

# Epigenetic characterization of adult rhesus monkey spermatogonial stem cells identifies key regulators of stem cell homeostasis

Rui Bi<sup>1,2,3,†</sup>, Lin-Nuo Pan<sup>4,5,†</sup>, Hao Dai<sup>6,†</sup>, Chunli Sun<sup>1,†</sup>, Cong Li<sup>1,†</sup>, Hui-Juan Lin<sup>1,2,3</sup>, Lan-Ping Xie<sup>1,5</sup>, Huai-Xiao Ma<sup>1,2,3</sup>, Lin Li<sup>1,5</sup>, Heng Xie<sup>1,5</sup>, Kun Guo<sup>1</sup>, Chun-Hui Hou<sup>1,5</sup>, Yong-Gang Yao<sup>1,2,3,5,6,\*</sup>, Luo-Nan Chen<sup>4,7,\*</sup> and Ping Zheng<sup>1,2,3,5,6,\*</sup>

<sup>1</sup>State Key Laboratory of Genetic Evolution and Animal Models, Kunming Institute of Zoology, Chinese Academy of Sciences, No.17 Longxin Road, Kunming, Yunnan 650204, China

<sup>2</sup>Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, No.17 Longxin Road, Kunming, Yunnan 650204, China

<sup>3</sup>National Resource Center for Non-Human Primates, National Research Facility for Phenotypic & Genetic Analysis of Model Animals (Primate Facility), Kunming Institute of Zoology, Chinese Academy of Sciences, Baohua Road, Kunming 650107, China

<sup>4</sup>Key Laboratory of Systems Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, No. 320 Yue Yang Road, Shanghai 200031, China

<sup>5</sup>University of Chinese Academy of Sciences, No.1 Yanqihu East Rd, Huairou District, Beijing 101408, China

<sup>6</sup>KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming Institute of Zoology, Chinese Academy of Sciences, No.17 Longxin Road, Kunming, Yunnan 650204, China

<sup>7</sup>Key Laboratory of Systems Biology, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Chinese Academy of Sciences, No. 1 Xiangshan Branch Lane, Xihu District, Hangzhou 310024, China

\*To whom correspondence should be addressed. Tel: +86 871 68125423; Fax: +86 871 68125423; Email: zhengp@mail.kiz.ac.cn

Correspondence may also be addressed to Luo-Nan Chen. Tel: +86 021 54920100; Fax: +86 021 54920100; Email: lnchen@sibcb.ac.cn

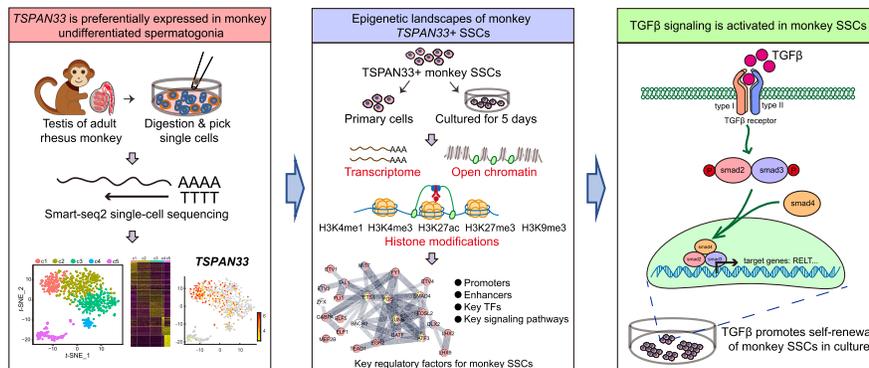
Correspondence may also be addressed to Yong-Gang Yao. Tel: +86 871 68125440; Fax: +86 871 68125440; Email: yaoyg@mail.kiz.ac.cn

<sup>†</sup>The first five authors should be regarded as Joint First Authors.

## Abstract

Spermatogonial stem cells (SSCs) play crucial roles in the preservation of male fertility. However, successful *ex vivo* expansion of authentic human SSCs remains elusive due to the inadequate understanding of SSC homeostasis regulation. Using rhesus monkeys (*Macaca mulatta*) as a representative model, we characterized SSCs and progenitor subsets through single-cell RNA sequencing using a cell-specific network approach. We also profiled chromatin status and major histone modifications (H3K4me1, H3K4me3, H3K27ac, H3K27me3 and H3K9me3), and subsequently mapped promoters and active enhancers in TSPAN33<sup>+</sup> putative SSCs. Comparing the epigenetic changes between fresh TSPAN33<sup>+</sup> cells and cultured TSPAN33<sup>+</sup> cells (resembling progenitors), we identified the regulatory elements with higher activity in SSCs, and the potential transcription factors and signaling pathways implicated in SSC regulation. Specifically, TGF- $\beta$  signaling is activated in monkey putative SSCs. We provided evidence supporting its role in promoting self-renewal of monkey SSCs in culture. Overall, this study outlines the epigenetic landscapes of monkey SSCs and provides clues for optimization of the culture condition for primate SSCs expansion.

## Graphical abstract



Received: May 24, 2024. Revised: September 12, 2024. Editorial Decision: October 14, 2024. Accepted: October 17, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

## Introduction

Spermatogonial stem cells (SSCs), which localize to the membrane basement of seminiferous tubules, are instrumental for life-long spermatogenesis post-puberty, achieving a balance in self-renewal and differentiation (homeostasis) to continuously generate mature sperms. Pre-puberty, spermatogenesis halts at the meiosis phase, with testicular structure resembling that of the early post-natal stage (1,2). Given the lack of sperm production in childhood, SSC transplantation holds great promise in fertility preservation for pediatric male patients undergoing gonadotoxic treatments (e.g. radiotherapy and chemotherapy). With childhood cancer survival rates now exceeding 80% at 5 years (3), and disease and treatments often depleting germ cells and causing infertility (4), SSC expansion and transplantation are of paramount importance for preserving fertility in these patients. Despite recent advancements in *in vitro* expansion of human undifferentiated spermatogonia (5,6), the regenerative potential of these cells remains uncertain. Consequently, successful long-term *ex vivo* expansion of authentic SSCs has not yet been achieved in humans or non-human primates, impacted by insufficient understanding of the regulatory networks governing SSC self-renewal and differentiation.

Several recent studies have explored the continuum of the cellular and molecular landscapes of spermatogenesis in humans (2,7–10) and monkeys (11,12) using high-throughput single-cell RNA sequencing (scRNA-seq). The purification of human undifferentiated (including SSCs and progenitor spermatogonia) and differentiated spermatogonia via specific cell surface markers (e.g. SSEA4 and PLPPR3 for undifferentiated spermatogonia, and KIT for differentiated spermatogonia) and subsequent bulk RNA-seq and ATAC-seq have further enriched our understanding of gene expression and molecular regulation in such cells (13,14). Notably, several undifferentiation states of spermatogonia with specific gene expression signatures were identified in humans and monkeys, reflecting high plasticity and metastable behavior (8,11,12). In addition, the FGF and BMP signaling pathways have been implicated in regulating human undifferentiated spermatogonia (7), while the AKT activation was proposed to induce spermatogonia differentiation (14). Various transcription factors (TFs), including CTCF, CTCFL, DMRT1, NFY, SOX3, SOX9, FOXP1 and FOXA2, have also been predicted to function in human undifferentiated spermatogonia (13).

Epigenetic modifications are pivotal in governing gene expression and stem cell differentiation and development (15,16). While epigenetic regulation of SSC homeostasis has been intensively studied in mice (17), research on DNA methylation and histone modification in SSC homeostasis in humans and non-human primates remains limited due to a lack of testicular samples. One study has assessed the DNA methylome of human SSEA4<sup>+</sup> cells (enriched for undifferentiated spermatogonia) in relation to mature sperms, finding no markedly alterations in DNA methylation, with further exploration using ATAC-seq revealing intricate open chromatin regions and specific TFs potentially binding to these regions (13). Most recently, one study investigated the chromatin accessibility, DNA methylome and transcriptome landscapes in 1097 human testicular cells throughout the spermatogenesis by using single-cell multi-omics sequencing (10). To the best of our knowledge, no study has reported on histone modifications relevant to undifferentiated spermatogonia in humans

or monkeys, which is critical for understanding the properties and development of SSCs.

Highlighting the parallels in spermatogenesis regulation between monkeys and humans (11,12), this study used a cell-specific network (CSN) approach to characterize SSC and progenitor subsets by scRNA-seq. Using a membrane protein TSPAN33, putative rhesus monkey (*Macaca mulatta*) SSCs were enriched and, after 5 days of culture, initiated differentiation to molecularly resemble progenitors. Comprehensive RNA expression (RNA-seq), chromatin accessibility (ATAC-seq) and major histone modification (H3K4me3, H3K27me3, H3K9me3, H3K27ac and H3K4me1) analyses were conducted for the two cell populations. Through the integration of multi-omics analyses, the epigenetic landscapes of rhesus monkey putative SSCs were revealed. Importantly, we uncovered that TGF- $\beta$  signaling was active in monkey SSCs and inclusion of TGF- $\beta$  in culture stimulated the *in vitro* expansion of monkey SSCs.

## Materials and methods

### Key resources table

The antibodies and primer pairs used in this study were presented in Supplementary Key Resources Table.

### Collection and digestion of testis tissues from rhesus monkey

Testis samples from adult male rhesus monkeys (9–14 years old) were donated from the National Resource Center for Non-Human Primates, National Science & Technology Infrastructure of China (<http://nhp.kiz.ac.cn>). Monkey testis single-cell suspensions were prepared as previously reported (18), with some modifications. In brief, testis samples were washed five times with phosphate-buffered saline (PBS) containing penicillin and streptomycin solution (Thermo Fisher Scientific, 15140–122). After cutting into small pieces, the testes were digested in a solution containing 2 mg/mL collagenase (Gibco, 17104–019) and 0.4 mg/mL DNase (Sigma, DN-25) at 37°C for 30 min. After washing with PBS three times, the cells were digested in solution containing 0.05% trypsin (Thermo Fisher Scientific, 25200–072) and 0.2 mg/mL DNase for 10 min. The digestion reaction was then terminated by the addition of 10% fetal bovine serum (FBS, Thermo Fisher Scientific, 10091–148). The digested cell suspension was filtered through a 70- $\mu$ m cell strainer (BD, 352350) to obtain the final single-cell suspension for subsequent experiments.

### Smart-seq2 sequencing

Primary testis single-cell suspensions were subjected to two rounds of differential cell adhesion (each for 6 h) in culture medium to remove somatic cells. The obtained single-cell suspensions, or cells under *in vitro* culture for one or two weeks, were washed three times in PBS containing 0.1% bovine serum albumin (BSA, Macklin, B824162). Cells with a diameter of 10–15  $\mu$ m were manually selected and added to lysis buffer containing ERCC spike-in (19). Single-cell cDNA was prepared using Smart-seq2 (20) as previously described (21), with some modifications. In brief, full-length poly(A)-tailed RNAs of each single cell were reverse-transcribed and amplified with 23 PCR cycles to increase cDNA amount. PCR analysis was performed to detect the mRNA expression of *ZBTB16* (marker gene for undifferentiated spermatogonia).

nia) in cDNA libraries. Libraries positive (~60%) or negative (~40%) for *ZBTB16* mRNA expression were subjected to smart-seq2 sequencing. Sequencing was performed using the Illumina X Ten platform with 150-bp paired-end reads.

### Smart-seq2 single-cell raw data processing

UMI-tool (<https://github.com/CGATOxford/UMI-tools>) dealing with unique molecular identifiers (UMIs) and scRNA-seq cell barcodes (22), was applied to process raw FASTQ files and generate a gene-cell UMI matrix, using *Macaca mulatta* Mmul\_8.0.1.92 as a reference genome. The output matrix was then imported into the Seurat (v4.0.2) R toolkit for quality control and downstream analysis. For each cell, gene counts, UMIs and mitochondrial gene expression proportions were quantified. Cells meeting the following criteria were used for further analysis: detection threshold of 2000–10 000 genes, 10 000–100 000 UMIs and mitochondrial gene percentage <10%. Finally, 713 single cells were retained for further analysis.

### CSN method and unsupervised clustering analysis

The CSN method (23) was used to transform the initial gene expression matrix into a network degree matrix before clustering analysis. The network degree matrix can obtain better clustering performance based on dynamic theory. An unsupervised graph-based clustering algorithm implemented in Seurat 4 (v4.0.2) was then used to cluster all single cells based on their network degree (24). The *FindNeighbors* and *FindClusters* functions across 30 principal components (PCs) at a resolution of 0.5 were used to cluster all single cells. Five clusters were obtained, i.e. c1, c2, c3, c4 and c5.

### Dimensionality reduction and visualization using *t*-distributed stochastic neighbor embedding and uniform manifold approximation and projection

For dimension reduction and visualization, the *RunPCA* function in Seurat was used to perform principal component analysis (PCA) of the normalized gene expression matrix, which yielded 30 PCs. The *RunTSNE* and *RunUMAP* functions were then used to perform *t*-distributed stochastic neighbor embedding (*t*-SNE) and uniform manifold approximation and projection (UMAP) analyses of the 30 PCs, followed by two-dimensional visualization.

### Pseudo-trajectory inference

Diffusion pseudotime (DPT), a diffusion mapping method, was used to perform pseudo-trajectory inference (25). This method can measure transitions between cells using diffusion-like random walks and robustly estimate cell order. The MATLAB package *DPT* (v1.0) was used to perform analysis, with the top 30 PCs from PCA used as input data. The DPT parameters were: *transition matrix construction method* = 'nearest neighbors', *knn* = 20, *gstatmin* = 1.01, *nsig* = 10.

### Network flow entropy analysis

The network flow entropy (NFE) approach was used to estimate the differentiation potency of each cell (26). NFE was calculated based on a gene-gene association network and illustrated the lineage dynamics of cell differentiation by quantifying the differentiation potency of the single cells. Undifferentiated cells exhibit higher differentiation potency and higher

entropy values, while differentiated cells exhibit lower entropy values. The undifferentiated spermatogonia were then identified using NFE analysis.

### Identification of DEGs in different cell clusters

Differentially expressed genes (DEGs) were identified based on the Wilcoxon rank-sum test implemented using the *FindMarkers* and *FindAllMarkers* functions in the Seurat package. Unless noted otherwise, DEGs were selected based on the following criteria:  $P < 0.001$ , absolute log<sub>2</sub> fold-change (after Laplace transformation) > 0.5 and minimum fraction > 0.1. The DEGs were further classified into several groups. For gene *x*, the five clusters were sorted in descending order based on average gene expressions, e.g.  $c1 > c2 > c3 > c4 > c5$ . If the difference between  $c1/c2$  was not significant, but the differences between  $c2/c3$  and between  $c1c2/c3c4c5$  were significant, the gene was considered to be highly expressed in both  $c1$  and  $c2$ , rather than just highly expressed in  $c1$ .

### Enrichment analysis of DEGs

The Database for Annotation, Visualization and Integrated Discovery (DAVID, <https://david.ncifcrf.gov>) was used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of DEGs (27,28). DAVID provides a comprehensive set of functional annotation tools to understand the biological meaning behind large lists of genes. *Metascape* (v3.5, <https://metascape.org>) was also used to perform enrichment analysis of multiple gene lists (29).

### Enrichment of TSPAN33<sup>+</sup> cells

After two rounds of differential cell adhesion to remove the testicular somatic cells, TSPAN33<sup>+</sup> cells were enriched by magnetic activated cell sorting (MACS) or fluorescence activated cell sorting (FACS). For MACS, cells were re-suspended in MACS buffer [PBS containing 0.5% BSA and 2 mM ethylenediaminetetraacetic acid (EDTA)], then incubated with TSPAN33 antibodies (1:50 volume ratio of antibody/cell suspension, ABclonal, A5222) at 4°C for 30 min. After washing with MACS buffer three times, cells were incubated with anti-rabbit IgG MicroBeads (Miltenyi, 130–048–602) (1:4 volume ratio of beads/cell suspension) at 4°C for 15 min. After washing with MACS buffer three times, the TSPAN33<sup>+</sup> cells were enriched and obtained using LS Separation columns (Miltenyi, 130–042–401) combined with a MidiMASC Separator (Miltenyi, 130–042–302). For FACS, cells were re-suspended in FACS buffer (PBS containing 3% FBS and 10 mM EDTA), then incubated with TSPAN33 antibodies (1:50 volume ratio of antibody/cell suspension, ABclonal, A5222) at 4°C for 30 min. After washing with FACS buffer three times, the cells were incubated with goat anti-rabbit secondary antibodies (Thermo Fisher Scientific, A11034) labeled with green-fluorescent dye at 4°C for 15 min. The TSPAN33<sup>+</sup> cells were obtained using flow cytometry (BD, Influx, USA). Antibody information is listed in the [Supplementary Key Resources Table](#).

### Cell culture

TSPAN33<sup>+</sup> cells obtained by MACS were maintained in culture medium, as previously described (18). In brief, cells were cultured in StemPro-34 SFM with StemPro supplement (Thermo Fisher Scientific, 10639–011), supplemented with 10% FBS (Thermo Fisher Scientific, 10091–148), 40 ng/mL

recombinant human glial cell line-derived neurotrophic factor (GDNF, R&D, Lot# VQ3222081), 10 ng/mL human epidermal growth factor (EGF, Thermo Fisher Scientific, PHG0311), 1000 U human leukemia inhibitory factor (LIF, Millipore, LIF1050), 2 ng/mL human basic fibroblast growth factor (bFGF, Millipore, GF003), 5 ng/mL recombinant human Wnt3a (R&D, 5036-WN-010), 2 mM L-glutamine (Sigma, 56-85-9), 1 mM  $\beta$ -mercaptoethanol (Sigma, M3148), 1 mM sodium pyruvate (Gibco, 11360-070), 1% Modified Eagle Medium (MEM) non-essential amino acid solution (Gibco, 11140-050) and 60  $\mu$ M putrescine (Sigma, V900377). Tree shrew Sertoli cells, treated with 10  $\mu$ g/mL mitomycin C (Sigma, M4287), were used as feeder cells. Cells were cultured in 24-well plates (Corning, 3524) and passaged every 7 days. The culture medium was changed every other day. The culture was maintained at 37°C and 5% CO<sub>2</sub> atmosphere. For TGF- $\beta$  treatment, 1 ng/mL of TGF- $\beta$  (R&D, 7754-BH-005/CF) was added to the culture medium.

### Immunofluorescence staining

Testes were fixed in 4% paraformaldehyde (Sigma, 158127-500G), sequentially dehydrated with 15% and 30% sucrose (A502792-0500), embedded in optimal cutting temperature compound (OCT, 4583), sectioned (10  $\mu$ m) and subjected to immunofluorescence staining. In brief, testicular sections were washed three times in PBS, followed by permeabilization with 0.25% Triton X-100 (Sigma, Lot# SLBW6818) for 15 min at room temperature. For cell immunofluorescence staining, cultured cells were fixed in 4% paraformaldehyde for 15 min, and permeabilized with 0.2% Triton X-100 in PBS for 15 min at room temperature.

Testis sections and cells were blocked with 5% BSA for 1 h at room temperature and stained with primary antibodies diluted in 1% BSA at 4°C overnight. The cells and sections were washed three times with PBS, then stained with Alexa Fluor conjugated secondary antibodies for 1 h at room temperature. The nuclei were counterstained with 4',6-diamidino-2-phenylindole (Sigma-Aldrich, D9542). The cells and sections were sealed with 80% glycerin (Solarbio, G8190) and photographed using confocal laser microscope (FV1000) or high-resolution fluorescence microscopy system (LSM880). Antibody information is listed in the [Supplementary Key Resources Table](#).

Immunofluorescence staining was quantified by counting the positively stained cells in field of views (for sorted cells), and cell colonies (for cultured cells, colonies containing >4 cells were included in the analysis). The numbers of positively stained cells are indicated in the respective figure legends.

### Immunohistochemistry analysis

Testes were fixed in 4% paraformaldehyde for 24~72 h, and were dehydrated in a series wash of ethanol at different concentrations (50%, 75%, 95% and 100%). Tissues were then embedded in paraffin blocks, and were sectioned at 6  $\mu$ m thickness. The sections were deparaffinized three times in xylene (each for 20 min), and were rehydrated in a descending ethanol series and water. Heat-induced epitope retrieval was performed using a water bath (95 °C) in 1 M EDTA buffer (pH 8.0) for 20 min. The sections were then incubated with 3% hydrogen peroxide in PBS for 10 min to quench the endogenous peroxidase activity. Sections were blocked using 2.5% BSA and 5% donkey serum in PBS for 60 min at room

temperature, followed by incubation with the primary antibody against TSPAN33 (Abclonal, A5222, 1:100) overnight at 4°C. After washing three times with tris-buffered saline with 0.1% Tween 20 (Roche, 11332465001) (TBST), the sections were incubated with streptavidin-conjugated HRP-IgG goat anti-rabbit secondary antibody for 60 min at room temperature. The sections were further visualized with the 3,3-diaminobenzidine solution, followed by a counterstaining of nuclei with hematoxylin. The sections were then dehydrated in a series wash of ethanol at different concentrations (50%, 75%, 95% and 100%, each for 5 min), and were immersed in xylene two times (each for 10 min). The sections were viewed with OLYMPUS VS120 microscope.

### Quantitative RT-PCR

Total cell RNA was extracted using TRNzol reagent (Tiangen, DP424). Reverse transcription was performed using a PrimeScript™ RT Reagent Kit with gDNA Eraser (TaKaRa, RR047B). Quantitative RT-PCR (qRT-PCR) was performed using a TB Green™ Premix Ex Taq™ II Kit (Takara, RR820A) and CFX96TM-Real Time System (Bio-Rad, CFX96Touch). The expression level of each gene was normalized by *GAPDH*. The primers for analyzed genes are provided in the [Supplementary Key Resources Table](#).

