

## 2 Postnatal characterization of mouse spermatogonial cells reveals distinct chromatin regulatory landscapes in the developing and adult testis

Irina Lazar-Contes<sup>1,2,\*</sup>, Deepak K. Tanwar<sup>1,2,3,\*</sup>, Pierre-Luc Germain<sup>2,3</sup>, Isabelle M. Mansuy<sup>1,2,#</sup>

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

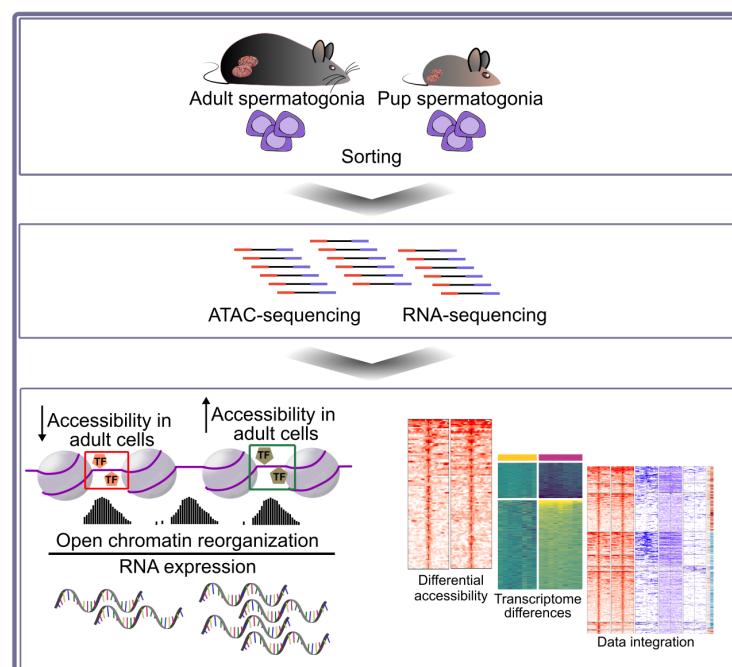
\*Authors share equal contribution

#Correspondence: [mansuy@hifo.uzh.ch](mailto:mansuy@hifo.uzh.ch)

Available on *BioRxiv*, doi: <https://doi.org/10.1101/2020.08.20.259374>

In preparation for *Developmental Cell*

### Graphical Abstract



### Highlights

Open chromatin reorganization in spermatogonial cells between early postnatal and adult stage

Data integration reveals distinct chromatin landscapes at genes important for spermatogonial cell maintenance and proliferation

Regions of differential accessibility are enriched in numerous transcription factor binding sites

Less accessible LTR-ERVKs and more accessible LINE L1s associate with distinct gene families in adult spermatogonial cells

## 2.1 Abstract

In mammals, spermatogonial cells are undifferentiated male germ cells that exit quiescence after birth, and self-renew or differentiate to produce spermatogenic cells and functional sperm across life. Recent rodent studies showed that spermatogonial cells have a highly dynamic transcriptome between early postnatal life and adulthood. However, what drives this developmental transition at the chromatin level is not fully understood. We characterized chromatin accessibility in early postnatal and adult spermatogonial cells in mice using ATAC-seq, and integrated the data with a range of transcriptomic and epigenomic features across age. We show that extensive chromatin remodelling occurs in spermatogonial cells across postnatal age, and that this correlates with distinct biological pathways and transcription factor (TF) motif enrichment. We further identify genomic regions with significantly different chromatin accessibility that are marked by distinct histone modifications, and are situated in proximity of transcription start sites (TSSs) of genes important for spermatogonial cell maintenance and proliferation. Some of the regions with increased accessibility correspond to transposable elements (TEs) enriched in multiple TF motifs, and with increased RNA expression in adult spermatogonia. Taken together, our results underscore the dynamic nature of open chromatin in spermatogonial cells across postnatal life, and reveal novel profiles of chromatin organization, histone modifications and gene expression between developing and adult spermatogonia.

## 2.2 Introduction

Spermatogonial cells are the initiators and supporting foundation of spermatogenesis in various species, including mammals. 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 subpopulation of spermatogonial cells proliferates and gives rise to undifferentiated  $A_{\text{single}}$  ( $A_s$ ),  $A_{\text{paired}}$  ( $A_{\text{pr}}$ ) and  $A_{\text{aligned}}$  ( $A_{\text{al}}$ ) 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 at PND 42-48 (Kubota and Brinster, 2018; Oatley and Griswold, 2017; De Rooij, 2017).

Recent work showed that distinct transcriptional profiles characterize spermatogonial cells in early postnatal life (Green et al., 2018; Hammoud et al., 2014, 2015; Hermann et al., 2018; Law et al., 2019). During the first week of postnatal development, spermatogonia display unique features necessary for the rapid establishment and expansion of the cell population along the basement membrane. These include high expression of genes involved in cell cycle regulation, stem cell proliferation, transcription and RNA processing (Grive et al., 2019). In comparison, in the adult testis, the focus lies in the maintenance of a steady cell population, which balances proliferation and differentiation capabilities to ensure sperm formation across life. Previous reports have revealed that adult spermatogonial cells prioritize pathways related to paracrine signalling and niche communication, as well as mitochondrial function and oxidative phosphorylation (Grive et al., 2019; Hermann et al., 2018).

Concomitant with gene expression changes, histone tail modifications and DNA methylation differences in spermatogonial cells from distinct postnatal stages have also been described (Hammoud et al., 2014, 2015). However, little is known about the accessible chromatin landscape of spermatogonial cells during the transition from early postnatal to adult stage, and how it could facilitate these dynamic transcriptional changes.

To investigate open chromatin reorganization in the transition from early postnatal to adult spermatogonia, we employed the Omni-ATAC protocol, an improved version of the Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) (Corces et al., 2017). Our results revealed extensive chromatin remodelling, in particular an increase in accessibility, of open chromatin regions in adult spermatogonia compared to PND15 cells. We further characterized novel histone signatures at regions of differentially accessible chromatin associated with genes dynamically expressed across postnatal age, by integrating published chromatin immunoprecipitation sequencing (ChIP-seq), bisulphite sequencing (BS) and transcriptome data. Lastly, by investigating chromatin accessibility at transposable elements (TEs), we described previously uncharacterized changes in accessibility at LTR and LINE L1 subtypes between developing and adult spermatogonial cells. Taken together our results suggest an important contribution of open chromatin reorganization to the diverse transcriptome of developing and adult spermatogonial cells.

## 2.3 Results

### 2.3.1 Fluorescence activated cell sorting (FACS) enriches spermatogonial cells from developing and adult mouse testis

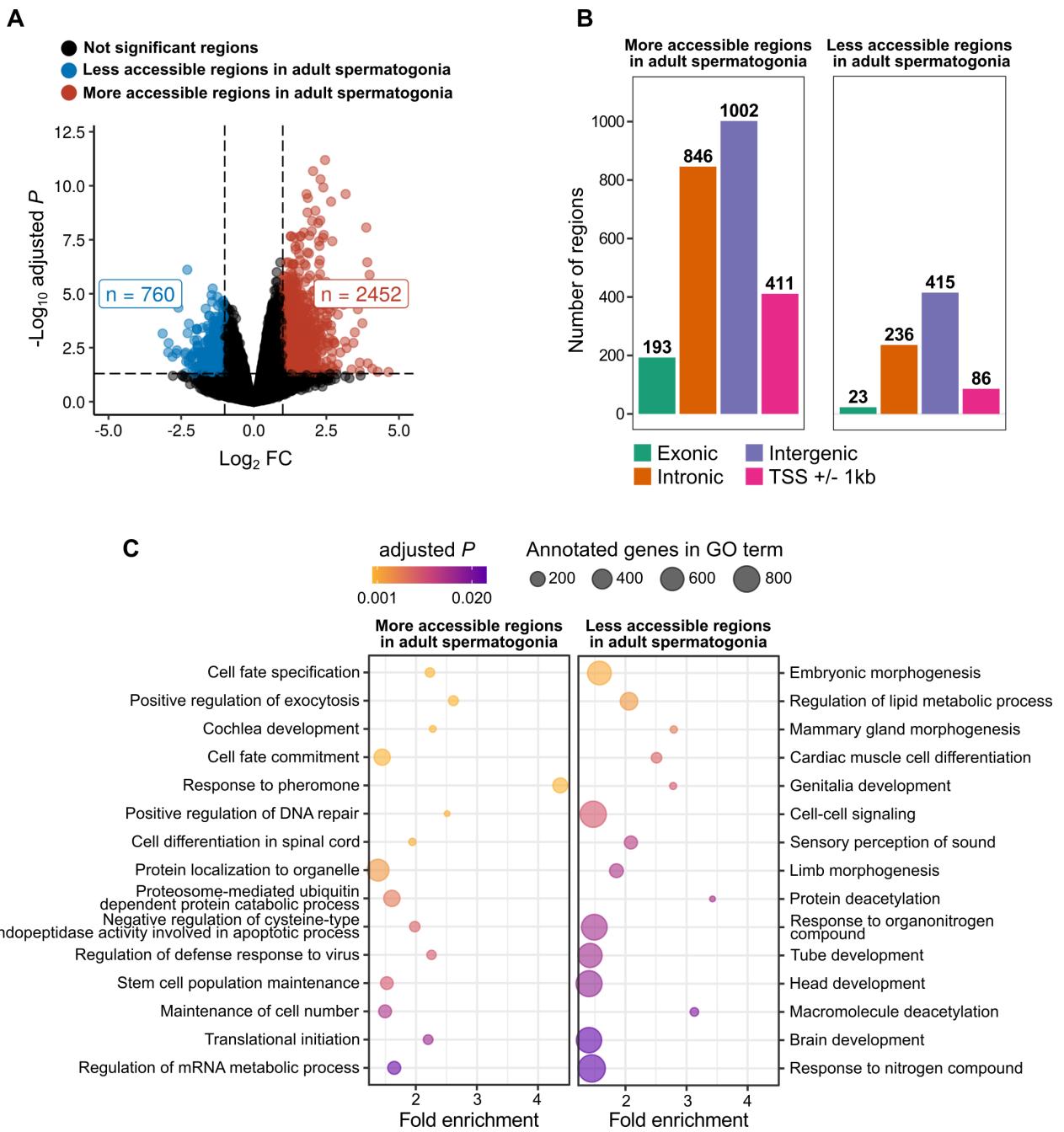
We collected testes from 8- and 15-days old pups (PND8 and 15), and from adult males at postnatal week (PNW) 20, and prepared cell suspensions by enzymatic digestion of the testes. To achieve high enrichment of spermatogonial cells in our cell populations, we employed fluorescence-activated cell sorting (FACS) using the surface phenotype established by Kubota et al. (Figure S 2-1A) (Kubota et al., 2004b). The enriched spermatogonial populations were further used for Omni-ATAC ( $n = 6$  PND15 and  $n = 5$  adult samples) and RNA-seq ( $n = 9$  PND8 and  $n = 8$  PND15 samples) approaches. To evaluate the purity of our sorted samples, we performed immunocytochemistry using PLZF, a well-established marker of undifferentiated spermatogonia (Costoya et al., 2004). This revealed that our FACS strategy generates cell populations that are 85-95% PLZF<sup>+</sup>, compared to only 3-6% PLZF<sup>+</sup> cells identified in our samples before FACS (Figure S 2-1B). Using our newly generated RNA-seq data, we also examined the expression of known pluripotency, spermatogonial stem cell and somatic markers (Leydig and Sertoli cells). We observed that cells in our sorted preparations expressed high level of stem cell and undifferentiated spermatogonial markers, but low level of somatic cells markers, confirming spermatogonial cell enrichment (Figure S 2-1C).

