nature portfolio

Corresponding author(s): PG Maass

Last updated by author(s): 2024_10_01

Reporting Summary

Nature Portfolio 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 Portfolio 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	Data was downloaded from GEO, SRA (SRA tool kit), and NIMH (package: synapserutils / Synapse API client).
Data analysis	Mapping Hi-C reads with bwa version 0.7.17-r1198; mapping statistics with Samtools version 1.5; filtering for valid Hi-C and Omni-C alignments by using Pairtools version 0.3.0; Indexing of reads with Pairix version 0.3.7; Contact matrix normalization with Cooler 0.8.11; Python version 3.9.2_pbalin; Perl version 5.28.0; Python version 2.79; Bedtools version 2.24.0; GO term enrichments https://metascape.org/gp/index.html; R version 4.2.1; Genome Topology maps with Gephi 0.10.1 (https://gephi.org); 3D model of community detection results with Helios Web 0.7.9 (https://heliosweb.io); statistics with R version 4.2.1 and Python version 3.8.10

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Sequencing and all Signature data are deposited at GEO (accession numbers: GSE217358, GSM7757606- GSM7757610, GSE242273). Source data are provided in supplementary data files S1-S13. The code of Signature (LWPR & Community detection), its documentation, and a demo of how to utilize Signature, as well as computational analysis required for graphical visualization are available at https://github.com/MaassLab/Signature.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	We considered sex (female and male datasets) in our study. We determined the sex of Hi-C datasets by logistic regression (see Methods), and provide our training dataset to determine sex here: https://github.com/MaassLab/Signature
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	SickKids Research Ethics Board

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

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 for Hi-C datasets. Sixty-two diploid Hi-C datasets with a total of 161 billion reads provide sufficient sample power for our measurements. For FISH, we analyzed > 100 nuclei in each of the three different cell lines, which represents a 'gold standard' in cytogenetics. Oligopainting was performed in two independent experiments, each with > 300 nuclei. MSCs were differentiated in two independent replicates across chondrogenesis which is sufficient to determine variation.
Data exclusions	No data was excluded.
Replication	We used LWPR across all 62 datasets twice to ensure invariability of Signature (yielding 100 % reproducibility). We repeated community detection ten times and were able to replicate results. Oligopainting was repeated twice and replicated results. The chondrogenic differentiation was done in two biological replicates, of which three independent technical Omni-C replicates were prepared. Upon inter-replicate analysis, we pooled the data. The replication attempts were all successful.
Randomization	Samples were split into two main groups dependent on their sex (determined by logistic repression). Single datasets were assigned to a 'tissue group' based on their physiological makeup. We did not control for co-variates, because our study design did not require to do this.
Blinding	Blinding was not relevant for our study because we analyzed only one condition for Hi-C parameters, oligopainting and FISH and did not compare between groups. To compare our observations with random sampling, we applied various randomization tests (see Methods).

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
☒	☐ 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 and Sex and Gender in Research

Cell line source(s)	ATCC
Authentication	Cell lines were not authenticated.
Mycoplasma contamination	Mycoplasma testing was negative for all cells. Testing was performed every eight weeks with LookOut Mycoplasma PCR detection kit (Sigma-Aldrich).
Commonly misidentified lines (See ICLAC register)	no commonly misidentified cell lines were used in the study