nature portfolio

Corresponding author(s): Elisabetta Ferretti

Last updated by author(s): May 4, 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	No software was used to collect data
Data analysis	R/Bioconductor (v4.0.5) (https://www.R-project.org), bcl2fastq (v2.19.1), umis (v1.0.3), Cutadapt (v1.18), HISAT2 (v2.1.0), Subread (v1.5.3), UMI_tools (v1.0.0), Seurat (v3.1), Monocle (v2.16.0), STAR (v2.5.3a), FastQC (v0.11.7), MultiQC (v1.7), DESeq2 (v1.24.0), Panther (v15.0), Bowtie2 (v2.3.4.1), bedtools (v2.27.1), Samtools (v1.7 and v1.9), Picard (v2.9.1), bamCoverage (v3.2.0), MACS (v2.1.1.20160309), deepTools (v3.2.0), Homer (v4.7.2), UCSC genome browser (https://genome.ucsc.edu/index.html), GraphPad Prism (v8.4.3), Imaris (v9.3.0), rGreat (v4.0.4). For single-cell RNA-seq (scRNA-seq) experiments, R1 and R2 fastq files were generated using bcl2fastq (v2.19.1), and the pooling and well information were extracted from the sequence using umis (v1.0.3) [https://github.com/vals/umis] into a unique fastq file. The reads were then filtered based on the pooling barcodes with 1 mismatch allowed. The poly-Ts at the end of the reads were trimmed using Cutadapt (v1.18). The reads were mapped to the mouse genome (GRCm38/mm10 together with ERCC92) using HISAT2 (v2.1.0), the alignments were processed with Samtools (v1.7), and the reads were counted with featureCounts (Subread (v1.5.3)) using Ensembl v93, and the umis using UMI_tools (v1.0.0). Expression data were analysed using Seurat (v3.1). Data filtering, normalisation and scaling were performed using the standard pre-processing workflow, as described in Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R., Nat Biotechnol 36, 411-420 (2018). Integration of different datasets was performed as described in Stuart et al., Cell 177, 1888-1902 (2019). Spiked-in EpiSCs were used as a reference to correct the batch-effect between integrated datasets. Marker genes of each cell cluster were outputted for GO-term analyses to define the cell type. The Monocle package (v2.16.0) was used to perform pseudotime analyses. For bulk RNA-seq experiments, Fastq files were aligned to the GRCm38/mm10 genome using STAR (v2.5.3a). Transcript expression levels were estimated with the --quantMode GeneCounts option and GRCm38p5.vM15 annotations. FastQC (v0.11.7) (http://

www.bioinformatics.babraham.ac.uk/projects/fastqc) was used for QC metrics, and MultiQC (v1.7) for reporting. Data analysis was then performed with R/Bioconductor (v4.0.5) (https://www.R-project.org). Differentially expressed genes were identified using DESeq2 (v1.24.0), which applies negative binomial generalised linear model fitting and Wald statistics on the count data. The results were filtered for Benjamini–Hochberg adjusted P≤0.05 and fold-change ≥ 1.5, as indicated in the figure legends.

For GO-term analyses, biological pathway enrichments were determined using Panther (v15.0). Statistical significance was assessed using Fisher's Exact test with Bonferroni correction for multiple testing (P≤0.05).

For ATAC-seq experiments, reads were trimmed with Cutadapt (v1.18) (Cutadapt -a CTGTCTCTTATA -A CTGTCTCTTATA -q 20 -m 5), mapped to the GRCm38/mm10 mouse reference genome with Bowtie2 (v2.3.4.1). Mitochondrial reads and reads overlapping blacklisted regions (i.e. regions that are considered artefacts of ATAC-seq and other chromatin assays in mm10, https://github.com/MayurDivate/GUAVA/blob/master/lib/blacklists/JDB_blacklist.mm10.bed and http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/mm10-mouse) were removed using bedtools (v2.27.1), intersect and only uniquely mapped reads were kept using Samtools (v1.9). Bam files were sorted, deduplicated with Picard (v2.9.1) (http://broadinstitute.github.io/picard/) and indexed. Bigwig files were generated with bamCoverage (v3.2.0) from deepTools and the RPGC option. MACS (v2.1.1.20160309) was used to call peaks.

