Whole-genome fingerprint of the DNA methylome during human B cell differentiation

Marta Kulis, Angelika Merkel, Simon Heath, Ana C. Queirós, Ronald P. Schuyler, Giancarlo Castellano, Renée Beekman, Emanuele Raineri, Anna Esteve, Guillem Clot, et al.

▶ **To cite this version:**

Marta Kulis, Angelika Merkel, Simon Heath, Ana C. Queirós, Ronald P. Schuyler, et al.. Whole-genome fingerprint of the DNA methylome during human B cell differentiation. Nature Genetics, 2015, 47 (7), pp.746-756. 10.1038/ng.3291 . hal-01163753

HAL Id: hal-01163753

https://univ-rennes.hal.science/hal-01163753

Submitted on 4 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Whole-genome fingerprint of the DNA methylome during human B-cell differentiation

Marta Kulis[1], Angelika Merkel[2], Simon Heath[2], Ana C. Queirós[1], Ronald P. Schuyler[2], Giancarlo Castellano[1], Renée Beekman[1], Emanuele Raineri[2], Anna Esteve[2], Guillem Clot[1], Nuria Verdaguer-Dot[1], Martí Duran-Ferrer[1,2], Nuria Russiñol[1], Roser Vilarrasa-Blasi[1], Simone Ecker[3], Vera Pancaldi[3], Daniel Rico[3], Lidia Agueda[2], Julie Blanc[2], David Richardson[4], Laura Clarke[4], Avik Datta[4], Marien Pascual[5], Xabier Agirre[5], Felipe Prosper[5,6], Diego Alignani[7], Bruno Paiva[6,7], Gersende Caron[8], Thierry Fest[8], Marcus O. Muench[9,10], Marina E. Fomin[9,10], Seung-Tae Lee[11], Joseph L. Wiemels[12], Alfonso Valencia[3], Marta Gut[2], Paul Flicek[4], Hendrik G. Stunnenberg[13], Reiner Siebert[14], Ralf Küppers[15], Ivo G. Gut[2], Elías Campo[1], José I. Martín-Subero[1]

[1]Unidad de Hematopatología, Servicio de Anatomía Patológica, Hospital Clínic, Universitat de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona 08036, Spain. [2]Centro Nacional de Análisis Genómico, Parc Científic de Barcelona, Barcelona 08028, Spain. [3]Structural Biology and Biocomputing Program, Spanish National Cancer Research Centre (CNIO), Spanish National Bioinformatics Institute, Madrid 28029, Spain. [4]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus Hinxton, Cambridge CB10 1SD, United Kingdom. [5]Area de Oncología, Centro de Investigación Médica Aplicada (CIMA), Universidad de Navarra, Pamplona 31008, Spain. [6]Servicio de Hematologia, Clinica Universidad de Navarra, Pamplona 31008, Spain. [7]Flow Cytometry Core, Centro de Investigación Médica Aplicada (CIMA), Universidad de Navarra, Pamplona 31008, Spain. [8]Université de Rennes 1, INSERM U917, hemato-biologique CHU Pontchaillou, Rennes 35043, France. [9]Blood Systems Research Institute, San Francisco CA 94118, USA. [10]Department of Laboratory Medicine, University of California San Francisco, San Francisco CA 94143, USA. [11]Department of Laboratory Medicine, Yonsei University College of Medicine, Seoul 120-752, Republic of Korea. [12]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco CA 94158, USA. [13]Molecular Biology, NCMLS, FNWI, Radboud University Nijmegen 6500 HB, The Netherlands. [14]Institute of Human Genetics, Christian-Albrechts-University, Kiel 24105, Germany. [15]Institute of Cell Biology (Cancer Research), Medical School, University of Duisburg-Essen, Essen 45122, Germany.

Correspondence should be addressed to:
José I. Martín-Subero
Centre Esther Koplowitz
Rosselló 153, 08036-Barcelona, Spain
Phone: +34-93-2275400
E-mail: imartins@clinic.ub.es

Abstract

We analyzed the DNA methylome of ten subpopulations spanning the entire B-cell differentiation program by whole-genome bisulfite sequencing and high-density microarrays. We observed that non-CpG methylation disappeared upon B-cell commitment whereas CpG methylation changed extensively during B-cell maturation, showing an accumulative pattern and affecting around 30% of all measured CpGs. Early differentiation stages mainly displayed enhancer demethylation, which was associated with upregulation of key B-cell transcription factors and affected multiple genes involved in B-cell biology. Late differentiation stages, in contrast, showed extensive demethylation of heterochromatin and methylation gain of polycomb-repressed areas, and did not affect genes with apparent functional impact in B cells. This signature, which has been previously linked to aging and cancer, was particularly widespread in mature cells with extended life span. Comparing B-cell neoplasms with their normal counterparts, we identified that they frequently acquire methylation changes in regions undergoing dynamic methylation already during normal B-cell differentiation.

The multitude of cell types and tissues of an organism can be defined by their particular epigenetic makeup[1-2]. DNA methylation is an important component of the epigenome which is extensively modulated during regulatory and developmental processes, both in the context of physiological and pathological conditions[3-5]. Although recent reports have analyzed the DNA methylation profiles of various cell types at the whole-genome scale[1,6-16], the DNA methylome of a single human cell type during its complete differentiation process has not been defined so far. The B-cell lineage represents a paradigmatic cellular model to study the dynamic epigenome during cell development and specification because major B-cell maturation stages have distinct phenotypic and gene expression features and can be isolated in sufficient numbers from hematopoietic tissues[17-19].

B-cell lymphopoiesis is a complex and tightly coordinated process guided by a hierarchical expression of different stage-specific transcription factors and microenvironmental influences[20-21]. The process starts in the bone marrow, where hematopoietic stem cells differentiate into multipotent progenitors and common lymphoid progenitors, which then commit to the B-cell lineage and give rise to precursor B cells. These precursors gradually rearrange their immunoglobulin genes and differentiate into mature naive B cells, which leave the bone marrow to enter the blood stream. Resting naive B cells transit through lymph nodes and, eventually, they are activated by specific antigens via activation of the B-cell receptor, which induces the germinal center reaction. Germinal center B cells further rearrange and mutate their immunoglobulin genes, rapidly proliferate and differentiate. Finally, the germinal center reaction gives rise to plasma cells producing large amounts of high-affinity antibodies and memory B cells. Plasma cells exiting the lymph nodes migrate to the bone marrow where they can reside for extended periods of time, and long-lived memory B cells recirculate through the blood and lymphoid organs, providing the basis

for enduring humoral immunity[22-23]. Hence, an interesting feature of the B-cell maturation process is that it entails a variety of cells with different functional features, proliferation abilities, microenvironmental influences and life spans, providing an exceptional opportunity to study the epigenome in the context of different biological processes, and to provide insights into the fields of cell differentiation, B-cell biology, cancer and aging.

RESULTS

Whole-genome DNA methylation maps of B-cell subpopulations

We generated unbiased DNA methylation maps of uncommitted hematopoietic progenitor cells (HPCs) and five B-cell lineage subpopulations, including pre-B-II cells (preB2Cs), naive B cells from peripheral blood (naiBCs), germinal center B cells (gcBCs), memory B cells from peripheral blood (memBCs) and plasma cells from bone marrow (bm-PCs), by whole-genome bisulfite sequencing (WGBS) (**Fig. 1a** and **Supplementary Table 1**). We sequenced two biological replicates of each subpopulation and a total of 2,217 billion base pairs (bp) of which 85–95% could be mapped (mean depth of 54-fold per sample) (**Supplementary Table 2**). On average, we measured methylation levels of 22.7 million CpGs per sample (ranging from 21 to 25 million). Unsupervised principal component analysis (PCA) of CpG methylation levels showed that B-cell subpopulations segregate according to their developmental stage (**Fig. 1b**). Globally, B-cell differentiation is accompanied by a gradual widespread demethylation of the genome, which was more pronounced at late differentiation stages such as memBC and bm-PC (**Fig. 1c-e**). The global methylation status of CpGs was largely bimodal in all sorted cell populations and the level of partially methylated regions increased to 19–24% in advanced maturation stages (**Fig. 1e**). This result contrasts to other WGBS studies using whole tissues, in which the proportion of partially methylated regions is usually high[24], and highlights the importance of using purified cell subpopulations for DNA methylation studies.

The results obtained by WGBS were complemented with the analysis of 3 to 9 replicates of 10 different B-cell subpopulations by high-density DNA methylation microarrays[25] (**Fig. 1a**). These subpopulations included those analyzed by WGBS as well as pre-B-I cells (preB1Cs), immature B cells (iBC), tonsillar naive B cells (t-naiBCs) and tonsillar plasma cells (t-PCs) (**Supplementary Table 1**). The biological replicates of each subpopulation analyzed by WGBS or microarrays showed high reproducibility (correlation coefficient > 0.95) (**Supplementary Fig. 1**), and both high-throughput techniques were further validated by bisulfite pyrosequencing (**Supplementary Fig. 2**). In line with WGBS, PCA of microarray data separated B-cell subpopulations mostly according to their developmental stage (**Fig. 1f**) and we observed a gradual global methylation loss throughout B-cell maturation (**Fig. 1g**). Interestingly, samples clustered into "antigen-inexperienced" and "antigen-experienced" cells (**Supplementary Fig. 3**). This finding indicates that proliferative gcBCs start a massive reconfiguration of the DNA methylome[18-19] which continues in cell subpopulations with extended life span such as memBCs and bm-PCs.

Demethylation of non-CpG sites upon B-cell commitment

Cytosine methylation in mammals can occur outside CpGs, a phenomenon commonly observed in embryonic stem cells (ESCs) and neurons[7-8,16]. In our study, considering non-CpG sites with high-confidence methylation estimates and having ruled out suboptimal bisulfite conversion and sequence variants, we did observe non-CpG methylation, which was primarily confined to the most undifferentiated cells (i.e. HPCs) and occurred mainly in a CpApC context (**Fig. 2a-c**). Significant non-CpG methylation in HPCs was detected by WGBS at 25,763 sites in replicate 1 and 16,838 in replicate 2, with a mean methylation levels of 25.1% and 24.7%, respectively. Non-CpG methylation in HPCs frequently targeted the same sites in the two biological replicates, fact that was also confirmed by bisulfite pyrosequencing in independent

samples (**Fig. 2d** and **Supplementary Fig. 4**). Non-CpG methylation in HPCs was preferentially located in gene bodies (both introns and exons) and depleted in lamina-associated domains (**Supplementary Fig. 5**). Although the methylation microarrays only measured 3,091 non-CpG sites, we could confirm the WGBS results (**Fig. 2e,f**). Similarly to reports in ESCs [26], methylated non-CpGs were flanked by methylated CpGs. However, demethylation of non-CpGs and CpGs was not simultaneous. We observed a dramatic demethylation of non-CpGs upon B-cell commitment in preB2Cs, but 97% of these demethylated non-CpGs remained flanked by methylated CpGs (**Supplementary Fig. 6**). These data were confirmed by bisulfite pyrosequencing (**Supplementary Fig. 4**) and indicate that non-CpG methylation is passively erased from HPCs to preB2Cs without simultaneous demethylation of flanking CpGs (**Fig. 2g**). Although previous studies have reported that high expression of *DNMT3A* and *DNMT3B* is associated with non-CpG methylation[26-27], we did not identify consistent differences in the expression levels of DNMTs between HPCs and preBCs (**Supplementary Fig. 7**).

