Supplemental Materials

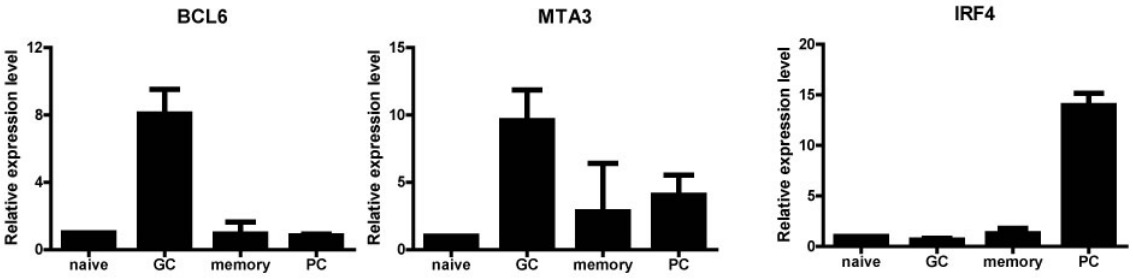
DNA methylation profiling in human B cells reveals immune regulatory elements and epigenetic plasticity at *Alu* elements during B cell activation

Lai et al., 2013

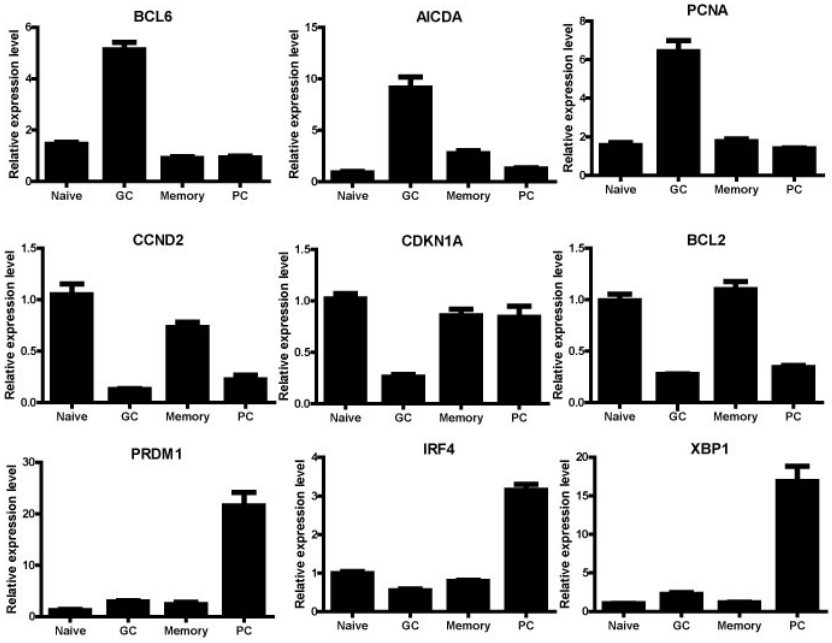
Contents

Lai et al., Supplemental Figure 1

A



B



C

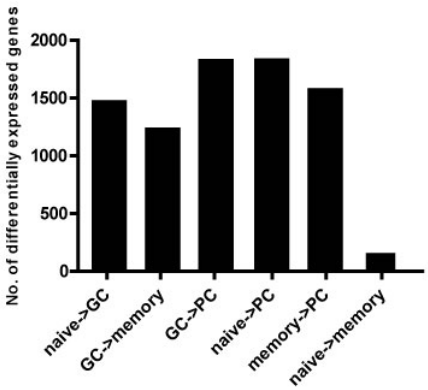


Figure S1. Gene expression analysis of purified B cell subsets from tonsil

A. Quantitative RT-PCR analysis of GC-specific transcripts (BCL6 and MTA3) and a PC-specific transcript IRF4 in FACS sorted cell populations as indicated in bar graph. Gene expression levels are normalized to actin expression level in each cell type. The expression level of each gene in naive B cells is arbitrarily set as 1. Error bars denote standard deviation from 3 biological replicates.

B. Expression pattern of GC-specific transcripts (top panels), known targets repressed by BCL6 (middle panels), and PC-specific transcripts (bottom panels) among the 4 cell types. The bar graphs represent the relative expression level between the cell types based on normalized intensities in the gene expression microarray. Expression level of a naive B cell sample is arbitrarily set as 1 for each gene. Error bars denote the standard deviation of 8 biological replicates.

C. The column graph depicts quantitation of differentially expressed transcripts from the gene expression microarray. Comparisons were made between each pair of cell types by t-test, adjusted $p<0.01$, with fold change > 2.

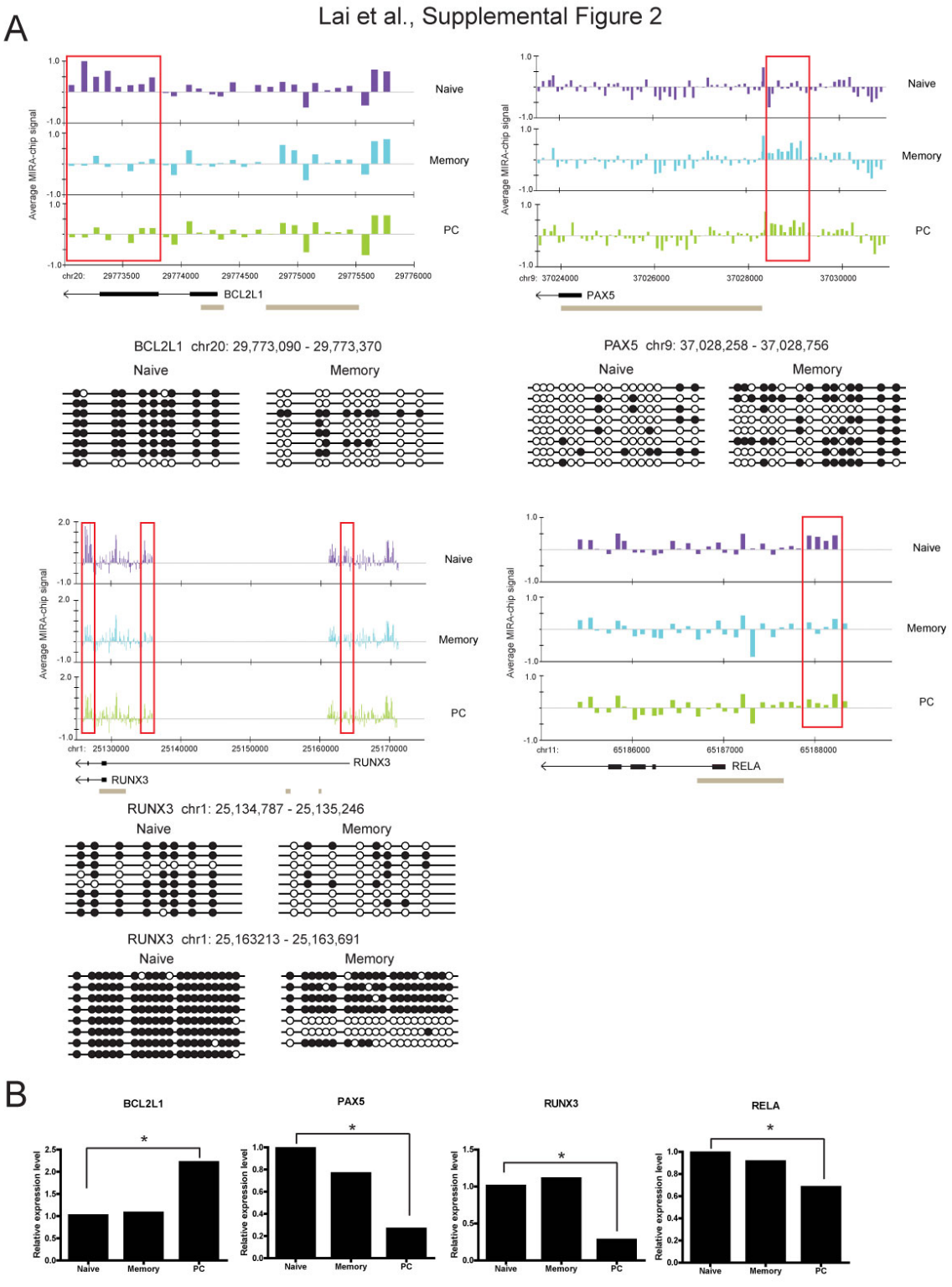

A
Lai et al., Supplemental Figure 2
Average MIRA-chip signal
Naive
Memory
PC
chr20: 29773500 29774000 29774500 29775000 29775500 29776000
BCL2L1
chr9: 37024000 37026000 37028000 37030000
PAX5
BCL2L1 chr20: 29,773,090 - 29,773,370
PAX5 chr9: 37,028,258 - 37,028,756
chr1: 25130000 25140000 25150000 25160000 25170000
RUNX3
chr11: 65186000 65187000 65188000
RELA
RUNX3 chr1: 25,134,787 - 25,135,246
RUNX3 chr1: 25,163213 - 25,163,691
B
BCL2L1
PAX5
RUNX3
RELA
Relative expression level
Naive
Memory
PC

Figure S2. Validation of DMRs by genomic bisulfite sequencing.

