methylated in transcriptionally competent genes.[6] Conversely, methylation of promoter-associated CpGs is frequently associated with transcriptional repression and gene silencing.[6]

There is now compelling evidence that aberrant DNA methylation is involved in a variety of human diseases, most notably in cancer etiology and, more recently, a growing spectrum of autoimmune disorders that include rheumatoid arthritis (RA). Distinct leukocyte populations, particularly lymphocytes and their subsets, play a pivotal role in RA pathogenesis, driving autoimmunity and chronic inflammation. Levels of global DNA methylation have been found to be altered in lymphocytes and peripheral blood mononuclear cells from patients with RA.[7-9] Moreover, others have also reported aberrant methylation in RA, but in these cases have examined individual genes and using DNA samples derived from whole blood or the mononuclear cell fraction.[10-15] Although these studies have identified genes potentially involved in RA pathogenesis, they are not able to consider the distinct roles of individual leukocyte populations. Furthermore, analysis of mixed cell samples may preclude the identification of important cell type-specific and disease-specific methylation changes

due to the inherent cellular heterogeneity. Indeed, this view is supported by recent evidence indicating that leukocyte heterogeneity may be a confounder for DNA methylation analyses on whole blood,[16] and that individual leukocyte populations show evidence of differences in methylation.[17] In this context, a recent genome-wide study in RA by Liu *et al*[18] utilized a statistical algorithm to estimate and adjust for blood leukocyte proportions, concluding this to be a critical step for downstream bioinformatic analyses.

It is clear therefore that the examination of individual leukocyte populations that drive disease is an important step towards better understanding the role of DNA methylation in conditions such as RA. A prerequisite to this goal must first be to define the intrinsic differences in methylation between individual leukocyte populations. In the current work, we have taken a stringent methodological approach to address these issues by defining genome-wide DNA methylation profiles in highly purified matched pairs of T-lymphocytes and B-lymphocytes isolated from the blood of healthy individuals.

Results

450K array technical validation

Genome-wide DNA methylation at 485,577 CpG sites (99% of RefSeq genes), was examined using the HumanMethylation450 BeadChip (henceforth referred to as "450K array" or "array"; Illumina Inc.) to identify differences in methylation between matched pairs of T-lymphocytes and B-lymphocytes isolated from the blood of ten healthy females. We first performed a technical validation of the array data using an independent technique, in this case, sodium bisulfite Pyrosequencing. To this end, five CpGs were selected at random, from three separate genes (*AMN*, *HMOX2* and *PM20D1*), which represented the spectrum of possible array β-values (0 to 1). These sites, which were analyzed in paired samples from seven of the ten individuals chosen randomly, revealed excellent correlation between array β-values and methylation determined by Pyrosequencing (Spearman's $r = 0.952$, $p < 0.00001$) (**Figure 1**). Since the correlation was apparent across the spectrum of β-values it is unlikely to be consequent to the dominance of extreme values.[19]

Assessment of global DNA methylation

Methylation of the LINE-1 retrotransposon is frequently used as a surrogate measure of global DNA methylation levels.[20,21] Using sodium bisulfite Pyrosequencing we quantified methylation across 3 adjacent CpG sites within the LINE-1 sequence. Mean LINE-1 methylation was similar in each individual for each specific cell type; however, a subtle difference between the cell types was observed, 65.2% in T-lymphocytes and 69.8% in B-lymphocytes (**Figure 2**). Similar levels of methylation have been reported for peripheral blood leukocytes.[22] We noted that LINE-1 methylation was consistently

higher in B-lymphocytes for each of the three CpGs inspected. Although these differences were subtle they were statistically significant at individual sites and also collectively ($p \leq 0.01$, Wilcoxon Signed Rank Test; **Figure 2**). A more gene-focused assessment of global methylation by calculating the average β-value across all array CpGs revealed no significant difference between the cell types ($\beta = 0.562$ and 0.558 in T-lymphocytes and B-lymphocytes, respectively, $p > 0.2$).

A restricted set of differentially methylated CpGs distinguish T-lymphocytes and B-lymphocytes

After initial array processing (described in the Methods), the remaining dataset comprised 484,616 CpGs. For robust identification of CpG loci displaying differential methylation between the two cell types, we first performed a series of filtering steps in which only those CpGs meeting stringent criteria were retained. The filtering criteria and number of CpGs remaining at each of the steps are presented in **Figure 3**. As a first step to improve the power of subsequent statistics, non-variable sites (CpGs for which β-values were either ≤ 0.2, or ≥ 0.8, in all 20 samples) were removed, an approach we and others have previously employed.[23-25] For initial identification of CpGs showing a difference in methylation, we calculated the difference in β-value between B-lymphocytes and T-lymphocytes for each of the ten paired samples. We next calculated the mean β-value difference and retained only those CpGs with a mean difference ≥ 0.2 (step 2; **Figure 3**). Implementation of the subsequent steps resulted in the identification of 679 gene-associated CpGs displaying significant differences in methylation between T-lymphocytes and B-lymphocytes from the ten healthy individuals examined.

Hierarchical clustering and validation of candidate CpGs

We next used hierarchical clustering to establish whether the β-value methylation profile for the 679 CpGs identified above was able to distinguish between the two cell types across the ten individuals. As presented in the heatmap and dendrograms in **Figure 4**, a perfect dichotomy was observed, with T-lymphocyte and B-lymphocyte samples segregating into two distinct clusters (indicated in the heatmap by green and red bars, respectively). Moreover, we observed this dichotomy when clustering was performed on the number of CpGs identified at each successive stage of the filtering process, from step 3 onward and identifying in this case 2,306 statistically significant CpGs (**Supplementary Figure 1**). The data therefore clearly show that the observed differences in methylation are highly consistent between the T- and B-lymphocytes in each of the ten individuals examined.