### 2.3.2 Accessible chromatin is reorganised in adult spermatogonia compared to early postnatal stage

Previous RNA-seq analyses in rodents showed that spermatogonial cell transcriptome changes during postnatal life, with distinct transcriptional programs revealed in neonatal, juvenile and adult stages (Grive et al., 2019; Hammoud et al., 2015). To determine if changes in chromatin organization underlie the dynamic transcriptome of spermatogonial cells as mice age, we profiled chromatin accessibility of spermatogonial cells at PND15, and from adult testis. We relied on the Omni-ATAC protocol, which shows a higher signal-to-noise ratio compared to the original ATAC-seq, and allows a lower number of starting cells (Corces et al., 2017). Accessible regions were identified by peak-calling on the merged nucleosome-free fragments (NFF) from all PND15 and adult samples. Following the removal of lowly enriched

regions, we included 158,978 regions in our downstream analyses (see Methods section for details). Most of the Tn5-accessible regions were intergenic (38%), located in gene bodies (33%) or in proximity of a gene transcription starting site (TSS) (+/-1 kb from TSS, 28%) (Figure S 2-2A). Differential accessibility analysis revealed 3212 differentially accessible regions between PND15 and adult spermatogonia (FDR  $\leq$  0.05, abs Log<sub>2</sub> fold change (FC)  $\geq$  1), with the majority of the regions showing a gain in accessibility in adult stage (Figure 2-1A and Table S1). The regions of differential accessibility were predominantly localized in intergenic regions and introns (34% in introns and 45% in intergenic regions). Only 15% of all differentially accessible regions resided +/- 1kb from the TSS of a gene (Figure 2-1B).

To investigate the biological significance of the regions with differential accessibility, we performed Gene Ontology (GO) analysis using the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010). Overall, regions of increased accessibility in adult spermatogonia were enriched in pathways associated with cell fate and stem cell population maintenance, protein metabolism and RNA metabolic processes (Figure 2-1C and Table S1). Given the high number of more accessible regions situated at different genomic locations, we asked if this genomic distribution is associated with distinct biological functions. We found that regions located in gene bodies (mainly introns) were specifically enriched for GO terms related to reproduction and protein metabolism, whilst regions close to (+/- 1 kb), or overlapping TSSs of genes, were enriched for cell fate specification and tissue morphogenesis (Figure S 2-2B and Table S1). We observed that the regions of less accessibility in adult spermatogonial cells were predominantly localized in intergenic regions (Table S1). GO enrichment in less accessible chromatin regions revealed enrichment of pathways related to embryonic development (Figure 2-1C and Table S1).



**Figure 2-1 Regions of differential chromatin accessibility from PND15 to adult spermatogonial cells associate with distinct gene pathways.**

(A) Volcano plot of differentially accessible ATAC-seq regions (adjusted  $P \leq 0.05$  and absolute Log<sub>2</sub> FC  $\geq 1$ ) between adult and PND15 spermatogonial cells;

(B) Bar plot illustrating the genomic distribution of differentially accessible ATAC-seq regions between adult and PND15 spermatogonial cells. Genomic regions were categorized in intronic, exonic, intergenic and +/- 1kb from the TSS of a gene;

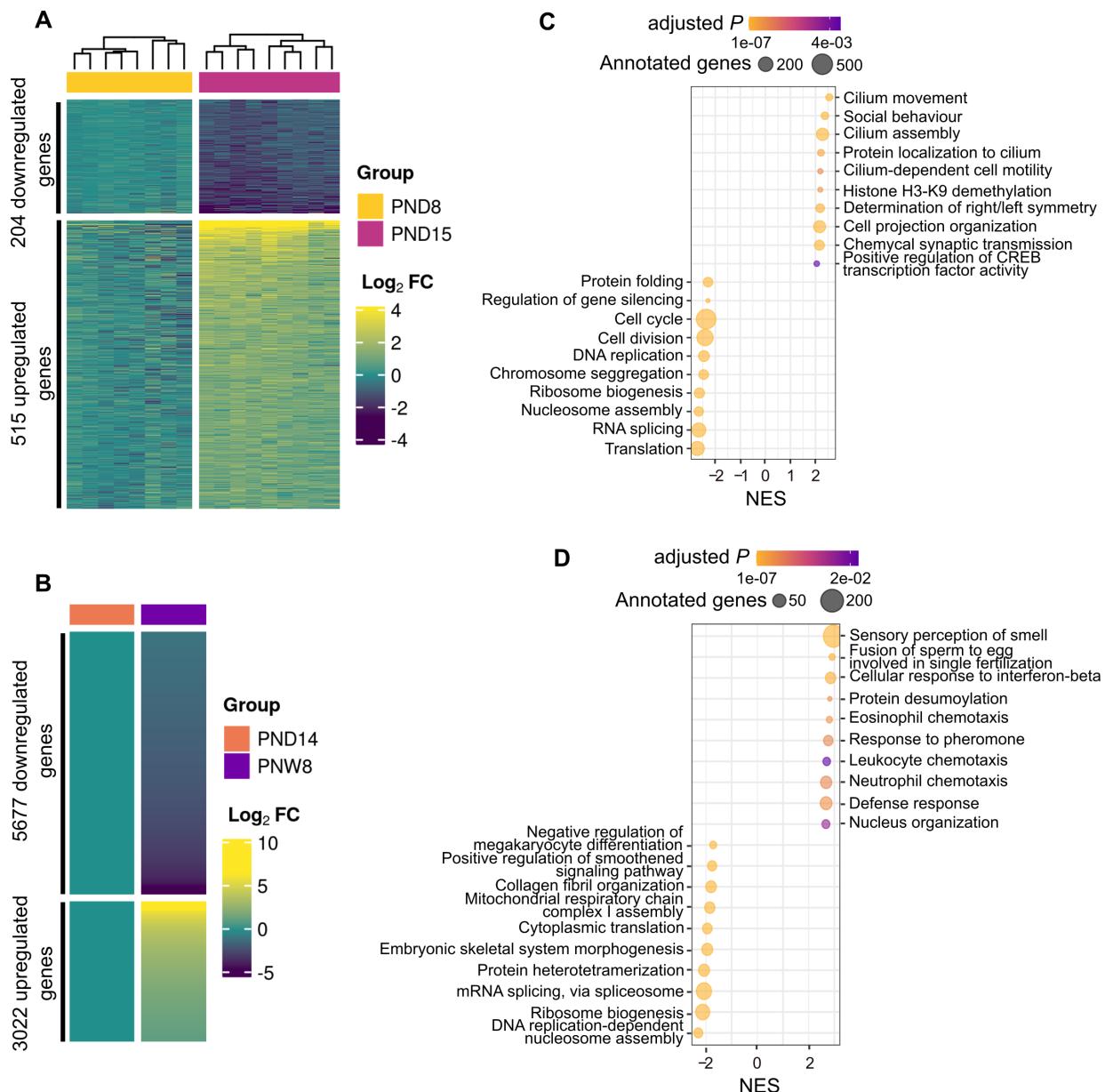
(C) Dot plots of top enriched GO biological processes (BP) terms for regions with increased and decreased chromatin accessibility in adult spermatogonia compared to PND15. The size of the dot indicates the number of genes in the term, and the colour of each dot corresponds to the adjusted  $P$  value of the term's enrichment.

### **2.3.3 Differentially accessible chromatin regions associate with distinct gene expression dynamics and histone marks**

To better understand the chromatin accessibility – transcriptome relationship across postnatal age, we conducted RNA-seq on spermatogonial cells from PND8 and PND15 testis, and further used previously generated literature datasets from PND14 and postnatal week (PNW) 8 THY1<sup>+</sup> spermatogonial cells which were categorized as adult spermatogonia (Hammoud et al., 2014, 2015). We found 719 differentially expressed genes ( $FDR \leq 0.05$ ,  $\text{abs Log}_2 FC \geq 1$ ) between PND8 and PND15 cells, indicative of dynamic transcriptional programs during early phases of spermatogonial cell proliferation and differentiation (Figure 2-2A and Table S2).

The comparison between PND14 and adult spermatogonia transcriptomes, while not affording the same type of analysis due to the low sample size, suggested even broader transcriptional changes taking place between early postnatal and steady - state adult spermatogonial cells ( $\text{Log}_2 CPM \geq 1$  and  $\text{abs Log}_2 FC \geq 1$ ) (Figure 2-2B and Table S2). To identify the biological significance of these dynamic gene expression profiles, we conducted pathway analysis using Fast Gene Set Enrichment Analysis (FGSEA) (Korotkevich et al., 2016; Subramanian et al., 2005). Between PND8 and PND15 spermatogonia we revealed a downregulation of pathways related to RNA processing and splicing, cell cycle, redox homeostasis and protein catabolism, and an upregulation of terms associated with cellular transport, exocytosis and signal transduction (Figure 2-2C and Table S2).

GSEA from PND14 and adult spermatogonia revealed that, similarly to the transition from PND8 to PND15, pathways related to RNA processing, ribosome biogenesis, and cell cycle were downregulated in adult spermatogonia (Figure 2-2D and Table S2). We also found a downregulation of pathways related to developmental programs and mitochondrial functions. In contrast, upregulated pathways were related to spermatogenesis, as well as numerous processes involving cytokine signalling (Figure 2-2D and Table S2).



**Figure 2-2 Transcriptome dynamics between early postnatal and adult stage spermatogonial cells.**

(A) Heatmap of differentially expressed genes (adjusted *P* ≤ 0.05 and abs Log<sub>2</sub> FC ≥ 1) between PND15 (*n* = 8) and PND8 (*n* = 9) spermatogonial cells. Shown are the Log<sub>2</sub> FC with respect to the average of the PND8;

(B) Heatmap of differentially expressed genes (Log<sub>2</sub> CPM ≥ 1 and abs Log<sub>2</sub> FC ≥ 1) between adult (PNW8) spermatogonia (*n* = 1) and PND14 (*n* = 1) from literature RNA-seq datasets;

(A, B) Genes are ordered by Principal Component Analysis (PCA) method using seriation (R package);

(C, D) Dot plots of top 10 enriched GO BP terms (adjusted *P* ≤ 0.05) from GSEA analysis of PND15 vs PND8 and PNW8 vs PND14 comparison, respectively. GO terms are summarized by REVIGO and ordered by their normalized enrichment scores (NES). The size of the dot indicates the number of expressed genes annotated in the GO term, and the colour corresponds to the adjusted *P* value.

Next, we integrated the chromatin accessibility and transcriptome findings from early postnatal to adult spermatogonial stage. Additionally, we mined previously published ChIP-seq and BS data from THY1<sup>+</sup> spermatogonia and investigated histone mark (H3K4me3, H3K27ac, H3K27me3) distribution and DNAm patterns (Hammoud et al., 2014, 2015) at the regions of differentially accessible chromatin. Notably, ChIP-seq data was only available for adult spermatogonial cells (PNW8), whilst DNAm was available for PND7, PND14 and adult stages.

For integrating our ATAC-seq data with the RNA-seq, ChIP-seq and BS 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 (Harrow et al., 2012a; Myers et al., 2011; Thurman et al., 2012). We further grouped proximal regions based on the change in chromatin accessibility and expression of the nearest gene between PND14 and adult spermatogonia, which led to 6 distinct categories. The first 2 most abundant categories contained regions with increased chromatin accessibility and upregulated nearby genes - Category 1, and increased chromatin accessibility and downregulated nearby genes - Category 2. Category 3 and 4 comprised proximal regions of decreased chromatin accessibility for which gene expression was either downregulated or upregulated, respectively. Categories 5 and 6 were proximal regions with increased and decreased chromatin accessibility respectively, for which the expression of the nearest gene was not detected in spermatogonial cells (Figure 2-3A and Table S3).

For a subset of regions in Category 1, ChIP-seq overlap revealed the presence of active H3K4me3, H3K27ac or dual H3K4me3/H3K27ac, and an overall lack of H3K27me3 (Figure 2-3B and Table S3). Notably, several of the genes in Category 1 for which chromatin was also marked by histone modifications, are known regulators of stem cell potential. *Pdk1* promoter region was marked by a dual H3K4me3/H3K27ac, while *Pdk1* mRNA showed a slight upregulation in adult spermatogonia (Figure S 2-3A). *Pdk1* (phosphoinositide-dependent protein kinase 1) is a glycolysis factor, important for spermatogonial stem cell self-renewal (Chen et al., 2020; Kanatsu-Shinohara et al., 2016). In contrast, *Gata2* promoter region was marked by a bivalent H3K4me3/H3K27me3 mark, while its expression showed a slight upregulation across testis maturation (Figure S 2-3A). *Gata2* is a known target of

NANOS2, an essential regulator of spermatogonial stem cell potential (Barrios et al., 2010; Sada et al., 2009). Other exemplary genes in Category 1 included pyruvate cellular carriers *Slc25a18*, *Slc23a1* and *Slc2a5*, suggesting differences in glycolysis regulation in adult spermatogonial cells (Table S3). Notably, we found an increased chromatin accessibility at the TSS of GDNF receptor *Gfra2* (Figure 2-3C). At 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 (Figure 2-3C) (Grive et al., 2019; Hammoud et al., 2015).