Identification of dynamic DNA methylation patterns

Next, we focused our analysis on the modulation of CpG methylation throughout the complete B-cell maturation program. We identified dynamic methylation levels in 4.93 million CpGs (> 0.25 methylation change in one set of samples and > 0.1 in the other), which represent 30.6 % of the 16.1 million CpGs with methylation estimates in all 12 samples analyzed. The cell subtypes showing the most pronounced methylation changes compared to the preceding stage were gcBCs, memBCs and bm-PCs (**Fig. 3a** and **Supplementary Fig. 8**). Interestingly, multiple genes directly involved in B-cell differentiation (e.g. *ARID3A*, *BCL2*, *BLK*, *EBF1* and *IRF4*) show a complex modulation of their DNA methylation profile across the gene length, with different regulatory elements losing methylation at distinct maturation stages (**Fig. 3b** and **Supplementary Fig. 9**). In a previous study, 5.6 million dynamic CpGs were identified to be

differentially methylated in a wide range of human cell types and tissues[1]. Only one third of the dynamic CpGs in B-cell differentiation defined in our study overlapped with these CpGs (**Supplementary Fig. 10**). Although this result may in part be caused by differences in coverage and bioinformatic pipelines used in the two studies, it suggests that the majority of dynamic methylation in B cells may entail CpGs not previously detected to be differentially methylated in other cell types and tissues.

Similarly to WGBS, although at lower scale, microarray data revealed that 22.4% of the measured CpGs showed variable methylation levels during B-cell differentiation (n = 106,562, see online methods for details). Due to the more comprehensive set of samples, we used these data to define sets of CpGs sharing similar DNA methylation dynamics during B-cell differentiation. To that end, we based our next analysis on a linear model of B-cell maturation, from HPC to terminal differentiation of gcBC into t-PC and long-lived bm-PC. We defined 20 major modules containing at least 500 CpGs that could be classified according to four general DNA methylation modulation patterns during differentiation: i) methylation levels decrease (9 modules), ii) methylation levels first decrease and then increase (3 modules), iii) methylation levels first increase and then decrease (2 modules) and iv) methylation levels increase (6 modules) (**Supplementary Data 1**). Three of these 20 modules covered 57.8% of all dynamic CpGs: module 8 defined by CpG demethylation starting in gcBCs and continuing in bm-PCs (n = 34,604), module 9 showing demethylation only in bm-PCs (n = 13,044) and module 20 characterized by hypermethylation in bm-PCs only (n = 13,949) (**Fig. 3c**). These data confirmed that DNA hypomethylation mainly occurs in the gcBCs and bm-PCs. Furthermore, it could be appreciated that hypermethylation is a late event, mainly occurring in bm-PCs. Overall, we may conclude that 84.5% of the dynamic CpGs either gain or lose methylation (DNA methylation modulation patterns i and iv, n = 90,070) as B-cell differentiation progresses. Hence, each B-cell differentiation stage has its specific DNA methylation pattern, but furthermore retains an epigenetic memory of the previous stages. Interestingly, although the

methylation levels of the 20 modules are able to separate B-cell maturation stages, we performed a complexity reduction step and selected 5 CpGs in genes important for B-cell differentiation such as *BLK*, *SEMA4B*, *ARID3A*, *AICDA* and *PRDM1* whose methylation levels could accurately classify each maturation stage (**Supplementary Fig. 11**).

Functional analysis of dynamically-methylated regions

In general, CpGs losing methylation at any B-cell maturation stage were preferentially located in introns, intergenic regions and repetitive elements (e.g. LINEs, SINEs and LTRs), and were enriched for genomic areas lacking CpG islands (CGIs) (**Supplementary Figs. 12 and 13**). In contrast, CpGs gaining methylation were enriched for CGIs and promoter regions (**Supplementary Fig. 12**). Additionally, we classified differentially methylated sites using a categorization of the genome into different chromatin states in immortalized mature B cells[28] (**Supplementary Fig. 14**). Both WGBS and microarrays revealed that the majority of dynamic CpGs during B-cell differentiation were enriched for enhancer regions (mainly intragenic), polycomb-repressed regions or heterochromatin (**Fig. 3d,e** and **Supplementary Figs. 15 and 16**). Demethylation in precursor B cells was mostly related to enhancer elements whereas that occurring exclusively from gcBCs onwards was preferentially located in heterochromatic regions. Gain of CpG methylation was a rare event in early B-cell differentiation but rather frequent in mature B cells, especially in bm-PCs. Such CpG hypermethylation targeted preferentially polycomb-repressed regions (**Fig. 3d,e**).

Next, we studied the mechanisms underlying enhancer demethylation in the B-cell differentiation process. We identified a significant enrichment (FC > 2, $P < 0.01$) in transcription factor binding sites (TFBSs) of key B-cell transcription factors such as BCL11A, EBF1, IRF4, MEF2A, MEF2C, PAX5 or TCF3 (E2A) (**Fig. 4a, Supplementary Fig. 17** and **Supplementary Data**

2 and 3). As B-cell commitment is associated with the expression of lineage-specific TFs[20], we analyzed the transition from HPC to preB1C (module 1) in detail and, globally, we observed an inverse correlation between TF expression and the methylation levels of their binding sites (**Fig. 4b,c**). We further investigated the association between TF expression and the methylation status of their binding sites during the entire differentiation program, and in general, once a TFBS becomes demethylated at any B-cell differentiation stage, it remains unmethylated in subsequent stages, suggesting an epigenetic memory of TF binding[24] (**Supplementary Fig. 18**).

At the functional level, genes within microarray-based methylation modules enriched for enhancer elements were involved in multiple immune system-related functions (**Supplementary Fig. 19** and **Supplementary Data 4**). In contrast, CpGs in modules enriched for heterochromatin or polycomb-repressed regions, targeted genes not involved in the immune system but rather affected terms such as development, locomotion or behavior (**Supplementary Fig. 19** and **Supplementary Data 4**). Based on this observation, we hypothesized that differential methylation in enhancer elements may be globally associated with gene expression, while that affecting inactive elements (i.e. heterochromatin or polycomb-repressed regions) may not. We initially explored the transcriptome of B-cell subpopulations and observed that they clustered separately using an unsupervised approach (**Fig. 4d**). We then calculated the number of genes differentially expressed and differentially methylated by comparing adjacent cell subpopulations. Globally, there was a poor association, and large transcriptional changes could be related to a minor modulation of the DNA methylome and vice versa (**Fig. 4e**). We further explored the association between DNA methylation patterns and gene expression by exploring genes with dynamic methylation in enhancers, heterochromatin and polycomb-repressed regions. We observed that both variability and mean expression levels of genes containing dynamic CpGs in enhancer elements were much higher in comparison with genes showing modulation of CpG methylation in non-functional chromatin states such as heterochromatin and polycomb-repressed regions ($P <$

0.001; **Fig. 4f-h**). However, regardless of the DNA methylation pattern throughout B-cell maturation, dynamic CpGs target genes with higher expression levels and variation throughout the whole B-cell differentiation. Thus, these data suggest that there is no direct correlation between DNA methylation and gene expression, as previously shown elsewhere[6,10], but rather that dynamic CpGs affecting functional elements target immune system-related genes whose expression is modulated during B-cell differentiation.

Next, we sought to analyze the functions of genes with dynamic enhancer methylation in more detail. As B-cell receptor signaling is a key element specific for B-cell differentiation, we studied the methylation pattern of 41 genes involved in this function. Out of those, 38 (93%) had a total of 234 CpGs with dynamic methylation and were preferentially located in enhancers (**Supplementary Fig. 20**). Furthermore, we observed that 38% of all enhancers ($P <$ 0.001) belonged to a B-cell specific functional gene network[29] (**Supplementary Fig. 21)**.

As our study comprises cell subpopulations isolated from three different compartments (bone marrow, peripheral blood and tonsil), we aimed at detecting particular DNA methylation imprints related to these different locations but we did not identify any consistent pattern. However, we did observe that naiBCs isolated from different compartments (i.e. peripheral blood or tonsils) showed a massive modulation of their transcriptome while their methylomes remained virtually identical (**Supplementary Figs. 22 and 23**). Based on this observation, we may hypothesize that the gene expression changes in t-naiBCs are essential to optimize antigen recognition in the tonsil, followed by successful B-cell activation. However, when no antigen is found, the t-naiBCs will re-enter the blood stream. Hence, the t-naiBCs state has to be reversible, which may be the reason why no changes were observed at the level of the DNA methylome.

Epigenetic link among long-lived B cells, cancer and aging

We observed that long-lived B cells such as memBCs and bm-PCs[23] show an extensive perturbation of their DNA methylome as compared to gcBCs (**Supplementary Fig. 8**). Remarkably, although bm-PCs and newly generated t-PCs share a similar transcriptome, they have different methylomes (**Supplementary Fig. 22**). The major DNA methylation changes in memBCs and bm-PCs affect presumably non-functional elements such as heterochromatin and polycomb-repressed regions. Thus, our findings suggest that part of the epigenetic makeup of memBCs and bm-PCs is caused by an epigenetic drift associated with their long life span[23]. To determine whether such drift may be related to the expression of DNMTs, we analyzed *DNMT1*, *DNMT3A* and *DNMT3B* by qRT-PCR in sorted gcBCs, memBCs, t-PCs and bm-PCs. *DNMT1* showed low levels in memBCs, t-PCs and bm-PCs (**Supplementary Fig. 7**). As these cell types are considered to be non-proliferative, our finding implies that demethylation occurs either through an active mechanism or passively if these cell types still proliferate at a low rate[30]. Gain of methylation of polycomb-repressed regions has been linked to the activity of *DNMT3A* and *DNMT3B*[31-32]. We detected that as compared to gcBCs, *DNMT3A* was upregulated in t-PCs and bm-PCs, and at a lower extent also in memBCs. Thus, these results may suggest that minor increases of this enzyme in the context of long-lived cells could result in hypermethylation of polycomb-repressed areas. To obtain further insights into the mechanisms of chromatin repression by hypermethylation at polycomb-repressed regions, we performed bisulfite sequencing of chromatin immunoprecipitated DNA (ChIP-BS) with a H3K27me3 antibody[33-34]. This experiment suggests that in memBCs, H3K27me3 and DNA methylation coexist both in CpG-rich and CpG-poor regions, and that the DNA methylation levels within nucleosomes containing H3K27me3 were lower than regions outside such nucleosomes (**Supplementary Fig. 24**).

Remarkably, demethylation of heterochromatin (in part bound to the nuclear lamina) and hypermethylation of polycomb-repressed regions are among the most frequent epigenetic changes in solid and hematological tumors[11,35-38]. To analyze whether those CpGs showing differential methylation in long-lived B cells overlap with those becoming differentially methylated in cancer, we used methylation data from various lymphoid neoplasms including acute lymphoblastic leukemia[39] (ALL, n = 46), chronic lymphocytic leukemia[10] (CLL, n = 139), diffuse large B-cell lymphoma (DLBCL, n = 40) and multiple myeloma[40] (MM, n = 104). We analyzed the methylation levels of CpGs losing methylation (modules 8 and 9) and those gaining methylation in memBCs and bm-PCs (modules 19 and 20). Indeed, we observed that, considering these CpGs, neoplastic cells show a DNA methylation profile similar to memBCs and bm-PCs (**Fig. 5a-d** and **Supplementary Fig. 25**). To further evaluate the epigenetic link between normal B-cell differentiation and neoplastic transformation, we compared the DNA methylome of B-cell neoplasms with their normal cell counterparts, e.g. ALL[39] vs. preBCs, DLBCL-GCB type vs. gcBCs, and MM[40] vs. PCs (**Fig. 6a** and **Supplementary Data 5**). The results indicate that a large fraction of the CpGs differentially methylated in tumors are dynamically methylated during normal B-cell differentiation, ranging from 53% to 82% for hypermethylated and from 29% to 84% for hypomethylated sites (**Fig. 6b**). Interestingly, we identified that hypomethylation in ALL was enriched in enhancers whereas hypomethylation in DLBCL and MM predominantly affected heterochromatin (**Fig. 6c**). Additionally, although ALL cells are arrested at the preBCs stage, they acquire hypermethylation in polycomb-repressed regions, which is characteristic of more mature differentiation stages (**Fig. 6d**). MM cells, in contrast, do not acquire hypermethylation of polycomb-repressed regions because their cellular origin already shows this feature, but, as they downregulate the B-cell program, they acquire hypermethylation of B cell-specific enhancers[40].

