2590–2602 *Nucleic Acids Research, 2015, Vol. 43, No. 5* *Published online 17 February 2015*
doi: 10.1093/nar/gkv103

Epigenetic remodeling in B-cell acute lymphoblastic leukemia occurs in two tracks and employs embryonic stem cell-like signatures

Seung-Tae Lee[1,2], Marcus O. Muench[3,4], Marina E. Fomin[3], Jianqiao Xiao[1], Mi Zhou[1], Adam de Smith[1], José I. Martín-Subero[5], Simon Heath[6], E. Andres Houseman[7], Ritu Roy[8], Margaret Wrensch[9], John Wiencke[9], Catherine Metayer[10] and Joseph L. Wiemels[1,9,*]

[1]Department of Epidemiology and Biostatistics, University of California at San Francisco, San Francisco, CA 94158, USA, [2]Department of Laboratory Medicine, Yonsei University College of Medicine, Seoul 120752, Republic of Korea, [3]Blood Systems Research Institute, University of California at San Francisco, San Francisco, CA 94158, USA, [4]Liver Center and Department of Laboratory Medicine, University of California at San Francisco, San Francisco, CA 94158, USA, [5]Unidad de Hematopatología, Servicio de Anatomía Patológica, Hospital Clínic, Universitat de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona 08036, Spain, [6]Centro Nacional de Análisis Genómico, Parc Científic de Barcelona, Barcelona 08036, Spain, [7]Department of Public Health, Oregon State University, Corvallis, OR, 97331 USA, [8]Cancer Research Institute, University of California at San Francisco, San Francisco, CA 94158, USA, [9]Department of Neurological Surgery, University of California at San Francisco, San Francisco, CA 94158, USA and [10]Division of Epidemiology, School of Public Health, University of California at Berkeley, Berkeley, CA 94720, USA

Received November 04, 2014; Revised January 28, 2015; Accepted January 31, 2015

ABSTRACT

We investigated DNA methylomes of pediatric B-cell acute lymphoblastic leukemias (B-ALLs) using whole-genome bisulfite sequencing and high-definition microarrays, along with RNA expression profiles. Epigenetic alteration of B-ALLs occurred in two tracks: *de novo* methylation of small functional compartments and demethylation of large inter-compartmental backbones. The deviations were exaggerated in lamina-associated domains, with differences corresponding to methylation clusters and/or cytogenetic groups. Our data also suggested a pivotal role of polycomb and CTBP2 in *de novo* methylation, which may be traced back to bivalency status of embryonic stem cells. Driven by these potent epigenetic modulations, suppression of polycomb target genes was observed along with disruption of developmental fate and cell cycle and mismatch repair pathways and altered activities of key upstream regulators.

INTRODUCTION

Our understanding of the cancer epigenome has expanded from the canonical view of promoter DNA methylation and gene silencing to incorporate an intricate landscape (1). *De novo* methylation in promoter CpG islands (CGIs) and global demethylation in intergenic repeat sequences and gene bodies are now recognized as hallmarks of cancer (2). Connections between DNA methylation and histone, polycomb and transcription factor (TF) proteins are increasingly apparent (1,3). The characterization of an abnormal departure from normal developmental states is typically challenged by the difficulty in obtaining appropriate control sets of early relevant stages. Here, we investigated genome-wide DNA methylomes of B-cell acute lymphoblastic leukemias (B-ALLs), the most common cancer in children, in reference to their normal B-cell precursors (pre-B).

MATERIALS AND METHODS

B-ALL samples and nucleic acids extraction

A total of 231 pediatric patients with B-ALL were utilized from the California Childhood Cancer Study. After exclusion by quality check described below, 227 patients were finally analyzed, including 206 patients with unequivocal cytogenetic classification. After informed consent, bone marrow specimens packed with leukemic blasts were obtained for each patient. Mononuclear cells were purified using Ficoll. Genomic DNA and total RNA were isolated using the AllPrep DNA/RNA Mini Kit (Qiagen).

*To whom correspondence should be addressed. Tel: +1 415 514 0577; Fax: +1 415 502 7411; Email: joe.wiemels@ucsf.edu

Control pre-B cells and other blood cells

Expression and methylation data of purified pre-B cells in four developmental stages were retrieved from our previous study (4), and by the established protocol, pre-B-II cells for whole genome bisulfite sequencing (WGBS) were purified. Purified normal mature B-cells, T-cells, regulatory T-cells, NK-cells, monocytes and granulocytes were purchased from the Allcells Corp. DNAs were extracted as described above.

WGBS and library construction

WGBS was conducted as described previously (5). Briefly, unmethylated λ DNA (Promega) was spiked into genomic DNA. We sheared the DNA by sonication to 50–500 bp with a Covaris E220, followed by selecting 150- to 300-bp fragments using AMPure XP beads (Agencourt Bioscience). Genomic DNA libraries were constructed using the TruSeq Sample Preparation kit (Illumina). After adaptor ligation, two rounds of sodium bisulfite-conversion of DNA using the EpiTexy Bisulfite Kit (Qiagen) were performed to get a conversion efficiency of >99%. We enriched adaptor-ligated DNA through seven cycles of PCR using the PfuTurboCx Hotstart DNA polymerase (Stratagene). Library quality was monitored using the Agilent 2100 Bioanalyzer (Agilent) and KAPA Library Quantification Kit (Kapa Biosystem). Paired-end sequencing (2 × 100 bp) was then carried out using the Illumina Hi-Seq 2000.

Read mapping and methylation% estimation

From the human genome (GRCh37) and NCBI viral genome (v35), two reference sequences including reference C2T (all Cs were replaced by T) and reference G2A (all Gs were replaced by A) were generated, and to these, each paired-end read was aligned using the GEM software (6). Up to four mismatches were allowed, and read pairs with unique compatible mappings (to the same chromosome and in the expected orientation) were selected. After mapping, most likely genotype and methylation% in each C were estimated, using the software developed at the Centro Nacional de Análisis Genómico, which takes into account the observed bases, base quality scores and the strand origin of each read pair. The confidence in the genotype call was estimated by the Phred-scaled likelihood. CpG sites were selected where both bases were called as homozygous CC followed by GG with a Phred score of ≥20. Sites with >500× coverage depth were excluded to avoid highly repetitive regions (5).

High-definition methylation array

Genomic DNA was bisulfite-converted using the EZ DNA Methylation Kit (Zymo Research), followed by hybridization onto the Infinium HumanMethylation450 Beadchip (Illumina). Raw data was normalized and further processed using the GenomeStudio software (Illumina) and R software packages including methylumi and others. CpG sites with poor quality (detection $P > 1.0 \times 10^{-4}$) were removed from analysis, and four samples with high proportion (>15%) of poor CpG sites were excluded. Additional 450k array data of relapsed B-ALLs was obtained from a clinical set of patients (GSE49031) (7).

Genomic annotation and definition of regions

Data was annotated from the UCSC Genome Browser database (version hg19), using RefSeq Genes, lincRNA Transcripts, CpG Islands, Vista Enhancers, ENCODE Transcription Factor ChIP-Seq, DNase Clusters, LaminB1 (LAD), and RepeatMasker tracks. ChIP-Seq data of H1 ESC was retrieved from the UCSF-UBC Human Reference Epigenome Mapping Project (GSE16368). For histone mark information in somatic tissues, we retrieved the Broad Histone track from the USCS browser. For genes with multiple isoforms, the longest one was used as the reference. A 'promoter' was designated to be from −1.5 kb upstream to the transcriptional start site of each gene. We defined 'CGI shores' as 0–2 kb from the CGI edge and CGI shelf as 2–4 kb. By scaling from the whole gene body length, we defined '5′-body' as 0–0.1 fractional region and 'main body' as 0.1–1.0. After removing all promoters, 5′-body, exon, exon–intron boundary (±100 bp), CGI, CGI shore, DNase HS, TF-binding and enhancer sites, we designated the remaining region as 'backbone.'

DMR statistics

For both WGBS and array data, 'DMR' was defined as a CpG with >20% difference of methylation. When comparing two groups, array methylation β values were converted by the equation of 'arcsine[square root(β)]' to perform parametric statistical analysis. Moderated T-statistics and false discovery rate (FDR)-correction were performed, and from this, $P < 0.01$ was used as an additional criteria for defining DMRs. The occurrence of DMRs in specified regions was examined by χ^2 or Fisher-exact tests.

TF and motif enrichment

Using WGBS data, we constructed DMR-blocks by selecting CpGs with >25% mean difference between case and control and merging consecutively changed CpGs within 500 bp of distance into a single DMR-block. Then we examined positional enrichment of DMR-blocks around 148 ENCODE TF binding sites using Homer software (8). For array data, the enrichment was assessed by calculating the enrichment rate of single DMRs and by Fisher-exact test. After retrieving sequences around each DMR CpG (±25 bp) along with those around non-DMR CpGs randomly chosen as a background, motif analysis was carried out using Homer software with CpG-density normalization.

