Intermediately Methylated Regions in Normal Cells Are Epimutation Hotspots in Cancer

Mohamed Mahgoub[1], Haley Abel[1], Nidhi Davarapalli[1], Heidi Struthers[1], Emilee Kotnik[1], Brittany Johnson[1], Christopher Markovic[2], Catrina Fronick[2], Robert Fulton[2], Michael P. Meers[3], and David H. Spencer*[1,2,4]

[1]Department of Medicine, Division of Oncology, Washington University School of Medicine; St. Louis, MO USA

[2]McDonnell Genome Institute, Washington University School of Medicine; St. Louis, MO USA

[3]Department of Genetics, Washington University School of Medicine; St. Louis, MO USA

[4]Department of Pathology and Immunology, Washington University School of Medicine; St. Louis, MO USA

* Corresponding author: dspencer@wustl.edu

Abstract

DNA methylation is altered in all cancers, but the mechanisms responsible for these changes are not well understood. Using data from 100 primary samples, we show that regions with intermediate methylation in normal hematopoietic cells are hotspots for stable, clonal epimutations in acute myeloid leukemia. Analysis of hematopoietic stem cell clones demonstrated that intermediate methylation at epimutation hotspots reflects random allele-specific methylation at the single cell level, representing a previously unrecognized form of somatically acquired imprinting. We identified somatically imprinted regions in other tissues and show they are epimutation hotspots in multiple cancer types. This demonstrates that random allele-specific methylation is both a general property of normal cells and a vulnerability that renders them susceptible to cooperating events and clonal selection during cancer.

29 **Main text**

30 Cancer is caused by heritable somatic events that endow neoplastic cells with a selective
31 advantage. While the most extensively studied somatic events in cancer are genetic mutations,
32 cytosine methylation of DNA is also heritable through cell divisions and exhibits tissue-specific
33 changes in malignant vs. normal cells. These include genome-wide methylation patterns(*1–3*)
34 and signatures that reflect a tumor's cell of origin, subtype, or the effects of somatic
35 mutations(*4–15*). Genome-wide studies have shown that abnormally methylated CpGs in cancer
36 frequently cluster into differentially methylated regions (DMRs)(*4*, *5*, *16*, *17*). However, factors
37 that determine why certain regions acquire these changes are poorly understood. This has made it
38 difficult to define the mechanisms responsible for cancer-associated methylation changes, and to
39 distinguish early events that could initiate cancer from alterations that are consequences of it(*4*,
40 *18*).

41 Epimutations are spontaneous, heritable epigenetic abnormalities that can affect gene
42 activity and cause human disease, including cancer(*19*, *20*). The best-characterized epimutations
43 are inherited events that cause the imprinting disorders Beckwith-Wiedemann, Angelman, and
44 Prader-Willi syndromes(*21*, *22*), or predispose patients to cancer via silencing of tumor
45 suppressors(*23*, *24*). Epimutations can also be somatically acquired, most notably in cancer.
46 These can also silence tumor suppressors, producing a second epigenetic 'hit' in tumors with a
47 co-occurring somatic mutation(*25*). Although the mechanisms and timing of somatic
48 epimutations are not well understood, there is evidence they may pre-exist in normal tissues(*26–*
49 *30*) and could therefore be selected for during cancer development.

50 Although methylation in cancer has been extensively studied, most analysis has focused
51 on changes associated with clinical or molecular features(*4*, *9*, *31–35*). Such approaches may be

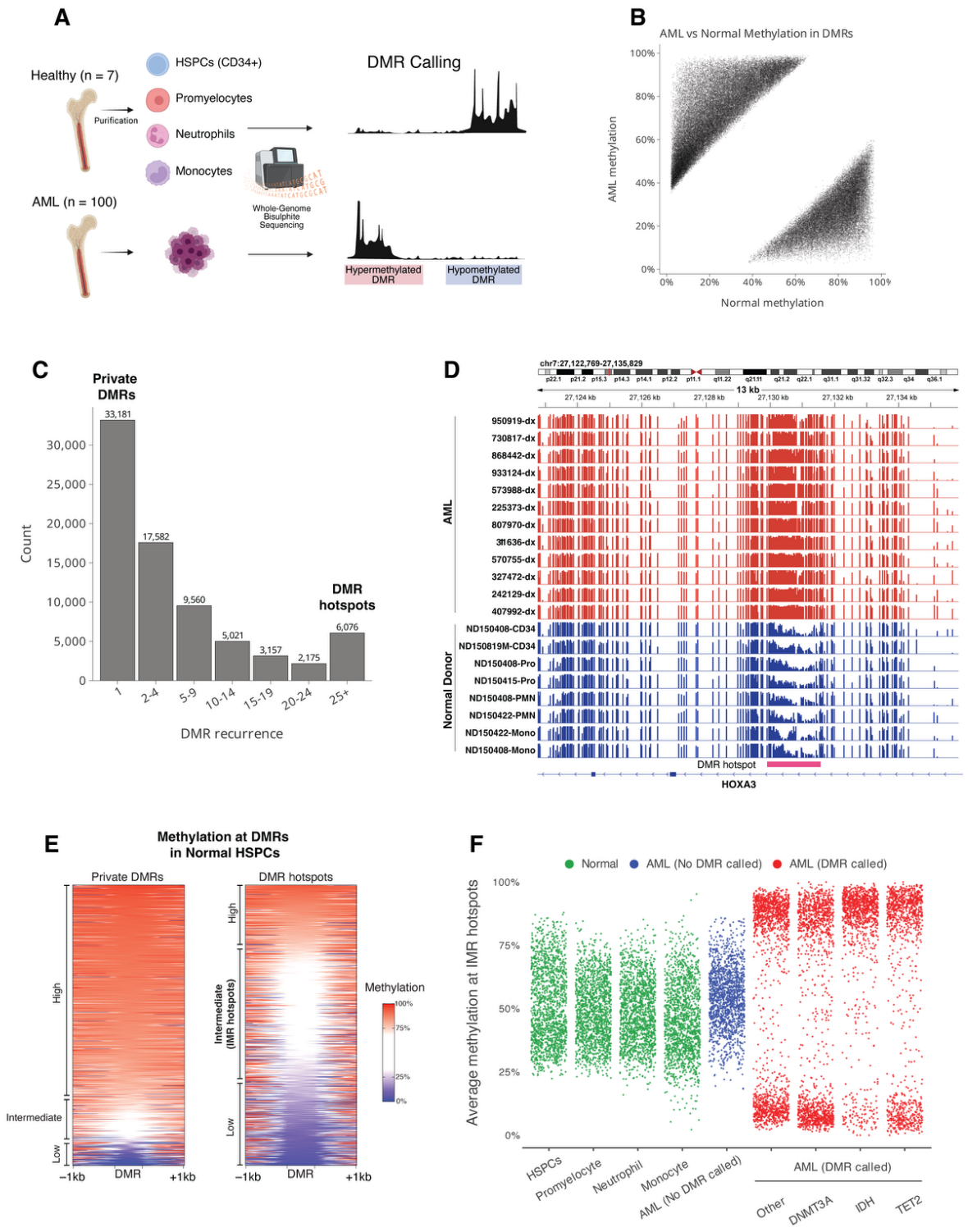
52 less sensitive to detecting spontaneously arising epimutations, which could be independent of
53 these variables. Here, we used a genome-wide analysis to identify somatically acquired DNA
54 methylation epimutations in acute myeloid leukemia (AML). Surprisingly, these events
55 frequently occur in regions with intermediate methylation in normal hematopoietic cells, which
56 harbor blocks of CpGs with random allele-specific methylation in single cells. We demonstrate
57 that methylation at these blocks affects chromatin and gene expression, and that these
58 "somatically imprinted" regions are epimutation hotspots in AML and other cancers.
59

60 **Intermediately methylated regions in normal hematopoietic cells are hotspots for altered**
61 **methylation in AML**

62 To identify DNA methylation changes in AML that could be somatically acquired
63 epimutations, we compared whole-genome bisulfite sequencing (WGBS) data from 100 primary
64 AML samples to WGBS from 16 purified hematopoietic cell populations from 7 healthy
65 donors(*4*) (Figure 1A, S1A). Pairwise comparisons were performed between each AML sample
66 to pooled data from stem/progenitor and mature myeloid cells (see Methods) to detect common
67 and rare events and account for maturation-related methylation changes in AML. Between 752
68 and 19,961 autosomal differentially methylated regions (DMRs) with a methylation difference
69 >0.3 were identified in each sample, for a total of 575,783 DMRs (Figure 1B, S1B). Merging of
70 all DMRs produced 76,752 unique regions, of which 33,181 (43.2%) were "private DMRs" in a
71 single AML sample, and 43,571 (56.8%) regions that overlapped DMRs identified at least twice
72 (Figure 1C). Applying a threshold of ≥25 AML samples resulted in 6,076 regions overlapping
73 recurrent DMRs that we defined as DMR "hotspots" that included regions known to have altered
74 methylation in AML (e.g., *HOXA3*, Figure 1D).

We next investigated the normal epigenetic features at hotspots by annotating them with chromatin states(*36*) from hematopoietic stem/progenitor cells (HSPCs) (Figure S1C, see Methods). Hotspots were >10-fold enriched for bivalent chromatin and transcriptional initiation compared to private DMRs, which were evenly distributed across all states (Figure S1D-F). Interestingly, states with the highest enrichment for hotspots had intermediate methylation in normal HSPCs, an uncommon genomic feature that marks tissue-specific regulatory elements(*37*). Direct analysis showed that 48% of hotspots had intermediate methylation in normal HPSCs, and there was a 5-fold enrichment for intermediately methylated CpGs in hotspots compared to their genome-wide frequency (Figure 1E).

We next defined hematopoietic-specific intermediately methylated regions (IMRs) with methylation between 0.3 and 0.7 in normal hematopoietic cells and evaluated the hotspots in these IMRs (i.e., "IMR hotspots") for features of somatic epimutations. Interestingly, methylation at IMR hotspots remained consistently intermediate in promyelocytes, monocytes, and neutrophils (Figure 1F). Methylation at IMR hotspots was also stable between presentation and relapse in 6 AML patients (Figure S1H). Further, AML samples with DMRs at IMR hotspots showed consistent hyper- or hypomethylation at each hotspot, including samples with mutations in *DNMT3A*, *IDH1*, *IDH2*, or *TET2* (Figures 1F, S1I) that cause directional methylation changes in AML(*4*, *5*, *9*, *35*). This implies that DMRs at IMR hotspots are not simply driven by the presence of these mutations. Finally, there was a modest but significant negative correlation between methylation at promoter-associated IMR hotspots and expression of nearby genes, supporting their functional role in gene regulation (Figure S1J). Together this demonstrates that intermediately methylated regions in normal cells are hotspots for methylation changes in AML that resemble somatically acquired epimutations.


A
Healthy (n = 7)
Purification
HSPCs (CD34+)
Promyelocytes
Neutrophils
Monocytes
AML (n = 100)
Whole-Genome Bisulphite Sequencing
DMR Calling
Hypermethylated DMR
Hypomethylated DMR
B
AML vs Normal Methylation in DMRs
AML methylation
Normal methylation
C
Private DMRs
DMR hotspots
Count
DMR recurrence
33,181
17,582
9,560
5,021
3,157
2,175
6,076
D
chr7:27,122,769-27,135,829
13 kb
AML
Normal Donor
DMR hotspot
HOXA3
E
Methylation at DMRs in Normal HSPCs
Private DMRs
DMR hotspots
High
Intermediate
Low
Intermediate (IMR hotspots)
Methylation
−1kb
DMR
+1kb
F
Normal
AML (No DMR called)
AML (DMR called)
Average methylation at IMR hotspots
HSPCs
Promyelocyte
Neutrophil
Monocyte
AML (No DMR called)
AML (DMR called)
Other
DNMT3A
IDH
TET2

98

99 **Fig. 1**. **Recurrent differentially methylated regions in AML are enriched for intermediately**
100 **methylated regions in normal hematopoietic cells.**
101 (**A**) Analysis of differentially methylated regions (DMRs) in 100 primary AML samples. WGBS
102 data from each AML was compared to a pooled dataset of 16 hematopoietic cell populations
103 purified from bone marrow aspirates from 7 distinct healthy donors. (**B**) Scatter plot showing the
104 methylation level of each DMR identified in all AML samples versus mean methylation in normal
105 hematopoietic cells. (**C**) Recurrence of DMRs across AMLs after merging overlapping regions
106 from all samples. Private DMRs are regions with differential methylation in one AML sample,
107 while DMR hotspots are regions overlapping highly recurrent DMRs identified in at least 25 AML
108 samples. (**D**) DNA methylation at a DMR hotspot in the *HOXA3* locus from normal hematopoietic
109 cells and selected AML samples. (**E**) Heatmap showing methylation in normal hematopoietic
110 stem/progenitors (HSPCs) at regions overlapping private DMRs identified in only one AML
111 sample (left) and DMR hotspots (25+ samples) (right). DMRs are length-normalized and shown
112 with 1 kb of flanking sequence. (**F**) Methylation levels at intermediately methylated regions in
113 normal hematopoietic cells that are hotspots for differential methylation in AML samples (IMR
114 hotspots). Each point shows the mean methylation of IMR hotspot in the indicated normal
115 hematopoietic cell population (green), AMLs with no DMR called at the hotspot (blue), and AMLs
116 in which a DMR was called (red). AML samples with a DMR at the hotspots are grouped by the
117 presence of mutations in selected genes (minimum of 2 AML samples with the DMR hotspot per
118 mutation group).
119
120
121

122 **Intermediately methylated regions contain CpG blocks with allelic methylation in AML**

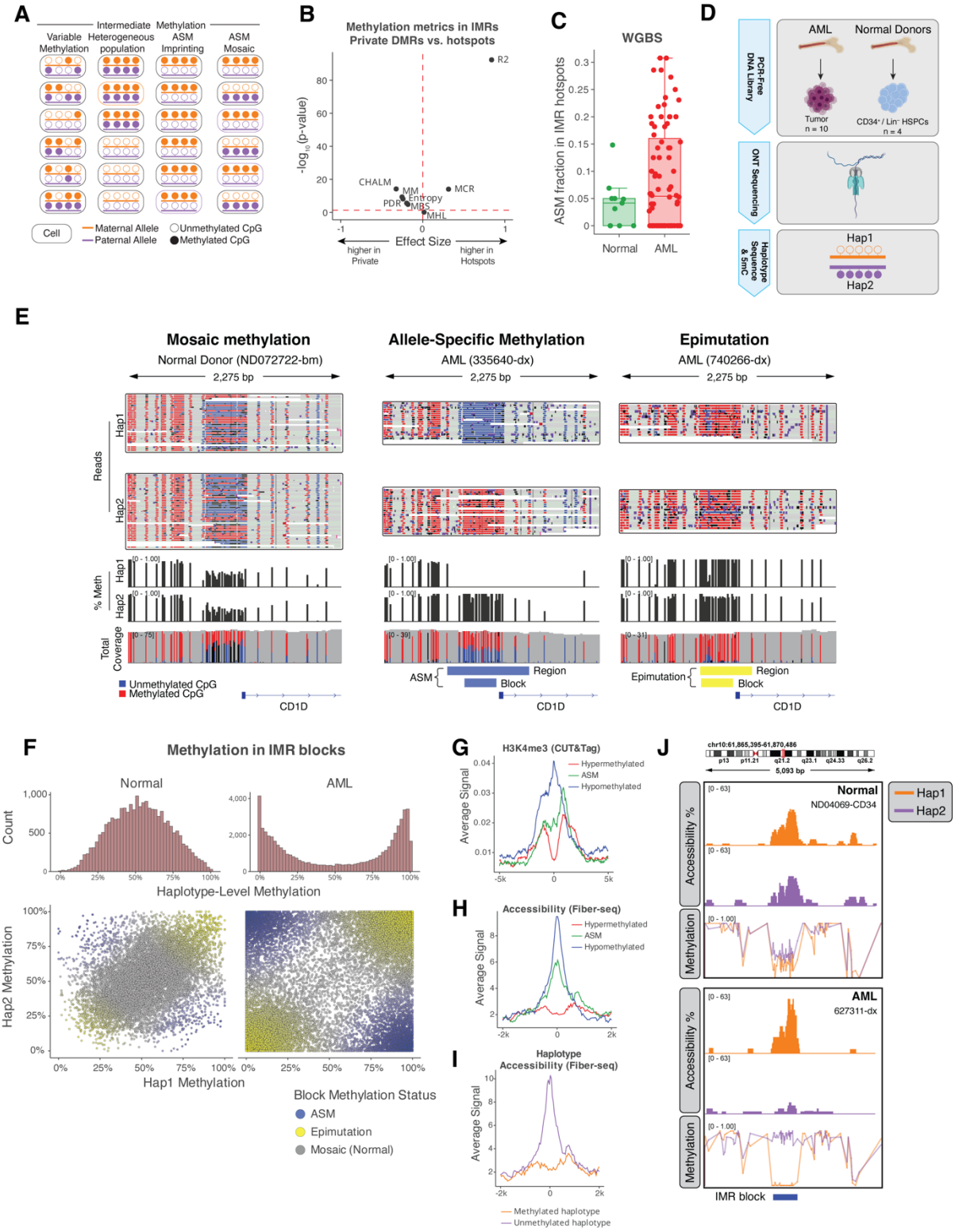
123 Intermediate methylation could be due to cellular heterogeneity, discordant methylation
124 across CpGs, or allele-specific methylation (Figure 2A). To distinguish between these
125 possibilities, we analyzed fragment-level methylation (*38*, *39*) at IMR hotspots in WGBS data
126 from purified normal hematopoietic cells, which showed that methylation at consecutive CpGs
127 was highly correlated compared to non-hotspot DMRs with intermediate methylation ($p<10^{-10}$,
128 Figure 2B, S2A). We next tested for allele-specific methylation (ASM) at IMR hotspots that
129 contained heterozygous SNPs in the normal and AML samples. Surprisingly, this showed that
130 ASM was rare in normal samples but more common in AML, especially in samples with hotspots
131 that retained intermediate methylation (Figure 2C, S2B-C). To further explore observation, we
132 used Oxford Nanopore sequencing (ONT) to obtain haplotype-resolved methylation from 10
133 primary AML samples and CD34+ HSPCs or immature (lineage marker-negative) bone marrow
134 cells from four healthy donors (Figure 2D). We identified 1,264-4,497 autosomal regions of
135 ASM per AML sample, after excluding regions in known imprinted genes, near structural
136 variants, or with high SNP density; in comparison, ~300 ASM regions were identified in normal
137 samples (Figure S2D). Interestingly, 30% of the IMR hotspots overlapped an ASM region in the
138 ONT data from AML samples, supporting our findings from WGBS.

