nature research

Corresponding author(s): Andrew Adey

☐ Initial submission ☒ Revised version ☐ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

No statistical analysis was performed to predetermine sample size. The number of single-cell libraries to be generated was selected based on required sequencing depths from previous publications on single-cell methylomes, as well as the analyses to be performed. We determined the sample size for individuals in the mouse cortex to be sufficient at n=3, based on the minimal variance of cell type apportionment between individuals (z-score < |0.1|). Cell number was determined sufficient for both human cell line and mouse cortex experiments due to our ability to confidently identify cell line identity (human) and the three primary expected populations of cells in mouse (non-neuronal, excitatory, and inhibitory).

2. Data exclusions

Describe any data exclusions.

Reads were excluded if the three associated indexes did not match a predetermined barcode sequence. Barcodes to be analyzed were filtered based on final uniquely aligned read count and <5% nonCG methylation (a filter previously used in the field). Final uniquely aligned read count threshold was set independently for each library preparation both by using a local minima analysis to define a true single nuclei population dependent on sequencing depth (Supplemental figs. 1-3) and the intended analysis per library.

3. Replication

Describe whether the experimental findings were reliably reproduced.

The sci-MET protocol was used to generate a total of five library pools independently across four separate cell lines and two species. This demonstrates the reproducibility of the technique. The ability to reproduce single-cell methylomes was demonstrated by the clustering analysis in which hundreds of single cells showed a strong cluster identity with nuclei from the same cell line. Independently-generated replicate library preparations following the same conditions were not produced.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Nuclei were randomized at two stages. First, following isolation from cells, nuclei across the three human cell line types were mixed to generate a randomization at the first tagmentation step of the sci-MET protocol (referred to as the "Mix" population in the manuscript). Secondly, inherent to the sci-MET protocol of random subsampling of pooled nuclei, nuclei identity was randomized during the second nuclei sorting. For mouse cortex samples, the biological replicate was encoded in the first round of indexing and randomized during the pooling and second round.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

As stated in the Randomization response above, nuclei are randomly sampled during the second FANS in the sci-MET protocol. The nuclei identities within each reaction is unknown to the experimenter throughout bisulfite conversion, library construction and sequencing. Reads were processed using the same pipeline without barcodes assigned to a specific cell line until final analysis. Further the "Mix" population did not have an assigned cell type until the clustering analysis.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

FastQ files were generated from BCL files using bcl2fastq (Illumina Inc., v2.19.0). Reads were assigned to a 3-index barcode using a custom perl (v5.10.1) script. Reads were then trimmed using TrimGalore! (v0.4.0) and quality checked with FastQC (v0.11.3). Trimmed reads were then aligned to a reference genome using Bismark (v0.14.3). Further analysis was performed with bedtools (v2.22.0) and custom R and python scripts (v3.4.0, and v2.7.9, respectively). Determination of bona-fide single-cell libraries based on read depth distributions was carried out using MASS (v7.3-45) and mixtools (v1.1.0). Cell projection into two-dimensional space was performed using RtSNE (v0.13), and clustered using DBSCAN (v1.1-1). Plotting of data extensively utilized the R library ggplot2 (v2.2.1) or gplots (v3.0.1) and additional R libraries for data processing included GenomicRanges (v1.28.4), Genomation (v1.8.0), qvalue (v2.8.0), and ComplexHeatmap (v1.14.0). Code for custom perl, R and python scripts will be made available on GitHub by the time of publication and are available at any time upon request.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

The transposase reagent used in this study was provided by Illumina Inc. through a material transfer agreement (MTA). As with previous work published using this reagent (PMID: 25953818), Illumina will provide the transposase complexes through an MTA upon request. Alternatively, the transposase reagent can be synthesized according to protocols published by Illumina (PMID: 28095828) or another published method (PMID: 25079858), that has been demonstrated in combinatorial indexing assays by other groups (https://doi.org/10.1101/128520, https://doi.org/10.1101/159137).

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used in this study.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

GM12878, Coriell
NIH/3T3, ATCC CRL-1658
HEK293, ATCC CRL-1554
Primary Fibro., inguinal fibroblast, GM05756, Coriell

b. Describe the method of cell line authentication used.

