Single-molecule simultaneous profiling of DNA methylation and DNA-protein interactions with Nanopore-DamID

Seth W. Cheetham[1,4*], Yohaann M. A. Jafrani[1,4], Stacey B. Andersen[2], Natasha Jansz[1], Adam D. Ewing[1] and Geoffrey J. Faulkner[1,3*]

1. Mater Research Institute – University of Queensland, TRI Building, Woolloongabba QLD 4102, Australia.

2. Genome Innovation Hub, The University of Queensland, Brisbane, QLD 4072, Australia

3. Queensland Brain Institute, University of Queensland, Brisbane QLD 4072, Australia.

4. Equal contribution

*Correspondence

Abstract

DNA-protein interactions and cytosine methylation control eukaryotic gene expression. Here, we present an approach to simultaneously detect cytosine methylation and DNA-protein interactions from single molecules, through selective sequencing of adenine-labelled DNA. Applying this approach to LaminB1-associated heterochromatin domains, we identify strict CpG methylation maintenance at transcriptional start sites amidst a generalised relaxation of methylation, potentially to prevent ectopic and aberrant heterochromatic gene expression.

Main

Cytosine methylation modulates the interactions of chromatin-associated proteins with DNA[1]. DNA adenine methylase identification (DamID) is an approach to profile DNA-protein interactions[2,3]. In DamID, the *E. coli* adenine methylase Dam is fused to a chromatin-associated protein (**Fig. 1a**). The Dam-fusion protein methylates adenines in GATC motifs proximal to the site of chromatin-association. Methylated GATC sites are then cleaved by the methylation-specific restriction enzyme DpnI, followed by adaptor ligation, amplification and Illumina-sequencing[4]. DamID has been adapted to detect chromatin-accessibility[5], RNA-chromatin interactions[6],

chromatin topology[7,8], nuclear lamina interactions[9], and has been successfully implemented in various model organisms[2,10–14].

Despite the utility of DamID, the approach has drawbacks. Each molecule must have GATC methylation on both ends to be amplified, reducing signal and rendering DamID incompatible with bisulphite sequencing. The protocol is labour-intensive (~five days[4]) and the extensive use of PCR has the potential to introduce biases.

Oxford Nanopore Technologies (ONT) long-read sequencing can directly detect both cytosine and adenine methylation, including in regions refractory to short-read analysis[15–17]. Exogenous promiscuous adenine methylases have recently been used to profile chromatin-accessibility[18] and, by fusion to protein-A, regions of transcription factor occupancy bound by a cognate antibody[19,20]. However, direct detection of adenine methylation requires high sequencing depth, which is prohibitively expensive in mammals. Here, we develop a Nanopore-based DamID which enables rapid and cost-effective profiling of single-molecular transcription factor-cytosine methylation interactions.

Nanopore sequencing can be used to selectively analyse adaptor-ligated DNA without purification or amplification[21]. Thus, if adaptors are only ligated to DNA cleaved by the methylated-GATC-specific enzyme DpnI these fragments can be selectively sequenced from a heterogenous mixture of largely unligated fragments (**Fig. 1a**). To enable selective sequencing of Dam-methylated fragments, we stably transduced the mouse neuroblastoma cell line, N2A, with a doxycycline-inducible Dam-LaminB1 fusion protein expression cassette (**Supplementary figure 1a)** and constitutively expressed m6A-tracer[22] (GFP fused to catalytically inactive DpnI, **Supplementary figure 1b**). Characteristic fluorescent rings at nuclear lamina were evident in cells upon doxycycline treatment, confirming that Dam-LaminB1 can efficiently modify GATC sites in lamin-associated chromatin (**Fig. 1b**). To simultaneously detect DNA-protein interactions and CpG methylation on single molecules, we expressed Dam-LaminB1 in N2A cells and extracted genomic DNA. We dephosphorylated double-strand breaks and incubated with DpnI to cleave methylated

