nature portfolio

Corresponding author(s): Marc-Henri Stern

Last updated by author(s): Oct 20, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	FACSDiva Software v9.0.1 (BD Biosciences), INSPIRE for ISX mkII Software v200.1.620.0 (Cytek Biosciences), Zen Microscopy Software v3.7 (Zeiss), Odyssey Imaging System Application v3.0.25 (LI-COR)
Data analysis	RNAseq: FastQC v0.11.8, TrimGalore v0.6.2, STAR v2.6.1, Picard MarkDuplicates v2.18.15, Samtools v1.9, featureCounts v1.6.2 WES/WGS: FastQC v0.11.9, TrimGalore v0.6.10, BWA-MEM v0.7.17, Sambamba v1.0, Samtools v1.17, GATK v4.4.0.0, VEP v104.3 WGBS: FastQC v0.12.0, TrimGalore v0.6.10, Bismark v0.21.0, Bowtie2 v2.2.9 CUT&RUN: Pipeline based on nf-core/chipseq, using FastQC v0.11.8, TrimGalore v0.6.2, BWA-MEM v0.7.17, Picard MarkDuplicates v2.19.0, Samtools v1.9, phantompeakqualtools v1.2.2 Mutational signatures and data integration: bedtools v2.27.1, R 4.2.1 with packages signature.tools.lib v2.4.1, MutationalPatterns v3.8.1, annotatr v1.24.0, ggpubr v0.6.0, tidyverse v2.0.0 Flow cytometry: FlowJo v10.8.1 Imaging flow cytometry: IDEAS Image Analysis Software v6.2 (Cytek Biosciences)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Newly generated WGBS, WGS, WES, and RNAseq data are deposited in the European Genome-phenome Archive (EGA) database under accession EGAS50000000536 [https://ega-archive.org/studies/EGAS50000000536]. Data can be made accessible upon request to the DACs EGAC50000000356 (Institut Curie) or EGAC00001002078 (Max Planck Institute for Molecular Genetics). Newly generated WGS data on the HGG from Hôpital La Pitié Salpêtrière cannot be made available due to ethical approval restrictions. Newly generated CUT&RUN data are deposited in Gene Expression Omnibus (GEO) under accession GSE275181 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE275181]. Previously published WES from Institut Curie patients is available from EGA under accessions EGAD00001004554 [https://ega-archive.org/datasets/EGAD00001004554] and EGAD00001006988 [https://ega-archive.org/datasets/EGAD00001006988]. Previously published WGS data from external sources are available from EGA under accessions EGAD00001003568 [https://ega-archive.org/datasets/EGAD00001003568] and EGAD00001005454 [https://ega-archive.org/datasets/EGAD00001005454]. WGS data from GEL is not available upon request, but accessible by registered researchers in the Trusted Research Environment (https://www.genomicsengland.co.uk/research/research-environment). All GEL analyses must take place within the Trusted Research Environment (https://www.genomicsengland.co.uk/understanding-genomics/data). Registration involves an online application, verification by the applicant's institution, completion of a short information, governance course, and verification of approval by Genomics England. Please see https://www.genomicsengland.co.uk/research/academic/ for more information. Previously published WGBS data on human tissues are available in EGA under accession EGAD00001009789 [https://ega-archive.org/datasets/EGAD00001009789] and in GEO under accession GSE65196 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65196]. Previously published WGBS data on mouse cells are available in GEO under accessions GSM1382253 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1382253] and GSM1382256 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1382256]. Publicly available human epigenomic data from ENCODE [https://www.encodeproject.org/] are described in Supplementary Data 10. Previously published single nuclei RNAseq data are available from the Broad Institute Single Cell Portal (dataset SCP2298 [https://singlecell.broadinstitute.org/single_cell/study/SCP2298]). Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	This information has not been collected.
Reporting on race, ethnicity, or other socially relevant groupings	This information has not been collected.
Population characteristics	We analyzed a population of individuals with a diagnosis of cancer. Only tumor type and genotypic information were collected.
Recruitment	Selection of patients from the Genomics England series (GEL) was based on sequencing quality and diagnosis information, as provided by GEL cancer analysis Table v17, which provides cross-referenced and updated classification aligned with The Cancer Genome Atlas (TCGA) nomenclature. All samples external to the GEL series were chosen due to their known biallelic loss of MBD4. This selection bias is not expected to impact the results presented.
Ethics oversight	Institut Curie Review Board CRI-DATA (Project DATA190061) and Charité - Universitätsmedizin Berlin Institutional Review Board (EA4/063/13).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. Sample size of human tumors was based on availability of sequencing data.
Data exclusions	Human tumors from the GEL series were excluded if not meeting pre-defined requirements of sequencing quality, tumor cell content, and minimal signature exposure.