Closer inspection of the 679 CpGs revealed that the majority (520 CpGs; 76.6%) showed higher methylation in B-lymphocytes (summary details are provided in **Table 1**). The observed differences in mean β-value were in many cases pronounced (maximum difference of 0.86) and approximately half of the CpGs identified demonstrated a β-value difference of at least 0.5 (326; 48.0%) and/or were associated with a *bona fide* CpG island (326; 48.0%).[26] In total, 250 genes were represented by the 679 CpGs, with a maximum of 18 CpGs in any one gene (a complete list of CpGs/genes and associated β-values is provided in **Supplementary Table 1**). **Table 2** identifies the 20 genes displaying the largest methylation differences between the cell types (top 10 hypermethylated genes in each cell type). Of particular note, numerous genes showing high methylation in B-lymphocytes and low methylation in T-lymphocytes were those

with known T-lymphocyte-specific expression/function, such as the CD3D and CD3G genes (subunits of the CD3 T-cell co-receptor).

For independent validation of the candidate genes identified in our genome-wide approach, we selected eight which showed large methylation differences between the cell types and which also contained at least 2 differentially methylated CpGs that could be sequenced simultaneously; four genes that were hypermethylated in B-lymphocytes (*CD3G/D*, *DENND2D, GIMAP7* and *PITPNC1*) and four that were hypermethylated in T-lymphocytes (*CLPTM1L, DDAH2, RNH1* and *SLC7A5*). Validation was performed by sodium bisulfite Pyrosequencing (assay details are provided in **Supplementary Table 2**). **Figure 5** displays data for one representative CpG from each gene examined. In each case, the large statistically significant differences observed between the cell types by array β-values were accurately replicated by Pyrosequencing ($p < 0.01$ for each CpG/gene). This was also true for each of the CpGs analyzed in each gene, as exemplified by the multiple CpGs in *RNH1* and *PITPNC1* ($p < 0.01$; **Figure 6**). Validation across these eight candidate genes reinforces the robustness of the array data and our approach to identify sites differentially methylated between T-lymphocytes and B-lymphocytes.

Differentially methylated genes show enrichment for immune and leukocyte-specific characteristics.

To explore the functional characteristics of genes identified as differentially methylated between T-lymphocytes and B-lymphocytes, we examined gene ontologies and biological pathways using the online software DAVID (Database for Annotation,

Visualization and Integrated Discovery).[27,28] Results from these analyses are provided in Supplementary Tables 3-5. Analysis of the 250 candidate genes revealed significant enrichment for multiple ontology terms (**Supplementary Table 3**). Thirty-two terms remained significant following adjustment for multiple comparisons and, of note, half of these (16/32) specifically related to 'T-cell', 'Lymphocyte' or 'Leukocyte' categories. Similar results identifying T-lymphocyte, leukocyte and immune related categories were obtained when we performed clustering on annotation terms (for example 'T-cell receptor complex' and 'T-cell activation'; **Supplementary Table 4**), and also when we interrogated the gene set for enrichment of biological pathways (for example 'T-cell receptor signaling pathway' and 'T helper cell surface molecules'; **Supplementary Table 5**). Genes displaying disparate methylation between the cell types are thus enriched for lymphocyte-related functions.

Discussion

DNA methylation represents one of several principal targets for epigenetic change and acts in concert with other epigenetic modifications to regulate gene expression. A burgeoning literature describes these types of change in multiple diseases, including those with an autoimmune component such as RA, in which leukocytes are known to play a pivotal role. The majority of previous studies examining methylation in leukocytes from RA patients have employed DNA extracted from whole blood or the mononuclear cell fraction.[7-15,18] However, in these cases cellular heterogeneity may be a confounding factor. To determine if intrinsic differences in methylation are a feature of individual leukocyte populations, we compared genome-wide DNA methylation profiles in matched pairs of T-lymphocytes and B-lymphocytes isolated from the blood of healthy individuals. Through this approach, our unequivocal observation is that a restricted set of CpGs in functionally-related genes show differential methylation between T-lymphocytes and B-lymphocytes, and define a DNA methylation signature that accurately distinguishes the two lymphocyte types.

We identified a set of 679 gene-associated CpGs, representing 250 unique genes, which displayed altered methylation between T-lymphocytes and B-lymphocytes. For many of these sites striking differences in methylation were observed (β-value differences up to 0.864), resulting in the perfect segregation of T- and B-lymphocytes into two distinct clusters. In common with other tissues/cell types, the majority of CpG islands in T-lymphocytes and B-lymphocytes are unmethylated.[34,35] There is also evidence from a variety of different tissues and cell types, including leukocytes, that some CpG islands show tissue-specific methylation.[36,37] In our study, a small proportion (9%) of the 679

CpGs differentially methylated between T- and B-lymphocytes were located within a *bona fide* CpG island (as defined by Takai and Jones[26]). Of interest however, approximately three-times as many CpGs (25.3%) resided within the associated CpG island shores, with a further 13.7% within island shelves. Thus, our data indicate that the majority of intrinsic differences in methylation between T-lymphocytes and B-lymphocytes occur not within CpG islands *per se* but within the shores surrounding these islands (≤2 kb from the island). Of particular note, the importance of these CpG island "associated regions" in the context of gene expression has recently been described[38] and is worthy of further investigation. Interestingly, analysis of global DNA methylation, using methylation of the LINE-1 retrotransposon as a surrogate measure, revealed marginal, yet statistically significant higher methylation in B-lymphocytes. To our knowledge this comparative increase has not previously been identified, and it is currently unclear as to its significance.

