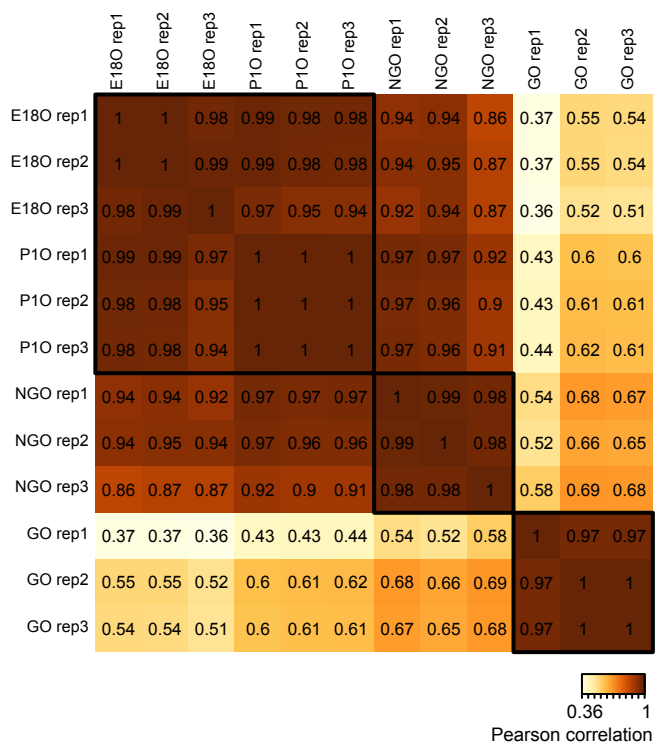
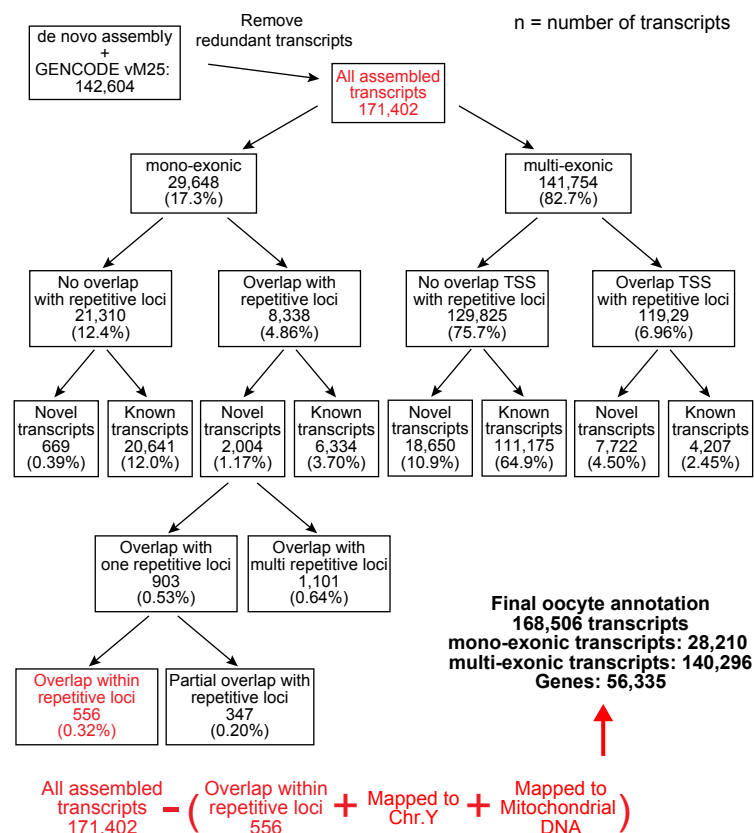


