Transcriptome and epigenome characterization of mouse spermatogonial cells reveals distinct chromatin regulatory landscapes in postnatal and adult testis

Irina Lazar-Contes[1,2,*], Deepak K. Tanwar[1,2,3,*], Pierre-Luc Germain[2,3], Niharika Gaur[1], Isabelle M. Mansuy[1,2,+]

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

Abstract

In mammals, spermatogonial cells exit quiescence immediately after birth and initiate a life-long equilibrium between proliferation and differentiation for the production of spermatogenic cells and functional sperm. Although it is generally accepted that the transcriptome of spermatogonial cells is different in postnatal life and adulthood, it is not known if the genome landscape at the chromatin level is also different across age. We characterized the transcriptome and chromatin landscape of spermatogonial cells by RNA-seq and ATAC-seq using enriched populations at different postnatal stages and in adulthood, and established a comprehensive multiparametric profiling building on published epigenetic datasets. Our results validate previously described changes in gene expression between postnatal and adult stages, and reveal novel transcriptomic transitions. We show that extensive chromatin remodeling takes place in adult spermatogonial cells that is accompanied by distinct transcription factor dynamics and histone modifications. The data identify chromatin and transcription factor dynamics at common repeat elements in postnatal and

adult spermatogonia. Overall, our results provide new knowledge about the transcriptional and epigenetic landscape of spermatogonial cells across postnatal stages until adulthood, and provide a valuable resource for germ cell studies.

Introduction

Spermatogonial cells are the initiators and supporting foundation of mammalian spermatogenesis. In mice, they become active at postnatal day (PND) 1 to 2 when they exit mitotic arrest and start cycling, in order to populate the basement membrane of the seminiferous tubules. During the first week after birth, a population of spermatogonial cells continue to proliferate and give rise to the undifferentiated A_{single} (A_s), A_{paired} (A_{pr}) and $A_{aligned}$ (A_{al}) spermatogonial pool. The remaining spermatogonial cells proceed into differentiation and form chains of daughter cells that become primary and secondary spermatocytes around PND 10 to 12. A second round of meiosis gives rise to haploid spermatids, which then mature to form adult spermatozoa capable of fertilization by PND 35 to 37 (Oatley and Brinster, 2006, 2012; De Rooij, 2017). Recent work has shown distinct epigenetic and transcriptional profiles in spermatogonial cell populations in postnatal compared to adult testis (Green et al., 2018; Hammoud et al., 2014, 2015; Hermann et al., 2018; Law et al., 2019). The pool of spermatogonial cells in early postnatal testis displays unique characteristics necessary for rapid establishment and expansion. In contrast, in adulthood the focus lies on the maintenance of a steady spermatogonial population which balances proliferation and differentiation capabilities to ensure adult sperm formation across life (Law and Oatley, 2020). However, little is known about chromatin remodeling and how it associates with transcriptional regulation of spermatogonial cells during the transition from postnatal to adult stage.

We characterized the transcriptome and chromatin accessibility landscape of enriched populations of mouse spermatogonial cells from postnatal and adult testis by RNA-seq and

ATAC-seq. The results identify chromatin transitions and their association with pathways important for stem cell proliferation and differentiation, and transcription factor dynamics between postnatal and adult spermatogonial cells. Furthermore, we suggest yet unknown regulatory roles for repeat elements in the germline. Our RNA-seq revealed specific transcriptional signatures of spermatogonial cells across postnatal testis development, mainly characterized by differences in cell cycle and proliferation, driven by dynamic gene expression of factors involved in RNA metabolism and DNA replication. Interestingly, neuronal development, DNA replication, and DNA and nucleosome packaging transcriptional programs all undergo a downregulation in the transition from postnatal to adult stages, whilst mitochondrial processes, electron transport chain and spermatogenesis associated genes show peak expression in adult spermatogonial cells. Chromatin accessibility profiling using OmniATAC-seq (Corces et al., 2017) revealed an extensive change, in particular increase in accessibility, in adult spermatogonia compared to PND15 cells. Furthermore, motif enrichment analysis identified candidate transcription factors (TFs) specifically enriched in the regions of differential accessibility, suggesting distinct regulatory mechanisms contributing to the differential chromatin organization between pup and adult spermatogonial cells. By combining our datasets with previously published chromatin immunoprecipitation sequencing (ChIP-seq) and whole genome bisulfite sequencing (WGBS) data from postnatal and adult spermatogonia (Hammoud et al., 2014, 2015), we were able to identify chromatin regions for which the change in accessibility was marked by distinct histone modifications and DNA methylation profiles, providing novel insight into the epigenetic programming associated with specific classes of differentially expressed genes. Lastly, by performing chromatin accessibility analysis specifically on repeat elements (REs), we revealed changes in accessibility at all main families of repeats in adult spermatogonial cells, including in

promoters of genes with differential expression, suggesting a potential role for REs in regulating gene expression between postnatal and adult stages.

Results

Fluorescence activated cell sorting highly enriches for spermatogonial cells from postnatal and adult mouse testis

To enrich for spermatogonial cells from postnatal and adult mice testis, we performed surface marker-based fluorescence activated-cell sorting (FACS) following the protocol previously established by Kubota et al (Kubota et al., 2003, 2004a, 2004b). To validate the enrichment of spermatogonial cells in our sorted populations, we performed immunocytochemistry using PLZF, a well-established marker of undifferentiated spermatogonia (Costoya et al., 2004). Following staining of unsorted and sorted cell fractions, we identified between 85-99% $PLZF^+$ cells in the sorted fractions compared to 3-6% $PLZF^+$ cells in the unsorted samples (Fig. S1A). To further confirm the high enrichment of spermatogonial cells in our sorted samples, we also compared expression of known pluripotency and spermatogonial stem cell markers between our sorted fractions and previously published datasets from Oct4-GFP and Id4-GFP transgenic mice (Garcia and Hofmann, 2012; Helsel et al., 2017) and found similar expression levels for key stem cell and undifferentiated spermatogonial markers (Fig. S1B). Noteworthy, our sorted populations still showed low expression of genes expressed in differentiating spermatogonial cells and scarce expression of somatic contaminants. Additionally, to exclude biological variability from our inferential analysis and detect statistically significant differences, we have included between 5 and 9 biological replicates for each individual timepoint in both RNA-seq and ATAC-seq analyses.

Dynamic transcriptome states characterize spermatogonial cells between early postnatal development and adult stage

Extensive work in rodents has underlined the unique characteristics of the first wave of spermatogenesis, including distinct transcriptome features of early postnatal spermatogonial cells (Law et al., 2019; Yoshida et al., 2006). Moreover, discernable differences between gene expression profiles of spermatogonia during the first two weeks of postnatal development have also been documented (Hammoud et al., 2015). To capture the early postnatal transcriptome transitions, and to obtain a comprehensive view of the gene expression programs in the developing testis, we performed RNA-seq on FACS sorted spermatogonial cells from 2 distinct postnatal stages (PND8 and PND15) and from adult 20 weeks old (PNW20) mouse testis. Pairwise comparisons revealed a high correlation between biological replicates at each of the three timepoints (Fig. S2), while principal component analysis (PCA*)* pointed to an age-dependent grouping of the samples, with postnatal spermatogonial cells more closely related and distanced from adult spermatogonia (Fig. 1A). Initially, we performed gene set enrichment analysis (GSEA) to identify distinct classes of over / under represented GO terms between PND8 and PND15 spermatogonial cells. Our analysis revealed a downregulation of genes involved in cell cycle and cell proliferation pathways, consistent with the increased proliferative and colonization capacities of spermatogonial cells specific to the first week of postnatal development, and essential for successful colonization of the testis basement membrane (Hammoud et al., 2015). Among other candidates, the most downregulated GO terms were RNA splicing, DNA replication, and ribosome biogenesis (Fig. S3A). GSEA between PND15 and adult spermatogonial cells revealed an upregulation of genes related to mitochondrial function, mRNA metabolism and RNA splicing (Fig. S3B). To better understand the relationship between the two transitions, from PND8 to PND15 and from PND15 to adult stage, we continued by performing unsupervised k-means clustering of all differentially expressed genes across the three stages,

which led to five clusters with distinct gene expression profiles (Fig. 1B). Over-representation analysis (ORA) revealed distinct sets of Gene Ontology (GO) terms enriched within each gene cluster (Fig. 1C). Genes with a marked downregulation in adult spermatogonia compared to postnatal stages were enriched for DNA packaging, protein-DNA complex assembly and nucleosome assembly (Cluster 1). A similar expression trajectory was detected for genes involved in GTPase-mediated signal transduction and activity, and histone and chromatin modifications (Cluster 2). Genes involved in synapse organization, synaptic vesicle exocytosis, axon guidance and modulation of chemical synaptic transmission were maximally expressed at PND15, followed by a downregulation in adult spermatogonia (Cluster 3). In contrast, mitochondrial function, genes involved in ribosome biogenesis and ncRNA processing (Cluster 4) displayed an upregulation, reaching peak expression in adult spermatogonial cells. Cluster 5, which showed the most elevated expression levels in adults, consisted of genes involved in the electron transport chain, spermatogenesis and ribonucleoprotein complex (Fig. 1C). Collectively, our RNA-seq data reveal the main transcriptional changes accompanying the transition from postnatal to adult spermatogonia, and support similar findings from recent single cell profiling studies of rodent and human spermatogonia (Green et al., 2018; Guo et al., 2017; Hermann et al., 2018; Law et al., 2019; Wang et al., 2018)

Spermatogonia in postnatal and adult testes have different expression profiles of signaling factors important for self-renewal and differentiation

Next, we focused on the expression of factors important for spermatogonial cell fate. Several key transcription factors involved in stem cell maintenance and proliferation exhibited highly dynamic gene expression profiles between postnatal and adult stages. The majority of genes known to be involved in spermatogonial stem cell maintenance and self-renewal (*Lhx1, Etv5, Id4, Cxcr4, Bcl6b, Pou5f1, Pax7*) showed a gradual downregulation

from PND8 to PND15 to adult stage (Fig. 1D). Signaling factors important for spermatogonial cell maintenance *in vitro,* such as LIF receptors (*Lifr*), showed a transient and moderate upregulation between PND8 and 15, followed by downregulation in adulthood, whereas FGF receptor genes (*Fgfr1* and *Fgfr2*) were maximally expressed at PND8, followed by a gradual decrease at PND15, and an even more pronounced decrease in adult spermatogonial cells (Fig. 1D). Self-renewal factors belonging to the TGFb family, *Tgfb1*, *Tgfb2* and *Tgfb3* were also significantly reduced in adult spermatogonial cells (Fig1. D). Altogether these findings indicate a switch in the stem cell maintenance and renewal strategies employed by spermatogonial cells between postnatal and adult stages, reinforcing previous findings suggesting a change from autocrine to mainly paracrine (relying on the niche) signaling in adult spermatogonial cells (Hammoud et al., 2015). One other class of genes with important roles in spermatogonial cell maintenance and proliferation is the testis-specific melanoma-associated antigen (MAGE) family, specifically the *Mage-a* and *Mage-b* subclasses, which were recently found to protect the adult germline against environmental stressors (Fon Tacer et al., 2019; Lee et al., 2020). Notably, our RNA-seq data revealed high levels of MAGE genes in adult spermatogonial cells, with all MAGE genes going through a significant upregulation in adult spermatogonial cells compared to postnatal stages (Fig. 1D). These findings may suggest that gene programs meant to ensure steady-state sperm production and protect against potentially damaging environmental fluctuations are preferentially upregulated and utilized in adult spermatogonial cells, compared to early postnatal development, when the emphasis falls on rapid and successful expansion and proliferation.