Finally, global hypomethylation accompanied by local hypermethylation of polycomb targets is also a molecular hallmark of aging[41-45]. We analyzed DNA methylation values of CpGs

within modules 8, 9, 19 and 20 using data from 694 peripheral blood samples from donors ranging from 0 to 101 years[43-44]. Indeed, we detected a significant correlation with age (P < 0.01) following the same tendency as in long-lived cells and cancer cells, i.e. methylation of heterochromatin diminished with age whereas methylation of polycomb targets increased (**Fig. 5e-f; Supplementary Fig. 25**). We next compared B cells with short and long lifespan (naiBCs and memBCs, respectively) isolated simultaneously from individuals of different age and we detected that memBCs acquire hypomethylation of heterochromatin and hypermethylation of polycomb targets regardless of the chronological age of the donor (**Supplementary Fig. 26**). As the cellular composition of blood changes with age, our results imply that a relative increase of long-lived cells in older individuals may represent a confounding variable in age-related methylation studies[46].

Discussion

The B-cell maturation process is an orchestrated program integrating internal and environmental signals to finally give rise to plasma cells and memory B cells that play an essential role in adaptive immunity. Although previous reports have studied epigenetic changes in the context of B-cell differentiation, they only studied partial DNA methylomes either of precursor or mature B cells[17-19,47]. With the exception of few cell subpopulations such as transitional B cells, CD5+ B cells and splenic marginal zone B cells, our study comprises all major B-cell differentiation stages and represents the first whole-genome epigenetic characterization of a complete human cell lineage from progenitor to terminally-differentiated cells. The comprehensive nature of our study has allowed us to provide epigenetic insights into different scientific fields and offers a resource for researchers working in different areas of cell differentiation, B-cell biology and related diseases, cancer and aging, both at single gene and genome-wide levels.

Our study points to a massive perturbation of the DNA methylome during B-cell differentiation, affecting 30% of all autosomal CpG sites. These changes follow an accumulative pattern in which each B-cell maturation stage, although characterized by a particular signature, keeps an epigenetic memory of past differentiation stages. In contrast to other reports in hematopoietic precursors[48-49], we did observe non-CpG methylation in HPCs, which virtually disappeared upon B-cell commitment in regions lacking simultaneous CpG demethylation. Precursor B cells showed relatively small losses of CpG methylation, which mostly affected enhancers containing binding sites of B cell-specific TFs. The functional link between TF binding, CpG demethylation and enhancer activation has been recently analyzed during stem cell differentiation[50-51] as well as in hematopoietic cells[52-53], non-hematopoietic cells[54] and cancer[40,55].

Interestingly, more than half of all enhancers defined in immortalized mature B cells show dynamic DNA methylation throughout the B-cell differentiation process, and 38% of all genes with dynamically-methylated enhancers are included in a regulatory network associated with human B cells[29]. Although mainstream research on DNA methylation still remains centered on promoter regions, our results imply that DNA methylation changes in enhancers seems to be more closely related to cell specification and maturation[1,9,56]. However, similar to other recent studies[6,10,57], we rarely observed a direct correlation between gene expression and DNA methylation, even in regulatory elements. Our study also suggests that at later stages of B-cell differentiation (from naiBC onwards), DNA methylation changes are more guided by other mechanisms than the intrinsic program of B-cell TFs. Upon antigen encounter, the germinal center reaction is induced and, at this stage, gcBCs start experiencing a wave of global demethylation, mostly affecting late-replicating regions such as heterochromatin and DNA repeats, and local hypermethylation of polycomb-repressed regions. This finding can be partially explained by the high proliferation rate, as normal proliferative tissues tend to lose methylation at late-replicating regions[58]. However, downstream B-cell subpopulations derived

from gcBCs, such as non-proliferative memBCs and PCs, which recirculate through the body and reside in bone marrow, respectively, acquire additional epigenetic changes in heterochromatin and polycomb-repressed regions. We postulate that these additional changes may be related to a potential epigenetic drift in the context of longevity that may be mediated by downregulation of *DNMT1* and slight upregulation of *DNMT3A*.

Hypomethylation of heterochromatin and hypermethylation of polycomb-repressed regions have been previously described as an epigenetic hallmark of organismal aging, cellular senescence and cancer[4,41-44,59-61]. Here, we observe that this signature starts in proliferating gcBCs and becomes particularly enhanced in non-proliferative long-lived B cells. Based on our results, we hypothesize that not all the cells of the organism are subjected to epigenetic drift as an influence of time but only those with long life span. At last, one of the most relevant implications of our study is related to the field of cancer. We demonstrate that B-cell tumors and long-lived cells share in part similar DNA methylation signatures. Furthermore, comparing various B-cell tumors with their normal cellular counterparts, we observed that a large proportion of the differentially methylated sites overlap with those undergoing dynamic methylation during normal differentiation, especially with those changing in memBCs and bm-PCs. Interestingly, similarly to preBCs, hypomethylation in ALL is enriched for enhancer elements whereas in line with gcBCs and bm-PCs, hypomethylation in DLBCL and MM is mostly enriched for heterochromatin. In general, this finding suggests that the epigenetic configuration of a particular maturation stage influences the DNA methylation changes acquired during its clonal expansion and neoplastic transformation. This novel strategy of analyzing the DNA methylome of B-cell tumors in the context of the entire differentiation program may allow us to provide new insights into the role of DNA methylation in cancer. We postulate that changes shared during neoplastic transformation and normal differentiation may represent epigenetic passengers whereas those exclusively taking place in tumor cells should contain epigenetic drivers with a potential functional impact in the disease.

URLs. European Genome-Phenome Archive, http://www.ebi.ac.uk/ega/studies/; Blueprint project, http://www.blueprint-epigenome.eu; The R Project for Statistical Computing, http://www.r-project.org/; UCSC Genome Browser (http://genome.ucsc.edu/).

Accession codes. WGBS data have been deposited in the European Genome-Phenome Archive (EGA) under the accession numbers EGAD00001001304 and EGAS00001000272. DNA methylation and gene expression microarray data are available from the EGA under accession numbers EGAS00001001196 and EGAS00001001197, respectively. Additional published data from precursor B cells are available through Gene Expression Omnibus under accession number GSE45461.

ACKNOWLEDGEMENTS

We thank Carlos López-Otín for critical reading of this manuscript, Marc Dabad Castellà for his assistance with the WGBS data analysis and Miguel A. Peinado for providing RNA from HCT116 DKO cells. This work was funded by the European Union's Seventh Framework Programme through the Blueprint Consortium (grant agreement 282510), the Spanish Ministry of Economy and Competitivity (MINECO) (project SAF2009-08663). Methylation microarrays were outsourced to the Spanish "Centro Nacional de Genotipado" (CEGEN-ISCIII). M.K. is supported by the AGAUR (Generalitat de Catalunya), EC is an Academia Researcher of the Institució Catalana de Recerca i Estudis Avançats, and J.I.M.-S. is a Ramon y Cajal researcher of the MINECO.

AUTHOR CONTRIBUTIONS

M.K., A.C.Q., N.R., M.P., X.A., F.P., D.A., B.P., G.Car., T.F., M.O.M., M.E.F., S.-T.L., J.L.W., provided samples from healthy donors and/or purified B-cell subpopulations.

M.K., A.C.Q., G.Cas., R.B. and G.Clo. analyzed DNA methylation and gene expression arrays. L.A., J.B. and M.G. performed WGBS library preparation and sequencing. A.M., S.H., R.P.S., E.R., A.E. and M.D.-F. processed and analyzed WGBS data. M.K., N.V.-D. and R.V.-B. performed validation experiments. M.K., G.Cas., S.E., V.P., D.Rico. and A.V functionally characterized dynamically methylated genes. D.Rich., L.C., A.D. and P.F. were in charge of data management. I.G.G. and H.G.S. coordinated sequencing efforts and performed primary data analysis. H.G.S., R.S., R.K. and E.C participated in the study design and data interpretation. J.I.M.-S. conceived the study, J.I.M.-S. led the experiments and wrote the paper with predominant assistance of M.K. and R.B.

COMPETING FINANCIAL INTEREST

The authors declare no competing financial interests.

REFERENCES

1. Ziller, M.J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477-81 (2013).
2. Bernstein, B.E., Meissner, A. & Lander, E.S. The mammalian epigenome. *Cell* **128**, 669-81 (2007).
3. Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14**, 204-20 (2013).
4. Bergman, Y. & Cedar, H. DNA methylation dynamics in health and disease. *Nat Struct Mol Biol* **20**, 274-81 (2013).
5. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21 (2002).
6. Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537-41 (2014).
7. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
8. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-22 (2009).
9. Gifford, C.A. *et al.* Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* **153**, 1149-63 (2013).
10. Kulis, M. *et al.* Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat Genet* **44**, 1236-42 (2012).
11. Berman, B.P. *et al.* Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet* **44**, 40-6 (2012).
12. Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360-9 (2013).
13. Hansen, K.D. *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* **43**, 768-75 (2011).
14. Li, Y. *et al.* The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol* **8**, e1000533 (2010).
15. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134-48 (2013).
16. Varley, K.E. *et al.* Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res* **23**, 555-67 (2013).
17. Lee, S.T. *et al.* 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**, 11339-51 (2012).
18. Lai, A.Y. *et al.* DNA methylation profiling in human B cells reveals immune regulatory elements and epigenetic plasticity at Alu elements during B cell activation. *Genome Res* (2013).
19. Shaknovich, R. *et al.* DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. *Blood* **118**, 3559-69 (2011).
20. Matthias, P. & Rolink, A.G. Transcriptional networks in developing and mature B cells. *Nat Rev Immunol* **5**, 497-508 (2005).
21. Kurosaki, T., Shinohara, H. & Baba, Y. B cell signaling and fate decision. *Annu Rev Immunol* **28**, 21-55 (2010).
22. Manz, R.A., Thiel, A. & Radbruch, A. Lifetime of plasma cells in the bone marrow. *Nature* **388**, 133-4 (1997).
23. Slifka, M.K., Antia, R., Whitmire, J.K. & Ahmed, R. Humoral immunity due to long-lived plasma cells. *Immunity* **8**, 363-72 (1998).
24. Hon, G.C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet* **45**, 1198-206 (2013).