### Bulk RNA sequencing

RNA sequencing (RNA-seq) of the primary and cultured TSPAN33<sup>+</sup> germ cells with or without TGF- $\beta$  treatment was performed. Two independent biological samples were prepared. Total RNA was extracted and subjected to RNA-seq using the NovaSeq 6000 platform. Bulk RNA-seq data were aligned to the RheMac10 genome reference by STAR (30). Gene expression was calculated according to the ENSEMBL database using htseq-count. Read counts were used as input in the DESeq2 (v1.36.0) (31) package for differential expression analysis. DEGs between different conditions were defined based on adjusted  $P < 0.05$  and fold-change > 1.5. GO term and KEGG pathway enrichment analyses of DEGs were performed using DAVID web-tools (<https://david.ncifcrf.gov/tools.jsp>). Heatmaps were drawn using the R package (v4.2.1), ggplot2 (v3.4.2) and pheatmap (v1.0.12, <https://github.com/raivokolde/pheatmap>), with transcripts per million (TPM) instead of raw counts shown.

### ATAC-seq library construction

The ATAC-seq libraries of primary and cultured cells were prepared using the TruePrep® DNA Library Prep Kit (Vazyme, TD501) following the manufacturer's instructions. Briefly,  $1 \times 10^4$  cells were lysed in 50  $\mu$ L of cold lysis buffer (Sigma-Aldrich, NUC101) to generate nuclei, followed by centrifugation at 500 relative centrifugal force for 10 min at 4°C to remove the supernatant. The nuclear pellet was subjected to transposition reaction with Tn5 transposase at 37°C for 30 min and purified using KAPA Pure Beads (KAPA Biosystems, ks8002). The transposed DNA fragments were amplified by the following procedure: 72°C for 5 min, 98°C for 1 min, and 22 cycles (each cycle: 98°C for 15 s, 60°C for 30 s and 72°C for 1 min). The amplified PCR products were purified using the KAPA Pure Beads to obtain the ATAC-seq libraries and were sequenced on the NovaSeq 6000 platform to generate 150-bp paired-end reads.

## CUT&Tag library construction

The CUT&Tag histone modification libraries (H3K4me1, H3K4me3, H3K27ac, H3K27me3 and H3K9me3) of primary and cultured cells were prepared using a Hyperactive Universal CUT&Tag Assay Kit (Vazyme, TD904) following the manufacturer's instructions. Briefly,  $1 \times 10^4$  cells were incubated with ConA Beads at room temperature for 10 min. The cell-bead complex was collected on a magnetic rack. Pre-chilled antibody buffer (50  $\mu$ L) containing 1  $\mu$ g of primary antibody (see the [Supplementary Key Resources Table](#)) was added to each sample. The samples were gently mixed and incubated at 4°C overnight. The cell-bead complex was collected using a magnetic rack, then incubated with secondary antibodies at room temperature for 1 h. After twice washing with washing buffer, the cell-bead complex was incubated with Tn5 transposase at room temperature for 1 h, then subjected to fragmentation at 37°C for 1 h. The reaction was terminated by the addition of 1  $\mu$ L of 10% sodium dodecyl sulfate (SDS) and incubated at 55°C for 10 min. The products were then purified with DNA extract beads. The extracted DNA was amplified using the following procedure: 72°C for 3 min, 98°C for 30 s, 22 cycles (each cycle: 98°C for 15 s, 60°C for 20 s and 72°C for 10 s) and maintenance at 72°C for 2 min. The amplified PCR products were purified using DNA Clean Beads to obtain the CUT&Tag libraries and sequenced on the NovaSeq 6000 platform to generate 150-bp paired-end reads.

## ATAC-seq and CUT&Tag sequencing analysis

Reads from ATAC-seq and CUT&Tag datasets were mapped to the RhesusMac10 reference genome using Bowtie2 (v2.3.1) (32). Multiple mapped reads and PCR duplicates of ATAC-seq were removed using the Sambamba (v0.6.6) (33) markdup command. Peaks were called by MACS2 (34) with  $P < 0.01$ . Duplicates in the CUT&Tag datasets were removed using the Picard MarkDuplicates (v2.26.10, <https://broadinstitute.github.io/picard/>) command. MACS2 was used for peak calling. After removing the duplicates, the reads were normalized by bamCoverage (v1.5.11) (35) to calculate reads per genome coverage. The consistency of replications was calculated with the Spearman correlation analysis using plotCorrelation (v3.5.1) (35). The WashU Epigenome Browser (<http://epigenomegateway.wustl.edu/browser/>), R package ggplot2 (v3.4.2) and heatmaply (v1.4.0) (36) were used to visualize the normalized data. The R package Diffbind (v3.6.5) was used to identify differential peaks between the primary and cultured cells ( $P < 0.05$ ; fold-change  $> 1.5$ ). ChIPseeker (v1.32.1) (37) was used to annotate peaks to their nearest genes and corresponding genomic regions. GO term and KEGG pathway enrichment analyses were performed using DAVID web-tools (<https://david.ncifcrf.gov/tools.jsp>). Promoters were defined as regions with strong H3K4me3 signal and  $\pm 3$  kb from the transcription start sites (TSS). Active enhancers were defined as regions simultaneously positive for H3K4me1 and H3K27ac modifications, but negative for H3K4me3 modification.

## Identification of bivalent promoters

ChIPseqSpikeInFree (38) was used to normalize the enrichment of H3K4me3 and H3K27me3 in the promoter regions ( $\pm 3$  kb from the TSS region). Inactive genes (TPM  $< 5$ ) showing  $> 80\%$  overlap of H3K4me3 and H3K27me3 peaks in promoter regions were identified as bivalent genes. Promoters

were identified as strongly bivalent based on the following: (1) H3K4me3 and H3K27me3 levels in the promoter region exceeded average levels found in housekeeping genes; and (2) RNA expression of related genes was low (TPM  $< 5$ ). Housekeeping genes were identified using CHIPIN (39). By analyzing the mean and standard deviation (SD) of the TPM values for each gene from RNA-seq dataset, the least variable genes with the smallest SD values between primary and cultured cells were obtained, classified as housekeeping genes.

## TF motif analysis

The *'findMotifsGenome.pl'* script in HOMER (40) was used to identify potential TF motifs in specific regions. The protein-protein association network and functional enrichment of the identified TFs were analyzed using the STRING database (41). Gene regions with potential SMAD3-binding motifs were predicted using the *'scanMotifGenomeWide.pl'* script.

## Immunoblotting analysis

The TSPAN33<sup>+</sup> germ cells were collected by MACS. Primary and 5 day cultured TSPAN33<sup>+</sup> cells were lysed in RIPA buffer (Beyotime, P0013J) supplemented with proteinase inhibitor (Beyotime, P1006) and phosphatase inhibitor (Beyotime, P1082). Protein concentration was determined using a BCA Protein Assay Kit (Beyotime, P0012). The protein lysate was separated using 12% SDS-polyacrylamide gel electrophoresis and transferred to a polyvinylidene difluoride membrane (Bio-Rad, 162-0177). After blocking with 5% (*w/v*) non-fat dry milk in TBST for 2 h at room temperature, the membrane was incubated with respective primary antibodies overnight at 4°C, followed by incubation with the corresponding HRP-conjugated secondary antibodies for 1 h at room temperature. After washing with TBST buffer three times, bands were detected using the ECL reagent (Beyotime, P0018FS) from the ProteinSimple FluorChem system (protein simple, Fluorchem M FM0561). All antibody information is listed in the [Supplementary Key Resources Table](#).

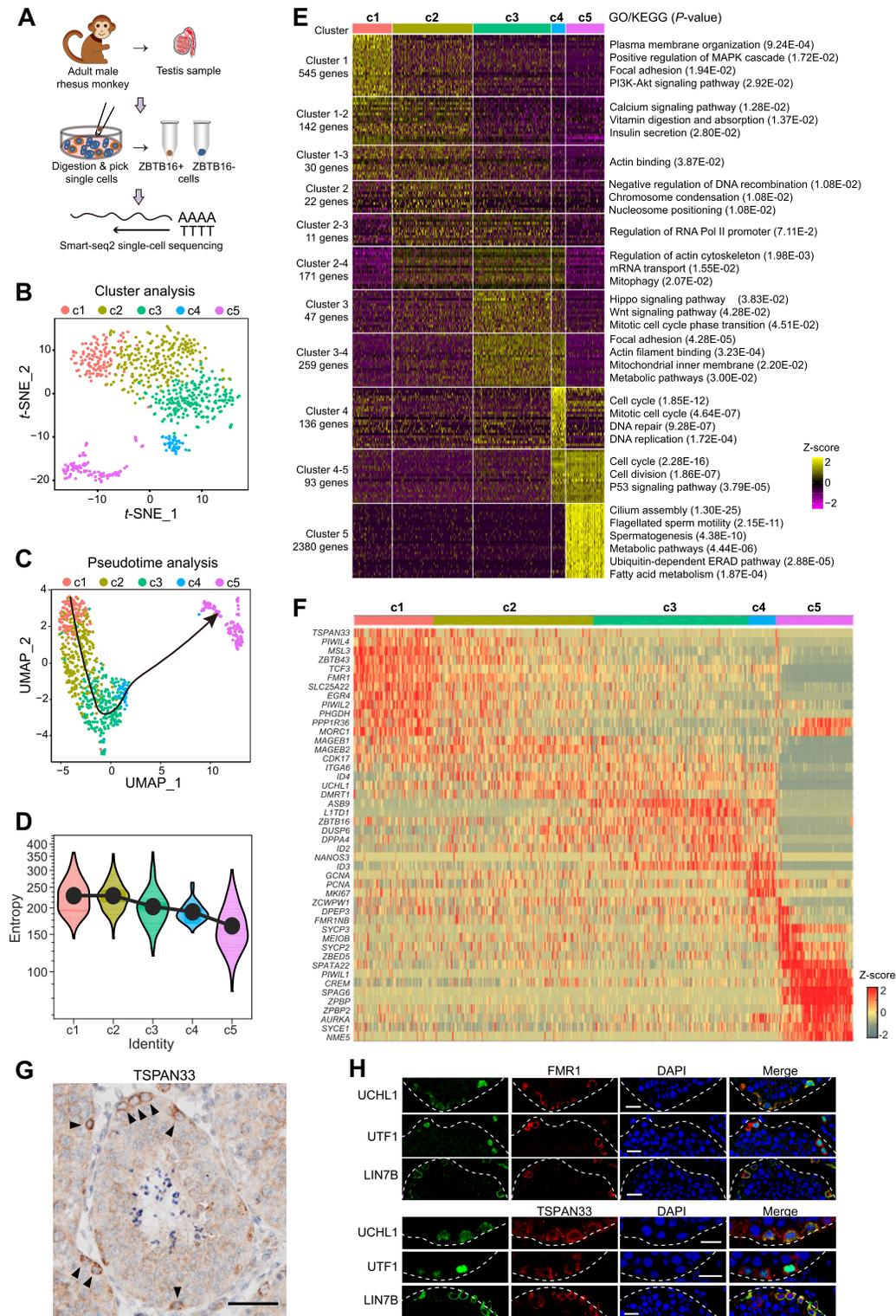
## Statistical analysis

Differences in mRNA and protein expression levels between primary and cultured cells, or in cells under TGF- $\beta$  treatment, were analyzed using GraphPad Prism v8 (GraphPad Software, La Jolla, CA, USA) through the two-tailed Student's *t*-test.  $P < 0.05$  was considered significant, with all data provided as mean  $\pm$  SD.

## Results

### Characterization of undifferentiated spermatogonia cell types in rhesus monkeys by scRNA-seq analysis based on a CSN approach

To elucidate the molecular characteristics of undifferentiated and differentiated spermatogonia cells, a single-cell suspension was prepared from adult rhesus monkey testes. After removing most somatic cells, single germ cells with a diameter of 10–15  $\mu$ m were manually collected, and cDNA was amplified using the Smart-seq2 protocol (20). Upon examining amplified cDNA via conventional PCR, germ cells either expressing or lacking *ZBTB16* (also called *PLZF*), a conserved marker of undifferentiated spermatogonia, were subjected to scRNA-seq (Figure 1A). In total, 713 cell samples exhibited high sequenc-



**Figure 1.** Characterization of spermatogonia cell types by scRNA-seq analysis in rhesus monkeys based on a CSN method. **(A)** Experimental workflow. **(B)** *t*-SNE analysis of single-cell transcriptome identified five cell clusters (c1–c5) by the CSN method. **(C)** DPT analysis was used to infer pseudo-trajectories of the five cell clusters. **(D)** Entropy of each cell was estimated based on NFE analysis. Cells in c1 showed the highest entropy, while cells in c5 showed the lowest entropy. **(E)** DEGs were identified by pairwise comparison of gene expression across all clusters [ $P < 0.001$ ,  $\log_2$  (fold-change)  $> 0.5$ , minimum fraction  $> 0.1$ ], and were clustered into 11 groups. Enriched GO terms and KEGG pathways for each group of DEGs were shown. **(F)** Expression heatmap of marker genes for different cell clusters. **(G)** Immunohistochemistry analysis of TSPAN33 in testicular sections of adult monkey. The triangles indicate TSPAN33 positive cells. Scale bar, 50  $\mu$ m. **(H)** Immunofluorescence staining of testicular sections of adult monkeys with indicated antibodies: FMR1 (in red), TSPAN33 (in red), UCHL1 (a broad marker for undifferentiated spermatogonia), UTF1 and LIN7B (markers for undifferentiated spermatogonia). Scale bar, 50  $\mu$ m. Dashed line outlines the boundary of seminiferous tubules. Partial enlarged views were shown here, and the whole staining sections were seen in [Supplementary Figure S2](#). Also see [Supplementary Figure S1](#) and [S2](#).

ing quality (~5000 genes/cell) were used for downstream data analysis. In our previous work, a CSN method was formulated to analyze scRNA-seq data (23), whereby a distinct network is constructed for each cell, with cell clustering and downstream analyses performed based on these single-cell networks (23). Compared to conventional gene expression-based computational methodologies, the CSN approach transforms the data from an 'unstable' gene expression form to a 'stable' network form, which is more reliable for characterizing cellular state. The CSN method can also identify important network hub genes that are often ignored by traditional differential gene expression analyses (23). Through the application of CSN-based single-cell analysis, we identified five cell clusters (c1–c5) with distinct transcriptional states (Figure 1B). Most of the cells (>97%) expressed germ cell genes *DAZL* and/or *DDX4* (Supplementary Figure S1A and B), confirming their germ cell identity. Pseudotime ordering showed a developmental trajectory from c1 to c5 (Figure 1C). Consistently, estimation of cellular NFE in each cluster revealed a continuous decrease in entropy from c1 to c5 (Figure 1D). These findings suggest that c1 cells, with the highest differentiation potency, may be the putative SSCs.

To better understand the developmental phase of each cell cluster and the variations in gene expression throughout germ cell development, a pairwise comparison of gene expression among clusters was performed. In total, 3836 DEGs either unique to one of the five cell clusters or shared by adjacent cell clusters along pseudotime were yielded. These DEGs were clustered into 11 groups (Figure 1E and Supplementary Data 1), and the representative pathways enriched for each group were identified. Notably, genes highly expressed in c5 (2380 genes, cluster 5 group) were enriched in cilium assembly, flagellated sperm motility and spermatogenesis (Figure 1E and Supplementary Data 1). Specifically, meiosis-specific markers, including *SYCP2/3* and *SYCE*, were predominantly detected in c5, but not in the c1–c4 cells (Figure 1F). Marker genes (*PSMA8*, *PIWIL1*, *CCNA1*, *PRM2*) for mouse pachytene spermatocytes to spermatids (42) were also highly expressed in c5 (Supplementary Figure S1C). In contrast, the pan undifferentiated spermatogonia markers including *ZBTB16*, *UCHL1* and *LIN7B* were expressed in c1–c4 (Supplementary Figure S1D–G). These observations suggested that c1–c4 cells were undifferentiated spermatogonia, whereas c5 cells likely initiated meiosis. Of note, *KIT*, a well-established marker for differentiated spermatogonia in humans and mice (11), showed no enriched expression pattern in any of the cell clusters (Supplementary Figure S1H). *KIT*-expressing spermatogonia accounted for about 0.1% of the starting cell preparation (Supplementary Figure S1I). We reasoned that *KIT*<sup>+</sup> spermatogonia were not picked for analysis, and lack of *KIT*<sup>+</sup> germ cell population might partly account for the uncontinuity between c4 and c5 cells in the developmental trajectory.

Based on further analysis of the c1–c4 undifferentiated spermatogonia cells, we found that the expression of markers indicative of early progenitor spermatogonia *LITD1* and *ASB9* (12) was initiated in c3 cells and persisted in c4 cells, whereas late progenitor spermatogonia markers (*NANOS3* and *GCNA*) and cell proliferation markers (e.g. *PCNA* and *MKI67*) peaked in the c4 cells (Figure 1F). In contrast, the non-canonical putative SSC markers, including *TSPAN33*, *PIWIL4*, *MSL3*, *FMR1* and *MORC1*, were dominantly expressed in the c1–c2 cells (Figure 1F). These patterns sug-

gested that c1–c2 cells were potential SSCs, whereas c3 and c4 cells were early and late progenitor spermatogonia, respectively. These findings are consistent with previous reports demonstrating the existence of different SSC states in humans (8). In line with the higher entropy value, c1 cells expressed higher levels of putative SSC markers (e.g. *TSPAN33*, *MSL3*, *ZBTB43*, *FMR1*, *MORC1*, *PIWIL4* and *PHGDH*) (8,11) compared to c2 cells (Figure 1F and Supplementary Figure S1J–P). In contrast, compared to c1 cells, a small part of c2 cells initiated expression of early progenitor markers (e.g. *LITD1* and *ASB9*) (9) (Figure 1F). Thus, we concluded that c1 cells were more primitive than c2 cells. Similarly, previous studies in humans (8) and monkeys (11,12) have also reported higher expression of these genes in a specific SSC subtype, considered as the earliest/most naive SSCs. Overall, we identified four subtypes of undifferentiated spermatogonia (c1–c4). The transition from c1 to c4 represents a progressive differentiation process from primitive SSCs to late progenitors.

We next explored the potential molecular events facilitating the c1–c4 transition. Notably, c1 primitive SSCs were enriched in the MAPK, cell adhesion and phosphatidylinositol 3-kinase (PI3K) signaling pathways (Figure 1E). Mitochondrial inner-membrane and metabolic pathways were enriched in c3–c4 cells (progenitors) (Figure 1E), suggesting a potential metabolic pattern shift from glycolysis to oxidative phosphorylation upon early SSC differentiation. The gradual changes in gene expression during c1 (primitive SSCs) to c3 (early progenitors) transition were also assessed (Supplementary Data 2). Comparison between the c1 and c2 cells identified 533 DEGs, including 272 that were specifically or highly expressed in c1 cells. In total, 747 DEGs were identified between c2 and c3. Intriguingly, GO terms associated with genes that were upregulated during c1 to c3 transition were enriched in mitochondrial and metabolic processes (Supplementary Figure S1Q), supporting the significant role of metabolism in regulating SSC homeostasis.

### TSPAN33<sup>+</sup> undifferentiated spermatogonia are enriched for putative SSCs

This and other studies consistently revealed that in monkeys and humans, the potential SSCs highly transcribed a list of non-canonical marker genes including *TSPAN33*, *FMR1*, *MORC1* and *PIWIL4* (Supplementary Figure S1J, M, N and O) (8,11). The protein expressions of these genes were detected in a subset of human spermatogonia which express undifferentiated spermatogonia markers UTF1, UCHL1, but not differentiated spermatogonia marker *KIT* (8,9,11,13). Thus, the putative human SSCs are enriched in the undifferentiated spermatogonia expressing these non-canonical marker proteins. Because *FMR1* and especially *TSPAN33* were predominantly transcribed in putative monkey SSCs (c1 and c2 cells) (Supplementary Figure S1J and M), we examined the identity of spermatogonia expressing the two proteins. In testicular sections of rhesus monkey, cells expressing *TSPAN33* or *FMR1* were located in the basal compartment of testis tubules (Figure 1G and H, and Supplementary Figure S2). Moreover, cells expressing *FMR1*/*TSPAN33* also expressed undifferentiated spermatogonia markers UTF1, UCHL1 and *LIN7B* [*LIN7B* was identified as a marker of human SSCs (9)] (Figure 1H and Supplementary Figure S2). These observations suggested that germ cells expressing *TSPAN33*/*FMR1* were undifferentiated spermatogonia and enriched for puta-

tive SSCs in monkeys. Autologous transplantation or xenotransplantation into recipient immunodeficient mouse testis is required for functional assessment of SSCs (43). In xenotransplantation experiments, putative SSCs can migrate and colonize the basement membrane of the seminiferous tubules, where they proliferate and survive for months but do not initiate differentiation (43). In humans, TSPAN33<sup>+</sup> spermatogonia were capable of colonizing the basement membrane of seminiferous tubules after xenotransplantation, indicating that TSPAN33<sup>+</sup> population is enriched for transplantable SSCs (11). Of note, TSPAN33<sup>-</sup> human testicular cells also displayed colonization potential (11). This observation reconciles to the two states of SSCs with distinct TSPAN33 expression patterns. Similar to human SSCs, monkey putative SSCs (c1–c2) also displayed heterogeneity in TSPAN33 expression. We propose that monkey TSPAN33<sup>+</sup> population is enriched for transplantable SSCs, even though it does not capture the entire population due to the heterogeneity of SSCs. It should be mentioned that monkey functional SSCs can only be validated by autologous transplantation. However, due to the large amount of cells (about  $5\text{--}8 \times 10^7$  cells) required for transplantation into recipient monkeys (44), it is impractical to evaluate the *in vivo* spermatogenesis potential of monkey TSPAN33<sup>+</sup> spermatogonia by conducting autologous transplantation.