Chromatin accessibility transitions to a more open state from postnatal to adult spermatogonia

To investigate if the transcriptome changes we found between postnatal and adult spermatogonial cells are accompanied by changes in chromatin organization, we profiled chromatin accessibility of 6 biological replicates from PND15 and 5 biological replicates from adult spermatogonial cells using Omni-ATAC-seq. Omni-ATAC-seq assesses chromatin accessibility with a higher signal-to-noise ratio compared to the original ATAC-seq protocol, starting from as low as few hundred cells (Corces et al., 2017). Accessible regions were identified by peak-calling on the merged nucleosome-free fragments (NFF). Following the removal of lowly enriched regions, we included 158,978 peaks in downstream analyses (see Methods section for details). Most of the Tn5-accesible regions were intergenic (38%), in gene bodies (33%) and in proximity of gene transcription starting sites (+/-1 kbp from TSS, 28%) (Fig. S4). Differential accessibility analysis revealed 3212 differentially accessible regions between PND15 and adult spermatogonia, with the majority of regions showing a gain in accessibility at adult stage (Fig. 2A). Accessibility changes were localized predominantly to intergenic regions and gene bodies, whilst 15% of all differentially accessible regions resided +/- 1kbp from the TSS of a gene (Fig. 2B).

Next, to investigate the functional significance of the regions with differential accessibility between PND15 and adult spermatogonia, we performed GO analysis using the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010). Overall, regions of increased accessibility in adults were associated with cell fate and stem cell population maintenance, protein metabolism and RNA metabolic processes (Fig. 2C). Interestingly, when performing GREAT analysis at specific genomic locations where regions with increased accessibility where situated (intergenic, gene bodies, +/- 1kbp from TSS) we found that regions located in gene bodies (exons and introns) were enriched for GO terms related mainly to spermatogenesis, protein metabolism and stem cell population maintenance,

whilst regions close to, or overlapping with TSSs, were strongly enriched for developmental processes and tissue morphogenesis (Fig. 2D). Regions with decreased accessibility in adult spermatogonia featured an overabundance of terms related to embryonic development (Fig. 2C), with the majority of these regions located in intergenic spaces (Table S1). Collectively, our differential accessibility analysis and functional analysis of differentially accessible regions suggest that chromatin undergoes extensive remodeling from postnatal to adult stage, and that increased and decreased accessibility regions associate with distinct gene functions, depending on their genomic localization.

Transcription factor motif enrichment reveals potential regulators of chromatin accessibility from postnatal to adult spermatogonial cells

Previous work has underlined the importance of transcription factors (TFs) in establishing and maintaining distinct transcriptional signatures across the developmental trajectory of a cell population or subpopulation (Fushan et al., 2015; Shavlakadze et al., 2019). To predict which TF interactions may mediate the chromatin accessibility remodeling we detected between postnatal and adult spermatogonia, we performed motif enrichment analysis using the Hypergeometric Optimization of Motif EnRichment (HOMER) tool (Heinz et al., 2010a). In regions that showed an increased chromatin accessibility at adult stage, we identified 41 enriched TF motifs (q-value < 0.05) (Fig. 3A). The top candidate list (q-value<0.0001) was dominated by members of the Fos/ Jun family (FOS, FOSB, FOSL1 and FOSL2, JUN, JUNB and JUND) (Fig. 3B). Notably, almost all TFs with enriched motifs displayed stage-specific differences in mRNA levels as well: *Fos, Jun* and *Junb* were upregulated in adult spermatogonial cells, while the other FOS/JUN family TFs displayed a decrease in mRNA levels (Fig. 3A). 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 (Martínez-

Zamudio et al., 2020; Shaulian and Karin, 2002). Specific members of this complex, such as JUND and c-FOS have also been studied in the context of spermatogonial stem cell proliferation (He et al., 2008; Wang et al., 2018). Noteworthy, we identified binding sites for JUND and c-FOS in proximity of genes essential for stem-cell self-renewal (*Gata2* and *Psmd2*), spermatogonial cell differentiation (*Kit*), chromatin dynamics (*Chd2*) and growth factor adaptor proteins (*Frs2*) (Fig. 3E and Table S2). These genes also displayed differential expression between postnatal and adult stage, with *Psmd2* showing an increased expression, whilst *Gata2, Kit, Chd2* and *Frs2* had a decreased expression in adult spermatogonial cells (Fig. 3D). TF motif analysis using HOMER also revealed enriched binding sites for retinoic acid receptors such as RXRA and RAR (Fig 3A). Notably, among the regions of increased accessibility, we observed motifs for RXRA and RAR situated in proximity of chromatin remodeling factors *Satb1, Hdac4* and *Eed,* with all 3 genes displaying differential expression between postnatal and adult spermatogonial cells. (Fig 3D-E). These results suggest a role for nuclear receptors in the stage-specific chromatin remodeling that we found in adult spermatogonial cells, potentially via transcriptional regulation of HDAC4 and Polycomb Repressive Complex 2 (PRC2) core component EED. We also performed motif enrichment analysis for the more accessible chromatin regions situated in gene bodies, intergenic regions and +/- 1kbp from TSS, and identified several TF motifs specifically enriched in intergenic regions and not present in gene bodies such as the ubiquitously expressed NF-Y complex, NF-YA, NF-YB and NF-YC (Fig. 3C). Interestingly, in distal regions, NF-Y has been shown to facilitate a permissive chromatin conformation, and play an important role in the expression of core embryonic stem cell (ESC) pluripotency genes (Oldfield et al., 2014). Its enrichment in intergenic regions of increased chromatin accessibility could suggest a similar role for NF-Y trimers in adult spermatogonial cells. Furthermore, NF-YA/B motif enrichment

has also been found in regions of open chromatin in human spermatogonial cells (Guo et al., 2017), suggesting a potentially similar role for both rodent and human spermatogonia.

Although regions of increased chromatin accessibility predominate in adult spermatogonia compared to PND15, we have also identified TFs with enriched motifs in the regions of decreased accessibility (Fig. 3A). Interestingly, almost all of these TFs were not found in the regions of increased chromatin accessibility. 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. 3B). The gene expression levels of most of these TFs were also decreased in adult spermatogonial cells (Fig. 3A). FOXO1 is a pivotal regulator in the self-renewal and differentiation of spermatogonial stem cells in both pup and adult testis, acting via the PI3K-Akt signaling pathway (Chan et al., 2014; Goertz et al., 2011). 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 (Cheng et al., 2019; Law et al., 2019). Furthermore, motif enrichment in gene bodies and intergenic regions with decreased chromatin accessibility revealed a binary pattern for certain TFs, with motifs specifically enriched in intergenic regions mainly relating to developmental functions (FOXC1, FOXJ2, FOXM1, OTX2, LHX6) (Fig. 3C). This is consistent with the enrichment in embryonic development-related GO terms we identified in intergenic regions of decreased chromatin accessibility in adult spermatogonial cells (Fig. 2C). Taken together, our findings support major roles for FOS/JUN-transcription factors, RA receptors, NFYA/B and FOX family factors in the extensive chromatin remodeling taking place from postnatal to adult spermatogonial states.

Chromatin accessibility differences between postnatal and adult spermatogonia associate with distinct histone modifications and gene programs

Next, we investigated how the differentially accessible chromatin regions associate with the transcriptome changes that we have identified from our RNA-seq data. To aid the functional interpretation of our findings, we divided differentially accessible regions in distal - situated more than 2.5 kb from the TSS of a gene, and proximal - situated less than 2.5 kb from the TSS of a gene. Additionally, we made use of published ChIP-seq and WGBS data from Hammoud et al (Hammoud et al., 2014, 2015) and investigated histone mark (H3K4me3, H3K27ac, H3K27me3) distribution and DNAme patterns in the distal and proximal regions. Notably, ChIP-seq data was only available for adult spermatogonial cells (PNW8), whilst DNAme was available for PND7, 14 and adult stages.

Proximal regions with differential chromatin accessibility between postnatal and adult spermatogonia separated in 6 distinct categories, depending of the paired chromatin accessibility / gene expression profiles (Fig. S5A). The most abundant 2 categories consisted of regions with increased chromatin accessibility and upregulated nearby genes (Category 1) and increased chromatin accessibility and downregulated nearby genes (Category 2) at adult stage (Fig. 4A). Half of the regions in Category 1 also showed active H3K4me3/H3K27ac marking, or H3K4me3 and H3K27ac alone, and were generally depleted of H3K27me3 (Fig. 4B). Functional enrichment analysis using GREAT revealed that Category 1 regions associated with genes important for mitochondrial function and spermatogenesis (Fig. S5B). Exemplary genes in Category 1 included *Polg2*, *Tomm22* and *Tomm34*, mitochondrial factors reflective of increased mitochondrial activity, pyruvate cellular carriers *Slc25a41* and *Slc2a5*, and spermatogenic transposon silencer *Mael* (Fig. 4C and Table S3). Genes important in cell cycle regulation and stem cell maintenance were also present in Category 1, including DAZL-mediated *Psmd1* and *Psmd2* (Mikedis et al., 2020), and GDNF signaling receptor *Gfra2* (Hammoud et al., 2015) (Table S3). Category 2 included proximal regions with

increased chromatin accessibility and decreased expression of nearby genes in adult spermatogonia, indicative of active repression taking place (Fig. 4A). Notably, regions in Category 2 were enriched in binding motifs for FOS and JUND, suggesting a role for these TFs in mediating the transcriptional repression of the nearby genes (Fig. S5D). Additionally, a subset of Category 2 regions, which mainly associated with developmental genes, were marked by H3K27me3 together with H3K4me3, suggesting a poised state of these genes, which ensures silencing in the germline, but a competent transcriptional state in future embryos (Hammoud et al., 2014) (Fig. 4B). Examples include developmental factors *Satb1, Hmx1, Hoxb5* and *Gata2 (*Fig. 4C and Table S3). GO analysis using GREAT confirmed that regions in Category 2 were mainly associated with genes involved in regulation of developmental processes (Fig. S5C). Not surprisingly, DNAme levels in both Category 1 and 2 regions were inversely correlated with the enrichment for histone marks. Regions in Category 3 displayed decreased chromatin accessibility and a downregulation of nearby genes in adult spermatogonia and were relatively few in number compared to Categories 1 and 2 (Fig. 4A). Notably, most of the regions in this category did not show ChIP-signal for any of the 3 histone marks investigated. Category 4 comprised an even smaller number of proximal regions with decreased accessibility (<20 regions) at genes which were upregulated in adult SCs (Fig. 4A). Categories 5 and 6 consisted of regions with an increased and decreased chromatin accessibility respectively, for which nearby gene expression was not detected in PND15 and adult spermatogonial cells (Fig. S5A). Notably, DNAme profiles from postnatal to adult stages did not show drastic changes across any of the 4 categories of proximal regions, suggesting a relatively stable DNAme profile in the transition from postnatal to adult stage. Aside from proximal regions, we also identified regions of differential chromatin accessibility between PND15 and adult spermatogonia located distal from TSSs of genes (Fig. 5A). The majority of these regions did not show enrichment for any

of the evaluated histone modifications (Fig. 5B). Nevertheless, in distal regions of increased accessibility in adult spermatogonia for which ChIP-seq signal was present, most regions were enriched in H3K4m3, H3K27ac and coupled H3K4me3/H3K27ac, indicative of a potential distal enahncer role of these regions (Fig. 5B). Furthermore, TF motif analysis revealed enrichment of binding sites for nuclear receptors RXRA, RFX3, NR4A1 and RFX2 in these regions (Fig. 5C). In distal regions which were less accessible in adult spermatogonia, a subset of regions displayed H3K27me3 alone or together with H3K4me3 (Fig. 5B). Similar to proximal regions, DNAme levels were decreased in the presence of histone marks, and did not display major changes in the transition from postnatal to adult stage (Fig. 5B). Altogether, our data integration reveals distinct epigenetic landscapes associated with differentially expressed genes in postnatal and adult spermatogonial cells, which may be required to coordinate specific signaling pathways important for spermatogonial cell maintenance and spermatogenic potential.