25. Bibikova, M. *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98**, 288-95 (2011).
26. Ziller, M.J. *et al.* Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet* **7**, e1002389 (2011).
27. Arand, J. *et al.* In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet* **8**, e1002750 (2012).
28. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-9 (2011).
29. Lefebvre, C. *et al.* A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol Syst Biol* **6**, 377 (2010).
30. Tooze, R.M. A Replicative Self-Renewal Model for Long-Lived Plasma Cells: Questioning Irreversible Cell Cycle Exit. *Front Immunol* **4**, 460 (2013).
31. Vire, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871-4 (2006).
32. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* **10**, 295-304 (2009).
33. Brinkman, A.B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res* **22**, 1128-38 (2012).
34. Statham, A.L. *et al.* Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res* **22**, 1120-7 (2012).
35. Martin-Subero, J.I. *et al.* New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. *Blood* **113**, 2488-97 (2009).
36. Ohm, J.E. *et al.* A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* **39**, 237-42 (2007).
37. Schlesinger, Y. *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* **39**, 232-6 (2007).
38. Widschwendter, M. *et al.* Epigenetic stem cell signature in cancer. *Nat Genet* **39**, 157-8 (2007).
39. Busche, S. *et al.* Integration of high-resolution methylome and transcriptome analyses to dissect epigenomic changes in childhood acute lymphoblastic leukemia. *Cancer Res* **73**, 4323-36 (2013).
40. Agirre, X. *et al.* Whole-epigenome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers. *Genome Res* (2015).
41. Lopez-Otin, C., Blasco, M.A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194-217 (2013).
42. Horvath, S. *et al.* Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol* **13**, R97 (2012).
43. Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* **49**, 359-67 (2013).
44. Heyn, H. *et al.* Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A* **109**, 10522-7 (2012).
45. Maegawa, S. *et al.* Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res* **20**, 332-40 (2010).
46. Jaffe, A.E. & Irizarry, R.A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* **15**, R31 (2014).
47. Deaton, A.M. *et al.* Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res* **21**, 1074-86 (2011).

48. Hodges, E. *et al.* Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol Cell* **44**, 17-28 (2011).
49. Jeong, M. *et al.* Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat Genet* **46**, 17-23 (2014).
50. Tsankov, A.M. *et al.* Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**, 344-9 (2015).
51. Stadler, M.B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490-5 (2011).
52. Schmidl, C. *et al.* Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res* **19**, 1165-74 (2009).
53. Tagoh, H. *et al.* Dynamic reorganization of chromatin structure and selective DNA demethylation prior to stable enhancer complex formation during differentiation of primary hematopoietic cells in vitro. *Blood* **103**, 2950-5 (2004).
54. Wiench, M. *et al.* DNA methylation status predicts cell type-specific enhancer activity. *EMBO J* **30**, 3028-39 (2011).
55. Taberlay, P.C., Statham, A.L., Kelly, T.K., Clark, S.J. & Jones, P.A. Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. *Genome Res* **24**, 1421-32 (2014).
56. Schlesinger, F., Smith, A.D., Gingeras, T.R., Hannon, G.J. & Hodges, E. De novo DNA demethylation and noncoding transcription define active intergenic regulatory elements. *Genome Res* **23**, 1601-14 (2013).
57. Aran, D., Sabato, S. & Hellman, A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol* **14**, R21 (2013).
58. Aran, D., Toperoff, G., Rosenberg, M. & Hellman, A. Replication timing-related and gene body-specific methylation of active human genes. *Hum Mol Genet* **20**, 670-80 (2011).
59. Teschendorff, A.E. *et al.* Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* **20**, 440-6 (2010).
60. Cruickshanks, H.A. *et al.* Senescent cells harbour features of the cancer epigenome. *Nat Cell Biol* **15**, 1495-506 (2013).
61. Rakyan, V.K. *et al.* Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* **20**, 434-9 (2010).

FIGURE LEGENDS

Figure 1. Analysis of the DNA methylome of different B-cell subpopulations by WGBS and microarrays. (a) Description of the B-cell subpopulations and techniques used in this study. (b) Unsupervised principal component analysis (PCA) of WGBS data of two biological replicates per cell subpopulation. (c) Circular representation of DNA methylation levels for HPC, preB2C, naiBC, gcBC, memBC and bm-PC measured by WGBS. CpG methylation levels were averaged in 10-Mbp genomic windows and represented as histogram tracks. The heatmap indicates the DNA methylation change with respect to the sample in the next-innermost track. (d) Boxplot summarizing the distribution of DNA methylation levels per sample of the 16.1 million CpGs with methylation estimates in all 12 samples. (e) Global methylation status of samples measured by WGBS. Percentage of methylated (M, in red), partially methylated (PM, in yellow) and unmethylated (UM, in blue) CpGs. (f) Unsupervised PCA of microarray methylation data of all samples used in the study. (g) Median values of DNA methylation data measured by microarrays. HPC: hematopoietic progenitor cell. preB1C: pre-B-I cell. preB2C: pre-B-II cell. iBC: immature B cell. naiBC: naive B cell from peripheral blood. t-naiBC: naive B cell from tonsil. gcBC: germinal center B cell. t-PC: plasma cell from tonsil. memBC: memory B cell from peripheral blood. bm-PC: plasma cell from bone marrow. In panels d and e, R1 and R2 refer to the two biological replicates.

Figure 2. Non-CpG methylation detected during B-cell differentiation. (a) Browser representation of non-CpG methylation, which takes place mostly in the CpApC sequence context (only chromosome 1 is shown). Methylation in the reverse strand is marked in blue whereas that in the forward strand appears in red. (b) Number of non-CpG sites with non-zero methylation in different B-cell subpopulations detected by WGBS. Methylated cytosines in the CpApC context are marked in dark red and those in other contexts in pale red. (c) Scatter plot showing the numbers of methylated non-CpG sites (*using only the 3,437 non-CpGs with

methylation estimates in all 12 samples and methylated in at least one of them) and median CpG methylation levels. In this analysis, 99% of the non-CpGs methylated in one HPC sample were also methylated in the biological replicate. (d) Validation of non-CpG methylation by bisulfite pyrosequencing in two independent biological replicates of each subpopulation. For this analysis, we used a CpApC site (chr2: 85,933,406) shown to be methylated in HPCs by WGBS. (e) Heatmap representation of 26 methylated non-CpGs measured by microarrays (mean methylation = 34.7%). (f) Percentage of methylated non-CpGs in distinct sequence contexts detected by microarrays. (g) Representation of CpG and non-CpG dynamics upon B-cell commitment. CpG methylation is marked in blue and non-CpG methylation in green. Regions with CpG methylation loss (enhancer region, blue box) and non-CpG methylation loss (heterochromatin and polycomb-repressed region, green box) are not coupled.

Figure 3. Dynamic DNA methylation during B-cell differentiation. (a) Differentially methylated CpGs detected by WGBS considering the two replicates per cell subpopulation (see online methods for an explanation of the criteria). (b) Smoothed DNA methylation data generated by WGBS across the promoter region and gene body of *ARID3A* and *BLK*. The DNA methylation pattern of these genes is widely modulated in different B-cell subpopulations, especially in enhancer regions. (c) Heatmap representation of 20 major modules of dynamic CpGs, divided in 4 different patterns, detected by microarrays. The number of CpGs within each module is given in brackets. (d) Chromatin state characterization of differentially methylated CpGs identified by WGBS. (e) Chromatin state characterization of the 20 major modules detected by microarrays. In panels d and e, numbers indicate the percentage of sites located in enhancers, polycomb-repressed regions or heterochromatin. The blue to red color scale represents log2 of enrichment values, with respect to the background. Green and blue bars represent

percentages of differentially methylated sites that reside in early- or late-replicating regions, respectively. RepT: replication timing.

Figure 4. Association between DNA methylation and gene expression in different chromatin states. (a) Heatmap representing significant ($P < 0.01$) enrichments for transcription factor binding sites (TFBSs) in different methylation modules identified by arrays. Below the heatmap, log2 enrichment for enhancer, heterochromatin and polycomb-repressed regions in each differentially methylated group is represented. (b) Correlation between the expression levels of TFs and the mean methylation levels of their binding sites using samples with available expression and methylation data from the same donors. For this analysis, we analyzed methylation data from module 1 (demethylation upon B-cell commitment) and gene expression data of precursor cells. The white to brown color scale represents the odds ratio for TFBS enrichments. (c) Scatter plots showing the correlation of expression levels of *PAX5*, *EBF1* and *IRF4* with the mean methylation of their binding sites in each sample (the number of TFBS associated with CpGs belonging to module M1 is shown below the TF name). (d) Unsupervised cluster analysis of gene expression data using the 687 tags (439 genes) with the highest variability (SD > 2) across the B-cell differentiation process. (e) Differentially expressed genes (upper part) and differentially methylated genes (lower part) in each comparison of adjacent cell subpopulations. (f-g) Mean expression levels (f) and expression variability (g) during B-cell differentiation of genes containing dynamic CpGs targeting enhancers, polycomb-repressed regions and heterochromatin. (h) Heatmaps showing DNA methylation levels (left) and gene expression levels (right) of representative genes with dynamic methylation in enhancers, heterochromatin and polycomb-repressed regions.

Figure 5. DNA methylation changes during B-cell differentiation in the context of cancer and aging. (a) Heatmap of a subset of CpGs from module M9 that lose methylation in heterochromatin regions. (b) Heatmap of subset of CpGs from module M20 that gain methylation in polycomb-repressed regions. (c-d) Scatter plots representing the mean methylation levels of CpGs in heterochromatin from M9 (c) and CpGs in polycomb-repressed regions from M20 (d) in different B-cell subsets and four types of hematological neoplasms. (e-f) Mean methylation levels of CpGs in heterochromatin from M9 (e) and of CpGs in polycomb-repressed regions from M20 (f) in whole blood samples isolated from donors of different age. ALL: acute lymphoblastic leukemia. CLL: chronic lymphocytic leukemia. DLBCL: diffuse large B-cell lymphoma. MM: multiple myeloma.

Figure 6. DNA methylation changes in various B-cell neoplasms as compared to their normal counterparts. (a) Differential methylation analysis was performed in three models of lymphoid neoplasms that arise from three distinct maturation stages of B-cell development: ALL vs. precursor B-cells (i.e. preB1C and preB2C), the GCB subgroup of DLBCL vs. gcBC, and MM vs. plasma cells (i.e. t-PC and bm-PC). (b) Barplots showing the proportion of dynamically methylated CpGs in B-cell differentiation that are also differentially methylated in hematological neoplasias as compared to their normal counterparts. (c) Percentage of hypermethylated (upper panels) and hypomethylated (lower panels) CpGs located in enhancers (left) , heterochromatin (middle) and polycomb-repressed regions (right). Pale orange, gray and burgundy represent fraction of CpGs that are dynamically methylated during B-cell differentiation while dark shades of the same colors correspond to CpGs with stable methylation throughout B-cell maturation. (d) Heatmap representing differentially methylated CpGs in ALL as compared to precursor B cells in the context of normal B-cell differentiation. ALL: acute lymphoblastic leukemia. DLBCL: diffuse large B-cell lymphoma. MM: multiple myeloma. Bkgr: background of 450k microarray data.

ONLINE METHODS

Isolation of B-cell subpopulations

Precursor B cells were isolated from fetal bone marrow (22 week fetuses) using flow cytometry sorting. Early progenitors were isolated based on high levels of CD34 protein expression (CD34hi) and a lack of expression of the B-cell marker CD19. This population, designated as uncommitted hematopoietic progenitors (HPC), contains predominantly multipotent progenitors before lineage commitment and also common lymphoid progenitors and hematopoietic 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 (preB1Cs). Two immature B-cell populations expressing CD19 and lacking CD34 were isolated and differentiated based on surface IgM (sIgM) expression: pre-B-II cells (preB2Cs) that express sIgM−/CD19+ and immature B cells that express sIgM+/CD19+ (iBCs). DNA methylation and gene expression data from these 4 subpopulations have been previously published[17].

