

A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network

Seung-Tae Lee^{1,2}, Yuanyuan Xiao¹, Marcus O. Muench^{3,4}, Jianqiao Xiao¹, Marina E. Fomin³, John K. Wiencke⁵, Shichun Zheng⁵, Xiaoqin Dou⁵, Adam de Smith¹, Anand Chokkalingam⁶, Patricia Buffler⁵, Xiaomei Ma⁷ and Joseph L. Wiemels^{1,*}

¹Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA 94158, USA, ²Department of Laboratory Medicine & Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 135710, Korea, ³Blood Systems Research Institute, ⁴Liver Center and Department of Laboratory Medicine, ⁵Department of Neurological Surgery, University of California San Francisco, San Francisco, CA 94158, ⁶Division of Epidemiology, School of Public Health, University of California Berkeley, Berkeley, CA 94720 and ⁷Department of Epidemiology and Public Health, Yale University, New Haven, CT 06520, USA

Received June 15, 2012; Revised September 14, 2012; Accepted September 22, 2012

ABSTRACT

The epigenetic changes during B-cell development relevant to both normal function and hematologic malignancy are incompletely understood. We examined DNA methylation and RNA expression status during early B-cell development by sorting multiple replicates of four separate stages of pre-B cells derived from normal human fetal bone marrow and applied high-dimension DNA methylation scanning and expression arrays. Features of promoter and gene body DNA methylation were strongly correlated with RNA expression in multipotent progenitors (MPPs) both in a static state and throughout differentiation. As MPPs commit to pre-B cells, a predominantly demethylating phenotype ensues, with 79% of the 2966 differentially methylated regions observed involving demethylation. Demethylation events were more often gene body associated rather than promoter associated; predominantly located outside of CpG islands; and closely associated with *EBF1*, *E2F*, *PAX5* and other functional transcription factor (TF) sites related to B-cell development. Such demethylation events were accompanied by TF occupancy. After commitment,

DNA methylation changes appeared to play a smaller role in B-cell development. We identified a distinct development-dependent demethylation signature which has gene expression regulatory properties for pre-B cells, and provide a catalog reference for the epigenetic changes that occur in pre-B-cell leukemia and other B-cell-related diseases.

INTRODUCTION

B-lymphopoiesis is a highly coordinated process initiating from pluripotent hematopoietic stem cells (HSCs) and involves multiple developmental stages. Multipotent progenitors (MPPs) develop into common lymphoid progenitors (CLPs), B-progenitor (pro-B), B-precursor (pre-B), immature B-cell and mature B-cell stages, each of which is characterized by distinct biological features (1,2). The key events during early stages include commitment to B-lineage and suppression of non-B and stem cell components, which are gradually intensified as the developmental stages proceed through hierarchical lineage priming by different transcription factors (TFs) (3,4). It is now believed that an orchestrated network of TFs has a fundamental role in B-cell development (5,6). These TFs are shown to be indispensable in expressing functioning B-cell proteins and altering the genetic landscape, including immunoglobulin V(D)J recombination.

*To whom correspondence should be addressed. Tel: +415 514 0577; Fax: +415 502 7411; Email: joe.wiemels@ucsf.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Recent evidence suggests that TF networks are closely related to DNA methylation, an important means of gene regulation (7,8). During tissue-specific development, DNA hypermethylation has been implicated in the stable silencing of stem cell-associated TFs, such as POU5F1 (also known as OCT4), SOX2 and NANOG, whereas DNA demethylation has been implicated in the activation of differentiation-associated TFs and their target genes (9–11). In the case of B-cell development, the key TF genes include *SP1* (PU.1), *BCL11A*, *TCF3* (E2A), *IKZF1* (Ikaros), *EBF1*, *PAX5* and *FOXO1*, and their target genes include *IL7R*, *RAG1*, *RAG2*, *VPREB1*, *IGLL1* ($\lambda 5$), *CD79A* (mb-1), *CD19*, *BLNK*, *IRF4* and many others (3–5). Although key B-cell regulators have generally been studied on a gene-by-gene basis in mouse models, genome-wide studies in human cells are urgently needed for comprehensive understanding of the transcriptional and epigenetic signatures of B-cell development.

In the present study, we used genome-wide arrays to analyse the dynamic DNA methylation and expression changes during B-cell development, including the use of isolated and purified MPPs, pre-B-I cells, pre-B-II cells and immature B-cells from bone marrow isolates. We provide a reference methylation and expression map as well as a catalog of key DNA methylation changes during B-cell development.

MATERIALS AND METHODS

Acquisition and sorting of human fetal B-cells

Human fetal bone marrow (FBM) was obtained from elective abortions at San Francisco General Hospital with the consent of the women undergoing the surgical procedures and with the approval of the University of California San Francisco's Committee on Human Research. FBM was extracted as previously described (12) from specimens from eight foetuses ranging in age between 20 weeks and 24 weeks of gestation as estimated based on foot length. Four stages of B-cell precursors were isolated via flow cytometry sorting (Supplementary Figure S1). Early progenitors were isolated based on high levels of CD34 protein expression (CD34⁺⁺) and a lack of expression of the B-cell marker CD19. This population, designated as stage 1 (S1), contains MPPs before lineage commitment (predominantly MPPs) but also containing CLPs and stem cells. B-cell-committed progenitors were isolated based on their expression of CD19 and CD34 (CD19⁺CD34⁺), which were predominantly pre-B-I cells and were designated as stage 2 (S2). Two immature B-cell populations were isolated that expressed CD19, but not CD34, and were differentiated based on surface IgM (sIgM) expression: stage 3 cells (S3) were predominantly pre-B-II cells that express sIgM⁻CD19⁺; stage 4 cells (S4) were predominantly immature B-cells that express sIgM⁺CD19⁺.

Infinium DNA methylation profiling and data preprocessing

Genomic DNA was isolated from eight FBM specimens using the AllPrep DNA/RNA Mini Kit (Qiagen).

For each specimen, both DNA and total RNA was extracted for four purified cell populations representing S1–S4 (Supplementary Figure S1), giving rise to a total of 32 isolations. DNA was modified by sodium bisulfite to convert unmethylated cytosines to uracil using the EZ DNA Methylation kit (Zymo Research, Orange, CA). The bisulfite-treated DNA of six specimens (24 samples) was amplified and hybridized onto the Illumina HumanMethylation450 Beadchip (HM450; Illumina, San Diego, CA) according to the manufacturer's specifications. Raw data were processed using the GenomeStudio software (Illumina), and the average methylation beta values (β) ranging from 0 (unmethylated) to 1 (fully methylated) were retrieved for each probe. CpG sites with detection P -values $>1.0 \times 10^{-4}$ were removed from analysis. Methylation samples with a high proportion ($>4\%$) of suboptimal data were eliminated from analysis. This resulted in the removal of two samples with poor data quality.

RNA expression and data preprocessing

Total RNA was isolated from the four purified cell populations, S1–S4 (Supplementary Figure S1), from each of the eight specimens as described above. Genome-wide gene expression analysis was performed using the GeneChip Human Gene 1.0 ST Array (Affymetrix, Santa Clara, CA). For each sample, 1 μ g of high-quality total RNA was amplified and labeled (NuGen Ovation) and hybridized onto the microarray according to the manufacturer's specifications. Data were processed and normalized using the Expression Console software (Affymetrix) with the standard Robust Multi-array Average (RMA) algorithm (13). One sample was excluded because of poor data quality, resulting in a total of 31 samples from 4 developmental stages in 8 fetuses.

ChIP-PCR analysis

Chromatin immunoprecipitation (ChIP) was done as previously described (14) with modifications. Briefly, sorted S1 and S3 cells (see above) from normal bone marrow were crosslinked for 15 min at room temperature with 1% (wt/vol) formaldehyde (sigma, F8775), then lysed and sheared to 100–1000 bp in size with ChIP-IT Express Enzymatic Shearing Kit (Active Motif, Cat# 53035) according to manufacturer's instruction. Sheared chromatin was immunoprecipitated with 10 μ g anti-EBF1 antibody (Abnova, H00001879-D01P) or 10 μ g anti-BCL11A antibody (abcam, Cat# ab19489) or non-specific IgG (Santa Cruz Biotechnology, Cat# sc-2025, performed as a negative control) using EZ-ChIPTM Chromatin Immunoprecipitation Kit (Millipore, Cat# 17-371) following the kit instruction manual. The agarose beads were washed and the bound chromatin was eluted. The ChIPed and input chromatin were incubated overnight at 65°C for reversal of crosslinking. Samples were then treated with RNase A and proteinase K and purified with a Spin Filter from EZ ChIP—a Chromatin Immunoprecipitation Kit (Millipore, Cat# 17-371). Chromatin was end repaired, ligated to a linker

and amplified by linker-mediated PCR (LM-PCR) using GeneAmp High Fidelity PCR System (ABI, Cat# 4328212). Amplified ChIP and input DNA were purified with MinElute Reaction Cleanup Kit (Qiagen, Cat# 28204) and normalized to ~ 30 ng/ μ l.