A. The bar graphs display the average MIRA-chip signals from 8 biological replicates in naive, memory, and PC cells at regions near TSS at the indicated gene. The x-axis shows the genomic location (UCSC genome browser, HG18). The location of the gene is shown below the bar graphs relative to the genomic coordinates. Orientation of arrow indicates the direction of transcription, black rectangles indicate location of exons. Gray rectangles below indicate the presence of CpG islands. Bisulfite sequencing of the same genomic region from a randomly selected naive and memory B cell sample is shown below the bar graph. Each line of circles indicates an individual clone sequenced following bisulfite treatment and PCR. Open circles indicate CpG sites at which no DNA methylation is detected. Blackened circles indicate CpG sites which are methylated.

B. Relative expression level of transcripts shown in panel A in naive, memory, and PC cells from gene expression microarray data. * denotes statistical significance by student t-test.

Lai et al., Supplemental Figure 3

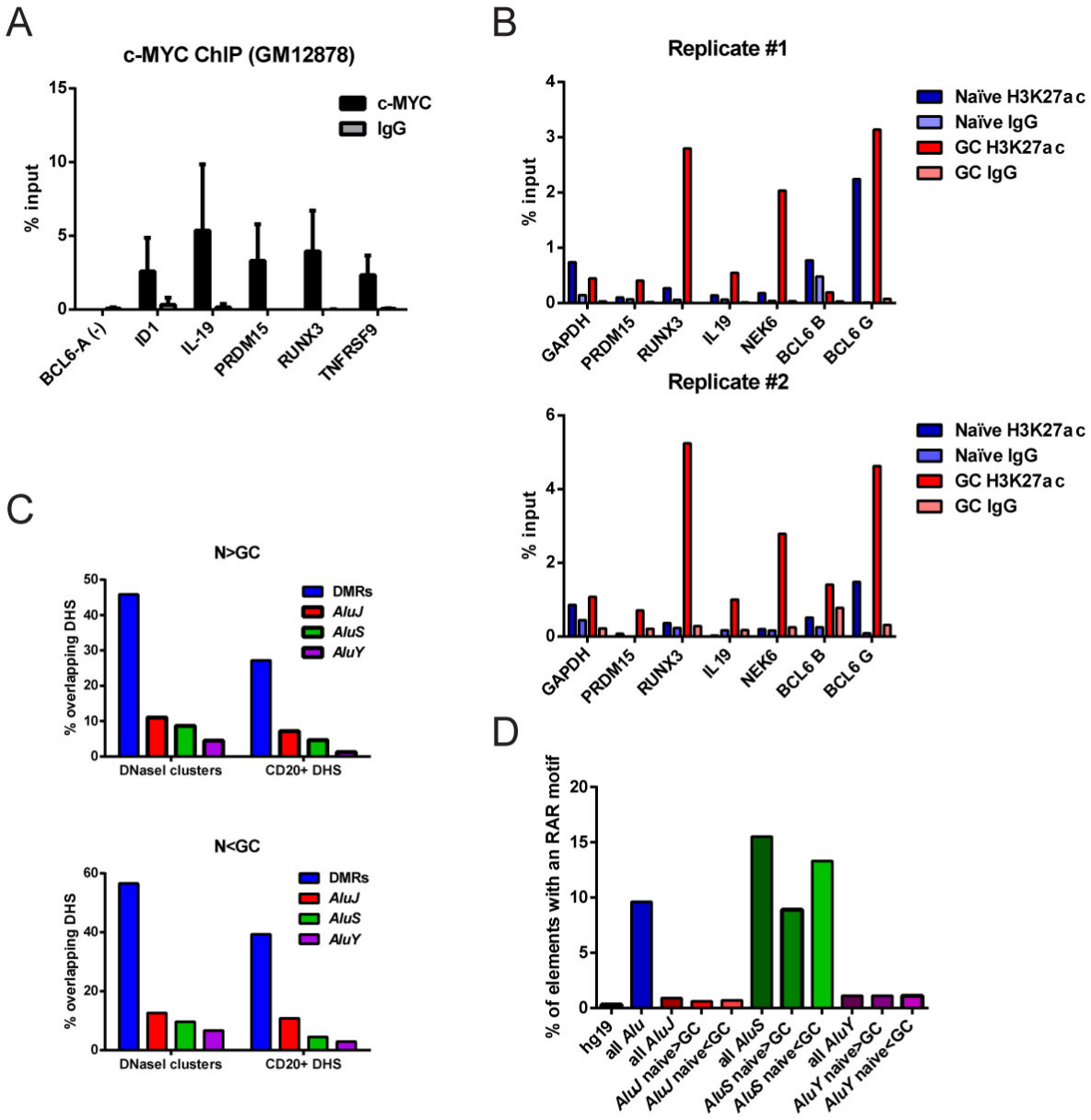


Figure S3. Enrichment of cis-regulatory elements at DMRs.

A. Myc ChIP was performed in B lymphoblastoid cell line GM12878. The % input is shown from primer sets overlapping 5 DMRs (labelled as the nearest annotated gene). Error bars denote standard deviation of three replicates. BCL6-A is a region upstream of the BCL6 promoter and serves as a negative control region based on Myc ChIP-seq data (performed by Iyer's group at UT Austin) from ENCODE.

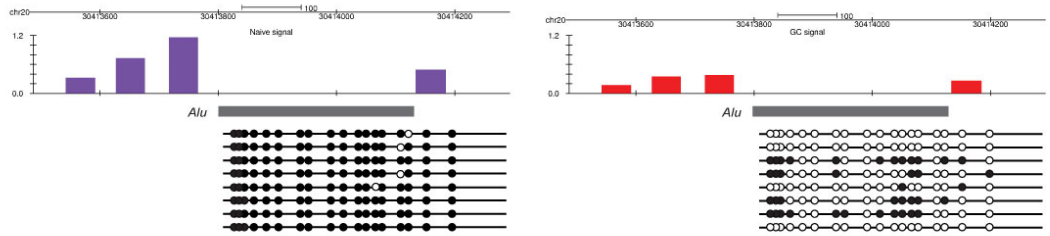
B. H3K27ac or IgG control ChIPs were performed in 2 biological replicates of naive and GC B cells. The % input at various DMRs analyzed by qPCR (labelled as the nearest annotated gene) is shown. GAPDH and BCL6 primer sets serve as control regions. H3K27ac levels at GAPDH promoter remain constant in naive and GC B cells, since its expression is ubiquitous. BCL6 B is a region upstream of the BCL6 promoter and enrichment of H3K27ac is also expected to remain unchanged in the two cell types. BCL6 G overlaps with the BCL6 promoter, and H3K27ac level is expected to become elevated in GC B cells accompanying increased gene expression.

C. Overlap analysis of DMRs or differentially methylated *Alu*s (from MIRA-seq) with DHS. The percentage of loss-of-methylation (N>GC) or gain-of-methylation (N<GC) DMRs and differentially methylated *Alu*s intersecting DNaseI clusters from any cell type or DHS from primary CD20+ B cells are shown.

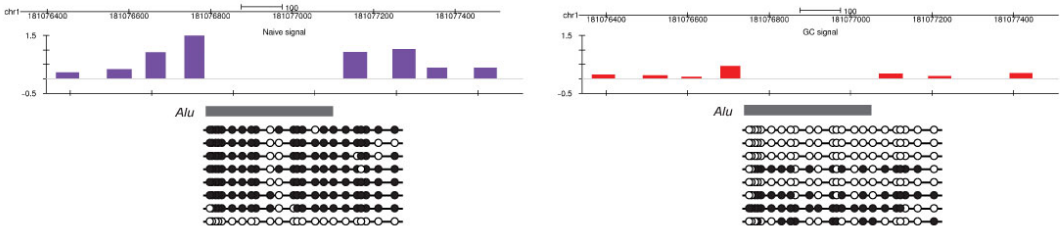
D. The percentage naive>GC or naive<GC differentially methylated elements in each *Alu* family containing the RAR motif (RGKTCAN1-5RGKTCA) on either strand is shown.