Peripheral blood B-cell subpopulations, i.e. naive B cell (naiBCs) and memory B cells (memBCs) were obtained from buffy coats from healthy adult donors of age ranging between 28 and 66 years. After Ficoll-Isopaque density centrifugation CD19+ B-cells were isolated by positive magnetic cell separation by using AutoMACS system (Milteny Biotec, Auburn, CA). CD19+ cells were labeled with anti-CD27, anti-IgD, anti-IgM, anti-IgG and anti-IgA for 15 min at room temperature in staining buffer (PBS with 0.5% BSA). naiBCs (CD19+/CD27-/IgD+) and memBCs (CD19+/CD27+/IgA+ or IgG+) were obtained by FACS sorting on FACSAriaII (BD Biosciences).

Plasma cells (t-PCs), germinal centre B cells (gcBCs) and naive B cells (t-naiBCs) were separated from tonsils of children undergoing tonsillectomies (ranging in age between 2 and 13), obtained from the Clinica Universidad de Navarra (Pamplona, Spain) or Children's Clinique La Sagesse (Rennes, France). Tonsils were minced extensively and after Ficoll-Isopaque density centrifugation, enrichment of B cells was performed with the AutoMACS system either by positive selection of CD19+ cells or by the B Cell Isolation Kit II (Milteny Biotec, Auburn, CA). t-PCs (CD20med/CD38high), gcBCs (CD20high/CD38med) and t-naiBCs (CD20+/CD23+) were separated by FACS sorting. In part, t-naiBCs were also selected using a slightly different marker combination (i.e. CD19+/CD27-/IgD+ or IgD+/CD38low/CD27-). gcBCs were also selected by this marker combination: IgD-/CD38hi/CD10+/CXCR4+. t-naiBCs and gcBCs isolated with different markers identified the same cell subpopulations, as evidenced by the fact that each subpopulation shows homogeneous DNA methylation and transcriptional profiles (**Supplementary Fig. 22**).

Plasma cells from bone marrow (bm-PCs) were selected from healthy donors ranging from 20 to 30 years. After density gradient centrifugation, we performed a selective depletion of CD3+, CD14+ and CD15+ cells by immunomagnetic selection (Miltenyi Biotec, Germany), followed by flow cytometry cell sorting of CD45+/CD138+/CD38+ cells using a FACSAriaII (BD Biosciences) device.

The purity of each of the isolated B-cell subpopulations exceeded 90% in all samples. DNA was extracted from purified samples by using a Qiagen kit (QIAamp DNA Mini Kit), following manufacturer's instructions, and was quantified using a Nanodrop ND-100 spectrophotometer. DNA samples for WGBS, 450k array and BPS experiments were derived from individual donors with the exception of those from bm-PCs, which were pooled from 4 different donors. Total

RNA was extracted with TRIzol (Invitrogen) following the manufacturer's recommendations. RNA quality was assessed with the Agilent 2100 Bioanalyzer.

Whole-genome bisulfite sequencing

We performed the WGBS on two independent sets of biological replicates from six B-cell differentiation stages. Briefly, genomic DNA (1–2 μg) was spiked with unmethylated λ DNA (5 ng of λ DNA per μg of genomic DNA) (Promega). The DNA was shared by sonication to 50–500 bp using a Covaris E220 and fragments of size 150–300 bp were selected using AMPure XP beads (Agencourt Bioscience Corp.). Genomic DNA libraries were constructed using the Illumina TruSeq Sample Preparation kit (Illumina Inc.) following the lllumina standard protocol: end repair was performed on the DNA fragments, an adenine was added to the 3' extremities of the fragments and Illumina TruSeq adapters were ligated at each extremity. After adaptor ligation, the DNA was treated with sodium bisulfite using the EpiTexy Bisulfite kit (Qiagen) following the manufacturer's instructions for formalin-fixed and paraffin-embedded (FFPE) tissue samples. Two rounds of bisulfite conversion were performed to assure a conversion rate of over 99%. Enrichment for adaptor-ligated DNA was carried out through 7 PCR cycles using the PfuTurboCx Hotstart DNA polymerase (Stratagene). Library quality was monitored using the Agilent 2100 BioAnalyzer (Agilent), and the concentration of viable sequencing fragments (molecules carrying adaptors at both extremities) estimated using quantitative PCR with the library quantification kit from KAPA Biosystem. Paired-end DNA sequencing (2x100bp) was then performed using the Illumina Hi-Seq 2000. Amounts of sequence reads and the proportion of aligned reads are shown in **Supplementary Table 2**.

Read mapping and estimation of cytosine methylation levels

Read mapping was carried out using the GEM aligner (v1.242)[62] against a composite reference containing two copies of the human GRCh37 reference and two copies of the NCBI viral genome database (v35). For both the human and viral references, one copy had all C bases replaced by T and the other had all G bases replaced by A. The names of the contigs in the combined reference FASTA file were modified by adding #C2T or #G2A to the end of the contig names depending on the conversion performed. Before mapping was performed the original sequence of each read was stored. The first read of each pair then had all C bases replaced by T, and the second read had all G bases replaced by A. Read mapping with GEM was performed allowing up to 4 mismatches per read from the reference. After read mapping the original sequence of each read was restored.

Estimation of cytosine levels was carried out on read pairs where both members of the read mapped to the same contig with consistent orientation, and there was no other such configuration at the same or less edit distance from the reference. After mapping, we restored the original read data in preparation for the inference of genotype and methylation status. We estimated genotype and DNA methylation status simultaneously using software developed at the Centro Nacional de Análisis Genómico, taking into account the observed bases, base quality scores and the strand origin of each read pair. For each genome position, we produced estimates of the most likely genotype and the methylation proportion (for genotypes containing a C on either strand). A phred scaled likelihood ratio for the confidence in the genotype call was estimated for the called genotype at each position. For each sample, CpG sites were selected where both bases were called as homozygous CC followed by GG with a

Phred score of at least 20, corresponding to an estimated genotype error level of <= 1%. Sites with >500x coverage depth were excluded to avoid centromeric/telomeric repetitive regions. A common set of called CpG sites for all analyzed samples was generated, and all subsequent analyses used this common set.

Microarray-based DNA methylation analysis with 450k arrays

We used the EZ DNA Methylation Kit (Zymo Research) for bisulfite conversion of 500 ng genomic DNA. Bisulfite-converted DNA was hybridized onto the HumanMethylation 450K BeadChip kit (Illumina) which covers 99% of RefSeq genes and 96% of CpG islands. The Infinium methylation assay was carried out as described previously[25,63]. Data from the 450k Human Methylation Array were analyzed in R using the minfi package[64] (version: 1.6.0), available through the Bioconductor open source software. To exclude technical and biological biases that might produce false results in further analyses, we developed and optimized an analysis pipeline with several filters (i.e. CpGs with low detection p values, sex-specific and individual-specific methylation or overlapping with SNPs). Taking into account the different performance of Infinium I and Infinium II assays we used the subset-quantile within array normalization (SWAN)[65] that corrects for the technical differences between the Infinium I and II assay designs and produces a smoother overall beta value distribution.

Detection of non-CpG methylation and differential methylation analysis

Cytosines in non-CpG context were identified as two adjacent nucleotides where the genotype of the first nucleotide was called with high confidence as homozygous C and the second was called with high confidence as a genotype not containing a G. Non-CpG cytosines were called methylated if they had at least two non-converted reads, at least six reads informative for methylation status, and a methylation probability greater than twice its standard deviation. The significance of the change in methylation levels between samples was assessed using the numbers of converted and non-converted reads in both samples with a chi square test, or Fisher's exact test when the chi square approximation was not appropriate.

The difference in methylation levels between different stages of B-cell differentiation was calculated using 16.1 million CpGs with methylation estimates in all 12 samples analyzed by WGBS. The normal approximation to the binomial was used to test for significant differences of individual CpGs between samples. As we sequenced two biological replicates per cell subpopulation, we defined consistent DNA methylation changes between two differentiation stages if in one set of samples the methylation difference was above 0.25 and in the second set was at least a 0.1 difference in the same direction (hyper or hypomethylation). CpGs with dynamic methylation were defined as with differential methylation in comparisons of adjacent stages and in the comparison between HPCs and bm-PCs.

To calculate the overlap between dynamically methylated CpGs in B cells and by Ziller et al[1], we downloaded all the DMRs with dynamic CpGs from that study (GEO number GSE46644) and determined how many of the 4.93 million dynamic CpGs in B cells are located within DMRs identified by Ziller et al.

We also defined CpGs that show variable methylation levels throughout B-cell development using 450k microarray data for HPC, preB1C, preB2C, iBC, naiBCs, t-naiBCs, gcBC, t-PC and bm-PCs. From gcBCs on, B-cell differentiation is branched into memBCs or PCs and therefore memBCs were not included in this linear analysis. We performed pair-wise comparisons between all these subsets of B-cell differentiation. We defined as “dynamic” those CpGs that

presented a mean DNA methylation difference above 0.25 in at least one comparison (FDR<0.1, Wilcoxon test). Furthermore, we detected CpGs whose methylation shows a similar modulation pattern during the entire B-cell differentiation process and we group them into distinct modules. To allow for the identification of gradual changes (but consistent in different replicates) throughout the differentiation program, we applied a mean methylation difference between adjacent subpopulations of 0.1. The modules were grouped into four groups depending on their methylation tendency during differentiation (decrease, increase, decrease-increase and increase-decrease). Differentially methylated sites between specific B-cell tumor entities and their normal cellular counterpart were identified by a mean DNA methylation difference above 0.25 and an FDR < 0.05.

Bisulfite pyrosequencing studies

Validation of DNA methylation levels generated by WGBS and 450k microarrays was validated by bisulfite pyrosequencing (BPS). Briefly, 500 ng of genomic DNA was bisulfite converted using the EpiTect 96 Bisulfite kit or the EpiTect Plus Bisulfite Conversion Kit (Qiagen) according to manufacturer's instructions. PCR amplification of the bisulfite treated DNA was done using the specific primers for each of the selected CpGs and non-CpGs (**Supplementary Table 3**). These primers were determined with the PyroMark Assay Design software (Qiagen). BPS and DNA methylation data analysis were performed with the pyrosequencer PyroMark Q96 ID and PyroMark CpG software (Qiagen).

Genomic and functional annotation of CpG sites

Both WGBS and 450k microarray data were annotated using the UCSC Genome Browser database (hg19). For the location relative to a gene, we used these categories: TSS 1500 (from 201 to 1,500 bp upstream of the transcriptional start site (TSS)), TSS 200 (from 1 to 200 bp upstream of the TSS), 5' UTR, first exon, exon (all exons excluding exon 1), intron, 3' UTR and intergenic regions. Owing to the presence of alternative transcription start sites and regions containing more than one gene, some of the CpGs were assigned multiple annotations. For the location relative to a CpG island (CGI), we used these groups: within CGI, in CGI shore (0–2 kb from the CGI edge), in CGI shelf (>2 kb to 4 kb from the CGI edge) and outside CGI.

We also annotated all CpG probes using a recent categorization of chromatin and transcriptional states from the lymphoblastoid B cell line GM12878[28] (ChromHMM track of the UCSC Genome Browser), which has a DNA methylome similar to memory B cells and plasma cells (**Supplementary Fig. 14**). Regions with chromatin states 1–3 (*active, weak and poised promoter*) were considered as "Promoter regions", states 4–7 (*strong* and *weak enhancer*) as "Enhancer regions", state 8 as "Insulator", states 9 "Transcriptional transition*",* state 10 "Transcriptional elongation", state 11 as "Weak transcription", state 12 as "Polycomb-repressed regions" and state 13 as "Heterochromatin (nuclear lamina)".

