Dynamic chromatin accessibility in spermatogonial cells for transcriptional programming from early postnatal to adult stages

Irina Lazar-Contes[1*], Deepak K. Tanwar[1*], Pierre-Luc Germain[1,2], Isabelle M. Mansuy[1#]

1. Laboratory of Neuroepigenetics, Brain Research Institute, Medical Faculty of the University of Zurich, and Institute for Neuroscience, Department of Health Science and Technology of the ETH Zurich, Zurich, Switzerland
2. Statistical Bioinformatics Group, Swiss Institute of Bioinformatics, Zurich, Switzerland

*Authors share equal contribution

[#]Correspondence: mansuy@hifo.uzh.ch

Abstract

Stem cells have the capacity to self-renew and differentiate to give rise to tissues. In mammals, spermatogonial stem cells are unique undifferentiated cells in the male germline that are present throughout life and produce functional sperm. They have a remarkable dynamic transcriptome between postnatal life and adulthood, thought to allow adaptation to the maturing gonadal environment. To assess if chromatin remodelling is associated with this dynamic, we characterized chromatin accessibility in postnatal and adult spermatogonial cells in mice using ATAC-seq. Extensive changes in chromatin accessibility were observed in spermatogonial cells across postnatal development, that correlate with distinct gene expression profiles and transcription factor motif enrichment. We identify genomic regions with differential chromatin accessibility in adult spermatogonial cells, that are marked by distinct histone modifications and are situated in proximity to the transcription start site of genes important for cell maintenance and proliferation. Some of these regions correspond to transposable element subtypes enriched in multiple transcription factor motifs and with increased transcription. Together, our results reveal profiles of

chromatin organization, histone modifications and gene expression in spermatogonial cells, and underscore the dynamic nature of the germline genome.

Introduction

Spermatogonial cells are cells in male gonads that actively self-renew and differentiate into spermatogenic cells to produce adult gametes. In mice, they become active one to two days after birth, when they exit mitotic arrest and start dividing, to populate the basement membrane of seminiferous tubules. During the first week of postnatal life, a population of spermatogonial cells continues to proliferate and gives rise to undifferentiated A_{single} (A_s), A_{paired} (A_{pr}) and $A_{aligned}$ (A_{al}) cells. The remaining spermatogonia differentiate to form chains of daughter cells that become primary and secondary spermatocytes around postnatal day (PND) 10 to 12. Spermatocytes undergo meiosis and give rise to haploid spermatids that develop into spermatozoa. Spermatozoa are then released in the lumen of the seminiferous tubules and continue to mature in the epididymis until becoming capable of fertilization by PND 42-48 [1–3].

Recent work using population or single-cell RNA sequencing (RNA-seq) showed that distinct transcriptional profiles characterize spermatogonial cells in postnatal and adult life [4–8]. During the first week after birth, spermatogonia have unique features necessary for their rapid establishment and expansion along the basement membrane. This includes high expression of genes involved in cell cycle regulation, stem cell proliferation, transcription and RNA [9]. In comparison, in the adult testis, genes are involved in the maintenance of a steady spermatogonial cells population, with a balance between proliferation and differentiation to ensure sperm formation predominates. This includes pathways related to paracrine signalling and niche communication, as well as mitochondrial function and oxidative phosphorylation [6,9].
Epigenetic changes such as histone tail posttranslational modifications and DNA methylation accompany transcriptional differences in spermatogonial cells across postnatal stages [4,5]. However, little is known about the accessible chromatin landscape, during the transition from postnatal to the adult stage.

We examined chromatin reorganization in the transition from postnatal to adult spermatogonia and its functional relevance by profiling chromatin accessibility by ATAC-seq (assay for transposase-accessible chromatin with high-throughput sequencing) using an Omni-ATAC protocol [10] and integrating data with transcriptomic and epigenetic profiles. The results show that chromatin is extensively remodelled in spermatogonial cells during development, and its accessibility is increased at several genomic regions particularly transposable elements (TEs), that correlate with transcriptional and epigenetic signatures.

Results

Enrichment of spermatogonial cells from postnatal and adult mouse testis

We collected testes from mouse pups at postnatal days (PND) 8 and 15 and from males at postnatal week (PNW) 20, and prepared cell suspensions by enzymatic digestion. Spermatogonial cells were enriched by fluorescence-activated cell sorting (FACS) using surface markers (Fig. S1A) [11]. Immunocytochemistry using PLZF, a well-established marker of undifferentiated spermatogonia [12], confirmed that our preparation was enriched in spermatogonial cells, with 85-95% $PLZF^+$ cells after FACS compared to 3-6% $PLZF^+$ cells before (Fig. S1B). Transcriptomic analyses by RNA-seq validated the spermatogonial identity of the cells, showing high expression of stem cell and undifferentiated spermatogonial markers, and low expression of somatic cells (Leydig and Sertoli cells) markers both at PND 8 and 15 (Fig. S1C).

Chromatin is remodelled in spermatogonial cells during the development

We profiled chromatin accessibility in postnatal and adult spermatogonial cells using an Omni-ATAC protocol [10]. Omni-ATAC has a higher signal-to-noise ratio than classical ATAC-seq and can be used with low input material (a few hundred cells). Accessible regions in the genome were identified by peak-calling on merged nucleosome-free fragments (NFF). Following the removal of lowly enriched regions,

158,978 regions were selected for downstream analyses (see Methods section for details). Most Tn5-accessible regions are intergenic (38%) and located in gene bodies (33%) or in the proximity of a transcription start site (TSS) (28% +/-1 kb from TSS) (Fig. S2A). 3212 differentially accessible regions were identified between PND15 and adult spermatogonia with the majority showing a gain in accessibility in adult cells (Fig. 1A and Table S1). Regions of differential accessibility are predominantly intergenic (45%) and intronic (34%), and 15% of all differentially accessible regions are +/- 1kb from a TSS (Fig. 1B). Gene ontology (GO) analysis, showed that regions with increased accessibility in adult spermatogonia are associated with cell fate and stem cell population maintenance, protein metabolism and RNA metabolic processes (Fig. 1C and Table S1). Separate GO analyses depending on genomic location showed that regions located in gene bodies (mainly introns) are enriched for terms related to reproduction and protein metabolism, whilst regions close to or overlapping with a TSS relate to cell fate specification and tissue morphogenesis (Fig. S2B and Table S1). Regions with decreased accessibility in adult spermatogonia are predominantly located in intergenic regions and associated with multiple terms related to embryonic development (Fig. 1C and Table S1).

Differentially accessible chromatin regions associated with distinct gene expression dynamics

To assess the relevance of changes in chromatin accessibility, we first integrated our ATAC-Seq data with transcriptomic datasets and examined the correlation between chromatin states and transcriptome. We conducted RNA-seq analyses on postnatal spermatogonial cells and used published data from Thy1+ adult spermatogonial cells. We found that 719 genes are differentially expressed in spermatogonial cells between PND8 and PND15, with 515 being upregulated and 212 downregulated (Fig. 2A and Table S2). Fast Gene Set Enrichment Analysis (FGSEA) showed that pathways related to RNA processing and splicing, cell cycle, redox homeostasis and protein catabolism are downregulated between PND8 and PND15 while pathways associated with cellular transport, exocytosis and signal transduction are upregulated (Fig. 2B and Table S2). Comparable analyses using published RNA-seq data from PND14 and adult spermatogonia cells [4,5] similarly identified transcriptional changes

between postnatal and adult stages (Fig. S?A and Table S2) and downregulation of pathways related to RNA processing, ribosome biogenesis and cell cycle in adults compared to postnatal cells. Pathways related to developmental programs and mitochondrial functions were also downregulated and those related to spermatogenesis or involving cytokine signalling were upregulated (Fig. S? and Table S2).