Lai et al., Supplemental Figure 4

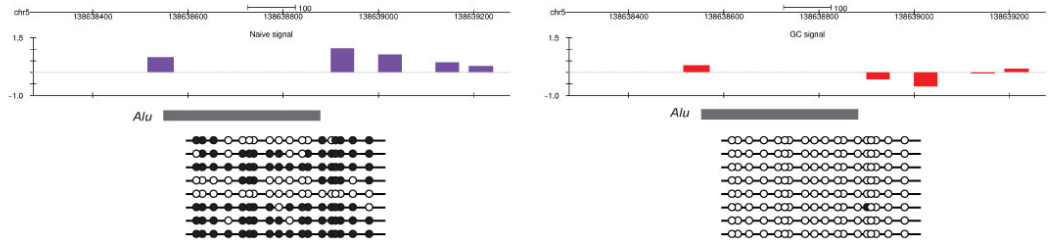
AluY: chr20:030413818-030414133



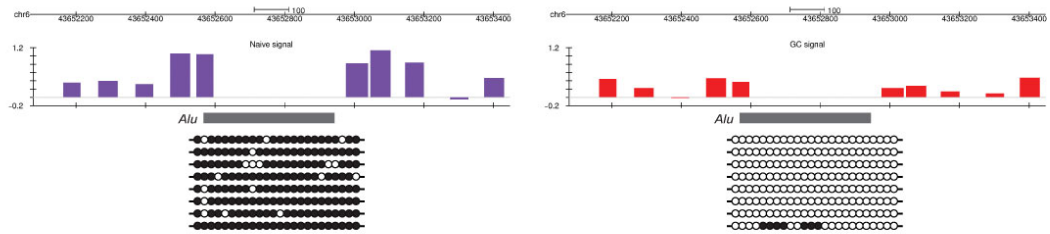
AluSc:chr1:181076355-181076057



AluY: chr5:138638884-138638569



AluJo: chr6:043658561-043658263



AluSq: chr10:001025354-001025054

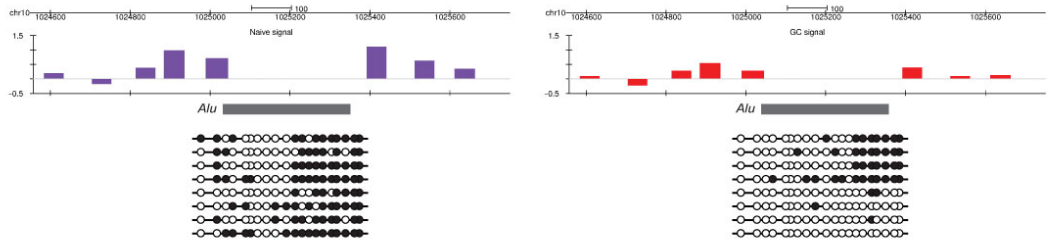


Figure S4. Validation of differentially methylated *Alu* elements by genomic bisulfite sequencing

The methylation status adjacent to and within 5 *Alu* elements was analyzed in naive and GC cell populations. The bar graphs illustrate the average MIRA-chip signal from 8 biological replicates in naive or GC B cells at probes neighbouring each *Alu* element. The x-axis shows the genomic location (UCSC genome browser, Hg18) of the repeat. Bisulfite sequencing of the same genomic region of a randomly selected naive or GC B cell sample is shown below each bar graph. Each line of circles indicates an individual clone sequenced following bisulfite treatment and PCR. Open circles indicate CpG sites at which no DNA methylation is detected. Blackened circles indicate CpG sites which are methylated.

Lai et al., Supplemental Figure 5

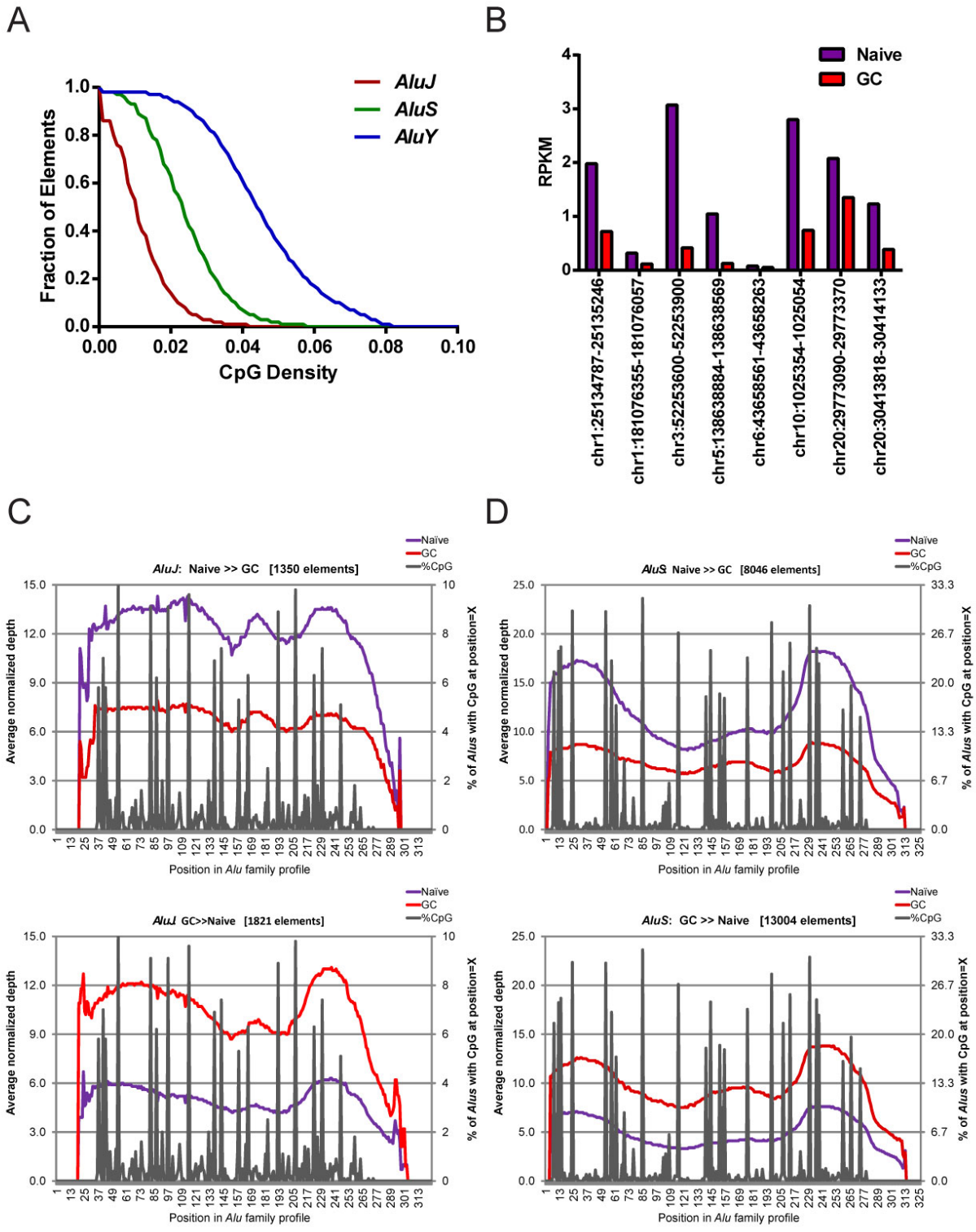


Figure S5. Analysis of DNA methylation at *Alu* elements.

A. Spectrum of CpG density in different family members of *Alu*. CpG density is defined as the number of CpG sites within a given element divided by element width. The line graph plots the fraction of elements in *Alu* family with CpG density $\geq$ X.

B. Sequencing coverage (RPKM) at 8 different *Alu*s identified from MIRA-chip analysis that have lost methylation in GC B cells compared to naïve B cells. The genomic locations of these elements are listed on the x-axis. Higher RPKM was observed in naive compared to GC B cell sample in all cases. The methylation status of these *Alu* elements was also validated by bisulfite sequencing (Fig. S4 and data not shown).

C. Average depth of mapped reads from MIRA-seq across a consensus profile for *AluJ*, the position in base pair is shown on the x-axis. % of *Alu*s with CpG at each position within the *AluJ* subfamily is also shown. The top panel shows the average depth of *AluJ*s that have lost methylation in GC B cells, while the bottom panel shows the average depth of *AluJ*s that have gained methylation during in GC B cells compared to naive B cells.

D. Average normalized depth of mapped reads from MIRA-seq across a consensus profile for *AluS*, the position in base pair is shown on the x-axis. % of *Alu*s with CpG at each position with *AluS* is also shown. The top panel shows the average depth of *AluS*s that have lost methylation in GC B cells, while the bottom panel shows the average depth of *AluS*s that have gained methylation during in GC B cells compared to naive B cells.