For Whole Genome Bisulfite Sequencing experiments, adapter trimming, sequence alignment, and methylation calls were performed using Bismark (v0.18.1). Differential methylation analyses were performed using methylKit (v1.25.0), applying Fisher's exact test for comparing the fraction of methylated Cs in differentially probed samples, with cutoff for the absolute value of methylation percentage change (25%) and qvalue of differential methylation statistic (0.01). Differentially methylated Cs were annotated with rGREAT (v4.0.4) using single nearest gene approach and defined as differentially methylated regions near genes (DMGs). The output from Bismark was also reported into SeqMonk (v1.46.0; Babraham Bioinformatics) for parallel DSS-based methylation validation analyses and WGBS data visualization over key genes of interest.

Homer (v4.7.2) motif enrichment analysis was performed on bed files generated by MACS (v2.1.1.20160309) peak caller on the GRCm38/mm10 reference genome with default settings.

For single-cell spatial transcriptomics data, final image analysis was performed in ImageJ using genexyz Polylux tool plugin from Resolve BioSciences.

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

E-MTAB-13407 - ATAC-seq data; E-MTAB-13574 - Whole Genome Bisulfite Sequencing (WGBS) data; E-MTAB-13618 - RNAseq data; E-MTAB-13660 – single-cell RNAseq data. Published ChIP-seq datasets were used in the paper (E-MTAB-977530). Published ATAC-seq datasets were used to draw gene lists for lineages (E-MTAB-63376, E-MTAB-977630).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.* *Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.* *Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*

Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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 pre-determine sample size. All experiments were included with multiple biological replicates based on preliminary experiments and previous experiences: three or more to enable calculation of standard deviation and significance. Exceptions to this were for some ChIP-qPCR and RT-qPCR experiments where two samples were analysed for each condition. For experiments with mouse embryos, at least three embryos per genotype were used.
Data exclusions	No samples were excluded from analyses.
Replication	All differentiation experiments were performed using two or more independent mutant EpiSC clones, and all attempts at replication were successful. All IF experiments were performed at least three times independently, with similar results. All RT-qPCR, ChIP-qPCR, WB and EMSA results were tested and confirmed with at least two independent biological experiments.
Randomization	Samples were allocated to the experimental groups based on their genotype. Appropriate controls were included in each figure.
Blinding	No blinding method was applied as no subjective measurements were taken.

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
☒	☐ Plants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Sheep polyclonal anti-FOXC2, R&D Systems, Cat#AF6989, RRID:AB_10973139, Lot#CAFS011901A, 1:100 for IF of adherent cells and of embryos. Mouse monoclonal anti-Histone H3, abcam, Cat#ab10799, RRID:AB_470239, Lot#GR3238785-1, 1:30,000 for WB. Rabbit polyclonal anti-IgG, Cell Signaling Technology, Cat#2729, RRID:AB_1031062, Lot#9, 1:1,000 for ChIP. Rabbit polyclonal anti-PBX1, Cell Signaling Technology, Cat#4342, RRID:AB_2160295, Lot#2, 1:100 for IF of embryos and for ChIP, 1:1,000 for WB. Goat polyclonal anti-TBX6, R&D Systems, Cat#AF4744, RRID:AB_2200834, Lot#CAPT0217111, 1:50 for IF of adherent cells and of embryos. Donkey anti-Goat IgG Alexa Fluor 568, Thermo Fisher Scientific, Cat#A-11057, RRID:AB_2534104, Lot#17711491, 1:200 for IF of adherent cells and 1:500 for IF of embryos. Donkey anti-Rabbit IgG Alexa Fluor 488, Thermo Fisher Scientific, Cat#A-21206, RRID:AB_2535792, Lot#1674651, 1:200 for IF of adherent cells and 1:500 for IF of embryos. Rabbit anti-phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (D13.14.4E) XP® Cell Signaling Technology, 4370, 1:1,000 for WB.

