

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

Marta Kulis¹, Angelika Merkel², Simon Heath², Ana C Queirós¹, Ronald P Schuyler², Giancarlo Castellano¹, Renée Beekman¹, Emanuele Raineri², Anna Esteve², Guillem Clot¹, Núria Verdaguer-Dot¹, Martí Duran-Ferrer^{1,2}, Nuria Russiñol¹, Roser Vilarrasa-Blasi¹, Simone Ecker³, Vera Pancaldi³, Daniel Rico³, Lidia Agueda², Julie Blanc², David Richardson⁴, Laura Clarke⁴, Avik Datta⁴, Marien Pascual⁵, Xabier Agirre⁵, Felipe Prosper^{5,6}, Diego Alignani⁷, Bruno Paiva^{6,7}, Gersende Caron⁸, Thierry Fest⁸, Marcus O Muench^{9,10}, Marina E Fomin^{9,10}, Seung-Tae Lee¹¹, Joseph L Wiemels¹², Alfonso Valencia³, Marta Gut², Paul Flück⁴, Hendrik G Stunnenberg¹³, Reiner Siebert¹⁴, Ralf Küppers¹⁵, Ivo G Gut², Elías Campo¹ & José I Martín-Subero¹

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 CpG sites. 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 at Polycomb-repressed areas, and genes with apparent functional impact in B cells were not affected. This signature, which has previously been linked to aging and cancer, was particularly widespread in mature cells with an extended lifespan. Comparing B cell neoplasms with their normal counterparts, we determined that they frequently acquire methylation changes in regions already undergoing dynamic methylation 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 on a whole-genome scale^{1,6–16}, the DNA methylome of a single human cell type throughout its complete differentiation process has not been defined thus 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 the 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 bloodstream. Resting naive B cells transit through lymph nodes, and they are eventually 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

¹Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Department of Anatomic Pathology, Pharmacology and Microbiology, University of Barcelona, Barcelona, Spain. ²Centro Nacional de Análisis Genómico (CNAG), Parc Científic de Barcelona, Barcelona, Spain. ³Structural Biology and Biocomputing Program, Spanish National Cancer Research Centre (CNIO), Spanish National Bioinformatics Institute, Madrid, Spain. ⁴European Molecular Biology Laboratory–European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, UK. ⁵Área de Oncología, Centro de Investigación Médica Aplicada (CIMA), Universidad de Navarra, Pamplona, Spain. ⁶Servicio de Hematología, Clínica Universidad de Navarra, Pamplona, Spain. ⁷Flow Cytometry Core, Centro de Investigación Médica Aplicada (CIMA), Universidad de Navarra, Pamplona, Spain. ⁸Université de Rennes 1, INSERM U917, Hematology Laboratory, University Hospital of Rennes, Rennes, France. ⁹Blood Systems Research Institute, San Francisco, California, USA. ¹⁰Department of Laboratory Medicine, University of California, San Francisco, San Francisco, California, USA. ¹¹Department of Laboratory Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea. ¹²Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, USA. ¹³Molecular Biology, Nijmegen Centre for Molecular Life Sciences (NCMLS), Faculties of Science and Medicine, Radboud University, Nijmegen, the Netherlands. ¹⁴Institute of Human Genetics, Christian Albrechts University, Kiel, Germany. ¹⁵Institute of Cell Biology (Cancer Research), Medical School, University of Duisburg-Essen, Essen, Germany. Correspondence should be addressed to J.I.M.-S. (imartins@clinic.ub.es).

Received 20 October 2014; accepted 3 April 2015; published online 8 June 2015; doi:10.1038/ng.3291

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 cell types with different functional features, proliferation abilities, microenvironmental influences and lifespans, 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 sorted human cell populations—uncommitted hematopoietic progenitor cells (HPCs) and five B cell-lineage subpopulations, including pre-BII cells, naive B cells from peripheral blood, germinal center B cells, memory B cells from peripheral blood and plasma cells from bone marrow—by whole-genome bisulfite sequencing (WGBS) (Fig. 1a and Supplementary Table 1). We sequenced two biological replicates of each subpopulation for a total of 2,217 billion base pairs, of which 85–95% could be mapped (mean depth of 54-fold per sample) (Supplementary Table 2). On average, we measured the methylation levels of 22.7 million CpG sites per sample (range of 21 to 25 million). In unsupervised principal-component analysis (PCA) of CpG methylation levels, B cell subpopulations were segregated according to their developmental stage (Fig. 1b). Globally, B cell differentiation was accompanied by gradual, widespread demethylation of the genome, with this demethylation more pronounced at late differentiation stages, such as in memory B cells and bone marrow plasma cells (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 is in contrast to other WGBS studies using whole tissues, in which the proportion of partially methylated regions is usually high²⁴, and highlights the importance of using purified cell subpopulations for DNA methylation studies.

The results obtained by WGBS were complemented with the analysis of three to nine replicates of ten different B cell subpopulations by high-density DNA methylation microarray²⁵ (Fig. 1a). These subpopulations included those analyzed by WGBS as well as pre-BI cells, immature B cells, tonsillar naive B cells and tonsillar plasma cells (Supplementary Table 1). The biological replicates of each subpopulation analyzed by WGBS or microarray showed high reproducibility (correlation coefficient > 0.95; Supplementary Fig. 1), and we further validated both high-throughput techniques by bisulfite pyrosequencing (BPS; Supplementary Fig. 2). In line with the WGBS data, PCA of the microarray data separated B cell subpopulations mostly according to their developmental stage (Fig. 1f), and we observed 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 germinal center B cells start a massive reconfiguration of the DNA methylome^{18,19}, which continues in cell subpopulations with extended lifespan, such as memory B cells and bone marrow plasma cells.

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, when 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 (HPCs) and occurred mainly in a CpApC sequence context (Fig. 2a–c). We detected substantial levels of non-CpG methylation in HPCs by WGBS at 25,763 sites in replicate 1 and at 16,838 sites in replicate 2, with 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, a finding that was also confirmed by BPS 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 was depleted in laminin-associated domains (Supplementary Fig. 5). Although the methylation microarray only measured methylation at 3,091 non-CpG sites, we could confirm the WGBS results (Fig. 2e,f). Similarly to reports in ESCs²⁶, methylated non-CpGs were flanked by methylated CpGs. However, demethylation of non-CpGs and CpGs was not simultaneous. We observed dramatic demethylation of non-CpGs upon B cell commitment in pre-BII cells, but 97% of these demethylated non-CpGs remained flanked by methylated CpGs (Supplementary Fig. 6). These data were confirmed by BPS (Supplementary Fig. 4) and indicate that non-CpG methylation is passively erased in the transition from HPCs to pre-BII cells, 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 *de novo* methyltransferases (DNMTs) in HPCs and pre-B cells (Supplementary Fig. 7).

Identification of dynamic DNA methylation patterns

We next focused our analysis on the modulation of CpG methylation throughout the complete B cell maturation program. We identified dynamic methylation levels at 4.93 million CpGs (>0.25 methylation change in one set of samples and >0.1 methylation change 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 in comparison to the preceding stage were germinal center B cells, memory B cells and bone marrow plasma cells (Fig. 3a and Supplementary Fig. 8). Interestingly, multiple genes directly involved in B cell differentiation (for example, *ARID3A*, *BCL2*, *BLK*, *EBF1* and *IRF4*) showed complex modulation of their DNA methylation profiles 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 found to be differentially methylated in a wide range of human cell types and tissues¹. 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 have been caused by differences in coverage and the bioinformatics 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 data, although on a smaller scale, microarray data showed that 22.4% of the measured CpGs had variable methylation levels during B cell differentiation ($n = 106,562$; Online Methods). Because the set of samples was more comprehensive for these analyses, we used these data to define sets of CpGs showing similar DNA methylation dynamics during B cell differentiation. To this end, we based our next analysis on a linear model of B cell maturation, from HPCs to the terminal differentiation of germinal center B cells into tonsillar plasma cells and long-lived bone marrow plasma cells. We defined 20 major modules containing at least 500 CpGs each that could be

classified according to 4 general DNA methylation modulation patterns during differentiation: (i) methylation levels decreased (9 modules), (ii) methylation levels first decreased and then increased (3 modules), (iii) methylation levels first increased and then decreased

(2 modules) and (iv) methylation levels increased (6 modules) (**Supplementary Data Set 1**). Three of these 20 modules covered 57.8% of all dynamic CpGs: module 8 defined by CpG demethylation starting in germinal center B cells and continuing in bone marrow