To integrate chromatin accessibility and transcriptomic datasets, we first divided differentially-accessible regions into proximal (situated less than +/- 2.5 kb from a TSS) and distal (situated more than +/- 2.5 kb from a TSS) following ENCODE practice [13–15]. For proximal regions, we further defined 6 different categories based on the change in expression of the nearest gene. The first two include proximal regions with increased chromatin accessibility and upregulated (Category 1) or downregulated (Category 2) expression of the nearest gene. Category 3 and 4 include regions of less chromatin accessibility with the nearest gene either downregulated or upregulated, respectively. Category 5 and 6 include regions with increased or decreased chromatin accessibility respectively, whose nearest gene is not expressed in spermatogonial cells (Fig. 3A and Table S3).

Differentially accessible chromatin regions associated with distinct epigenetic profiles

Next, we examined the relationship between chromatin accessibility and epigenetic marks using published ChIP-seq and bisulfite sequencing (BS) datasets from Thy1$^+$ spermatogonia [4,5]. Profiles of histone marks including H3K4me3, H3K27ac and H3K27me3 in adult spermatogonial cells, and DNA methylation in PND7, PND14 and adult spermatogonial cells were used. …..analyses showed that for a subset of regions in Category 1, there is an overlap with active H3K4me3, H3K27ac or dual H3K4me3/K27ac modifications, and an overall lack of H3K27me3 (Fig. 3B and Table S3). Notably, several of the genes in Category 1 with chromatin opening marked by histone modification(s) are known regulators of stem cell potency. *Pdpk1* promoter region was marked by dual H3K4me3/K27ac, while *Pdpk1* mRNA was upregulated

in adult spermatogonia (Fig. S3A). *Pdpk1* (phosphoinositide-dependent protein kinase 1) is a glycolysis factor important for stem cell self-renewal [16,17]. In contrast, the *Gata2* promoter region was marked by the bivalent H3K4me3/K27me3 mark, while *Gata2* expression showed an upregulation across testis maturation (Fig. S3). *Gata2* (GATA-binding factor 2) is a known target of NANOS2, an essential regulator of spermatogonial stem cell potential [18,19]. Other exemplary genes in Category 1 include pyruvate cellular carriers *Slc25a18, Slc23a1* and *Slc2a5*, suggesting differences in glycolysis regulation in adult spermatogonial cells (Table S3). Notably, we found increased chromatin accessibility at the TSS of GDNF receptor *Gfra2* (Fig. 3C). At the mRNA level, *Gfra2*displayed a marked upregulation in adult spermatogonial cells, indicating an increased utilization of GFRA2 receptors in adult spermatogonial cells compared to early postnatal stages, in which GFRA1-mediated signalling is dominant (Fig. 3C) [5,9]. We also observed increased chromatin accessibility and upregulation of DNA damage responsive gene *Fus*, and the multifunctional redox gene *Prdx4* (Table S3).

Interestingly, the highest number of differentially accessible chromatin regions were in Category 2, and included proximal regions with increased chromatin accessibility and decreased expression of nearby genes in adult spermatogonia, indicative of active repression taking place (Fig. 3A and Table S3). A subset of Category 2 regions was marked by H3K27me3 together with H3K4me3 and mainly associated with developmental genes such as *Satb1* and *Hmx1, (*Fig. S3B). Other developmental genes such as *Tbx4* also displayed a decreased accessibility and an increase in mRNA expression (Fig. 3C). GO enrichment analysis revealed that regions in Category 2 are associated with regulation of cell cycle, RNA processing, DNA repair and cell division (Table S4). Such an example is is *Fgf8*, important for *Fgf8-Fgfr1* mediated maintenance of undifferentiated spermatogonia [20]. *Fgf8* showed increased chromatin accessibility close to its TSS and a downregulated expression in adult spermatogonia, in agreement with recent findings from scRNA-seq data suggesting downregulation of *Fgfr1*-mediated signalling with age(Fig. S3B) [9,20].

Regions in Category 3 displayed decreased chromatin accessibility and downregulation of nearby genes in adult spermatogonia and were mostly depleted of

any of the 3 histone marks investigated (Fig. 3A and 3B). GO enrichment on the nearby genes revealed an association with developmental processes and WNT signalling (Table S4). A notable example we identified in this category is *Pdgfra,* a gene involved in the hepatic stellate cell activation pathway, which was recently identified by scRNA-seq to be upregulated in spermatogonial stem cells in the immature testis compared to the adult stage [6]. Pdgfra displayed a marked downregulation in adult spermatogonia, and a decrease in chromatin accessibility overlapping its TSS (Fig. 3C). Another example of a gene important for early postnatal spermatogonial cell establishment is Dap2ip, which we also found to carry a marked decrease in chromatin accessibility at its TSS and lower expression in adult spermatogonial cells (Fig. S3C). Surprisingly, we also identified a category of proximal regions (Category 4) with decreased accessibility (<20 regions) at genes that were upregulated in adult spermatogonial cells and with a previously uncharacterized role in spermatogonial cells (Fig. 3A). Notably, DNAme profiles across postnatal stages did not show any significant change across any of the 6 categories of proximal regions, suggesting a relatively stable DNAme profile in the transition from early postnatal to adult stage (Fig. S3D). Aside from proximal regions, we also identified numerous distal regions with differential chromatin accessibility between PND15 and adult spermatogonial cells (Fig. S4A). Similar to proximal regions, accessibility in distal regions mainly increased in adult spermatogonial cells compared to the early postnatal stage. When integrating the literature ChIP-seq data, we observed enrichment for H3K4me3, H3K27ac and H3K27me3 at a small number of the differentially accessible distal regions, indicative of potential regulatory roles (Fig. S4B and Table S3). Similar to proximal regions, DNAme levels did not display major changes in the transition from early postnatal to adult stage (Fig. S4B). Taken together, our data integration reveals novel associations between chromatin regions of differential accessibility, histone marks and gene transcription in spermatogonial cells from postnatal to adult stages of development. They suggest that chromatin accessibility is a mechanism of control of transcriptional programs and of certain signalling pathways during development.

Accessibility changes at open chromatin regions carry binding sites for distinct families of transcription factors

Transcription factors (TFs) are essential for establishing and maintaining transcriptional programs across developmental stages in cells [21,22]. To examine if regions of different chromatin accessibility between PND15 and adult spermatogonia are enriched in regulatory elements such as TF binding motifs, we performed motif enrichment analysis using the Hypergeometric Optimization of Motif EnRichment (HOMER) tool [23]. In regions with increased chromatin accessibility, we identified 41 enriched TF motifs (q-value ≤ 0.05) (Fig. 4A). Top candidates by significance (q-value ≤ 0.0001) are members of the Fos/Jun family (FOS, FOSB, FOSL1 and FOSL2, JUN, JUNB and JUND) (Fig. 4B). Notably, at the mRNA level, some of the TFs displayed age-specific differences (Log_2CPM ≥ 1 and abs Log_2 fold change ≥ 1): *Fos, Junb* and *Jund* were downregulated in adult spermatogonial cells (Fig. 4C). JUN, FOS and CREB are all part of the AP-1 (activating protein-1) superfamily, and play an important role in regulating cell proliferation and death, by mediating the senescence-associated chromatin and transcriptional landscape [24,25]. JUND and c-FOS specifically promote the proliferative potential of spermatogonial stem cells[26,27]. USF1 and POU3F1, *2* factors important in the maintenance of the spermatogonial stem cell pool also displayed enriched binding motifs in the more accessible regions. However, their mRNA levels were downregulated in adult spermatogonial cells (Fig. 4C) [28,29]. POU3F1 is a GDNF-regulated TF, which has been shown to play an important role in promoting spermatogonial cell self-renewal capacity [29,30]. Notably, *Pou3f1* downregulation in adult spermatogonial cells coincides with marked downregulation of *Gfra1* and upregulation of *Gfra2* (Fig. 3C and Table S2). TF motif analysis using HOMER also revealed enriched binding sites for retinoic acid receptors such as RXRα and RARα (Fig. 4A). Recently, expression of RA receptors *Rxrα* and *Rarα* was reported in the stem cell-containing population of spermatogonial cells in both pup and adult testis, together with evidence that their utilization in spermatogonial cells is vastly dependent on the niche microenvironment [31].