Thirty nanograms DNA (pull-down and input DNA, normalized for DNA concentration) was analysed by SYBR green real-time PCR and the CFX384 Real-Time PCR System (BioRad, CA, USA). Primers were designed to four predicted EBF1 targets and four independent BCL11A targets (predicted from ENCODE data). Primer sequences and amplicon locations are available on request. A two-step PCR cycling method was used: Preheat at 95°C for 10 min, and repeat 40 times at 95°C for 15 s, 60°C for 1 min. We calculated the fold differences in eight candidate loci between ChIP DNA and the input DNA using the following expression: $\text{Fold Difference} = 2^{(Cq_{\text{Input DNA}} - Cq_{\text{Ip DNA}})}$. The experiment was performed in quadruplicate, and a separate experiment starting from chromatin was repeated once.

Statistical analysis

Statistical analysis and data visualization were carried out using the *R* statistical software with Bioconductor packages. Methylation or expression levels between two groups were compared using moderated *t*-statistics in a mixed-effect model to accommodate correlations between samples from the same donor (15). *P*-values were adjusted with the Benjamini–Hochberg correction (16). To investigate the association between differential expression and differential methylation, a robust linear regression model was used with the differential expression level (\log_2 fold changes) as the outcome and the differential methylation level (beta differences) as the predictor, adjusted for baseline expression and methylation levels at the preceding stage. Confidence intervals and *P*-values were derived for the regression coefficients using 2000 bootstrapped samples. Fisher's exact test was used to compare the equality of two binomial probabilities. Enrichment for TF motifs was done using the Analysis of Motif Enrichment (AME) (17) and MEME Suite (18); sequences 100 bp (± 50 bp) around significant CpGs were retrieved and analysed with core motifs from the TRANSFAC Professional (<http://www.gene-regulation.com>) database. *E*-values below 0.001 were regarded as significant. DNA sequences, positional information of the NCBI reference genes, the ENCODE chromatin immunoprecipitation sequencing (ChIP-Seq) data, and CpG island (CGI) annotations were retrieved from the UCSC database (<http://genome.ucsc.edu/>). Chip-Seq enrichment analysis was performed by examining whether differentially methylated regions (DMRs) are located within TF binding sites of the ENCODE Chip-Seq database and by comparing their frequencies with those of all CpG sites on HM450 (Fisher's exact test). Pathway enrichment analysis was done using the Ingenuity Pathway Analysis software (Ingenuity Systems, Redwood City, CA).

RESULTS

Static methylation and expression patterns in MPPs (S1)

The MPPs (S1) serve as the baseline DNA methylation levels before the cells are committed to any lineages. At S1, similar to a previous report (19), we observed lower methylation levels near the promoter region compared to the body of a gene (Supplementary Figures S2A and B). Interestingly, while CGIs were in general depleted of methylation, we found that CGIs located in the body region were more methylated and more variable than CGIs sites lying within promoters (Supplementary Figures S2B).

We next compared methylation levels in the S1 samples with the histone modification ChIP-seq data of the lymphoblastoid cell line GM12878 from the ENCODE project (20) as DNA methylation is believed to interact with histone modifications to regulate gene expression (7,8). There was a strong negative correlation between DNA methylation levels and the presence of histone marks that target actively transcribed genes. The median methylation value was low in H3K4Me1, H3K4Me3 and H3K79me2 peaks, which preferentially mark enhancers, promoters and transcriptional transition regions, respectively, whereas the value was high in repressive mark, H3K27me3. In addition, Polycomb target genes tended to be hypomethylated whereas TET1 target genes (derived from human homologs of mouse TET1 targets) tended to be semimethylated (Figure 1A) at S1. TET2 targets would be more appropriate to examine in hematopoietic cells but have not been described. Similar methylation profiles were also observed in the other stages (results not shown).

Although it has been well-established that methylation in the promoter region is involved in transcriptional silencing (21), little is known about how body methylation affects gene expression and whether promoter and body methylation regulate expression in concert. To examine this, we partitioned promoter and body methylation levels into hypo- ($\beta \leq 0.3$), semi- ($0.3 < \beta \leq 0.7$) and hypermethylated ($\beta > 0.7$) groups and compared gene expression levels according to these different combinations in Figure 1B. The most highly expressed group was that with both promoter and body hypomethylated whereas the lowest was with both regions hypermethylated, and the mean difference in gene expression between these two groups was 2.8-fold ($P < 2.2 \times 10^{-16}$, two-sample *t*-test). Interestingly, when the gene body region is devoid of CpG sites, the repressive nature of promoter methylation is at its most potent with a striking 3.6-fold difference in expression between the hypo- and hypermethylated promoter groups whereas the same difference is 2-fold in genes with body CpG sites (Figure 1B, right). We found many housekeeping genes such as GUK1, CENPB, RPS2, NME2 and UBE2M belong to this group of genes lacking body methylation sites. To formally model the relationship between expression and methylation, a robust linear regression model was fit to the expression levels with promoter and body methylation and their interactions as predictors, and the regression coefficients were -1.55 (promoter, $P < 0.001$ using 2000

