

# Activation-induced deaminase is critical for the establishment of DNA methylation patterns prior to the germinal center reaction

Francesc Català-Moll<sup>1,2</sup>, Anna G. Ferreté-Bonastre<sup>1,2,†</sup>, Tianlu Li<sup>1,2,†</sup>, Dieter Weichenhan<sup>3</sup>, Pavlo Lutsik<sup>3</sup>, Laura Ciudad<sup>1,2</sup>, Ángel F. Álvarez-Prado<sup>4</sup>, Javier Rodríguez-Ubreva<sup>1,2</sup>, Christian Klemann<sup>5</sup>, Carsten Speckmann<sup>5,6</sup>, Amaya Vilas-Zornoza<sup>7</sup>, Hassan Abolhassani<sup>8</sup>, Mónica Martínez-Gallo<sup>9</sup>, Romina Dieli-Crimi<sup>9</sup>, Jacques G. Rivière<sup>10,11,12</sup>, Andrea Martín-Nalda<sup>10,11,12</sup>, Roger Colobran<sup>9,12,13</sup>, Pere Soler-Palacín<sup>10,11,12</sup>, Sven Kracker<sup>14</sup>, Lennart Hammarström<sup>8</sup>, Felipe Prosper<sup>7</sup>, Anne Durandy<sup>14</sup>, Bodo Grimbacher<sup>5,15,16,17,18</sup>, Christoph Plass<sup>3</sup> and Esteban Ballestar<sup>1,2,\*</sup>

<sup>1</sup>Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), 08916 Badalona, Barcelona, Spain, <sup>2</sup>Chromatin and Disease Group, Cancer Epigenetics and Biology Programme (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), 08908 L'Hospitalet de Llobregat, Barcelona, Spain, <sup>3</sup>Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany, <sup>4</sup>B Cell Biology Laboratory, Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), 28029 Madrid, Spain, <sup>5</sup>Institute for Immunodeficiency, Center for Chronic Immunodeficiency (CCI), Medical Center, Faculty of Medicine, University of Freiburg, 79110 Freiburg, Germany, <sup>6</sup>Faculty of Medicine, Center for Pediatrics and Adolescent Medicine, Medical Center, University of Freiburg, Germany, <sup>7</sup>Center for Applied Medical Research (CIMA), University of Navarra, 31008 Pamplona, Spain, <sup>8</sup>Division of Clinical Immunology, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital Huddinge, SE 14186 Stockholm, Sweden, <sup>9</sup>Immunology Division, Hospital Universitari Vall d'Hebron and Diagnostic Immunology Research Group, Vall d'Hebron Research Institute (VHIR), Barcelona, Spain, <sup>10</sup>Pediatric Infectious Diseases & Immunodeficiencies Unit, Hospital Universitari Vall d'Hebron, Vall d'Hebron Research Institute (VHIR), Autonomous University of Barcelona, Barcelona, Spain, <sup>11</sup>Infection in Immunocompromised Pediatric Patients Research Group, Vall d'Hebron Institut de Recerca (VHIR), Hospital Universitari Vall d'Hebron, 08035 Barcelona, Spain, <sup>12</sup>Jeffrey Modell Diagnostic and Research Center for Primary Immunodeficiencies, Barcelona, Spain, <sup>13</sup>Department of Cell Biology, Physiology and Immunology, Universitat Autònoma de Barcelona (UAB), Bellaterra, Catalonia, Spain, <sup>14</sup>Laboratory of Human Lymphohematopoiesis, Imagine Institute, INSERM UMR 1163, Université de Paris, Paris, France, <sup>15</sup>German Center for Infection Research (DZIF), Satellite Center Freiburg, Germany, <sup>16</sup>Centre for Integrative Biological Signalling Studies (CIBSS), Albert-Ludwigs University, Freiburg, Germany, <sup>17</sup>RESIST, Cluster of Excellence 2155 to Hanover Medical School, Satellite Center Freiburg, Germany and <sup>18</sup>Institute of Immunity & Transplantation, Royal Free Hospital, University College London, UK

Received February 07, 2020; Revised April 15, 2021; Editorial Decision April 16, 2021; Accepted April 19, 2021

## ABSTRACT

Activation-induced deaminase (AID) initiates antibody diversification in germinal center B cells by deaminating cytosines, leading to somatic hypermutation and class-switch recombination. Loss-of-function mutations in AID lead to hyper-IgM syndrome type 2 (HIGM2), a rare human primary anti-

body deficiency. AID-mediated deamination has been proposed as leading to active demethylation of 5-methylcytosines in the DNA, although evidence both supports and casts doubt on such a role. In this study, using whole-genome bisulfite sequencing of HIGM2 B cells, we investigated direct AID involvement in active DNA demethylation. HIGM2 naïve and memory B cells both display widespread DNA methy-

\*To whom correspondence should be addressed. Tel: +34 935572800; Fax: +34 934651472; Email: eballestar@carrerasresearch.org

†The authors wish it to be known that, in their opinion, the second and third authors should be regarded as Joint Second Authors.

Present address: Christian Klemann, Department of Paediatric Pulmonology, Allergy and Neonatology, Hannover Medical School, Hannover, Germany.

lation alterations, of which ~25% are attributable to active DNA demethylation. For genes that undergo active demethylation that is impaired in HIGM2 individuals, our analysis indicates that AID is not directly involved. We demonstrate that the widespread alterations in the DNA methylation and expression profiles of HIGM2 naïve B cells result from premature overstimulation of the B-cell receptor prior to the germinal center reaction. Our data support a role for AID in B cell central tolerance in preventing the expansion of autoreactive cell clones, affecting the correct establishment of DNA methylation patterns.

## INTRODUCTION

Hyper-IgM syndrome type 2 (HIGM2) is a rare primary antibody deficiency, with autosomal recessive inheritance, characterized by loss-of-function mutations in activation-induced deaminase (AID) (1), an enzyme required for several crucial steps of B cell terminal differentiation. AID converts deoxycytosines (dCs) into deoxyuracils (dUs), producing dU:dG mismatches that are removed by mismatch repair and base-excision repair (2). Deaminase activity is required for somatic hypermutation (SHM) and class-switch recombination (CSR) of immunoglobulin (Ig) genes, which are necessary processes for affinity maturation and antibody diversification within the germinal centers (GC) (3,4). AID deficiency results in the absence of CSR and SHM, and leads to lymphoid hyperplasia (1). HIGM2 patients have normal or elevated serum IgM levels with severe reduction of IgG, IgA and IgE, resulting in considerable susceptibility to bacterial infections (1).

In addition to its role in CSR and SHM, AID has been proposed to participate in active DNA demethylation through deamination of 5-methylcytosine (5mC), leading to a mismatch that is converted to G:C by thymine DNA glycosylase (TDG), followed by base-excision repair. The potential role of AID in active DNA demethylation was first proposed by Petersen-Mahrt and colleagues (5). Initial studies in this topic mainly focusing on non-lymphoid cells, such as including zebrafish embryos and heterokaryon-based reprogramming, supported that model (6–8). These studies subsequently led to its study in B cells, given that activated B cells display the highest levels of AID expression (9). During the past decade, conflicting reports have both supported and discounted a role in active demethylation for AID in that context [reviewed in (10)]. For instance, Fritz *et al.* (11) performed reduced-representation bisulfite sequencing (RRBS) of mouse splenic naïve B cells from wild type and AID-deficient mice, activated *ex vivo* for 72 h and found no significant differences in their DNA methylation profiles. Similar conclusions were obtained looking at GC B cells from AID-deficient mice and using MethylCap-Seq (12). However, more recently, Dominguez and colleagues showed that the transit of B cells through the GC is associated with marked locus-specific loss of methylation and increased methylation diversity, both of which are lost in *Aicda*<sup>−/−</sup> animals (13). Methodological aspects could explain the discrepancies between these studies, including the limited coverage and resolution of the three aforementioned

analyses, or the study of different cells (*in vitro* activated or isolated GC B cells).

On the other hand, different *in vitro* studies suggest that 5mC is a poorer substrate than C, although early studies showed that human AID can deaminate 5mC (5). For instance, Abdouni and colleagues showed that the efficient deamination of 5mC by zebrafish AID is due to the flexibility of its structure, in comparison with that of other AID orthologs, including human AID (14). In addition, the comparison of AID with other members of the AID/APOBEC family have shown that human AID deaminates 5mC only weakly because the 5-methyl group fits poorly in its DNA-binding pocket (15). Larijani *et al.* have shown that methylated-CpG motifs, but not their unmethylated counterparts, are in fact protected from AID-mediated deamination (16). Nabel and colleagues have also shown that steric requirements for cytosine deamination are one intrinsic barrier to the proposed function of deaminases in DNA demethylation (17).

In parallel, the discovery of alternative enzymatic pathways that lead to *bona fide* active DNA demethylation through ten-eleven translocation methylcytosine dioxygenase (TET)-mediated oxidation of methylcytosines (18,19) raised more doubts about the possibility that AID redundantly plays such a role. There is currently no consensus about whether AID is involved in mediating DNA demethylation in specific cell contexts.

Whole-genome analysis has shown the occurrence of a vast amount of demethylation associated with B cell differentiation. Changes occur mostly during naïve B cell activation, yielding memory B cells (20,21) that coincide with the highest peak of AID expression (3). Naïve B cells start to proliferate upon activation by antigen encounter. Then they express AID which triggers the secondary diversification of antibodies by SHM and CSR. This is followed by affinity maturation which finally leads to (a) a new cycle of SHM or (b) terminal differentiation into memory or plasma B cells depending on the affinity of the B cell receptor (BCR) for the cognate antigen (22). It has been originally reported that AID is targeted to Ig genes with the involvement of RNA polymerase II (23). Nevertheless, it has since been described that AID is also recruited to non-Ig genes to mediate recurrent mutations and these genes are collectively termed ‘AID off-targets’ (24,25).

In this study, we took advantage of the exceptional possibility to investigate the direct role of AID in active demethylation by comparing the complete DNA methylomes of naïve and memory B cells of HIGM2 patients with those of healthy individuals. By studying two sibling patients with a homozygous mutation for AID that results in a severely truncated enzyme, we were able to determine its direct link with DNA methylation defects and infer its catalytic activity in relation to active DNA demethylation.

Our results show that the absence of AID catalytic activity affects DNA methylation in naïve and memory B cells. The majority of the changes occurring in the transition from naïve to memory B cells arise from passive demethylation and are linked to late-replicating domains. However, for those potentially associated with active demethylation, we found no evidence of direct involvement of AID, and our analysis suggests that TET enzymes are responsi-

ble for DNA methylation changes in this cell context. The increased DNA demethylation noted in naïve B cells of HIGM2 patients is associated with premature demethylation of BCR downstream genes prior to the GC reaction. Indeed, we found that these changes are related to the expansion of autoreactive clones, which suggests a major role for AID in preventing the expansion of such clones under normal conditions.

## MATERIALS AND METHODS

### Human samples

The patients included in the study fulfilled the diagnostic criteria for hyper-IgM syndrome type 2, based on ESID clinical diagnostic criteria (26) and genetic confirmation of *AICDA* mutation and exclusion of other primary and secondary causes of immunodeficiencies. Samples come from the Medical Center of the University Hospital, University of Freiburg, Freiburg, Germany and Hospital Universitari Vall d'Hebron, Barcelona, Spain. The Committees for Human Subjects of the local hospitals approved the study, which was conducted in accordance with the ethical guidelines of the 1975 Declaration of Helsinki. All samples were in compliance with the guidelines approved by the local ethics committee and all donors (and/or their parents) received oral and written information about the possibility that their blood would be used for research purposes.

**Isolation of B cell populations.** Peripheral blood mononuclear cells (PBMCs) were obtained from blood. After Ficoll-Isopaque density centrifugation (Rafer, Zaragoza, Spain), collected cells were washed twice with ice-cold PBS, followed by centrifugation at 2000 rpm for 5 min. Next, cells were labeled with antibodies to CD19 – FITC (Miltenyi Biotec, clone LT19), CD27 – APC (Miltenyi Biotec, clone M-T271), IgD – PE (SouthernBiotech, Cat. No. 2032-09) and IgM – PerCP/Cy5.5 (BioLegend, clone MHM-88) for 20 min on ice in a staining buffer (PBS with 4% FBS and 2 mM EDTA). Naïve B cells (CD19<sup>+</sup> CD27<sup>−</sup> IgD<sup>+</sup>) and unswitched memory B cells (CD19<sup>+</sup> CD27<sup>+</sup> IgD<sup>+</sup>) were obtained by FACS sorting on a MoFlo Astrios (Beckman Coulter). Purified samples were pelleted and stored at −80°C.

