



OPEN ACCESS

EDITED BY

Huan Zhang,
University of Science and Technology of
China, China

REVIEWED BY

Radha Chaube,
Institute of Science, Banaras Hindu
University, India
Kang Zou,
Nanjing Agricultural University, China

*CORRESPONDENCE

Haiquan Yu,
hyu@imu.edu.cn

†These authors have contributed equally
to this work

SPECIALTY SECTION

This article was submitted to
Reproduction,
a section of the journal
Frontiers in Physiology

RECEIVED 21 May 2022

ACCEPTED 28 November 2022

PUBLISHED 08 December 2022

CITATION

Shi K, Wang B, Dou L, Wang S, Fu X and
Yu H (2022), Integrated bioinformatics
analysis of the transcription factor-
mediated gene regulatory networks in
the formation of spermatogonial
stem cells.

Front. Physiol. 13:949486.

doi: 10.3389/fphys.2022.949486

COPYRIGHT

© 2022 Shi, Wang, Dou, Wang, Fu and
Yu. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Integrated bioinformatics analysis of the transcription factor-mediated gene regulatory networks in the formation of spermatogonial stem cells

Kesong Shi[†], Baoluri Wang[†], Le Dou, Shu Wang, Xinrui Fu and
Haiquan Yu^{*†}

State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock (RRBGL), Inner
Mongolia University, Hohhot, China

Background: *In vitro* induction of spermatogonial stem cells (SSCs) from embryonic stem cells (ESCs) provides a promising tool for the treatment of male infertility. A variety of molecules are involved in this complex process, which needs to be further clarified. Undoubtedly, the increased knowledge of SSC formation will be beneficial to facilitate the currently complex induction process.

Methods: Based on ATAC-seq, DNase-seq, RNA-seq, and microarray data from GEO datasets, chromatin property data (ATAC-seq, DNase-seq) and gene expression data (RNA-seq, microarray data) were combined to search for SSC-specific transcription factors (TFs) and hub SSC-specific genes by using the WGCNA method. Then, we applied RNA-seq and microarray data screening for key SSC-specific TFs and constructed key SSC-specific TF-mediated gene regulatory networks (GRNs) using ChIP-seq data.

Results: First, after analysis of the ATAC-seq and DNase-seq data of mouse ESCs, primordial germ cells (PGCs), and SSCs, 33 SSC-specific TFs and 958 targeting genes were obtained. RNA-seq and WGCNA revealed that the key modules (turquoise and red) were the most significantly related to 958 SSC-specific genes, and a total of 10 hub SSC-specific genes were identified. Next, when compared with the cell-specific TFs in human ESCs, PGCs, and SSCs, we obtained five overlapping SSC-specific TF motifs, including the NF1 family TF motifs (NFIA, NFIB, NFIC, and NFIX), GRE, Fox:Ebox, PGR, and ARE. Among these, *Nfib* and *Nfix* exhibited abnormally high expression levels relative to mouse ESCs and PGCs. Moreover, *Nfib* and *Nfix* were upregulated in the testis sample with impaired spermatogenesis when compared with the normal group. Finally, the ChIP-seq data results showed that NFIB most likely targeted the hub SSC-specific genes of the turquoise module (*Rpl36a1*, *Rps27*, *Rps21*, *Nedd8*, and *Sec61b*) and the red module (*Vcam1* and *Ccl2*).

Conclusion: Our findings preliminarily revealed cell-specific TFs and cell-specific TF-mediated GRNs in the process of SSC formation. The hub SSC-specific genes and the key SSC-specific TFs were identified and suggested

complex network regulation, which may play key roles in optimizing the induction efficiency of the differentiation of ESCs into SSCs *in vitro*.

KEYWORDS

gene regulatory networks, transcription factor, spermatogonial stem cells, ATAC-seq, DNase-seq, WGCNA

Introduction

Infertility is a major health problem worldwide, and it is estimated that infertility affects approximately 10% of couples globally, with the male factor being a primary or contributing cause in approximately 50% of couples (Li et al., 2019; Ishikura et al., 2021). Male infertility is a multifactorial pathological condition in which genetic factors are highly complex, and azoospermia is the most common genetic factors that contributes to male infertility (Krausz and Riera-Escamilla, 2018). Spermatogonial stem cells (SSCs) are the ancestral cells of sperm and are the basis for spermatogenesis and fertility in men (Kossack et al., 2009). Therefore, SSCs are considered a promising alternative for the regeneration of impaired or damaged spermatogenesis, and SSCs transplantation is a promising technique for male infertility treatment (Abdelaal et al., 2021). Nevertheless, the number of SSCs is very scarce, and long-term culture and expansion of SSCs have not yet been available (Zhou et al., 2018). At present, both in mice and humans, several studies have found that ESCs have the ability to form putative primordial germ cells (PGCs), and these ESC-derived PGCs could further differentiate into SSCs (Nagamatsu and Hayashi, 2017; Li et al., 2019; Ishikura et al., 2021). However, these reports either involve a complex induction process with undefined induction factors or show a low induction efficiency, while the reconstitution of SSC formation *in vitro* remains a key challenge (Nagamatsu and Hayashi, 2017).

The dynamic reorganization of chromatin is accompanied by a genome-wide transcriptional change. In recent years, chromatin accessibility profiling has become an important tool for studying genetic and epigenetic regulation (Maezawa et al., 2018), allowing us to dissect the pangenomic regulatory landscape of cells and tissues in both time and space dimensions by detecting specific chromatin states and their corresponding TFs (Ma and Zhang, 2020). Moreover, chromatin accessibility profiling is expected to be a powerful tool for the identification of regulatory DNA elements that gene regulatory networks (GRNs), and changes in chromatin accessibility can be interpreted in the context of these dynamic regulatory networks (Nikolaev et al., 2007). The assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) and DNase I hypersensitive sites sequencing (DNase-seq) are a technology that maps the landscape of chromatin accessibility (Shashikant and Etensohn, 2019). This method not only recognizes different cell types but also reveals cell-type-specific regulatory regions and

detects the chromatin accessibility of related genes and putative TF-binding sites (Buenrostro et al., 2013; Shashikant and Etensohn, 2019). Chromatin accessibility is closely correlated with the differential expression of genes, and it can potentially be a transcription factor regulator (Huang et al., 2020). Recently, ATAC-seq and ChIP-seq were combined to profile the change in chromatin accessibility in spermatogenesis. Namekawa et al. revealed the genome-wide, dynamic reorganization of open chromatin during spermatogenesis and detected possible regulatory elements for spermatogenesis-specific gene expression (Maezawa et al., 2018). They also found distinct chromatin environments of autosomes and sex chromosomes during spermatogenesis, suggesting that poised chromatin and the formation of bivalent domains underlie genome-wide epigenetic changes during late spermatogenesis (Sin et al., 2015).

Gene regulatory networks of the cells can also be revealed through chromatin profiling assays. Currently, GRNs have been used in many different areas, such as B-cell differentiation in the mammalian immune system (Singh et al., 2005), ankylosing spondylitis (Yu et al., 2020), plasma cell function (Treize and Nutt, 2021) and immune cells associated with cancer (Han et al., 2017). Among them, through systematic analysis of the GRNs of immune cells, our group's previous study found that the network size of the GRNs of tumour-infiltrating immune T cells was reduced when compared to the GRNs of their corresponding immune cells in blood (Han et al., 2017). It is already known that the transcription factors Plzf and Taf4b have been implicated in regulating SSC functions, and these molecules are part of a robust gene network controlling SSC fate decisions (Oatley and Brinster, 2008). However, the intrinsic GRNs that control SSC fate decisions and that disrupt these networks in clinical cases of human male infertility have yet to be determined (Oatley and Brinster, 2008).