RNA expression analysis

Total RNA was amplified and labeled (NuGen Ovation) and hybridized onto the GeneChip Human Gene 1.0 ST Array (Affymetrix), according to standard protocols. Data was normalized using the Expression Console software (Affymetrix) with the Robust Multi-array Average (RMA) algorithm. Further analysis was done using the R software package. DEGs were defined by fold-change >1.5 and

FDR-corrected $P < 0.01$. Gene Set Enrichment Analysis was done using B-cell developmental gene sets retrieved from our previous work (Up-DEGs from S2 to S3) (4) and polycomb target gene sets from a study by Lee *et al.* (9). Pathway analysis was performed using the Ingenuity Pathway Analysis software (Ingenuity System). Additionally, the expression array data of mature peripheral B-cells was obtained from another study (GSE46261).

ROC curve analysis

To identify CpGs or genes that best discriminate one group from another, we performed receiver operating characteristic (ROC) curve analysis for each CpG or gene. After filtering by average differences, CpGs or genes were ranked according to area under the curve (AUC) values of the ROC analysis.

Data access

Raw data for the current study is available in the NCBI Gene Expression Omnibus (GEO) (GSE56602).

RESULTS

Whole DNA methylomes at single base-pair resolution

We first bisulfite sequenced, at single-base pair resolution, two common pediatric B-ALL cancers including *ETV6/RUNX1* rearrangement (ETV6-ALL) from a 3-year-old girl and high hyperdiploid (HD-ALL) from a 1-year-old boy (representing clusters III and I described below, respectively; Supplementary Table S1), in parallel with normal pre-B-II cells obtained from the same source as those used in our previous study (4) and processed as a part of the Blueprint Consortium. Bisulfite-converted DNA was sequenced with average coverage of 46.6 and 49.6 for ETV6-ALL and HD-ALL, respectively, yielding >700 million high-quality reads and >80% unique alignment (Supplementary Table S2). About 20 million CpGs among all 28 million human genomic CpGs were uniquely mapped with high quality and further analyzed (Supplementary Table S3). WGBS data was validated with high correlation ($r > 0.97$) to Illumina 450k array data (Supplementary Figure S1A). Unlike embryonic stem cells (ESCs) (10) and like other somatic cells (11), methylation at non-CpG sites was unnoticeable in our samples.

The majority of CpGs were hypermethylated, and consecutive CpGs with similar methylation levels constituted methylation blocks. Methylated blocks in HD-ALL were larger in number but smaller in size, and as a result, covered less genomic area than the methylated blocks of ETV6-ALL (Supplementary Figure S1B–D). The net methylation level of ETV6-ALL was almost equal to that of the pre-B cell control, while that of HD-ALL was 4.4% lower (Figure 1A; Supplementary Table S3). Compared to pre-B cells, *de novo* methylation mostly occurred in originally hypomethylated CpGs while demethylation did so in originally hypermethylated CpGs, with demethylation being more profound in HD-ALL (Figure 1B and C). We identified 1 315 701 and 2 065 605 differentially methylated regions (DMRs) in ETV6-ALL and HD-ALL, respectively; these were highly concordant with the Illumina 450k array DMRs (Supplementary Table S4). The vast majority of genes in the genome had one or more DMR(s) within their promoter or body regions (Figure 1D). Demethylated DMRs (de-DMRs) outnumbered *de novo* methylated DMRs (*de novo* DMRs) especially in HD-ALL (1 741 261 versus 324 344 in HD-ALL; 826 237 versus 489 464 in ETV6-ALL). There were a number of DMRs that overlap between ETV6-ALL and HD-ALL (89.5% among the two WGBS samples) but also a substantial number of sample-specific DMRs (Supplementary Tables S3 and S4).

We retrieved potentially regulatory regions including CGIs, DNase hypersensitive sites (HS), and transcription factor (TF) binding sites, by combining all regions from various cell types from the ENCODE project, and investigated methylation changes in each type. In both cancers, CpGs in promoter, 5′-body (0–0.1 in fractional region of gene body) and CGI are more enriched for *de novo* DMRs while those in 'CGI shelf' and repeat regions were enriched for de-DMRs. TF binding sites and DNase HS are more enriched for *de novo* methylation in ETV6-ALL and for demethylation in HD-ALL (Figure 1E; Supplementary Table S5). This distribution can be interpreted in regard to methylation levels in normal cells; CGIs and promoters are mostly hypomethylated in pre-B cells, and accordingly, *de novo* DMRs at these sites occur in large part at CpGs with low methylation. Main body and repeat regions are mostly methylated in normal cells so de-DMRs are often at such methylated sites (Supplementary Table S6). On average, CGIs were methylated to similar degrees (∼+10%), regardless of genic locations. However, due to the dense localization of CGIs in promoters and 5′-bodies, *de novo* DMRs were far more enriched in such regions (Supplementary Figure S2). DNase HS in promoters and 5′-bodies were profoundly methylated but such a trend was only modest or negligible in main body and intergenic regions (Figure 2A). Exons in main gene bodies were less altered while introns, especially the deep portion of long introns, were profoundly demethylated in HD-ALL but not in ETV6-ALL (Figure 2B). By excluding all small functional compartments including promoters, 5′-bodies, exons, CGIs, CGI shores, DNase HS, TF-binding sites and enhancer sites, we retrieved the remaining large frame-like regions and defined them as 'backbones'. The regions largely overlap with those often referred to as 'open sea', and cover the vast majority of CpGs, especially those in introns and intergenic regions in the human genome, and are mostly hypermethylated. Backbones of ETV6-ALL were maintained in a hypermethylated state but those of HD-ALL were globally demethylated (∼−5%) (Figure 2C and D). Unlike other cultured cell-lines and normal placental tissues but comparable to most other cells from somatic tissues (12,13), partially methylated domains (PMDs) were absent both in our B-ALLs and normal pre-B cells. However from a whole genome view, we observed a trend that some large regions are slightly demethylated in HD-ALL. The subtle change overlapped with the lamina associated domains (LADs), as defined by (14), as well as PMDs in other cancers which share many regions with LADs (11) (Supplementary Figure S3). Although the definition of LADs were retrieved from normal fibroblasts, the majority of LADs were shown