For isolation of naïve autoreactive B cells. Total B cells were isolated from PBMCs using positive selection with MACS CD19 microbeads (Miltenyi Biotec). Next, cells were stained with CD27-APC (Miltenyi Biotec, clone M-T271), IgD – PE (SouthernBiotech, Cat. No. 2032-09), HLA-DR – PE-Cy7 (eBioscience, clone LN3), 9g4 primary ab (igm Bioscience) and donkey anti-rat IgG (H + L) – Alexa Fluor 488 (invitrogen). 9g4<sup>+</sup> naïve B cells (CD27<sup>−</sup> IgD<sup>+</sup> 9g4<sup>+</sup>) and 9g4<sup>−</sup> naïve B cells (CD27<sup>−</sup> IgD<sup>+</sup> 9g4<sup>−</sup>) were obtained by FACS sorting on a BD FACSARIA II (BD Biosciences). Purified samples were pelleted and stored at −80°C.

**Genomic DNA extraction.** For whole-genome bisulfite sequencing, DNA was extracted with a QIAamp DNA micro kit (Qiagen) according to the manufacturer's protocol. For pyrosequencing analysis (detailed below), DNA was

extracted with a Maxwell RSC Cultured Cells DNA kit (Promega), following the manufacturer's protocol.

**Tagmentation-based whole-genome bisulfite sequencing.** For whole-genome bisulfite sequencing, 30 ng of genomic DNA was used to produce four independent barcoded sequencing libraries per DNA sample using the tagmentation method (27). Sequencing of the TWGBS libraries was done on a HiSeq 2000, PE 125 bp mode. Bisulfite sequencing reads were processed by the DKFZ bisulfite analysis workflow. In brief, the reads were trimmed using Trimmomatic, pre-processed and aligned using MethylTools, with default parameters (V. Hovestadt, S. Picelli, B. Radlwimmer, M.Z. and P.L., unpublished data), which uses the Burrows-Wheeler alignment algorithm (28). Following quality control of bisulfite conversion (>99.5% in all samples) and of read-mapping (80–90% could be mapped on average), we performed methylation calling using methylTools. A summary of the sequencing data for each sample is provided in Supplementary Table S1.

**Differentially methylated region (DMR) calling.** DMRs were detected with the DeNovoDMR algorithm included in the Specific Methylation Analysis and Report Tool (SMART2) (29) using all the default parameters except for the number of CpGs per segment, which was set to 4, the absolute mean methylation difference, which was set to 0.2, and a threshold value of *P* of 0.01. Only those CpGs with a coverage of ≥5 in all samples were considered in the construct of the SMART input matrix. DMR calling was performed on all possible comparisons between naïve and memory B cells for both control and HIGM2 patients.

**Bisulfite pyrosequencing.** 500 ng of genomic DNA was converted with an EZ DNA Methylation-Gold kit (Zymo Research), following the manufacturer's instructions. Bisulfite-treated DNA was PCR amplified using primers (see Supplementary Table S2) designed with PyroMark Assay Design 2.0 software (Qiagen). Finally, PCR amplicons were pyrosequenced with the PyroMark Q24 system and analyzed with PyroMark CpG software (Qiagen).

### Processing of chromatin immunoprecipitation (ChIP) assays with sequencing (ChIP-Seq) datasets

Sequencing reads from ChIP-seq experiments from the BLUEPRINT consortium (21) were mapped to the hg19 assembly of human reference genome using Burrows-Wheeler Aligner (BWA) v0.7.13 (with parameters -q 5, -l 32, -k 2). After removing reads with MAPQ <30 with Sequence Alignment/Map (SAMtools) v1.2, PCR duplicates were eliminated using the Picard function available in MarkDuplicates software v1.126. Peak calling was performed using macs2 (with parameters -p 1e-2 -nomodel -shift 0 -B -SPMR). Only peaks with an overlap of ≥0.5 between replicates were considered. Histone mark signals around DMR sets were extracted with the annotatePeaks.pl algorithm available in Hypergeometric Optimization of Motif Enrichment (HOMER) software v4.10.3 (with parameters: size = 10 000, hist = 10).



**Super-enhancer identification.** H3K27ac ChIP-seq data from BLUEPRINT database were used to identify the super-enhancer regions, as described previously (30) using Rank-Ordering of Super-Enhancers (ROSE) software. An enhancer stitching distance of 15 kb was used along with a 2.5 kb transcriptional start site (TSS)-exclusion window.

**DNA methylation data analysis.** Hierarchical clustering was carried out based on Pearson correlation distance metrics and average linkage criteria. For low-dimensional analysis, we used the *t*-distributed stochastic neighbor embedding (*t*-SNE) method implemented in the Rtsne v0.15 package.

Transcription factor motifs were enriched for each set of DMRs using HOMER software v4.10.3. Specifically, we used findMotifsGenome.pl algorithm (with parameters -size given -cpg) to search for significant enrichment against a background sequence adjusted to have similar CpG and GC contents.

Transcription factor binding analysis was performed interrogating the overlap between the different sets of DMRs with ChIP-seq data for transcription factors available for GM12878 lymphoblastoid cell line from the ENCODE Project (31). The enrichment factor was calculated against random regions as a background, and *P* values were calculated using Fisher's exact test. Finally, the transcription factors downstream of the BCR signaling pathway were manually annotated from a curated database (32).

Chromatin states and histone mark enrichments analysis for NBC, GC B cells and ncsMBC were assessed using a custom adaptation of the EpiAnnotator R package (33) using BLUEPRINT data (21). DMRs were converted to hg38 assembly with the liftOver function in the rtracklayer v1.42 R package.

Replication timing data in the GM12878 lymphoblastoid cell line were obtained from the UW Repli-seq track of the UCSC Genome Browser. Genomic replication timing values were binned in deciles where the first decile contained the regions of latest replication and the last decile the regions of earliest replication. Identified DMRs were then overlapped to these regions by genomic location.

DMR annotation for genetic context location was performed using the annotatePeaks.pl algorithm in the HOMER software v4.10.3. To determine the location relative to a CpG island (CGI), we used 'hg19.cpgs' annotation in the annotatr v1.8 R package.

GREAT software (34) was used to enrich downstream pathways and gene ontologies. We used the single nearest gene option for the association between genomic regions with genes.

Genes obtained with HOMER annotation software were used to study correlations between their associated DMRs and gene expression data from healthy naïve, memory and GC B cells from BLUEPRINT database, as well as with the AID off-target genes dataset obtained from Álvarez-Prado *et al.* (35).

All statistical analysis (excluding T-WGBS and ChIP-seq analyses) were done in R v3.5.1. Data distributions were tested for normality. Normal data were tested using two-tailed unpaired Student's *t*-tests; non-normal data were analyzed with the appropriate non-parametric

statistical test. Levels of significance are indicated as: \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001; \*\*\*\**P* < 0.0001. Non-significance (*P* ≥ 0.05) was indicated as 'ns'.

**Public RRBS of B cell activation.** Data of EBV and CD40/IL4 B cell activation were downloaded from the NCBI Gene Expression Omnibus (GSE49629) (36). Methylation calls from RRBS data were filtered so that only those CpGs with a minimum of five reads per position in all samples were retained. Since RRBS genomic coverage is significantly lower than T-WGBS we only tested the methylation status of positions common to two datasets.

**BCR activation and EBV infection.** For naïve B cell activation and infection experiments, we obtained buffy coats from anonymous donors through the Catalan Blood and Tissue Bank (CBTB). The CBTB follows the principles of the World Medical Association (WMA) Declaration of Helsinki. Before providing the first blood sample, all donors received detailed oral and written information and signed a consent form at the CBTB. PBMCs were isolated using Ficoll-Paque gradient centrifugation. Total B cells were isolated from PBMCs using positive selection with MACS CD19 microbeads (Miltenyi Biotec). Next, cells were stained with CD27-APC (Miltenyi Biotec, clone M-T271) and IgD – PE (SouthernBiotech, Cat. No. 2032–09) and naïve B cells were sorted as CD27–IgD<sup>+</sup>. For EBV infection, pure naïve B cells were incubated with B95–8 cell supernatant for 3 h at 37°C in order to infect them with EBV, they were harvested after 30 days. For BCR activation, pure naïve B cells were cultured in RPMI medium 1640 GlutaMAX™-1 (Gibco, Life Technologies) containing 20% fetal bovine serum (Gibco, Life Technologies), 1% sodium pyruvate and 1% Penicillin Streptomycin (Gibco Thermo Fisher Scientific, MA, USA). Cells were activated with 10 µg/ml anti-IgM (SouthernBiotech), 0.1 µg/ml MEGACD40L<sup>®</sup> Protein (Enzo) and 50 ng/ml IL21 (Tebu-bio) for 24 h (ChIP) or 7 days (bisulfite pyrosequencing).

**RNA-sequencing.** RNA-sequencing libraries from total mRNA were prepared and purified using Illumina Stranded mRNA Prep, Ligation kit (Illumina). Sequencing was performed on a NextSeq500 (Illumina) with a NextSeq 500/550 High Output Kit v2.5 (75 Cycles, Illumina, 20024906) to generate 38-nucleotide paired-end reads at a read depth of at least 20 million reads per sample.

**RNA-seq data analysis.** For bioinformatics analyses, raw sequence reads were initially processed using FastQC (Babraham Institute, Cambridge, UK) for quality control, and then adapter sequences and poor quality reads were removed using TrimGalore that uses Cutadapt for trimming. Quality-filtered reads were then mapped to the human genome (hg19) using STAR, and only the uniquely mapped reads were kept. Read counts were calculated using HTSeq-count. Differentially expressed genes were identified using R package DESeq2 (fold change ≥ 2 and adjusted *P*-value < 0.05). Inference of TF activities from expression values were calculated using DoRothEA (37). Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to perform functional enrichment analysis of differentially expressed genes.

**ChIP-PCR.** Cell-sorter isolated naïve B cells from healthy donors were activated as described above for 24 h.  $1 \times 10^6$  unstimulated and stimulated B cells were crosslinked with 1% methanol-free formaldehyde (Thermo Fisher) for 10 min subjected to 15 min of sonication utilizing Covaris M220 Focused-ultrasonicator (Covaris, MA, USA). ChIP experiments were performed using the LowCell# ChIP kit<sup>TM</sup> protein A, according to manufacturer's instructions (Digenode, Liege, Belgium). Anti-BATF antibody (#8638) and corresponding rabbit IgG (#2729) were both obtained from Cell Signaling. Protein binding was analyzed by real-time quantitative PCR, and data are represented as a ratio of the enriched fraction with respect to input. ChIP primers are shown in Supplementary Table S2.

## RESULTS

### Study strategy

We obtained peripheral blood from two sibling HIGM2 patients, both with the same homozygous mutation for the *AICDA* gene, and two sex-matched healthy controls. Specifically, the patients carried a deletion (Exon 2 c.22.40del19) that generates a frameshift variant (p.Arg8Asnfs\*19) that affects the DNA binding region of AID located within its nuclear localization signal domain (Figure 1A). This deletion impairs both CSR and SHM (38). We confirmed the impact on CSR by inspecting the peripheral B cell compartment by flow cytometry. The two HIGM2 patients were characterized by the absence of class-switched memory B cells (csMBC; CD19<sup>+</sup>CD27<sup>+</sup>IgM<sup>+</sup>IgD<sup>-</sup>), as described (1) (Figure 1B). In contrast, classic non-class switched memory B cells (ncsMBC; CD19<sup>+</sup>CD27<sup>+</sup>IgM<sup>+</sup>IgD<sup>+</sup>) and naïve B cells (NBC; CD19<sup>+</sup>CD27<sup>-</sup>IgM<sup>+</sup>IgD<sup>+</sup>) were present in patients (Figure 1B). It has been shown that, under physiological conditions, ncsMBC cells display certain levels of SHM at the immunoglobulin locus (20), which supports the expression of AID during their maturation in GCs. Hence, the comparison between DNA methylation profiles of NBC and ncsMBC isolated from healthy and HIGM2 individuals is an adequate model to investigate the potential role of AID in DNA demethylation.

