nature portfolio

Corresponding author(s): Housheng Hansen He
Trevor Pugh
Mathieu Lupien
Yong Zeng

Last updated by author(s): Jan 7, 2026

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	No specific software was used for data collection in this study.
Data analysis	All data analysis details, including method and detailed parameters are described in Methods. All codes for this study have been deposited at GitHub (https://github.com/HansenHeLab/cfMeDIP-seq_Data_Resource_Codes). The analytical pipeline will be also deposited on the CoBE platform (www.pmcobe.ca). Scripts used for the fragmentomics analysis were based on scripts originally developed for these projects: https://github.com/pughlab/TGL48_Uveal_Melanoma and https://github.com/pughlab/LFS-early-detection-ctdna. The following software or packages were used: MEDIPIPE (version 1.1.0), ComBat-seq (sva version 3.52.0), DESeq2 (version 1.44.0), MEDEStrand (version 0.0.0.9000), QSEA (Beta, version 1.20.0), methyclock (version 1.4.0), annotatr (version 1.24.0), ClusterProfiler (version 4.6.2), missMethyl (version 1.32.1), ChIPseeker (version 1.34.1), Picard(version 4.0.1.2), pwr.t.test (v.1.3-0), PERMANOVA (adonis2, vegan v.2.6-10), mvnormtest (v0.1.9.3).

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

The raw cfMeDIP-seq data that support the findings of this study from TCGE-CFMe-MCA, TCGE-CFMe-BCA, TCGE-CFMe-HNSC, and TCGE-CFMe-SCLC were obtained directly from the corresponding authors of the respective cohorts upon request. TCGE-CFMe-MCA and TCGE-CFMe-BCA data can be requested from Dr. Daniel D. De Carvalho (ddecarv@uhnresearch.ca), TCGE-CFMe-HNSC data from Dr. Scott V. Bratman (scott.bratman@rmp.uhn.ca), and TCGE-CFMe-SCLC data from Dr. Benjamin H. Lok (benjamin.lok@rmp.uhn.ca) via submission of a data access application. The other cfMeDIP-seq datasets were deposited in the European Genome-Phenome Archive (EGA): TCGE-CFMe-AML (EGAS00001005069), TCGE-CFMe-PRAD (EGAS00001005522), TCGE-CFMe-UM (EGAD00001008998), TCGE-CFMe-HBC (EGAS00001006539), TCGE-CFMe-LFS (EGAS00001006539), TCGE-CFMe-INSPIRE (EGAD00001011312), and TCGE-CFMe-HCC (EGAD50000000652). Access to all seven UHN-generated datasets is made available upon completion of the required data access agreement, which will be reviewed by the UHN genomics data access committee (dac@uhn.ca). Data access will be granted to qualified investigators for appropriate and compliant use. The source data for DMRs analyses (Fig. 2b; Extended Data Fig. 3e, f; Extended Data Fig. 4d, e, g; Extended Data Fig. 48a) and methylation-based PCA and UMAP plots (Fig. 2a; Extended Data Fig. 2a-c, f; Extended Data Fig. 3a, c; Extended Data Fig. 7i), together with BED files for pan-cancer and cancer-specific DMRs as well as age- and sex-associated regions, are available on Zenodo (https://zenodo.org/records/15191455). All other source data is provided in this paper. The remaining data are available within the article and Supplementary Tables.

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	For samples collected in this study, all sex information was self-reported. For the primary dataset, samples came from 390 male donors and 305 female donors. The remaining 223 individuals lacked study-reported sex information. For the validation dataset, sample were from 38 male and 71 female donors. Although potential confounding effects of sex was evaluated, it was not included as covariate in our analyses.
Reporting on race, ethnicity, or other socially relevant groupings	None of this information was included in this study.
Population characteristics	The population characteristics (age and clinical features) and summary for our primary and validation samples were provided in Supplementary Table 1 and Supplementary Table 9, respectively.
Recruitment	Recruitment is not applicable in this case as all datasets analysed here were derived from published studies.
Ethics oversight	All samples obtained in this study complied with the relevant ethical regulations approved by the institutional ethics committee and Research Ethics Board at the University Health Network (UHN).

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	Sample sizes were not predetermine by any statistical methods. Sample size was based on sample availability, and was demonstrated the sufficient power to obtain definitive results.
Data exclusions	All data were uniformly processed, provided as a resource. While the primary analyses in this study were restricted to the high quality and baseline time point samples.
Replication	One replication per sample was obtained
Randomization	The work requires no randomization: Human specimen were allocated into groups according to disease types

Blinding | Blinding is not relevant as no treatment for patient is needed

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

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.*