To check if some TF binding motifs are preferentially enriched in certain genomic locations, we performed motif enrichment analysis for more accessible chromatin regions situated in gene bodies, intergenic regions and in regions +/- 1kb from TSS. We identified several TF motifs specifically enriched in intergenic regions, specifically members of the ubiquitously expressed NF-Y complex, NF-YA, NF-YB and NF-YC (Fig. 4D). In mESCs, NF-Y TF family members located in distal regions facilitate a permissive chromatin conformation and play an important role in the expression of core ESC pluripotency genes [32]. Furthermore, NF-YA/B motif enrichment has also been found in regions of open chromatin in human spermatogonial cells [33].

Although regions of more accessible chromatin encompass the majority of the differentially accessible regions in adult spermatogonia compared to PND15, less accessible chromatin also displayed a high number of enriched TF binding motifs (Fig. 4A). Notably, almost all of these TF motifs were uniquely enriched in the regions of decreased chromatin accessibility and predominantly associated with developmental factors. Top hits included members of the FOX family (FOXO1, FOXO3, FOXP2, FOXK1, FOXA2) and members of the ETS and ETS-related families (ETS1, GABPA, ETV4, ELF1, ELF3) (Fig. 4B). The gene expression levels of most of these TFs were decreased in adult spermatogonial cells (Fig. 4A). FOXO1 is a pivotal regulator of the self-renewal and differentiation of spermatogonial stem cells in both pup and adult testis, via the PI3K-Akt signalling pathway [34,35]. The roles of the various ETS-related TFs in spermatogonial cells have not been clarified, however recently published data found ETV4 in the stem-cell enriched fraction of the spermatogonial population, particularly during the spermatogonial stem cell pool establishment immediately after birth [7,36]. Motif enrichment analysis on the regions with decreased chromatin accessibility situated in the gene body and intergenic regions revealed that TFs are important in numerous developmental processes (FOXC1, FOXJ2, FOXM1, LHX6) were specifically enriched in intergenic regions of decreased chromatin accessibility (Fig. 4D). This is consistent with the association of intergenic regions with embryonic development-related GO terms that we previously detected (Fig. 1C). Our findings provide a detailed characterization of the enriched TF motifs present at the regions of dynamic accessibility between PND15 and adult

spermatogonial populations and point towards novel candidate regulators of the differential transcriptome of the pup and adult mouse spermatogonia.

Chromatin accessibility at transposable elements undergoes significant remodelling in the transition from postnatal to adult spermatogonia

Transposable elements (TEs) are under tight control in the germline, through coordinated epigenetic mechanisms involving DNA methylation, chromatin silencing and PIWI proteins – piRNA pathway [37,38]. Recent evidence suggests an important role for long terminal repeat (LTR) - type elements, specifically for ERVKs, the youngest class of endogenous retroviruses (ERVs), in the transcriptional regulation of mRNAs and long non-coding RNAs (lncRNAs) during mitosis-to-meiosis transition [39,40]. Furthermore, accessibility analysis at LTRs in mitotic and meiotic germ cells, revealed a unique chromatin accessibility landscape in spermatogonial cells, compared to the rest of the germ cells in the testis [39]. To explore potential differences in TEs regulation driven by postnatal age, we compared the accessibility of TEs in PND15 and adult spermatogonia. For this purpose, we quantified the ATAC-seq reads overlapping TEs defined by UCSC RepeatMasker and performed differential accessibility analysis at the subtype level (see Methods section). Our results revealed that the transition from PND15 to the adult stage is accompanied by significant chromatin accessibility differences at 135 TE subtypes (Fig. 5A and 5B and Table S5). Although most of the differentially accessible TE subtypes displayed a decrease in chromatin accessibility between PND15 and adult stages (68,9%, 93/135) (Fig. 5A), we also observed 42 TE subtypes which increased in accessibility in adult spermatogonia (Fig. 5B). Of note, more accessible subtypes also displayed an increased expression in adult spermatogonia compared to the early postnatal stage (Fig. 5B). TE loci within the subtypes harbouring changes in chromatin accessibility were situated in intergenic and intronic regions (68% intergenic and 25% intronic), and around 6% were located in the proximity of a gene (+/- 1kb from a TSS) (Fig. 5C). LTRs were the most abundant TEs to display changes in chromatin accessibility, specifically ERVK and ERV1 subtypes (Fig. 5A and 5B). Exemplary ERVK subtypes harbouring less accessible chromatin included RLTR17, RLTR9A3, RLTR12B and RMER17B (Table S5). Enrichment of RLTR17 and RLTR9 repeats

have been reported previously in mESCs, specifically at TFs important for pluripotency maintenance such as *Oct4* and *Nanog* [41]. Interestingly, we identified the promoter region of the lncRNA *Lncenc1*, an important regulator of pluripotency in mESCs [41,42], harboring several LTR loci with decreased accessibility in our adult spermatogonia, with the RLTR17 locus falling within the TSS of *Lncenc1*. This decrease in accessibility correlated with a marked decrease in expression of *Lncenc1* in adult spermatogonia (Fig. 5D). *Lncenc1* (also known as *Platr18*) is part of the pluripotency-associated transcript (*Platr*) family of lncRNAs which were recently identified as potential regulators of the pluripotency-associated genes *Oct4*, *Nanog* and *Zfp42* in mESCs [43–45]. We were also able to identify several other *Platr* genes, such as *Platr27* and *Platr14*, for which the TSS overlapped LTRs with reduced accessibility, RLTR17 and RLTR16B_MM, respectively (Fig. 5D and Table S5). These 2 pluripotency-associated transcripts also showed a decrease in mRNA expression in adult spermatogonia, while their expression was unchanged between PND8 and PND15 (Fig. 5D and Table S5). The remaining LTR subtypes with decreased accessibility in adult spermatogonia belonged to the ERV1, ERVL and MaLR families (Fig. 5A). Only very few other non-LTR TEs showed a decrease in chromatin accessibility, with 7 DNA element subtypes, 2 Satellite subtypes and 1 LINE subtype, respectively (Fig. 5A and Table S5). Emerging evidence suggests an important contribution of TEs in providing tissue-specific substrates for TF binding [41,46,47]. To investigate the regulatory potential of the less accessible LTR subtypes, we assessed the enrichment of TF motifs in these regions using HOMER. To do so, we focused on the family level and grouped together all LTR subtypes coming from one family (EVK, ERV1, ERVL and ERVL-MaLR families). Among the less accessible LTR families, ERVKs showed the highest number of enriched TF motifs in adult spermatogonial cells. Top hits included TFs with known regulatory roles in cell proliferation and differentiation such as FOXL1 and FOXQ1, stem cell maintenance factors ELF1, EBF1 and THAP11 and TFs important in spermatogenesis PBX3, ZNF143 and NFYA/B (Fig. 5E and S6A). ERVLs displayed motif enrichment for very few TFs, among which the previously undescribed ETV2, newly reported spermatogonial stem cell factor ZBTB7A and the testis-specific CTCF paralog CTCFL (Fig. 5E) [8].

