

1 **Simultaneous quantification of protein-DNA contacts and transcriptomes in**
2 **single cells**

3
4 Koos Rooijers^{1,5}, Corina M. Markodimitraki^{1,5}, Franka J. Rang^{1,6}, Sandra S. de Vries^{1,6}, Alex
5 Chialastri^{2,3}, Kim de Luca¹, Dylan Mooijman^{1,4}, Siddharth S. Dey^{2,3} * and Jop Kind¹ *

6
7 ¹ Oncode Institute, Hubrecht Institute–KNAW and University Medical Center Utrecht,
8 Utrecht, The Netherlands.

9 ² Department of Chemical Engineering, University of California Santa Barbara, Santa
10 Barbara, CA 93106, USA.

11 ³ Center for Bioengineering, University of California Santa Barbara, Santa Barbara, CA
12 93106, USA.

13 ⁴ Present address: Genome Biology Unit, European Molecular Biology Laboratory,
14 Heidelberg, Germany.

15 ⁵ These authors contributed equally to this work

16 ⁶ These authors contributed equally to this work

17
18 *Corresponding authors: S.S.D. (sdey@ucsb.edu) and J.K. (j.kind@hubrecht.eu).
19

20 **Abstract**

21 The epigenome plays a critical role in regulating gene expression in mammalian cells.
22 However, understanding how cell-to-cell heterogeneity in the epigenome influences gene
23 expression variability remains a major challenge. Here we report a novel method for
24 simultaneous single-cell quantification of protein-DNA contacts with DamID and
25 transcriptomics (scDamID&T). This method enables quantifying the impact of protein-DNA
26 contacts on gene expression from the same cell. By profiling lamina-associated domains
27 (LADs) in human cells, we reveal different dependencies between genome-nuclear lamina
28 (NL) association and gene expression in single cells. In addition, we introduce the *E. coli*
29 methyltransferase, Dam, as an *in vivo* marker of chromatin accessibility in single cells and
30 show that scDamID&T can be utilized as a general technology to identify cell types *in silico*
31 while simultaneously determining the underlying gene-regulatory landscape. With this
32 strategy the effect of chromatin states, transcription factor binding, and genome organization
33 on the acquisition of cell-type specific transcriptional programs can be quantified.

34 **Main**

35 mRNA output is tightly regulated at many levels to ensure the precise coordination of cell-
36 type specific gene expression programs. On the transcriptional level, packaging of DNA into
37 chromatin can control access of transcriptional regulators to functional DNA elements like
38 enhancers and promoters. Higher levels of organization that contribute to the regulation of
39 gene expression involve the spatial segmentation of the genome into compartments with
40 transcriptionally permissive or repressive gene regulatory activities. Failure to integrate and
41 coordinate the multi-layered regulatory control of gene expression can result in
42 developmental defects and the commencement of disease. To understand the regulation of
43 gene expression it is key to dissect the direct relationships between epigenetic and
44 transcriptomic heterogeneity. To this end, it is pivotal to develop techniques that enable
45 simultaneous measurements of the epigenome together with the transcriptome from the same
46 cell.

47 Recent advances in measuring genome architecture (HiC, DamID)¹⁻⁴, chromatin
48 accessibility (ATAC-seq and DNaseI-seq)⁵⁻⁷, DNA methylation (5mC)⁸⁻¹⁰, DNA
49 hydroxymethylation (5hmC)¹¹ and histone PTMs post-translational modifications (ChIP-
50 seq)¹² in single cells have enabled studies to characterize cell-to-cell heterogeneity at the
51 gene-regulatory level. More recently, multiomics methods to study direct single-cell
52 associations between genomic or epigenetic variations and transcriptional heterogeneity¹³⁻¹⁶
53 have provided the first methods to directly link upstream regulatory elements to
54 transcriptional output from the same cell. Protein-DNA interactions play a critical role in
55 regulating gene expression and therefore we have developed a new technology to
56 simultaneously quantify these interactions in conjunction with transcriptomic measurements
57 from the same cell without requiring physical separation of the nucleic acids.

58 DamID involves the fusion of the *E.coli* Dam adenine methyltransferase to a protein
59 of interest, followed by the *in vivo* expression of the fusion protein to enable detection of
60 protein-DNA interactions. For single-cell applications, a major advantage of the DamID
61 method is that it minimizes biochemical losses arising from antibody-based pulldowns or
62 degradation of genomic DNA (gDNA) that occurs in bisulfite-based methods. Further, as
63 DamID is an *in vivo* method, protein-DNA interactions can be measured over varying time
64 windows and can also be used to record cumulative protein-DNA interactions¹⁷. Currently, no
65 methods exist to quantify protein-DNA interactions for an arbitrary protein-of-interest and
66 transcriptomes in single cells. We therefore chose to benchmark scDamID&T and compare it
67 to the previously reported single-cell DamID (scDamID) method where lamina-associated

68 domains (LADs) were detected using a Dam-LmnB1 fusion protein². Furthermore, we
69 exploited the expression of untethered Dam to obtain DNA accessibility profiles
70 simultaneously with transcriptome measurements and employed the scDamID&T technology
71 to generate combined and allele-resolved single-cell measurements in hybrid mouse
72 embryonic stem cells.

73 To improve the scDamID method and make it compatible with simultaneous mRNA
74 measurement in single cells, we optimized several shortcomings of the previously developed
75 protocol². The improvements include (1) the requirement of one, rather than two ligation
76 events to amplify fragmented gDNA molecules, (2) switching from PCR to linear
77 amplification through *in vitro* transcription, (3) inclusion of unique molecule identifiers
78 (UMI) for both gDNA- and mRNA-derived reads, and (4) the use of liquid-handling robots
79 that result in rapid and higher processing throughputs of thousands of single cells per day
80 together with reduced reaction volumes, and a more consistent sample quality. As described
81 previously², KBM7 cells (a near haploid myeloid leukemia cell line, except for chr8 and parts
82 of chr15) expressing either untethered Dam or a Dam-LmnB1 fusion protein and the 2-colour
83 Fucci reporter system¹⁸ are sorted by FACS at the G1/S cell cycle transition 15 hours post-
84 induction of Dam with Shield1². After single cells are sorted into 384-well plates, poly-
85 adenylated mRNA is reverse transcribed using primers that contain a T7 promoter, P5
86 Illumina adapter, a random UMI sequence, and mRNA- and cell-specific barcodes in the
87 overhang, as described previously for the CEL-Seq protocol¹⁹⁻²⁰ (Fig. 1a). Second strand
88 synthesis is then performed to generate double-stranded cDNA. Next, the reaction mixture,
89 containing tagged cDNA molecules and gDNA, is digested with the restriction enzyme DpnI.
90 DpnI recognizes adenine residues that are methylated by Dam in a GATC context and creates
91 blunt double-stranded cuts in gDNA. Double-stranded adapters are then ligated to digested
92 gDNA molecules (Fig. 1a). Similar in design to the RT primers, the double-stranded adapters
93 contain a T7 promoter, P5 Illumina adapter, UMI, and gDNA- and cell-specific barcodes.
94 Single cells are then pooled, and cDNA and ligated gDNA molecules, both containing T7
95 promoter sequences, are simultaneously amplified by *in vitro* transcription. The amplified
96 RNA molecules are then used to prepare Illumina libraries, as described previously²⁰ (Fig.
97 1a). Thus, this new method enables genome-wide quantification of protein-DNA interactions
98 and mRNA from the same cell without requiring physical separation steps, thereby
99 minimizing losses and making it easily adaptable to automated liquid handlers that can
100 process thousands of single-cells per day in a high-throughput format.

101 To determine the efficiency of the combined method, we benchmarked scDamID&T
102 to previous data in KBM7 cells; a clonal line for which single-cell genome-NL interaction
103 maps (scDamID) and single-cell transcriptomes are already available². We successfully
104 detected reads corresponding to both DamID and mRNA. We detected a median of 60,348
105 unique DamID reads per cell, identifying all major LADs, as previously reported from bulk
106 and single-cell sequencing². As illustrated for chromosome 17, observed over expected (OE)
107 scores² calculated based on the combined method not only detected all LADs but also
108 captured the cell-to-cell heterogeneity in genome-NL interactions as observed previously
109 (Fig. 1b and Supplementary Fig. 1a). This is further illustrated by the high concordance
110 (Pearson $r = 0.97$) in the contact frequencies (CFs), the percentage of cells, which at a given
111 position in the genome are in contact with the NL (Fig. 1c). Altogether this shows that
112 scDamID&T can successfully capture the dynamics of genome-NL interactions in single
113 cells. A crucial improvement in the scDamID&T method is that the cell- and nucleic acid-
114 specific barcoding enables single cells to be pooled prior to amplification and library
115 preparation, as opposed to the individual cell library preparation and sample selection in
116 scDamID. This significantly contributes to increased throughput and cost reduction. Although
117 single cells are pooled in scDamID&T prior to amplification without selection for cells with
118 the highest signal, the complexity of the single-cell libraries, quantified as the number of
119 unique reads per read sequenced in a cell, is comparable between both methods
120 (Supplementary Fig. 1b). Further, the loss of reads with incorrect adapter sequences is
121 substantially reduced in the new method (Supplementary Fig. 1c). The previously developed
122 scDamID is biased against detection of GATC sites that were separated by over 1 kb in the
123 genome; a drawback that is overcome by a single ligation event in scDamID&T which
124 captured the genome-wide distribution of GATC sites more faithfully (Fig. 1d and
125 Supplementary Fig. 1d).