139 We next used the long-read ONT methylation data from all AML samples and normal
140 HSPCs to identify blocks of consecutive, highly correlated CpGs(*40*) that may be driving allelic
141 methylation at IMR hotspots. We also included published IMRs(*37*) and ASM regions from the
142 ONT data in this analysis, which we reasoned could share similar genomic features. This resulted
143 in 11,534 total methylation haplotype blocks, of which 1,575 had consistent intermediate
144 methylation in normal HSPCs (between 0.3 and 0.7), which we defined as "IMR blocks". These

145 blocks contained a mean of 12 CpGs, spanned 186 bp, and showed highly correlated—but
146 mosaic—patterns in HSPCs (Figure 2E). Analysis of haplotype-level methylation at all IMR
147 blocks from HSPCs showed that both haplotypes had intermediate methylation, matching the
148 aggregate methylation levels from ONT and WGBS (Figure 2F). In contrast, most IMR blocks in
149 AML samples either had ASM (Figure 2E and 2F: indicated in blue) or were fully
150 methylated/unmethylated on both alleles, reflecting an epimutation (Figure 2E and 2F: indicated
151 in yellow). These findings demonstrate the unique properties of IMR blocks, which have mosaic
152 haplotype-level methylation in normal cells that converts to either ASM or fully
153 methylated/unmethylated epimutations in AML.
154
155 **Methylation at IMR blocks is associated with differential chromatin activity**
156 We next investigated whether methylation at IMR blocks is associated with differential
157 histone modifications using CUT&Tag and haplotype-level chromatin accessibility via Fiber-
158 seq(*41*). IMR blocks in HSPCs possessed both H3K27 and H3K4 trimethylation, consistent with
159 the bivalent chromatin states at hotspots from the HSPC-defined chromHMM model (Figure
160 S2E). In AML samples, both modifications were negatively correlated with methylation at IMR
161 blocks, with ASM blocks having intermediate signal consistent with allelic histone modification
162 differences (Figure 2G, S2F). Likewise, a negative correlation in aggregate accessibility at IMR
163 blocks was observed via Fiber-seq in two AML samples (Figure 2H). This was also apparent at
164 the haplotype level, with unmethylated haplotypes having higher accessibility in aggregate and at
165 individual regions (Figure 2I, 2J). Overall, these findings are consistent with previous analysis of
166 histone modifications at IMRs(*37*) and support a functional relationship where unmethylated

167 alleles at IMR blocks have greater accessibility and increased histone methylation compared to

168 methylated blocks.


A
Variable Methylation
Intermediate Heterogeneous population
Methylation ASM Imprinting
ASM Mosaic
Cell
Maternal Allele
Paternal Allele
Unmethylated CpG
Methylated CpG
B
Methylation metrics in IMRs
Private DMRs vs. hotspots
-log10 (p-value)
R2
CHALM
MM
PDR
Entropy
MBS
MHL
MCR
Effect Size
higher in Private
higher in Hotspots
C
WGBS
ASM fraction in IMR hotspots
Normal
AML
D
PCR-Free DNA Library
AML
Normal Donors
Tumor n = 10
CD34+ / Lin- HSPCs n = 4
ONT Sequencing
Haplotype Sequence & 5mC
Hap1
Hap2
E
Mosaic methylation
Normal Donor (ND072722-bm)
Allele-Specific Methylation
AML (335640-dx)
Epimutation
AML (740266-dx)
2,275 bp
Reads
Hap1
Hap2
% Meth
Total Coverage
Unmethylated CpG
Methylated CpG
ASM
Region
Block
Epimutation
CD1D
F
Methylation in IMR blocks
Normal
AML
Count
Haplotype-Level Methylation
Hap2 Methylation
Hap1 Methylation
Block Methylation Status
ASM
Epimutation
Mosaic (Normal)
G
H3K4me3 (CUT&Tag)
Hypermethylated
ASM
Hypomethylated
Average Signal
H
Accessibility (Fiber-seq)
I
Haplotype Accessibility (Fiber-seq)
Methylated haplotype
Unmethylated haplotype
J
chr10:61,865,395-61,870,486
5,093 bp
Normal
ND04069-CD34
AML
627311-dx
Hap1
Hap2
Accessibility %
Methylation
IMR block

169

170 **Fig. 2. Intermediately methylated regions in normal hematopoietic cells display allelic**
171 **methylation patterns in AML**.
172 **(A)** Models of possible fragment-level methylation patterns at intermediately methylated regions
173 (IMRs). **(B)** Comparison of WGBS fragment-level methylation metrics at IMRs overlapping DMR
174 hotspots vs. non-hotspot DMRs in normal hematopoietic cells. Each point represents a different
175 metric, with the x-axis indicating the effect size (standardized difference in means between
176 hotspots and non-hotspots) and the y-axis indicating statistical significance as $-\log_{10}$(p-value) from
177 a t-test comparing the two groups. Metrics include PDR (Proportion of Discordant Reads), Entropy
178 (Shannon entropy), CHALM (Cell Heterogeneity-Adjusted cLoNal Methylation), R^2 (linkage
179 disequilibrium R squared), MHL (Methylation Haplotype Load), MCR (Methylation Concordance
180 Ratio), and MBS (Methylation Block Score). **(C)** Fraction of IMR hotspots exhibiting allele-
181 specific methylation (ASM) per sample from WGBS in normal hematopoietic cells and AML
182 samples, calculated among all IMR hotspots with a heterozygous SNP and sufficient coverage.
183 Only samples with at least 5 informative IMR hotspots are included. **(D)** ONT sequencing and
184 haplotype-resolved methylation analysis of normal HSPCs (N=4) and AML samples (N=10). **(E)**
185 Example of haplotype-resolved methylation in an IMR hotspot (at the promoter of *CD1D*),
186 showing mosaic methylation in normal hematopoietic cells, allele-specific methylation (ASM) in
187 AML sample 335640, and full methylation (epimutation) in AML sample 740266. Tracks show
188 the DMR region identified in each AML sample vs. normal cells using DSS R package and the
189 "IMR block" defined by correlation of fragment-level CpG methylation using ONT data(*40*). **(F)**
190 ONT haplotype-level methylation at IMR blocks in all normal HSPCs (N=4) and AML samples
191 (N=10). Top panels show the distribution haplotype-level methylation at IMR blocks. Bottom
192 panels show scatter plots comparing average methylation of Haplotype 1 versus Haplotype 2 for

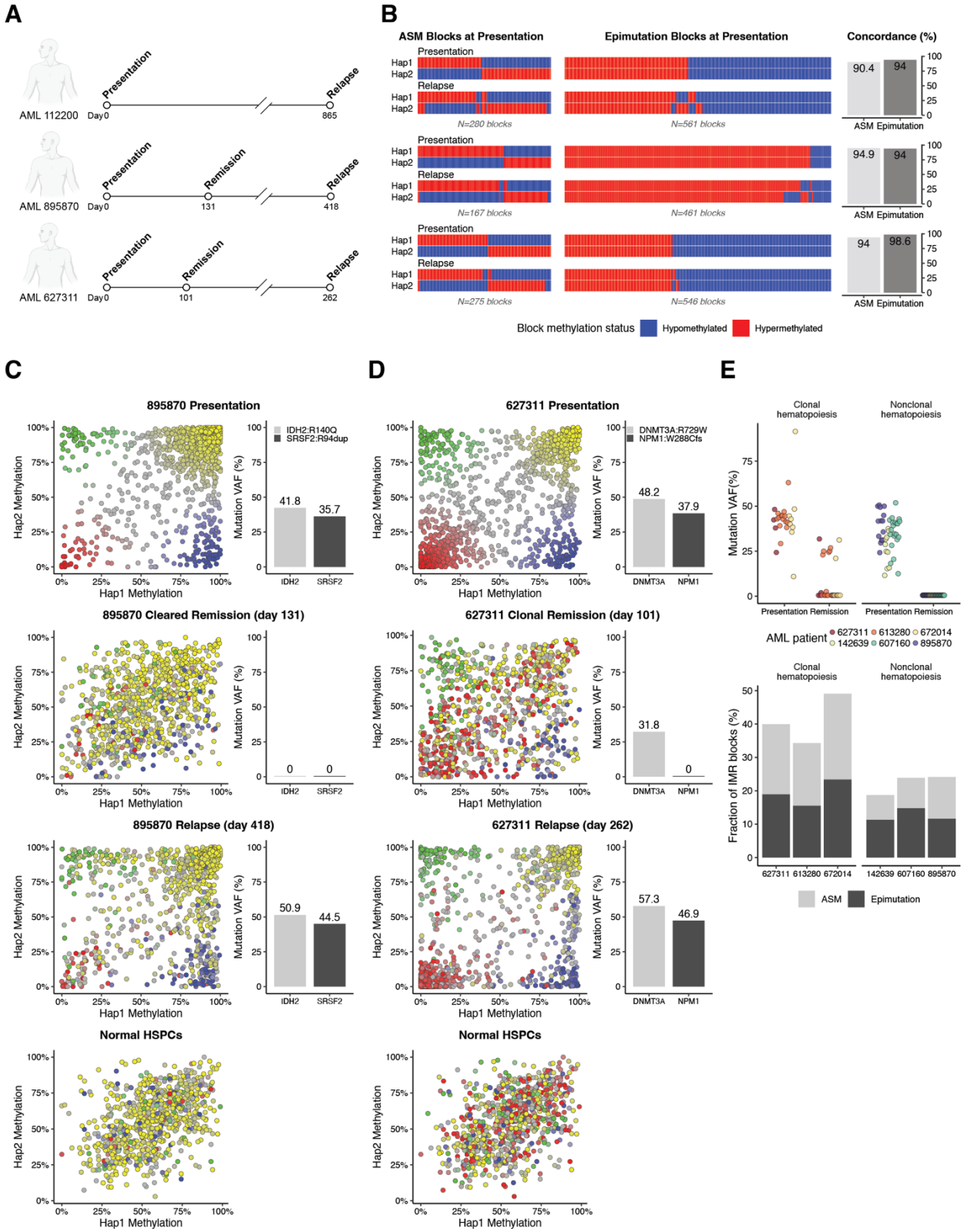
193 each IMR block. **(G and H)** Aggregated H3K4me3 CUT&Tag signal (G) and chromatin
194 accessibility (Fiber-seq) (H) at IMR blocks in AML samples, with blocks grouped by methylation
195 status (hypomethylated, intermediately methylated, or hypermethylated). **(I)** Chromatin
196 accessibility (Fiber-seq) at IMR blocks with intermediate methylation (green in panel H) and
197 exhibiting ASM in AML samples, comparing accessibility between the methylated and
198 unmethylated haplotypes. **(J)** Example of haplotype-resolved methylation and chromatin
199 accessibility at an IMR block in an intergenic region on chromosome 10 (chr10:61,866,122-
200 61,869,606).
201

202 **Allelic methylation at IMR blocks is stable over time and marks clonal hematopoiesis**

203 To determine the stability of allelic methylation at IMR blocks, we used ONT sequencing
204 to analyze paired presentation/relapse samples from 3 AML patients who relapsed after
205 chemotherapy with a related AML clone (Figure 3A, S3A). We compared IMR blocks with
206 allelic methylation (i.e., either ASM or a methylated or unmethylated epimutation) between
207 presentation and relapse samples for each patient (using synchronized haplotype designations),
208 which demonstrated that >92.5% of the ASM blocks and >98.9% of the blocks with epimutations
209 were stable at relapse. We next tested whether allelic methylation at IMR blocks is specific to
210 leukemic cells by analyzing remission samples from two of the AML patients. Whole-genome
211 sequencing(*42*) confirmed a molecular remission in patient 895870, while patient 627311 had a
212 persistent *DNMT3A* mutation and several noncoding exome variants, indicating the presence of
213 clonal hematopoiesis (Figure S3B). Patient 895870 had allelic methylation at IMR blocks at
214 presentation that nearly completely normalized in the remission sample before returning at
215 relapse (Figure 3C). Interestingly, allelic methylation only partially normalized in the remission

216 sample from patient 627311 that had clonal hematopoiesis, which instead retained a pattern
217 closer to the presentation and relapse samples (Figure 3D).

218 To further examine the relationship between allelic methylation at IMR blocks and clonal
219 hematopoiesis, we used ONT sequencing to measure methylation in additional remission
220 samples from 4 patients who had either complete clearance of somatic mutations or multiple
221 persistent mutations by exome sequencing, indicating clonal hematopoiesis(*42*) (Figure 3E,
222 S3C). Quantification of IMR blocks with allelic methylation showed that patients with clonal
223 hematopoiesis had a significantly higher fraction of allelic blocks compared to samples with
224 complete mutation clearance (41% vs. 22%; p=0.034, Student's *t*), supporting an association of
225 ASM and epimutations with clonal hematopoiesis.


A
AML 112200
Presentation
Relapse
Day 0
865
AML 895870
Presentation
Remission
Relapse
Day 0
131
418
AML 627311
Presentation
Remission
Relapse
Day 0
101
262
B
ASM Blocks at Presentation
Epimutation Blocks at Presentation
Concordance (%)
Presentation
Relapse
Hap1
Hap2
N=280 blocks
N=561 blocks
90.4
94
N=167 blocks
N=461 blocks
94.9
94
N=275 blocks
N=546 blocks
94
98.6
ASM
Epimutation
Block methylation status
Hypomethylated
Hypermethylated
C
895870 Presentation
IDH2:R140Q
SRSF2:R94dup
41.8
35.7
895870 Cleared Remission (day 131)
0
0
895870 Relapse (day 418)
50.9
44.5
Normal HSPCs
D
627311 Presentation
DNMT3A:R729W
NPM1:W288Cfs
48.2
37.9
627311 Clonal Remission (day 101)
31.8
0
627311 Relapse (day 262)
57.3
46.9
Normal HSPCs
Hap2 Methylation
Hap1 Methylation
Mutation VAF (%)
IDH2
SRSF2
DNMT3A
NPM1
E
Clonal hematopoiesis
Nonclonal hematopoiesis
Mutation VAF(%)
Presentation Remission
AML patient
627311
613280
672014
142639
607160
895870
Fraction of IMR blocks (%)
ASM
Epimutation

226

227 **Fig. 3. IMR blocks are stable, clonal events in hematopoietic cell populations. (A)**
228 Longitudinal analysis of methylation at IMR blocks in 3 AML patients (112200, 895870, and
229 627311) who achieved a complete morphologic remission with standard "7+3" induction
230 chemotherapy. ONT sequencing was performed on the samples at the timepoints indicated by the
231 open circles. **(B)** Heatmap representation of haplotype-level methylation from ONT sequencing of
232 paired presentation/relapse sample pairs. IMR blocks were classified as either having ASM (left)
233 or a fully methylated/unmethylated epimutation (right) via Fisher's exact test on methylation read
234 counts either between haplotypes (for ASM) or compared to normal HSPCs (for epimutations).
235 Methylated and unmethylated IMR blocks are shown in red and blue, respectively. Only IMR
236 blocks with statistically significant classifications at both presentation and relapse are shown. The
237 percent of concordant ASM and epimutation blocks between presentation and relapse is shown on
238 the right. **(C)** Haplotype-level methylation at IMR blocks from AML patient 895870 at
239 presentation, complete morphologic and molecular remission (i.e., blasts <1% and no leukemia-
240 associated mutations detected), and relapse. Points show the methylation of IMR blocks in
241 haplotype 2 vs. haplotype 1 and are colored by their methylation status in the presentation sample.
242 Bar plots show the variant allele fraction (VAF) of selected recurrent mutations in each sample.
243 Bottom panel shows haplotype-level methylation at the same IMR blocks from a representative
244 CD34+ HSPC sample and are colored based on the methylation status of the block in the
245 presentation AML sample. **(D)** Haplotype-level methylation in AML patient 627311 at
246 presentation, complete morphologic remission (bone marrow blasts <1%), but with a persistent
247 *DNMT3A*, and relapse. Scatter plots are shown as in (C). Bar plots show mutation VAFs in each
248 sample and indicate the presence of a *DNMT3A*^{R729W} variant in the remission sample, consistent
249 with persistent clonal hematopoiesis. **(E)** Correlation between clonal hematopoiesis in AML

250 remission and the presence of ASM and epimutations at IMR blocks from ONT sequencing. Top
251 panels show mutation VAFs at presentation and complete remission for 6 patients, 3 with clonal
252 hematopoiesis in remission (left) and 3 with nonclonal hematopoiesis. Bottom panel shows the
253 fraction of all IMR blocks (N=1,575) that met criteria for being ASM or fully
254 methylated/unmethylated epimutations compared to normal HSPCs in each sample using a
255 Fisher's exact test on haplotype-level methylation read counts. All remission samples were bone
256 marrow aspirates with normal trilineage hematopoiesis and <1% blasts. See reference (*42*) .
257

258 **Random allele-specific methylation at IMR blocks is somatically acquired in normal**
259 **hematopoietic stem cells**

260 We reasoned that intermediate methylation at IMR blocks could reflect random ASM that
261 is developmentally established in HSPCs at the single cell level and thus can be observed in
262 clonal AML samples and clonal hematopoiesis. To test this, we examined IMR block methylation
263 in ONT data from human embryonic stem cells (ESC). This showed that most blocks were fully
264 methylated or unmethylated on both haplotypes (Figure S4A), implying that intermediate
265 methylation emerges during hematopoietic development. IMR blocks also reverted to being fully
266 methylated or unmethylated in induced pluripotent stem cells derived from CD34+ HSPCs(*43*)
267 (Figure S4B), indicating these states are intrinsic to pluripotent stem cell identity.