Among the TE subtypes which increased in chromatin accessibility, members of the ERVK, ERVL and ERV1 families were predominant (57,1%, 24/42) (Fig. 5B). Interestingly, we also found a considerable number of LINE L1 subtypes with increased chromatin accessibility in adult spermatogonial cells (Fig. 5B). When parsing the data for more accessible loci within the L1 subtypes, we found several L1 loci situated less than +/- 5 kb from the TSS of numerous olfactory (*Olfr*) genes.Most of them were located in *Olfr* gene clusters on chromosomes 2, 7 and 11 (Table S5). Furthermore, the increase in accessibility of the L1 loci correlated with an increase in mRNA expression of the nearby *Olfr* gene in adult spermatogonial cells (Fig. 6A). Representative examples were *Olfr362* and *Olfr1307*, both situated in the *Olfr* gene cluster on Chr2 (Fig. 6B). Interestingly, when visualizing the data in IGV, we also observed that the *Olfr* gene cluster on chromosome 2 exhibited a higher density of L1 loci compared to neighbouring regions (Fig. S6B). Similar to before, we performed TF motif enrichment analysis at the family level by grouping together all differentially accessible TE subtypes coming from one family. More accessible LINE L1s were highly enriched in TF motifs, particularly in multiple members of the ETS, E2F and FOX families (Fig. 6C). The most significant motifs belonged to spermatogonial stem cell maintenance and stem cell potential regulators FOXO1 and ZEB1, as well as TFs which have been recently associated with active enhancers of the stem cell-enriched population of spermatogonia such as ZBTB17 and KLF5 (Fig. 6C and S6C) [36]. More accessible ERV1s also displayed enrichment of several TF binding sites, including spermatogenesis-related TFs (PBX3, PRDM1, NFYA/B), hypoxia-inducible HIF1A and cytokine regulators STAT5A/B, suggestive of different spermatogonial cell metabolic demands between early postnatal and adult stage (Fig. 6C and S6C). Overall, we provide an extensive characterization of the chromatin accessibility landscape of TEs in PND15 and adult spermatogonia, reveal differences in accessibility and TF motif landscape at distinct subtypes of TEs between these 2-time points, and suggest potential gene programs that may be regulated by these changes.

Discussion

As initiators of the spermatogenic cascade, spermatogonial cells are essential in germ cell proliferation and differentiation throughout postnatal life. Although recent studies employing bulk and scRNA-seq have revealed distinct transcriptional signatures of spermatogonial cells across postnatal life, very few have focused on describing the underlying landscape of open chromatin, and the extent to which it can contribute to the gene expression dynamics [5,6,9].

Our ATAC-seq revealed a reorganization of open chromatin in adult spermatogonia compared to the PND15 cell population. This striking reorganization of accessible chromatin, which we detected mainly in intergenic and intronic regions, may be indicative of novel regulatory regions governing at least in part the vast transcriptome changes that spermatogonial cells undergo in their transition from early postnatal to adult stage [5,9]. Indeed, we found that the regions of differential accessibility were associated with distinct gene pathways, with morphogenesis and developmental pathways associated to the regions of decreased chromatin accessibility, while regions of increased chromatin openness were enriched for DNA repair pathways, stem cell maintenance, RNA processing and protein metabolic processes. Furthermore, we identified numerous enriched TF motifs in the regions of differential accessibility, indicative of their potential regulatory role. AP-1 TFs with previously described roles in spermatogonial cell proliferation such as JUND and c-FOS, and pluripotency factors such as POU3F1, displayed enriched binding sites in the regions of increased chromatin accessibility. In contrast, FOX and ETS TF motifs, known regulators of developmental pathways, mainly mapped to regions that decreased in accessibility in adult spermatogonial cells. For some of the enriched TF motifs, a preference for certain genomic locations was evident: NF-YA and B binding sites exhibited enrichment specifically in intergenic regions of more accessible chromatin which interestingly, were also associated with spermatogenesis-related pathways. NF-YA/B is also localized in intergenic regions of open chromatin in humans spermatogonial cells [33], prompting an additional investigation of their roles in

regulating spermatogonial cell programs, with potential consequences for sperm formation.

Our comparison of the gene expression changes from PND8 to PND15 spermatogonial cells confirmed the dynamic transcriptome associated with developing spermatogonial cell states [5,9]. Notably, PND15 corresponds to the maturation of the Sertoli cell niche in the mouse testis [48,49]. Upregulation of pathways associated with signal transduction and cellular transport in PND15 spermatogonia suggests an increased cellular communication with the somatic niche, compared to the first week of postnatal development. We complemented these findings with literature RNA-seq data from PND14 and adult spermatogonia and confirmed further transcriptome changes in signalling pathways related to mitochondria, developmental processes and cell-to-cell signalling. Furthermore, we revealed numerous pathways related to cytokine signalling upregulated in adult spermatogonia, which point to the recently suggested role of testis resident macrophages in maintaining and regulating spermatogonial proliferation [50,51]. These findings suggest that, as the testis matures and the somatic niche develops, spermatogonial cells rely more on paracrine signalling and undergo vast changes in gene expression programs.

To obtain a comprehensive profile of the chromatin and the transcriptome differences between early postnatal and adult spermatogonial cells, we have integrated the chromatin accessibility and gene expression, with known histone H3 modifications and global DNAme patterns of THY1$^+$ spermatogonial cells from [4,5]. This allowed us to identify 4 distinct categories of differentially accessible chromatin regions for which the nearest gene was dynamically expressed between early postnatal and adult stages. Interestingly, we found a similar number of genes with increased chromatin accessibility around their TSS which exhibited either an up- or a downregulated expression between pup and adult spermatogonia. In the category of upregulated genes with increased nearby chromatin accessibility, we identified several factors associated with redox processes, mitochondria function and cell proliferation. In contrast, genes marked by an increase in chromatin accessibility and downregulated

expression in adult spermatogonia comprised factors important for cell cycle, RNA processing and developmental genes, suggesting that active repression is taking place at these genes [52]. For some of the developmental genes, more accessible chromatin was also marked by a bivalent H3K4me3/H3K27me3, indicative of a poised state. Notably, previous findings in $THY1^+$ adult spermatogonial cells and in sperm also revealed a poised state at promoters of developmental genes [4,53,54]. Therefore, our findings suggest that open chromatin reorganization may contribute to the poised status that is already established at the spermatogonial cell stage for certain developmental genes. We also identified a category of regions for which the decrease in chromatin accessibility correlated with a decreased expression, the category which also included developmental factors. The stable methylation patterns we detected at the differentially accessible chromatin regions, in both distal and proximal genomic regions, suggest a minimal impact for DNAme in regulating gene expression dynamics of spermatogonial cells across postnatal age.

Lastly, by investigating chromatin accessibility specifically at TEs, we revealed that distinct TE subtypes undergo changes in chromatin accessibility between PND15 and adult spermatogonial populations. ERVK and ERV1 subtypes were the most abundant categories of TEs to become less accessible in adult spermatogonia, whilst LINE L1 subtypes gained in accessibility. Although the majority of these TEs resided in intergenic and intronic regions, we were able to detect specific loci belonging to the differentially accessible ERVK and LINE L1 subtypes, which localized nearby TSS of distinct gene families. RLTR17, one of the LTR subtypes with decreased chromatin accessibility in adult spermatogonial cells, overlapped the TSS of several downregulated long-non coding RNAs from the *Platr* family. *Platr* genes, including the ones identified in our study, *Lncenc1* and *Platr14,* are LTR-associated long non-coding RNAs important for embryonic stem cell gene expression programs [45]. Interestingly, RLTR17 has also been previously linked to pluripotency maintenance. In mouse embryonic stem cells, RLTR17 is highly expressed and enriched in open chromatin regions and has been shown to provide binding sites for pluripotency factors Oct4 and Nanog [55]. On the basis of these findings, we suggest that RLTR17 chromatin organization may play a significant role

in regulating pluripotency programs between early postnatal and adult spermatogonial cells. In contrast to the decreased accessibility of LTRs, LINE L1 subtypes displayed an increase in chromatin accessibility in adult spermatogonial cells. Some of these L1 loci were situated in the vicinity of olfactory receptor genes with upregulated mRNA expression in adult spermatogonia. Recent findings in mouse and human embryonic stem cells have suggested a non-random genomic localization for L1 elements, specifically at genes that encode proteins with specialized functions [56]. Among these, the *Olfr* gene family was the most enriched in L1 elements [56]. Although their role in spermatogonial cells is currently not established, Olfr proteins have been implicated in the swimming behaviour of sperm [57,58]. Given their dynamic regulation across age in spermatogonial cells, we speculate that *Olfr* genes could play additional roles in spermatogenesis, other than in sperm physiology. This data together with the high number of enriched TF motifs identified at the differentially accessible ERVKs and LINE L1 elements underscores previously undescribed regulatory roles for chromatin organization of TEs in spermatogonial cells during the transition from developing to adult stages [46,47].