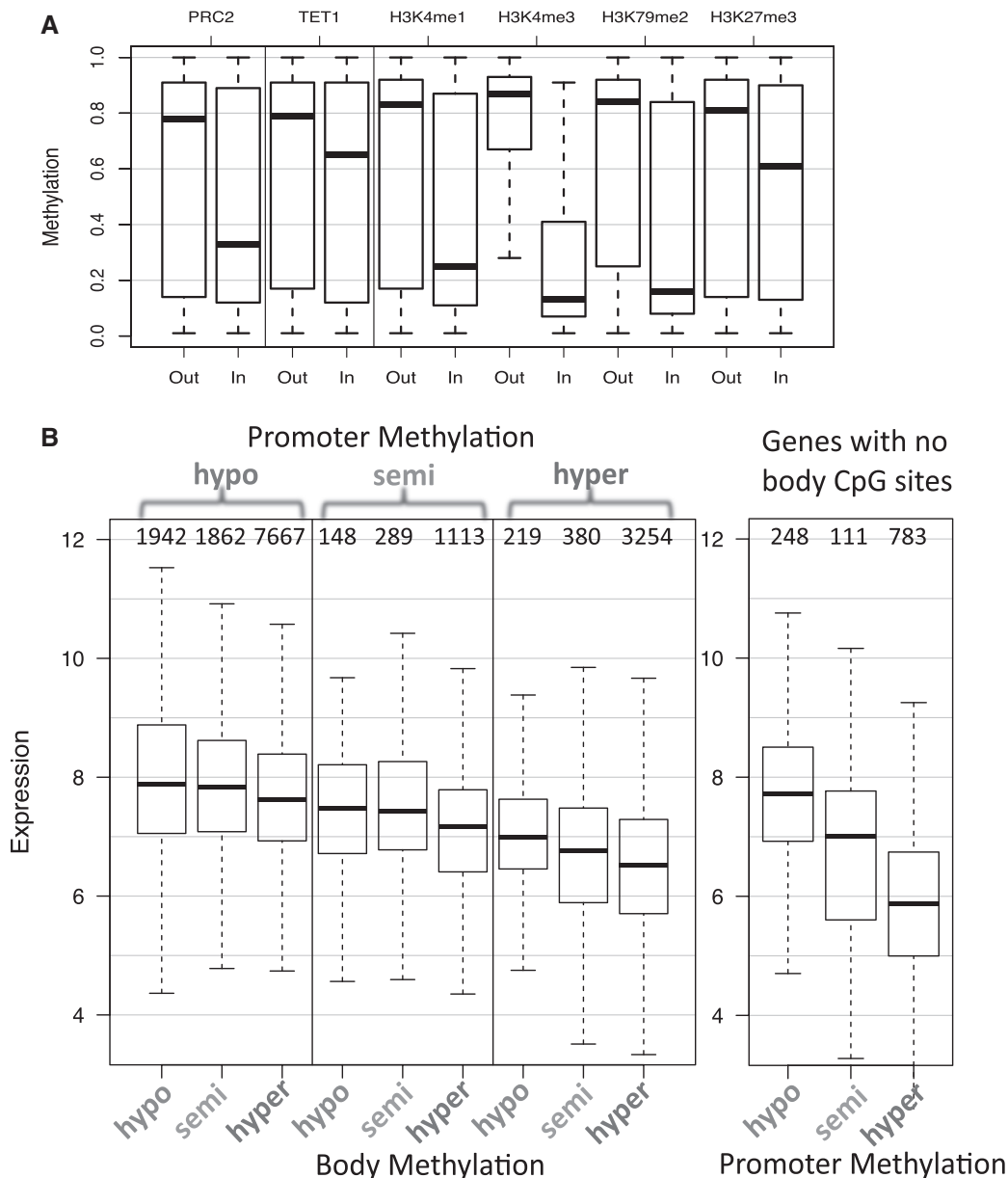


Figure 1. Baseline methylation and expression levels at S1. (A) Methylation levels inside and outside of annotation categories, including histone modifications, Polycomb target genes and TET1 target genes. (B) Association of DNA methylation with gene expression. Median methylation levels of CpG sites mapped to the promoter region and the body region of a gene were obtained together with its gene expression level as measured by the Affymetrix Gene 1.0 ST Array. Methylation levels were then categorized as hypomethylated, semimethylated and hypermethylated based on the following beta cutoffs, respectively: $\beta \leq 0.3$, $0.3 < \beta \leq 0.7$ and $\beta > 0.7$. Genes that have no body-associated CpG sites (mainly small housekeeping genes) are plotted at the right side of the figure. The number of genes in each group was labeled at the top of the graph. The most highly expressed group was that with both promoter and body hypomethylated whereas the lowest was with both regions hypermethylated, and the mean difference in gene expression between these two groups was 2.8-fold ($P < 2.2 \times 10^{-16}$, two-sample *t*-test).

bootstrapping), -0.46 (body, $P < 0.001$) and 0.05 (interaction, $P = 0.36$), respectively. This quantification contrasts the extent to which promoter and body methylation controls expression and confirms that promoter methylation exerts stronger control over transcription. But intriguingly, when promoter methylation is held constant, increasing body methylation is associated with more repressed expression, suggesting a fine-tuning mechanism of body methylation for transcription regulation.

Alterations in DNA methylation associated with pre-B-cell development reveals a dominant demethylation signature

After examining the static methylation and expression profiles of the lymphoid precursor (S1), we next sought to identify DMRs between any two subsequent stages (S1→S2, S2→S3 and S3→S4). We chose the CpG sites that had Benjamini–Hochberg-adjusted P -values ≤ 0.05 for the moderated *t*-test and β -difference ($\Delta\beta$) ≥ 0.2 . Although two distinct types of DMRs were observed during stage progression, those that had a loss of

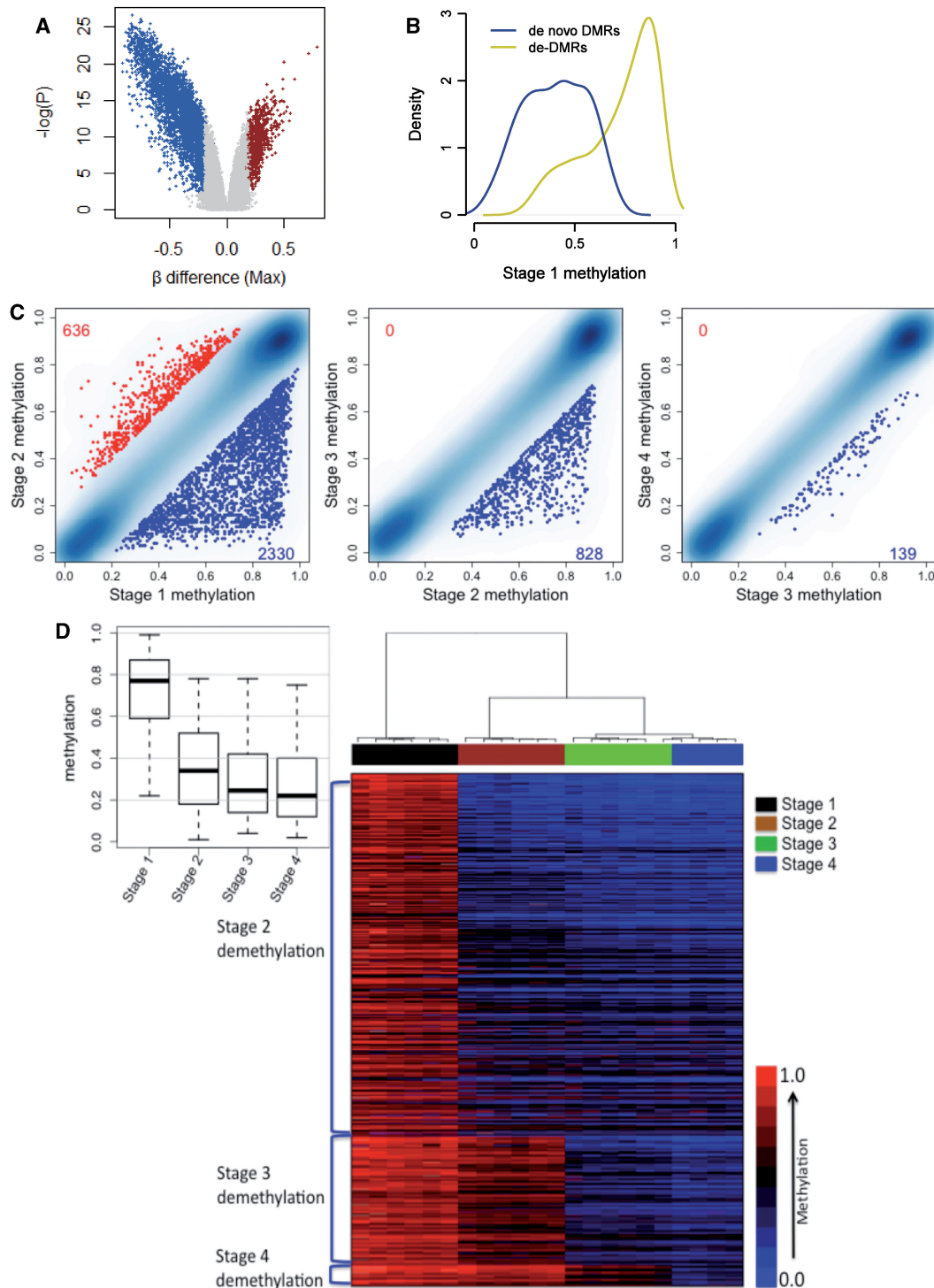


Figure 2. Differential methylation analysis of B-cell developmental stages (stages 1–4; S1–S4). DMRs between any two subsequent stages were identified if the FDR-adjusted P -value for the t -test was ≤ 0.05 and the $|\Delta\beta|$ value was ≥ 0.2 . (A) A volcano plot with maximal $-\log P$ and maximal $|\Delta\beta|$ between any two subsequent stages for each CpG site shows a negative skewness of $\Delta\beta$. (B) Density plots of the methylation level at S1 for CpG loci that are identified as de-DMRs and *de novo* DMRs. (C) Scatter plots of methylation levels of any two subsequent stages (S1 versus S2, S3 versus S2 and S4 versus S3). Red and blue dots highlight *de novo* and de-DMRs, respectively. (D) A hierarchical clustering of the 22 samples using the core de-DMRs-classified samples according to their stages. (E) Association of core de-DMRs with non-CGI regions and gene body regions. (F) ChIP-PCR analysis of four BCL11A (top) four EBF1 (bottom) targets. ENCODE-described TF binding sites which were also de-DMRs were pulled down using ChIP techniques in S1 and S3 cells. The amount of each targeted sequence was compared in the pull-down relative to the input DNA using quantitative PCR, and the fold enrichment in S1 and S3 cells shown in the figure. The analysis demonstrated that S3 cells harbored between 10-fold (for CD22/EBF1 pull-down) and 3073-fold (for NACA2/BCL11a pull-down) more sequence bound to the indicated TFs compared to input, in most cases orders of magnitude more than in S1 cells. Non-specific IgG control pull-down displayed at maximum only a 3.1-fold enrichment of target sequence compared to control (data not shown). The experiment was performed in technical quadruplicate and repeated once from the chromatin, with similar results.