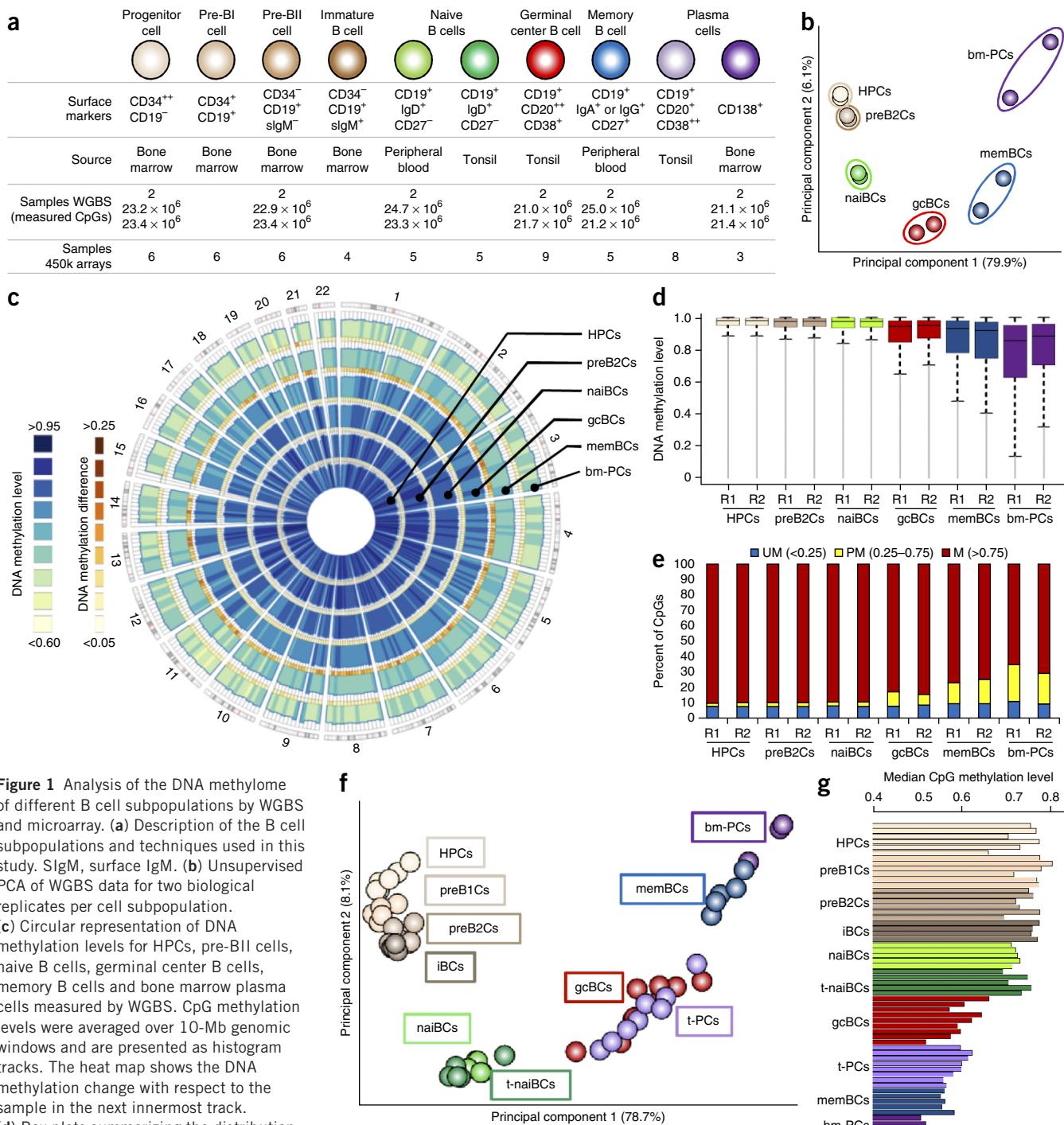


Figure 1 Analysis of the DNA methylome of different B cell subpopulations by WGBS and microarray. **(a)** Description of the B cell subpopulations and techniques used in this study. IgM, surface IgM. **(b)** Unsupervised PCA of WGBS data for two biological replicates per cell subpopulation. **(c)** Circular representation of DNA methylation levels for HPCs, pre-BII cells, naïve B cells, germinal center B cells, memory B cells and bone marrow plasma cells measured by WGBS. CpG methylation levels were averaged over 10-Mb genomic windows and are presented as histogram tracks. The heat map shows the DNA methylation change with respect to the sample in the next innermost track. **(d)** Box plots summarizing the distribution of DNA methylation levels per sample for the

16.1 million CpGs with methylation estimates in all 12 samples. The lower and upper limits of the colored boxes represent the first and third quartiles, respectively, and the black horizontal line is the median. Whiskers indicate the variability outside the upper and lower quartiles, and outliers are plotted as gray points. (e) Global methylation status of samples measured by WGBS. Percentages of methylated (M; red), partially methylated (PM; yellow) and unmethylated (UM; blue) CpGs are shown. (f) Unsupervised PCA of microarray methylation data for all samples used in the study. (g) Median values of DNA methylation data measured by microarray. HPCs, hematopoietic progenitor cells; preB1Cs, pre-BI cells; preB2Cs, pre-BII cells; iBCs, immature B cells; naiBCs, naive B cells from peripheral blood; t-naiBCs, naive B cells from tonsil; gcBCs, germinal center B cells; t-PCs, plasma cells from tonsil; memBCs, memory B cells from peripheral blood; bm-PCs, plasma cells from bone marrow. In d and e, R1 and R2 refer to the two biological replicates.