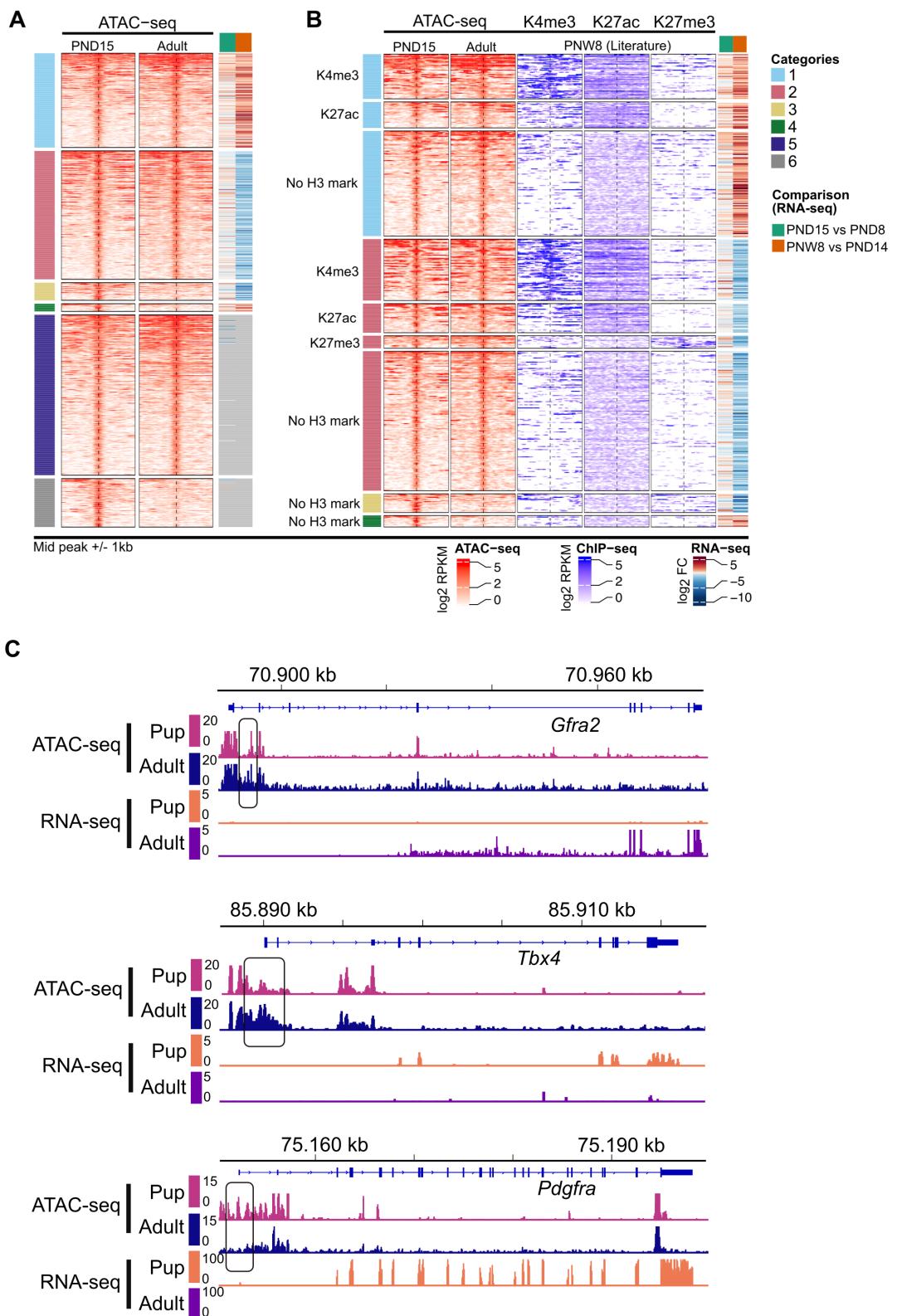
Interestingly, the highest number of differentially accessible chromatin regions were in Category 2, and included proximal regions with increased chromatin accessibility and decreased expression of nearby genes in adult spermatogonia, indicative of active repression taking place (Figure 2-3A and Table S3). Members of this category included developmental genes such as *Tbx4*, *Satb1* and *Hmx1* (Figure 2-3B and Figure S 2-3B). ChIP-seq overlap revealed that some of the regions nearby developmental genes displayed a poised H3K27me3/H3K4me3 mark (Figure S 2-3B). In addition, GO enrichment analysis on genes in Category 2 found an enrichment for pathways related to regulation of cell cycle, RNA processing, DNA repair and cell division (Table S4). An exemplary gene is *Fgf8*, important for maintenance of undifferentiated spermatogonia in the testis, via FGFR1 signalling (Hasegawa and Saga, 2014). *Fgf8* showed increased chromatin accessibility at the TSS and a downregulated expression in adult spermatogonia, in agreement with recent findings from scRNA-seq data which reported a downregulation of FGFR1-mediated signalling with age (Figure S 2-3B) (Grive et al., 2019; Hasegawa and Saga, 2014).

Regions in Category 3 displayed decreased chromatin accessibility and a downregulation of nearby genes in adult spermatogonia and were mostly depleted of any of the 3 histone marks investigated (Figure 2-3B). GO enrichment on genes in this category genes revealed enrichment for developmental pathways 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 adult stage (Hermann et al., 2018). *Pdgfra* displayed a marked downregulation in adult spermatogonia, and a decrease in chromatin accessibility overlapping the TSS

(Figure 2-3C). Another example of a gene important for early postnatal spermatogonial cell establishment is *Dap2ip*, which displayed a marked decrease in chromatin accessibility at its TSS and a decreased expression in adult spermatogonial cells (Figure S 2-3C). Surprisingly, we also identified a category of proximal regions (Category 4) with decreased accessibility (<20 regions) at genes which were upregulated in adult spermatogonial cells and with previously uncharacterized roles in spermatogonial cells (Figure 2-3A and Table S3).

Notably, DNAm profiles across postnatal stages did not show drastic changes in any of the 6 categories of proximal regions, suggesting a relatively stable DNAm profile in the transition from early postnatal to adult stage (Figure S 2-4). Aside from proximal regions, we also identified numerous distal regions with differential chromatin accessibility between PND15 and adult (Figure S 2-5A and B). Similar to proximal regions, accessibility in distal regions mainly increased in adult spermatogonial cells compared to 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 (Figure S 2-5A and B and Table S3). Similar to proximal regions, DNAm levels did not display major changes in distal regions (Figure S 2-5A and B).

Taken together, our data integration reveals novel associations between open chromatin regions of differential accessibility, histone marks and dynamically expressed genes in spermatogonial cells during testis maturation, and reveal how chromatin accessibility may contribute to the differential utilization of signalling pathways across age.



**Figure 2-3 Chromatin accessibility and histone modifications at proximal regions of genes dynamically expressed between adult and PND15 spermatogonial cells.**

(A) Heatmap of chromatin accessibility differences deduced by Omni-ATAC between PND15 and adult spermatogonial cells. Differentially accessible regions were grouped based on the correlation with the

expression of the nearest gene between PND14 and adult spermatogonia in 6 distinct categories: proximal regions of increased chromatin accessibility and increased gene expression (Category 1, n = 171), increased chromatin accessibility and decreased gene expression (Category 2, n = 233), decreased chromatin accessibility and decreased gene expression (Category 3, n = 32) and decreased chromatin accessibility and increased gene expression (Category 4, n = 14). Proximal inactive regions were defined as regions of increased accessibility (Category 5, n = 291) or decreased accessibility (Category 6, n = 88) for which the nearest gene expression was not detected from RNA-seq;

(B) Heatmaps showing the overlap between Category 1-4 regions and genes, and literature ChIP-seq data in PNW8 spermatogonia for H3K4me3, H3K27ac and H3K27me3. For each of the Category 1-4 the following sub-categorization was applied: regions that are enriched for H3K4me3 (with or w/o H3K27ac and/or H3K27me3), regions that are enriched for H3K27ac (and lack both H3K4me3 and H3K27me3) and regions that are enriched for H3K27me3 (and lack both H3K4me3 and H3K27ac);

(A, B) Each line represents a peak region and the regions are ordered within a category by the ATAC-seq signal. Mid-x-axis corresponds to the middle of a peak region and is extended to +/- 1 kb. The colour-key of the ATAC-seq heatmap represents the ATAC-seq signal in Log<sub>2</sub> Reads Per Kilobase per Million (RPKM) reads sequenced. For RNA-seq, log<sub>2</sub> FC is shown from PND15 vs PND8 and PNW8 vs PND14 comparisons;

(C) Genomic snapshots from the Integrative Genomics Viewer (IGV, Broad Institute) of exemplary genes from Category 1 (*Gfra2*), Category 2 (*Hmx1*) and Category 3 (*Pdgfra*) showing relative abundance of transcripts from RNA-seq and chromatin accessibility from ATAC-seq. RNA-seq data corresponds to literature PND14 and adult (PNW8) spermatogonial cells and ATAC-seq data to PND15 and adult (PNW20) spermatogonial cells, respectively.

### **2.3.4 Differentially accessible chromatin regions are marked by binding sites for distinct families of transcription factors (TFs)**

TFs are important in establishing and maintaining distinct transcriptional signatures across the developmental trajectory of a cell population (Fushan et al., 2015; Shavlakadze et al., 2019). TF predominantly bind regions of open chromatin throughout the genome and allow for a dynamic regulation of gene expression (Klemm et al., 2019). To investigate if the regions of open chromatin with differential accessibility between PND15 and adult spermatogonia are enriched in TF binding motifs, we performed motif enrichment analysis using the Hypergeometric Optimization of Motif EnRichment (HOMER) tool (Heinz et al., 2010).

In regions with increased chromatin accessibility, we identified 41 enriched TF motifs ( $q\text{-value} \leq 0.05$ ) (Figure 2-4A and Table S4). The top candidate list ( $q\text{-value} \leq 0.0001$ ) was dominated by members of the Fos/Jun family (FOS, FOSB, FOSL1 and FOSL2, JUN, JUNB and JUND) (Figure 2-4B). Notably, at mRNA level, some of the TFs displayed age-specific differences ( $\text{Log}_2\text{CPM} \geq 1$  and  $\text{abs Log}_2\text{FC} \geq 1$ ): *Fos*, *Junb* and *Jund* were downregulated in adult spermatogonial cells (Figure 2-4C). JUN, FOS and CREB are 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 (Martínez-Zamudio et al., 2020; Shaulian and Karin, 2002). JUND and c-FOS specifically promote the proliferative potential of spermatogonial stem cells (He et al., 2008; Wang et al., 2018). USF1 and POU3F1, 2 important factors in the maintenance of the spermatogonial stem cell pool displayed enriched motifs in the regions of increased chromatin accessibility and a decrease in expression (Figure 2-4A and C). POU3F1 is a GDNF-regulated TF, which has been shown to play an important role in promoting spermatogonial cell self-renewal capacity (Niu et al., 2011; Wu et al., 2010). Our analysis also revealed enriched binding sites for retinoic acid receptors such as RXR $\alpha$  and RAR $\alpha$  (Figure 2-4A). Recent reports have revealed that *Rxra* and *Rara* utilization in spermatogonial cells is vastly dependent on the niche microenvironment (Lord et al., 2018).

To check if some of 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 (Figure 2-4D). In mESCs, NF-Y TFs facilitate a permissive chromatin conformation, and play an important role in the expression of core ESC pluripotency genes (Oldfield et al., 2014). Furthermore, NF-YA/B motif enrichment has also been found in regions of open chromatin in human spermatogonial cells (Guo et al., 2017).

Less accessible chromatin regions in adult spermatogonial cells also displayed a high number of enriched TF binding motifs (Figure 2-4A and Table S4). Notably, almost all of these TF motifs were uniquely enriched in the regions of decreased chromatin accessibility and predominantly associated with developmental functions. Top hits included members of the FOX family (FOXO1, FOXO3, FOXP2, FOXK1, FOXA2) and members of the ETS and ETS-related TFs (ETS1, GABPA, ETV4, ELF1, ELF3) (Figure 2-4B). Expression levels of most of these TFs were decreased in adult spermatogonial cells (Figure 2-4A). FOXO1 is also a pivotal regulator of the self-renewal and differentiation of spermatogonial stem cells, via the PI3K-Akt signalling pathway (Chan et al., 2014; Goertz et al., 2011). The roles of ETS-related TFs in spermatogonial cells have yet to be clarified, however recently published data found high expression of *Etv4* in the stem-cell enriched fraction of the spermatogonial population, particularly during the spermatogonial stem cell pool establishment, immediately after birth (Cheng et al., 2020; Law et al., 2019). Motif enrichment analysis on the regions with decreased chromatin accessibility situated in gene body and intergenic regions revealed that TFs important in numerous developmental processes (FOXC1, FOXJ2, FOXM1, LHX6) were specifically enriched in intergenic regions of decreased chromatin accessibility (Figure 2-4D). This is consistent with the enrichment in developmental pathways that we found at the regions of decreased chromatin accessibility in adult spermatogonia (Figure 2-1C). Taken together, our TF motif analyses reveal that a shifting repertoire of TFs accompanies the chromatin reorganization taking place from early postnatal to adult spermatogonia.