126 Next, we benchmarked the transcriptomic measurements from scDamID&T to
127 previously obtained single-cell CEL-Seq data for KBM7 cells². Both methods detected the
128 expression of comparable number of genes (Median: CEL-Seq = 2509, scDamID&T = 2052)
129 (Fig. 1e), and the number of unique transcripts detected per cell was similar for both methods
130 (Median: CEL-Seq = 4920, scDamID&T = 3743) (Supplementary Fig. 2a). The efficiency of
131 mRNA detection appears to reduce with higher DamID double-stranded adapter
132 concentrations; we find that the quality of the transcriptome libraries can be further increased
133 by lowering the double-stranded adapter concentrations, without compromising the quality of
134 the DamID libraries (Fig. 1f and Supplementary Fig. 2b). Hierarchical clustering of the

135 single-cell transcriptomes showed that samples from both methods cluster together
136 (Supplementary Fig. 2c), emphasizing the concordance between the transcriptomes captured
137 by both techniques.

138 To verify scDamID&T in an independent cell line, we also established the system in
139 hybrid (129/Sv:Cast/EiJ) mouse embryonic stem (mES) cells²¹ where DamID expression is
140 controlled via the auxin-AID degron system²² (Supplementary Fig. 3a). The quality of the
141 scDamID&T libraries in mES cells expressing Dam or Dam-LmnB1 is comparable to KBM7
142 cells except that the single-cell Dam-LmnB1 data is of lower complexity (Supplementary Fig.
143 3b). The reduction in DamID complexity is likely a reflection of the shorter induction time of
144 Dam-LmnB1 in mES cells and difference in cell cycle characteristics. Nevertheless,
145 measurements with scDamID&T from these samples show strong DamID signals in
146 previously reported²³ bulk LAD domains (Supplementary Fig. 3c).

147 Extrapolating the technology that we developed for the detection of genome-NL
148 interactions and mRNA from the same cell, we hypothesized that KBM7 cells expressing
149 untethered Dam could be used to quantify both DNA accessibility and the transcriptome on a
150 genome-wide scale from single cells. To explore the possibility of using Dam as a DNA
151 accessibility marker, we first quantified the levels of Dam GATC methylation of averaged
152 single-cell profiles around transcription start sites (TSS) of actively transcribed genes and
153 observed a sharp peak at these sites (Fig. 2a). As a control, we also performed these single-
154 cell experiments using the non-methylation sensitive restriction enzyme AluI. We did not
155 observe signatures of accessibility around TSS of actively expressed genes (Fig. 2b),
156 indicating that the observed Dam accessibility patterns are the result of *in vivo* Dam
157 methylation at accessible regions of the genome, and not a consequence of restriction enzyme
158 accessibility. Similar to active TSSs, we also observe strong Dam enrichment at active
159 enhancers (Fig. 2c).

160 Nucleosomes are known to be regularly spaced on active TSS^{24,25} and CTCF sites,
161 and this can be observed in DNA accessibility data pooled across 96 single cells obtained
162 using scDamID&T (Fig. 2d and 2e and Supplementary Fig. 4a). The observed periodicity of
163 178bp is in general agreement with the reported spacing of nucleosomes in human cells²⁵
164 (Supplementary Fig. 4b). Remarkably, these nucleosome positioning profiles are also
165 apparent in data from single cells (Fig. 2f), indicating that Dam can serve to determine
166 nucleosome positioning *in vivo* in single cells. This feature could be especially powerful
167 when scDamID&T is combined with single-cell CRISPR/Cas9 to screen for factors involved
168 in nucleosome positioning²⁶. When comparing Dam-mediated DNA accessibility data to bulk

169 DNaseI-seq data, we find that the dynamic range of Dam-mediated DNA accessibility is
170 larger; for a substantial fraction of the genome only baseline levels of DNaseI are detected,
171 while Dam indicates intermediate levels of accessibility (Fig. 2g). Further analyses showed
172 that these regions are typified by genes with low expression, indicating that Dam is more
173 sensitive than DNaseI and allows discrimination between inactive and lowly transcribed
174 genes. This feature may be attributed to the advantage of Dam detecting both active
175 promoters (H3K4me3) and gene bodies (H3K36me3) (Supplementary Fig. 4c) and the *in vivo*
176 accumulation of Dam signal over time.

177 As scDamID&T enables simultaneous quantification of protein-DNA interactions and
178 mRNA from the same cell, we next investigated how variations in genome-NL association
179 directly influence gene expression. Further, as dissociation of genomic loci from the NL has
180 been shown to result in an increase in active histone modifications for some of those loci¹⁷,
181 we hypothesized that the propensity of a region in the genome to associate with the NL could
182 result in differentially regulated gene expression. To test this hypothesis, we first quantified
183 heterogeneity in genome-NL associations for each 500 kb region using CFs². While single-
184 cell samples generally show a large degree of concordance, certain regions are found in
185 contact with the NL in only a small fraction of cells (“low CF”). We found that gene
186 expression in that small fraction of cells that exhibit NL contact is generally lower compared
187 to cells that do not show NL contact (for example genomic region 839, Fig. 3a). In contrast,
188 for regions with intermediate CF (for example genomic region 317, Fig. 3a), gene expression
189 was independent of NL-positioning (Fig. 3a “middle CF”). Performing this analysis on a
190 genome-wide scale and stratifying bins by their CF values, we found a significant decrease of
191 gene expression upon NL association in regions with low CF values (Fig. 3b), whereas
192 genomic regions with CF values greater than 20% appear to be insensitive to NL association.
193 Interestingly, the impact on gene expression does not seem to vary with the (mean) gene
194 expression levels (Supplementary Fig. 5a). Taken together, these results suggest that the CF
195 of a region biases the sensitivity of gene expression to NL positioning. To our knowledge,
196 this is the first report to show that heterogeneity in spatial positioning of the genome directly
197 impacts gene expression in single cells. Finally, this differential sensitivity in transcriptional
198 output of genomic regions upon NL association may explain the varied outcomes of three
199 previous studies showing that artificial targeting of genomic regions to the NL resulted in
200 reduced, mixed or unchanged expression levels of the genes²⁷⁻²⁹.

201 Next, we applied this analysis to explore how variability in DNA accessibility relates
202 to heterogeneity in gene expression in KBM7 cells. We found that for regions that were in

203 contact with Dam in a large fraction of the cells (CF > 40%), expression was significantly
204 higher in cells showing Dam contact (Fig. 3c and Supplementary Fig. 5b). These results
205 suggest that gene expression heterogeneity between single cells is more sensitive to
206 variability in DNA accessibility within open chromatin regions. Consistent with the results of
207 KBM7 cells, we also observed the same relationship in the hybrid mES cells, suggesting that
208 the observed relationship between DNA accessibility and gene expression is generalizable to
209 other mammalian systems (Supplementary Figs. 5c and 5d)

210 To expand upon the analysis presented above, we investigated how DNA accessibility
211 tunes gene expression at an allelic resolution. For this, we used a hybrid mES cell line of
212 129/Sv:Cast/EiJ genotype²¹⁻³⁰ which is known to harbor a duplication of Cast/EiJ
213 chromosome 12. In order to carefully karyotype this cell line prior to application of
214 scDamID&T, we modified our technique to detect copy number variations in single cells, by
215 using the Dam-methylation insensitive restriction enzyme AluI instead of DpnI. This
216 demonstrates that scDamID&T can also be easily extended to quantify the genome and
217 transcriptome from the same cell, using minor modifications to the protocol presented
218 above^{13,14}. The AluI data showed that the hybrid mES cell line harbors a systematic
219 duplication of the Cast/EiJ chromosome 12 in most but not all single cells (Supplementary
220 Fig. 6a). When we performed scDamID&T using untethered Dam to measure single-cell
221 DNA accessibility profiles we also detected increased Dam contacts for the Cast/EiJ
222 chromosome 12, and a chromosome-wide mRNA bias towards Cast/EiJ transcripts
223 (Supplementary Fig. 6b and 6c). Surprisingly, we also detected a small fraction of cells that
224 displayed increased DNA accessibility for the 129/Sv allele over the Cast/EiJ allele for
225 chromosome 12, and a corresponding increase in 129/Sv derived transcripts for one cell
226 (Supplementary Figure 6c). After excluding the confounding effects of CNVs on
227 chromosome 12 as well as chromosomes 5 and 8 in this hybrid mES cell line, we observed a
228 significant positive correlation between allele-specific DNA accessibility and gene
229 expression (Fig. 3d). Taken together, these results demonstrate that scDamID&T can also be
230 used to directly quantify the allele-specific relationship between DNA accessibility and the
231 transcriptome (Supplementary Figs. 6a-c).