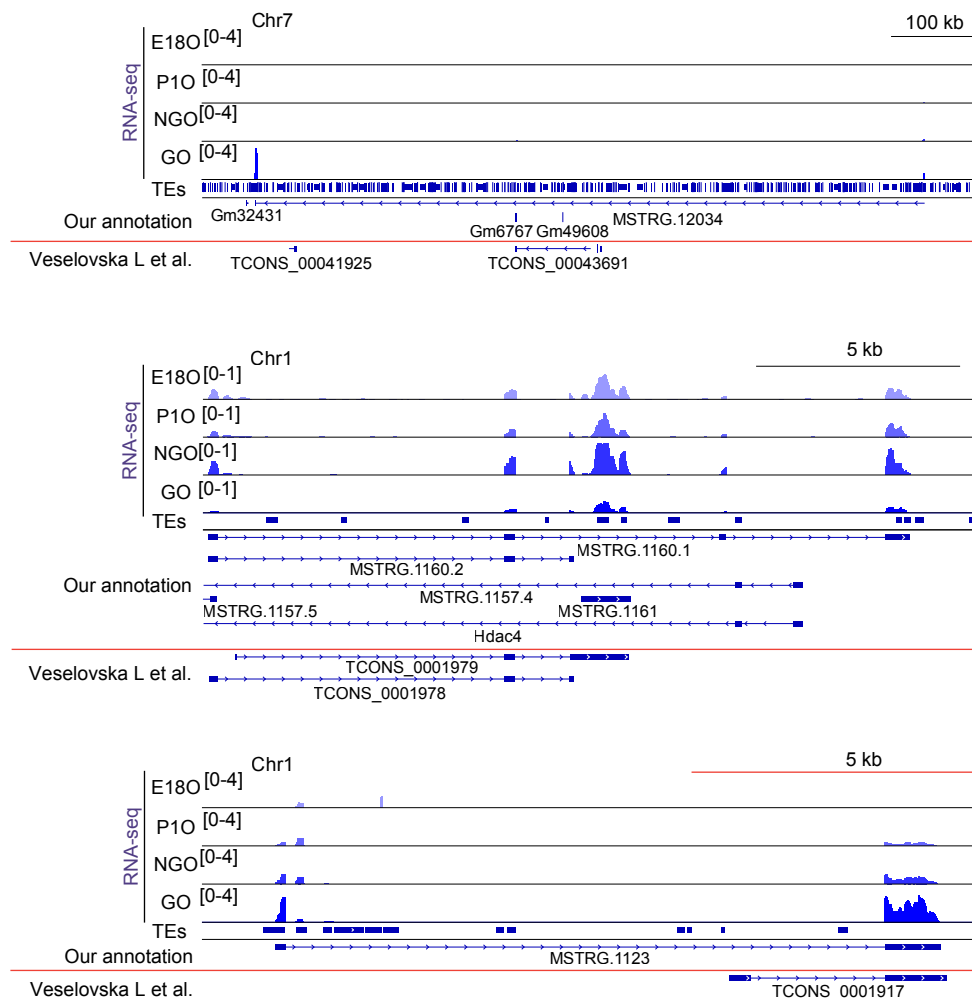
a RNA-seq correlation based on TPM



b Workflow of transcriptome assembly



c RNA-seq

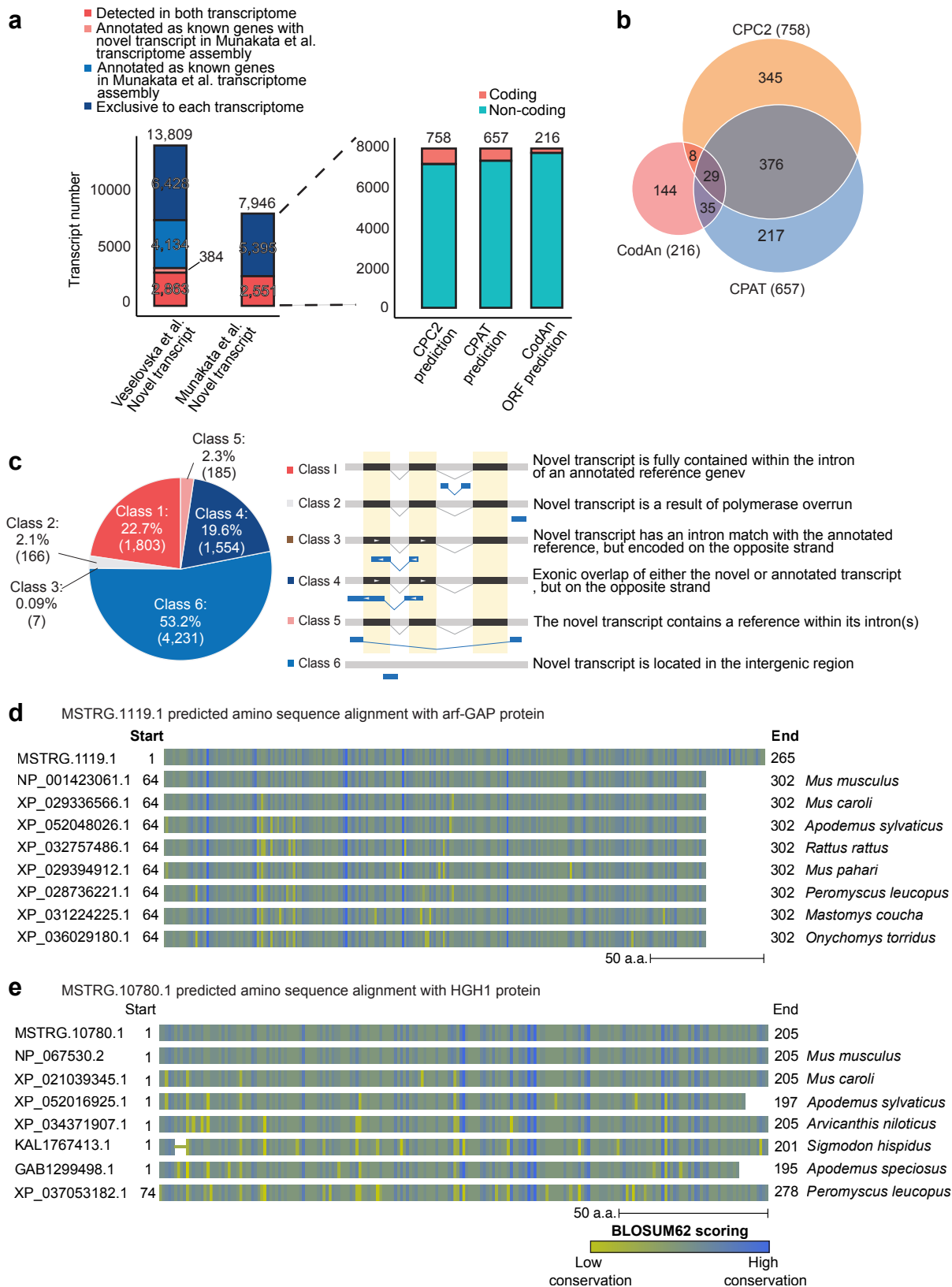


Extended Data Figure 1. De novo transcriptome assembly during perinatal oogenesis.

a, Heatmaps showing Pearson correlations of TPM values among biological replicates in strand-specific total RNA-seq datasets. Pearson correlation coefficients are indicated in the panels.

b, Flow chart showing the number of transcripts in each category: monoexonic versus multiexonic, known versus novel, and overlapping with repeats versus non-overlapping. Redundant transcripts were removed by merging de novo assemblies with the GENCODE reference annotation. Monoexonic transcripts fully overlapping with RepeatMasker-annotated TEs were excluded to eliminate non-uniquely mappable TE-derived transcripts.

c, Representative track views showing RNA-seq signals, the oocyte transcript annotation generated in this study, and the previously published oocyte annotation from Veselovska et al.²⁰ Data ranges are shown in brackets.



