nature portfolio

Corresponding author(s): Ilyas Singec

Last updated by author(s): Jul 1, 2022

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	Whole exome were collected using the Illumina HiSeq 500 platform.
Data analysis	Statistical tests were performed in Graph Pad Prism 9.0.0 and R 3.6.0. A customized fully automated digital image analysis algorithm segmented each cell as an individual object and measured the shape. Cell roundness is quantified per cell as deviation from perfect mathematical roundness. Comet assay data was analyzed with Trevigen Comet Analysis Software Version 1.3d. BD FACSAriaIII Fusion (BD Biosciences) used for single cell deposition was controlled with FACSDiva v9.0. Whole exome sequencing analysis used the following software packages: Burrows-Wheeler Aligner (BWA) 0.7.17, the Genome Analysis Toolkit (GATK) 3.8-1, ANNOVAR 2018-06-16, R 3.6.0, UCSC Genome Browser (build GRCh38 and GENCODE track v32).

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

The main data discussed in this protocol were generated as part of the studies published in the supporting primary research papers in refs. 21 and 22. Whole-exome files have been deposited to the Sequence Read Archive under BioProject PRJNA552890. Source data are provided with this paper.

Human research participants

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

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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 statistical method was used to predetermine the sample size. Sample sizes for experiments were estimated based on the experience of the investigators and similar studies. Three or more independent replicates were performed for most experiments.
Data exclusions	No data are excluded.
Replication	The experimental findings were independently validated by multiple team members.
Randomization	Cells were imaged randomly to measure roundness after treatment. All cells within a given field of view were analyzed, therefore there was no requirement for randomization
Blinding	Investigators were not blinded during collection or analysis because each assay is conducted by one investigator from the beginning to the end. Instead, the findings were confirmed by independent replications by multiple team members.

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

Antibodies

Antibodies used

Santa Cruz Biotechnology
Anti-OCT4(sc-5279)

R&D Systems
Anti-SOX2(AF2018)

Cell Signaling Technology
Anti-NANOG(4903S), Anti-SOX17(81778)

Novus
Anti-beta Tubulin(NB600-936)

Biolegend
Anti-PAX6(901301)

Thermo Fisher Scientific
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 568(A10037), Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488(A21206), Donkey anti-Goat IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 647(A32849)

Validation

Santa Cruz Biotechnology
Anti-OCT4(sc-5279) WB/IP/IF/IHC/FCM/ELISA www.scbt.com/p/oct-3-4-antibody-c-10

R&D Systems
Anti-SOX2(AF2018) WB/ELISA www.rndsystems.com/products/human-mouse-rat-sox2-antibody_af2018

Cell Signaling Technology
Anti-NANOG(4903S) WB/IHC/IF/F www.cellsignal.com/products/primary-antibodies/nanog-d73g4-xp-rabbit-mab/4903
Anti-SOX17(81778) WB/IF www.cellsignal.com/products/primary-antibodies/sox17-d1t8m-rabbit-mab/81778

Novus
Anti-beta Tubulin(NB600-936) WB/Simple Western/ICC/IF/IHC/ICH-P www.novusbio.com/products/beta-tubulin-antibody_nb600-936

Biolegend
Anti-PAX6(901301) WB/IHC-P/IHC-F www.biolegend.com/fr-fr/products/purified-anti-pax-6-antibody-11511

Thermo Fisher Scientific
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 568(A10037) IF www.thermofisher.com/antibody/product/Donkey-anti-Mouse-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A10037
Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488(A21206) IF www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21206

Donkey anti-Goat IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 647(A32849) IF www.thermofisher.com/antibody/product/Donkey-anti-Goat-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A32849

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

All human ESCs lines (WAO1, WA07, WA09) were purchased from WiCell, and iPSC lines (LiPSC-GRl.1 and NCRM-5) were generated by the NIH Regenerative Medicine Program. IPSC lines (GM25256). JHU078i was purchased from WiCell. HUES53 was purchased from Harvard Stem Cell Institute. ESl-035 was purchased from ESIBIO.

Authentication	Cells were not further authenticated in our lab.
Mycoplasma contamination	All cells lines tested negative for mycoplasma contanimation.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

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	Cells were harvested for FACS using TrypLE. The cell suspension was filtered through a 35-µm mesh filter into polystyrene round-bottomed tubes.
Instrument	FACSArialll Fusion (BD Biosciences)
Software	FACSDiva v9.0
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	Forward scatter and side scatter were used to exclude doublets.

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