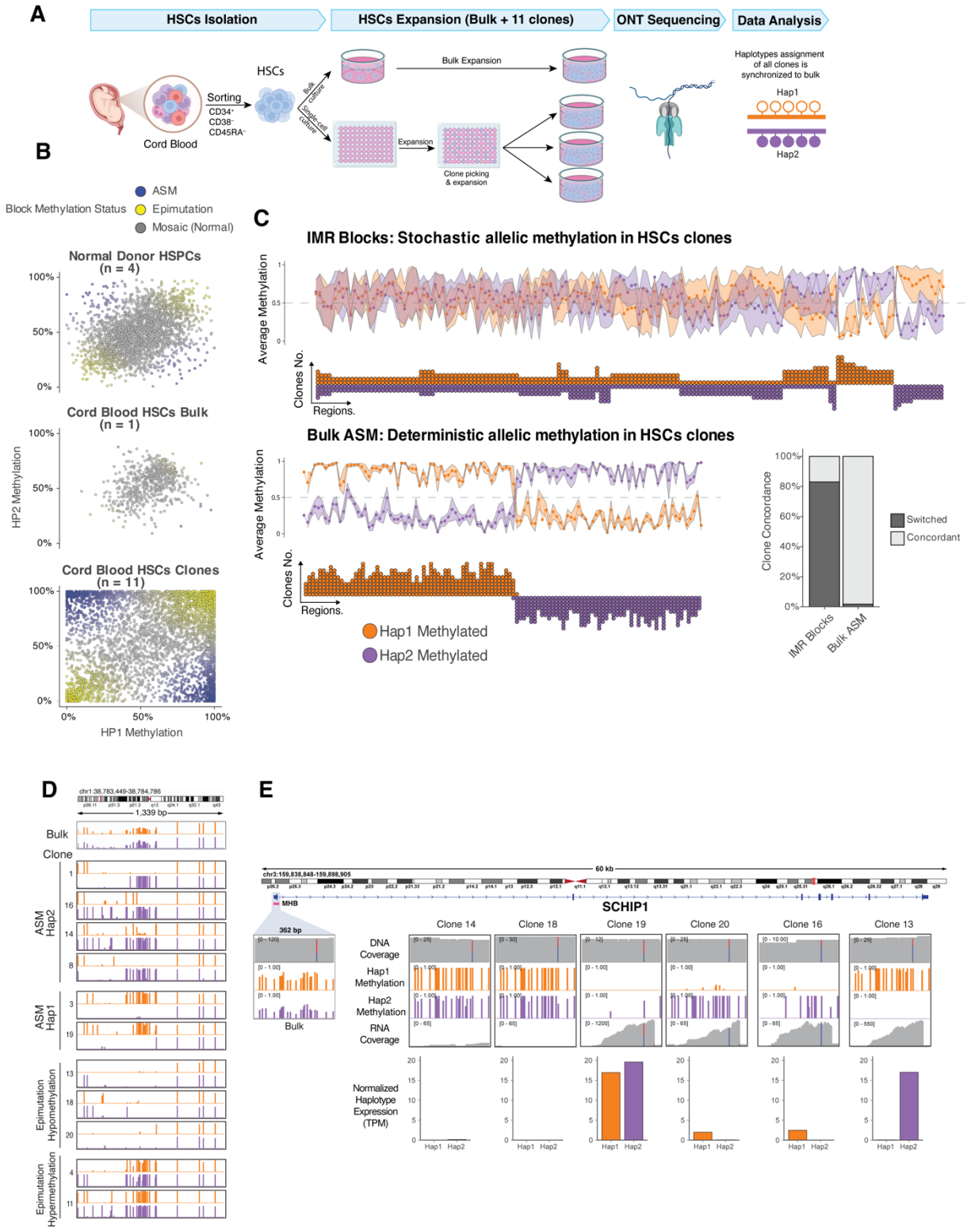
268 To determine whether mosaic methylation at IMR blocks reflects random ASM, we
269 sorted CD34+/CD38-/CD45RA- HSCs from primary cord blood for expansion of bulk HSCs and
270 11 single-cell HSC clones and analysis via ONT sequencing (Figure 4A). This showed that
271 methylation at most IMR blocks with intermediate methylation in bulk HSCs (684 of 1,575 of
272 blocks) lost this pattern in single HSC clones: 69% of blocks had ASM in at least one clone and

273 most were either ASM or fully methylated/unmethylated on both haplotypes (Figure 4B).
274 Further, 83% of IMR blocks with ASM in ≥4 clones (N=164) showed 'switching' of the
275 methylated haplotype, implying stochastic selection of the methylated state (Figure 4C, 4D). In
276 comparison, haplotype-level methylation was consistent in 98% of regions with ASM in bulk
277 HSCs (Figure 4C), indicating that the random methylation is unlikely to be due to technical or
278 biologic variability. IMR blocks with hyper or hypomethylated epimutations in the HSC clones
279 were also random, but to a lesser extent, with 44% of blocks switching between full hyper and
280 hypomethylated epimutations (Figure S4C, S4D). Most of the epimutation blocks had increased
281 methylation, which is consistent with methylation changes previously observed in HSCs forced
282 to rapidly divide in culture(*4*).

283 We next tested whether random ASM could be linked to allele-specific expression (ASE)
284 by performing bulk RNAseq on 6 HSC clones. Although we were limited to examining only
285 genes with coding SNPs near IMR blocks, we identified a block in an internal promoter of
286 *SCHIP1* with intermediate methylation in bulk HSCs that was either fully
287 methylated/unmethylated or had ASM in all 6 clones (Figure 4E). Haplotype-level expression of
288 *SCHIP1* isoforms using this promoter was variable and correlated with IMR block haplotype
289 methylation. Similar effects were seen for *GATA2*, a critical gene in hematopoietic development
290 previously shown to have ASE in AML(*25*, *28*); several ASM regions were present at the *GATA2*
291 locus in our data, but these did not meet all IMR block criteria. However, *GATA2* had striking
292 monoallelic expression in two clones, and the expression level correlated with haplotype-level
293 methylation in a block-like region (Figure S4E), indicating that both ASM and ASE at *GATA2*
294 are pre-existing states in normal HSCs. Together these observations imply that IMR blocks like

295 those at the *SCHIP1* and *GATA2* loci undergo a form of somatically acquired "imprinting" during

296 hematopoietic development.


A
HSCs Isolation
HSCs Expansion (Bulk + 11 clones)
ONT Sequencing
Data Analysis
Cord Blood
Sorting
CD34+
CD38-
CD45RA-
HSCs
Bulk culture
Single-cell culture
Bulk Expansion
Expansion
Clone picking & expansion
Haplotypes assignment of all clones is synchronized to bulk
Hap1
Hap2
B
Block Methylation Status
ASM
Epimutation
Mosaic (Normal)
Normal Donor HSPCs (n = 4)
Cord Blood HSCs Bulk (n = 1)
Cord Blood HSCs Clones (n = 11)
HP2 Methylation
HP1 Methylation
C
IMR Blocks: Stochastic allelic methylation in HSCs clones
Average Methylation
Clones No.
Regions.
Bulk ASM: Deterministic allelic methylation in HSCs clones
Hap1 Methylated
Hap2 Methylated
Clone Concordance
Switched
Concordant
IMR Blocks
Bulk ASM
D
1,339 bp
Bulk
Clone
ASM Hap2
ASM Hap1
Epimutation Hypomethylation
Epimutation Hypermethylation
E
60 kb
SCHIP1
MHB
362 bp
Bulk
Clone 14
Clone 18
Clone 19
Clone 20
Clone 16
Clone 13
DNA Coverage
Hap1 Methylation
Hap2 Methylation
RNA Coverage
Normalized Haplotype Expression (TPM)
Hap1
Hap2

297

298 **Fig. 4. Random allele-specific methylation at IMR blocks is established in single**
299 **hematopoietic stem cells and is associated with allele-specific gene expression.**
300 **(A)** HSC expansion and sequencing to analyze haplotype-level methylation in single-cell-derived
301 HSC clones. **(B)** Haplotype-level methylation at IMR blocks in normal donor HSPCs, bulk cord
302 blood HSCs, and single-cell-derived HSC clones. Each point corresponds to the average
303 methylation of one IMR block within a single clone. Only IMR blocks with intermediate
304 methylation (0.3–0.7) in the bulk HSCs were included. **(C)** Haplotype-level methylation patterns
305 in single-cell-derived HSC clones at IMR blocks exhibiting intermediate methylation in bulk HSCs
306 (top, labelled "IMR Blocks") and regions showing allele-specific methylation (ASM) in bulk
307 HSCs (bottom left, "Bulk ASM"). In both panels, line plots display the average methylation of
308 haplotype 1 (orange) and haplotype 2 (purple) across all clones for each region. The dot plots
309 below indicate the number of individual clones in which Haplotype 1 or Haplotype 2 was
310 methylated, which is indicated by the Y-axis and the dot color. The bar plot (bottom right)
311 quantifies the percentage of regions where the methylated allele is concordant across clones versus
312 switched (i.e., random) for both IMR blocks and bulk ASM regions. Only regions with ASM
313 observed in at least four clones were included in this analysis. **(D)** Example of an IMR block
314 exhibiting random methylation patterns in HSC clones. The tracks display haplotype-level
315 methylation at the block that shows intermediate methylation in bulk HSCs. Individual clones
316 demonstrate random allele-specific methylation (ASM) on either haplotype 1 (orange) or
317 haplotype 2 (purple), or an epimutation showing either fully hypomethylated or hypermethylated
318 states. **(E)** Allele-specific methylation is associated with allele-specific expression in HSC clones
319 at an IMR block located within an internal promoter of the *SCHIP1* gene. For several
320 representative clones, tracks display DNA coverage, haplotype 1 methylation (orange), haplotype

321 2 methylation (purple), and RNA coverage. The bar plots quantify normalized, haplotype-specific
322 expression in Transcripts Per Million (TPM). RNA Reads are assigned to haplotypes based on a
323 heterozygous SNP, which is indicated by the vertical blue/grey (haplotype 1) and red (haplotype
324 2) lines in the coverage tracks. While DNA coverage is consistently biallelic, the level of RNA
325 expression from each haplotype is inversely correlated with its methylation status, demonstrating
326 that higher methylation is linked to transcriptional repression.
327

328 **Intermediately methylated blocks in multiple normal tissues are epimutation hotspots in**
329 **cancer**

330 We were next interested in whether IMR blocks in other tissues displayed epigenetic
331 features like those in hematopoietic cells and were associated with abnormal methylation in
332 tumors. For this analysis, we used data from a recent study of methylation in normal tissues that
333 defined tissue-specific CpG blocks with bimodal methylation (i.e., IMR blocks) (*44*, *45*), as well
334 as separate studies of primary tumors using WGBS or ONT(*46–49*) (Figure 5A). Interestingly,
335 the published set of IMR blocks were enriched for bivalent chromatin and transcriptional start
336 sites in all tissues, similar to the blocks we identified in hematopoietic cells (Figure 5B, see also
337 Figure S1). Likewise, analysis of haplotype-level methylation from primary tumors showed that
338 these blocks lost intermediate methylation and frequently had ASM. Haplotype-level analysis
339 showed that 25-71% of IMR blocks with retained intermediate methylation had ASM across
340 tumor types, further indicating that ASM in these regions is a property of clonal cell populations
341 (Figure 5C, 5D).

342 Finally, we sought to determine whether IMR blocks are hotspots for altered methylation
343 in solid tumors and could therefore represent epimutations. Indeed, methylation at tissue-specific

344 IMR blocks was frequently altered in both ONT and WGBS tumor data (Figure 5E, S5A). We
345 then quantified the frequency of altered methylation at IMR blocks by comparing block
346 methylation in each tumor sample to its respective normal tissue. For comparison, the same
347 analysis was performed using other genomic regions commonly associated with differential
348 methylation in cancer, including CpG islands, gene promoters and tissue-specific enhancers
349 defined via chromatin states(*36*). Surprisingly, the fraction of IMR blocks with differential
350 methylation was the highest among these regions, with a mean of 40% to 65% having
351 statistically significant differential methylation vs. normal tissue (Figure 5F, S5B). By
352 comparison, 3.6% to 22% of the other regions had differential methylation across tumor types.
353 This demonstrates that tissue-specific CpG blocks with intermediate methylation are "hotspots"
354 for altered methylation in multiple tumor types.
355

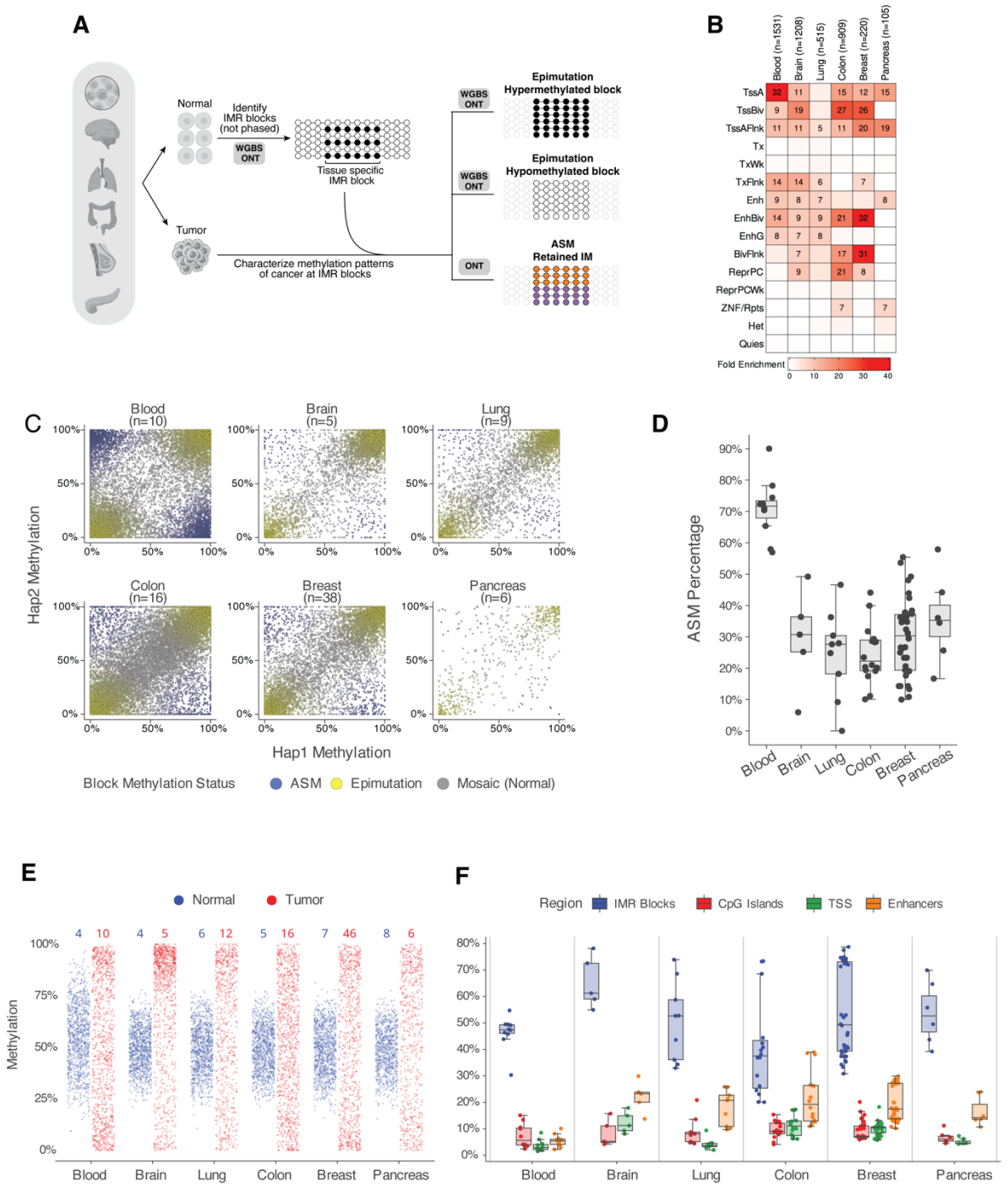


Fig. 5. Tissue-specific IMRs in normal cells are hotspots for epimutations in cancer.

(A) Identifying methylation changes at tissue-specific intermediately methylated regions (IMRs) in multiple cancers. IMRs for hematopoietic cells were identified using ONT data from this study, while IMRs for other tissues were defined using a normal tissue WGBS atlas(*44*). For each tissue,

361 epimutations (hyper- or hypo-methylation) were identified by comparing tumor samples to their
362 normal counterparts. Allele-specific methylation (ASM) was called using haplotype-phased ONT
363 data. This analysis integrated data for AML generated in this study with published datasets for
364 other cancers, including ONT(*46*) and WGBS(*47*, *49*) data. For details, see the Methods section.
365 **(B)** Heatmap showing the fold enrichment of tissue-specific IMRs within tissue-specific chromatin
366 states. For each tissue (columns), the enrichment of its specific IMRs (n = number of blocks) was
367 calculated for each chromatin state (rows) relative to a genomic background. Values displayed
368 represent the fold enrichment for regions with an enrichment of ≥5. Chromatin state annotations
369 were obtained from the Roadmap Epigenomics Consortium(*50*) . **(C)** Comparison of Haplotype 1
370 versus Haplotype 2 methylation at tissue-specific IMRs in cancer samples using ONT data**. (D)**
371 Percentage of tissue-specific IMRs with retained intermediate methylation (0.3-0.7) that exhibit
372 allele-specific methylation (ASM) per sample in various cancer types, using phased ONT data.
373 Each point represents an individual tumor sample. The box plots summarize the distribution of
374 ASM percentages for each cancer type shown. **(E)** Distribution of average methylation levels at
375 tissue-specific IMRs in normal tissues (blue) and their corresponding primary tumors (red) from
376 ONT data. The numbers at the top of the plot indicate the number of samples included in each
377 group. Each point represents the average methylation of a single IMR block within a sample, with
378 data from multiple samples pooled for each tissue type. For visualization, 1,000 points were
379 randomly sampled for each group. Data for tumor samples and normal blood were derived from
380 ONT, while data for other normal tissues were from WGBS. **(F)** Percentage of various genomic
381 features exhibiting epimutations across different cancer types from ONT data for individual
382 primary tumor samples. Epimutations were defined as a significant difference in aggregate
383 methylation via a Fisher's exact test on methylation counts for each region between tumor and its

respective normal. Boxplots show the percentage of tissue-specific IMRs (IMR Blocks), CpG islands, tissue-specific transcriptional start sites (TSS), and tissue-specific enhancers that have significant count-based differential methylation in each tumor sample compared to its corresponding normal tissue.