GATC fragments. Blunt ends produced by DpnI were A-tailed and ligated to Nanopore adapters. Library preparation here takes <1.5 day, compared to 4-5 days for Illumina DamID[4]. From sequencing on a single MinION flow cell, we generated 12.8 million reads, enriched in the broad domains characteristic of LaminB1[9,23] (**Fig. 1c**). The nanopore reads are expected to terminate at GATC motifs (**Fig. 1c**). Indeed, *de novo* motif enrichment analysis on read termini identified profound enrichment of GATC (p=3.3e-703, **Fig. 1d**). At least one end in 97% of reads terminated within 10bp of a GATC motif (**Fig. 1e**) and were thus derived from a DpnI-cleaved Dam-methylated site. These results provide proof-of-principle confirmation that selective sequencing of adenine methylated DNA, without amplification or purification, is rapid, efficient and cost-effective.
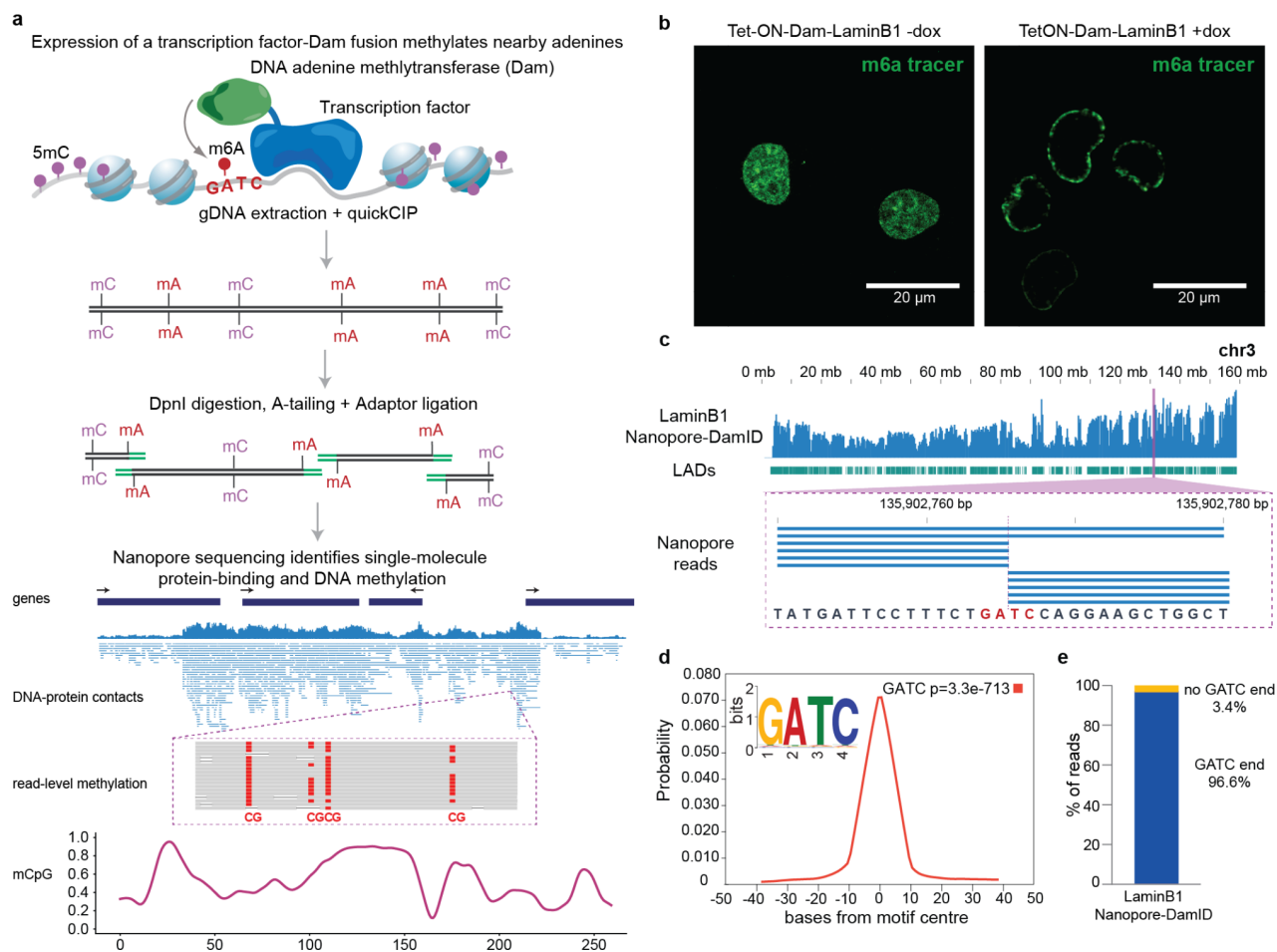

