

EXTENDED DATA

Extended Data Fig. 1 | Validation of RA-synchronized spermatid isolation and purity, related to Fig. 1.

a. Flow cytometric gating strategy for the isolation of synchronized spermatids at the indicated developmental stages. RS* denotes a subsequent cohort of unsynchronized spermatids arising under physiological conditions. LS* designates a distinct cell population observed at the LS9–10 stage, which was excluded from downstream analyses.

b. Immunofluorescence (IF) validation of sorted cell populations using DAPI and PNA staining in testis seminiferous tubules and in FACS-isolated spermatids. The scale bar is 10 μm in FACS-sorted IF images and 20 μm in testis IF images.

c. Quantitative assessment of spermatid purity at each stage, based on morphological evaluation of at least 100 cells per stage.

Extended Data Fig.2| Characterization of RNA-seq data and gene expression programs, related to Fig. 1.

a. Pearson correlation coefficients of RNA-seq between biological replicates of all spermatid stages.

b. Temporal expression trajectories of gene clusters calculated by RNA-seq during spermatid differentiation. Data are presented as mean $\pm$ s.d. across genes within each cluster.

c. Heatmap depicting RNA expression levels of genes associated with transcription-related Gene Ontology (GO) terms across developmental stages: Initiation (GO:0006352), Elongation (GO:0006354), Termination (GO:0006353), RNA Processing (GO:0006397), mRNA 3'-end Processing (GO:0031124), and Epigenetic regulation (GO:0040029).

d. Expression profiles of canonical stage-specific marker genes. Data are presented as mean $\pm$ s.d. across biological replicates.

e. Dynamics of transition protein and protamine gene expression. Data are presented as mean $\pm$ s.d. across biological replicates.

f. Mean expression of SWI/SNF chromatin remodeling subunits during spermatid differentiation.

g. Heatmap depicting mean expression of individual TE families.

h. Comparative TE expression among spermatids, mESCs, and myoblasts. Data are presented as mean $\pm$ s.e. across individual TE elements within each class.

i. Overview of major TE class expression dynamics. Each block represents the z-score of the mean expression of the class.

Extended Data Fig.3 | Refined annotation of spermatid subpopulations in scRNA-seq data using bulk RNA-seq signatures, related to Fig. 1.

a. Workflow for deriving stage-specific markers from bulk RNA-seq and applying them to scRNA-seq annotation.

b PCA of RNA-seq datasets, incorporating our synchronized spermatid samples together with earlier-stage cells reported by Lin et al.¹, with replicate numbers indicated in parentheses, performed on the basis of all expressed genes. Undiff, Undifferentiated spermatogonia; A1, A1 spermatogonia; B, B spermatogonia; mPL, mid-S phase preleptotene; L, leptotene; Z, zygotene; mP, mid pachytene; D, diplotene.

c. Temporal expression trajectories of stage markers calculated from RNA-seq data. Values are presented as mean $\pm$ s.d. across biological replicates.

d. t-SNE plots showing the expression distribution of representative stage-specific markers.

e. Bubble plot analysis similar to Fig. 1f showing that marker genes identified by Xia et al.² based on circumstantial presumption exhibit unqualified stage-specific markers, underscoring the importance of our rigorous direct measurement of highly purified cells.

Extended Data Fig.4 | Quality control and cross-validation of nascent transcription datasets, related to Fig. 2.

a. TT-seq workflow schematic.

b. Pearson correlation matrix of TT-seq replicates across spermatid stages.

c. Reproducibility of EU-seq replicates at RS* and LS15 stages.

d. Temporal expression curves of TT-seq-defined gene clusters. Data are presented as mean $\pm$ s.d. across genes within each cluster.

e. Representative IGV genome browser tracks comparing TT-seq and RNA-seq.

f. Representative IGV genome browser tracks for LS15-expressed genes across RNA-seq, TT-seq, and EU-seq.

g. Correlation of differential gene expression between TT-seq (RS1 versus LS15) and EU-seq (RS* versus LS15).

h. Heatmap of TE family dynamics of nascent transcription measured by TT-seq. Each block represents the z-score of the mean expression of the class.

Extended Data Fig.5 | Genomic distribution of TT-seq-defined gene clusters along autosomes and chromosome X, related to Fig. 2.

Each dot represents a gene ordered by TSS position and colored by TT-seq cluster assignment.

Extended Data Fig.6 | Chromatin accessibility dynamics during spermatid differentiation, related to Fig. 3.

a. Agarose gel analysis of PCR-amplified ATAC-seq library products across stages.

b. PCA of chromatin accessibility dynamics across all ATAC peaks during spermiogenesis.

c. Metagene profiles of promoter accessibility (measured by ATAC-seq) for TT-seq-defined clusters.

d. Heatmap of promoter accessibility (± 3 kb of TSS) measured by ATAC-seq for TT-seq-defined clusters.

e. Representative tracks for RS-enriched (*Slc2a8*) and LS-enriched (*Tex101*) genes.

f. Quantification of ATAC-seq signal at TSS ± 1 kb in myoblasts. Values are presented as mean $\pm$ s.e. across the grouped genes.