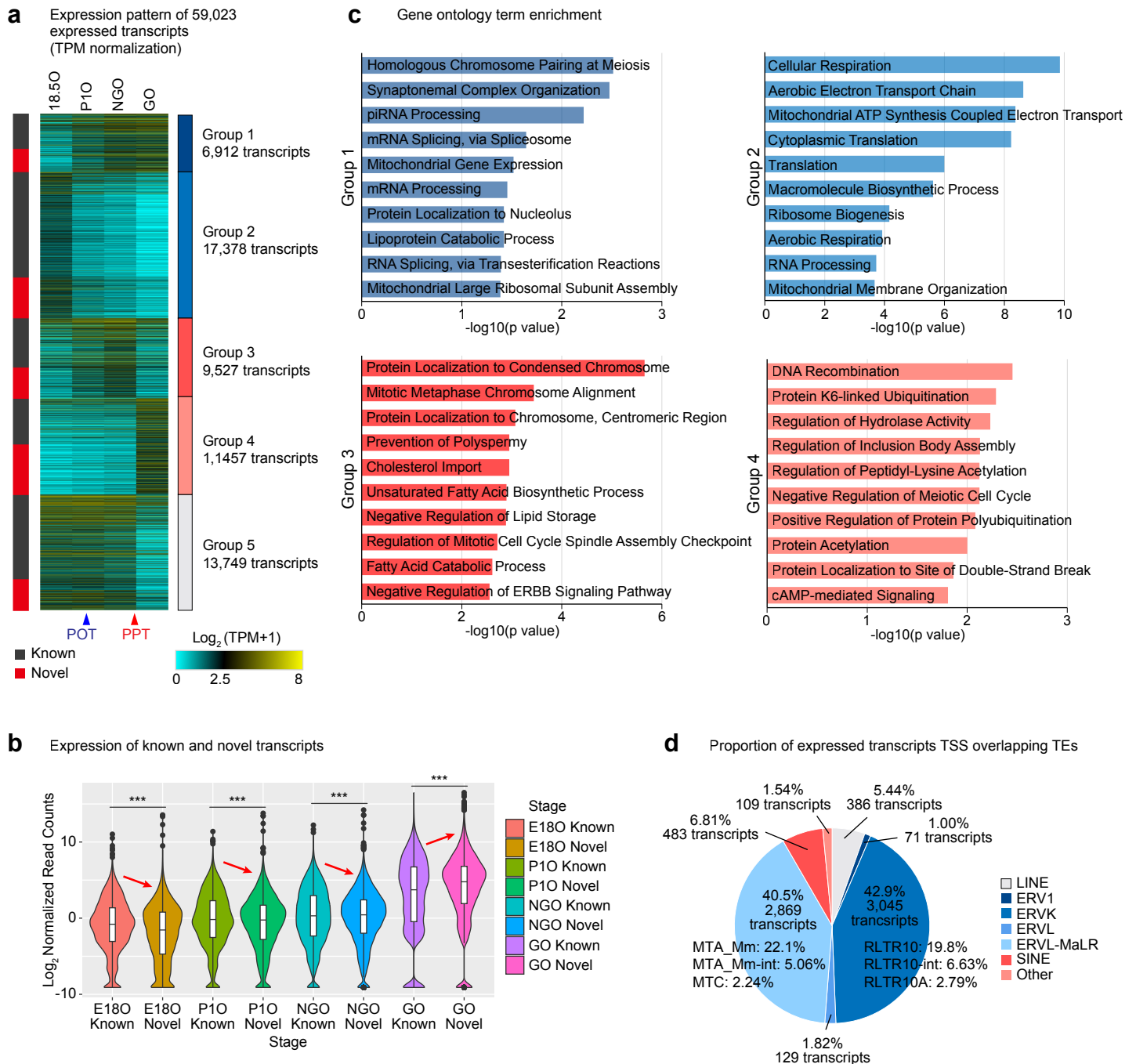
Extended Data Figure 2. Characterization of novel transcripts.

a, Comparison between novel transcripts identified in Veselovska et al. and in this study (left). Proportions of novel transcripts with coding potential predicted by CPC2, CPAT, and CodAn (right).

b, Venn diagram showing the overlap of transcripts predicted to contain coding sequences by CPC2, CPAT, and CodAn. Numbers of total predictions and overlaps between programs are indicated.