Table S1: Distribution of Methylation Peaks at Genomic Features

	Naive			GC			Memory			PC		
	# Peaks	**probes**	**bp**	**# Peaks**	**probes**	**bp**	**# Peaks**	**probes**	**bp**	**# Peaks**	**probes**	**bp**
All Peaks	5374 (2149)	34431 (17473)	9069211 (4621829)	4043 (1006)	23948 (6676)	6101894 (2504010)	5188 (1414)	32088 (10595)	7893186 (3054787)	4849 (1155)	30004 (9348)	7357600 (3095753)
Promoter[a]	788 (366)	5200 (3019)	830787 (479382)	624 (156)	3767 (1065)	901732 (575715)	802 (258)	5043 (1941)	810145 (420193)	746 (204)	4699 (1702)	725963 (302767)
CpG island[b]	1363 (611)	9714 (5092)	6272787 (3579831)	997 (499)	6627 (3614)	4082991 (2405027)	1240 (644)	8628 (4675)	5420106 (2893536)	1124 (554)	7812 (4264)	4902531 (2750091)
CpG shore[c]	2149 (933)	14317 (7512)	6861464 (3842774)	1625 (515)	10056 (3722)	4562921 (2323938)	2038 (682)	13154 (5173)	5925342 (2906729)	1920 (608)	12344 (4976)	5448788 (2855860)
Alu[d]	2539 (1231)	16043 (9540)	7266575 (3842603)	1900 (496)	10874 (3026)	4835088 (2128437)	2466 (909)	14839 (6173)	6219569 (2654327)	2372 (667)	14228 (4789)	5825192 (2646017)

[a] 2kb window centered at TSS

[b] island start and end site

[c] 1kb upstream of island start and 1kb downstream of island end site

[d] start and end site extended by 250bp

*Values in ( ) represent standard deviation

Table S1. Distribution of methylation peaks at genomic features

The average number of methylation peaks overlapping genomic features from eight subjects is listed for each B cell subset. The average number of probes and the average number of base pairs (bp) are also reported. Standard deviation is shown in parentheses. The genomic features are defined based on annotation from the hg18 assembly in UCSC Genome Browser, with modification listed below the table.

Table S2: Differential Methylated Regions (DMRs) Between Cell Types

Comparison (Cell Type 1 vs. Cell Type 2)	Number of DMRs	Genomic Coverage of DMRs (probes)	Genomic Coverage of DMRs (bp)	DMR Probes Overlapping Promoter[a] %	Promoter[a] p-value*	*Alu*[b] %	*Alu*[b] p-value*	CpG island[c] %	CpG island[c] p-value*	CpG shore[d] %	CpG shore[d] p-value*
Cell Type 1 > Cell Type 2											
Naïve vs. GC	766	3930	453384	11.04	<.001	18.98	<.001	11.68	<.001	22.67	<.001
GC vs. Memory	75	324	27501	19.44	0.164	26.23	0.050	8.02	0.269	15.43	0.812
GC vs. PC	68	286	24695	11.54	0.022	13.99	0.001	4.20	0.001	14.34	0.767
Naïve vs. PC	416	2099	285415	12.48	<.001	20.00	0.054	11.33	0.023	22.19	<.001
Naïve vs. Memory	454	2327	263096	12.98	<.001	26.51	<.001	5.89	<.001	21.87	<.001
Memory vs. PC	52	657	56954	10.65	<.001	26.03	0.008	2.28	<.001	8.68	<.001
Cell Type 1 < Cell Type 2											
Naïve vs. GC	268	1313	117417	27.40	<.001	16.89	<.001	18.65	<.001	45.51	<.001
GC vs. Memory	159	702	60176	12.80	0.007	29.45	<.001	5.83	<.001	12.52	0.069
GC vs. PC	166	727	62825	8.80	<.001	36.45	<.001	1.38	<.001	14.86	0.936
Naïve vs. PC	332	1566	137598	20.18	<.001	29.25	<.001	6.58	<.001	32.12	<.001
Naïve vs. Memory	414	2016	183088	23.89	<.001	22.75	0.272	8.62	0.063	39.64	<.001
Memory vs. PC	211	966	84458	12.22	<.001	30.85	<.001	1.55	<.001	19.25	<.001

[a] 2kb window centered at TSS (overlaps 16.57% probes)

[b] start and end site extended by 250bp (overlaps 21.74% probes)

[c] island start and end site (overlaps 9.86% probes)

[d] 1kb upstream of island start and 1kb down stream of island end site (overlaps 14.96% probes)

* p-value is calculated as “single population proportion test” (Biostatistics, sixth edition by Wayne Daniel)

Table S2. Differential methylated regions (DMRs) between cell types

The number and genomic coverage of DMRs (see Supplemental Methods for definition) are reported for each pair of cell types compared. The percentage and p-value of DMRs overlapping defined genomic features (defined based on annotation from the hg18 assembly in UCSC Genome Browser, with modification listed below the table) are also shown. The percentage of probes in microarrays overlapping with each genomic feature is listed below the table.

Table S3: Genomic Features with Significant* Differential Methylation Scores Between Cell Types

Comparison (Cell Type 1 vs. Cell Type 2)	Genomic Features: Promoter[a] (N = 666) # (%)	CpG island[b] (N = 695) # (%)	CpG shore[c] (N = 1008) # (%)	*Alu*[d] (N = 2181) # (%)
Total				
Naïve vs. GC	427 (64.11%)	455 (65.47%)	727 (72.12%)	1017 (46.63%)
GC vs. Memory	214(32.13%)	223 (32.09%)	175 (17.36%)	636 (29.16%)
GC vs. PC	99 (14.86%)	123 (1.08%)	85 (8.43%)	427 (19.58%)
Naïve vs. PC	**271 (40.69%)**	**229 (32.95%)**	**469 (46.53%)**	**964 (44.20%)**
Naïve vs. Memory	**287 (43.09%)**	**265 (38.13%)**	**581 (57.64%)**	**1089 (49.93%)**
Memory vs. PC	**84 (12.61%)**	**73 (10.50%)**	**95 (9.42%)**	**375 (17.19%)**
Cell Type 1 > Cell Type 2				
Naïve vs. GC	156 (23.42%)	185 (26.62%)	385 (38.19%)	710 (32.55%)
GC vs. Memory	170 (25.53%)	177 (25.47%)	118 (11.71%)	250 (11.46%)
GC vs. PC	78 (11.71%)	96 (13.81%)	52 (5.16%)	148 (6.79%)
Naïve vs. PC	**133 (19.97%)**	**133 (19.14%)**	**288 (28.57%)**	**573 (26.27%)**
Naïve vs. Memory	**132 (19.82%)**	**128 (18.42%)**	**291 (28.87%)**	**663 (30.40%)**
Memory vs. PC	**32 (4.80%)**	**26 (3.74%)**	**48 (4.76%)**	**180 (8.25%)**
Cell Type 1 < Cell Type 2				
Naïve vs. GC	271 (40.69%)	270 (38.85%)	342 (33.93%)	307 (14.08%)
GC vs. Memory	44 (6.61%)	46 (6.62%)	57(5.65%)	386 (17.70%)
GC vs. PC	21 (3.15%)	27 (3.88%)	33 (3.27%)	279 (12.79%)
Naïve vs. PC	**138 (20.72%)**	**96 (13.81%)**	**181 (17.96%)**	**391 (17.93%)**
Naïve vs. Memory	**155 (23.27%)**	**137 (19.71%)**	**290 (28.77%)**	**426 (19.53%)**
Memory vs. PC	**52 (7.81%)**	**47 (6.76%)**	**47 (4.66%)**	**195 (8.94%)**

*post-hoc pairwise comparison p-value <= 0.05

[a]2kb window centered at TSS

[b]island start and end site

[c]1kb upstream of island start and 1kb down stream of island end site

[d]start and end site extended by 250bp

Table S3. Genomic Features with Significant Differential Methylation Scores Between Cell Types

Pair-wise comparison of average DNA methylation signal was analyzed at each genomic feature (defined based on annotation from hg18 assembly in UCSC Genome Browser, with modification listed below the table) using a mixed-affect model as described in supplemental methods.

Table S4: Pathway Analysis of genes containing DMRs within 10 kb of TSS in Naïve vs. Memory B cells

	Naïve > Memory (N = 142)		Naïve < Memory (N = 179)	
Function	*No. of molecules*	*P value**	*No. of molecules*	*P value**
Antigen Presentation	10	9.15E-05	1	1.52E-02
Cellular Movement	16	9.15E-05	33	4.31E-03
Immune Cell Trafficking	11	9.15E-05	6	5.97E-03
Cell Death	61	1.45E-04	33	7.72E-03
Cellular Growth & Proliferation	44	4.07E-04	59	1.35E-03
Cellular Assembly & Organization	15	4.72E-04	25	1.83E-04
Cellular Development	37	6.72E-04	55	1.35E-03
Hematological System Development & Function	29	6.72E-04	14	3.31E-03
Hematopoiesis	21	9.36E-04	9	2.38E-03
Cell Cycle	29	9.34E-04	34	6.87E-04
DNA Replication, Recombination, & Repair	14	9.36E-04	9	1.45E-02
Gene Expression	47	1.23E-03	60	4.46E-08
Nucleic Acid Metabolism	6	1.55E-03	7	7.70E-03
Cell Signaling	8	4.23E-03	12	8.07E-03
Cellular Function & Maintenance	20	4.23E-03	21	1.83E-04
Cell-mediated Immune Response	15	6.15E-03	5	6.05E-03
Cell Morphology	13	6.87E-03	25	2.81E-03
Cellular Compromise	10	9.64E-03	10	3.29E-04
Amino Acid Metabolism	1	1.26E-02	1	1.52E-02
Cell-To-Cell Signaling & Interaction	16	1.26E-02	43	1.22E-04
Humoral Immune Response	3	1.26E-02	1	1.52E-02
Lymphoid Tissue Structure & Development	1	2.51E-02	4	1.14E-02

Source: Ingenuity Pathway Analysis

* Fisher's Exact Test p-value

Table S4: Pathway Analysis of genes containing DMRs within 10 kb of TSS in Naive vs. Memory B cells

A list of genes with TSS located within 10 kb to DMRs (naive>memory and naive<memory DMRs) was analyzed by Ingenuity Pathway Analysis. Significantly enriched functional categories are shown.