The purpose of our study was to screen for factors that induce the differentiation of ESCs into SSCs *in vitro* by constructing TF-mediated GRNs during the formation of SSCs. First, we searched for SSC-specific TFs and hub SSC-specific genes based on chromatin property data (ATAC-seq, DNase-seq) and gene expression data (RNA-seq, microarray data). Then, we obtained overlapping SSC-specific TFs between humans and mice. Next, key SSC-specific TFs were obtained by comparing gene expression among ESCs, PGCs, and SSCs. Finally, we analysed the gene expression levels of key SSC-specific TFs in testis samples with impaired spermatogenesis compared with the normal group and constructed key SSC-specific TF-mediated GRNs using ChIP-seq data (Figure 1). Our study provides

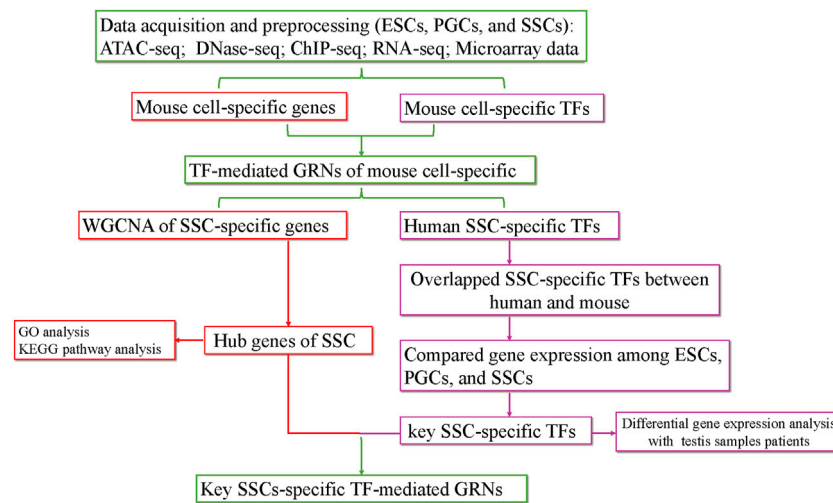


FIGURE 1

The flow chart of the analysis process. KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology.

potential induction factors for further optimizing the induction efficiency of the differentiation of ESCs into SSCs *in vitro*.

Materials and methods

Data sources

The ATAC-seq, DNase-seq, RNA-seq, and microarray data were obtained from the GEO database (<http://www.ncbi.nlm.nih.gov/geo>). ChIP-seq data from the Cistrome (<http://www.cistrome.org/db>). The details are shown in [Supplementary Table S1](#).

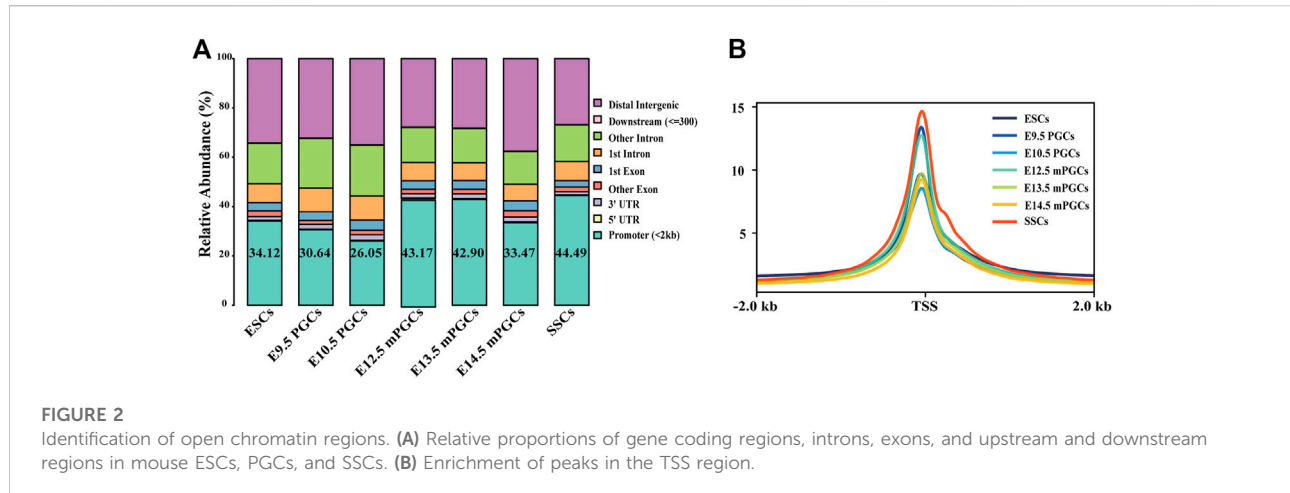
ATAC-seq, DNase-seq and regulatory network construction

To explore the accessibility of chromatin and obtain cell-specific TFs, raw sequence reads were initially processed by FastQC v0.11.9 for quality control, and quality filtering and adapter trimming were performed using Cutadapt v1.9. Then, reads were aligned to the reference genome (mm10 or hg38) with Bowtie2 (v2.4.4) (Langmead and Salzberg, 2012). Bam files from the resulting sam files by using Samtools v1.9 (Li et al., 2009). The peaks that replicated across each sample were merged into a single file using the bedtools software v2.30.0 (Quinlan and Hall, 2010). Peaks with an initial threshold q -value of 0.05 as the cut-off in experimental bam files were called by using MACS2 v2.2.7.1 (Zhang et al., 2008), and the peaks were identified with the parameters `callpeak --shift -100 --extsize 200`. DeepTools v2.0 (Ramírez et al., 2014) was used to enrich

peaks in the transcriptional start site (TSS) region by using the `computeMatrix` and `plotHeatmap` functions. Annotation of peaks was performed using the ChIPSeeker package (Yu et al., 2015). Library to annotate the peaks to genomic features using a cut-off of $2\text{ kb} \pm$ from TSS. Subsequently, we applied HOMER's FindMotifGenome.pl tool v4.11 (Heinz et al., 2010) to obtain the list of TFs that putatively bind to the peaks. The TFs were filtered based on the p -value, and the final results were filtered based on the p -value $\leq 1 \times 10^{-4}$. To discover the TFs that can bind to cell-specific genes, we used ChIP-seq data from Cistrome Data (score >1). The TFs and their regulated downstream genes were confirmed, and GRNs were constructed. GRNs and regulatory circuit analyses were conducted by Cytoscape v3.7.2 software. For the visualization of read count data, bam files, bigwig files, and genome browser images were made using the Integrative Genomics Viewer (IGV) tools (Thorvaldsdóttir et al., 2013). The UpSet diagram was generated using the R package UpSetR v1.4.0.

RNA-seq and microarray data analysis

The raw reads were processed as given above. High-quality reads were then mapped to the mm10 reference genome using HISAT2 v2.2.1 (Kim et al., 2015). Aligned RNA-seq reads were quantified and annotated using `featureCounts` (Liao et al., 2014), and gene expression levels were calculated based on RNA-seq data as transcripts per kilobase million (TPM) values, and TPM values from RNA-seq samples were averaged. Microarray data form GSE145467 (<http://www.ncbi.nlm.nih.gov/geo>), which contained 20 samples (10 testis samples with normal spermatogenesis and 10 testis samples with impaired



spermatogenesis). Differential gene expression analysis was performed using GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r>). Volcano plots were plotted using the ggplot2 package (<http://ggplot2.org/>). Heatmaps were created using the pheatmap package v1.0.12 (<https://CRAN.R-project.org/package=pheatmap>).

WGCNA, hub gene selection and enrichment analysis

In this study, the coexpressed gene module and the hub module correlated with SSCs were analysed using the R package WGCNA (Langfelder and Horvath, 2008), and we selected 958 SSC-specific genes targeted by SSC-specific TFs for WGCNA network construction. A power of $\beta = 9$ and the adjacencies between all the filtered genes were transformed into the corresponding dissimilarity. The hub module and hub genes were visualized with the plug-in MCODE of Cytoscape v3.7.2 with a cut-off MCODE score of ≥ 4.5 . Afterwards, the functional annotation and pathway enrichment for the genes in the hub module was conducted using DAVID (false-discovery rate, $FDR < 0.05$) (Huang da et al., 2009), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) analyses were plotted using R package GOplot v1.0.2 (<https://cran.r-project.org/web/packages//GOplot/>).