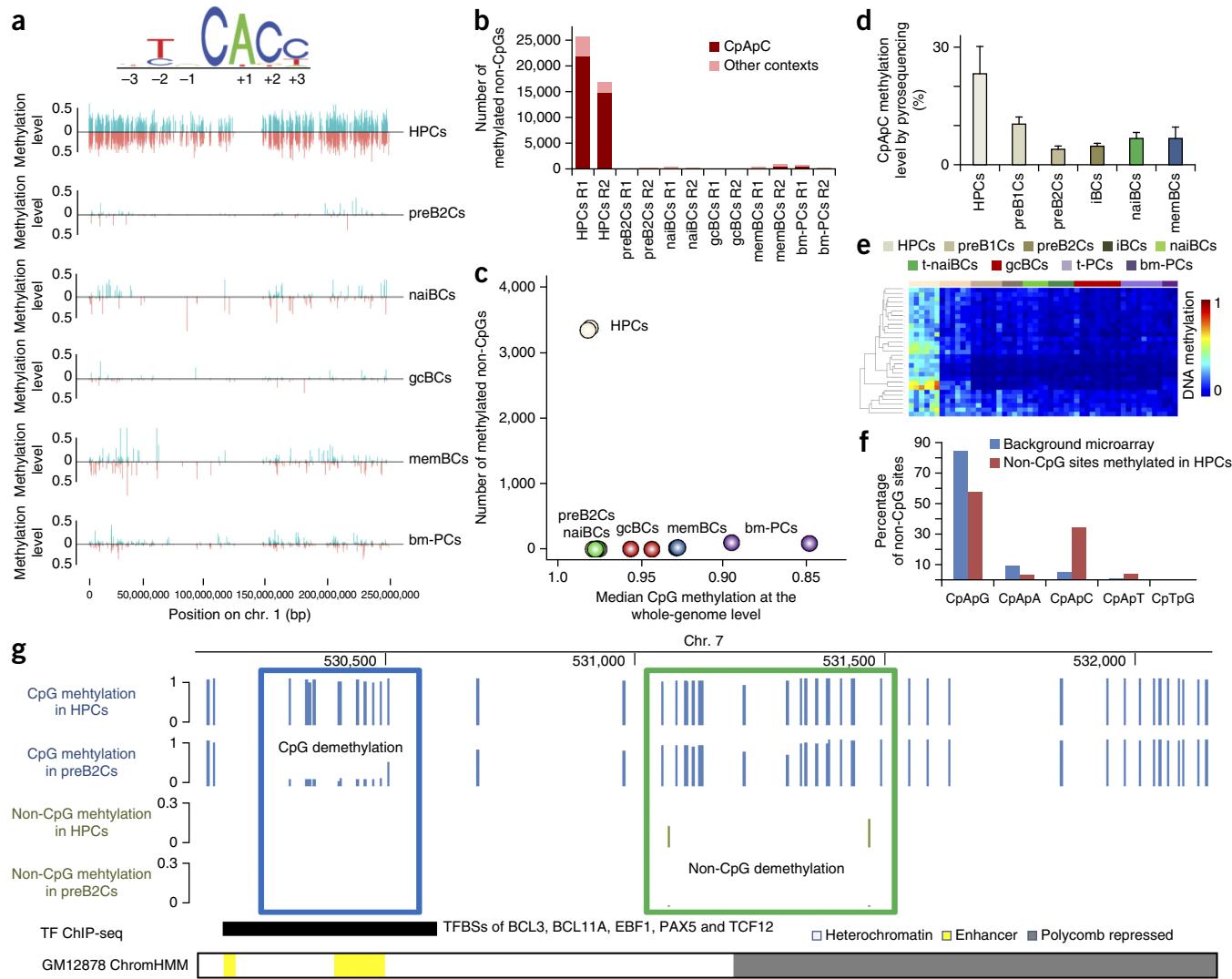


Figure 2 Non-CpG methylation detected during B cell differentiation. **(a)** Browser representation of non-CpG methylation, which takes place mostly in the CpApC sequence context (only chromosome 1 is shown). Methylation on the reverse strand is marked in blue, whereas that on the forward strand appears in red. **(b)** Number of non-CpG sites with nonzero 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 are marked 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 BPS in two independent biological replicates of each subpopulation. For this analysis, we used a CpApC site (chr. 2: 85,933,406) shown to be methylated in HPCs by WGBS. Error bars, s.d. **(e)** Heat-map representation of 26 methylated non-CpGs measured by microarray (mean methylation = 34.7%). **(f)** Percentage of methylated non-CpGs in distinct sequence contexts detected by microarray. **(g)** Representation of CpG and non-CpG dynamics upon B cell commitment. CpG methylation is marked in blue, and non-CpG methylation is marked 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. The ChromHMM track, available at the UCSC Genome Browser, shows the chromatin states of the displayed genomic region in the lymphoblastoid B cell line GM12878. TF, transcription factor.

plasma cells ($n = 34,604$), module 9 showing demethylation only in bone marrow plasma cells ($n = 13,044$) and module 20 characterized by hypermethylation in only bone marrow plasma cells ($n = 13,949$) (Fig. 3c). These data confirmed that DNA hypomethylation mainly occurs in germinal center B cells and bone marrow plasma cells. Furthermore, it could be appreciated that hypermethylation is a late event, mainly occurring in bone marrow plasma cells. Overall, we may conclude that 84.5% of 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 B cells can be separated by maturation stage on the basis of the methylation levels of the 20 modules, 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 B cells into each maturation stage (Supplementary Fig. 11).

Functional analysis of dynamically methylated regions

In general, the CpGs losing methylation at any B cell maturation stage were preferentially located in introns, intergenic regions and repetitive elements (for example, long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs) and long

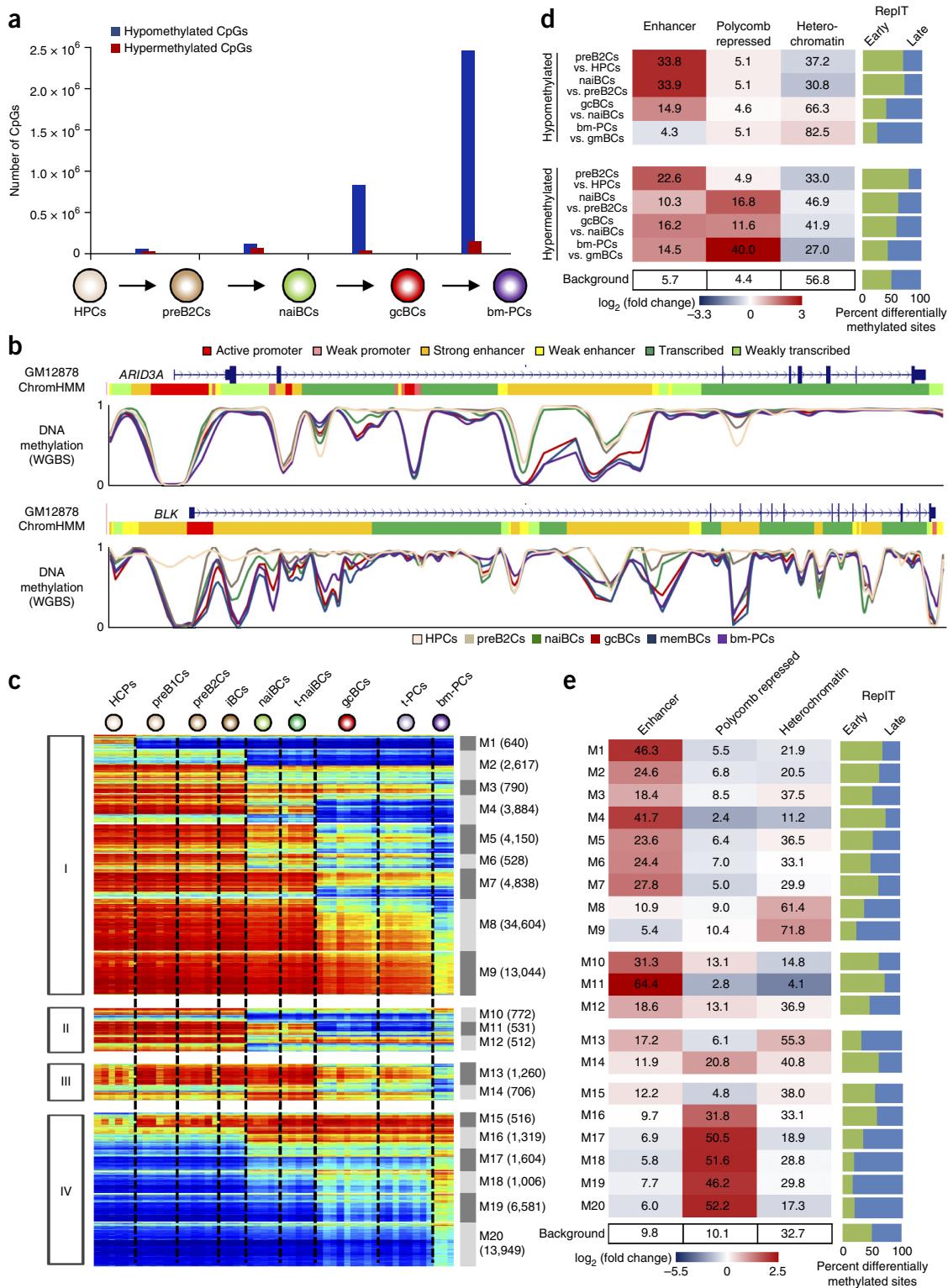


Figure 3 Dynamic DNA methylation during B cell differentiation. **(a)** Differentially methylated CpGs detected by WGBS considering the two replicates per cell subpopulation (see the Online Methods for an explanation of the criteria). **(b)** Smoothed DNA methylation data generated by WGBS across the promoter regions and gene bodies of *ARID3A* and *BLK*. The DNA methylation patterns for these genes are differentially modulated in different B cell subpopulations, especially in enhancer regions. **(c)** Heat-map representation of 20 major modules of dynamic CpGs, divided into 4 different patterns, detected by microarray. The number of CpGs within each module is given in parentheses. **(d)** Chromatin state characterization of differentially methylated CpGs identified by WGBS. **(e)** Chromatin state characterization of the 20 major modules detected by microarray. In **d** and **e**, numbers indicate the percentage of sites located in enhancers, Polycomb-repressed regions or heterochromatin. The blue-to-red color scale represents \log_2 (fold change), with respect to the background. Green and blue bars represent the percentages of differentially methylated sites that reside in early- or late-replicating regions, respectively. ReplT, replication timing.

