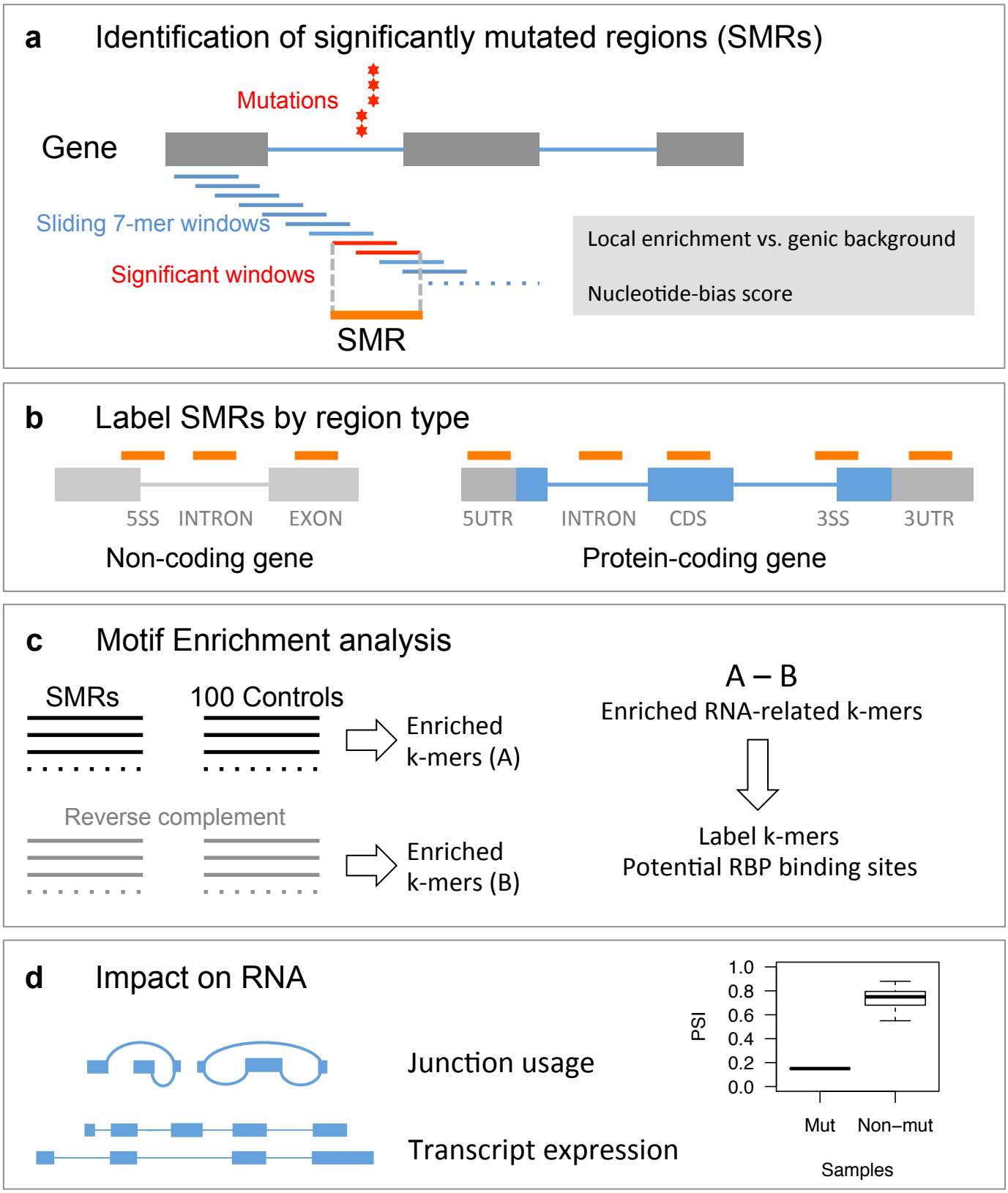
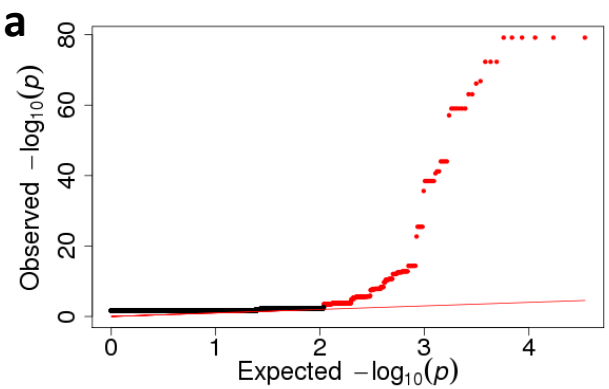
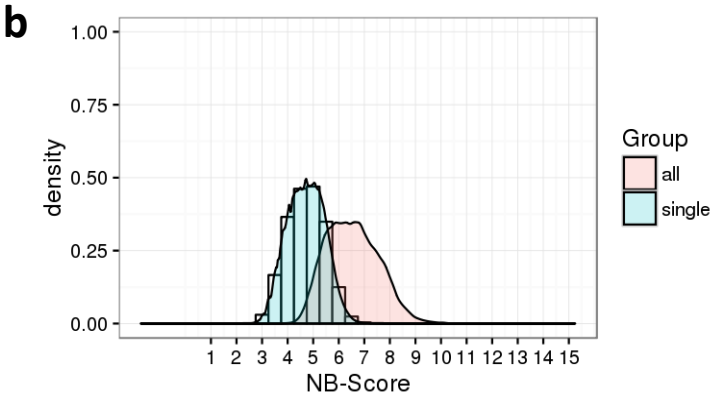


Figure S1. MIRA. Systematic identification of RNA-related significantly mutated regions. **(a)** Short k-mer windows (k=7 in our study) along genes are tested for the enrichment in mutations with respect to the gene mutation rate and the local nucleotide biases. **(b)** Significant windows are clustered by region type into significantly mutated regions (SMRs). **(c)** SMRs are validated by comparison with replication timing and expression, as well as with other methods and a list of candidate cancer gene drivers. **(d)** Association with RNA processing motifs is assessed by identifying enriched motifs that are specific of the gene strands. **(e)** The impact on RNA is evaluated using an outlier statistic to test significant difference between the patient with mutated SMR vs. the patients from the same tumor type but without mutations in the same SMR.





c

	6-mer	7-mer	8-mer	9-mer
Total k-mers	115398	120932	183053	222264
Significant kmers	103689	78352	141440	163121

SMRs	6mer	7mer	8mer	9mer
3UTR	319	295	366	392
5UTR	129	120	141	149
CDS	260	209	312	338
EXON	443	336	490	521
INTRON	23576	17475	26345	27857
5SS	22	16	28	30
3SS	11	10	13	18

Figure S2. Identification of significantly mutated regions (SMRs). **(a)** QQ-plot comparing the distribution of p-values in our calculated 7- mer windows, using all windows with 1 or more mutations, with the uniform distribution. In red we indicate those that we are taking as significant: 3 or mutations and corrected p-value < 0.05. This plot was built with 20 protein-coding genes. **(b)** Comparison of the distributions of nucleotide bias (NB) scores in 7-mer windows with 1 mutation (blue) and in 7- mer windows with 3 or more mutations. We selected 7-mer windows with 3 or more mutations and with NB-score >= 6. **(c)** For k=6,7,8 and 9 we give the total number of k-mer windows with 3 or more mutations (Total k-mers), total number of significant k-mers (corrected p-value < 0.05 and NB-score > 6), and the resulting number of SMRs for each k, running MIRA with the same parameters.

