nature research

Corresponding author(s): Bing Ni and 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. SICER(v1.1) was used for ChIP-seq peaks calling. Homer(v4.10.1) was used calculate the TSS density profile. SC3 was used for clustering cells. Seurat was used for pre-processing scRNA-seq data. ScImpute was used for imputing zeros. ChIPseeker was used for annotating H3K27me3 and H3K4me3 peaks. ChromVAR was used to identify enriched TFs. Other data analysis, e.g. removing batch effect and construction of co-methylation network, was performed using custom codes and fully described in Methods. Matlab and R were used for writing scripts. Custom codes for generating every steps of the analysis are available at https://github.com/wailimku/scChIC-seq.git

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 scChIC-seq data sets were deposited in the Gene Expression Omnibus database with accession number GSE105012.

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 profiled H3K4me3/H3K27me3 marks in the single cell level using scChIC-seq in 285/106 single human white blood cells. The sample size selected here was similar to a previous study (titled "Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing") which is sufficient to validate the quality of the technique. In addition, we generated several other libraries using scChIC-seq in small cell numbers 1) H3K4me3 libraries using 3000, 1000, 300, 100 NIH-3T3 cells using H3K4me3 Ab-MNase 2) H3K4me3 libraries using 3000 NIH-3T3 cells, 3000 mES cells, 3000 mouse naive T cells (using H3K4me3 Ab-MNase) 3) H3K4me3 library using H3K4me3 Ab-MNase 3000 NIH-3T3 cells 4) H3K4me3 library using H3K4me3 Ab+pA-MNase 3000 NIH-3T3 cells 5) H3K4me3 libraries using H3K4me3 Ab-MNase 3000 NIH-3T3 cells under different washing conditions:1) 200mM salt wash 2) 300mM salt wash 3) 400 mM salt wash 4) RIPA+150 mM salt wash 6) H3K4me3 libraries using H3K4me3 Ab+PA-MNase 3000 NIH-3T3 cells under different washing conditions:1) 200mM salt wash 2) 300mM salt wash 3) 400 mM salt wash 4) RIPA+150 mM salt wash 7)H327me3 library using H3K27me3 Ab+pA-MNase 3000 NIH-3T3 cells The information of all libraries including libraries sizes, mapping statistics are included in Supplemental table S2. The information about the data set that has been used to generate figures is also included inSupplemental table S2.
Data exclusions	All data were included for assessing or characterizing the scChIC-seq libraries. Among the 285 H3K4me3 single cells, 4 of them are considered to be outliers since their log-library sizes are more than 3 median absolute deviations (MAD), and was larger than the averaged log-library sizes. Then the 281 single cells are pooled to call peaks and visualized in the genome browser. A similar method was used in removing outliers in the scRNA-seq data (e.g., in the study of the paper with title "A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor") To further select the informative cells,the number of reads and the generated TSS density profile of each single cell were used. A single cell was considered to be informative if 1) The Pearson correlation between the average TSS profile over all cells and the TSS profile of the single cell was higher than a cutoff, and the cutoff was set to be 0.9. 2) For each TSS density profile, we defined a quantity referring to the signal-to-noise ratio, in which the ratio was equal to the ratio between the maximum and the minimum values of the TSS density profile. The signal-to-noise ratio of an informative single cell should be larger than the average signal-to-noise ratio (obtained from the average TSS density profile) multiplied by a correction factor (was set to 0.7). 3) The number of reads in an information single cell was between 4000 and 100000. Finally, 242 single cells were selected from the 281 cells for downstream analysis. For H3K27me3 single cell, due to the smaller number of cells and the different nature of antibody, we used a different strategy to exclude the cells. First we pool all 106 cells to call peaks and visualize them via the WasU genome browser. Then 22 single cells are excluded due to the low number of reads (<1000). Then the remaining 84 single cells are used for other analysis. All information about data exclusion is included in the Supplemental Methods.
Replication	For H3K4me3 profiling in the single cell level using scChIC-seq, the mixed population was separated using computational clustering analysis. We focused on the identified Monocytes and T cells. In both identified cell types, we observed the same biological findings. 1) Co-methylation between the H3K4me3 level of peaks is positively related to co-expression of gene expression level of genes that are

associated to the H3k4me3 peaks.

2) Variability of H3K4me3 level at peaks is positively related to the heterogeneity in gene expression for genes that associated to the H3k4me3 peaks.

3) Highly co-methylated peaks and highly variable peaks (their definitions were described in Supplemental Methods) show distinct gene expression patterns, and this distinct gene expression pattern is cell type specific.

4) The enriched TFs in highly co-methylated peaks and highly variable peaks, obtained by the software chromVAR, show distinct biological functions supported by using the gene expression data of TF and published literature findings.