***Figure 1: Nanopore-DamID identifies Lamin-associated domains. a**, In Nanopore-DamID, a fusion Dam-chromatin-associated protein is expressed in cells and methylates adenines in GATC motifs proximal to sites of DNA-protein interaction. Genomic DNA is isolated from cells, dephosphorylated and digested with DpnI. Digested DNA ends are A-tailed and then Nanopore sequencing adaptors are ligated. Nanopore sequencing identifies DNA-protein*

*contacts and CpG methylation on single DNA molecules. **b,** Upon doxycycline treatment, the m6A-tracer (an inactive DpnI fragment fused to GFP) localises to the nuclear lamina in cells transfected with TetON-Dam-LaminB1. **c,** Sequencing of LaminB1-Dam-expressing cells shows enrichment in domains identified by RSEG[24] characteristic of LaminB1-DamID. Raw nanopore reads terminate at GATC sites . **d,** De novo motif discovery at the ends of Nanopore reads using MEME-ChIP[25] revealed the motif GATC was strongly enriched. **e,** Most reads from Nanopore-DamID terminate at a GATC motif (±10 bp).*

To compare Nanopore and Illumina-based approaches we sequenced conventional LaminB1-DamID libraries from N2A cells. LaminB1-occupancy profiles were comparable between technologies (**Fig. 2a**). Lamin-associated domains (LADs) identified by RSEG[24] largely overlap. 79% of bases in Illumina LADs fall within Nanopore-DamID LADs (**Supplementary tables 1 and 2**). Nanopore-detected domains were significantly broader and more contiguous than those detected by Illumina (**Fig. 2b**). The higher length of Nanopore reads (median length 996bp vs 75bp for Illumina) enables alignment of >4.5x more bases to regions of low mappability (**Supplementary figure 2**).
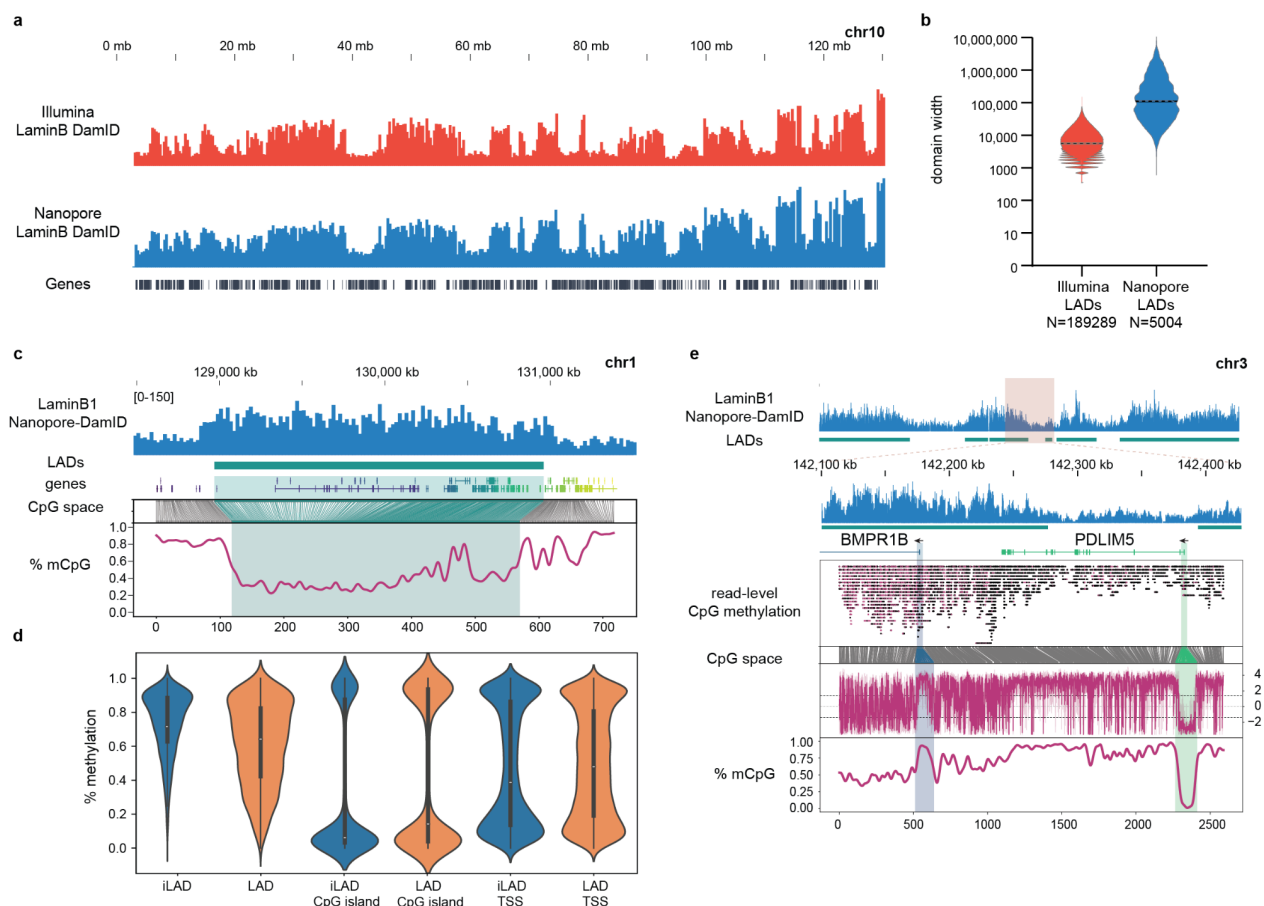