Chromatin accessibility dynamics at repeat elements in adult spermatogonial cells compared to postnatal development

Repeat elements (REs) are under tight control in the germline, achieved through coordinated epigenetic mechanisms involving DNA methylation, chromatin silencing and the PIWI proteins-piRNA pathway (Deniz et al., 2019; Thompson et al., 2016a). To explore potential differences in REs regulation between postnatal and steady-state adult spermatogonial cells, we performed differential accessibility analysis of REs using our ATAC-seq data. Interestingly, we found that the transition from postnatal to adult stage spermatogonia is accompanied by chromatin accessibility changes at members of all major RE families including LTRs, LINEs and SINEs (Fig. 6A and Table S4). Most of the REs harboring changes in chromatin accessibility were situated in intergenic and intronic regions, and around 10% were located in proximity of a gene (+/- 1kbp from a TSS) (Fig. 6B). Next,

we investigated if the distribution of REs harboring changes in chromatin accessibility may relate to distinct gene functions. GO enrichment using GREAT revealed that differentially accessible LINE elements were mainly associated with genes involved in embryogenesis and nervous system development and synaptic function, differentially accessible LTRs associated with genes important for immune responses, while SINEs were strongly linked to genes involved in chromatin organization and transcriptional regulation (Fig. 6C). Moreover, for differentially accessible SINEs inserted in proximity of a gene, GREAT analysis identified a strong enrichment for RNA-related functions including transport, processing and rRNA metabolism (Fig. S6A). Similar gene functions were recently found for promoter-situated SINEs in mESCs, suggesting a conserved regulatory role for this repeat family in different types of stem cells (Lu et al., 2020). LTRs integrated in promoters may also contribute to the transcriptional regulation of nearby genes (Lowe et al., 2007; Thompson et al., 2016b). 25% of the differentially accessible LTR elements in our data resided in proximity of genes (Fig. S6B). For example, the promoter region of the lncRNA Lncenc1, an important regulator of pluripotency in mESCs (Fort et al., 2014; Sun et al., 2018), harbored two LTR repeats with decreased accessibility in adult spermatogonial cells, which correlated with a decreased expression of Lncenc1 in adult spermatogonia (Fig. 6D). Other examples of promoter-integrated LTRs which displayed changes in accessibility in adult spermatogonia included miR-21, an important regulator of stem-cell renewal capacity of spermatogonial cells (Niu et al., 2011), and miR-34, and miR-146, small RNAs preferentially expressed in spermatogonial stem cell-enriched populations compared to somatic cells (Niu et al., 2011) (Table S4).

To further understand the regulatory potential of differentially accessible REs, we performed TF motif enrichment analysis on LINEs, SINEs and LTRs which harbored changes in chromatin accessibility, and localized in proximity of a gene (+/- 1kbp from TSS). Differentially accessible LINE elements were highly enriched in binding sites for

transcription initiating factors KLF3/4/5, E2F family, spermatogonial stem cell maintenance factor SALL4 and FOX family members FOXJ2 and FOXP2 (Fig S7). LTRs displayed high enrichment for several TF motifs mainly important for embryonic development such as HXB8, HXA9, ALX1 and ZN143 (Fig S7). Differentially accessible SINEs comprised the highest number of enriched TF motifs, showing specific enrichment for DMRT1/B binding sites (Fig S7). By analyzing chromatin accessibility and TF motif enrichment specifically at REs, we were able to reveal extensive chromatin transitions taking place at numerous SINEs, LINEs and LTRs between postnatal and adult spermatogonial cells. Furthermore, the non-random distribution in the vicinity of genes with distinct functions, and their enrichment in TF binding sites suggests an important contribution of REs to shaping gene regulatory networks in spermatogonial cells during postnatal life.

Discussion

Mammalian spermatogenesis is critical for ensuring reproductive potential across life. As initiators of the spermatogenic cascade, spermatogonial cells are essential in germ cell proliferation and differentiation throughout postnatal life. By profiling the chromatin accessibility and transcriptome of mouse spermatogonia during postnatal life, we have uncovered extensive changes in transcriptome and chromatin structure which accompany the transition from postnatal to adult spermatogonia. We have identified a down-shift in cell-cycle and proliferation programs already from PND8 to PND15 and further to adult stage, while observing a gradual increase in expression levels of genes ensuring enhanced metabolic demand such as mitochondrial complexes, electron transport chains and ATP synthesis factors. Interestingly, when looking at specific genes, we observed an increased expression of the MAGE gene family, recently shown to protect the germline from environmental stressors (Fon Tacer et al., 2019; Lee et al., 2020). Overall, our transcriptome profiling suggests a decrease in spermatogonial proliferation capabilities in the adult testis, while upregulating

expression of gene networks able to respond to increased metabolic needs and protect the germline against fluctuating environmental conditions. We next speculated that chromatin accessibility changes between postnatal and adult stages could in part contribute to the observed transcriptional dynamics. Indeed, ATAC-seq revealed a reorganization of chromatin accessibility in intergenic, gene bodies and proximity of TSSs, mainly leading to more open chromatin regions in adult spermatogonia. Further analyses showed that these accessibility changes are related to distinct gene categories – with morphogenesis and developmental pathways strongly enriched in less open chromatin regions, while regions of increased chromatin openness were enriched for DNA repair pathways, stem cell maintenance and protein metabolic processes. Notably, we identified motif enrichment for distinct transcription factor families between regions of increased and decreased accessibility. AP-1 transcription factor binding sites were enriched in the more open chromatin, whilst FOX and ETS transcription factor motifs mainly mapped to regions which decreased in accessibility in adult spermatogonial cells. Such findings suggest an important contribution of these TF families to the chromatin reorganization underpinning the transition from postnatal to adult spermatogonial state.

Prior work in postnatal spermatogonial stem cell-enriched populations has revealed the genomic distribution of H3K4me3, H3K27ac, H3K27me3 and of DNAme patterns in the mouse (Hammoud et al., 2014). To better understand the spermatogonial cell epigenome transitions between postnatal and adult stages, we have used our ATAC-seq data from PND15 and adult spermatogonial cells and integrate it with the active and repressive histone modifications and DNAme data publicly available for mouse spermatogonial cells. In addition, we have integrated this multi-level analysis with our transcriptome data and determined the epigenetic content of the differentially expressed genes between PND15 and adult stage. Notably, we found comparable number of genes with increased chromatin

accessibility proximal to their TSS and an up or downregulation in expression levels, suggesting that active transcriptional repression may play an important role in the differences we observe between postnatal and adult spermatogonial cell transcriptomes. While the upregulated genes related to mitochondrial functions and reproductive processes and were mainly marked by the presence of activating H3K4me3 and H3K27ac, we found downregulated developmental genes in a poised state, suggested by the presence of coupled H3K4me3/ H3K27me3. Interestingly, previous work in adult sperm also revealed a poised state at promoters of developmental genes (Erkek et al., 2013; Jung et al., 2017), indicating that such poised status may already be established at the spermatogonial cell stage and maintained throughout spermatogenesis. Regarding DNAme changes, we observed stable methylation patterns in the transition from postnatal to adult stage. Nevertheless, DNAme levels were inversely correlated with the presence of histone modifications, both distal and proximal to TSSs of genes, suggesting a mutually exclusive role in regulating gene expression. By integrating a vast amount of epigenetic information with transcriptomic data, we were able to identify novel epigenetic mechanisms regulating the gene signaling programs with dynamic expression between postnatal and adult spermatogonial cells.

Lastly, by investigating chromatin accessibility specifically at REs, we were able to show that REs undergo changes in chromatin accessibility between postnatal and adult stages. We initially revealed a non-random distribution of the differentially accessible repeats, with LINEs and LTRs more likely to associate with genes important for specialized functions such as nervous system development and immune processes respectively, whereas SINEs were mainly associated with genes important for general cellular functions and were particularly enriched in promoters of genes responsible for RNA processing and metabolism. Furthermore, we identified promoters of genes important for spermatogonial stem cell renewal enriched in differentially accessible LTRs, supporting a novel role for repeat

elements in stage-specific regulation of genes with distinct functions. Similar roles for SINEs and L1 elements have been found in gene and chromatin regulation in mESCs (Lu et al., 2020; Thompson et al., 2016c). Additionally, the abundant enrichment in TF motifs we identified in the differentially accessible repeats points towards an important contribution of REs to the regulation of gene expression dynamics between postnatal and adult spermatogonial cells (Sundaram and Wysocka, 2020; Sundaram et al., 2014).

To our knowledge, this is the first time chromatin accessibility has been profiled in adult mouse spermatogonial cells. However, we cannot entirely exclude the influence of rare contaminating cells on the transcriptome and chromatin accessibility data interpretation. Therefore, we acknowledge that our findings are representative for the broader spermatogonial cell population in the testis, and that further refinement of our investigation will clarify how these stage-specific mechanisms are fine-tuned in certain spermatogonial subpopulations. Collectively, our cross-functional analyses reveal novel informational content of the spermatogonial chromatin states at different postnatal stages, integrate it with specific transcriptome changes, and provide a vast resource for further analysis of mouse postnatal spermatogonial cell programming.

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, and 20 weeks of age (PNW20). Testicular single-cell suspensions were prepared as previously described with slight modifications (Kubota et al., 2003, 2004a, 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 37ºC. The suspension was vigorously pipetted up and down 10 times and incubated again for 3 min at 37ºC. 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 4ºC. 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 adults, one adult male was used for each sample. The tunica was removed from testes, 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 37ºC 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 40µm strainer (Corning Life Sciences) and washed with HBSS. After centrifugation at 600g for 7 min at 4ºC, the cells were resuspended in PBS-S, layered on a 30% Percoll solution (Sigma-Aldrich) and

centrifuged at 600g for 8 min at 4ºC without braking. The top 2 layers (HBSS and Percoll) were removed and the cell pellets resuspended in PBS-S and used for sorting.

Spermatogonial cell enrichment by FACS

For pup testis, dissociated cells were stained with BV421-conjugated anti-β2M, biotin-conjugated anti-Thy-1 (53-2.1), and PE-conjugated anti-αv-integrin (RMV-7) antibodies. Thy-1 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-Thy-1 (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 4ºC 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 4ºC and using a 85µm 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 µL PBS-S, immediately pelleted by centrifugation and snap frozen in liquid N_2. Cells pellets were stored at -80ºC until DNA/RNA extraction. For RNA-seq on adult spermatogonia, 1000 cells from each animal were collected for processing using the SmartSeq2 protocol (Picelli et al., 2014). For OmniATAC-seq on PND15 spermatogonia, 25’000 cells were collected in a separate tube, pelleted by centrifugation and immediately processed using OmniATAC-seq library preparation protocol (Corces et al., 2017). For OmniATAC-seq on adult spermatogonia, 5000 cells from each animal were collected in a separate tube and further processed.

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 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 RNA-seq library preparation

For RNA-seq at PND8 and PND15, 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 long RNA sequencing libraries using the SMARTer® Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian (Takara Bio USA, Inc.) according to the manufacturer's instructions. For adult spermatogonia, 5000 cells from each sample were lysed in 10 uL buffer RLT Plus (Qiagen) immediately after sorting and processed following the SmartSeq2 protocol (Picelli et al., 2014). The number of biological replicates sequenced at each timepoint was 8 for PND8, 9 for PND15 and 6 for adult spermatogonia.

OmniATAC-seq library preparation and sequencing

Chromatin accessibility was profiled from PND15 and adult spermatogonial cells. The libraries were prepared according to the OmniATAC-seq 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 $MgCl_2$, 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 for 12 cycles 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 biological replicates were sequenced from PND15 and 5 biological replicates from adult spermatogonial cells.