Those genes identified as unmethylated in T-lymphocytes but hypermethylated in B-lymphocytes included a significant number with known T-lymphocyte expression and functionality, such as *ITK* (a T-lymphocyte kinase), *CD2* (early T-lymphocyte marker), and the genes for the delta, gamma and epsilon subunits of the CD3 T-cell co-receptor (*CD3D*, *CD3G* and *CD3*, respectively) which is expressed exclusively on T-lymphocytes. Although not experimentally addressed in this study, our data suggest methylation-based epigenetic regulation of the CD3 locus, a finding that has recently been reported for leukocyte subsets.[17] Analysis of functional enrichment by examining gene ontologies and pathways confirmed the predominance of genes associated with lymphocyte and leukocyte functioning amongst those differentially methylated between

the lymphocyte types. Similar enrichment of leukocyte/lymphocyte genes was previously reported by Reinius and colleagues,[17] but in individual comparisons of each cell type with methylation in whole blood.

In our investigation and for the purpose of defining intrinsic genome-wide differences in methylation between T-lymphocytes and B-lymphocytes, we adopted a stringent methodological approach. First, we isolated the individual cell populations using a validated technique that achieved mean purity in excess of 99% and 93% for each cell type respectively. Second, we implemented a series of stringent filtering criteria and independent validation to robustly identify those sites/genes displaying the most significant differences in methylation between the lymphocyte types. Finally, we only considered a gene to show altered methylation between the cell types if at least two CpGs from the same gene demonstrated statistically significant β-value differences ≥ 0.2 and in the same direction (*i.e.* higher/lower in the same cell type), as used with success in a previous investigation from our laboratory.[23] Independent validation of 19 CpGs from eight selected candidate genes confirmed the large differences observed, and emphasizes the robustness of our data.

Previous studies have reported subtle but significant differences in DNA methylation between tissue-types and between individuals,[25] and additional reports provide evidence for associations of this epigenetic change with demographical factors such as age, gender and ethnicity,[39,40] and lifestyle factors including smoking.[41,42] Herein, and partly prompted by this literature, we focused our analyses on two lymphocyte populations in a carefully selected and closely matched, albeit small, cohort of healthy individuals

(Caucasian, non-smoking females, approximately 50-55 years of age). With regard to the limited variation in subject age, inspection of the results from hierarchical clustering revealed no evidence to suggest that younger and older individuals clustered separately. We are confident therefore, that our findings are not unduly influenced by potentially confounding variables such as those discussed above.

The important issue of cellular heterogeneity as a confounder for DNA methylation analyses of blood-derived DNA has previously been recognised.[16-18] Indeed, in the context of disease, underlying changes in the proportion of the different leukocyte populations may bias the identification of disease-specific changes in methylation. To address this issue a statistical algorithm has recently been described[43] and provides an accurate method to estimate and adjust for leukocyte proportions in genome-wide DNA methylation studies.[18,43] However, although this approach represents a significant advance, it is not able to determine the particular leukocyte populations in which disease-specific changes in DNA methylation occur. In these cases these changes represent an important question for diseases where different, specific, leukocyte populations play central roles in the disease process. In our own study, we have first set out to establish the intrinsic differences in methylation between matched pairs of purified T- and B-lymphocytes from healthy individuals. In this way we provide a foundation from which disease-specific changes of DNA methylation in these cell types may be determined.

In summary, we have presented a unique genome-wide DNA methylation signature that accurately distinguishes healthy T-lymphocytes from B-lymphocytes. We recognize that

it will be important for future efforts to define methylation levels and disease-specific aberrations in these cell populations in RA patients, and to perhaps extend these types of analyses to individual cell subsets. The different T-helper subsets, for example, will be of particular interest given their distinct and multifaceted role in RA.

Patients and Methods

Patients

Peripheral blood samples (30 ml, EDTA) were collected from 10 healthy female Caucasians. Donors were recruited from the hospital staff population at the Haywood Rheumatology Centre, Stoke-on-Trent, and gave prior written informed consent. Each donor completed a short questionnaire to establish basic demographical information and to provide information on their current health status. The average age of the group was 52.1 ± 5.0 years (mean ± SD) and none were current smokers. Individuals who reported any musculoskeletal or inflammatory-related conditions were excluded. The study was approved by the East Midlands (Derby) Research Ethics Committee.

Isolation of individual lymphocyte populations

Peripheral blood mononuclear cells were isolated from fresh whole blood samples by density gradient centrifugation using Histopaque-1077 (Sigma-Aldrich). $CD3^+$ T-lymphocytes were purified from the mononuclear cell fraction using anti-CD3 magnetic microbeads, with subsequent isolation of $CD19^+$ B-lymphocytes from the negative portion using anti-$CD19^+$ magnetic microbeads (MiniMACS® separation system; Miltenyi Biotec). We chose to use positive selection as this has previously been demonstrated to be superior for array-based analyses of leukocytes[44] and has minimal effects on leukocyte activation.[45-48] All separations were performed according to the manufacturer's instructions (Miltenyi Biotec). Cell preparations were immediately pelleted, lysed and homogenized (QIAShredder columns; Qiagen). Lysates were stored at –80°C.

Antibody staining and flow cytometry

$CD3^+$ T-lymphocytes and $CD19^+$ B-lymphocytes from the ten donors were assessed for purity using flow cytometry. T-lymphocytes were stained with a murine phycoerythrin (PE)-labeled anti-CD3 IgG monoclonal antibody, and B-lymphocytes with a murine PE-anti-CD20 IgG monoclonal antibody. Each sample was separately stained with the appropriate PE-labeled isotype-matched control antibody. Antibodies were purchased from Miltenyi Biotec. Samples were analyzed on a FACSort flow cytometer and raw data collected using CellQuest software (BD Biosciences). A forward scatter threshold of 180 was set and 10,000 events/sample were counted. Data were analyzed using Cyflogic software (version 1.2.1; CyFlo Ltd). Representative histograms of cell purity for each cell type are provided in **Supplementary Figure 2**. The mean ± SD purity for T-lymphocytes and B-lymphocytes from the ten donors was 99.6 ± 0.4% and 93.6 ± 3.4%, respectively.