*Figure 2: Nanopore-DamID identifies widespread CpG hypomethylation in lamin-associated DNA. **a**, Comparison of LaminB1-DamID signal from Illumina and Nanopore-based approaches on chromosome 10. **b,** Illumina-DamID identifies more numerous, smaller Lamin-associated domains (LADs), compared to the large domains identified by Nanopore-DamID. **c**, Analysis of CpG-methylation from the Nanopore-DamID data with Methylartist[26] reveals pronounced hypomethylation within a LAD on chromosome 1. **d**, Genome-wide analysis of methylation within LADs or between LADs (iLADs) reveals hypomethylation of LADs, other than at TSSs, which maintain higher methylation. **e**, Example of the interaction of LaminB1 occupancy with DNA-methylation on chromosome 3.*

To determine the interactions between LaminB1 and DNA methylation we identified CpG methylation from the LaminB1-Nanopore-DamID reads. Consistent with previous reports[27,28] we found that, despite the heterochromatic structure of LADs, CpG methylation was globally depleted in these regions compared to non-Lamin-associated DNA, (inter-LADs, iLADs) (**Fig. 2c-d, Supplementary figure 3**). However, CpG islands and, most strikingly, the regions surrounding (± 1 kb) transcriptional start sites are resistant to LAD hypomethylation (**Fig 2c-d**). The maintenance of high methylation at LAD TSSs suggests that despite tethering to the nuclear periphery, DNA methylation at the proximal promoter region may prevent aberrant heterochromatic gene expression in LADs.

Overall, Nanopore-DamID efficiently selectively sequences Dam-labelled native DNA enabling co-detection of DNA-protein contacts and DNA methylation. This selectively sequencing approach can be extended to other adenine methylase-based approaches[18–20] to increase sequencing depth over methylated regions. Nanopore-DamID is less labour-intensive, enables superior coverage of regions with low mappability, and identifies more contiguous binding domains than short-read approaches. Nanopore-DamID facilitates elucidation of the interactions between CpG methylation and chromatin-associated proteins that control gene regulation.