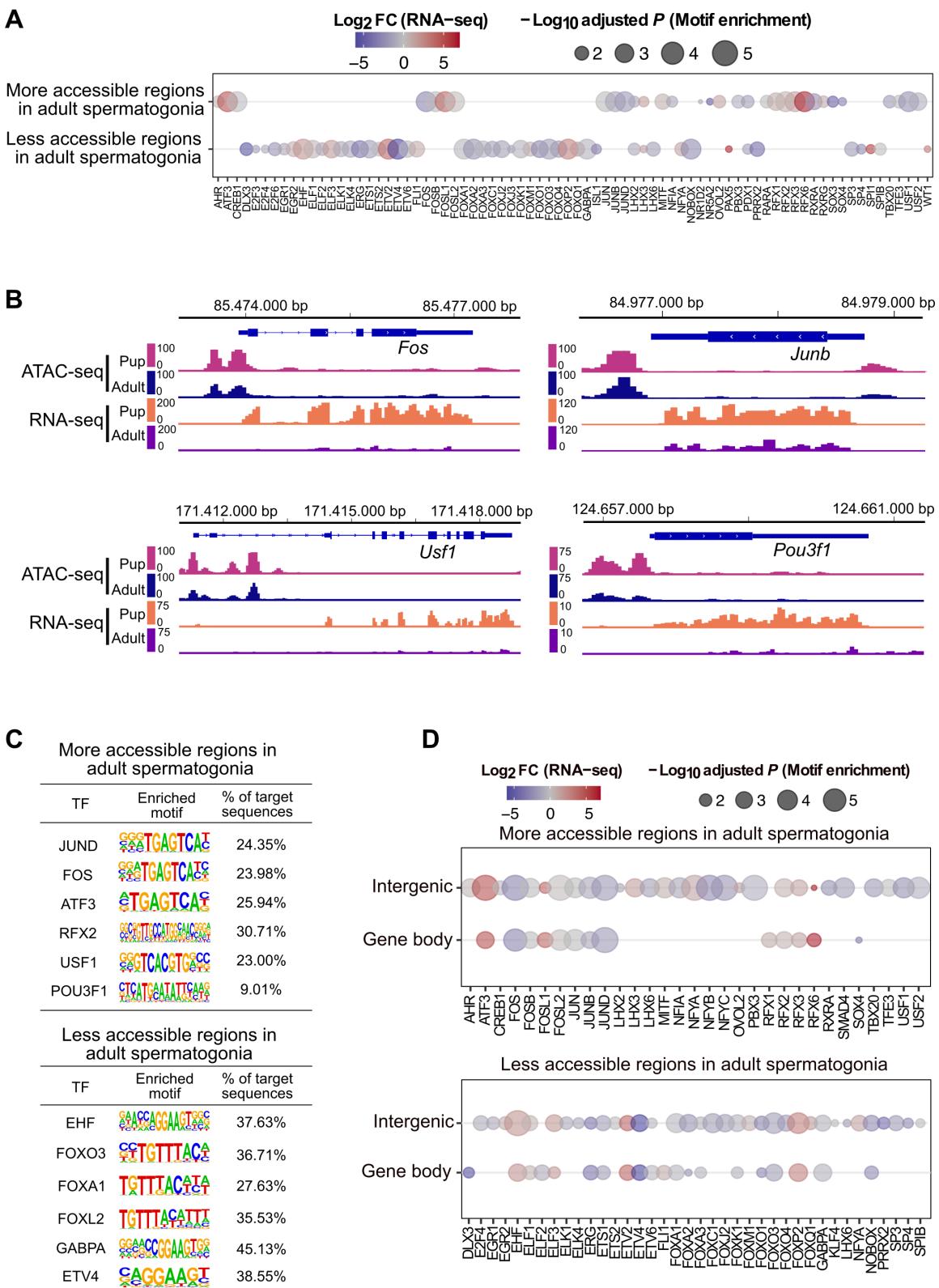


Figure 2-4 Transcription factor dynamics at differentially accessible regions as predicted by motif enrichment in adult spermatogonial cells.

(A) Dot plot of all transcription factor motifs enriched in the regions of decreased and increased accessibility between PND15 and adult spermatogonia;

- (B) Genomic snapshots from the Integrative Genomics Viewer (IGV, Broad Institute) of mRNA expression levels of representative enriched TFs in the regions of increased chromatin accessibility. RNA-seq data corresponds to literature PND14 and adult (PNW8) spermatogonial cells and ATAC-seq data to PND15 and adult (PNW20) spermatogonial cells, respectively;
- (C) HOMER extracted consensus sequences for each transcription factor motif. Representative examples from the most enriched transcription factor families are depicted;
- (D) Dot plots of top transcription factor motifs enriched in differentially accessible chromatin regions situated in gene bodies and in intergenic areas of the genome;
- (A, D) Each dot corresponds to a motif. The differential gene expression of each transcription factor was extracted from the PND14 and adult spermatogonia RNA-seq from literature, and is shown as colour-coded Log<sub>2</sub> FC. The size of the dot indicates the HOMER motif enrichment adjusted *P* of each motif.

### **2.3.5 Open chromatin at transposable elements (TEs) undergoes significant remodelling in the transition from early postnatal to adult spermatogonia**

Recent evidence suggests an important role for long terminal repeat (LTR) - type elements, specifically for ERVKs, in the transcriptional regulation of mRNAs and long non-coding RNAs (lncRNAs) during spermatogenesis (Davis et al., 2017; Sakashita et al., 2020). Accessibility analysis at LTRs in mitotic and meiotic germ cells, revealed a unique chromatin accessibility landscape in spermatogonial cells, compared to the rest of germ cells in the testis (Sakashita et al., 2020).

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 adult stage is accompanied by significant chromatin accessibility differences at 135 TE subtypes (FDR  $\leq 0.05$ , abs Log<sub>2</sub> FC  $\geq 0.5$ ) (Figure 2-5A and B and Table S5). Although most of the differentially accessible TE subtypes displayed a decrease in chromatin accessibility between PND15 and adult spermatogonia (68.9%, 93/135) (Figure 2-5A), we also observed 42 TE subtypes which increased in accessibility in adult spermatogonia (Figure 2-5B). Of note, the increase in accessibility correlated with an increased expression in adult spermatogonial cells compared to early postnatal stage (Figure 2-5B). TE loci within the subtypes harbouring changes in chromatin accessibility were predominantly situated in intergenic and intronic regions (68% intergenic and 25% intronic), and only 6% were located in proximity of a gene (+/- 1kb from a TSS) (Figure 2-5C).

LTRs were the most abundant TEs with altered chromatin accessibility, specifically ERVK and ERV1 subtypes (Figure 2-5A and B). Exemplary ERVK subtypes harbouring less accessible chromatin included RLTR17, RLTR9A3, RLTR12B and RMER17B (Table S5). Enrichment of RLTR17 and RLTR9 repeats has been reported previously in mESCs, specifically at TFs important for pluripotency maintenance such as *Oct4* and *Nanog* (Fort et al., 2014). Interestingly, we identified the promoter region of the lncRNA *Lncenc1*, an important regulator of pluripotency in mESCs (Fort et al., 2014; Sun et al., 2018b), harbouring several LTR loci with decreased accessibility in adult spermatogonia. One of these LTR loci, RLTR17, overlapped the TSS of *Lncenc1*.

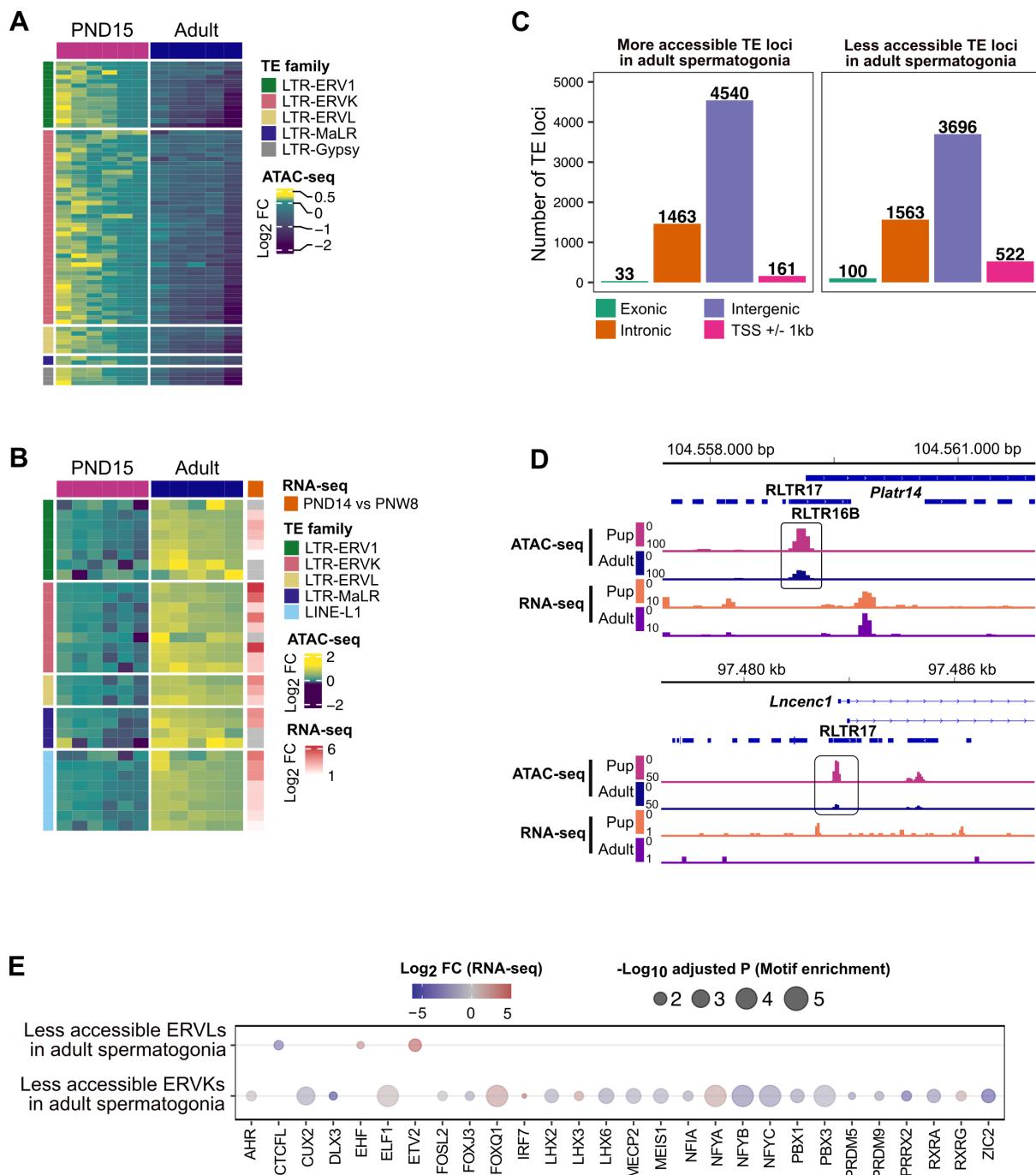
This decrease in accessibility correlated with a marked decrease in expression of *Lncenc1* in adult spermatogonia (Figure 2-5D). *Lncenc1* (also known as *Platr18*) is part of the pluripotency-associated transcript (*Platr*) family of lncRNAs which were recently identified as potential regulators of the pluripotency-associated genes *Oct4*, *Nanog* and *Zfp42* in mESCs (Bergmann et al., 2015; Dann et al., 2008; Wu et al., 2018). We also identified several other *Platr* genes, such as *Platr27* and *Platr14*, for which the TSS overlapped LTRs with reduced accessibility, RLTR17 and RLTR16B\_MM, respectively (Figure 2-5D and Table S5). *Platr27* and *Platr14* also showed a decrease in mRNA expression in adult spermatogonia, while their expression was unchanged between PND8 and PND15 (Figure 2-5D and Table S5). The remaining LTR subtypes with decreased accessibility in adult spermatogonia belonged to the ERV1, ERVL and MaLR families (Figure 2-5A). 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 (Table S5).