(continued)

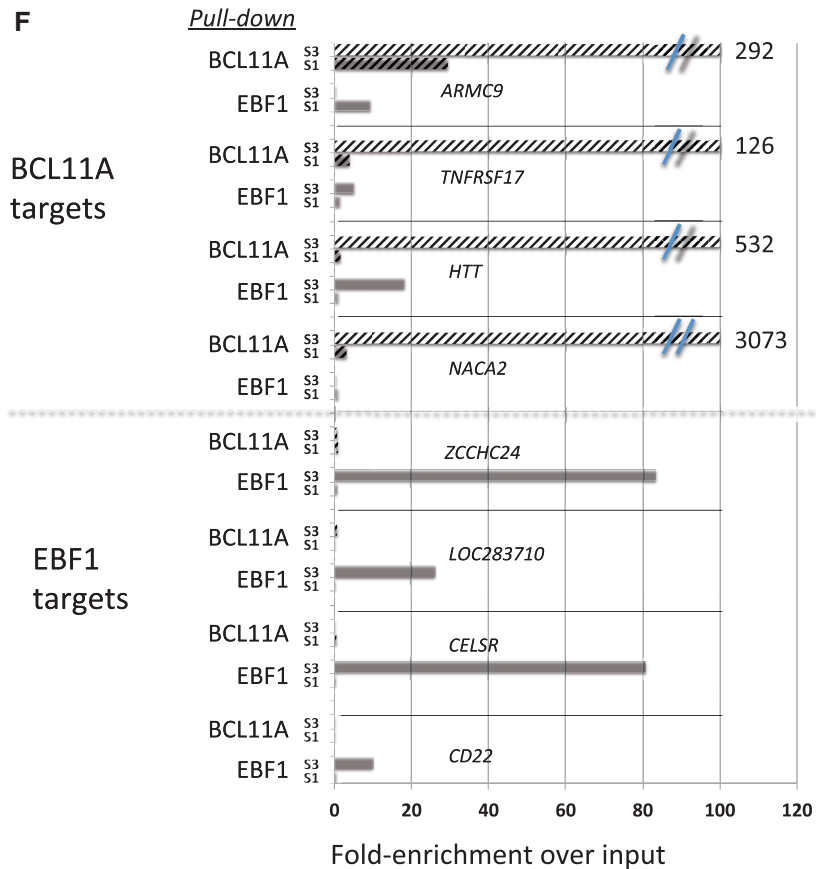
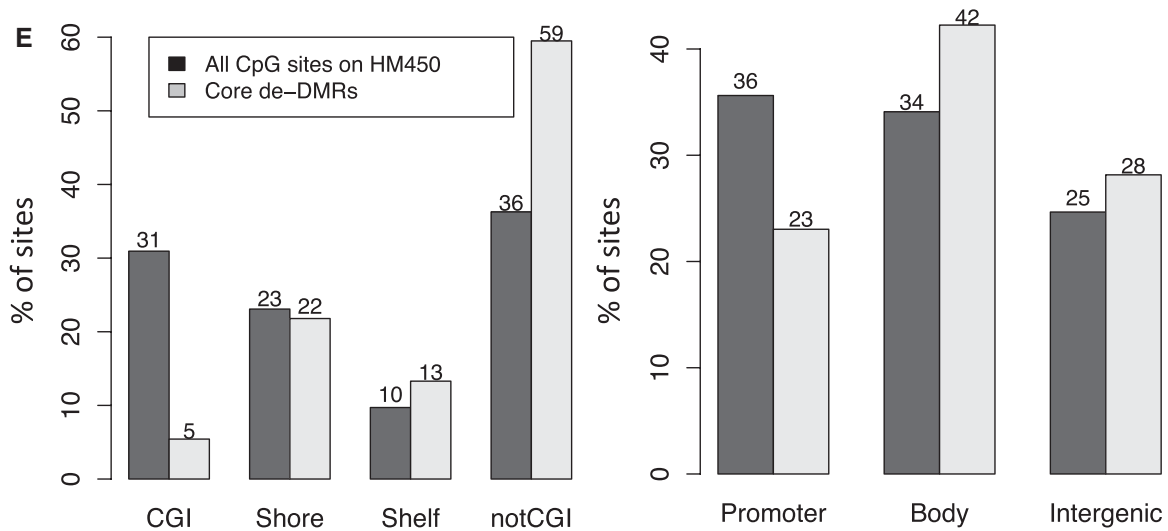


Figure 2. Continued.

methylation, which we designate as *de-DMRs*, and those that obtained more methylation, which we designate as *de novo* DMRs, *de-DMRs* were conspicuously dominant during stage progression. A scatter plot between $-\log_{10} P$ and $\Delta\beta$ of any two sequential stages showed that significant CpG sites were skewed to negative $\Delta\beta$ values (Figure 2A). We next assembled the 2987 unique *de-DMRs* along the stage progression and designated them as the ‘core *de-DMRs*’; as expected, these loci were

likely to be more methylated at S1 than *de novo* DMRs (Figure 2B). During S1→S2, *de-DMRs* (2330 CpG sites; 1039 genes) outnumbered *de novo* DMRs (636 CpG sites; 421 genes) by almost 4 to 1. DMRs between S2→S3, and S3→S4 are exclusively *de-DMRs*, 828 and 139 CpG sites (455 and 81 genes), respectively (Figure 2C). Not only did *de-DMRs* far outnumber *de novo* DMRs loci, there was also a significant overlap between these two sets from S1→S2 ($P < 2.2 \times 10^{-16}$), which suggests that

demethylation plays a more significant role during pre-B-cell development but *de novo* methylation likely plays only a secondary and accessory role.

We observed only a small number of DMRs between S3 and S4, suggesting that these two stages are very similar in terms of their methylation profiles. A hierarchical clustering of the 22 samples using the 2987 core de-DMRs perfectly classified each sample into its relevant stage, with S3 and S4 samples closely assigned into the same sub-tree (Figure 2D). Interestingly, both the clustering heatmap and the boxplots indicate that loss of methylation is a continuous event during pre-B-cell development and once a CpG site incurs a loss of methylation, it remains or becomes even more demethylated throughout subsequent stages (Figure 2D).

The core de-DMRs in B-cell development predominantly occurred in the CpG sites that are located in the gene body regions and outside of the CGIs, and depleted in promoter regions within the CGIs (Figure 2E). Among the core de-DMRs, 23% and 42% map to the promoter and body regions of a gene, respectively; a 1.8-fold preference for the body regions, in contrast to the background frequencies by all the CpG sites on the HM450 array, which have roughly equal amount of promoter or body CpG sites (36% and 34%, respectively). Interestingly, only 35 among the >1000 genes demethylated from S1 to S2 acquired both promoter and gene body changes, indicating that methylation status in these two regions were altered by independent processes. The partiality toward CpG sites outside of CGIs is even more striking with only 5% of the core de-DMRs located inside of the CGIs as opposed to 31% background frequency on the array (Figure 2E).

Using Chip-Seq enrichment data from the ENCODE project which combines data from 95 different cell lines and 152 transcription-factor (TF) targeting antibodies, DMRs were frequently observed in locations associated with specific TFs; with specific TF enrichment being stage dependent. From S1 to S2, EBF binding sites are very significantly enriched for de-DMRs (fold enrichment = 10.94; Fisher's exact test $P < 2.2 \times 10^{-16}$, Table 1), supporting its principal role in B-cell development. On the other hand, *de novo* DMRs were significantly enriched for CEBPB, c-Jun and p300. From S2 to S3, de-DMRs were preferentially enriched with IRF4 and MEF2A binding sequences (Table 1). This specific enrichment of TF binding sites in B-cell development DMRs was further confirmed by motif enrichment analysis using AME and the TRANSFAC motifs. From S1 to S2, EBF-related motifs, V\$EBF_Q6 and V\$OLF1_01, were the most significantly associated with de-DMRs followed by other ETS and E2A motifs, whereas *de novo* DMRs were significantly enriched for PU.1, SPI1, ELF1 and SPIB motifs (Supplementary Table S1). To test whether the demethylation events at TF binding sites was functionally significant, we performed ChIP-PCR analysis with two of the most significant TFs, EBF1 and BCL11A, comparing multipotential cells (S1) to committed pre-B cells (S3). ChIP-PCR analysis revealed that predicted targets were not occupied in S1 cells but strongly and specifically occupied for each TF during S3 (Figure 2F).

