

Figure 1. Regions of differential chromatin accessibility in early postnatal and adult spermatogonial cells associate with distinct gene programs

- A. Volcano plot of differentially accessible regions (adjusted $P \leq 0.05$ and absolute $\text{Log}_2 \text{FC} \geq 1$) between PND15 and adult spermatogonial cells identified by ATAC-seq ($n = 6$ for PND15 and 5 for adult stage);
- B. Bar plot illustrating the genomic distribution of differentially accessible chromatin regions in PND15 and adult spermatogonial cells. Genomic regions are categorized in exonic, intronic, intergenic and +/- 1kb from the TSS of a gene;
- C. Dot plots of top enriched GO biological processes for regions with increased and decreased chromatin accessibility in adult spermatogonia. The size of dots 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.

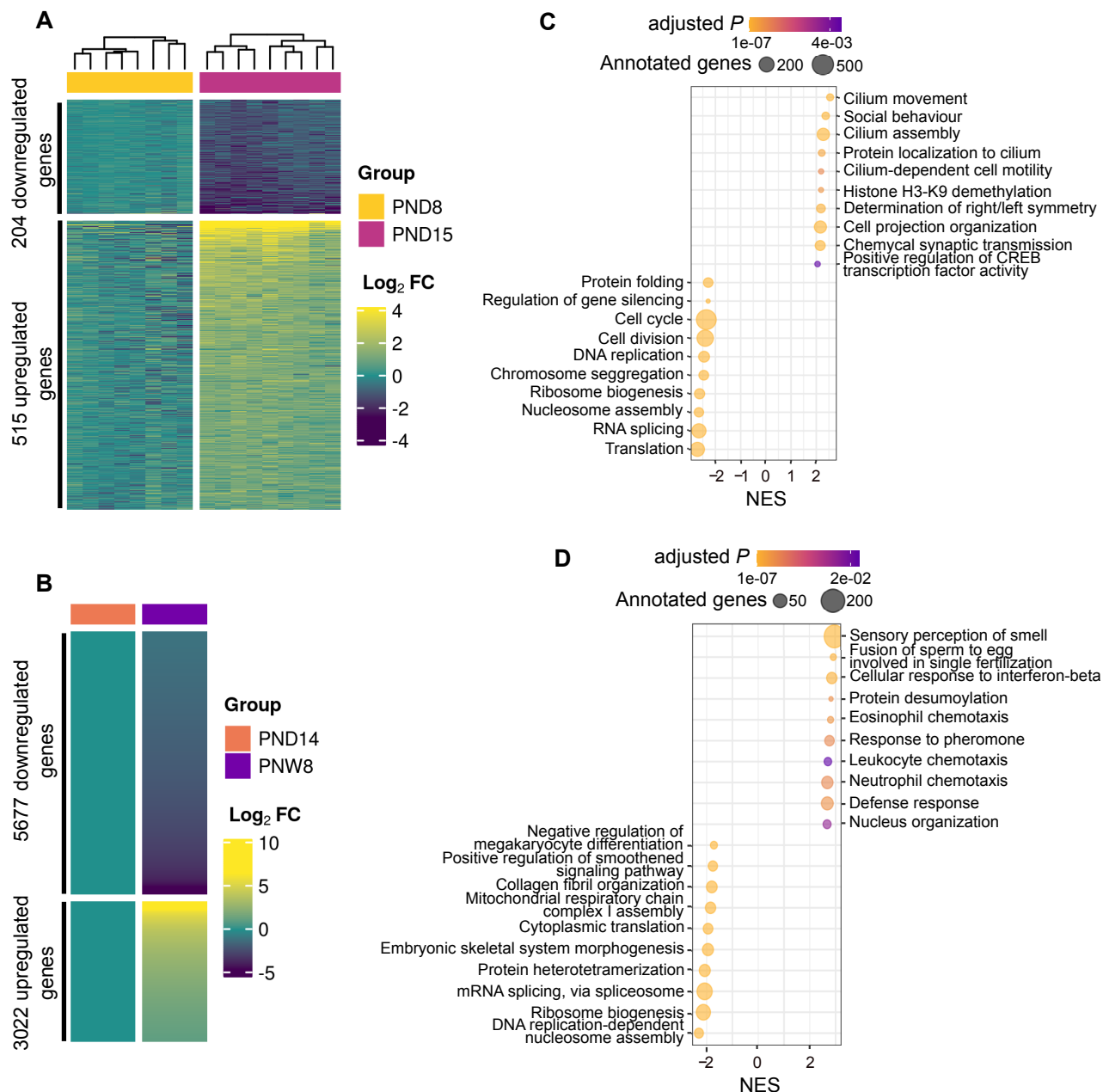


Figure 2. Transcriptomic profile in early postnatal and adult spermatogonial cells

A. Heatmap of differentially expressed genes (adjusted $P \leq 0.05$ and abs Log₂ FC ≥ 1) between PND8 ($n = 9$) and PND15 ($n = 8$) spermatogonial cells. Shown are the Log₂ FC with respect to the average of the PND8. Samples are clustered using Ward's method;

B. Heatmap of expressed genes (Log₂ CPM ≥ 1 and abs Log₂ 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 20 enriched GO biological processes (adjusted $P \leq 0.05$) from GSEA analysis of PND15 vs PND8 and PND14 vs PNW8 comparisons, 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 color corresponds to the adjusted *P* value.