Emerging evidence suggests an important contribution of TEs in providing tissue-specific substrates for TF binding (Fort et al., 2014; Sundaram and Wysocka, 2020; Sundaram et al., 2014). To investigate the regulatory potential of the less accessible LTR subtypes, we assessed the enrichment of TF motifs in these regions using HOMER (Table S6). 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 (Figure 2-5E and Figure S 2-6A). 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 (Figure 2-5E) (Green et al., 2018).

Among the TE subtypes which increased in accessibility, members of the ERVK, ERVL and ERV1 families were predominant (57,1%, 24/42) (Figure 2-5B). Interestingly, we also found a considerable number of LINE L1 subtypes with increased chromatin accessibility in adult spermatogonial cells (Figure 2-5B). 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 chromosome 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 (Figure 2-6A). Representative examples were *Olfr362* and *Olfr1307*, both situated in the *Olfr* gene cluster on Chr2 (Figure 2-6B). 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 (Figure S 2-6B).

Similar to before, we performed TF motif enrichment analysis at the family level by grouping together the more accessible TE subtypes coming from one family (Table S6). More accessible LINE L1s were highly enriched in TF motifs, particularly in multiple members of the ETS, E2F and FOX families (Figure 2-6C). 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 (Figure 2-6C and Figure S 2-6A) (Cheng et al., 2020). 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 (Figure 2-6C and Figure S 2-6A). 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 timepoints, and suggest potential gene programs that may be regulated by these changes.



**Figure 2-5 Differential chromatin accessibility at transposable elements (TEs) in adult spermatogonial cells compared to PND15**

(A) Heatmap of the LTR and LINE subtypes with decreased accessibility between adult and PND15 spermatogonia extracted from the Omni-ATAC data (adjusted  $P \leq 0.05$  and  $\text{Log}_2 \text{ FC} \geq 0.5$ );

(B) Heatmap of the LTR and LINE subtypes with increased accessibility between adult and PND15 spermatogonia extracted from the Omni-ATAC data (adjusted  $P \leq 0.05$  and  $\text{Log}_2 \text{ FC} \geq 0.5$ ). Expression changes of these subtypes between PND14 and adult spermatogonia RNA-seq from literature is represented as  $\text{Log}_2 \text{ FC}$ ;

(A, B) For both accessibility heatmaps, Log<sub>2</sub> FC are shown with respect to the average of the PND15 samples. Samples are clustered using Ward's method. Subtypes are ordered by principal component analysis (PCA) method using seriation (R package);

(C) Bar plot illustrating the genomic distribution of differentially accessible TEs between adult and PND15 spermatogonial cells;

(D) Genomic snapshots from the Integrative Genomics Viewer (IGV, Broad Institute) of *Lncenc1* and *Platr14* showing LTRs from RepeatMasker and the average normalized RNA-seq and ATAC-seq coverage (RPKM). LTR loci were extracted using RepeatMasker. RNA-seq data corresponds to literature PND14 and adult (PNW8) spermatogonial cells and ATAC-seq data to PND15 and adult (PNW20) spermatogonial cells, respectively;

(E) Dot plots of top transcription factor motifs enriched in the less accessible ERVKs and ERVL subtypes. Each dot corresponds to a motif. The differential gene expression of each transcription factor was extracted from the PNW8 vs PND14 comparison from literature RNA-seq, and is shown as colour-coded Log<sub>2</sub> FC. The size of the dot indicates the HOMER motif enrichment adjusted *P* of each motif.

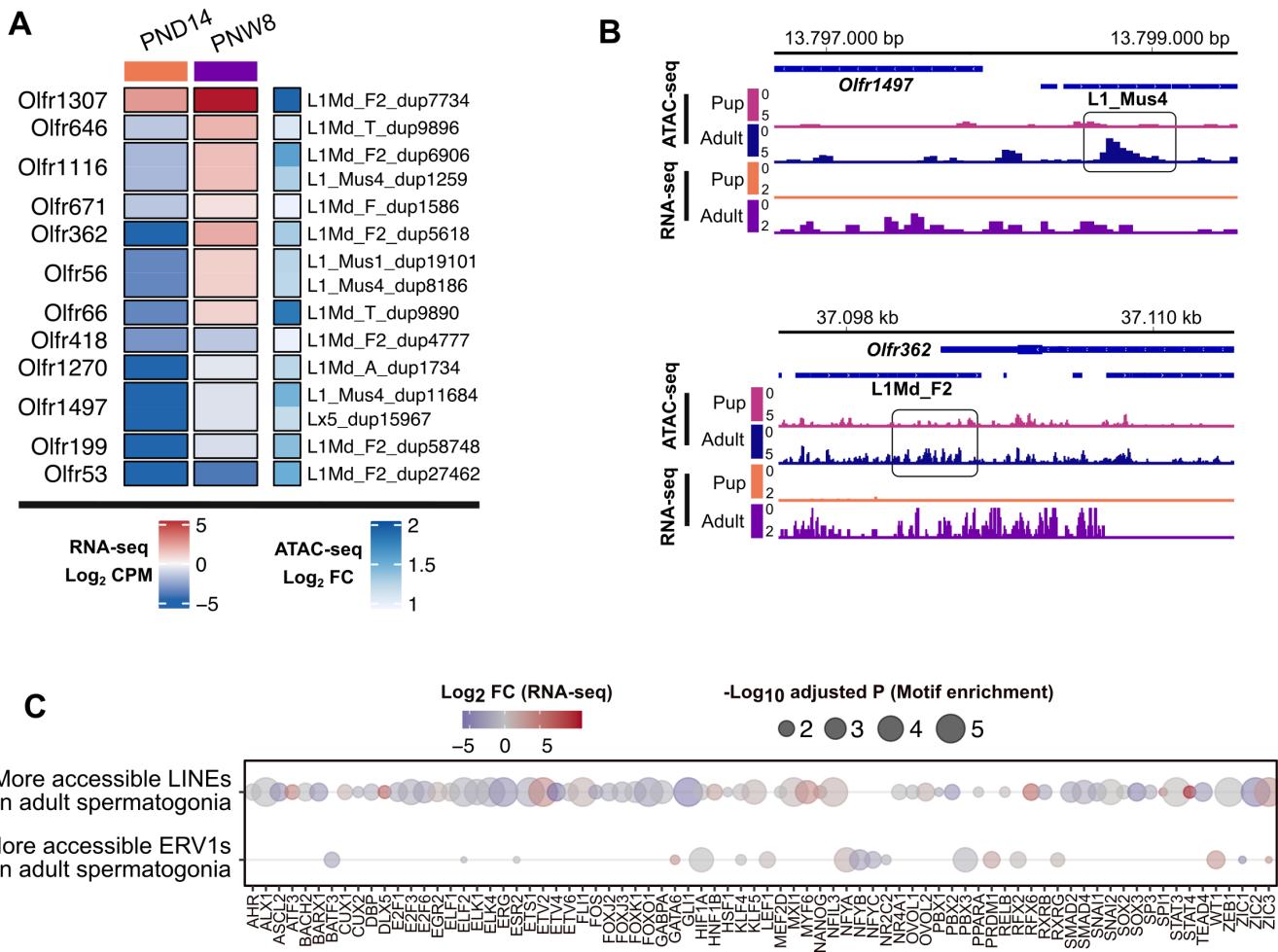


Figure 2-6 Increased accessibility at LINE L1 subtypes located near *Olfr* gene clusters

(A) Heatmap of the *Olfr* genes for which we identified an upregulated expression from PND14 to PNW8 timepoints and an increase in accessibility at a nearby L1 locus. RNA expression levels are expressed as Log<sub>2</sub> CPM at each timepoint. Accessibility changes at each of the corresponding L1 locus situated within +/- 5kbp from the gene are expressed as Log<sub>2</sub> FC calculated from the ATAC-seq analysis of the differentially accessible TEs between adult and PND15 spermatogonial cells;

(B) Genomic snapshots from the Integrative Genomics Viewer (IGV, Broad Institute) of exemplary genes *Olfr1497* and *Olfr362* showing relative abundance of transcripts from RNA-seq and chromatin accessibility from ATAC-seq. LINE loci were extracted using Repeat Masker. RNA-seq data corresponds to literature PND14 and adult (PNW8) spermatogonial cells and ATAC-seq data to PND15 and adult (PNW20) spermatogonial cells, respectively;

(C) Dot plots of top transcription factor motifs enriched in the more accessible L1 and ERV1 subtypes. Each dot corresponds to a motif. The differential gene expression of each transcription factor was extracted from the PNW8 vs PND14 comparison from literature RNA-seq, and is shown as colour-coded Log<sub>2</sub> FC. The size of the dot indicates the HOMER motif enrichment adjusted *P* of each motif.

## 2.4 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 (Grive et al., 2019; Hammoud et al., 2015; Hermann et al., 2018).

Our ATAC-seq revealed a reorganization of open chromatin in adult spermatogonia compared to the PND15 cell population. We found that the regions of differential accessibility were associated with distinct gene pathways, with morphogenesis and developmental pathways enriched in regions of decreased chromatin accessibility, while regions of increased chromatin openness in adult spermatogonia were enriched for DNA repair pathways, stem cell maintenance, RNA processing and protein metabolism.

Notably, distinct families of TFs were enriched at the regions of increased chromatin accessibility compared to the regions which became less accessible in adult spermatogonial cells. As such, TFs known to regulate spermatogonial cell proliferation, pluripotency potential and niche-dependent signalling such as JUND and c-FOS, POU3F1, and retinoic acid receptors, 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 which decreased in accessibility in adult spermatogonia. For some of the enriched TF motifs, we observed a preference for certain genomic locations: 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 also localize in intergenic regions of open chromatin in human spermatogonial cells (Guo et al., 2017), prompting additional investigation of their roles in regulating spermatogonial cell programs, with potential consequences for sperm formation. Altogether, our pathway and TF analyses reveal potential regulatory elements for the age-specific gene expression of spermatogonial cells.

Our comparison of the gene expression changes from PND8 to PND15 spermatogonial cells, together with the reanalysis of literature RNA-seq data from PND14 and adult THY1+ spermatogonia, confirmed the dynamic transcriptome associated with postnatal spermatogonial cell states (Grive et al., 2019; Hammoud et al., 2015). The age-dependent upregulation of pathways associated with signal transduction, cellular communication and cytokine signalling, supports previous findings suggesting that, as the testis matures and the somatic niche develops, spermatogonial cells rely more on paracrine signalling, and undergo vast changes in gene expression. To obtain a comprehensive profile of the chromatin landscape and the transcriptome differences between early postnatal and adult spermatogonial cells, we integrated the chromatin accessibility and gene expression profiles with known histone H3 modifications and global DNAmc patterns of THY1+ spermatogonial cells from literature (Hammoud et al., 2014, 2015). This allowed us to identify several distinct categories of differentially accessible chromatin regions for which the nearest gene was dynamically expressed between early postnatal and adult stage.