#### Analyses of chromatin open status of TSPAN33<sup>+</sup> spermatogonia identified the potential TFs implicated in putative SSCs regulation

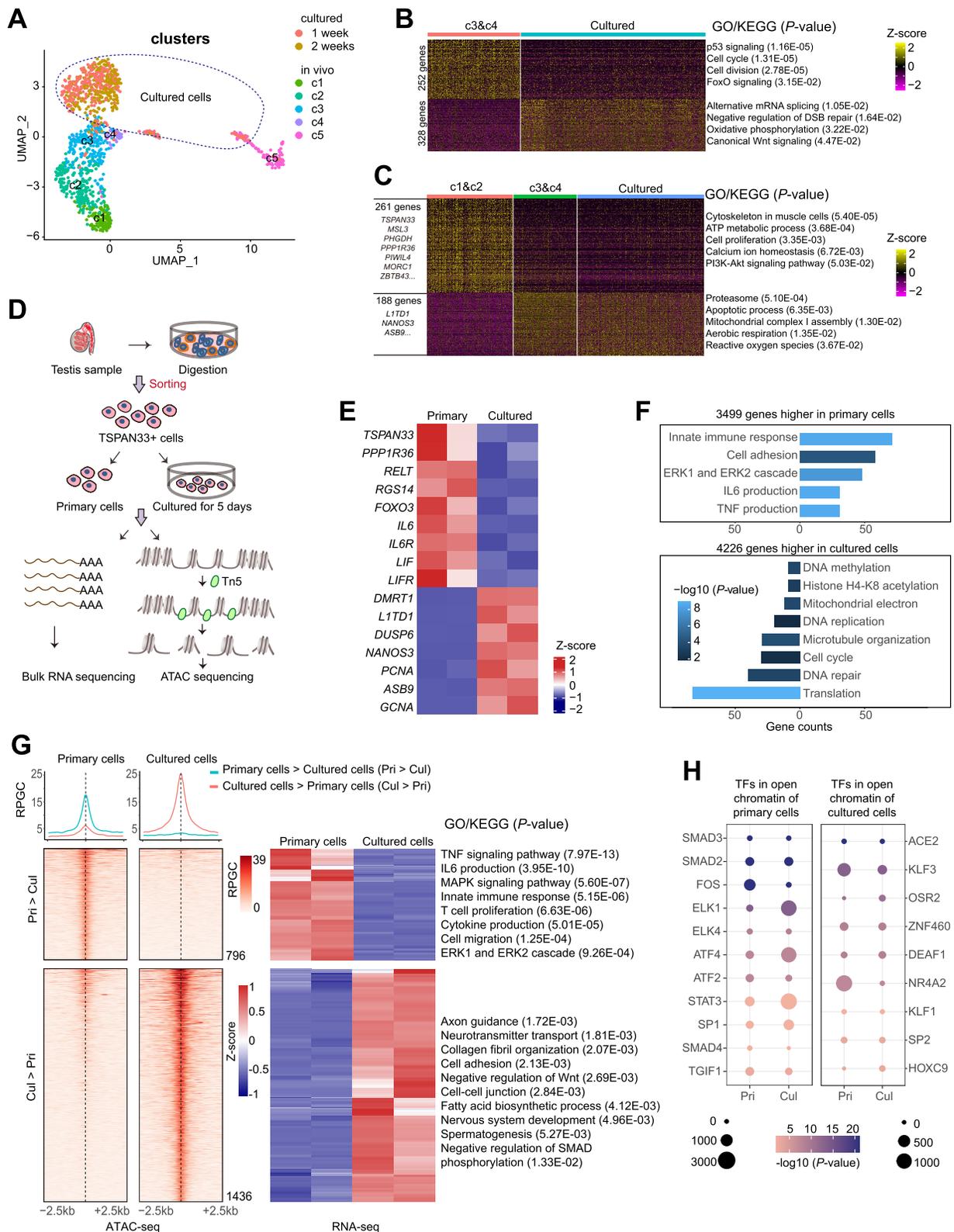
TSPAN33 is a membrane protein and TSPAN33-expressing spermatogonia are enriched for potential SSCs. We then purify the TSPAN33<sup>+</sup> spermatogonia by FACS or MACS in order to explore their epigenetic properties. Because TSPAN33 mRNA expression was also detected in two types of testicular somatic cells including macrophage and peritubular cells (45,46), we conducted rounds of differential cell adhesion to remove the testicular somatic cells before cell purification. Immunofluorescence staining with TSPAN33, UCHL1, LIN7B and DDX4 on the sorted cell samples (Supplementary Figure S3A) revealed that > 80% cells expressed DDX4, UCHL1 and LIN7B (Supplementary Figure S3B and C), indicating that these purified cells were predominantly SSCs. In addition, RNA-seq analysis of purified TSPAN33<sup>+</sup> cells revealed that differentiation genes, including *LITD1*, *GCNA*, *MKI67*, *NANOS3*, *KIT* and *SYCP2* were barely expressed (Supplementary Figure S3D), further supporting that TSPAN33 labels a population of SSCs.

To understand the epigenetic properties of TSPAN33<sup>+</sup> spermatogonia (enriched for putative SSCs), we considered to include the *in vitro* differentiating spermatogonia for comparison because there was no strategy to purify the early *in vivo* differentiating progenitors (c3 and c4). To this end, we first investigated whether cultured spermatogonia could resemble the state of differentiating spermatogonia. Germ cells were cultured in a medium optimized for the long-term expansion of SSCs from tree shrews (18,47), a species phylogenetically proximate to primates (48–50). The tree shrew SSC culture medium with mouse LIF replaced by human LIF supported the growth of undifferentiated monkey spermatogonia for around 1 month (Supplementary Figure S3E). After culture for about 1 week, undifferentiated spermatogonia formed small colonies containing 2–4 cells. These cells were manually picked and subject to same scRNA-seq after culture for one or two weeks.

Intriguingly, all cultured cells were homogenous and aligned to c3 and c4 (Figure 2A), suggesting that cultured cells were molecularly similar and close to *in vivo* differentiating progenitors. Indeed, these cultured cells retained the expression of *UCHL1* and *ZBTB16*, two markers of undifferentiated spermatogonia (Supplementary Figure S3F and G). To further understand the differences between *in vivo* progenitors (c3&c4) and *in vitro* cultured spermatogonia, we compared their gene expression profiles. Of note, only 580 DEGs were identified between c3&c4 and cultured cells (Supplementary Data 3). Genes decreased in cultured cells (252 genes) were prevalently related to cell cycle regulation, probably reflecting the cellular responses to culture conditions. Genes upregulated in cultured cells (328 genes) were enriched in RNA splicing, DSB repair and oxidative phosphorylation (Figure 2B and Supplementary Data 3). We also investigated the common genes which showed down- or up-regulation during *in vivo* and *in vitro* differentiation. Compared to *in vivo* c1&c2 SSCs, 261 common genes, including SSC signature genes (e.g. *TSPAN33*, *MSL3*, *PIWIL4*, *MORC1*, *ZBTB43*, *PPP1R36*), were downregulated in both c3&c4 progenitors and *in vitro* cultured spermatogonia (Figure 2C, Supplementary Figure S3H and Supplementary Data 3). These downregulated genes participate in the processes including ATP metabolism, cell proliferation and PI3K-AKT signaling. Meanwhile, 188 common genes (e.g. *LITD1*, *NANOS3*, *ASB9*) were upregulated and involved in the key processes of SSC differentiation (e.g. mitochondrial respiratory chain complex I assembly and aerobic respiration) (Figure 2C, Supplementary Figure S3H and Supplementary Data 3). Taken together, these analyses supported that short-term culture of undifferentiated spermatogonia could largely mimic the *in vivo* progenitor state.

Therefore, we cultured TSPAN33<sup>+</sup> cells for 5 days to capture the early differentiation (progenitor-like) state. Changes in gene expression of putative SSC markers (*TSPAN33*, *ZBTB43*, *FMR1* and *MSL3*) and differentiation markers (*DNMT1* and *MKI67*) indicated that 5-day culture induced the initiation of SSC differentiation (Supplementary Figure S3I). Furthermore, bulk RNA-seq of primary and cultured TSPAN33<sup>+</sup> cells was compared (Figure 2D). Specifically, a panel of putative SSC and progenitor markers identified by scRNA-seq were respectively downregulated and upregulated after culture (Figure 2E). Pearson correlation analysis confirmed that the cultured TSPAN33<sup>+</sup> cells showed the highest similarity to the *in vivo* c3 (early progenitors) and c4 cells (late progenitors) (Supplementary Figure S3J). These findings again demonstrated that TSPAN33<sup>+</sup> cells cultured for 5 days showed high resemblance to *in vivo* progenitor cells. Comparing the molecular properties of TSPAN33<sup>+</sup> cells prior to and after 5 days of culture could help understand the epigenetic characteristics of putative SSCs.

Gene expression profile comparisons between monkey primary and cultured TSPAN33<sup>+</sup> cells identified 7725 DEGs [3499 genes higher in primary cells and 4226 genes higher in cultured cells (adjusted  $P < 0.05$ , fold-change  $\geq 1.5$ )] (Supplementary Figure S3K and Supplementary Data 4). GO term analysis revealed that genes showing higher expression in primary cells were involved in processes such as immune response, cell adhesion and ERK cascade, whereas genes with higher expression in cultured cells were enriched in processes related to translation, cell cycle, DNA replication and mito-



**Figure 2.** Analyses of chromatin open status of undifferentiated spermatogonia identified the potential TFs for stem cell maintenance. **(A)** scRNA-seq analysis showed that *in vitro* cultured cells were close to *in vivo* c3 and c4 cells. **(B)** DEGs and enriched GO terms/KEGG pathways between *in vivo* c3&c4 cells and *in vitro* cultured cells. **(C)** DEGs and enriched GO terms/KEGG pathways among *in vivo* c1&c2, c3&c4 cells and *in vitro* cultured cells. **(D)** Experimental workflow. **(E)** Expression changes of marker genes of spermatogonia in TSPAN33<sup>+</sup> germ cells before and after 5 days of culture. **(F)** GO terms and KEGG pathways enriched in DEGs between primary and cultured TSPAN33<sup>+</sup> cells. **(G)** Heatmap of differential open chromatin signals in promoter regions and associated gene expression levels. Enriched GO terms and KEGG pathways for DEGs are shown. **(H)** TFs identified by motif enrichment analysis of differential open chromatin regions in primary and cultured TSPAN33<sup>+</sup> cells, respectively. Also see [Supplementary Figures S3–S5](#).

chondrial electron transport, which were highly relevant to spermatogonia differentiation (Figure 2F and Supplementary Data 4). Notably, *IL6*, *IL6R*, *LIF* and *LIFR* were highly expressed in primary cells (Figure 2E), suggesting that the signals evoked by *IL6* and *LIF* were more active in primary TSPAN33<sup>+</sup> cells (enriched for SSCs).

ATAC-seq was then conducted to profile open chromatin. On average, 53 806 and 42 826 peaks with an open chromatin signal were identified in primary and cultured spermatogonia, respectively (Supplementary Figure S4A). Most peaks (> 80%) were distributed at the distal intergenic regions, while 15% peaks were located at the promoter region (Supplementary Figure S4A). In total, 11 877 differential peaks were identified between the primary and cultured TSPAN33<sup>+</sup> spermatogonia (2689 peaks with a stronger signal in primary cells, and 9188 peaks with a stronger signal in cultured cells,  $P < 0.05$ , fold-change > 1.5). Among these differential peaks, about 40% were in the promoter regions, 30% were in the distal intergenic regions and 20% were in the intron regions (Supplementary Figure S4B).

We next focused on the differential open chromatin peaks located in the promoter regions ( $\pm 3$  kb from the TSS) and associated with DEGs. In total, 796 and 1436 chromatin regions predominantly opened in the primary and cultured TSPAN33<sup>+</sup> cells, respectively. Intriguingly, the open regions in the primary cells regulated genes involved in tumor necrosis factor (TNF) signaling, immune responses, *IL6* production, MAPK signaling and ERK cascades (Figure 2G and Supplementary Data 5). The immune responses, such as *IL-6* and *IL-1 $\beta$*  signaling, have been shown to play critical role in reprogramming and protect stem cells from cytotoxic stress (51,52). Motif enrichment analysis of the specific open regions in primary TSPAN33<sup>+</sup> cells identified various TFs including SMAD2/3 (Figure 2H and Supplementary Data 5). The genome browser snapshots of ATAC-seq results for several genes (*RELT*, *NCF2*, *MFNG*, *SLC43A3*) containing SMAD2/3 binding motifs at promoter regions were shown (Supplementary Figure S5A–D). Notably, GO term and KEGG pathway analyses of the TFs identified several potentially active pathways, including TGF- $\beta$  and MAPK, in the primary cells (Supplementary Figure S5E). For instance, SMAD2, TGIF1, SMAD4, SMAD3 and SP1 are key components of TGF- $\beta$  signaling, while ELK1, ELK4, ATF2, ATF4 and FOS are involved in MAPK signaling (Supplementary Figure S5E and Supplementary Data 5).

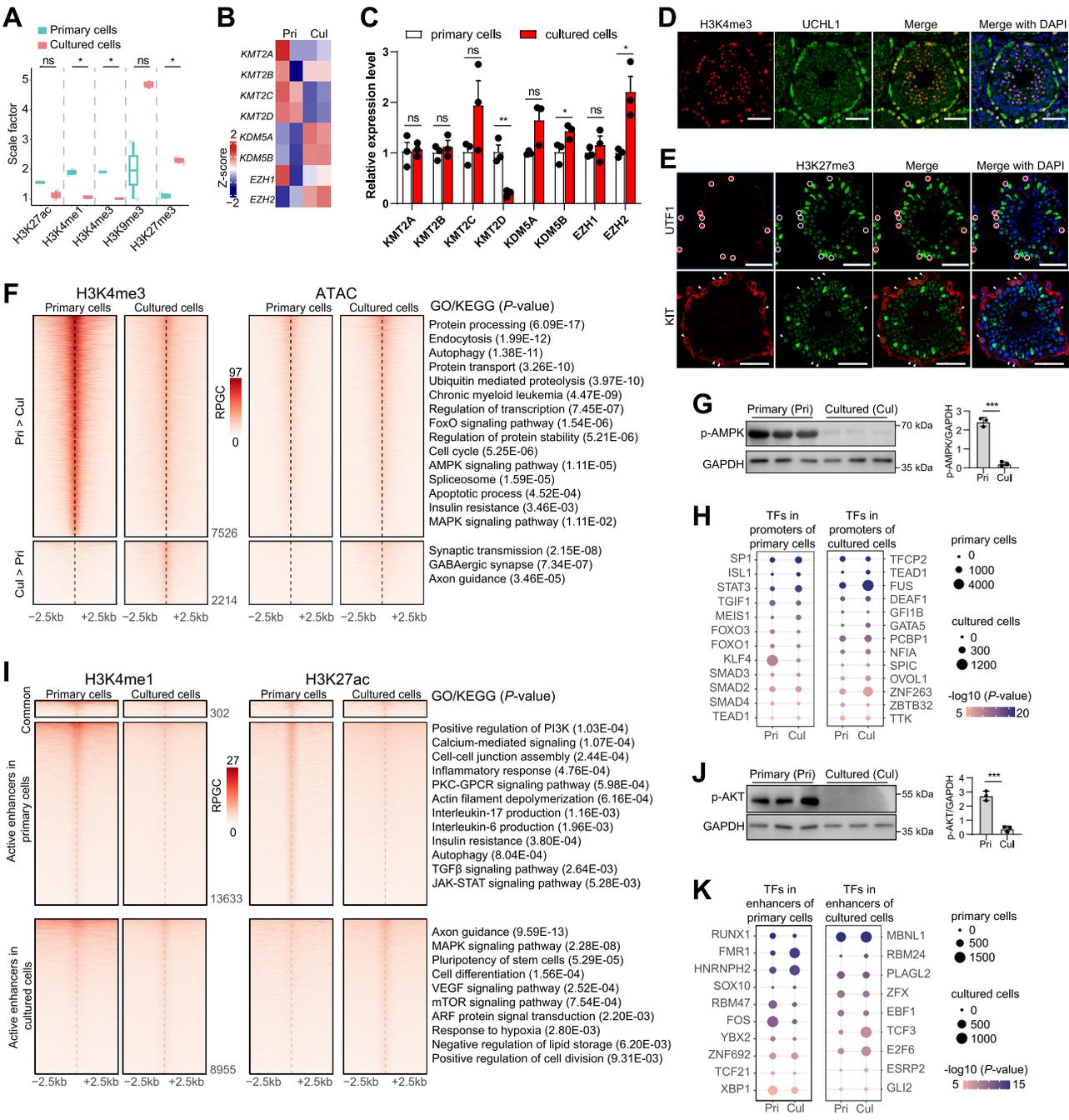
We then went on to investigate whether some of these regulators were preferentially expressed or activated in subset of putative monkey SSCs *in vivo*. Co-immunofluorescent staining showed that phosphorylated SMAD2/3 (p-SMAD2/3), which bind to DNA, were widely detected in germ cells. However, some UCHL1-expressing cells showed stronger p-SMAD2/3 signal (Supplementary Figure S5F). *LIN7B* is another marker of c1–c4 undifferentiated spermatogonia (Supplementary Figure S1G) and encodes a membrane protein (Figure 1H). We enriched two subtypes of undifferentiated spermatogonia, LIN7B<sup>+</sup>TSPAN33<sup>+</sup> and LIN7B<sup>+</sup>TSPAN33<sup>-</sup>. Immunoblotting analysis revealed that LIN7B<sup>+</sup>TSPAN33<sup>+</sup> spermatogonia, which is more primitive than LIN7B<sup>+</sup>TSPAN33<sup>-</sup> cells (Supplementary Figure S1G and J), contained higher level of p-SMAD2/3 and FOS (Supplementary Figure S5G). These observations suggested that some of the identified regulators, for instance p-SMAD2/3 and FOS, were in-

deed preferentially expressed in primitive undifferentiated spermatogonia.

### Primary TSPAN33<sup>+</sup> spermatogonia enriched for SSCs contain high levels of active histone modifications and low levels of repressive histone modifications

Gene expression is regulated by epigenetic modifications to DNA and histones. To elucidate the landscapes of major histone H3 modifications in monkey putative SSCs and identify the epigenetic characteristics relevant to stem cell identity, CUT&Tag sequencing was used to analyze H3K4me1, H3K4me3, H3K27me3, H3K27ac and H3K9me3 modifications in primary and cultured TSPAN33<sup>+</sup> spermatogonia (Supplementary Figure S6A). The profiles of histone modifications, with the exception of H3K9me3, exhibited high consistency across two biological replicates (Supplementary Figure S6B). The distribution patterns of each histone modification in both primary and cultured cells were summarized (Supplementary Figure S6C). Most peaks of H3K4me3, which is indicative of active gene transcription (53), were located in the promoter regions, accounting for > 60% in primary cells and > 30% in cultured cells (Supplementary Figure S6C). Most peaks for H3K4me1 and H3K27ac, which are indicative of enhancer regions (54), were identified in the distal intergenic (H3K4me1 and H3K27ac) and intronic regions (H3K27ac) (Supplementary Figure S6C). Although H3K27me3 and H3K9me3 act to silence genes (53), they displayed distinct genomic distribution patterns (Supplementary Figure S6C).