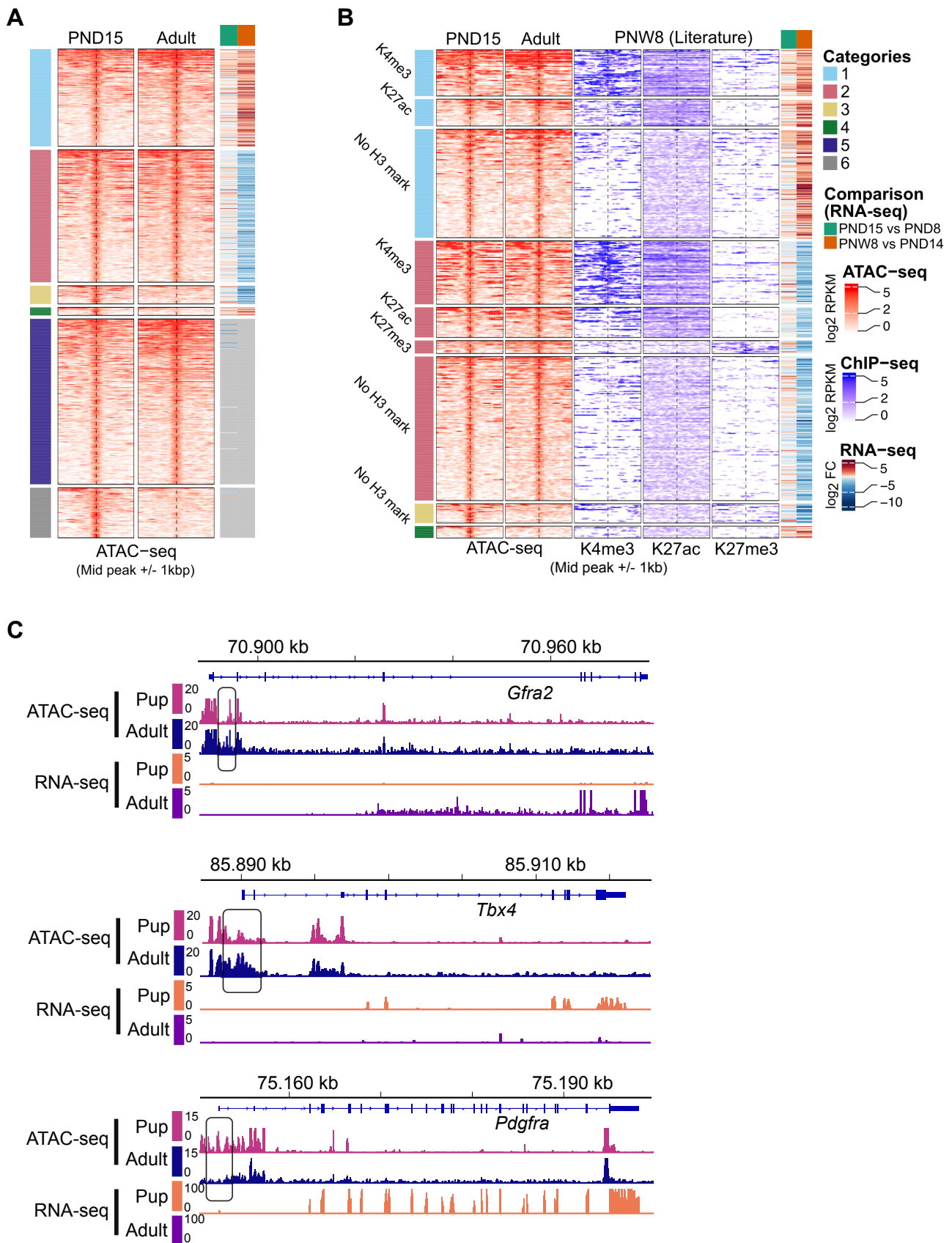


Figure 3. Chromatin accessibility and histone modifications at proximal regions of genes dynamically expressed between adult and PND15 spermatogonial cells

A. Heatmap of average coverage across condition showing 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, $n = 14$) when comparing adult with PND15 spermatogonia. 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. (continues on the next page)

B. Heatmaps showing the overlap between Category 1-4 regions 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 kbp. The color-key of the ATAC-seq heatmap represents the ATAC-seq signal in Log₂ Reads Per Kilobase per Million (RPKM) reads sequenced. For RNA-seq, log₂ FC is shown from PND15 vs PND8 and PNW8 vs PND14 comparisons. Rows are ordered by the enriched scores of ATAC-seq data;

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.

