nature research

Corresponding author(s): Keji Zhao

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 Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Bowtie 2 (v2.2.5) was used for sequencing reads mapping. MACS2 (v2.1.0) was used for ChIP-seq peaks calling. Cistrome SeqPos was used to do motif analysis. PIQ (v1.3) was used to predict TF binding sites in DHSs. Other data analysis, e.g. nucleosome spacing uniformness quantification and nucleosome occupancy score, was performed using custom codes and fully described in Methods. Custom codes for nucleosome spacing uniformness quantification and nucleosome occupancy score are available at https://github.com/binbinlai2012/scMNase.

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/reviewers upon request. 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 scMNase-seq data sets have been deposited in the Gene Expression Omnibus database with accession number GSE96688.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We generated scMNase-seq datasets for 48 NIH3T3 single cells, 198 mouse embryonic stem cells (mESCs), and 278 mouse naïve CD4 T cells. In addition, we generated three groups of NIH3T3 scMNase-seq libraries (10 single cells for each group) with different MNase concentrations for titration purpose, and 12 haploid mESC scMNase-seq libraries. Statistical tests were not used to pre-determine sample size. ~500 single cell libraries covering 3 cell types generates total 500 million high quality non-redundant sequencing reads used for determining nucleosome position, subnucleosome-sized particle position and nucleosome-nucleosome distance from the same cells or across cells, within or across cell types. This data set fully supports the findings described in the manuscript.
Data exclusions	All data were included for assessing or characterizing the scMNase-seq libraries. Single cells with low library size were excluded for the analysis of "clustering single cells based on nucleosome similarity around DHSs" and the analysis of "identifying subgroup of naive CD4 T cells or mESCs with depleted nucleosomes at de novo enhancers for differentiation". Details were described in the methods section.
Replication	We have three cell types and in total ~500 single cells to support our findings related to the single cell nucleosome profiling. In all three cell types, we observed that nucleosome spacing uniformness is smaller in silent chromatin region than active chromatin region. In both mESCs and naive CD4 T cells, which are undifferentiated cells, we found a subgroup of primed cells show nucleosome depletion at downstream lineage specific de novo enhancers. For the analysis of bimodal nucleosome spacing distribution at DHSs and its link to heterogeneity in chromatin accessibility and gene expression, we only used NIH3T3 cells because (1) only NIH3T3 cell libraries have sufficient nucleosome coverage to provide sufficient nucleosome pairs around DHSs, and (2) only NIH3T3 cells have published scDNase-seq and single cell Drop-seq data sets available for this integrative analysis.
Randomization	For each cell type, single cells were FACS sorted from similar cell samples. There was no randomization in this in vitro experiment.
Blinding	This study is in vitro experiment and no blinding is required. The investigators were not blinded during data collection and analysis.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

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)	NIH3T3 cells and mESCs were obtained from ATCC (ATCC® CRL-1658™ and ATCC® SCRC-1036™).
Authentication	In addition to the authentic commercial source, our own data (RNA-seq and MNase-seq) from these cell lines validate the identity of these cells.
Mycoplasma contamination	The cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.