Genome-wide DNA methylation analysis by Infinium 450K array

Genomic DNA was extracted from thawed lysates using an AllPrep DNA/RNA Mini kit (Qiagen) according to the manufacturer's instructions. Further purification and concentration of DNA was performed using ethanol-based precipitation reactions. DNA concentration was assessed using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific).

Quantification of genome-wide DNA methylation was conducted using Infinium-based HumanMethylation450 BeadChips. Each array enables simultaneous quantification of methylation at 485,577 unique CpGs across the genome, representing 21,231 RefSeq

genes (99% coverage), in 12 independent samples in parallel.[49] Array hybridizations were performed by Gen-Probe Life Sciences Ltd (Manchester, UK) according to Illumina recommended protocols. Briefly, 650 ng of genomic DNA was sodium bisulfite converted using an EZ DNA Methylation Kit (Zymo Research). Subsequently, 4 µl of bisulfite-converted DNA was hybridized to 450K arrays following the Illumina Infinium HD Methylation protocol. Each sample was arrayed once. Arrays were scanned using an Illumina iScan and image intensities extracted using the GenomeStudio Methylation Module (version 1.8.5). The methylation status of each CpG/probe was reported as a methylation β-value, where β is defined as the ratio of the methylated signal intensity over the summed intensity of the methylated and unmethylated signals + 100.[49] β-values were reported on a continuous linear scale ranging from 0 (unmethylated) to 1 (completely methylated).

Whole genome amplification and touchdown PCR

Sodium bisulfite-converted DNA samples used on the arrays were separately used for validation of array data and to confirm the methylation status of identified candidate CpGs/genes. To increase template quantity, samples were first subjected to whole genome amplification according to a protocol described by Mill *et al*.[50] Briefly, 1 µl bisulfite-converted DNA was amplified by polymerase chain reaction (PCR) using GoTaq Flexi DNA polymerase and buffer (Promega), dNTPs (Bioline) and random 15-mer oligonucleotide primers (200 pmoles; Life Technologies). Thermal cycling conditions were: 98°C for 3 minutes, followed by 50 cycles of 98°C for 1 minute, 37°C for 2 minutes and 55°C for 4 minutes, with a ramping rate of 0.1°C/second prior to each denaturation and extension step.

PCR amplicons that encompassed CpGs of interest were generated from whole genome amplified DNA using touchdown PCR.[51,52] Forward, reverse and sequencing primers specific for bisulfite-converted DNA were designed using PyroMark Assay Design software (version 2.0; Qiagen) and were purchased from Biomers.net (primer sequences are provided in **Supplementary Table 2**). In each 25 µl PCR reaction, 2-4 µl whole genome-amplified DNA was amplified with the appropriate forward and reverse primers, as we have described previously.[53]

Validation by sodium bisulfite Pyrosequencing

Methylation at individual CpG sites was quantified by Pyrosequencing of bisulfite-converted PCR-amplified DNA using a PyroMark Q24 instrument (Qiagen). Pyrosequencing was performed according to the manufacturer's instructions and as we have previously described.[23,53] The sequence to analyze for each CpG/gene is provided in **Supplementary Table 2** and all assays included one or more control dispensations to confirm sequence identity and completeness of bisulfite conversion. Data were analyzed using PyroMark Q24 software (v 2.0.6., build 20; Qiagen).

Data analysis

Array data were analyzed using NIMBL software.[54] Each array passed initial quality control assessment based on the performance of internal array controls, and the distribution of β-values across all 485,577 CpGs was found to be similar in each sample. All CpGs for which one or more of the 20 samples displayed detection p-values > 0.05 (indicating an unreliable site) or presented with missing β-values were excluded.

As we have described previously, we considered it a more robust approach to remove from the dataset CpGs that failed in any one of the samples rather than excluding individual failed CpGs from specific samples.[24] To adjust for the observation that β-values derived from Infinium II probes on the 450K array can be less sensitive than those generated from Infinium I probes, NIMBL was used to perform 'peak-based correction' as devised by Dedeurwaerder *et al*.[55]

Criteria for identification of differentially methylated CpGs are described in detail in the Results. Statistical analysis of DNA methylation at individual CpG sites was compared between T-lymphocytes and B-lymphocytes using paired t-tests. Correction for multiple testing was performed using the false discovery rate (FDR) adjustment of Benjamini and Hochberg,[29] with adjusted p-values < 0.05 considered significant. Euclidian-based hierarchical clustering of differentially methylated CpGs was performed using Genesis software (v1.7.6).[56] Differentially methylated genes were assessed for enrichment of gene ontology terms and biological pathways using DAVID online software.[27,28] Additional statistical analyses were performed using NCSS 2000 (NCSS LCC). P-values < 0.05 were considered significant.

Acknowledgements

The authors wish to thank Cuong V. Duong and Dr. Mark Kitchen for assistance with Pyrosequencing assays. This work was supported by funding provided by the Haywood Rheumatism Research and Development Foundation.

References

1. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 1992; 69:915-26.
2. Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. Nature 1993; 366:362-5.
3. Mohandas T, Sparkes RS, Shapiro LJ. Reactivation of an inactive human X-chromosome: evidence for X inactivation by DNA methylation. Science 1981; 211:393-6.
4. Bird AP. CpG-rich islands and the function of DNA methylation. Nature 1986; 321:209-13.
5. Antequera F, Bird A. Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci USA 1993; 90:11995-9.
6. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev 2002; 16:6-21.
7. Richardson B, Scheinbart L, Strahler J, Gross L, Hanash S, Johnson M. Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis. Arthritis Rheum 1990; 33:1665-73.
8. Corvetta A, Della Bitta R, Luchetti MM, Pomponio G. 5-Methylcytosine content of DNA in blood, synovial mononuclear cells and synovial tissue from patients affected by autoimmune rheumatic diseases. J Chromatogr 1991; 566:481-91.
9. Liu CC, Fang TJ, Ou TT, Wu CC, Li RN, Lin YC, et al. Global DNA methylation, DNMT1, and MBD2 in patients with rheumatoid arthritis. Immunol Lett 2011; 135:96-9.