Discussion

The factors that predispose certain regions to DNA methylation changes in cancer are not well understood. Here we showed that discrete blocks of CpGs with intermediate methylation in normal cells are *a priori* hotspots for methylation changes in AML and multiple solid tumors. In AML, methylation changes at these intermediately methylated regions (IMRs) resemble epimutations: they are stable, clonal events that are less dependent on mutation-driven phenotypes and can be observed in non-leukemic cells. Further, intermediate methylation at these blocks in normal hematopoietic cells is caused by random allele-specific methylation at the single cell level. The highly tissue-specific and stable nature of methylation at these regions implies they normally undergo a type of developmental "somatic imprinting", and epimutations in AML are due to the absence of this allele-specific state. Further, these regions possess active and bivalent chromatin features, and some occur near important tissue-specific genes, which suggests they are involved in gene regulation and could be subject to selection during AML pathogenesis.

The concept of allele-specific somatic imprinting is consistent with previous epigenetic studies in primary human tissues. In hematopoietic cells, clonal cell populations exhibit heritable, random monoallelic gene expression(*27*, *28*), and separate studies have used methylation signatures as clonal markers(*51*, *52*). The blocks of random allele-specific methylation described here provide a direct link between these observations by identifying

408 discrete loci where methylation can both contribute to gene regulation and encode stable markers
409 of clonal identity. While it is possible that the mosaic methylation pattern of these regions means
410 they are less functionally relevant, their relationship to chromatin activity and expression
411 suggests otherwise. The striking consistency and stability of IMRs in HSCs also implies that
412 establishing intermediate methylation via somatic imprinting could be a defining feature of HSC
413 identity; epimutations could then be caused by a failure or spontaneous loss of imprinting. While
414 the forces that make ASM advantageous are unknown, it could enable precise gene dosage or
415 simultaneous transcription of two closely spaced genes from opposite alleles. Notably, the
416 *GATA2*, *SCHIP1* and *HOXA* loci with ASM all encode multiple overlapping transcripts and could
417 be subject to such constraints.
418 One consequence of somatic imprinting is that it generates a substrate for selection
419 during cancer development. For example, genes with ASM and monoallelic expression would be
420 vulnerable to somatic mutations as a second hit(*25*, *53*). Likewise, the presence of epimutations
421 is a source of epigenetic heterogeneity that could confer an advantage to some cells. We found
422 that IMR blocks were differentially methylated more often than CpG islands, promoters, and
423 enhancers in multiple cancers, and in AML, individual IMR blocks tended to become either fully
424 methylated or unmethylated, hinting that certain epimutations may be under selection. The
425 discrete nature of these events means they may be more amenable to direct measurements and
426 functional characterization to understand how pre-existing epigenetic variability contributes to
427 the somatic processes that lead to cancer.
428

429 **Supplementary Materials**

430 **Materials and Methods**

431 *Primary AML and normal hematopoietic cell samples*

432 Primary AML and normal hematopoietic cell samples used in this study were collected from
433 consented subjects using IRB-approved protocols (WashU HRPO#201011766, FHCC #985.03).
434 AML samples were either peripheral blood or bone marrow aspirate specimens obtained from
435 patients with active disease at presentation or relapse and were processed without fractionation,
436 purification, or enrichment via buffy coat preparation into cryopreserved cell suspensions or
437 directly used for DNA extraction (see Tables S1 and S4). Remission samples were unfractionated
438 whole bone marrow aspirates or peripheral blood samples obtained from patients in morphologic
439 complete remission at the timepoints indicated. Normal adult CD34+ HSPCs, promyelocytes,
440 neutrophils, and monocytes from healthy donor bone marrow aspirate samples and were
441 described previously(*4*, *5*). Additional data were generated from peripheral blood mobilized
442 CD34+ HSPCs from 3 donors from the Fred Hutchinson Cancer Center (FHCC) Cooperative
443 Centers of Excellence in Hematology Cell Procurement & Processing Core and enriched for
444 lineage-negative cells from bone marrow aspirates from two healthy donors via bead-based
445 depletion of lineage-positive cells (Miltenyi cat# 130-092-211) using an autoMACS instrument
446 (Miltenyi). Finally, commercially available purified CD34+ HSPCs from cord blood samples
447 (STEMCELL) were used for the clonal expansion experiment.

448

449 *Cord blood HSC sorting and expansion*

450 Cryopreserved umbilical cord blood CD34+ cells (10^6 cells from a single donor purchased from
451 STEMCELL Technologies) were thawed in a ThawSTAR instrument (BioLife) and then cultured

452 overnight in 10 wells of a 96-well U-bottom plate in StemSpan media with CD34+ expansion
453 supplement, UM729 (STEMCELL) and 1% pen/strep antibiotic. After overnight culture, cells
454 were collected and stained with CD34-APC (BD Pharmingen, #555824), CD38-PE (BD
455 Pharmingen, #560981), CD45RA-FITC (BD Pharmingen, #561882) and Near-IR LIVE/DEAD
456 stain (Invitrogen) for FACS on a MoFlo instrument. Single CD34+/CD38-/CD45RA- cells were
457 sorted into two 96-well flat-bottom plates seeded with 25,000 cells/well of irradiated (2000 cGy)
458 AFT024 stroma (ATCC) and containing 100 uL of prepared StemSpan media. An additional
459 30,000 cells with the same immunophenotype were sorted into an additional well for expansion
460 of a bulk (nonclonal) cell population. On days 4 and 10 after sorting 50 uL of prepared StemSpan
461 media was added to all wells. The bulk HSCs were harvested in day 15 of culture and yielded
462 approximately 10^7 cells. Expanded single-cell clones were transferred to 24-well plates without
463 AFT024 stroma on days 21 and 23 of culture and then expanded in 0.5 mL of prepared
464 StemSpan media for an additional 4 days before harvesting for DNA and RNA extraction.
465
466 *Human embryonic stem cell and iPS cell culture*
467 Human WA0, WA09/H9 ESCs and BM9 iPSCs were cultured in mTESR media (STEMCELL)
468 and passaged using dispase before harvesting for DNA extraction.
469
470 *Whole-genome bisulfite sequencing data production*
471 WGBS data from 51 AML samples and 15 normal hematopoietic cell populations were
472 repurposed from prior studies(*4*, *5*). Additional datasets were generated as described
473 previously(*5*) using the Zymo Gold Bisulfite conversion (Zymo Research) and Accel DNA
474 methylation kit (IDT) kits to generate WGBS libraries. QC was performed on Agilent

475 Bioanalyzer or Tapestation instruments prior to sequencing on NovaSeq 6000 or X plus
476 instruments to generate 2x150 paired end reads to a target genome coverage of 10x. All WGBS
477 data were processed using the DRAGEN software suite (v4.2.4) to obtain genome-wide 5mC
478 read counts and methylation fractions.
479
480 *ONT sequence data production and processing*
481 All DNA for ONT sequencing was extracted using either the NEB Monarch or Qiamp DNA mini
482 kit. DNA was fragmented and size-selected to eliminate small (<5 kbp) fragments and then
483 prepared for sequencing via the LSK114 kit and sequenced using R10.4.1 flowcells on a
484 PromethION sequencer (Oxford Nanopore Technologies, Oxford, UK) with up to two
485 wash/reload cycles. For data obtained from expanded HSCs, adaptive sampling was performed
486 for all IMRs and ASM regions identified in AML samples. Data processing and analysis,
487 including basecalling, was performed via a Nextflow workflow (https://github.com/dhslab/nf-
488 core-wgsnano, commit 8fe4655) using dorado (v0.5.2+7969fab) for basecalling and read
489 mapping, and the PEPPER-Margin-Deepvariant pipeline for phased germline variant calling of
490 SNV and small indels (v0.8.0)(*54*). Structural variants were identified after data processing using
491 Sniffles(*55*). Haplotagging was performed using Whatshap (v2.2)(*56*), which used a common
492 phased VCF file for presentation/remission/relapse samples from AML patients and the bulk
493 HSC data for the HSC expansion experiment to synchronize the haplotype designations.
494 Mosdepth (v.0.3.3)(*57*) was used to calculate coverage. Combined and haplotype-specific 5mC
495 counts were generated with modkit (https://github.com/nanoporetech/modkit).
496
497 *Fiber-seq on normal HSPCs and primary AML samples*

498 Fiber-seq was performed on cryopreserved mobilized CD34+ HSPCs from two donors (obtained
499 from FHCC) and presentation/relapse sample pairs from two AML patients using the published
500 protocol(*41*) with slight modifications. Cells were thawed in a ThawSTAR instrument and 2e6
501 cells were collected at 250xg for 5 minutes, washed once in 1mL cold PBS and resuspended in
502 120 uL CUT&TAG nuclear extraction buffer (20 mM HEPES-KOH, pH 7.9, 10 mM KCl, 0.1%
503 Triton X-100, 20% glycerol, and 0.5mM spermidine) and incubated for 10 minutes on ice.
504 Nuclei were then collected at 350xg for 5 minutes and resuspended in 60 uL Buffer A (15mM
505 Tris-HCl, 15 mM NaCl, 60 mM KCl, and 1 mM EDTA, pH 8.0) and transferred to a PCR tube
506 before adding 1.5 uL 32 mM s-adenosyl methionine (SAM) and 0.5 uL Hia5 enzyme (200U/uL,
507 a gift from A. Stergachis). Reactions were incubated at 25C in a thermocycler for 10 min and
508 then stopped with 6uL of 10% SDS (final concentration of 1%). The lysate was transferred to a
509 1.5mL tube with wide bore tips and stored at -80C prior to DNA extraction and sequencing.
510 DNA was extracted with a MagAttract DNA extraction kit, fragmented with a megaruptor and
511 size-selected on a HTpippen to 18-22kbp prior to SMRTbell library preparation. Libraries were
512 sequenced on a PacBio Revio instrument to generate CCS reads containing 5mC and kinetic data
513 for prediction of 6mA via Fibertools(*58*). The FIRE pipeline was then used to obtain haplotype-
514 level accessibility for all samples.
515
516 *Histone CUT&Tag on CD34+ HSPCs and AML samples*
517 Histone profiling used a CUT&Tag-direct protocol(*59*) with modifications relating to cryo-
518 preserved primary AML cells and the release of tagmented DNA from Concanavalin A coated
519 (Con-A) magnetic beads. Cryopreserved HSPC and AML cells were thawed at 37C, washed once
520 in cold PBS, then pelleted for 3 minutes at 600g and resuspended in hypotonic NE1 buffer

521 (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 0.5 mM spermidine, 10% Triton X-100 and 20%
522 glycerol) followed by incubation on ice for between 45 seconds and 2 minutes for nuclei
523 extraction. The mixture was pelleted for 4 minutes at 1,300g, resuspended in 1× PBS and nuclei
524 were counted using the Trypan blue staining or a ViCell Automated Cell Counter (Beckman
525 Coulter). CUT&Tag-Direct was then carried out as previously described
526 (https://doi.org/10.17504/protocols.io.bcuhiwt6). In brief, nuclei were bound to washed
527 paramagnetic concanavalin A (ConA) beads (Bangs Laboratories) at a concentration of 5 µL
528 washed beads per 50K nuclei and incubated with primary antibody (H3K27me3: Cell Signaling
529 9733S, H3K4me3: Millipore CS200580) at 4°C overnight in Wash Buffer (10 mM HEPES pH
530 7.5, 150 mM NaCl, 0.5 mM spermidine and Roche Complete Protease Inhibitor Cocktail) with
531 2 mM EDTA. Bound nuclei were washed and incubated with 1:100 secondary antibody for
532 1 hour at room temperature and then washed 2x and incubated in Wash-300 Buffer (Wash Buffer
533 with 300 mM NaCl) with 1:50 5 µM loaded pA–Tn5 for 1 hour at room temperature. Nuclei
534 were washed 2x and tagmented in Wash-300 Buffer with 10 mM MgCl2 for 1 hour at 37 °C and
535 then resuspended sequentially in 50 µl of 10 mM TAPS and 5 µl of 10 mM TAPS with 0.1% SDS
536 and incubated for 1 hour at 58 °C. The resulting suspension was mixed well with 16 µl of
537 0.9375% Triton X-100, and then 2 µL each of F and R primers and 25 µL 2× NEBNext Master
538 Mix (New England Biolabs) were added for direct amplification with the following conditions:
539 (1) 58 °C for 5 minutes, (2) 72 °C for 5 minutes, (3) 98 °C for 30 seconds, (4) 98 °C 15 seconds,
540 (5) 60 °C for 15 seconds, (6) repeat steps 4–5 13 times, (7) 72 °C for 1 minute and (8) hold at
541 8 °C. DNA from amplified product was purified via AMPure bead cleanup and resuspended in
542 25 µl of 10 mM Tris-HCl with 1 mM EDTA. Libraries were assessed using the Qubit (Thermo
543 Fisher Scientific) and TapeStation (Agilent) systems and sequencing was performed on an

544 Illumina NovaSeq Xplus to obtain a minimum of 2 million 2x150 bp paired end reads. All data
545 were analyzed via the nf-core(*60*) CUT&Run analysis pipeline to obtain normalized signal in
546 bigwig format.
547
548 *Bulk RNAseq on AML samples and HSC clones*
549 RNA was extracted from 10^6 cells using Qiaquick RNA micro kit (Zymo Research). RNA
550 integrity was assessed by RIN scores by Agilent Tape Station 4150 (Agilent Technologies).
551 Quantification of RNA eluates obtained by Qubit RNA BR Assay kit (Invitrogen-Thermo-
552 Fisher). RNAseq libraries were generated via rRNA/Globin HMR depletion followed by
553 ligation-based library preparation (Watchmaker Genomics). Libraries were assessed using
554 Agilent Tape Station 4150 with High Sensitivity D1000 screen tapes (Agilent Technologies) and
555 Qubit dsDNA HS Assays (Invitrogen-Thermo-Fisher). Sequencing was performed on Illumina
556 NovaSeq Xplus to obtain a minimum of 50 million paired end 2x150 bp reads. Data processing
557 used the Illumina DRAGEN RNA pipeline to obtain normalized gene and transcript-level
558 expression values in transcripts per million (TPM).
559
560 *Differentially methylated region (DMR) identification from WGBS data*
561 The DSS(*61*) and bsseq(*62*) packages in R (version 4.3) were used to identify DMRs between
562 each AML sample (N=100) and 17 normal hematopoietic cell methylation datasets (9 CD34+
563 HSPC samples, 3 promyelocyte samples, 3 neutrophil samples, and 2 monocyte samples, all
564 purified as described above from normal bone marrow aspirates from adults aged 20-50). An
565 FDR cutoff of 0.05 and minimum methylation difference of 0.3 were used within DSS without
566 additional filtering. Analysis of DMR recurrence used *bedtools*(*63*) to merge all DMRs and count

567 overlaps with the DMR set from each AML sample; initial merging and overlapping operations
568 used default settings with no minimum overlap. DMR hotspots were then defined using *bedtools*
569 *multiintersect* to obtain intervals that had >=25 overlapping DMRs across the 100 AML samples
570 and had at least 5 CpG. Coverage-weighted mean methylation of all 'multi-intersect' DMRs was
571 calculated for each AML sample and the pooled normal hematopoietic cell data using *methfast*, a
572 custom c program for aggregating CpG methylation counts over bed intervals (code available
573 upon request). These unmanipulated methylation fractions were used for all figures, statistical
574 calculations, and comparisons, except for the heatmaps in Figure 1E where smoothing was
575 performed to impute missing data for visualization (smoothed values were not used for sorting or
576 determining methylation levels).
577
578 *DNA fragment-level and allele-specific methylation analysis from WGBS data*
579 Fragment-level methylation metrics were calculated in private and recurrent DMRs using all
580 normal hematopoietic cell datasets using mHapSuite(*38*). Allele-specific methylation (ASM)
581 analysis was performed by first identifying common single-nucleotide polymorphisms (SNPs)
582 from the 1000 Genomes Project (1000G_phase1.snps.high_confidence.hg38.vcf.gz) that
583 overlapped with IMR hotspots. Sequencing aligned reads covering these "hotspot SNPs" were
584 then partitioned into two groups: those matching the reference allele and those matching the
585 specific alternate allele defined in the 1000 Genomes Project. For the subsequent analysis of an
586 IMR hotspot in an individual sample, each allele group was required to have a minimum of 8
587 reads, with each read spanning at least 4 CpGs. Individual reads were classified as methylated
588 (≥70% of CpGs methylated) or unmethylated (≤30% of CpGs methylated); reads falling between
589 these thresholds were designated "undetermined". Finally, Fisher's exact test was used to