Upon comparing global histone modification between primary and cultured TSPAN33<sup>+</sup> spermatogonia, we found that primary cells contained higher levels of active histone modifications (H3K27ac, H3K4me1 and H3K4me3) and lower levels of repressive histone modifications (H3K9me3 and H3K27me3). Notably, the levels of H3K4me1, H3K4me3 and H3K27me3 differed significantly (Figure 3A). The KMT2A/B/C/D enzymes catalyze methyl transfer at histone H3K4 at promoters (KMT2A/B) or enhancers (KMT2C/D), while the KDM5A/B enzymes facilitate rapid demethylation of H3K4me3. Polycomb repressive complex 2 (PRC2) is a writer of H3K27me3 (16). In line with the histone modification change patterns, RNA-seq data revealed that *KMT2C/D* showed relatively higher expression in primary TSPAN33<sup>+</sup> cells than in cultured cells, whereas *KDM5A/B* showed much lower expression in primary than in cultured cells. Additionally, an increase in the expression of *EZH2*, the catalytic subunit of PRC2, was observed after culture (Figure 3B). The expression changes of *KMT2D*, *KDM5B* and *EZH2* were validated by qRT-PCR analysis (Figure 3C). Intriguingly, analysis using scRNA-seq data also showed that *KDM5B* and *EZH2* expression increased during the transition from c1 (putative SSCs) to c5 (Supplementary Figure S7). These findings suggest that similar patterns of histone modification changes may occur during *in vivo* SSC differentiation. To test this hypothesis, we performed co-immunostaining on monkey testicular sections. Intriguingly, H3K4me3 was detected in UCHL1<sup>+</sup> cells, and the intensity of UCHL1 and H3K4me3 exhibited positive correlation (Figure 3D). While H3K27me3 signal was barely detected in UTF1<sup>+</sup> cells, obvious staining was seen in differentiated germ cells including KIT<sup>+</sup> cells and sertoli cells (Figure 3E). These results support the notion that monkey SSC differ-



**Figure 3.** Histone modification changes between primary and cultured monkey TSPAN33<sup>+</sup> cells. **(A)** Comparison of overall histone modification levels between primary and cultured monkey TSPAN33<sup>+</sup> cells. Scale factor was counted by ChIPseqSpikInFree. **(B)** Expression levels of key genes regulating histone modifications in primary and cultured monkey TSPAN33<sup>+</sup> cells. **(C)** qRT-PCR analysis of genes regulating histone modifications in primary and 5-day cultured TSPAN33<sup>+</sup> cells. The relative expression levels of each gene were normalized by *GAPDH*. **(D)** Immunofluorescence staining of testicular sections of adult monkeys with H3K4me3 and UCHL1 (a broad marker for undifferentiated spermatogonia). Scale bars, 50  $\mu$ m. **(E)** Immunofluorescence staining of testicular sections of adult monkeys with H3K27me3, UTF1 (marker for undifferentiated spermatogonia) and KIT (marker for differentiated spermatogonia). Scale bars, 50  $\mu$ m. **(F)** Heatmap of differential H3K4me3 signals (fold-change > 1.5,  $P < 0.05$ ) at promoter regions between primary and cultured monkey TSPAN33<sup>+</sup> cells. Related chromatin open status and gene-enriched GO/KEGG pathways are shown. **(G)** Immunoblotting analysis revealed that AMPK phosphorylation (p-AMPK) was higher in primary TSPAN33<sup>+</sup> germ cells than in cultured TSPAN33<sup>+</sup> germ cells. Left panel shows images, right panel shows quantification of p-AMPK level normalized to *GAPDH*. **(H)** TFs enriched in differential H3K4me3 promoter regions in primary and cultured TSPAN33<sup>+</sup> cells, respectively. **(I)** Heatmap of active enhancer regions in primary and cultured monkey TSPAN33<sup>+</sup> cells. GO/KEGG pathways enriched in enhancer-associated genes. **(J)** Immunoblotting analysis revealed that AKT phosphorylation (p-AKT) level was higher in primary TSPAN33<sup>+</sup> germ cells than in cultured TSPAN33<sup>+</sup> germ cells. Left panel shows image, right panel shows quantification of p-AKT level normalized to *GAPDH*. **(K)** TFs enriched in specific enhancers of primary and cultured TSPAN33<sup>+</sup> cells, respectively. In panels (C), (G) and (J), data are shown as mean  $\pm$  SD from three independent replicates. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; ns, not significant; two-tailed Student's *t*-test. Also see [Supplementary Figures S6–S8](#).

entiation is accompanied by the loss of H3K4me3 and gain of H3K27me3 modification. In sharp contrast, H3K27me3 displays global loss during spermatogonia differentiation in mice. H3K27me3 is an epigenetic hallmark of mouse SSCs and maintains the self-renewal state by suppressing the differentiation program (55). The different change patterns of H3K27me3 between primates and rodents highlight the distinct regulatory mechanisms of spermatogonia differentiation. Future studies should be performed to illustrate whether H3K27me3 regulates spermatogonia differentiation by suppressing self-renewal program in primates.

### Promoters, active enhancers and potential regulators in putative SSCs

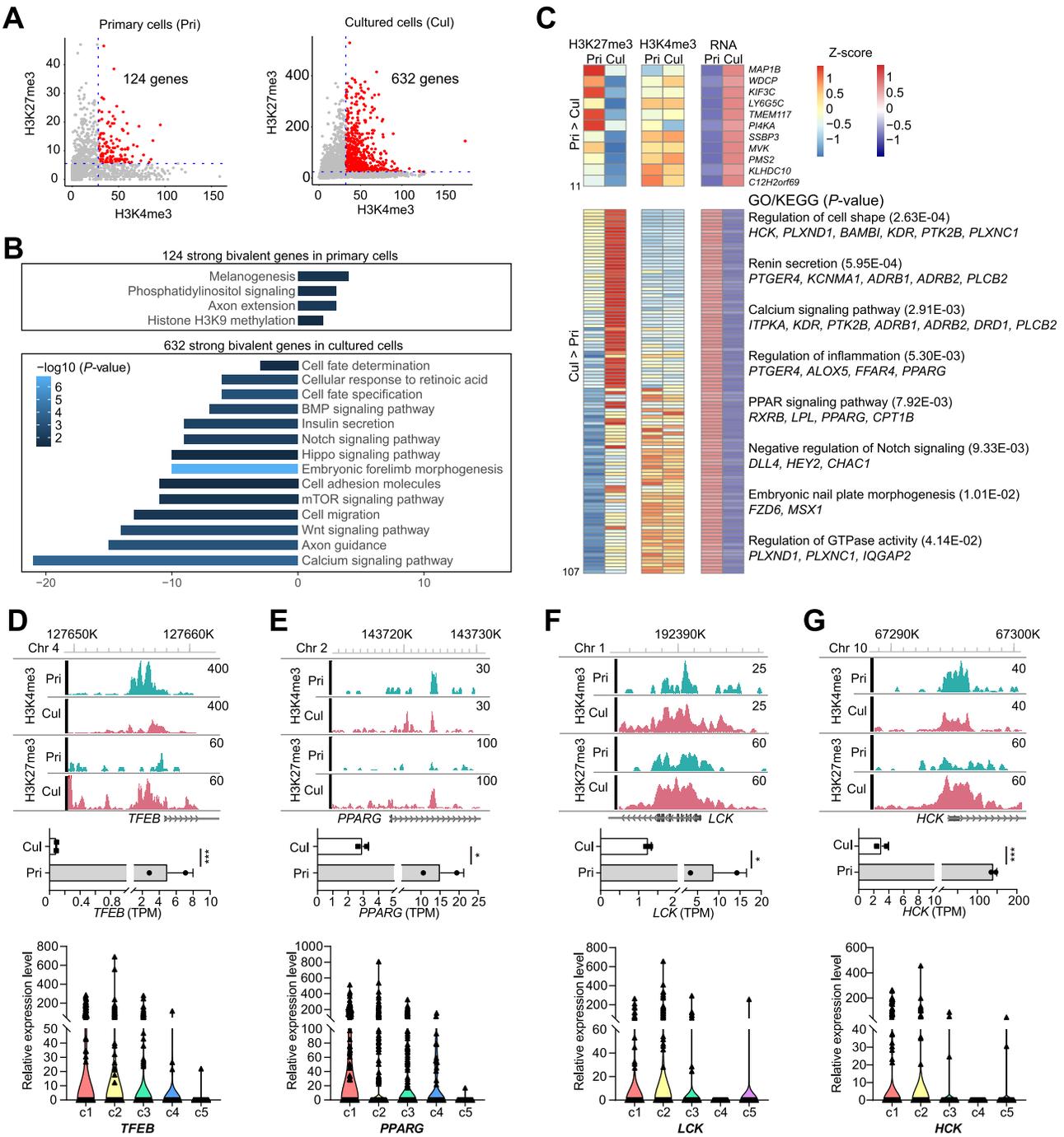
H3K4me3 at promoter regions is instrumental in the regulation of RNA polymerase II pause-release, ensuring transcription elongation and output (56). Moreover, H3K4me3 counteracts DNA methylation and other repressive histone modifications, such as H3K9me3 (57,58). Thus, H3K4me3 modification at the promoter regions was closely investigated. In total, 9740 promoter regions showed significantly different H3K4me3 levels between primary and cultured TSPAN33<sup>+</sup> cells (fold-change > 1.5,  $P < 0.05$ ). Among these regions, 7526 promoters (corresponding to 6288 genes) with higher levels of H3K4me3 were detected in the primary cells (Figure 3F, Supplementary Figure S8A and Supplementary Data 6). These promoters were associated with genes involved in various biological processes, including autophagy, FOXO, AMPK and MAPK signaling and insulin resistance pathways (Figure 3F and Supplementary Data 6). Some of these processes are enriched in undifferentiated spermatogonia (7). Specifically, AMPK is a central regulator of metabolic homeostasis by promoting catabolic pathways (e.g. autophagy) (59). Immunoblotting analysis validated that AMPK activity was higher in primary TSPAN33<sup>+</sup> spermatogonia than in cultured spermatogonia (Figure 3G), suggesting that *in vivo* putative SSCs might possess higher AMPK activity compared to germ cells at advanced developmental stages. Intriguingly, although these promoter regions contained differential level of H3K4me3 between primary and cultured cells, they showed similar open chromatin status (Figure 3F), indicating that loss of H3K4me3 modification may be an early event driving gene expression changes and SSC differentiation. Compared to primary cells, 2214 promoters (corresponding to 1417 genes) exhibiting high levels of H3K4me3 were identified in the cultured cells (Figure 3F, Supplementary Figure S8A and Supplementary Data 6). These promoters are associated with genes involved in synaptic transmission and axon guidance (Figure 3F and Supplementary Data 6). Subsequent motif enrichment analysis of these differentially modified promoter regions identified a list of potential motifs and binding TFs in the primary and cultured cells, respectively (Figure 3H and Supplementary Data 6). Interestingly, TFs enriched in primary cells were predicted to participate in TGF- $\beta$  signaling, mitophagy and JAK-STAT cascades (Supplementary Figure S8B). For instance, SP1 transcriptional activity can be modulated by a wide range of signaling pathways, including MAPK/ERK, MAPK/p38, JAK/STAT and PI3K/AKT (60,61), while STAT3 is an integral component of the JAK-STAT pathway.

Enhancers are pivotal for the regulation of gene expression and can function at long distance. Active enhancers can

be distinguished by the combined presence of H3K4me1 and H3K27ac modifications. Here, 13 633 and 8955 active enhancers (enhancers hereafter) were identified in the primary and cultured cells, respectively (Figure 3I, Supplementary Figure S8C and Supplementary Data 7). As no 3D-chromatin interaction data are currently available for undifferentiated spermatogonia of rhesus monkeys, we preliminarily defined nearby genes ( $\pm 150$  kb) as potential target genes of these enhancers, yielding 5067 and 3647 associated genes in the primary and cultured TSPAN33<sup>+</sup> cells, respectively (Supplementary Data 7). The associated genes in the primary cells were involved in PI3K signaling, inflammatory response, IL6 and IL17 production, autophagy and TGF- $\beta$  signaling (Figure 3I and Supplementary Data 7). Among these signals, we validated that PI3K activity, monitored by the phosphorylation of its effector AKT (p-AKT), was higher in primary TSPAN33<sup>+</sup> spermatogonia than in cultured cells (Figure 3J). The associated genes in the cultured cells were enriched in axon guidance, MAPK signaling, cell differentiation, VEGF pathway and mTOR signaling (Figure 3I and Supplementary Data 7). Of note, elevated mTOR signaling is known to induce SSC differentiation in mice (62). We then performed motif enrichment analysis and identified potential binding TFs on these enhancers in primary and cultured cells (Figure 3K and Supplementary Data 7). Notably, RUNX1, which is expressed in primary TSPAN33<sup>+</sup> cells and the transcription level decreased after culture (Supplementary Figure S8D), was significantly enriched in enhancers of primary cells (Figure 3K). RUNX1 binds to enhancers and fine-tunes the balance among cell proliferation, differentiation and cell cycle exit in multiple tissue stem cells, including hematopoietic, hair follicle, mammary, gastric and neural crest stem cells (63). Thus, we speculated that RUNX1 may also play roles in controlling the behavior of monkey putative SSCs. Taken together, our study elucidated the landscape of several key histone modifications and identified promoters and active enhancers in rhesus monkey putative SSCs. Importantly, we predicted potential TFs and signaling pathways that may play essential roles in sustaining putative SSC identities.

### Bivalent modification of stem cell regulators during early-stage differentiation

Bivalency refers to the co-occurrence of H3K4me3 and H3K27me3 at the promoters of developmental genes characterized by low expression, and is a key feature of germline cells, embryonic stem cells and other tissue stem and progenitor cells (64,65). Bivalent modification of promoters primes key regulatory genes for future activation or repression, serving as a mechanism to maintain epigenetic plasticity (64,65). Here, we analyzed bivalent modifications in primary and cultured TSPAN33<sup>+</sup> cells. Bivalent promoters were determined based on the following criteria: (1) Over 80% overlap of H3K4me3 and H3K27me3 peaks within the promoter region ( $\pm 3$  kb from the TSS region); and (2) Low RNA expression level of related genes (TPM < 5). A total of 1530 and 4850 regions (corresponding to 1514 and 4686 genes) were categorized as bivalent promoters in the primary and cultured cells, respectively (Figure 4A and Supplementary Data 8). An in-depth examination of the strong bivalent modifications was conducted, focusing on signals surpassing the average peaks of H3K4me3 and H3K27me3 in housekeeping genes (66). A total of 124 and 634 promoter regions (corresponding to 124



**Figure 4.** Bivalent modification of stem cell regulators during early differentiation. **(A)** Plots of H3K4me3 and H3K27me3 levels for bivalent genes identified in primary and cultured monkey TSPAN33<sup>+</sup> cells. Highlighted spots represent genes with strong bivalent modification. **(B)** Pathways enriched in strong bivalent genes in primary and cultured monkey TSPAN33<sup>+</sup> cells. **(C)** Heatmap of H3K27me3, H3K4me3 and RNA expression levels for strong bivalent genes exhibiting bivalency status change (H3K27me3 fold-change > 1.5) between primary and cultured TSPAN33<sup>+</sup> cells. GO/KEGG pathways enriched in corresponding genes are shown. **(D–G)** Snapshots of genome browser showing distributions of histone modifications and RNA levels of genes including *TFEB* **(D)**, *PPARG* **(E)**, *LCK* **(F)** and *HCK* **(G)**. Genome browser view scales were adjusted based on global data range. Cul, cultured cells; Pri, primary cells. \* $P < 0.05$ ; \*\*\* $P < 0.001$ , from RNA-seq data showed in [Supplementary Data 4](#).

and 632 genes) with strong bivalent modification were identified in the primary and cultured cells, respectively (Figure 4A and Supplementary Data 8). The strong bivalent genes identified in the primary TSPAN33<sup>+</sup> cells were involved in developmental events and H3K9 methylation, while those identified in the cultured cells were enriched in calcium, Wnt, mTOR, Hippo, Notch, insulin and BMP signaling (Figure 4B and Supplementary Data 8).

Within the cohort of strong bivalent genes, genes with alterations in bivalency status (H3K27me3 fold-change > 1.5) between primary and cultured cells were further assessed. Specifically, 11 genes presented a decline in H3K27me3 modification and a corresponding increase in RNA expression following culture (Figure 4C), with several implicated in differentiation events. For instance, *MAP1B* is a key regulator of axon guidance and neuronal differentiation (67), *PI4KA* is a key regulator of erythroid cell differentiation (68) and *MVK* is a primary modulator of keratinocyte differentiation (69). Thus, bivalency appears to repress the expression of differentiation genes in monkey undifferentiated spermatogonia enriched for SSCs. In contrast, 107 genes displayed an increase in H3K27me3 modification and a decrease in RNA expression following culture (Figure 4C and Supplementary Data 8). Intriguingly, several biological processes significantly represented in primary cells (e.g. inflammatory response, calcium and PPAR signaling) were also enriched in these genes (Figure 4C and Supplementary Data 8). Specifically, c1 cell (putative SSCs)-prevalent genes *TFEB* and *PPARG* (Supplementary Data 1), which regulate lysosomal function and metabolism (70,71), exhibited increased bivalent modification post-culture (Figure 4D, E). *TFEB* is a major effector of AMPK (59), whose activity was high in TSPAN33<sup>+</sup> undifferentiated spermatogonia (Figure 3G). Previous studies on hematopoietic stem cells have underscored the critical roles of *TFEB* and *PPARG* in preserving stem cell identity and promoting stem cell self-renewal (70,71). Moreover, genes in the Src family, including *LCK* and *HCK*, also showed increased bivalent modification post-culture (Figure 4F and G). Src family kinases were shown to be activated by GDNF, and played roles in self-renewal of mouse SSCs (72). These observations suggest that bivalency functions as a means to repress stem cell maintenance regulators and pathways in a reversible manner during early differentiation, potentially granting early progenitor cells with plasticity to de-differentiate and revert to a stem cell state.

### Key TFs regulating signature gene expression in putative SSCs

Using *in vitro* cultured TSPAN33<sup>+</sup> cells (progenitor-like cells) for comparison, we identified a list of TFs and signaling pathways implicated in regulation of putative SSCs. *In vivo* scRNA expression data were also used to elucidate the regulators of SSCs. ScRNA-seq identified 545 signature genes in c1 cells (primitive SSCs) (Supplementary Data 1). Signature genes are highly relevant to cell identity and their upstream regulators may provide insights into cell identity regulation. Thus, the promoters and enhancers of the 545 signature genes of c1 cells were investigated using the H3K4me3, H3K4me1 and H3K27ac profiling data obtained from fresh TSPAN33<sup>+</sup> spermatogonia (primary cells).

H3K4me3 peaks at promoter regions were detected in 456 of the 545 c1 signature genes (Supplementary Data 9). Mo-

tif enrichment analysis identified 24 motifs and TFs in these regions (Supplementary Data 9). Of interest, the motif for peroxisome proliferator-activated receptor- $\alpha$  (PPARA) ranked the highest (Figure 5A). PPARA, along with PPARG, belongs to the PPAR family and regulates the transcription of genes involved in lipid and glucose metabolism (73). It also activates autophagy by directly binding to and regulating the transcription of core autophagy machinery gene loci (74). For instance, *TFEB*, a master regulator of lysosomal biogenesis, is a direct target of PPARA (75). To further understand the regulatory modules of primitive undifferentiated spermatogonia, a protein-protein network was constructed for these TFs using the STRING database (41). Functional enrichment analysis of the network indicated that the TFs participated in metabolic processes, SMAD binding and cell fate commitment (Figure 5B). Similarly, active enhancers were detected for 185 of the 545 c1 signature genes (Supplementary Data 9), with subsequent motif enrichment analysis identifying 68 motifs and TFs (Supplementary Data 9). The most significantly enriched TFs, including *HOXA2*, *FOSL2*, *JUN*, *FOS*, *JUNB* and *NFKB1* (Figure 5C), were also hub factors in the protein-protein regulatory network, with predominant enrichment in metabolic processes (Figure 5D).

Overlapping the regulatory TFs in putative SSCs enriched by using *in vitro* cultured TSPAN33<sup>+</sup> cells for comparison or by analyzing the signature genes of c1 cells, 7 and 24 common regulatory TFs were identified to potentially bind to the promoter or enhancer regions of monkey putative SSCs, respectively (Figure 5E). Subsequent protein-protein network construction and functional enrichment analysis revealed that hub TFs, including *FOS*, *ETS1*, *JUNB*, *ATF3* and *SMAD4*, were significantly enriched in the TGF- $\beta$  signaling pathway (Figure 5F).

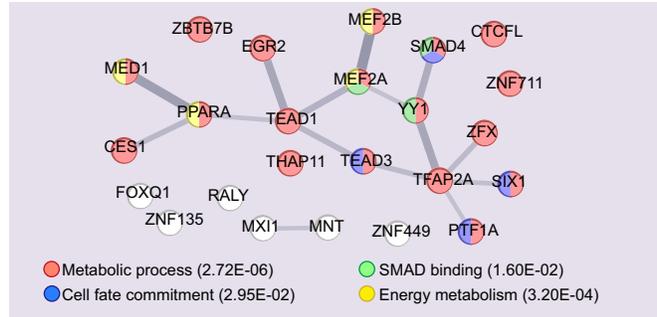
### Multi-omics analyses reveal activation of TGF- $\beta$ signaling in monkey putative SSCs

Among the TFs identified through motif enrichment analyses, SMAD2/3/4, key components of the TGF- $\beta$  signaling pathway, were consistently and predominantly enriched in putative SSCs (Figures 2H, 3H, 5B and 5F), suggesting that TGF- $\beta$  signaling may be active in monkey SSCs. Similarly, in humans, primitive SSCs exhibit elevated expression of TGF- $\beta$  receptors (*TGFBR1* and *TGFBR2*) (14), whereas, in mice, TGF- $\beta$  receptors are detected in differentiated spermatogonia preparing for meiotic entry (11). These differential expression patterns suggest diverse functions of TGF- $\beta$  signaling in spermatogenesis in rodents and primates. To further characterize TGF- $\beta$  signaling, we analyzed scRNA-seq data and showed that *TGFBR1* expression was highest in the c1 cells (putative SSCs) compared to the other cell clusters (Figure 6A). Multi-omics analysis of primary versus cultured monkey TSPAN33<sup>+</sup> spermatogonia also suggested the active expression of *TGFBR1* in primary cells (Figure 6B), as validated by qRT-PCR (Figure 6C) and immunoblotting (Figure 6D). The TGF- $\beta$  pathway activation can be monitored through SMAD2/3 phosphorylation (p-SMAD2/3) (76). Indeed, the phosphorylation of SMAD2/3 (p-SMAD2/3) was higher in primary TSPAN33<sup>+</sup> cells than in cultured cells (Figure 6D). Altogether, these lines of evidence support the activation of TGF- $\beta$  signaling in monkey putative SSCs.