Extended Data Fig.7 | Comparative and dynamic analysis of H3 and H3K36me3 occupancy, related to Fig. 4.

a. Metagene comparison of H3K36me3 occupancy measured by CUT&Tag between spermatids and myoblasts.

b. Representative tracks for RS- and LS-specific genes, illustrating the correlation between H3K36me3-marked genes and RNA expression.

c. Quantification of H3K36me3 signal (CUT&Tag) in myoblast gene clusters. Values are presented as mean $\pm$ s.e. across the grouped genes.

d. Metagene profiles of H3 occupancy (CUT&Tag) stratified by TT-seq clusters.

e. Heatmap of intergenic H3 occupancy across stages.

Extended Data Fig.8 | Biochemical and biophysical validation of the histone-to-protamine transition, related to Fig. 5.

a. FACS profiles of adult spermatids (RS: round spermatids; ES: elongating spermatids; CS: condensed spermatids).

b. Immunofluorescence validation of sorted cell populations using DAPI and PNA staining for FACS-isolated spermatids.

c. Quantitative assessment of spermatid purity at each developmental stage, based on morphological evaluation of at least 100 cells per stage.

d. Western blot of histone retention in RS and CS isolated by FACS.

e. Quantification of histone signal by Western blot, normalized to total genomic DNA and spike-in DNA.

f. Heatmap of pairwise overlap coefficients (OV) for the ratio distributions among RS, LS, and sperm based on SRS. The overlap coefficient quantifies the similarity between two probability distributions (OV = 1 indicates identical distributions).

g. Three-dimensional Raman spectral mapping of sperm, with annotated peaks corresponding to proteins, lipids, and secondary structures.

Extended Data Fig.9 | Alternative splicing dynamics and transcription-coupled repair analysis, related to Fig. 6.

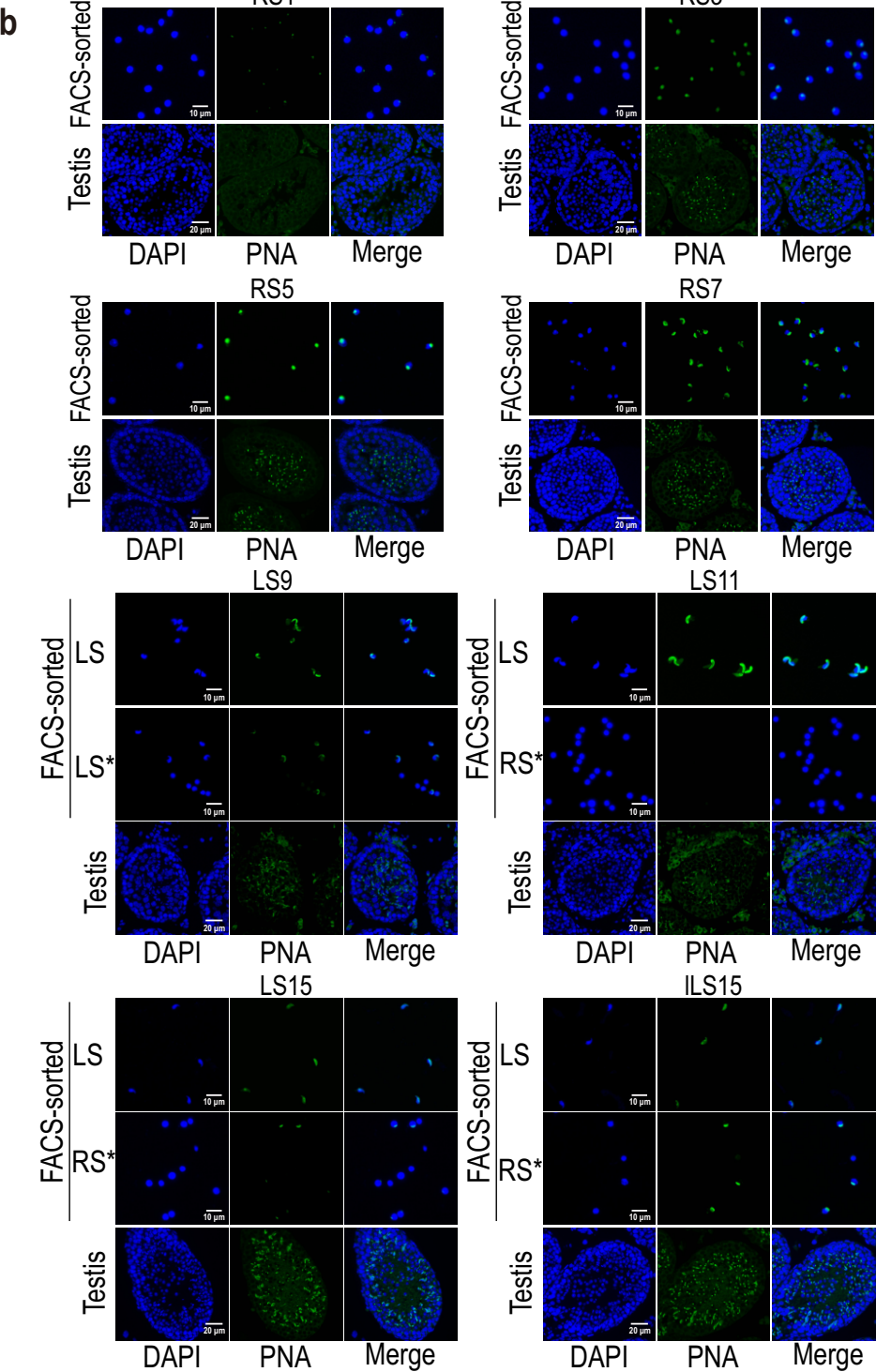
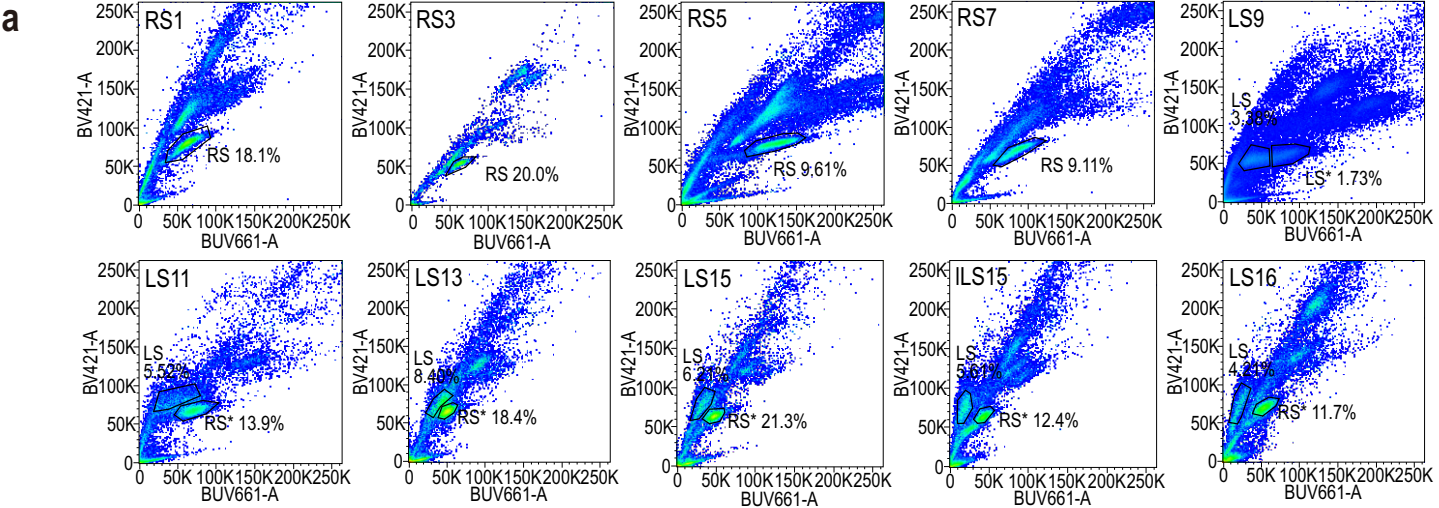
a. Number of protein-coding genes harboring alternatively spliced exons (PSI 20–80) at each developmental stage, as determined by RNA-seq analysis.