High-throughput sequencing data analysis

Data availability: The datasets used in this study are available from the following GEO accessions: **GSE_____**, GSE49621, 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), Adult (6)
GSE_____	ATAC-Seq	PND15 (6), Adult (5)
GSE49621	ChIP-Seq (H3K4me3)	PNW8 (1)
GSE49621	ChIP-Seq (H3K27me3)	PNW8 (1)
GSE49621	ChIP-Seq (H3K27ac)	PNW8 (1)
GSE49623	BS-Seq	PND7 (1), PND14 (1), PNW8 (7)

RNA-seq

Quality control and alignment: Single-end (SE) sequencing was performed on the PND8, PND15, and adult samples on the Illumina HiSeq4000, Illumina HiSeq4000, and NovaSeq platform, respectively. PND8 raw data (FASTQ files) was merged from two runs for the RNA-Seq samples. The FASTQ files were assessed for the quality using FastQC (Andrews et al., 2012) (version 0.11.8). TrimGalore(Krueger, 2012) (version 0.6.2) was used to trim adapters, low-quality ends from reads with Phred score less than 30 (-q 30), and 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 the GENCODE (Harrow et al., 2012) annotation (version M18) with additional piRNA precursors, and transposable elements from repeat masker (concatenated by family), 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). Genes that had less than 15 mapped reads in at least 40% of the samples were filtered out. Surrogate variable analysis was performed using thesva package (Leek, 2014) (version 3.32.1) on variance-stabilized data from the DESeq2 package (Love et al., 2014) (version 1.24.0), identifying 3 variables which were then included in the model for differential expression. Normalization factors were obtained using TMM normalization (Robinson and Oshlack, 2010) from the edgeR package (Robinson et al., 2009) (version 3.28.1) and differential gene expression (DGE) analysis was performed using the limma-voom (Law et al., 2014) pipeline from the limma (Ritchie et al., 2015) (version 3.42.2), adjusting for the three surrogate variables.

ATAC-seq

Quality control, alignment, and peak calling: Paired-end (PE) sequencing was performed on the PND15 and adult samples on the Illumina HiSeq2500 platform. The FASTQ files were assessed for the quality using FastQC (Andrews et al., 2012) (version 0.11.8). Quality control (QC) was performed using TrimGalore (Krueger, 2012) (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 of the QC data was performed using Bowtie2 (Langmead and Salzberg, 2012) (version 2.3.5) with 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), minimum acceptable alignment score with respect to the read length (--score-min L,-0.4,-0.4) on a GRCm38. 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 center of the transposon binding event (--ATACshift), and filtered for reads with a high mapping quality (--minMappingQuality 30). Reads mapping to the mitochondrial chromosome and blacklisted regions, obtained on 10.10.2019 from http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/mm10mouse/mm10.blacklist.bed.gz and https://www.encodeproject.org/files/ENCFF547MET/@@download/ENCFF547MET_.bed.gz (merged together) 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 (NFF) were obtained using awk (awk 'function abs(v) {return v < 0 ? -v : v}; {if(abs($9)>=40 && abs($9)<=140 || $1 ~ /^@/){ print $0; }}'). Peak calling (identifying areas in a genome that have been enriched for transcription factors) on the NFF 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).

Downstream analysis: The downstream analysis was performed in R (R Core Team, 2020) (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., 2012) (version M18) transcript and/or the distance to the nearest transcription start site (https://gist.github.com/dktanwar/f873af6c4acd8c38ee928aa6f3d5b81d). The number of extended reads overlapping in the peaks regions was calculated using the csaw package (Lun and Smyth, 2015) (version 1.20.0). The peak regions that had less than 15 mapping 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 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). The peak regions which had an absolute log-fold change value of greater than or equal to 1 and an FDR of less than or equal to 0.05 were categorized as the differentially accessible regions (DARs). Functional analysis (Gene Ontology 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., 2010b) (version 4.11.1).

ChIP-seq

Quality control, alignment, and peak calling: ChIP-Seq SE data for adults were obtained from GEO accession GSE49621 (Hammoud et al., 2014) The FASTQ files were assessed for the 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). Aligned to the GRCm38 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) using Perl code (perl -ne "print if((/XM:i:[0-3][^0-9]/) || (/^@/));"), and reads with a low mapping quality (--minMappingQuality 30) or mapping to the mitochondrial chromosome or blacklisted regions (obtained on 10.10.2019 from http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/mm10mouse/mm10.blacklist.bed.gz and https://www.encodeproject.org/files/ENCFF547MET/@@download/ENCFF547MET_.bed.gz) 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).

WGBS

Quality control and alignment: BS paired end data for PND7, PND14, and PNW8 (adult) were obtained from GEO accession GSE49623 (Hammoud et al., 2014).The FASTQ files were assessed for the quality using FastQC (Andrews et al., 2012) (version 0.11.8). QC was performed using TrimGalore (Krueger, 2012) (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). The 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). The methylation information for individual cytocines was extracted using the bismark_methylation_extractor tool from the Bismark package (version 0.22.3).

Figures

All the figures in this study were generated using the ggplot2 (Wickham, 2016) ggpubr (Kassambara, 2020) EnrichedHeatmap (Gu et al., 2018) ggforce (Pedersen, 2020) and ComplexHeatmap (Gu et al., 2016) packages, and using base plotting functions in R (R Core Team, 2020).

Author contributions: ILC, DKT and IMM designed and conceived the study. ILC performed all RNA-seq, ICC and ATAC-seq experiments, and DKT analyzed the RNA-seq, ATAC-seq, ChIP-seq and WGBS data with significant support from PLG. NG prepared the adult spermatogonia RNA-seq libraries. ILC and DKT prepared all the figures. ILC interpreted the data with significant input from DKT, PLG and IMM. All authors read and accepted the final version of the manuscript.

Acknowledgments: We thank Francesca Manuella and Martin Roszkowski 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 Rodrigo Arzate for conceptual support and critical reading of the manuscript. We thank Catherine Aquino and Emilio Yángüez from the Functional Genomics Center (ETH/UZH) Zurich for support and advice with sequencing and library preparation. We thank Jon Oatley for advice and for providing the immunocytochemistry protocol for spermatogonial cells. We thank S3IT of UZH (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 Zürich, ETH Zürich, 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. Supplemental materials and all other data are available from the corresponding author upon reasonable request. The code for data analysis is available from https://github.com/mansuylab/SC_postnatal_adult.

References

Amezquita, R.A. (2018). marge: An {API} for Analysis of Motifs Using {HOMER} in R.

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., and Wingett, S. (2012). {FastQC}.

Chan, F., Oatley, M.J., Kaucher, A. V, Yang, Q.-E., Bieberich, C.J., Shashikant, C.S., and Oatley, J.M. (2014). Functional and molecular features of the Id4+ germline stem cell population in mouse testes. Genes Dev. *28*, 1351–1362.

Cheng, K., Chen, I.-C., Hale, B.J., Hermann, B.P., Geyer, C.B., Oatley, J.M., and McCarrey, J.R. (2019). Unique Epigenetic Programming Distinguishes Regenerative Spermatogonial Stem Cells in the Developing Mouse Testis. BioRxiv 674457.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Costoya, J.A., Hobbs, R.M., Barna, M., Cattoretti, G., Manova, K., Sukhwani, M., Orwig, K.E., Wolgemuth, D.J., and Pandolfi, P.P. (2004). Essential role of Plzf in maintenance of spermatogonial stem cells. Nat. Genet. *36*, 653–659.

Deniz, Ö., Frost, J.M., and Branco, M.R. (2019). Regulation of transposable elements by DNA modifications. Nat. Rev. Genet. *20*.

Erkek, S., Hisano, M., Liang, C.Y., Gill, M., Murr, R., Dieker, J., Schübeler, D., Vlag, J. Van Der, Stadler, M.B., and Peters, A.H.F.M. (2013). Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. Nat. Struct. Mol. Biol. *20*, 868–875.

Fon Tacer, K., Montoya, M.C., Oatley, M.J., Lord, T., Oatley, J.M., Klein, J., Ravichandran, R., Tillman, H., Kim, M., Connelly, J.P., et al. (2019). MAGE cancer-testis antigens protect the mammalian germline under environmental stress. Sci. Adv. *5*, eaav4832.

Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C.A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A., et al. (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. Nat. Genet. *46*, 558–566.

Fushan, A.A., Turanov, A.A., Lee, S.G., Kim, E.B., Lobanov, A. V., Yim, S.H., Buffenstein, R., Lee, S.R., Chang, K.T., Rhee, H., et al. (2015). Gene expression defines natural changes

in mammalian lifespan. Aging Cell *14*, 352–365.

Gapp, K., van Steenwyk, G., Germain, P.L., Matsushima, W., Rudolph, K.L.M., Manuella, F., Roszkowski, M., Vernaz, G., Ghosh, T., Pelczar, P., et al. (2018). Alterations in sperm long {RNA} contribute to the epigenetic inheritance of the effects of postnatal trauma. Mol. Psychiatry.

Garcia, T., and Hofmann, M.-C. (2012). Isolation of undifferentiated and early differentiating type A spermatogonia from Pou5f1-GFP reporter mice. Methods Mol. Biol. *825*, 31–44.

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

Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C., Gurczynski, S.J., Moore, B.B., et al. (2018). A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. Dev. Cell *46*, 651-667.e10.

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

Gu, Z., Eils, R., Schlesner, M., and Ishaque, N. (2018). {EnrichedHeatmap}: an R/Bioconductor package for comprehensive visualization of genomic signal associations. {BMC} Genomics *19*.

Guo, J., Grow, E.J., Yi, C., Goriely, A., Hotaling, J.M., and Cairns Correspondence, B.R. (2017). Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling and Metabolic Transitions during Human Spermatogonial Stem Cell Development.

Hammoud, S., Low, D.P., Yi, C., Carrell, D., Guccione, E., and Cairns, B. (2014). Chromatin and Transcription Transitions of Mammalian Adult Germline Stem Cells and Spermatogenesis. Cell Stem Cell *15*, 239–253.

Hammoud, S.S., Low, D.H.P., Yi, C., Lee, C.L., Oatley, J.M., Payne, C.J., Carrell, D.T., Guccione, E., and Cairns, B.R. (2015). Transcription and imprinting dynamics in developing postnatal male germline stem cells.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). {GENCODE}: The reference human genome annotation for The {ENCODE} Project. Genome Res. *22*, 1760–1774.

He, Z., Jiang, J., Kokkinaki, M., Golestaneh, N., Hofmann, M.-C., and Dym, M. (2008). Gdnf Upregulates c-Fos Transcription via the Ras/Erk1/2 Pathway to Promote Mouse Spermatogonial Stem Cell Proliferation. Stem Cells *26*, 266–278.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010a). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010b). Simple Combinations of Lineage-Determining

Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.

Helsel, A.R., Yang, Q.-E., Oatley, M.J., Lord, T., Sablitzky, F., and Oatley, J.M. (2017). ID4 levels dictate the stem cell state in mouse spermatogonia. Development *144*, 624–634.

Hermann, B.P., Cheng, K., Singh, A., Roa-De La Cruz, L., Mutoji, K.N., Chen, I.C., Gildersleeve, H., Lehle, J.D., Mayo, M., Westernströer, B., et al. (2018). The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. Cell Rep. *25*, 1650-1667.e8.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods *12*, 115–121.

Jung, Y.H., Sauria, M.E.G., Lyu, X., Cheema, M.S., Ausio, J., Taylor, J., and Corces, V.G. (2017). Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. Cell Rep. *18*, 1366–1382.

Kassambara, A. (2020). ggpubr: ggplot2 Based Publication Ready Plots.

Krueger, F. (2012). {TrimGalore}.

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

Kubota, H., Avarbock, M.R., and Brinster, R.L. (2003). Spermatogonial stem cells share some, but not all, phenotypic and functional characteristics with other stem cells. Proc. Natl. Acad. Sci. U. S. A. *100*, 6487–6492.

Kubota, H., Avarbock, M.R., and Brinster, R.L. (2004a). Culture Conditions and Single Growth Factors Affect Fate Determination of Mouse Spermatogonial Stem Cells. Biol. Reprod. *71*, 722–731.

Kubota, H., Avarbock, M.R., and Brinster, R.L. (2004b). Growth factors essential for self-renewal and expansion of mouse spermatogonial stem cells. Proc. Natl. Acad. Sci. U. S. A. *101*, 16489–16494.