**A** Top 5 TFs enriched in promoters of c1 signature genes

TFs	Type	P-value	Motif
PPARA	known	1.00E-03	
MNT	known	1.00E-02	
MXI1	de novo	1.00E-13	
MED1	de novo	1.00E-12	
ZBTB7B	de novo	1.00E-10	

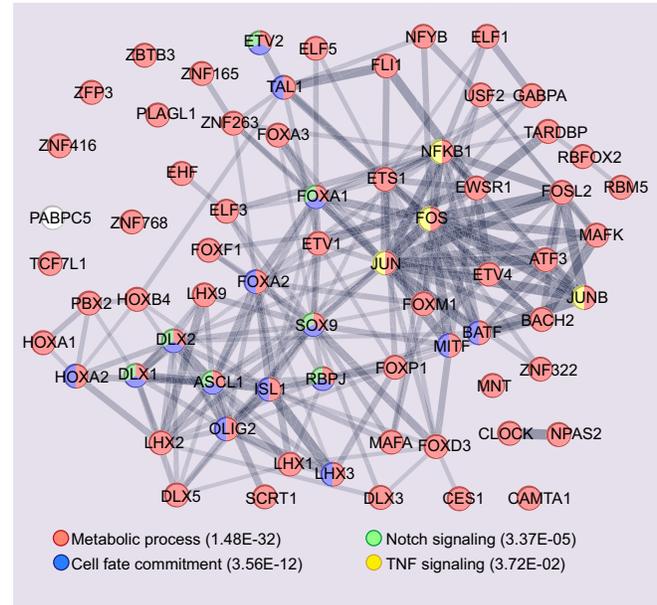
**B** 24 TFs enriched in promoters of c1 signature genes



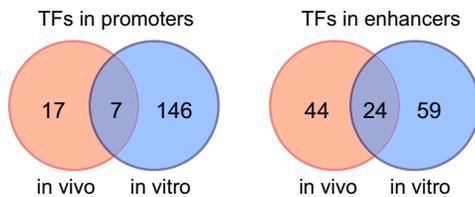
**C** Top 10 TFs enriched in enhancers of c1 signature genes

TFs	Type	P-value	Motif
HOXA2	known	1.00E-17	
FOSL2	known	1.00E-14	
JUN	known	1.00E-13	
FOS	known	1.00E-13	
JUNB	known	1.00E-13	
RBM5	de novo	1.00E-65	
NFKB1	de novo	1.00E-62	
PABPC5	de novo	1.00E-62	
PLAGL1	de novo	1.00E-56	
NFYB	de novo	1.00E-45	

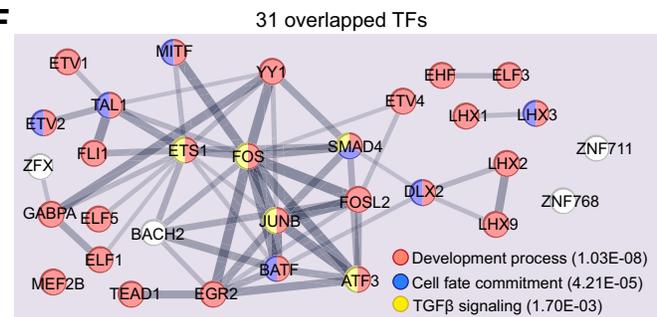
**D** 68 TFs enriched in enhancers of c1 signature genes



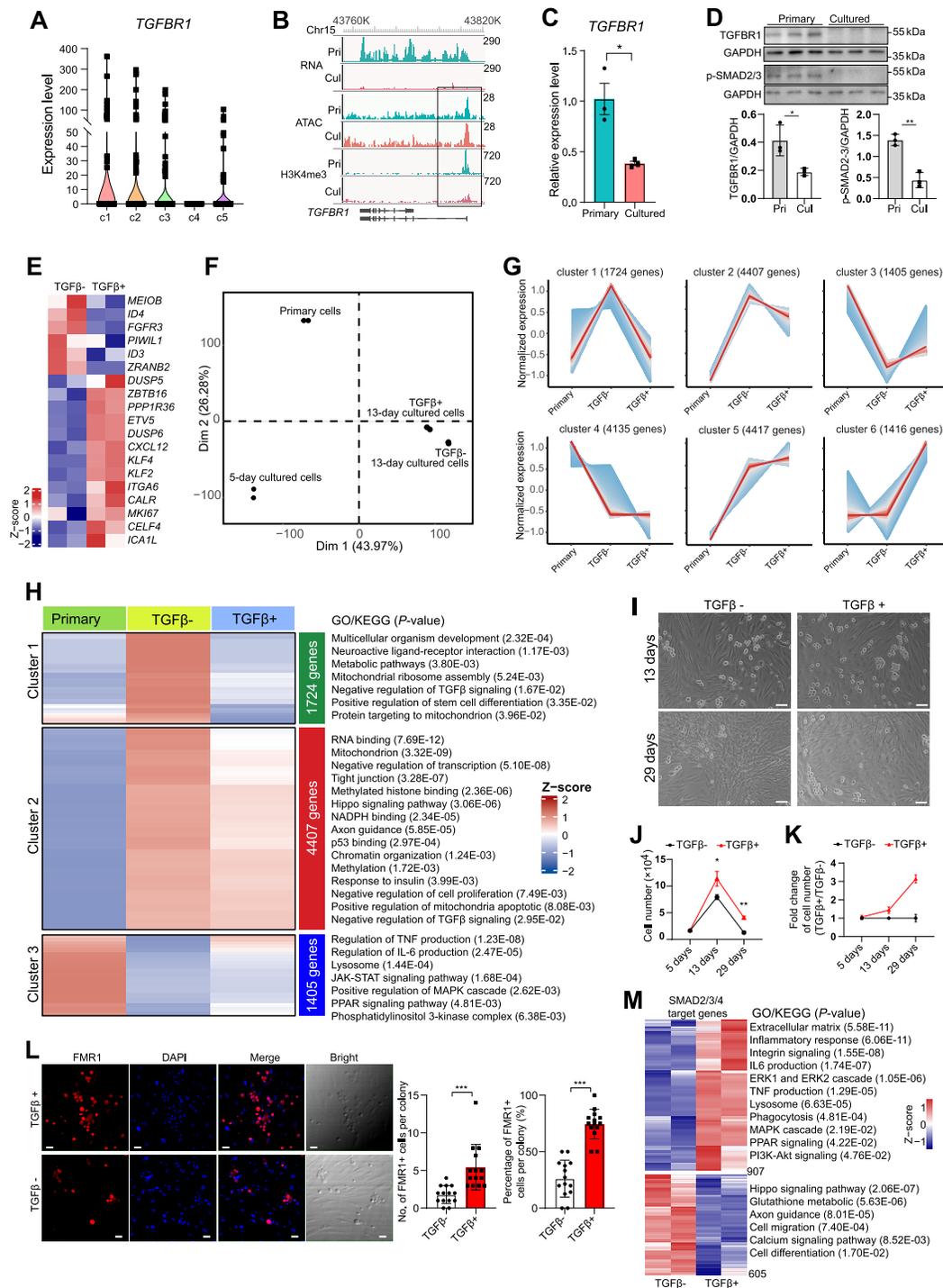
**E**



**F**



**Figure 5.** Key TFs regulating signature gene expression in primitive undifferentiated spermatogonia (c1 cells) in monkeys. **(A)** Top TFs enriched in promoters of c1 signature genes. **(B)** Protein–protein association and functional enrichment analyses of TFs potentially enriched at the c1 signature gene promoters. Value of false discovery rate (FDR) is shown for pathways. **(C)** Top TFs enriched in active enhancers of c1 signature genes. **(D)** Protein–protein association and functional enrichment analyses of TFs identified at active enhancers of c1 signature genes. FDR is indicated for pathways. **(E)** Venn diagram of TFs. *In vivo*, TFs identified at promoters/enhancers of c1 signature genes; *in vitro*, TFs enriched in specific promoters/enhancers of primary TSPAN33<sup>+</sup> cells compared with cultured TSPAN33<sup>+</sup> cells. **(F)** Protein–protein association and functional enrichment analyses of common regulatory TFs identified in **(E)**. FDR is shown for enriched pathways.



**Figure 6.** TGF- $\beta$  signaling promotes the *in vitro* expansion of undifferentiated spermatogonia. **(A)** Expression plots of *TGFBR1* in five cell clusters. Expression data were from our scRNA-seq. **(B)** Snapshots of genome browser showing distributions of histone modifications, chromatin open status and RNA levels of *TGFBR1* in primary (Pri) and cultured (Cul) monkey TSPAN33<sup>+</sup> cells. Genome browser view scales were adjusted based on global data range. **(C)** qRT-PCR analysis of relative *TGFBR1* expression (normalized by *GAPDH*) in primary and 5-day cultured TSPAN33<sup>+</sup> cells. **(D)** Immunoblotting analysis of TGFBR1 and phosphorylated SMAD2/3 levels in primary TSPAN33<sup>+</sup> germ cells and cultured TSPAN33<sup>+</sup> germ cells. Upper panel shows image and lower panel shows quantification normalized to *GAPDH*. **(E)** Expression changes of spermatogonia markers in TSPAN33<sup>+</sup> cells cultured for 13 days with or without TGF- $\beta$  treatment (1 ng/mL). **(F)** PCA of primary TSPAN33<sup>+</sup> cells, TSPAN33<sup>+</sup> cells cultured for 5 days and TSPAN33<sup>+</sup> cells cultured for 13 days with or without TGF- $\beta$  treatment. **(G)** DEGs among three cell groups (primary TSPAN33<sup>+</sup> cells, TSPAN33<sup>+</sup> cells cultured for 13 days with or without TGF- $\beta$  treatment) were classified into six clusters (c1–c6). **(H)** GO terms/KEGG pathways enriched in DEGs in c1–c3. **(I)** Bright views of cells cultured with or without TGF- $\beta$  treatment. Scale bars, 50  $\mu$ m. **(J, K)** TGF- $\beta$  treatment increased spermatogonia proliferation in culture. Cell numbers **(J)** and cell number fold changes between groups with or without TGF- $\beta$  treatment **(K)** were shown. **(L)** Immunofluorescence staining of FMR1 protein in germ cells cultured with or without TGF- $\beta$  for 29 days (*left*). The numbers and the percentages of FMR1<sup>+</sup> cells in colonies (14 cell colonies for each culture condition) were analyzed (*right*). Scale bars, 50  $\mu$ m. **(M)** Heatmap of RNA expression levels of TGF- $\beta$  target genes. GO terms and KEGG pathways enriched in TGF- $\beta$  up or downregulated targets. In panels (C), (D), (J), (K) and (L), data are shown as mean  $\pm$  SD from three independent replicates. \* $P$  < 0.05; \*\* $P$  < 0.01; \*\*\* $P$  < 0.001; two-tailed Student's *t*-test. Also see [Supplementary Figure S9](#).

## TGF- $\beta$ signaling promotes the *in vitro* expansion of putative SSCs

To clarify the function of TGF- $\beta$  signaling in monkey putative SSCs, TGF- $\beta$  was added in the culture medium, with its impacts on gene expression and spermatogonia proliferation then examined. TSPAN33<sup>+</sup> spermatogonia were enriched by magnetic beads and cultured in medium supplemented with or without 1 ng/mL TGF- $\beta$ . Cells were harvested after 5 and 13 days in culture, respectively, for downstream examination. Based on qRT-PCR analysis, supplementing TGF- $\beta$  for 13 days significantly increased the transcription of undifferentiated spermatogonia marker genes, including *TSPAN33* and *MSL3* (Supplementary Figure S9A). RNA-seq was further conducted to compare the gene expression profiles of cells cultured for 13 days with or without TGF- $\beta$  treatment. In total, 1622 DEGs were identified, including 676 downregulated genes and 946 upregulated genes in the presence of TGF- $\beta$  (adjusted  $P < 0.05$ , fold-change  $> 1.5$ ) (Supplementary Figure S9B and Supplementary Data 10). Specifically, TGF- $\beta$  treatment increased the expression of multiple markers of undifferentiated spermatogonia (Figure 6E). *MKI67*, a marker of cell proliferation, was also increased by TGF- $\beta$  treatment (Figure 6E). Genes upregulated by TGF- $\beta$  were involved in phagocytosis, IL6 and TNF production, ERK cascade and immune response (Supplementary Figure S9B and Supplementary Data 10). To better understand the effect of TGF- $\beta$  on culture, we included the RNA-seq of primary TSPAN33<sup>+</sup> cells for comparison. PCA showed that compared to cultured cells without TGF- $\beta$  treatment, cells cultured with TGF- $\beta$  were closer to primary cells (Figure 6F). DEGs among the three groups were classified into six clusters (Figure 6G and Supplementary Data 11). Compared to primary TSPAN33<sup>+</sup> cells, culture for 13 days without TGF- $\beta$  induced the expression changes of 16 088 genes (c1–c5). Of note, supplying TGF- $\beta$  in culture could completely (c1) or partially (c2, c3) reverse the expression changes of 47% affected genes (c1–c3). Specifically, c3 genes involved in biological processes including IL6 and TNF production, lysosome, JAK-STAT signal, PI3K and PPAR pathways, were highly expressed in primary TSPAN33<sup>+</sup> spermatogonia and activated by TGF- $\beta$  (Figure 6H). These lines of evidence suggested that TGF- $\beta$  signal might play beneficial role on monkey SSCs.

We then assessed the effect of TGF- $\beta$  signal on the proliferation of monkey undifferentiated spermatogonia in culture. Magnetic beads enriched TSPAN33<sup>+</sup> cells were cultured with or without 1 ng/mL TGF- $\beta$ , and cell counts were analyzed at days 5, 13 and 29. Cell division and colonies were observed at about day 8, and active cell proliferation could be observed during days 5–15 (Figure 6I–K). At days 13 and 29, the addition of TGF- $\beta$  led to an increase in total cell number when compared to culture without TGF- $\beta$  at each time-point (Figure 6J and K). The cultured TSPAN33<sup>+</sup> cells were all positive for *DDX4* at day 29 (Supplementary Figure S9C). Moreover, the numbers and proportions of *FMR1*<sup>+</sup> undifferentiated spermatogonia in colonies were significantly increased by treatment with TGF- $\beta$  (Figure 6L). Consistently, TGF- $\beta$  treatment for 29 days resulted in an increase in the mRNA expression levels of undifferentiated spermatogonia markers (*TSPAN33*, *ZBTB43* and *MSL3*) (Supplementary Figure S9D). Taken together, these findings demonstrate that TGF- $\beta$  signaling stimulates the self-renewal of monkey undifferentiated spermatogonia. With addition of TGF- $\beta$ , monkey spermatogonia colonies

could be seen after 8–9 weeks of culture. However, few colonies were observed after 5–6 weeks of culture without TGF- $\beta$ .

The TGF- $\beta$  signaling pathway is mediated by SMAD2/3/4. We identified 1512 potential SMAD2/3/4 target genes, which contained either one of the SMAD2/3/4 binding motifs in the promoter region and displayed altered expression upon TGF- $\beta$  treatment in culture (Figure 6M and Supplementary Data 10). Among the targets, 907 genes were upregulated by TGF- $\beta$  and were enriched in phagocytosis, TNF production, MAPK activity, lysosome, PPAR signaling and PI3K activity (Figure 6M and Supplementary Data 10). Notably, these biological processes were also highly represented in TSPAN33<sup>+</sup> undifferentiated spermatogonia (Figure 2F). Conversely, 605 potential target genes were downregulated by TGF- $\beta$  and enriched in Hippo signaling, glutathione metabolism, axon guidance, cell migration, calcium signaling and cell differentiation (Figure 6M). Taken together, these results suggest that TGF- $\beta$  might function by activating and suppressing numerous gene expressions to regulate self-renewal of monkey SSCs.

## Discussion

Due to the scarcity of human testicular samples, studies on the regulation of human SSC homeostasis are very limited. Although several recent studies have depicted gene expression changes during the whole spermatogenesis process in humans and monkeys (7–9,11,12), little is known regarding epigenetic characteristics in human or monkey SSCs, despite similarities in the properties and spermatogenesis processes of SSCs between these species (10–12). In this study, the open chromatin status, major histone modifications (H3K4me1, H3K4me3, H3K27me3, H3K27ac and H3K9me3) and distribution of genome-wide promoters and active enhancers in rhesus monkey putative SSCs were explored. Epigenetic differences between primary TSPAN33<sup>+</sup> cells (enriched for putative SSCs) and cultured cells (resembling differentiating progenitors) were assessed, as were the regulatory elements of monkey putative SSC signature genes. Results identified a list of TFs and signaling pathways (e.g. PI3K-AKT, JAK-STAT and TGF- $\beta$  signaling pathways) potentially implicated in regulating SSC self-renewal. Notably, the activation of TGF- $\beta$  signaling in monkey SSCs was confirmed. Evidence supporting its role in promoting self-renewal of monkey SSCs was also provided. Taken together, our study generated comprehensive information on the regulation of monkey SSC homeostasis, offering critical insights into the optimization of culture conditions for long-term expansion of human SSCs. In addition, we identified the distinct role of H3K27me3 in regulating spermatogenesis in monkey and mouse. During spermatogonia differentiation, H3K27me3 level is decreased in mouse, but increased in monkey. H3K27me3 is an epigenetic hallmark of mouse SSCs and maintains the self-renewal state by suppressing the differentiation program (55). However, H3K27me3 is low in monkey SSCs and might play roles in differentiation.

After 5 days of culture, TSPAN33<sup>+</sup> cells lost their stem cell properties, instead resembling *in vivo* progenitor cells. These progenitor-like cells underwent several rounds of proliferation in culture. Comparing putative SSCs to these progenitor-like cells, rather than *KIT*<sup>+</sup> differentiated spermatogonia, revealed the immediate changes that occurred during early differentiation of SSCs, resulting in the following observations: (i) The early differentiation stage exhibited marked epigenetic

changes, consistent with the dynamic epigenetic programming observed between SSCs and progenitors freshly isolated from mouse testes (77). Notably, the primary TSPAN33<sup>+</sup> cells obtained from monkeys exhibited higher levels of active histone modifications and lower levels of repressive histone modifications compared to cultured cells, suggesting that putative SSCs possess relatively open chromatin that compacts upon differentiation. This was validated by immunostaining of monkey testicular sections with H3K27me3 and H3K4me3. H3K4me3 undergoes rapid turnover, dependent on KDM5 (56), while H3K27me3 is catalyzed by the PRC2 complex (53). Consistent with the global changes in H3K4me3 and H3K27me3 in cultured cells, *KDM5A/B* and *EZH2* expression were elevated during culture. Based on these observations, we speculated that suppressing the activities of KDM5 and EZH2 by inhibitors may facilitate the maintenance of epigenetic status and SSC identity in culture. (ii) Certain stem cell-prevalent genes were silenced in cultured TSPAN33<sup>+</sup> cells (progenitor-like cells) through bivalency, which is intriguing as bivalent modification often occurs in developmental regulators. Bivalency is plastic and gene expression can be reversed by suitable cues (64,65). Bivalency can also prevent inappropriate and irreversible silencing of genes by DNA methylation or heterochromatin formation (65). Therefore, we hypothesized that bivalent modification of stem cell genes may serve as a mechanism to render progenitor cells with developmental plasticity.

Multi-omics analysis identified several signaling pathways implicated in the regulation of monkey putative SSCs. Both scRNA-seq and active enhancer analysis identified enrichment of PI3K-AKT signaling in putative SSCs, a finding supported by the prevalent PI3K activation observed in primary TSPAN33<sup>+</sup> cells. However, Tan *et al.* (14) reported increased AKT activation in human KIT<sup>+</sup> differentiated spermatogonia compared to primitive SSCs, noting that AKT signaling inhibition promoted *in vitro* expansion of human undifferentiated SSCs. In mice, PI3K activation occurs in a specific subset of SSCs to drive the cell cycle, and the PI3K-AKT-mTORC1 axis plays a critical role in mouse SSC self-renewal and differentiation (62). Although SSCs can be categorized into multiple subsets, the specific subset of primate SSCs in which the PI3K pathway is activated remains undetermined. Beyond AKT, PI3K regulates other downstream targets and executes distinct functions (78). For instance, *ex vivo* expansion of human hematopoietic stem cells requires PI3K but not AKT activation (79). Although AKT activation drives SSC differentiation (14,62), whether PI3K activation is essential for primate SSC self-renewal warrants further investigation.

The TFs SMAD2/3/4 are core components of TGF- $\beta$  signaling, binding to DNA to regulate target gene expression upon pathway activation (80,81). Multi-omics analysis revealed that the binding motifs for SMAD2/3/4 were repeatedly enriched in primary TSPAN33<sup>+</sup> cells, especially within open chromatin, promoter and enhancer regions. In addition, the TGF- $\beta$  receptor TGFBR1 exhibited higher expression in primary TSPAN33<sup>+</sup> cells than in cultured progenitor-like cells. Higher expression of TGF- $\beta$  receptors TGFBR1 and TGFBR2 in SSCs compared to differentiated KIT<sup>+</sup> spermatogonia has also been reported in humans (14). Further experiments were performed to examine the activation and function of TGF- $\beta$  signaling in monkey putative SSCs, which indicated that the TGF- $\beta$  pathway stimulated the self-renewal of monkey SSCs. Conversely, in mice, the TGF- $\beta$  receptors

are expressed in differentiated spermatogonia (11). This suggests that species-specific differences in TGF- $\beta$  functionality and SSC homeostasis regulation exist between rodents and primates. Although TGF- $\beta$  signaling plays beneficial role in promoting self-renewal of monkey SSCs, it is not sufficient to support their long-term expansion. Additional factors are required to achieve the successful long-term expansion of monkey SSCs *in vitro*.

