nature portfolio

Corresponding author(s): Nozomu Yachie (The Uniersity of British Columbia)

Last updated by author(s): Mar 5, 2025

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	We used software tools provided by the manufacturers for the analyses conducted with the TECAN Infinite 200 PRO plate reader, InCellAnalyzer 6000, Illumina MiSeq, HiSeq 2500, and NovaSeq 6000.
Data analysis	We used QUEEN v1.2, bartender-1.1, symspellpy v6.7, Dropseq Tools v2.5.1, Picard v2.18.14, STAR v2.7, Seurat v3, kneed v0.8.1, starcode v1.4, cutadapt v4.1, PySanger v1, NCBI BLAST+ v2.6.0, bcl2fastq2 v2.20.0, QuantStudio v1.4.1, ImageMagick v7.1.0-20, Fiji v1.0, flowWorkspace v0.5.40, flowCore v1.11.20, CytoExploreR v1.1.00, FlowCytometryTools v0.5.0, R v4.2.0, R v4.3.1, FlowJo v10.7.2, STAR v2.7.10a, HTSeq v2.0.2, DESeq2 v1.34.0, IGV v2.16.2, deepTools v3.5.4, pheatmap v1.0.12, GSEApy v1.1.1, networkx v3.2.1, Cytoscape v3.10.1, Trim Galore v0.6.10, Bismark v0.24.1, bedGraphToBigWig v2.10, methylKit v0.9.7, pyBigWig v0.3.22, and Tecan i-control v1.10.4.0. We also shared all the custom scripts developed in this study in the GitHub repository (https://github.com/yachielab/CloneSelect_v1).

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 information on all oligonucleotides, plasmids, and Illumina indices used for multiplexing sequencing libraries is available in Supplementary Tables 2-3. Most of the plasmids constructed in this study will be available at Addgene (depositor: Nozomu Yachie https://www.addgene.org/depositing/82198/), and the other plasmids can be requested. The barcode count data obtained from this study is provided in Supplementary Table 1. All high-throughput sequencing datasets acquired in this study are available in the NCBI Sequence Read Archive (PRJNA901977). High-throughput sequencing data analysis was performed with NCBI hg38 (human) and mm10 (mouse) reference genome.

Human research participants

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

Reporting on sex and gender	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

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	We selected the sample sizes based on the research community standard and our previous experiences. The sample sizes are displayed in the figure panel or figure legends. The sample size values are vary depending on the experiment.
Data exclusions	For the mESC clone isolation experiment (Fig. 3j), Clone 153 was excluded from the analysis because we could not obtain enough input cells for the high-throughput amplicon sequencing (we could sort only 26 cells).
Replication	The replicate assays were all performed independently as separate experimental batches. We provided exact sample numbers for each figure or legend. All attempts at replication were sucessfull.
Randomization	Randomization is not applicable to those kind of experiments. We did not use subjective quantification.
Blinding	Not applicable. The blinding is not relevant to this study because it is not subjective trial. All results are objective description of our experimentations.

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

- PE mouse monoclonal anti-SUSD2 (clone W5C5), BioLegend Cat# 327406, 1:200 dilution
- APC mouse monoclonal anti-CD24 (clone eBioSN3 (SN3 A5-2H10)), Thermo Fisher Scientific Cat# 17-0247-42, 1:200 dilution
- PE mouse monoclonal anti-CD249 (ENPEP) (clone 2D3/APA), BD Cat# 564533, 1:200 dilution
- Biotin human monoclonal TROP2 Antibody (clone REA916), Miltenyi Biotec Cat# 130-115-054, 1:200 dilution
- Streptavidin-APC, BioLegend Cat# 405207, 1:1000 dilution