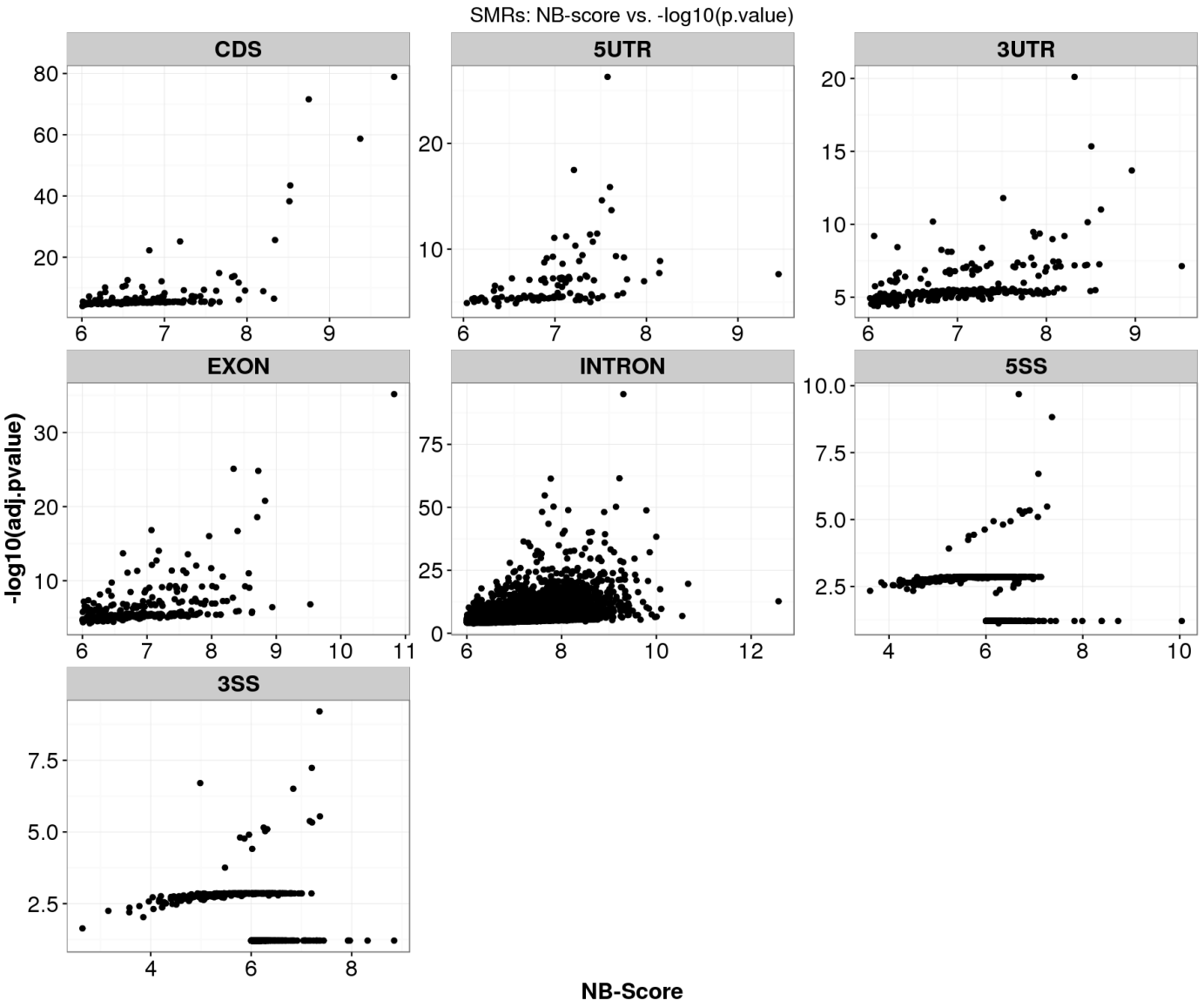


Figure S3. Significantly mutated regions (SMRs). Distribution of the nucleotide-bias (NB) scores (x axis) and p-values (in –log10 scale) (y axis) of the identified SMRs from the PAN505 dataset separated by region type.

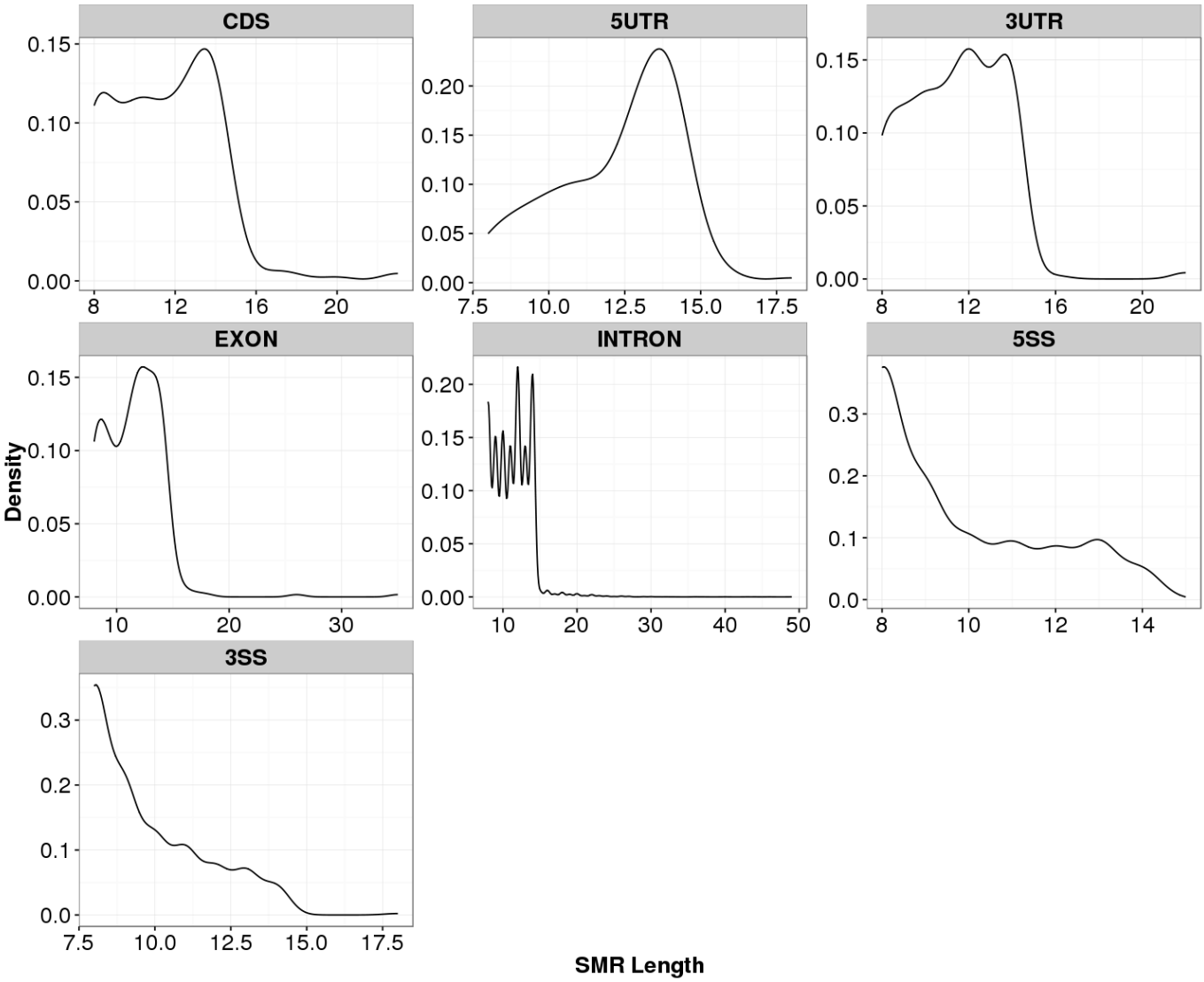


Figure S4. Length distribution of SMRs. Length distribution of the identified significantly mutated regions (SMRs) for the PAN505 dataset separately for each region type.

