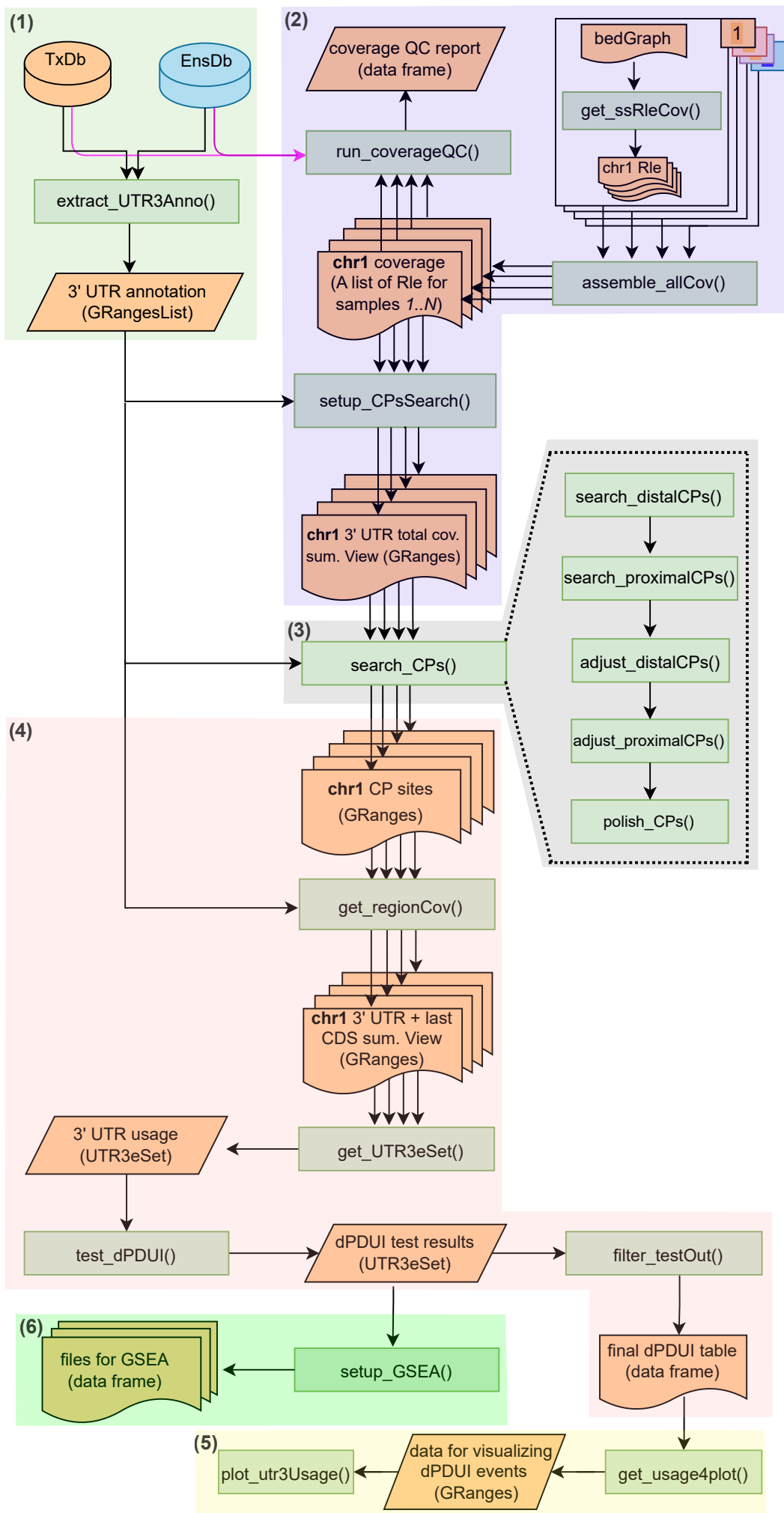
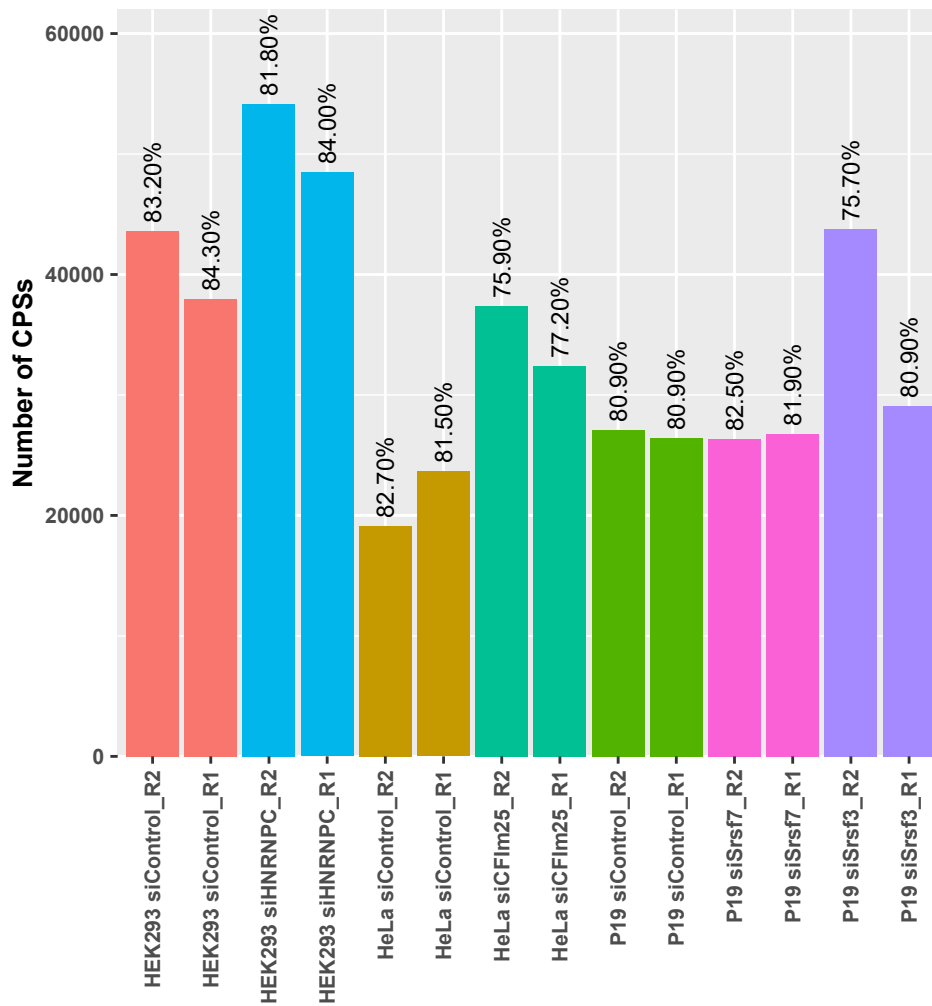


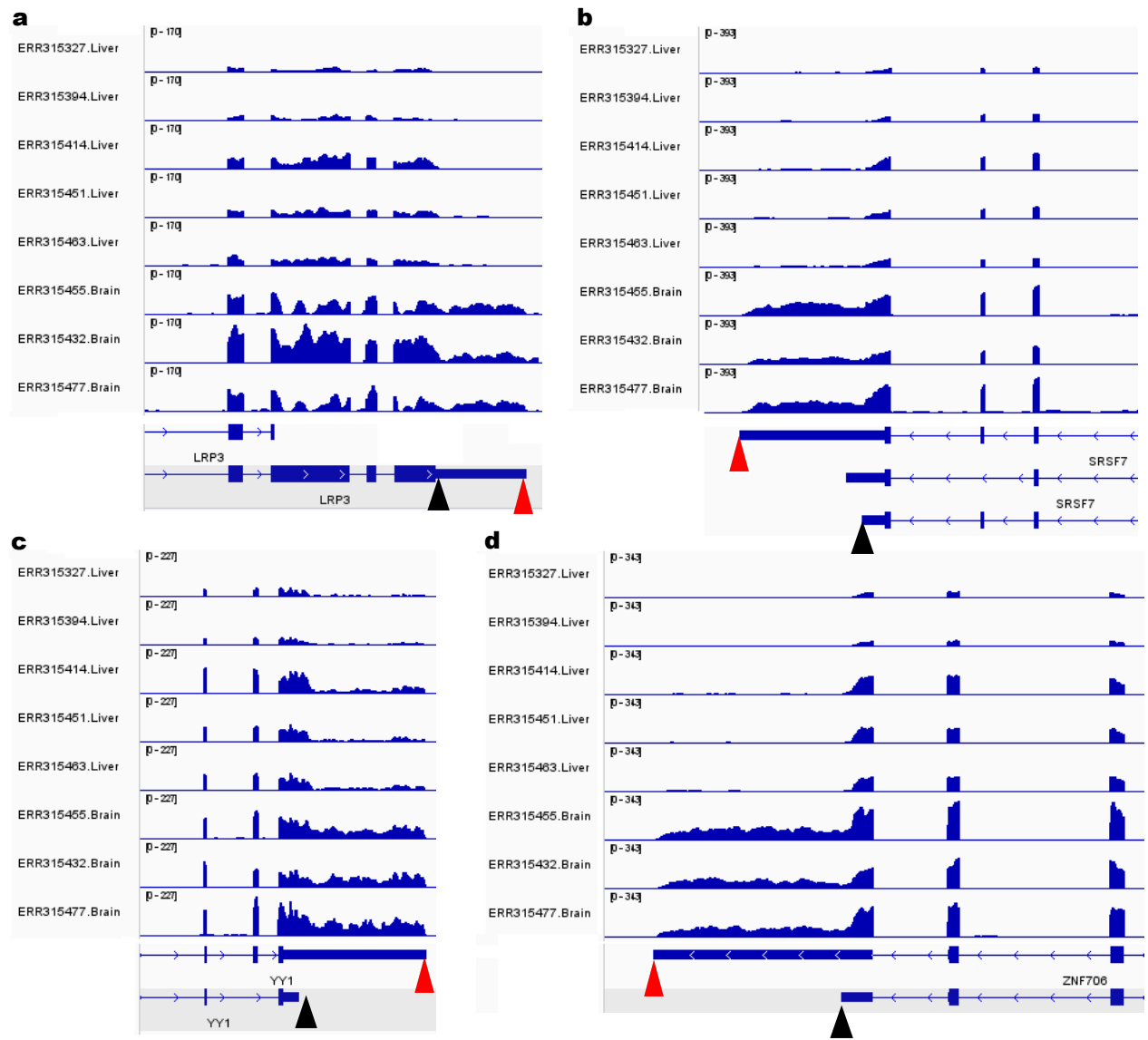
**Fig. S1. Schematic diagram outlining the process of identifying CPSs from RNA-seq data by InPAS.** Shown are the four key steps: (1) Extraction of 3' UTR annotation, (2) Manipulation of coverage data, (3) CPS search, and (4) differential APA analysis. Briefly, the gene model is displayed in the top track, depicting the terminal exon, part of the coding sequence (CDS), and 3' UTR of Gene A, the UTR of a downstream gene (Gene B), and the intergenic region between the two genes. InPAS extracts 3' UTR annotation from the genome annotation file and utilizes the last exons of each transcript as proxies of 3' UTRs, along with the adjacent downstream intergenic region up to 10 kb as extended regions for de novo CPS detection. Raw RNA-seq read coverage from four samples across two experimental groups is depicted in Tracks 2-5, which are merged and normalized by group, and displayed in Tracks 6 and 7. The detection of dCPSs and pCPSs are performed sequentially as depicted in Tracks 8-13. The merged, normalized coverage data along the extracted 3' UTRs and the extended region is scanned from 5' to 3' (Track 8). dCPSs are inferred as described in APATrap. The coverage data in the range from the start of 3' UTR to the predicted dCPS is used for pCPS search (Track 9). pCPSs are determined using an optimal segmentation method, which fits a series of regression models for coverage data assuming each nucleotide in this range as a hypothetical CPS. Mean squared errors (MSEs) of the regression models are shown in Track 10. The locus with a global minimum of MSEs is predicted as the potential pCPS. The predicted dCPS and pCPS can be further adjusted with the naïve Bayes classifier from the CleanUpdTSeq package (Tracks 11 and 12). The final predicted short and long 3' UTRs are shown in Track 13. Differential APA analysis is performed for the short and long 3' UTRs between conditions.