590 compare the number of methylated and unmethylated reads between the reference and alternate
591 allele groups at each IMR hotspot. P-values were adjusted for multiple testing using the
592 Benjamini-Hochberg method, and hotspots with an FDR < 0.05 were considered significant
593 ASM sites. The code used for this analysis is available at https://github.com/dhslab/methtools.
594
595 *ASM analysis from ONT data*
596 ASM regions were identified from ONT data by first importing haplotype-resolved 5mC counts
597 into *bsseq* objects in R and then calling DMRs between haplotypes with *DSS* using a minimum
598 FDR of 0.05 and methylation difference of 0.4. ASM regions were then further filtered to
599 exclude intervals meeting any of the following criteria: 1) on the X or Y chromosome, 2)
600 overlaps a known imprinted region (N=782, from (*64*, *65*)), 3) occurs within 1,000 bp of an
601 identified structural variant in the ONT data, or 4) overlaps a 10,000 bp region with a SNP
602 density that is in the top 1% across the genome.
603
604 *Identification and analysis of IMR blocks in ONT data from normal HSPCs and AML samples*
605 To define IMR blocks we pooled three sets of genomic coordinates: DMR hotspots from our
606 AML WGBS data, ASM regions from our AML ONT data, and known IMRs in normal
607 tissues(*37*). Coordinates from these sources were merged to create a unified set of intervals.
608 Within these intervals, we identified methylation haplotype blocks (MHBs) by adapting a
609 previously described method(*40*) for long read ONT data. Specifically, per-read CpG
610 methylation calls were extracted from the ONT BAM files of four normal donors and 10 AML
611 samples using modkit (https://github.com/nanoporetech/modkit). From this read-level
612 methylation data, we calculated the methylation linkage disequilibrium (mLD) between adjacent

613 CpG sites, measured as the squared Pearson correlation coefficient (r^2), requiring a minimum of
614 10 reads covering a CpG pair for the calculation. An MHB was then defined as a contiguous
615 region containing at least four consecutive CpG sites where the r^2 value between any two
616 adjacent CpGs was ≥ 0.5. The code used for this analysis is available at
617 https://github.com/dhslab/methtools. To identify allele-specific methylation (ASM) and
618 epimutations within IMR blocks using ONT data, we applied Fisher's exact test. For the ASM
619 analysis, we compared the counts of methylated and unmethylated CpGs between the two
620 haplotypes. For the epimutation analysis, we compared the aggregated counts of methylated and
621 unmethylated CpGs (regardless of haplotype) between AML or HSC samples (for data from
622 expanded HSC bulk cells and clones) and pooled HSPCs from normal donors (n=4). For a block
623 to be included, we required a minimum of four covered CpGs. Furthermore, we mandated a
624 minimum adjusted CpG coverage (defined as total CpG counts divided by the number of CpGs
625 in the block) of 5 per haplotype for ASM and 5 per sample for epimutation. P-values for both
626 analyses were adjusted for multiple testing using the Benjamini-Hochberg method across all
627 evaluable blocks. A significant ASM call was defined by an FDR < 0.05 and an absolute
628 methylation difference between haplotypes > 0.4. A significant epimutation was defined by an
629 FDR < 0.05 and an absolute methylation difference between the test and normal samples > 0.3.
630

631 *Histone analysis at IMR blocks*

632 Histone analysis at IMR blocks was performed using H3K4me3 and H3K27me3 CUT&Tag data
633 from 7 AML samples and 3 mobilized CD34+ HSPC samples. All analysis used normalized
634 signal from bigwig files produced by the nf-core pipeline. Mean methylation at each block was
635 classified as being intermediate (i.e., having ASM) or being hyper or hypomethylated on both

636 haplotypes. Aggregate normalized signal for each modification in 200 equally spaced bins was
637 then calculated for a 5 kbp window centered on the block location for analysis by methylation-
638 defined groups.
639
640 *Fiber-seq FIRE analysis*
641 Haplotype-resolved chromatin accessibility was assessed using the Fiber-seq Inferred Regulatory
642 Elements (FIRE) analysis pipeline(*66*), available at https://github.com/fiberseq/FIRE. The
643 workflow was run using aligned BAM files as input, and accessibility was quantified as the
644 percentage of accessible fibers per region. Analysis of FIRE signal at IMR blocks used either
645 combined or haplotype-level estimates of percent accessibility and were performed at IMR
646 blocks grouped by methylation status (see histone analysis above).
647
648 *Short read genome and exome sequencing of AML samples*
649 Whole-genome sequencing (WGS) was performed on representation and relapse AML samples
650 from patients 112200, 627311, and 895870 with matched normal tissue from skin biopsies. All
651 sequencing was performed using 500ng of DNA input to generate libraries with the Illumina
652 DNA library preparation kit, which were quantified by qPCR and sequenced to obtain a target
653 genome coverage of 60x in 2x150 bp paired end reads on an Illumina NovaSeq Xplus sequencer.
654 Data were analyzed using paired somatic mode with the Illumina DRAGEN software. Identified
655 somatic variants were filtered to retain single nucleotide variants with PASS filter flag and total
656 coverage between 20x and 90x to minimize sequencing artifacts. Whole-exome data for the
657 remission samples was described previously(*42*); variants and their associated variant allele
658 fractions were used directly from that publication.

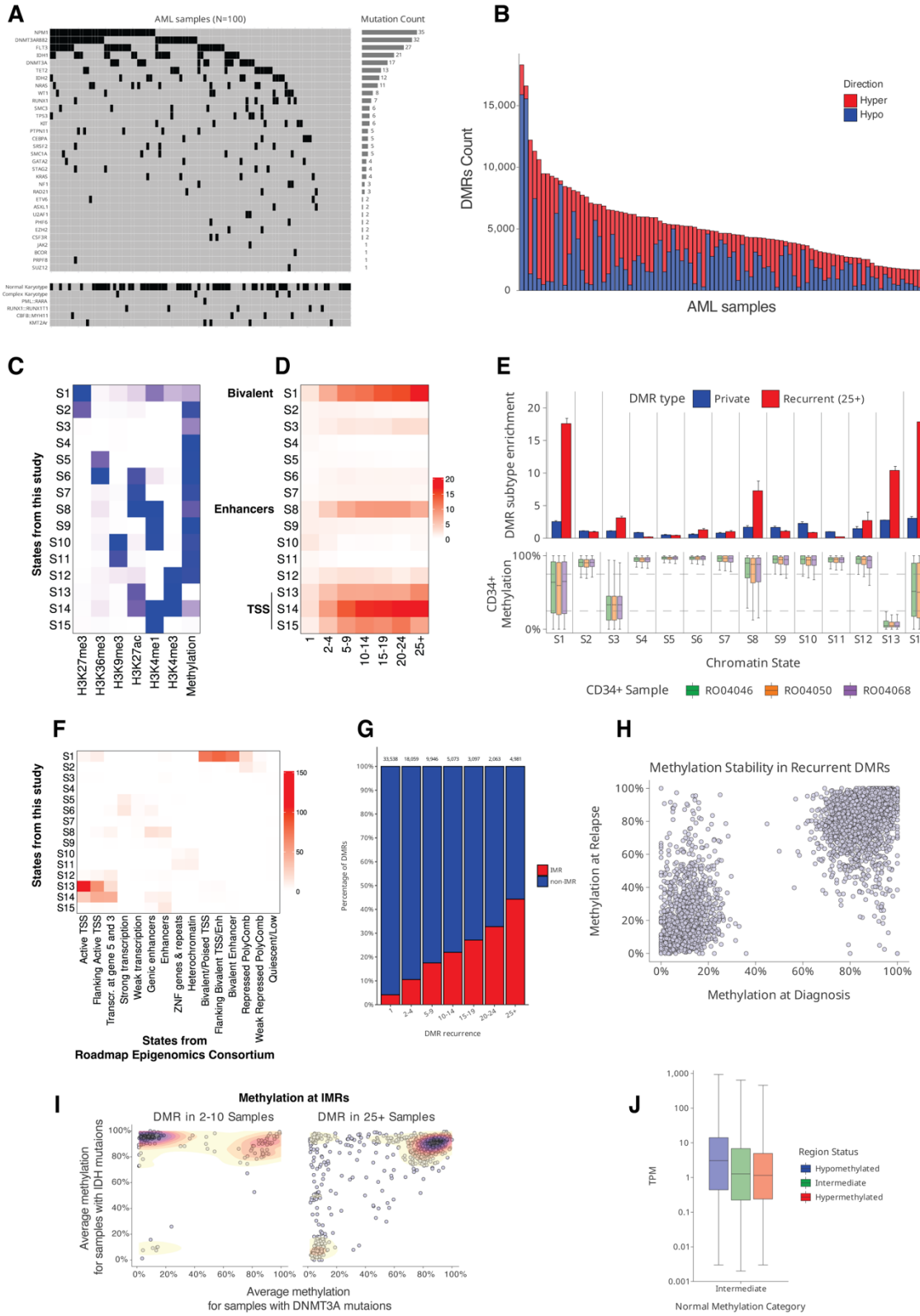
IMR block identification in WGBS from normal tissues

To identify IMR blocks in normal tissues (brain, lung, colon, breast, and pancreas), we began with a set of previously defined regions of bimodal methylation "bimodal blocks" from WGBS data(*44*). We then applied a multi-step filtering process to refine these regions and isolate those specific to a single tissue type. First, to ensure consistency within a given tissue, we only retained bimodal blocks that were present in all normal samples for each tissue type. Next, these consistent blocks from all tissue types were merged to create a non-redundant block list (minimum block size of 50 bp; minimum distance between blocks of 50 bp). To isolate truly tissue-specific blocks and exclude germline imprinted regions, this non-redundant list was subjected to a second round of filtering. For each block, we calculated the average methylation within its matched-tissue samples and the average methylation across all other (non-matched) normal samples. A block was kept only if it exhibited intermediate methylation specifically in its matched tissue, defined by two criteria: 1) an average methylation level between 0.3 and 0.7 with a coefficient of variation (CV) ≤ 0.5 within the matched tissue, and 2) an absolute difference in average methylation of ≥ 0.2 between the matched and non-matched tissues. This process resulted in a final set of tissue-specific IMR blocks in hg19 coordinates. For compatibility with the solid-tumor ONT data, a second version of these block coordinates was generated in hg38 using liftOver.

Analysis of IMR blocks and other regions in WGBS and ONT data from solid tumors

To analyze methylation patterns in solid tumor WGBS and ONT data, we first defined several sets of genomic regions for analysis. These included the IMR blocks described above, CpG

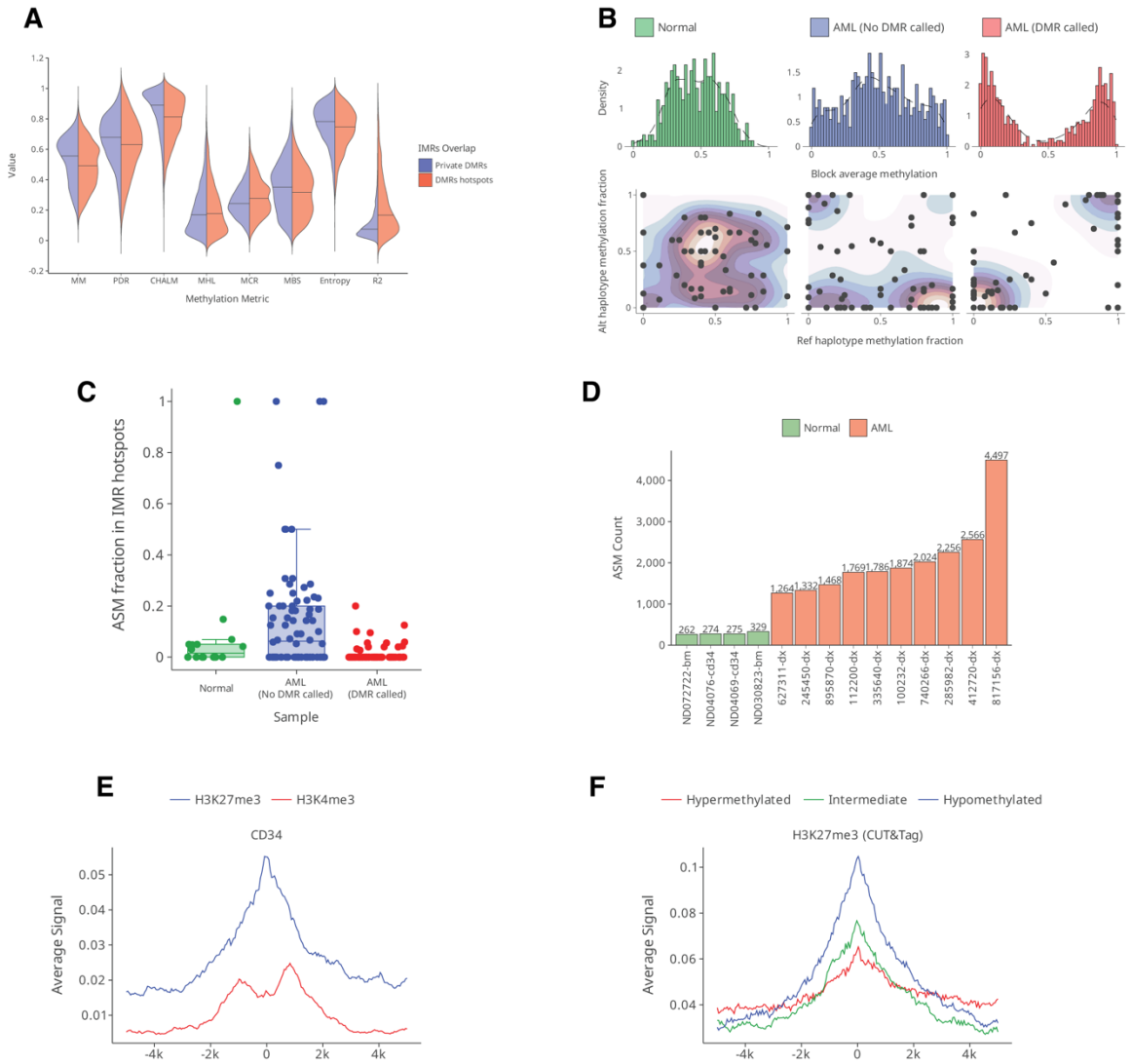
682 islands downloaded from the UCSC genome browser, and tissue-specific TSS and enhancers
683 extracted from Roadmap Epigenomics Consortium ChromHMM annotations (model identifiers:
684 E035 for blood, E073 for brain, E096 for lung, E075 for colon, E027 for breast, E098 for
685 pancreas)(*50*). For the tumor methylation data, we used datasets from GEO: GSE121721
686 (WGBS: brain), GSE212391 (WGBS: other solid tumors), and GSE270257 (ONT: all solid
687 tumors). For the matched normal samples, we downloaded beta value files from the DNA
688 methylation atlas(*45*). All datasets were processed into a unified CpG-level bedmethyl format;
689 for the normal samples, this conversion was performed using wgbstools(*67*). We then used
690 methfast (https://github.com/dhslab/methfast) to aggregate CpG methylation counts from these
691 bedmethyl files over all genomic regions of interest. For the blood cell type analyses, we used
692 the same WGBS and ONT data for normal and AML samples as previously described. In Figure
693 5 and Figure S5, these normal hematopoietic cells and AML tumor samples are represented as
694 the 'blood' normal and tumor category, respectively. Finally, to identify epimutations (in WGBS
695 and ONT) and ASM (in ONT), we applied the same Fisher's exact test-based approaches used
696 for the hematopoietic cells, which involved comparing aggregated counts between tumor and
697 normal samples for epimutation, and between the two haplotypes for ASM


A
AML samples (N=100)
Mutation Count
B
Direction
Hyper
Hypo
DMRs Count
AML samples
C
States from this study
H3K27me3
H3K36me3
H3K9me3
H3K27ac
H3K4me1
H3K4me3
Methylation
D
Bivalent
Enhancers
TSS
E
DMR type
Private
Recurrent (25+)
DMR subtype enrichment
CD34+ Methylation
Chromatin State
CD34+ Sample
RO04046
RO04050
RO04068
F
States from this study
Active TSS
Flanking Active TSS
Transcr. at gene 5 and 3
Strong transcription
Weak transcription
Genic enhancers
Enhancers
ZNF genes & repeats
Heterochromatin
Bivalent/Poised TSS
Flanking Bivalent TSS/Enh
Bivalent Enhancer
Repressed PolyComb
Weak Repressed PolyComb
Quiescent/Low
States from
Roadmap Epigenomics Consortium
G
Percentage of DMRs
DMR recurrence
IMR
non-IMR
H
Methylation Stability in Recurrent DMRs
Methylation at Relapse
Methylation at Diagnosis
I
Methylation at IMRs
DMR in 2-10 Samples
DMR in 25+ Samples
Average methylation
for samples with IDH mutaions
Average methylation
for samples with DNMT3A mutaions
J
TPM
Region Status
Hypomethylated
Intermediate
Hypermethylated
Normal Methylation Category