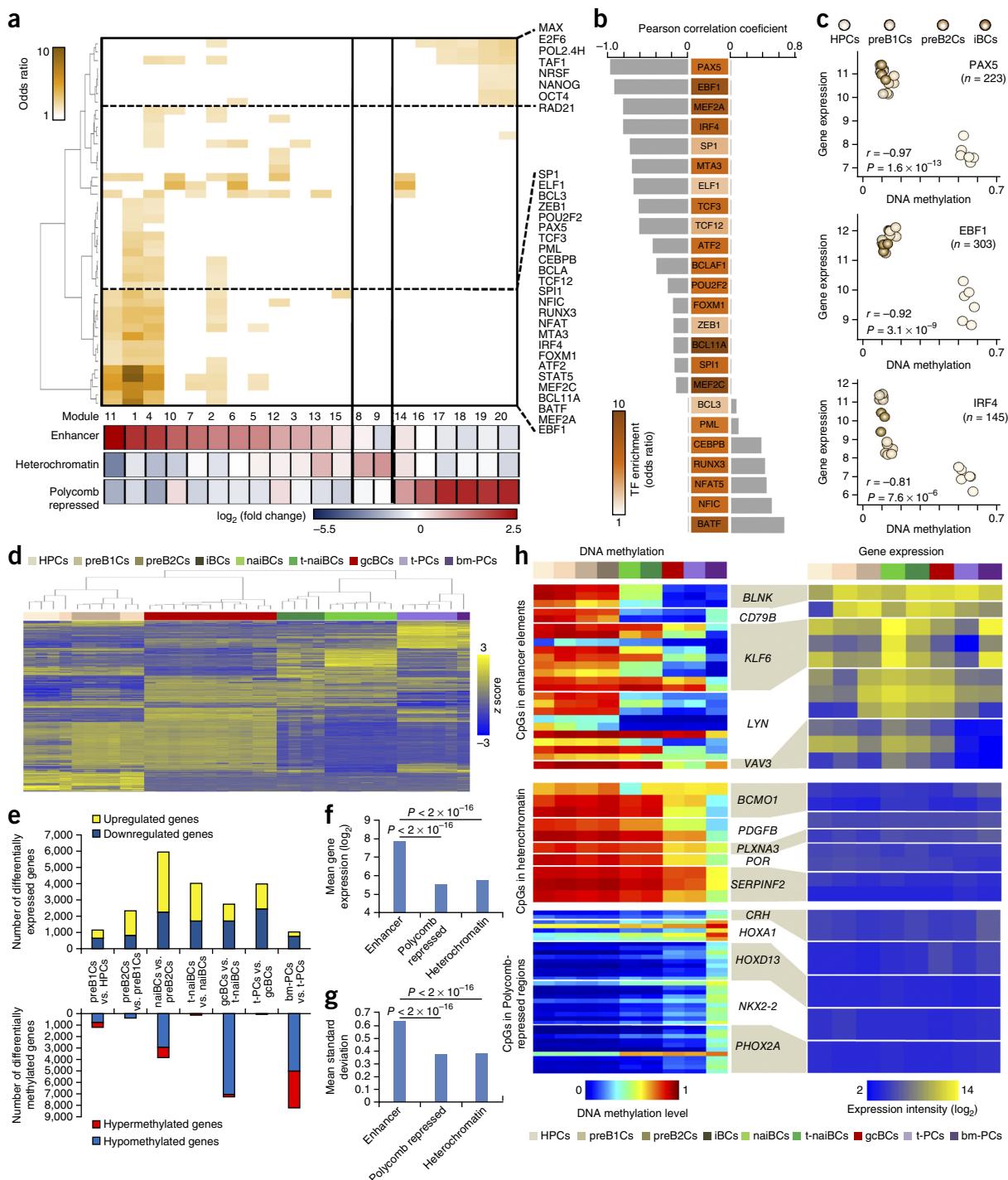


Figure 4 Association between DNA methylation and gene expression in different chromatin states. (a) Heat map presenting significant ($P < 0.01$) enrichments for TFBSs in the different methylation modules identified by methylation array. Below the heat map, \log_2 (fold change) for enhancer, heterochromatin and Polycomb-repressed regions in each differentially methylated group is presented. (b) Correlation between the expression levels of transcription factors and the mean methylation levels of their binding sites using samples with available expression and methylation data for the same donors. We applied this analysis to methylation data for module 1 (demethylation upon B cell commitment) and gene expression data for precursor cells. The white-to-brown color scale represents the odds ratio for TFBS enrichment. (c) Scatter plots showing correlation of the expression levels of PAX5, EBF1 and IRF4 with the mean methylation levels of their binding sites in each sample (the number of TFBSs associated with CpGs belonging to module 1 is shown below the transcription factor name). (d) Unsupervised clustering analysis of gene expression data using the 687 tags (439 genes) with the highest variability (s.d. >2) across the B cell differentiation process. (e) Differentially expressed genes (top) and differentially methylated genes (bottom) in each comparison of adjacent cell subpopulations. (f,g) Mean expression levels (f) and expression variability (g) during B cell differentiation for genes containing dynamic CpGs targeting enhancers, Polycomb-repressed regions and heterochromatin. (h) Heat maps showing the DNA methylation levels (left) and gene expression levels (right) of representative genes with dynamic methylation in enhancers, heterochromatin and Polycomb-repressed regions.

terminal repeats (LTRs)) and were enriched for genomic areas lacking CpG islands (CGIs) (**Supplementary Figs. 12 and 13**). In contrast, CpG-gaining methylation were enriched for CGIs and promoter regions (**Supplementary Fig. 12**). Additionally, we classified differentially methylated sites using categorization of the genome into the different chromatin states observed in immortalized mature B cells²⁸ (**Supplementary Fig. 14**). Both WGBS and microarray data showed 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 germinal center B cells onward was preferentially located in heterochromatic regions. Gain of CpG methylation was a rare event in early B cell differentiation but was rather frequent in mature B cells, especially in bone marrow plasma cells. Such CpG hypermethylation preferentially targeted Polycomb-repressed regions (**Fig. 3d,e**).

We next studied the mechanisms underlying enhancer demethylation in the B cell differentiation process. We identified a significant enrichment (fold change > 2, $P < 0.01$) in these enhancers of the 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 Sets 2** and **3**). As B cell commitment is associated with the expression of lineage-specific transcription factors²⁰, we analyzed the transition from HPCs to pre-BI cells (module 1) in detail and observed globally an inverse correlation between the expression of transcription factors and the methylation levels of their binding sites (**Fig. 4b,c**). We further investigated the association between expression of transcription factors and the methylation status of their binding sites over the entire differentiation program; in general, once a TFBS became demethylated at any B cell differentiation stage, it remained unmethylated in subsequent stages, suggesting an epigenetic memory of transcription factor binding²⁴ (**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 Set 4**). In contrast, CpGs in modules enriched for heterochromatin or

Polycomb-repressed regions did not target genes involved in the immune system but rather affected terms such as development, locomotion or behavior (**Supplementary Fig. 19** and **Supplementary Data Set 4**). On the basis of these observations, we hypothesized that differential methylation in enhancer elements might be globally associated with gene expression, whereas that affecting inactive elements (heterochromatin or Polycomb-repressed regions) might 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 numbers of genes differentially expressed and differentially methylated by comparing adjacent cell subpopulations. Globally, there was poor association between these numbers, and large transcriptional changes could be related to minor modulation of the DNA methylome and vice versa (**Fig. 4e**). We further explored the association between DNA methylation patterns and gene expression by focusing on genes with dynamic methylation in enhancers, heterochromatin and Polycomb-repressed regions. We observed that both the variability in expression and mean expression levels of genes containing dynamic CpGs in enhancer elements were much higher than for 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 targeted genes with higher expression levels and greater variation in expression throughout the entire B cell differentiation process. 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.

We next sought to analyze the functions of genes with dynamic enhancer methylation in more detail. As B cell receptor signaling is a key