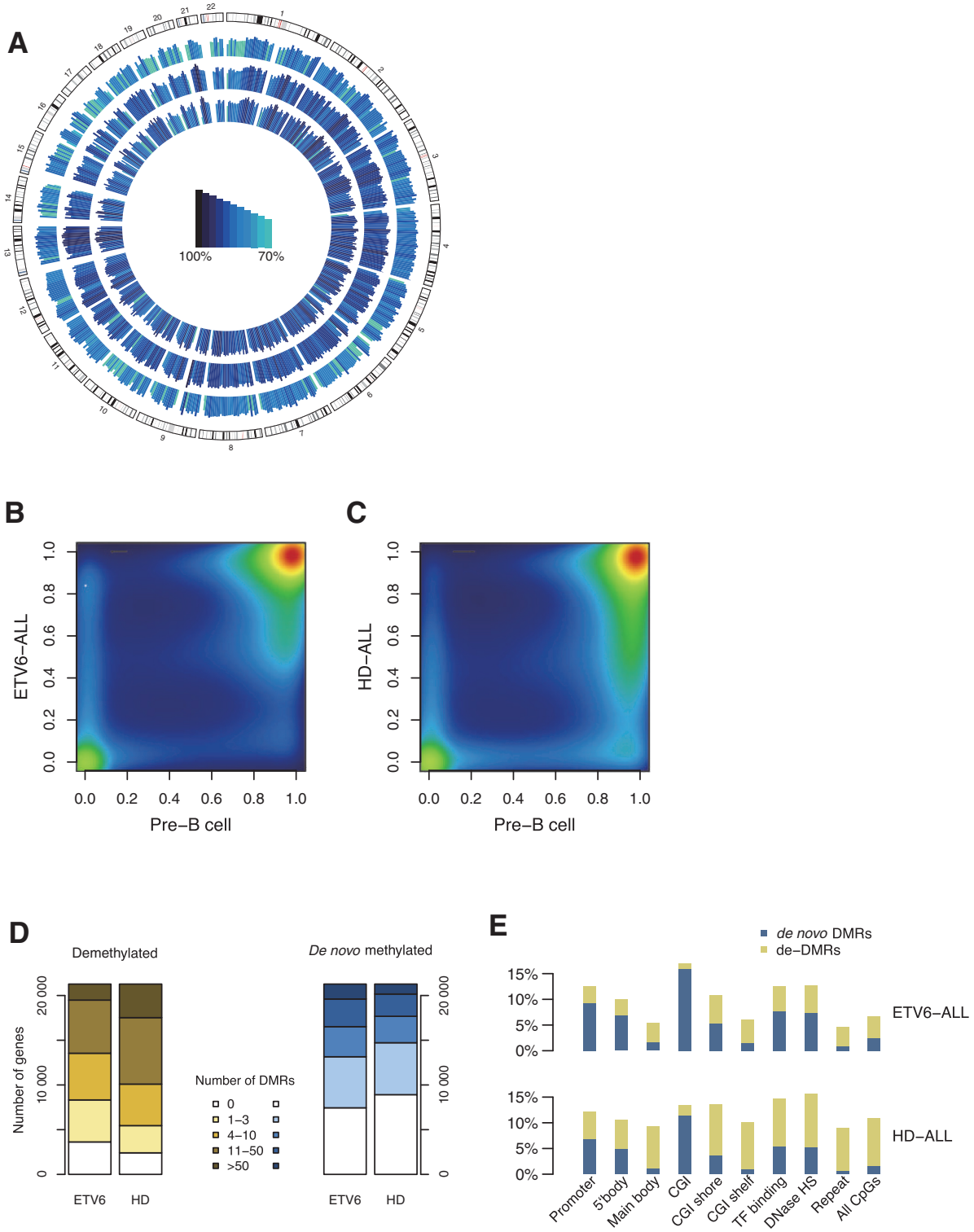

Figure 1. Overview and statistics of whole-genome bisulfite sequencing data of B-ALLs and normal pre-B cells. (**A**) Circular barplots of genome-wide average methylation levels in pre-B cells, ETV6-ALL and HD-ALL (from inside to outside, bin size = 5 Mb) showing slight demethylation of HD-ALL throughout genomic location. (**B** and **C**) Density plots of DNA methylation levels of all autosomal CpGs between B-ALL and pre-B cells. Individual CpG site methylation of pre-B cell and B-ALL are expressed on the X- and Y-axes. Hotter colors indicate higher density of data. (**D**) Counts of differentially methylated regions (DMRs; CpGs with methylation difference >20%) within genes or promoters. Most genes have more than one DMR inside the genic region or promoters, with some genes having numerous DMRs. HD-ALL had more de-DMRs than ETV6-ALL. (**E**) Relative distribution of DMRs according to specified regions. In both tumors, CpGs in promotes, 5′-body (0–0.1 in fractional region of gene body) and CpG island (CGI) are more enriched for *de novo* DMRs while CGI shelf and repeat regions were so for de-DMRs. Transcription factor (TF) binding sites and DNase hypersensitive sites (HS) are more enriched for *de novo* methylation in ETV6-ALL and for demethylation in HD-ALL.

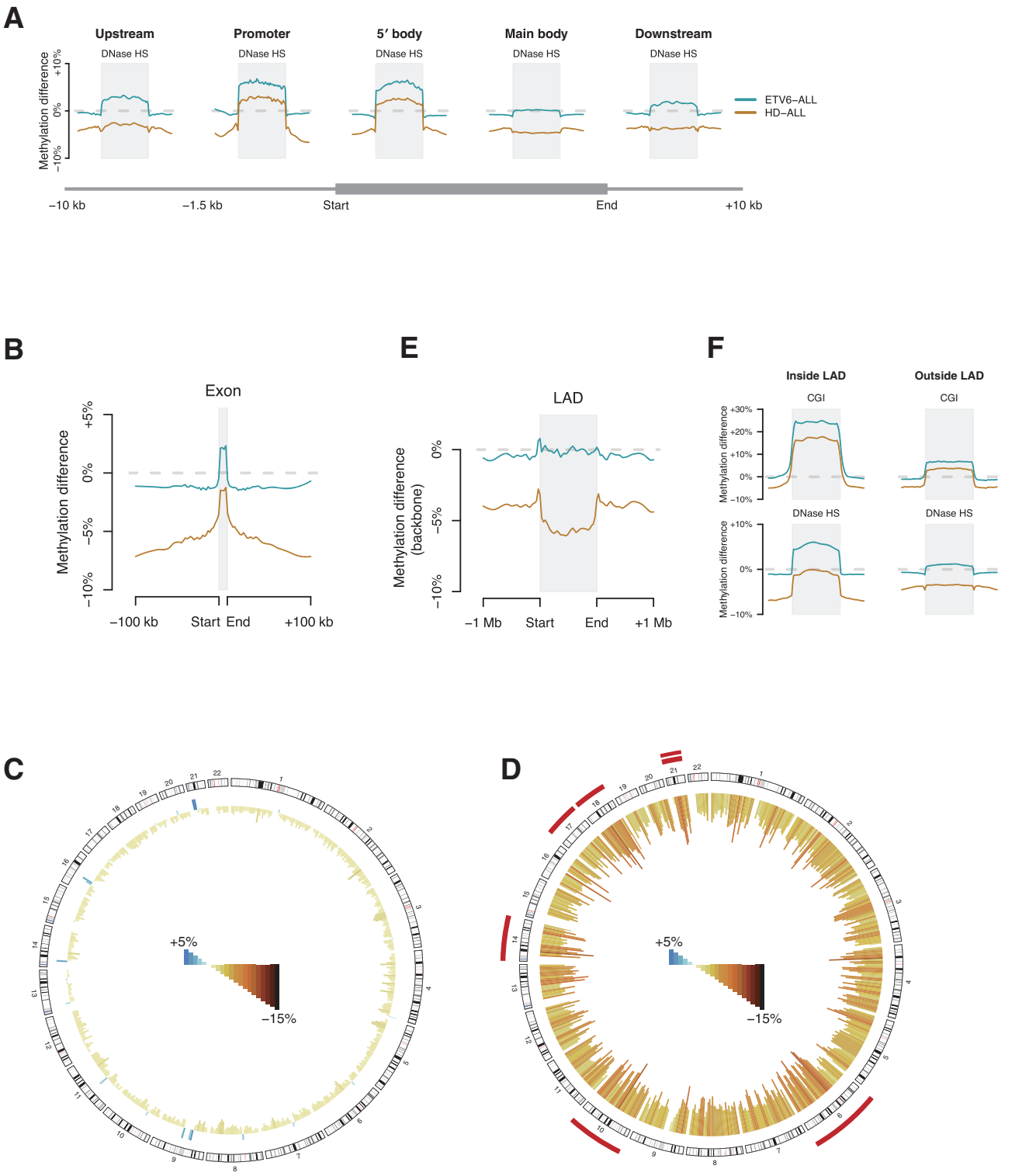

Figure 2. DNA methylation changes of B-ALLs compared to pre-B cell according to specified regions. (**A**) Local regression showing methylation difference in DNase HS, stratified by genic location. DNase HS in promoter and 5′-body are more likely to be methylated, while the trend was slight or negligible in main body and downstream regions. (**B**) Methylation difference in exons and introns in gene main bodies. Exons are less altered while introns, especially the deep portion of long introns, are profoundly demethylated in HD-ALL but not in ETV6-ALL. (**C**) Genome-wide overview of methylation changes in backbones defined by regions devoid of promoters, 5′-bodies, exons, CGIs, CGI shores, DNase HS, TF-binding sites and enhancer sites. ETV6-ALL demonstrates minimal change when compared with the pre-B cell control. (**D**) Genome-wide view of methylation changes in backbones of HD-ALL show profound demethylation across the whole genome. Red bars indicate chromosomes with extra copies. (**E**) Methylation changes of backbones according to location to lamina-associated domain (LAD) show more demethylation of regions inside LADs in HD-ALL. (**F**) Methylation differences in CGI and DNase HS, stratified by location to LAD show profound *de novo* methylation of CGI and DNase HS inside LAD.

to be constitutively conserved in various tissues (15) and so we used this defined region in further analysis as a surrogate for the large repressive epigenetic domains. Backbone demethylation was exaggerated in LADs, especially in HD-ALL, which contrasts to the trend that small functional sites including CGIs and DNAse HS are more *de novo* methylated in LADs (Figure 2E and F). Most repeat families were demethylated but centromere, acrocentric, satellite, tRNA, rRNA, simple repeat and low complexity families were preferentially *de novo* methylated (Supplementary Table S7).

Gene expression correlated negatively with DNA methylation in promoters and 5′-bodies, and even in those regions, the trend was particularly remarkable in CGIs or DNase HS (Figure 3A and B; Supplementary Figure S4A and B). Demethylation of main bodies of low-level expressed genes was noted in HD-ALL but not in ETV6-ALL (Figure 3A; Supplementary Figure S4A). Plotting methylation values of CpGs in *RB1*, an exemplary tumor suppressor gene down-regulated in B-ALLs, illustrated typical patterns: ETV6-ALL preferentially had promoter CGIs methylated while HD-ALL rather had gene body, including deep introns, demethylated (Supplementary Figure S4C). Highly expressed genes had a tendency to maintain hypomethylated promoters and hypermethylated gene bodies, in concordance with the theory that gene body hypermethylation helps to dampen transcriptional noise and maintain high expression of intact mRNA (16,17). However, the situation may vary in different genes and it appears that expression is confounded by many factors (Supplementary Figure S4D–G).