Alterations in RNA expression associated with pre-B-cell development

We next analysed the concomitant RNA expression changes within the same samples in order to examine

Table 1. Fold enrichment of DMRs for TF binding sites derived from ENCODE ChIP-Seq data

TFs	Known functions	Fold enrichment ^a			
		<i>de novo</i> DMR (S1–S2)	de-DMR (S1–S2)	de-DMR (S2–S3)	de-DMR (S3–S4)
EBF1	Early B-cell development	1.96	10.94	8.28	5.65
BCL11A	Late B-cell development. Interact with BCL6	5.36	8.48	16.06	15.44
BATF	Interacts with Jun family proteins	8.12	5.84	8.91	11.60
PAX5	Early B-cell development	1.13	4.08	6.00	3.46
MEF2A	Control cell growth and apoptosis	2.21	3.17	5.88	6.58
IRF4	Late B-cell development. Regulate interferon	2.95	3.14	6.44	7.00
TCF12	Development of B- and T-cells	0.96	2.82	3.66	5.97
PU.1	Myeloid and B-cell development	4.73	1.94	4.18	7.58
p300	Cell proliferation and differentiation	9.98	1.65	2.01	0.92
c-Jun	Bind to DNA and regulate expression	6.07	1.24	1.48	1.64
CEBPB	Myeloid development	5.13	0.46	0.62	0.53
TAF1	General TF (TFIID)	0.55	0.33	0.37	0.61
GABP	Control of mitochondrial function	0.91	0.21	0.15	0.33
Sin3Ak	Co-repressor interacting with HDAC1 and MECP2	0.38	0.19	0.28	0.39
HEY1	Neurogenesis and cardiovascular development	0.54	0.16	0.14	0.08
IRF1	Activator of interferon- α and - β	0.50	0.11	0.08	0.24
Pol2	RNA polymerase II, large subunit	0.23	0.09	0.05	0.00
E2F1	Cell cycle control	0.30	0.09	0.08	0.26
E2F6	Cell cycle control	0.74	0.08	0.09	0.00
HMGN3	Control chromatin fiber	0.80	0.08	0.14	0.28
CCNT2	Cell cycle control	1.10	0.07	0.12	0.24
E2F4	Cell cycle control	0.13	0.00	0.00	0.00

^aFold enrichments were calculated by dividing the percentage of DMRs overlapping with a specific TF binding sites to the percentage of all loci on HM450 overlapping with the same TF binding sites. Significant enrichments/depletions (Bonferroni-adjusted Fisher's exact test P -values < 0.05) are bolded.

RNA expression changes associated with pre-B-cell development, and moreover, to distil functional methylation changes that had transcriptional consequences. To nominate differentially expressed genes (DEGs) between any two ensuing stages, we used the thresholds of adjusted P -values ≤ 0.05 (Benjamini–Hochberg adjustment) for moderated t -test and fold-change ≥ 1.5 , which resulted in a total of 3913 DEGs. Unlike DNA methylation that was predominantly one-directional with hypomethylation being the driving force, RNA expression was a bi-directional process with similar numbers of up- and down-regulated genes in total. There was, however, a stage dependent preference toward up-regulation or down-regulation. During transition from S1 to S2, up-regulated DEGs were about two times more prevalent than down-regulated ones (1996 versus 675 genes); this together with the concomitant predominant demethylation leading to S2 suggests a possible role of methylation-regulated transcriptional changes during the initial B-cell lineage commitment. In contrast, during S3–S4 progression, down-regulated genes significantly outnumbered up-regulated genes (949 versus 364 genes; Figure 3A), this in combination with just 81 genes (corresponding to 139 DMRs) that are demethylated

from S3 suggests that DNA methylation might not play a major role in regulating gene expression leading to the progression to immature B-cells. As expected, hierarchical clustering using these 3913 DEGs accurately classified the 31 samples into 4 different stages, and S2 and S3 were the most similar in expression profiles (Figure 3B). The putative B-cell-related TF and surface antigen genes such as *TCF3*, *IKZF1*, *EBF1*, *PAX5*, *CD19*, *CD79A* and *BLNK* were up-regulated throughout all stages. Genes involved in immunoglobulin rearrangement or functioning as temporary surrogate markers, such as *RAG1*, *RAG2*, *VPREB1* and *IGLL1*, were activated during S2 and S3 and down-regulated again at S4. The early progenitor genes, *SOX2* and *SPI1*, were gradually down-regulated as well as non-B lineage genes such as *GF11*, *GATA1* and *NOTCH3*. DNA methyltransferase genes including *DNMT1*, *DNMT3A* and *DNMT3B* were progressively down-regulated (Supplementary Figure S3).

Pathway enrichment analysis on the up-regulated and down-regulated genes between the three stage progressions (S1→S2, S2→S3 and S3→S4) identified significant pathways specifically in up-regulated genes between S1 and S2, and S2 and S3, and down-regulated genes between S3 and S4. From S1 to S2, antigen presentation

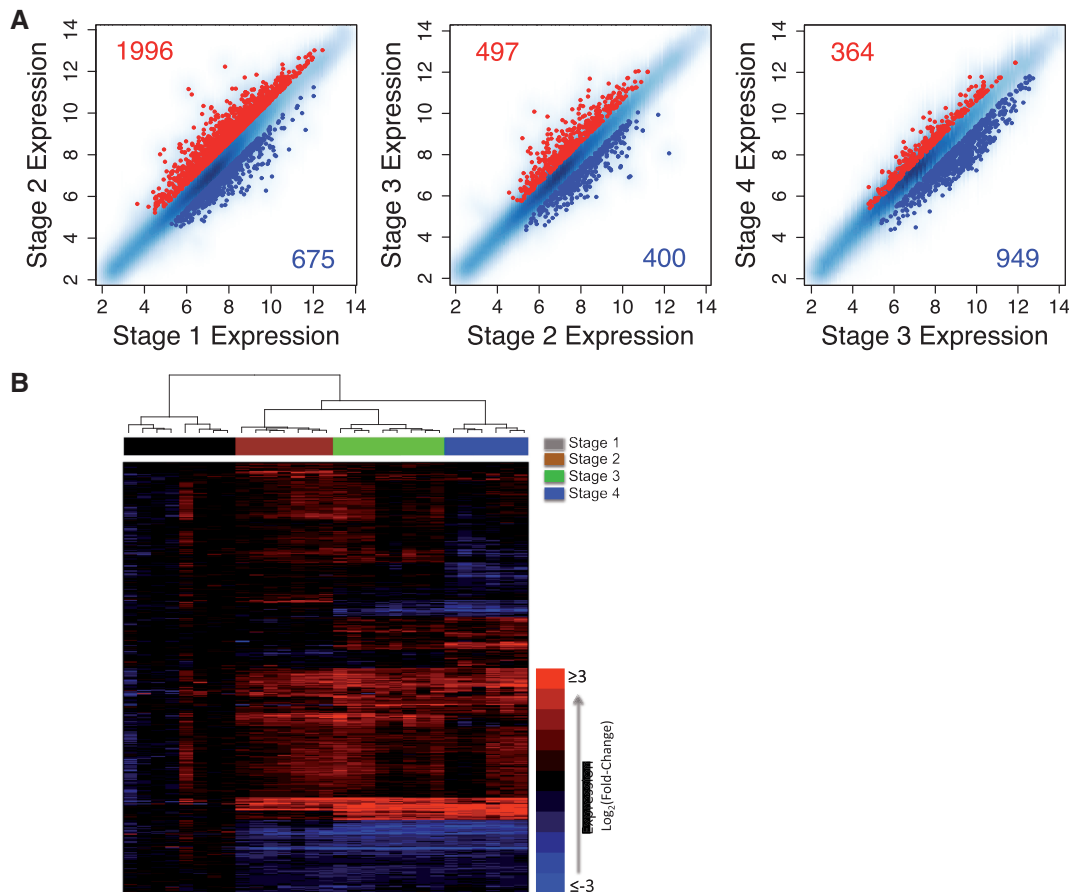


Figure 3. Differential expression analysis during B-cell development. (A) Scatter plots of gene expression levels of two subsequent stages during B-cell development. DEGs (FDR-adjusted $P \leq 0.05$ and fold changes ≥ 1.5) were highlighted with red and blue dots (up- and down-regulated, respectively). (B) A hierarchical clustering of 31 samples based on the expression profiles of the 3913 core DEGs classified each sample into its relevant stages. Heatmap colors indicate log-ratios against expression levels at S1.

and B-cell receptor (BCR) pathways associated with B-cell development were enriched in up-regulated genes, as well as polo-like kinase (PLK) and G2/M checkpoint pathways-associated cell proliferation. Nuclear factor of activated T-cells (NFAT), hepatocyte growth factor (HGF), interleukin-3, protein kinase A, estrogen receptor, glucocorticoid receptor signaling pathways and protein ubiquitination pathways were also modestly enriched during these stages. From S2 to S3, Phosphoinositide 3-kinase (PI3K) and interferon (mediated by JAK1, JAK2 and TYK2) signaling pathways were specifically enriched in up-regulated genes. During S3->S4 progression, many genes linked to cell cycle, cell division and mismatch repair were down-regulated. The associated signaling pathways include BRCA1 (most significant, $P = 1.6 \times 10^{-23}$), ataxia telangiectasia mutated (ATM), checkpoint kinase (CHK), PLK, cyclin and purine/pyrimidine pathways (Supplementary Table S2).