None of the cell lines used in this study were authenticated.

c. Report whether the cell lines were tested for mycoplasma contamination.

Cell lines were not tested for mycoplasma contamination.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

HEK293 cells were used in this study and are listed as a commonly misidnetified cell line maintained by ICLAC. These cells are commonly mistaken with HeLa cells, a cell line excluded from this study. HEK293 cells were chosen because of both the established cell culture techniques as well as the publicly available data sets used to validate our approach. Of note, a misidentification of HEK293 for another human cell line would not change any of the major conclusions drawn from this study.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Mouse Samples. All animal studies were approved by the Oregon Health and Science University Institutional Animal Care and Use Committee. 60-day old C57BL/6J male mice (n=3) were obtained from Jackson Laboratory (stock number 000664).

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

This study did not involve human research participants.

nature research

Corresponding author(s): Andrew Adey

☐ Initial submission ☒ Revised version ☐ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

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

☒ 2. 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).

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

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

Methodological details

5. Describe the sample preparation.

Tissue Culture: Tissue culture cell lines (GM12878, Coriell; NIH/3T3, ATCC CRL-1658; HEK293, ATCC CRL-1554; Primary Fibro., inguinal fibroblast, GM05756, Coriell, passage 7) were cultured in 5% CO2 at 37˚C. GM12878 cells were grown in Roswell Park Memorial Institute media (RPMI, Gibco, Cat. 11875093) supplemented with 15% (by volume) fetal bovine serum (FBS, Gibco, Cat. 10082147), 1X L-glutamine (Gibco, Cat. 25030081), 1X Penicillin-Streptomycin (Gibco, Cat. 15140122), and gentamicin (Gibco, Cat. 15750060). HEK293 cells were grown in Dulbecco's Modified Eagle's media (DMEM, Gibco, Cat. 11995065), supplemented with 10% FBS, and 1X L-glutamine. NIH/3T3 cells were grown in the same preparation of DMEM as HEK293 cells. Primary Fibroblasts were cultured in a growth medium comprised of DMEM/F12 (with GlutaMax; Thermo Fisher), 10% fetal bovine serum (FBS; Thermo Fisher), 1% MEM Non-Essential Amino Acids (Thermo Fisher), and 1% Penicillin/Streptomycin (Thermo Fisher). Adherent cell lines were grown to ~90% confluency at the time of harvest.
Mouse Samples: All animal studies were approved by the Oregon Health and Science University Institutional Animal Care and Use Committee. C57BL/6J mice were obtained from Jackson Laboratory (stock number 000664). Sixty-day-old C57BL/6J mice were deeply anesthetized using isoflurane. After decapitation the brain was removed and the entire cortex isolated and placed in ice cold PBS.
Sample preparation and nuclei isolation: For library preparation, cells were pelleted if cultured in suspension, or trypsinzed (Gibco, Cat. 25200056), if adherent. Cell were washed once with ice cold PBS and carried through cross-linking (for the xSDS method) or directly into nuclei preparation using nuclei isolation buffer (NIB, 10mM TrisHCl pH 7.4, 10 mM NaCl, 3mM MgCl2, 0.1% Igepal, 1X protease inhibitors (Roche, Cat. 1187358001)). Cortical samples were cut with a sterile razor blade and resuspended in a chilled 5mL modified nuclei isolation buffer (NIB-HEPES, 20mM HEPES, 10mM NaCl, 3mM MgCl2, 0.1% Igepal, 1X protease inhibitors). Cells were given 5 minutes to equilibrate to the salt solution before 5 loose strokes in a dounce homogenize, another 5 minutes to equilibrate, another 5 loose strokes and 10 tight strokes. Nuclei were then spun in a pre-chilled 4C centrifuge for 5 minutes at 600xg.
Nucleosome Depletion: Detailed step-by-step protocol for nucleosome depletion and all subsequent steps can be found in the Supplementary Protocol. Nucleosome depletion and combinatorial indexing strategies