Table S5: Correlation between DNA methylation (Naïve vs. GC) and gene expression by Gene Set Enrichment Analysis

Gene set	No. of genes	No. of probes	Average local stat	ES	P-value	NES	FDR	Global p-value
Naïve<GC	275	323	-0.296331	-0.180034	0.794682	-0.891748	0.832092	0.680749557
Naïve>GC	200	252	-0.704393	-0.368369	0.001475	-1.804696	0.015282	0.001308187
Naïve<PC	325	387	-0.153206	-0.166452	0.916312	-0.846125	0.892741	0.765874016
Naïve>PC	284	353	-0.6767	-0.345502	0.001449	-1.735879	0.02472	0.002767885
Naïve<Memory	17	22	0.328169	0.400488	0.131148	1.31192	0.387262	0.089982235
Naïve>Memory	9	13	-1.338249	-0.51867	0.094718	-1.386293	0.162546	0.052046344
GC<PC	265	314	-0.300147	-0.208829	0.322963	-1.039305	0.586996	0.383014455
GC>PC	308	385	-0.429075	-0.238497	0.058239	-1.207106	0.327912	0.158465835
GC<Memory	166	203	-0.455718	-0.269841	0.04769	-1.286384	0.24352	0.099024116
GC>Memory	208	251	-0.37261	-0.218764	0.335799	-1.056295	0.556396	0.353424294
PC<Memory	245	303	-0.444914	-0.257329	0.033923	-1.269821	0.258987	0.109503453
PC>Memory	280	332	-0.215737	-0.178483	0.786337	-0.893583	0.829393	0.676984475

No. of genes- number of gene symbols or entrez ids that belong to the gene set on the microarray platform

No. of probes- Number of probes that corresponds to genes within the gene set

Average local stat- Average value local statistics for probes that corresponds to genes within the gene set

ES- Global statistics (Enrichment score, Naïve-GC)

P-value- Tail probability from density plot of ES values across all permutations

NES- Normalized Enrichment Score

FDR- False discovery rate as defined by Subrmanian et al., 2002

Global p-value- Tail probably from density plot of NES values across all gene sets and all permutations

Table S5. Correlation between DNA methylation (Naive vs. GC) and gene expression by Gene Set Enrichment Analysis

Pairwise up and downregulated transcripts in each pair of cell types, determined from gene expression microarray data with both overall and pair-wise Benjamini-Hochberg adjusted p-value <=0.001 and pair-wise fold change >=3, were used as the 12 gene sets for the analysis: Naive>GC, Naive<GC, Naive>PC, Naive<PC, Naive>Memory, Naive<Memory, GC>PC, GC<PC, GC>Memory, GC<Memory, Memory>PC, and Memory<PC. The 12 gene sets were added into "C4-CM" section of MolSigDB (v3). Differential DNA methylation signal (average signal from all probes overlapping -7kb upstream and 3kb downstream of TSS of each refseq promoter) was first computed for each pair of cell types. The null-mean one sample t-statistic was used as local statistics for promoter significance. 1000 random permutations of promoter labels were used to compute enrichment p-value and raw/normalized enrichment statistics as described (Subramanian et al. 2005). The global p-value and the false discovery rate estimate were computed across all sections of MolSigDB.

Table S6: Gain and loss of methylation in *Alu* elements by MIRA-seq

	***Alu*J**		***Alu*S**		***Alu*Y**	
	No.	**%**	**No.**	**%**	**No.**	**%**
Total elements	308,193	100%	676,649	100%	141,006	100%
RPKM > 2 in at least one sample	7,568	2.5%	63,738	9.4%	55,022	39.0%
GC/Naïve > 1.5	85,984	27.9%	257,988	38.1%	32,195	22.8%
GC/Naive >1.5 and RPKM > 2 in at least one sample	**1,821**	**0.6%**	**13,004**	**1.9%**	**5,031**	**3.6%**
Naïve/GC > 1.5	83,741	27.2%	137,185	20.3%	22,569	16.0%
Naïve/GC > 1.5 and RPKM > 2 in at least one sample	**1,350**	**0.4%**	**8,046**	**1.2%**	**9,346**	**6.6%**
GC/Naïve >1.5 or Naïve/GC > 1.5	169,725	55.1%	395,173	58.4%	54,764	38.8%
GC/Naïve > 1.5 or Naïve/GC > 1.5 and RPKM > 2 in at least one sample	3,171	1.0%	21,050	3.1%	14,377	10.2%

Table S6. Gain and loss of methylation in *Alu* elements by MIRA-seq

Criteria for gain or loss of methylation by MIRA-seq at specific *Alu* elements were an RPKM ratio > 1.5 (GC/naive or naive/GC), with at least one cell type having RPKM > 2. Average depth across all members of a given *Alu* family was determined by calculating the positional depth for individual elements aligned against a consensus profile for that family.

Table S7: Cytosine coverage and methylation status in Naïve and GC B cells by MIRA-seq

% Methylation*	CpG		CHG		CHH	
	Naïve	GC	Naïve	GC	Naïve	GC
0%	0.70%	0.89%	97.54%	98.22%	98.48%	98.98%
0.1-33.3%	0.99%	1.58%	2.34%	1.68%	1.41%	0.92%
33.4-66.6%	5.29%	8.40%	0.06%	0.05%	0.05%	0.05%
66.7-99.9%	44.26%	46.30%	0.03%	0.03%	0.03%	0.03%
100%	48.76%	42.83%	0.02%	0.02%	0.02%	0.02%
Positions passing depth filter (minimum depth at position = 5)	**7,448,491**	**8,068,366**	**14,068,133**	**15,569,993**	**28,697,229**	**31,051,801**

* %Methylation = methyl-C count / position depth

(data is summed for + and – strands)

Table S7. Cytosine coverage and methylation status in Naive and GC B cells by MIRA-seq

The % of methylation in each sequence read is calculated as methyl-C count / position depth. Only positions with a minimum depth of 5 are analyzed. The data is summed for + and – strands.

Table S8: Correlation of *Alu* methylation between MIRA-seq and MIRA-chip data

Naïve > GC *Alus*

Filtering criteria (seq, array)	No. of *Alus* in both platforms	Naïve/GC > 1.5 (seq)	Naïve>GC (array)	% agreement
No RPKM filter, any p-val	15,774	2,451	1,517	61.9%
RPKM>2, any p-val	7,589	963	687	71.3%
No RPKM filter, p-val < 0.1	2,143	388	316	81.4%
RPKM>2, array p-val < 0.1	1,112	174	158	90.8%

GC > Naïve *Alus*

Filtering criteria (seq, array)	No. of *Alus* in both platforms	GC/Naive > 1.5 (seq)	GC>Naive (array)	% agreement
No RPKM filter, any p-val	15,774	3,763	1,907	50.7%
RPKM>2, any p-val	7,589	767	391	51.0%
No RPKM filter, p-val < 0.1	2,143	464	250	53.9%
RPKM>2, array p-val < 0.1	1,112	106	66	62.3%

Table S8. Correlation of *Alu* methylation between MIRA-seq and MIRA-chip data

The percentage of naive>GC or naive<GC differentially methylated *Alu*s defined in both the sequencing and microarray platforms are reported. Differentially methylated *Alu*s from MIRA-chip were identified by a mixed effect model based on the average methylation signal at these elements in naive and GC B cells (see Supplemental Methods for details).