Our data also suggested the involvement of other signaling pathways, including MAPK and JAK-STAT, in regulating monkey SSCs. Similarly, these pathways have been implicated to function in human SSCs (13,14). Of note, cytokines, such as FGF2, LIF, EGF, BMP and TGF- $\beta$ , are reported to modulate MAPK, JAK-STAT and PI3K-AKT (82,83). Thus, the effects of these cytokines on monkey SSC expansion warrant further examination. In addition, analyses of differential promoters and enhancers revealed significant enrichment in autophagy and AMPK signaling in the primary TSPAN33<sup>+</sup> cells versus cultured cells. Moreover, several TFs, including PPARG, PPARA and TFEB, which regulate autophagy, were highly expressed/enriched in primary TSPAN33<sup>+</sup> cells. The AMPK signaling pathway is a central metabolic regulator and plays critical roles in modulating autophagy and glucose and lipid metabolism (59). These observations suggest that monkey SSCs may exhibit a high level of catabolism. In alveolar stem cells, AMPK signaling is required for alveolar regeneration, and stimulating AMPK activation in aged alveolar stem cells can promote alveolar regeneration and repair (84). Hence, evaluating the role of AMPK in the long-term expansion of monkey SSCs would be a constructive avenue for future research.

## Data availability

All data are available in the main text or the supplementary materials. The sequencing data was deposited at GSA (<https://ngdc.cncb.ac.cn/gsa/>) under accession number: CRA016786.

## Supplementary data

Supplementary Data are available at NAR Online.

## Acknowledgements

We thank Yu Li, Ru-Yue Wang, Yong Wu and Min Xu for helpful discussions and technical assistance.

*Author contributions:* P.Z. designed the study. P.Z., Y-G.Y. and L-N.C. supervised the work. P.Z. and R.B. wrote the manuscript. R.B., C-L.S. and H-X.M. performed cell collection and constructed the libraries for ATAC-seq and CUT&Tag. C-L.S. performed single-cell collection and library construction for scRNA-seq. L-N.P., H.D. and C-H.H. performed computational analyses for sequencing data. R.B., C.L., H-J.L., L-P.X. and L.L. performed cellular experiments on cell culture, qRT-PCR and immunoblotting assay. C.L., L-P.X., L.L. and H.X. performed immunofluorescence staining and data analysis. K.G. provided partial funding support. All authors approved the final version of the manuscript.

## Funding

STI2030-Major Projects [2021ZD0200900]; National Natural Science Foundation of China [U2102202, 31801239,

62132015, 32100508, 31930022, 12131020, T2341007, T2350003]; Youth Innovation Promotion Association of CAS [2020000017]; Science and Technology Department of Yunnan Province [202305AH340006, 202305AH340007]; JST Moonshot R&D [JPMJMS2021].

## Conflict of interest statement

None declared.

## References

- Koskenniemi,J.J., Virtanen,H.E. and Toppari,J. (2017) Testicular growth and development in puberty. *Curr. Opin. Endocrinol Diabetes Obes.*, **24**, 215–224.
- Guo,J., Nie,X., Giebler,M., Mlcochova,H., Wang,Y., Grow,E.J., DonorConnect, Kim,R., Tharmalingam,M., Matilionyte,G., *et al.* (2020) The dynamic transcriptional cell atlas of testis development during human puberty. *Cell Stem Cell*, **26**, 262–276.
- Winther,J.F., Kenborg,L., Byrne,J., Hjorth,L., Kaatsch,P., Kremer,L.C., Kuehni,C.E., Auquier,P., Michel,G., de Vathaire,F., *et al.* (2015) Childhood cancer survivor cohorts in Europe. *Acta Oncol.*, **54**, 655–668.
- Masliukaite,I., Ntemou,E., Feijen,E.A.M., van de Wetering,M., Meissner,A., Soufan,A.T., Repping,S., Kremer,L.M.C., Jahnukainen,K., Goossens,E., *et al.* (2023) Childhood cancer and hematological disorders negatively affect spermatogonial quantity at diagnosis: a retrospective study of a male fertility preservation cohort. *Hum. Reprod.*, **38**, 359–370.
- Bhang,D.H., Kim,B.J., Kim,B.G., Schadler,K., Baek,K.H., Kim,Y.H., Hsiao,W., Ding,B.S., Rafii,S., Weiss,M.J., *et al.* (2018) Testicular endothelial cells are a critical population in the germline stem cell niche. *Nat. Commun.*, **9**, 4379.
- Guo,Y., Liu,L., Sun,M., Hai,Y., Li,Z. and He,Z. (2015) Expansion and long-term culture of human spermatogonial stem cells via the activation of SMAD3 and AKT pathways. *Exp. Biol. Med. (Maywood)*, **240**, 1112–1122.
- Wang,M., Liu,X., Chang,G., Chen,Y., An,G., Yan,L., Gao,S., Xu,Y., Cui,Y., Dong,J., *et al.* (2018) Single-cell RNA sequencing analysis reveals sequential cell fate transition during human spermatogenesis. *Cell Stem Cell*, **23**, 599–614.
- Guo,J., Grow,E.J., Mlcochova,H., Maher,G.J., Lindskog,C., Nie,X., Guo,Y., Takei,Y., Yun,J., Cai,L., *et al.* (2018) The adult human testis transcriptional cell atlas. *Cell Res.*, **28**, 1141–1157.
- Sohni,A., Tan,K., Song,H.W., Burrow,D., de Rooij,D.G., Laurent,L., Hsieh,T.C., Rabah,R., Hammoud,S.S., Vicini,E., *et al.* (2019) The neonatal and adult human testis defined at the single-cell level. *Cell Rep.*, **26**, 1501–1517.
- Huang,Y., Li,L., An,G., Yang,X., Cui,M., Song,X., Lin,J., Zhang,X., Yao,Z., Wan,C., *et al.* (2023) Single-cell multi-omics sequencing of human spermatogenesis reveals a DNA demethylation event associated with male meiotic recombination. *Nat. Cell Biol.*, **25**, 1520–1534.
- Shami,A.N., Zheng,X., Munyoki,S.K., Ma,Q., Manske,G.L., Green,C.D., Sukhwani,M., Orwig,K.E., Li,J.Z. and Hammoud,S.S. (2020) Single-cell RNA sequencing of human, macaque, and mouse testes uncovers conserved and divergent features of mammalian spermatogenesis. *Dev. Cell*, **54**, 529–547.
- Lau,X., Munusamy,P., Ng,M.J. and Sangrithi,M. (2020) Single-cell RNA sequencing of the cynomolgus macaque testis reveals conserved transcriptional profiles during mammalian spermatogenesis. *Dev. Cell*, **54**, 548–566.
- Guo,J., Grow,E.J., Yi,C., Mlcochova,H., Maher,G.J., Lindskog,C., Murphy,P.J., Wike,C.L., Carrell,D.T., Goriely,A., *et al.* (2017) Chromatin and single-cell RNA-seq profiling reveal dynamic signaling and metabolic transitions during human spermatogonial stem cell development. *Cell Stem Cell*, **21**, 533–546.
- Tan,K., Song,H.W., Thompson,M., Munyoki,S., Sukhwani,M., Hsieh,T.C., Orwig,K.E. and Wilkinson,M.F. (2020) Transcriptome profiling reveals signaling conditions dictating human spermatogonia fate *in vitro*. *Proc. Natl Acad. Sci. U.S.A.*, **117**, 17832–17841.
- Corso-Diaz,X., Jaeger,C., Chaitankar,V. and Swaroop,A. (2018) Epigenetic control of gene regulation during development and disease: a view from the retina. *Prog. Retin. Eye Res.*, **65**, 1–27.
- Atlati,Y. and Stunnenberg,H.G. (2017) The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.*, **18**, 643–658.
- Zhou,S., Feng,S., Qin,W., Wang,X., Tang,Y. and Yuan,S. (2021) Epigenetic regulation of spermatogonial stem cell homeostasis: from DNA methylation to histone modification. *Stem Cell Rev. Rep.*, **17**, 562–580.
- Li,C.H., Yan,L.Z., Ban,W.Z., Tu,Q., Wu,Y., Wang,L., Bi,R., Ji,S., Ma,Y.H., Nie,W.H., *et al.* (2017) Long-term propagation of tree shrew spermatogonial stem cells in culture and successful generation of transgenic offspring. *Cell Res.*, **27**, 241–252.
- Baker,S.C., Bauer,S.R., Beyer,R.P., Brenton,J.D., Bromley,B., Burrill,J., Causton,H., Conley,M.P., Elespuru,R., Fero,M., *et al.* (2005) The external RNA controls consortium: a progress report. *Nat. Methods*, **2**, 731–734.
- Picelli,S., Bjorklund,A.K., Faridani,O.R., Sagasser,S., Winberg,G. and Sandberg,R. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods*, **10**, 1096–1098.
- Liu,D., Wang,X., He,D., Sun,C., He,X., Yan,L., Li,Y., Han,J.J. and Zheng,P. (2018) Single-cell RNA-sequencing reveals the existence of naive and primed pluripotency in pre-implantation rhesus monkey embryos. *Genome Res.*, **28**, 1481–1493.
- Smith,T., Heger,A. and Sudbery,I. (2017) UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res.*, **27**, 491–499.
- Dai,H., Li,L., Zeng,T. and Chen,L. (2019) Cell-specific network constructed by single-cell RNA sequencing data. *Nucleic Acids Res.*, **47**, e62.
- Hao,Y., Hao,S., Andersen-Nissen,E., Mauck,W.M., Zheng,S., Butler,A., Lee,M.J., Wilk,A.J., Darby,C., Zager,M., *et al.* (2021) Integrated analysis of multimodal single-cell data. *Cell*, **184**, 3573–3587.
- Haghverdi,L., Buttner,M., Wolf,F.A., Buettner,F. and Theis,F.J. (2016) Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods*, **13**, 845–848.
- Li,L., Dai,H., Fang,Z. and Chen,L. (2021) c-CSN: single-cell RNA sequencing data analysis by conditional cell-specific network. *Genomics Proteomics Bioinformatics*, **19**, 319–329.
- Huang da,W., Sherman,B.T. and Lempicki,R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
- Sherman,B.T., Hao,M., Qiu,J., Jiao,X., Baseler,M.W., Lane,H.C., Imamichi,T. and Chang,W. (2022) DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.*, **50**, W216–W221.
- Zhou,Y., Zhou,B., Pache,L., Chang,M., Khodabakhshi,A.H., Tanaseichuk,O., Benner,C. and Chanda,S.K. (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.*, **10**, 1523.
- Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
- Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
- Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
- Tarasov,A., Vilella,A.J., Cuppen,E., Nijman,I.J. and Prins,P. (2015) Sambamba: fast processing of NGS alignment formats. *Bioinformatics*, **31**, 2032–2034.

34. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
35. Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F. and Manke, T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*, **44**, W160–W165.
36. Galili, T., O'Callaghan, A., Sidi, J. and Sievert, C. (2018) heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*, **34**, 1600–1602.
37. Yu, G., Wang, L.G. and He, Q.Y. (2015) ChIPseeker: an R/bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, **31**, 2382–2383.
38. Jin, H., Kasper, L.H., Larson, J.D., Wu, G., Baker, S.J., Zhang, J. and Fan, Y. (2020) ChIPseqSpikeInFree: a ChIP-seq normalization approach to reveal global changes in histone modifications without spike-in. *Bioinformatics*, **36**, 1270–1272.
39. Polit, L., Kerdivel, G., Gregoricchio, S., Esposito, M., Guillouf, C. and Boeva, V. (2021) CHIPIN: chIP-seq inter-sample normalization based on signal invariance across transcriptionally constant genes. *BMC Bioinformatics*, **22**, 407.
40. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
41. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., *et al.* (2023) The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.*, **51**, D638–D646.
42. Chen, Y., Zheng, Y., Gao, Y., Lin, Z., Yang, S., Wang, T., Wang, Q., Xie, N., Hua, R., Liu, M., *et al.* (2018) Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. *Cell Res.*, **28**, 879–896.
43. Brinster, R.L. (2002) Germline stem cell transplantation and transgenesis. *Science*, **296**, 2174–2176.
44. Hermann, B.P., Sukhwani, M., Winkler, F., Pascarella, J.N., Peters, K.A., Sheng, Y., Valli, H., Rodriguez, M., Ezzelarab, M., Dargo, G., *et al.* (2012) Spermatogonial stem cell transplantation into rhesus testes regenerates spermatogenesis producing functional sperm. *Cell Stem Cell*, **11**, 715–726.
45. Karlsson, M., Zhang, C., Mear, L., Zhong, W., Digre, A., Katona, B., Sjostedt, E., Butler, L., Odeberg, J., Dusart, P., *et al.* (2021) A single-cell type transcriptomics map of human tissues. *Sci. Adv.*, **7**, eabh2169.
46. Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., *et al.* (2015) Proteomics. Tissue-based map of the human proteome. *Science*, **347**, 1260419.
47. Li, C., Bi, R., Wang, L., Ma, Y.H., Yao, Y.G. and Zheng, P. (2023) Characterization of long-term ex vivo expansion of tree shrew spermatogonial stem cells. *Zool. Res.*, **44**, 1080–1094.
48. Fan, Y., Huang, Z.Y., Cao, C.C., Chen, C.S., Chen, Y.X., Fan, D.D., He, J., Hou, H.L., Hu, L., Hu, X.T., *et al.* (2013) Genome of the Chinese tree shrew. *Nat. Commun.*, **4**, 1426.
49. Fan, Y., Ye, M.S., Zhang, J.Y., Xu, L., Yu, D.D., Gu, T.L., Yao, Y.L., Chen, J.Q., Lv, L.B., Zheng, P., *et al.* (2019) Chromosomal level assembly and population sequencing of the Chinese tree shrew genome. *Zool. Res.*, **40**, 506–521.
50. Ye, M.S., Zhang, J.Y., Yu, D.D., Xu, M., Xu, L., Lv, L.B., Zhu, Q.Y., Fan, Y. and Yao, Y.G. (2021) Comprehensive annotation of the Chinese tree shrew genome by large-scale RNA sequencing and long-read isoform sequencing. *Zool. Res.*, **42**, 692–709.
51. Li, L., Cui, L., Lin, P., Liu, Z., Bao, S., Ma, X., Nan, H., Zhu, W., Cen, J., Mao, Y., *et al.* (2023) Kupffer-cell-derived IL-6 is repurposed for hepatocyte dedifferentiation via activating progenitor genes from injury-specific enhancers. *Cell Stem Cell*, **30**, 283–299.
52. Liu, C., Xu, Y., Yang, G., Tao, Y., Chang, J., Wang, S., Cheung, T.H., Chen, J. and Zeng, Y.A. (2024) Niche inflammatory signals control oscillating mammary regeneration and protect stem cells from cytotoxic stress. *Cell Stem Cell*, **31**, 89–105.
53. Macrae, T.A., Fothergill-Robinson, J. and Ramalho-Santos, M. (2023) Regulation, functions and transmission of bivalent chromatin during mammalian development. *Nat. Rev. Mol. Cell Biol.*, **24**, 6–26.
54. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. U.S.A.*, **107**, 21931–21936.
55. Hu, M., Yeh, Y.H., Maezawa, S., Nakagawa, T., Yoshida, S. and Namekawa, S.H. (2024) PRC1 directs PRC2-H3K27me3 deposition to shield adult spermatogonial stem cells from differentiation. *Nucleic Acids Res.*, **52**, 2306–2322.
56. Wang, H., Fan, Z., Shliaha, P.V., Miele, M., Hendrickson, R.C., Jiang, X. and Helin, K. (2023) H3K4me3 regulates RNA polymerase II promoter-proximal pause-release. *Nature*, **615**, 339–348.
57. Zhang, T., Cooper, S. and Brockdorff, N. (2015) The interplay of histone modifications - writers that read. *EMBO Rep.*, **16**, 1467–1481.
58. Li, Y., Chen, X. and Lu, C. (2021) The interplay between DNA and histone methylation: molecular mechanisms and disease implications. *EMBO Rep.*, **22**, e51803.
59. Mihaylova, M.M. and Shaw, R.J. (2011) The AMPK signalling pathway coordinates cell growth, autophagy and metabolism. *Nat. Cell Biol.*, **13**, 1016–1023.
60. Ivanenko, K.A., Prassolov, V.S. and Khabusheva, E.R. (2022) Transcription factor sp1 in the expression of genes encoding components of MAPK, JAK/STAT, and PI3K/Akt signaling pathways. *Mol. Biol. (Mosk)*, **56**, 832–847.
61. Beishline, K. and Azizkhan-Clifford, J. (2015) Sp1 and the 'hallmarks of cancer'. *FEBS J.*, **282**, 224–258.
62. Suzuki, S., McCarrey, J.R. and Hermann, B.P. (2021) An mTORC1-dependent switch orchestrates the transition between mouse spermatogonial stem cells and clones of progenitor spermatogonia. *Cell Rep.*, **34**, 108752.
63. Wang, C.Q., Mok, M.M., Yokomizo, T., Tergaonkar, V. and Osato, M. (2017) Runx family genes in tissue stem cell dynamics. *Adv. Exp. Med. Biol.*, **962**, 117–138.
64. Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, **125**, 315–326.
65. Kumar, D., Cinghu, S., Oldfield, A.J., Yang, P. and Jothi, R. (2021) Decoding the function of bivalent chromatin in development and cancer. *Genome Res.*, **31**, 2170–2184.
66. Xiang, Y., Zhang, Y., Xu, Q., Zhou, C., Liu, B., Du, Z., Zhang, K., Zhang, B., Wang, X., Gayen, S., *et al.* (2020) Epigenomic analysis of gastrulation identifies a unique chromatin state for primed pluripotency. *Nat. Genet.*, **52**, 95–105.
67. Villarroel-Campos, D., Henriquez, D.R., Bodaleo, F.J., Oguchi, M.E., Bronfman, F.C., Fukuda, M. and Gonzalez-Billault, C. (2016) Rab35 functions in axon elongation are regulated by P53-related protein kinase in a mechanism that involves Rab35 protein degradation and the microtubule-associated protein 1B. *J. Neurosci.*, **36**, 7298–7313.
68. Ziyad, S., Riordan, J.D., Cavanaugh, A.M., Su, T., Hernandez, G.E., Hilfenhaus, G., Morselli, M., Huynh, K., Wang, K., Chen, J.N., *et al.* (2018) A forward genetic screen targeting the endothelium reveals a regulatory role for the lipid kinase pi4ka in myelo- and erythropoiesis. *Cell Rep.*, **22**, 1211–1224.
69. Zhang, S.Q., Jiang, T., Li, M., Zhang, X., Ren, Y.Q., Wei, S.C., Sun, L.D., Cheng, H., Li, Y., Yin, X.Y., *et al.* (2012) Exome sequencing identifies MVK mutations in disseminated superficial actinic porokeratosis. *Nat. Genet.*, **44**, 1156–1160.