For 3000 cells H3K4me3 profiling, we have more than one replicate of H3K4me3 Ab-MNase scChIC-seq library using 3000 NIH-3T3 cells to confirm the specificity of our method. For other small cell number scChIC-seq, we have used 1000 cells, 300 cells, and 100 cells to show that scChIC-seq works for profiling H3K4me3 marker using small cell number.

For application of using scChIC-seq to profileH3K4me3 marks in different cell types, we have used two different cell types (mES and Naive CD4 T cells) to confirm that scChIC-seq is workable for different cell types. Also, we have pooled single cell data to support that scChIC-seq works in human white blood cells.

For optimizing washing conditions of scChIC-seq (3000 cells NIH-3T3 cells) using H3K4me3 Ab-MNase, we found that the optimal condition was "RIPA+150 mM salt wash". There were more than one replicates for this condition. The data of one replicate with lower sequencing depth was shown in Extended Data Fig. 4 while the data of another replicate with higher sequencing depth was shown in Extended Data Fig. 6. However, we do not have replicates for other non-optimal washing conditions since in our experiments, we observe that the results of "RIPA +150 mM salt wash" are highly reproducible while the other other conditions are not always workable. This reproducibility of the results is also taken account into searching the optimal conditions.

For optimizing washing conditions of scChIC-seq (3000 cells NIH-3T3 cells) using H3K4me3 Ab++PA-MNase, we found that the optimal condition was "400 mM salt wash". There were more than one replicates for this condition. The data of one replicate with lower sequencing depth was shown in Extended Data Fig. 5 while the data of another replicate with higher sequencing depth was shown in Extended Data Fig. 6. However, we do not have replicates for other non-optimal washing conditions since in our experiments, we observe that the results of "400 mM salt wash" are highly reproducible while the other other conditions are not always workable. This reproducibility of the results is also taken account into searching the optimal conditions.

For application of using scChIC-seq to profile H3K27me3 marks in the single human WBC, we have in total 106 single cells to confirm that scChIC-seq is workable in H3K27me3 at the single cell level. Also we have NIH-3T3 cells and human WBC to confirm that our protocol can reproducibly profile H3K27me3 using small cell number in different cell types.

Randomization

For human white blood cells, single cells were randomly FACS sorted.

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

Antibodies

Antibodies used

Histone H3 trimethyl Lys4 antibody (Millipore, cat no. 17-614) and histone H3 trimethyl Lys27 antibody (Diagenode, cat no. pAb-069-050) were used. The dilution conditions were described in the supplemental protocols.

Validation

The qPCR is performed before library preparation for scChIC-seq for both H3K4me3 Ab and H3K27me3 Ab.

H3K4me3 Ab is applied to scChIC-seq for profiling H3K4me3 marks in 3000 NIH-3T3 cells, and the specificity of the H3K4me3 Ab is compared to published H3K4me3 ChIP-seq data (GSM1246686) in NIH-3T3 cells. Our data reveals that H3K4me3 profiling from the two methods are highly consistent.

H3K27me3 Ab is applied to scChIC-seq for profiling H3K27me3 marks in 3000 NIH-3T3 cells, and the specificity of the H3K27me3 Ab is compared to published H3K27me3 ChIP-seq data (GSM1246690) in NIH-3T3 cells. Our data reveals that H3K27me3 profiling from the two methods are highly consistent.

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 (scChIC-seq using 3000 cells and ChIP-seq) from these cell lines validate the identity of these cells. ChIP-seq is the method for profiling chromatin modification traditionally.
Mycoplasma contamination	The cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Human research participants

Policy information about studies involving human research participants

Population characteristics	N/A
Recruitment	These cells were obtained from NIH blood bank and were donated by unknown healthy donors.

Flow Cytometry

Plots

Confirm that:

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

☐ 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).

☐ All plots are contour plots with outliers or pseudocolor plots.

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

Methodology

Sample preparation	After the cells were bound with the Ab+PA-MNase or Ab-MNase, the cells were re-suspended in 200 µL Sorting Buffer (10 mM Tris-Cl buffer [pH7.5], 1 mM EDTA, 10 mM sodium chloride, 0.1% [v/v] triton X-100). Then, the cells were filtered to generate single-cell suspension using the 40-µm cell strainer (Thermo Fisher Scientific) and then sorted to collect one single cell per tube in the Sorting Buffer using BD FACSAria III cell sorter (BD Biosciences, San Jose, CA, USA). The single cell was collected in a PCR tube containing 20 µL Sorting Buffer.
Instrument	BD FACSAria III
Software	BD FACS DIVA 8.0
Cell population abundance	HUMAN WHITE BLOOD CELLS
Gating strategy	Using physical properties of cells, FSC and SSC parameters were adjusted to focus WBCs on FSC and SSC axis on linear scales. After excluding debris, single cells were selected by excluding doublet cells having higher width parameters on FSC and SSC scales. These single cells were sorted, using single cell sorting criteria for sort, in to PCR tubes containing 20 ul of cell sorting buffer and were subjected to Sequencing as described in text.

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