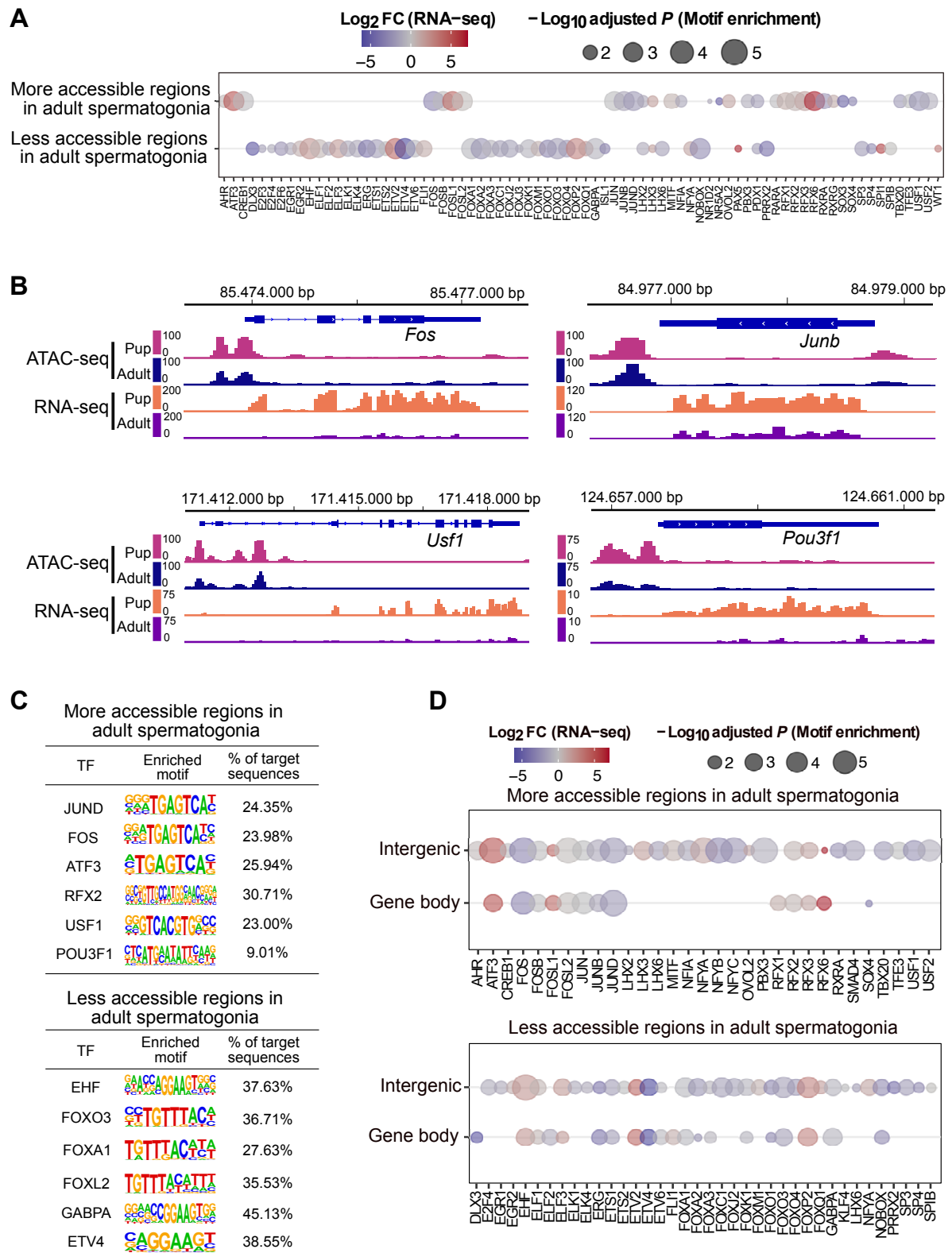


Figure 4. Transcription factor dynamics at differentially accessible regions as predicted by motif enrichment in adult spermatogonial cells

- A. Dot plot of all transcription factor motifs enriched in the regions of decreased and increased accessibility between adult and PND15 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; (*continues on the next page*)

- 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 PNW8 literature RNA-seq, and is shown as color coded Log_2 FC. The size of the dot indicates the HOMER motif enrichment adjusted P of each motif.