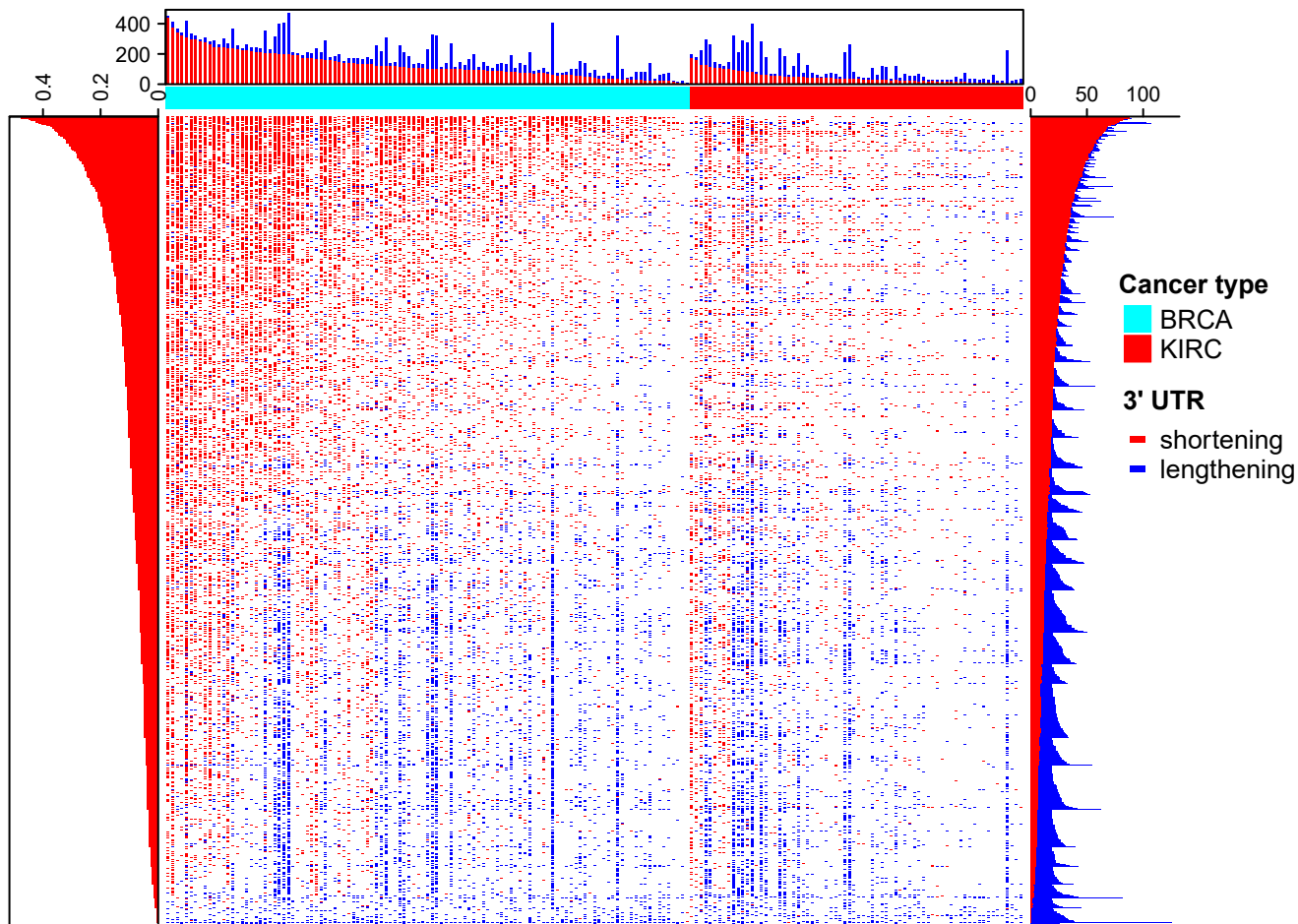
**Fig. S2. Workflow for InPAS-based APA analysis.** The analysis involves a few steps which are color coded: (1) preparation of the 3' UTR annotation, (2) manipulation and quality control of RNAseq coverage data, (3) CPS search, (4) differential APA event detection, (5) APA event visualization, and (6) generation of files for GSEA. Briefly, (1) The 3' UTR annotation is extracted from a TxDb object as well as an EnsDb object and processed into a GRangesList object using the extract\_UTR3Anno function. This annotation is used for subsequent quality control of coverage data and CPS search. This procedure also sets up an RSQLite database for storing metadata and names of intermediate files generated later. (2) The RNA-seq coverage data in the bedGraph format for each sample is then converted into a list of Rle objects with one object per chromosome using the get\_ssRleCov function. The resulting lists of sample-wise Rle objects are re-organized into chromosome-wise lists of Rle objects for each sample using the assemble\_allCov function. Background coverage depth for each transcript with 3' UTR annotation in the GRangesList is calculated, and coverage for 3' UTRs and their potential extension gaps is summed up by condition. Subsequently, the quality of RNA-seq coverage data is assessed using the run\_coverageQC function to check the percentage of genes and 3' UTRs covered by RNA-seq data. (3) CPSs are searched using the search\_CP function. In detail, dCPSs are searched based on the cumulative, normalized coverage across conditions with respect to background cutoffs using the search\_distalCP function. pCPSs are searched subsequently using the search\_proximalCP function. The resulted distal and pCPSs are optionally adjusted with the naïve Bayes classifier from the cleanUpdTSeq package using the adjust\_distalCPs and adjust\_proximalCPs functions. CPSs are further filtered using the polish\_CP function. (4) For each entry of 3' UTR annotation with predicted pCPSs and dCPSs, per sample coverage for the short and long 3' UTRs, and the last CDS segments is calculated using the get\_regionCov function. The coverage data is organized into a UTR3eSet object using the get\_UTR3eSet function and differential APA analysis is performed using the test\_dPDI function. The test outcome is further filtered based on the given cutoff thresholds. (5) Differential APA events of interest can be visualized by using the plot\_utrUsage function. (6) In addition, files for GSEA can be generated using the setup\_GSEA function to facilitate functional analysis of differential APA events.