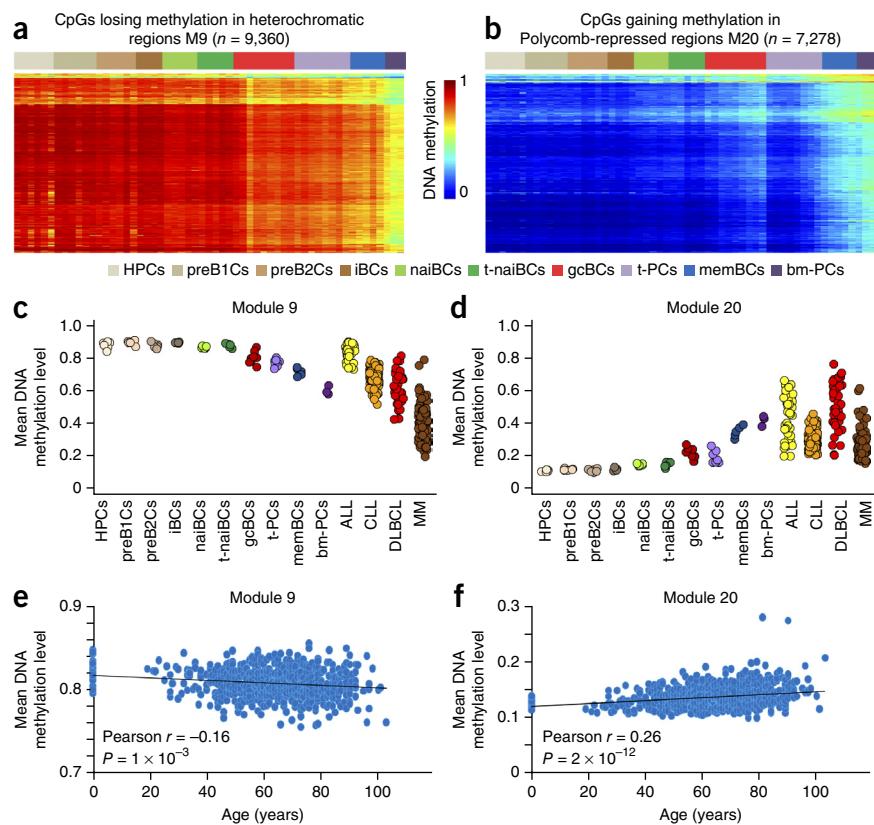


Figure 5 DNA methylation changes during B cell differentiation in the context of cancer and aging. **(a)** Heat map of a subset of CpGs from module 9 that lose methylation in heterochromatic regions. **(b)** Heat map of a subset of CpGs from module 20 that gain methylation in Polycomb-repressed regions. **(c,d)** Scatter plots presenting the mean methylation levels of CpGs in heterochromatin from module 9 (**c**) and in Polycomb-repressed regions from module 20 (**d**) in different B cell subsets and four types of hematological neoplasm. **(e,f)** Mean methylation levels of CpGs in heterochromatin from module 9 (**e**) and in Polycomb-repressed regions from module 20 (**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 in comparison to their normal counterparts. (a) Analysis of differential methylation was performed for three models of lymphoid neoplasm that arise from three distinct maturation stages in B cell development: ALL versus precursor B cells (pre-BI and pre-BII cells), the GCB subgroup of DLBCL versus germinal center B cells, and MM versus plasma cells (from the tonsil and bone marrow). (b) Bar plots showing the proportion of dynamically methylated CpGs in B cell differentiation that are also differentially methylated in hematological neoplasias in comparison to their normal counterparts. (c) Percentages of hypermethylated (top) and hypomethylated (bottom) CpGs located in enhancers (left), heterochromatin (middle) and Polycomb-repressed regions (right). (d) Heat maps showing differentially methylated CpGs in ALL in comparison 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 for 450k microarray data.

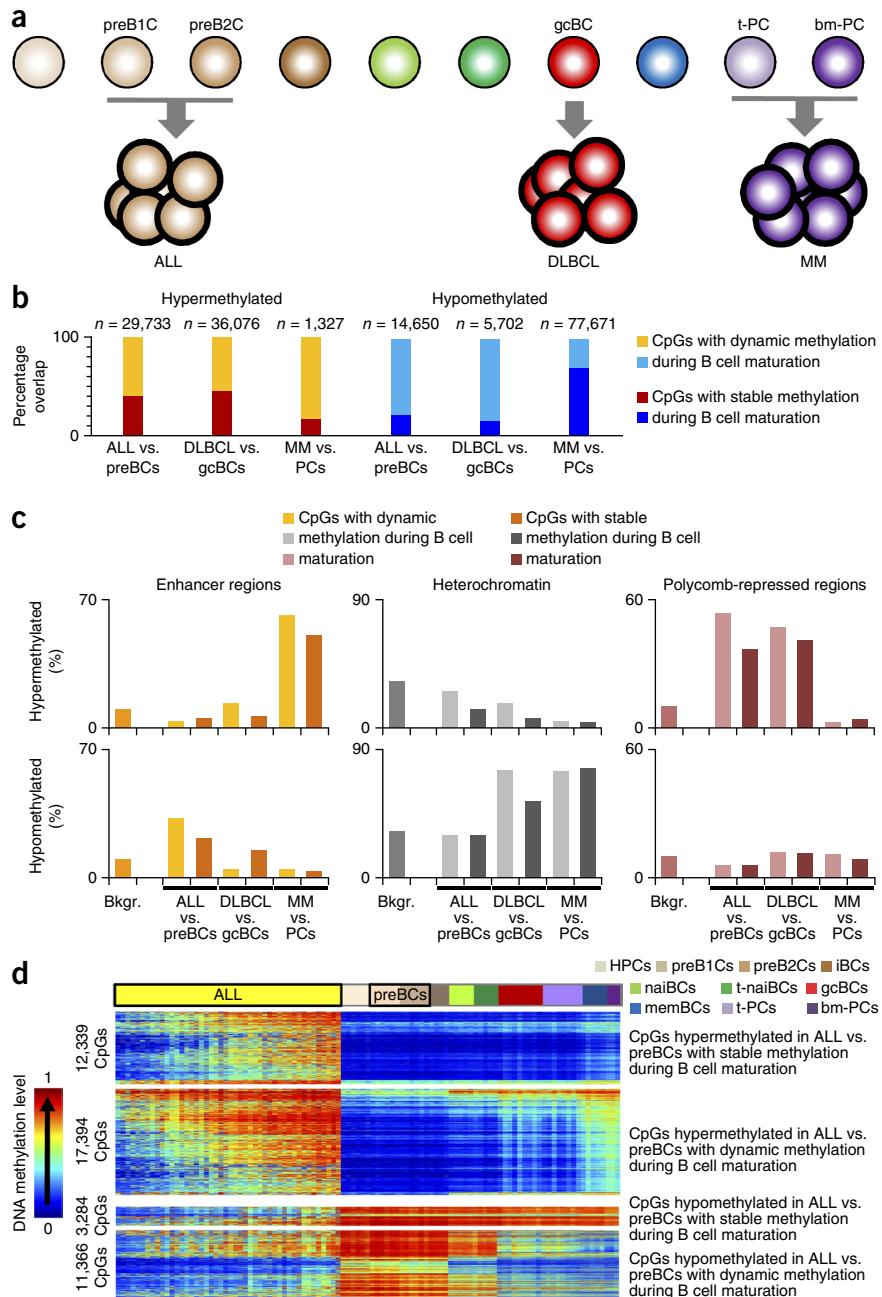
element specific for B cell differentiation, we studied the methylation patterns of 41 genes involved in this function. Of these, 38 (93%) had a total of 234 CpGs with dynamic methylation that were preferentially located in enhancers (**Supplementary Fig. 20**). Furthermore, we observed that 38% of all enhancers with dynamic methylation ($P < 0.001$) belonged to a B cell-specific functional gene network²⁹ (**Supplementary Fig. 21**).

As our study comprised cell subpopulations isolated from three different compartments (bone marrow, peripheral blood and tonsil), we aimed to detect particular DNA methylation imprints related to these different locations, but we did not identify any consistent pattern. However, we did observe that naive B cells isolated from different compartments (peripheral blood or tonsils) showed a pronounced change in their transcriptomes, whereas their methylomes remained virtually identical (**Supplementary Figs. 22**

and **23**). On the basis of this observation, we may hypothesize that the gene expression changes in naive tonsillar B cells are essential to optimize antigen recognition in the tonsil, followed by successful B cell activation. However, when no antigen is found, the naive tonsillar B cells will reenter the bloodstream. Hence, the naive B cell state in tonsils 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 memory B cells and bone marrow plasma cells²³ showed extensive perturbation of their DNA methylomes in comparison to germinal center B cells (**Supplementary Fig. 8**). Remarkably, although bone marrow plasma cells and newly generated tonsillar plasma cells had similar transcriptomes, their methylomes were different (**Supplementary Fig. 22**). The major DNA



methylation changes in memory B cells and bone marrow plasma cells affected presumably non-functional elements such as heterochromatin and Polycomb-repressed regions. Thus, our findings suggest that part of the epigenetic makeup of memory B cells and bone marrow plasma cells is caused by epigenetic drift associated with their long lifespans²³. To determine whether such drift might be related to the expression of DNMTs, we analyzed *DNMT1*, *DNMT3A* and *DNMT3B* transcript levels by quantitative RT-PCR (qRT-PCR) in sorted germinal center B cells, memory B cells, tonsillar plasma cells and bone marrow plasma cells. There were low levels of *DNMT1* in memory B cells, tonsillar plasma cells and bone marrow plasma cells (**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 rate³⁰. Gain of methylation in Polycomb-repressed regions has been