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

Law, N.C., and Oatley, J.M. (2020). Developmental underpinnings of spermatogonial stem cell establishment. Andrology *8*, 852–861.

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

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

Lee, A., Klein, J., Tacer, K.F., Lord, T., Oatley, M., Oatley, J., Porter, S., Pruett-Miller, S., Tikhonova, E., Karamyshev, A., et al. (2020). Enhanced stress tolerance through reduction of

G3BP and suppression of stress granules. BioRxiv 2020.02.03.925677.

Leek, J.T. (2014). svaseq: removing batch effects and other unwanted noise from sequencing data. Nucleic Acids Res. *42*, e161--e161.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and {SAMtools}. Bioinformatics *25*, 2078–2079.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for {RNA}-seq data with {DESeq}2. Genome Biol. *15*.

Lowe, C.B., Bejerano, G., and Haussler, D. (2007). Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. Proc. Natl. Acad. Sci. U. S. A. *104*, 8005–8010.

Lu, J.Y., Shao, W., Chang, L., Yin, Y., Li, T., Zhang, H., Hong, Y., Percharde, M., Guo, L., Wu, Z., et al. (2020). Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. Cell Rep. *30*, 3296-3311.e5.

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

Martínez-Zamudio, R.I., Roux, P.F., de Freitas, J.A.N.L.F., Robinson, L., Doré, G., Sun, B., Belenki, D., Milanovic, M., Herbig, U., Schmitt, C.A., et al. (2020). AP-1 imprints a reversible transcriptional programme of senescent cells. Nat. Cell Biol. *22*, 842–855.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., de Rooij, D.G., and Page, D.C. (2020). DAZL mediates a broad translational program regulating expansion and differentiation of spermatogonial progenitors. Elife *9*.

Niu, Z., Goodyear, S.M., Rao, S., Wu, X., Tobias, J.W., Avarbock, M.R., and Brinster, R.L. (2011). MicroRNA-21 regulates the self-renewal of mouse spermatogonial stem cells. Proc. Natl. Acad. Sci. U. S. A. *108*, 12740–12745.

Oatley, J.M., and Brinster, R.L. (2006). Spermatogonial Stem Cells. pp. 259–282.

Oatley, J.M., and Brinster, R.L. (2012). The Germline Stem Cell Niche Unit in Mammalian Testes. Physiol. Rev. *92*, 577–595.

Oldfield, A.J., Yang, P., Conway, A.E., Cinghu, S., Freudenberg, J.M., Yellaboina, S., and Jothi, R. (2014). Histone-Fold Domain Protein NF-Y Promotes Chromatin Accessibility for Cell Type-Specific Master Transcription Factors. Mol. Cell *55*, 708–722.

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

Pedersen, T.L. (2020). ggforce: Accelerating ggplot2.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

R Core Team (2020). R: A Language and Environment for Statistical Computing.

Ramirez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). {deepTools}2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160--W165.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for {RNA}-sequencing and microarray studies. Nucleic Acids Res. *43*, e47--e47.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of {RNA}-seq data. Genome Biol. *11*, R25.

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

De Rooij, D.G. (2017). The nature and dynamics of spermatogonial stem cells.

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

Shaulian, E., and Karin, M. (2002). AP-1 as a regulator of cell life and death. Nat. Cell Biol. *4*, E131–E136.

Shavlakadze, T., Morris, M., Fang, J., Wang, S.X., Zhu, J., Zhou, W., Tse, H.W., Mondragon-Gonzalez, R., Roma, G., and Glass, D.J. (2019). Age-Related Gene Expression Signature in Rats Demonstrate Early, Late, and Linear Transcriptional Changes from Multiple Tissues. Cell Rep. *28*, 3263-3273.e3.

Sun, Z., Zhu, M., Lv, P., Cheng, L., Wang, Q., Tian, P., Yan, Z., and Wen, B. (2018). The Long Noncoding RNA Lncenc1 Maintains Naive States of Mouse ESCs by Promoting the Glycolysis Pathway. Stem Cell Reports *11*, 741–755.

Sundaram, V., and Wysocka, J. (2020). Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. Philos. Trans. R. Soc. B Biol. Sci. *375*.

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M.P., and Wang, T. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res. *24*, 1963–1976.

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

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

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

Wang, M., Liu, X., Chang, G., Chen, Y., An, G., Yan, L., Gao, S., Xu, Y., Cui, Y., Dong, J., et al. (2018). Single-Cell RNA Sequencing Analysis Reveals Sequential Cell Fate Transition during Human Spermatogenesis. Cell Stem Cell *23*, 599-614.e4.

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

Yang, Q.-E., Racicot, K.E., Kaucher, A. V, Oatley, M.J., and Oatley, J.M. (2013). MicroRNAs 221 and 222 regulate the undifferentiated state in mammalian male germ cells. Development *140*, 280–290.

Yoshida, S., Sukeno, M., Nakagawa, T., Ohbo, K., Nagamatsu, G., Suda, T., and Nabeshima, Y. (2006). The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. Development *133*, 1495–1505.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of {ChIP}-Seq ({MACS}). Genome Biol. *9*, R137.

Zuguang, G. (2020). No Title. Https://Github.Com/Jokergoo/RGREAT.

Figure Legends:

Figure 1. Transcriptome dynamics between postnatal and adult spermatogonial cells

(A) Principal Component Analysis (PCA) of the transcriptome of PND8 (n = 9), PND15 (n = 8) and adult (n = 6) spermatogonial cells;

(B) K-means clustering (k = 5) of genes with differential expression in postnatal versus adult spermatogonial cells (cluster 1: n = 325 genes, cluster 2: n = 2064 genes, cluster 3: n = 5131 genes, cluster 4: n = 3122 genes, cluster 5: n = 1386 genes). Each row depicts a gene and each column depicts a sample. Gene expression fold-change is calculated by subtracting $\log_2$ counts per million (CPM) of PND15 and adults from PND8. Samples are clustered using Ward's method and genes are ordered by "MDS_angle method" using seriation (R package);

(C) Dot plot of top 10 Gene Ontology (GO) Biological Processes (BP) terms enriched for each gene cluster selected based on the adjusted *P*. The dot color is representative of the number of genes annotated for each GO term;

(D) Line plots depicting expression patterns of key genes at PND8, PND15 and in adult spermatogonial cells.

Figure 2. Chromatin accessibility changes between PND15 and adult spermatogonial cells

(A) Volcano plot of differentially accessible ATAC regions (adjusted $P \leq 0.05$ and $\text{Log}_2\text{FC} \geq 1$) between PND15 and adult spermatogonial cells. More acc. in Adult = regions which gain in accessibility between PND15 and adult spermatogonial cells, Less acc. in Adult = regions which loose in accessibility between PND15 and adult spermatogonial cells;

(B) Bar plot illustrating the genomic distribution of differentially accessible ATAC regions in adult spermatogonia compared to PND15 cells;

(C) Dot plots of top enriched GO BP terms for regions with increased and decreased chromatin accessibility in adult spermatogonia. GO analysis was performed using GREAT. The size of the dot is representative of the number of genes annotated in the GO term, and the color corresponds to the adjusted *P*;

(D) Dot plots of top enriched GO BP for more accessible regions located in gene body and in proximity of TSS (+/- 1kb) in adult spermatogonia. GO analysis was performed using GREAT. The size of the dot is representative of the number of genes annotated in the GO term, and the color corresponds to the adjusted *P*.

Figure 3. Transcription factor dynamics as predicted by motif enrichment at differentially accessible regions in adult spermatogonial cells

(A) Dot plot of all transcription factor motifs enriched in regions of decreased and increased accessibility between PND15 and adult stage. Each dot corresponds to a

motif. The differential gene expression of the corresponding transcription factor between PND15 and adult spermatogonia was extracted from the RNA-seq data, and is shown as color coded Log2 fold-change. The size of the dot indicates the expression adjusted *P* of each motif. More acc. in Adult = regions which gain in accessibility between PND15 and adult spermatogonial cells, Less acc. in Adult = regions which loose in accessibility between PND15 and adult spermatogonial cells;

(B) HOMER extracted consensus sequences for each transcription factor motif. Representative examples from the most enriched transcription factor families are depicted;

(C) Dot plots of top transcription factor motifs enriched in differentially accessible chromatin regions situated in gene bodies or in intergenic areas of the genome. Each dot corresponds to a motif. The differential gene expression of the corresponding transcription factor between PND15 and adult spermatogonia was extracted from the RNA-seq data, and is shown as color coded Log2 fold-change. The size of the dot indicates the expression adjusted *P* of each motif;

(D) Heatmap of exemplary differentially expressed genes between PND15 and adult spermatogonia displaying motifs of JUND/cFOS and RFX1, RXRa and RXRg in their proximity. Each row depicts a gene and each column depicts a sample. Gene expression fold-change was calculated by subtracting $\log_2$CPM of Adult from PND15. Samples are clustered using Ward's method and genes are ordered by "PCA method" using seriation (R package);

(E) Genomic snapshots of exemplary genes *Gata2* and *Eed* showing relative abundance of: transcripts from RNA-seq, chromatin accessibility from ATAC-seq, 3 and TF motifs from HOMER analysis, in PND15 and adult spermatogonial cells.

Figure 4. Epigenetic profiling at proximal regions of genes expressed in PND15 and adult spermatogonial cells

(A) Enriched heatmaps showing proximal active regions of increased chromatin accessibility and increased gene expression (Category 1, n = 193), increased chromatin accessibility and decreased gene expression (Category 2, n = 246), decreased chromatin accessibility and decreased gene expression (Category 3, n = 39) and decreased chromatin accessibility and increased gene expression (Category 4, n = 17). Each line represents a peak region and the regions are ordered by the ATAC signal. Mid-x-axis corresponds to the middle of a peak region and is extended to +/- 2.5 kbp. The color-key of the ATAC heatmap represents ATAC signal. For RNA-seq, gene expression fold-change is calculated by subtracting log2 CPM of PND15 and Adults from PND8.

(B) Enriched heatmaps showing Category 1 regions without any H3 mark (n = 114), Category 1 regions with H3K4me3 (n = 48), Category 1 regions with H3K27ac (n = 29), Category 2 regions without any H3 mark (n = 138), Category 2 regions with H3K27me3 (n = 13), Category 2 regions with H3K4me3 (n = 63), Category 2 regions

with H3K27ac (n = 32), Category 3 regions without any H3 mark (n = 25). Each line represents a peak region and the regions are ordered by the ATAC signal. Mid-x-axis corresponds to the middle of a peak region and is extended to +/- 2.5 kbp. The color-key of the ATAC and ChIP heatmaps represent ATAC and ChIP-seq signal respectively. For WGBS the color key shows the percentage of methylation. For RNA-seq, gene expression fold-change is calculated by subtracting log2 CPM of PND15 and Adults from PND8;

(C) Genomic snapshots of exemplary genes from Category 1 (Mael) and Category 2 (Gata2 and Hmx1) showing relative abundance of: transcripts from RNA-seq, chromatin accessibility from ATAC-seq, 3 different histone marks (H3K27ac, H3K4me3 and H3K27me3) from ChIP-seq and TF motifs from HOMER analysis, in PND15 and adult spermatogonial cells.

Figure 5. Epigenetic profiling of distal regions in PND15 and adult spermatogonial cells

(A) Enriched heatmaps showing distal regions of decreased chromatin accessibility (Less acc.; n = 626) and increased chromatin accessibility (More acc.; n = 1757). Each line represents a peak region and the regions are ordered by the ATAC signal. Mid-x-axis corresponds to the middle of a peak region and is extended to +/- 2.5 kbp. The color-key of the ATAC heatmap represents ATAC signal;

(B) Enriched heatmaps showing distal regions of decreased chromatin accessibility and no H3 mark (n = 533), less accessibility with H3K27me3 (n = 30), less chromatin accessibility with H3K4me3 (n = 55), increased chromatin accessibility and no H3 mark (n = 1472), increased chromatin accessibility with H3K27ac (n = 115), increased chromatin accessibility with H3K27me3 (n = 49), and increased chromatin accessibility with H3K4me3 (n = 121). Each line represents a peak region and the regions are ordered by the ATAC signal. Mid-x-axis corresponds to the middle of a peak region and is extended to +/- 2.5 kbp. The color-key of the ATAC and ChIP heatmaps represents ATAC and ChIP-seq signal, respectively. For WGBS the color key shows the percentage of methylation;

(C) HOMER extracted consensus sequences for each transcription factor motif. Representative examples from the most enriched transcription factor families are depicted.