Validation

Validation statements for the antibodies used in this study are available on the manufacturers' websites as follows:
- SUSD2 (327406): https://www.biolegend.com/ja-jp/products/pe-anti-human-susd2-antibody-4354
- CD24 (17-0247-42): https://www.thermofisher.com/antibody/product/CD24-Antibody-clone-eBioSN3-SN3-A5-2H10-Monoclonal/17-0247-42
- ENPEP (564533): https://www.bdbiosciences.com/ja-jp/products/reagents/flow-cytometry-reagents/research-reagents/single-colorantibodies-ruo/pe-mouse-anti-human-cd249.564533
- TROP2 (130-115-054): https://www.biocompare.com/9776-Antibodies/11622540-Anti-TROP2-Biotin-Antibody/
- Streptavidin-APC (405207): https://www.biolegend.com/ja-jp/products/apc-streptavidin-1470?GroupID=GROUP23

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

HEK293Ta (Genecopia), HEK293T Lenti-X (Takara), mouse ESC, and Human ESC lines H1 (WiCell Research Institute, Madison, WI, USA).

Authentication

The HEK293Ta cell line (Genecopia) and HEK293T Lenti-X (Takara) were authenticated by the vendor (Genecopia or Takara). The human ESC line H1 has been verified by original sources and also authenticated in-house through observations of colony morphology, RT-qPCRs, immunostaining, RNA-seq, and/or in vitro differentiation.

Mycoplasma contamination

Routinely tested negative by PCR.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell line was 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

All of the detailed sample preparation protocols are provided in Materials and Methods. The following is the same information in sample preparations.

HEK293T: Cells were detached with 0.25% w/v Trypsin-EDTA (Wako #201-18841), incubated at 37C for 5 min, collected into a 1.5-mL tube or a 96-well round-bottom plate, and centrifuged at 1,000 rpm at room temperature for 5 min. After aspirating the supernatant, cell pellets were gently resuspended in 150–500 µL of ice-cold FACS buffer (2% FBS in 1x PBS). Samples were immediately placed on ice until flow cytometry analysis.

mESC: Three days after transduction with a query gRNA, each cell sample was expanded in a 10-cm cell culture dish. Cells were detached with 0.25% w/v Trypsin-EDTA (Gibco #25200072), incubated at 37C for 5 min, collected into a 1.5-mL tube, and centrifuged at 1,000 rpm at room temperature for 5 min. The cells were then resuspended to approximately 1 × 106 cells

	in PBS containing 2% FBS and transferred to a 5-mL polystyrene round-bottom tube (Falcon #352054). The cell suspension was immediately placed on ice until sorting. hPSC: Cell samples were washed with 1x D-PBS (-) and detached using 2 mL of Accutase (SIGMA #A6964-500ML) to create a single-cell suspension. Cells were resuspended in FACS buffer, composed of 450 mL MilliQ water, 50 mL 10× HBSS (no calcium, no magnesium, no phenol red) (Invitrogen #14185052), and 5 g Bovine Serum Albumin (Sigma #A2153-100G), and kept on ice for 30 minutes.
Instrument	We used the FACSVerse Cell Analyzer (BD Biosciences) or the CytoFLEX Flow Cytometer (Beckman Coulter) for flow cytometry analysis and employed the FACSJazz (BD Biosciences) or MoFlo Astrios EQ Cell Sorter (Beckman Coulter) for flow cytometry cell sorting.
Software	We used software tools provided by the manufacturers and exported raw FSC files for downstream data analysis and visualization. The detailed flow cytometry data analysis is described in the Materials and Methods section. We also included the custom scripts used for analyzing and visualizing flow cytometry data in the GitHub repository (https://github.com/yachielab/CloneSelect_v1/tree/main/FACS).
Cell population abundance	Cell population sizes and the counts of positive vells varied across the isolation experiments. For the GFP reporter assays, at least 10-20K cells were analyzed to derive GFP positives and negatives. We included the information regarding cell population abundance in the manuscript.
Gating strategy	We provided the gating strategy for each cell lines and experiments in the Supplementary Information.

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