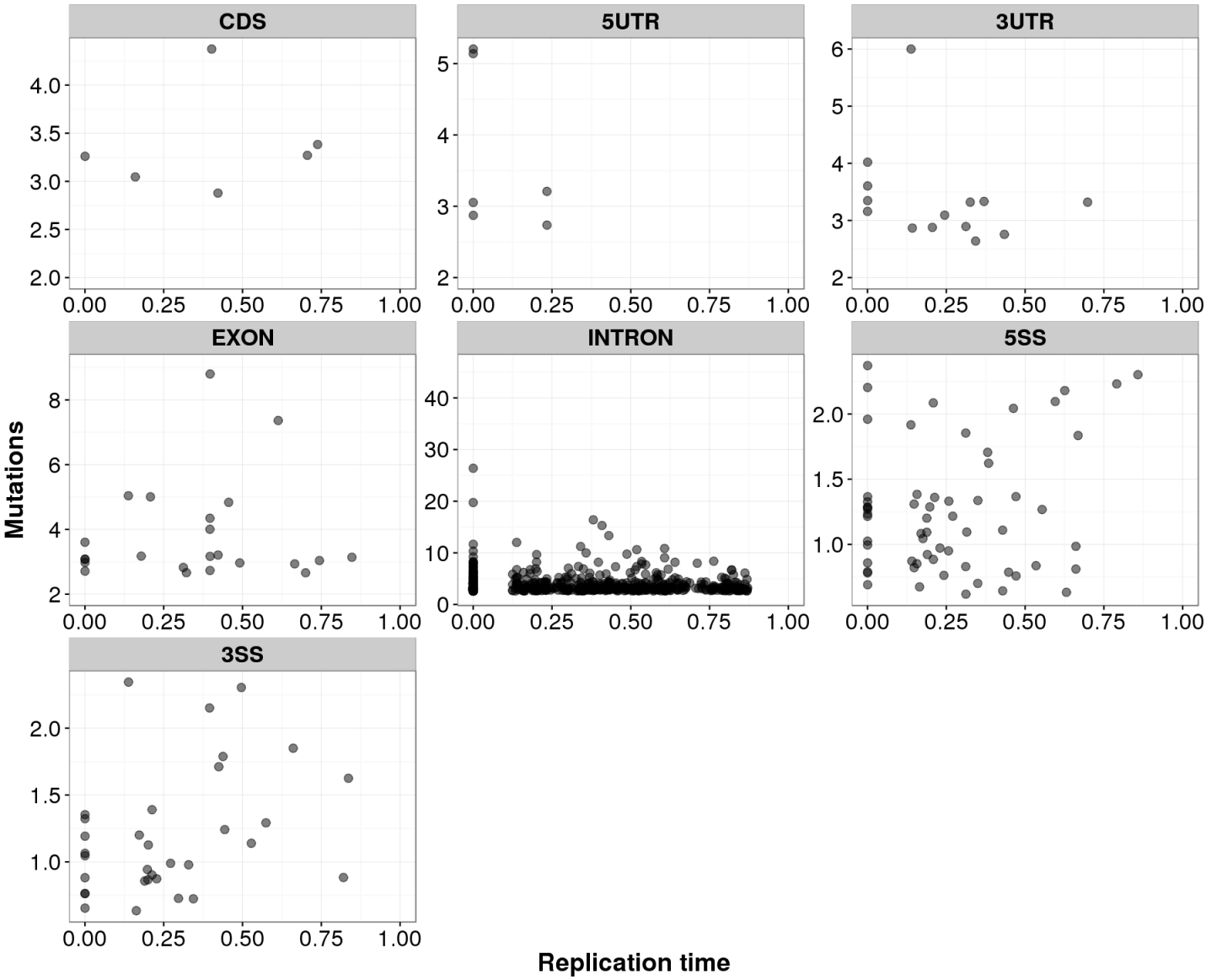


Figure S5. Comparison of mutation count with replication timing. Relation between replication timing (x-axis) and number of mutations (y-axis) in the SMRs from the PAN505 cohort falling in regions with replication timing data from (Lochovsky et al. 2015).

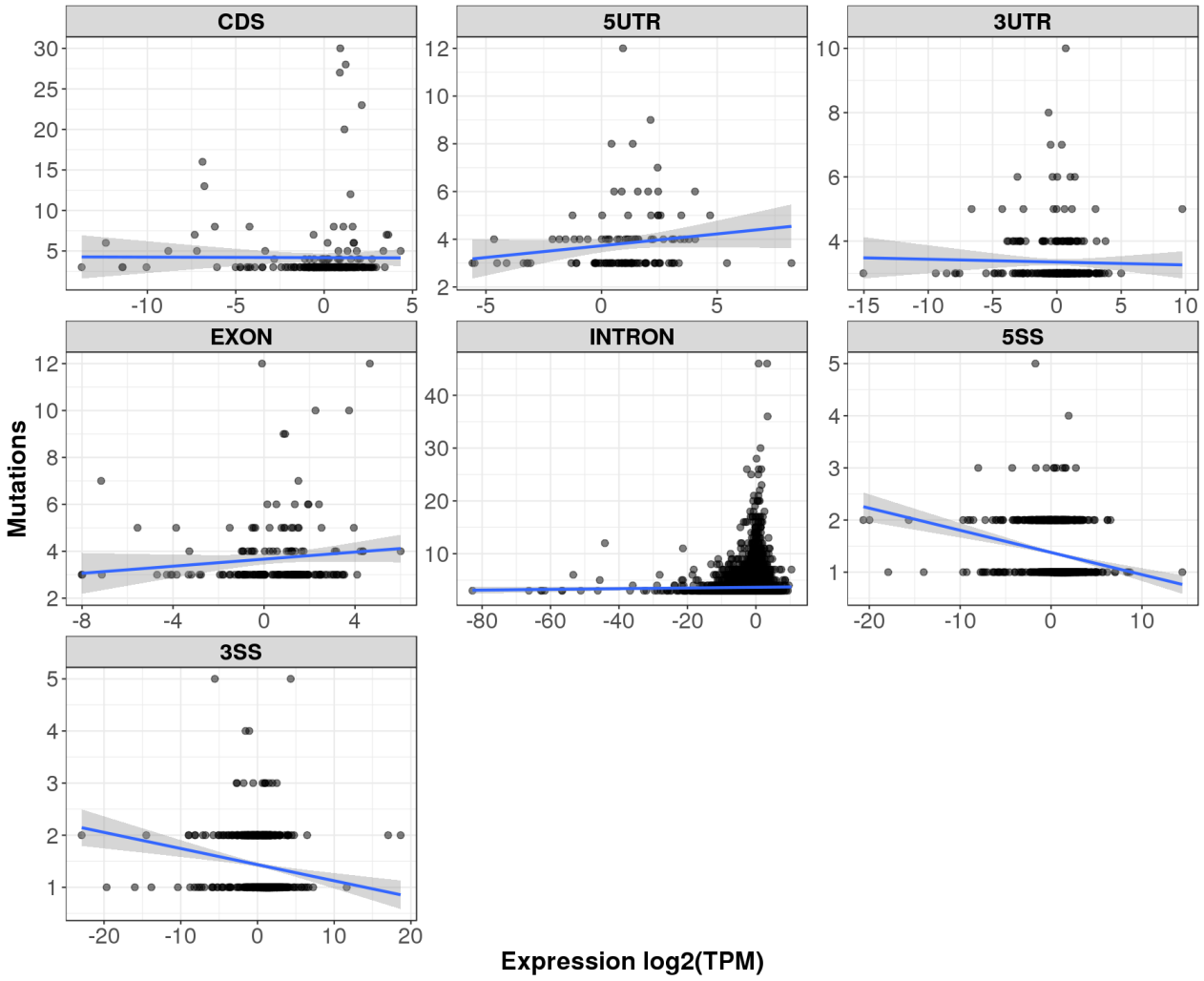


Figure S6. Comparison between expression and mutation count. For every SMR in the PAN505 cohort, we compared the average expression of the transcripts, in log10(TPM) units (x axis), with the mutation count (y axis), within the same patients harboring the mutations in the SMR. For each transcript including an SMR, we considered the expression of the transcript in the same patients that harbor the mutation. These expression values were averaged over all patients. The Pearson's correlation values (R) of the comparisons were: -0.16 (3SS), -0.22 (5SS), -0.02 (3UTR), 0.15 (5UTR), 0.00 (CDS), 0.11 (EXON), 0.01 (INTRON).