698

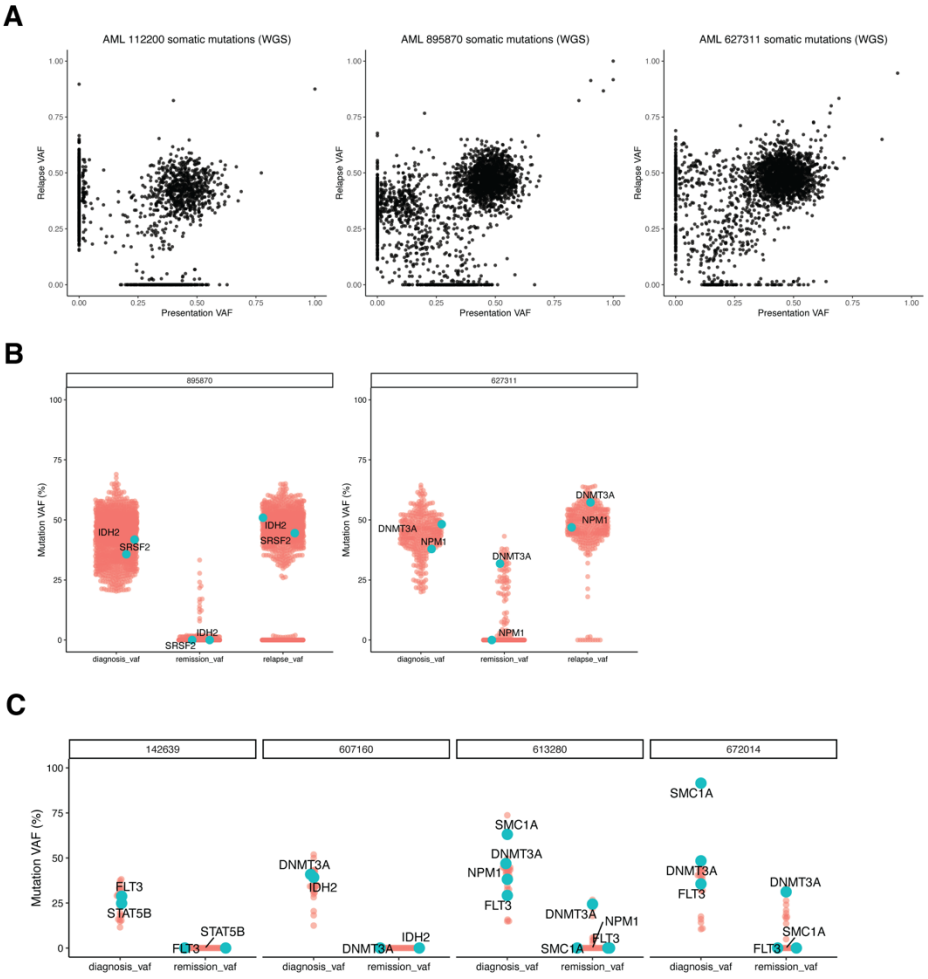
699 **Fig. S1. Identification of differentially methylated regions in AML and their enrichment in**
700 **intermediately methylated regions in normal hematopoietic cells.**
701 **(A)** Mutation and karyotype status of AML samples used for WGBS analysis. (**B**) Number of
702 DMRs identified in each AML sample compared to CD34+ cells. **(C)** Emission parameters for
703 states S1 through S15 of a chromHMM model trained on six histone modifications (H3K4me1,
704 H3K4me3, H3K27me3, H3K36me3, H3K9me3, and H3K27ac) and DNA methylation data from
705 purified CD34+ HSPCs. (**D**) Heatmap of DMR enrichment grouped by recurrence frequency in
706 chromatin states defined in HSPCs. (**E**) (Top) Relative enrichment of private and recurrent
707 DMRs and (Bottom) Methylation levels in CD34+ (n=3) at chromatin states. For the methylation
708 analysis, all segments within each state are included. (**F**) Heatmap showing the enrichment
709 mapping of predefined functional genomic regions, annotated in published CD34+ data from the
710 Roadmap Epigenomics Consortium (histone modifications only), within the chromatin states
711 (S1-S15) defined in this study using both histone modifications and DNA methylation in CD34+
712 cells. (**G**) Overlap percentage of AML DMRs (grouped by recurrence frequency) with previously
713 published intermediately methylated regions (IMRs)(*37*). (**H**) Methylation levels at IMR
714 hotspots with altered methylation (DMRs called), identified in the presentation sample,
715 comparing the methylation levels between presentation and relapse samples from the same
716 patient. (**I**) Comparison of average methylation levels between samples with DNMT3A and IDH
717 mutations, including only those with an identified differentially methylated region (DMR). The
718 analysis distinguishes between two categories of normally intermediately methylated regions
719 (IMRs) based on DMR recurrence: high-recurrence hotspots (DMRs in ≥25 samples) and low-
720 recurrence regions (DMRs in 2–10 samples). Only IMRs with at least two differentially
721 methylated samples in each mutation group were included in the comparison. Regions with low

722 recurrence exhibit methylation patterns that appear to be mutation-driven, while highly recurrent
723 regions are consistently hypermethylated or hypomethylated regardless of mutation type. (**J**)
724 RNA expression levels (TPM) of genes with promoters overlapping recurrent IMR hotspots in
725 AML samples, with samples grouped by promoter methylation status into hypomethylated
726 (<30%), intermediately methylated (30–70%), and hypermethylated (>70%) promoter categories.
727 $P < 10^{-8}$ for a difference in expression across these groups, based on an ANOVA.
728


A
Value
MM
PDR
CHALM
MHL
MCR
MBS
Entropy
R2
Methylation Metric
IMRs Overlap
Private DMRs
DMRs hotspots
B
Normal
AML (No DMR called)
AML (DMR called)
Density
Block average methylation
Alt haplotype methylation fraction
Ref haplotype methylation fraction
C
ASM fraction in IMR hotspots
Normal
AML (No DMR called)
AML (DMR called)
Sample
D
Normal
AML
ASM Count
262
274
275
329
1,264
1,332
1,468
1,769
1,786
1,874
2,024
2,256
2,566
4,497
ND072722-bm
ND04076-cd34
ND04069-cd34
ND030823-bm
627311-dx
245450-dx
895870-dx
112200-dx
335640-dx
100232-dx
740266-dx
285982-dx
412720-dx
817156-dx
E
H3K27me3
H3K4me3
CD34
Average Signal
-4k
-2k
0
2k
4k
F
Hypermethylated
Intermediate
Hypomethylated
H3K27me3 (CUT&Tag)
Average Signal
-4k
-2k
0
2k
4k

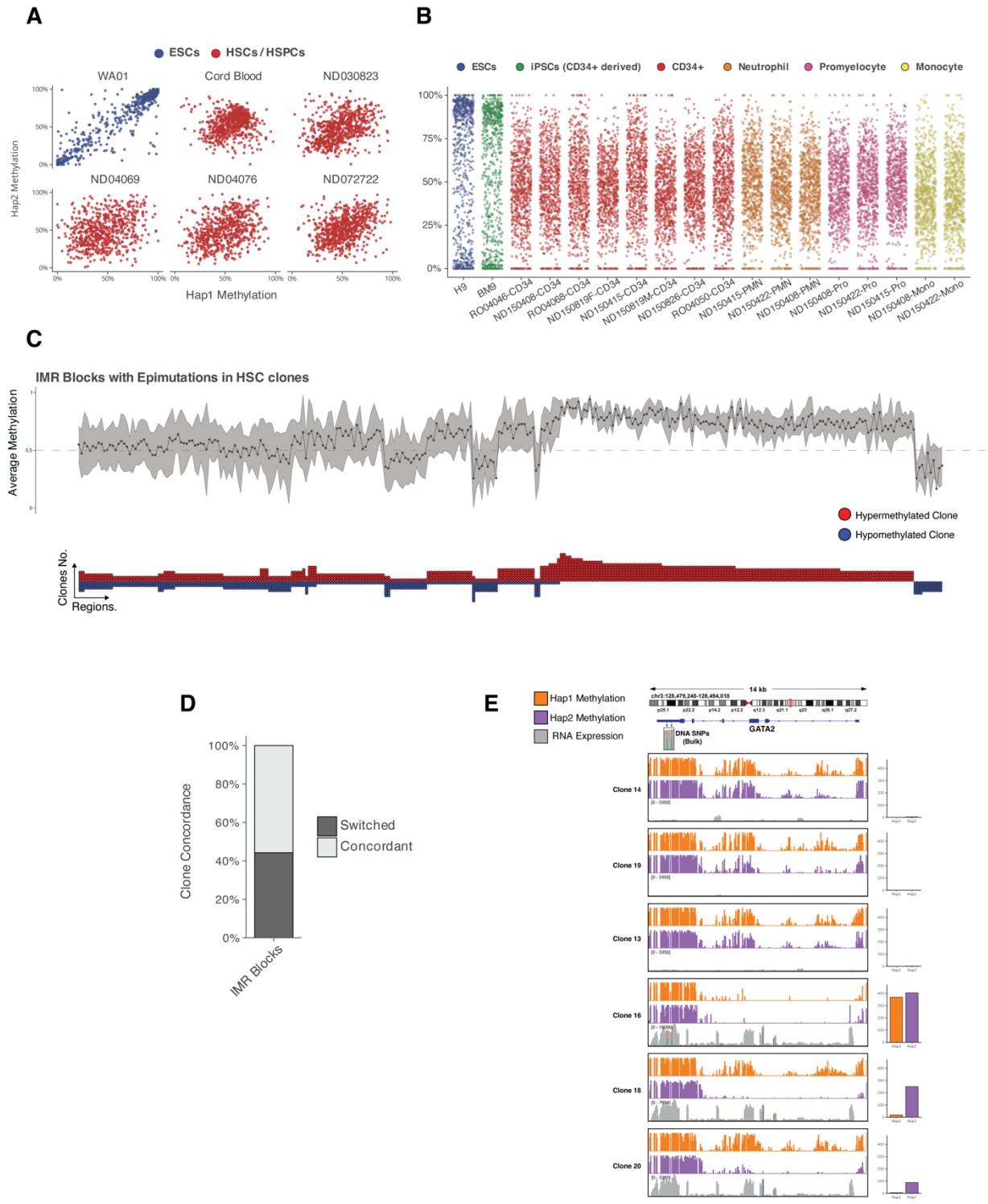
729

730 **Fig. S2. Intermediately methylated regions in normal hematopoietic cells display allelic**
731 **methylation patterns in AML.**
732 **(A)** Fragment-level WGBS methylation metrics statistics at IMRs in normal hematopoietic cells,
733 comparing DMR hotspots vs. non-hotspot DMRs. **(B)** Methylation at IMR hotspots in normal
734 hematopoietic cells, AML samples without differential methylation (no DMR called), and AML
735 samples with DMRs (DMR called). Top: distribution histograms of average methylation at IMR
736 hotspots; bottom: scatter plots comparing the average methylation of reference versus alternate
737 haplotypes for each IMR hotspot with phased WGBS reads above the coverage threshold. **(C)**
738 Fraction of IMR hotspots exhibiting allele-specific methylation (ASM) from WGBS in normal
739 hematopoietic cells, AML without DMRs, and AML with DMRs. **(D)** Number of ASM regions
740 identified from ONT haplotype-resolved methylation using the DSS R package in normal
741 hematopoietic cells and AML samples. **(E)** Aggregated CUT&Tag signal for H3K27me3 and
742 H3K4me3 modifications at IMR blocks in normal hematopoietic cells. **(F)** Aggregated
743 H3K27me3 CUT&Tag signal at IMR blocks in AML samples, with blocks grouped by
744 methylation status (hypomethylated, intermediate, or hypermethylated).
745
746


A
AML 112200 somatic mutations (WGS)
AML 895870 somatic mutations (WGS)
AML 627311 somatic mutations (WGS)
Relapse VAF
Presentation VAF
B
895870
627311
Mutation VAF (%)
IDH2
SRSF2
DNMT3A
NPM1
diagnosis_vaf
remission_vaf
relapse_vaf
C
142639
607160
613280
672014
FLT3
STAT5B
SMC1A

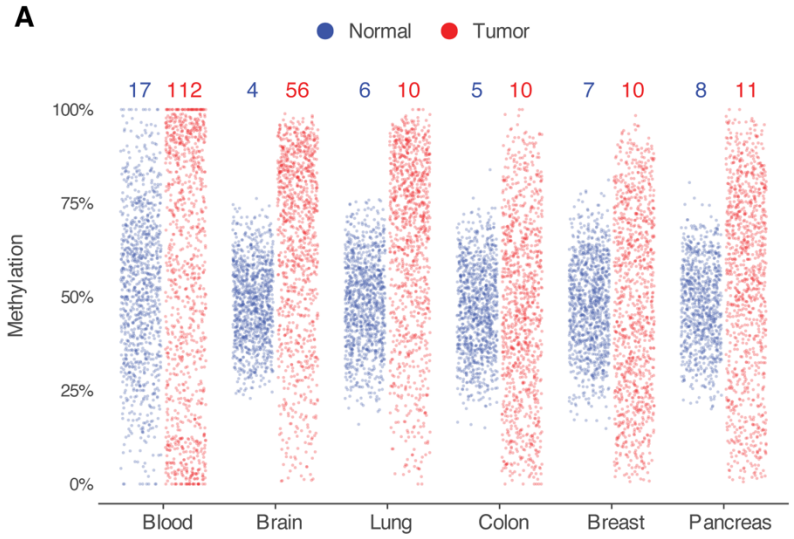
747

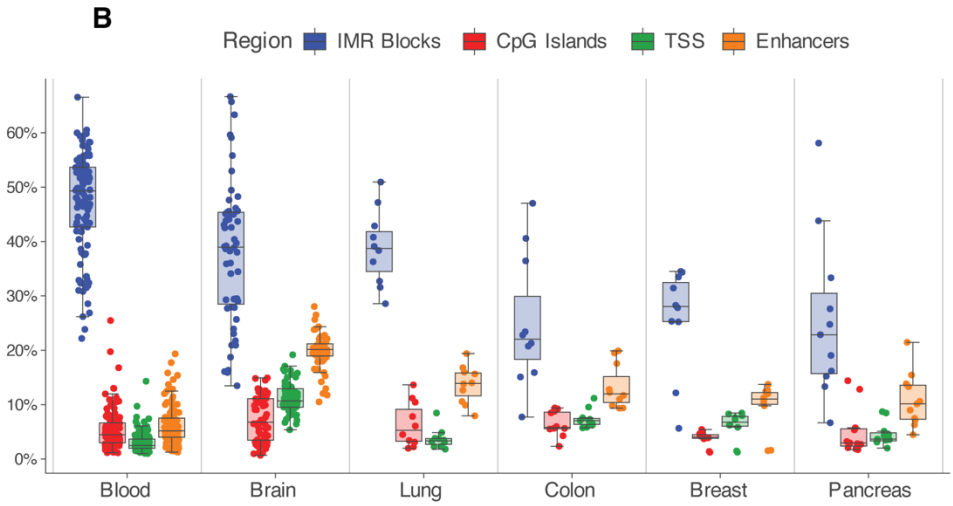
748 **Fig. S3. IMR blocks are clonal events in AML and normal hematopoietic cell populations.**
749 **(A)** Somatic mutations from whole-genome sequencing of AML patients 112200, 895870, and
750 627311 at presentation and relapse. Shown are the VAFs at relapse (Y-axis) and presentation (X-
751 axis) demonstrating their clonal relationships. **(B)** VAFs for somatic mutations from AML patient
752 895870 with nearly complete clearance of all variants in remission (left) vs. AML patient 627311
753 who had persistent variants, including *DNMT3A*^{R729W}, indicating clonal hematopoiesis at the
754 remission timepoint. Both patients had *de novo* AML and were treated with standard "7+3"
755 induction chemotherapy and were in complete morphologic remission at the remission time
756 point. **(C)** Mutation clearance plots from whole-exome sequencing of 4 additional AML patients
757 who achieved a complete morphologic remission after induction chemotherapy(*42*) .
758
759


A
ESCs
HSCs / HSPCs
WA01
Cord Blood
ND030823
ND04069
ND04076
ND072722
Hap2 Methylation
Hap1 Methylation
B
ESCs
iPSCs (CD34+ derived)
CD34+
Neutrophil
Promyelocyte
Monocyte
H9
BM9
RO04046-CD34
ND150408-CD34
RO04068-CD34
ND150819F-CD34
ND150415-CD34
ND150819M-CD34
ND150826-CD34
RO04050-CD34
ND150415-PMN
ND150422-PMN
ND150408-PMN
ND150408-Pro
ND150422-Pro
ND150415-Pro
ND150408-Mono
ND150422-Mono
C
IMR Blocks with Epimutations in HSC clones
Average Methylation
Hypermethylated Clone
Hypomethylated Clone
Clones No.
Regions.
D
Clone Concordance
Switched
Concordant
IMR Blocks
E
Hap1 Methylation
Hap2 Methylation
RNA Expression
14 kb
GATA2
DNA SNPs (Bulk)
Clone 14
Clone 19
Clone 13
Clone 16
Clone 18
Clone 20

760

761 **Fig. S4. Random allele-specific methylation at IMR blocks is established in single**
762 **hematopoietic stem cells and is associated with allele-specific gene expression. (A)**
763 Comparison of haplotype 1 versus haplotype 2 ONT methylation at IMR blocks with
764 intermediate methylation in bulk HSCs. Scatter plots are shown for WA01 embryonic stem cells
765 (ESCs), cord blood HSCs, and normal donor hematopoietic stem and progenitor cells (HSPCs).
766 **(B)** Average WGBS methylation levels at IMR blocks for H9 embryonic stem cells (ESCs), BM9
767 induced pluripotent stem cells (iPSCs) derived from CD34+ cells, CD34+ cells, and
768 differentiated myeloid lineages. Each point represents the non-phased average WGBS
769 methylation at a single single IMR for the indicated sample. **(C)** Analysis of epimutations in
770 single HSC clones at IMR blocks with intermediate methylation. The top line plot displays the
771 average methylation across all clones for each block. The dot plot below indicates individual
772 clones exhibiting either hyper-methylation (red) or hypo-methylation (blue) epimutations at these
773 blocks relative to normal HSPCs. Only IMR blocks with epimutations observed in at least four
774 clones were included in this analysis. **(D)** Bar plot quantifying the concordance of epimutation
775 type across clones for the IMR blocks shown in (C). "Switched" indicates regions where both
776 hyper- and hypo-methylation events were observed in different clones, while "Concordant"
777 indicates regions where all observed epimutations were of the same type. **(E)** Allele-specific
778 methylation is associated with allele-specific expression at the *GATA2* locus. Tracks for
779 representative single HSC clones show Haplotype 1 methylation (orange), Haplotype 2
780 methylation (purple), and RNA expression coverage (grey), with two heterozygous SNPs in the
781 last exon used for phasing RNA expression. The bar plots on the right quantify haplotype-
782 specific expression (TPM).
783


A
Normal
Tumor
17
112
4
56
6
10
5
10
7
10
8
11
100%
75%
50%
25%
0%
Methylation
Blood
Brain
Lung
Colon
Breast
Pancreas


B
Region
IMR Blocks
CpG Islands
TSS
Enhancers
60%
50%
40%
30%
20%
10%
0%
Blood
Brain
Lung
Colon
Breast
Pancreas

784

785 **Fig. S5. Epimutations at tissue-specific IMRs using WGBS cancer datasets.**

786 **(A)** Distribution of average methylation levels at tissue-specific IMRs in normal tissues (blue)
787 and their corresponding primary tumors (red), based on WGBS data. The numbers at the top of
788 the plot indicate the number of samples included in each group. Each point represents the
789 average methylation at a single IMR block within a sample. For visualization, 1,000 points were
790 randomly sampled for each group. **(B)** Percentage of various genomic features exhibiting
791 epimutations across different cancer types, based on WGBS data. The box plots compare the
792 percentage of epimutations within tissue-specific IMRs (IMR Blocks), CpG islands, tissue-
793 specific transcription start sites (TSS), and tissue-specific enhancers. Each point represents an
794 individual tumor sample.