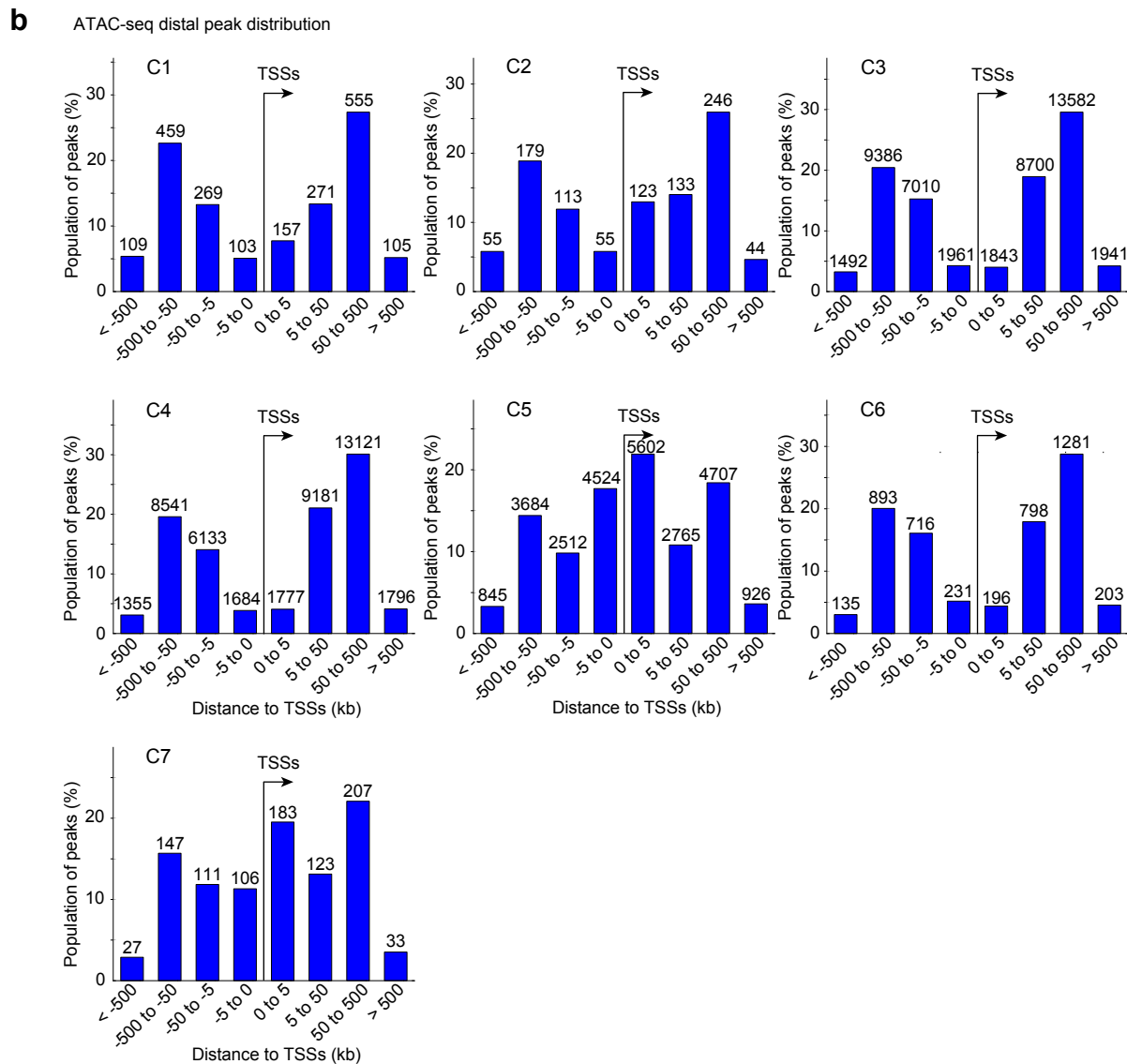
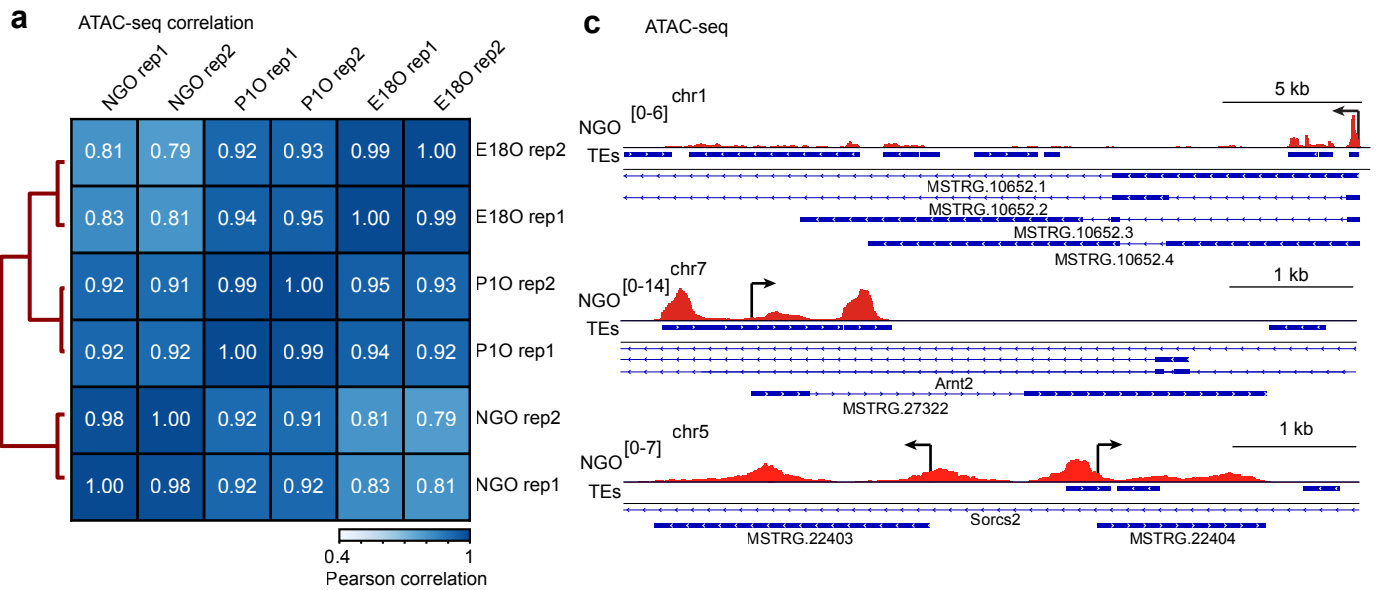
c, Genomic location of novel transcripts relative to the GENCODE vM25 annotation, classified as classes 1–6. The schematic representation of each class is shown on the right. Class 1, novel transcript located within an intron of an annotated gene; class 2, novel transcript likely resulting from polymerase overrun; class 3, novel transcript sharing an intron with an annotated gene on the opposite strand; class 4, novel and annotated transcripts with overlapping exons on opposite strands; class 5, annotated gene located within the boundary of the novel transcript; class 6, novel transcript located in the intergenic region.