Results

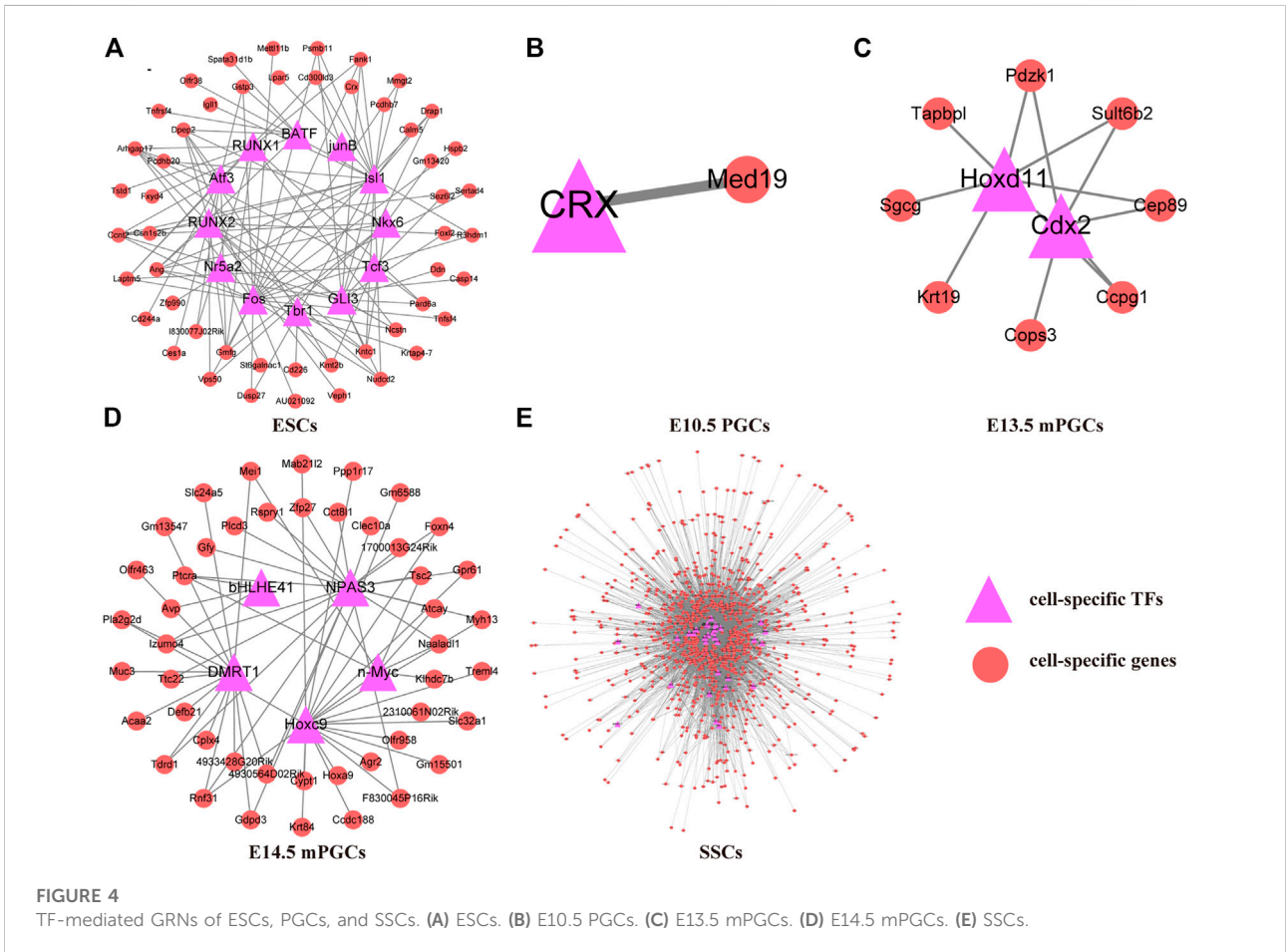
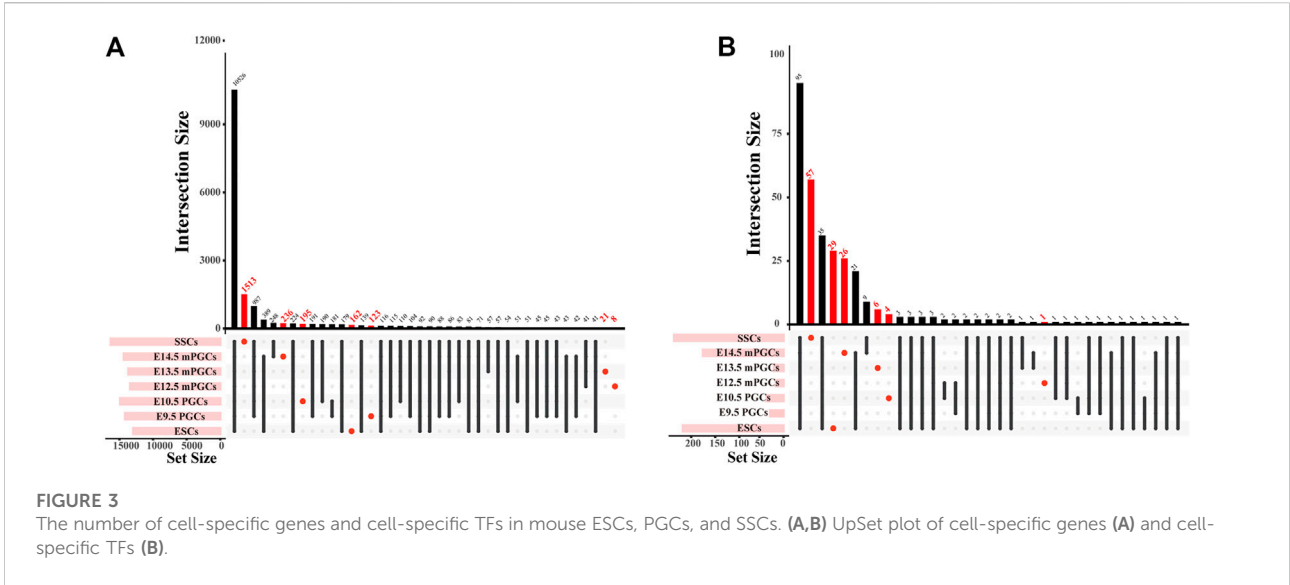
Mapping accessible chromatin in mouse ESCs, PGCs, and SSCs

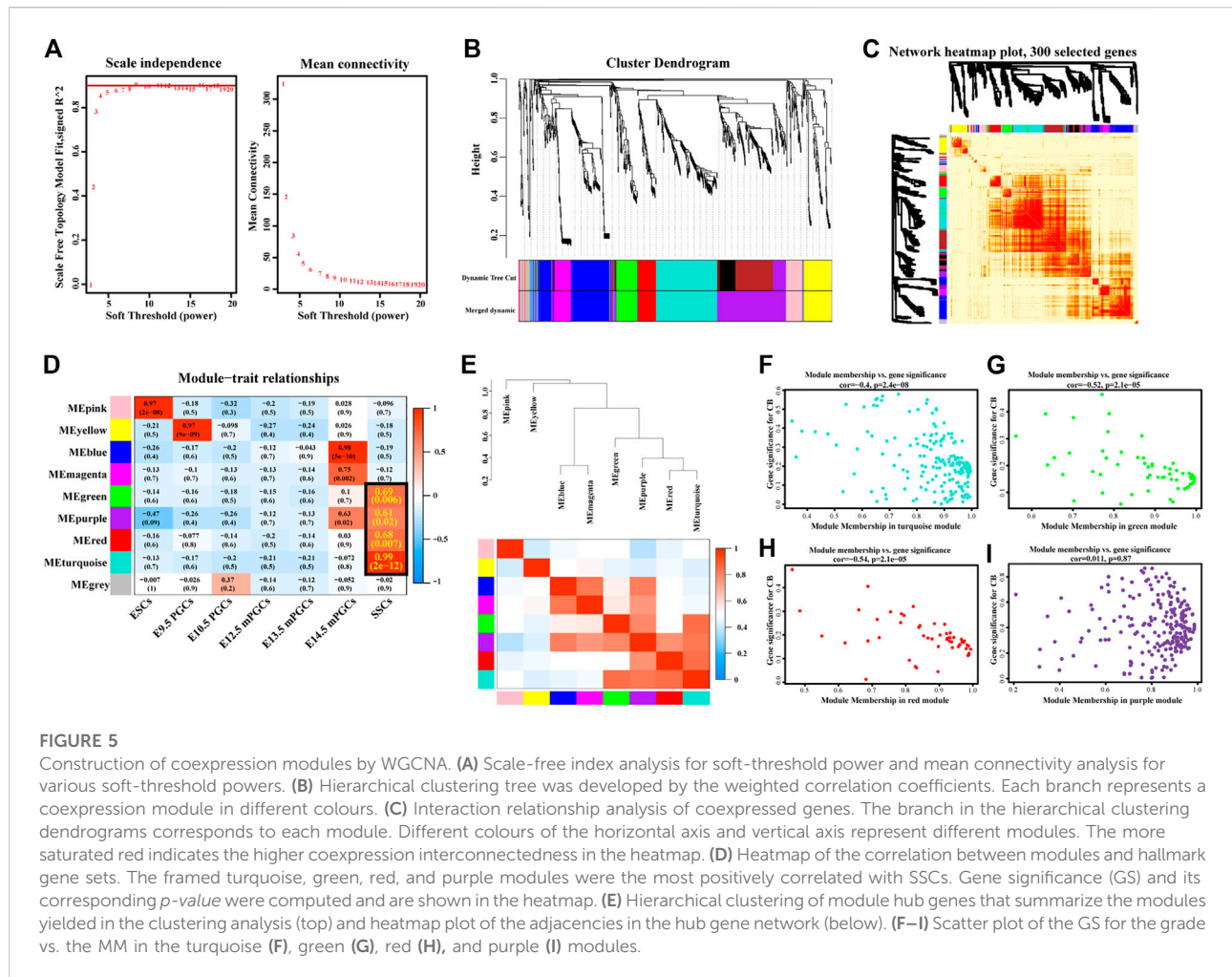
ATAC-seq and DNase-seq have emerged as powerful methods to study accessible chromatin and GRNs (Han et al.,