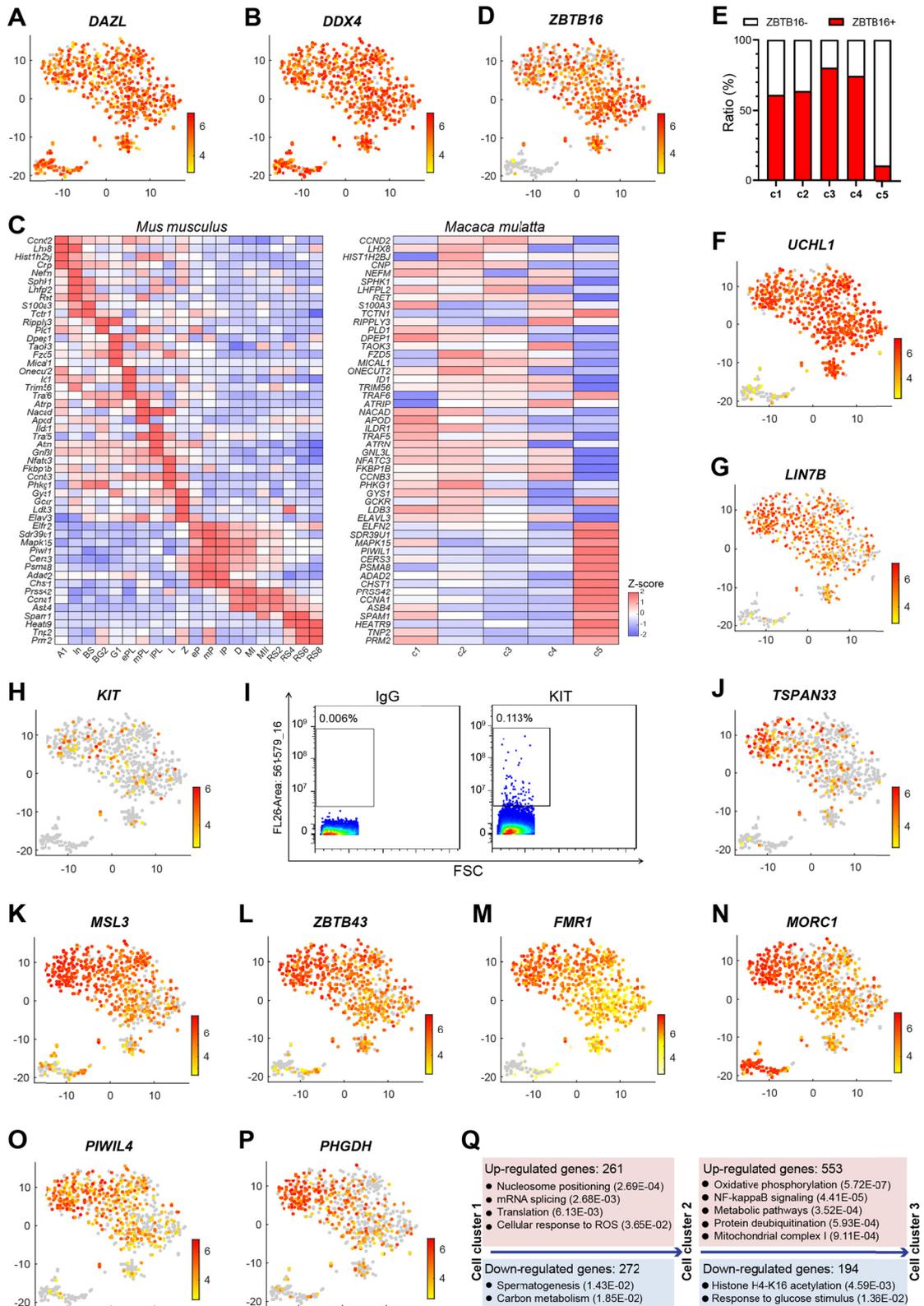
70. Garcia-Prat,L., Kaufmann,K.B., Schneiter,F., Voisin,V., Murison,A., Chen,J., Chan-Seng-Yue,M., Gan,O.I., McLeod,J.L., Smith,S.A., *et al.* (2021) TFEB-mediated endolysosomal activity controls human hematopoietic stem cell fate. *Cell Stem Cell*, **28**, 1838–1850.
71. Ito,K., Carracedo,A., Weiss,D., Arai,F., Ala,U., Avigan,D.E., Schafer,Z.T., Evans,R.M., Suda,T., Lee,C.H., *et al.* (2012) A PML-PPAR-delta pathway for fatty acid oxidation regulates hematopoietic stem cell maintenance. *Nat. Med.*, **18**, 1350–1358.
72. Oatley,J.M., Avarbock,M.R. and Brinster,R.L. (2007) Glial cell line-derived neurotrophic factor regulation of genes essential for self-renewal of mouse spermatogonial stem cells is dependent on Src family kinase signaling. *J. Biol. Chem.*, **282**, 25842–25851.
73. Montaigne,D., Butruille,L. and Staels,B. (2021) PPAR control of metabolism and cardiovascular functions. *Nat. Rev. Cardiol.*, **18**, 809–823.
74. Lee,J.M., Wagner,M., Xiao,R., Kim,K.H., Feng,D., Lazar,M.A. and Moore,D.D. (2014) Nutrient-sensing nuclear receptors coordinate autophagy. *Nature*, **516**, 112–115.
75. Ghosh,A., Jana,M., Modi,K., Gonzalez,F.J., Sims,K.B., Berry-Kravis,E. and Pahan,K. (2015) Activation of peroxisome proliferator-activated receptor alpha induces lysosomal biogenesis in brain cells: implications for lysosomal storage disorders. *J. Biol. Chem.*, **290**, 10309–10324.
76. Mullen,A.C. and Wrana,J.L. (2017) TGF-beta family signaling in embryonic and somatic stem-cell renewal and differentiation. *Cold Spring Harb. Perspect. Biol.*, **9**, a022186.
77. Cheng,K., Chen,I.C., Cheng,C.E., Mutoji,K., Hale,B.J., Hermann,B.P., Geyer,C.B., Oatley,J.M. and McCarrey,J.R. (2020) Unique epigenetic programming distinguishes regenerative spermatogonial stem cells in the developing mouse testis. *iScience*, **23**, 101596.
78. De Santis,M.C., Gulluni,F., Campa,C.C., Martini,M. and Hirsch,E. (2019) Targeting PI3K signaling in cancer: challenges and advances. *Biochim. Biophys. Acta Rev. Cancer*, **1871**, 361–366.
79. Sakurai,M., Ishitsuka,K., Ito,R., Wilkinson,A.C., Kimura,T., Mizutani,E., Nishikii,H., Sudo,K., Becker,H.J., Takemoto,H., *et al.* (2023) Chemically defined cytokine-free expansion of human haematopoietic stem cells. *Nature*, **615**, 127–133.
80. Budi,E.H., Duan,D. and Derynck,R. (2017) Transforming growth factor-beta receptors and smads: regulatory complexity and functional versatility. *Trends Cell Biol.*, **27**, 658–672.
81. Derynck,R. and Budi,E.H. (2019) Specificity, versatility, and control of TGF-beta family signaling. *Sci. Signal*, **12**, eaav5183.
82. Sakaki-Yumoto,M., Katsuno,Y. and Derynck,R. (2013) TGF-beta family signaling in stem cells. *Biochim. Biophys. Acta*, **1830**, 2280–2296.
83. Mossahebi-Mohammadi,M., Quan,M., Zhang,J.S. and Li,X. (2020) FGF signaling pathway: a key regulator of stem cell pluripotency. *Front. Cell Dev. Biol.*, **8**, 79.
84. Wang,Z., Wei,D., Bin,E., Li,J., Jiang,K., Lv,T., Mao,X., Wang,F., Dai,H. and Tang,N. (2023) Enhanced glycolysis-mediated energy production in alveolar stem cells is required for alveolar regeneration. *Cell Stem Cell*, **30**, 1028–1042.

Received: May 24, 2024. Revised: September 12, 2024. Editorial Decision: October 14, 2024. Accepted: October 17, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

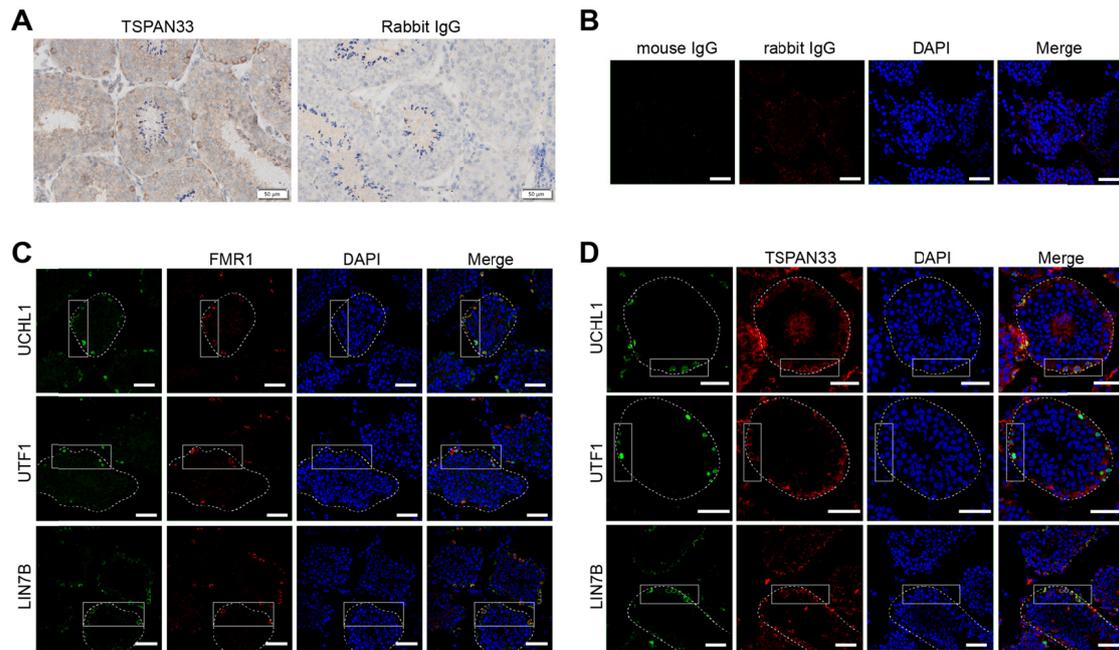
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [reprints@oup.com](mailto:reprints@oup.com) for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

## Supplementary Figures



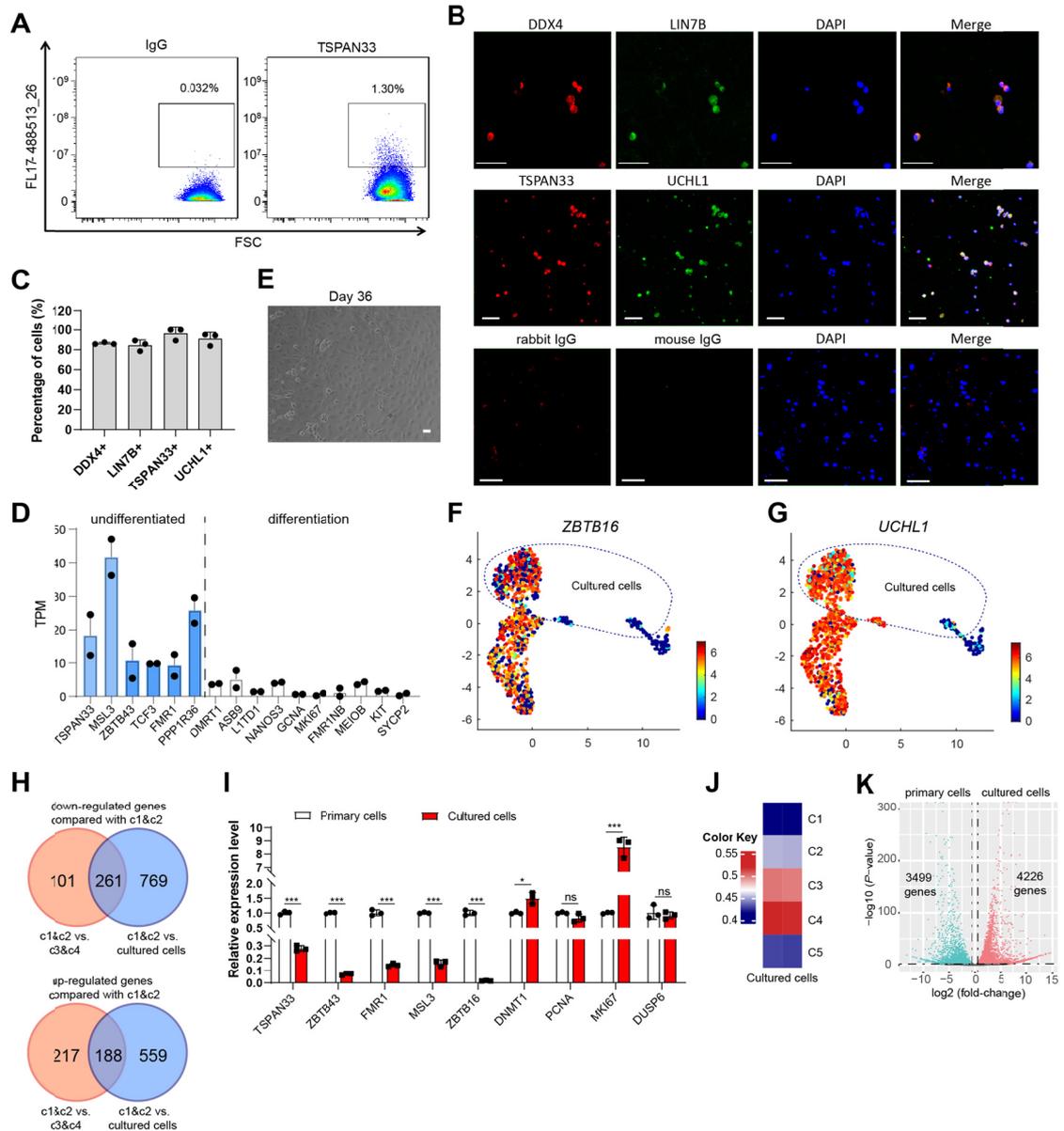
Supplementary Figure S1. Expression profiles of selected genes and pathway enrichment of DEGs between cell clusters in scRNA-seq, related to Figure 1.

**(A)** *DAZL* and **(B)** *DDX4* were expressed in c1 to c5 cells. **(C)** Expression comparison of marker genes for differential stage of spermatogenesis between monkey and mouse. A1, type A1 spermatogonia; In, intermediate spermatogonia; BS, S phase type B spermatogonia; BG2, G2/M phase type B spermatogonia; G1, G1 phase preleptotene; ePL, early S phase preleptotene; mPL, middle S phase preleptotene; lPL, late S phase preleptotene; L, leptotene; Z, zygotene; eP, early pachytene; mP, middle pachytene; lP, late pachytene; D, diplotene; MI, metaphase I; MII metaphase II; RS, round spermatids. **(D)** *ZBTB16* was predominantly expressed in c1 to c4 cells. **(E)** Percentages of *ZBTB16*-expressing cells in each cell cluster. **(F)** *UCHL1* and **(G)** *LIN7B* were prevalently expressed in c1-c4 cells. **(H)** *KIT* showed no enriched expression pattern. **(I)** FACS showed the *KIT*-expressing spermatogonia accounted for about 0.1% of the starting cell preparation. **(J-P)** Expression profiles of *TSPAN33* **(J)**, *MSL3* **(K)**, *ZBTB43* **(L)**, *FMRI* **(M)**, *MORC1* **(N)**, *PIWIL4* **(O)**, and *PHGDH* **(P)**. **(Q)** Pathway enrichment of DEGs between c1 vs. c2, and c2 vs. c3.



**Supplementary Figure S2.** Immuno-staining of testicular sections of adult monkeys, related to Figure 1.

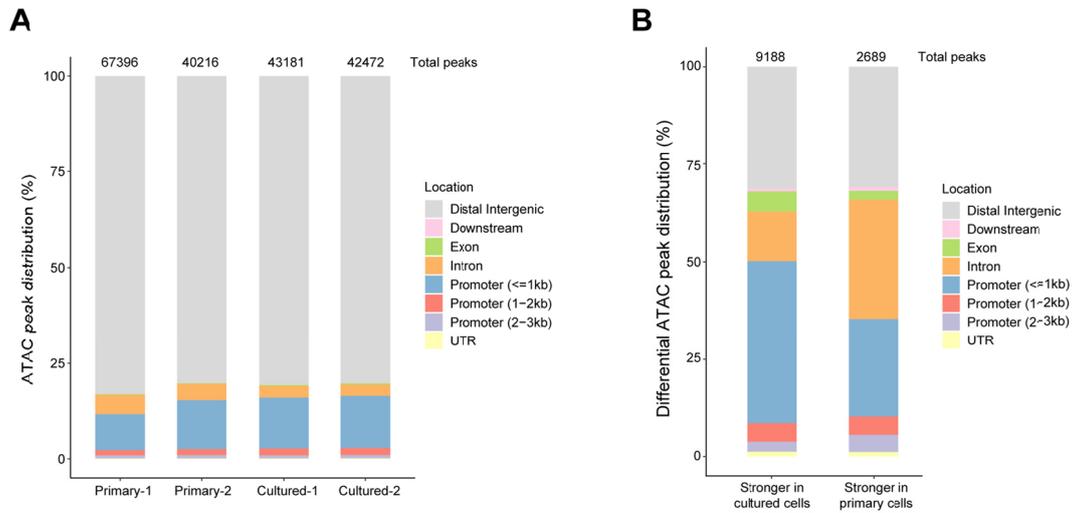
**(A)** Immunohistochemistry analysis of TSPAN33 in testicular sections of adult monkey (*left*), and rabbit IgG control (*right*). Scale bars, 50  $\mu$ m. **(B)** IgG controls of immunofluorescence staining. **(C-D)** Co-staining of FMR1 (red) or TSPAN33 (red) with UCHL1 (green), UTF1 (green), or LIN7B (green), respectively. Scale bars, 50  $\mu$ m. Rectangle regions were enlarged and showed in Figure 1H.



**Supplementary Figure S3.** Short-term culture of monkey TSPAN33<sup>+</sup> germ cells initiated differentiation, related to Figure 2.

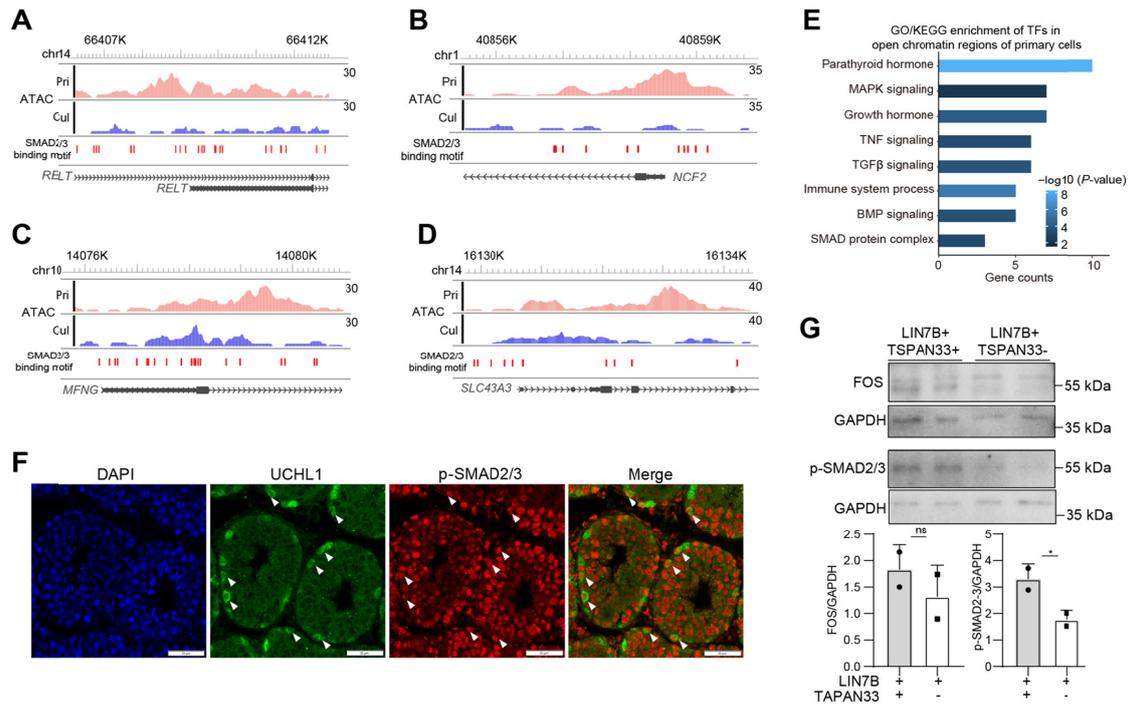
(A) TSPAN33<sup>+</sup> cells were sorted and enriched by flow cytometry. (B) Immunofluorescence staining of TSPAN33<sup>+</sup> cells with indicated antibodies: DDX4 (in red), TSPAN33 (in red), LIN7B (in green), UCHL1 (in green), and IgG controls. Scale bars, 50  $\mu$ m. (C) The percentages of DDX4<sup>+</sup>, LIN7B<sup>+</sup> cells in sorted TSPAN33<sup>+</sup> cells were analyzed in three independent fields of views containing a total of 77 cells. The percentages of TSPAN33<sup>+</sup>, UCHL1<sup>+</sup> cells in sorted TSPAN33<sup>+</sup> cells were analyzed in three independent fields of views containing a total of 76 cells. (D) TSPAN33<sup>+</sup> cells were subject to RNA-seq analysis. The expression values (TPM) of genes

marking SSCs and spermatogonia differentiation were shown. **(E)** Spermatogonia cultured for 36 days in tree shrew culture medium. Scale bars, 50  $\mu$ m. **(F, G)** Cultured germ cells (one or two weeks in culture) expressed pan undifferentiated spermatogonia markers *ZBTB16* **(F)** and *UCHL1* **(G)**. **(H)** Venn diagram showing common genes down-regulated or up-regulated during *in vivo* or *in vitro* spermatogonia differentiation. **(I)** Relative expression level of marker genes in TSPAN33<sup>+</sup> cells prior to and after culture for 5 days. Gene expression was normalized with *GAPDH*. Bars are mean  $\pm$  SD; \*,  $P < 0.05$ ; \*\*\*,  $P < 0.001$ ; ns, not significant; two-tailed Student's *t*-test. **(J)** Pearson correlation analysis was performed to compare the similarity between five-day cultured TSPAN33<sup>+</sup> cells and *in vivo* c1-c5 cells. **(K)** Volcano plot showing DEGs between TSPAN33<sup>+</sup> germ cells before and after 5 days of culture.