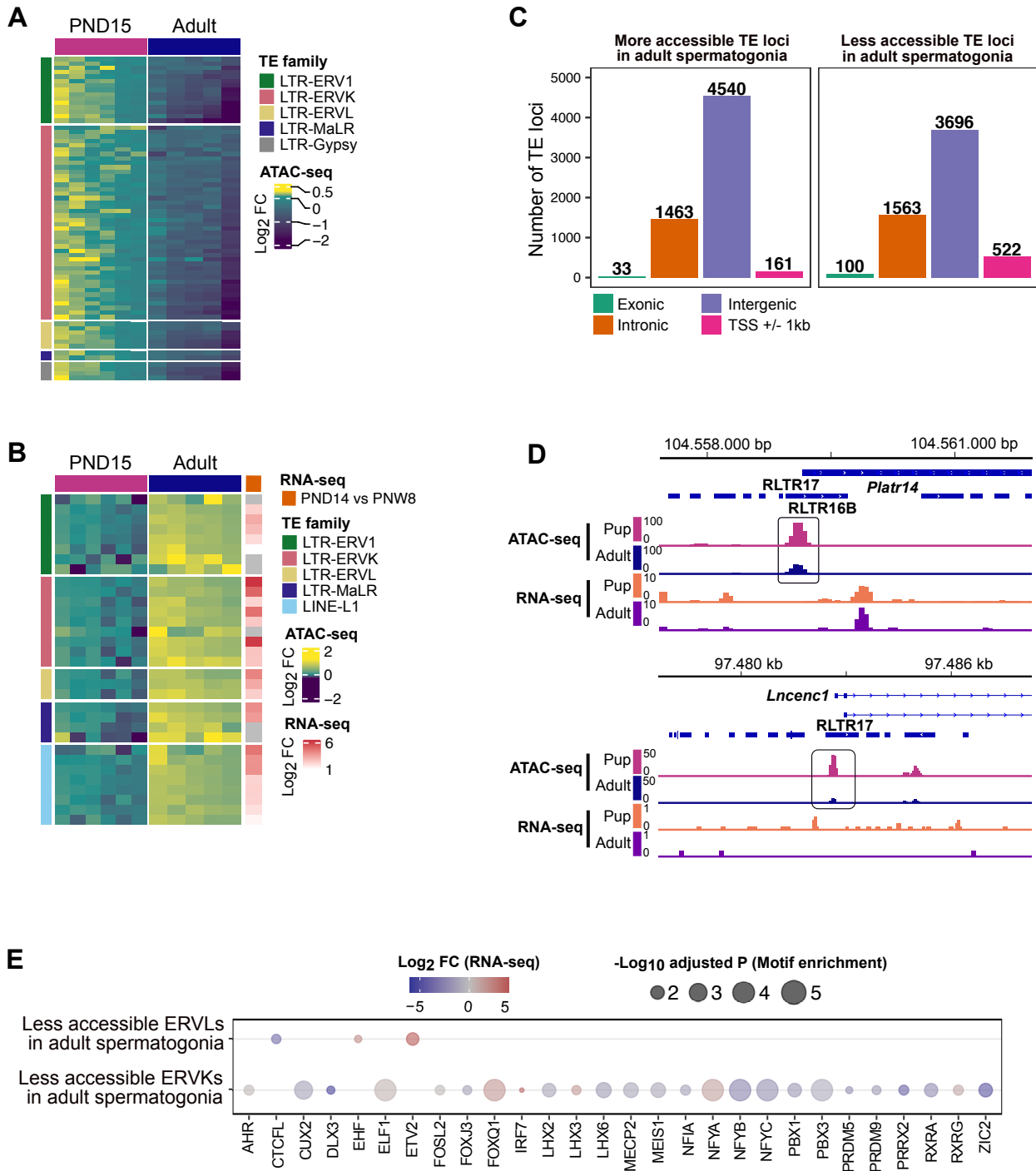


Figure 5. Differential chromatin accessibility at transposable elements (TEs) in adult spermatogonial cells compared to PND15

- A. Heatmap of the LTR and LINE subtypes with decreased accessibility between adult and PND15 spermatogonia (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 (adjusted $P \leq 0.05$ and $\text{Log}_2 \text{FC} \geq 0.5$). Expression changes of these subtypes between PNW8 and PND14 spermatogonia literature RNA-seq is represented as $\text{Log}_2 \text{FC}$;
- (A, B) $\text{Log}_2 \text{FC}$ 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; (continues on the next page)

- C. 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;
- D. 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 color coded Log_2 FC. The size of the dot indicates the HOMER motif enrichment adjusted P of each motif.