### HIGM2 patients display an aberrant methylation profile in naïve and unswitched memory B cells

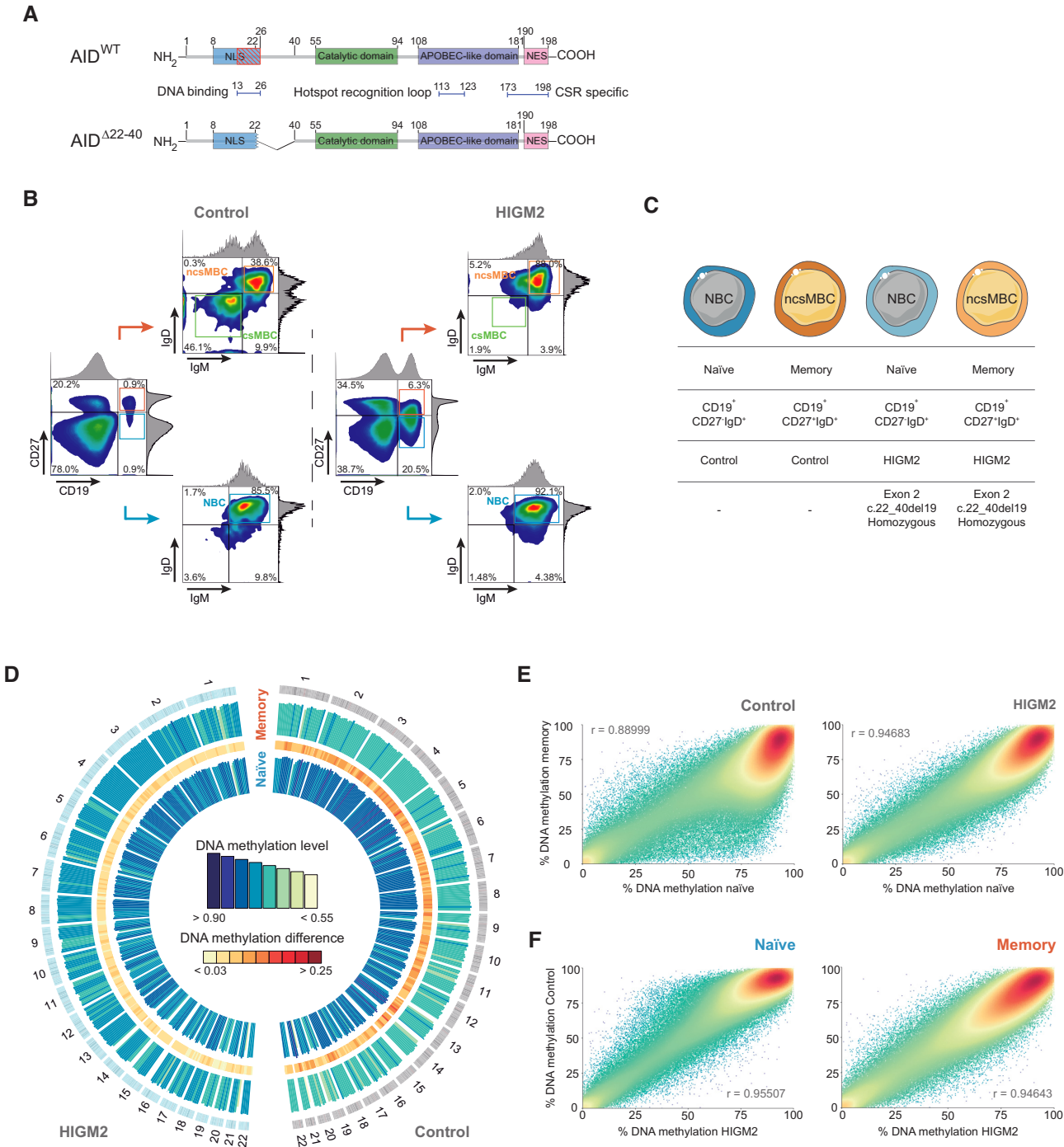
We performed tagmentation-based whole-genome bisulfite sequencing (T-WGBS), a version of the WGBS method that allows analysis of limited DNA amounts (27), in NBC and ncsMBC (henceforth referred to as 'naïve' and 'memory' B cells) isolated from two HIGM2 siblings and two sex-matched healthy controls (Figure 1C). Pearson correlation and *t*-distribution stochastic neighbor embedding (t-SNE) between samples were highly reproducible between replicates with correlation coefficient to be  $>0.9$  (Supplementary Figure S1A, B). We also compared our DNA methylation data from healthy controls with public data from the International Cancer Genome Consortium (ICGC) (39) and Oakes *et al.* (20), in which we observed high correlation (Pearson correlation  $>0.85$ ), thereby confirming the robustness of our data (Supplementary Figure S1C, D).

Global inspection of DNA methylation confirmed previous studies (20,21) that the transition from naïve to memory B cells is accompanied by global demethylation of the genome in healthy controls (Figure 1D). However, HIGM2 patients showed a partial impairment of demethylation during naïve-to-memory B cell differentiation (Figure 1D, E), which is compatible with a potential role of AID as a demethylating enzyme. The transition from naïve to memory B cells in healthy controls is characterized by the demethylation of 30175 DMRs. In contrast, comparing naïve to memory B cells in HIGM2 patients only identified 4803 hypomethylated DMRs (Supplementary Table S3). Furthermore, we also observed that naïve B cells were more demethylated in HIGM2 patients than in healthy controls (a total of 2936 hypomethylated DMRs) (Figure 1F). Taken together, these global observations suggest that AID loss not only affects the DNA methylation patterns in the transition from naïve to memory B cells but also has a potential role in establishing the B cell DNA methylome in earlier stages of development.

### A high proportion of DNA demethylation events identified in HIGM2 are due to passive demethylation of late-replicating domains

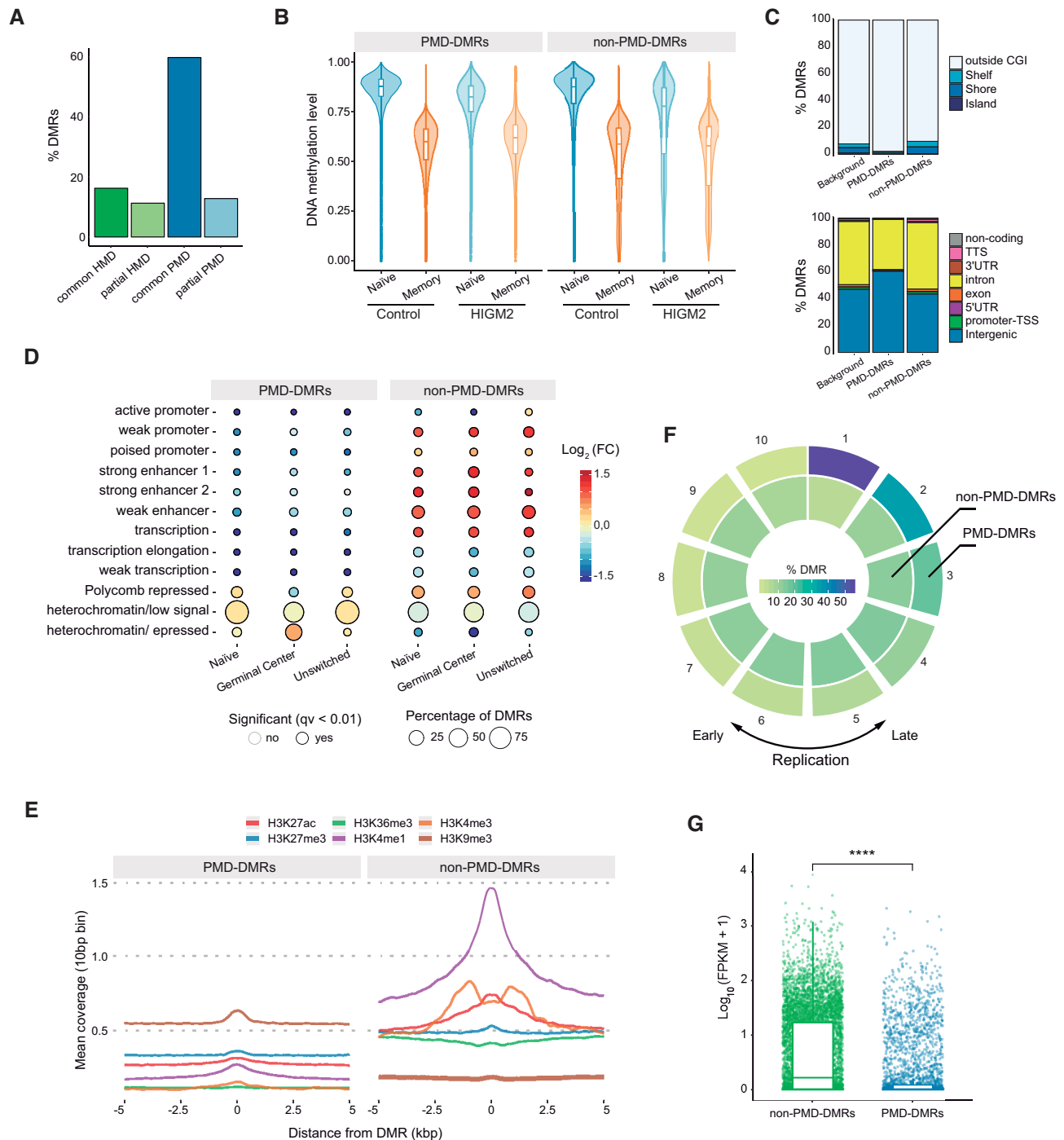
Recent studies have shown that a high proportion of the demethylation events occurring in cancer and in differentiation processes are associated with high proliferation rates (40). Such demethylation events take place in regions known as 'partially methylated domains' (PMDs), which are characterized by repressive chromatin and low-GC density (40), and do not appear to occur in regions of high methylation (HMD: highly methylated domains). PMDs encompass regions that undergo late replication and their demethylation is a passive event, which is a result of inefficient DNA remethylation during DNA replication (36,40,41). Recent reanalysis of the B cell lineage DNA methylation profiles published by the BLUEPRINT consortium (21) has shown the occurrence of demethylation of PMDs in the transition towards memory B cells and antibody-secreting plasma cells (42). This highlights the importance of separating DNA methylation analysis into PMD and non-PMD regions when interrogating the occurrence of active demethylation processes in order to exclude those changes due to DNA replication-dependent or passive demethylation.

To address this matter, we first identified total differentially methylated regions (DMRs) by performing pairwise comparisons of DNA methylation between all samples. Specifically, we compared naïve to memory B cells in both control and HIGM2 patients (control\_NBC versus control\_ncsMBC, HIGM2\_NBC versus HIGM2\_ncsMBC), as well as comparing HIGM2 patients to controls for both naïve and memory B cells (control\_NBC versus HIGM2\_NBC, control\_ncsMBC versus HIGM2\_ncsMBC). We then divided the DMRs into PMD and HMD regions, which were previously identified by Zhou *et al.* (40) to be present either in all cell types (common) or cell-type specific (partial). We found that the majority of DMRs overlapped with PMDs that were either common or partially shared by cell types (72.5%, Figure 2A). We then classified DMRs into two groups: the first com-



**Figure 1.** DNA methylomes of HIGM2 B cell subpopulations determined by T-WGBS. (A) Graphical representation of wild type (WT) *AICDA* gene and truncated AID of HIGM2 patients used in this study. The mutation in HIGM2 AID harbors a deletion from 22–40 which encompasses the DNA binding domain. (B) Representative examples of cell sorting strategy of B cell populations (naïve B cells, NBC; classic non-class-switched memory B cells, ncsMBC; class-switched memory B cells, csMBC). (C) Description of the B cell subpopulations analyzed, including surface markers and details of mutation. (D) Circos plot of DNA methylation of naïve B cells (inner circle) and unswitched memory B cells (outer circle) for controls (right) and HIGM2 patients (left). Histogram tracks represent the average methylation levels of 10 Mb windows. Heatmap shows the DNA methylation differences between naïve and memory B cells. (E) Density scatter plots of average DNA methylation in windows of 500 bp in naïve and memory B cells in controls (left) and HIGM2 patients (right). (F) Density scatter plots of average DNA methylation in windows of 500 bp comparing HIGM2 and control for naïve (left) and memory (right) B cells.