Enrichment of specific TFs related to ESC bivalency

After reconstituting *de novo* and de-DMR blocks, we screened them against the binding sites of 148 TFs from ENCODE. Two TFs, SUZ12 and CTBP2, were remarkably enriched around *de novo* DMRs, and most TFs were poorly enriched around de-DMRs (Figure 4A and B; Supplementary Figure S5A and B). We further investigated DNA methylation changes in association with histone signatures by obtaining ChIP-Seq data of H1 human embryonic stem cell and found that bivalent domains of ESCs, marked by H3K4me3 and H3K27me3 co-occupancy (18), are more likely to be *de novo* methylated in B-ALLs (Figure 4C; Supplementary Figure S5C). We performed similar analysis using histone marks from different types of somatic tissues. Compared to ESCs, bivalent regions in the somatic cells were much smaller but still showed higher enrichment rates for *de novo* DMRs. Rather, those cells had large H3K27me3 domains, which should have replaced the bivalent domains during differentiation (19), and as to absolute numbers, *de novo* DMRs were far enriched in the H3K27me3 single-occupancy domain (Supplementary Figure S6). Positional comparisons between histone signatures and TF-binding sites suggested that CTBP2 is actually linked to the H3K4me3 portion and SUZ12 is to the H3K27me3 portion of the bivalent domain in ESC (Figure 4D). An exemplary gene is illustrated in Figure 4E; the promoter of *ALX4* is occupied with both H3K4me3 and H3K27me3 histone marks in human ESC, and when differentiated, the gene is activated in normal skeletal muscle (HSMM cell-line) by H3K4me3 single-occupancy (i.e. by loss of H3K27me3) while suppressed in normal lung fibroblasts (NHLF cell-line) by H3K27me3 single-occupancy (i.e. by loss of H3K4me3). In both ALLs, *de novo* DMRs are highly enriched in such bivalent regions at the promoter, which also overlap with CGIs as well as SUZ12- and CTBP2-binding sites. It is noteworthy that the H3K4me3 peaks of the bivalent domain frequently coincide with the CTBP2-binding sites (Figure 4E).

Motifs of REST/NRSF, PU.1, HOX and GATA families were significantly enriched in *de novo* DMRs, especially in ETV6-ALL, whereas motifs of onco-proteins including ERG and MYC and B-cell developmental TFs including EBF were more enriched in de-DMRs, especially in HD-ALL. We also discovered novel motifs enriched in DMRs (Figure 4F; Supplementary Table S8).

Genome-wide methylation profiles of B-ALLs

We next analyzed methylation profiles of 227 pediatric B-ALLs using Illumina 450k array (Supplementary Table S9). From our previous work (4), we also included data of purified pre-B cells representing four ordered developmental stages including multipotent progenitor, pre-B-I, pre-B-II and immature B cells. Using 5021 differentiation-associated DMRs, we calculated similarity scores to find that most B-ALLs are concordant to the last three stages but not the first stage (Supplementary Figure S7A). Accordingly, the last three stage cells were used as controls. Among the 472 499 autosomal CpGs that passed quality control, 26 874 (5.7%) around 4671 genes were *de novo* DMRs and 23 909 (5.1%) around 9854 genes were de-DMRs (Supplementary Figure S7B). *De novo* DMRs were overrepresented in the promoter and 5′-body, and de-DMRs were overrepresented in the main body (Supplementary Figure S7C and S8).

An unsupervised hierarchical clustering using the 500 most variable CpGs divided samples into four distinct clusters (Figure 5A). Cluster I mainly contained hyperdiploid cases, cluster II had ‘others’ cases and a small number of *ETV6/RUNX1* cases, cluster III mostly had *ETV6/RUNX1* cases, and cluster IV had ‘others’ and *ETV6/RUNX1* cases. In each cytogenetic group, methylation clusters represented different epigenetic status in CGI, with clusters III and IV having more methylation (Figure 5B). The cytogenetics groups or methylation clusters largely shared DMRs that may be common for all B-ALLs but each group/cluster also had specific DMRs (Figure 5C and Supplementary Figure S9A). The numbers of DMRs in individual leukemias varied across broad ranges (*de novo* DMRs from 3070 to 89 511 and de-DMRs from 9677 to 89 033) and were related to methylation clusters; cluster IV had larger numbers of DMRs while cluster II had smaller numbers (Supplementary Figure S9B). We also identified top CpGs discriminating hyperdiploid and *ETV6/RUNX1* (Supplementary Figure S10).

As a replication of the findings from WGBS, we could also observe *de novo* methylation of promoter, 5′-body and CGIs and demethylation of main body and CGI shelves but the degrees differed in each leukemia and related to methylation clusters (Supplementary Figure S8A). We also

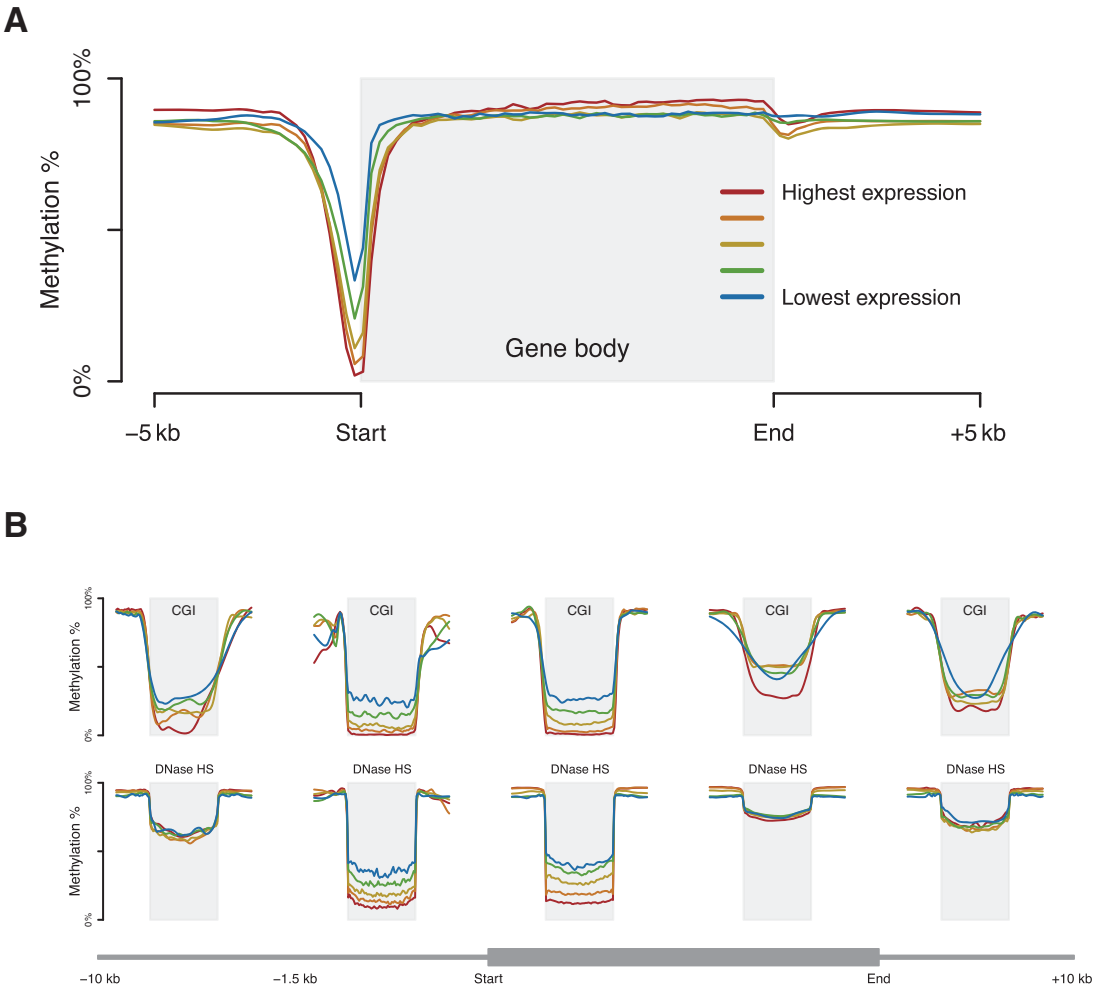

Figure 3. Association between DNA methylation and gene expression. (**A**) Local regression showing methylation levels of HD-ALL according to genic locations, stratified by expression quintile. Genes in the highest expression quintile tend to have their promoter unmethylated and gene body mostly methylated. (**B**) Methylation levels of HD-ALL in CGI and DNase HS, stratified by genic location and expression quintile. Methylation in CGI and DNase HS located around promoter and 5′-body apparently correlated with expression while not in the remaining regions.

observed the trend that CGIs are methylated regardless of genic location and DNase HS are more methylated around promoters and the 5′-body. However, the trends in DNase HS were quite differing in each leukemia and each cluster, with some leukemias having negative average differences (Supplementary Figure S8B). As observed in WGBS cases, *de novo* methylation in CGIs and DNase HS are more remarkable inside LADs (Supplementary Figure S8C).

A total of 8738 CpGs around 182 putative imprinted genes were selected and found to be less altered during normal development but more likely to be *de novo* methylated in B-ALLs, compared to those around non-imprinted genes (10.3% versus 5.6%) (Supplementary Table S10).