2017; Maher et al., 2018). Thus, we obtained a comprehensive landscape of the accessible chromatin in mouse ESCs, PGCs, and SSCs. We first mapped the distribution map of transposase hypersensitive sites (THSs) or peaks by either ATAC-seq or DNase-seq data. Figure 2A summarizes the relative enriched proportions of promoter, 5' UTR, 3' UTR, exons, downstream regions, and distal intergenic regions. Most of the peaks in ESCs, PGCs, and SSCs were found to be located within the promoter (peaks were located within 2 kb upstream of transcriptional start site (TSS)) and distal intergenic regions. The proportion of promoters in the SSCs was the highest (44.49%), while the proportion of promoters on embryonic day (E) 10.5 PGCs was the lowest (26.05%). We next examined the signal of peaks located within 2 kb of TSS using average plots (Figure 2B). The results revealed that a large proportion of peaks are located close to TSS, which means that the TSS tends to bind to TFs.

Specific genes and specific TFs of mouse ESCs, PGCs, and SSCs

The identified peaks can be used to predict motifs generally recognized by TFs, and GRNs of the cells can also be revealed through chromatin profiling assays (Nikolaev et al., 2007). Accordingly, we first obtained cell-specific genes and cell-specific TFs and then preliminarily constructed cell-specific GRNs. A total of 2,258 cell-specific genes were identified from DNase-seq and ATAC-seq for mouse ESCs, PGCs, and SSCs (Figure 3A, Supplementary Table S2). The largest number of cell-specific genes was expressed in SSCs, while the fewest cell-specific genes were expressed in E12.5 male (m) PGCs (Figure 3A). We also scanned the DNA-binding motifs and their associated TF motifs by using HOMER. We illustrate the total number of TF motifs that can be found across mouse ESCs, PGCs, and SSCs (Supplementary Table S2), and the distribution of TF motifs across the cell subsets is





shown in Figure 3B. The ascending order ranked by the number of cell-specific TF motifs expressed was E9.5 PGCs, E12.5 mPGCs, E10.5 PGCs, E13.5 mPGCs, E14.5 mPGCs, ESCs, and SSCs. Of these TFs, no cell-specific TFs were found in E9.5 PGCs. The results above imply the different developmental stages of male germ cells as the level of analysis because genes and TFs can vary in spatial and temporal expression.

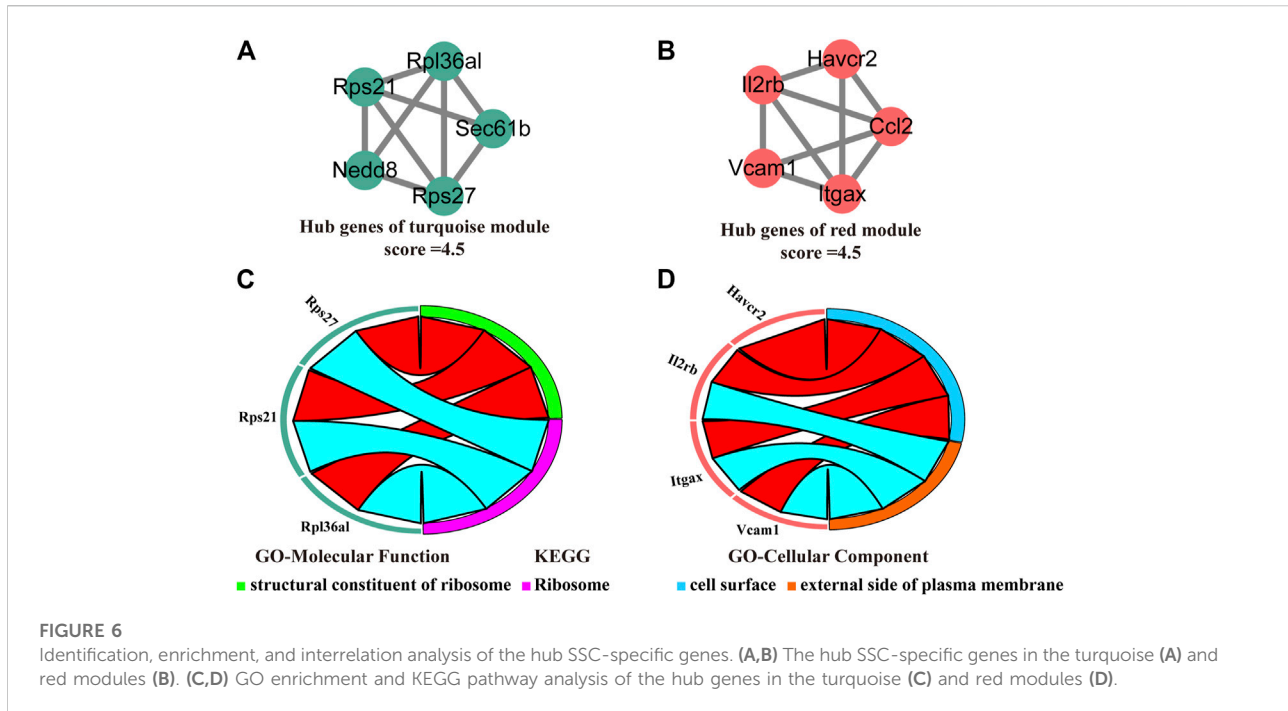
TF-mediated gene regulatory networks of mouse ESCs, PGCs, and SSCs

To preliminarily construct the TF-mediated GRNs of mouse ESCs, PGCs, and SSCs, we analysed the cell-specific genes related to the cell-specific TF motifs according to ChIP-seq. We found that GRNs of mouse ESCs, PGCs, and SSCs differ between different cell types. Therefore, 50 cell-specific genes are targeted by cell-specific TFs in ESCs (Figure 4A), and only one cell-specific gene is targeted by cell-specific TFs in E10.5 PGCs (Figure 4B). E13.5 mPGCs (Figure 4C) and E14.5 mPGCs (Figure 4D), containing 8 and

46 cell-specific genes, respectively, were targeted by cell-specific TFs. However, no cell-specific genes are targeted by cell-specific TFs in E9.5 PGCs and E12.5 mPGCs. We also found that SSCs had the largest number of cell-specific TFs and their target genes, and 958 cell-specific genes were targeted by 33 cell-specific TFs, comprising 7,326 regulatory relationships in SSCs, after removing the redundant and unannotated TFs (Supplementary Table S3, Figure 4E). Taken together, these results showed that cell-specific TFs and their targeting cell-specific genes are different between mouse ESCs, PGCs, and SSCs. The results above implied that cell-specific TFs and their targeted cell-specific genes in mouse ESCs, PGCs, and SSCs could be used to determine their identity.