Figure 6. Chromatin accessibility profiling at repeat elements in PND15 and adult spermatogonial cells

(A) Volcano plots of repeat elements with differential chromatin accessibility between PND15 and adult spermatogonia (adjusted $P \leq 0.05$ and $Log_2FC \geq 1$). Illustrated are the 3 most abundant repeat families detected in spermatogonial cells (LINEs, LTRs and SINEs). Less accessible repeat elements in adults are represented by red points and more accessible repeat elements in adults are represented by blue points;

(B) Bar plot illustrating the genomic distribution of differentially accessible repeat elements between PND15 and adult spermatogonial cells;

(C) Dot plots of top enriched GO BP terms for differentially accessible LINEs, LTRs and SINEs. GO analysis was performed using GREAT. The size of the dot is representative of the number of genes annotated in the GO term, and the color corresponds to the adjusted *P*;

(D) Genomic snapshots of exemplary gene Lncenc1 showing relative abundance of: transcripts from RNA-seq, chromatin accessibility from ATAC-seq, 3 different histone marks (H3K27ac, H3K4me3 and H3K27me3) from ChIP-seq in PND15 and adult spermatogonial cells. LTRs and SINEs were extracted using Repeat Masker.

Figure 1

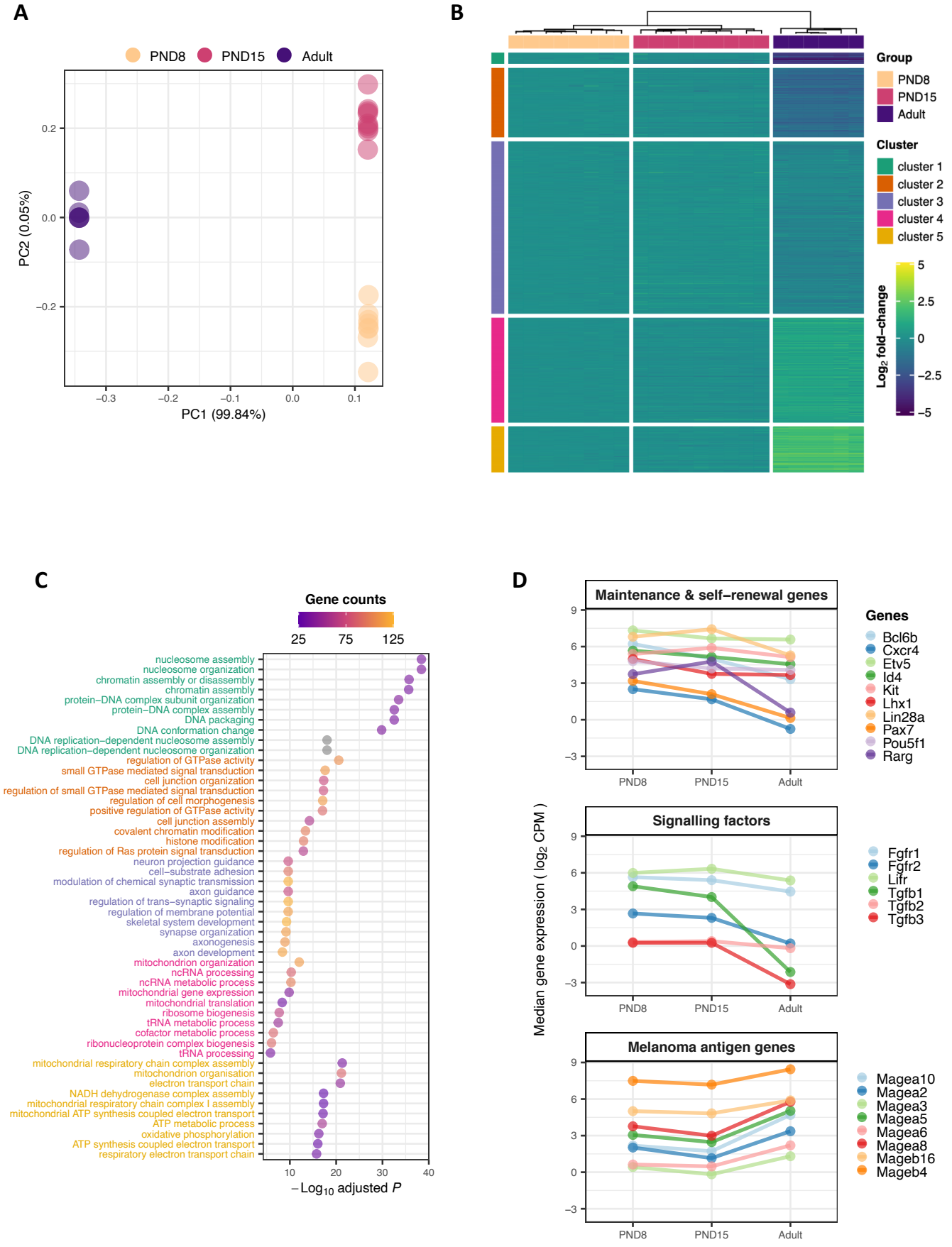

A
PND8
PND15
Adult
PC2 (0.05%)
PC1 (99.84%)
B
Group
PND8
PND15
Adult
Cluster
cluster 1
cluster 2
cluster 3
cluster 4
cluster 5
Log2 fold−change
C
Gene counts
25
75
125
nucleosome assembly
nucleosome organization
chromatin assembly or disassembly
chromatin assembly
protein−DNA complex subunit organization
protein−DNA complex assembly
DNA packaging
DNA conformation change
DNA replication−dependent nucleosome assembly
DNA replication−dependent nucleosome organization
regulation of GTPase activity
small GTPase mediated signal transduction
cell junction organization
regulation of small GTPase mediated signal transduction
regulation of cell morphogenesis
positive regulation of GTPase activity
cell junction assembly
covalent chromatin modification
histone modification
regulation of Ras protein signal transduction
neuron projection guidance
cell−substrate adhesion
modulation of chemical synaptic transmission
axon guidance
regulation of trans−synaptic signaling
regulation of membrane potential
skeletal system development
synapse organization
axonogenesis
axon development
mitochondrion organization
ncRNA processing
ncRNA metabolic process
mitochondrial gene expression
mitochondrial translation
ribosome biogenesis
tRNA metabolic process
cofactor metabolic process
ribonucleoprotein complex biogenesis
tRNA processing
mitochondrial respiratory chain complex assembly
mitochondrion organisation
electron transport chain
NADH dehydrogenase complex assembly
mitochondrial respiratory chain complex I assembly
mitochondrial ATP synthesis coupled electron transport
ATP metabolic process
oxidative phosphorylation
ATP synthesis coupled electron transport
respiratory electron transport chain
−Log10 adjusted P
D
Maintenance & self−renewal genes
Signalling factors
Melanoma antigen genes
Median gene expression (log2 CPM)
PND8
PND15
Adult
Genes
Bcl6b
Cxcr4
Etv5
Id4
Kit
Lhx1
Lin28a
Pax7
Pou5f1
Rarg
Fgfr1
Fgfr2
Lifr
Tgfb1
Tgfb2
Tgfb3
Magea10
Magea2
Magea3
Magea5
Magea6
Magea8
Mageb16
Mageb4