We retrieved methylation values of the CGI and backbone regions from the array data. Four methylation clusters behaved differently in terms of CGI methylation and backbone demethylation (Supplementary Figure S11A). As to cytogenetic groupings, hyperdiploid leukemias had backbones more demethylated than *ETV6/RUNX1* (Supplementary Figures S11B and S12A). These trends were far exaggerated in CpGs in LADs (Supplementary Figure S11C and D). Analysis on a clinical cohort (7) illustrated that relapsed clones tend to have intensified CGI methylation and slight backbone demethylation; 20 among the 24 cases in the cohort had CGI altered methylation by more than 0.05, and 11 among the 24 cases had backbone demethylation by more than 0.01 (Supplementary Figure S11E). The average methylation values of backbones in some LADs correlated strongly with each other (Supplementary Figure S12B and C).

We further scanned the ENCODE TF tracks and again found significant enrichment of SUZ12 and CTBP2 (enrichment rates 11.5 and 4.8, respectively) in *de novo* DMRs (Supplementary Figures S13A and S14). Plotting enrichment rates in each case indicated that virtually all B-ALLs are enriched for binding sites of the two TFs, with higher rates in clusters III and IV (Supplementary Figure S13B). The mean methylation at SUZ12-binding sites significantly correlated with mean expression of polycomb target genes ($r = -0.276$; $P = 0.020$), suggesting a substitutive and/or

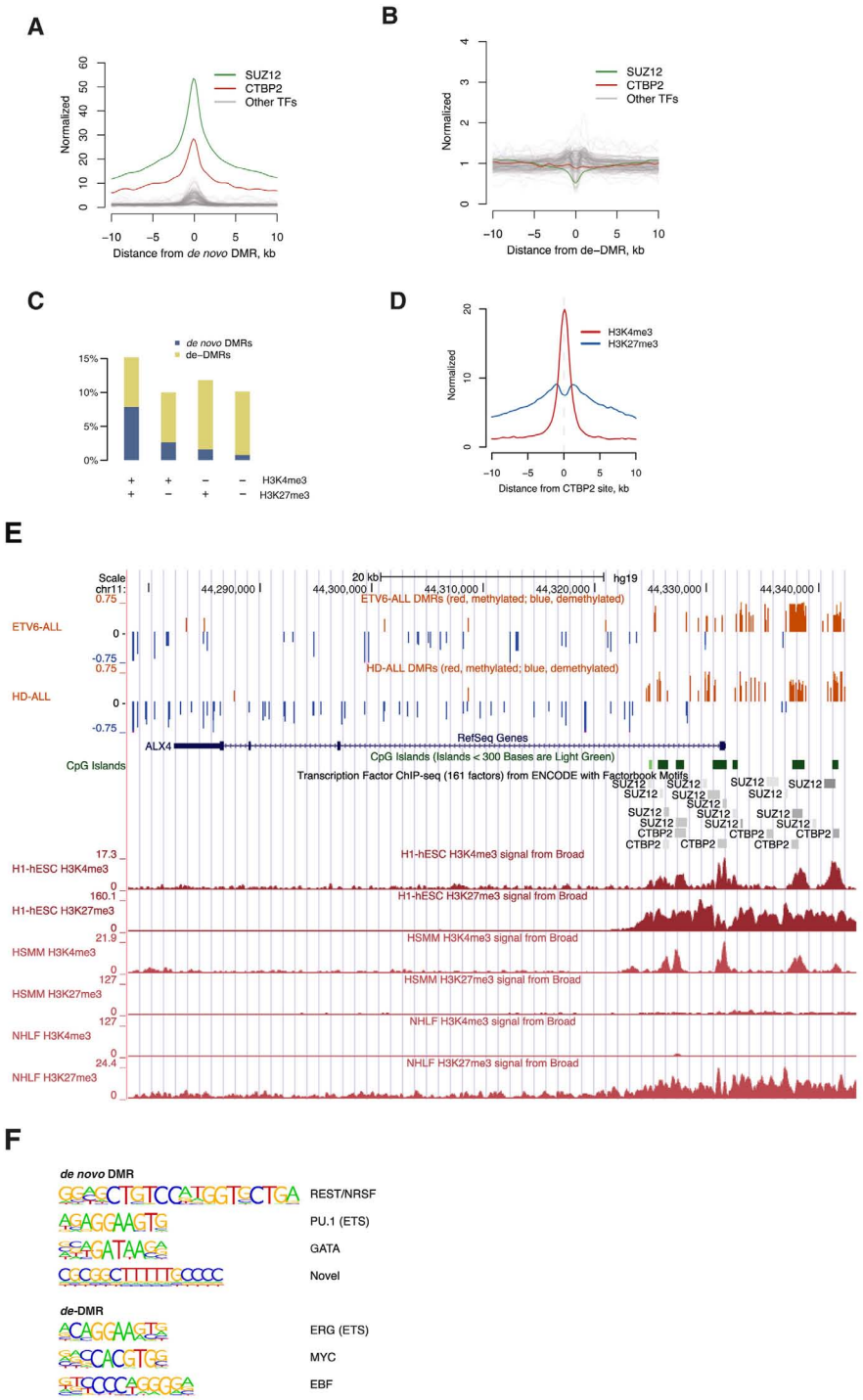

Figure 4. Enrichment of WGBS DMRs according to specific TFs, histone marks or motifs. (**A** and **B**) Positional enrichment of *de novo* DMRs and de-DMRs of HD-ALL against 148 ENCODE TF-binding sites. Two transcription factors, SUZ12 and CTBP2, are highly enriched around *de novo* DMRs while no specific TFs are enriched around de-DMRs. (**C**) Enrichment of DMRs of HD-ALL according to H3K4me3 (active) and H3K27me3 (repressive) histone marks of H1 embryonic stem cell (ESC) shows remarkable enrichment of *de novo* DMRs in bivalent domain characterized by co-occupancy of both histone marks (parallel analysis for ETV6-ALL is presented in Supplementary Figure S5A–C). (**D**) Positional analysis for histone and TF sites in ESC shows enrichment of H3K4me3 at the center of CTBP2-binding site and dual peak of H3K27me3 around the ±1 kb region from the center, suggesting CTBP2 constitute H3K4me3 part of the bivalent domain. (**E**) Methylation changes in an exemplary gene, *ALX4*. In human ESC (H1-hESC), the promoter is poised bivalently with H3K4me3 (active) and H3K27me3 (repressive) histones marks. At differentiation, the gene is activated by H3K4me3 in normal skeletal muscle (HSMM) and suppressed by H3K27me3 in normal lung fibroblasts (NHLF). In ALLs, promoters are mostly methylated (red bars) while gene body is demethylated (blue bars). *De novo* DMRs frequently coincide with the bivalent regions, CGIs and SUZ12- and CTBP2-binding sites. Note that H3K4me3 marks mostly peak around the CTBP2 sites. (**F**) Novel and known motifs significantly enriched around DMRs. Regions around *de novo* DMRs are enriched for REST/NRSF, PU.1 and GATA motifs and others, while those around de-DMRs are so for proto-oncogenes, ERG and MYC and a developmental gene, EBF.