One limitation of our study is the incomplete purification achieved using FACS, which doesn't fully remove other testis cell types from our cell preparations. Therefore, we cannot entirely exclude the influence of contaminating cells on some of the transcriptome and chromatin accessibility data interpretation. Secondly, differences can also stem from the literature datasets which involve similar but not identically enriched populations of spermatogonial cells. Nevertheless, by comparing open chromatin landscape between developing and adult spermatogonial cells, our results reveal for the first time that there is an age-dependent dynamic reorganization of chromatin accessibility in spermatogonial cells. By integrating this newly generated data with gene expression profiles and known histone modifications, we provide novel insight into the chromatin - transcriptome dynamics of mouse spermatogonial cells between developing and adult stages and compile an information-rich resource for further germline studies.

Methods

Mouse husbandry

Male C57Bl/6J mice were purchased from Janvier Laboratories (France) and bred in-house to generate male mice used for experiments. All animals were kept on a reversed 12-h light/12-h dark cycle in a temperature- and humidity-controlled facility, with food (M/R Haltung Extrudat, Provimi Kliba SA, Switzerland) and water provided ad libitum. Cages were changed once weekly. Animals from 2 independent breedings were used for the experiments.

Germ cells isolation

Germ cells were isolated from male mice at postnatal day (PND) 8 or 15 for RNA-seq and ATAC-seq experiments, and adults at 20 weeks of age (PNW20) for ATAC-seq. Testicular single-cell suspensions were prepared as previously described with slight modifications [11,59]. For preparations using PND8 and PND15 pups, testes from 2 animals were pooled for each sample. Pup testes were collected in sterile HBSS on ice. Tunica albuginea was gently removed from each testis, making sure to keep the seminiferous tubules as intact as possible. Tubules were enzymatically digested in 0.25% trypsin-EDTA (ThermoFisher Scientific) and 7mg/ml DNase I (Sigma-Aldrich) solution for 5 min at 37°C. The suspension was vigorously pipetted up and down 10 times and incubated again for 3 min at 37°C. The digestion was stopped by adding 10% fetal bovine serum (ThermoFisher Scientific) and the cells were passed through a 20µm-pore-size cell strainer (Miltenyi Biotec) and pelleted by centrifugation at 600g for 7 min at 4°C. Cells were resuspended in PBS-S (PBS with 1% PBS, 10 mM HEPES, 1 mM pyruvate, 1mg/ml glucose, 50 units/ml penicillin and 50 µg/ml streptomycin) and used for sorting. For preparations from adult testis, one adult male was used for each sample. The tunica was removed and seminiferous tubules were digested in 2 steps. The first consisted of incubation in 1mg/ml collagenase type IV (Sigma-Aldrich) for 5 min at 37°C and vigorous swirling until the tubules were completely separated. Then tubules were placed on ice for 5 min to sediment, the supernatant was removed and washed with HBSS.

Washing/sedimentation steps were repeated 3 times and were necessary to remove interstitial cells. After the last washing step, sedimented tubule fragments were digested again with 0.25% trypsin-EDTA and 7mg/ml DNase I solution, and the digestion was stopped by adding 10% FBS. The resulting single-cell suspension was filtered through a 20µm strainer (Corning Life Sciences) and washed with HBSS. After centrifugation at 600g for 7 min at 4°C, the cells were resuspended in PBS-S, layered on a 30% Percoll solution (Sigma-Aldrich) and centrifuged at 600g for 8 min at 4°C without braking. The top 2 layers (HBSS and Percoll) were removed and the cell pellets were resuspended in PBS-S and used for sorting.

Spermatogonial cells enrichment by FACS

For pup testis, dissociated cells were stained with BV421-conjugated anti-β2M, biotin-conjugated anti-THY1 (53-2.1), and PE-conjugated anti-αv-integrin (RMV-7) antibodies. THY1 was detected by staining with Alexa Fluor 488-Sav. For adult testes, cells were stained with anti-α6-integrin (CD49f; GoH3), BV421-conjugated anti-β2 microglobulin (β2M; S19.8), and R-phycoerythrin (PE)-conjugated anti-THY1 (CD90.2; 30H-12) antibodies. α6-Integrin was detected by Alexa Fluor 488-SAv after staining with biotin-conjugated rat anti-mouse IgG1/2a (G28-5) antibody. Prior to FACS, 1 µg/ml propidium iodide (Sigma) was added to the cell suspensions to discriminate dead cells. All antibody incubations were performed in PBS-S for at least 30 min at 4°C followed by washing in PBS-S. Antibodies were obtained from BD Biosciences (San Jose, United States) unless otherwise stated. Cell sorting was performed at 4°C on a FACS Aria III 5L using an 85µm nozzle at the Cytometry Facility of the University of Zurich. For RNA-seq on PND8 and PND15 spermatogonia, cells were collected in 1.5 ml Eppendorf tubes in 500 µL PBS-S, immediately pelleted by centrifugation and snap-frozen in liquid N_2. Cell pellets were stored at -80°C until RNA extraction. For OmniATAC on PND15 spermatogonia, 25’000 cells were collected in a separate tube, pelleted by centrifugation and immediately processed using the OmniATAC library preparation protocol [10]. For OmniATAC on adult spermatogonia, 5000 cells from each animal were collected in a separate tube and further processed using the same protocol.

Immunocytochemistry

The protocol used for assessing spermatogonial cell enrichment after sorting was kindly provided by the Oatley Lab at Washington State University, Pullman, USA [60]. Briefly, 30,000-50,000 cells adhered to poly-L-Lysine coated coverslips (Corning Life Sciences) in 24-well plates for 1 h. Cells were fixed in freshly prepared 4% PFA for 10 min at room temperature then washed in PBS with 0.1% Triton X-100 (PBS-T). Non-specific antibody binding was blocked by incubation with 10% normal goat serum for 1 h at room temperature. Cells were incubated overnight at 4°C with mouse anti-PLZF (0.2 µg/ml, Active Motif, clone 2A9) primary antibody. Alexa488 goat anti-mouse IgG (1 µg/mL, ThermoFisher Scientific) was used for secondary labelling at 4°C for 1 h. Coverslips were washed 3x and mounted onto glass slides with VectaShield mounting medium containing DAPI (Vector Laboratories) and examined by fluorescence microscopy. Stem cell enrichment was determined by counting $PLZF^+$ cells in 10 random fields of view from each coverslip and dividing by the total number of cells present in the field of view (DAPI-stained nuclei).

RNA extraction and library preparation for RNA-seq

For RNA-seq on PND8 and PND15 spermatogonial cells, total RNA was extracted from sorted cells using AllPrep RNA/DNA Micro kit (Qiagen). RNA quality was assessed using a Bioanalyzer 2100 (Agilent Technologies). Samples were quantified using Qubit RNA HS Assay (ThermoFisher Scientific). 10 ng of total RNA from each sample were used to prepare total long RNA sequencing libraries using SMARTer® Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian (Takara Bio USA, Inc.) at the Functional Genomics Center Zurich (FGCZ) according to the manufacturer's instructions.

Library preparation for Omni-ATAC

Chromatin accessibility was profiled in PND15 and adult spermatogonial cells. Libraries were prepared starting from 25 000 PND15 and 5000 adults sorted spermatogonia, respectively [10]. Briefly, sorted cells were lysed in cold lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP40, 0.1% Tween-20, and

0.01% digitonin) and nuclei were pelleted and transposed using Nextera Tn5 (Illumina) for 30 min at 37°C in a thermomixer with shaking at 1000 rpm. Transposed fragments were purified using the MinElute Reaction Cleanup Kit (Qiagen). Following purification, libraries were generated by PCR amplification using the NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs), and purified using Agencourt AMPure XP magnetic beads (Beckman Coulter) to remove primer dimers (78bp) and fragments >1000bp. Library quality was assessed on an Agilent High Sensitivity DNA chip using the Bioanalyzer 2100 (Agilent Technologies).