**Figure 2.** Detection and characterization of partially methylated regions. (A) Bar plot showing the percentage of DMRs in PMD and HMD (partially and highly methylated domains, respectively) regions annotated by Zhou and colleagues (40). 'Common' refers to PMD and HMD regions common to all cell types analyzed by Zhou and colleagues. 'Partial' refers to those regions detected only in a fraction of the cell types analyzed. (B) Box and violin plots summarizing the distribution of DNA methylation levels for controls and HIGM2 patients, in which the DMRs were characterized as inside (PMD-DMRs) and outside (non-PMD-DMRs) of PMDs. (C) Location proportions of PMD-DMRs and non-PMD-DMRs in the context of CpG islands (CGIs) and gene-related regions. (D) Bubble chart depicting the enrichment (red) or depletion (blue) of chromatin states defined from ChromHMM ChIP-seq data of healthy naive, GC and unswitched memory B cells. Color scale represents the logarithmic fold change, circle size indicates the percentage of DMRs in the chromatin state, and the edge color indicates the statistical significance of the enrichment (black: significant, none: not significant;  $q < 0.01$ ). (E) ChIP-seq data of the represented histone marks in healthy GC B cells were downloaded from the BLUEPRINT database and mean coverage was calculated for 10bp bins  $\pm 5$  kb centered around the midpoints of PMD-DMRs (left) and non-PMD-DMRs (right). (F) Circular representation of the proportion of PMD- (outer circle) and non-PMD-DMRs (inner circle) in deciles of replication timing data of GM12878 cell line. Color scale represents the proportion of DMRs in each decile. The first and last deciles correspond to the regions of latest and earliest replication, respectively. (G) Box and dot plots showing the distribution of expression values of associated genes of PMD- and non-PMD-DMRs. FPKM refers to fragments per kb of transcript per million mapped reads. Expression data from healthy GC B cells downloaded from BLUEPRINT database. Statistical tests: two-tailed Fisher's exact (D) and unpaired Wilcoxon's (G) tests (\*\*\*\*  $P$ -value  $< 0.0001$ ).

prising DMRs that coincided with common PMDs (PMD-DMRs), and the second containing DMRs that did not overlap with common PMDs (non-PMD-DMRs). To validate the identity of these PMD-DMRs, inspection of the methylation values showed that the two groups of DMRs had intermediate methylation values in memory B cells, although non-PMD-DMRs had slightly lower methylation levels (Figure 2B). Utilizing functional genomic feature analysis, we found that most PMD-DMRs were located outside of CpG islands and in intergenic regions (Figure 2C). In relation to chromatin states, we observed that DMRs occurring in PMDs were mainly associated with heterochromatic regions, while non-PMD-DMRs were highly enriched at enhancers and active promoters, in concordance to what was previously described as generally associated with replication-independent demethylation (43) (Figure 2D). This was also confirmed by overlapping DMRs with ChIP-seq data of healthy GC B cells obtained from the BLUEPRINT database, in which non-PMD-DMRs were enriched in enhancer marks, including H3K4me1, H3K4me3 and H3K27ac (Figure 2E). It has been previously described that PMDs are characterized by their association with late-replication domains (40), hence we analyzed DMRs based on their association with late or early phase replication. First, genomic density data obtained from GM12878 Repli-seq (UW Repli-seq, UCSC Genome Browser) were divided into deciles by their replication timing and then overlapped by genomic location with identified DMRs. We confirmed that PMD-DMRs were mainly found in late-replicating regions (Figure 2F) and were accompanied by lower expression of associated genes in GC B cells (Figure 2G). Taking all these observations into account, our results suggest that most of the DNA methylation changes occur in *bona fide* PMDs. However, the existence of a set of non-PMD-DMRs (~27%) located in highly active regions suggests the potential participation of DNA replication-independent or, in other words, active demethylation events, and AID could directly drive these events.

### HIGM2-associated defects in DNA methylation in the transition from naïve to memory B cells do not have the features of AID targets

Given all the previous considerations, including the removal of DNA methylation changes related to DNA replication (PMD-DMRs), our model allows us to examine whether AID has a direct role in mediating demethylation in the B cell lineage. In this context, GC B cells, in the transition from naïve to memory, displayed the highest *AICDA* mRNA levels (Supplementary Figure S2A) (as obtained from the BLUEPRINT database). Hence, it is plausible that defective AID in HIGM2 patients could play a role in aberrant DNA demethylation occurring during the transition from naïve to memory B cells within the GC.

To explore this possibility, we first identified DMRs whose DNA demethylation could potentially be driven by AID (P-AID DMRs). These DMRs are demethylated in the transition from naïve to memory B cells in healthy controls but not in HIGM2 patients. Furthermore, DMRs that were already aberrantly demethylated in naïve HIGM2 B cells compared to controls were excluded (Figure 3A). A total of

522 DMRs (containing 450 different genes) fulfilled these conditions (Supplementary Table S4).

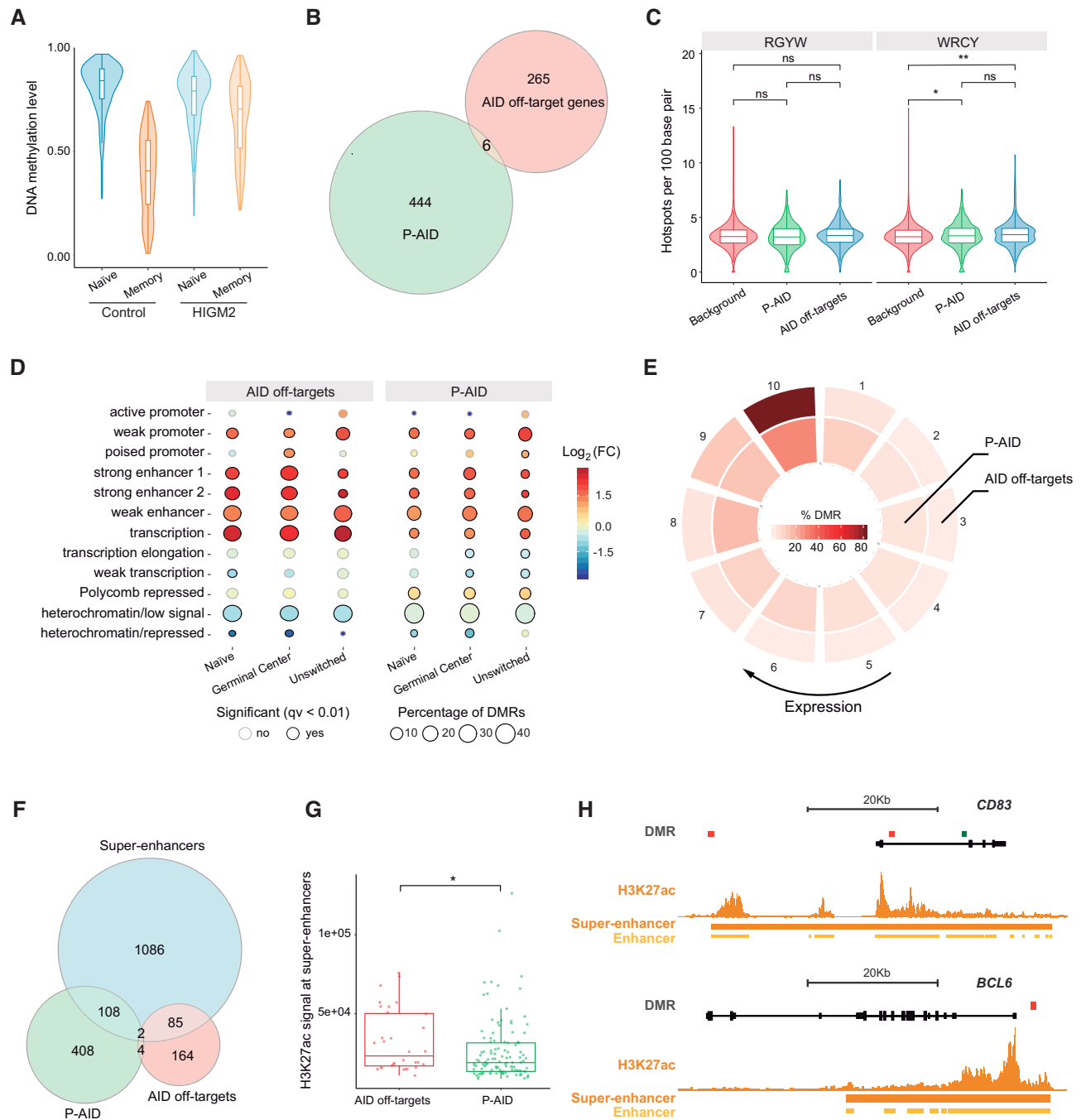
We then compared P-AID DMRs associated genes with data obtained from a mouse model that allows the identification of targets of AID deamination events (35), in which 271 non-Ig genes were identified to be dependent on AID catalytic activity and referred to as AID off-target genes. We found low correspondence between these two sets of genes in which only 6 genes (corresponding to 6 DMRs) overlap with the identified AID off-target genes (Figure 3B). We also tested for the presence of described AID hotspot sequences (RGYW and WRCY) (44) in our DMRs and found a significant increase for the WRCY hotspot with respect to the background, although it appeared too low (<1 hotspot per 100 bp) to be of biological relevance (Figure 3C). On the other hand, we found that DMRs associated with AID off-target genes underwent demethylation in HIGM2 patients (Supplementary Figure S2B).

Two recent studies have characterized the genomic and epigenomic features of AID off-target regions in mice. Independently, they found that AID targets regions with convergent transcription from intragenic super-enhancers (45,46). In this sense, there were no significant differences regarding gene localization between the DMRs associated with AID off-targets and P-AID DMRs (Supplementary Figure S2C and Supplementary Table S4). However, AID off-target DMRs exhibited greater enrichment of enhancer regions, as observed by both ChromHMM (Figure 3D) and histone mark (Supplementary Figure S2D) analyses, and were associated with more transcriptional activity of associated genes in GC B cells (Figure 3E). We also found that, although the two DMR groups had a similar percentage of overlap with super-enhancers (P-AID 21%, AID off-target 33.3%; Figure 3F), the super-enhancers of the AID off-targets had a stronger signal for H3K27ac (Figure 3G, H; Supplementary Figure S2E) and greater transcriptional activity of their associated genes than P-AID (Supplementary Figure S2F). Finally, we hypothesized that if AID had a role mediating active DNA demethylation, we would expect to see differences in CpG sites containing WRCY hotspots at the super-enhancers of AID off-target genes. However, such differences were not observed (Supplementary Figure S2G).

Taken together, our results suggest that, despite the differences in DNA methylation associated with B cell activation between wild type and AID-deficient B cells, if we assume that mechanisms and targets are conserved between mouse and human, such demethylation may not be directly associated with AID catalytic activity.