Rabbit anti-phospho-MEK1/2 (Ser217/221) (41G9), Cell Signaling Technology, 9154, 1:1,000 for WB.
Rabbit anti-p44/42 MAPK (total Erk1/2) (137F5) ,Cell Signaling Technology,, 4695, 1:1,000 for WB.
Rabbit monoclonal anti-CLDN6, Cell Signaling Technology, Cat#62831, 1:100 for IF of adherent cells and of embryos.
Goat polyclonal anti-HAND1, R&D Systems, Cat#AF3168, 1:100 for IF of adherent cells and of embryos.
Goat polyclonal anti-ECHADERIN, R&D Systems, Cat#AF748, 1:1000 for IF of adherent cells.
Goat polyclonal anti-T-BRA, R&D Systems, Cat#AF2085, RRID:AB_2200235, Lot#KQP0616121, 1:100 for IF of embryos.
Rabbit monoclonal anti-LEF1, Cell Signaling Technology, Cat#2230, RRID:AB_823558, Lot#8, 1:100 for ChIP.
Rat monoclonal anti-NANOG, eBioscience, Cat#14-5761, 1:1000 for IF of embryos.
Goat polyclonal anti-SOX17, R&D Systems, Cat#NL1924R, 1:200 for IF of embryos.
Goat polyclonal anti-GATA6, R&D Systems, Cat#AF1700, 1:100 for IF of embryos.
CLDN6 Alexa 517 Fluor 488-conjugated antibody, R&D Systems, Cat#FAB3656G, 1:200 for FACS.
Donkey anti-Mouse IgG (H+L) HRP, Thermo Fisher Scientific, Cat#A-16011, RRID:AB_2534685, Lot#42−141−070714, 1:10,000 for WB.
Donkey anti-Rabbit IgG (H+L) HRP, Thermo Fisher Scientific, Cat#A-16023, RRID:AB_2534697, Lot#50−23−112315, 1:10,000 for WB.

Validation

Sheep polyclonal anti-FOXC2, https://www.rndsystems.com/cn/products/mouse-foxc2-antibody_af6989.
Mouse monoclonal anti-Histone H3, https://www.abcam.com/histone-h3-antibody-mabcam-10799-chip-grade-ab10799.html.
Rabbit monoclonal anti-LEF1, https://www.cellsignal.com/products/primary-antibodies/lef1-c12a5-rabbit-mab/2230.
Rabbit polyclonal anti-PBX1, https://www.cellsignal.com/products/primary-antibodies/pbx1-antibody/4342.
Goat polyclonal anti-TBX6, https://www.rndsystems.com/cn/products/human-tbx6-antibody_af4744.
Rabbit anti-phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204), https://www.cellsignal.com/products/primary-antibodies/phospho-p44-42-mapk-erk1-2-thr202-tyr204-d13-14-4e-xp-rabbit-mab/4370.
Rabbit anti-phospho-MEK1/2 (Ser217/221) (41G9), https://www.cellsignal.com/products/primary-antibodies/phospho-mek1-2-ser217-221-41g9-rabbit-mab/9154.
Rabbit anti-p44/42 MAPK (total Erk1/2) (137F5), https://www.cellsignal.com/products/primary-antibodies/p44-42-mapk-erk1-2-137f5-rabbit-mab/4695.
Rabbit monoclonal anti-CLDN6, https://www.cellsignal.com/products/primary-antibodies/claudin-6-e2s5m-rabbit-mab/62831.
Goat polyclonal anti-HAND1, https://www.rndsystems.com/cn/products/human-hand1-antibody_af3168,
Goat polyclonal anti-ECHADERIN, https://www.rndsystems.com/cn/products/human-mouse-e-cadherin-antibody_af748.
Goat polyclonal anti-T-BRA, https://www.rndsystems.com/cn/products/human-mouse-brachyury-antibody_af2085.
Rat monoclonal anti-NANOG, https://www.thermofisher.com/antibody/product/Nanog-Antibody-clone-eBioMLC-51-Monoclonal/14-5761-80.
Goat polyclonal anti-SOX17, https://www.rndsystems.com/cn/products/human-sox17-antibody_af1924.
Goat polyclonal anti-GATA6, https://www.rndsystems.com/cn/products/human-gata-6-antibody_af1700.
CLDN6 Alexa 517 Fluor 488-conjugated antibody, https://www.rndsystems.com/cn/products/human-claudin-6-alexa-fluor-488-conjugated-antibody-342927_fab3656g.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Epiblast stem cell (EpiSC) lines were derived from E6.5 129S6/SvEvTac
Authentication	None of the cell lines used were authenticated. Cells were functionally validated and sub-cloned.
Mycoplasma contamination	All cell lines used tested negative for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No misidentified cell lines were used.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Mouse, 129S6/SvEvTac and mixed C57Bl/6, male and female, E3.5, E4.5, E6.5, E9.0, E13.5.
Wild animals	The study did not involve wild animals.
Reporting on sex	The study did not apply to only one sex.
Field-collected samples	The study did not involve animals collected from the field.
Ethics oversight	All animal work was carried out in accordance with European legislation. All work was authorised by and carried out under the Project License 2017-15-0201-01255 issued by the Danish Regulatory Authority.

