

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 hard drives.
Data analysis	For WGBS and EM-seq, all the raw paired fastq files were processed in parallel through two pipelines. In the first pipeline, the adapters and low-quality reads were trimmed using fastp47 v0.23.4 (https://github.com/OpenGene/fastp), aligned to the human reference genome (build GRCh38p14, ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna.gz) with Bismark-Bowtie248 v0.23.0 (https://github.com/FelixKrueger/Bismark), and duplicated reads from aligned bam files were removed with Bismark. In the other pipeline, Trim Galore v0.6.2 (https://github.com/FelixKrueger/TrimGalore), BWA-MEM2 v2.2.1 (https://github.com/bwa-mem2/bwa-mem2), bwa-meth v0.2.7 (https://github.com/brentp/bwa-meth), and GATK4 v 4.5.0.0 (https://broadinstitute.github.io/picard/) were used for trimming, alignment, and deduplication. For TAPS, two pipelines were engaged. In the first pipeline, the Bismark-bowtie2 aligner was replaced with BWA-MEME v1.0.6 (https://github.com/kaist-ina/BWA-MEME), and GATK4 was used for deduplication. After removing duplicated reads from different methods, the per-cytosine methylation information was called by the asTair v3.3.2 (https://bitbucket.org/bsblabludwig/astair/src/master/) with in silico SNP exclusion using Quartet SNP reference datasets, ensuring CpG-specific signal detection before Bismark-mediated pipeline consensus merging. FastQC v0.12.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), Qualimap49 v2.3 (http://qualimap.conesalab.org/), and MultiQC50 v1.10.1 (https://seqla.io/multiqc/) were used as quality control tools without further separate explanations and were all run with default parameters. The reference of lambda DNA (Enterobacteria phage lambda, NC_001416.1) as negative base conversion control was spiked into the GRCh38p14 human reference genome fasta file (NCBI, GCA_000001405.15). The ALT contigs were removed from the fasta file. For the ONT platform, Megalodon v2.2.9 (https://nanoporetech.github.io/megalodon) and Tombo v1.5.151 (https://github.com/nanoporetech/tombo) were utilized to call the bases and modifications from the raw fast5 files. DeepSignal252 v0.1.3 and f5c v1.1 (https://github.com/hasindu2008/f5c) were used for downstream methylation calling and correction. The configuration file of the MinION DNA R9.4.1

5mC model was retrieved from the Rerio database.

For PacBio, pbccs v6.4.0 (<https://github.com/nlhepler/pbccs>) and pbmm2 v1.9.0 (<https://github.com/PacificBiosciences/pbmm2>) were used for base calling and alignment, respectively. After that, ccsmeth v0.3.2 (<https://github.com/PengNi/ccsmeth>), Primrose v1.3.0 (<https://github.com/mattosimp/primrose>), and pb-cpg-tools v2.2.0 (<https://github.com/PacificBiosciences/pb-CpG-tools>) were used to call the methylation information.

The source codes for the data analysis are available at Compute Capsule's provisional DOI: <https://doi.org/10.24433/CO.1027879.v1>

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 sequence data used in this paper have been deposited in the Genome Sequence Archive (GSA) under accession HRA011205 [<https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA037414>]. The microarray data used in this paper have been deposited in the Gene Expression Omnibus (GEO) under accession GSE241900 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE241900>]. Reference datasets are available at 10.6084/m9.figshare.29481713.v1 [<https://doi.org/10.6084/m9.figshare.29481713.v1>]; a mirror is provided at the Quartet Data Portal [<https://reference-datasets.chinese-quartet.org/index.html?prefix=develop/Methylation/v20250409/>] under the Administrative Regulations of the People's Republic of China on Human Genetic Resources. The raw sequence data, CpG call sets, and microarray IDATs are also available in NODE OEP00004817 [<https://www.biosino.org/node/project/detail/OEP00004817>]. 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

Quartet DNA reference materials were derived from four healthy volunteers of a Chinese Quartet family from the Fudan Taizhou Cohort, including father (F7), mother (M8), and monozygotic twin daughters (D5 and D6).

Reporting on race, ethnicity, or other socially relevant groupings

East Asians.

Population characteristics

See above.

Recruitment

The recruitment has been previously approved and accomplished (Zheng, Y., Liu, Y., Yang, J. et al. Multi-omics data integration using ratio-based quantitative profiling with Quartet reference materials. *Nat Biotechnol* 42, 1133–1149 (2024). <https://doi.org/10.1038/s41587-023-01934-1>).

Ethics oversight

This study was approved by the Institutional Review Board (IRB) of the School of Life Sciences, Fudan University (BE2050). It was conducted under the principles of the Declaration of Helsinki.

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](https://www.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 statistical method was used to predetermine the sample size. Sizes were fixed by the certified Quartet reference set and a factorial design sufficient to estimate variance components and stabilize metrics. We sequenced three replicates of each of the four Quartet DNA reference materials using three commercially available short-read sequencing protocols: WGBS, EMseq, and TAPS, generating nine data batches. Each batch included twelve libraries, resulting in a total of 108 libraries across the nine batches. These sample sizes are sufficient to provide within-batch technical replication, cross-protocol comparisons, and cross-batch/laboratory reproducibility assessment. Details are illustrated explicitly in Fig. 1.

Data exclusions

All data from planned experiments have been included. All attempts at replication were successful.

Replication

Each batch included twelve libraries (three replicates of each of the four Quartet DNA reference materials), resulting in a total of 108 libraries across the nine batches.

Randomization

Samples were allocated by a pre-specified, balanced design rather than randomization. Each batch contained 12 libraries—three technical replicates per donor (D5, D6, F7, M8)—loaded in an interleaved order (D5-1, D6-1, F7-1, M8-1, D5-2, D6-2, F7-2, M8-2, D5-3, D6-3, F7-3, M8-3) to minimize run-order effects. Wet-lab personnel were blinded to donor identities and the allocation key. Because groups are defined by donor and protocol (not treatment), covariates were controlled by including all donors in every batch under identical preparation and sequencing conditions.

Blinding

Data collection was blinded. Analysis was not, because donor-specific references (Quartet SNP datasets) and ground-truth comparisons required linking each library to its known identity; analyses were fully scripted with fixed parameters and objective metrics, so blinding was not relevant.

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

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	Human immortalized B-lymphoblastoid cell lines from the Quartet family cohort of the Taizhou Longitudinal Study (Taizhou, Jiangsu, China): D5 (female), D6 (female), F7 (male), M8 (female). Not obtained from a vendor or cell bank; catalog numbers not applicable.
Authentication	The cell lines have been authenticated by STR profile, karyotype, PCR mycoplasma and sterility testing.
Mycoplasma contamination	No mycoplasma contamination found.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this 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, mosaicism, off-target gene editing) were examined.