Table S9. Differential methylation at *Alu* elements (bisulfite sequencing)

Alu at chr20 30950157-30950472

Position in *Alu*	% Methylation Naive	% Methylation GC
48	100	50
53	100	50
57	100	50
64	100	25
78	100	0
98	100	0
142	100	37.5
150	100	12.5
174	100	0
188	100	25
198	100	37.5
204	100	50
206	87.5	50
213	100	50
230	75	0
236	87.5	25
264	100	25

Alu at chr1 182809434-182809732

Position in *Alu*	% Methylation Naive	% Methylation GC
4	100	62.5
8	100	50
10	100	50
20	100	62.5
39	100	62.5
52	100	37.5
57	25	37.5
97	25	12.5
108	75	62.5
141	87.5	37.5
149	87.5	50
153	87.5	25
173	100	62.5
195	100	37.5
212	100	50
237	87.5	50
259	87.5	50
267	100	37.5

275	62.5	37.5

Alu at chr5 138610670-138610985

Position in *Alu*	**% Methylation Naive**	**% Methylation GC**
4	62.5	0
8	75	0
10	75	0
20	25	0
57	75	0
64	75	0
78	75	0
98	62.5	0
109	62.5	0
160	50	0
164	37.5	0
172	50	0
196	62.5	0
235	87.5	11.11111
252	75	0
258	87.5	0
298	87.5	0

Alu at chr3 52278600-52278884

Position in *Alu*	**% Methylation Naive**	**% Methylation GC**
4	75	10
20	87.5	0
48	87.5	10
53	87.5	0
57	87.5	0
80	87.5	0
111	75	0
135	87.5	0
147	87.5	0
189	50	0
223	50	0

Alu at chr10 1035054-1035354

Position in *Alu*	% Methylation Naive	% Methylation GC
39	50	7.142857
61	42.85714	14.28571
69	42.85714	7.142857
93	7.142857	0
95	14.28571	14.28571
103	28.57143	21.42857
125	21.42857	21.42857
145	14.28571	7.142857
172	71.42857	21.42857
189	78.57143	21.42857
202	35.71429	7.142857
220	92.85714	35.71429
234	85.71429	35.71429
247	78.57143	64.28571
252	100	50
280	78.57143	50
292	100	50
296	100	42.85714

Alu at chr6 43544611-43544931

Position in *Alu*	% Methylation Naive	% Methylation GC
4	100	0
8	50	0
10	100	0
20	87.5	0
47	87.5	12.5
53	100	12.5
57	100	12.5
80	100	12.5
111	100	0
166	87.5	0
171	62.5	12.5
176	87.5	12.5
178	87.5	12.5
198	100	0
220	87.5	0
230	100	0
237	100	0
253	100	0

259	100	0
261	100	0
283	87.5	0
291	87.5	0
299	87.5	0

Alu at chr6 74160430-74160742

Position in *Alu*	% Methylation Naive	% Methylation GC
4	62.5	55.55556
8	62.5	66.66667
10	75	33.33333
20	75	33.33333
48	62.5	44.44444
53	50	44.44444
57	87.5	55.55556
78	75	33.33333
98	62.5	11.11111
109	50	11.11111
140	62.5	33.33333
152	62.5	22.22222
156	62.5	22.22222
176	25	22.22222
200	37.5	11.11111
208	50	22.22222
215	50	44.44444
232	37.5	0
240	37.5	0
265	25	22.22222
270	37.5	22.22222
278	37.5	22.22222

Alu at chr19 6592774-6593074

Position in *Alu*	% Methylation Naive	% Methylation GC
8	70	0
20	70	12.5
57	90	25
100	90	62.5
112	70	62.5
141	90	50
153	90	37.5
157	90	50
201	80	25

209	90	62.5
233	90	37.5
241	80	50
279	80	87.5

Alu at chr8 613758-614058

Position in *Alu*	% Methylation Naive	% Methylation GC
4	88.88889	0
10	88.88889	11.11111
20	88.88889	0
48	88.88889	11.11111
57	88.88889	11.11111
67	88.88889	11.11111
78	88.88889	22.22222
98	88.88889	22.22222
138	88.88889	22.22222
142	88.88889	22.22222
150	100	22.22222
154	88.88889	22.22222
236	88.88889	0
260	88.88889	0
268	88.88889	0

Table S9. Differential methylation of *Alu* elements by genomic bisulfite sequencing.

The table gives the genomic location of the 9 *Alu* elements validated by manual genomic bisulfite sequencing along with the location of each CpG assessed within the *Alu* and its percent methylation in the naive and germinal center cell populations.

Table S10: Functional enrichment of genes near *Alus* that have lost methylation (naïve>GC) in GC B cells compared to naïve B cells*

Fuctional category	p-value	No. of genes (out of 2998)
Cellular Growth and Proliferation	$1.41x10^{-5}$-$1.68x10^{-2}$	606
Cellular Development	$4.85x10^{-10}$-$1.68x10^{-2}$	538
Gene Expression	$2.02x10^{-8}$-$1.28x10^{-2}$	499
Cellular Assembly and Organization	$9.97x10^{-7}$-$1.66x10^{-2}$	357
Cellular Function and Maintenance	$2.43x10^{-5}$-$1.68x10^{-2}$	281

***A list of genes associated with differentially methylated *Alus* (gene with TSS closes to the *Alu*) was analyzed by Ingenuity Pathway analysis.**

Table S10. Functional enrichment of genes near *Alu*s that have lost methylation (naive>GC) in GC B cells compared naive B cells.

A list of genes proximal to naive>GC differentially methylated *Alu*s defined by MIRA-seq (defined as the annotated gene with TSS closest in distance (bps) to each Naive>GC differentially methylated *Alu*s) was analyzed by Ingenuity Pathway Analysis. Significantly enriched functional categories are shown.

Table S11. Differential methylation at regions neighboring differentially methylated *Alu* elements (bisulfite sequencing)

Alu at chr1 182809434-182809732

Position relative to 5' end of *Alu*	% Methylation Naive	% Methylation GC
708	50	0
711	58.33333	0
715	75	10
719	83.33333	0
721	83.33333	0
731	66.66667	10
750	75	10
763	83.33333	0
768	75	10
808	50	0
819	33.33333	10
852	83.33333	10
860	83.33333	20
864	83.33333	10
884	75	10
906	83.33333	30
923	83.33333	20
948	75	20
970	83.33333	30
978	83.33333	10
986	83.33333	30

Alu at chr20 30950157-30950472

Position relative to 5' end of *Alu*	% Methylation Naive	% Methylation GC
-567	75	33.33333
-561	100	66.66667
-551	87.5	55.55556
-523	75	11.11111
-514	87.5	22.22222
-491	75	11.11111
-460	37.5	0
-393	62.5	0
-361	75	0
-313	75	44.44444
-310	75	22.22222
-208	75	77.77778

Alu at chr3 52278600-52278884

Position relative to 5' end of *Alu*	% Methylation Naive	% Methylation GC
-69	25	20
-91	100	50
-99	87.5	50
-132	100	60
-147	100	70
-164	100	60
-180	100	50

Alu at chr10 1035054-1035354

Position relative to 5' end of *Alu*	% Methylation Naive	% Methylation GC
-287	0	0
-280	0	0
-271	0	0
-254	0	0
-235	0	0
-230	0	0
-223	0	0
-210	0	0
-194	0	0
-161	0	0
-152	0	0
-142	0	14.28571

Table S11. Differential methylation of genomic regions neighboring *Alu* elements.

The table gives the genomic coordinates for four regions neighboring a differentially methylated *Alu* element whose methylation status was determined by manual bisulfite sequencing. The table depicts the genomic coordinates, location of each CpG measured relative to the *Alu* element 5' end, and percent methylation in the naive and germinal center cell populations.

Table S12: Clinical data of tonsil samples

Case	Age (yrs)	Gender	Clinical Diagnoses	Histopathology
1	2	F	Obstructive sleep apnea, pyriform aperture stenosis, adenotonsillar hypertrophy	Follicular hyperplasia
2	6	F	Chromosome 18 abnormalities, sleep disordered breathing, tonsillar hypertrophy	Follicular hyperplasia
3	1	M	Chronic adenotonsillitis, adenotonsillar hypertrophy	Follicular hyperplasia
4	5	F	Premature, hypothyroid, failure to thrive, adenotonsillar hypertrophy	Follicular hyperplasia, actinomyces-like organisms
5	4	F	Adenotonsillar hypertrophy	Follicular hyperplasia
6	2	M	Sickle cell anemia, adenotonsillar hypertrophy	Follicular hyperplasia, actinomyces-like organisms
7	2	M	Obstructive sleep apnea, adenotonsillar hypertrophy	Follicular hyperplasia
8	3	F	Sleep disordered breathing, adenotonsillar hypertrophy	Follicular hyperplasia

Table S12. Clinical data of tonsil samples

Available information of subjects used in this study is listed.

3. Supplemental Methods

MIRA-chip data analysis

Data normalization

A two-step normalization approach was employed where the first step is designed to correct for GC bias and dye bias within a chip (intra-chip correction); and the second step corrects for variations across chips (inter-chip correction).

(i) Within chip normalization: First all probes were binned according to their GC content. GC content was computed as a ratio of number of C + G nucleotides to the total number of nucleotides in the probe sequence. The overall variability in GC content values, computed by looking at all probe sequences on the chip, was used to compute bin width according to the zero-stage rule (Wand 1997). These bin widths are proven to be approximate L2-optimal i.e. they minimize Mean Integrated Square Error (MISE). The bins with fewer probes are then merged to ensure that each bin contains at least 500 probes. Within each bin Lowess regression was used (Cleveland et al. 1988; Hastie and Tibshirani 1990) to predict $\log_2$(cy5) values as a smooth function of $\log_2$(cy3) values. The residuals from bin-wise lowess regression were rescaled using mean absolute deviation (to ensure homogeneity across bins) and were used by subsequent analysis.

(ii) Between chip normalization: Once the data was corrected for dye and GC bias as described above, quantile normalization was used to correct for between sample variations. The resulting dataset is referred to as 'normalized data' and is used for further investigations.