232 Finally, we sought to test scDamID&T as an *in silico* cell sorting strategy to
233 distinguish and group cell types based on the transcriptomes and thereafter, uncover the
234 underlying cell-type specific gene-regulatory landscape by DamID. Such a strategy to obtain
235 cell-type specific protein-DNA interaction maps is particularly attractive for complex tissues

236 and tumors with unknown cellular constitution, or for certain cell types that cannot be
237 isolated with sufficient purity due to a lack of discriminating surface markers or a lack of
238 high quality antibodies.

239 To demonstrate that our new technology can be used as an *in silico* cell sorting
240 technique that enables generation of cell-type specific DNA accessibility profiles, we
241 performed a proof-of-principle experiment where mES cells cultured under 2i or serum
242 conditions were sorted and quantified using scDamID&T. Single-cell transcriptomes obtained
243 using scDamID&T could be used to readily separate the population into two distinct clusters,
244 corresponding to 2i and serum grown cells (Fig. 4a). Expression analysis showed signature
245 genes differentially expressed between the two conditions (Supplementary Fig. 7a). DNA
246 accessibility profiles generated from the two *in silico* transcriptome clusters showed
247 differential accessibility patterns on a genome-wide scale. For example, DNA accessibility
248 tracks along Peg10, a gene strongly upregulated under serum conditions, showed increased
249 accessibility at the TSS and along the length of the gene (Fig. 4b). Interestingly, the increased
250 accessibility in the serum condition extends beyond the Peg10 gene locus, encompassing the
251 entire length of a large topologically associated domain (TAD). Indeed, the overall
252 expression of neighboring genes within this TAD is higher in serum conditions (Fig. 4b).
253 Generalizing this to all differentially expressed genes, we found that upregulation of gene
254 expression in 2i or serum conditions correlated with increased DNA accessibility over the
255 entire gene body (Figs. 4c and 4d and Supplementary Fig. 7b). Similarly, we observed that
256 differentially upregulated genes in each condition showed an increase in DNA accessibility at
257 the TSS for those genes (Fig. 4d). Thus, these results demonstrate that scDamID&T can be
258 used to effectively generate cell-type specific DNA accessibility profiles. Finally, we found
259 that upregulated gene expression also correlated with increased accessibility at the single-cell
260 level, highlighting that scDamID&T can be used to study changes in cellular identities in
261 direct relationship with the accompanying gene-regulatory mechanisms that shape cell type-
262 specific gene expression programs (Fig. 4e).

263 In summary, we have developed a new technology to simultaneously quantify
264 genome-NL interactions (Dam-LmnB1), DNA accessibility (Dam) or genome CNVs (AluI)
265 with the transcriptome from the same cell. scDamID&T enables dissection of the relationship
266 between the direct impact of spatial genome organization and chromatin accessibility on gene
267 expression. Further, it can be applied to sort cell types *in silico* and obtain their associated
268 gene-regulatory landscapes. Excitingly, in the future, scDamID&T can be employed to obtain

269 combined single-cell quantifications of many distinct nuclear regulatory mechanisms via the
270 coupling of Dam to transcription factors, various constituents of different chromatin types
271 (for example, Polycomb-group proteins and HP1) or the DNA replication or DNA damage
272 machineries³¹. Applied to dynamic biological processes, this technique should prove
273 especially powerful to dissect the order and sequence of epigenetic changes that are necessary
274 for the acquisition of different cell fates in heterogeneous tissues and differentiation systems.

275

276 **Acknowledgements**

277 We would like to thank the members of the JK and AvO labs for their comments on the
278 manuscript. We would also like to thank Mauro Muraro and Lennart Kester for valuable
279 input setting up this technique. This work was supported by an European Research Council
280 Starting grant (ERC-STG 678423-EpiID), Advanced grant (ERC-AdG 742225-IntScOmics)
281 and a Nederlandse organisatie voor Wetenschappelijk Onderwijs (NWO) open grant
282 (824.15.019) and TOP award (NWO-CW 714.016.001).

283

284 **Competing interests statement.** The authors declare that they have no competing financial
285 interests.

286

287 Correspondence and requests for materials should be addressed to S.S.D. (sdey@ucsb.edu) or
288 J.K. (j.kind@hubrecht.eu)

289

290 **Data availability.** The sequencing DamID data from this study are available from the Gene
291 Expression Omnibus, accession number GSE108639
292 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108639>). The data can be
293 accessed with the use of the token: ytsvcsiqhzoppux.

294

295 **References**

- 296 1 Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome
297 structure. *Nature* **502**, 59-64, doi:10.1038/nature12593 (2013).
- 298 2 Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human
299 cells. *Cell* **163**, 134-147, doi:10.1016/j.cell.2015.08.040 (2015).
- 300 3 Flyamer, I. M. *et al.* Single-nucleus Hi-C reveals unique chromatin reorganization at
301 oocyte-to-zygote transition. *Nature* **544**, 110-114, doi:10.1038/nature21711 (2017).
- 302 4 Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by
303 single-cell Hi-C. *Nature* **544**, 59-64, doi:10.1038/nature21429 (2017).

- 304 5 Cusanovich, D. A. *et al.* Multiplex single cell profiling of chromatin accessibility by
305 combinatorial cellular indexing. *Science (New York, N.Y.)* **348**, 910-914,
306 doi:10.1126/science.aab1601 (2015).
- 307 6 Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of
308 regulatory variation. *Nature* **523**, 486-490, doi:10.1038/nature14590 (2015).
- 309 7 Jin, W. *et al.* Genome-wide detection of DNase I hypersensitive sites in single cells
310 and FFPE tissue samples. *Nature* **528**, 142-146, doi:10.1038/nature15740 (2015).
- 311 8 Guo, H. *et al.* Single-cell methylome landscapes of mouse embryonic stem cells and
312 early embryos analyzed using reduced representation bisulfite sequencing. *Genome*
313 *research* **23**, 2126-2135, doi:10.1101/gr.161679.113 (2013).
- 314 9 Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing
315 epigenetic heterogeneity. *Nature methods* **11**, 817-820, doi:10.1038/nmeth.3035
316 (2014).
- 317 10 Farlik, M. *et al.* Single-cell DNA methylome sequencing and bioinformatic inference
318 of epigenomic cell-state dynamics. *Cell reports* **10**, 1386-1397,
319 doi:10.1016/j.celrep.2015.02.001 (2015).
- 320 11 Mooijman, D., Dey, S. S., Boisset, J. C., Crosetto, N. & van Oudenaarden, A. Single-
321 cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables
322 lineage reconstruction. *Nature biotechnology* **34**, 852-856, doi:10.1038/nbt.3598
323 (2016).
- 324 12 Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by
325 chromatin state. *Nature biotechnology* **33**, 1165-1172, doi:10.1038/nbt.3383 (2015).
- 326 13 Dey, S. S., Kester, L., Spanjaard, B., Bienko, M. & van Oudenaarden, A. Integrated
327 genome and transcriptome sequencing of the same cell. *Nature biotechnology* **33**,
328 285-289, doi:10.1038/nbt.3129 (2015).
- 329 14 Macaulay, I. C. *et al.* G&T-seq: parallel sequencing of single-cell genomes and
330 transcriptomes. *Nature methods* **12**, 519-522, doi:10.1038/nmeth.3370 (2015).
- 331 15 Angermueller, C. *et al.* Parallel single-cell sequencing links transcriptional and
332 epigenetic heterogeneity. *Nature methods* **13**, 229-232, doi:10.1038/nmeth.3728
333 (2016).
- 334 16 Clark, S. J. *et al.* scNMT-seq enables joint profiling of chromatin accessibility DNA
335 methylation and transcription in single cells. *Nature communications* **9**, 781,
336 doi:10.1038/s41467-018-03149-4 (2018).
- 337 17 Kind, J. *et al.* Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**,
338 178-192, doi:10.1016/j.cell.2013.02.028 (2013).
- 339 18 Sakaue-Sawano, A. *et al.* Visualizing spatiotemporal dynamics of multicellular cell-
340 cycle progression. *Cell* **132**, 487-498, doi:10.1016/j.cell.2007.12.033 (2008).
- 341 19 Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by
342 multiplexed linear amplification. *Cell reports* **2**, 666-673,
343 doi:10.1016/j.celrep.2012.08.003 (2012).
- 344 20 Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq.
345 *Genome biology* **17**, 77, doi:10.1186/s13059-016-0938-8 (2016).
- 346 21 Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F. & Gribnau, J. X inactivation
347 counting and choice is a stochastic process: evidence for involvement of an X-linked
348 activator. *Cell* **132**, 410-421, doi:10.1016/j.cell.2007.12.036 (2008).
- 349 22 Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An
350 auxin-based degron system for the rapid depletion of proteins in nonplant cells.
351 *Nature methods* **6**, 917-922, doi:10.1038/nmeth.1401 (2009).