10. Schwab J, Illges H. Silencing of CD21 expression in synovial lymphocytes is independent of methylation of the CD21 promoter CpG island. Rheumatol Int 2001; 20:133-7.
11. Fu LH, Cong B, Zhen YF, Li SJ, Ma CL, Ni ZY, et al. [Methylation status of the IL-10 gene promoter in the peripheral blood mononuclear cells of rheumatoid arthritis patients]. Yi Chuan 2007; 29:1357-61.
12. Nile CJ, Read RC, Akil M, Duff GW, Wilson AG. Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis. Arthritis Rheum 2008; 58:2686-93.
13. Kitamura T, Kabuyama Y, Kamataki A, Homma MK, Kobayashi H, Aota S, et al. Enhancement of lymphocyte migration and cytokine production by ephrinB1 system in rheumatoid arthritis. Am J Physiol Cell Physiol 2008; 294:C189-96.
14. Fu LH, Ma CL, Cong B, Li SJ, Chen HY, Zhang JG. Hypomethylation of proximal CpG motif of interleukin-10 promoter regulates its expression in human rheumatoid arthritis. Acta Pharmacol Sin 2011; 32:1373-80.
15. Ishida K, Kobayashi T, Ito S, Komatsu Y, Yokoyama T, Okada M, et al. Interleukin-6 gene promoter methylation in rheumatoid arthritis and chronic periodontitis. J Periodontol 2012; 83:917-25.
16. Adalsteinsson BT, Gudnason H, Aspelund T, Harris TB, Launer LJ, Eiriksdottir G, et al. Heterogeneity in white blood cells has potential to confound DNA methylation measurements. PLoS One 2012; 7:e46705.
17. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One 2012; 7:e41361.

18. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol 2013; 31:142-7.
19. Roessler J, Ammerpohl O, Gutwein J, Hasemeier B, Anwar SL, Kreipe H, et al. Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. BMC Res Notes 2012; 5:210.
20. Yang AS, Estecio MR, Doshi K, Kondo Y, Tajara EH, Issa JP. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. Nucleic Acids Res 2004; 32:e38.
21. Weisenberger DJ, Campan M, Long TI, Kim M, Woods C, Fiala E, et al. Analysis of repetitive element DNA methylation by MethyLight. Nucleic Acids Res 2005; 33:6823-36.
22. Zhu ZZ, Hou L, Bollati V, Tarantini L, Marinelli B, Cantone L, et al. Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis. Int J Epidemiol 2012; 41:126-39.
23. Duong CV, Emes RD, Wessely F, Yacqub-Usman K, Clayton RN, Farrell WE. Quantitative, genome-wide analysis of the DNA methylome in sporadic pituitary adenomas. Endocr Relat Cancer 2012; 19:805-16.
24. Fryer AA, Emes RD, Ismail KM, Haworth KE, Mein C, Carroll WD, et al. Quantitative, high-resolution epigenetic profiling of CpG loci identifies associations with cord blood plasma homocysteine and birth weight in humans. Epigenetics 2011; 6:86-94.
25. Byun HM, Siegmund KD, Pan F, Weisenberger DJ, Kanel G, Laird PW, et al. Epigenetic profiling of somatic tissues from human autopsy specimens identifies

tissue- and individual-specific DNA methylation patterns. Hum Mol Genet 2009; 18:4808-17.

26. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci USA 2002; 99:3740-5.
27. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc 2009; 4:44-57.
28. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009; 37:1-13.
29. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B 1995; 57:289-300.
30. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000; 28:27-30.
31. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 2006; 34(Database issue):D354-7.
32. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res 2008; 36(Database issue):D480-4.
33. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res 2013; 41(Database issue):D377-86.

34. Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP. A human B cell methylome at 100–base pair resolution. Proc Natl Acad Sci USA 2009; 106:671-8.
35. Hughes T, Webb R, Fei Y, Wren JD, Sawalha AH. DNA methylome in human CD4+ T cells identifies transcriptionally repressive and non-repressive methylation peaks. Genes Immun 2012; 11:554-60.
36. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 2006; 38:1378-85.
37. Illingworth R, Kerr A, DeSousa D, Jorgensen H, Ellis P, Stalker J, et al. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol 2008; 6:e22
38. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 2009; 41:178-86.
39. Liu J, Morgan M, Hutchison K, Calhoun VD. A study of the influence of sex on genome wide methylation. PLoS One 2010; 5:e10028.
40. Zhang FF, Cardarelli R, Carroll J, Fulda KG, Kaur M, Gonzalez K, et al. Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. Epigenetics 2011; 6:623-9.
41. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. Am J Hum Genet 2011; 88:450-7.