While the proposed mechanism of DNA demethylation by AID implies that the 5mC conversion in thymine could be repaired and replaced with an unmethylated cytosine (5), the removal of methyl groups of cytosines mediated by TET enzymes involves the generation of oxidation intermediates, including 5hmC (18) (Supplementary Figure S3A). It is therefore plausible that AID-dependent demethylation would associate with lower levels of 5hmC compared to TET-dependent demethylation. Hence, we analyzed publicly-available DNA methylation data obtained from GC and naïve B cells of *Aicda*<sup>-/-</sup> mice (13). We selected three sets of CpGs: (i) CpGs demethylated in the transition





**Figure 3.** Indirect involvement of AID in DNA demethylation dynamics during B cell activation. (A) Box and violin plot representation of DNA methylation of DMRs that are potentially demethylated directly by AID (P-AID). These DMRs were filtered by the presence of demethylation between control naïve and memory B cells which were not present in HIGM2 B cells. Furthermore, DMRs that displayed demethylation in HIGM2 naïve B cells compared to controls were excluded. (B) Venn diagram showing the overlap between associated genes of P-AID DMRs and the human orthologues of mouse AID off-target genes as defined by Álvarez-Prado *et al.* (35). (C) Box and violin plots displaying the frequency of WRCY/RGYW (W = A/T; R = G/A; Y = C/T) hotspots per 100 bp in DMRs of each subset. Background was generated as 100bp windows of the whole genome. (D) Bubble plot depicting the enrichment (red) or depletion (blue) of chromatin states. Color represents the logarithmic-fold change, size shows the percentage of DMRs in the chromatin state, and the edge indicates the statistical significance of the enrichment (black: significant, none: not significant,  $q$ -value < 0.01). Chromatin state data was obtained from ChromHMM, generated in GC B cells. (E) Circular representation of proportion of DMRs in deciles of genes whose expression is ranked from low to high expression in GC B cells. The first and last deciles correspond to the least and most highly expressed genes, respectively. Color scale represents the proportion of DMRs in each decile. Gene expression data was obtained from the BLUEPRINT database. (F) Venn diagram of the overlap between P-AID and AID off-target DMRs with super-enhancers in GC B cells. Super-enhancer identification is described in the Materials section utilizing H3K27ac ChIP-seq data from the BLUEPRINT database. (G) Box plots showing H3K27ac ChIP-seq peaks of super-enhancers associated with AID off-target (red) and P-AID (green) DMRs. One-sided unpaired Wilcoxon's test was used to examine signal intensity differences. (H) DNA methylation and H3K27ac profiles in the vicinity of two representative genes. DMR color indicates the type of DMR, either P-AID (green) or AID off-target DMRs (red). H3K27ac ChIP-seq signals from GC B cells are shown in dark orange. Super-enhancers and enhancers are depicted. Statistical tests: two-tailed unpaired Wilcoxon's (C) and Fisher's exact (D) tests (\*  $P$ -value < 0.05, \*\*  $P$ -value < 0.01; ns is not significant).

from naïve to GC B cells in WT but not in *Aicda*<sup>-/-</sup> mice, termed mouse potential AID targets (mouse P-AID); (ii) CpGs significantly demethylated in naïve-to-GC transition in both WT and knockout mice and, therefore, more likely to be TET-dependent, termed positive control; (iii) CpGs that did not present demethylation in the transition from naïve to GC B cells neither in WT nor in *Aicda*<sup>-/-</sup> mice, termed negative control (differential DNA methylation < 0.05) (Supplementary Figure S3B). Next, we merged this data with public hydroxy-meDIP-seq data of mouse B cell activation (47) to determine the hydroxymethylation status of these three groups of CpGs. We found that mouse P-AID CpGs presented similar levels of hydroxymethylation to those that are TET-dependent (positive control) compared to negative control (Supplementary Figure S3C). On the other hand, we only found 21 genes associated with mouse P-AID CpGs that overlapped with the aforementioned AID off-target genes (18), which represents <8% of the AID-off target genes (Supplementary Figure S3D). Taken together, these results suggest that, at least in a mouse model, the majority of demethylation associated with B cell differentiation may not be directly driven by AID activity.

#### AID deficiency results in premature demethylation of the BCR pathway of naïve B cells

Our initial analysis suggested that alterations in DNA methylation not only occurred in the transition from naïve to memory B cells, but were already present in naïve B cells of HIGM2 individuals in comparison with healthy controls. Specifically, HIGM2 naïve B cells appeared to be more demethylated than those of healthy controls (Figures 1F and 4A). AID expression has customarily been associated with the GC reaction (3), however, more recent evidence suggests that AID may have a role in earlier stages of B cell development (48,49).

Comparing naïve B cells from HIGM2 to control, we detected 2152 hypomethylated DMRs (Figure 4A) and 127 hypermethylated DMRs (Supplementary Figure S4A), excluding PMDs. Both hypo- and hypermethylated DMRs were mostly found outside CpG islands in intergenic regions and introns (Supplementary Figure S4B). Genetic annotation of DMRs in CpG islands reveals an association with promoters (Supplementary Figure S4C and Supplementary Table S5). However, while hypomethylation was associated with enhancer regions, hypermethylation was mainly enriched in promoters (Supplementary Figure S4D), in agreement with its reported regulatory role in the 'spurious' initiation of transcription (50).

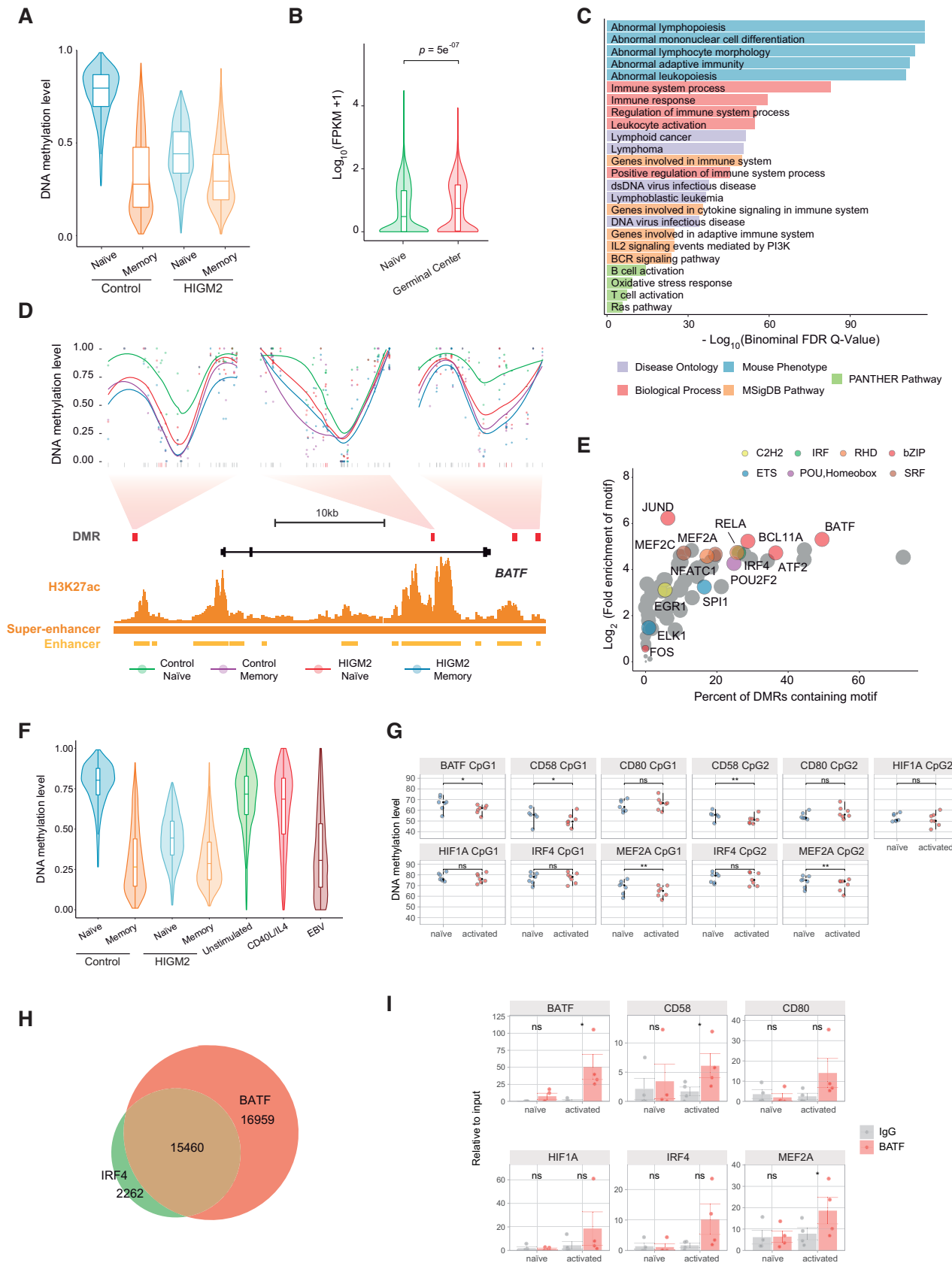
We observed that the DMRs that were hypomethylated in naïve B cells from HIGM2 patients underwent normal demethylation during the transition from naïve to memory cells in healthy controls (Figure 4A). Our results suggest that the demethylation of these DMRs may be essential for correct naïve-to-memory B cell differentiation, and premature demethylation observed in HIGM2 naïve B cells may indicate aberrant pre-activation outside of the GC. This is consistent with the finding that genes associated with these DMRs became upregulated during the activation of B cells in the GC (Figure 4B) and were associated with functional categories related to B cell activation via BCR (Figure 4C).

Some genes that have altered DMRs, such as *BATF* (Figure 4D) and *MEF2A* (Supplementary Figure S4E), are crucial to B cell development (47,51).

To explore the possibility that aberrant DNA methylation in HIGM2 naïve B cells may be due to pre-activation outside the GC, we first analyzed the enrichment for TF binding motifs with HOMER software in HIGM2 naïve DMRs. Some of the most enriched TFs are downstream of the BCR pathway, including *BATF*, *IRF8* and *RELA* (32) (Supplementary Figure S4F). We then validated these results through enrichment analysis of the ChIP-seq data available for GM12878 cells from the ENCODE consortium (31) and we again observed enrichment of TFs downstream of BCR activation in DMRs found to be altered in naïve B cells of HIGM2 patients (Figure 4E).

The type III latency state of the Epstein-Barr virus (EBV) is characterized by the constitutive activation of the BCR and CD40 pathways (52,53), both of which are major signaling pathways essential for B cell activation (54). In that sense, the B lymphoblastoid cell line GM12878 presents a type III latency state (55) and is therefore a good model to investigate the involvement of BCR/CD40 in aberrant DNA methylation of naïve B cells in HIGM2 patients. Firstly, we represented DNA methylation levels of publicly-available data of EBV-transformed and CD40L/IL-4-activated B cells (36) in regions that overlapped with DMRs identified in naïve B cells of HIGM2 patients and observed that EBV transformation effectively reproduced the aberrant DNA demethylation presented in HIGM2 patients (Figure 4F). Conversely, such changes did not take place when activation was performed with CD40/IL-4 alone, suggesting that the BCR pathway has a significant role in driving aberrant DNA methylation (Figure 4F). Furthermore, these DNA methylation effects were not reproduced in hypermethylated DMRs (Supplementary Figure S5A). To confirm these results, we transformed naïve B cells isolated from healthy donors with EBV and, after 30 days, we interrogated DNA methylation of the set of CpGs that we had previously identified to be aberrantly demethylated in HIGM2. We observed that these CpGs underwent demethylation following EBV-mediated transformation of naïve B cells (Supplementary Figure S5B). Additionally, we directly stimulated BCR in isolated healthy naïve B cells and observed similar demethylation of the same CpGs, which supported the direct role of BCR in their demethylation (Figure 4G). Taken together, our findings suggest that alterations in DNA methylation of HIGM2 naïve B cells may at least be partly due to the aberrant pre-activation of the BCR at some point of B cell development prior to the GC reaction.

Since ChIP-seq analysis in GM12878 cells identified *BATF* to bind to almost 50% of regions that correspond to DMRs in HIGM2 naïve B cells (Figure 4E), we therefore hypothesize that *BATF* might be the main mediator of demethylation in naïve-to-memory transition, and its aberrant recruitment may drive the aberrant demethylation observed in HIGM2 naïve B cells. *BATF*, together with *IRF4* (56,57), is a regulator of B and T cell activation via the BCR and TCR pathways respectively. We observed that 87% of the ChIP-seq peaks of *IRF4* overlapped with *BATF* peaks in GM12878 (Figure 4H). As ex-



**Figure 4.** Role of AID in early B cell development. (A) Box and violin plot representation of DNA methylation of hypomethylated DMRs identified comparing control and patient naïve B cells. (B) Box plot showing the distribution of expression values of DMR-associated genes in naïve (green) and GC



pected, we confirmed that regions with BATF and IRF4 binding had lower DNA methylation levels in naïve B cells of HIGM2 patients compared to controls (Supplementary Figure S5C). This is also confirmed for regions that contained JUND binding, another TF downstream of BCR (Supplementary Figure S5D). This did not occur in regions enriched for other B cell-intrinsic TF binding motifs, such as RAD21, ZNF274 and MYC (Supplementary Figure S5D). Using mRNA transcription data from *IRF4* and *BATF* knockouts from GM12878 cells (55), we determined that genes with binding motifs for both TFs displayed expression changes for many of these genes associated with DMRs (Supplementary Figure S5E). Furthermore, we observed that BATF was specifically recruited to genes that displayed defective DNA demethylation in HIGM2 patients when naïve B cells were activated via BCR stimulation (Figure 4I). Altogether, these results suggest that a significant fraction of DMRs identified in HIGM2 naïve B cells may be associated with the recruitment of the BATF/IRF4 complex to these genomic sites.