- 352 23 Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear
353 lamina interactions during differentiation. *Molecular cell* **38**, 603-613,
354 doi:10.1016/j.molcel.2010.03.016 (2010).
- 355 24 Schones, D. E. *et al.* Dynamic regulation of nucleosome positioning in the human
356 genome. *Cell* **132**, 887-898, doi:10.1016/j.cell.2008.02.022 (2008).
- 357 25 Valouev, A. *et al.* Determinants of nucleosome organization in primary human cells.
358 *Nature* **474**, 516-520, doi:10.1038/nature10002 (2011).
- 359 26 Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout.
360 *Nature methods* **14**, 297-301, doi:10.1038/nmeth.4177 (2017).
- 361 27 Kumaran, R. I. & Spector, D. L. A genetic locus targeted to the nuclear periphery in
362 living cells maintains its transcriptional competence. *The Journal of cell biology* **180**,
363 51-65, doi:10.1083/jcb.200706060 (2008).
- 364 28 Reddy, K. L., Zullo, J. M., Bertolino, E. & Singh, H. Transcriptional repression
365 mediated by repositioning of genes to the nuclear lamina. *Nature* **452**, 243-247,
366 doi:10.1038/nature06727 (2008).
- 367 29 Finlan, L. E. *et al.* Recruitment to the nuclear periphery can alter expression of genes
368 in human cells. *PLoS genetics* **4**, e1000039, doi:10.1371/journal.pgen.1000039
369 (2008).
- 370 30 Loda, A. *et al.* Genetic and epigenetic features direct differential efficiency of Xist-
371 mediated silencing at X-chromosomal and autosomal locations. *Nature*
372 *communications* **8**, 690, doi:10.1038/s41467-017-00528-1 (2017).
- 373 31 Filion, G. J. *et al.* Systematic protein location mapping reveals five principal
374 chromatin types in Drosophila cells. *Cell* **143**, 212-224,
375 doi:10.1016/j.cell.2010.09.009 (2010).
- 376 32 Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development.
377 *Cell* **171**, 557-572.e524, doi:10.1016/j.cell.2017.09.043 (2017).

378

379 **Figure legends**

380 **Figure 1 | Quantitative comparison of scDamID, CEL-Seq and scDamID&T applied to** 381 **KBM7 cells**

382 **a)** Schematic representation of the scDamID&T method. **b)** Binary representation of OE
383 values of Dam-LmnB1 signal measured with scDamID&T and scDamID² in single cells on
384 chromosome 17. Unmappable regions are marked in grey. **c)** Comparison of CFs for
385 scDamID (y-axis) and scDamID&T (x-axis). CF distributions are depicted in the margins.
386 Pearson's $r = 0.97$. **d)** Distribution of inter-GATC distances of mappable GATC fragments
387 genome-wide (dotted line), and inter-GATC distances of GATCs observed with scDamID
388 (orange line) and scDamID&T (blue lines) for Dam-LmnB1. **e)** Distributions of the number
389 of unique genes detected using CEL-Seq² (green line) and scDamID&T (blue line). **f)**
390 Distribution of the number of unique transcripts detected by CEL-Seq data² (green line) and
391 scDamID&T (blue line) for Dam and Dam-LmnB1, and for different DamID adapter
392 concentrations.

393

394 **Figure 2 | Untethered Dam marks accessible chromatin in single cells**

395 **a)** Transcription start site (TSS) alignment of the single-cell average (n=96 cells) Dam signal
396 stratified by gene expression into four categories of expression levels (category 1 most active
397 or highly expressed; category 4 least active or not expressed). **b)** TSS alignment as for (a),
398 showing the density of AluI-derived genomic fragments. **c)** Alignment plot of the single-cell
399 average (n=96 cells) Dam signal at active enhancers. **d)** TSS alignment of the single-cell
400 average (n=96 cells) Dam signal for active genes at 10bp resolution for OE values (orange),
401 observed reads (brown) and density of mappable GATCs (black). The red arrow highlights an
402 example of periodicity in the DNA accessibility signal. **e)** Single-cell average (n=96 cells)
403 Dam signal alignment at CTCF sites, stratified in four regimes of increasing CTCF binding
404 activity (see computational methods for details on stratification). **f)** Example of Dam signal at
405 CTCF sites for a single cell with the highest CTCF binding activity. **g)** Scatter plot of bulk
406 DNaseI (y-axis) and single-cell average Dam data (x-axis). The left panel displays the density
407 of 20kb bins as a function of DNaseI (y-axis) and Dam (x-axis) signal. The middle panel
408 displays the density of 20kb bins with at least a single TSS. The right panel depicts the mean
409 expression for all genes in all 20kb regions for each point in the plot. Note that for baseline
410 DNaseI signal (red arrow), genes that are expressed at low levels display elevated Dam signal
411 (green arrow).

412

413 **Figure 3 | Parallel transcriptomic and DamID measurements link transcriptional**
414 **dependencies with heterogeneity in DamID contacts**

415 **a)** Examples of regions with low (left) and intermediate (right) CFs. The black filled boxes
416 indicate single-cell 500kb NL contacts (OE value > 1); white boxes indicate no NL contact
417 (OE value < 1). Boxplots in the right panels display gene expression levels in these bins,
418 stratified by NL contacts. For the low CF bin, note the increased expression levels in cells
419 with no NL contacts. Bin 839 corresponds to genomic region chr2:170000000-170500000.
420 Bin 317 corresponds to genomic region chr1:158500000-159000000. **b)** Top panel:
421 distribution of CF values across the genome for Dam-LmnB1 data in KBM7 cells. Red lines
422 indicate the segmentation of the genomic regions in low, intermediate and high CF bins.
423 Bottom panel: distributions of log2 fold-change (FC) in gene expression between cells
424 exhibiting contact vs. cells not exhibiting contact. *=p<0.05, two-sided t-test. **c)** Analysis as
425 in b, for untethered Dam in KBM7 cells. *=p<0.05, two-sided t-test. **d)** Scatter plot of the
426 measured mES cell allelic bias (129/Sv vs. Cast/EiJ) in transcription (y-axis) vs. the allelic

427 bias in chromatin accessibility (x-axis), measured in 100kb bins. Chromosomes 5, 8 and 12,
428 as well as the sex chromosomes were excluded from this analysis.

429

430 **Figure 4 | scDamID&T enables *in silico* cell sorting and reconstruction of corresponding**
431 **cell type specific gene regulatory landscapes.**

432 **a)** Principle component (left) and principal components-linear discriminant (right) analysis on
433 Dam expressing mES cells cultured in 2i (blue) or serum conditions (orange). **b)** DNA
434 accessibility profiles in 2i and serum conditions. Arrowheads indicate genes with log₂FC of
435 ≥ 1.25 in serum condition. Arrowheads with black outline were found to be significantly
436 differentially expressed (with FDR < 5%). The lower panel shows HiC data obtained from
437 mESCs³² displayed with the 3D genome browser {DOI:[10.1101/112268](https://doi.org/10.1101/112268)}. **c)** log₂ FC in
438 DNA accessibility between serum and 2i conditions for genes that are differentially up
439 (orange), down (blue) or unaffected (green) in serum conditions compared to 2i. **d)** DNA
440 accessibility at TSSs of differentially up- (top panel) or down-regulated (bottom panel) genes
441 in serum (orange line) conditions compared to 2i (blue line). **e)** DNA accessibility for the top
442 5 induced genes in serum compared to 2i condition in single cells (cells are represented by
443 dots).

444

445 **Supplementary Figure 1 | Quantitative comparison between scDamID and ScDamID&T**

446 **a)** Comparison between the binarized single cell (horizontal tracks) contact frequency maps
447 for scDamID (top panel 118 cells) and scDamID&T (bottom panel 93 cells) **b)** Comparison
448 of sample complexities with scDamID (orange) and scDamID&T (blue) depicted by unique
449 reads (y-axis) with increasing sequencing depth (x-axis) in single-cell samples. **c)** Overview
450 of losses during processing of raw sequencing data in scDamID (orange bars) and
451 scDamID&T (blue bars). The raw reads are first filtered on the correct adapter structure, then
452 aligned to the human genome, where reads not yielding a unique alignment are filtered out, as
453 well as reads not aligning immediately adjacent to GATCs. Finally, duplicate reads are
454 removed, on account of the haploid nature of the KBM7 cell-line. **d)** Distribution of inter-
455 GATC distances of mappable GATC fragments genome-wide (dotted line), and inter-GATC
456 distances of GATCs observed with scDamID (orange line) and scDamID&T (blue lines) for
457 Dam.

458

459 **Supplementary Figure 2 | Quantitative comparison between CEL-Seq and scDamID&T**

460 **a)** Distributions of the number of unique transcripts detected using CEL-Seq² (green line) and
461 scDamID&T (blue line). **b)** Overview of losses during processing of transcriptomic data
462 obtained with CEL-Seq (green bars) or scDamID&T (blue bars). The raw reads are aligned to
463 the human genome, reads that do not yield unique alignments are filtered, as well as reads
464 that do not match exons. Finally, duplicate reads are removed based on the UMIs. **c)**
465 Hierarchical clustering of the transcriptomes obtained with CEL-Seq (green) and
466 scDamID&T (blue).