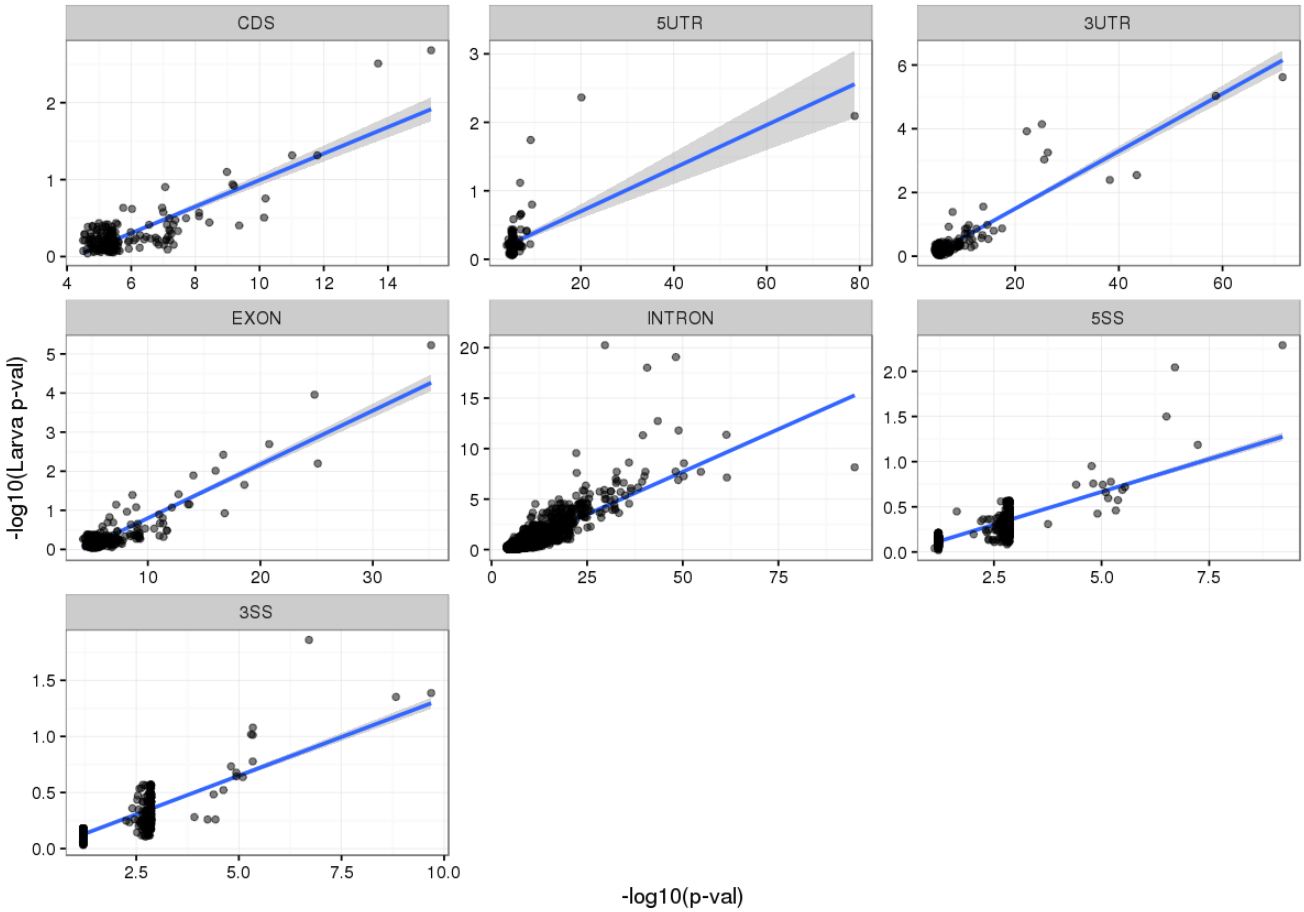


Figure S7. Comparison with LARVA. For each of our predicted SMRs, we compare in –log10 scale the p-value we provided with the p-value provided by LARVA (Lochovsky et al. 2015) using a model that accounts for over-dispersion of the mutation rate and replication timing (p.bbd.cor, see Methods). The Pearson's correlation values (R) of the comparisons were 0.84 (CDS), 0.87 (5UTR), 0.82 (3UTR), 0.90 (EXON), 0.83 (INTRON), 0.83 (5SS), 0.80 (3SS).