Figure 2

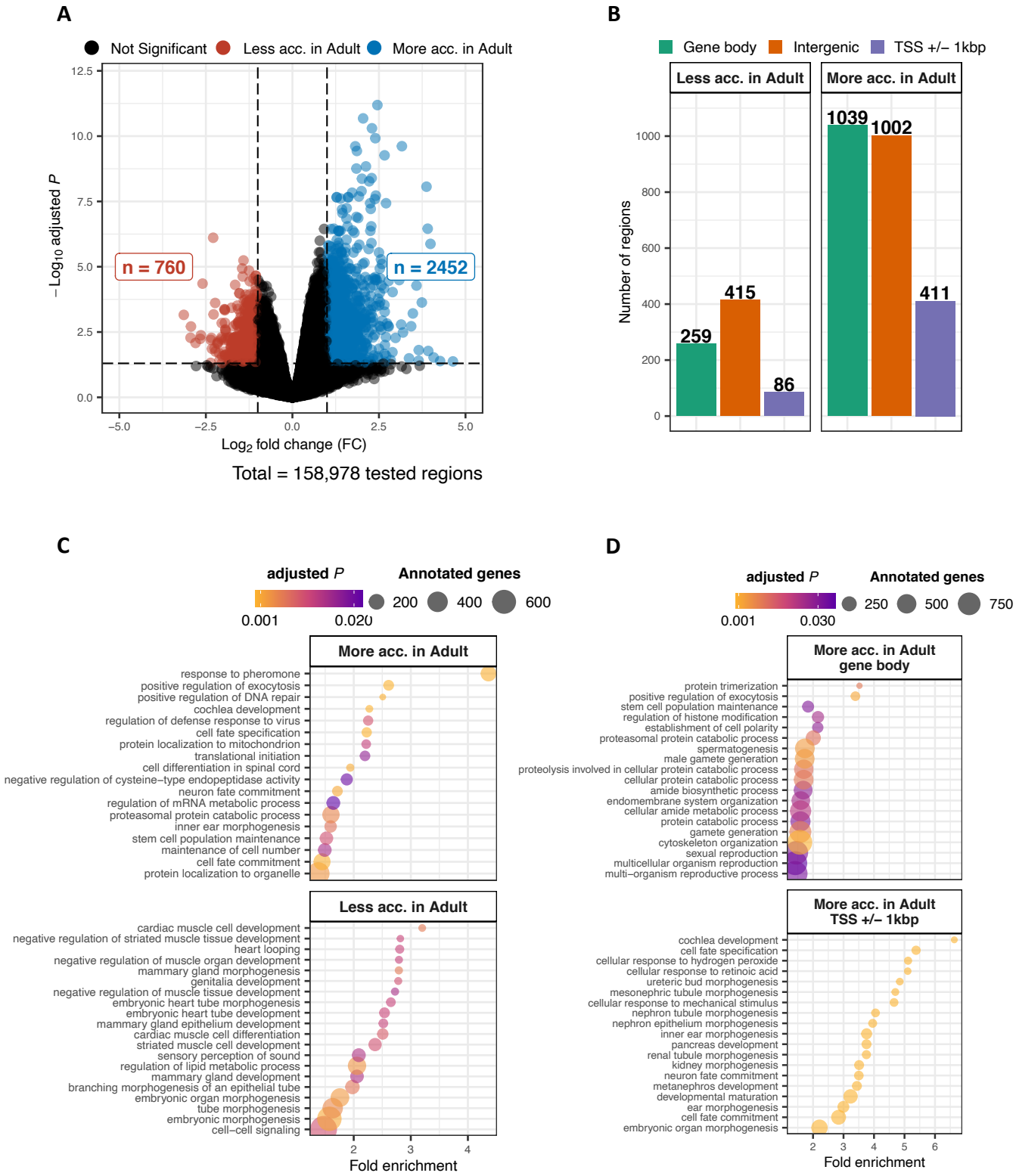

A
Not Significant
Less acc. in Adult
More acc. in Adult
n = 760
n = 2452
−Log10 adjusted P
Log2 fold change (FC)
Total = 158,978 tested regions
B
Gene body
Intergenic
TSS +/− 1kbp
Less acc. in Adult
More acc. in Adult
Number of regions
259
415
86
1039
1002
411
C
adjusted P
Annotated genes
0.001
0.020
More acc. in Adult
response to pheromone
positive regulation of exocytosis
positive regulation of DNA repair
cochlea development
regulation of defense response to virus
cell fate specification
protein localization to mitochondrion
translational initiation
cell differentiation in spinal cord
negative regulation of cysteine−type endopeptidase activity
neuron fate commitment
regulation of mRNA metabolic process
proteasomal protein catabolic process
inner ear morphogenesis
stem cell population maintenance
maintenance of cell number
cell fate commitment
protein localization to organelle
Less acc. in Adult
cardiac muscle cell development
negative regulation of striated muscle tissue development
heart looping
negative regulation of muscle organ development
mammary gland morphogenesis
genitalia development
negative regulation of muscle tissue development
embryonic heart tube morphogenesis
embryonic heart tube development
mammary gland epithelium development
cardiac muscle cell differentiation
striated muscle cell development
sensory perception of sound
regulation of lipid metabolic process
mammary gland development
branching morphogenesis of an epithelial tube
embryonic organ morphogenesis
tube morphogenesis
embryonic morphogenesis
cell−cell signaling
Fold enrichment
D
adjusted P
Annotated genes
0.001
0.030
More acc. in Adult gene body
protein trimerization
positive regulation of exocytosis
stem cell population maintenance
regulation of histone modification
establishment of cell polarity
proteasomal protein catabolic process
spermatogenesis
male gamete generation
proteolysis involved in cellular protein catabolic process
cellular protein catabolic process
amide biosynthetic process
endomembrane system organization
cellular amide metabolic process
protein catabolic process
gamete generation
cytoskeleton organization
sexual reproduction
multicellular organism reproduction
multi−organism reproductive process
More acc. in Adult TSS +/− 1kbp
cochlea development
cell fate specification
cellular response to hydrogen peroxide
cellular response to retinoic acid
ureteric bud morphogenesis
mesonephric tubule morphogenesis
cellular response to mechanical stimulus
nephron tubule morphogenesis
nephron epithelium morphogenesis
inner ear morphogenesis
pancreas development
renal tubule morphogenesis
kidney morphogenesis
neuron fate commitment
metanephros development
developmental maturation
ear morphogenesis
cell fate commitment
embryonic organ morphogenesis
Fold enrichment

Figure 3

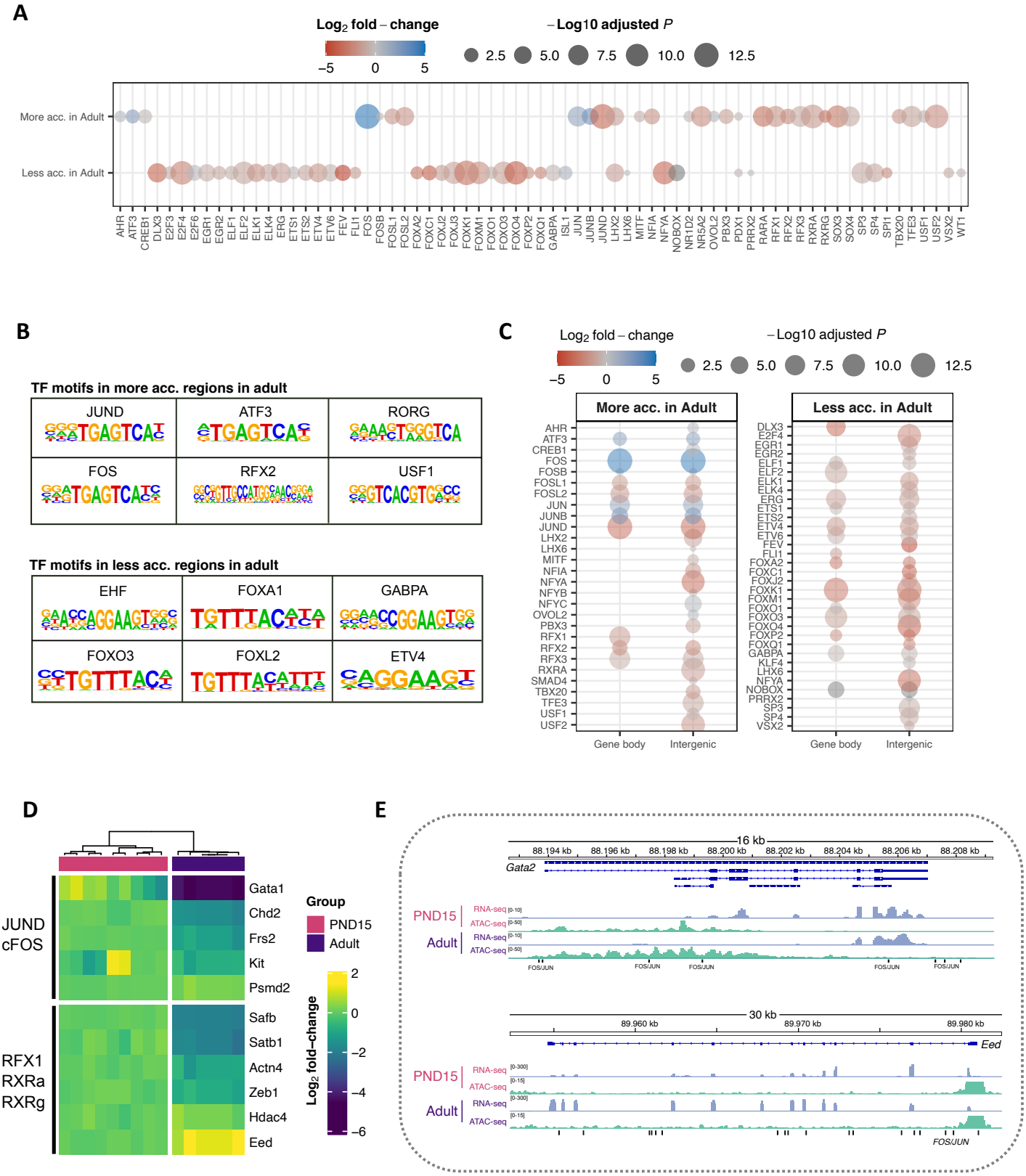

A
Log2 fold – change
–5 0 5
–Log10 adjusted P
2.5 5.0 7.5 10.0 12.5
More acc. in Adult
Less acc. in Adult
AHR ATF3 CREB1 DLX3 E2F3 E2F4 E2F6 EGR1 EGR2 ELF1 ELF2 ELK1 ELK4 ERG ETS1 ETS2 ETV4 ETV6 FEV FLI1 FOS FOSB FOSL1 FOSL2 FOXA2 FOXC1 FOXJ2 FOXJ3 FOXK1 FOXM1 FOXO1 FOXO3 FOXO4 FOXP2 FOXQ1 GABPA ISL1 JUN JUNB JUND LHX2 LHX6 MITF NFIA NFYA NOBOX NR1D2 NR5A2 OVOL2 PBX3 PDX1 PRRX2 RARA RFX1 RFX2 RFX3 RXRA RXRG SOX3 SOX4 SP3 SP4 SPI1 TBX20 TFE3 USF1 USF2 VSX2 WT1
B
TF motifs in more acc. regions in adult
JUND
ATF3
RORG
FOS
RFX2
USF1
TF motifs in less acc. regions in adult
EHF
FOXA1
GABPA
FOXO3
FOXL2
ETV4
C
Log2 fold – change
–5 0 5
–Log10 adjusted P
2.5 5.0 7.5 10.0 12.5
More acc. in Adult
Less acc. in Adult
AHR ATF3 CREB1 FOS FOSB FOSL1 FOSL2 JUN JUNB JUND LHX2 LHX6 MITF NFIA NFYA NFYB NFYC OVOL2 PBX3 RFX1 RFX2 RFX3 RXRA SMAD4 TBX20 TFE3 USF1 USF2
DLX3 E2F4 EGR1 EGR2 ELF1 ELF2 ELK1 ELK4 ERG ETS1 ETS2 ETV4 ETV6 FEV FLI1 FOXA2 FOXC1 FOXJ2 FOXK1 FOXM1 FOXO1 FOXO3 FOXO4 FOXP2 FOXQ1 GABPA KLF4 LHX6 NFYA NOBOX PRRX2 SP3 SP4 VSX2
Gene body
Intergenic
D
JUND
cFOS
RFX1
RXRa
RXRg
Gata1
Chd2
Frs2
Kit
Psmd2
Safb
Satb1
Actn4
Zeb1
Hdac4
Eed
Group
PND15
Adult
Log2 fold–change
2 0 –2 –4 –6
E
16 kb
88.194 kb 88.196 kb 88.198 kb 88.200 kb 88.202 kb 88.204 kb 88.206 kb 88.208 kb
Gata2
PND15
Adult
RNA-seq
ATAC-seq
FOS/JUN
30 kb
89.960 kb 89.970 kb 89.980 kb
Eed

Figure 4

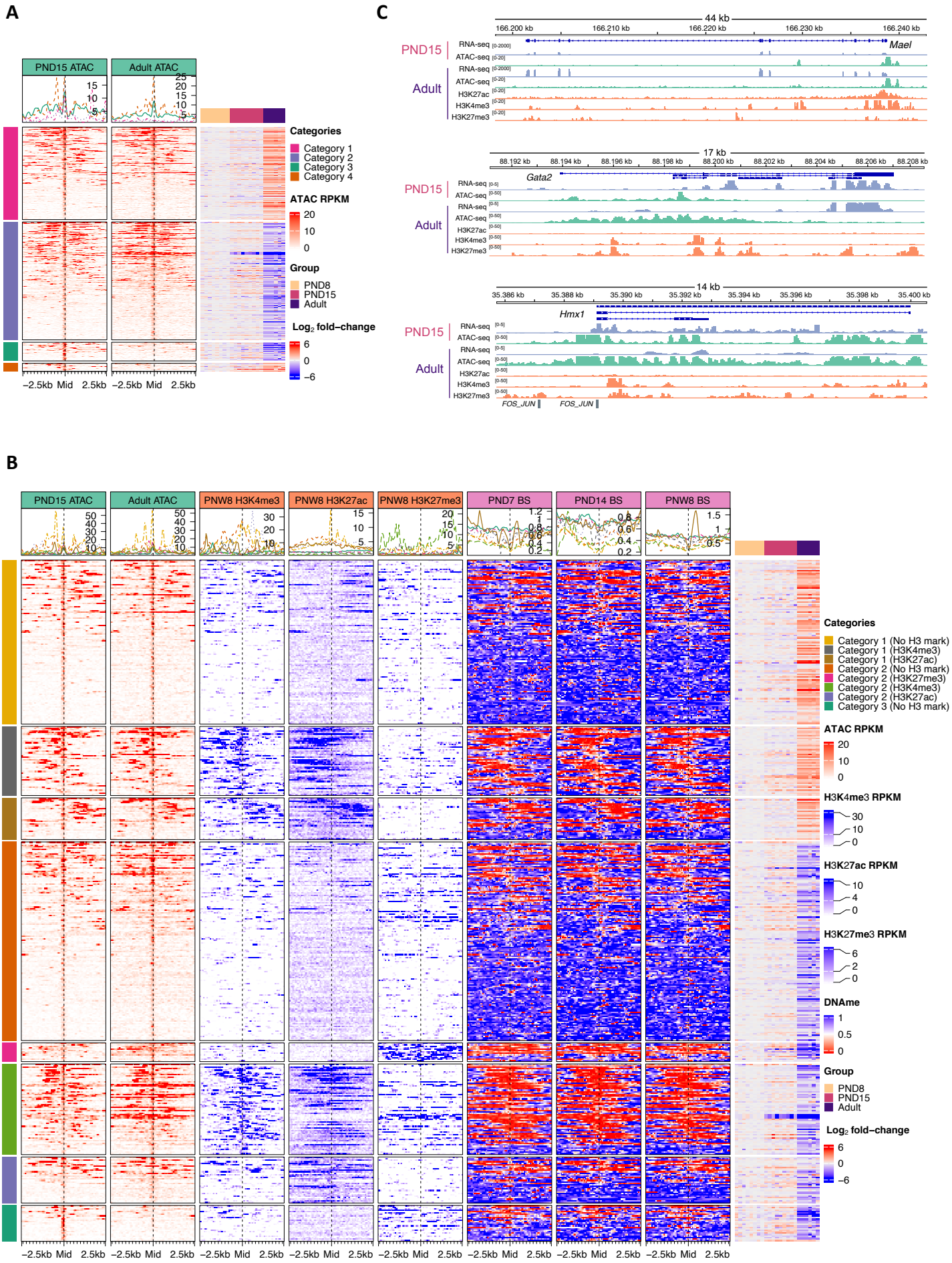

A
PND15 ATAC
Adult ATAC
Categories
Category 1
Category 2
Category 3
Category 4
ATAC RPKM
Group
PND8
PND15
Adult
Log2 fold-change
−2.5kb Mid 2.5kb
C
44 kb
Mael
17 kb
Gata2
14 kb
Hmx1
FOS_JUN
PND15
Adult
RNA-seq
ATAC-seq
H3K27ac
H3K4me3
H3K27me3
B
PND15 ATAC
Adult ATAC
PNW8 H3K4me3
PNW8 H3K27ac
PNW8 H3K27me3
PND7 BS
PND14 BS
PNW8 BS
Category 1 (No H3 mark)
Category 1 (H3K4me3)
Category 1 (H3K27ac)
Category 2 (No H3 mark)
Category 2 (H3K27me3)
Category 2 (H3K4me3)
Category 2 (H3K27ac)
Category 3 (No H3 mark)
H3K4me3 RPKM
H3K27ac RPKM
H3K27me3 RPKM
DNAme