467

468 **Supplementary Figure 3 | ScDamID&T in hybrid mES cells**

469 **a)** Auxin mediated control of AID-Dam and AID-Dam-LmnB1 cell lines. DamID PCR
470 products of cells 24- and 48hours after auxin washout (top panel). Time course and
471 quantitative PCR analysis of auxin induction for a locus within a LAD, 0-, 8-, 10-, 12- and 24
472 hours after auxin washout (bottom panel). Quantification of the ^{m6}A levels as described for
473 the DpnII assay¹⁷. **b)** Overview of losses during data processing as in Supplementary Figure
474 2a for the scDamID&T libraries obtained in mES cells. **c)** mES Dam-LmnB1 OE values
475 projected on the upstream (top panel) and downstream (bottom panel) of LAD-boundaries
476 defined previously²³.

477

478 **Supplementary Figure 4 | Untethered Dam enzyme marks accessible chromatin in single 479 cells**

480 **a)** TSS alignment of the single-cell average (n=96 cells) Dam signal for inactive genes at
481 10bp resolution for OE values (orange), observed reads (brown) and mappable GATCs
482 (black). **b)** 10bp resolution frequency spectrum of single-cell average (n=96 cells) Dam-
483 signal stratified in four regimes of increasing CTCF binding activities. Note the peak signal
484 for the CTCF sites with the highest binding activities corresponds to 178bp (red arrow). **c)**
485 Distribution of 20kb bins as function of bulk H3K4me3 (y-axis, left panel) or bulk
486 H3K36me3 (y-axis, right panel) and single-cell average Dam data (x-axis). Increasing grey-
487 level intensity represents increasing 20kb bin density.

488

489 **Supplementary Figure 5 | Single-cell associations between transcription and Dam or 490 Dam-LmnB1 contacts**

491 **a)** log₂ FCs in expression levels (y-axis) between Dam-LmnB1 contact (OE > 1) and no
492 contact (OE < 1) samples, measured in 500kb bins, versus log-scaled expression levels (x-
493 axis). Note that negative log₂ FCs indicate higher expression in the “no NL-contact” samples

494 compared to “NL-contact” samples. The dotted line indicates a locally-weighted regression
495 (“lowess”). **b)** log₂ FCs in expression levels (y-axis) calculated between contact and no
496 contact samples in KBM7 cells expressing untethered Dam, as in **a**. Note that positive log₂
497 FCs indicate higher expression in the “Dam contact” samples compared to the “no Dam
498 contact” samples. **c)** Violin plot for the log₂ FC expression levels between contact and no-
499 contact samples obtained with Dam-expressing hybrid mES cells, as **Fig. 3b** and **Fig. 3c**.
500 *= $p < 0.05$, two-sided t-test. **d)** Same as for **b**, but in Dam expressing hybrid mES cells.

501

502 **Supplementary Figure 6 | Allelic associations between single-cell transcription and Dam** 503 **contacts**

504 **a)** AluI signal obtained from 74 129/Sv:Cast/EiJ mES cells. Each row represents a single cell;
505 each column a 100kb bin along the genome. The checkered black box indicates the
506 duplication of the Cast/EiJ chromosome 12. The track below the plot shows allelic bias for
507 the maternal 129/Sv allele in purple and the paternal Cast/EiJ allele in green, as determined
508 using partial least squares regression. **b)** Plot as in A, showing DamID signals obtained from
509 67 129/Sv:Cast/EiJ mES cells. **c)** Allelic bias in transcription (y-axis) in relationship to the
510 allelic bias in Dam signal (x-axis) for chromosome 12. One single cell (named #12) exhibits
511 about 2-fold lower Dam signal and transcriptional output from the Cast/EiJ allele (right
512 panel), while exhibiting a 2-fold increase in Dam and transcriptional signals originating from
513 the 129/Sv allele (left panel).

514

515 **Supplementary Figure 7 | *In silico* sorting of cell identities and corresponding** 516 **regulatory landscapes with scDamID&T**

517 **a)** log₂-transformed expression values for the top five differentially up-regulated genes in 2i
518 (left) and serum (right) conditions. The horizontal line for Gpx2 in serum conditions indicates
519 no expression. **b)** Density plot of genes relating the log₂ FC in Dam accessibility (x-axis) to
520 log₂ FC in gene expression (y-axis), showing only genes that were found to be differentially
521 expressed between 2i and serum conditions (FDR < 5%).

522

523 **Supplementary table 1 | scDamID double-stranded adapters**

524 **Supplementary table 2 | CEL-Seq2 primers**

525 **Supplementary table 3 | Statistical details per figure**

526

527

528 **Methods**

529 Cell culture. Haploid KBM7 cells were cultured in suspension in IMDM (Gibco)
530 supplemented with 10% FBS and 1% Pen/Strep. The same Shield1-inducible Dam-LmnB1
531 and Dam-only stable clonal KBM7 cell lines were used as in ¹. Cells were split every 3 days.
532 F1 hybrid 129/Sv:Cast/Eij mouse embryonic stem cells (mESCs)² were cultured on primary
533 mouse embryonic fibroblasts (mEFs), in ES cell culture media; G-MEM (Gibco)
534 supplemented with 10% FBS, 1% Pen/Strep, 1x GlutaMAX (Gibco), 1x non-essential amino
535 acids (Gibco), 1x sodium pyruvate (Gibco), 143 μ M β -mercaptoethanol and 1:1000 hLIF (in-
536 house production). Cells were split every 3 days. Expression of constructs was suppressed by
537 addition of 0.5 μ M and indole-3-acetic acid (IAA; Sigma, I5148). 2i F1 hybrid
538 129/Sv:Cast/Eij mESCs cells were cultured for 2 weeks on primary mEFs in 2i ES cell
539 culture media; 48% DMEM/F12 (Gibco) and 48% Neurobasal (Gibco), supplemented with
540 1x N2 (Gibco), 1x B27 supplement (Gibco), 1x non-essential amino acids, 1% Pen/Strep, 143
541 μ M β -mercaptoethanol, 0.5% BSA, 1 μ M PD0325901 (Axon Medchem, 1408), 3 μ M
542 CHIR99021 (Axon Medchem, 1386) and 20 ng/mL hLIF (in-house production). Cells were
543 split every 3 days. Expression of constructs was suppressed by addition of 0.5 μ M IAA.

544 **Generating cell lines.** Stable clonal Dam and Dam-LmnB1 F1 hybrid mESC lines were
545 created by co-transfection of the EF1alpha-Tir1-neo and hPGK-AID-Dam-mLmnB1 or
546 hPGK-AID-Dam plasmids in a ratio of 1:5. Cells were trypsinized and 0.5×10^6 cells were
547 plated directly with Effectene transfection mixture (Qiagen, 301427) on 0.1% gelatin (in-
548 house production) in 60% BRL- conditioned medium. The transfection was according to the
549 kit protocol. Cells were selected for 10 days with 250 μ g/mL G418 and selection of the
550 clones was based on methylation levels, determined by DpnII-qPCR assays as previously
551 described ³ To reduce the background methylation levels in the presence of 1.0 mM IAA
552 (Sigma, I5148), we transduced the selected clones of both AID-Dam-LmnB1 and Dam-only
553 with extra hPGK-Tir1-puro followed by selection with 0.8 μ g/mL puromycin. Positive clones
554 were screened for IAA induction in the presence and absence of IAA by DpnII-qPCR assays
555 and DamID PCR products.

556 **DamID induction.** Expression of Dam-LmnB1 or Dam-only constructs was induced in the
557 KBM7 cells with 0.5 nM Shield1 (Glaxo laboratories, 02939) 15 hours prior to harvesting as
558 described previously ¹. Expression of Dam-LmnB1 or Dam-only constructs was induced in
559 the F1 mESCs by IAA washout 12 hours prior to harvesting. Based on the growth curve of
560 cells counted at time points 0, 12, 24, 30, 36, 42, 48, 54, 60, 72 and 84 after plating, the
561 generation time of both the Dam-LmnB1 and Dam-only cell lines was estimated at ~12 hours

562 (data not shown). Considering that 55% of the cells are in G1 and early S, the estimated time
563 these cells reside in G1 and early S is 6,75 hours.

564 **Cell harvesting and sorting.** KBM7 cells were harvested in PBS (in-house production),
565 stained with 0.5 µg/mL DAPI for live/dead selection. Small haploid Single cells were sorted
566 based on forward and side-scatter properties (30% of total population) and selected for
567 double positive FUCCI profile as described before¹. F1 mES cells were collected in plain or
568 2i ES cell culture media, stained with 30 µg/mL Hoechst 34580 for 45 minutes at 37°C. mES
569 cell singlets were sorted based on forward and side-scatter properties, and in mid-S phase of
570 the cell cycle based on DNA content histogram. One cell per well was sorted into 384-well
571 plates (Biorad, HSP3801) using the BD FACSJazz cell sorter. Wells contained 4 µL mineral
572 oil (Sigma) and 100 nL of 15 ng/µL unique CELseq primer.