linked to the activities of DNMT3A and DNMT3B^{31,32}. We detected that, in comparison to germinal center B cells, tonsillar plasma cells and bone marrow plasma cells, and to a lesser extent memory B cells, had upregulated *DNMT3A* levels. These results may suggest that minor increases in the levels 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 in Polycomb-repressed regions (defined by the presence of histone H3 trimethylated at lysine 27, H3K27me3), we performed bisulfite sequencing of immunoprecipitated chromatin (ChIP-BS) using an antibody to H3K27me3 (refs. 33,34). This experiment suggested that, in memory B cells, H3K27me3 and DNA methylation coexist both in CpG-rich and CpG-poor regions and that the DNA methylation levels within nucleosomes containing H3K27me3 are lower than in 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 determine whether the CpGs showing differential methylation in long-lived B cells overlapped with those becoming differentially methylated in cancer, we used methylation data from various lymphoid neoplasms, including acute lymphoblastic leukemia³⁹ (ALL; $n = 46$), chronic lymphocytic leukemia¹⁰ (CLL; $n = 139$), diffuse large B cell lymphoma (DLBCL; $n = 40$) and multiple myeloma⁴⁰ (MM; $n = 104$). We analyzed the methylation levels of CpGs losing methylation (modules 8 and 9) and those gaining methylation in memory B cells and bone marrow plasma cells (modules 19 and 20). Indeed, we observed that, for these CpGs, neoplastic cells showed a DNA methylation profile similar to those of memory B cells and bone marrow plasma cells (**Fig. 5** and **Supplementary Fig. 25**). To further evaluate the epigenetic link between normal B cell differentiation and neoplastic transformation, we compared the DNA methylomes of B cell neoplasms with those of their normal cell counterparts, for example, ALL³⁹ versus pre-B cells, germinal center B cell-like DLBCL versus germinal center B cells and MM⁴⁰ versus plasma cells (**Fig. 6a** and **Supplementary Data Set 5**). The results indicate that a large fraction of the CpGs differentially methylated in cancers are dynamically methylated during normal B cell differentiation, with the overlap ranging from 53 to 82% for hypermethylated sites and from 29 to 84% for hypomethylated sites (**Fig. 6b**). Interestingly, we found that hypomethylation in ALL was enriched for CpGs in enhancers, whereas hypomethylation in DLBCL and MM predominantly affected CpGs in heterochromatin (**Fig. 6c**). Additionally, although ALL cells are arrested at the pre-B cell stage, they acquired hypermethylation in Polycomb-repressed regions, which is characteristic of more mature differentiation stages (**Fig. 6d**). MM cells, in contrast, did not acquire hypermethylation of Polycomb-repressed regions, as their cell of origin already shows this feature, but, as they downregulate the B cell program, they acquire hypermethylation of CpGs in B cell-specific enhancers⁴⁰.

Finally, global hypomethylation accompanied by local hypermethylation of Polycomb targets is also a molecular hallmark of aging^{41–45}. We analyzed the DNA methylation values of CpGs within modules 8, 9, 19 and 20 using data from 694 peripheral blood samples obtained from donors ranging in age from 0 to 101 years^{43,44}. Indeed, we detected a significant correlation with age ($P < 0.01$), following the same trend in long-lived cells and cancer cells—that is, methylation of heterochromatin diminished with age, whereas methylation of Polycomb targets increased (**Fig. 5e,f** and **Supplementary Fig. 25**). We next compared B cells with short and long lifespans (naive

B cells and memory B cells, respectively) isolated simultaneously from individuals of different age, and we found that memory B cells acquired 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 in the fraction of long-lived cells in older individuals may represent a confounding variable in age-related methylation studies⁴⁶.

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 have 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 a 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 previous differentiation stages. In contrast to other reports for 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 in CpG methylation, which mostly affected enhancers containing binding sites for B cell-specific transcription factors. The functional link between transcription factor binding, CpG demethylation and enhancer activation has recently been analyzed during stem cell differentiation^{50,51} as well as in hematopoietic cells^{52,53}, non-hematopoietic cells⁵⁴ 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 cells²⁹. Although mainstream research on DNA methylation still remains centered on promoter regions, our results imply that DNA methylation changes in enhancers seem to be more closely related to cell specification and maturation^{1,9,56}. However, similarly 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 naive B cells onward), DNA methylation changes are guided more by other mechanisms than by the intrinsic program of B cell transcription factors. Upon antigen encounter, the germinal center reaction is induced and, at this stage, germinal center B cells 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 of germinal center B cells, as normal proliferative tissues tend to lose

methylation at late-replicating regions⁵⁸. However, downstream B cell subpopulations derived from germinal center B cells, such as non-proliferative memory B cells and plasma cells, 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 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 has previously been 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 germinal center B cells and becomes particularly enhanced in non-proliferative, long-lived B cells. On the basis of our results, we hypothesize that not all cells in an organism are subject to epigenetic drift as a consequence of time, only those with a long lifespan. Finally, 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 have similar DNA methylation signatures. Furthermore, comparing various B cell neoplasms with their normal cellular counterparts, we observe that a large proportion of the differentially methylated sites in cancers overlap with those undergoing dynamic methylation during normal differentiation, especially with those altered in memory B cells and bone marrow plasma cells. Interestingly, as in pre-B cells, hypomethylation in ALL is enriched for enhancer elements, whereas, in line with germinal center B cells and bone marrow plasma cells, hypomethylation in DLBCL and MM is mostly enriched for heterochromatin. In general, these findings suggest that the epigenetic configuration of a cell from a particular maturation stage influences the DNA methylation changes acquired during its clonal expansion and neoplastic transformation. This new strategy of analyzing the DNA methylome of B cell tumors in the context of the entire differentiation program may allow new insights into the role of DNA methylation in cancer. We postulate that methylation changes shared by neoplastic transformation and normal differentiation may represent epigenetic passengers, whereas those exclusively taking place in tumor cells should constitute epigenetic drivers with a potential functional impact in the disease.

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

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. WGBS data have been deposited in the European Genome-phenome Archive (EGA) under accessions EGAD00001001304 and EGAS00001000272. DNA methylation and gene expression microarray data are available from the EGA under accessions EGAS00001001196 and EGAS00001001197, respectively.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank C. López-Otín for critical reading of this manuscript, M. Dabad Castellà for his assistance with the WGBS data analysis and M.A. Peinado (Institute of Predictive and Personalized Medicine of Cancer, Barcelona) 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) and the Spanish Ministry of Economy and Competitiveness (MINECO; project SAF2009-08663). Methylation microarrays were outsourced to the Spanish Centro Nacional de Genotipado (CEGEN-ISCIII). We are indebted to the Genomics core facility of the Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS) for technical help. This work was partially developed at the Centro Esther Koplowitz (CEK; Barcelona, Spain). M.K. is supported by Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR; Generalitat de Catalunya), E.C. is an Academia Researcher of the Institució Catalana de Recerca i Estudis Avançats and J.I.M.-S. is a Ramón y Cajal researcher of MINECO.