Integrative analysis of methylation and expression changes during pre-B-cell development

To investigate whether methylation alterations affect transcription, we examined the expression changes of genes with altered methylation levels during stage progression. In most cases, methylation changes did not lead to discernable expression differences in associated genes. 12.4% of the genes with one or more DMRs from S1 to S2 resulted in corresponding transcriptional alterations, which represented a highly significant overlap (Fisher's exact $P = 1.9 \times 10^{-6}$). Similarly, from S2 to S3, 10% of genes with ≥ 1 DMRs had concurrent methylation and expression changes (Fisher's exact $P < 2.2 \times 10^{-16}$), whereas the corresponding number from S3 to S4 was only 2.6% ($P = 0.77$). Looking among the genes that incurred stage-wise expression changes, the percentages that were accompanied with methylation changes were 10.1, 11.9 and 0.25% for stage progression from S1 to S2, from S2 to S3 and from S3 to S4, respectively. This suggests that methylation plays a pivotal role in B-cell lineage commitment, but mechanisms other than methylation also play an important role, especially toward progression into immature B-cells. We listed the 224, 60 and 7 genes in that had undergone concurrent methylation and expression changes from S1 to S2, from S2 to S3 and from S3 to S4, respectively, in Supplementary Table S3.

Given the enrichment of DMRs detected at intragenic regions outside of CGIs, we classified all stage-wise DMRs into different regulatory subgroups according to their relative locations in the gene (Promoter, Body) and to CGIs (non-Island, Island, Shelf/Shore) and examined the relationship between differential methylation and differential expression within each subgroup. If there were multiple DMRs mapped to the same gene and same regulatory subgroup, we selected the DMRs with the largest absolute methylation changes to represent the unit. In a robust linear regression analysis using the expression changes as the outcome and the methylation changes as the predictor, adjusted for baseline expression and methylation levels at the preceding stage, DMRs in all regions,

except for body/CGI, showed an inverse correlation with expression changes. Interestingly, the strongest association was found in the group of promoter/non-CGI (coefficient = -0.61 , $P = 0.001$, $n = 350$); a $\Delta\beta$ of 0.75 in this region was estimated to correspond to about a 1.37-fold change of expression levels. Even though the majority of the methylation changes (810 genes) were detected in body/non-CGI regions, their effects on expression were much weaker (coefficient = -0.21 , $P = 0.002$). DMRs in body/CGI were small in number but intriguingly showed the only positive correlation with expression (coefficient = 0.11 , $P = 0.32$, $n = 87$; Figure 4A). In a subset of CpGs and genes, we observed delayed or lagged expression changes in later stages after methylation changes in prior stages. A total of 48 de-DMRs from S1 to S2 were associated with up-regulated DEGs from S2 to S3, and 56 de-DMRs from S1 to S2 were so with those from S3 to S4 (Supplementary Table S4). Characteristic CpGs include those near the *CD19* gene, which are substantially demethylated from S1 to S2, with accompanying substantial expression changes in later stages (Supplementary Figure S4).

Among the 350 genes that had one or more DMRs detected in the promoter/non-CGI group, 25% (89) genes had gene expression changes of at least 1.5-fold. Focusing on these 89 genes, we observed that 65 of them (74%) had methylation and expression changes occurring in inverse correlation (Figure 4B). Methylation and gene expression changes of this 65-gene set are listed in Supplementary Table S5. Examining the CpG dinucleotide density of these 65 genes, we found that a surprisingly high percentage (75.8%) do not contain any CGI regions in their entire genic region (including promoter region), furthermore, 86.4% had no CGI at their promoter region (Supplementary Figure S5A). In contrast, only 27.1% of the transcripts on HM450 do not contain any overlapping CGIs in their genic region, and only 34.1% have non-CGI promoters (Supplementary Figure S5B). While the role of CGIs located in the promoter regions in transcriptional silencing is well-recognized, little is known about the regulation of the genes whose promoters contain no CGIs, our results show that demethylation at low-density CpG sites in the promoter regions has a functional significance in the establishment of pre-B-cell lineage.

Specific histone modifications have been previously shown to be associated with active CGI promoters (22). To characterize the chromatin structure associated with these 65 genes, we examined the histone modification ChIP-seq data of the lymphoblastoid cell line GM12878 from the ENCODE project. A large majority, 75% and 66%, of these 65 genes have their non-CGI promoter regions marked with active histone modifications H3K4me1 and H3K4me3, respectively. Interestingly, only 37% and 28% of all the non-CGI promoter regions on HM450 are marked with these two active histone modifications. The percentages are notably higher, at 50% and 85%, for CGI promoter regions (Supplementary Figure S6). Our results indicate that the non-CGI promoters associated with expression changes display a histone modification profile more similar to

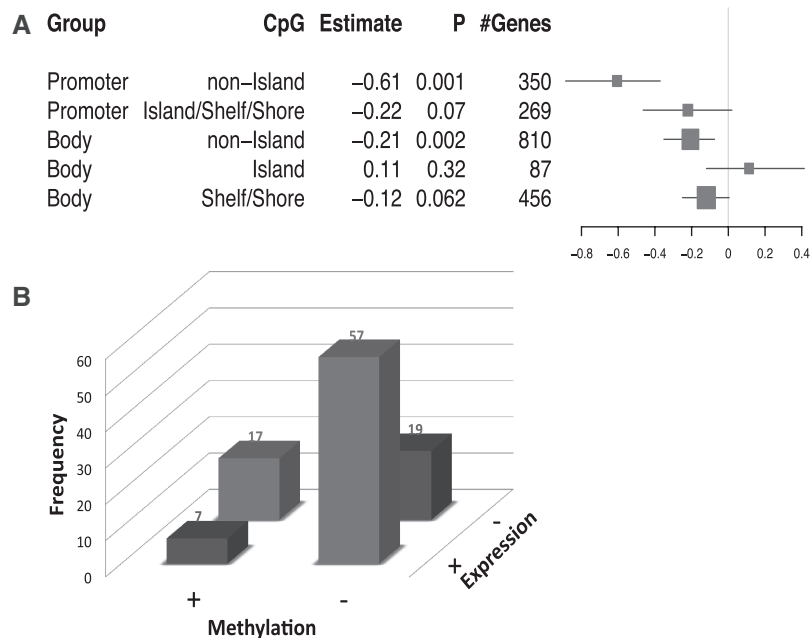


Figure 4. Association of RNA differential expression with DNA differential methylation during B-cell development. **(A)** Forest plot of the coefficient of differential methylation in a robust linear regression analysis using the gene-wise methylation changes as the outcome and the gene-wise expression changes as the predictor. Methylation changes were stratified with respect to annotation categories and model fitting was carried out within each category. **(B)** Differential promoter/non-CGI methylation negatively correlates with gene expression. The 89 genes showing both a change in one or more DMRs in its promoter/non-CGI region and a change in gene expression (>1.5-fold) were plotted. (+) and (-) denote an increase and a decrease in expression or methylation, respectively. Numbers above the bars are percentages.

that of CGI promoters than that of the average non-CGI promoters.

Differences of methylation and expression changes in a gene-specific context

Although some trends were noted in methylation and expression changes, the situation differed gene by gene. A critical gene for pre-B-cell formation is *CD19*, whose promoter region was demethylated during gene activation. These regions and DMRs exactly coincided with EBF binding sites (Supplementary Figure S7A). A second example is *TCF3* (E2A), which has two distinct TSSs. The upstream TSS (TSS1; encoding isoform E12) was hypomethylated at S1 and its methylation level remained unchanged during B-cell development whereas regions around the second TSS encoding isoform E47 underwent demethylation (Supplementary Figure S7B). This suggests a role of methylation changes in isoform regulation. In the case of *PAX5*, its TSS overlaps with CGIs and these regions were hypomethylated at S1. During gene activation, the CGI region in the middle of intron 5 became hypermethylated (Supplementary Figure S7C). In a mouse study, the 5'-part of *PAX5* intron 5 was proven to have enhancer elements that were affected by EBF1 binding and were demethylated during B-cell development (23). An example of down-regulated gene is *NR1H3*, a key regulator of macrophage function, whose CpG sites in the promoter region were subjected to *de novo* methylation whereas CpG sites located in the 3'-part of the gene body were demethylated (Supplementary Figure S7D).

DISCUSSION

In the present study, we compared the methylation profiles of early progenitor cells starting from a point prior to lineage commitment (predominantly MPPs and CLPs) through pre-B-I, pre-B-II and immature B-cells, and have comprehensively cataloged methylation changes associated with pre-B-cell development. The dense probe placement enabled us to examine methylation changes within and outside CGIs and in various locations within individual genes to gain insights on loco-regional epigenetic control of human early B-cell development. Our analysis created a reference set for those studies investigating malignancies originating from B-cell progenitors which are the most common cancer in children. In combination with genome-wide RNA expression profiling, we have identified a distinct development-dependent demethylation signature which has gene expression regulatory properties. Altered DNA methylation occurred predominantly within gene bodies outside of CGIs, but less frequently at annotated gene promoters or with CGIs.