Replication timing in GM12878 data was obtained from the UW Repli-seq track of the UCSC Genome Browser. Replication timing values for all sites from the background was divided into 3 bins: early-, mid- and late-replicating regions. Only early- and late-replicating regions were used for the analysis.

The annotation of repeat elements was done based on the RepeatMasker Annotation, available at the UCSC Genome Browser.

B-cell network analysis

We used the B-cell specific functional interaction network of Lefebvre et al.,[29] containing 5,748 nodes (genes) and 64,600 unique edges (interactions) based on Entrez gene identifiers. We selected the 5,668 dynamic enhancer genes and mapped them to Entrez gene identifiers resulting in 5,658 unique Entrez gene identifiers. 2,154 of these Entrez genes are contained in the B-cell network and 1,993 of them are directly connected in the network by 11,741 edges. This subnetwork of 1,993 nodes and 11,741 edges was investigated further. We identified 9 communities in the subnetwork by using Gephi[66] and Louvain's method[67].

Chromatin immunoprecipitation coupled with bisulfite sequencing

B cells were cross-linked with 1% formaldehyde during 8 min (RT) before FACS separation of memBCs. Chromatin preparation and ChIP were performed according to the BLUEPRINT Histone ChIP protocol with a anti-H3K27me3 antibody (C15410195, Diagenode). For whole-genome bisulfite library construction and sequencing the immunoprecipitated DNA (50 ng) was sheared on a Covaris™ E220 in order to reach the fragment size of 50–500 bp and size selected 150 to 500 bp fragments using AMPure XP beads (Agencourt Bioscience Corp.). 500 ng of unmethylated λ DNA (Promega) was treated in parallel on Covaris™ E220 and also size selected by AMPure XP beads to reach the same fragment sizes as the DNA sample. The unmethylated λ DNA was spiked into fragmented and size selected immunoprecipitated DNA (5 ng of λ DNA per 1 µg of DNA) and TruSeq Sample Preparation kit (Illumina Inc.) was used to prepare the Illumina library in order to add platform specific adaptors. After adaptor ligation 450 ng of fragmented and size selected unmethylated λ DNA was added to the library. Two rounds of bisulfite conversion was performed to reach > 99% conversion following the manufacturer's instructions for formalin-fixed and paraffin-embedded (FFPE) tissue samples (EpiTect Bisulfite kit; Qiagen). Adaptor-ligated DNA was enriched through ten cycles of PCR using the KAPA HiFi Uracil+ polymerase (Kapa Biosystems). The library was run in a fraction of a lane of HiSeq2000 flow cell (to reach 35 million paired end reads) with read length of 2x100 bp, according to standard Illumina operation procedures. Primary data analysis was carried out with the standard Illumina pipeline.

The sequence reads were passed through the same read mapping and genotype/methylation calling pipeline as the conventional WGBS samples. In addition, the aligned reads were analyzed with the NucHunter package[68], to provide predictions of the positions of H3K27me3-containing nucleosomes. The average methylation of cytosines at all distances from 0-500 bp of each peak was calculated, taking into account the strand of the cytosine. The same analysis was performed on WGBS data from the same cell type (memBC), to allow comparison of the enriched (ChIP-BS) and non-enriched (WGBS) results. This analysis was repeated using only predicted nucleosome peaks that fell within predicted polycomb repressed regions from the Broad ChromHMM analysis of the ENCODE cell line GM12878.

Gene ontology analysis

The GOstat package[69] available through Bioconductor was used to determine the enrichment of individual ontology terms in the different methylation modules as compared to all the genes analyzed in the 450k array. In **Supplementary Data 4**, the top 20 significant terms for each module ($P < 0.001$) are shown.

Analysis of transcription factor binding sites

Transcription factor binding site (TFBS) information was obtained by ChIP-seq data in the ENCODE project and available at the UCSC Genome Browser. A total of 79 TFBS were used for the analysis. The relative enrichment of each TFBS was calculated as compared to the background. A fisher exact test was used and an Odds Ratio and a P value were assigned to each comparison.

Gene expression analyses

RNA samples of HPCs, preB1Cs, preB2Cs, naiBCs, t-naiBCs, gcBCs, memBCs, t-PCs and bm-PCs were hybridized to Affymetrix Human Genome U219 Array Plates according to Affymetrix standard protocols. The analysis of scanned images for each probe set of the array was obtained with GeneChip Operating Software (GCOS, Affymetrix). Raw CEL files were processed and normalized with the robust multichip average (RMA) algorithm using R statistical software in conjunction with the Affy library[70] available through Bioconductor. The GeneChip Human Gene 1.0 ST Array data of progenitor B cells were downloaded through Gene Expression Omnibus under accession number GSE45461 and normalized using RMA (these data were only used for **Fig. 4b,c**).

To evaluate the gene expression variability among B-cell subpopulations, we calculated the inter-cell subtype standard deviation per each of the Affymetrix tags. A global measure of the variability for a particular set of genes (within modules and chromatin states) was then calculated as the average of all the standard deviations. Differential expression between t-naiBCs and naiBCs was calculated using Limma library available through Bioconductor, using the criteria of FC > 1 (log2) between two groups and adjusted $P < 0.05$.

Reverse transcription and quantitative PCR for DNMTs expression was done as follows: cDNAs from 100 ng RNA samples were synthesized in 20 µl reaction mix using oligo dT primers and SuperScript III enzyme according to the manufacturer recommendations (Invitrogen). Primers for DNMTs were taken from Fang et al.[71] whereas those for the housekeeping EEF2 gene are provided in **Supplementary Table 3**. PCR amplification was carried out with 1 µl of the 1:2 diluted reverse transcription sample with Power SYBR green PCR master mix according to manufacturer recommendations (Applied Biosystems). PCR reactions were run in triplicates on a StepOne System (Applied Biosystems).

REFERENCES ONLINE METHODS

62. Marco-Sola, S., Sammeth, M., Guigo, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat Methods* **9**, 1185-8 (2012).
63. Bibikova, M. *et al.* Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics* **1**, 177-200 (2009).
64. Aryee, M.J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-9 (2014).
65. Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol* **13**, R44 (2012).

66. Bastian, M., Heymann, S. & Jacomy, M. Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media* (2009).
67. Blondel, V.D., Guillaume, J.L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* **10**, P10008 (2008).
68. Mammana, A., Vingron, M. & Chung, H.R. Inferring nucleosome positions with their histone mark annotation from ChIP data. *Bioinformatics* **29**, 2547-54 (2013).
69. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257-8 (2007).
70. Gautier, L., Cope, L., Bolstad, B.M. & Irizarry, R.A. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307-15 (2004).
71. Fang, J. *et al.* Epigenetic changes mediated by microRNA miR29 activate cyclooxygenase 2 and lambda-1 interferon production during viral infection. *J Virol* **86**, 1010-20 (2012).

FIGURE-1

a

	Progenitor cell	PreB-I cell	PreB-II cell	Immature B cell	Naive B cells		Germinal center B cell	Memory B cell	Plasma cells	
Surface markers	$CD34^{++}$ $CD19^{-}$	$CD34^{+}$ $CD19^{+}$	$CD34^{-}$ $CD19^{+}$ $sIgM^{-}$	$CD34^{-}$ $CD19^{+}$ $sIgM^{+}$	$CD19^{+}$ IgD^{+} $CD27^{-}$	$CD19^{+}$ IgD^{+} $CD27^{-}$	$CD19^{+}$ $CD20^{++}$ $CD38^{+}$	$CD19^{+}$ IgA/IgG^{+} $CD27^{+}$	$CD19^{+}$ $CD20^{+}$ $CD38^{++}$	$CD138^{+}$
Source	Bone marrow	Bone marrow	Bone marrow	Bone marrow	Peripheral blood	Tonsil	Tonsil	Peripheral blood	Tonsil	Bone marrow
Samples WGBS (measured CpGs)	2 ($23.2x10^6$) ($23.4x10^6$)		2 ($22.9x10^6$) ($23.4x10^6$)		2 ($24.7x10^6$) ($23.3x10^6$)		2 ($21.0x10^6$) ($21.7x10^6$)	2 ($25.0x10^6$) ($21.2x10^6$)		2 ($21.1x10^6$) ($21.4x10^6$)
Samples 450k arrays	6	6	6	4	5	5	9	5	8	3