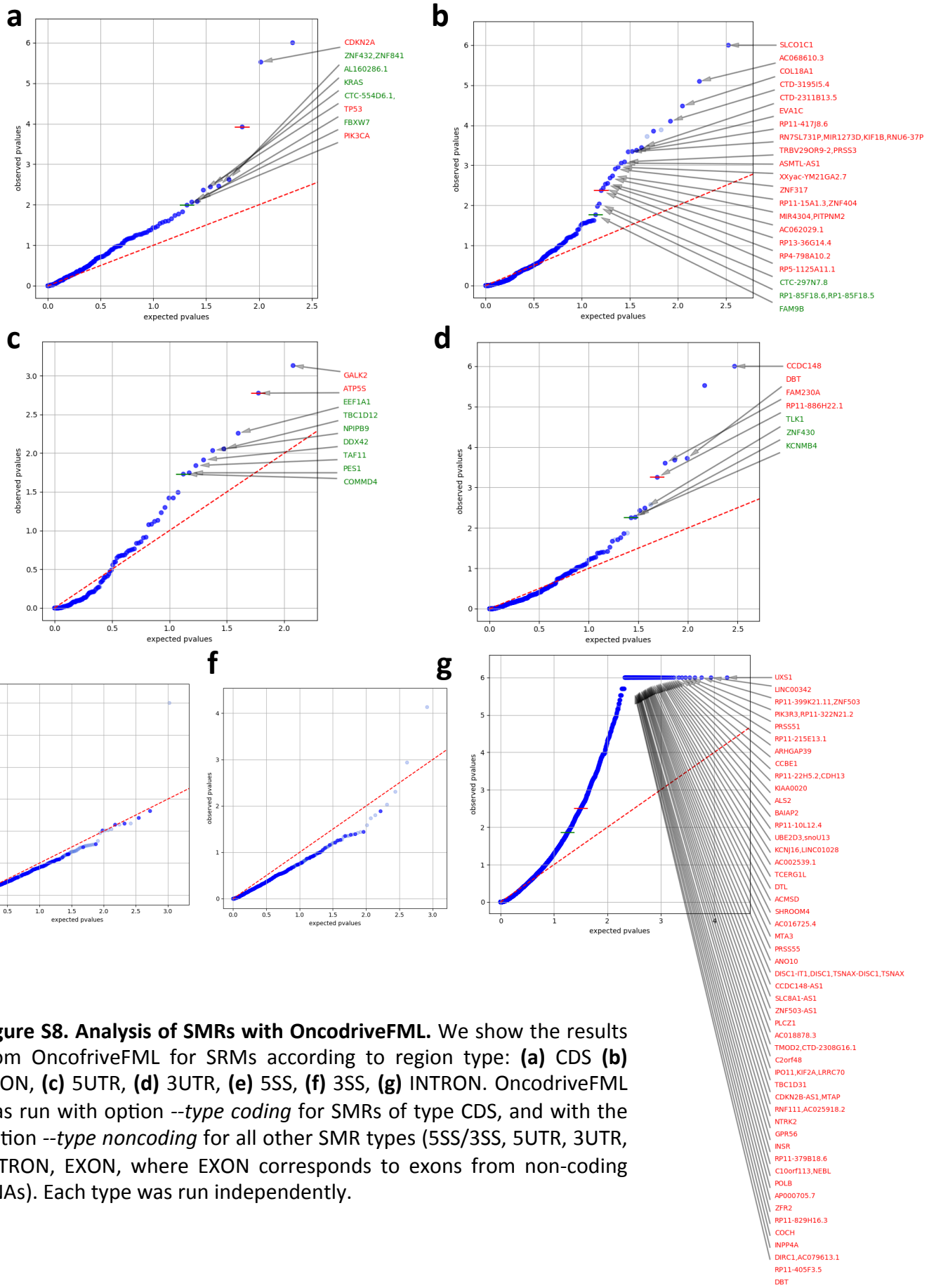


Figure S8. Analysis of SMRs with OncodriveFML. We show the results from OncofriveFML for SRMs according to region type: **(a)** CDS **(b)** EXON, **(c)** 5UTR, **(d)** 3UTR, **(e)** 5SS, **(f)** 3SS, **(g)** INTRON. OncodriveFML was run with option --*type coding* for SMRs of type CDS, and with the option --*type noncoding* for all other SMR types (5SS/3SS, 5UTR, 3UTR, INTRON, EXON, where EXON corresponds to exons from non-coding RNAs). Each type was run independently.

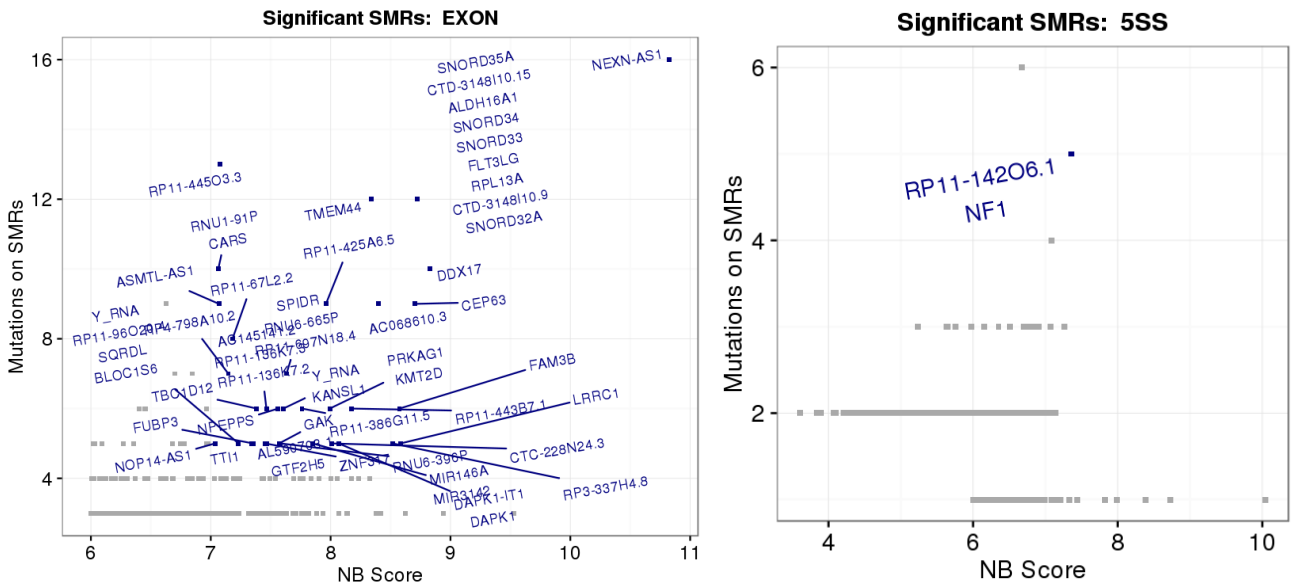


Figure S9. **EXON and 5SS SMRs.** For EXON and 5SS SMRs we give the nucleotide biase (NB) score (x axis) and the number of mutations (y axis). We show the gene name for the SMRs with NB-score > 6 and with 5 or more mutations.