WGCNA of the SSC-specific genes associated with SSC-Specific TFs

Previously, we learned that 958 SSC-specific genes are targeted by 33 SSC-specific TFs, comprising 7,326 regulatory relationships. To further explore the possible target genes for



SSC-specific TFs and reveal the hub SSC-specific genes, we conducted WGCNA based on the RNA-seq (Supplementary Table S1) results. We first identified the relatively balanced scale independence and mean connectivity of the WGCNA, and the soft threshold $\beta = 9$ was adopted to achieve the scale-free topology criterion of the network (Figure 5A) and then obtained a hierarchical clustering tree using the dynamic cutting method (Figure 5B). The 9 modules marked were identified, and the interactions of the 9 modules were visualized as a network heatmap. The results revealed that each module was an independent validation to each other, which indicated that genes in the same module had a highly coexpressed relationship with each other (Figure 5C). The turquoise, green, red, and purple modules were significantly positively (p -value < 0.05) correlated with the SSCs, indicating that the turquoise, green, red, and purple modules may play an important role in the formation of SSCs (Figure 5D). Of these, the turquoise module was the most positively correlated with SSCs, including 181 genes, the green module contained 60 genes, the red module contained 55 genes and the purple module contained 127 genes. However, no module was significantly negatively correlated with the SSCs (Figure 5D). Then, we calculated eigengenes and clustered them according to their correlation to explore the coexpression similarity of all modules, and similar results were demonstrated by the heatmap plotted according to adjacencies (Figure 5E). Next, we evaluated the correlation between the gene significance (GS) and module membership (MM) in the turquoise, green, red, and purple modules. The correlation was significant in the turquoise ($R = -0.4$, p -value = $2.4e-08$),

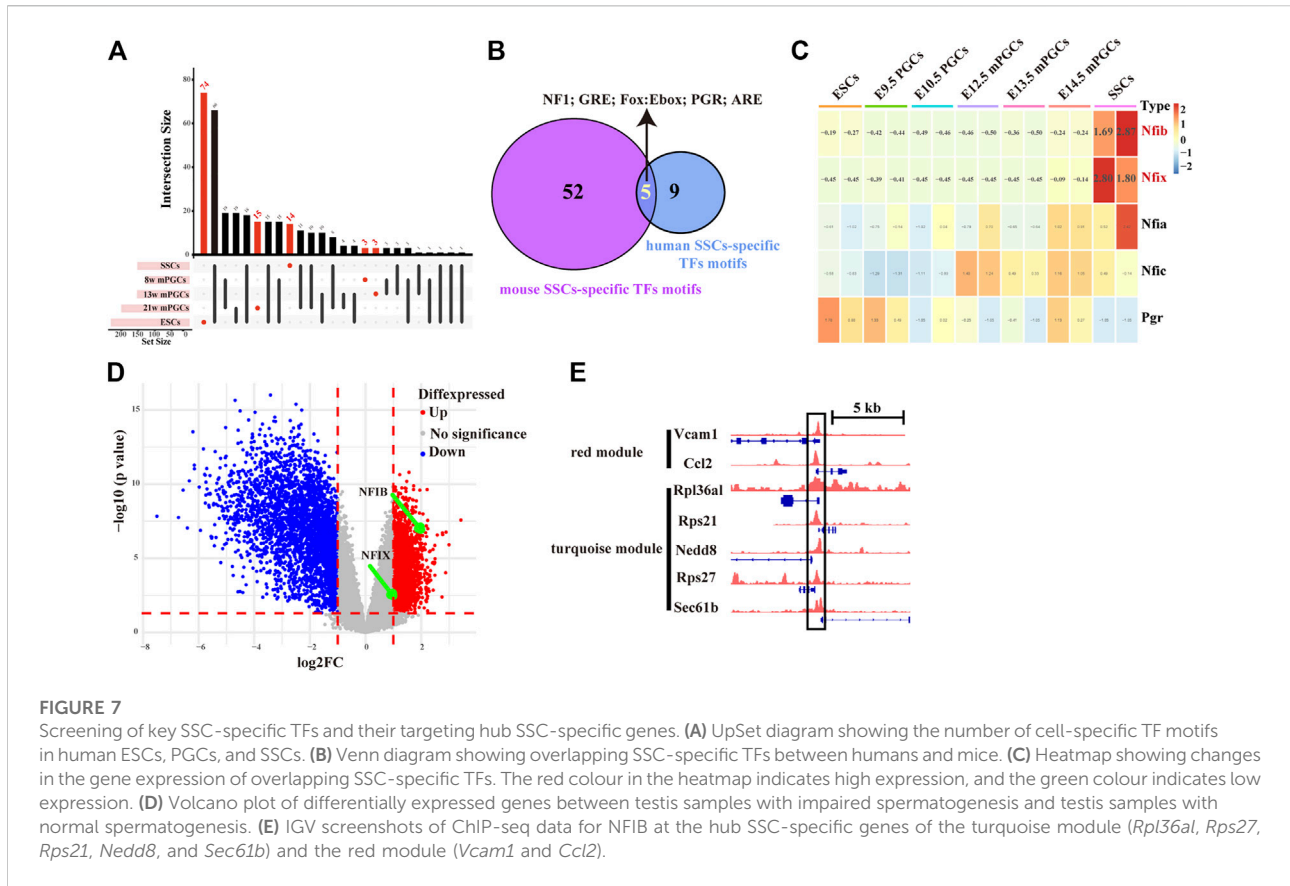
green ($R = -0.52$, $p = 2.1e-05$), and red ($R = -0.54$, p -value = $2.1e-05$) modules (Figures 5F–H), but no significant in the purple ($R = 0.011$, p -value = 0.87) module (Figure 5I).

Hub genes, GO enrichment and KEGG pathway analyses of the turquoise and red modules

We visualized the turquoise (Figure 6A) and red (Figure 6B) as networks in MCODE and screened out significant modules (score ≥ 4.5). However, the green and purple module are not a significant module in MCODE. Subsequently, the functional annotation for the hub genes in the turquoise and red modules was performed using DAVID. As the results show, the hub genes of the turquoise module were primarily enriched in structural constituent of ribosome (GO-Molecular Function) and ribosome (KEGG, Figure 6C). The hub genes of the red module, on the other hand, the results of enrichment analysis were mainly concerned with the cell surface and external side of plasma membrane (GO-Cellular Component, Figure 6D).

Inferring key SSC-Specific TFs and constructing key SSC-Specific TF-Mediated GRNs

To assess whether SSC-specific TFs in mice also have roles in human SSCs, we further screened key SSC-specific TFs.



Combining the landscape of the accessible chromatin in human ESCs, PGCs, and SSCs (Figure 7A), we finally obtained five overlapping SSC-specific TF motifs, including NF1 family TF motifs (NFIA, NFIB, NFIC, and NFIX), GRE, Fox:Ebox, PGR, and ARE, whereas no overlapping ESC-specific TF motifs and PGC-specific TF motifs were found (Figure 7B). Of these TFs, *Nfib* and *Nfix* exhibited abnormally high gene expression levels relative to mouse ESCs and PGCs (Figure 7C). Moreover, we also found, from the analysis of public databases (GSE145467), that *Nfib* ($\log_2\text{fold change (log2FC)} = 2.19, p\text{-value} = 3.38\text{E-}07$) and *Nfix* ($\log_2\text{FC} = 1.24, p\text{-value} = 1.15\text{E-}03$) were upregulated in testis samples with impaired spermatogenesis in comparison with testis samples with normal spermatogenesis (Figure 7D). Finally, to construct the key SSC-specific TF-mediated GRNs, we found in the ChIP-seq database (Cistrome Data Browser: 69,127) that NFIB most likely targeted the hub SSC-specific genes of the turquoise module (*Rpl36al*, *Rps27*, *Rps21*, *Neddd8*, and *Sec61b*) and the red module (*Vcam1* and *Ccl2*). (Figure 7E). However, for NFIX, no ChIP-Seq data were available. The results preliminarily implied that key SSC-specific TFs could be involved in the regulation of SSC formation and spermatogenesis by regulating their mediated GRNs. Nevertheless, experimental validation is required for confirmation.