b, c. Comparative analysis of alternative splicing levels across datasets and cell types.

d. Germline SNV ratios stratified by gene expression level at each stage.

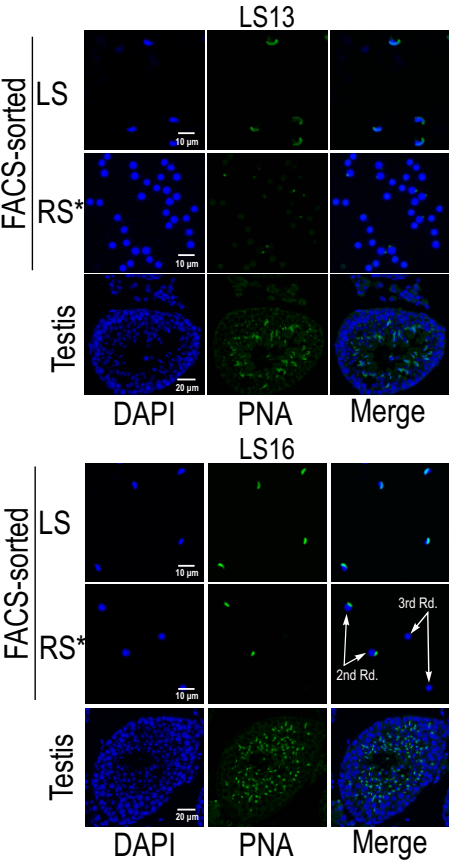
References:

- 1 Lin, Z. *et al.* SETD1B-mediated broad H3K4me3 controls proper temporal patterns of gene expression critical for spermatid development. *Cell Research* **35**, 345-361 (2025). <https://doi.org/10.1038/s41422-025-01080-0>
- 2 Xia, B. *et al.* Widespread Transcriptional Scanning in the Testis Modulates Gene Evolution Rates. *Cell* **180**, 248-262 e221 (2020). <https://doi.org/10.1016/j.cell.2019.12.015>

