nature research

Corresponding author(s): Fang Fang

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 Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒ An indication of 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 statistics 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
☐	☒ Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	no software was used.
Data analysis	For ChIP-seq analysis: Bowtie (1.1.2), MACS (1.4.1). For RNA-seq analysis: TopHat (2.1.1), Cufflink (2.2.1), StrandNGS. For Motif analysis: MEME-ChIP(4.12.0). Statistical analysis: GraphPad Prism 5 (5.03). Flow cytometry analysis: Flowjo (10.4.2)

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/reviewers upon request. 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 sequencing data that support the findings of this study have been deposited in NCBI's GEO under accession number GSE100639. Source data for Fig. 3g, h; 4a, d, e, g, h; 5d, e, g, h, I; 6a-c and Supplementary Fig. 2b; 3a, d, f; 4c-d; 6a, b; 7a, c, d, e, g, h is provided in Supplementary Table 2. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

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

Sample size	No statistical was used to predetermine sample size. The sample size was determined based on the previous publications and experimental designs that have similar objectives(Ramathal et al., Cell Rep., 2014; Ramathal et al., Sci. Rep., 2015; Durruthy et al., Hum. Mol. Genet., 2014; Kee et al., Nature, 2009). Three to more independent results were used to perform statistical analyses. If less, no statistics were performed from these samples. All source data required for statistical tests are indicated in Supplementary Table 2 (Statistics data source).
Data exclusions	Data were excluded when negative or positive controls were not working.
Replication	For each experiment, we performed at least two independent biological replicates and all attempts were successful.
Randomization	For the xenotransplantation studies, animals were randomly allocated into groups receiving various cell line injections. Randomization (formal or otherwise) was not relevant for other data included in the manuscript.
Blinding	Investigators were not blinded to group allocation. For the counting analysis for xenotransplantation, investigators were blinded to the samples.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☒	☐ Unique materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☐	☒ Research animals
☐	☒ Human research participants

Antibodies

Antibodies used	anti-OCT4 (sc-8628, Santa Cruz Biotechnology, Lot#: C2613); anti-PAX5 (sc-1974, Santa Cruz Biotechnology, Lot#: A3113); anti-PRDM1 (C14A4, Cell Signaling Technology, Lot:4); anti-DDX4 (R&D #AF2030, Lot: KPX0 4051); anti-GFP (Abcam; ab18373, Lot: GR254056-13); anti-DAZL (Novus; NB100-2437; Lot: P1); anti-CKIT (A4502; DAKO); anti-beta-ACTIN (8H10D10, Cell Signaling Technology, Lot:15) HRP-conjugated anti-rabbit IgG secondary antibody, Supplier: Santa Cruz Biotechnology, Cat.: sc-2004, Lot: H0913. HRP-conjugated anti-goat IgG secondary antibody, Supplier: Santa Cruz Biotechnology, Cat.: sc-2020, Lot: B0613. HRP-conjugated anti-mouse IgG secondary antibody, Supplier: Santa Cruz Biotechnology, Cat.: sc-2005.

Alexa Fluor® 488 AffiniPure Donkey Anti-Goat IgG (H+L), Jackson ImmunoResearch Laboratories, Code: 705-545-003
Alexa Fluor® 594 AffiniPure Donkey Anti-Rabbit IgG (H+L), Jackson ImmunoResearch Laboratories, Code: 711-585-152.
The commercial antibodies were validated based on the information on the manufacturers' instructions.
For ChIP and Co-IP assay, antibodies were used at 1:100.
For immunostaining, primary antibodies were used at 1:200 and secondary antibodies were used at 1:300.
For western blot, primary antibodies were used at 1:1000 and secondary antibodies were used at 1:5000.
anti-OCT4 (sc-8628, Santa Cruz Biotechnology, Lot#: C2613);
anti-PAX5 (sc-1974, Santa Cruz Biotechnology, Lot#: A3113);
anti-PRDM1 (C14A4, Cell Signaling Technology, Lot:4);
anti-DDX4 (R&D #AF2030, Lot: KPX0 4051);
anti-GFP (Abcam; ab18373, Lot: GR254056-13);
anti-DAZL (Novus; NB100-2437; Lot: P1);
anti-CKIT (A4502; DAKO);
anti-beta-ACTIN (8H10D10, Cell Signaling Technology, Lot:15)
HRP-conjugated anti-rabbit IgG secondary antibody, Supplier: Santa Cruz
Biotechnology, Cat.: sc-2004, Lot: H0913.
HRP-conjugated anti-goat IgG secondary antibody, Supplier: Santa Cruz
Biotechnology, Cat.: sc-2020, Lot: B0613.
HRP-conjugated anti-mouse IgG secondary antibody, Supplier: Santa Cruz
Biotechnology, Cat.: sc-2005.
Alexa Fluor® 488 AffiniPure Donkey Anti-Goat IgG (H+L), Jackson ImmunoResearch Laboratories, Code: 705-545-003
Alexa Fluor® 594 AffiniPure Donkey Anti-Rabbit IgG (H+L), Jackson ImmunoResearch Laboratories, Code: 711-585-152.
For ChIP and Co-IP assay, antibodies were used at 1:100.
For immunostaining, primary antibodies were used at 1:200 and secondary antibodies were used at 1:300.
For western blot, primary antibodies were used at 1:1000 and secondary antibodies were used at 1:5000.