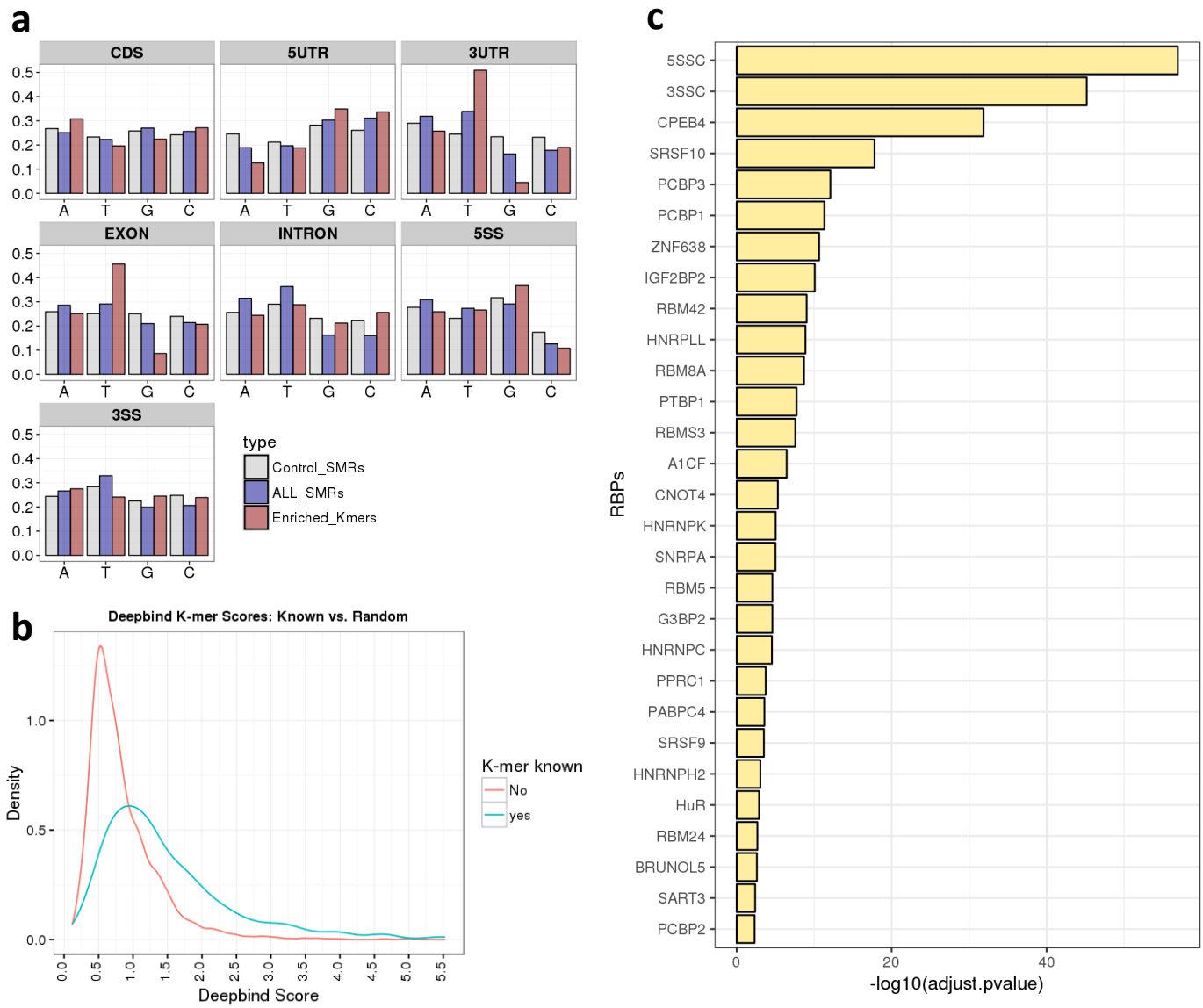


Figure S10. **Motif enrichment in SMRs.** Association with RNA processing motifs is assessed by identifying enriched motifs that are specific of the gene strands (Figure S1c). **(a)** For each region type we give the proportion of A, C, G and T (y axis) for found enriched k-mers in SMRs (k=6), all SMRs, and control regions built by sampling sequences of same length and GC content as the SMRs from the same region type (Methods). **(b)** Distribution of Deepbind scores on all 6-mers separated according to whether they have been identified previously as RBP binding motif (yes) or not (no). To label k-mers as known motif we checked whether each 6-mer was identical or included in any of the motifs from AttractDB (https://attract.cnic.es/) (Giudice et al. 2016). **(c)** RBP labels assigned to SMRs that show significant association to CLIP signals (see Methods).

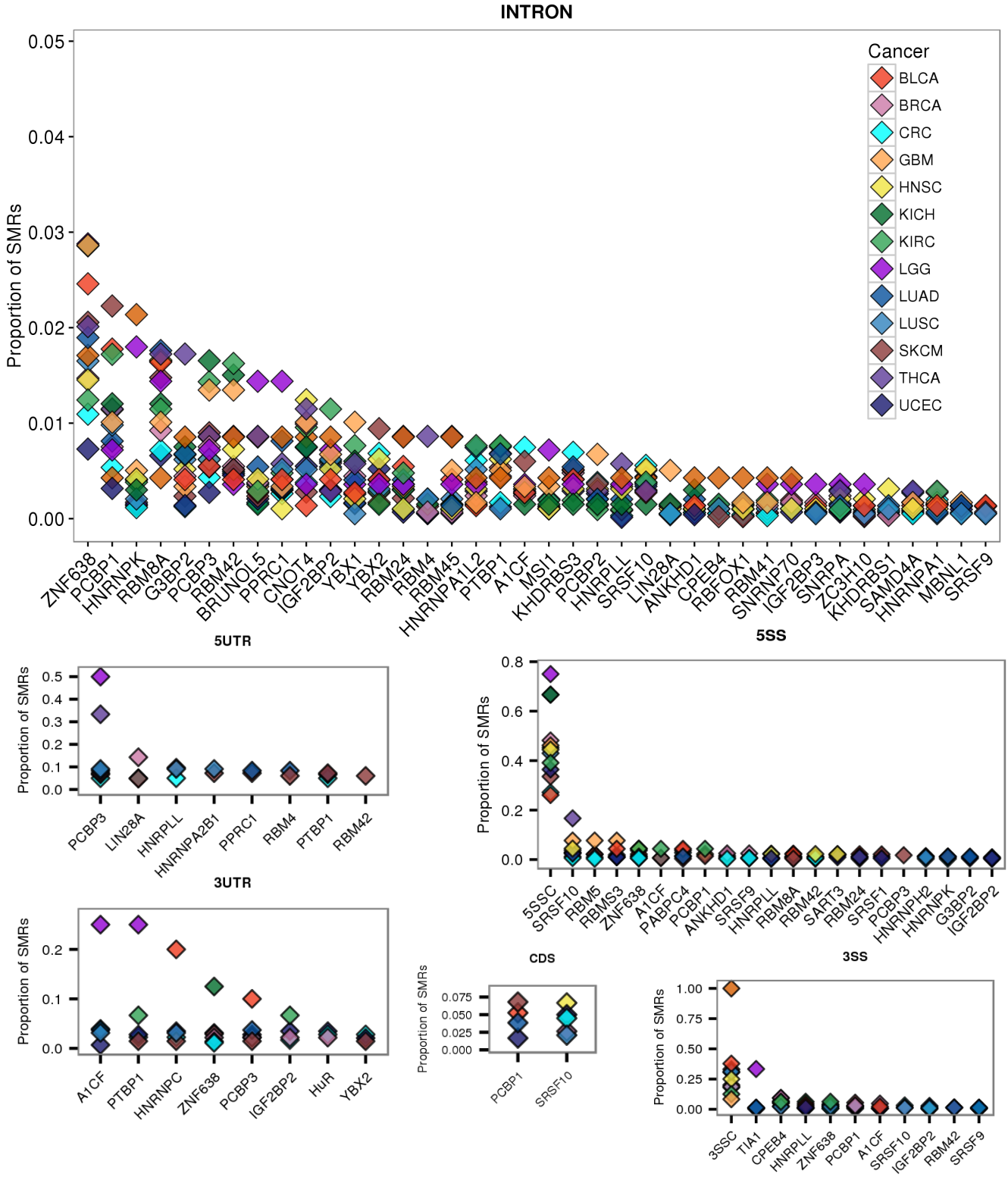


Figure S11. Proportion of SMRs with enriched motifs. For the region types INTRON, 5SS, 3SS, tUTR and 3UTR, and for each cancer type (coded by color), we give the proportion (y axis) of SMRs that contain at least one mutated RBP motif (x axis).