573 **scDamID&T.** Robotic preparation: 4 µL mineral oil was dispensed manually into each well
574 of a 384-well plate using a multichannel pipet. 100 nL of unique CEL-seq primer was
575 dispensed per well using the mosquito HTS robot (TTP Labtech). The NanodropII robot
576 (BioNex) was used for all subsequent dispensing steps at 12 p.s.i. pressure. After sorting, 100
577 nL lysis mix was added (0.8 U RNase inhibitor (Clontech, 2313A), 0.07% Igepal, 1mM
578 dNTPs, 1:500000 ERCC RNA spike-in mix (Ambion, 4456740)). Each single cell was lysed
579 at 65°C for 5 min and 150 nL reverse transcription mix was added (1x First Strand Buffer
580 (Invitrogen, 18064-014), 10 mM DTT (Invitrogen, 18064-014), 2 U RNaseOUT
581 Recombinant Ribonuclease Inhibitor (Invitrogen, 10777019), 10 U SuperscriptII (Invitrogen,
582 18064014)) and the plate was incubated at 42°C for 1 h, 4°C for 5 min and 70°C for 10 min.
583 Next, 1.92 µL of second strand synthesis mix was added (1x second strand buffer (Invitrogen,
584 10812014), 192 µM dNTPs, 0.006 U *E. coli* DNA ligase (Invitrogen, 18052019), 0.013 U
585 RNaseH (Invitrogen, 18021071)) and the plate was incubated at 16°C for 2 h. 500 nL of
586 protease mix was added (1x NEB CutSmart buffer, 1.21 mg/mL ProteinaseK (Roche,
587 000000003115836001)) and the plate was incubated at 50°C for 10 hr and 80°C for 20 min.
588 Next, 230 nL DpnI mix was added (1x NEB CutSmart buffer, 0.2 U NEB DpnI) and the plate
589 was incubated at 37°C for 4 hr and 80°C for 20 min. Finally, 50 nL of DamID2 adapters were
590 dispensed (final concentrations varied between 2 and 128 nM), together 450 nL of ligation
591 mix (1x T4 Ligase buffer (Roche, 10799009001), 0.14 U T4 Ligase (Roche, 10799009001))
592 and the plate was incubated at 16°C for 12 hr and 65°C for 10 min. Contents of all wells with
593 different primers and adapters was pooled and incubated with 0.8x 1:4 diluted magnetic
594 beads (CleanNA, CPCR-0050) for 10 min, washed twice with 80% ethanol and resuspended

595 in 7 μ L nuclease-free water before *in vitro* transcription at 37°C for 14 hr using the
596 MEGAScript T7 kit (Invitrogen, AM1334). Library preparation was done as described in
597 the CEL-seq protocol with minor adjustments⁴. Amplified RNA (aRNA) was cleaned and
598 size-selected by incubating with 0.8x magnetic beads (CleanNA, CPCr-0050) for 10 min,
599 washed twice with 80% ethanol and resuspended in 22 μ L nuclease-free water, and
600 fragmented at 94°C for 2 min in 0.2x fragmentation buffer (200 mM Tris-acetate, pH 8.1, 500
601 mM KOAc, 150 mM MgOAc). Fragmentation was stopped by addition of 0.1x fragmentation
602 STOP buffer (0.5 M EDTA pH8) and quenched on ice. Fragmented aRNA was incubated
603 with 0.8x magnetic beads (CleanNA, CPCr-0050) for 10 min, washed twice with 80%
604 ethanol and resuspended in 12 μ L nuclease-free water. Thereafter, library preparation was
605 done as previously described⁴ using 5 μ L of aRNA and PCR cycles varied between 8 and 10.
606 Libraries were run on the Illumina NextSeq platform with high output 75bp paired-end
607 sequencing.

608 **DamID adapters.** The adapter was designed (5' to 3') with a 4 nt fork, a T7 promoter, the 5'
609 Illumina adapter (as used in the Illumina small RNA kit), a 3 nt UMI (unique molecular
610 identifier), a 6 nt unique barcode and half a NlaIII digestion site (CA) such that NlaIII cutting
611 site is reconstituted upon self-ligation of adapters (CATG). The barcodes were designed with
612 a hamming distance of two. Bottom sequences contained a phosphorylation site at the 5' end.
613 Adapters were produced as standard desalted primers. Top and bottom sequences were
614 annealed at a 1:1 ratio in annealing buffer (10 mM Tris pH 7.5–8.0, 50 mM NaCl, 1 mM
615 EDTA) by immersing tubes in boiling water, then let to cool to room temperature. The oligo
616 sequences can be found in [Supplementary Table 1](#).

617 **CEL-seq primers.** The RT primer was designed according to the Yanai protocol⁴ with an
618 anchored polyT, a 8nt unique barcode, a 6nt UMI (unique molecular identifier), the 5'
619 Illumina adapter (as used in the Illumina small RNA kit) and a T7 promoter. The barcodes
620 were designed such that each pair is different by at least two nucleotides, so that a single
621 sequencing error will not produce the wrong barcode. Primers are desalted at the lowest
622 possible scale, stock solution 1 μ g/ μ L. The oligo sequences can be found in [Supplementary](#)
623 [Table 2](#).

624 **Raw data preprocessing.** First mates in the raw read pairs (i.e. "R1" or "read1") conform to
625 a layout of either:

626 5'-[3 nt UMI][8 nt barcode]CA[gDNA]-3'

627 in the case of gDNA (DamID and AluI restriction) reads, or

628 5'-[6 nt UMI][8 nt barcode][unalignable sequence]-3'

629 in the case of transcriptomic reads.

630 In the case of transcriptomic reads, the second mate in the read pair contains mRNA
631 sequence.

632 Raw reads were processed by demultiplexing on barcodes (simultaneously using the DamID
633 and transcriptomic barcodes), allowing no mismatches. The UMI sequences were extracted
634 and stored alongside the names of the reads for downstream processing.

635 **Sequence alignments.** After demultiplexing of the read pairs using the first mate and
636 removal of the UMI and barcode sequences, the reads were aligned. In the case of gDNA-
637 derived reads, a 'GA' dinucleotide was prepended to the sequences of read1 ('AG' in the case
638 of AluI), and read1 was then aligned to a reference genome using bowtie2 (v.2.3.2) using
639 parameters --seed 42 --very-sensitive -N 1. For transcriptome-derived reads, read2 was
640 aligned using tophat2 (v2.1.1) using parameters --segment-length 22 --read-mismatches 4 --
641 read-edit-dist 4 --min-anchor 6 --min-intron-length 25 --max-intron-length 25000 --no-novel-
642 juncs --no-novel-indels --no-coverage-search --b2-very-sensitive --b2-N 1 --b2-gbar 200 and
643 using transcriptome-guiding (options --GTF and --transcriptome-index). Human data was
644 aligned to hg19 (GRCh37) including the mitochondrial genome, the sex chromosomes and
645 unassembled contigs. Transcriptomic reads were aligned by making additional use of
646 transcript coordinates obtained from GENCODE (v26)
647 https://www.gencodegenes.org/releases/grch37_mapped_releases.html supplemented with
648 ERCC mRNA spike-in sequences [https://assets.thermofisher.com/TFS-](https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_095047.txt)
649 [Assets/LSG/manuals/cms_095047.txt](https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_095047.txt). mESC data was aligned to reference genomes
650 generated by imputing 129S1/SvImJ and CAST/EiJ SNPs obtained from the Sanger Mouse
651 Genomes project [<http://www.sanger.ac.uk/science/data/mouse-genomes-project>]⁵, onto the
652 mm10 reference genome. The mitochondrial genome, sex chromosome and unassembled
653 contigs were used in the alignments. Transcriptomic reads were aligned using a GTF file with
654 transcript annotations obtained from ENSEMBL (release 89)
655 [ftp://ftp.ensembl.org/pub/release-89/gtf/mus_musculus/Mus_musculus.GRCm38.89.gtf.gz].
656 Both human and mouse references were supplemented with ERCC mRNA spike-in
657 sequences [https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_095047.txt]. For
658 both genomic and transcriptomic data, reads that yielded an alignment with mapping quality
659 (BAM field 'MAPQ') lower than 10 were discarded. For the genomic data, reads not aligning
660 exactly at the expected position (5' of the motif, either GATC in the case of DpnI restriction,
661 or AGCT in the case of AluI restriction) were discarded. For the transcriptomic data, reads
662 not aligning to an exon of a single gene (unambiguously) were discarded. The mESC reads

663 were assigned to the 129S1/SvImJ or CAST/EiJ genotype by aligning reads to both
664 references. Reads that align with lower edit-distance (SAM tag 'NM') or higher alignment-
665 score (SAM tag 'AS') in case of equal edit-distance to one of the genotypes were assigned to
666 that genotypes. Reads that aligned with equal scores to both genotypes were considered of
667 'ambiguous' genotype.