AUTHOR CONTRIBUTIONS

M.K., A.C.Q., N.R., M.P., X.A., F.P., D.A., B.P., G. Caron, T.F., M.O.M., M.E.F., S.-T.L. and J.L.W. provided samples from healthy donors and/or purified B cell subpopulations. M.K., A.C.Q., G. Castellano, R.B. and G. Clot 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. Castellano, S.E., V.P., D. Rico and A.V. functionally characterized dynamically methylated genes. D. Richardson, 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 manuscript with predominant assistance from M.K. and R.B.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Ziller, M.J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477–481 (2013).
- Bernstein, B.E., Meissner, A. & Lander, E.S. The mammalian epigenome. *Cell* **128**, 669–681 (2007).
- Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
- Bergman, Y. & Cedar, H. DNA methylation dynamics in health and disease. *Nat. Struct. Mol. Biol.* **20**, 274–281 (2013).
- Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
- Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537–541 (2014).
- Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
- Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
- Gifford, C.A. *et al.* Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* **153**, 1149–1163 (2013).
- Kulis, M. *et al.* Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 1236–1242 (2012).
- 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**, 40–46 (2012).
- Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).
- Hansen, K.D. *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.* **43**, 768–775 (2011).
- Li, Y. *et al.* The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol.* **8**, e1000533 (2010).
- Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
- Varley, K.E. *et al.* Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* **23**, 555–567 (2013).
- 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**, 11339–11351 (2012).
- 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.* **23**, 2030–2041 (2013).
- Shaknovich, R. *et al.* DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. *Blood* **118**, 3559–3569 (2011).
- 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**, 133–134 (1997).
23. Slifka, M.K., Antia, R., Whitmire, J.K. & Ahmed, R. Humoral immunity due to long-lived plasma cells. *Immunity* **8**, 363–372 (1998).
24. Hon, G.C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* **45**, 1198–1206 (2013).
25. Bibikova, M. *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98**, 288–295 (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**, 43–49 (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. Viré, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871–874 (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**, 1128–1138 (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**, 1120–1127 (2012).
35. Martín-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**, 2488–2497 (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**, 237–242 (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**, 232–236 (2007).
38. Widschwenter, M. *et al.* Epigenetic stem cell signature in cancer. *Nat. Genet.* **39**, 157–158 (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**, 4323–4336 (2013).
40. Aguirre, X. *et al.* Whole-epigenome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers. *Genome Res.* **25**, 478–487 (2015).
41. López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (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**, 359–367 (2013).
44. Heyn, H. *et al.* Distinct DNA methylomes of newborns and centenarians. *Proc. Natl. Acad. Sci. USA* **109**, 10522–10527 (2012).
45. Maegawa, S. *et al.* Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res.* **20**, 332–340 (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**, 1074–1086 (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**, 344–349 (2015).
51. Stadler, M.B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495 (2011).
52. Schmidl, C. *et al.* Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res.* **19**, 1165–1174 (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**, 2950–2955 (2004).
54. Wiench, M. *et al.* DNA methylation status predicts cell type-specific enhancer activity. *EMBO J.* **30**, 3028–3039 (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**, 1421–1432 (2014).
56. Schlesinger, F., Smith, A.D., Gingras, T.R., Hannon, G.J. & Hodges, E. *De novo* DNA demethylation and noncoding transcription define active intergenic regulatory elements. *Genome Res.* **23**, 1601–1614 (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., Toporoff, G., Rosenberg, M. & Hellman, A. Replication timing-related and gene body-specific methylation of active human genes. *Hum. Mol. Genet.* **20**, 670–680 (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**, 440–446 (2010).
60. Cruickshanks, H.A. *et al.* Senescent cells harbour features of the cancer epigenome. *Nat. Cell Biol.* **15**, 1495–1506 (2013).
61. Rakyan, V.K. *et al.* Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res.* **20**, 434–439 (2010).

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 on the basis of high levels of CD34 protein expression (CD34^{hi}) and lack of expression of the B cell marker CD19. This population, designated uncommitted HPCs, contained predominantly multipotent progenitors before lineage commitment and also common lymphoid progenitors and hematopoietic stem cells. B cell-committed progenitors were isolated on the basis of their expression of CD19 and CD34 (CD19⁺CD34⁺) and were predominantly pre-BI cells. Two immature B cell populations expressing CD19 and lacking CD34 were isolated and differentiated on the basis of sIgM expression: pre-BII cells that were sIgM⁻CD19⁺ and immature B cells that were sIgM⁺CD19⁺. DNA methylation and gene expression data for these four subpopulations have been published previously¹⁷.

Peripheral blood B cell subpopulations—naive B cells and memory B cells—were obtained from buffy coats for healthy adult donors ranging in age between 28 and 66 years. After Ficoll-Isopaque density centrifugation, CD19⁺ B cells were isolated by positive magnetic cell separation using the AutoMACS system (Miltenyi Biotec, 130-050-301). CD19⁺ cells were labeled with antibodies to CD27 (BD Biosciences, clone M-T271), IgD (BD Biosciences, clone IA6-2), IgM (BD Biosciences, clone G20-27), IgG (BD Biosciences, clone G18-145) and IgA (DacoCytomation, F0188) for 15 min at room temperature in staining buffer (PBS with 0.5% BSA). Naive B cells (CD19⁺CD27⁻IgD⁺) and memory B cells (CD19⁺CD27⁺IgA⁺ or CD19⁺CD27⁺IgG⁺) were obtained by FACS sorting on a FACSAria II (BD Biosciences).

Plasma cells, germinal center B cells and naive B cells were isolated from the tonsils of children undergoing tonsillectomy (ranging in age between 2 and 13 years), obtained from the Clínica Universidad de Navarra (Pamplona, Spain) or Clinique Mutualiste La Sagesse (Rennes, France). Tonsils were minced extensively; after Ficoll-Isopaque density centrifugation, enrichment of B cells was performed with the AutoMACS system either by positive selection of CD19⁺ cells or using B Cell Isolation Kit II (Miltenyi Biotec). Tonsillar plasma cells (CD20^{med}CD38^{hi}), germinal center B cells (CD20^{hi}CD38^{med}) and naive B cells (CD20⁺CD23⁺) were separated by FACS sorting using antibodies from BD Biosciences (CD20, clone 2H7; CD38, clone HIT-2; CD23, clone M-L233). In part, naive B cells were also selected using a slightly different marker combination (CD19⁺CD27⁻IgD⁺ or IgD⁺CD38^{low}CD27⁻). Germinal center B cells were also selected by the marker combination IgD-CD38^{hi}CD10⁺CXCR4⁺ (CD38 and CD10 antibodies from Beckman Coulter, clones LS198-4-3 and ALB1, respectively; CXCR4 antibody from BD Biosciences, clone 51505). Naive B cells and germinal center B cells isolated with different markers constituted the same cell subpopulations, as evidenced by the fact that each subpopulation showed homogeneous DNA methylation and transcriptional profiles (Supplementary Fig. 22).

Bone marrow plasma cells were selected from healthy donors ranging from 20 to 30 years of age. After density gradient centrifugation, we performed selective depletion of CD3⁺, CD14⁺ and CD15⁺ cells by immunomagnetic selection (Miltenyi Biotec) followed by flow cytometry cell sorting for CD45⁺CD138⁺CD38⁺ cells (CD38 and CD45 antibodies from Miltenyi Biotec, clones HIT-2 and 5B1, respectively; CD138 antibody from BD Biosciences, clone 44F9) using a FACSAria II device.

The purity of each of the isolated B cell subpopulations exceeded 90% in all samples. DNA was extracted from purified samples using a Qiagen kit (QIAamp DNA Mini kit), following the 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 bone marrow plasma cells, which were pooled from four different donors. Total RNA was extracted with TRIzol (Invitrogen) following the manufacturer's recommendations. RNA quality was assessed with the Agilent 2100 Bioanalyzer (Agilent Technologies). The use of the samples analyzed in the present study was approved by the ethics committees of the Hospital Clinic de Barcelona, Hospital Universidad de Navarra and the University Hospital of Rennes as well as the University of California San Francisco Committee on Human Research.

Whole-genome bisulfite sequencing. We performed WGBS on two independent sets of biological replicates for six B cell differentiation stages. Briefly, genomic

DNA (1–2 µg) was spiked with unmethylated λ DNA (5 ng of λ DNA per microgram of genomic DNA; Promega). DNA was sheared by sonication to 50–500 bp in size using a Covaris E220 sonicator, and fragments of 150–300 bp were selected using AMPure XP beads (Agencourt Bioscience). Genomic DNA libraries were constructed using the Illumina TruSeq Sample Preparation kit following Illumina's standard protocol: end repair was performed on the DNA fragments, an adenine was added to the 3' end of each fragment and Illumina TruSeq adaptors were ligated to both ends. After adaptor ligation, DNA was treated with sodium bisulfite using the EpiTExy Bisulfite kit (Qiagen), following the manufacturer's instructions for formalin-fixed, paraffin-embedded tissue samples. Two rounds of bisulfite conversion were performed to ensure a conversion rate of over 99%. Enrichment for adaptor-ligated DNA was carried out through seven PCR cycles using PfuTurboCx Hot-Start DNA polymerase (Stratagene). Library quality was monitored using the Agilent 2100 Bioanalyzer, and the concentration of viable sequencing fragments (molecules carrying adaptors at both ends) was estimated using quantitative PCR with the library quantification kit from Kapa Biosystems. Paired-end DNA sequencing (2 × 100 bp) was then performed using the Illumina HiSeq 2000 platform. The amounts of sequence reads and proportions 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 genome and two copies of the NCBI viral genome database (v35). For both the human and viral references, one copy had all cytosine bases replaced by thymine bases and the other had all guanine bases replaced by adenine bases. 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 for each read was stored. The first read from each pair then had all cytosine bases replaced by thymine bases, and the second read had all guanine bases replaced by adenine bases. Read mapping with GEM was performed, allowing up to four mismatches per read with respect to the reference. After read mapping, the original sequence for each read was restored.