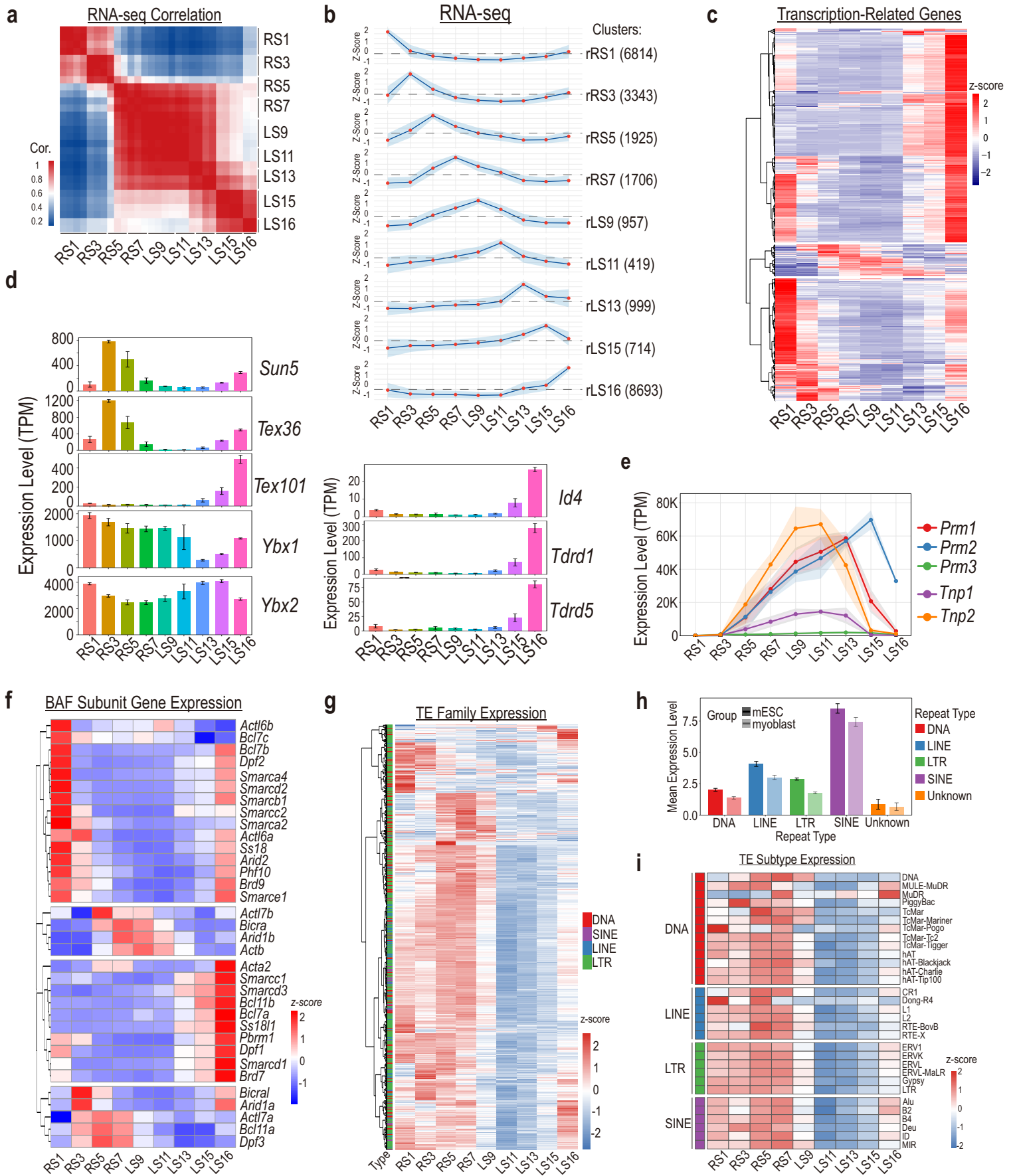


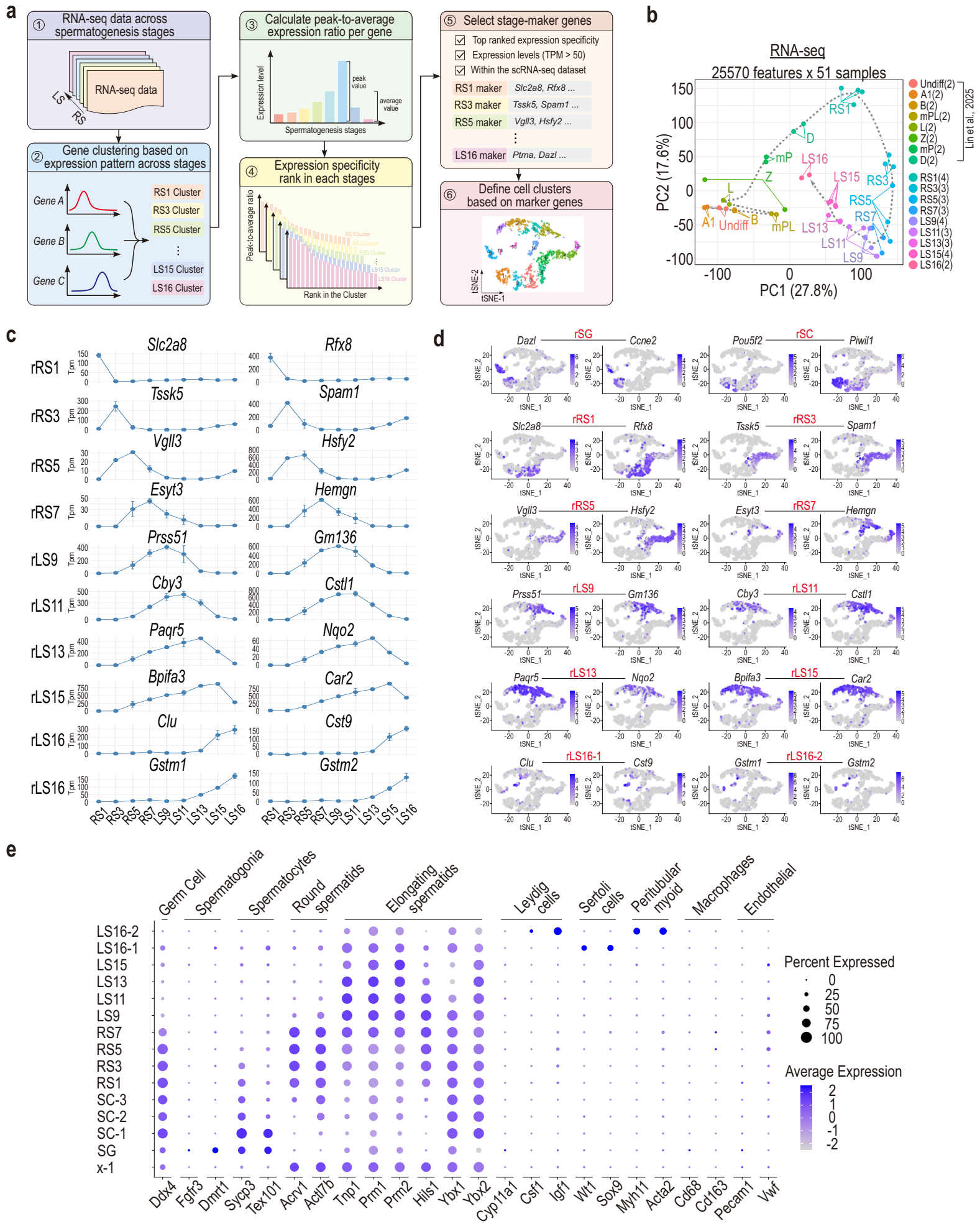
c

Cell Stage	Purity (%)
RS1	96.5±2
RS3	96.5±2
RS5	93.5±2
RS7	92.5±2
LS9	97.5±1
LS11	96.0±1
LS13	96.0±1
LS15	96.0±1
ILS15	98.5±1
LS16	97.5±1