Identification of peaks of enrichment:

We modified the ACME algorithm (Scacheri et al. 2006) to identify methylation "peaks" within a sample as further described below. Our modified ACME algorithm depends on three user-

specified tuning parameters: window size (w), signal threshold (s) and p-value threshold (p). Any probes in the data that are above threshold (s) are considered positive probes. Enrichment p-value is computed using hyper geometric distribution by looking at observed number of positive probes (probes with signal > s) within a sliding window of size w centered on each probe as described below:

$$p(x) = \sum_{j=x}^{n} \binom{K}{j} \binom{N-K}{n-j} \Big/ \binom{N}{n}$$

where N denotes total number of probes, K denotes total number of probes with signal > s (signal threshold defaults to 90th percentile), n denotes number of probes in sliding window of size w (defaults to 500) and *x* denotes number of probes inside the sliding window with signal >s. Next, the binding sites are identified as runs of positive enrichment p-values i.e. below threshold (defaults is $p < 0.001$). Each positive run of this sequence is considered a binding site. We do not correct the enrichment p-values for multiple comparisons as they are only used as a means of finding regions of interest in the genome rather than a strict statistical significance level. Peaks identified here represent genomic regions that display significantly high degree of methylation in the sample under consideration compared to the sample's genome wide baseline.

Identification of DMRs: To identify regions that represent differential methylation in one cell type compared to another cell type of the same individual (DMRs), a modified ACME algorithm was used to spool data across the 8 independent subjects (replicates). Differences in methylation signals for each pairwise cell type comparisons was computed using the window size (*w*), signal threshold (s) and p-value threshold (*p*) as described for the identification of methylation peaks (Lai et al. 2010). To identify up (or down) regulated peaks, the number (*x*) of signal values within window of size *w* (centered at probe) that are greater than 100sth percentile (or less than

100(1-*s*)th percentile) was first computed across all replicates. Next, the enrichment p-value for probe was computed using hypergeometric distribution as following:

$$p(x) = \frac{\sum_{i=x}^{kr} \binom{Nr(1-s)}{i} \binom{Nrs}{kr-i}}{\binom{Nr}{kr}}$$

where *N* denotes total number of probes, *k* denotes number of probes in window and *r* denotes the number of replicates. DMRs were then identified as runs of enrichment p-values that are less than p-value threshold (*p*). The analysis presented here correspond to signal threshold (*s*=0.95), window size (*w*=500) and p-value threshold (*p*=0.001) with peaks containing less than four probes excluded. DMRs were mapped to various genomic features, the types and their definitions are listed in Table S1.

Analysis of average methylation signal at defined genomic features

We first generated the various feature level methylation (feature score) datasets by averaging methylation signal from probes that overlap corresponding genomic features which are defined as follows: Promoters- 2 kbps window centered at TSS (n=22,386), *Alu*s- *Alu* start and end sites extended by 250 bps (n=143,020), CpG Islands- island start and end sites (n=27,441), CpG Island Shores- 1kbps upstream of island start site and 1 kbps downstream of island end site (n=24,489).

Cell Type Signatures: The genomic features that are differentially methylated by cell types were identified using the following mixed effect model

$$y_{ij} = \mu + \alpha_i + b_j + \varepsilon_{ij}$$

$i \in \{GC, PC, Memory, Naive\}$ j=1,2,..,8

where y_{ij} denotes the methylation signal for i^{th} cell type from j^{th} subject; μ denotes overall mean; α_i denotes fixed effect of cell type *i*; b_j denotes random effect for subject j and ε_{ij} denotes random experimental error. The random effects (b_j) and residual errors (ε_{ij}) were assumed to be independent and normally distributed with means 0 and variances σ_s^2 and σ^2 respectively.

Subject Signatures: The features that are differentially methylated by subject were identified using following mixed effect model

$$\mathbf{y}_j = \mu_j \mathbf{1} + \boldsymbol{\varepsilon}_j \quad \text{j=1,2,...,8}$$

where $\mathbf{y}_j$ denotes the methylation signal from j^{th} subject; μ_j denotes overall mean for subject j; $\boldsymbol{\varepsilon}_j$ denotes random error that are assumed to be multivariate normal with mean **0** and variance covariance matrix $\sigma^2[(1-\rho)\mathbf{I} + \rho\mathbf{J}]$; **I** denotes identity matrix and **J** is matrix with all elements equal to 1.

Analysis of average methylation signal at *Alu*s: The methylation score at each *Alu* was first calculated by averaging methylation signal from probes that overlap 250 bps upstream and downstream of each element. The *Alu*s that are differentially methylated by cell types were identified using following mixed effect model

$$y_{ij} = \mu + \alpha_i + b_j + \varepsilon_{ij}$$

$i \in \{GC, PC, Memory, Naive\}$ j=1,2,..,8

where y_{ij} denotes the methylation signal for *i*th cell type from *j*th subject; μ denotes overall mean; α_i denotes fixed effect of cell type *i*; b_j denotes random effect for subject j and ε_{ij} denotes

random experimental error. The random effects (b_j) and residual errors (ε_{ij}) were assumed to be independent and normally distributed with means 0 and variances σ_s^2 and σ^2 respectively.

MIRA-seq analysis

Three lanes of data per cell type were merged. For each data set, both mates were trimmed to 60 bases to remove a region of poor base quality at the 3' end of reads. The raw data was further filtered to retain only read pairs meeting the following criteria: (a) average base quality score ≥ 20 for each mate, and (b) not a match to the sequencing primer/adapter. These filters removed 35.5% and 27.5% of the raw input read pairs for the naive and GC cell types, respectively. Filtered read pairs were then aligned to the human genome (GRCh37 / hg19, excluding haplotype chromosomes) via Bismark (v0.7.7, with parameters -n 2 -l 40 -e 70 -I -X 10000) (Krueger and Andrews 2011), resulting in 64.2% (naive) and 57.6% (GC) of read pairs were uniquely mapped. The deduplication script associated with the Bismark tool was used to remove duplicate sequenced fragments, corresponding to 33.9% (naive) and 31.4% (GC) of uniquely-mapped read pairs. A total of 31,156,289 (naive) and 33,121,073 (GC) uniquely-mapped, non-duplicate read pairs were used for analysis. Median insert size (as calculated by CollectInsertSizeMetrics.jar, Picard tools suite) of uniquely-mapped non-duplicate read pairs was 287 (naive) and 274 (GC), with 88.1% (naive) and 77.2% (GC) in the 200-400nt range, confirming that the observed mapped fragment size is consistent with the expected fragment size of the libraries.

Coverage of *Alu* Elements by MIRA-seq

Genomic coordinates of *Alu* elements were retrieved from the UCSC Genome Browser records identified by RepeatMasker (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/, file timestamp 27-Apr-2009). Coverage of each *Alu* element by MIRA-seq was calculated as RPKM

(reads per kilobase per million reads mapped), using aggregate mapped bases divided by trimmed read length in place of read count per element. Criteria for gain or loss of methylation by MIRA-seq at specific *Alu* elements were an RPKM ratio > 1.5 (GC/naive or naive/GC), with at least one cell type having RPKM > 2. Average depth across all members of a given *Alu* family was determined by calculating the positional depth for individual elements aligned against a consensus profile for that family. *Alu* family profiles were constructed with the hmmbuild tool from the HMMER3.0 package (http://hmmer.janelia.org/), trained on a multiple sequence alignment of randomly-selected family members (10,000 elements for *AluY* and 2,000 elements each for *AluJ* and *AluS*) generated by MUSCLE (Edgar 2004).

4. Supplemental References

Blanco-Betancourt CE, Moncla A, Milili M, Jiang YL, Viegas-Pequignot EM, Roquelaure B, Thuret I, Schiff C. 2004. Defective B-cell-negative selection and terminal differentiation in the ICF syndrome. *Blood* **103**(7): 2683-2690.

Bock C, Beerman I, Lien WH, Smith ZD, Gu H, Boyle P, Gnirke A, Fuchs E, Rossi DJ, Meissner A. 2012. DNA Methylation Dynamics during In Vivo Differentiation of Blood and Skin Stem Cells. *Molecular cell* **47**(4): 633-647.

Broske AM, Vockentanz L, Kharazi S, Huska MR, Mancini E, Scheller M, Kuhl C, Enns A, Prinz M, Jaenisch R et al. 2009. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nature genetics* **41**(11): 1207-1215.

Bultman SJ, Klebig ML, Michaud EJ, Sweet HO, Davisson MT, Woychik RP. 1994. Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (a(t)) and white-bellied agouti (Aw) reveals alternative forms of agouti transcripts. *Genes & development* **8**(4): 481-490.

Cleveland WS, Devlin SJ, Grosse E. 1988. Regression by Local Fitting. *Journal of Econometrics* **37**: 87-114.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**(50): 21931-21936.

Cuddapah S, Barski A, Zhao K. 2010. Epigenomics of T cell activation, differentiation, and memory. *Current opinion in immunology* **22**(3): 341-347.