42. Zochbauer-Muller S, Lam S, Toyooka S, Virmani AK, Toyooka KO, Seidl S, et al. Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. Int J Cancer 2003; 107:612-6.
43. 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.
44. Lyons PA, Koukoulaki M, Hatton A, Doggett K, Woffendin HB, Chaudhry AN, et al. Microarray analysis of human leucocyte subsets: the advantages of positive selection and rapid purification. BMC Genomics 2007; 8:64.
45. Semple JW, Allen D, Chang W, Castaldi P, Freedman J. Rapid separation of CD4+ and CD19+ lymphocyte populations from human peripheral blood by a magnetic activated cell sorter (MACS). Cytometry 1993; 14:955-60.
46. Zahler S, Kowalski C, Brosig A, Kupatt C, Becker BF, Gerlach E. The function of neutrophils isolated by a magnetic antibody cell separation technique is not altered in comparison to a density gradient centrifugation method. J Immunol Methods 1997; 200:173-9.
47. Lea T, Smeland E, Funderud S, Vartdal F, Davies C, Beiske K, et al. Characterization of human mononuclear cells after positive selection with immunomagnetic particles. Scand J Immunol 1986; 23:509-19.
48. Funderud S, Erikstein B, Asheim HC, Nustad K, Stokke T, Blomhoff HK, et al. Functional properties of CD19+ B lymphocytes positively selected from buffy coats by immunomagnetic separation. Eur J Immunol 1990; 20:201-6.
49. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG resolution. Genomics 2011; 98:288-95.

50. Mill J, Yazdanpanah S, Guckel E, Ziegler S, Kaminsky Z, Petronis A. Whole genome amplification of sodium bisulfite-treated DNA allows the accurate estimate of methylated cytosine density in limited DNA resources. Biotechniques 2006; 41:603-7.
51. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 1991; 19:4008.
52. Hecker KH, Roux KH. High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. Biotechniques 1996; 20:478-85.
53. Haworth KE, Farrell WE, Emes RD, Ismail KM, Carroll WD, Borthwick HA, et al. Combined influence of gene-specific cord blood methylation and maternal smoking habit on birth weight. Epigenomics 2013; 5:37-49.
54. Wessely F, Emes RD. Identification of DNA methylation biomarkers from Infinium arrays. Front Genet 2012; 3:161.
55. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. Epigenomics 2011; 3:771-84.
56. Sturn A, Quackenbush J, Trajanoski Z. Genesis: cluster analysis of microarray data. Bioinformatics 2002; 18:207-8.

Figure legends

Figure 1. Technical validation of 450k array data by sodium bisulfite Pyrosequencing. Correlation of DNA methylation as measured by array β-value and by bisulfite Pyrosequencing for 70 individual sites in T-lymphocytes and B-lymphocytes from seven of the ten healthy individuals (5 CpGs were selected at random, from three separate genes: *HMOX2:* cg14951292, *AMN*: cg09616556, *PM20D1*: cg24503407, cg07167872, cg11965913). T- and B-lymphocytes are shown as circles and crosses, respectively. Spearman's $r = 0.952$, $p < 0.00001$.

Figure 2. LINE-1 DNA methylation in matched pairs of T-lymphocytes and B-lymphocytes. Sodium bisulfite Pyrosequencing was used to quantify methylation at three CpG sites within LINE-1 repetitive sequences in purified T-lymphocytes and B-lymphocytes from ten healthy individuals. T- and B-lymphocytes are shown as circles and crosses, respectively. The mean level of methylation for each CpG site in each cell population is shown by the short black horizontal bar in each case.

* $p \leq 0.01$ (Wilcoxon Signed-Rank Test).

Figure 3. Criteria for filtering and identification of CpGs differentially methylated between paired T-lymphocyte and B-lymphocyte samples. The starting number of 484,616 CpGs was derived through removal of CpGs with high detection p-values ($p > 0.05$) and those with missing β-values, as described in the Methods.

† 'Change' refers to a methylation difference which fulfilled all preceding filtering steps.

‡ 'Change in the same direction' refers to multiple CpGs from one gene all showing either hyper- or hypomethylation in one cell type as compared with the other cell type.

Figure 4. Heatmap and clustering for the 679 CpGs identified as differentially methylated. CpG sites identified by 450K array analysis as differentially methylated between paired T-lymphocytes (green bar) and B-lymphocytes (red bar) were analyzed by hierarchical clustering. Each row represents an individual CpG site and each column a different sample (listed beneath the heatmap). Branches of the dendrograms indicate similarity between CpGs (rows) and samples (columns). Color gradation from yellow to blue represents low to high DNA methylation respectively, with β-values ranging from 0 (no methylation; yellow) to 1 (complete methylation; blue). Sample order, left-to-right: B-lymphocytes: HC03, 10, 04, 09, 11, 14, 08, 12, 16, 17; and T-lymphocytes: HC03, 08, 09, 11, 12, 04, 14, 16, 10, 17.

Figure 5. Validation of differentially methylated CpG candidates identified by 450K array analysis. Sodium bisulfite Pyrosequencing was used to confirm candidate genes/CpGs that were differentially methylated between T-lymphocytes and B-lymphocytes in all ten healthy individuals. CpG sites in four genes that were hypermethylated in T-lymphocytes (A) and four genes that were hypermethylated in B-lymphocytes (B) were selected from the array candidates. T- and B-lymphocytes are shown by circles and crosses, respectively. Gene names are shown on the x-axis, where for each gene methylation values are shown for the array (left) and Pyrosequencing (right). The mean values are defined by the horizontal black bar in each case. The methylation differences observed between T-lymphocytes and B-lymphocytes by

Pyrosequencing analysis were statistically significant for each of the eight CpGs/genes examined ($p < 0.01$, Wilcoxon Signed-Rank Test).

Abbreviations: Pyro., bisulfite Pyrosequencing.

Figure 6. Representative plots displaying validation of multiple array CpGs within two candidate genes by sodium bisulfite Pyrosequencing analysis. Four CpGs within the RNH1 gene (A) and three within PITPNC1 (B), all of which demonstrated significant differences in methylation between T-lymphocytes and B-lymphocytes by array analysis, were examined. T- and B-lymphocytes are shown as circles and crosses, respectively. CpG identifiers are shown on the x-axis (in each case, the first site shown for RNH1 (A) and PITPNC1 (B) is the site depicted in Figure 5A and 5B, respectively). For each site, methylation values are shown for the array (left) and Pyrosequencing (right). The mean values are defined by the horizontal black bar in each case. The CpG sites are displayed in the 5' to 3' direction, with the four sites in RNH1 covering 20-bp and three sites in PITPNC1 covering 45-bp. The methylation differences observed between T-lymphocytes and B-lymphocytes by Pyrosequencing analysis were statistically significant for each of the CpGs in each gene examined ($p < 0.01$, Wilcoxon Signed-Rank Test).