RNA sequencing

Quality control and alignment: Single-end (SE) sequencing was performed using an Illumina HiSeq4000 at the FGCZ. PND8 raw data (FASTQ files) was merged from two individual runs. For analysis of published RNA-seq data (PND14 and PNW8 spermatogonia), FASTQ files were obtained using fastq-dump (version 2.10.8), quality-controlled using FastQC [61] (version 0.11.8). TrimGalore [62] (version 0.6.2) was used to trim adapters and low-quality ends from reads with Phred score less than 30 (-q 30), and for discarding trimmed reads shorter than 30 bp (--length 30). Trimmed reads were pseudo-aligned using Salmon [63] (version 0.9.1) with automatic detection of the library type (-l A), correcting for sequence-specific bias (--seqBias) and correcting for fragment GC bias correction (--gcBias) on a transcript index prepared for the Mouse genome (GRCm38) from GENCODE (version M18) [13], with additional piRNA precursors and transposable elements (concatenated by family) from Repeat Masker as in [64].

Downstream analysis: Analysis was conducted with R (R Core Team, 2020) (version 3.6.2) using packages from The Comprehensive R Archive Network (CRAN) (https://cran.r-project.org) and Bioconductor [65]. Pre-filtering of genes was performed using the filterByExpr function from edgeR [66] (version 3.28.1) with a design matrix and requiring at least 15 counts (min.counts = 15). Normalization factors were obtained using TMM normalization [67] from edgeR package and differential gene expression (DGE) analysis was performed using limma-voom [68] pipeline from limma

[69] (version 3.42.2). Log_2 fold change between samples was calculated by subtracting Log_2 normalized expression values. Gene ontology (GO) analysis was performed on expressed genes with fGSEA (version 1.15.2) using fGSEAMultilevel function on sets with 10 to 1000 annotated genes (minSize = 10, maxSize = 1000), and p-values boundary of 1E-100 (eps = 1e-100) [70]. For PND8 versus PND15 comparison, genes were pre-ranked using t-statistic; for PND14 versus PNW8 comparison, Log_2 fold change was used due to the lack of multiple replicates for each time point. REVIGO was used to summarize GO terms obtained following fGSEA [71].

Omni-ATAC

Quality control, alignment, and peak calling: Paired-end (PE) sequencing was performed on PND15 and adult spermatogonial cells samples on an Illumina HiSeq2500 platform (FGCZ). FASTQ files were assessed for quality using FastQC [61] (version 0.11.8). Quality control (QC) was performed using TrimGalore [62] (version 0.6.2) in PE mode (--paired), trimming adapters, low-quality ends (-q 30) and discarding reads < 30 bp after trimming (--length 30). Alignment on the GRCm38 genome was performed using Bowtie2 [72] (version 2.3.5) with the following parameters: fragments up to 2 kb were allowed to align (-X 2000), entire read alignment (--end-to-end), suppressing unpaired alignments for paired reads (--no-mixed), suppressing discordant alignments for paired reads (--no-discordant) and minimum acceptable alignment score with respect to the read length (--score-min L,-0.4,-0.4). Using alignmentSieve (version 3.3.1) from deepTools [73] (version 3.4.3), aligned data (BAM files) were adjusted for the read start sites to represent the centre of the transposon cutting event (--ATACshift), and filtered for reads with a high mapping quality (--minMappingQuality 30). Reads mapping to the mitochondrial chromosome and ENCODE blacklisted regions were filtered out. To call nucleosome-free regions, all aligned files were merged within groups (PND15 and adult), sorted, and indexed using SAMtools [74] (version 0.1.19), and nucleosome-free fragments (NFFs) were obtained by selecting alignments with a template length between 40 and 140 inclusively. Peak calling (identifying areas in a genome that have been enriched for transcription factors) on the NFFs was

performed using MACS2 [75] (version 2.2.7.1) with mouse genome size (-g 2744254612) and PE BAM file format (-f BAMPE).

Differential accessibility analysis: The downstream analysis was performed in R(version 3.6.2), using packages from CRAN (https://cran.r-project.org) and Bioconductor [65]. The peaks were annotated based on overlap with GENCODE (version M18) [13] transcript, and/or the distance to the nearest transcription start site (available at the following link: https://github.com/mansuylab/SC_postnatal_adult/bin/annoPeaks.R).The number of extended reads overlapping in the peak regions was calculated using the csaw package [76] (version 1.20.0). Peak regions that did not have at least 15 reads in at least 40% of the samples were filtered out. Normalization factors were obtained on the filtered peak regions using the TMM normalization method [67] and differential analysis on the peaks (adults vs PND15) was performed using the Genewise Negative Binomial Generalized Linear Models with Quasi-likelihood (glmQLFit) Tests from the edgeR package [66] (version 3.28.1). Peak regions that had an absolute Log_2 fold change ≥ 1 and an FDR ≤ 0.05 were categorized as differentially accessible regions. GO analysis was performed on DARs with the rGREAT package [77] (version 1.18.0), which is a wrapper around the GREAT tool [78] (version 4.0). Transcription factor motif enrichment analysis was performed using the marge package [79] (version 0.0.4.9999), which is a wrapper around the Homer tool [23] (version 4.11.1).

Differential accessibility analysis at transposable elements: TE gene transfer format (GTF) file was obtained from http://labshare.cshl.edu/shares/mhammelllab/www-data/TEtranscripts/TE_GTF/mm10_rmsk_TE.gtf.gz_on_03.02.2020. The GTF file provides hierarchical information about TEs: **Class** (level 1, eg. LTR), **Family** (level 2, eg. LTR L1), **Subtype** (level 3, eg. LTR L1 L1_Rod), and **Locus** (level 4, eg. LTR L1 L1_Rod L1_Rod_dup1). TE loci were annotated based on overlap with GENCODE (version M18) as described above for ATAC-seq peaks. Filtered BAM files (without reads mapping to blacklisted or mitochondrial regions) were used for analyzing TEs. Mapped reads were assigned to TEs using featureCounts from the R package Rsubread [80] (version 2.0.1) and were summarized to Subtypes (level 3), allowing for

multi-overlap with fractional counts, while ignoring duplicates. The number of extended reads overlapping at the TE loci were obtained using the csaw package [76] (version 1.20.0). Subtypes which did not have at least 15 reads, and loci that did not have at least 5 reads in at least 40% of the samples, were filtered out. Normalization and differential accessibility analysis were performed as described above. Subtypes that had an absolute Log_2fold change ≥ 0.5 and an FDR ≤ 0.05 were categorized as differentially accessible subtypes and the loci with an absolute Log_2 fold change ≥ 1 and an FDR ≤ 0.05 were categorized as differentially accessible loci. For further downstream data analysis, only the differentially accessible loci of differentially accessible subtypes were considered. GO and motif enrichment analysis were performed as described above.

Chromatin immunoprecipitation sequencing

Quality control, alignment, and peak calling: ChIP-Seq SE data for PNDW8 (adults) were obtained from GEO accession GSE49621 [4]. FASTQ files were obtained using fastq-dump (version 2.10.8), and different runs were merged. The FASTQ files were assessed for quality using FastQC [61] (version 0.11.8). Quality control (QC) was performed using TrimGalore [61] (version 0.6.0), trimming adapters, low-quality ends (-q 30) and discarding trimmed reads shorter than 30 bp (--length 30). Alignment to the GRCm38 genome was performed using Bowtie2 [72] (version 2.3.5). Reads with more than 3 mismatches were removed from the aligned data, as suggested in [81], and reads with low mapping quality (--minMappingQuality 30) or mapping to the mitochondrial chromosome or aforementioned blacklisted regions were filtered out. Peak calling was performed using MACS2 [75] (version 2.2.7.1) with mouse genome size (-g 2744254612) and SE BAM file format (-f BAM).