Replication	CUT&RUN experiment was performed once. CUT&RUN FLAG reactions on FLAG-tagged MBD4 clones were performed in biological duplicates. CUT&RUN control reactions (FLAG on parental HAP1 cells, IgG negative controls, and H3K4me3 positive control) were performed without replicates. Long-term culturing of HAP1 cells was performed once, followed by WGS of one clone (D0) and four subclones (D120 biological replicates) per genotype. Real-time quantitative reverse transcription PCR was performed once, without biological replicates and with three technical replicates per genotype. Doubling times of HAP1 single-cell clones were obtained once, with three biological replicates per genotype.
Randomization	This is not relevant to our study. Allocation of patient data was based on tumor type and mutational signature contribution. Allocation of cell line models data was based on known and precisely controlled genotypic information.
Blinding	Investigators were not blinded to group allocation during data collection and/or analysis. Blinding was not relevant to the analysis of patient data, as we focused on a subset of samples with a known contribution of specific mutational signatures. Blinding was not relevant to the analysis of experimental cell line models, as genotypic information was known and precisely controlled for each sample.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Anti-MBD4 (Abcam ab227625; lot GR3230875-5), anti-MBD4 (Abcam ab224809; lot 1017919-4), anti-MBD4 (Abcam ab12187; lot GR21754-20), anti-TDG (Invitrogen PA5-29140; lot YA3812362), anti-FLAG (Cell Signaling 14793S; lot 7), anti-MLH1 (Sigma HPA052707; lot R69680), IgG negative control (Cell Signaling 66362S; lot 2), anti-H3K4me3 (EpiCypher 13-0041; lot 13-0041k), anti-Tubulin (Invitrogen 14-4502-80; lot 2003406), anti-Histone H3 (Abcam ab1791; lot GR252388-1), anti-mouse Alexa Fluor Plus 555 (Invitrogen, A32727; lot XH350742), anti-rabbit Alexa Fluor Plus 647 (Invitrogen, A32733; lot XG349344), anti-rabbit 800CW (LI-COR, 926-32213; lot D11005-09).
Validation	- MBD4 and TDG antibodies were validated by immunofluorescence and/or western blotting using HAP1 cells wild-type or knock-out for MBD4, TDG, or both (dKO). - FLAG antibody was validated by immunofluorescence and western blotting using parental HAP1 cells or clones overexpressing C- or N-terminally FLAG-tagged MBD4. - MLH1 antibody has been extensively validated by the Human Protein Atlas (https://www.proteinatlas.org/ENSG00000076242-MLH1/summary/antibody). - IgG negative control is validated for CUT&RUN applications by manufacturer (https://www.cellsignal.com/products/primary-antibodies/rabbit-da1e-mab-igg-xp-isotype-control-cut-amp-run/66362?srsltid=AfmBOooFn-Rv35UuosW3kjrfXUWniSQ6amko7sh7vZZs2IUeMIwWqNKc). - H3K4me3 antibody is validated for CUT&RUN applications by manufacturer (https://www.epicypher.com/products/antibodies/snap-chip-certified-antibodies/histone-h3k4me3-antibody-snap-chip-certified-cutana-cut-run-compatible). - Tubulin antibody has been largely used by the scientific community, as indicated in the manufacturer's website (https://www.thermofisher.com/antibody/product/alpha-Tubulin-Antibody-clone-DM1A-Monoclonal/14-4502-82). - Histone H3 antibody has been largely used by the scientific community, as indicated in the manufacturer's website (https://www.abcam.com/en-us/products/primary-antibodies/histone-h3-antibody-nuclear-marker-and-chip-grade-ab1791).