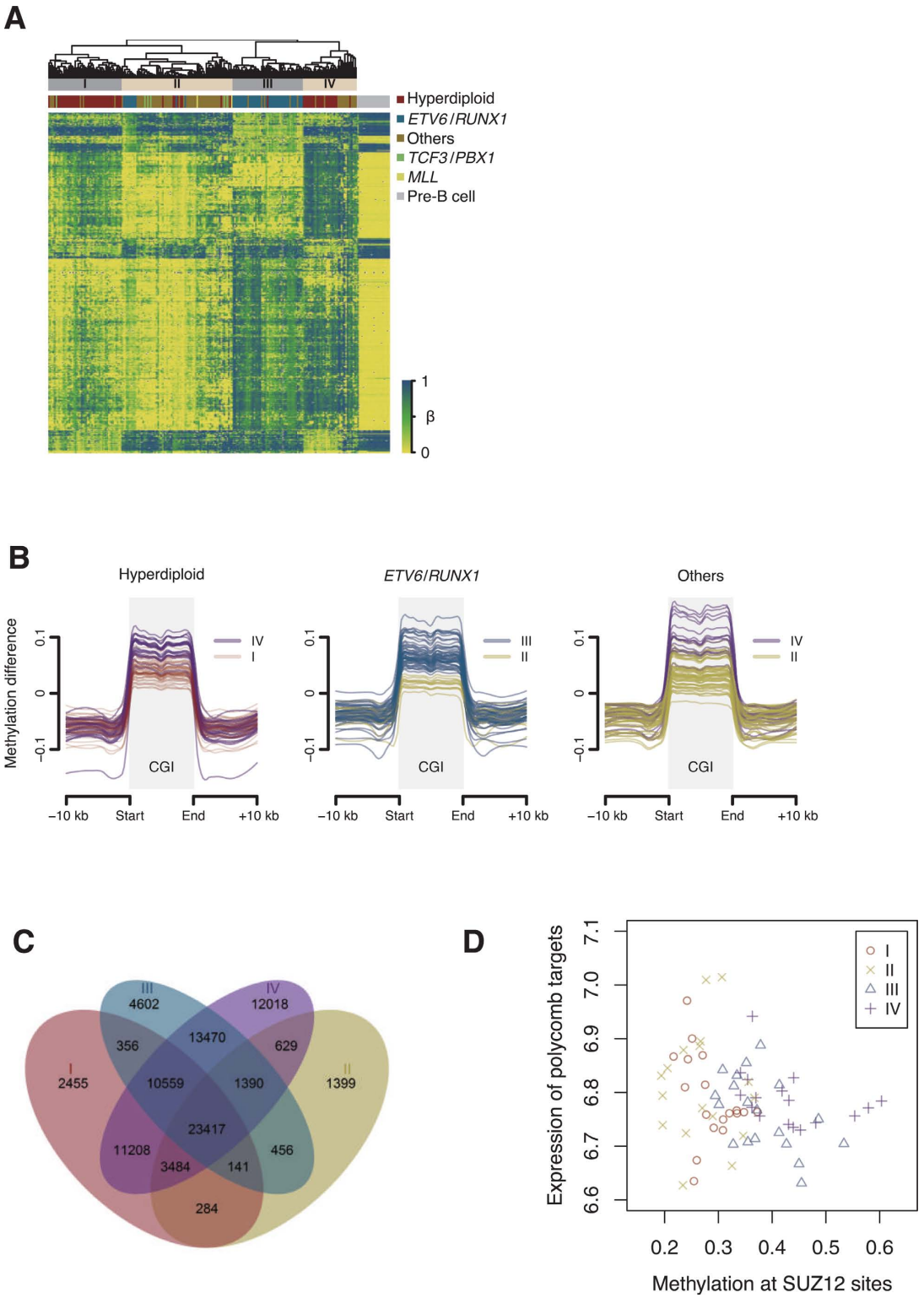

Figure 5. Methylation profiles of 227 B-ALL cases analyzed by Illumina 450k array. (**A**) A hierarchical clustering analysis using 500 most variable CpGs classifies the tumors into four distinct clusters. Some clusters are dominated by specific cytogenetic groups; cluster I by hyperdiploid, cluster II by others, cluster III by *ETV6/RUNX1*, and cluster IV by hyperdiploid/others. (**B**) Local regression of the methylation differences illustrates that each methylation cluster has different degrees of CGI methylation in each cytogenetics group. (**C**) Among the four methylation clusters, many DMRs are shared while some DMRs were cluster-specific. (**D**) Mean methylation at SUZ12-binding sites significantly correlates with mean expression of polycomb target genes (9) ($r = -0.276$; $P = 0.020$), suggesting a role of *de novo* methylation at polycomb sites for repressing their target genes.

additive role of *de novo* methylation at polycomb sites for repressing their target genes (Figure 5D).

Expression profiles in relation with methylation

For a randomly selected 81 B-ALLs, we analyzed expression profiles by microarray and paired them with DNA methylation profiles. A negative correlation of CGIs around promoters and a slight positive correlation of backbones in main bodies were observed (Supplementary Figure S15A). We have listed differentially expressed genes (DEGs) best discriminating B-ALLs from pre-B cells including *ELL2*, *DDIT4*, *DRAM1, NR4A2, GAFA1* and some globin genes (Supplementary Figure S16A). We also identified genes specific to *ETV6/RUNX1* (*CLIC5*, *ACVR1C*, *IGF2BP1*, *DSC2* and *PCLO*) and hyperdiploid (*ZC3H12C*, *SH3BP5*, *ZNF507*, *IL1B* and *PDGFA*) groups (Supplementary Figure S16B and C). In hyperdiploid cases, the average expression levels did not differ between genes in disomy and trisomy chromosomes, suggesting that copy number gain contributes little to gene expression (Supplementary Figure S15B).

We further performed pathway enrichment analysis using DEGs in comparison with pre-B-I and pre-B-II cells, where most B-ALLs originate. Many pathways related to hematopoietic cell development, cell cycle control and mismatch repair were shown to be significantly dysregulated, with more significant enrichment of methylation cluster III for B-cell development and clusters II and IV for cell cycle and mismatch repair. Glucocorticoid receptor signaling was more significantly dysregulated in clusters I and II (Figure 6A).

Using the Ingenuity database, the activation *z*-scores of numerous upstream regulators were calculated based on the expression patterns of their downstream target genes. Chromatin modifiers including NUPR1, SMARCB1 and EP400 and cell cycle controllers including CCND1 and TBX2 were shown to be significantly dysregulated in addition to the well-studied molecules, MYC, RB1, CDKN2A and TP53 (Figure 6B). NUPR1 was the most significantly activated molecule with estimated actions to its target molecules to be increased, and the activation score was less high in methylation cluster III (Supplementary Figure S17). Certain oncogenic transcription networks related to tumor suppression like CDKN2A, RB1 and TP53 showed positive *z*-scores and proto-oncogenes like MYC and TBX2 showed negative z scores. Those activities did not always correlate with the expression levels of the molecule itself, reflecting the complexity between gene expression and function from various interactions and feedback regulations (Supplementary Figure S18). Since these molecules dynamically change during normal development, we further performed comparisons against each developmental stage of B-cell. Despite the relatively higher activities of the tumor suppressors in B-ALLs compared with normal pre-B cells, the leukemia cells still possessed much lower activities than mature peripheral B-cells. The inverse situation was the case for the proto-oncogenes (Supplementary Figure S19). Many cytokines including interferons, interleukins, and tumor necrosis factors and toll-like receptors were also significantly dysregulated, especially in cluster I. To our interest, a conventional drug for B-ALL, doxorubicin, and other trial drugs for B-ALLs, including PD98059, SB203580, SP600125 and CpG oligonucleotide, ranked among the most significant chemicals. This indicates that a substantial proportion of genes and pathways pathobiologically dysregulated in cancers are also downstream effectors of such drugs, and so administration of these drugs may revert some dysregulated genes. Although evidence is not sufficient at this time, the different z-scores according to methylation clusters might indicate that response to such drugs can be different according to methylation clusters. This may be further investigated in well controlled studies in the future.

DISCUSSION

Our analyses suggest that epigenetic modulations in B-ALLs operate in two separate manners; small functional compartments including CGIs, promoters, TF-binding sites and DNase HS are preferentially *de novo* methylated, and large inter-compartmental backbones tend to be demethylated in a subset of leukemias. We found the former modification is in large part related with the bivalent signatures of ESCs that repress lineage-specific genes as well as those repressive signatures of differentiated somatic cells that prevent differentiation to other lineages. Since bivalent histone mark sites in ESCs are known to largely overlap with polycomb binding sites and also with many CGIs, the *de novo* methylation of these three sites have some commonality. Interestingly, CGIs in leukemias were *de novo* methylated throughout promoters and bodies in contrast to the normal physiologic state in which CGIs in gene bodies are more demethylated in down-regulated genes (16,17). This may point toward a crude but still potent means of gene silencing in leukemogenesis.

We found an enrichment of CTBP2 sites in regions that are hypermethylated in B-ALL. This suggests that CTBP2 may recruit factors that drive the observed hypermethylation. CTBP2 along with CTBP1 also function as transcriptional corepressors by recruitment of histone deacetylases or by their intrinsic enzymatic activities (20), and in cancer cells, antagonize apoptosis by repressing tumor suppressor genes (21). Interestingly, CTBPs serve as receptors for oncogenic viruses such as Epstein-Barr virus and adenoviruses (22). REST/NRSF, another TF remarkably enriched in motif scanning, was also recently identified as a polycomb recruiter (23).

The global demethylation of backbones was observed only in a subset of patients (clusters I and IV) and was even subtle compared to the PMDs observed in other solid cancers and cultured cell-lines (11,24,25). Nonetheless, it is intriguing that the demethylation in B-ALLs occurs preferentially in LADs where PMDs in other cancers largely coincide. LADs are large contiguous regions contacting nuclear periphery, characterized by transcriptionally repressive environment from the unique epigenetic nature (14). LADs also largely overlap with the large organized chromatin K9 modifications (LOCKs) (26). The majority of LADs were shown to be constitutively conserved in various tissues (15). In cancers, there is growing evidence that DNA methylation becomes ‘destabilized’ through the hypomethylated blocks or PMDs, giving rise to unregulated