Estimation of cytosine levels was carried out on read pairs where both members of the pair mapped to the same contig with consistent orientation and there was no other such configuration at the same or a smaller 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 cytosine base 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 >500× coverage depth were excluded to avoid centromeric or 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 of genomic DNA. Bisulfite-converted DNA was hybridized to the HumanMethylation 450k BeadChip (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 HumanMethylation 450k array were analyzed in R using the minfi package⁶⁴ (version: 1.6.0), available through 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., removing CpGs with low detection *P* values, with sex-specific or individual-specific methylation, or overlapping with SNPs). Taking into account the different performance of Infinium I and Infinium II assays we used subset-quantile within-array

normalization (SWAN)⁶⁵, which corrects for technical differences between the Infinium I and Infinium II assay designs and produces a smoother overall β -value distribution.

Detection of non-CpG methylation and differential methylation analysis.

Cytosines in a non-CpG context were defined as two adjacent nucleotides where the genotype of the first nucleotide was called with high confidence as homozygous C and the second nucleotide was called with high confidence as a genotype other than 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 χ^2 test or with a Fisher's exact test when the χ^2 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 by one set of samples having a methylation difference above 0.25 and the second set of samples having a difference of at least 0.1 in the same direction (hyper- or hypomethylation). CpGs with dynamic methylation were defined as those with differential methylation in comparisons of adjacent stages and in comparison of HPCs and bone marrow plasma cells.

To calculate the overlap between dynamically methylated CpGs in B cells and those identified by Ziller *et al.*¹, we downloaded all the differentially methylated regions (DMRs) with dynamic CpGs from that study (Gene Expression Omnibus (GEO), [GSE46644](#)) and determined how many of the 4.93 million dynamic CpGs in B cells were located within DMRs identified by Ziller *et al.*

We also defined CpGs that showed variable methylation levels throughout B cell development using 450k microarray data for HPCs, pre-BI cells, pre-BII cells, immature B cells, naïve B cells, tonsillar naïve B cells, germinal center B cells, tonsillar plasma cells and bone marrow plasma cells. From the germinal center B cell stage on, B cell differentiation is branched into memory B cells or plasma cells, and memory B cells were therefore not included in this linear analysis. We performed pairwise 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 (false discovery rate (FDR) < 0.1, Wilcoxon test). Furthermore, we detected CpGs whose methylation showed a similar modulation pattern over the entire B cell differentiation process, and we grouped 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 placed into four groups depending on their methylation tendency during differentiation (decreased, increased, decreased then increased, and increased and then decreased). 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 FDR < 0.05.

Bisulfite pyrosequencing studies. The DNA methylation levels generated by WGBS and 450k microarray were validated by 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 the manufacturer's instructions. PCR amplification of the bisulfite-treated DNA was performed using specific primers for each of the selected CpGs and non-CpGs (**Supplementary Table 3**). These primers were selected using PyroMark Assay Design software (Qiagen). BPS and DNA methylation data analysis were performed with the PyroMark Q96 ID pyrosequencer 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 location of a site relative to a gene, we used these categories: TSS

1,500 (from 201 to 1,500 bp upstream of the 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. Owing to the presence of alternative TSSs and regions containing more than one gene, some of the CpGs were assigned multiple annotations. For location of a site relative to a 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 (ref. 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 promoter, weak promoter and poised promoter) defined as 'promoter regions', states 4–7 (strong enhancer and weak enhancer) 'enhancer regions', state 8 an 'insulator', state 9 a 'transcriptional transition', state 10 'transcriptional elongation', state 11 'weak transcription', state 12 'Polycomb-repressed regions' and state 13 '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 three bins: early-, mid- and late-replicating regions. Only early- and late-replicating regions were used for the analysis.

Annotation of repeat elements was carried out on the basis of 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 genes with dynamically methylated enhancers and mapped them to Entrez gene identifiers, resulting in 5,658 unique Entrez gene identifiers. Of these genes, 2,154 are contained in the B cell network and 1,993 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 using Gephi⁶⁶ and Louvain's method⁶⁷.

Chromatin immunoprecipitation coupled with bisulfite sequencing. B cells were cross-linked with 1% formaldehyde for 8 min (room temperature) before FACS separation of memory B cells. Chromatin preparation and ChIP were performed according to the Blueprint histone ChIP protocol with an antibody to H3K27me3 (C15410195, Diagenode). For whole-genome bisulfite library construction and sequencing, the immunoprecipitated DNA (50 ng) was sheared on a Covaris E220 to a fragment size of 50–500 bp and size selected for 150- to 500-bp fragments using AMPure XP beads (Agencourt Bioscience). Unmethylated λ DNA (500 ng; Promega) was treated in parallel on a Covaris E220 and also size selected with AMPure XP beads to the same fragment sizes as for the DNA sample. The unmethylated λ DNA was spiked into fragmented and size-selected immunoprecipitated DNA (5 ng of λ DNA per 1 μ g of DNA), and the TruSeq Sample Preparation kit (Illumina) was used to prepare the Illumina library by adding 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 were performed to obtain >99% conversion, following the manufacturer's instructions for formalin-fixed, paraffin-embedded tissue samples (EpiTect Bisulfite kit, Qiagen). Adaptor-ligated DNA was enriched through ten cycles of PCR with the Kapa HiFi Uracil⁺ polymerase (Kapa Biosystems). The library was run on a fraction of a lane of a HiSeq 2000 flow cell (to generate 35 million paired-end reads) with read length of 2 × 100 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, aligned reads were analyzed with the NucHunter package⁶⁸, to provide predictions of the positions of H3K27me3-modified nucleosomes. The average methylation of cytosines within 500 bp of each peak was calculated, taking into account the strand on which the cytosine was present. The same analysis was performed on WGBS data from the same cell type (memory B cells), 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 in the Broad ChromHMM analysis of the ENCODE cell line GM12878.

Gene ontology analysis. The GOstat package⁶⁹ available through Bioconductor was used to determine the enrichment of individual ontology terms in the different methylation modules as compared to all genes analyzed on the 450k array. The top 20 most significant terms for each module ($P < 0.001$) are shown in **Supplementary Data Set 4**.

Analysis of transcription factor binding sites. TFBS information was obtained using ChIP-seq data from the ENCODE Project and available through the UCSC Genome Browser. A total of 79 TFBSs were used for the analysis. The relative enrichment of each TFBS was calculated in comparison to background. A Fisher's exact test was used to assign an odds ratio and P value to each comparison.

Gene expression analyses. RNA samples from HPCs, pre-BI cells, pre-BII cells, naive B cells, tonsillar naive B cells, germinal center B cells, memory B cells, tonsillar plasma cells and bone marrow plasma cells were hybridized to Affymetrix Human Genome U219 arrays according to Affymetrix standard protocols. Analysis of scanned images for each probe set of the array was performed with GeneChip Operating Software (GCOS, Affymetrix). Raw CEL files were processed, and signals were normalized with the robust multichip average (RMA) algorithm using R statistical software in conjunction with the affy library⁷⁰ available through Bioconductor. GeneChip Human Gene 1.0 ST array data for progenitor B cells were downloaded from GEO (GSE45461) and normalized using RMA (these data were only used for **Fig. 4b,c**).

To evaluate the variability in gene expression among B cell subpopulations, we calculated the standard deviation between cell subtypes for 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 tonsillar naive B cells and naive B cells was calculated using the

limma library available through Bioconductor, requiring fold change > 1 (\log_2) between two groups and adjusted $P < 0.05$.

Reverse transcription and quantitative PCR for DNMT expression was carried out as follows. cDNA was synthesized from 100 ng of RNA sample in 20 μ l of reaction mix using oligo(dT) primers and SuperScript III enzyme according to the manufacturer's recommendations (Invitrogen). Primer sequences for the DNMTs were taken from Fang *et al.*⁷¹, whereas the sequences for those amplifying the housekeeping gene *EEF2* are provided in **Supplementary Table 3**. PCR amplification was carried out with 1 μ l of the reverse transcription sample diluted 1:2 using Power SYBR Green PCR Master Mix according to the manufacturer's recommendations (Applied Biosystems). PCR reactions were run in triplicate on a StepOne System (Applied Biosystems).

62. Marco-Sola, S., Sammeth, M., Guigo, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat. Methods* **9**, 1185–1188 (2012).
63. Bibikova, M. *et al.* Genome-wide DNA methylation profiling using Infinium® 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**, 1363–1369 (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. in *Proc. 3rd Int. AAAI Conf. Weblogs and Social Media* 361–362 (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**, 2547–2554 (2013).
69. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
70. Gautier, L., Cope, L., Bolstad, B.M. & Irizarry, R.A. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
71. Fang, J. *et al.* Epigenetic changes mediated by microRNA miR29 activate cyclooxygenase 2 and λ -1 interferon production during viral infection. *J. Virol.* **86**, 1010–1020 (2012).