

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	The demultiplexed FASTQ files from the sequencing center were shared with us via FTP, or downloaded using Illumina BaseSpace manager.
Data analysis	Samples were demultiplexed either onboard Illumina HiSeq2500 or offline using bcl2fastq Conversion Software v1.8.4 (Illumina) Sequencing output was quality checked with FastQC v0.11.9 (Babraham Institute, https://www.babraham.ac.uk/) and adapter sequences were trimmed with TrimGalore version 0.6.5, and reads are aligned to the UCSC hg19 reference assembly using Bismark v0.10.1 (Krueger and Andrews, 2011). Read deduplication for NuGen was carried out using NuDup, (NuGEN, https://github.com/nugentechnologies/nudup). For hybridization capture methods PICARD CollectHsMetrics tool v2.3.0 was used to obtain statistics on HC efficiency. BEDTools v2.25.0 and custom R, sed and awk scripts were used to calculate and visualize the target region coverage as a function of read depth. MultiQC v1.10.1 was used for visualizing aggregated quality metrics. R version 3.4.1 was used with following packages: 'plyr' (1.4.0), 'stringr' (>1.12.0), 'data.table' (1.9.6), To compare target breadth of coverage UpSet plot from 'ComplexHeatmap' (2.10.0) and Venn diagram from 'ChipPeakAnno' (3.28.1) were used to visualize the intersection of genomic ranges and CpG sites. Feature annotation was done using 'anotatr' R package, and analyzed with 'GenomicRanges' (1.46.1) and 'GenomicFeatures' (1.46.5). Bismark produced strand-specific methylation calls were analyzed with 'MethylKit' (1.20.0). The genic annotations are determined by functions from 'GenomicFeatures' (1.18.13) and data from the 'TxDb' (3.2.2) and 'org.hg19.eg.db' packages. Jaccard similarity index was calculated using functions from 'HelloRanges' (1.3.1) Intervene shiny app (https://asntech.shinyapps.io/intervene/) was used to visualize the Jaccard similarity index matrix. BoostMe (https://github.com/lulizou/boostme) was used for imputation. The custom code used for annotation, differential methylation analysis, plotting and imputation is available from Github repository at https://github.com/ucl-medical-genomics/EpiCapture .

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 datasets generated and analyzed in the current study including all raw targeted BS-seq, WGBS of Ref.gDNA and Nanopore sequencing data have been deposited in the European Nucleotide Archive repository under accession no. PRJEB46506 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB46506?show=reads>) and are freely available. Raw WGBS sequencing data for Coriell NA12878 WGBS_EC sample generated by the ENCODE Project Consortium²⁶ was downloaded from the Encode Project (experiment: ENCSR890UQO, library: ENCLB898WPW) (<https://www.encodeproject.org/experiments/ENCSR890UQO/>), and CpG count files for WGBS_IL sample were downloaded from Illumina BaseSpace Hub (<https://basespace.illumina.com/datacentral>) under sample name WGBS_P3 from HiSeq 4000: TruSeq DNA Methylation (NA12878, 2x76) dataset.

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	For each TBS platform a total of 16 libraries were generated (Fig. 1a) using the same source DNA from 11 biological replicates generating a total of 80 TBS libraries, in addition to n=2 WGBS (Coriell-NA12878 and Ref.gDNA) and n=1 Nanopore (Coriell-NA12878) libraries. The primary goal of the study was to measure the technical (random) variability and correlation of called DNA methylation levels and other metrics within and between TBS platforms over the same set of samples. To determine the number of samples we followed a statistical rule of thumb based on the fact that the width of the confidence interval (CI) is a function of the square root of the number of observations. The width of a CI involving estimation of variability and sample size decreases rapidly until 12 observations are reached and approaches the asymptote at 15 observations (Gerald van Belle, Statistical Rules of Thumb, 2008). The half-width of the 95% confidence interval at n=15 is 0.56 assuming normal distribution with n-1 degrees of freedom.
Data exclusions	Library prep failed for samples Agilent DNAm-10pct and Illumina DNAm-5pct, while Roche Hela-1 library failed target capture and was excluded from downstream analysis.
Replication	For each TBS platform a total of 16 libraries were generated (Figure 1a) using the same source DNA from 11 biological replicates that consisted of: reference gDNA isolated from human peripheral blood cells (Ref.gDNA) at two different DNA inputs (recommended by the manufacturer and 500 ng) in duplicate; Coriell NA12878 and Hela cell lines processed in duplicate; four DNA methylation standards generated from ZYMO fully methylated and unmethylated control samples; a pair of genetically and phenotypically divergent bladder cancer cell lines and a pair of isogenic bladder cancer cell lines with different sensitivity to cisplatin treatment. All replication attempts were successful, except for Hela sample processed with Roche TBS platform.
Randomization	Not relevant to our study. We used the same source DNA to generate libraries used for comparison, sequenced them on the same instrument (HiSeq 2500) and analyzed using the same bioinformatic pipeline.
Blinding	Not relevant to our study. Blinding was not required since all samples processed by different TBS platforms came from the same source DNA and the data was processed using the same computational pipeline designed to be unbiased in their evaluation of methods.

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
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

Cell lines 253J , T24 and RT112 were obtained from the American Type Culture Collection (#TCP-1020). RT112-CP cell line derived from RT112 by growing the cell under selection for cisplatin resistance was a gift from Prof Jim Catto, Sheffield University. The Coriell-NA12878 DNA sample was purchased from Coriell Cell Repositories. Hela cell genomic DNA was purchased from New England Biolabs. The Human Methylated & Non-methylated DNA Set were purchased from ZYMO Research.

Authentication

None of the cell lines were authenticated.

Mycoplasma contamination

Cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines
(See [ICLAC](#) register)

None of these misidentified lines were used.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

The reference genomic DNA sample (Ref.gDNA) was obtained by pooling DNA extracted from peripheral whole blood cells of six adult healthy volunteers (four male and two female), age range (28-62).

Recruitment

Adult healthy UCL CI staff members were approached to donate blood samples providing informed consent under Ethics approval 15/YH/0311 of the UCL BioBank for Health and Disease. The plasma samples were pooled to ensure anonymity and make a reference genomic DNA sample (Ref.gDNA) representing an admixture of healthy blood plasma samples.

Ethics oversight

Informed consent was obtained under study 15/YH/0311 approved by the UCL BioBank for Health and Disease ethics committee.

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