d, e, Sequence alignments of predicted protein products from novel transcripts MSTRG.1119.1 and MSTRG.10780.1 showing partial homology to arf-GAP and HGH1 protein sequences, respectively. The color key reflects the average BLOSUM62-based similarity of each residue to all others at the same alignment position: blue indicates higher similarity, green indicates lower similarity.



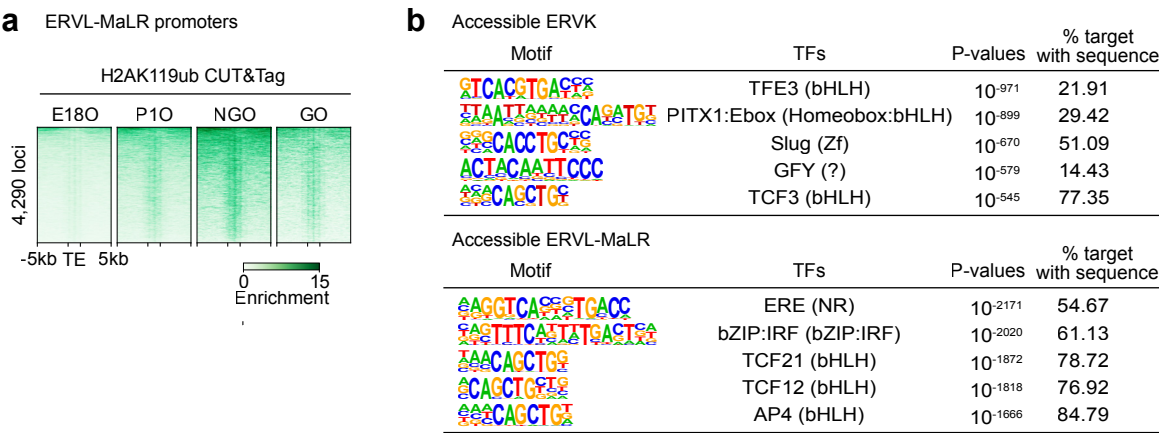
Extended Data Figure 3. Expression profiles of novel transcripts.

- a**, Heatmaps showing transcript expression during perinatal oogenesis (>1 TPM in at least one of the four developmental stages). Expressed transcripts were grouped into five k-means clusters.
- b**, Violin plots with a box plot showing log₂-transformed transcript expression normalized by ERCC spike-in during perinatal oogenesis. Central lines represent medians; boxes indicate the interquartile range (IQR; 25th–75th percentile); whiskers extend to 1.5× IQR from the hinges. ****P* < 0.001, Wilcoxon rank-sum test.
- c**, Gene ontology term enrichment analysis of genes in each group shown in Figure 1e.
- d**, Pie charts showing the proportion of TE families and types among TEs overlapping with TSSs of expressed transcripts.