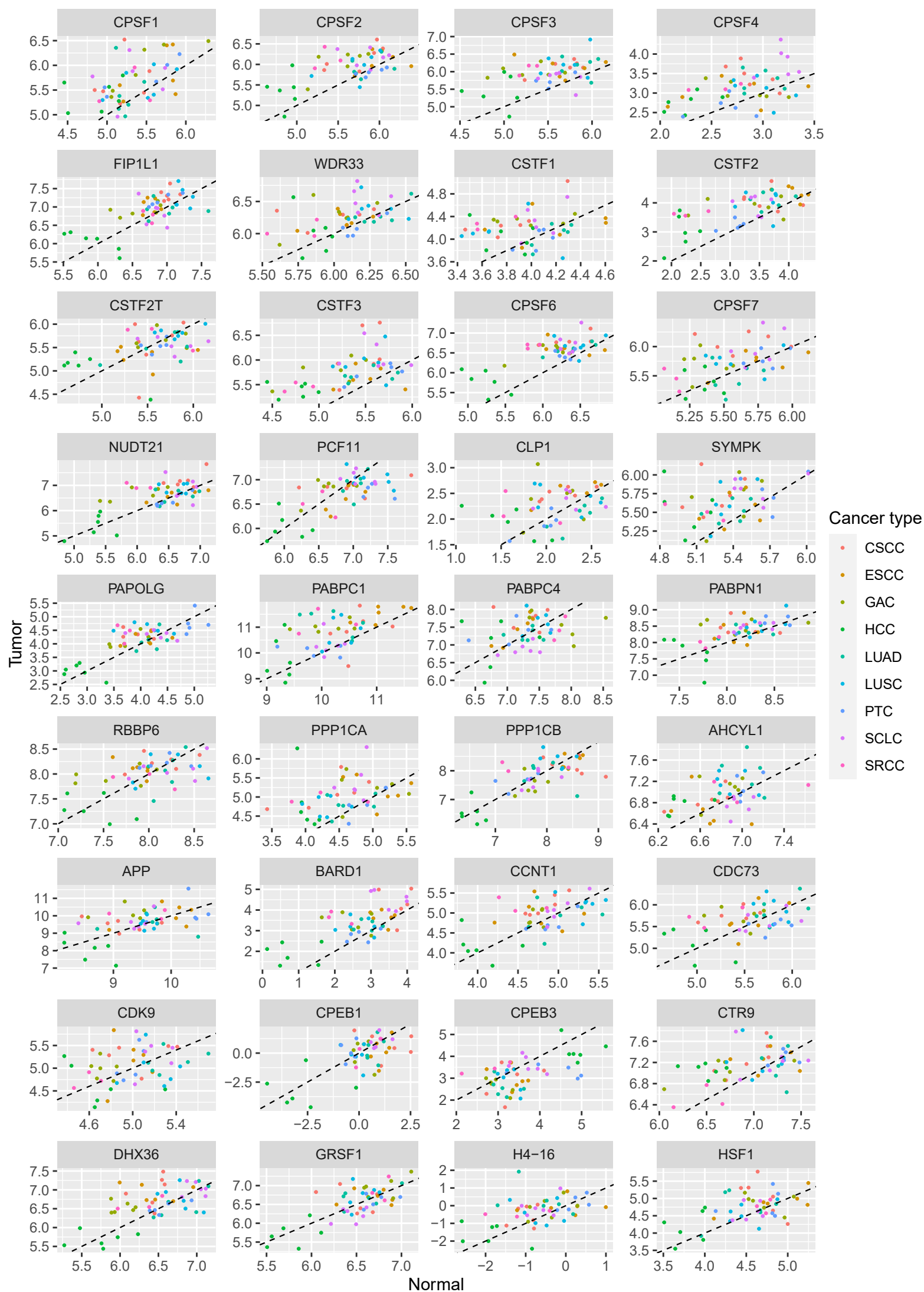
**Fig. S3. Barplot showing the number of CPSs identified by 3'-end RNA-seq in each sample.** Percentages of CPSs overlap the entities in the PolyASite 2.0 database are labelled above each bar for each sample.