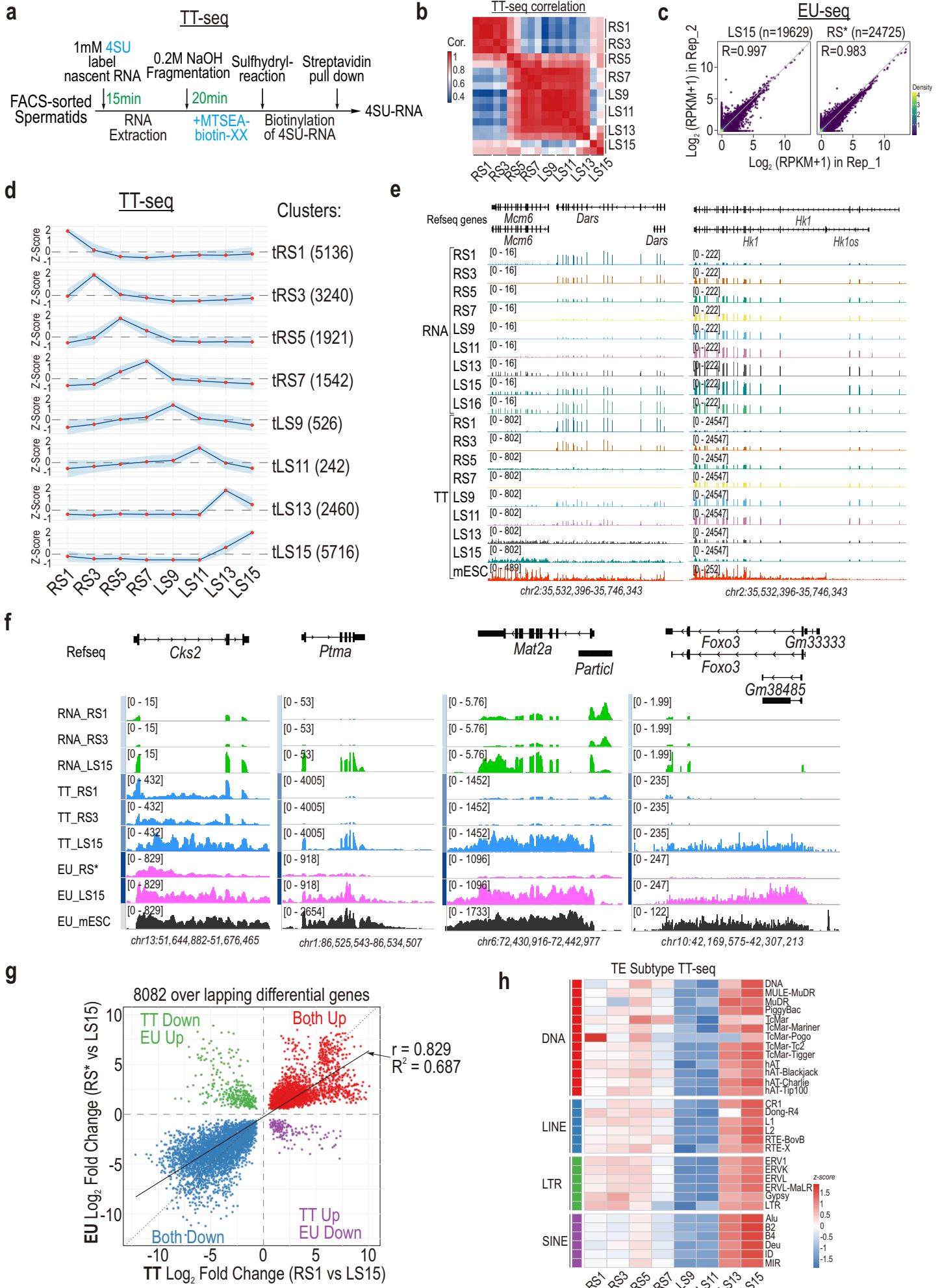


Extended Data Fig. 2



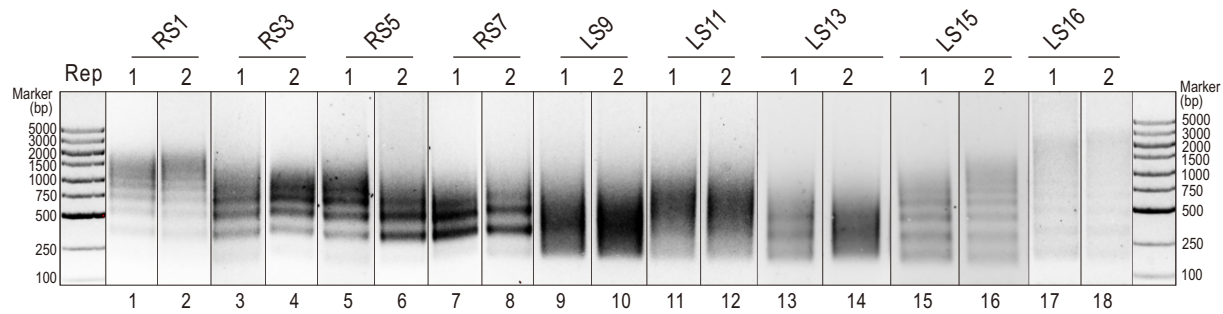
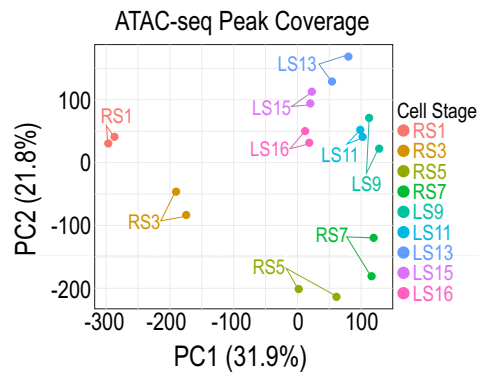