In the category of upregulated genes with increased nearby chromatin accessibility, we identified several factors associated with redox processes, mitochondria function and cell proliferation. Interestingly, a similar number of genes displayed an increase in chromatin accessibility and a downregulated expression in adult spermatogonia, suggesting that active repression is taking place at these genes (Starks et al., 2019). This category comprised factors important for cell cycle and RNA processing, as well as developmental genes. 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 (Erkek et al., 2013; Hammoud et al., 2014; Jung et al., 2017). Therefore, our findings suggest that open chromatin reorganization may play a role in regulating the poised status of certain developmental genes in the germline. We also identified a category of regions for which the decrease in chromatin accessibility correlated with a decreased expression, which also included developmental factors. The stable methylation patterns we detected at the differentially accessible chromatin regions, suggest a minimal impact for DNAmc 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 spermatogonia. ERVK and ERV1 subtypes were the largest categories of TEs to become less accessible in adult spermatogonia, whilst LINE L1 subtypes displayed an increase in chromatin accessibility. Although the majority of these TEs resided in intergenic and intronic regions, we identified 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 pluripotency potential of mouse and human embryonic stem cells (Bergmann et al., 2015; Fort et al., 2014). In mouse embryonic stem cells, RLTR17 is highly expressed and is enriched in open chromatin regions, and provides binding sites for pluripotency factors OCT4 and NANOG (Fort et al., 2014). 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 *Oifr* 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 which encode proteins with specialized functions (Lu et al., 2020). Among these, the *Oifr* gene family was the most enriched in L1 elements (Lu et al., 2020). Although their role in spermatogonial cells is currently not established, *Oifr* proteins have been implicated in the swimming behaviour of sperm (Fukuda and Touhara, 2005; Vanderhaeghen et al., 1997). Given the dynamic chromatin profile of LINE L1 elements that we found at *Oifr* genes between early postnatal and adult spermatogonia, we speculate that *Oifr* genes could play additional roles in spermatogenesis, other than in sperm physiology. In addition, we found a high number of enriched TF motifs present at the differentially accessible ERVKs and LINE L1 families. These findings reveal that chromatin accessibility at TEs is reorganized in

spermatogonia cells during the transition from developing to adult stages, and may contribute to distinct gene regulatory networks (Sundaram and Wysocka, 2020; Sundaram et al., 2014).

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 arise 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.

## 2.5 Methods

### 2.5.1 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.

### 2.5.2 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 (Kubota et al., 2004b). 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% foetal 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 in an 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 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 resuspended in PBS-S and used for sorting.

### **2.5.3 Spermatogonial cell 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

$\mu\text{g}/\text{ml}$  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 excess. Antibodies were obtained from BD Biosciences (San Jose, United States) unless otherwise stated. Cell sorting was performed on a FACS Aria III 5L, at 4°C and using an 85 $\mu\text{m}$  nozzle, at the Cytometry Facility of University of Zurich. For RNA-seq on PND8 and PND15 spermatogonia, cells were collected in 1.5 ml Eppendorf tubes in 500  $\mu\text{L}$  PBS-S, immediately pelleted by centrifugation and snap frozen in liquid N<sub>2</sub>. Cells 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 (Corces et al., 2017). For OmniATAC on adult spermatogonia, 5000 cells from each animal were collected in a separate tube and further processed using the same protocol.

#### **2.5.4 Immunocytochemistry**

The protocol used for assessing spermatogonial cell enrichment after sorting was kindly provided by the Oatley Lab at Washington State University, Pullman, USA (Yang et al., 2013). Briefly, 30,000-50,000 cells were 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  $\mu\text{g}/\text{ml}$ , Active Motif, clone 2A9) primary antibody. Alexa488 goat anti-mouse IgG (1  $\mu\text{g}/\text{mL}$ , 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<sup>+</sup> 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).

#### **2.5.5 RNA extraction and RNA-seq library preparation**

For RNA-seq on PND8 and PND15 spermatogonial cells, total RNA was extracted from sorted cells using the 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 the SMARTer® Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian (Takara Bio USA, Inc.) at the Functional Genomics Centre (FGC) Zurich, according to the manufacturer's instructions. The number of samples sequenced at each timepoint was 8 for PND8 and 9 for PND15 spermatogonia (each sample was representative of 4 testes from 2 pups).

### 2.5.6 Omni-ATAC library preparation

Chromatin accessibility was profiled from PND15 and adult spermatogonial cells. The libraries were prepared according to the Omni-ATAC protocol, starting from 25 000 PND15 and 5000 adult sorted spermatogonia, respectively (Corces et al., 2017). Briefly, sorted cells were lysed in cold lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin) and the nuclei were pelleted and transposed using the Nextera Tn5 (Illumina) for 30 min at 37°C in a thermomixer with shaking at 1000 rpm. The 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) in order to remove primer dimers (78bp) and large fragments of 1000-10,000 bp in length. Library quality was assessed on an Agilent High Sensitivity DNA chip using the Bioanalyzer 2100 (Agilent Technologies). 6 samples were sequenced from PND15 and 5 from adult spermatogonial cells.

### 2.5.7 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)

#### 2.5.7.1 RNA-seq data analysis

*Quality control and alignment:* Single-end (SE) sequencing was performed on the PND8, and PND15 spermatogonia samples on the Illumina HiSeq4000 at the FGC Zurich. PND8 raw data (FASTQ files) was merged from two individual runs. For the analysis of literature RNA-seq data, FASTQ files of PND14 and PNW8 samples were obtained using fastq-dump (version 2.10.8). Quality assessment of the FASTQ files was performed using FastQC (Andrews et al., 2012) (version 0.11.8). TrimGalore (Krueger, 2015) (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 then pseudo-aligned using Salmon (Patro et al., 2017) (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 (Harrow et al., 2012a) (version M18) with additional piRNA precursors, and transposable elements (concatenated by family) from Repeat Masker as in (Gapp et al., 2018).

*Downstream analysis:* the downstream analysis was performed using 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 (Huber et al., 2015). Pre-filtering of genes was performed using the filterByExpr function from edgeR (Robinson et al., 2009) (version 3.28.1) with a design matrix and requiring at least 15

counts (min.counts = 15). Normalization factors were obtained using TMM normalization (Robinson and Oshlack, 2010) from the edgeR package and differential gene expression (DGE) analysis was performed using the limma-voom (Law et al., 2014) pipeline from limma (Ritchie et al., 2015) (version 3.42.2). The Log<sub>2</sub> fold change between PNW8 and PND14 spermatogonial samples was calculated by subtracting Log<sub>2</sub> normalized expression values of PND14 from PNW8 samples. 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) (Korotkevich et al., 2016). For the PND15 vs PND8 comparison, genes were pre-ranked using their t-statistic; for the PNW8 vs PND14 comparison the Log<sub>2</sub> fold change was used due to the lack of multiple replicates for each timepoint. REVIGO was used to summarize the GO terms obtained following fGSEA (Supek et al., 2011).

#### 2.5.7.2 Omni-ATAC data analysis

*Quality control, alignment, and peak calling:* Paired-end (PE) sequencing was performed on PND15 and adult spermatogonial cell samples on the Illumina HiSeq2500 platform at the FGC Zurich. FASTQ files were assessed for quality using FastQC (Andrews et al., 2012) (version 0.11.8). Quality control (QC) was performed using TrimGalore (Krueger, 2015) (version 0.6.2) in PE mode (--paired), trimming adapters, low-quality ends (-q 30) and discarding reads that become shorter than 30 bp after trimming (--length 30). Alignment on the GRCm38 genome was performed using Bowtie2 (Langmead and Salzberg, 2012) (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 (Ramirez et al., 2016) (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 (ENCODE accession ENCF547MET), were filtered out. To call nucleosome-free regions, all aligned files were merged within groups (PND15 and adult), sorted, and indexed using SAMtools

(Li et al., 2009) (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 (Zhang et al., 2008) (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 (Huber et al., 2015). The peaks were annotated based on overlap with GENCODE (Harrow et al., 2012b) (version M18) 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](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 (Lun and Smyth, 2015) (version 1.20.0). Peak regions which 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 (Robinson and Oshlack, 2010) 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 (Robinson et al., 2009) (version 3.28.1). Peak regions which had an absolute Log<sub>2</sub> fold change  $\geq 1$  and an FDR  $\leq 0.05$  were categorized as differentially accessible regions (DARs). GO analysis was performed on DARs with the rGREAT package (Zuguang, 2020) (version 1.18.0), which is a wrapper around the GREAT tool (McLean et al., 2010) (version 4.0). Transcription factor motif enrichment analysis was performed using the marge package (Amezquita, 2018) (version 0.0.4.9999), which is a wrapper around the Homer tool (Heinz et al., 2010) (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](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, e.g. LTR), **Family** (level 2, e.g. LTR → L1), **Subtype** (level 3, e.g. LTR → L1 → L1\_Rod), and **Locus** (level 4, e.g. 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 analysing TEs. Mapped reads were assigned to TEs using featureCounts from the R package Rsubread (Liao et al., 2019b) (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 csaaw package (Lun and Smyth, 2015) (version 1.20.0). Subtypes which did not have at least 15 reads, and loci which 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 which had an absolute  $\text{Log}_2$  fold change  $\geq 0.5$  and an FDR  $\leq 0.05$  were categorized as differentially accessible subtypes and the loci with an absolute  $\text{Log}_2$  fold change  $\geq 1$  and an FDR  $\leq 0.05$  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.

#### 2.5.7.3 ChIP-seq data analysis

*Quality control, alignment, and peak calling:* ChIP-Seq SE data for PNDW8 (adults) were obtained from GEO accession GSE49621 (Hammoud et al., 2014). 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 (Andrews et al., 2012) (version 0.11.8). Quality control (QC) was performed using TrimGalore (Andrews et al., 2012) (version 0.6.0), trimming adapters, low-quality ends (-q 30) and discarding trimmed reads shorter than 30 bp (--length 30). Alignment to the GRCh38 genome was performed using Bowtie2 (Langmead and Salzberg, 2012) (version 2.3.5). Reads with more than 3 mismatches were removed from the aligned data, as suggested in (Royo et al., 2016), 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 (Zhang et al., 2008) (version 2.2.7.1) with mouse genome size (-g 2744254612) and SE BAM file format (-f BAM).

#### 2.5.7.4 Bisulphite sequencing (BS) data analysis

*Quality control and alignment:* BS paired-end data for PND7, PND14, and PNW8 (adults) were obtained from GEO accession GSE49623 (Hammoud et al., 2015). FASTQ files were obtained using fastq-dump (version 2.10.8), and different runs were

merged. FASTQ files were assessed for quality using FastQC (Andrews et al., 2012) (version 0.11.8). QC was performed using TrimGalore (Krueger, 2015) (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 (Krueger and Andrews, 2011) (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).

## 2.6 Authors contributions

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

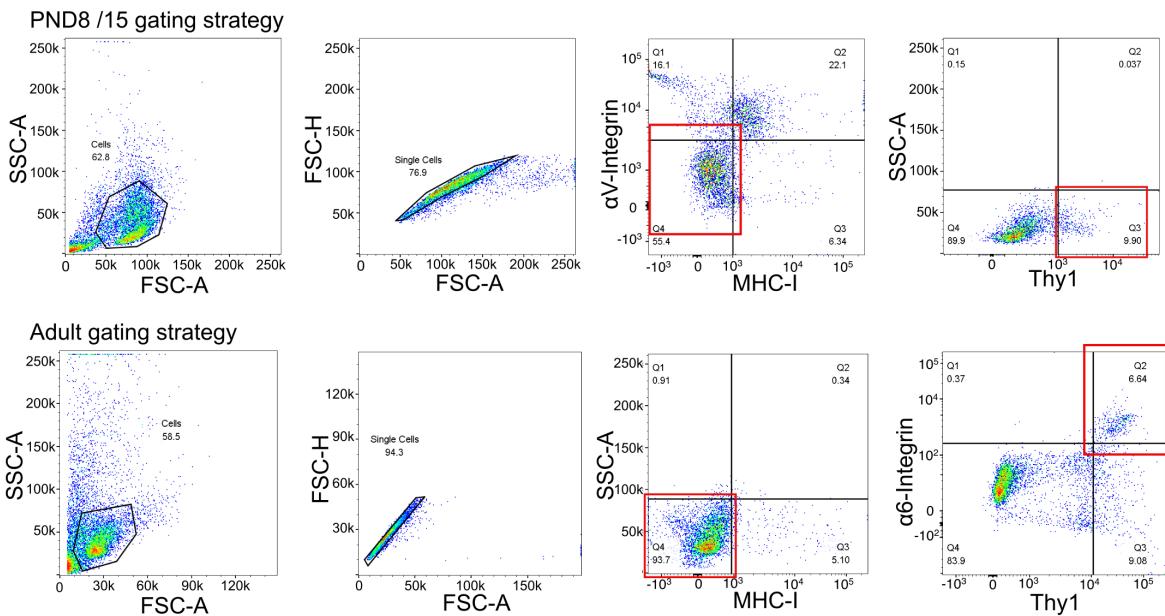
## 2.7 Acknowledgements

We thank Martin Roszkowski and Francesca Manuella for assisting with the breedings, Yvonne Zipfel for animal care in Zurich, Silvia Schelbert and Alberto Corcoba for taking care of the animal licenses and lab organization in Zurich. We thank Niharika Gaur for technical support with the Omni-ATAC protocol. We thank Rodrigo Arzate for conceptual support and critical reading of the manuscript. We thank Catherine Aquino and Emilio Yángüez from the FGC (ETH/UZH) Zurich for support and advice with sequencing and library preparation. We thank Jon Oatley, Tessa Lord and Nathan Law for advice and for providing detailed protocols of the testis dissection and preparation and for the immunocytochemistry of spermatogonial cells. We thank Zuguang Gu for his support with the heatmaps generation. We thank S3IT of UZH ([www.s3it.uzh.ch](http://www.s3it.uzh.ch)) for computational infrastructure.