### The transcriptome of AID-deficient naïve B cells also reflects alterations in the BCR pathway

We then inspected whether loss of function of AID had an impact on the transcriptomic profiles of both naïve and memory B cells. To this end, we performed RNA-seq of both naïve and memory B cells from the same two aforementioned HIGM2 patients. In parallel, we performed RNAseq with two healthy controls.

Comparison of the naïve B cells transcriptomes revealed statistically significant expression changes of a number of genes in HIGM2 compared to controls (DEGs: differentially expressed genes). DEGs were considered to be statistically significant when fold change  $\geq 2$  and adjusted *P*-value  $< 0.05$ . 3223 genes became upregulated in HIGM2 naïve B cells in relation to healthy controls, whereas 3146 genes were downregulated. (Figure 5A and Supplementary Table S6). We then performed Gene Ontology analysis of DEGs using GREAT software which revealed enrichment of categories related to EBV activation and BCR signaling pathway (Figure 5B), consistent with our previous findings related to DNA methylation alterations in naïve B cells of HIGM2. Interestingly, we also observed differences in gene expression in memory B cells (2199 upregulated genes, 2384 downregulated) (Supplementary Figure S6A and Supplementary Table S6) which were also enriched in the aforemen-

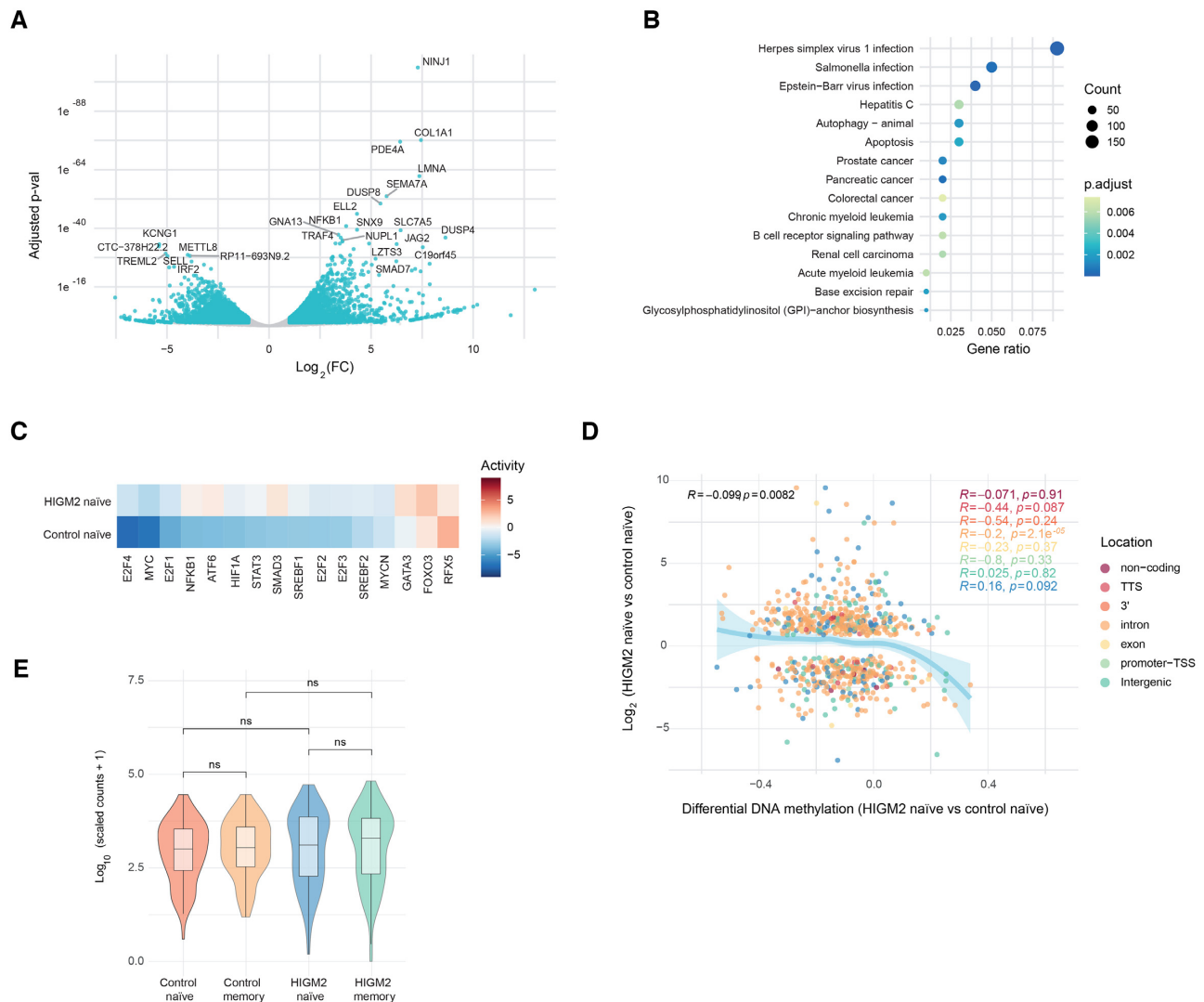
tioned categories (Supplementary Figure S6B). Utilizing DoRothEA TF analysis, we observed that DEGs identified in naïve B cells of HIGM2 patients also displayed increased TF activity compared to controls (Figure 5C), where similar results were observed for memory B cells (Supplementary Figure S6C). Integration of the DNA methylation and expression datasets in naïve B cells did not show a clear correlation when comparing HIGM2 versus controls (Figure 5D). Interestingly, hypomethylated DMRs in HIGM2 naïve B cells showed slightly increased expression levels in HIGM2 naïve B cells in respect to control naïve B cells. The naïve B cell expression levels of hypomethylated DMRs of HIGM2 patients were more similar to those of memory B cells, reinforcing the notion of a premature activation of these cells in HIGM2 (Figure 5E).

### AID deficiency causes blockade of central B cell tolerance with an expansion of pre-activated autoreactive B cells

Our results suggest that naïve B cells are pre-activated in AID deficient patients. However, the stage of B cell differentiation at which this alteration is produced remains to be established. Two independent studies reported a potential role for AID in removing autoreactive B cells during the central B cell tolerance process in the bone marrow (48,49). Specifically, immature B cells with auto-reactive BCR were activated and went into a secondary receptor editing process with an increase in AID and recombination-activating gene 2 (*RAG2*). If the autoreactive BCR did not lose self-antigen affinity, the genomic instability induced by the over-exposure to high levels of AID led to apoptosis. However, AID deficiency reduces the genomic damage that causes the expansion of autoreactive B cells (48,49). With that in mind, we hypothesized that the presence of naïve B cells with a pre-activation methylation signature in HIGM2 patients is a consequence of the impairment of the central B cell tolerance mechanism that causes autoreactive naïve B cells to accumulate. In fact, it has been reported that 21% of HIGM2 patients suffer some type of autoimmune disease (58).

To assess this hypothesis, we first checked whether there is an expansion of naïve autoreactive B cells in HIGM2 patients with respect to controls. To this end, we used a commercial antibody against 9G4<sup>+</sup> IgG used to detect autoreactive clones in autoimmune diseases like systemic lupus erythematosus and rheumatoid arthritis (59,60). We observed an expansion of the naïve B cell compartment in HIGM2 patients in comparison with healthy controls (Fig-

(red) B cells. Gene expression data obtained from BLUEPRINT database. FPKM refers to fragments per kilobase of transcript per million mapped reads. (C) Results of gene set enrichment analysis using GREAT software. The plot depicts the top five enriched terms for five annotation databases ranked by *P*-values from the binominal distribution. All the depicted enriched terms passed a significance threshold of adjusted *P*-value  $< 0.05$ . (D) Smoothed DNA methylation data of *BATF*-associated DMRs in HIGM2 naïve B cells and H3K27ac ChIP-seq signal, super-enhancer and enhancer location in GC B cells. H3K27ac ChIP-seq data obtained from BLUEPRINT database. (E) Bubble scatter plot of transcription factor ChIP-seq peaks determined in GM12878 lymphoblastoid cell line in regions that correspond to DMRs identified in HIGM2 naïve B cells. Transcription factors downstream of BCR signaling are colored according to the transcription factor family. Bubble size corresponds to the logarithm of adjusted *P*-values. ChIP-seq data were obtained from ENCODE consortium (31). FDR refers to false discovery rate. (F) Box and violin plots summarizing the distribution of DNA methylation of regions that correspond to HIGM2-identified DMRs in resting B cells (Unstimulated), B cells activated with CD40L/IL4 and B cells infected with Epstein-Barr virus (EBV). Data of resting and activated B cells obtained from Hansen *et al.* (36) (G) Dot plot showing the DNA methylation values determined by pyrosequencing of naïve B cells (naïve) and B cells activated with anti-BCR/CD40L/IL-21 (activated) from healthy donors. (H) Venn diagram representing the overlap between BATF and IRF4 ChIP-seq peaks in the GM12878 cell line. (I) Dot and bar plot representation of BATF ChIP-qPCR results of naïve B cells (naïve) and naïve B cells activated with anti-BCR/CD40L/IL-21 (activated) from healthy donors. Statistical tests: Student's *t*-test (G, I) (\* *P*-value  $< 0.05$ , \*\* *P*-value  $< 0.01$ , ns is not significant).



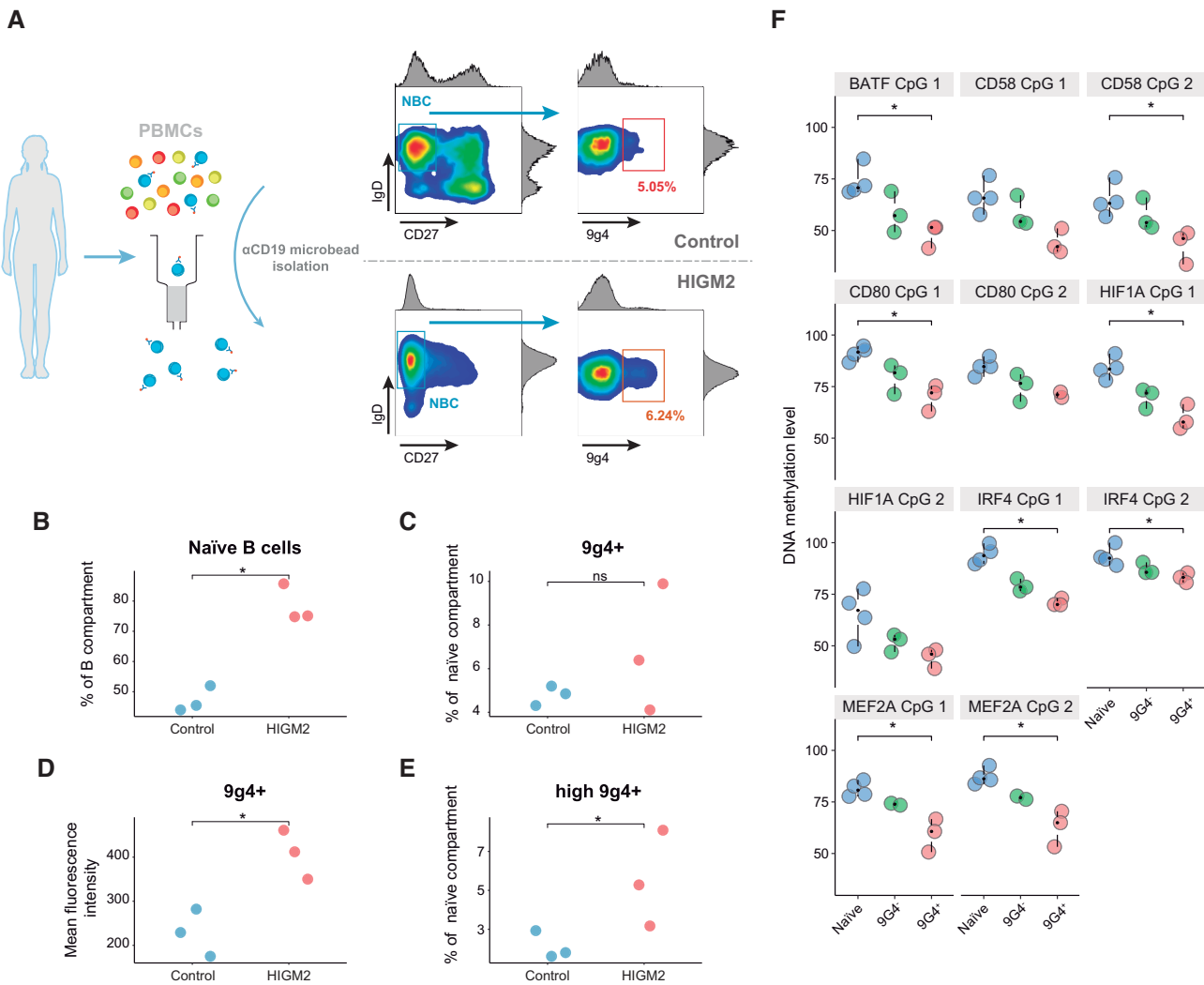
**Figure 5.** RNA sequencing analysis of naïve B cells transcriptomes from HIGM2 patients and healthy donors. (A) Volcano plot representation of the transcriptomic comparison of HIGM2 with control naïve B cells. In blue, genes that are differentially expressed (absolute of fold change  $\geq 2$  and adjusted  $P$ -value  $< 0.05$ ). (B) Gene Ontology analysis representation of differentially expressed genes. All represented categories are statistically significantly enriched (adjusted  $P$ -value  $< 0.05$ ). Gene ratio refers to the ratio between the unique genes present in the DMRs that are associated with a Gene Ontology term and all the genes potentially included in that ontology gene set. Dot size represents the number of times each term is represented in the DMRs. Dot color represents adjusted  $P$ -value. (C) Heatmap representation showing TF activity predicted using DoRothEA from mRNA expression of differentially expressed genes in control and HIGM2 naïve B cells. Color represents normalized enrichment scores (NES) where blue and red represent a decreased and increased activity, respectively, in comparison to background. (D) Correlation between DNA methylation differences and gene expression scaled counts ratio in naïve B cells from HIGM2 and controls. Colors represent genetic context location. (E) Box and violin plot representation of expression values per sample group from the genes associated with hypomethylated DMRs in HIGM2 naïve B cells. Statistical tests: Spearman's rank correlation test (D) and two-tailed unpaired Wilcoxon's (E) (ns is non significant).