A

Pathway	Methylation cluster I	II	III	IV
DNA damage-induced 14-3-3σ Signaling	4.7	**6.3**	3.6	**5.2**
ATM Signaling	**6.5**	**8.4**	**5.9**	**7.5**
B Cell Development	1.8	0.8	**5.0**	2.7
B Cell Receptor Signaling	**6.7**	3.8	2.8	**5.2**
CD28 Signaling in T Helper Cells	4.7	3.2	**6.3**	2.8
Cell Cycle Control of Chromosomal Replication	**9.4**	**11.6**	**9.4**	**10.3**
Cell Cycle: G1/S Checkpoint Regulation	3.3	**5.2**	4.8	4.0
Cell Cycle: G2/M DNA Damage Checkpoint Regulation	**7.6**	**9.6**	**7.4**	**8.6**
Communication between Innate and Adaptive Immune Cells	**5.3**	3.3	1.9	3.2
Cyclins and Cell Cycle Regulation	3.4	**5.4**	**6.1**	4.3
DNA Double-Strand Break Repair by Homologous Recombination	3.7	**6.4**	3.3	**5.1**
Estrogen-mediated S-phase Entry	**5.7**	**6.4**	**6.5**	**6.3**
GADD45 Signaling	**9.0**	**8.6**	**7.5**	**7.3**
Glucocorticoid Receptor Signaling	**5.1**	**5.1**	2.3	4.0
Granzyme A Signaling	2.9	**8.5**	**5.1**	4.0
Hereditary Breast Cancer Signaling	**7.9**	**10.9**	**9.4**	**8.4**
iCOS-iCOSL Signaling in T Helper Cells	5.0	3.7	**8.1**	3.5
Mismatch Repair in Eukaryotes	4.9	**5.4**	4.5	**5.3**
Mitotic Roles of Polo-Like Kinase	**6.3**	**7.0**	**5.9**	**7.4**
OX40 Signaling Pathway	4.1	0.9	**5.1**	2.7
p53 Signaling	**7.9**	**7.9**	**8.7**	**6.1**
Role of BRCA1 in DNA Damage Response	**13.7**	**15.6**	**13.1**	**14.4**
Role of CHK Proteins in Cell Cycle Checkpoint Control	**6.0**	**7.9**	**7.3**	**7.7**
Role of NFAT in Regulation of the Immune Response	3.5	4.3	**7.2**	3.1
T Helper Cell Differentiation	**5.4**	2.5	4.9	**5.2**
TREM1 Signaling	4.5	3.6	2.4	**5.4**
Type I Diabetes Mellitus Signaling	**5.6**	1.4	3.3	2.7

B

Upstream regulator	Methylation cluster I	II	III	IV	Type
CCND1	-4.9	**-5.1**	-4.4	-4.4	transcription regulator
CDKN2A	**5.1**	**6.0**	**6.1**	**5.8**	transcription regulator
IRF7	5.0	4.5	**5.1**	4.6	transcription regulator
MYC	**-5.5**	**-5.5**	-4.8	-4.8	transcription regulator
NUPR1	**11.8**	**11.8**	**9.9**	**11.5**	transcription regulator
RB1	5.0	4.8	4.8	**5.4**	transcription regulator
SMARCB1	**5.0**	4.5	4.3	4.8	transcription regulator
TBX2	**-5.2**	**-6.0**	**-5.2**	**-5.0**	transcription regulator
TP53	**5.9**	**5.5**	**6.1**	**6.1**	transcription regulator
IFNA2	**5.3**	**5.5**	4.6	4.7	cytokine
IFNG	**7.7**	**5.5**	**6.9**	**6.4**	cytokine
IL1A	**6.0**	**5.2**	4.6	**5.5**	cytokine
IL1B	**7.5**	**6.9**	**7.2**	**7.0**	cytokine
IL2	**5.2**	4.5	4.5	4.6	cytokine
IL3	**5.7**	**5.1**	4.9	4.4	cytokine
IL6	**5.5**	**5.1**	**5.2**	4.6	cytokine
TNF	**8.3**	**6.7**	**7.5**	**7.1**	cytokine
TGFB1	**5.3**	5.0	4.2	4.6	growth factor
TLR3	**6.3**	**6.4**	**5.4**	**6.0**	transmembrane receptor
TLR4	**5.8**	**5.0**	**5.2**	**5.3**	transmembrane receptor
TLR9	**5.7**	**6.0**	4.7	**5.6**	transmembrane receptor
IKBKB	**5.4**	4.8	4.5	**5.1**	kinase
RAF1	**5.1**	4.6	4.1	4.2	kinase
TGM2	**5.9**	**5.4**	**5.6**	**5.0**	enzyme
NFkB (complex)	**7.3**	**6.2**	**6.0**	**5.5**	complex
PDGF BB	**6.1**	**5.5**	4.6	**5.4**	complex
IL1	**5.2**	**5.2**	4.3	4.2	group
Interferon alpha	**5.8**	5.0	**5.4**	4.6	group
EP400	**-5.1**	**-5.1**	**-5.1**	**-5.1**	other
tretinoin	**6.0**	**5.6**	**6.4**	**5.1**	chemical - endogenous
uric acid	**5.3**	**5.3**	4.3	**5.1**	chemical - endogenous
PD98059	**-5.3**	-4.5	-4.6	-4.0	chemical - kinase inhibitor
SB203580	**-6.7**	**-6.3**	**-5.9**	**-6.7**	chemical - kinase inhibitor
SP600125	**-5.3**	**-5.5**	-4.8	**-5.5**	chemical - kinase inhibitor
calcitriol	4.6	**5.7**	**6.0**	**5.2**	chemical drug
CpG oligonucleotide	**5.1**	4.9	3.6	4.6	chemical drug
doxorubicin	**5.8**	**5.5**	**5.8**	**5.3**	chemical drug
lipopolysaccharide	**10.4**	**8.6**	**8.3**	**8.9**	chemical drug
phorbol myristate acetate	**8.7**	**8.1**	**7.7**	**8.4**	chemical drug
poly rI:rC-RNA	**7.9**	**7.7**	**6.7**	**7.4**	chemical reagent

Figure 6. Pathway and upstream regulators dysregulated in different clusters. (**A**) Pathway enrichment analysis for differentially expressed genes (DEGs) compared to pre-B-I and pre-B-II cells were done using the Ingenuity database. Values in boxes indicate -log(FDR-corrected *P*). B-cell development pathway is more enriched for methylation cluster III, cell cycle and mismatch repair for clusters II and IV, and glucocorticoid receptor signaling for clusters I and II. (**B**) Upstream regulators significantly dysregulated. Values in boxes display activation z scores calculated from the expression patterns of their downstream target genes. Chromatin modifiers including NUPR1, SMARCB1 and EP400 and cell cycle controllers including CCND1 and TBX2 are significantly dysregulated as well as MYC, RB1, CDKN2A and TP53. Among chemicals, doxorubicin and other investigative drugs for B-ALLs ranked at the top, with different degrees according to methylation clusters.

genome plasticity. The fact that LADs, LOCKs and PMDs significantly overlap suggests that these domains may be related mechanistically (11,24,27). From an experiment treating B-ALL cell-lines with 5-aza-2′-deoxycytidine, we observed demethylation does not lead to restoration of suppressed genes in all cell lines, but can lead to preferential restoration of polycomb identified genes when compared to non-polycomb genes. This result is consistent with the polycomb DNA methylation signature found in primary leukemias.

Due to the limited availability of public data, functional regions identified in various cell types were used for the enrichment analysis and thus concerns about cell type-specific bias may exist. Although these should be resolved in further analyses using exact controls, our observation may still provide relevant information since many regions like LADs are shared across different cell types. Moreover, we retrieved regions merged from all available tracks from the UCSC genome database, as exemplified by DNase HS concatenated from a number of cell types, which may relieve biases to some extent. Some regions may have specific implications only in specific cell types, e.g. polycomb proteins like SUZ12 are hallmarks of ESCs, and seeing if the tumor cells ectopically gain signatures of other cell types may be sometimes pertinent.

Through rigorous screening for upstream regulators using expression data, we identified candidate key molecules and drugs that may have some therapeutic implications. The most significant was NUPR1, a chromatin-binding protein upregulated in response to cellular stress. NUPR1 has been shown to play complicated roles as both inducer and suppressor in cell and tumor growth and is also associated with resistance to chemotherapeutic agents including doxorubicin (28). NUPR1 was less activated in methylation cluster III, the *ETV6/RUNX1*-dominant cluster. It is interesting that B-ALLs with *ETV6/RUNX1* actually showed superior *in vitro* drug sensitivity to doxorubicin (29). SMARCB1 is a core component of the BAF (hSWI/SNF) complex that relieves repressive chromatic structures, allowing the transcriptional machinery to access its targets effectively. Decreased expression of SMARCB1 is reported to be related with increased steroid resistance in B-ALLs (30). CCND1 or cyclin D1 is a protein involved in cell cycle progression from G1 to S phases. This molecule was also suggested as a therapeutic target of farnesyltransferase inhibitors in B-ALL (31). Since many regulators and chemical drugs showed different degrees of dysregulation according to methylation clusters, one may suggest tailoring drug therapies to such DNA methylation subtypes. Such an approach will require further research and validation.