795
796

797 **References**

798 1. A. P. Feinberg, B. Vogelstein, Hypomethylation distinguishes genes of some human cancers
799 from their normal counterparts. *Nature* **301**, 89–92 (1983).

800 2. A. P. Feinberg, B. Vogelstein, Hypomethylation of ras oncogenes in primary human cancers.
801 *Biochem Bioph Res Co* **111**, 47–54 (1983).

802 3. M. A. Gama-Sosa, V. A. Slagel, R. W. Trewyn, R. Oxenhandler, K. C. Kuo, C. W. Gehrke,
803 M. Ehrlich, The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res.* **11**,
804 6883–6894 (1983).

805 4. D. H. Spencer, D. A. Russler-Germain, S. Ketkar, N. M. Helton, T. L. Lamprecht, R. S.
806 Fulton, C. C. Fronick, M. O'Laughlin, S. E. Heath, M. Shinawi, P. Westervelt, J. E. Payton, L.
807 D. Wartman, J. S. Welch, R. K. Wilson, M. J. Walter, D. C. Link, J. F. DiPersio, T. J. Ley, CpG
808 Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression. *Cell*
809 **168**, 801-816.e13 (2017).

810 5. E. R. Wilson, N. M. Helton, S. E. Heath, R. S. Fulton, J. E. Payton, J. S. Welch, M. J. Walter,
811 P. Westervelt, J. F. DiPersio, D. C. Link, C. A. Miller, T. J. Ley, D. H. Spencer, Focal disruption
812 of DNA methylation dynamics at enhancers in IDH-mutant AML cells. *Leukemia* **36**, 935–945
813 (2022).

814 6. D. A. Russler-Germain, D. H. Spencer, M. A. Young, T. L. Lamprecht, C. A. Miller, R.
815 Fulton, M. R. Meyer, P. Erdmann-Gilmore, R. R. Townsend, R. K. Wilson, T. J. Ley, The
816 R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-Type DNMT3A by
817 Blocking Its Ability to Form Active Tetramers. *Cancer Cell* **25**, 442–454 (2014).

818 7. M. Sasaki, C. B. Knobbe, J. C. Munger, E. F. Lind, D. Brenner, A. Brüstle, I. S. Harris, R.
819 Holmes, A. Wakeham, J. Haight, A. You-Ten, W. Y. Li, S. Schalm, S. M. Su, C. Virtanen, G.
820 Reifenberger, P. S. Ohashi, D. L. Barber, M. E. Figueroa, A. Melnick, J.-C. Zúñiga-Pflücker, T.
821 W. Mak, IDH1(R132H) mutation increases murine haematopoietic progenitors and alters
822 epigenetics. *Nature* **488**, 656 (2012).

823 8. C. Lu, P. S. Ward, G. S. Kapoor, D. Rohle, S. Turcan, O. Abdel-Wahab, C. R. Edwards, R.
824 Khanin, M. E. Figueroa, A. Melnick, K. E. Wellen, D. M. O'Rourke, S. L. Berger, T. A. Chan,
825 R. L. Levine, I. K. Mellinghoff, C. B. Thompson, IDH mutation impairs histone demethylation
826 and results in a block to cell differentiation. *Nature* **483**, 474 (2012).

827 9. J. L. Glass, D. Hassane, B. J. Wouters, H. Kunimoto, R. Avellino, F. E. Garrett-Bakelman, O.
828 A. Guryanova, R. Bowman, S. Redlich, A. M. Intlekofer, C. Meydan, T. Qin, M. Fall, A.
829 Alonso, M. L. Guzman, P. J. M. Valk, C. B. Thompson, R. Levine, O. Elemento, R. Delwel, A.
830 Melnick, M. E. Figueroa, Epigenetic Identity in AML Depends on Disruption of Nonpromoter
831 Regulatory Elements and Is Affected by Antagonistic Effects of Mutations in Epigenetic
832 Modifiers. *Cancer Discov* **7**, 868–883 (2017).

10. M. E. Figueroa, O. Abdel-Wahab, C. Lu, P. S. Ward, J. Patel, A. Shih, Y. Li, N. Bhagwat, A. Vasanthakumar, H. F. Fernandez, M. S. Tallman, Z. Sun, K. Wolniak, J. K. Peeters, W. Liu, S. E. Choe, V. R. Fantin, E. Paietta, B. Löwenberg, J. D. Licht, L. A. Godley, R. Delwel, P. J. M. Valk, C. B. Thompson, R. L. Levine, A. Melnick, Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell* **18**, 553–567 (2010).

11. A. J. Gentles, S. K. Plevritis, R. Majeti, A. A. Alizadeh, Association of a Leukemic Stem Cell Gene Expression Signature With Clinical Outcomes in Acute Myeloid Leukemia. *Jama* **304**, 2706–2715 (2010).

12. D. Capper, D. T. W. Jones, M. Sill, V. Hovestadt, D. Schrimpf, D. Sturm, C. Koelsche, F. Sahm, L. Chavez, D. E. Reuss, A. Kratz, A. K. Wefers, K. Huang, K. W. Pajtler, L. Schweizer, D. Stichel, A. Olar, N. W. Engel, K. Lindenberg, P. N. Harter, A. K. Braczynski, K. H. Plate, H. Dohmen, B. K. Garvalov, R. Coras, A. Hölsken, E. Hewer, M. Bewerunge-Hudler, M. Schick, R. Fischer, R. Beschorner, J. Schittenhelm, O. Staszewski, K. Wani, P. Varlet, M. Pages, P. Temming, D. Lohmann, F. Selt, H. Witt, T. Milde, O. Witt, E. Aronica, F. Giangaspero, E. Rushing, W. Scheurlen, C. Geisenberger, F. J. Rodriguez, A. Becker, M. Preusser, C. Haberler, R. Bjerkvig, J. Cryan, M. Farrell, M. Deckert, J. Hench, S. Frank, J. Serrano, K. Kannan, A. Tsirigos, W. Brück, S. Hofer, S. Brehmer, M. Seiz-Rosenhagen, D. Hänggi, V. Hans, S. Rozsnoki, J. R. Hansford, P. Kohlhof, B. W. Kristensen, M. Lechner, B. Lopes, C. Mawrin, R. Ketter, A. Kulozik, Z. Khatib, F. Heppner, A. Koch, A. Jouvet, C. Keohane, H. Mühleisen, W. Mueller, U. Pohl, M. Prinz, A. Benner, M. Zapatka, N. G. Gottardo, P. H. Driever, C. M. Kramm, H. L. Müller, S. Rutkowski, K. von Hoff, M. C. Frühwald, A. Gnekow, G. Fleischhack, S. Tippelt, G. Calaminus, C.-M. Monoranu, A. Perry, C. Jones, T. S. Jacques, B. Radlwimmer, M. Gessi, T. Pietsch, J. Schramm, G. Schackert, M. Westphal, G. Reifenberger, P. Wesseling, M. Weller, V. P. Collins, I. Blümcke, M. Bendszus, J. Debus, A. Huang, N. Jabado, P. A. Northcott, W. Paulus, A. Gajjar, G. W. Robinson, M. D. Taylor, Z. Jaunmuktane, M. Ryzhova, M. Platten, A. Unterberg, W. Wick, M. A. Karajannis, M. Mittelbronn, T. Acker, C. Hartmann, K. Aldape, U. Schüller, R. Buslei, P. Lichter, M. Kool, C. Herold-Mende, D. W. Ellison, M. Hasselblatt, M. Snuderl, S. Brandner, A. Korshunov, A. von Deimling, S. M. Pfister, DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469–474 (2018).

13. C. Koelsche, D. Schrimpf, D. Stichel, M. Sill, F. Sahm, D. E. Reuss, M. Blattner, B. Worst, C. E. Heilig, K. Beck, P. Horak, S. Kreutzfeldt, E. Paff, S. Stark, P. Johann, F. Selt, J. Ecker, D. Sturm, K. W. Pajtler, A. Reinhardt, A. K. Wefers, P. Sievers, A. Ebrahimi, A. Suwala, F. Fernández-Klett, B. Casalini, A. Korshunov, V. Hovestadt, F. K. F. Kommoss, M. Kriegsmann, M. Schick, M. Bewerunge-Hudler, T. Milde, O. Witt, A. E. Kulozik, M. Kool, L. Romero-Pérez, T. G. P. Grünewald, T. Kirchner, W. Wick, M. Platten, A. Unterberg, M. Uhl, A. Abdollahi, J. Debus, B. Lehner, C. Thomas, M. Hasselblatt, W. Paulus, C. Hartmann, O. Staszewski, M. Prinz, J. Hench, S. Frank, Y. M. H. Versleijen-Jonkers, M. E. Weidema, T. Mentzel, K. Griewank, E. de Álava, J. D. Martín, M. A. I. Gastearena, K. T.-E. Chang, S. Y. Y. Low, A. Cuevas-Bourdier, M. Mittelbronn, M. Mynarek, S. Rutkowski, U. Schüller, V. F. Mautner, J. Schittenhelm, J. Serrano, M. Snuderl, R. Büttner, T. Klingebiel, R. Buslei, M. Gessler, P. Wesseling, W. N. M. Dinjens, S. Brandner, Z. Jaunmuktane, I. Lyskjær, P. Schirmacher, A. Stenzinger, B. Brors, H. Glimm, C. Heining, O. M. Tirado, M. Sáinz-Jaspeado, J. Mora, J. Alonso, X. G. del Muro, S.

876 Moran, M. Esteller, J. K. Benhamida, M. Ladanyi, E. Wardelmann, C. Antonescu, A. Flanagan,
877 U. Dirksen, P. Hohenberger, D. Baumhoer, W. Hartmann, C. Vokuhl, U. Flucke, I. Petersen, G.
878 Mechtersheimer, D. Capper, D. T. W. Jones, S. Fröhling, S. M. Pfister, A. von Deimling,
879 Sarcoma classification by DNA methylation profiling. *Nat. Commun.* **12**, 498 (2021).

880 14. B. Giacopelli, M. Wang, A. Cleary, Y.-Z. Wu, A. R. Schultz, M. Schmutz, J. S. Blachly, A.-
881 K. Eisfeld, B. Mundy-Bosse, S. Vosberg, P. A. Greif, R. Claus, L. Bullinger, R. Garzon, K. R.
882 Coombes, C. D. Bloomfield, B. J. Druker, J. W. Tyner, J. C. Byrd, C. C. Oakes, DNA
883 methylation epitypes highlight underlying developmental and disease pathways in acute myeloid
884 leukemia. *Genome Res* **31**, gr.269233.120 (2021).

885 15. S. B. Baylin, P. A. Jones, Epigenetic Determinants of Cancer. *Csh Perspect Biol* **8**, a019505
886 (2016).

887 16. S. Saghafinia, M. Mina, N. Riggi, D. Hanahan, G. Ciriello, Pan-Cancer Landscape of
888 Aberrant DNA Methylation across Human Tumors. *Cell Rep.* **25**, 1066-1080.e8 (2018).

889 17. S. Hetzel, E. Hodis, E. T. Triglia, A. Kovacsovics, K. Steinmann, A. Gnirke, M. Cui, D.
890 McQuaid, R. Weigert, G. Pohl, M. D. Muzumdar, S. Leyvraz, U. Keilholz, M.-L. Yaspo, A.
891 Regev, H. Kretzmer, Z. D. Smith, A. Meissner, Direct genetic transformation bypasses tumor-
892 associated DNA methylation alterations. *Genome Biol.* **26**, 212 (2025).

893 18. S. Baylin, T. H. Bestor, Altered methylation patterns in cancer cell genomes: Cause or
894 consequence? *Cancer Cell* **1**, 299–305 (2002).

895 19. M. P. Hitchins, Constitutional epimutation as a mechanism for cancer causality and
896 heritability? *Nat. Rev. Cancer* **15**, 625–634 (2015).

897 20. B. Horsthemke, Heritable germline epimutations in humans. *Nat. Genet.* **39**, 573–574 (2007).

898 21. R. L. Poole, L. E. Docherty, A. A. Sayegh, A. Caliebe, C. Turner, E. Baple, E. Wakeling, L.
899 Harrison, A. Lehmann, I. K. Temple, D. J. G. Mackay, O. behalf of the I. C. I. Consortium,
900 Targeted methylation testing of a patient cohort broadens the epigenetic and clinical description
901 of imprinting disorders. *Am. J. Méd. Genet. Part A* **161**, 2174–2182 (2013).

902 22. K. Buiting, S. Groß, C. Lich, G. Gillessen-Kaesbach, O. El-Maarri, B. Horsthemke,
903 Epimutations in Prader-Willi and Angelman Syndromes: A Molecular Study of 136 Patients with
904 an Imprinting Defect. *Am. J. Hum. Genet.* **72**, 571–577 (2003).

905 23. M. P. Hitchins, R. L. Ward, Constitutional (germline) MLH1 epimutation as an aetiological
906 mechanism for hereditary non-polyposis colorectal cancer. *J. Méd. Genet.* **46**, 793 (2009).

907 24. C. M. Suter, D. I. K. Martin, R. L. Ward, Germline epimutation of MLH1 in individuals with
908 multiple cancers. *Nat. Genet.* **36**, 497–501 (2004).

909 25. R. Mulet-Lazaro, S. van Herk, C. Erpelinck, E. Bindels, M. A. Sanders, C. Vermeulen, I.
910 Renkens, P. Valk, A. M. Melnick, J. de Ridder, M. Rehli, C. Gebhard, R. Delwel, B. J. Wouters,
911 Allele-specific expression of GATA2 due to epigenetic dysregulation in CEBPA double-mutant
912 AML. *Blood* **138**, 160–177 (2021).

913 26. A. Gimelbrant, J. N. Hutchinson, B. R. Thompson, A. Chess, Widespread Monoallelic
914 Expression on Human Autosomes. *Science* **318**, 1136–1140 (2007).

915 27. O. Stewart, C. Gruber, H. E. Randolph, R. Patel, M. Ramba, E. Calzoni, L. H. Huang, J.
916 Levy, S. Buta, A. Lee, C. Sazeides, Z. Prue, D. P. H. van Konijnenburg, I. K. Chinn, L. A.
917 Pedroza, J. R. Lupski, E. G. Schmitt, M. A. Cooper, A. Puel, X. Peng, S. Boisson-Dupuis, J.
918 Bustamante, S. Okada, M. Martin-Fernandez, J. S. Orange, J.-L. Casanova, J. D. Milner, D.
919 Bogunovic, Monoallelic expression can govern penetrance of inborn errors of immunity. *Nature*,
920 1–12 (2025).

921 28. A. F. A. Seraihi, A. Rio-Machin, K. Tawana, C. Bödör, J. Wang, A. Nagano, J. A. Heward,
922 S. Iqbal, S. Best, N. Lea, D. McLornan, E. J. Kozyra, M. W. Wlodarski, C. M. Niemeyer, H.
923 Scott, C. Hahn, A. Ellison, H. Tummala, S. R. Cardoso, T. Vulliamy, I. Dokal, T. Butler, M.
924 Smith, J. Cavenagh, J. Fitzgibbon, GATA2 monoallelic expression underlies reduced penetrance
925 in inherited GATA2-mutated MDS/AML. *Leukemia* **32**, 2502–2507 (2018).

926 29. L. Marion-Poll, B. Forêt, D. Zielinski, F. Massip, M. Attia, A. C. Carter, L. Syx, H. Y.
927 Chang, A.-V. Gendrel, E. Heard, Locus specific epigenetic modalities of random allelic
928 expression imbalance. *Nat. Commun.* **12**, 5330 (2021).

929 30. Q. Deng, D. Ramsköld, B. Reinius, R. Sandberg, Single-Cell RNA-Seq Reveals Dynamic,
930 Random Monoallelic Gene Expression in Mammalian Cells. *Science* **343**, 193–196 (2014).

931 31. H. Noushmehr, D. J. Weisenberger, K. Diefes, H. S. Phillips, K. Pujara, B. P. Berman, F.
932 Pan, C. E. Pelloski, E. P. Sulman, K. P. Bhat, R. G. W. Verhaak, K. A. Hoadley, D. N. Hayes, C.
933 M. Perou, H. K. Schmidt, L. Ding, R. K. Wilson, D. V. D. Berg, H. Shen, H. Bengtsson, P.
934 Neuvial, L. M. Cope, J. Buckley, J. G. Herman, S. B. Baylin, P. W. Laird, K. Aldape, T. C. G. A.
935 R. Network, Identification of a CpG Island Methylator Phenotype that Defines a Distinct
936 Subgroup of Glioma. *Cancer Cell* **17**, 510–522 (2010).

937 32. T. J. Ley, C. Miller, L. Ding, B. J. Raphael, A. J. Mungall, A. G. Robertson, K. Hoadley, T.
938 J. Triche, P. W. Laird, J. D. Baty, L. L. Fulton, R. Fulton, S. E. Heath, J. Kalicki-Veizer, C.
939 Kandoth, J. M. Klco, D. C. Koboldt, K.-L. Kanchi, S. Kulkarni, T. L. Lamprecht, D. E. Larson,
940 L. Lin, C. Lu, M. D. McLellan, J. F. McMichael, J. Payton, H. Schmidt, D. H. Spencer, M. H.
941 Tomasson, J. W. Wallis, L. D. Wartman, M. A. Watson, J. Welch, M. C. Wendl, A. Ally, M.
942 Balasundaram, I. Birol, Y. Butterfield, R. Chiu, A. Chu, E. Chuah, H.-J. Chun, R. Corbett, N.
943 Dhalla, R. Guin, A. He, C. Hirst, M. Hirst, R. A. Holt, S. Jones, A. Karsan, D. Lee, H. I. Li, M.
944 A. Marra, M. Mayo, R. A. Moore, K. Mungall, J. Parker, E. Pleasance, P. Plettner, J. Schein, D.
945 Stoll, L. Swanson, A. Tam, N. Thiessen, R. Varhol, N. Wye, Y. Zhao, S. Gabriel, G. Getz, C.
946 Sougnez, L. Zou, M. D. M. Leiserson, F. Vandin, H.-T. Wu, F. Applebaum, S. B. Baylin, R.
947 Akbani, B. M. Broom, K. Chen, T. C. Motter, K. Nguyen, J. N. Weinstein, N. Zhang, M. L.