were performed similar to previously described, with some variations25. Lithium-assisted nucleosome depletion (LAND) was performed for generation of GM12878-only and Human/Mouse libraries. Prepared nuclei were pelleted and resuspended in NIB supplemented with 200 uL of 12.5 mM lithium 3,5-diiodosalicylic acid (Sigma, Cat D3635) for 5 minutes on ice before addition of 800 uL NIB and then taken directly into the combinatorial indexing protocol.
Cross-linking and SDS nucleosome depletion (xSDS): Cells were cross-linked by incubation in 10 mL of media with 1.5% formaldehyde (final conc. by vol.) and incubated at room temperature for 10 minutes with gentle agitation. Cross-linking was quenched with 800 uL 2.5 M glycine and incubated on ice for 5 minutes. Cells were then spun down, washed with ice-cold PBS, and resuspended in ice cold NIB for a 20 minute incubation on ice with gentle agitation. Cells were then pelleted, washed with 900 uL of 1X NEBuffer 2.1m and resuspended in 800 uL 1X NEBuffer 2.1 with 0.3% SDS (Sigma, Cat. L3771) and incubated at 42 ˚C with vigorous shaking for 30 minutes in a thermomixer (Eppendorf). 200 uL of 10% Triton-X was added to quench, and the solution was incubated at for another 30 minutes at 42 ˚C with vigorous shaking. Nuclei were then taken into the combinatorial indexing protocol. We were concerned that the crosslinking may affect the bisulfite conversion reaction; however, based on the methylation rates (particularly for those of nonCG methylation which were very low in concordance with expectations), we determined that not to be the case.
Nuclei were stained with 8uL of 5mg/mL DAPI (Thermo Fisher, Cat. D1306) and passed through a 35-um cell strainer. A 96-well plate was prepared with 10 uL of 1X TD buffer diluted with NIB in each well. Fluorescence-assisted nuclei sorting (FANS) was performed with a Sony SH800 flow sorter to sort 2,500 single nuclei into each well in fast sort mode (Supplementary Fig. 24). 4uL of 500 nM transposome, pre-loaded with cytosine-depleted, uniquely indexed, custom oligonucleotides was placed in each well (described in Supplementary table 1, transposomes assembled as described in Amini et. al. 2014, ref.17). Reactions were incubated at 55˚C for 20 minutes. All wells were then pooled and stained with DAPI as done for the first FANS sort. A second 96-well plate was prepared with each well containing digestion reagents as described by the manufacturer's protocol for the EZ-96 DNA Methylation MagPrep Kit (Zymo, Cat. D5040) at one-fifth the volumes (for 5 uL per well). 22 post-tagmentation nuclei from the pool of all reactions were sorted into each well using the single-cell sorting setting. Some wells were selected to receive only 10 nuclei, to allow for unmethylated controls.

The supplied flow-sorting strategy provided in supplementary fig. 24 reflects the xSDS preparation for the human/mouse split.

6. Identify the instrument used for data collection.

The Sony SH800 Flow Sorter was used for all sorting in this study.

7. Describe the software used to collect and analyze the flow cytometry data.

All collection and analysis of flow cytometry data was performed using Sony Cell Sorter Software (v2.1.1).

8. Describe the abundance of the relevant cell populations within post-sort fractions.

For the representative gating strategy figure supplied as supplementary fig. 24, 33.45% of events were classified as single nuclei for sorting as the first FANS procedure. For the second FANS procedure, 60.79% of events were found to be within the single-nuclei gate, reflecting the enrichment of single nuclei after the first sort and tagmentation.

9. Describe the gating strategy used.

Each preparation of a sci-MET library consisted of two FANS procedures which followed the same strategy to isolate pure single nuclei. First, events were gated on forward scatter-area by back scatter-area to remove small

debris (Gate A). Using that defined gate, events were then assessed by gating on DAPI staining of height by area. Singlet nuclei are expected to populate the diagonal (Gate B). Gate B was then purified to singlet nuclei by gating on DAPI-area by backscatter-height. Multiple subpopulations of events formed reflecting nuclei doublets and triplets. Gate C defined singlet nuclei and was used for the final gating for sorting. For the first FANS procedure, 2,000 nuclei were placed into each well using the "normal" sorting setting. For the second FANS procedure, we placed 22 or 10 nuclei into each well using the "single-cell" sorting setting.

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