Abbreviations: Pyro., bisulfite Pyrosequencing.

Supplementary Figure 1. Heatmap and clustering for the 2,306 CpGs identified as differentially methylated. Sites on the 450K array were identified as differentially methylated between paired T-lymphocyte (green bar) and B-lymphocyte (red bar) samples at step 3 of the filtering flow diagram. Each row represents an individual CpG

site and each column a different sample (listed beneath the heatmap). Branches of the dendrograms indicate similarity between CpGs (rows) and samples (columns). Color gradation from yellow to blue represents low to high DNA methylation respectively, with β-values ranging from 0 (no methylation; yellow) to 1 (complete methylation; blue). Sample order, left-to-right: B-lymphocytes: HC03, 10, 04, 09, 11, 14, 08, 12, 16, 17; and T-lymphocytes: HC03, 08, 09, 11, 10, 12, 04, 14, 16, 17.

Supplementary Figure 2. Representative flow cytometry histograms depicting cell purity for the T-lymphocyte and B-lymphocyte isolation procedures. The purity of each T-lymphocyte and B-lymphocyte isolation procedure was assessed quantitatively by flow cytometry. Shown for each of three healthy individuals – HC08 (red trace), HC12 (blue trace) and HC16 (green trace) – are cell purity histograms for the T-lymphocyte (A) and B-lymphocyte (B) isolations. Representative isotype-matched control data (HC12) is shown by the black trace with grey fill, where the horizontal bracket indicates background staining for 95% of the isotype control which was used as the threshold to determine CD3/CD20 positive staining. T-lymphocytes were stained with a murine PE-labeled anti-CD3 IgG monoclonal antibody, and B-lymphocytes with a murine PE-anti-CD20 IgG monoclonal antibody. The overall mean ± SD purity for the study cohort (n = 10) was 99.6 ± 0.4% and 93.6 ± 3.4% for T-lymphocytes and B-lymphocytes, respectively.

Supplementary Table 1. 450K array β-value matrix and accompanying calculations for the 679 CpGs identified as differentially methylated between T-lymphocytes and B-lymphocytes.

*Unless otherwise stated, values in the data matrix are methylation array β-values. The solid black horizontal line at row 522 separates CpGs hypermethylated in B-lymphocytes (above) from those hypermethylated in T-lymphocytes (below).

† FDR, false discovery rate adjustment for multiple comparisons.[29]

Supplementary Table 2. List of genes, associated CpGs and primer sequences interrogated by bisulfite Pyrosequencing.

*Further information that is not included here is available upon request. The horizontal dashed line at row 10 separates candidate genes hypermethylated in T-lymphocytes (above) and hypermethylated in B-lymphocytes (below).

†The prefix 'b-' denotes biotin labeling at the 5' end of the primer.

‡The sequence indicated is post-bisulfite conversion and the letters 'Y' and 'R' denote the cytosine of interrogated CpG sites ('Y' and 'R' refer to sequencing along the upper and lower strands respectively).

Abbreviations: bp, base pairs

Supplementary Table 3. Gene ontology categories identified by analysis using DAVID as significantly enriched in the 250 genes differentially methylated between T-lymphocytes and B-lymphocytes.

*DAVID, Database for Annotation, Visualization and Integrated Discovery.[27,28] Data shown are the output from running the Functional Annotation Chart tool on the three parent gene ontology domains (Biological Process, Cellular Component and Molecular Function) and all sub-categories contained within. Listed are all ontologies that were statistically significant at the $p < 0.05$ level using the EASE score.[27,28] The solid black

horizontal line at row 34 indicates the cut-off for enriched ontology terms that remained statistically significant following adjustment for multiple comparisons using the FDR method (False Discovery Rate[29]).

†Gene ontology terms listed belong to the parent ontology indicated, where BP, CC and MF indicate biological process, cellular component and molecular function, respectively.

‡P-values are calculated from the EASE score, a modified version of Fisher's Exact Test.[27,28]

Supplementary Table 4. Clustering using DAVID of similar gene ontology annotation terms, based on common membership of enriched genes, identified in the 250 genes differentially methylated between T-lymphocytes and B-lymphocytes.

*DAVID, Database for Annotation, Visualization and Integrated Discovery.[27,28] Data shown are the output from running the Functional Annotation Chart tool on the three parent gene ontology domains (Biological Process, Cellular Component and Molecular Function) and all sub-categories contained within. Listed are all ontologies statistically significant at the $p < 0.05$ level using the EASE score.[27,28] The solid black horizontal line at row 58 indicates the cut-off for clusters containing enriched ontology terms of which at least half remained statistically significant following adjustment for multiple comparisons using the FDR method (False Discovery Rate[29]).

†Gene ontology terms listed belong to the parent ontology indicated, where BP, CC and MF indicate biological process, cellular component and molecular function, respectively.

‡P-values are calculated from the EASE score, a modified version of Fisher's Exact Test.[27,28]

Supplementary Table 5. Pathways identified by analysis using DAVID as significantly enriched in the 250 genes differentially methylated between T-lymphocytes and B-lymphocytes.

*DAVID, Database for Annotation, Visualization and Integrated Discovery.[27,28] Data shown are the output from running the Functional Annotation Chart tool on selected pathway databases. Listed are all pathways that were statistically significant at the $p < 0.05$ level using the EASE score.[27,28] The solid black horizontal line at row 13 indicates the cut-off for enriched pathways that remained statistically significant following adjustment for multiple comparisons using the FDR method (False Discovery Rate[29]).