Delogu A, Schebesta A, Sun Q, Aschenbrenner K, Perlot T, Busslinger M. 2006. Gene repression by Pax5 in B cells is essential for blood cell homeostasis and is reversed in plasma cells. *Immunity* **24**(3): 269-281.

Ding LH, Xie Y, Park S, Xiao G, Story MD. 2008. Enhanced identification and biological validation of differential gene expression via Illumina whole-genome expression arrays through the use of the model-based background correction methodology. *Nucleic acids research* **36**(10): e58.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**(5): 1792-1797.

ENCODE Project Consortium 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS biology* **9**(4): e1001046.

Englander EW, Howard BH. 1995. Nucleosome positioning by human *Alu* elements in chromatin. *The Journal of biological chemistry* **270**(17): 10091-10096.

Englander EW, Wolffe AP, Howard BH. 1993. Nucleosome interactions with a human *Alu* element. Transcriptional repression and effects of template methylation. *The Journal of biological chemistry* **268**(26): 19565-19573.

Fujita N, Jaye DL, Geigerman C, Akyildiz A, Mooney MR, Boss JM, Wade PA. 2004. MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. *Cell* **119**(1): 75-86.

Good KL, Tangye SG. 2007. Decreased expression of Kruppel-like factors in memory B cells induces the rapid response typical of secondary antibody responses. *Proceedings of the National Academy of Sciences of the United States of America* **104**(33): 13420-13425.

Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, Gartler SM. 1999. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proceedings of the National Academy of Sciences of the United States of America* **96**(25): 14412-14417.

Harada H, Kawano MM, Huang N, Harada Y, Iwato K, Tanabe O, Tanaka H, Sakai A, Asaoku H, Kuramoto A. 1993. Phenotypic difference of normal plasma cells from mature myeloma cells. *Blood* **81**(10): 2658-2663.

Hastie TJ, Tibshirani RJ. 1990. *Generalized Additive Models*. Chapman & Hall, New York.

Hellmann-Blumberg U, Hintz MF, Gatewood JM, Schmid CW. 1993. Developmental differences in methylation of human *Alu* repeats. *Molecular and cellular biology* **13**(8): 4523-4530.

Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, Uren PJ, Park J, Butler J, Rafii S, McCombie WR et al. 2011. Directional DNA methylation changes and complex

intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Molecular cell* **44**(1): 17-28.

Jaye DL, Iqbal J, Fujita N, Geigerman CM, Li S, Karanam S, Fu K, Weisenburger DD, Chan WC, Moreno CS et al. 2007. The BCL6-associated transcriptional co-repressor, MTA3, is selectively expressed by germinal centre B cells and lymphomas of putative germinal centre derivation. *J Pathol* **213**(1): 106-115.

Jeong S, Liang G, Sharma S, Lin JC, Choi SH, Han H, Yoo CB, Egger G, Yang AS, Jones PA. 2009. Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. *Molecular and cellular biology* **29**(19): 5366-5376.

Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P, Lee H, Aryee MJ, Irizarry RA, Kim K et al. 2010. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature*.

Johnson EW, Li C. 2007. Adjusting batch effects in microarray expression data using Empirical Bayes methods. *Biostatistics* **8**(1): 118-127.

Jones PA, Liang G. 2009. Rethinking how DNA methylation patterns are maintained. *Nature reviews Genetics* **10**(11): 805-811.

Kersh EN, Fitzpatrick DR, Murali-Krishna K, Shires J, Speck SH, Boss JM, Ahmed R. 2006. Rapid demethylation of the IFN-gamma gene occurs in memory but not naive CD8 T cells. *J Immunol* **176**(7): 4083-4093.

Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, Kim J, Aryee MJ, Ji H, Ehrlich LI et al. 2010. Epigenetic memory in induced pluripotent stem cells. *Nature*.

Klein U, Tu Y, Stolovitzky GA, Keller JL, Haddad J, Jr., Miljkovic V, Cattoretti G, Califano A, Dalla-Favera R. 2003. Transcriptional analysis of the B cell germinal center reaction. *Proceedings of the National Academy of Sciences of the United States of America* **100**(5): 2639-2644.

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**(11): 1571-1572.

Lai AY, Fatemi M, Dhasarathy A, Malone C, Sobol SE, Geigerman C, Jaye DL, Mav D, Shah R, Li L et al. 2010. DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. *The Journal of experimental medicine*.

Lanzavecchia A, Sallusto F. 2009. Human B cell memory. *Current opinion in immunology* **21**(3): 298-304.

Laperriere D, Wang TT, White JH, Mader S. 2007. Widespread *Alu* repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution. *BMC genomics* **8**: 23.

Leonhardt H, Page AW, Weier HU, Bestor TH. 1992. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71**(5): 865-873.

Li E, Bestor TH, Jaenisch R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**(6): 915-926.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**(7271): 315-322.

Liu YJ, Barthelemy C, de Bouteiller O, Arpin C, Durand I, Banchereau J. 1995. Memory B cells from human tonsils colonize mucosal epithelium and directly present antigen to T cells by rapid up-regulation of B7-1 and B7-2. *Immunity* **2**(3): 239-248.

Liu YJ, Joshua DE, Williams GT, Smith CA, Gordon J, MacLennan IC. 1989. Mechanism of antigen-driven selection in germinal centres. *Nature* **342**(6252): 929-931.

Matthias P, Rolink AG. 2005. Transcriptional networks in developing and mature B cells. *Nat Rev Immunol* **5**(6): 497-508.

McHeyzer-Williams LJ, McHeyzer-Williams MG. 2005. Antigen-specific memory B cell development. *Annual review of immunology* **23**: 487-513.

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB et al. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**(7205): 766-770.

Messi M, Giacchetto I, Nagata K, Lanzavecchia A, Natoli G, Sallusto F. 2003. Memory and flexibility of cytokine gene expression as separable properties of human T(H)1 and T(H)2 lymphocytes. *Nature immunology* **4**(1): 78-86.

Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A. 2008. Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**(7200): 49-55.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**(7153): 553-560.

Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, Stivers JT, Zhang Y, Kohli RM. 2012. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nature chemical biology* **8**(9): 751-758.

Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**(3): 247-257.

Pascual V, Liu YJ, Magalski A, de Bouteiller O, Banchereau J, Capra JD. 1994. Analysis of somatic mutation in five B cell subsets of human tonsil. *The Journal of experimental medicine* **180**(1): 329-339.

Ponicsan SL, Kugel JF, Goodrich JA. 2010. Genomic gems: SINE RNAs regulate mRNA production. *Current opinion in genetics & development* **20**(2): 149-155.

Ranuncolo SM, Polo JM, Dierov J, Singer M, Kuo T, Greally J, Green R, Carroll M, Melnick A. 2007. Bcl-6 mediates the germinal center B cell phenotype and lymphomagenesis through transcriptional repression of the DNA-damage sensor ATR. *Nature immunology* **8**(7): 705-714.

Rauch T, Li H, Wu X, Pfeifer GP. 2006. MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer research* **66**(16): 7939-7947.

Scacheri PC, Crawford GE, Davis S. 2006. Statistics for ChIP-chip and DNase hypersensitivity experiments on NimbleGen arrays. *Methods in enzymology* **411**: 270-282.

Serre D, Lee BH, Ting AH. 2010. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic acids research* **38**(2): 391-399.

Shaknovich R, Cerchietti L, Tsikitas L, Kormaksson M, De S, Figueroa ME, Ballon G, Yang SN, Weinhold N, Reimers M et al. 2011. DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. *Blood* **118**(13): 3559-3569.

Shapiro-Shelef M, Calame K. 2005. Regulation of plasma-cell development. *Nat Rev Immunol* **5**(3): 230-242.

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D et al. 2011a. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**(7378): 490-495.

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK et al. 2011b. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**(7378): 490-495.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**(43): 15545-15550.

Tadokoro Y, Ema H, Okano M, Li E, Nakauchi H. 2007. De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. *The Journal of experimental medicine* **204**(4): 715-722.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**(4): 663-676.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B et al. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**(7414): 75-82.

Victora GD, Nussenzweig MC. 2012. Germinal centers. *Annual review of immunology* **30**: 429-457.

Wand MP. 1997. Data-Based Choice of Histogram Bin Width. *The American Statistician* **51**: 59-64.

Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature genetics* **39**(4): 457-466.

Weissman IL. 2000. Stem cells: units of development, units of regeneration, and units in evolution. *Cell* **100**(1): 157-168.

Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, Bernstein BE, Jaenisch R. 2007. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**(7151): 318-324.

Wu H, Coskun V, Tao J, Xie W, Ge W, Yoshikawa K, Li E, Zhang Y, Sun YE. 2010. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* **329**(5990): 444-448.

Wu SC, Zhang Y. 2010. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* **11**(9): 607-620.

Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Pequignot E. 1999. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**(6758): 187-191.

Yamane A, Resch W, Kuo N, Kuchen S, Li Z, Sun HW, Robbiani DF, McBride K, Nussenzweig MC, Casellas R. 2011. Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nature immunology* **12**(1): 62-69.