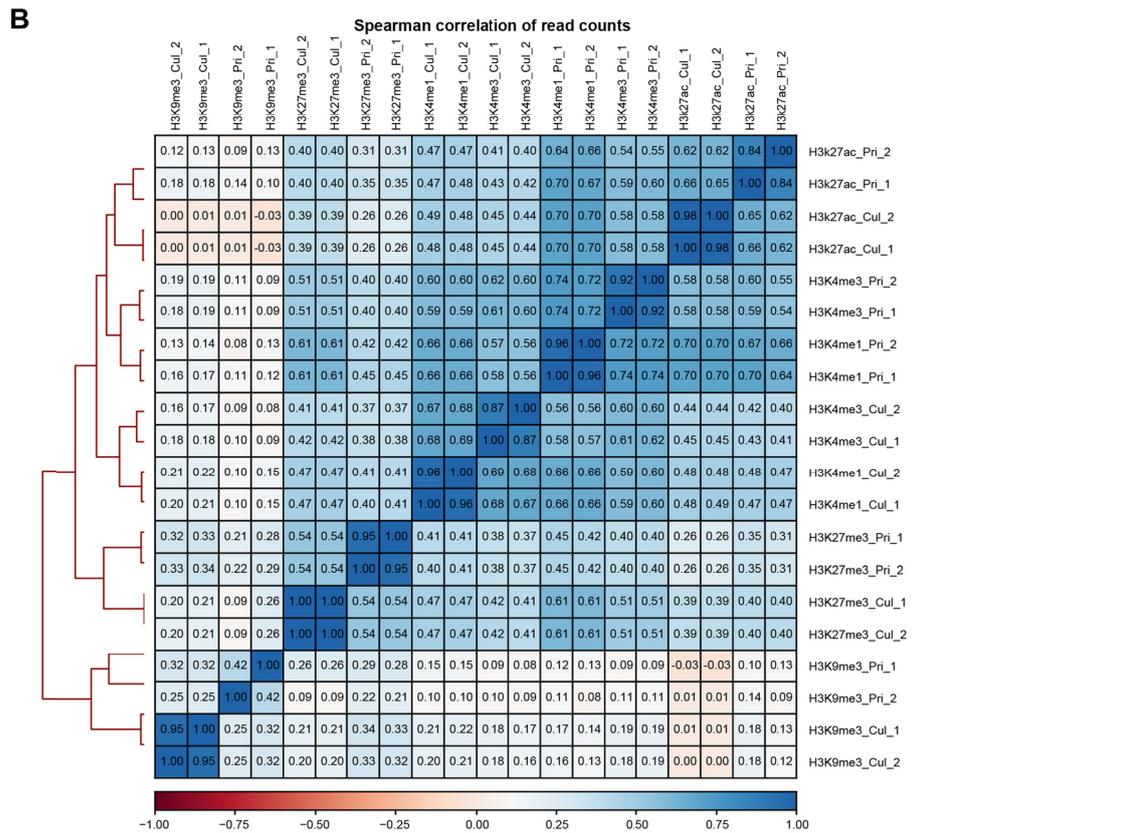
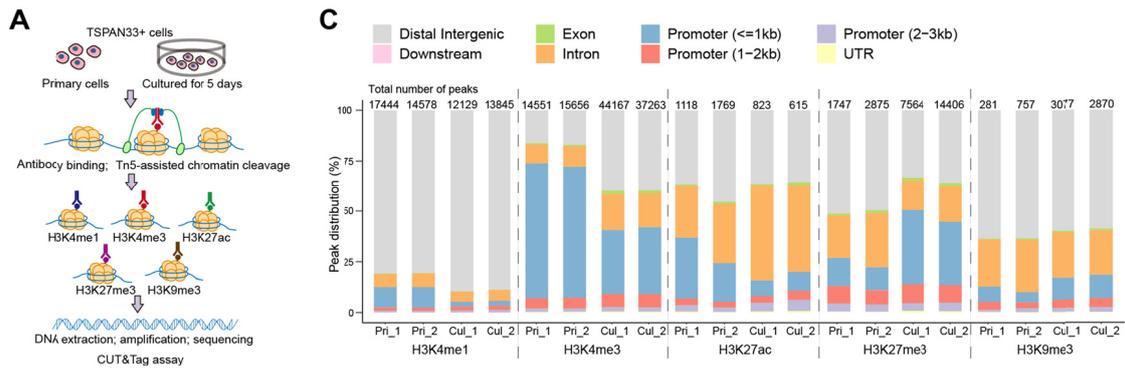
**Supplementary Figure S4.** Open chromatin distribution patterns in primary and cultured monkey TSPAN33<sup>+</sup> germ cells, related to Figure 2.

**(A)** Distribution patterns of open chromatin regions identified by ATAC-seq of primary and cultured monkey TSPAN33<sup>+</sup> cells. **(B)** Distribution patterns of chromatin regions predominantly opened in primary and cultured TSPAN33<sup>+</sup> cells.



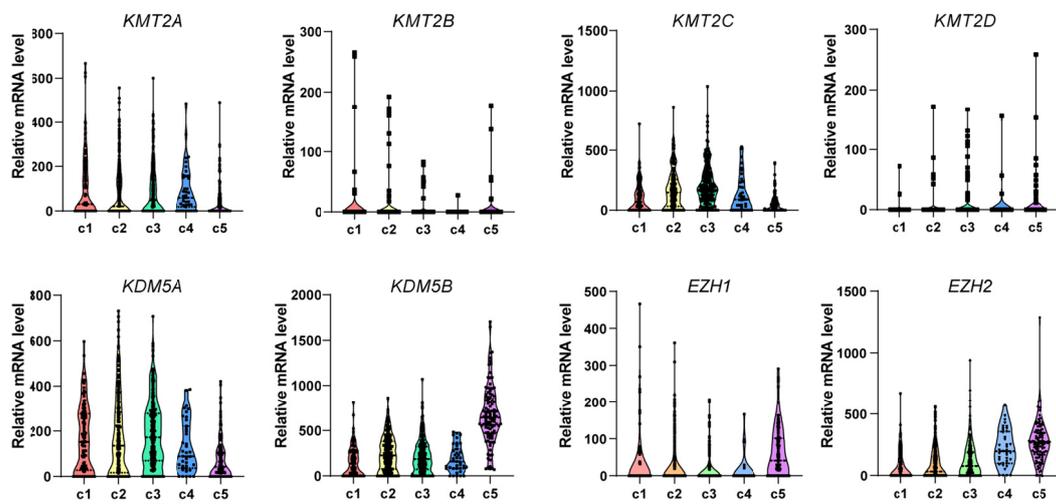
**Supplementary Figure S5.** Chromatin accessibility analysis identified the TFs and pathways highly active in primitive monkey SSCs, related to Figure 2.

(A-D) The genome browser snapshots of ATAC-seq results for genes *RELT* (A), *NCF2* (B), *MFNG* (C), and *SLC43A3* (D) containing SMAD2/3 binding motifs at promoter regions. (E) GO/KEGG enrichment of TFs in open chromatin regions of primary TSPAN33<sup>+</sup> spermatogonia. (F) Counter-immunofluorescent staining of p-SMAD2/3 with UCHL1. Note that some UCHL1-expressing cells showed stronger p-SMAD2/3 signal. Scale bar, 50 μm. (G) Immunoblotting analysis of FOS and p-SMAD2/3 in two subtypes of undifferentiated spermatogonia, LIN7B<sup>+</sup>TSPAN33<sup>+</sup> and LIN7B<sup>+</sup>TSPAN33<sup>-</sup>.

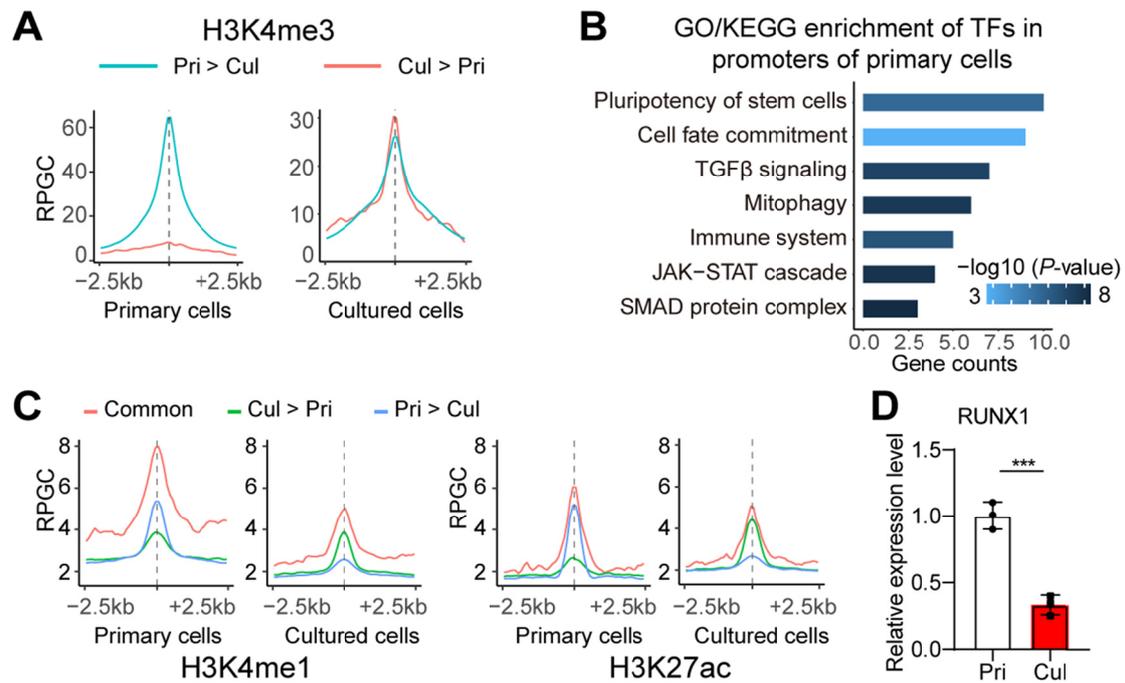


**Supplementary Figure S6.** Histone modification patterns in primary and cultured monkey TSPAN33<sup>+</sup> germ cells, related to Figure 3.

**(A)** Experimental workflow for examination of histone modifications. **(B)** Hierarchical clustering and Spearman correlation analysis of two replicates for histone modifications. Correlation coefficients are shown. **(C)** Distribution pattern of peaks identified in CUT&Tag seq. Primary cells: Pri\_1, Pri\_2; cultured cells: Cul\_1, Cul\_2.

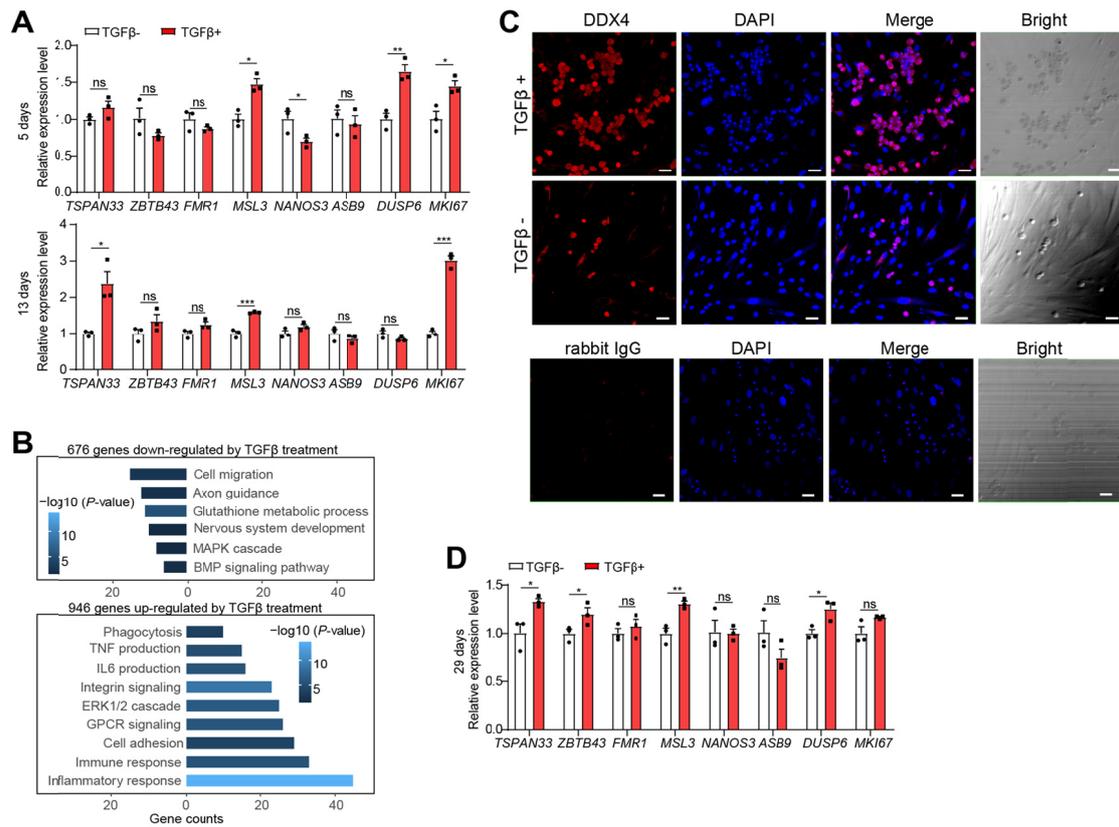


**Supplementary Figure S7.** Expression levels of key histone modification regulators in c1–c5 cells, related to Figure 3. Expression data are from scRNA-seq.



**Supplementary Figure S8.** Promoters, active enhancers, and potential TFs identified in primary and cultured monkey TSPAN33<sup>+</sup> germ cells, related to Figure 3.

(A) Plots of average H3K4me3 level in primary and cultured cells, related to Figure 3F. (B) GO and KEGG pathways enriched in TFs on promoter regions in primary TSPAN33<sup>+</sup> cells. (C) Plots of average H3K4me1 and H3K27ac levels in primary and cultured cells. (D) qRT-PCR analysis of *RUNX1* in primary or 5-day cultured TSPAN33<sup>+</sup> cells, the expression level was normalized by *GAPDH*. Data is shown as mean  $\pm$  SD from three independent replicates. \*\*\*,  $P < 0.001$ ; two-tailed Student's *t*-test.



**Supplementary Figure S9.** TGFβ treatment promotes the self-renewal of TSPAN33<sup>+</sup> cells in culture, related to Figure 6.

(A) TSPAN33<sup>+</sup> cells were cultured with or without TGFβ (1 ng/mL) for 5 or 13 days. Relative expression levels of different marker genes were examined by qRT-PCR. (B) GO terms and KEGG pathways enriched in TGFβ down- or up-regulated genes after treatment in culture for 13 days. (C) Immunofluorescence staining for DDX4 in TSPAN33<sup>+</sup> cells cultured for 29 days with or without TGFβ treatment (*upper*). Immunostaining using IgG was set as control (*lower*). Scale bars, 50 μm. (D) Relative expression levels of different marker genes were examined by qRT-PCR in TSPAN33<sup>+</sup> cells cultured for 29 days with or without TGFβ (1 ng/mL) treatment. Expression levels of each gene were normalized by *GAPDH*. Data are shown as mean ± SD from three independent replicates. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ; ns, not significant; two-tailed Student's *t*-test.

## Supplementary Key Resources Table

### Reagent or Resource

Antibodies	RRID	Source	Dilution (Application)
TSPAN33 Rabbit pAb	AB_2766065	Abclonal, Cat#A5222,	1:100 (IF); 1:50 (IHC); 1:50 (FACS/MACS)
LIN7B monoclonal antibody		Abmart, Cat#MG261554S	1:100 (IF); 1:100 (MACS)
TGF beta Receptor I (TGFBRI) Rabbit pAb	AB_2757355	Abclonal, Cat#A0708	1:1000 (IB)
Phospho-Smad2-S465/467 +Smad3-S423/425 Rabbit pAb	AB_2771541	Abclonal, Cat#AP0548	1:1000 (IB); 1:100 (IF)
Phospho-AKT1-S473 Rabbit pAb	AB_2770900	Abclonal, Cat#AP0140	1:1000 (IB)
Phospho-AMPKa1-T183/AMPKa2-T172 Rabbit pAb	AB_2771454	Abclonal, Cat#AP0432	1:1000 (IB)
Anti-UTF-1 Antibody, clone 5G10.2	AB_827541	Millipore, Cat#MAB4337	1:100 (IF)
Anti-DDX4/MVH antibody	AB_443012	Abcam, Cat#AB13840	1:100 (IF)
GAPDH (14C10) Rabbit mAb	AB_561053	Cell Signaling Technology, Cat#2118S	1:1000 (IB)
Human CD117/C-KIT Antibody	AB_355302	R&D system, Cat#AF332	1:50 (FACS) 1:100 (IF)
FMRP (Alias: FMR1) Rabbit pAb	AB_2766744	Abclonal, Cat#A6092	1:100 (IF)
Anti-PGP9.5 antibody [13C4 / I3C4] (antibody for UCHL1)	AB_306343	Abcam, Cat#AB8189	1:100 (IF)
Mono-Methyl-Histone H3 (Lys4) (D1A9) XP® Rabbit mAb (antibody for H3K4me1)	AB_10695148	Cell Signaling Technology, Cat#5326T	1:50 (CUT&Tag)
Tri-Methyl-Histone H3 (Lys4) (C42D8) Rabbit mAb (antibody for H3K4me3)	AB_2616028	Cell Signaling Technology, Cat#9751T	1:50 (CUT&Tag); 1:500 (IF)
Histone H3K27ac antibody (pAb)	AB_2722569	Active Motif, Cat#39034	1:50 (CUT&Tag)
Histone H3K27me3 antibody (pAb)	AB_2561020	Active Motif, Cat#39055	1:50 (CUT&Tag); 1:500 (IF)
Histone H3K9me3 antibody (pAb)	AB_2532132	Active Motif, Cat#39062	1:50 (CUT&Tag)
Mouse IgG Isotype Control	AB_10959891	Thermo Fisher, Cat#31903	1:500 (IF)
Rabbit IgG Isotype Control	AB_243593	Thermo Fisher, Cat#31235	1:1000 (IF); 1:500 (IHC); 1:500 (FACS)
Goat anti-Rabbit IgG (H+L) Secondary Antibody, Alexa Fluor 488	AB_2576217	Thermo Fisher, Cat#A11034	1:1000 (FACS)
Goat anti-Mouse IgG (H+L) Secondary Antibody, HRP	AB_228307	Thermo Fisher, Cat#31430	1:10000 (IB)
Goat anti-Rabbit IgG (H+L) Secondary Antibody, HRP	AB_228341	Thermo Fisher, Cat#31460	1:10000 (IB)
Donkey anti-Goat IgG (H+L) Secondary Antibody, Alexa Fluor™ 555	AB_141788	Thermo Fisher, Cat#A-21432	1:500 (IF) 1:1000 (FACS)

Donkey anti-mouse IgG (H+L) Secondary Antibody, Alexa Fluor 488	AB_141607	Thermo Fisher, Cat#A-21202	1:500 (IF)
Donkey anti-rabbit IgG (H+L) Secondary Antibody, Alexa Fluor 555	AB_162543	Thermo Fisher, Cat#A-31572	1:500 (IF)

---

**Oligonucleotides**

<b>RT-PCR Primers</b>		<b>Gene (for RT-PCR)</b>
5'-GGGAAACAGCAGTTGCACAG-3'	Forward	<i>TSPAN33</i>
5'-TGGTATCATGCCAACCGGAC-3'	Reverse	
5'-TGGCGTCTGCGGTAAGAAAT-3'	Forward	<i>ZBTB43</i>
5'-TCAGCCTTTGCAGCTTCGTA-3'	Reverse	
5'-TGCCAAAGAGTCGGCACATA-3'	Forward	<i>FMRI</i>
5'-TTAGTGCGCAGACTCCGAAA-3'	Reverse	
5'-CTGTGGGCATGAAGTCAGAGA-3'	Forward	<i>ZBTB16</i>
5'-CGCTGAATGAGCCAGTAAGTG-3'	Reverse	
5'-ACCCTCCCAAGTGCATTCAG-3'	Forward	<i>DNMT1</i>
5'-GCAAGTCAGTTTGTGCTGGG-3'	Reverse	
5'-TTAAACGGTTGCAGGCGTAGC-3'	Forward	<i>PCNA</i>
5'-GGAAGGAGGAAAGTCTAGCCG-3'	Reverse	
5'-GGTTGGAACAGAAGCTGGGATA-3'	Forward	<i>MSL3</i>
5'-TTCAGTGGGGAGGCCTTTTA-3'	Reverse	
5'-CAGTATCAGTGCCGGAGTCG-3'	Forward	<i>NANOS3</i>
5'-CAGAAAGAGCACAGGCGTTC-3'	Reverse	
5'-ACTGCATGCCAAGCACATTC-3'	Forward	<i>ASB9</i>
5'-GCTGGAAAATCCTGCCTCGAT-3'	Reverse	
5'-CAACAGACTCGGATGGCAGT-3'	Forward	<i>DUSP6</i>
5'-TGGGGGTGACGTTCAAGATG-3'	Reverse	
5'-GCAACCTGTAAAGCGGAAGC-3'	Forward	<i>MKI67</i>
5'-AAGCTATCAGCAGCACCGTT-3'	Reverse	
5'-GGACTCAGCTCTGGTTGGTG-3'	Forward	<i>TGFBR1</i>
5'-TGAGCAATGGCTGGCTTTC-3'	Reverse	
5'-ACAACCTCCAGGCAAAGGGTG-3'	Forward	<i>KMT2A</i>
5'-GATTCTCACATTTGGAATGGACC-3'	Reverse	
5'-GGAAGAAAGTCCTGGGCTCC-3'	Forward	<i>KMT2B</i>
5'-ATCTGGGGATGGAGGCTCTT-3'	Reverse	
5'-CCCCTTGAAACAGCAGCTC-3'	Forward	<i>KMT2C</i>
5'-GAAACCCCTTCCAGGGTAGC-3'	Reverse	
5'-CTGCCTGCTACAACCTGGGA-3'	Forward	<i>KMT2D</i>
5'-TTCTTAGGCTCAGTGCCTGC-3'	Reverse	
5'-AAAGTGGGTAGTCGCTTGGG-3'	Forward	<i>KDM5A</i>
5'-CTGCACACCCATAAGGCTCA-3'	Reverse	
5'-GACGTGTGCCAGTTTTGGAC-3'	Forward	<i>KDM5B</i>
5'-TCGAGGACACAGCACCTCTA-3'	Reverse	
5'-ACAACCTTAAACGGCTTCAGGC-3'	Forward	<i>EZH1</i>
5'-TGGTACAAGGCTTCATCGAC-3'	Reverse	
5'-TCCTTTTCATGCAACACCCA-3'	Forward	<i>EZH2</i>
5'-TTTGGTGGGGTCTTGATCCG-3'	Reverse	
5'-CTTCACAAACCCACCGCAAG-3'	Forward	<i>RUNX1</i>
5'-TCGGAAAAGGACAAGCTCCC-3'	Reverse	
5'-CTCCTGTTTCGAGAGTCAGCC-3'	Forward	<i>GAPDH</i>
5'-GCCCAATACGACCAAATCCG-3'	Reverse	

---