nature research

Corresponding author(s): Steven Henikoff

Last updated by author(s): Feb 3 2023

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.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

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 size distributions and molar concentration of libraries were determined using an Agilent 4200 TapeStation. Up to 48 barcoded CUT&RUN

Data analysis: Bedtools; Python Packages Used: Numpy, Pandas, Seaborn, Matplotlib, umap; R version 4.0.0, R libraries used: ggplot2, densityClust, archR;

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

All primary sequencing data have been deposited as paried-end fastq files in Gene Expression Omnibus under the accession code GSEXXXXX. Will update during Review

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

Life sciences study design

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

Sample size	In this study we collected single cell chromatin profiling data for comparative analysis from (1) a mixture of human cell lines (K562, H1, ML-2,
Data exclusions	Sequencing reads mapping to the mitochondrial genome were removed from all datasets. This was pre-established and is standard practice in
Replication	Each single cell experiment captured >10K cells, and the PBMC experiment was performed in biological duplicate with one replicate used for a
Randomization	n/a. The data and analysis for this study is objective and not prone to influence by the researchers bias.
Blinding	n/a. The data and analysis for this study is objective and not prone to influence by researchers bias.

Behavioural & social sciences study design

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

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing	
Data exclusions	
Non-participation	
Randomization	

Ecological, evolutionary & environmental sciences study design

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

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing and spatial scale	
Data exclusions	
Reproducibility	
Randomization	
Blinding	

Did the study involve field work? ○ Yes ○ No

Field work, collection and transport

Field conditions	
Location	
Access & import/export	
Disturbance	

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

Antibodies

Antibodies used: rabbit oligoclonal anti-H3K4me1 (1:10, Thermo Cat# 710795).

Validation: All antibodies are commercially available, and have been verified by Western blotting or by peptide ELISA described on the

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s): Human K562 cells were purchased from ATCC (Manassas, VA, Cat# CCL-243).

Authentication: All the cell lines used in this study are regularly submitted for karyotyping by the Fred Hutchinson Cancer Center Core Facilities.

Mycoplasma contamination: All cell lines were confirmed as mycoplasma negative on a tri-monthly basis.

Commonly misidentified lines (See ICLAC register): No commonly misidentified lines were used in this study.

Palaeontology and Archaeology

Specimen provenance:

Specimen deposition:

Dating methods:

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight:

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals:

Wild animals:

Field-collected samples:

Ethics oversight:

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics: All patient samples were obtained from Dr. Scott Furlan in accordance with the Declaration of Helsinki after written consent

Recruitment: Patients did not receive compensation for participation in this study.

Ethics oversight: The studies were overseen by the Institutional Review Boards at Fred Hutchinson Cancer Research Center (IR Protocol # XXXX).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration:

Study protocol	
Data collection	
Outcomes	

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
○	○	Public health
○	○	National security
○	○	Crops and/or livestock
○	○	Ecosystems
○	○	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
○	○	Demonstrate how to render a vaccine ineffective
○	○	Confer resistance to therapeutically useful antibiotics or antiviral agents
○	○	Enhance the virulence of a pathogen or render a nonpathogen virulent
○	○	Increase transmissibility of a pathogen
○	○	Alter the host range of a pathogen
○	○	Enable evasion of diagnostic/detection modalities
○	○	Enable the weaponization of a biological agent or toxin
○	○	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☑ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☑ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication*	All primary sequencing data have been deposited as paired-end fastq files in Gene Expression Omnibus under the accession code GSEXXXX (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1XXXXX). Will update during Review
Files in database submission	DJ_Hs_BY_K27me3_KHRM3_5cpw_220802 (BY_K27me3_KHRM3_5cpw)
Genome browser session (e.g. UCSC)	

Methodology

Replicates	For the human and mouse mixing experiment only one replicate was performed. For the analysis of human PBMCs we performed 2
Sequencing depth	All Experiments were paired-end. Sequencing depths and sampling is reported in the manuscript.
Antibodies	All antibodies and sources are provided in the Methods section.
Peak calling parameters	This manuscript does not include any Peak Calling
Data quality	Data quality assessment is the topic of this manuscript, and is reported.
Software	A link to custom code will be provided during review.

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

Instrument

Software

Cell population abundance

Gating strategy

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

Magnetic resonance imaging

Experimental design

Design type

Design specifications

Behavioral performance measures

Acquisition

Imaging type(s)

Field strength

Sequence & imaging parameters

Area of acquisition

Diffusion MRI ○ Used ○ Not used

Preprocessing

Preprocessing software

Normalization

Normalization template

Noise and artifact removal

Volume censoring

Statistical modeling & inference

Model type and settings

Effect(s) tested

Specify type of analysis: ○ Whole brain ○ ROI-based ○ Both

Statistic type for inference (See Eklund et al. 2016)

Correction

Models & analysis

n/a	Involved in the study
○	Functional and/or effective connectivity
○	Graph analysis
○	Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Graph analysis

Multivariate modeling and predictive analysis