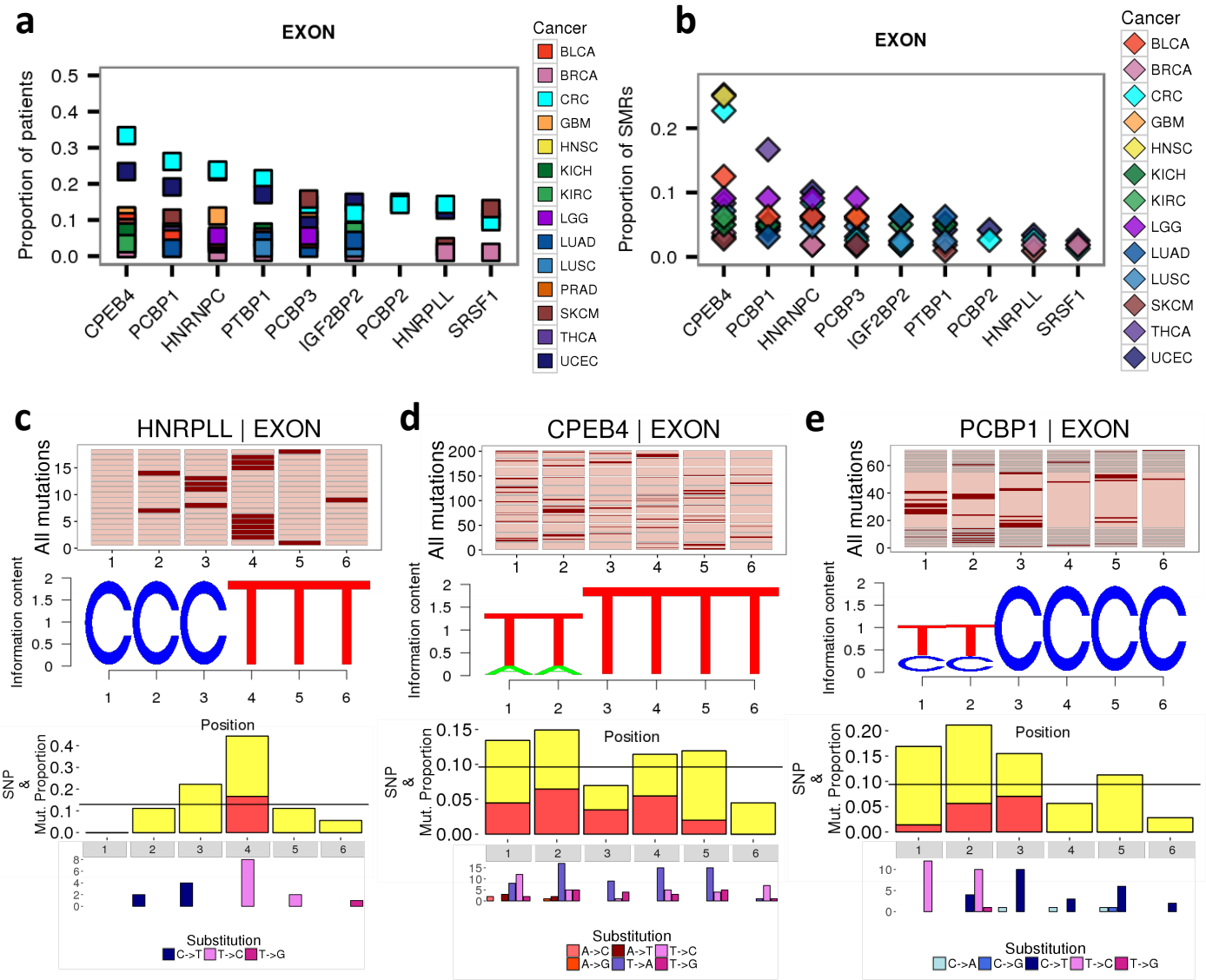


Figure S10. **Mutations on RBP motifs on EXON SMRs.** For the EXON (exons in long-non coding RNA genes) SMRs and for each cancer type (coded by color), we give in the y axis the proportion patients (a) and of SMRs (b) that contain at least one mutated RBP motif (x axis). We show he mutation patterns on HNRPLL **(c)**, CPEB4 **(d)** and PCBP1 **(e)** motifs. In the upper panels we indicate in red the positions covered by the motif and in dark red the position of the mutations. The barplots below show the proportion of somatic mutations (y axis) that fall on each position along the motif logo. In orange we indicate those somatic mutations that coincide with a germline SNP in position (with a different substitution pattern, as the exact matching substitutions were removed). Below we show for each position, the number of motifs with each type of substitution indicated with a color code below.

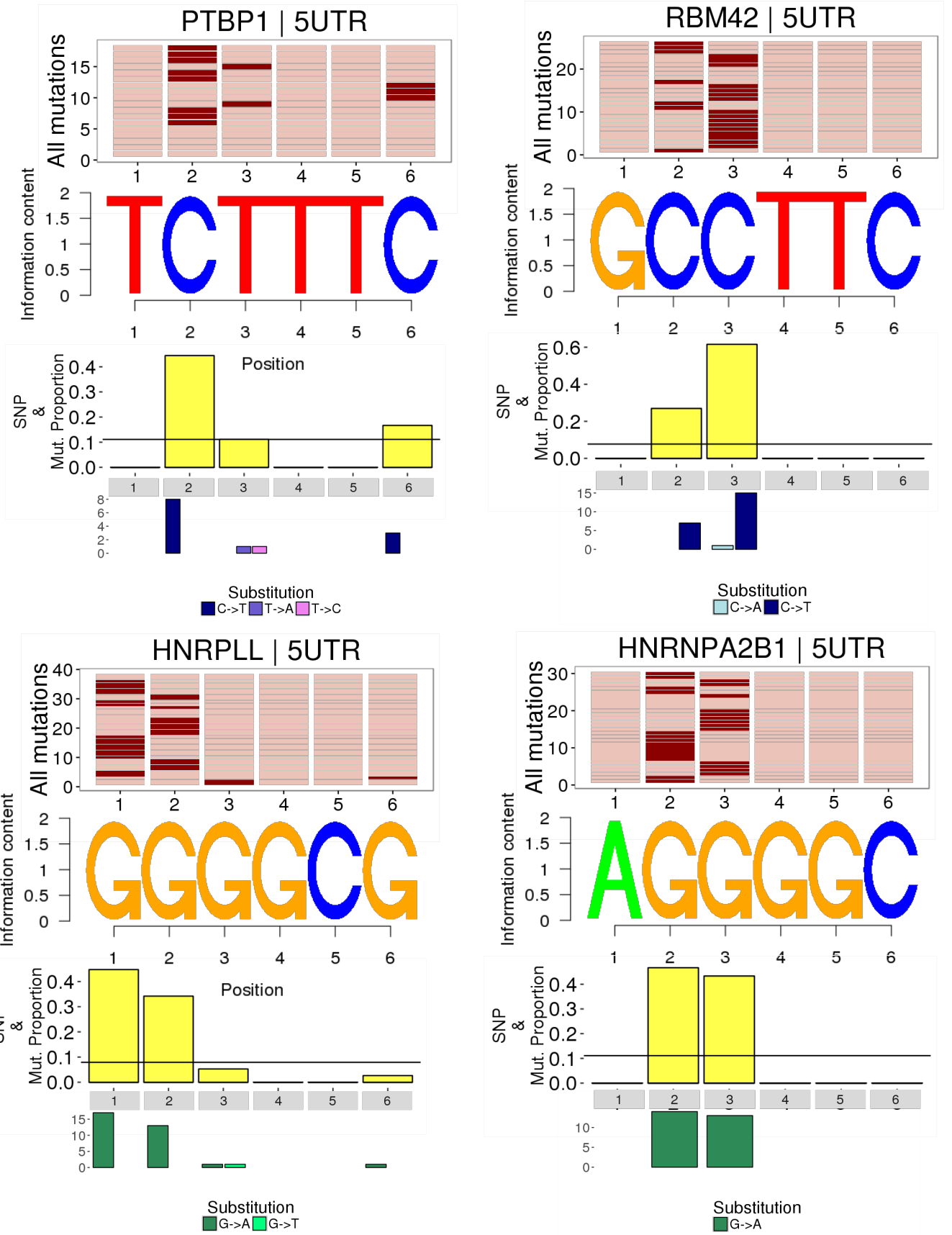


Figure S11. Mutation patterns on motifs in 5UTRs. We show the mutation patterns on PTBP1, RBM42, HNRPLL, HNRNPA2B1 motifs in 5UTR SMRs. In the upper panels we indicate in red the positions covered by the motif and in dark red the position of the mutations. The barplots below show the proportion of somatic mutations (y axis) that fall on each position along the motif logo. In orange we indicate those somatic mutations that coincide with a germline SNP in position (with a different substitution pattern, as the exact matching substitutions were removed). Below we show for each position, the number of motifs with each type of substitution indicated with a color code below.