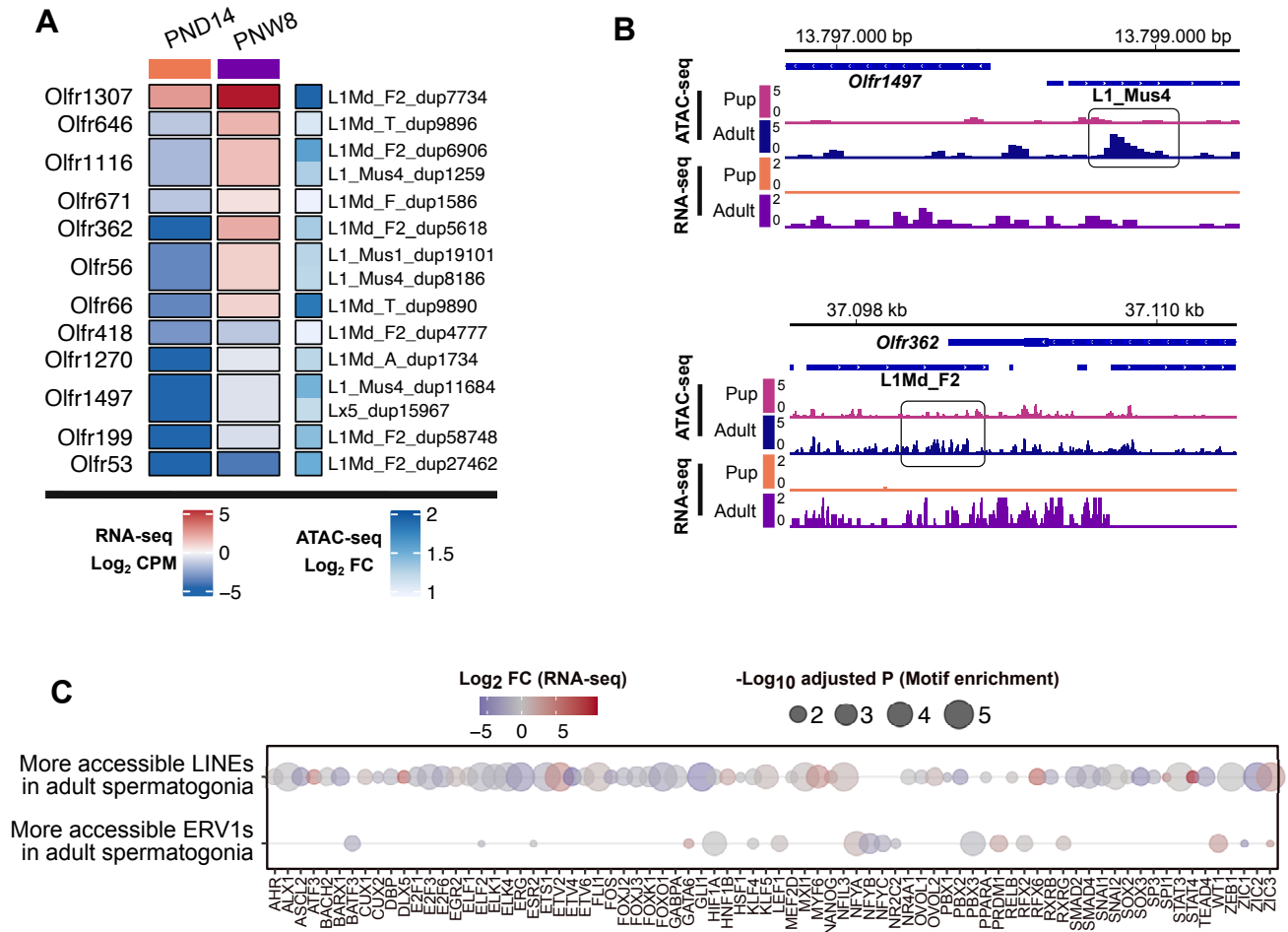


Figure 6. Increased accessibility at LINE L1 subtypes located near *Olfr* gene clusters

- 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 Log₂ CPM at each timepoint. Accessibility changes at each of the corresponding L1 locus situated within +/- 5kbp from the gene are expressed as Log₂ FC calculated from the ATAC-seq analysis of the differentially accessible TEs between adult and PND15 spermatogonial cells. Locus are ordered by PCA method using seriation (R package);
- 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;
- 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 color coded Log₂ FC. The size of the dot indicates the HOMER motif enrichment adjusted *P* of each motif.