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	HAP1 parental (Horizon Discovery C631) and knockout for MBD4 (Horizon Discovery HZGHC000921c002).
Authentication	HAP1 cell lines were authenticated by WGS.
Mycoplasma contamination	All cell lines tested negative for mycoplasma, both before and after long term culturing. Tests were performed by Eurofins Mycoplasmacheck service.

Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in the study.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	https://www.ncbi.xyz/geo/query/acc.cgi?acc=GSE275181
Files in database submission	Fastq raw files; Read depth normalized coverage bigwig files.
Genome browser session (e.g. UCSC)	No longer applicable.

Methodology

Replicates	CUT&RUN FLAG reactions on FLAG-tagged MBD4 clones were performed in biological duplicates. CUT&RUN control reactions (FLAG on parental HAP1 cells, IgG negative controls, and H3K4me3 positive control) were performed without replicates.
Sequencing depth	Paired-end (2 x 100 bp) libraries were sequenced with a target of 12-16 million fragments per sample.
Antibodies	FLAG (1:50; Cell Signalling, 14793S; lot 7), H3K4me3 (1:50; EpiCypher, 13-0041; lot 13-0041k), IgG (1:10; Cell Signaling, 66362S; lot 2).
Peak calling parameters	Not performed.
Data quality	FLAG enrichment over IgG control was compared between HAP1 clones expressing FLAG-tagged MBD4 versus parental HAP1 cells. H3K4me3 positive control reaction showed high correlation with ENCODE H3K4me3 data in the same cell type.
Software	Pipeline based on nf-core/chipseq, using FastQC v0.11.8, TrimGalore v0.6.2, BWA-MEM v0.7.17, Picard MarkDuplicates v2.19.0, Samtools v1.9, phantompeakqualtools v1.2.2.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Normal eye tissues from UVM or conjunctival melanoma patients who underwent eye enucleation surgery at Institut Curie were obtained. Normal uveal choroid was dissected from a region diametrically opposed to the tumor site. Tissue was enzymatically digested for 40 min at 37°C (DMEM/F12 [Gibco, 11330032], BFS 10% [BioSera, FB-1003], Penicillin-Streptomycin 100 U/mL [Gibco, 15140122], collagenase type IV 400 CDU/mL [Sigma, C1889], DNAse I 12 µg/mL [Sigma, D5025]) under gentle agitation. Cell suspensions were washed once with wash buffer (PBS without Ca2+ and Mg2+, BSA

	0.5%, EDTA 2 mM) and filtered with 30 µm strainers. During initial validation experiments, cells were additionally stained for 30 min on ice with anti-CD45 coupled with BB515 (1:20; BD Biosciences, 564585) in wash buffer. Cells were then resuspended in media (DMEM/F-12 [Gibco, 11039021]) containing 5 µM Vybrant DyeCycle Violet (Invitrogen, V35003) and incubated for 30 min at 37°C. Finally, cells were resuspended in wash buffer containing 5 µM Vybrant DyeCycle Violet and 1 µg/mL propidium iodide (PI; BioLegend, 421301) before flow cytometry analysis.
Instrument	FACSAria Fusion Cell Sorter (BD Biosciences), Amnis ImageStream Mk II (Cytek Biosciences).
Software	FACSDiva Software v9.0.1 (BD Biosciences), FlowJo v10.8.1, INSPIRE for ISX mkII Software v200.1.620.0 (Cytek Biosciences); IDEAS Image Analysis Software v6.2 (Cytek Biosciences).
Cell population abundance	Sorted uveal melanocytes showed purity >90%, as confirmed by direct brightfield melanin visualization using an imaging flow cytometer Amnis ImageStream Mk II (Cytek Biosciences) .
Gating strategy	Gating to sort uveal melanocytes was: Propidium iodide neg/CD45 neg/Vybrant Violet dim/Red autofluorescence bright.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.