nature research

Corresponding author(s): Alexander Meissner

Last updated by author(s): Apr 15, 2021

Reporting Summary

Nature Research 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 Research 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	The cell line used was HUES64. The genotype of the cell line is inducible DNMT1 KO as described in Liao et al, Nature Genetics, 2015. RNA-seq and RRBS libraries were sequenced using Illumina HiSeq 2500 instrument.
Data analysis	The quality of raw reads was assessed using FastQC (Andrews, 2010). The raw reads were aligned to the Homo sapiens genome (GRCh38.p2 genome build (Ensembl v79)) using STAR v2.7.3a (Dobin et al., 2013). The raw counts were computed using quantMode function in STAR. The obtained read counts are analogous to the expression level of each gene across all the samples. The single cell RNA-seq QC analysis was done using scater (McCarthy et al., 2017). RRBS-seq libraries were aligned to the bisulfite converted GRCh38.p2 genome using Bismark5 (v.0.18.2) with bowtie2 (Langmead et al 2012) as the aligner. The reads were trimmed using the Trim galore wrapper tool. Bismark methylation extractor (-bedgraph --buffer_size 50%) was used to determine the methylation state of each individual CpG across both the strands. Duplicated reads were removed, followed by estimating the number of CpGs. Here is the link to the terra pipeline is: https://app.terra.bio/#workspaces/aryee-lab/bisulfite-seq-tools-grch38. All the plots and graphs were generated in packages such as ggplot2, ggpubr, patchwork in the R environment.

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 Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

The sequencing data have been deposited to the NCBI GEO Archive under accession GSE157115 (GSE157113 for RNAseq and GSE157114 for RRBS data).

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. The sample size (n=96) for the technical demonstration data set was chosen because the protocol described is for a batch of n=96 single cells.
Data exclusions	We excluded data from 16 single cells that did not pass quality filters specified in Figure 3 from the performance metrics.
Replication	No experimental findings other than technical performance metrics based on a representative data set from a batch of 96 cells were reported. The performance metrics from other batches were very similar.
Randomization	Not relevant for this study which had only a single experimental group
Blinding	Blinding was not necessary for this merely technical performance study of a single experimental group

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)	The engineered human embryonic stem cell line was generated in the Meissner lab at Harvard and is a derivative of HUES 64 (NIHhESC-10-0067) provided by the Harvard Stem Cell Institute
Authentication	Cell line was generated and characterized in the Meissner lab at Harvard.
Mycoplasma contamination	The cell line tested negative for mycoplasma
Commonly misidentified lines (See ICLAC register)	Not applicable