Discussion

Currently, it is difficult to elucidate the mechanisms of human SSC formation because of the long period of the development process *in vivo* and restricted ethical issues (Martin and Seandel, 2013; Sohni et al., 2019). Alternatively, studies on the differentiation of ESCs to SSCs *in vitro* might be useful to understand human SSC formation and male infertility (Martin and Seandel, 2013; Li et al., 2019). Over the past years, a wide range of mouse models and *in vitro* cell culture systems have been used to explore the formation mechanisms of SSCs, inducing ESCs into PGCs and SSCs through transgenic technologies or by adding cytokines and chemical induction reagents to the culture medium (Geijsen et al., 2004; Wan et al., 2014; Li et al., 2019). However, the low efficiency and undefined induction factors of *in vitro* induction have always been important factors restricting this technology (Nagamatsu and Hayashi, 2017). Therefore, it is urgent to find key factors that influence the differentiation of ESCs into SSCs *in vitro* from a new perspective. In this study, by combining chromatin property data (ATAC-seq, DNase-seq, ChIP-seq) and gene expression data (RNA-seq, microarray data), cell-specific TFs and cell-specific TF-mediated GRNs in the process of SSC formation were identified. As a result, several key SSC-specific TFs and their

targeting hub SSC-specific genes were found to potentially be involved in regulating the differentiation of ESCs into SSCs *in vitro*.

By analysing the accessible chromatin of mouse ESCs, PGCs, and SSCs, we found that a large proportion of peaks were located close to TSS, and the spatial and temporal expression of genes varied in different developmental stages of male germ cells. Moreover, the integration of chromatin property data and gene expression data can contribute to decrypting transcriptional regulatory codes of male germ cell formation. In our analysis, through integrated chromatin property data (ATAC-seq, DNase-seq, ChIP-seq) and gene expression data (RNA-seq, microarray data), we found overlapping SSC-specific TFs between humans and mice and constructed TF-mediated GRNs in the process of SSC formation. It has previously been shown that the roles of TFs in GRNs are almost identical between humans and mice (Stergachis et al., 2014; Yue et al., 2014). Interestingly, Kim et al. found that the rewiring of GRNs contributes to the phenotypic discrepancies that occur between humans and mice (Ha et al., 2022).

We obtained five overlapping SSC-specific TF motifs between humans and mice, including NF1 family TF motifs (NFIA, NFIB, NFIC, and NFIX), GRE, Fox:Ebox, PGR, and ARE. Therein, *Nfib* and *Nfix* exhibited abnormally high gene expression levels relative to mouse ESCs and PGCs. Furthermore, *Nfib* and *Nfix* were upregulated in the testis sample with impaired spermatogenesis in comparison with the testis sample with normal spermatogenesis. These results suggest that NFIB and NFIX could be involved in the regulation of ESCs differentiation into SSCs and spermatogenesis.

The nuclear factor one (NFI) family of DNA binding proteins, previously also known as CCAAT box-binding transcription factors or TGGCA-binding proteins, has four members in vertebrates: NFIA, NFIB, NFIC, and NFIX (Borgmeyer et al., 1984; Zenker et al., 2019). In mice, *Nfib* and *Nfix* expression is most pronounced within the dorsal telencephalon and cerebellum. *Nfib* and *Nfix* knockout mice display severe brain phenotypes, suggesting that *Nfib* and *Nfix* are critical for brain development (Campbell et al., 2008; Bunt et al., 2015; Zenker et al., 2019). Other studies have found that *Nfib* knockout mice exhibit lung defects (Steele-Perkins et al., 2005), and *Nfic* knockout mice have abnormal teeth (Steele-Perkins et al., 2003). In humans, NFIB and NFIX have been related to human development and cancer (Zenker et al., 2019). NFIB and NFIX have been shown to act as either oncogenes or tumour suppressors across various cancers (Denny et al., 2016; Fane et al., 2017; Rahman et al., 2017; Zhao et al., 2021). Interestingly, evidence shows that NFIB is a key epigenetic regulator during development and within lung cancer (Denny et al., 2016). This study found that NFIB promotes metastasis of human small cell lung cancer (SCLC) cells through a widespread increase in chromatin accessibility (Denny et al., 2016). It has been reported that NFIX interference in human SSCs stimulates

propagation and suppresses early apoptosis of human SSCs; additionally, a study also found that NFIX negatively controls cyclin A2, cyclin B1, and cyclin E1 in human SSCs (Zhou et al., 2018). During PGCs development, it has been reported (Li et al., 2018) that the chromatin of mitotic-arrested male PGCs is permissive through nuclear transcription factor Y (NFY) binding in the distal regulatory regions, in contrast to that of meiotic female PGCs.

Notably, by analysing NFIB ChIP-seq data from the Cistrome database, we found that NFIB targets hub SSC-specific genes of the turquoise module (*Rpl36al*, *Rps27*, *Rps21*, *Nedd8*, and *Sec61b*) and the red module (*Vcam1* and *Ccl2*). *Rpl36al* is one of the X-linked human genes encoding ribosomal proteins. *Rpl36al* lacks introns in its coding regions, which was likely retrotransposed from X-linked genes (Uechi et al., 2002). Studies have found that *Rpl36al* may serve as a diagnostic biomarker based on immune infiltrates in Alzheimer's disease (Liu et al., 2021). Ribosomal protein S27 (RPS27) and RPS21 are part of the ribosomal protein. RPS27 might play a role in the initiation of translation (Feldheim et al., 2020). Furthermore, RPS27 protein was specifically expressed in tumour cells and neurons but not in healthy astrocytes (Feldheim et al., 2020). In contrast, fewer studies have examined RPS21, and a recent study found that RPS21 plays an essential role in the invasive behaviour of osteosarcoma cells through the inactivation of the mitogen-activated protein kinase (MAPK) pathway (Wang et al., 2020). Neural precursor cell expressed developmentally downregulated-8 (NEDD8) is a ubiquitin-like molecule that can be transferred to substrates to regulate protein function through a process termed protein neddylation (Kamitani et al., 1997), and *Nedd8* is expected to play a role as a therapeutic target in cancer (Jiang et al., 2020). The Sec61 translocon subunit beta (SEC61B) complex is the central component of the protein translocation apparatus of the endoplasmic reticulum membrane (Feng et al., 2017). Studies have reported that *Sec61b* was newly detected as a candidate gene involved in ovarian clear cell carcinogenesis (Yamada et al., 2021). Unlike the above genes, in addition to being involved in a range of cancers and diseases, *Vcam1* and *Ccl2* might play a crucial part in testicular interstitial tissues. Recently, among the 3D-reaggregated cultures of dissociated testicular cells from prepubertal mice, it was found that VCAM1, one of the ligands for integrins $\alpha 4\beta 1$ and $\alpha 9\beta 1$, is expressed mainly in CD34⁺ cells and adult Leydig cells but not in foetal Leydig cells, implying that the VCAM1- $\alpha 4\beta 1$ integrin interaction plays an important role in the formation of testicular interstitial tissues *in vitro* and *in vivo* (Abe et al., 2021). In autoimmune orchitis, as a chemotactic factor, CCL2 can induce attraction and extravasation of immune cells within the testicular interstitium (Guazzone et al., 2009). Although no studies have reported that these hub SSC-specific genes are involved in the differentiation of ESCs into SSCs, our results may provide some evidence for this aspect.

There are limitations in the study. First, we obtained five overlapping SSC-specific TF motifs between humans and mice, whereas no overlapping ESC-specific TF motifs and PGC-specific TF motifs were found. Thus, subsequent studies focused on the SSC-specific genes and associated SSC-specific TFs. Furthermore, we found that the overlapping SSC-specific TFs (NFIB and NFIX) exhibited abnormally high gene expression levels relative to ESCs and PGCs, but the expression level of *Nfib* and *Nfix* did not show an obvious change between ESCs and PGCs. These results may reflect the regulatory network from PGCs to SSCs, but not the upstream regulation from ESCs to PGCs. Second, no NFIB and NFIC ChIP-seq data from the mouse SSCs were found by searching the NCBI-GEO and PubMed. Ultimately, we found in a ChIP-seq database (Cistrome Data Browser: 69,127) that NFIB most likely targeted the hub SSC-specific genes of the turquoise module (*Rpl36a1*, *Rps27*, *Rps21*, *Neddd8*, and *Sec61b*) and the red module (*Vcam1* and *Ccl2*). It is undeniable that the evidence was not sufficient enough to use the ChIP-seq database (Cistrome Data Browser: 69,127) comes from the mammary gland. These results may only provide suggestions for future research.