The canonical view on epigenetic gene regulation is that DNA methylation in CGIs in gene promoters has a principle role for repressing gene expression, and the loss of methylation in this region is associated with gene activation (7,8). Recently, several groups have provided evidence linking gene body methylation and transcription, but outstanding questions remain (24). Our data suggest a complex mechanism of epigenetic gene regulation during pre-B-cell development via DNA methylation. We observed that DNA methylation changes more frequently

occur at gene body and remote upstream regions than promoter regions during early B-cell development although those at promoters still possess the most potent effect on gene expression. It is notable that at S1, before complete lineage commitment, the promoters of the majority of the genes are hypomethylated ($\beta < 0.3$), especially those having CGIs in their promoters (Supplementary Figure S2A). Our data indicate that, during progression through pre-B-cell stages, these promoter regions maintained their hypomethylation status whereas alterations in methylation occurred specifically to the gene body regions of many genes. Many of the DMRs in the body regions may correspond to cryptic TSS sites or enhancer regions, which need to be suppressed or activated for efficient gene transcription (25). We also observed demethylation of remote upstream regions in some up-regulated genes such as *NH1R3*. In fact, there is evidence that enhancer methylation in some genes is functionally critical for gene regulation (26,27). This may be understood in line with a previous model that primary TFs prime their target gene promoters at earlier stage in association with cell fate decisions and the gene expression is 'triggered' with a time lag after accumulation of secondary reinforcing events (5,11). On the contrary, some genes had their promoter methylated at baseline and were activated almost simultaneously with promoter demethylation, suggesting an instant triggering role of these genes. We provided one example of this in the pertinent role of demethylation at low-density CpG (non-Island) sites in the promoter regions having the strongest functional impact in triggering expression changes (Figure 3A) in the establishment of pre-B-cell lineage. We also found that these non-CGI promoters display a histone modification profile that is more similar to CGI promoters than to the average non-CGI promoters.

Strikingly, only demethylation processes were noted from stages S2 to S4, although many genes were down-regulated during this period. Furthermore, there were only seven genes that were altered in expression and in methylation from stages S3 to S4 (Supplementary Table S3); these genes included a negative regulator of BCR signaling, *BTLA* (which was increased in expression in S4) and the adenosine deaminase *ADARB1*, a gene distinct from the hypermutation and switch recombinase gene *AICDA* or activation-induced cytidine deaminase (Supplementary Table S3C). DNA methyltransferases including *DNMT1*, *DNMT3A* and *DNMT3B* were down-regulated as the stages progressed, supporting the diminished role of methylation in later stages. Therefore, it may be postulated that mechanisms other than methylation have the primary role in down-regulating genes in later stages, which may include changes in TF activity from protein modification, alterations in TF interactions and combinatory co-binding properties, changes in histone modification, effects of cytokines and miRNAs, and altered RNA decay speeds from RNA interference (28–32).

We noted that the *de novo* and demethylation processes occurred in regions enriched for specific TF binding sites, and these were identified *in silico* data using motif finding

algorithms (Supplementary Table S1) as well as in *in vivo* data using ChIP-Seq data from the ENCODE project (Table 1). From S1 to S2, EBF binding motifs were most significantly demethylated corresponding to its central role in B-cell development (3,4). This was followed by Ets, *RUNX1*, *TCF3* and *ELF5* motifs (*in silico*) and E2F family members and *PAX5* (*in vivo*). The Ets family TFs consists of over 30 members sharing common DNA binding (ETS) domain and have significant but heterogeneous roles on hematopoiesis (33). Accordingly, *de novo*-methylated regions were also enriched for Ets family TFs. *SPI1* (PU.1), an Ets family TF involved in the myeloid versus B-lymphoid lineage decision, were also enriched both in hyper- and hypomethylated regions *in vivo* suggesting its complex roles (3,5). The *TCF3* (E2A) is putatively known to activate the B-cell lineage-specific gene program synergistically with *EBF1* and its motifs were significantly enriched here (Table 1) (3,4,34). Gene families associated with hypermethylation include targets of the myeloid TFs *CEBPB* and *P300* (Table 1), a marker of lineage commitment. The *SPIB* motif was also enriched in hypermethylated regions; the molecule is specific to plasmacytoid dendritic cells and HSCs (35) suggesting that its repression is associated with binding site hypermethylation.

Using expression analysis, we found that many pathway-related genes were up-regulated from S1 to S2 (Supplementary Table S2). As expected, antigen presenting, BCR and B-cell developmental pathways were most significantly enriched. The *PI3K* pathway was suggested to be crucial for BCR/pre-BCR signaling (36), and protein ubiquitination was recently highlighted in *NFkB* signaling (37). The *HGF* β -chain was found to form a pre-pro-B-cell growth-stimulating factor with interleukin-7 (38), but may need further exploration. The role of *NFAT* family, *HGF*, *IL-3*, protein kinase A and glucocorticoid and estrogen receptor pathways were somewhat redundant although some were previously implicated in B-cell development (39–44). From S3 to S4, a number of genes related to cell division and mismatch repair were down-regulated with *BRCA1*-related pathway being most significant. The *ATM*, *CHK*, *PLK* and cyclin pathways are also related to cell cycle control and DNA repair. This may be due to the reduced need for cell division but may also be related to preparations for 'hypermutation' in mature and activated B-cells. Interestingly, a study using *BrcA*^{-/-} cell line found increased somatic hypermutation in the immunoglobulin genes (45).

Recently a global DNA demethylation signature was discovered in mouse erythropoiesis (46). This demethylation is associated with DNA replication and occurs at all DNA elements. This process is in contrast with the demethylation signature that Calvanese *et al.* (11) observe in non-erythropoietic cells, which are specific and targeted demethylation events with functional relevance, e.g. Figure 2F. In contrast to erythrocytes, a functional and specific DNA methylation pattern is critical for differentiated cell function in B-cells which remain nucleated and many of which are long-lived.

In summary, we have investigated DNA methylation changes in early human B-cell development in association with expression changes. DNA methylation changes were associated with profound effects on gene expression during early lineage commitment, especially DNA methylation changes in regions other than promoters. The changes were non-randomly located in terms of CGIs, alternative TSSs and TF binding sites. The impact of DNA methylation on gene regulation was reduced in later stages of B-cell development, suggesting that mechanisms other than DNA methylation may have a principal role after lineage commitment.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–5 and Supplementary Figures 1–7.

ACKNOWLEDGEMENTS

We thank the staff and faculty at San Francisco General Hospital Women's Options Center for assistance in the collection of human fetal tissues. We thank Margaret Wensch for helpful discussions and comments. The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding bodies. S.-T.L. and Y.X. performed all bio-statistical analyses, helped direct the analysis and drafted the manuscript. M.O.M. and M.E.F. conducted the cell isolation and sorting, and helped to draft the manuscript. J.X. performed nucleic acid isolations and DNA bisulfite modifications, and helped to draft the manuscript. J.K.W., S.Z., X.D., A.d.S., A.C., P.B. and X.M. helped to analyse data, and helped to draft the manuscript. J.L.W. provided overall direction and management, designed the study and drafted the manuscript.

FUNDING

National Institutes of Health [NIEHS/EPA P01ES018172 to J.L.W. and P.B., and NIDDK P01DK088760 to M.O.M.], [NCI R01CA155461 to J.L.W. and X.M.]; Tobacco-Related Disease Research Program [18CA-0127 to J.L.W.]; Leukemia and Lymphoma Society [6026-10 to J.L.W.]; and J.L.W. and P.B. funded the study. Funding for open access charge: NIH.

Conflict of interest statement. None declared.