Figure 5

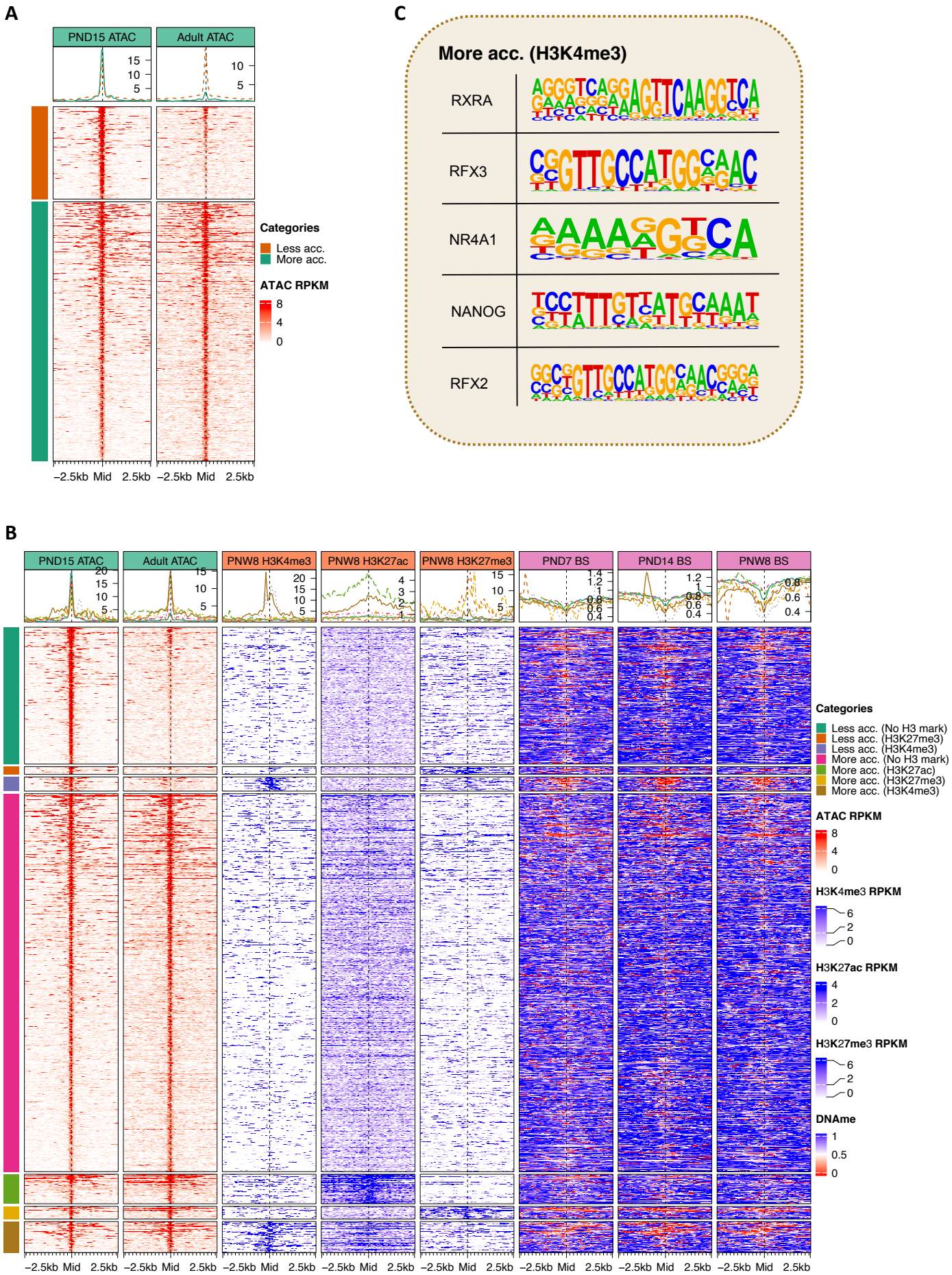
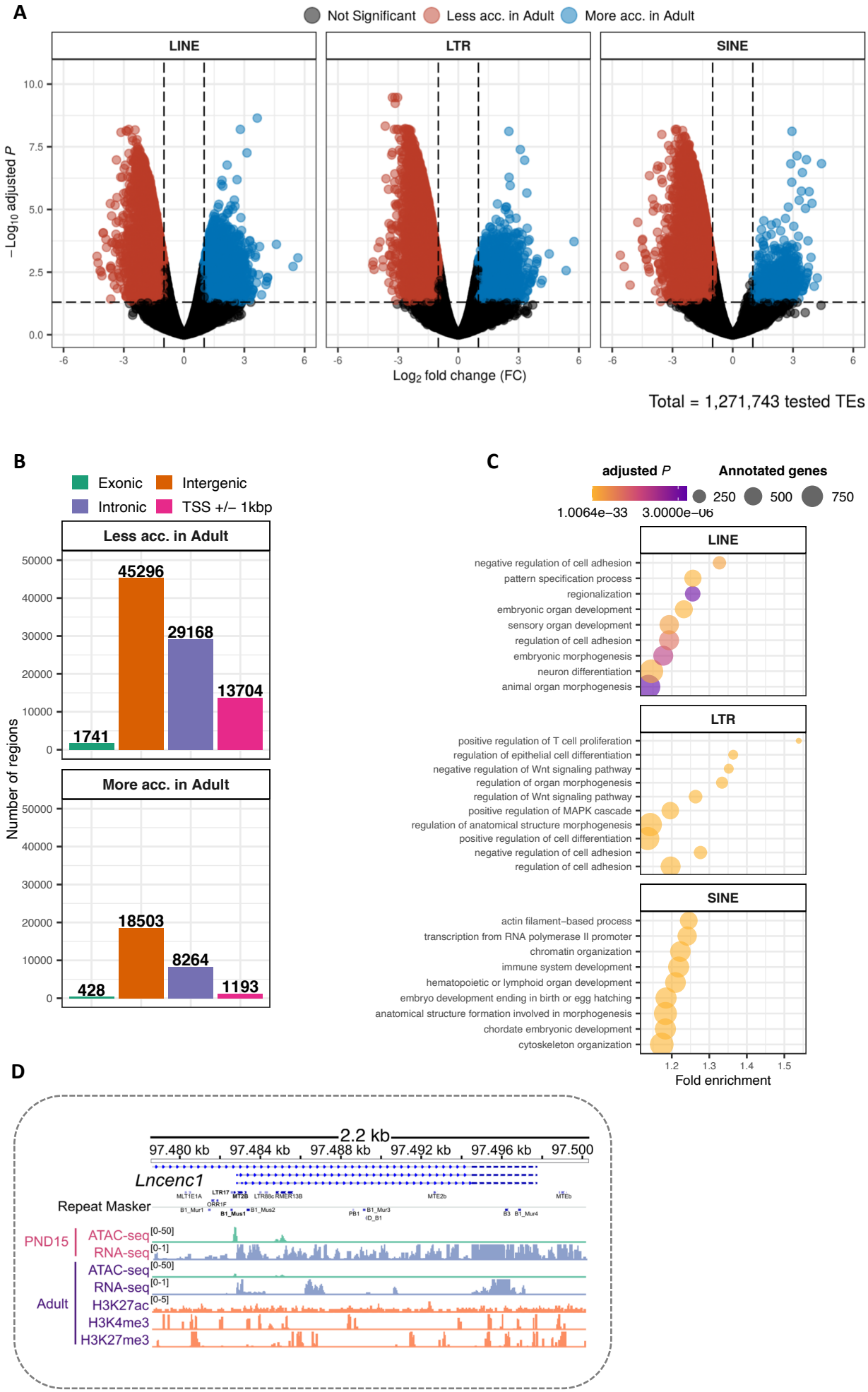

A
PND15 ATAC
Adult ATAC
Categories
Less acc.
More acc.
ATAC RPKM
8
4
0
−2.5kb Mid 2.5kb −2.5kb Mid 2.5kb
C
More acc. (H3K4me3)
RXRA
RFX3
NR4A1
NANOG
RFX2
B
PND15 ATAC
Adult ATAC
PNW8 H3K4me3
PNW8 H3K27ac
PNW8 H3K27me3
PND7 BS
PND14 BS
PNW8 BS
Categories
Less acc. (No H3 mark)
Less acc. (H3K27me3)
Less acc. (H3K4me3)
More acc. (No H3 mark)
More acc. (H3K27ac)
More acc. (H3K27me3)
More acc. (H3K4me3)
ATAC RPKM
H3K4me3 RPKM
H3K27ac RPKM
H3K27me3 RPKM
DNAme

Figure 6


A
Not Significant
Less acc. in Adult
More acc. in Adult
LINE
LTR
SINE
−Log10 adjusted P
Log2 fold change (FC)
Total = 1,271,743 tested TEs
B
Exonic
Intergenic
Intronic
TSS +/− 1kbp
Less acc. in Adult
More acc. in Adult
Number of regions
1741
45296
29168
13704
428
18503
8264
1193
C
adjusted P
Annotated genes
1.0064e−33
3.0000e−06
250
500
750
negative regulation of cell adhesion
pattern specification process
regionalization
embryonic organ development
sensory organ development
regulation of cell adhesion
embryonic morphogenesis
neuron differentiation
animal organ morphogenesis
positive regulation of T cell proliferation
regulation of epithelial cell differentiation
negative regulation of Wnt signaling pathway
regulation of organ morphogenesis
regulation of Wnt signaling pathway
positive regulation of MAPK cascade
regulation of anatomical structure morphogenesis
positive regulation of cell differentiation
negative regulation of cell adhesion
regulation of cell adhesion
actin filament−based process
transcription from RNA polymerase II promoter
chromatin organization
immune system development
hematopoietic or lymphoid organ development
embryo development ending in birth or egg hatching
anatomical structure formation involved in morphogenesis
chordate embryonic development
cytoskeleton organization
Fold enrichment
D
2.2 kb
97.480 kb
97.484 kb
97.488 kb
97.492 kb
97.496 kb
97.500
Lncenc1
Repeat Masker
PND15
ATAC-seq
RNA-seq
Adult
ATAC-seq
RNA-seq
H3K27ac
H3K4me3
H3K27me3