Methods

Transgenic cell line preparation

$1x10^5$ N2A (ATCC) cells were seeded in a well of a 6-well plate in 2ml of DMEM complete media (Gibco) supplemented with 10% Fetal Bovine Serum, 1% PenStrep and 1% L-glutamine and

maintained at 37°C with 5% CO_2 in a humidified incubator. 0.2μg Super piggyBac Transposase Expression Vector, 0.25μg m6A-tracer and 0.25μg XLONE-Dam-LamB1 (BlastR) were diluted in 100μl of Opti-MEM (Thermo Fisher). 2.1μl of FuGENE-HD® (Promega) was added and the reaction was briefly vortexed and incubated at room temperature for 10min prior to addition to the cells. The transfection mix was added dropwise to the well and incubated for 24hr. After 24hr the media was changed and the cells were selected for seven days in media containing 2.5μg/ml blasticidin.

Imaging

1×10^5 transgenic N2A cells were seeded in two 35mm glass bottom dishes (Cellvis™). Dam-LaminB1 fusion protein expression was induced with 2μg/ml doxycycline in one of the dishes. The m6A tracer in cells of both dishes were visualised and imaged at 488nm on the Olympus FV3000 microscope at 120X magnification with a 60X oil immersion objective followed by a 2X digital zoom. Images were analysed via Fiji[29].

Nanopore DamID

Dam-LaminB1 fusion protein expression was induced with 2μg/ml doxycycline and genomic DNA (gDNA) was extracted from Nunc™ T75 using the NEB Monarch® gDNA extraction kit according to manufacturer's instructions with the exception of using inversion rather than vortexing for mixing. 2.5μg of gDNA was dephosphorylated by the addition of 5μl of 10X NEB Cutsmart® buffer and 1μl of QuickCIP (NEB) in a total volume of 48μl. The reaction was mixed by gentle flicking and incubated at 37°C for 30 minutes. The QuickCIP was then heat inactivated at 80°C for 5 min. 2μl of DpnI (NEB) was added to the reaction, mixed by gently flicking and incubated at 37°C overnight. DNA was A-tailed by addition of 1μl 10mM dATPs and 1μl of Klenow exo- (NEB) and incubation at 37°C for 1 hr. The digested DNA was purified using the Qiagen QiaQuick™ PCR Purification kit (which purifies fragments <10kb) and eluted in 60μl of H_2O. The DNA was

prepared for ONT sequencing with the DNA Ligation Sequencing Kit (LSK110), starting from the adaptor ligation step and using the Short Fragment Buffer.

Illumina DamID

Illumina-DamID-seq was performed as previously described with minor modifications[4]. Briefly, gDNA was extracted using the NEB Monarch gDNA extraction kit and digested overnight with DpnI (NEB). DamID adaptors were then ligated to DpnI-digested DNA. The adaptor-ligated DNA was digested with DpnII (NEB) to cleave fragments containing internal unmethylated GATC sites. Dam-methylated DNA was then amplified by PCR with MyTaq polymerase (Bioline®), sonicated to an average size of 300bp and the DamID adaptors were removed by AlwI (NEB) digestion. 1µg of sonicated DNA was prepared for sequencing using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® without size selection and amplified with 3 cycles of PCR. Libraries were sequenced on the Illumina® NextSeq platform using single-end 75bp chemistry.

DamID analysis

Nanopore reads were aligned to the mouse genome (mm10) using Minimap2[30] with the options -a --MD --cs=long -x map-ont. Mapped reads were filtered for map quality using SAMtools[31] view -q 10 and converted to bed format using bedtools bamtobed and bigwig format using deepTools[32] bamCoverage. The coordinates of the ends of reads were extracted using awk, extended to 100bp and converted to FASTA using bedtools getfasta. The sequences surrounding read ends were surveyed for enriched 4bp motifs using MEME-ChIP[25]. GATC motifs in the mouse genome (mm10) were extracted using the damidseq_pipeline[33]. The intersection of Nanopore read ends with GATC motifs was calculated using bedtools intersect. Lamin-associated domains were identified using Rseg[24] with the option -i 20 and no correction for deadzones. Short-read DamID was mapped using bwa mem[34] and filtered for alignment quality using samtools view -q 10. CpG methylation was called from Nanopore-DamID data using megalodon v2.2.9 with the model

dna_r9.4.1_450bps_modbases_5mc_hac_prom.cfg. DNA methylation was visualised using methylartist[26].