Conclusion

In conclusion, after a comprehensive analysis of chromatin property data (ATAC-seq, DNase-seq, ChIP-seq) and gene expression data (RNA-seq, microarray data), we preliminarily identified SSC-specific TFs and constructed TF-mediated GRNs in the process of SSC formation. The key SSC-specific TFs (NFIB, NFIX) and their targeting hub SSC-specific genes were specifically analysed. The results also imply that NFIB and NFIX could be involved in the regulation of SSC formation and spermatogenesis. Our investigation establishes a foundation for future research aiming to elucidate the role of these TFs and their targeting hub genes in SSC formation and provides potential induction factors for further optimizing the induction efficiency of the differentiation of ESCs into SSCs *in vitro*.

Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

Ethics statement

All data in this study were collected from public databases. This article does not contain any studies with patients or animals performed by any of the authors.

Author contributions

HY, BW, and KS conceived and designed the overall study. HY, KS, and BW participated in data processing and program implementation. HY, KS, BW, LD, SW, and XF participated in the study design. HY, KS, BW, LD, SW, and XF drafted the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the Key Technology Research Plan Project of Inner Mongolia Autonomous Region (2021GG0153) and grants from the National Natural Science Foundation of China (31760333) to HY and a grant from Science and Technology Major Project of Inner Mongolia Autonomous Region of China to the State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock (2019ZD031).

Acknowledgments

We thank Dr. Jianming Zeng (University of Macau) and all the members of his bioinformatics team, biotrainee, for generously sharing their experience and codes. We also thank ShuFang for provided language help and writing assistance.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphys.2022.949486/full#supplementary-material>

References

- Abdelal, N. E., Tanga, B. M., Abdelgawad, M., Allam, S., Fathi, M., Saadeldin, I. M., et al. (2021). Cellular therapy via spermatogonial stem cells for treating impaired spermatogenesis, non-obstructive azoospermia. *Cells* 10 (7), 1779. doi:10.3390/cells10071779
- Abe, K., Kon, S., Kameyama, H., Zhang, J., Morohashi, K. I., Shimamura, K., et al. (2021). VCAM1- α 4 β 1 integrin interaction mediates interstitial tissue reconstruction in 3-D re-aggregate culture of dissociated prepubertal mouse testicular cells. *Sci. Rep.* 11 (1), 18332. doi:10.1038/s41598-021-97729-y
- Borgmeyer, U., Nowock, J., and Sippel, A. E. (1984). The TGGCA-binding protein: A eukaryotic nuclear protein recognizing a symmetrical sequence on double-stranded linear DNA. *Nucleic Acids Res.* 12 (10), 4295–4311. doi:10.1093/nar/12.10.4295
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10 (12), 1213–1218. doi:10.1038/nmeth.2688
- Bunt, J., Lim, J. W., Zhao, L., Mason, S., and Richards, L. J. (2015). PAX6 does not regulate Nfia and Nfib expression during neocortical development. *Sci. Rep.* 5, 10668. doi:10.1038/srep10668
- Campbell, C. E., Piper, M., Plachez, C., Yeh, Y. T., Baizer, J. S., Osinski, J. M., et al. (2008). The transcription factor Nfix is essential for normal brain development. *BMC Dev. Biol.* 8, 52. doi:10.1186/1471-213x-8-52
- Denny, S. K., Yang, D., Chuang, C. H., Brady, J. J., Lim, J. S., Grüner, B. M., et al. (2016). Nfib promotes metastasis through a widespread increase in chromatin accessibility. *Cell* 166 (2), 328–342. doi:10.1016/j.cell.2016.05.052
- Fane, M., Harris, L., Smith, A. G., and Piper, M. (2017). Nuclear factor one transcription factors as epigenetic regulators in cancer. *Int. J. Cancer* 140 (12), 2634–2641. doi:10.1002/ijc.30603
- Feldheim, J., Kessler, A. F., Schmitt, D., Salvador, E., Monoranu, C. M., Feldheim, J. J., et al. (2020). Ribosomal protein S27/metallopanstimulin-1 (RPS27) in glioma-A new disease biomarker? *Cancers (Basel)* 12 (5), 1085. doi:10.3390/cancers12051085
- Feng, S., Sekine, S., Pessino, V., Li, H., Leonetti, M. D., and Huang, B. (2017). Improved split fluorescent proteins for endogenous protein labeling. *Nat. Commun.* 8 (1), 370. doi:10.1038/s41467-017-00494-8
- Geijsen, N., Horoschak, M., Kim, K., Gribnau, J., Eggen, K., and Daley, G. Q. (2004). Derivation of embryonic germ cells and male gametes from embryonic stem cells. *Nature* 427 (6970), 148–154. doi:10.1038/nature02247
- Guazzzone, V. A., Jacobo, P., Theas, M. S., and Lustig, L. (2009). Cytokines and chemokines in testicular inflammation: A brief review. *Microsc. Res. Tech.* 72 (8), 620–628. doi:10.1002/jemt.20704
- Ha, D., Kim, D., Kim, I., Oh, Y., Kong, J., Han, S. K., et al. (2022). Evolutionary rewiring of regulatory networks contributes to phenotypic differences between human and mouse orthologous genes. *Nucleic Acids Res.* 50 (4), 1849–1863. doi:10.1093/nar/gkac050
- Han, P., Gopalakrishnan, C., Yu, H., and Wang, E. (2017). Gene regulatory network rewiring in the immune cells associated with cancer. *Genes (Basel)* 8 (11), 308. doi:10.3390/genes8110308
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., et al. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38 (4), 576–589. doi:10.1016/j.molcel.2010.05.004
- Huang, C., Huang, R., Chen, H., Ni, Z., Huang, Q., Huang, Z., et al. (2020). Chromatin accessibility regulates gene expression and correlates with tumor-infiltrating immune cells in gastric adenocarcinoma. *Front. Oncol.* 10, 609940. doi:10.3389/fonc.2020.609940
- 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 (1), 44–57. doi:10.1038/nprot.2008.211
- Ishikura, Y., Ohta, H., Sato, T., Murase, Y., Yabuta, Y., Kojima, Y., et al. (2021). *In vitro* reconstitution of the whole male germ-cell development from mouse pluripotent stem cells. *Cell Stem Cell* 28 (12), 2167–2179.e9. doi:10.1016/j.stem.2021.08.005
- Jiang, Y., Cheng, W., Li, L., Zhou, L., Liang, Y., Zhang, W., et al. (2020). Effective targeting of the ubiquitin-like modifier NEDD8 for lung adenocarcinoma treatment. *Cell Biol. Toxicol.* 36 (4), 349–364. doi:10.1007/s10565-019-09503-6
- Kamitani, T., Kito, K., Nguyen, H. P., and Yeh, E. T. (1997). Characterization of NEDD8, a developmentally down-regulated ubiquitin-like protein. *J. Biol. Chem.* 272 (45), 28557–28562. doi:10.1074/jbc.272.45.28557
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). Hisat: A fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi:10.1038/nmeth.3317
- Kossack, N., Meneses, J., Shefi, S., Nguyen, H. N., Chavez, S., Nicholas, C., et al. (2009). Isolation and characterization of pluripotent human spermatogonial stem cell-derived cells. *Stem Cells* 27 (1), 138–149. doi:10.1634/stemcells.2008-0439
- Krausz, C., and Riera-Escamilla, A. (2018). Genetics of male infertility. *Nat. Rev. Urol.* 15 (6), 369–384. doi:10.1038/s41585-018-0003-3
- Langfelder, P., and Horvath, S. (2008). Wgcna: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359. doi:10.1038/nmeth.1923
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352
- Li, J., Shen, S., Chen, J., Liu, W., Li, X., Zhu, Q., et al. (2018). Accurate annotation of accessible chromatin in mouse and human primordial germ cells. *Cell Res.* 28 (11), 1077–1089. doi:10.1038/s41422-018-0096-5
- Li, N., Ma, W., Shen, Q., Zhang, M., Du, Z., Wu, C., et al. (2019). Reconstitution of male germline cell specification from mouse embryonic stem cells using defined factors *in vitro*. *Cell Death Differ.* 26 (10), 2115–2124. doi:10.1038/s41418-019-0280-2
- Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30 (7), 923–930. doi:10.1093/bioinformatics/btt656
- Liu, Z., Li, H., and Pan, S. (2021). Discovery and validation of key biomarkers based on immune infiltrates in Alzheimer's disease. *Front. Genet.* 12, 658323. doi:10.3389/fgene.2021.658323
- Ma, S., and Zhang, Y. (2020). Profiling chromatin regulatory landscape: Insights into the development of ChIP-seq and ATAC-seq. *Mol. Biomed.* 1 (1), 9. doi:10.1186/s43556-020-00009-w
- Maewaza, S., Yukawa, M., Alavattam, K. G., Barski, A., and Namekawa, S. H. (2018). Dynamic reorganization of open chromatin underlies diverse transcriptomes during spermatogenesis. *Nucleic Acids Res.* 46 (2), 593–608. doi:10.1093/nar/gkx1052
- Maher, K. A., Bajic, M., Kajala, K., Reynoso, M., Pauluzzi, G., West, D. A., et al. (2018). Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. *Plant Cell* 30 (1), 15–36. doi:10.1105/tpc.17.00581
- Martin, L. A., and Seandel, M. (2013). Propagation of adult SSCs: From mouse to human. *Biomed. Res. Int.* 2013, 384734. doi:10.1155/2013/384734
- Nagamatsu, G., and Hayashi, K. (2017). Stem cells, *in vitro* gametogenesis and male fertility. *Reproduction* 154 (6), F79–f91. doi:10.1530/rep-17-0510
- Nikolae, L. G., Akopov, S. B., Chernov, I. P., and Sverdllov, E. D. (2007). Maps of cis-regulatory nodes in megabase long genome segments are an inevitable intermediate step toward whole genome functional mapping. *Curr. Genomics* 8 (2), 137–149. doi:10.2174/138920207780368178
- Oatley, J. M., and Brinster, R. L. (2008). Regulation of spermatogonial stem cell self-renewal in mammals. *Annu. Rev. Cell Dev. Biol.* 24, 263–286. doi:10.1146/annurev.cellbio.24.110707.175355
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26 (6), 841–842. doi:10.1093/bioinformatics/btq033
- Rahman, N. I. A., Abdul Murad, N. A., Mollah, M. M., Jamal, R., and Harun, R. (2017). NFIX as a master regulator for lung cancer progression. *Front. Pharmacol.* 8, 540. doi:10.3389/fphar.2017.00540
- Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., and Manke, T. (2014). DeepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, W187–W191. doi:10.1093/nar/gku365
- Shashikant, T., and Etensohn, C. A. (2019). Genome-wide analysis of chromatin accessibility using ATAC-seq. *Methods Cell Biol.* 151, 219–235. doi:10.1016/bs.mcb.2018.11.002
- Sin, H. S., Kartashov, A. V., Hasegawa, K., Barski, A., and Namekawa, S. H. (2015). Poised chromatin and bivalent domains facilitate the mitosis-to-meiosis transition in the male germline. *BMC Biol.* 13, 53. doi:10.1186/s12915-015-0159-8