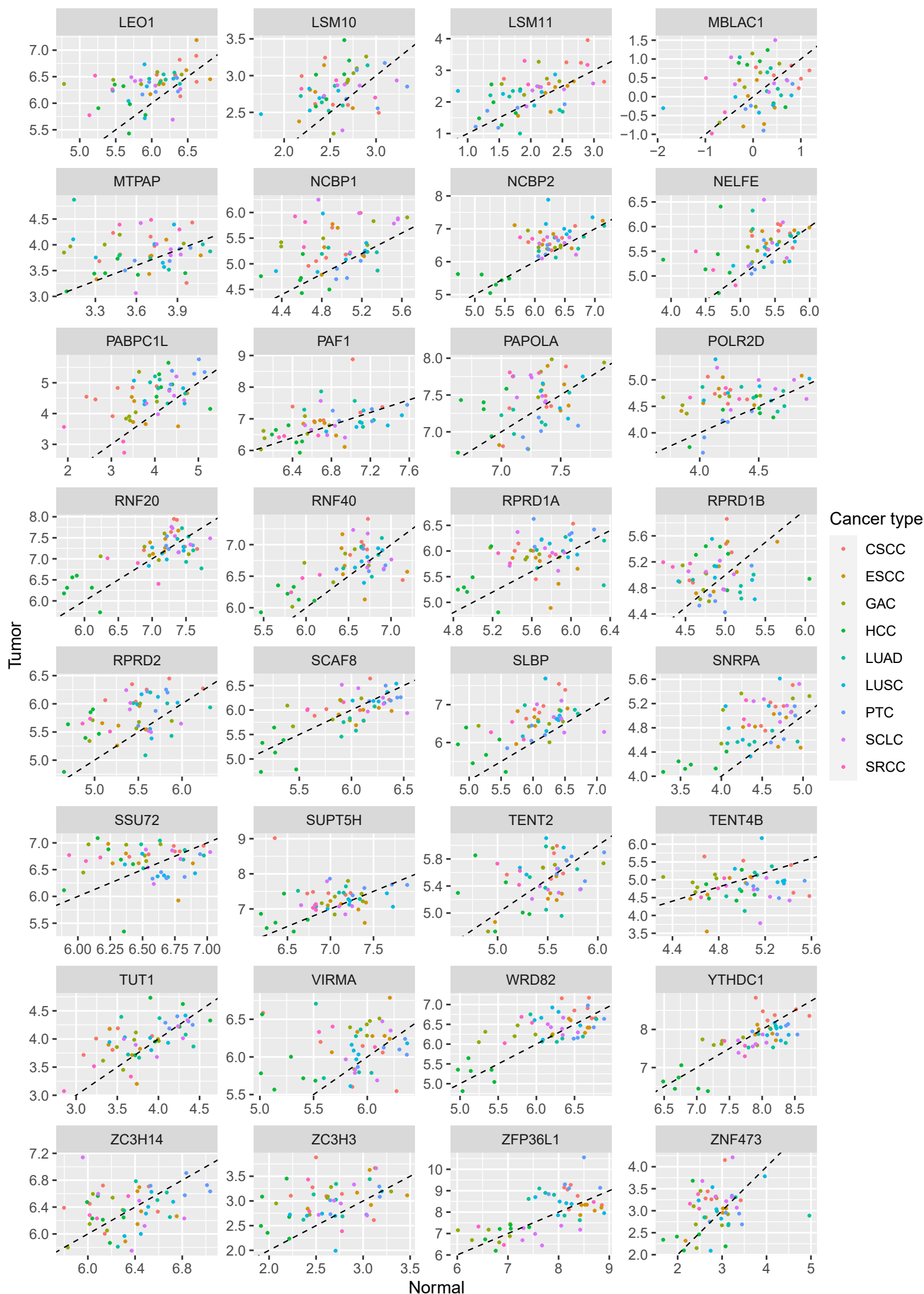
**Fig. S4. Representative differential APA events identified by InPAS between the brain and the liver.** (a-d) Differential APA detected for genes LRP3, S RSF7 (splicing factor), YY1 (transcription factor), and ZNF706 (transcription factor), respectively. In the brain, these genes show a preference for the usage of dCPSs compared to those in the liver. The top 5 tracks show the coverage for the liver samples, and Tacks 6-8 show the coverage for the brain samples. The bottom track shows the gene model. pCPSs and dCPSs determined by InPAS are indicated by black and red triangles, respectively.