Code availability

CpG methylation analysis was performed using megalodon available from https://github.com/nanoporetech/megalodon and methylartist[26] available from https://github.com/adamewing/methylartist.

Data availability

All sequencing data is deposited in the NCBI Short Read Archive (SRA) repository as BioProject PRJNA751437.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

S.W.C. designed the study. Y.M.A.J, S.B.A. and S.W.C. performed the experiments. A.D.E., and S.W.C. performed the analysis. S.W.C and G.J.F. funded the study. S.W.C., Y.M.A.J., N.J. and G.J.F. wrote the manuscript. All authors read and approved the final manuscript.

References

1. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).

2. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* **18**, 424–428 (2000).

3. Aughey, G. N., Cheetham, S. W. & Southall, T. D. DamID as a versatile tool for understanding gene regulation. *Development* **146**, dev173666 (2019).

4. Marshall, O. J., Southall, T. D., Cheetham, S. W. & Brand, A. H. Cell-type-specific profiling of protein–DNA interactions without cell isolation using targeted DamID with next-generation sequencing. *Nat. Protoc.* **11**, 1586–1598 (2016).

5. Aughey, G. N., Estacio Gomez, A., Thomson, J., Yin, H. & Southall, T. D. CATaDa reveals global remodelling of chromatin accessibility during stem cell differentiation in vivo. *Elife* **7**, e32341 (2018).

6. Cheetham, S. W. & Brand, A. H. RNA-DamID reveals cell-type-specific binding of roX RNAs at chromatin-entry sites. *Nat. Struct. Mol. Biol.* **25**, 109–114 (2018).

7. Lebrun, E., Fourel, G., Defossez, P.-A. & Gilson, E. A methyltransferase targeting assay reveals silencer-telomere interactions in budding yeast. *Mol. Cell. Biol.* **23**, 1498–1508 (2003).

8. Redolfi, J. *et al.* DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. *Nat. Struct. Mol. Biol.* **26**, 471–480 (2019).

9. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).

10. Germann, S., Juul-Jensen, T., Letarnec, B. & Gaudin, V. DamID, a new tool for studying plant chromatin profiling in vivo, and its use to identify putative LHP1 target loci. *Plant J.* **48**, 153–163 (2006).

11. Schuster, E. *et al.* DamID in C. elegans reveals longevity-associated targets of DAF-16/FoxO. *Mol. Syst. Biol.* **6**, 399 (2010).

12. Tosti, L. *et al.* Mapping transcription factor occupancy using minimal numbers of cells in vitro and in vivo. *Genome Res.* **28**, 592–605 (2018).

13. Cheetham, S. W. *et al.* Targeted DamID reveals differential binding of mammalian pluripotency factors. *Development* **145**, dev170209 (2018).

14. Wade, A. A., van den Ameele, J., Cheetham, S. W. & Yakob, R. Novel CHD8 genomic targets identified in fetal mouse brain by in vivo Targeted DamID. *bioRxiv* (2021).

15. Simpson, J. T. *et al.* Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **14**, 407–410 (2017).

16. Rand, A. C. *et al.* Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* **14**, 411–413 (2017).

17. Ewing, A. D. *et al.* Nanopore Sequencing Enables Comprehensive Transposable Element Epigenomic Profiling. *Mol. Cell* **80**, 915–928.e5 (2020).

18. Shipony, Z. *et al.* Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat. Methods* **17**, 319–327 (2020).

19. Weng, Z. *et al.* Long-range single-molecule mapping of chromatin modification in eukaryotes. *bioRxiv* (2021).

20. Altemose, N., Maslan, A., Smith, O. K. & Sundararajan, K. DiMeLo-seq: a long-read, single-molecule method for mapping protein-DNA interactions genome-wide. *bioRxiv* (2021).