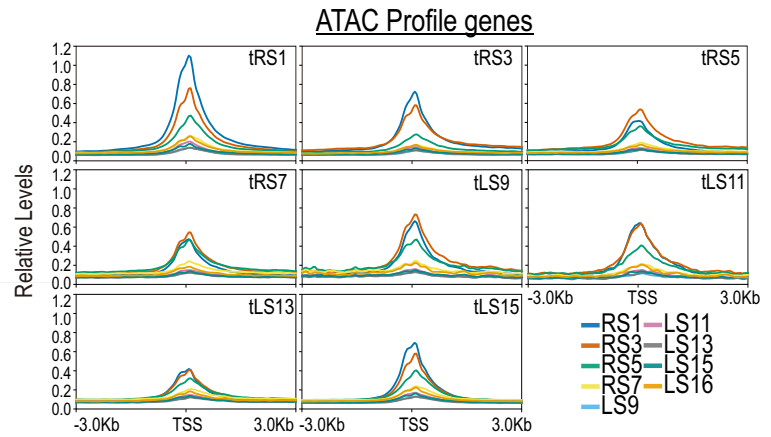
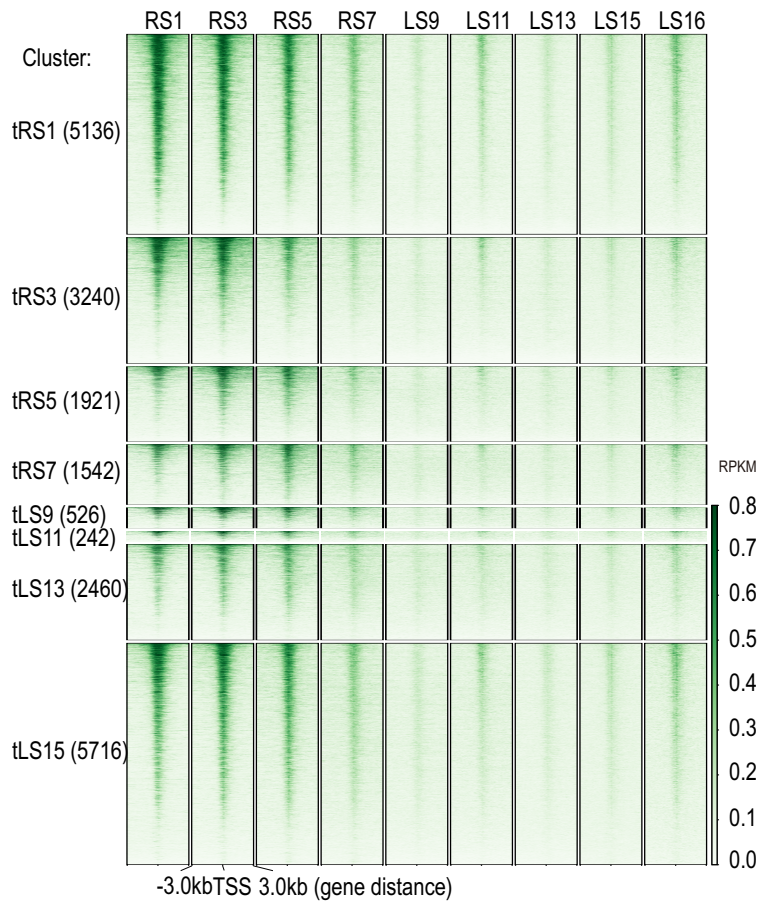
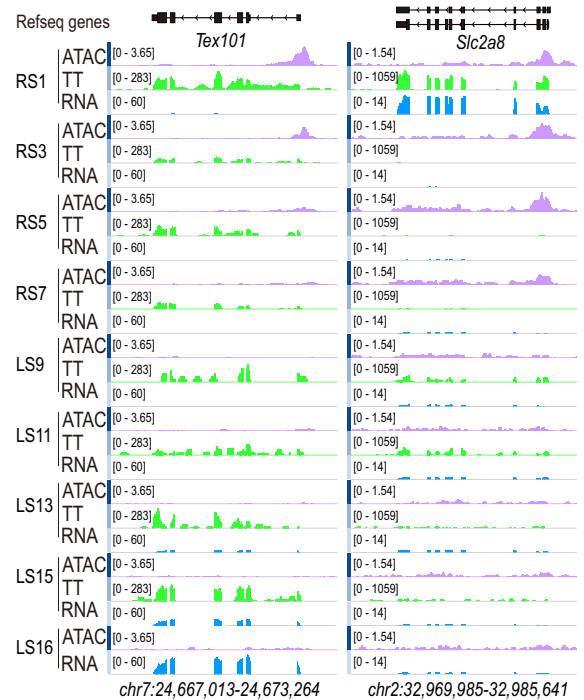
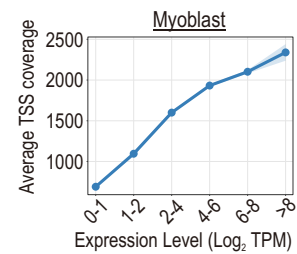