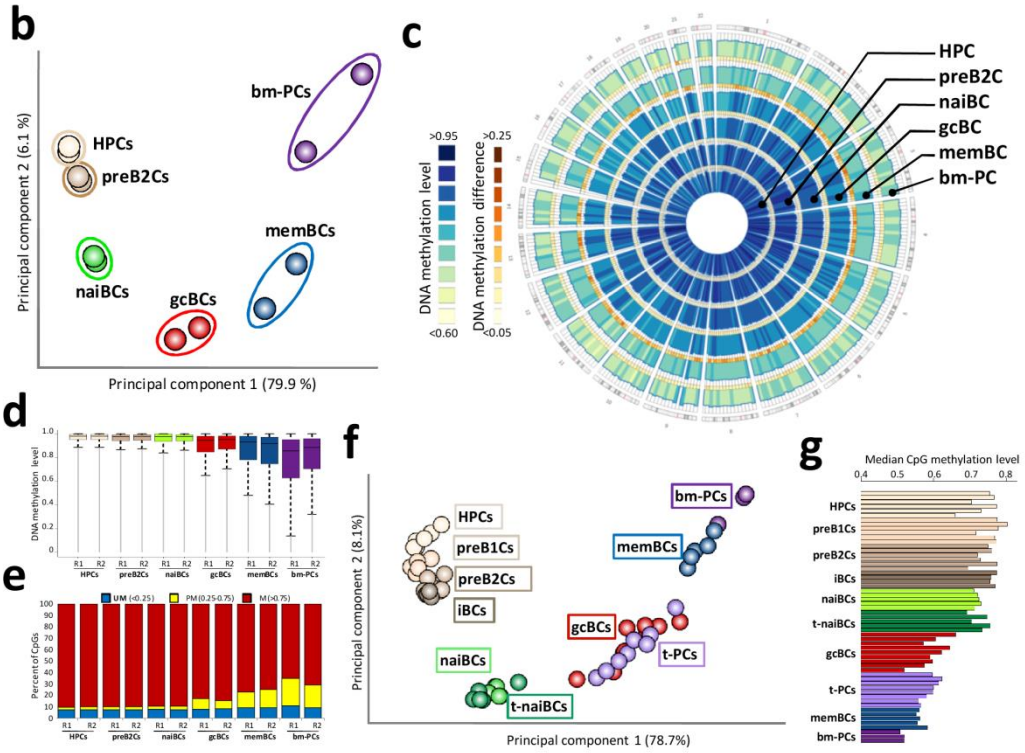


FIGURE-2

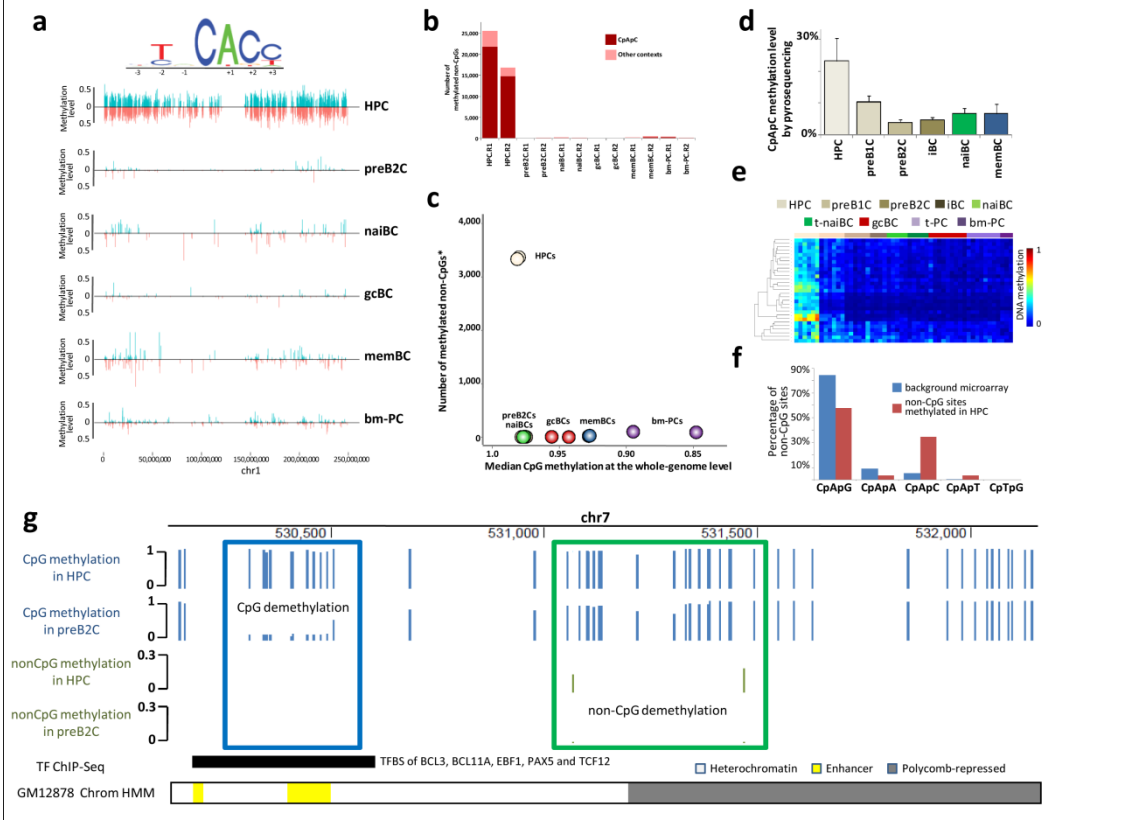

a
HPC
preB2C
naiBC
gcBC
memBC
bm-PC
Methylation level
chr1
b
CpApC
Other contexts
c
HPCs
preB2Cs
naiBCs
gcBCs
memBCs
bm-PCs
Number of methylated non-CpGs*
Median CpG methylation at the whole-genome level
d
CpApC methylation level by pyrosequencing
30%
0%
HPC
preB1C
preB2C
iBC
naiBC
memBC
e
HPC preB1C preB2C iBC naiBC
t-naiBC gcBC t-PC bm-PC
DNA methylation
f
Percentage of non-CpG sites
background microarray
non-CpG sites methylated in HPC
CpApG
CpApA
CpApC
CpApT
CpTpG
g
chr7
530,500
531,000
531,500
532,000
CpG methylation in HPC
CpG methylation in preB2C
nonCpG methylation in HPC
nonCpG methylation in preB2C
CpG demethylation
non-CpG demethylation
TF ChIP-Seq
TFBS of BCL3, BCL11A, EBF1, PAX5 and TCF12
GM12878 Chrom HMM
Heterochromatin
Enhancer
Polycomb-repressed

FIGURE-3

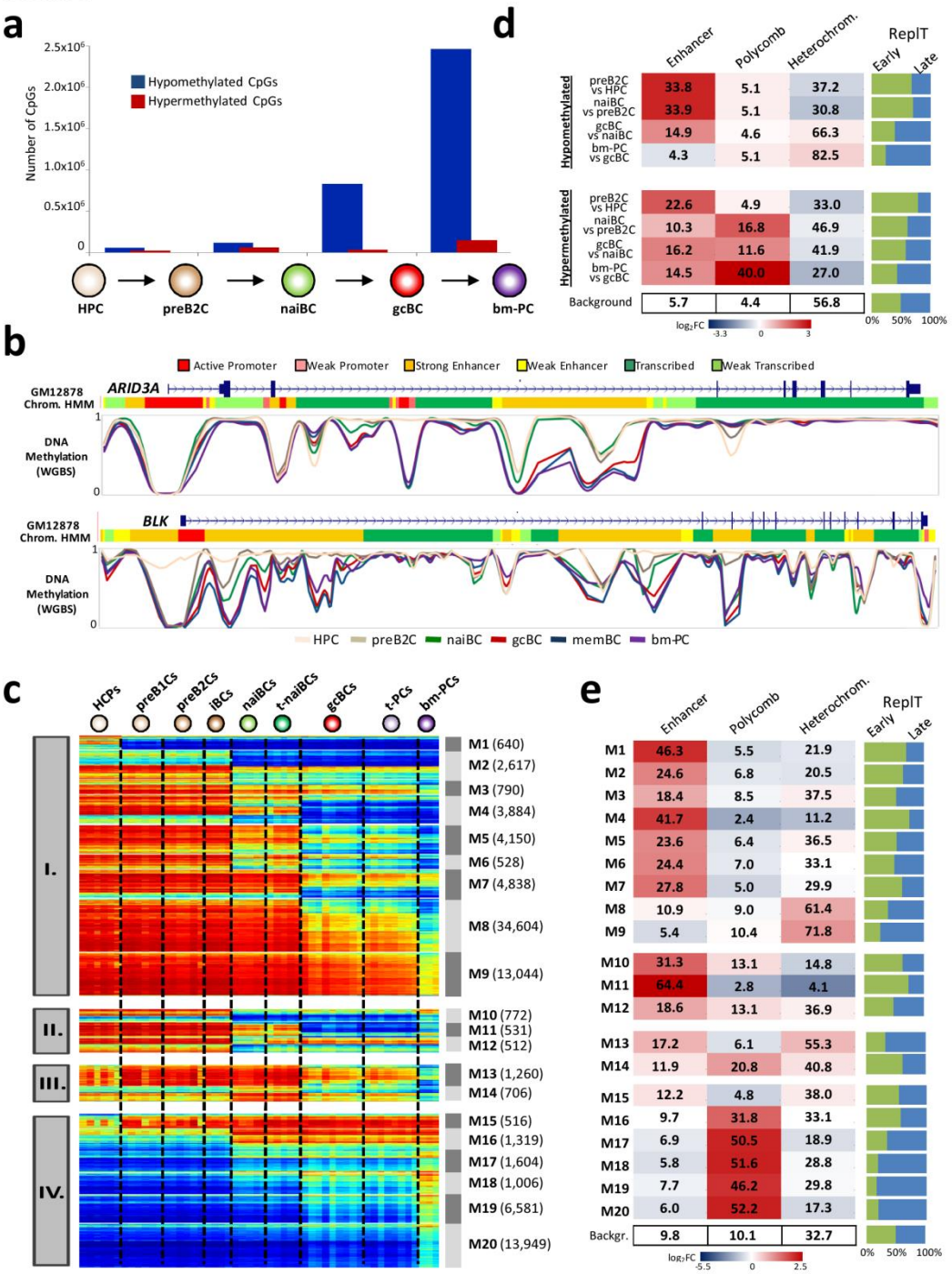

a
Number of CpGs
Hypomethylated CpGs
Hypermethylated CpGs
HPC
preB2C
naiBC
gcBC
bm-PC
d
Enhancer
Polycomb
Heterochrom.
ReplT
Early
Late
Hypomethylated
Hypermethylated
Background
log2FC
b
Active Promoter
Weak Promoter
Strong Enhancer
Weak Enhancer
Transcribed
Weak Transcribed
GM12878 Chrom. HMM
ARID3A
DNA Methylation (WGBS)
BLK
HPC
preB2C
naiBC
gcBC
memBC
bm-PC
c
HCPs
preB1Cs
preB2Cs
iBCs
naiBCs
t-naiBCs
gcBCs
t-PCs
bm-PCs
I.
II.
III.
IV.
M1 (640)
M2 (2,617)
M3 (790)
M4 (3,884)
M5 (4,150)
M6 (528)
M7 (4,838)
M8 (34,604)
M9 (13,044)
M10 (772)
M11 (531)
M12 (512)
M13 (1,260)
M14 (706)
M15 (516)
M16 (1,319)
M17 (1,604)
M18 (1,006)
M19 (6,581)
M20 (13,949)
e
Backgr.