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

Plants

Seed stocks: N/A

Novel plant genotypes: N/A

Authentication: N/A

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.

E-MTAB-9775. Previously published on Mariani et al., Nature Communications 2021.

Files in database submission

ChIPseq_PBX1_EpiSC_rep1.R1.fq.gz
ChIPseq_PBX1_EpiSC_rep2.R1.fq.gz
ChIPseq_PBX1_EpiSC_Input.R1.fq.gz

Genome browser session
(e.g. UCSC)

It is suggested to view the data in a local genome browser (e.g. UCSC; https://genome.ucsc.edu/index.html), genome mm10.

Methodology

Replicates: *Describe the experimental replicates, specifying number, type and replicate agreement.*

Sequencing depth: *Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*

Antibodies: *Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Peak calling parameters: *Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*

Data quality: *Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*

Software: *Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

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: EpiSCs were seeded at ~6 x 103 cells/cm2 in 6-well cell-culture plates pre-coated with 15 µg/ml human Fibronectin. Cells were dissociated using Accutase for 3 min at 37°C and counted. 150 x 106 cells per condition were incubated with the CLDN6 Alexa Fluor 488-conjugated antibody (R&D Systems, FAB3656G, 1:200) and Ghost Dye™ Red 780 (Tonbo, 13-0865, 1:1,000)

for 20 min on ice, washed with ice-cold FACS buffer (10% (v/v) FBS in PBS) and resuspended into 1 ml. Cells were sorted into MACSQuant Tyto Running Buffer using the MACSQuant Tyto sorter (Miltenyi Biotec) or into ice-cold FACS buffer using the SONY SH800S cell sorter (Sony Biotechnology) at the DanStem Flow Cytometry Platform (University of Copenhagen, Copenhagen, Denmark).

For intracellular staining and analysis experiments, EpiSCs were seeded at ~6 x 103 cells/cm2 in 6-well cell-culture plates pre-coated with 15 µg/ml human Fibronectin. Cells were dissociated using Accutase for 3 min at 37°C and counted. 1 x 106 cells per condition were resuspended in 4% PFA and incubated on ice for 20 min in Eppendorf tubes. Fixed cells were then permeabilized in PBS containing 0.3% (v/v) Triton X-100 (PBST) for 10 min at RT and blocked in 3% (v/v) Donkey serum in PBST containing 1% (w/v) BSA (blocking solution) for 15 min at RT. Primary antibody incubation was done overnight at 4°C, and secondary antibody incubation for 1 h at RT. Antibody incubations were followed by three quick washes in PBST, followed by two washes in PBST for 15 min at RT. Multiple centrifuging steps were done at 700G or 1500G for 5 min at 4°C, removing the supernatant using vacuum suction. After the final incubation, Eppendorf tubes containing the cells were wrapped in aluminium foil and stored in FACS Buffer (10% (v/v) FBS in PBS) containing 1 µg/ml DAPI at 4°C.

Instrument

BD LSRFortessa™ Cell Analyzer for cell analysis. MACSQuant Tyto sorter (S/N: 10143
Cap version: 1.0.0, Miltenyi Biotec) or the SONY SH800S cell sorter for the sorting experiments. All reseeding experiments were done using the MACSQuant Tyto sorter (Miltenyi Biotec) sorter.

Software

MACSQuant® Tyto® Cell Sorter Software and BD FACSDiva 9.0.

Cell population abundance

After each round of sorting and reseeding on the MACS Tyto sorter, cell population abundance and purity were assessed. Post-sort purity checks were done to ensure 90%+ purity sorts. Raw sorting data can be provided on request.

Gating strategy

For sorting and reseeding experiments, only singlets were considered, followed by gating for viable cells based on Ghost Dye™ Red 780 (Tonbo, 13-0865) staining. Following that, gating strategy for CLDN6 low and high sorting is shown in Fig.1c.

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