REFERENCES

- Hardy,R.R., Kincade,P.W. and Dorshkind,K. (2007) The protean nature of cells in the B lymphocyte lineage. *Immunity*, **26**, 703–714.
- Welner,R.S., Pelayo,R. and Kincade,P.W. (2008) Evolving views on the genealogy of B cells. *Nat. Rev. Immunol.*, **8**, 95–106.
- Ramirez,J., Lukin,K. and Hagman,J. (2010) From hematopoietic progenitors to B cells: mechanisms of lineage restriction and commitment. *Curr. Opin. Immunol.*, **22**, 177–184.
- Santos,P.M. and Borghesi,L. (2011) Molecular resolution of the B cell landscape. *Curr. Opin. Immunol.*, **23**, 163–170.
- Busslinger,M. (2004) Transcriptional control of early B cell development. *Annu. Rev. Immunol.*, **22**, 55–79.
- Mercer,E.M., Lin,Y.C. and Murre,C. (2011) Factors and networks that underpin early hematopoiesis. *Semin. Immunol.*, **23**, 317–325.
- Ooi,S.K., O'Donnell,A.H. and Bestor,T.H. (2009) Mammalian cytosine methylation at a glance. *J. Cell Sci.*, **122**, 2787–2791.
- Poetsch,A.R. and Plass,C. (2011) Transcriptional regulation by DNA methylation. *Cancer Treat. Rev.*, **37(Suppl. 1)**, S8–S12.
- Hattori,N., Imao,Y., Nishino,K., Ohgane,J., Yagi,S., Tanaka,S. and Shiota,K. (2007) Epigenetic regulation of Nanog gene in embryonic stem and trophoblast stem cells. *Genes Cells*, **12**, 387–396.
- Hattori,N., Nishino,K., Ko,Y.G., Ohgane,J., Tanaka,S. and Shiota,K. (2004) Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells. *J. Biol. Chem.*, **279**, 17063–17069.
- Calvanese,V., Fernandez,A.F., Urdinguio,R.G., Suarez-Alvarez,B., Mangas,C., Perez-Garcia,V., Bueno,C., Montes,R., Ramos-Mejia,V., Martinez-Cambor,P. *et al.* (2012) A promoter DNA demethylation landscape of human hematopoietic differentiation. *Nucleic Acids Res.*, **40**, 116–131.
- Golfier,F., Barcena,A., Harrison,M.R. and Muench,M.O. (2000) Fetal bone marrow as a source of stem cells for in utero or postnatal transplantation. *Br. J. Haematol.*, **109**, 173–181.
- Irizarry,R.A., Bolstad,B.M., Collin,F., Cope,L.M., Hobbs,B. and Speed,T.P. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.*, **31**, e15.
- Wiencke,J.K., Zheng,S., Morrison,Z. and Yeh,R.F. (2008) Differentially expressed genes are marked by histone 3 lysine 9 trimethylation in human cancer cells. *Oncogene*, **27**, 2412–2421.
- Smyth,G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article 3, pp.1–26.
- Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B (Methodological)*, **57**, 289–300.
- McLeay,R.C. and Bailey,T.L. (2010) Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinform.*, **11**, 165.
- Bailey,T. and Elkan,C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–39.
- Maunakea,A.K., Nagarajan,R.P., Bilenky,M., Ballinger,T.J., D'Souza,C., Fouse,S.D., Johnson,B.E., Hong,C., Nielsen,C., Zhao,Y. *et al.* (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, **466**, 253–257.
- Birney,E., Stamatoyannopoulos,J.A., Dutta,A., Guigo,R., Gingeras,T.R., Margulies,E.H., Weng,Z., Snyder,M., Dermitzakis,E.T., Thurman,R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
- Jones,P.A. (1999) The DNA methylation paradox. *Trends Genet.*, **15**, 34–37.
- Rakyan,V.K., Down,T.A., Thorne,N.P., Flicek,P., Kulesha,E., Graf,S., Tomazou,E.M., Backdahl,L., Johnson,N., Herberth,M. *et al.* (2008) An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res.*, **18**, 1518–1529.
- Decker,T., Pasca di Magliano,M., McManus,S., Sun,Q., Bonifer,C., Tagoh,H. and Busslinger,M. (2009) Stepwise activation of enhancer and promoter regions of the B cell commitment gene Pax5 in early lymphopoiesis. *Immunity*, **30**, 508–520.
- Ball,M.P., Li,J.B., Gao,Y., Lee,J.H., LeProust,E.M., Park,I.H., Xie,B., Daley,G.Q. and Church,G.M. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.*, **27**, 361–368.
- Deaton,A.M. and Bird,A. (2011) CpG islands and the regulation of transcription. *Genes Dev.*, **25**, 1010–1022.
- Wiench,M., John,S., Baek,S., Johnson,T.A., Sung,M.H., Escobar,T., Simmons,C.A., Pearce,K.H., Biddie,S.C., Sabo,P.J. *et al.* (2011) DNA methylation status predicts cell type-specific enhancer activity. *EMBO J.*, **30**, 3028–3039.

27. Shore, A., Karamitri, A., Kemp, P., Speakman, J.R. and Lomax, M.A. (2010) Role of Ucp1 enhancer methylation and chromatin remodelling in the control of Ucp1 expression in murine adipose tissue. *Diabetologia*, **53**, 1164–1173.
28. Golks, A., Tran, T.T., Goetschy, J.F. and Guerini, D. (2007) Requirement for O-linked N-acetylglucosaminyltransferase in lymphocytes activation. *EMBO J.*, **26**, 4368–4379.
29. Zamore, P.D., Tuschl, T., Sharp, P.A. and Bartel, D.P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, **101**, 25–33.
30. Gill, G. (2005) Something about SUMO inhibits transcription. *Curr. Opin. Genet. Dev.*, **15**, 536–541.
31. Zhu, J. and Emerson, S.G. (2002) Hematopoietic cytokines, transcription factors and lineage commitment. *Oncogene*, **21**, 3295–3313.
32. Navarro, F. and Lieberman, J. (2010) Small RNAs guide hematopoietic cell differentiation and function. *J. Immunol.*, **184**, 5939–5947.
33. Bartholdy, B. and Matthias, P. (2004) Transcriptional control of B cell development and function. *Gene*, **327**, 1–23.
34. Maier, H., Ostraat, R., Gao, H., Fields, S., Shinton, S.A., Medina, K.L., Ikawa, T., Murre, C., Singh, H., Hardy, R.R. *et al.* (2004) Early B cell factor cooperates with Runx1 and mediates epigenetic changes associated with mb-1 transcription. *Nat. Immunol.*, **5**, 1069–1077.
35. Schotte, R., Rissoan, M.C., Bendriss-Vermare, N., Bridon, J.M., Duhon, T., Weijer, K., Briere, F. and Spits, H. (2003) The transcription factor Spi-B is expressed in plasmacytoid DC precursors and inhibits T-, B-, and NK-cell development. *Blood*, **101**, 1015–1023.
36. Werner, M., Hobeika, E. and Jumaa, H. (2010) Role of PI3K in the generation and survival of B cells. *Immunol. Rev.*, **237**, 55–71.
37. Liu, S. and Chen, Z.J. (2011) Expanding role of ubiquitination in NF-kappaB signaling. *Cell Res.*, **21**, 6–21.
38. Lai, L. and Goldschneider, I. (2001) Cutting edge: identification of a hybrid cytokine consisting of IL-7 and the beta-chain of the hepatocyte growth factor/scatter factor. *J. Immunol.*, **167**, 3550–3554.
39. Berland, R. and Wortis, H.H. (2003) Normal B-1a cell development requires B cell-intrinsic NFATc1 activity. *Proc. Natl Acad. Sci. USA*, **100**, 13459–13464.
40. Tadmori, W., Feingersh, D., Clark, S.C. and Choi, Y.S. (1989) Human recombinant IL-3 stimulates B cell differentiation. *J. Immunol.*, **142**, 1950–1955.
41. Rennick, D., Jackson, J., Moulds, C., Lee, F. and Yang, G. (1989) IL-3 and stromal cell-derived factor synergistically stimulate the growth of pre-B cell lines cloned from long-term lymphoid bone marrow cultures. *J. Immunol.*, **142**, 161–166.
42. Miyamoto, K., Tsuji, K., Maekawa, T., Asano, S. and Nakahata, T. (2001) Inhibitory effect of interleukin 3 on early development of human B-lymphopoiesis. *Br. J. Haematol.*, **114**, 690–697.
43. Hill, L., Jeganathan, V., Chinnasamy, P., Grimaldi, C. and Diamond, B. (2011) Differential roles of estrogen receptors alpha and beta in control of B-cell maturation and selection. *Mol. Med.*, **17**, 211–220.
44. Grimaldi, C.M., Cleary, J., Dagtas, A.S., Moussai, D. and Diamond, B. (2002) Estrogen alters thresholds for B cell apoptosis and activation. *J. Clin. Invest.*, **109**, 1625–1633.
45. Longerich, S., Orelli, B.J., Martin, R.W., Bishop, D.K. and Storb, U. (2008) Brca1 in immunoglobulin gene conversion and somatic hypermutation. *DNA Repair (Amst)*, **7**, 253–266.
46. Shearstone, J.R., Pop, R., Bock, C., Boyle, P., Meissner, A. and Socolovsky, M. (2011) Global DNA demethylation during mouse erythropoiesis in vivo. *Science*, **334**, 799–802.