Collectively, we investigated DNA methylation profile of B-ALLs at a base pair resolution with further validation by DNA methylation arrays. The observations here suggest that B-ALLs inherit epigenetic patterns from their cells of origin but also aberrantly have tumorigenic and embryo-like signatures. Our data provide insight on epigenetic switching in cancer as well as on the epigenetic heterogeneity of pediatric B-ALL.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This research is supported by our clinical collaborators and participating hospitals, which include University of California Davis Medical Center (Dr Jonathan Ducore), University of California San Francisco (Dr Mignon Loh and Dr Katherine Matthay), Children's Hospital of Central California (Dr Vonda Crouse), Lucile Packard Children's Hospital (Dr Gary Dahl), Children's Hospital Oakland (Dr James Feusner), Kaiser Permanente Sacramento (Dr Vincent Kiley), Kaiser Permanente Santa Clara (Dr Carolyn Russo and Dr Alan Wong), Kaiser Permanente San Francisco (Dr Kenneth Leung), and Kaiser Permanente Oakland (Dr Stacy Month), and the families of the study participants. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Environmental Health Sciences, the National Cancer Institute or the Environmental Protection Agency.

Author contributions: S.T.L. and J.L.W. designed the study and wrote the manuscript, with help from all other authors. M.O.M., M.E.F., J.X., M.Z., A.d.S., J.I.M.-S., S.H. helped design and perform experiments. S.T.L. analyzed data with assistance from E.A.H. and R.R. M.W. and J.K. analyzed data. C.M. managed patient accrual and study quality control. All authors reviewed the manuscript during its preparation.

FUNDING

National Institute of Environmental Health Sciences (NIEHS) and Environmental Protection Agency (EPA) [P01ES018172]; NIEHS [R01ES09137, P42ES04705]; National Cancer Institute (NCI) [R01CA155461]; Tobacco-Related Disease Research Program [18CA-0127]; Leukemia and Lymphoma Society [6026-10]. Funding for open access charge: NIH.

Conflict of interest statement. None declared.

REFERENCES

1. Baylin,S.B. and Jones,P.A. (2011) A decade of exploring the cancer epigenome – biological and translational implications. *Nat. Rev. Cancer*, **11**, 726–734.
2. Jones,P.A. and Baylin,S.B. (2007) The epigenomics of cancer. *Cell*, **128**, 683–692.
3. Schlesinger,Y., Straussman,R., Keshet,I., Farkash,S., Hecht,M., Zimmerman,J., Eden,E., Yakhini,Z., Ben-Shushan,E., Reubinoff,B.E. *et al.* (2007) Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer. *Nat. Genet.*, **39**, 232–236.
4. Lee,S.T., Xiao,Y., Muench,M.O., Xiao,J., Fomin,M.E., Wiencke,J.K., Zheng,S., Dou,X., de Smith,A., Chokkalingam,A. *et al.* (2012) A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. *Nucleic Acids Res.*, **40**, 11339–11351.
5. Kulis,M., Heath,S., Bibikova,M., Queiros,A.C., Navarro,A., Clot,G., Martinez-Trillos,A., Castellano,G., Brun-Heath,I., Pinyol,M. *et al.* (2012) Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.*, **44**, 1236–1242.

6. Marco-Sola,S., Sammeth,M., Guigo,R. and Ribeca,P. (2012) The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat. Methods*, **9**, 1185–1188.
7. Nordlund,J., Backlin,C.L., Wahlberg,P., Busche,S., Berglund,E.C., Eloranta,M.L., Flaegstad,T., Forestier,E., Frost,B.M., Harila-Saari,A. *et al.* (2013) Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol.*, **14**, r105.
8. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
9. Lee,T.I., Jenner,R.G., Boyer,L.A., Guenther,M.G., Levine,S.S., Kumar,R.M., Chevalier,B., Johnstone,S.E., Cole,M.F., Isono,K. *et al.* (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell*, **125**, 301–313.
10. Lister,R., Pelizzola,M., Kida,Y.S., Hawkins,R.D., Nery,J.R., Hon,G., Antosiewicz-Bourget,J., O'Malley,R., Castanon,R., Klugman,S. *et al.* (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*, **471**, 68–73.
11. Berman,B.P., Weisenberger,D.J., Aman,J.F., Hinoue,T., Ramjan,Z., Liu,Y., Noushmehr,H., Lange,C.P., van Dijk,C.M., Tollenaar,R.A. *et al.* (2012) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.*, **44**, 40–46.
12. Ziller,M.J., Gu,H., Muller,F., Donaghey,J., Tsai,L.T., Kohlbacher,O., De Jager,P.L., Rosen,E.D., Bennett,D.A., Bernstein,B.E. *et al.* (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature*, **500**, 477–481.
13. Schroeder,D.I., Blair,J.D., Lott,P., Yu,H.O., Hong,D., Crary,F., Ashwood,P., Walker,C., Korf,I., Robinson,W.P. *et al.* (2013) The human placenta methylome. *Proc. Natl. Acad. U.S.A.*, **110**, 6037–6042.
14. Guelen,L., Pagie,L., Brasset,E., Meuleman,W., Faza,M.B., Talhout,W., Eussen,B.H., de Klein,A., Wessels,L., de Laat,W. *et al.* (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, **453**, 948–951.
15. Meuleman,W., Peric-Hupkes,D., Kind,J., Beaudry,J.B., Pagie,L., Kellis,M., Reinders,M., Wessels,L. and van Steensel,B. (2013) Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.*, **23**, 270–280.
16. Suzuki,M.M. and Bird,A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.*, **9**, 465–476.
17. Deaton,A.M. and Bird,A. (2011) CpG islands and the regulation of transcription. *Genes Dev.*, **25**, 1010–1022.
18. Bernstein,B.E., Mikkelsen,T.S., Xie,X., Kamal,M., Huebert,D.J., Cuff,J., Fry,B., Meissner,A., Wernig,M., Plath,K. *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, **125**, 315–326.
19. Voigt,P., Tee,W.W. and Reinberg,D. (2013) A double take on bivalent promoters. *Genes Dev.*, **27**, 1318–1338.
20. Chinnadurai,G. (2002) CtBP, an unconventional transcriptional corepressor in development and oncogenesis. *Mol. cell*, **9**, 213–224.
21. Chinnadurai,G. (2009) The transcriptional corepressor CtBP: a foe of multiple tumor suppressors. *Cancer Res.*, **69**, 731–734.
22. McClellan,M.J., Wood,C.D., Ojeniyi,O., Cooper,T.J., Kanhere,A., Arvey,A., Webb,H.M., Palermo,R.D., Harth-Hertle,M.L., Kempkes,B. *et al.* (2013) Modulation of enhancer looping and differential gene targeting by Epstein-Barr virus transcription factors directs cellular reprogramming. *PLoS Pathog.*, **9**, e1003636.
23. Dietrich,N., Lerdrup,M., Landt,E., Agrawal-Singh,S., Bak,M., Tommerup,N., Rappsilber,J., Sodersten,E. and Hansen,K. (2012) REST-mediated recruitment of polycomb repressor complexes in mammalian cells. *PLoS Genet.*, **8**, e1002494.
24. Hansen,K.D., Sabunciyan,S., Langmead,B., Nagy,N., Curley,R., Klein,G., Klein,E., Salamon,D. and Feinberg,A.P. (2014) Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization. *Genome Res.*, **24**, 177–184.
25. Hon,G.C., Hawkins,R.D., Caballero,O.L., Lo,C., Lister,R., Pelizzola,M., Valsesia,A., Ye,Z., Kuan,S., Edsall,L.E. *et al.* (2012) Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.*, **22**, 246–258.
26. Wen,B., Wu,H., Shinkai,Y., Irizarry,R.A. and Feinberg,A.P. (2009) Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat. Genet.*, **41**, 246–250.
27. Luperchio,T.R., Wong,X. and Reddy,K.L. (2014) Genome regulation at the peripheral zone: lamina associated domains in development and disease. *Curr. Opin. Genet. Dev.*, **25**, 50–61.
28. Chowdhury,U.R., Samant,R.S., Fodstad,O. and Shevde,L.A. (2009) Emerging role of nuclear protein 1 (NUPR1) in cancer biology. *Cancer Metastasis Rev.*, **28**, 225–232.
29. Frost,B.M., Forestier,E., Gustafsson,G., Nygren,P., Hellebostad,M., Jonsson,O.G., Kanerva,J., Schmiegelow,K., Larsson,R. and Lonnerholm,G. (2004) Translocation t(12;21) is related to in vitro cellular drug sensitivity to doxorubicin and etoposide in childhood acute lymphoblastic leukemia. *Blood*, **104**, 2452–2457.
30. Pottier,N., Cheok,M.H., Yang,W., Assem,M., Tracey,L., Obenauer,J.C., Panetta,J.C., Relling,M.V. and Evans,W.E. (2007) Expression of SMARCB1 modulates steroid sensitivity in human lymphoblastoid cells: identification of a promoter SNP that alters PARP1 binding and SMARCB1 expression. *Hum. Mol. Genet.*, **16**, 2261–2271.
31. Costa,C.B., Casalta-Lopes,J., Andrade,C., Moreira,D., Oliveira,A., Goncalves,A.C., Alves,V., Silva,T., Dourado,M., Nascimento-Costa,J.M. *et al.* (2012) Farnesyltransferase inhibitors: molecular evidence of therapeutic efficacy in acute lymphoblastic leukemia through cyclin D1 inhibition. *Anticancer Res.*, **32**, 831–838.