668 **PCR duplicate filtering.** For the genomic data (DamID and AluI-WGS), the number of
669 reads per motif, strand and UMI were counted. Read counts were collapsed using the UMIs
670 (i.e. multiple reads with the same UMI count as 1) after an iterative filtering step where the
671 most abundant UMI causes every other UMI sequence with a Hamming-distance of 1 to be
672 filtered out. E.g, observing the three UMIs 'AAA', 'GCG' and 'AAT' in decreasing order
673 would count as 2 unique events (with UMIs 'AAA' and 'GCG', since 'AAT' is within 1
674 Hamming distance from 'AAA'). For the data from KBM7 (a near-complete haploid cell line)
675 at most 1 unique event per motif and strand was kept. For the mESC data at most 1 unique
676 event per motif, strand and genotype was kept, or 2 unique events, if the genotype of the
677 reads at that position could not be resolved.

678 **Filtering of samples.** Only single-cell samples with at least $10^{3.7}$ unique DamID events or at
679 least 10^3 unique transcripts were taken into consideration for the analyses. These cutoffs
680 were applied jointly for analyses where both genomic and transcriptomic signals were used.

681 **Binning and calculation of OE values.** DamID and WGS data was binned using non-
682 overlapping bins. Binsizes were 100kbp for untethered Dam and 500kbp for Dam-LmnB1
683 DamID data, 100kbp for WGS data and 500kbp for all hybrid mESC data where genotype-
684 specific counts were used. For analyses at TSS and CTCF sites, binsizes were 10bp. In order
685 to calculate observed-over-expected (OE) values, the mappability of each motif (GATC or
686 AGCT) was determined by generating 65 nt. long sequences (in both orientations) from the
687 reference genome(s) and aligning and processing them identically to the data. By binning the
688 *in-silico* generated reads, the maximum amount of mappable unique events per bin was
689 determined. OE values were calculated using

690
$$OE = \frac{O + \psi}{E + \psi} \cdot \frac{T_E + B \cdot \psi}{T_O + B \cdot \psi}$$

691 where O is the number of observed unique methylation events per bin, E is the number of
692 mappable unique events per bin, ψ is the pseudocount (1, unless otherwise stated), T_O and T_E
693 are the total number of unique methylation events observed cq. mappable in the sample and B
694 is the number of bins. For analysis across multiple windows, e.g. windows around TSSs or

695 CTCF sites, *O* and *E* are summed across the windows, prior to calculation of the OE values.
696 For the definition of "contact", regions with OE values ≥ 1 were considered as "in-contact".
697 For further details and justification, see Kind et al., 2015¹ and FigS2A in particular.

698 **H3K4me3, H3K36me3 and DNase data (external datasets).** H3K4me3, H3K36me3 and
699 DNase data was obtained from ENCODE (GSM788087, GSM733714 and
700 GSE90334_ENCFF038VUM, respectively) as processed bigWig files. In order to calculate
701 OE values for these datasets, whole-genome mappability as determined by the ENCODE
702 project was used (wgEncodeCrgMapabilityAlign36mer).

703 **Independent transcription dataset.** For Fig2G independent expression data was used from
704 GSE56465. (only KBM7 haploid samples).

705 **Untethered Dam enrichment at TSSs and CTCF sites.** For the analyses at TSSs, one
706 isoform per gene was chosen from the gene annotations, by taking preferentially isoforms
707 that carry the GENCODE "basic" tag, have a valid, annotated CDS (start and stop codon, and
708 CDS length that is a multiple of 3nt.), and ties are broken by the isoform with longest CDS,
709 and shortest gene length (distance from first to last exon). As TSS, the most 5' position of the
710 first exon was taken. CTCF sites were obtained by integrating ENCODE ChIPseq data
711 (wgEncodeRegTfbsCellsV3, K562 CTCF ChIPseq tracks from GSE30263) with CTCF motif
712 sites (factorbookMotifPos obtained via the UCSC genome browser⁶). Only CTCF ChIPseq
713 peaks that contained a CTCF binding motif with score of at least 1.0 within 500nt. of the
714 center of the ChIPseq peak were considered. The ChIPseq peaks were subdivided by ChIPseq
715 binding score, and the group of peaks with maximum score (of 1000) was subdivided into
716 two groups by the motif score, such that 4 approximately equal-sized groups of CTCF-bound
717 loci were obtained.

718 **logFC between contact/no contact groups of samples.** logFCs between single-cell samples
719 that showed contact and those that show no contact (see Fig3A) was performed as follows:

720 In bins across the genome (500kb. for Dam-LmnB1, 100kb. for untethered Dam) the logFC
721 in expression was calculated between samples that have a DamID OE value ≥ 1 vs. samples
722 that have a DamID OE value lower than 1, for every bin that has (1) at least $10^{1.9}$ mappable
723 GATCs per 100kb and (2) contains at least 3 single-cell samples per group and (3) has a
724 mean transcriptional level of at least 10 RPM across all single-cell samples. Comparison
725 scDamID&T to Kind Cell 2015 data. For the comparisons with individual measurements of
726 single-cell DamID and single-cell transcriptomics (CELseq) with scID&T in Fig1 the
727 scID&T data was made comparable to the published data by (1) truncating the reads at the 3'

728 end such that after barcode (and in the case of scDamID adapters) removal the same number
729 of nt. of gDNA is remaining. Furthermore, UMIs were completely left out of the
730 consideration for the DamID measurements, and for the transcriptional measurements, the
731 UMIs were truncated to 4nt. to make the data comparable to the published CELseq data. The
732 data were obtained from GSE69423.

733

734 By figure details on the statistics can be found in [Supplementary Table 3](#). All computational
735 codes used for this study are available upon request.

736

737 **Methods Only References**

- 738 1 Kind, J. et al. Genome-wide maps of nuclear lamina interactions in single human
739 cells. *Cell* **163**, 134-147, doi:10.1016/j.cell.2015.08.040 (2015).
- 740 2 Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F. & Gribnau, J. X inactivation
741 counting and choice is a stochastic process: evidence for involvement of an X-linked
742 activator. *Cell* **132**, 410-421, doi:10.1016/j.cell.2007.12.036 (2008).
- 743 3 Kind, J. et al. Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**,
744 178-192, doi:10.1016/j.cell.2013.02.028 (2013).
- 745 4 Hashimshony, T. et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq.
746 *Genome Biol* **17**, 77, doi:10.1016/j.celrep.2012.08.003 (2016).
- 747 5 Keane, T.M. *et al.* Mouse genomic variation and its effect on phenotypes and gene
748 regulation. *Nature* **477**, 289-94, doi: 10.1038/nature10413 (2011).
- 749 6 Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Research* **12**, 996-
750 1006, doi: 10.1101/gr.229102 (2002).