- Singh, H., Medina, K. L., and Pongubala, J. M. (2005). Contingent gene regulatory networks and B cell fate specification. *Proc. Natl. Acad. Sci. U. S. A.* 102 (14), 4949–4953. doi:10.1073/pnas.0500480102
- Sohni, A., Tan, K., Song, H. W., Burow, D., De Rooij, D. G., Laurent, L., et al. (2019). The neonatal and adult human testis defined at the single-cell level. *Cell Rep.* 26 (6), 1501–1517. e1504. doi:10.1016/j.celrep.2019.01.045
- Steele-Perkins, G., Butz, K. G., Lyons, G. E., Zeichner-David, M., Kim, H. J., Cho, M. I., et al. (2003). Essential role for NFI-C/CTF transcription-replication factor in tooth root development. *Mol. Cell. Biol.* 23 (3), 1075–1084. doi:10.1128/mcb.23.3.1075-1084.2003
- Steele-Perkins, G., Plachez, C., Butz, K. G., Yang, G., Bachurski, C. J., Kinsman, S. L., et al. (2005). The transcription factor gene Nfib is essential for both lung maturation and brain development. *Mol. Cell. Biol.* 25 (2), 685–698. doi:10.1128/mcb.25.2.685-698.2005
- Stergachis, A. B., Neph, S., Sandstrom, R., Haugen, E., Reynolds, A. P., Zhang, M., et al. (2014). Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* 515 (7527), 365–370. doi:10.1038/nature13972
- Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* 14 (2), 178–192. doi:10.1093/bib/bbs017
- Treize, S., and Nutt, S. L. (2021). The gene regulatory network controlling plasma cell function. *Immunol. Rev.* 303 (1), 23–34. doi:10.1111/imr.12988
- Uechi, T., Maeda, N., Tanaka, T., and Kenmochi, N. (2002). Functional second genes generated by retrotransposition of the X-linked ribosomal protein genes. *Nucleic Acids Res.* 30 (24), 5369–5375. doi:10.1093/nar/kkf696
- Wan, Q., Lu, H., Wu, L. T., Liu, X., and Xiang, J. B. (2014). Retinoic acid can induce mouse embryonic stem cell R1/E to differentiate toward female germ cells while oleanolic acid can induce R1/E to differentiate toward both types of germ cells. *Cell Biol. Int.* 38 (12), 1423–1429. doi:10.1002/cbin.10380
- Wang, T., Wang, Z. Y., Zeng, L. Y., Gao, Y. Z., Yan, Y. X., and Zhang, Q. (2020). Down-regulation of ribosomal protein RPS21 inhibits invasive behavior of osteosarcoma cells through the inactivation of MAPK pathway. *Cancer Manag. Res.* 12, 4949–4955. doi:10.2147/cmar.S246928
- Yamada, Y., Miyamoto, T., Higuchi, S., Ono, M., Kobara, H., Asaka, R., et al. (2021). CDNA expression library screening revealed novel functional genes involved in clear cell carcinogenesis of the ovary *in vitro*. *J. Obstet. Gynaecol.* 41 (1), 100–105. doi:10.1080/01443615.2020.1716310
- Yu, G., Wang, L. G., and He, Q. Y. (2015). ChIPseeker: An R/bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31 (14), 2382–2383. doi:10.1093/bioinformatics/btv145
- Yu, H., Wu, H., Zheng, F., Zhu, C., Yin, L., Dai, W., et al. (2020). Gene-regulatory network analysis of ankylosing spondylitis with a single-cell chromatin accessible assay. *Sci. Rep.* 10 (1), 19411. doi:10.1038/s41598-020-76574-5
- Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515 (7527), 355–364. doi:10.1038/nature13992
- Zenker, M., Bunt, J., Schanze, I., Schanze, D., Piper, M., Priolo, M., et al. (2019). Variants in nuclear factor I genes influence growth and development. *Am. J. Med. Genet. C Semin. Med. Genet.* 181 (4), 611–626. doi:10.1002/ajmg.c.31747
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoutte, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.* 9 (9), R137. doi:10.1186/gb-2008-9-9-r137
- Zhao, L., Song, X., Guo, Y., Ding, N., Wang, T., and Huang, L. (2021). Long non-coding RNA SNHG3 promotes the development of non-small cell lung cancer via the miR-1343-3p/NFIX pathway. *Int. J. Mol. Med.* 48 (2), 147. doi:10.3892/ijmm.2021.4980
- Zhou, F., Yuan, Q., Zhang, W., Niu, M., Fu, H., Qiu, Q., et al. (2018). MiR-663a stimulates proliferation and suppresses early apoptosis of human spermatogonial stem cells by targeting NFIX and regulating cell cycle. *Mol. Ther. Nucleic Acids* 12, 319–336. doi:10.1016/j.omtn.2018.05.015