Validation

The commercial antibodies were validated based on the information on the manufacturers' instructions.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

H1 hESCs (WiCell) (feeder free)
293T (ATCC)

Authentication

Immunostaining of cell type specific markers was performed for cell line authentication.

Mycoplasma contamination

All lines were tested negative for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used

Research animals

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

Animals/animal-derived materials

In this study, we used busulfan-treated, immune-deficient nude male mice (NCr nu/nu; Taconic). Mice are treated with busulfan (40 mg/kg) at 5-6 weeks of age and then transplanted 5-12 weeks post busulfan.

Human research participants

Policy information about studies involving human research participants

Population characteristics

In this study, we obtained second trimester human fetal testis tissue. Note that the protocol for tissue procurement and use was approved by the Institutional Review Board of Montana State University (RR-P031014-EX); all procedures were compliant with all relevant ethical regulations regarding human research with unidentifiable banked tissue.

Method-specific reporting

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ Magnetic resonance imaging

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.*	https://www.ncbi.nlm.nih.gov/geo/submission/update/?acc=GSE100639
Files in database submission	1M_H9_OCT4_ChIP_CGATGT-04-30-2013.fastq 2nd_OCT4_SC_GCCAAT-04-30-2013.fastq 2nd_Test_OCT4_Ab_ACAGTG-04-30-2013.fastq IM_Mix_OCT4_TGACCA-04-30-2013.fastq testes_prdm1_chip-11-26-2013.fastq PAX5-testis-130909_BRISCOE_0116_BD2CRUACXX_L6_CTTGTA_pf.fastq.gz input-11-26-2013.fastq 1M_H9_OCT4.sorted.dedup.bam 2nd_OCT4_SC.sorted.dedup.bam 2nd_Test_OCT4_Ab.sorted.dedup.bam IM_Mix_OCT4.sorted.dedup.bam PAX5_TESTES_BRISCOE_0116_BD2CRUACXX_L6_CTTGTA_pf.bam TESTES_PRDM1.bowtie2.srt.bam
Genome browser session (e.g. UCSC)	no longer applicable

Methodology

Replicates	We did two replicates for OCT4 ChIP in human fetal testis, for the other ChIP, we only provided one
Sequencing depth	single-end, 36bp
Antibodies	anti-OCT4 (sc-8628, Santa Cruz Biotechnology); anti-PAX5 (sc-1974, Santa Cruz Biotechnology); anti-PRDM1 (C14A4, Cell Signaling Technology);
Peak calling parameters	Effective genome size 2700000000.0 Tag size 49 Band width 300 Pvalue cutoff for peak detection 1e-05 Select the regions with MFOLD high-confidence enrichment ratio against background to build model 32 Parse xls files into into distinct interval files False Save shifted raw tag count at every bp into a wiggle file wig Extend tag from its middle point to a wigextend size fragment. -1 Resolution for saving wiggle files 10 Use fixed background lambda as local lambda for every peak region False 3 levels of regions around the peak region to calculate the maximum lambda as local lambda 1000,5000,10000 Build Model create_model Diagnosis report no_diag Perform the new peak detection method (futurefdr) False
Data quality	OCT4 ChIP in human fetal testis: 27061 PAX5 ChIP in human fetal testis: 22728 PRDM1 ChIP in human fetal testis: 48399 OCT4 ChIP in hESCs:14799
Software	Aligment by Bowtie-illumina/Galaxy Peak call by MACS-ChIP-seq/Galaxy BiwWig generated by WigtoBigWig/Galaxy Motif detection by MEMEChIP

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	hESCs were dissociated in 0.25% trypsin--EDTA (Gibco BRL) at 37ºC for 5 min and collected by centrifugation at 200g in an Eppendorf 5702 R centrifuge. Then the cells were passed through the 70uM strainers (BD Biosciences) to make sure they were

	digested as single cells before they were subject to the flow cytometry. Mouse testis cells were dissociated with 0.25% Trypsin–EDTA for 5 min at 37 °C. Dissociated cells were passed through the 40uM strainers (BD Biosciences) and then incubated in 1% BSA in PBS containing primary antibodies on ice for 20 min.
Instrument	BD FACSAria 2.0
Software	For data collection: BD FACSDiva v 6.1.2 software; For data analysis: Flowjo (10.4.2)
Cell population abundance	Sorted cells were directly placed into DNA extraction buffer. The abundance for these samples could not be assessed.
Gating strategy	After cells were selected in the FSC/SSC dot plot to remove debris, they were gated to exclude cellular aggregates in the FSC/FSC dot plot. Gates of GFP-FITC, mOrange-PE, or CKIT-PE cells were set and compared with a control sample with no detectable fluorochrome expression.

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