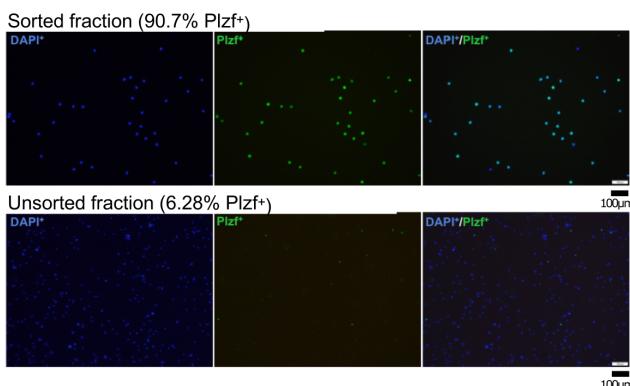
## 2.8 Supplemental figures

**Figure S1**

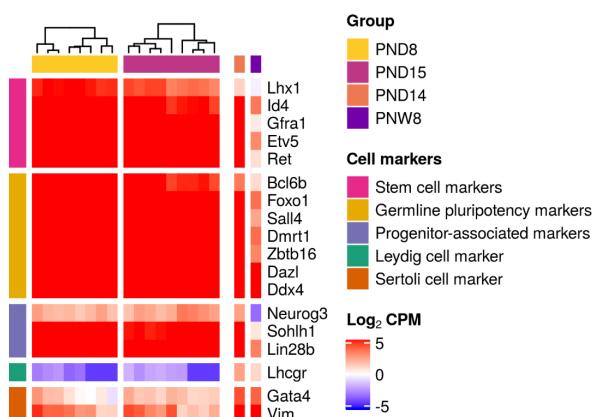
**A**



**B**



**C**



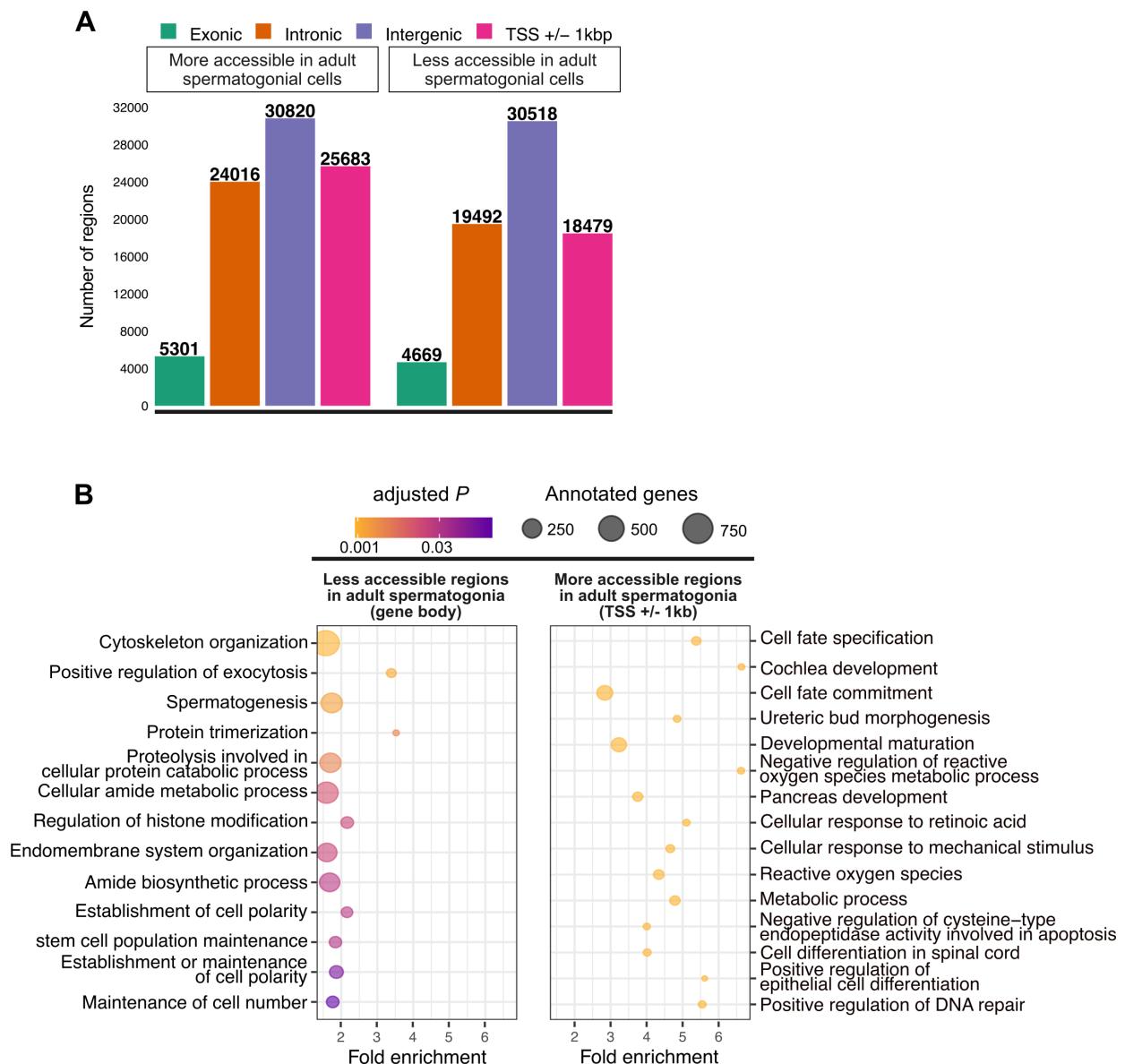
**Figure S 2-1 FACS of spermatogonial cells from pup testis leads to high enrichment of PLZF<sup>+</sup> cells.**

- (A) Representative dot plots of the sorting strategy for spermatogonial cell enrichment. Gating based on side scatter/forward scatter (SSC-A/FSC-A) and forward scatter – height/ forward scatter – area (FSC-H/FSC-A) was performed to exclude cell debris and cell clumps;
- (B) PLZF<sup>+</sup> cells are enriched following FACS, illustrated by immunocytochemistry on unsorted and sorted cell samples. Immunocytochemistry of PND15 unsorted and sorted testis cell suspension was performed by fixating the cells on poly-L-lysine coated slides. Cells were stained with anti-PLZF antibody (S19 clone, Active Motif) and VECTASHIELD (with DAPI) antifade mounting medium was used for mounting. Cells were visualized under a fluorescence

microscope and counted in 10 different fields of view / slide. The number of PLZF<sup>+</sup> and PLZF<sup>-</sup> cells from 10 different fields of view was averaged;

- (C) Heatmap of the expression profile of selected markers of spermatogonial and different testicular somatic cells extracted from the RNA-seq data on PND8, PND15 samples and on literature PND14 and PNW8 samples. Key genes for stem cell potential, stem and progenitor spermatogonia, and Leydig and Sertoli cells were chosen to assess the enrichment of spermatogonial cells in the sorted cell populations. Gene expression is represented in Log<sub>2</sub>CPM (counts per million).

**Figure S2**



**Figure S 2-2 Omni-ATAC profiles of PND15 adult spermatogonial cell samples and their genomic distribution.**

(A) Genomic distribution of the 158, 978 Omni-ATAC regions identified;

(B) Dot plots of top enriched GO biological processes for regions with increased chromatin accessibility in adult spermatogonia, within gene bodies and around transcription starting sites (TSSs) of nearby genes (TSS +/- 1kb). The size of dots indicates the number of genes in the term and the color of each dot corresponds to the adjusted *P* value of the term's enrichment.

## Figure S3

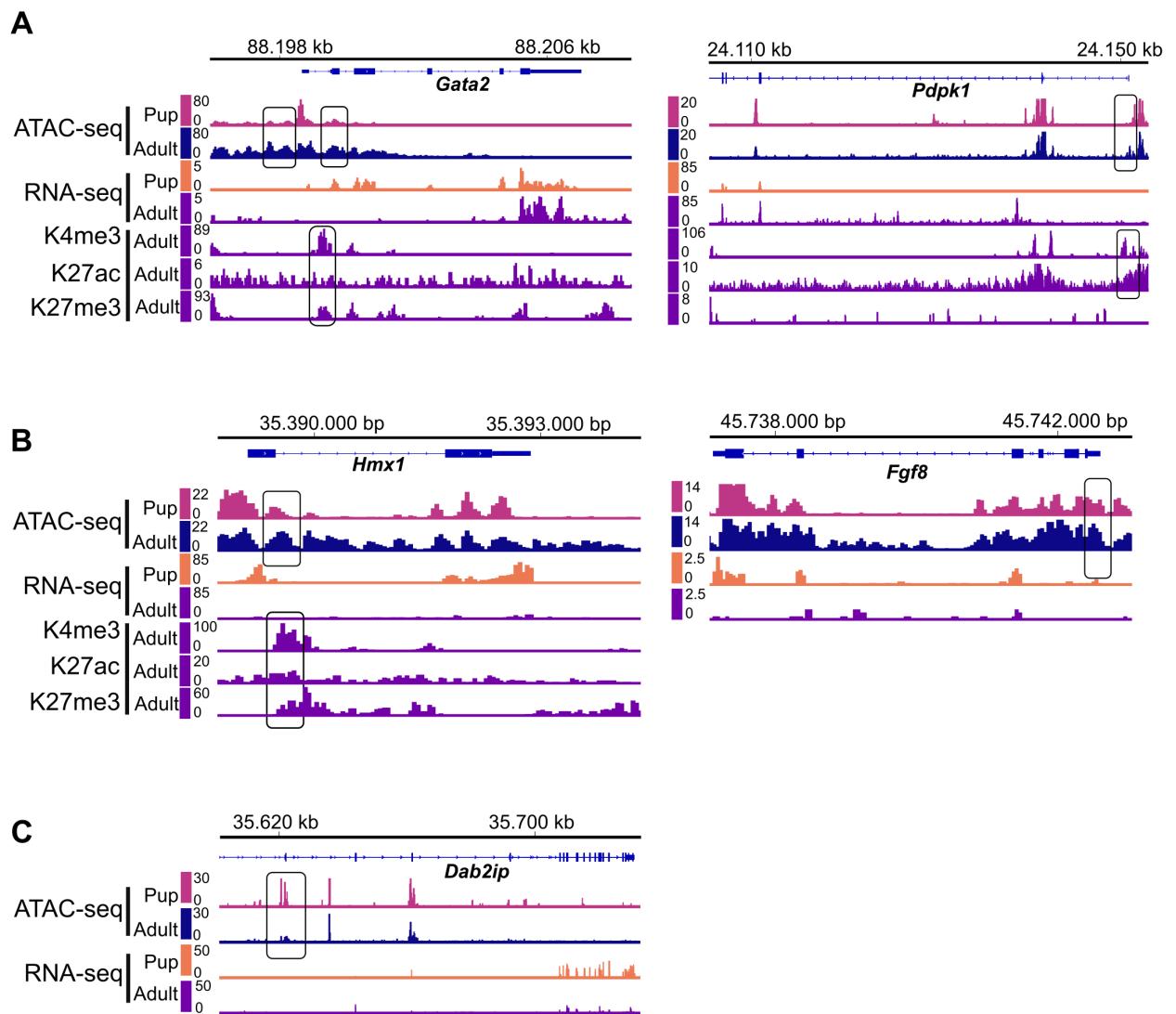
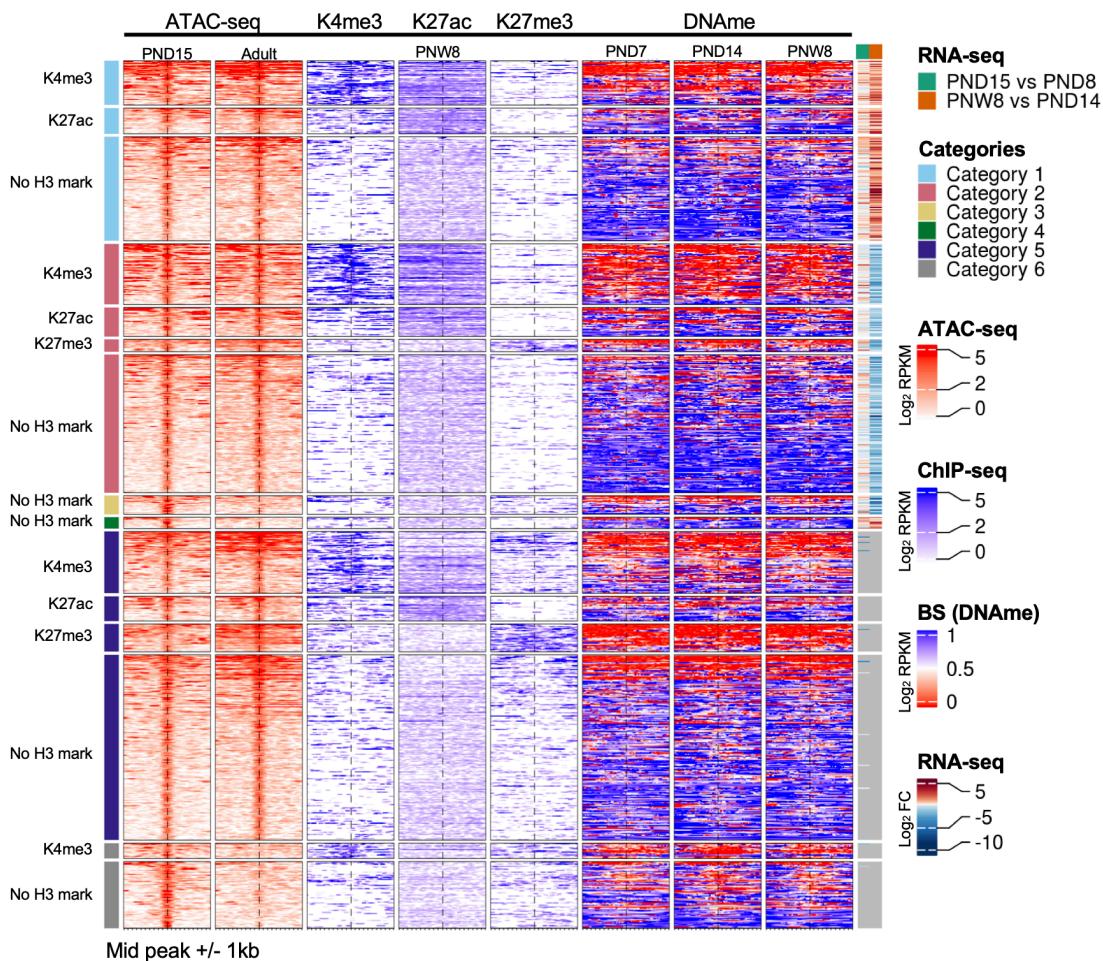