†Pathways listed are from the KEGG[30-32], Biocarta and PANTHER[33] databases, as indicated.

‡P-values are calculated from the EASE score, a modified version of Fisher's Exact Test.[27,28]

Table 1. Summary characteristics of the 679 CpGs identified as differentially methylated between paired T-lymphocyte and B-lymphocyte samples by genome-wide DNA methylation analysis using 450K arrays.*

	Hypermethylated in T-lymphocytes	Hypermethylated in B-lymphocytes	All differentially methylated CpGs
CpGs differentially methylated†	159 (23.4)	520 (76.6)	679 (100)
In CpG islands	27 (17.0)	34 (6.5)	61 (9.0)
In CpG island shores	60 (37.7)	112 (21.5)	172 (25.3)
In CpG island shelves	14 (8.8)	79 (15.2)	93 (13.7)
Non-island CpGs	58 (36.5)	295 (56.7)	353 (52.0)
Maximum β-value difference‡	0.712	0.864	0.864
CpGs with β-value difference ≥ 0.5	22 (13.8)	304 (58.5)	326 (48.0)
Genes represented	67	183	250
Maximum number of CpGs in a gene§	6	18	18

Unless otherwise indicated, all figures are the number (%).

*Differentially methylated CpGs were identified according to the criteria described in the Results section and Figure 3.

†Annotation describing CpG island, shore and shelf status is derived from the University of California at Santa Cruz (UCSC) database. A significant association between cell type and the distribution of CpGs according to CpG island status was observed ($p < 0.00001$, Chi-squared test).

‡Maximum β-value difference refers to the largest mean difference in β-value observed for any single CpG between T-lymphocytes and B-lymphocytes (mean value calculated from ten pairwise comparisons).

§Refers to the maximum number of differentially methylated CpGs identified within a single gene.

Table 2. Annotation for the top 10 hypermethylated genes in T-lymphocytes and top 10 hypermethylated genes in B-lymphocytes, as determined by genome-wide DNA methylation analysis.*

Gene symbol†	CpGs differentially methylated	Mean (range) β-value difference‡	Gene name†	Functional summary†
Hypermethylated in T-lymphocytes				
CTSZ	3	0.612 (0.55 - 0.70)	Cathepsin Z	Lysosomal cysteine proteinase
RNF11	2	0.586 (0.56 - 0.61)	Ring finger protein 11	Protein-protein interactions
PIK3R2	2	0.576 (0.55 - 0.60)	Phosphoinositide-3-kinase, regulatory subunit 2 (beta)	Lipid kinase, growth signaling pathways
FGR	3	0.566 (0.46 - 0.71)	Feline Gardner-Rasheed sarcoma viral oncogene homolog	Protein tyrosine kinase, cell migration and adhesion
ELMO1	2	0.533 (0.50 - 0.57)	Engulfment and cell motility 1	Phagocytosis and cell migration
LYL1	2	0.519 (0.47 - 0.57)	Lymphoblastic leukemia derived	Transcription factor, blood vessel

			sequence 1	maturation and hematopoiesis
WIPI2	5	0.504 (0.33 - 0.71)	WD repeat domain, phosphoinositide interacting 2	WD40 repeat protein, multiprotein complex assembly
CD82	2	0.481 (0.46 - 0.50)	CD82 molecule	Transmembrane 4 superfamily glycoprotein
LRP5	2	0.470 (0.27 - 0.67)	Low density lipoprotein receptor-related protein 5	Lipoprotein receptor, skeletal homeostasis
CLPTM1L	2	0.465 (0.46 - 0.47)	CLPTM1-like	Anti-apoptotic, lung cancer
Hypermethylated in B-lymphocytes				
SPATA13	2	0.814 (0.79 - 0.84)	Spermatogenesis associated 13	Guanine nucleotide exchange factor, cell migration and adhesion assembly
CD3D	4	0.808 (0.76 - 0.85)	CD3d molecule, delta (CD3-TCR complex)	T-cell co-receptor, T-cell receptor signalling
TNRC6B	3	0.788 (0.74 - 0.82)	Trinucleotide repeat containing 6B	RNA-mediated gene silencing
CD3G	5	0.783 (0.75 - 0.85)	CD3g molecule, gamma (CD3-	T-cell co-receptor, T-cell receptor signaling

			TCR complex)	
UBASH3A	7	0.775 (0.69 - 0.83)	Ubiquitin associated and SH3 domain containing A	T-cell ubiquitin ligand, T-cell signaling
FYB	4	0.767 (0.68 - 0.84)	FYN binding protein	T-cell adapter protein, IL-2 expression
MGAT4A	3	0.756 (0.72 - 0.82)	Mannosyl (alpha-1,3-)-glycoprotein beta-1, 4-N-acetylglucosaminyltransferase, isozyme A	Glycosyltransferase, Golgi structural branching
PLCG1	2	0.745 (0.69 - 0.80)	Phospholipase C, gamma 1	Intracellular transduction of receptor-mediated tyrosine kinase activators
EXOC1	2	0.743 (0.69 - 0.79)	Exocyst complex component 1	Exocytic vesicle targeting, cytoskeletal remodeling
MPHOSPH9	2	0.734 (0.72 - 0.74)	M-phase phosphoprotein 9	Cell cycle

*Differentially methylated CpGs were identified according to the criteria described in the Results section and Figure 3.

†The official gene symbol, gene name and stated function were retrieved from the NCBI Gene database (accessed May 2013).

‡The mean β-value difference was calculated as the average of the individual β-value differences determined for each differentially methylated CpG within the indicated gene (the difference for each individual CpG was calculated as described in the Results).