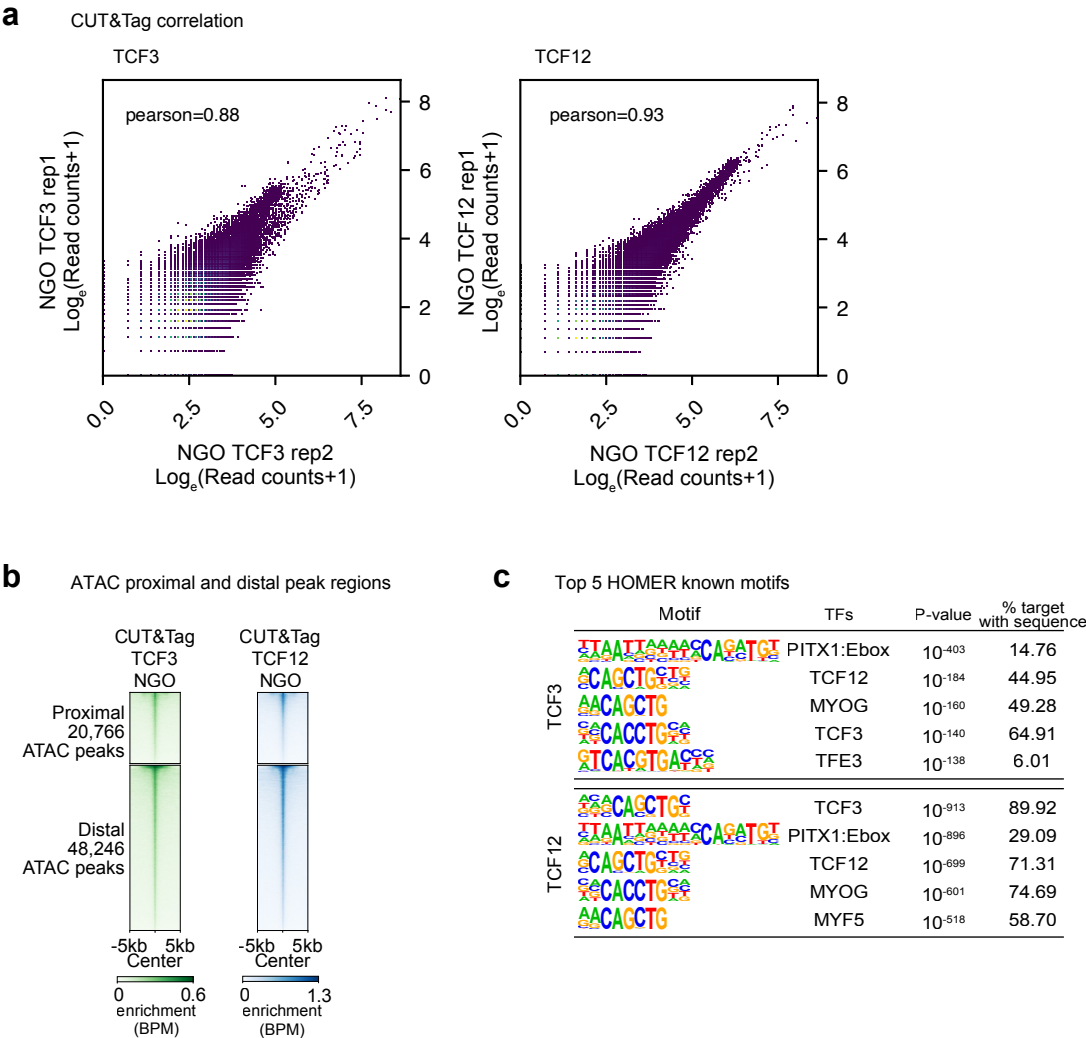


Extended Data Figure 4. ATAC-seq analysis during POT.

- a**, Heatmaps showing Pearson correlations of read counts among biological replicates in ATAC-seq datasets.
- b**, Bar chart showing the genomic distribution of ATAC-seq peaks according to their distance from TSSs.
- c**, Representative track views showing ATAC-seq signals at gene loci with accessible TEs downstream of TSSs. Arrows indicate the direction of transcription.



Extended Data Figure 5. Analysis of ERVs in accessible chromatin regions.
a, Heatmaps showing H2AK119ub enrichment at proximal accessible ERVL-MaLR regions during perinatal oogenesis.
b, HOMER known motif analyses of accessible ERVK (top) and ERVL-MaLR (bottom) loci in NGO.