751

752

753

754

Figure 1

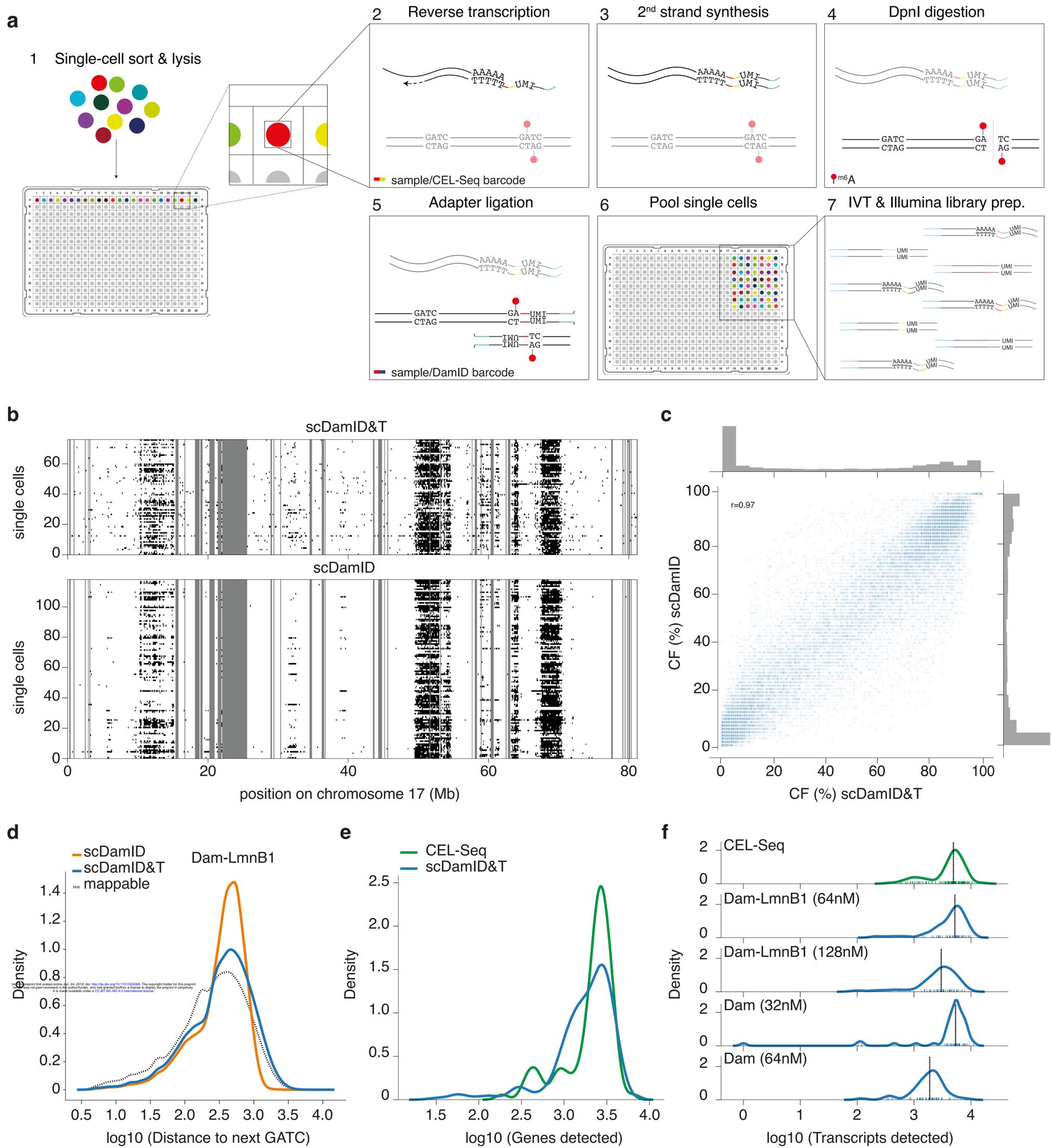


Figure 2

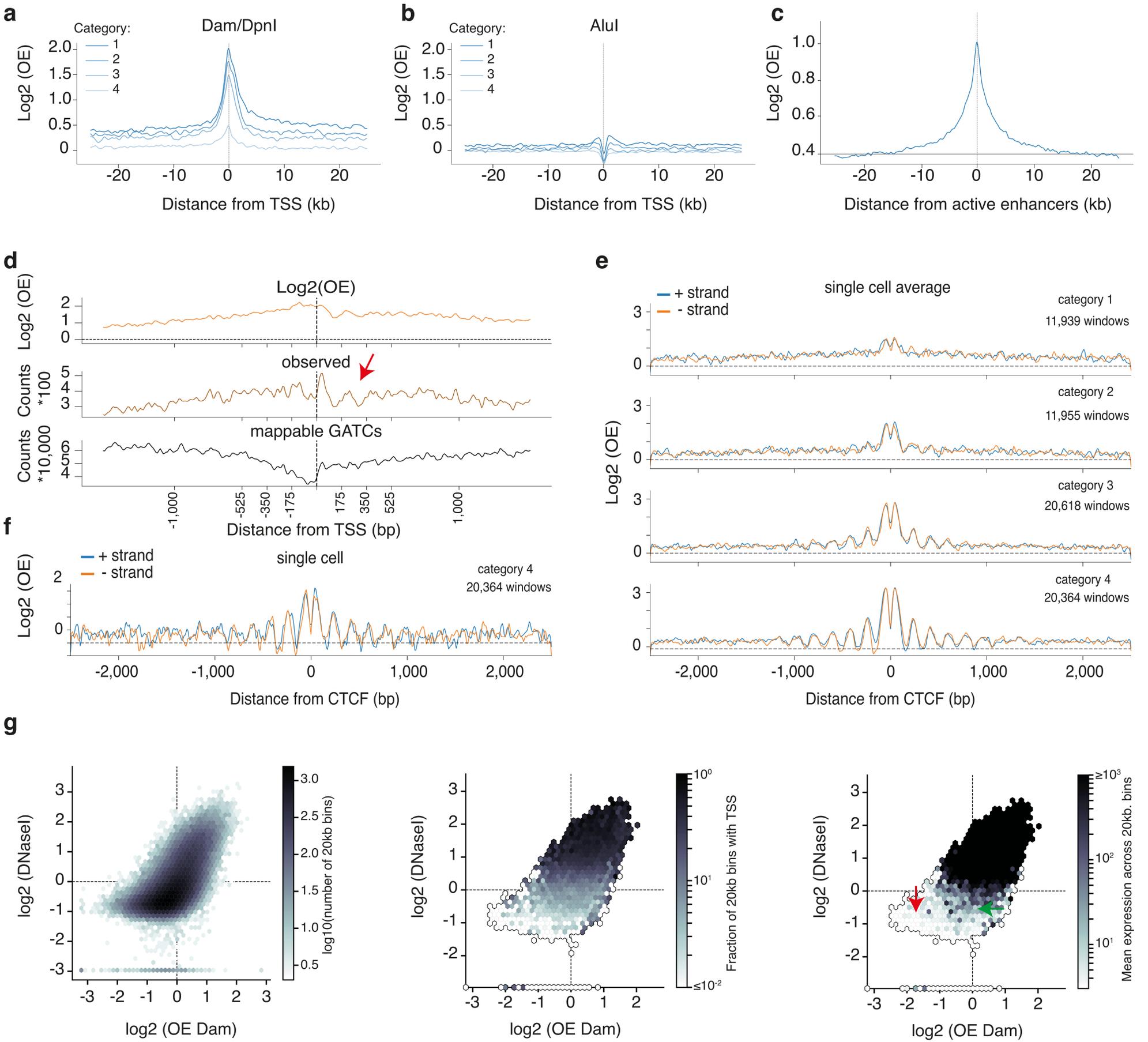


Figure 3

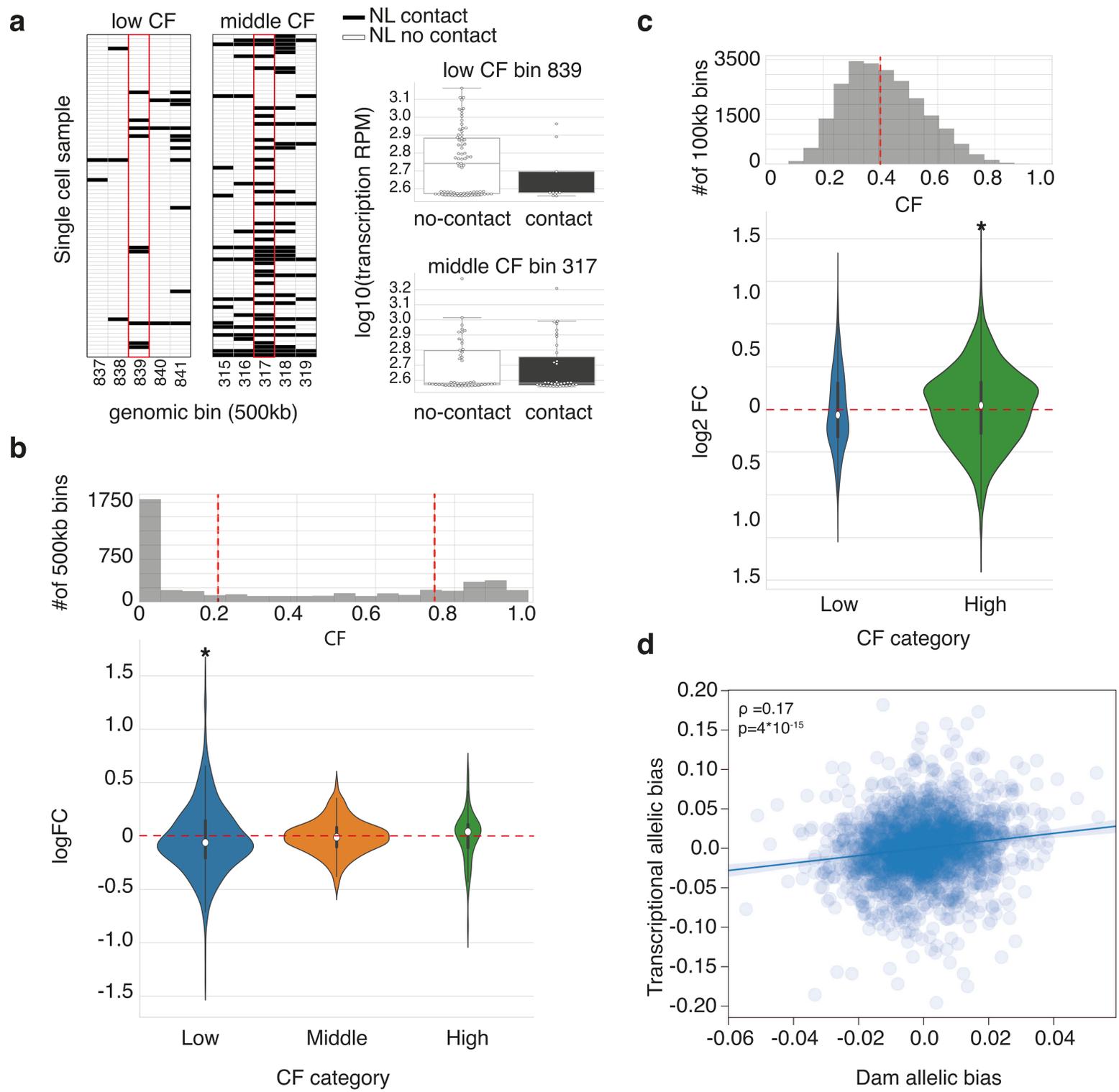


Figure 4