Figure S 2-3 Representative examples from Categories 1-3 resulted from the overlap of chromatin accessibility, gene expression and histone profiling datasets.

(A-C) Genomic snapshots from the Integrative Genomics Viewer (IGV, Broad Institute) of exemplary genes from Category 1 (*Gata2* and *Pdpk1*), Category 2 (*Hmx1* and *Fgf8*) and Category 3 (*Dab2ip*) showing relative abundance of: transcripts from RNA-seq, chromatin accessibility from ATAC-seq and enrichment of 3 different histone marks (H3K27ac, H3K4me3 and H3K27me3) from ChIP-seq. RNA-seq data corresponds to literature PND14 and adult (PNW8) spermatogonial cells and ATAC-seq data to PND15 and adult (PNW20) spermatogonial cells, respectively.

**Figure S4**



**Figure S 2-4 DNAme profiles of spermatogonial cells do not show major changes over the period of testis postnatal maturation.**

Enriched heatmaps showing the overlap between Category 1-4 regions, literature ChIP-seq data in PNW8 spermatogonia for H3K4me3, H3K27ac and H3K27me3, and DNAme data from BS in PND7, PND14 and PNW8 spermatogonia. For each of the Category 1-4 the following sub-categorization was applied:

- regions that are enriched for H3K4me3 (with or w/o H3K27ac and/or H3K27me3)
- regions that are enriched for H3K27ac (and lack both H3K4me3 and H3K27me3)
- regions that are enriched for H3K27me3 (and lack both H3K4me3 and H3K27ac)

Each line represents a peak region and the regions are ordered by the ATAC-seq signal. Mid-x-axis corresponds to the middle of a peak region and is extended to +/- 1 kbp. The color-key of the ATAC-seq, ChIP-seq and BS heatmaps represent ATAC-seq, ChIP-seq and BS signal, respectively. For RNA-seq, log<sub>2</sub> FC is shown from PND8 vs PND15 and PND14 vs PNW8 comparisons. For BS, the level of DNAme is between 0 and 1, with 0 representing a completely unmethylated locus and 1 a fully methylated locus, respectively.

## Figure S5

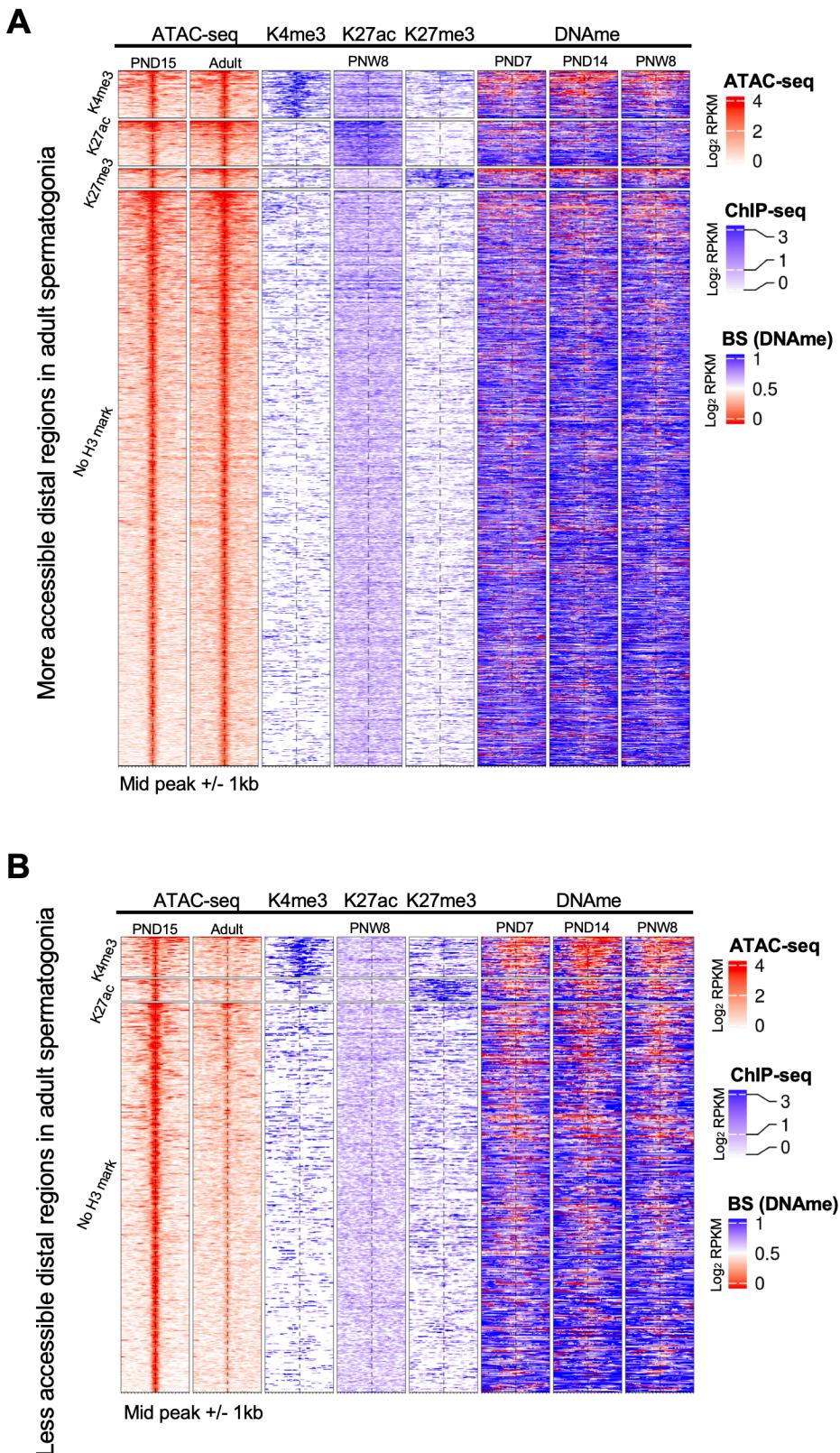


Figure S 2-5 Distinct chromatin profiles between PND15 and adult spermatogonia are present at distal regions across the genome. (legend continues on the next page)

(A-B) Enriched heatmaps showing the overlap between more accessible regions (A) and less accessible regions (B) situated in distal regions in spermatogonial cells, and literature ChIP-seq and DNAme data in PNW8 spermatogonia. The following sub-categorization was applied:

- regions that are enriched for H3K4me3 (with or w/o H3K27ac and/or H3K27me3)
- regions that are enriched for H3K27ac (and lack both H3K4me3 and H3K27me3)
- regions that are enriched for H3K27me3 (and lack both H3K4me3 and H3K27ac)

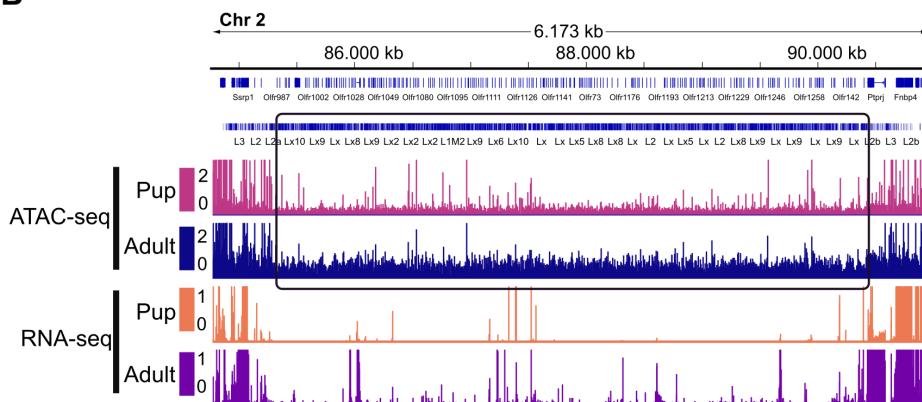
Each line represents a peak region and the regions are ordered by the ATAC-seq signal. Mid-x-axis corresponds to the middle of a peak region and is extended to +/- 1 kbp. The color-key of the ATAC-seq, ChIP-seq and BS heatmaps represent ATAC-seq, ChIP-seq and BS signal, respectively. For BS, the level of DNAme is between 0 and 1, with 0 representing a completely unmethylated locus and 1 a fully methylated locus, respectively.

**Figure S6**

**A**

Enriched TF motifs in less accessible ERVKs in adult spermatogonia			Enriched TF motifs in less accessible LINE L1s in adult spermatogonia		
TF	Enriched motif	% of target sequences	TF	Enriched motif	% of target sequences
PBX3		24.15%	FOXO1		15.02%
THAP11		22.25%	ZEB1		14.7%
ZNF143		27.21%	E2F3		13.79%
FOXL1		24.53%	KLF5		12.67%
NF-YA		25.88%	ZBTB17		12.74%

**B**



**Figure S 2-6 Differentially accessible TEs exhibit enriched TF motifs and correspond to regions nearby non-random gene families**

- (A) HOMER extracted consensus sequences for TF motifs enriched in the less accessible ERVK family and the more accessible LINE L1 family, respectively. Representative examples from the most enriched transcription factor families are depicted;
- (B) Genomic snapshots from the Integrative Genomics Viewer (IGV, Broad Institute) of a cluster of *Oifr* genes on Chr2 overlapping an increased density of LINE elements relative to the neighboring regions. The relative abundance of: transcripts from RNA-seq and chromatin accessibility from ATAC-seq are shown. RNA-seq data corresponds to literature PND14 and adult (PNW8) spermatogonial cells and ATAC-seq data to PND15 and adult (PNW20) spermatogonial cells, respectively.