948 Ferguson, C. Adams, A. Black, J. Bowen, J. Gastier-Foster, T. Grossman, T. Lichtenberg, L.
949 Wise, T. Davidsen, J. A. Demchok, K. R. M. Shaw, M. Sheth, H. J. Sofia, L. Yang, J. R.
950 Downing, G. Eley, Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid
951 Leukemia. *New Engl J Med* **368**, 2059–2074 (2013).

952 33. T. Mazor, A. Pankov, B. E. Johnson, C. Hong, E. G. Hamilton, R. J. A. Bell, I. V. Smirnov,
953 G. F. Reis, J. J. Phillips, M. J. Barnes, A. Idbaih, A. Alentorn, J. J. Kloezeman, M. L. M.
954 Lamfers, A. W. Bollen, B. S. Taylor, A. M. Molinaro, A. B. Olshen, S. M. Chang, J. S. Song, J.
955 F. Costello, DNA Methylation and Somatic Mutations Converge on the Cell Cycle and Define
956 Similar Evolutionary Histories in Brain Tumors. *Cancer Cell* **28**, 307–317 (2015).

957 34. T. C. G. A. R. Network, Comprehensive, Integrative Genomic Analysis of Diffuse Lower-
958 Grade Gliomas. *New Engl J Medicine* **372**, 2481–2498 (2015).

959 35. M. E. Figueroa, S. Lugthart, Y. Li, C. Erpelinck-Verschueren, X. Deng, P. J. Christos, E.
960 Schifano, J. Booth, W. van Putten, L. Skrabanek, F. Campagne, M. Mazumdar, J. M. Greally, P.
961 J. M. Valk, B. Löwenberg, R. Delwel, A. Melnick, DNA Methylation Signatures Identify
962 Biologically Distinct Subtypes in Acute Myeloid Leukemia. *Cancer Cell* **17**, 13–27 (2010).

963 36. J. Ernst, M. Kellis, Chromatin-state discovery and genome annotation with ChromHMM. *Nat*
964 *Protoc* **12**, 2478–2492 (2017).

965 37. G. Elliott, C. Hong, X. Xing, X. Zhou, D. Li, C. Coarfa, R. J. A. Bell, C. L. Maire, K. L.
966 Ligon, M. Sigaroudinia, P. Gascard, T. D. Tlsty, R. A. Harris, L. C. Schalkwyk, M. Bilenky, J.
967 Mill, P. J. Farnham, M. Kellis, M. A. Marra, A. Milosavljevic, M. Hirst, G. D. Stormo, T. Wang,
968 J. F. Costello, Intermediate DNA methylation is a conserved signature of genome regulation.
969 *Nat. Commun.* **6**, 6363 (2015).

970 38. Y. Ding, K. Cai, L. Liu, Z. Zhang, X. Zheng, J. Shi, mHapTk: a comprehensive toolkit for
971 the analysis of DNA methylation haplotypes. *Bioinformatics* **38**, 5141–5143 (2022).

972 39. Y. Feng, Z. Zhang, Y. Hong, Y. Ding, L. Liu, S. Gao, H. Fang, J. Shi, A DNA methylation
973 haplotype block landscape in human tissues and preimplantation embryos reveals regulatory
974 elements defined by comethylation patterns. *Genome Res.* **33**, 2041–2052 (2023).

975 40. S. Guo, D. Diep, N. Plongthongkum, H.-L. Fung, K. Zhang, K. Zhang, Identification of
976 methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor
977 tissue-of-origin mapping from plasma DNA. *Nat. Genet.* **49**, 635–642 (2017).

978 41. A. B. Stergachis, B. M. Debo, E. Haugen, L. S. Churchman, J. A. Stamatoyannopoulos,
979 Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* **368**,
980 1449–1454 (2020).

981 42. M. J. Slade, R. Ghasemi, M. O'Laughlin, T. Burton, R. S. Fulton, H. J. Abel, E. J.
982 Duncavage, T. J. Ley, M. A. Jacoby, D. H. Spencer, Persistent Molecular Disease in Adult

983 Patients With AML Evaluated With Whole-Exome and Targeted Error-Corrected DNA
984 Sequencing. *JCO Precis. Oncol.* **7**, e2200559 (2023).

985 43. K. Hu, J. Yu, K. Suknuntha, S. Tian, K. Montgomery, K.-D. Choi, R. Stewart, J. A.
986 Thomson, I. I. Slukvin, Efficient generation of transgene-free induced pluripotent stem cells
987 from normal and neoplastic bone marrow and cord blood mononuclear cells. *Blood* **117**, e109–
988 e119 (2011).

989 44. J. Rosenski, A. Peretz, J. Magenheim, N. Loyfer, R. Shemer, B. Glaser, Y. Dor, T. Kaplan,
990 Atlas of imprinted and allele-specific DNA methylation in the human body. *Nat. Commun.* **16**,
991 2141 (2025).

992 45. N. Loyfer, J. Magenheim, A. Peretz, G. Cann, J. Bredno, A. Klochendler, I. Fox-Fisher, S.
993 Shabi-Porat, M. Hecht, T. Pelet, J. Moss, Z. Drawshy, H. Amini, P. Moradi, S. Nagaraju, D.
994 Bauman, D. Shveiky, S. Porat, U. Dior, G. Rivkin, O. Or, N. Hirshoren, E. Carmon, A. Pikarsky,
995 A. Khalaileh, G. Zamir, R. Grinbaum, M. A. Gazala, I. Mizrahi, N. Shussman, A. Korach, O.
996 Wald, U. Izhar, E. Erez, V. Yutkin, Y. Samet, D. R. Golinkin, K. L. Spalding, H. Druid, P.
997 Arner, A. M. J. Shapiro, M. Grompe, A. Aravanis, O. Venn, A. Jamshidi, R. Shemer, Y. Dor, B.
998 Glaser, T. Kaplan, A DNA methylation atlas of normal human cell types. *Nature*, 1–10 (2023).

999 46. K. O'Neill, E. Pleasance, J. Fan, V. Akbari, G. Chang, K. Dixon, V. Csizmok, S.
1000 MacLennan, V. Porter, A. Galbraith, C. J. Grisdale, L. Culibrk, J. H. Dupuis, R. Corbett, J.
1001 Hopkins, R. Bowlby, P. Pandoh, D. E. Smailus, D. Cheng, T. Wong, C. Frey, Y. Shen, E. Lewis,
1002 L. F. Paulin, F. J. Sedlazeck, J. M. T. Nelson, E. Chuah, K. L. Mungall, R. A. Moore, R. Coope,
1003 A. J. Mungall, M. K. McConechy, L. M. Williamson, K. A. Schrader, S. Yip, M. A. Marra, J.
1004 Laskin, S. J. M. Jones, Long-read sequencing of an advanced cancer cohort resolves
1005 rearrangements, unravels haplotypes, and reveals methylation landscapes. *Cell Genom.* **4**,
1006 100674 (2024).

1007 47. Z. Zhang, Y. Hong, S. Zhang, X. Zhu, L. Liu, X. Liao, H. Gu, H. Fang, J. Shi, Toward the
1008 DNA methylation haplotype map of 11 common solid cancers. *Cell Rep.* **44**, 116197 (2025).

1009 48. A. L. Costa, D. Doncevic, Y. Wu, L. Yang, K. H. Man, A.-S. Spreng, H. Winter, M. N. C.
1010 Fletcher, B. Radlwimmer, C. Herrmann, A new IDH-independent hypermethylation phenotype is
1011 associated with astrocyte-like cell state in glioblastoma. *Genome Biol.* **26**, 192 (2025).

1012 49. Y. Wu, M. Fletcher, Z. Gu, Q. Wang, B. Costa, A. Bertoni, K.-H. Man, M. Schlotter, J.
1013 Felsberg, J. Mangei, M. Barbus, A.-C. Gaupel, W. Wang, T. Weiss, R. Eils, M. Weller, H. Liu,
1014 G. Reifenberger, A. Korshunov, P. Angel, P. Lichter, C. Herrmann, B. Radlwimmer,
1015 Glioblastoma epigenome profiling identifies SOX10 as a master regulator of molecular tumour
1016 subtype. *Nat. Commun.* **11**, 6434 (2020).

1017 50. A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P.
1018 Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D.
1019 Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. Pfenning, X. Wang, M.
1020 C. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D.

1021 Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K.
1022 Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S.
1023 Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph,
1024 V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-
1025 Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, N. Abdennur, M. Adli, M.
1026 Akerman, L. Barrera, J. Antosiewicz-Bourget, T. Ballinger, M. J. Barnes, D. Bates, R. J. A. Bell,
1027 D. A. Bennett, K. Bianco, C. Bock, P. Boyle, J. Brinchmann, P. Caballero-Campo, R. Camahort,
1028 M. J. Carrasco-Alfonso, T. Charnecki, H. Chen, Z. Chen, J. B. Cheng, S. Cho, A. Chu, W.-Y.
1029 Chung, C. Cowan, Q. A. Deng, V. Deshpande, M. Diegel, B. Ding, T. Durham, L. Echipare, L.
1030 Edsall, D. Flowers, O. Genbacev-Krtolica, C. Gifford, S. Gillespie, E. Giste, I. A. Glass, A.
1031 Gnirke, M. Gormley, H. Gu, J. Gu, D. A. Hafler, M. J. Hangauer, M. Hariharan, M. Hatan, E.
1032 Haugen, Y. He, S. Heimfeld, S. Herlofsen, Z. Hou, R. Humbert, R. Issner, A. R. Jackson, H. Jia,
1033 P. Jiang, A. K. Johnson, T. Kadlecek, B. Kamoh, M. Kapidzic, J. Kent, A. Kim, M.
1034 Kleinewietfeld, S. Klugman, J. Krishnan, S. Kuan, T. Kutyavin, A.-Y. Lee, K. Lee, J. Li, N. Li,
1035 Y. Li, K. L. Ligon, S. Lin, Y. Lin, J. Liu, Y. Liu, C. J. Luckey, Y. P. Ma, C. Maire, A. Marson, J.
1036 S. Mattick, M. Mayo, M. McMaster, H. Metsky, T. Mikkelsen, D. Miller, M. Miri, E. Mukame,
1037 R. P. Nagarajan, F. Neri, J. Nery, T. Nguyen, H. O'Geen, S. Paithankar, T. Papayannopoulou, M.
1038 Pelizzola, P. Plettner, N. E. Propson, S. Raghuraman, B. J. Raney, A. Raubitschek, A. P.
1039 Reynolds, H. Richards, K. Riehle, P. Rinaudo, J. F. Robinson, N. B. Rockweiler, E. Rosen, E.
1040 Rynes, J. Schein, R. Sears, T. Sejnowski, A. Shafer, L. Shen, R. Shoemaker, M. Sigaroudinia, I.
1041 Slukvin, S. Stehling-Sun, R. Stewart, S. L. Subramanian, K. Suknuntha, S. Swanson, S. Tian, H.
1042 Tilden, L. Tsai, M. Urich, I. Vaughn, J. Vierstra, S. Vong, U. Wagner, H. Wang, T. Wang, Y.
1043 Wang, A. Weiss, H. Whitton, A. Wildberg, H. Witt, K.-J. Won, M. Xie, X. Xing, I. Xu, Z. Xuan,
1044 Z. Ye, C. Yen, P. Yu, X. Zhang, X. Zhang, J. Zhao, Y. Zhou, J. Zhu, Y. Zhu, S. Ziegler, A. E.
1045 Beaudet, L. A. Boyer, P. L. D. Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W.
1046 Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang,
1047 R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M.
1048 Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis, A.
1049 Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z.
1050 Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G.
1051 Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y.
1052 Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D.
1053 Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K.
1054 Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S.
1055 Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph,
1056 V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-
1057 Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer,
1058 P. L. D. Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T.
1059 McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M.
1060 Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A.
1061 Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of
1062 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

1063 51. M. Chen, R. Fu, Y. Chen, L. Li, S.-W. Wang, High-resolution, noninvasive single-cell
1064 lineage tracing in mice and humans based on DNA methylation epimutations. *Nat. Methods*, 1–
1065 11 (2025).

1066 52. M. Scherer, I. Singh, M. M. Braun, C. Szu-Tu, P. S. Sanchez, D. Lindenhofer, N. A.
1067 Jakobsen, V. Körber, M. Kardorff, L. Nitsch, P. Kautz, J. Rühle, A. Bianchi, L. Cozzuto, R.
1068 Frömel, S. Beneyto-Calabuig, C. Lareau, A. T. Satpathy, R. Beekman, L. M. Steinmetz, S.
1069 Raffel, L. S. Ludwig, P. Vyas, A. Rodriguez-Fraticelli, L. Velten, Clonal tracing with somatic
1070 epimutations reveals dynamics of blood ageing. *Nature* **643**, 478–487 (2025).

1071 53. S. B. Simpkins, T. Bocker, E. M. Swisher, D. G. Mutch, D. J. Gersell, A. J. Kovatich, J. P.
1072 Palazzo, R. Fishel, P. J. Goodfellow, MLH1 Promoter Methylation and Gene Silencing is the
1073 Primary Cause of Microsatellite Instability in Sporadic Endometrial Cancers. *Hum. Mol. Genet.*
1074 **8**, 661–666 (1999).

1075 54. K. Shafin, T. Pesout, P.-C. Chang, M. Nattestad, A. Kolesnikov, S. Goel, G. Baid, M.
1076 Kolmogorov, J. M. Eizenga, K. H. Miga, P. Carnevali, M. Jain, A. Carroll, B. Paten, Haplotype-
1077 aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore
1078 long-reads. *Nat Methods* **18**, 1322–1332 (2021).

1079 55. M. Smolka, L. F. Paulin, C. M. Grochowski, D. W. Horner, M. Mahmoud, S. Behera, E.
1080 Kalef-Ezra, M. Gandhi, K. Hong, D. Pehlivan, S. W. Scholz, C. M. B. Carvalho, C. Proukakis, F.
1081 J. Sedlazeck, Detection of mosaic and population-level structural variants with Sniffles2. *Nat.*
1082 *Biotechnol.* **42**, 1571–1580 (2024).

1083 56. M. Martin, P. Ebert, T. Marschall, Haplotyping, Methods and Protocols. *Methods Mol. Biol.*
1084 **2590**, 127–138 (2022).

1085 57. B. S. Pedersen, A. R. Quinlan, Mosdepth: quick coverage calculation for genomes and
1086 exomes. *Bioinformatics* **34**, 867–868 (2017).

1087 58. A. Jha, S. C. Bohaczuk, Y. Mao, J. Ranchalis, B. J. Mallory, A. T. Min, M. O. Hamm, E.
1088 Swanson, D. Dubocanin, C. Finkbeiner, T. Li, D. Whittington, W. S. Noble, A. B. Stergachis, M.
1089 R. Vollger, DNA-m6A calling and integrated long-read epigenetic and genetic analysis with
1090 fibertools. *Genome Res.* **34**, 1976–1986 (2024).

1091 59. S. Henikoff, J. G. Henikoff, H. S. Kaya-Okur, K. Ahmad, Efficient chromatin accessibility
1092 mapping in situ by nucleosome-tethered tagmentation. *Elife* **9**, e63274 (2020).

1093 60. P. A. Ewels, A. Peltzer, S. Fillinger, H. Patel, J. Alneberg, A. Wilm, M. U. Garcia, P. D.
1094 Tommaso, S. Nahnsen, The nf-core framework for community-curated bioinformatics pipelines.
1095 *Nat. Biotechnol.* **38**, 276–278 (2020).

1096 61. H. Wu, T. Xu, H. Feng, L. Chen, B. Li, B. Yao, Z. Qin, P. Jin, K. N. Conneely, Detection of
1097 differentially methylated regions from whole-genome bisulfite sequencing data without
1098 replicates. *Nucleic Acids Res* **43**, gkv715 (2015).

1099 62. K. D. Hansen, B. Langmead, R. A. Irizarry, BSmooth: from whole genome bisulfite
1100 sequencing reads to differentially methylated regions. *Genome Biol* **13**, R83 (2012).

1101 63. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic
1102 features. *Bioinformatics* **26**, 841–842 (2010).

1103 64. V. Akbari, J.-M. Garant, K. O'Neill, P. Pandoh, R. Moore, M. A. Marra, M. Hirst, S. J.
1104 Jones, Genome-wide detection of imprinted differentially methylated regions using nanopore
1105 sequencing. *eLife* **11**, e77898 (2022).

1106 65. D. D. Jima, D. A. Skaar, A. Planchart, A. Motsinger-Reif, S. E. Cevik, S. S. Park, M.
1107 Cowley, F. Wright, J. House, A. Liu, R. L. Jirtle, C. Hoyo, Genomic map of candidate human
1108 imprint control regions: the imprintome. *Epigenetics* **17**, 1920–1943 (2022).

1109 66. M. R. Vollger, E. G. Swanson, S. J. Neph, J. Ranchalis, K. M. Munson, C.-H. Ho, Y. H. H.
1110 Cheng, A. E. Sedeño-Cortés, W. E. Fondrie, S. C. Bohaczuk, M. A. Dippel, Y. Mao, N. L.
1111 Parmalee, B. J. Mallory, W. T. Harvey, Y. Kwon, G. H. Garcia, K. Hoekzema, J. G. Meyer, M.
1112 Cicek, E. E. Eichler, W. S. Noble, D. M. Witten, J. T. Bennett, J. P. Ray, A. B. Stergachis, A
1113 haplotype-resolved view of human gene regulation. *bioRxiv*, 2024.06.14.599122 (2025).

1114 67. N. Loyfer, J. Rosenski, T. Kaplan, wgbstools: A computational suite for DNA methylation
1115 sequencing data representation, visualization, and analysis. *bioRxiv*, 2024.05.08.593132 (2024).

1116