Bisulfite sequencing (BS)

Quality control and alignment: BS paired-end data for PND7, PND14, and PNW8 (adults) were obtained from GEO accession GSE49623 [5]. FASTQ files were obtained using fastq-dump (version 2.10.8), and different runs were merged. FASTQ files were assessed for quality using FastQC [61] (version 0.11.8). QC was performed using TrimGalore [62] (version 0.6.4_dev) in PE mode (--paired), trimming adapters,

low-quality ends (-q 30) and discarding trimmed reads shorter than 30 bp (--length 30). Alignment of the QC data was performed using Bismark [82] (version 0.22.3) on a GRCm38 index built using bismark_genome_preparation (version 0.17.0). Methylation information for individual cytosines was extracted using the bismark_methylation_extractor tool from the Bismark package (version 0.22.3).

High-throughput sequencing data analysis

Data availability: the datasets used in this study are available from the following GEO accessions: **GSE_____**, GSE49621, GSE49622, GSE62355, and GSE49623. An overview of the datasets included in the study is shown in the following table:

Source	*Seq	Stages (n)
GSE_____	RNA-seq	PND8 (8), PND15 (9)
GSM1525703	RNA-seq	PND14 (1)
GSM1415671	RNA-seq	PNW8 (1)
GSE_____	ATAC-seq	PND15 (6), PNW20 (5)
GSM1202705	ChIP-seq (H3K4me3)	PNW8 (1)
GSM1202708	ChIP-seq (H3K27me3)	PNW8 (1)
GSM1202713	ChIP-seq (H3K27ac)	PNW8 (1)
GSM1202723	ChIP-seq (Input)	PNW8 (1)
GSE49623	BS-seq	PND7 (1), PND14 (1), PNW8 (7)

Figures

All figures in this study were generated using ggplot2 [83], EnrichedHeatmap [84], and ComplexHeatmap [85] packages, and using base plotting functions in R. Genomic tracks were generated in IGV and colour coded in Inkscape.

Authors contribution

ILC and IMM conceived and designed the study. ILC prepared samples, performed all RNA-seq, ICC and ATAC-seq experiments. DKT analyzed RNA-seq, ATAC-seq, ChIP-seq and BS data, with significant support from PLG. ILC and DKT prepared figures. ILC interpreted the data with significant input from DKT, PLG and IMM. ILC wrote the manuscript with significant help from DKT, PLG and IMM. All authors read and accepted the final version of the manuscript.

Acknowledgements

We thank Francesca Manuella and Martin Roszkowski for taking care of animal breeding, Yvonne Zipfel for animal care, Andrew McDonald for assistance with basic lab needs, Silvia Schelbert and Alberto Corcoba for taking care of the animal licenses and lab organization.We thank Rodrigo Arzate for conceptual support and critical reading of the manuscript. We thank Catherine Aquino and Emilio Yángüez from the Functional Genomics Center Zurich for support and advice with library preparation and sequencing. We are very grateful to Jon Oatley, Melissa Oatley, Tessa Lord and Nathan Law for conceptual advice, hands-on training, and for providing detailed protocols for testis dissection and preparation, and immunocytochemistry of spermatogonial cells. We thank Zuguang Gu for support with heatmaps. We thank Service and Support for Science IT (www.s3it.uzh.ch) for computational infrastructure.

Competing interest

The authors declare that they have no competing interests.

Funding

The lab is funded by the University Zurich, ETH Zurich, the Swiss National Science Foundation Grant No. 31003A_175742 / 1, ETH fellowship ETH-25 19-2. DKT is supported by a Swiss Government Excellence Scholarship.

Data and materials availability

Repository accession numbers will be available at publication or by request to the corresponding author. The code employed for the data analysis is available from https://github.com/mansuylab/SC_postnatal_adult.

References

1. Oatley, J. M. & Griswold, M. D. *The biology of mammalian spermatogonia*. *The Biology of Mammalian Spermatogonia* (2017). doi:10.1007/978-1-4939-7505-1.

2. De Rooij, D. G. The nature and dynamics of spermatogonial stem cells. *Development (Cambridge)* vol. 144 3022–3030 (2017).

3. Kubota, H. & Brinster, R. L. Spermatogonial stem cells. *Biology of Reproduction* vol. 99 52–74 (2018).

4. Hammoud, S. S. *et al.* Chromatin and Transcription Transitions of Mammalian Adult Germline Stem Cells and Spermatogenesis. *Cell Stem Cell* **15**, 239–253 (2014).

5. Hammoud, S. S. *et al.* Transcription and imprinting dynamics in developing postnatal male germline stem cells. *Genes & development* **29**, 2312–24 (2015).

6. Hermann, B. P. *et al.* The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. *Cell Reports* **25**, 1650-1667.e8 (2018).

7. Law, N. C., Oatley, M. J. & Oatley, J. M. Developmental kinetics and transcriptome dynamics of stem cell specification in the spermatogenic lineage. *Nature Communications* **10**, 2787 (2019).

8. Green, C. D. *et al.* A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Developmental Cell* **46**, 651-667.e10 (2018).

9. Grive, K. J. *et al.* Dynamic transcriptome profiles within spermatogonial and spermatocyte populations during postnatal testis maturation revealed by single-cell sequencing. *PLoS Genetics* **15**, e1007810 (2019).

10. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature Methods* **14**, 959–962 (2017).

11. Kubota, H., Avarbock, M. R. & Brinster, R. L. Culture Conditions and Single Growth Factors Affect Fate Determination of Mouse Spermatogonial Stem Cells1. *Biology of Reproduction* **71**, 722–731 (2004).

12. Costoya, J. A. *et al.* Essential role of Plzf in maintenance of spermatogonial stem cells. *Nature Genetics* **36**, 653–659 (2004).

13. Harrow, J. *et al.* GENCODE: The reference human genome annotation for the ENCODE project. *Genome Research* **22**, 1760–1774 (2012).

14. Myers, R. M. *et al.* A user's guide to the Encyclopedia of DNA elements (ENCODE). *PLoS Biology* **9**, e1001046 (2011).

15. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).

16. Kanatsu-Shinohara, M. *et al.* Myc/Mycn-mediated glycolysis enhances mouse spermatogonial stem cell self-renewal. *Genes and Development* **30**, 2637–2648 (2016).

17. Chen, W. *et al.* A bioenergetic shift is required for spermatogonial differentiation. *Cell Discovery* **6**, 1–17 (2020).

18. Barrios, F. *et al.* Opposing effects of retinoic acid and FGF9 on Nanos2 expression and meiotic entry of mouse germ cells. *Journal of Cell Science* **123**, 871–880 (2010).

19. Sada, A., Suzuki, A., Suzuki, H. & Saga, Y. The RNA-binding protein NANOS2 is required to maintain murine spermatogonia! Stem Cells. *Science* **325**, 1394–1398 (2009).

20. Hasegawa, K. & Saga, Y. FGF8-FGFR1 signaling acts as a niche factor for maintaining undifferentiated spermatogonia in the mouse. *Biology of Reproduction* **91**, 145–146 (2014).

21. Shavlakadze, T. *et al.* Age-Related Gene Expression Signature in Rats Demonstrate Early, Late, and Linear Transcriptional Changes from Multiple Tissues. *Cell Reports* **28**, 3263-3273.e3 (2019).

22. Fushan, A. A. *et al.* Gene expression defines natural changes in mammalian lifespan. *Aging Cell* **14**, 352–365 (2015).

23. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* **38**, 576–589 (2010).

24. Martínez-Zamudio, R. I. *et al.* AP-1 imprints a reversible transcriptional programme of senescent cells. *Nature Cell Biology* **22**, 842–855 (2020).

25. Shaulian, E. & Karin, M. AP-1 as a regulator of cell life and death. *Nature Cell Biology* vol. 4 E131–E136 (2002).

26. He, Z. *et al.* Gdnf Upregulates c-Fos Transcription via the Ras/Erk1/2 Pathway to Promote Mouse Spermatogonial Stem Cell Proliferation. *Stem Cells* **26**, 266–278 (2008).