FIGURE-4

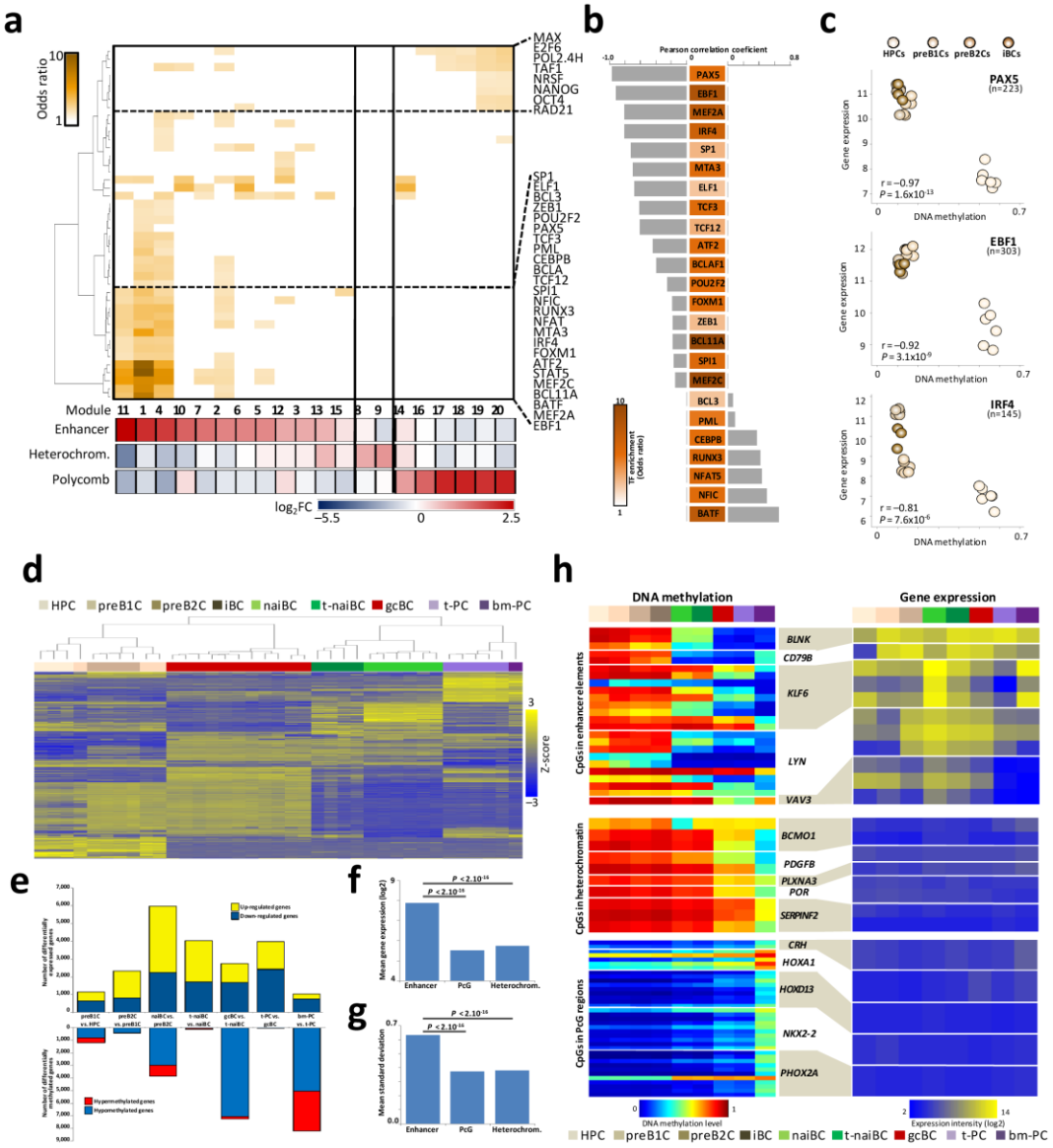

a
Odds ratio
10
1
MAX
E2F6
POL2.4H
TAF1
NRSF
NANOG
OCT4
RAD21
SP1
ELF1
BCL3
ZEB1
POU2F2
PAX5
TCF3
PML
CEBPB
BCLA
TCF12
SPI1
NFIC
RUNX3
NFAT
MTA3
IRF4
FOXM1
ATF2
STAT5
MEF2C
BCL11A
BATF
MEF2A
EBF1
Module 11 1 4 10 7 2 6 5 12 3 13 15 8 9 14 16 17 18 19 20
Enhancer
Heterochrom.
Polycomb
log2FC −5.5 0 2.5
b
Pearson correlation coeficient
-1.0 0 0 0.8
PAX5
EBF1
MEF2A
IRF4
SP1
MTA3
ELF1
TCF3
TCF12
ATF2
BCLAF1
POU2F2
FOXM1
ZEB1
BCL11A
SPI1
MEF2C
BCL3
PML
CEBPB
RUNX3
NFAT5
NFIC
BATF
TF enrichment (Odds ratio)
10
1
c
HPCs preB1Cs preB2Cs iBCs
PAX5 (n=223)
r = −0.97
P = 1.6x10-13
EBF1 (n=303)
r = −0.92
P = 3.1x10-9
IRF4 (n=145)
r = −0.81
P = 7.6x10-6
Gene expression
DNA methylation
0 0.7
d
HPC preB1C preB2C iBC naiBC t-naiBC gcBC t-PC bm-PC
Z-score
3
−3
e
Up-regulated genes
Down-regulated genes
Number of differentially expressed genes
Hypermethylated genes
Hypomethylated genes
Number of differentially methylated genes
f
P < 2.10-16
P < 2.10-16
Mean gene expression (log2)
Enhancer PcG Heterochrom.
g
P < 2.10-16
P < 2.10-16
Mean standard deviation
Enhancer PcG Heterochrom.
h
DNA methylation
Gene expression
CpGs in enhancer elements
BLNK
CD79B
KLF6
LYN
VAV3
CpGs in heterochromatin
BCMO1
PDGFB
PLXNA3
POR
SERPINF2
CpGs in PcG regions
CRH
HOXA1
HOXD13
NKX2-2
PHOX2A
DNA methylation level
Expression intensity (log2)
HPC preB1C preB2C iBC naiBC t-naiBC gcBC t-PC bm-PC

FIGURE-5

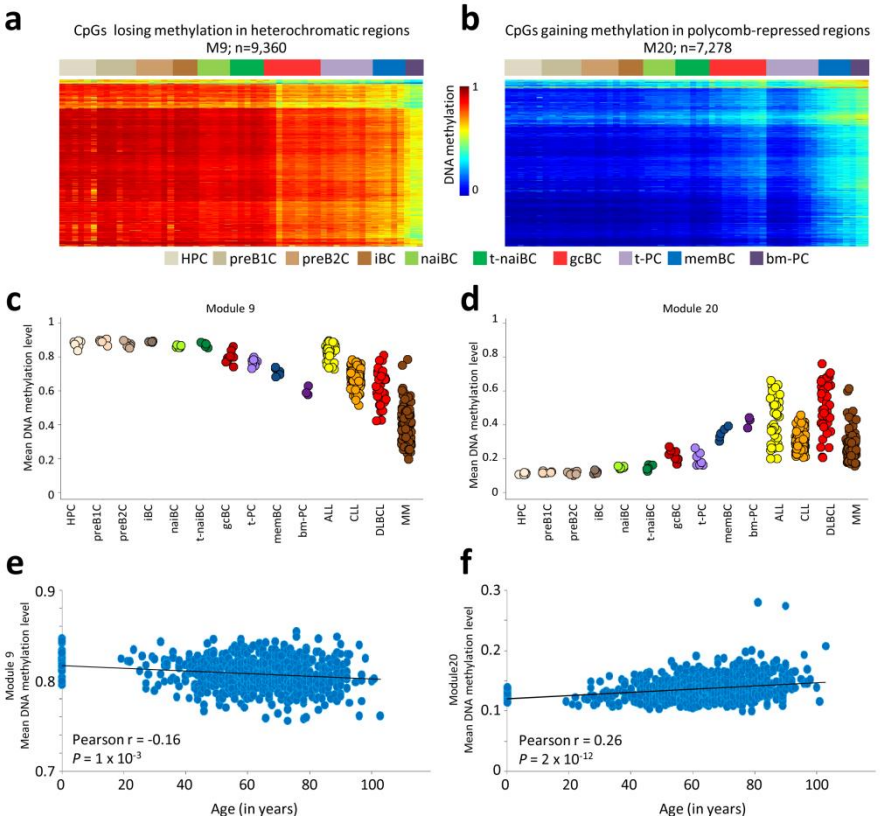
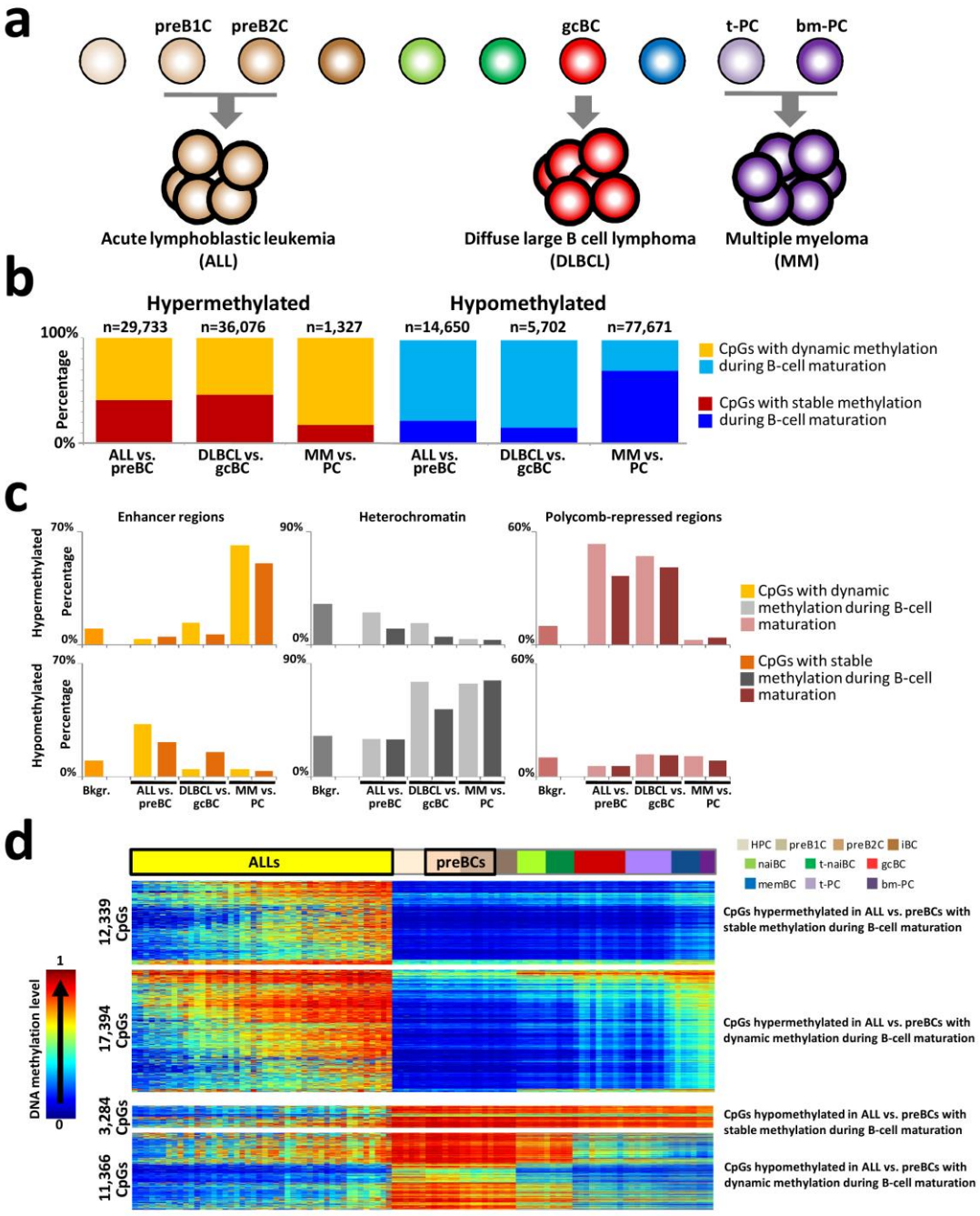

a
CpGs losing methylation in heterochromatic regions
M9; n=9,360
b
CpGs gaining methylation in polycomb-repressed regions
M20; n=7,278
DNA methylation
1
0
HPC preB1C preB2C iBC naiBC t-naiBC gcBC t-PC memBC bm-PC
c
Module 9
Mean DNA methylation level
HPC preB1C preB2C iBC naiBC t-naiBC gcBC t-PC memBC bm-PC ALL CLL DLBCL MM
d
Module 20
Mean DNA methylation level
HPC preB1C preB2C iBC naiBC t-naiBC gcBC t-PC memBC bm-PC ALL CLL DLBCL MM
e
Module 9
Mean DNA methylation level
Pearson r = -0.16
P = 1 x 10-3
Age (in years)
f
Module20
Mean DNA methylation level
Pearson r = 0.26
P = 2 x 10-12
Age (in years)

FIGURE-6


a
preB1C
preB2C
gcBC
t-PC
bm-PC
Acute lymphoblastic leukemia (ALL)
Diffuse large B cell lymphoma (DLBCL)
Multiple myeloma (MM)
b
Hypermethylated
Hypomethylated
n=29,733
n=36,076
n=1,327
n=14,650
n=5,702
n=77,671
100%
0%
Percentage
ALL vs. preBC
DLBCL vs. gcBC
MM vs. PC
CpGs with dynamic methylation during B-cell maturation
CpGs with stable methylation during B-cell maturation
c
Enhancer regions
Heterochromatin
Polycomb-repressed regions
Hypermethylated
Hypomethylated
Percentage
70%
90%
60%
0%
Bkgr.
ALL vs. preBC
DLBCL vs. gcBC
MM vs. PC
CpGs with dynamic methylation during B-cell maturation
CpGs with stable methylation during B-cell maturation
d
ALLs
preBCs
HPC
preB1C
preB2C
iBC
naiBC
t-naiBC
gcBC
memBC
t-PC
bm-PC
DNA methylation level
1
0
12,339 CpGs
17,394 CpGs
3,284 CpGs
11,366 CpGs
CpGs hypermethylated in ALL vs. preBCs with stable methylation during B-cell maturation
CpGs hypermethylated in ALL vs. preBCs with dynamic methylation during B-cell maturation
CpGs hypomethylated in ALL vs. preBCs with stable methylation during B-cell maturation
CpGs hypomethylated in ALL vs. preBCs with dynamic methylation during B-cell maturation