21. Gilpatrick, T. *et al.* Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* **38**, 433–438 (2020).

22. Kind, J. *et al.* Single-Cell Dynamics of Genome-Nuclear Lamina Interactions. *Cell* **153**, 178–192 (2013).

23. Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134–147 (2015).

24. Song, Q. & Smith, A. D. Identifying dispersed epigenomic domains from ChIP-Seq data. *Bioinformatics* **27**, 870–871 (2011).

25. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).

26. Cheetham, S. W., Kindlova, M. & Ewing, A. D. Methylartist: Tools for Visualising Modified Bases from Nanopore Sequence Data. *bioRxiv* (2021).

27. Berman, B. P. *et al.* Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina–associated domains. *Nat. Genet.* **44**, 40–46 (2011).

28. Xie, W. J. *et al.* Structural Modeling of Chromatin Integrates Genome Features and Reveals Chromosome Folding Principle. *Sci. Rep.* **7**, 2818 (2017).

29. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

30. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

31. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

32. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–91 (2014).

33. Marshall, O. J. & Brand, A. H. Damidseq-pipeline: An automated pipeline for processing DamID sequencing datasets. *Bioinformatics* **31**, 3371–3373 (2015).

34. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. in *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* 314–324 (2019).

35. Randolph, L. N., Bao, X., Zhou, C. & Lian, X. An all-in-one, Tet-On 3G inducible PiggyBac system for human pluripotent stem cells and derivatives. *Sci. Rep.* **7**, 1549 (2017).
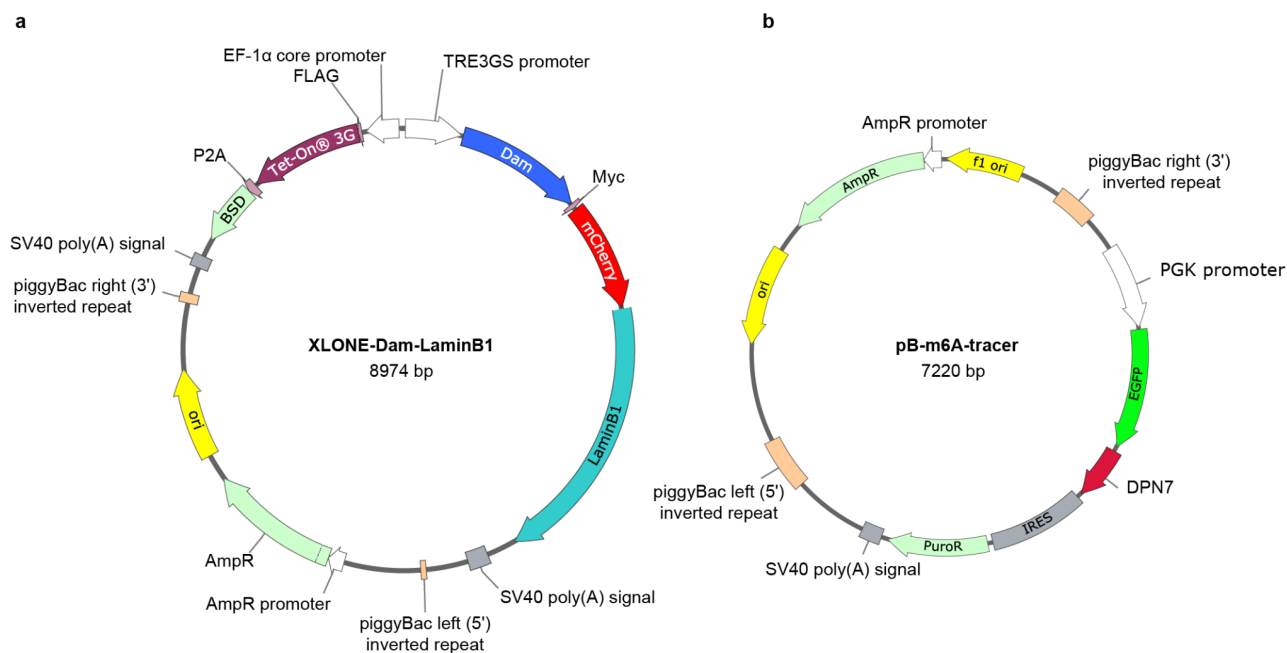
Supplementary figure 1: Constructs used for Nanopore-DamID. a, An all-in-one TetON-3G piggyBac construct based on XLONE[35] enables generation of a stable cell line with an integrated doxycycline-inducible Dam-LaminB1. **b,** A piggyBac construct enables generation of a stable cell line constitutively expressing a catalytically inactivate DpnI fragment (Dpn7) fused to eGFP. Dpn7-eGFP binds to methylated GATC sites enabling visualisation of Dam-LaminB1 activity.
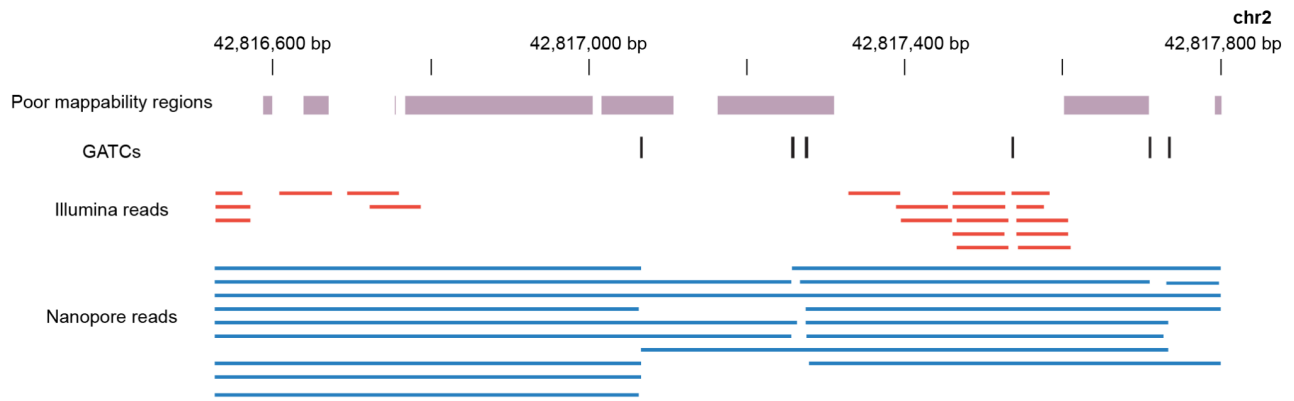
Supplementary figure 2: Superior alignment of Nanopore-DamID reads to regions of low mappability. Nanopore reads span regions that are difficult to align with 75bp Illumina reads (deadzones identified by RSEG).
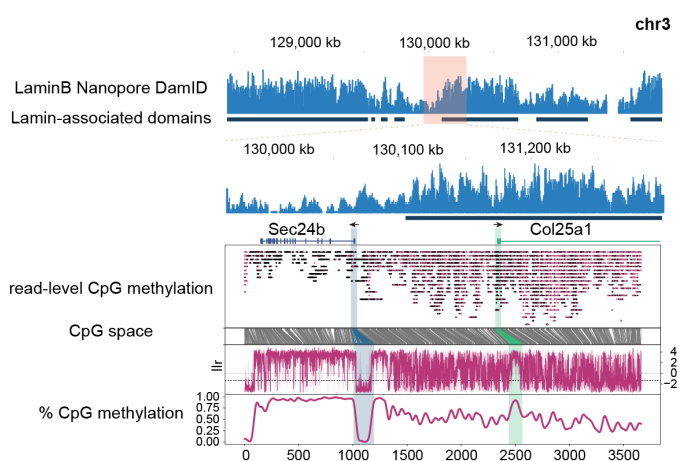
Supplementary figure 3: Resistance of a TSS on chr3 to LAD hypomethylation. Example of the interaction of LaminB1 occupancy with DNA-methylation.