27. Wang, M. *et al.* Single-Cell RNA Sequencing Analysis Reveals Sequential Cell Fate Transition during Human Spermatogenesis. *Cell Stem Cell* **23**, 599-614.e4 (2018).

28. Faisal, I. *et al.* Transcription Factor USF1 Is Required for Maintenance of Germline Stem Cells in Male Mice. *Endocrinology* **160**, 1119–1136 (2019).

29. Wu, X. *et al.* The POU Domain Transcription Factor POU3F1 Is an Important Intrinsic Regulator of GDNF-Induced Survival and Self-Renewal of Mouse Spermatogonial Stem Cells1. *Biology of Reproduction* **82**, 1103–1111 (2010).

30. Niu, Z. *et al.* MicroRNA-21 regulates the self-renewal of mouse spermatogonial stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 12740–12745 (2011).

31. Lord, T., Oatley, M. J. & Oatley, J. M. Testicular Architecture Is Critical for Mediation of Retinoic Acid Responsiveness by Undifferentiated Spermatogonial Subtypes in the Mouse. *Stem Cell Reports* **10**, 538–552 (2018).

32. Oldfield, A. J. *et al.* Histone-Fold Domain Protein NF-Y Promotes Chromatin Accessibility for Cell Type-Specific Master Transcription Factors. *Molecular Cell* **55**, 708–722 (2014).

33. Guo, J. *et al.* Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling and Metabolic Transitions during Human Spermatogonial Stem Cell Development. **21**, 533–546 (2017).

34. Goertz, M. J., Wu, Z., Gallardo, T. D., Hamra, F. K. & Castrillon, D. H. Foxo1 is required in mouse spermatogonial stem cells for their maintenance and the initiation of spermatogenesis. *Journal of Clinical Investigation* **121**, 3456–3466 (2011).

35. Chan, F. *et al.* Functional and molecular features of the Id4+ germline stem cell population in mouse testes. *Genes & development* **28**, 1351–62 (2014).

36. Cheng, K. *et al.* Unique Epigenetic Programming Distinguishes Regenerative Spermatogonial Stem Cells in the Developing Mouse Testis. *iScience* **23**, 101596 (2020).

37. Thompson, P. J., Macfarlan, T. S. & Lorincz, M. C. Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. *Molecular Cell* vol. 62 766–776 (2016).

38. Deniz, Ö., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nature Reviews Genetics* vol. 20 417–431 (2019).

39. Sakashita, A. *et al.* Endogenous retroviruses drive species-specific germline transcriptomes in mammals. *bioRxiv* 2020.03.11.987230 (2020) doi:10.1101/2020.03.11.987230.

40. Davis, M. P. *et al.* Transposon-driven transcription is a conserved feature of vertebrate spermatogenesis and transcript evolution. *EMBO reports* **18**, 1231–1247 (2017).

41. Fort, A. *et al.* Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nature Genetics* **46**, 558–566 (2014).

42. Sun, Z. *et al.* The Long Noncoding RNA Lncenc1 Maintains Naive States of Mouse ESCs by Promoting the Glycolysis Pathway. *Stem Cell Reports* **11**, 741–755 (2018).

43. Wu, F. *et al.* Long non-coding RNAs potentially function synergistically in the cellular reprogramming of SCNT embryos. *BMC Genomics* **19**, 631 (2018).

44. Dann, C. T. *et al.* Spermatogonial Stem Cell Self-Renewal Requires OCT4, a Factor Downregulated During Retinoic Acid-Induced Differentiation. *Stem Cells* **26**, 2928–2937 (2008).

45. Bergmann, J. H. *et al.* Regulation of the ESC transcriptome by nuclear long noncoding RNAs. *Genome Research* **25**, 1336–1346 (2015).

46. Sundaram, V. & Wysocka, J. Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 375 (2020).

47. Sundaram, V. *et al.* Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Research* **24**, 1963–1976 (2014).

48. Shinohara, T., Orwig, K. E., Avarbock, M. R. & Brinster, R. L. Remodeling of the postnatal mouse testis is accompanied by dramatic changes in stem cell number and niche accessibility. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 6186–6191 (2001).

49. Flickinger, C. J. The postnatal development of the Sertoli cells of the mouse. *Zeitschrift für Zellforschung und Mikroskopische Anatomie* **78**, 92–113 (1967).

50. DeFalco, T. *et al.* Macrophages Contribute to the Spermatogonial Niche in the Adult Testis. *Cell Reports* **12**, 1107–1119 (2015).

51. Garbuzov, A. *et al.* Purification of GFRα1+ and GFRα1– Spermatogonial Stem Cells Reveals a Niche-Dependent Mechanism for Fate Determination. *Stem Cell Reports* **10**, 553–567 (2018).

52. Starks, R. R., Biswas, A., Jain, A. & Tuteja, G. Combined analysis of dissimilar promoter accessibility and gene expression profiles identifies tissue-specific genes and actively repressed networks. *Epigenetics and Chromatin* **12**, 16 (2019).

53. Erkek, S. *et al.* Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nature Structural and Molecular Biology* **20**, 868–875 (2013).

54. Jung, Y. H. *et al.* Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell Reports* **18**, 1366–1382 (2017).

55. Fort, A. *et al.* Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nature Genetics* **46**, 558–566 (2014).

56. Lu, J. Y. *et al.* Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Reports* **30**, 3296-3311.e5 (2020).

57. Fukuda, N. & Touhara, K. Developmental expression patterns of testicular olfactory receptor genes during mouse spermatogenesis. *Genes to Cells* **11**, 71–81 (2005).

58. Vanderhaeghen, P., Schurmans, S., Vassart, G. & Parmentier, M. Specific repertoire of olfactory receptor genes in the male germ cells of several mammalian species. *Genomics* **39**, 239–246 (1997).

59. Kubota, H., Avarbock, M. R. & Brinster, R. L. Growth factors essential for self-renewal and expansion of mouse spermatogonial stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 16489–94 (2004).

60. Yang, Q.-E., Racicot, K. E., Kaucher, A. V, Oatley, M. J. & Oatley, J. M. MicroRNAs 221 and 222 regulate the undifferentiated state in mammalian male germ cells. *Development (Cambridge, England)* **140**, 280–90 (2013).

61. Andrews, S. *et al.* FastQC. A quality control tool for high throughput sequence data. (2012).

62. Krueger, F. Trim Galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, www.bioinformatics.babraham.ac.uk/projects/trim_galore/. (2015).

63. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**, 417–419 (2017).

64. Gapp, K. *et al.* Alterations in sperm long RNA contribute to the epigenetic inheritance of the effects of postnatal trauma. *Molecular Psychiatry* **25**, 2162–2174 (2018).

65. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods* **12**, 115–121 (2015).

66. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).

67. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**, R25 (2010).

68. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* **15**, R29 (2014).

69. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47--e47 (2015).

70. Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. *bioRxiv* 060012 (2016) doi:10.1101/060012.

71. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* **6**, e21800 (2011).

72. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).

73. Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**, W160--W165 (2016).

74. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

75. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).

76. Lun, A. T. L. & Smyth, G. K. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Research* **44**, e45 (2015).

77. Zuguang, G. No Title. *https://github.com/jokergoo/rGREAT* (2020).

78. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology* **28**, 495–501 (2010).

79. Amezquita, R. A. marge: An API for Analysis of Motifs Using HOMER in R. (2018) doi:10.1101/249268.

80. Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research* **47**, e47–e47 (2019).

81. Royo, H., Stadler, M. & Peters, A. Alternative Computational Analysis Shows No Evidence for Nucleosome Enrichment at Repetitive Sequences in Mammalian Spermatozoa. *Developmental Cell* **37**, 98–104 (2016).

82. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).

83. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

84. Gu, Z., Eils, R., Schlesner, M. & Ishaque, N. EnrichedHeatmap: An R/Bioconductor package for comprehensive visualization of genomic signal associations. *BMC Genomics* **19**, 234 (2018).

85. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).