ure 6A, B). We did not find an expansion of  $9G4^+$  cells in HIGM2 patients with respect to controls (Figure 6C). However, we observed a significant increase of mean fluorescence intensity for  $9g4$  staining (Figure 6A, D), as well as, an expansion of high  $9g4^+$  naïve B cells (Figure 6E). Next, we determined the methylation status of a selection of genes by pyrosequencing of  $9G4^-$  (non-autoreactive) and  $9G4^+$  (autoreactive) naïve B cells and found that autoreactive B cells had lower levels of DNA methylation than their non-autoreactive counterparts (Figure 6F). Overall, our results suggest that the demethylation in naïve B cells of HIGM2 patients compared with control donors is associated with

an expansion of pre-activated autoreactive naïve B cells as a consequence of central B cell tolerance impairment mediated by AID deficiency.

## DISCUSSION

Our results show that AID deficiency in HIGM2 syndrome results in the acquisition of aberrant DNA methylation profiles in naïve and memory B cells. First, we observed the presence of HIGM2-associated DNA methylation alterations occurring in the transition from naïve to memory B cell in which the majority are due to DNA replication-



**Figure 6.** Central B cell tolerance impairment by AID deficiency. (A) Representative example of strategy for sorting naïve autoreactive B cells. (B) Dot plot showing the percentage of naïve B cells in the B cell compartment of HIGM2 patients and controls. (C) Dot plot showing the percentage of autoreactive naïve B cells in the naïve B cell compartment of HIGM2 patients and controls. (D) Dot plot showing the mean fluorescence of 9g4<sup>+</sup> gate. (E) Dot plot showing the percentage of high 9g4<sup>+</sup> naïve B cells in the naïve B cell compartment of HIGM2 patients and controls. (F) Dot plot showing the DNA methylation values determined by pyrosequencing of control naïve B cells (naïve), HIGM2 patient 9g4<sup>-</sup> naïve B cells (9g4<sup>-</sup>) and HIGM2 patient 9g4<sup>+</sup> naïve B cells (9g4<sup>+</sup>). Statistical tests: student t-test (\* *P*-value < 0.05, ns is not significant).

mediated mechanisms. Second, compared to AID mouse models, we observed that AID may not be directly involved in the small proportion of those DNA methylation alterations attributable to active demethylation. Third, the comparison of naïve B cells in HIGM2 and healthy controls showed premature demethylation of genes downstream of the BCR in AID-deficient individuals, which was associated with the expansion of autoreactive B cell clones prior to the GC reaction. This reinforces a novel role for AID in preventing the expansion of autoreactive B cell clones, affecting the DNA methylation profiles of naïve B cells.

Our study indicates that AID does not play a direct role in mediating active demethylation in the transition from naïve to memory B cells. This transition is associated with the highest proportion of DNA methylation changes of the entire B cell differentiation process (20,21) and also coincides with the highest peak of AID expression. Previous studies

addressing the potential participation of AID in demethylation had not considered late-replicating domains or the relationship between DNA methylation changes and the genomic features of AID targets. In our study, we determined that most of the changes taking place during the transition from naïve to memory B cells occur through passive demethylation. It is possible that memory B cells of healthy controls undergo more divisions than HIGM2 patients, similarly to what has been described for *Aicda*<sup>-/-</sup> mice (61). This could explain the partial impairment of passive demethylation in HIGM2 patients. For active changes, no associations with AID targets were found. These findings are also in line with those by Álvarez-Prado *et al.* (35), which indicated that AID-mediated mutation frequencies are too low. Such low frequency would be unlikely to produce a perceptible effect at the level of DNA methylation. In relation to the results obtained by Dominguez and col-



leagues (13), it is remarkable that even without the removal of PMDs in their analysis, the overlap between the DNA methylation changes and AID-associated dsDNA breaks was very small, which again reinforces the notion that AID may not be directly involved in DNA demethylation events of B cell maturation. A limitation of our analysis was the lack of effective human models to study the direct contribution of AID activity to active demethylation. Therefore, we cannot discard the possibility that the use of mouse models may not be directly attributable to the functions of human AID in the context of HIGM2. However, comparing *Aicda*<sup>-/-</sup> mouse model with a mouse model deficient for BER and MMR pathways (*Ung*<sup>-/-</sup> *Msh2*<sup>-/-</sup>), we were able to deduce that DNA methylation during B cell maturation, at least in mice, was not dependent on AID enzymatic activity. Furthermore, there is a 92% identity between mouse and human AID protein sequences, in which the catalytic and hotspot recognition domains are conserved.

Another limitation of our analysis is that we work under the assumption that all deaminase activity is attributed to AID and that no other targets exist apart from AID off-targets and Ig genes. Without further evidence, we cannot effectively conclude that AID enzymatic activity is not essential to mediate DNA demethylation during B cell maturation. Moreover, AID hotspots were originally identified as genomic locations of SHM (29), but further experiments would be required to verify that they are direct targets of AID deaminase activity in the context of human B cells.

A second major conclusion of our study concerns the identification of DNA methylation defects in naïve B cells from HIGM2 patients in relation to healthy controls. Customarily, AID expression has been regarded as being restricted to GC B cells, but some evidence suggests that AID may also have a role in central B cell tolerance (48,49,62). Along this line, during B cell development, these cells not only become activated in the GC but also in previous stages of differentiation in the bone marrow, where self-reactive immature B cells are activated in a process characterized by the upregulation of both AID and recombination-activating gene 2 (RAG2) and the downregulation of the anti-apoptotic MCL-1 (63). In this context, AID activity increases the probability of genomic damage with the subsequent activation of apoptosis through p53, which is also enhanced by the inhibition of the anti-apoptotic proteins BCL2 and MCL-1 (49). At this point, self-reactive immature B cells that are unable to correct their affinity for self-antigens by receptor editing are eliminated. In patients with AID deficiency, this mechanism of cell removal is impaired and autoreactive cells accumulate (49). Indeed, we noted that HIGM2 patients accumulated more autoreactive B cells than healthy donors, a finding that is compatible with the previously described high frequency of autoimmune disorders in this type of patients (58). The failure in AID function in these patients could be responsible for the smaller degree of genomic damage that promotes the expansion of autoreactive naïve B cells. These self-reactive B cells, owing to the persistent activation of their BCR during negative selection in the bone marrow, display a more demethylated profile in genes downstream of the BCR compared with non-autoreactive naïve B cells. However, we have only replicated this data in a selected number of CpGs and

cannot assume that the pre-activation of autoreactive naïve B cells in HIGM2 accounts for all the observed alterations in genome-wide DNA methylation. It is possible that other mechanisms may actively contribute to this phenomenon. Our results, therefore, indicate that the enhanced demethylation of BCR downstream targets in HIGM2 naïve B cells may be the result of the expansion of autoreactive B cell clones as a consequence of the absence of AID. Furthermore, in order to generalize our findings to all variants of HIGM2 it would be important to extend these analyses to a bigger cohort of patients with different genetic mutations of the *AICDA* gene (38), as they can have a different impact on B cell function.

## DATA AVAILABILITY

DNA methylation and expression data for this publication have been deposited in the NCBI Gene Expression Omnibus and are accessible through GEO Series accession number GSE172012 (SuperSeries).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank CERCA Programme/Generalitat de Catalunya and the Josep Carreras Foundation for institutional support. This study makes use of data generated by the BLUEPRINT Consortium (Funding for the BLUEPRINT project provided by the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement no 282510—BLUEPRINT). A full list of the investigators who contributed to the generation of the data is available from [www.blueprint-epigenome.eu](http://www.blueprint-epigenome.eu). We are also very grateful to Dr Javier di Noia, Dr Almudena Ramiro, Dr Javier Carmona and Dr Biola M. Javierre for useful feedback.

## FUNDING

Spanish Ministry of Science, Innovation and Universities [SAF2017-88086-R to E.B.]; cofunded by FEDER funds/European Regional Development Fund (ERDF)—a way to build Europe. E.B. is supported by Instituto de Salud Carlos III (ISCIII), Ref. AC18/00057, associated with i-PAD project (ERARE European Union program); P.L. and C.P. are supported by the German Cancer Aid project CO-CLL [70113869]; B.G. is funded by the Deutsche Forschungsgemeinschaft [GR1617/14-1/iPAD, SFB1160/2.B5, RESIST-EXC 2155—Project ID 390874280, CIBSS-EXC-2189—Project ID 390939984]; BMBF [GAIN 01GM1910A]. Funding for open access charge: Spanish Ministry of Science, Innovation and Universities [SAF2017-88086-R].

*Conflict of interest statement.* None declared.