**Fig. S5. Differential APA events identified by InPAS in the tumor-normal pairs of the TCGA-BRCA and TCGA-KIRC RNA-seq data.** The central heatmap shows 941 recurrent APA events across the nine cancer types. The top stacking barplot shows the number of 3' UTR shortening (red) and lengthening (blue) events in each tumor sample compared to the matched normal tissue. The left barplot shows the proportions of samples across the nine cancer types with 3' UTR shortening, while the right stacking barplot shows the number of samples across the nine cancer types with 3' UTR shortening (red) and lengthening (blue). BRCA, Breast carcinoma; KIRC, Kidney renal clear cell carcinoma.

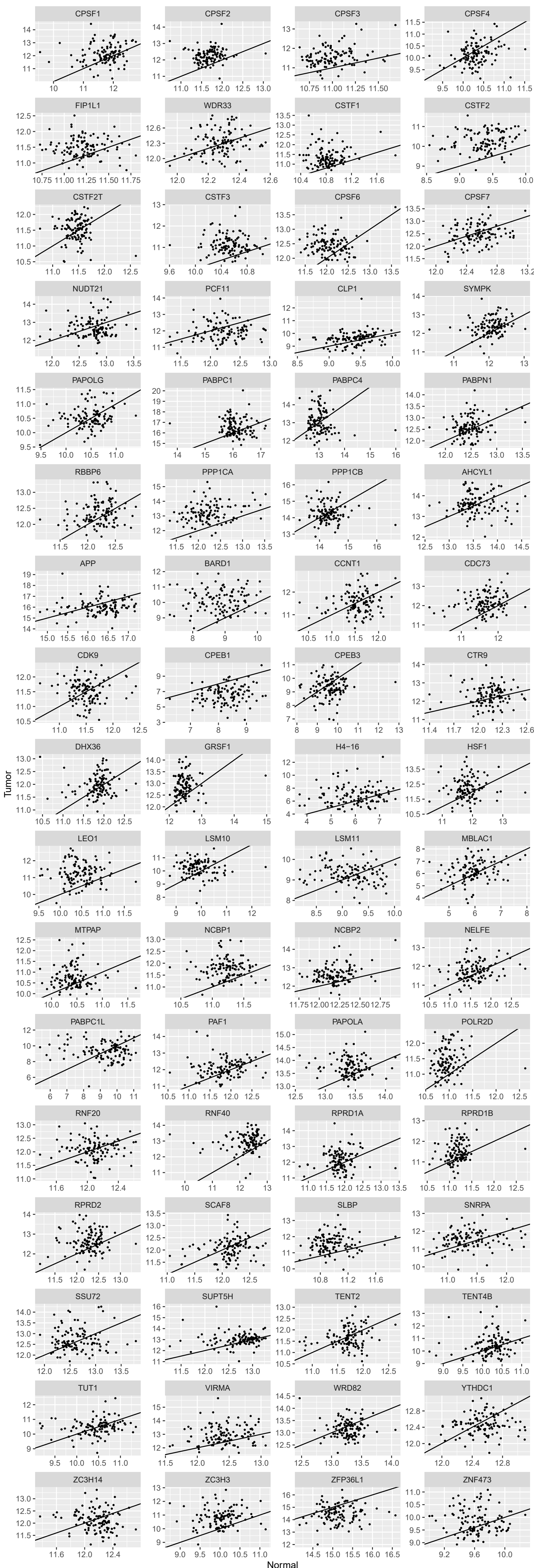


**Fig. S6 Altered expression of genes involved in mRNA 3'-end processing (GO0031124) in tumor samples across nine cancer types compared to the matched normal tissue. (To be continued on next page)**



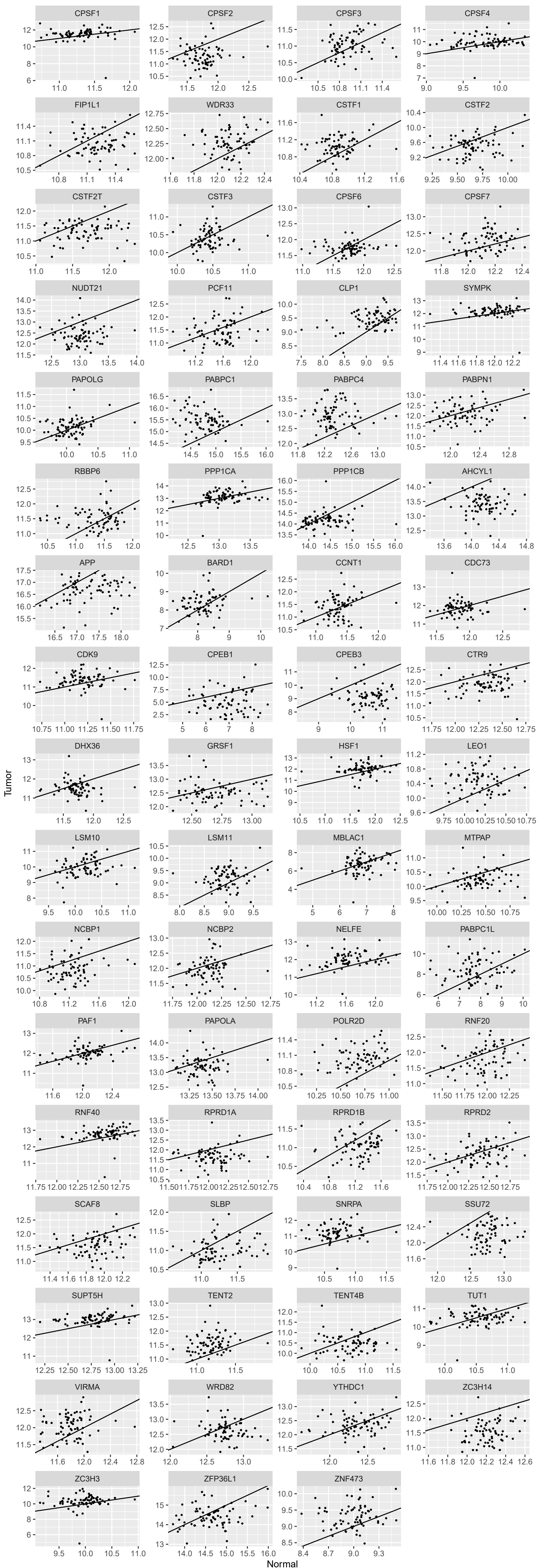
**Fig. S6. Altered expression of genes involved in mRNA 3'-end processing (GO0031124) in tumor samples across nine cancer types compared to the matched normal tissue.** Scatterplots show log2(count) in the normal tissues (x-axis) and tumor samples (y-axis). The line  $y = x$  is included in the graph as a reference. Each cancer type is color-coded as shown on the right. GAC, gastric adenocarcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; HCC, hepatocellular carcinoma; SCLC, small-cell lung carcinoma; CESC, cervical squamous cell carcinoma; ESCC, esophageal squamous cell carcinoma; SRCC, gastric signet-ring cell carcinoma; PTC, papillary thyroid carcinoma.





**Fig. S7. Altered expression of genes involved in mRNA 3'-end processing (GO0031124) in tumor samples of BRCA compared to the matched normal tissue.** Scatter plots show  $\log_2(\text{count})$  in the normal tissues (x-axis) and the tumor samples (y-axis). The line  $y = x$  is included in the graph as a reference.





**Fig. S8. Altered expression of genes involved in mRNA 3'-end processing (GO0031124) in tumor samples of KIRC compared to the matched normal tissue.** Scatter plots show  $\log_2(\text{count})$  in the normal tissues (x-axis) and the tumor samples (y-axis). The line  $y=x$  is included in the graph as a reference.