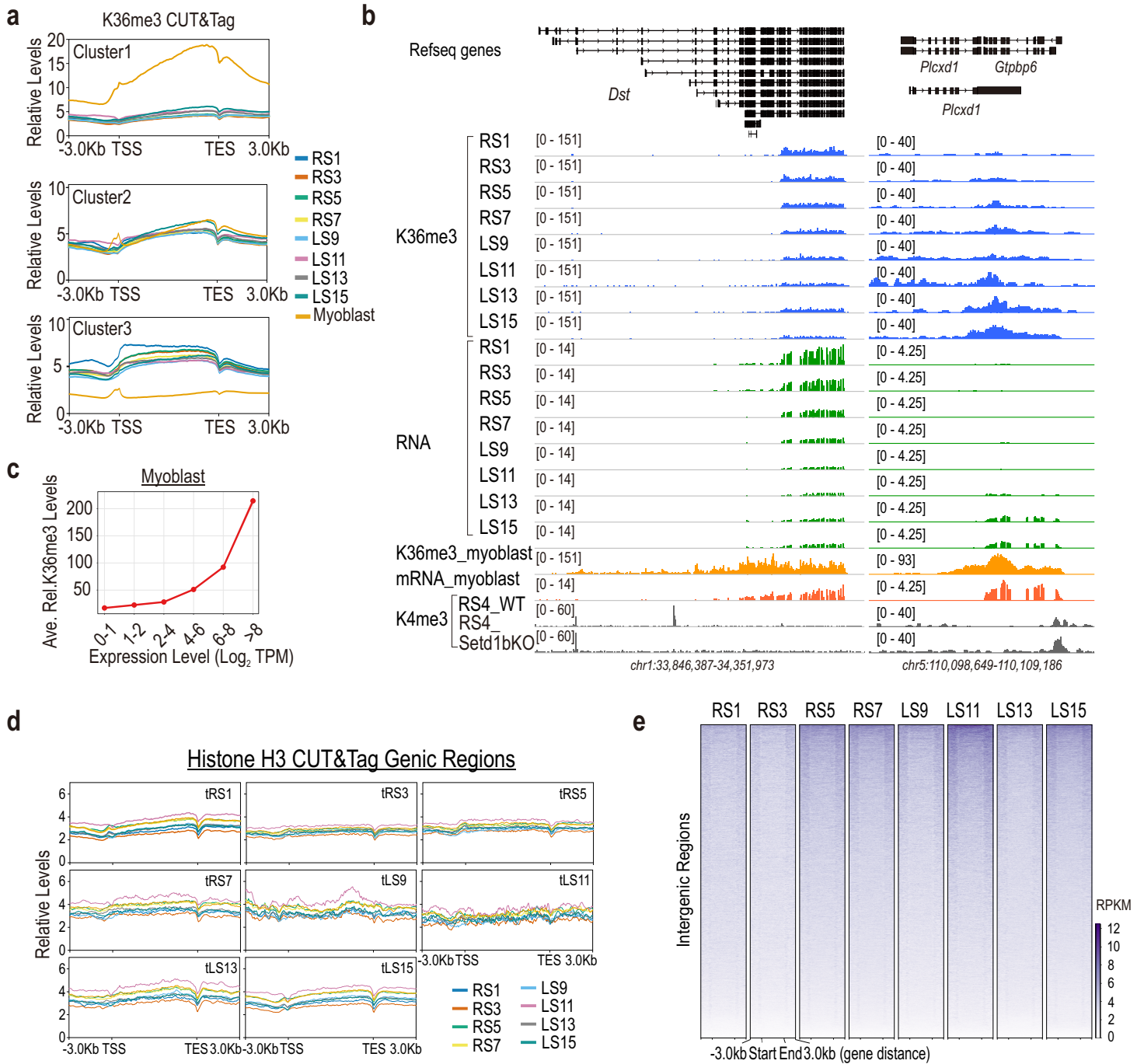
Extended Data Fig. 4

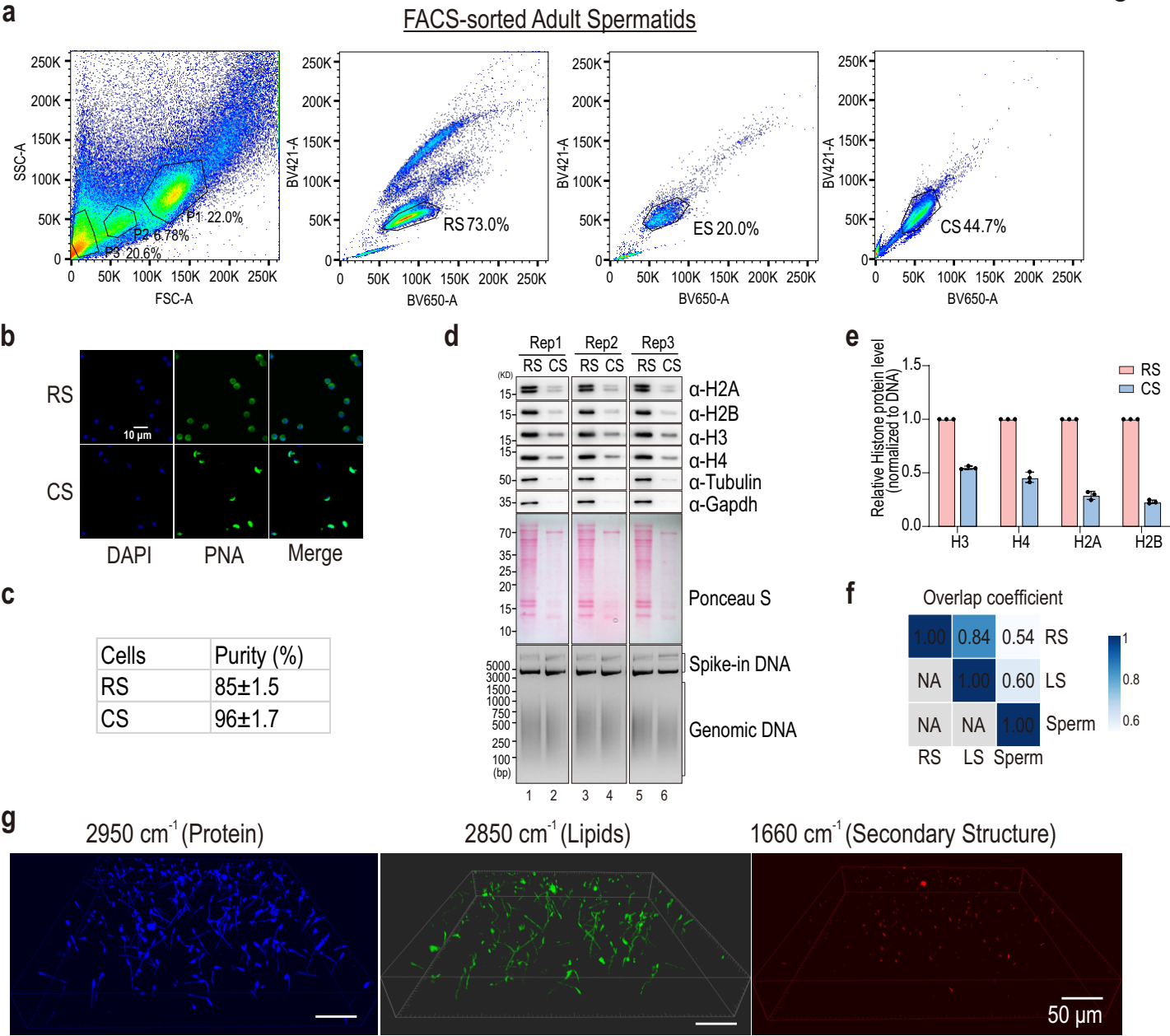




Locations of Stage-Specific Expressed Genes on Each Chromosome Grid

a.Electrophoretic analysis of ATAC-seq libraries after PCR enrichment**b.****c.****d.****e.****f.**





Extended Data Fig. 9