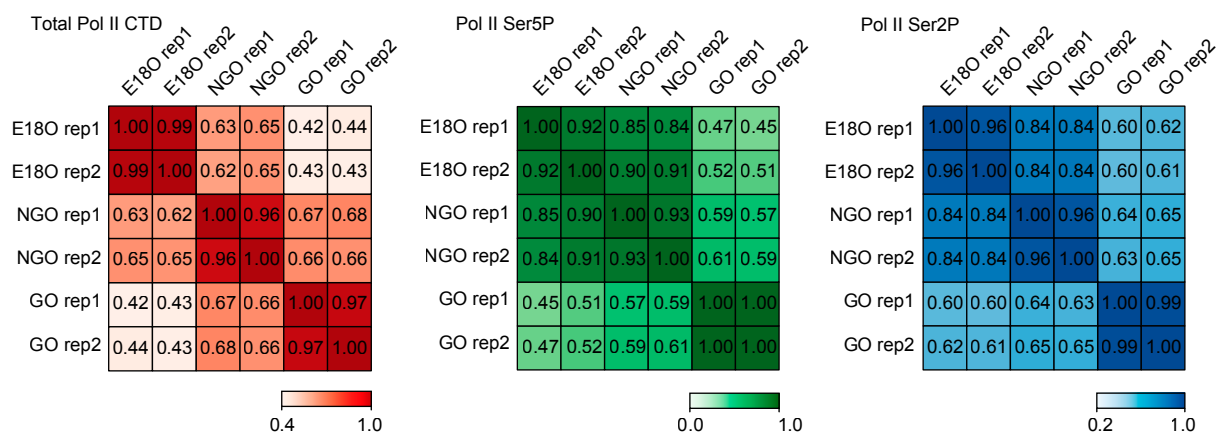
Extended Data Figure 6. TCF3 and TCF12 CUT&Tag analysis in NGO.

a, Scatter plots showing reproducibility between biological replicates in TCF3 (left) and TCF12 (right) CUT&Tag datasets. Pearson correlation coefficients are indicated in the panels.

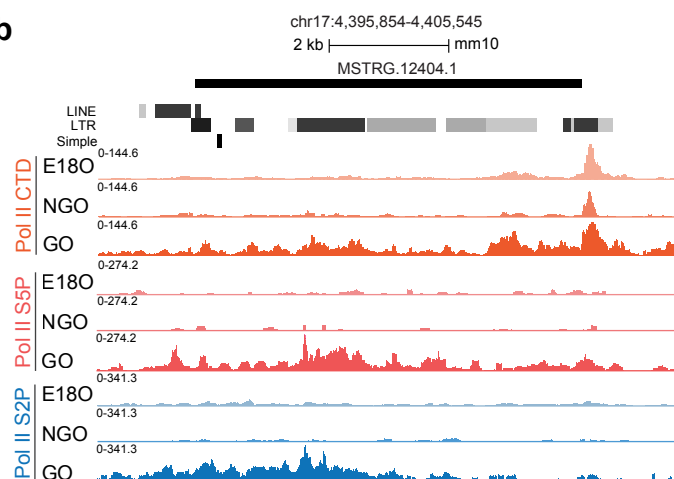
b, Heatmaps showing TCF3 (left) and TCF12 (right) CUT&Tag enrichment dynamics at proximal (TSS \pm 1 kb) and distal (>1 kb from TSSs) ATAC-seq peaks shown in Figure 2c.

c, HOMER known motif analyses of TCF3 (top) and TCF12 (bottom) CUT&Tag peaks.

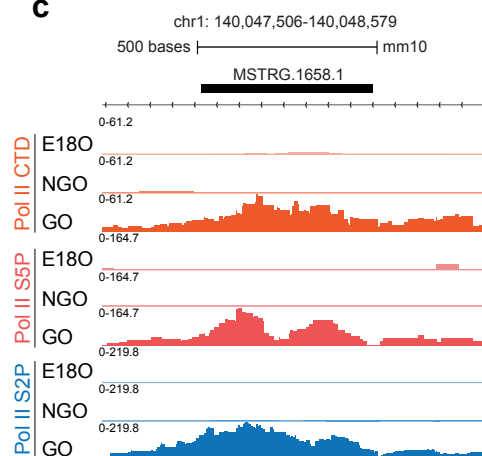
a CUT&Tag correlation



b



c



Extended Data Figure 7. Pol II CUT&Tag analysis during perinatal oogenesis.

a, Heatmaps showing Pearson correlations of read counts among biological replicates in total Pol II (left), Pol II Ser5P (center), and Pol II Ser2P (right) CUT&Tag datasets. Pearson correlation coefficients are indicated in the panels.

b, c, Track views showing total Pol II, Pol II Ser5P, and Pol II Ser2P enrichment at mono-exonic transcripts (MSTRG.12404.1 in **b** and MSTRG.1658.1 in **c**) derived from novel genes identified in this study during perinatal oogenesis.