## REFERENCES

1. Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A. *et al.*

- (2000) Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell*, **102**, 565–575.
2. Honjo, T., Kinoshita, K. and Muramatsu, M. (2002) Molecular mechanism of class switch recombination: linkage with somatic hypermutation. *Annu. Rev. Immunol.*, **20**, 165–196.
  3. Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y. and Honjo, T. (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, **102**, 553–563.
  4. Arakawa, H., Hauschi, L.D. and Buerstedde, J.M. (2002) Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science*, **295**, 1301–1306.
  5. Morgan, H.D., Dean, W., Coker, H.A., Reik, W. and Petersen-Mahrt, S.K. (2004) Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: Implications for epigenetic reprogramming. *J. Biol. Chem.*, **279**, 52353–52360.
  6. Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A. and Cairns, B.R. (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and Gadd45. *Cell*, **135**, 1201–1212.
  7. Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y. and Blau, H.M. (2010) Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature*, **463**, 1042–1047.
  8. Sabag, O., Zamir, A., Keshet, I., Hecht, M., Ludwig, G., Tabib, A., Moss, J. and Cedar, H. (2014) Establishment of methylation patterns in ES cells. *Nat. Struct. Mol. Biol.*, **21**, 110–112.
  9. Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O. and Honjo, T. (1999) Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.*, **274**, 18470–18476.
  10. Ramiro, A.R. and Barreto, V.M. (2015) Activation-induced cytidine deaminase and active cytidine demethylation. *Trends Biochem. Sci.*, **40**, 172–181.
  11. Fritz, E.L., Rosenberg, B.R., Lay, K., Mihailović, A., Tuschl, T. and Papavasiliou, F.N. (2013) A comprehensive analysis of the effects of the deaminase AID on the transcriptome and methylome of activated B cells. *Nat. Immunol.*, **14**, 749–755.
  12. Hogenbirk, M.A., Heideman, M.R., Velds, A., van den Berk, P.C.M., Kerkhoven, R.M., van Steensel, B. and Jacobs, H. (2013) Differential programming of B cells in AID deficient mice. *PLoS One*, **8**, e69815.
  13. Dominguez, P.M., Teater, M., Chambwe, N., Kormaksson, M., Redmond, D., Ishii, J., Vuong, B., Chaudhuri, J., Melnick, A., Vasanthakumar, A. et al. (2015) DNA Methylation dynamics of germinal center B cells are mediated by AID. *Cell Rep.*, **12**, 2086–2098.
  14. Abdouni, H., King, J.J., Suliman, M., Quinlan, M., Fifield, H. and Larijani, M. (2013) Zebrafish AID is capable of deaminating methylated deoxycytidines. *Nucleic Acids Res.*, **41**, 5457–5468.
  15. Wijesinghe, P. and Bhagwat, A.S. (2012) Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic Acids Res.*, **40**, 9206–9217.
  16. Larijani, M., Frieder, D., Sonbuchner, T.M., Bransteitter, R., Goodman, M.F., Bouhassira, E.E., Scharff, M.D. and Martin, A. (2005) Methylation protects cytidines from AID-mediated deamination. *Mol. Immunol.*, **42**, 599–604.
  17. Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y. and Kohli, R.M. (2012) AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.*, **8**, 751–758.
  18. Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* (80-.), **324**, 930–935.
  19. Ito, S., Dalessio, A.C., Taranova, O. V., Hong, K., Sowers, L.C. and Zhang, Y. (2010) Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, **466**, 1129–1133.
  20. Oakes, C.C., Seifert, M., Assenov, Y., Gu, L., Przekopowicz, M., Ruppert, A.S., Wang, Q., Imbusch, C.D., Serva, A., Koser, S.D. et al. (2016) DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat. Genet.*, **48**, 253–264.
  21. Kulis, M., Merkel, A., Heath, S., Queirós, A.C., Schuyler, R.P., Castellano, G., Beekman, R., Raineri, E., Esteve, A., Clot, G. et al. (2015) Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat. Genet.*, **47**, 746–756.
  22. Vitoria, G.D. and Nussenzweig, M.C. (2012) Germinal Centers. *Annu. Rev. Immunol.*, **12**, 117–139.
  23. Storb, U. (2014) Why does somatic hypermutation by AID require transcription of its target genes? In: *Advances in Immunology*. Academic Press Inc., Vol. **122**, pp. 253–277.
  24. Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J. et al. (2011) Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell*, **147**, 107–119.
  25. Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F. et al. (2011) Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell*, **147**, 95–106.
  26. Seidel, M.G., Kindle, G., Gathmann, B., Quinti, I., Buckland, M., van Montfrans, J., Scheible, R., Rusch, S., Gasteiger, L.M., Grimbacher, B. et al. (2019) The European Society for Immunodeficiencies (ESID) registry working definitions for the clinical diagnosis of inborn errors of immunity. *J. Allergy Clin. Immunol. Pract.*, **7**, 1763–1770.
  27. Wang, Q., Gu, L., Adey, A., Radlwimmer, B., Wang, W., Hovestadt, V., Bähr, M., Wolf, S., Shendure, J., Eils, R. et al. (2013) Tagmentation-based whole-genome bisulfite sequencing. *Nat. Protoc.*, **8**, 2022–2032.
  28. Li, H. and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, **26**, 589–595.
  29. Liu, H., Liu, X., Zhang, S., Lv, J., Li, S., Shang, S., Jia, S., Wei, Y., Wang, F., Su, J. et al. (2016) Systematic identification and annotation of human methylation marks based on bisulfite sequencing methylomes reveals distinct roles of cell type-specific hypomethylation in the regulation of cell identity genes. *Nucleic Acids Res.*, **44**, 75–94.
  30. Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I. and Young, R.A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307–319.
  31. Feingold, E.A., Good, P.J., Guyer, M.S., Kamholz, S., Liefer, L., Wetterstrand, K., Collins, F.S., Gingeras, T.R., Kampa, D., Sekinger, E.A. et al. (2004) The ENCODE (ENCyclopedia of DNA Elements) Project. *Science*, **306**, 636–640.
  32. Kandasamy, K., Sujatha Mohan, S., Raju, R., Keerthikumar, S., Sameer Kumar, G.S., Venugopal, A.K., Telikicherla, D., Navarro, D.J., Mathivanan, S., Pecquet, C. et al. (2010) NetPath: a public resource of curated signal transduction pathways. *Genome Biol.*, **11**, R3.
  33. Pageaud, Y., Plass, C. and Assenov, Y. (2018) Enrichment analysis with EpiAnnotator. *Bioinformatics*, **34**, 1781–1783.
  34. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M. and Bejerano, G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
  35. Álvarez-Prado, Á.F., Pérez-Durán, P., Pérez-García, A., Benguria, A., Torroja, C., de Yébenes, V.G. and Ramiro, A.R. (2018) A broad atlas of somatic hypermutation allows prediction of activation-induced deaminase targets. *J. Exp. Med.*, **215**, 761–771.
  36. Hansen, K.D., Sabuncian, S., Langmead, B., Nagy, N., Curley, R., Klein, G., Klein, E., Salamon, D. and Feinberg, A.P. (2014) Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization. *Genome Res.*, **24**, 177–184.
  37. Garcia-Alonso, L., Iorio, F., Matchan, A., Fonseca, N., Jaaks, P., Peat, G., Pignatelli, M., Falcone, F., Benes, C.H., Dunham, I. et al. (2018) Transcription factor activities enhance markers of drug sensitivity in cancer. *Cancer Res.*, **78**, 769–780.
  38. Durandy, A., Peron, S., Taubenheim, N. and Fischer, A. (2006) Activation-induced cytidine deaminase: structure-function relationship as based on the study of mutants. *Hum. Mutat.*, **27**, 1185–1191.
  39. Hudson, T.J., Anderson, W., Aretz, A., Barker, A.D., Bell, C., Bernabé, R.R., Bhan, M.K., Calvo, F., Eerola, I., Gerhard, D.S. et al.

- (2010) International network of cancer genome projects. *Nature*, **464**, 993–998.
40. Zhou, W., Dinh, H.Q., Ramjan, Z., Weisenberger, D.J., Nicolet, C.M., Shen, H., Laird, P.W. and Berman, B.P. (2018) DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.*, **50**, 591–602.
  41. Lister, R., Pelizzola, M., Downen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M.M. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.
  42. Durek, P., Nordström, K., Gasparoni, G., Salhab, A., Kressler, C., de Almeida, M., Bassler, K., Ulas, T., Schmidt, F., Xiong, J. *et al.* (2016) Epigenomic profiling of human CD4<sup>+</sup> T cells supports a linear differentiation model and highlights molecular regulators of memory development. *Immunity*, **45**, 1148–1161.
  43. Wu, X. and Zhang, Y. (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.*, **18**, 517–534.
  44. Rogozin, I.B. and Kolchanov, N.A. (1992) Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *BBA - Gene Struct. Expr.*, **171**, 11–18.
  45. Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K.R., Resch, W., Liang, G., Tang, Z., Mathé, E., Benner, C. *et al.* (2014) B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell*, **159**, 1524–1537.
  46. Meng, F.L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberg, D. *et al.* (2014) Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell*, **159**, 1538–1548.
  47. Lio, C.W.J., Shukla, V., Samaniego-Castruita, D., González-Avalos, E., Chakraborty, A., Yue, X., Schatz, D.G., Ay, F. and Rao, A. (2019) TET enzymes augment activation-induced deaminase (AID) expression via 5-hydroxymethylcytosine modifications at the Aicda superenhancer. *Sci. Immunol.*, **4**, eaau7523.
  48. Kuraoka, M., Holl, T.M., Liao, D., Womble, M., Cain, D.W., Reynolds, A.E., Kelsoe, G., Kuraoka, M., Cain, D.W., Womble, M. *et al.* (2011) Activation-induced cytidine deaminase mediates central tolerance in B cells. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11560–11565.
  49. Cantaert, T., Schickel, J.-N.N., Bannock, J.M.M., Ng, Y.-S.S., Massad, C., Oe, T., Wu, R., Lavoie, A., Walter, J.E.E., Notarangelo, L.D.D. *et al.* (2015) Activation-induced cytidine deaminase expression in human B cell precursors is essential for central B cell tolerance. *Immunity*, **43**, 884–895.
  50. Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F. and Oliviero, S. (2017) Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, **543**, 72–77.
  51. Herglotz, J., Unrau, L., Hauschildt, F., Fischer, M., Kriebitzsch, N., Alawi, M., Indenbirken, D., Spohn, M., Müller, U., Ziegler, M. *et al.* (2016) Essential control of early B-cell development by Mef2 transcription factors. *Blood*, **127**, 572–581.
  52. Cahir-McFarland, E.D., Carter, K., Rosenwald, A., Giltman, J.M., Henrickson, S.E., Staudt, L.M. and Kieff, E. (2004) Role of NF- $\kappa$ B in cell survival and transcription of latent membrane protein 1-expressing or Epstein-Barr virus latency III-infected cells. *J. Virol.*, **78**, 4108–4119.
  53. Mancao, C. and Hammerschmidt, W. (2007) Epstein-Barr virus latent membrane protein 2A is a B-cell receptor mimic and essential for B-cell survival. *Blood*, **110**, 3715–3721.
  54. Luo, W., Weisel, F. and Shlomchik, M.J. (2018) B cell receptor and CD40 signaling are rewired for synergistic induction of the c-Myc transcription factor in germinal center B cells. *Immunity*, **48**, 313–326.
  55. Ma, Y., Walsh, M.J., Bernhardt, K., Ashbaugh, C.W., Trudeau, S.J., Ashbaugh, I.Y., Jiang, S., Jiang, C., Zhao, B., Root, D.E. *et al.* (2017) CRISPR/Cas9 screens reveal Epstein-Barr virus-transformed B cell host dependency factors. *Cell Host Microbe*, **21**, 580–591.
  56. Ise, W., Kohyama, M., Schraml, B.U., Zhang, T., Schwer, B., Basu, U., Alt, F.W., Tang, J., Oltz, E.M., Murphy, T.L. *et al.* (2011) The transcription factor BATF controls the global regulators of class-switch recombination in both B cells and T cells. *Nat. Immunol.*, **12**, 536–543.
  57. Ochiai, K., Maienschein-Cline, M., Simonetti, G., Chen, J., Rosenthal, R., Brink, R., Chong, A.S., Klein, U., Dinner, A.R., Singh, H. *et al.* (2013) Transcriptional regulation of germinal center B and plasma cell fates by dynamical control of IRF4. *Immunity*, **38**, 918–929.
  58. Quartier, P., Bustamante, J., Sanal, O., Plebani, A., Debré, M., Deville, A., Litzman, J., Levy, J., Ferman, J.P., Lane, P. *et al.* (2004) Clinical, immunologic and genetic analysis of 29 patients with autosomal recessive hyper-IgM syndrome due to activation-induced cytidine deaminase deficiency. *Clin. Immunol.*, **110**, 22–29.
  59. Richardson, C., Chida, A.S., Adlowitz, D., Silver, L., Fox, E., Jenks, S.A., Palmer, E., Wang, Y., Heimburg-Molinaro, J., Li, Q.-Z. *et al.* (2013) Molecular basis of 9G4 B cell autoreactivity in human systemic Lupus erythematosus. *J. Immunol.*, **191**, 4926–4939.
  60. Moura, R.A., de la Torre, I., Leandro, M.J., Edwards, J.C.W., Fonseca, J.E. and Cambridge, G. (2012) 9G4 expression on autoantibodies to citrullinated peptides in patients with early inflammatory arthritis and established rheumatoid arthritis. *Ann. Rheum. Dis.*, **71**, A1–A93.
  61. Zaheen, A., Boulianne, B., Parsa, J.Y., Ramachandran, S., Gommerman, J.L. and Martin, A. (2009) AID constrains germinal center size by rendering B cells susceptible to apoptosis. *Blood*, **114**, 547–554.
  62. Meyers, G., Ng, Y.-S., Bannock, J.M., Lavoie, A., Walter, J.E., Notarangelo, L.D., Kilic, S.S., Aksu, G., Debre, M., Rieux-Laucat, F. *et al.* (2011) Activation-induced cytidine deaminase (AID) is required for B-cell tolerance in humans. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11554–11559.
  63. Nemazee, D. (2006) Receptor editing in lymphocyte development and central tolerance. *Nat. Rev. Immunol.*, **6**, 728–740.