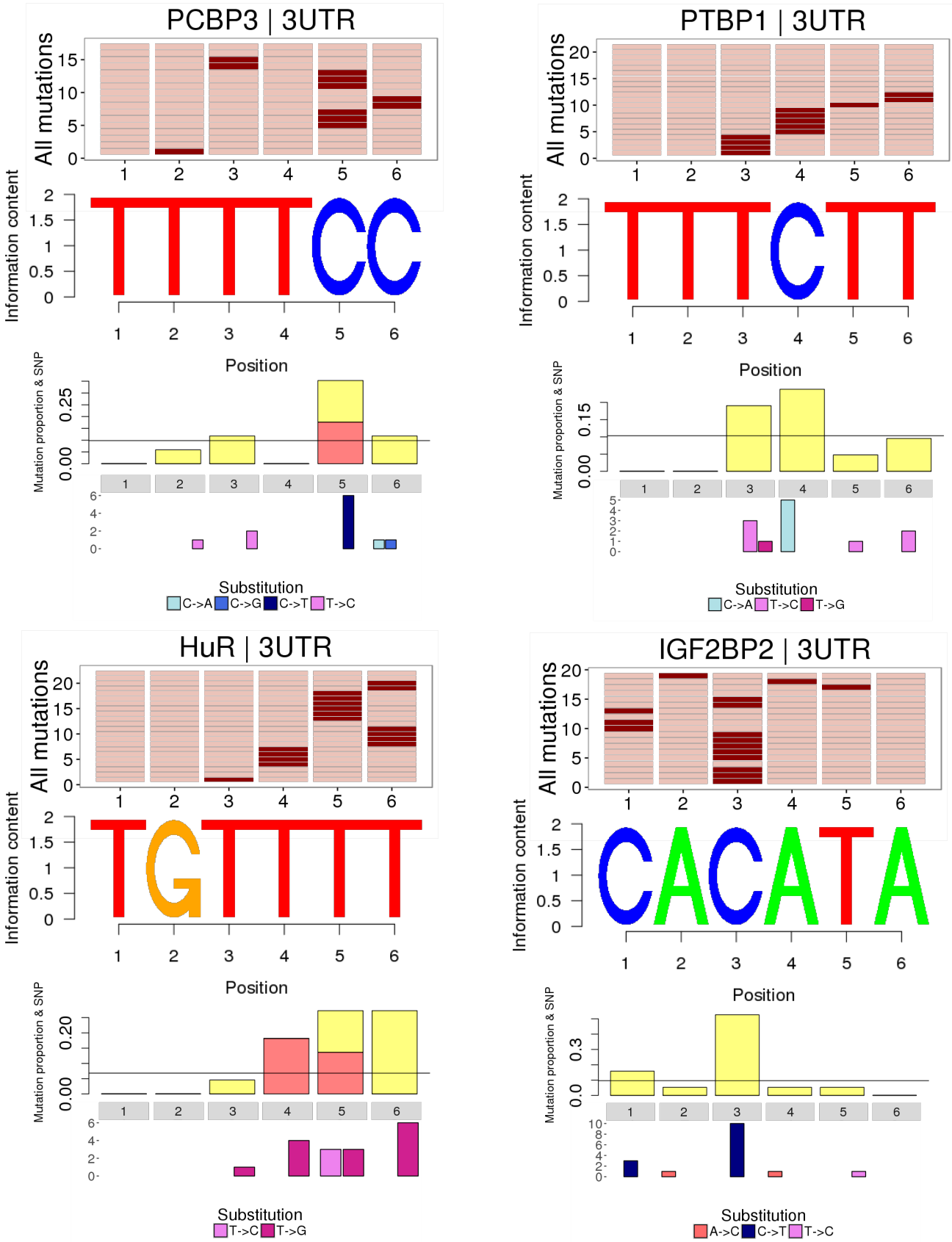


Figure S12. Mutation patterns on motifs in 3UTRs. We show the mutation patterns on PCBP3, PTBP1, HuR (ELAVL1), IGF2BP2 in 3UTR SMRs. In the upper panels we indicate in red the positions covered by the motif and in dark red the position of the mutations. The barplots below show the proportion of somatic mutations (y axis) that fall on each position along the motif logo. In orange we indicate those somatic mutations that coincide with a germline SNP in position (with a different substitution pattern, as the exact matching substitutions were removed). Below we show for each position, the number of motifs with each type of substitution indicated with a color code below.

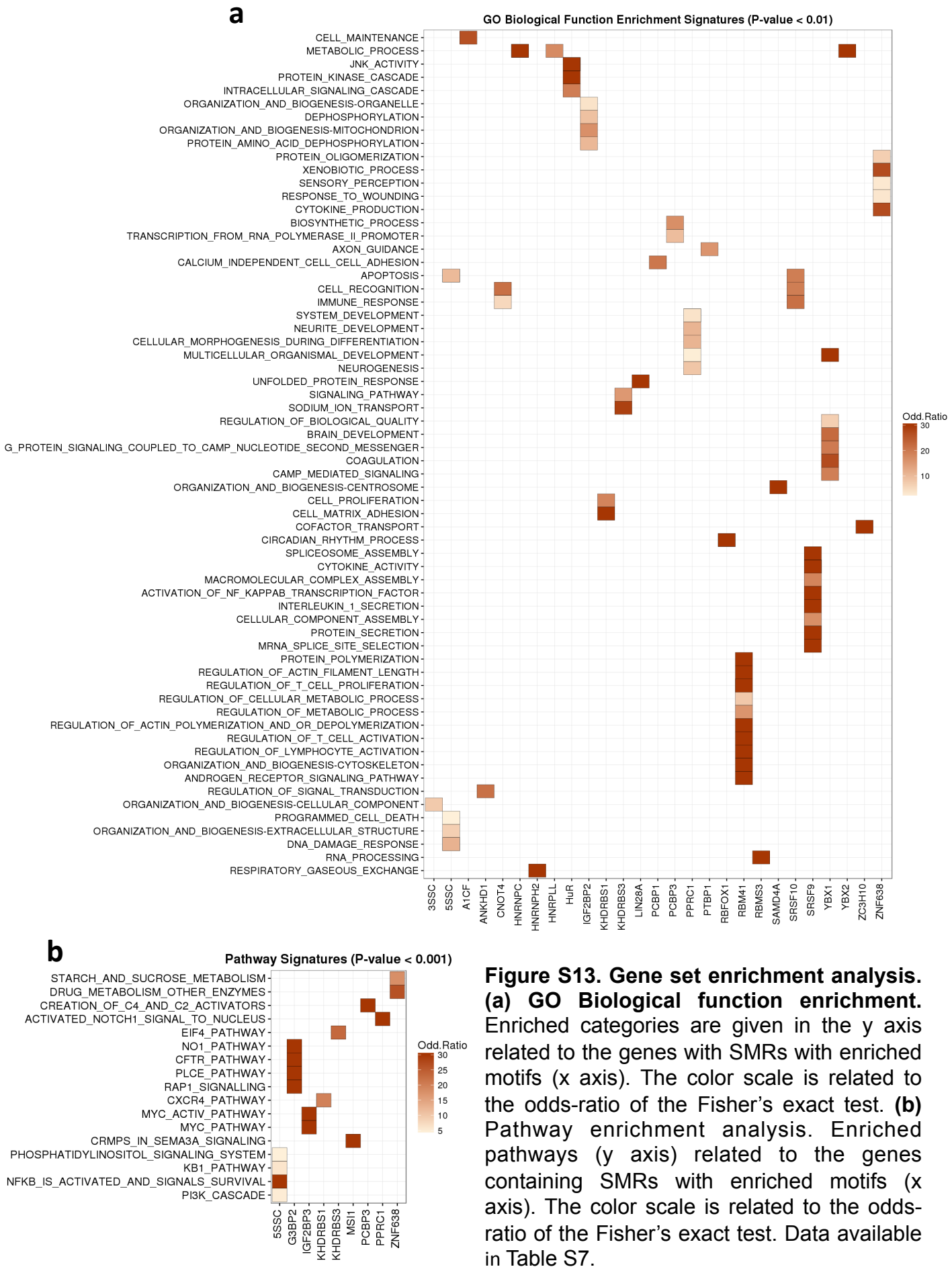


Figure S13. Gene set enrichment analysis. (a) GO Biological function enrichment. Enriched categories are given in the y axis related to the genes with SMRs with enriched motifs (x axis). The color scale is related to the odds-ratio of the Fisher's exact test. **(b)** Pathway enrichment analysis. Enriched pathways (y axis) related to the genes containing SMRs with enriched motifs (x axis). The color scale is related to the odds-ratio of the Fisher's exact test. Data available in Table S7.

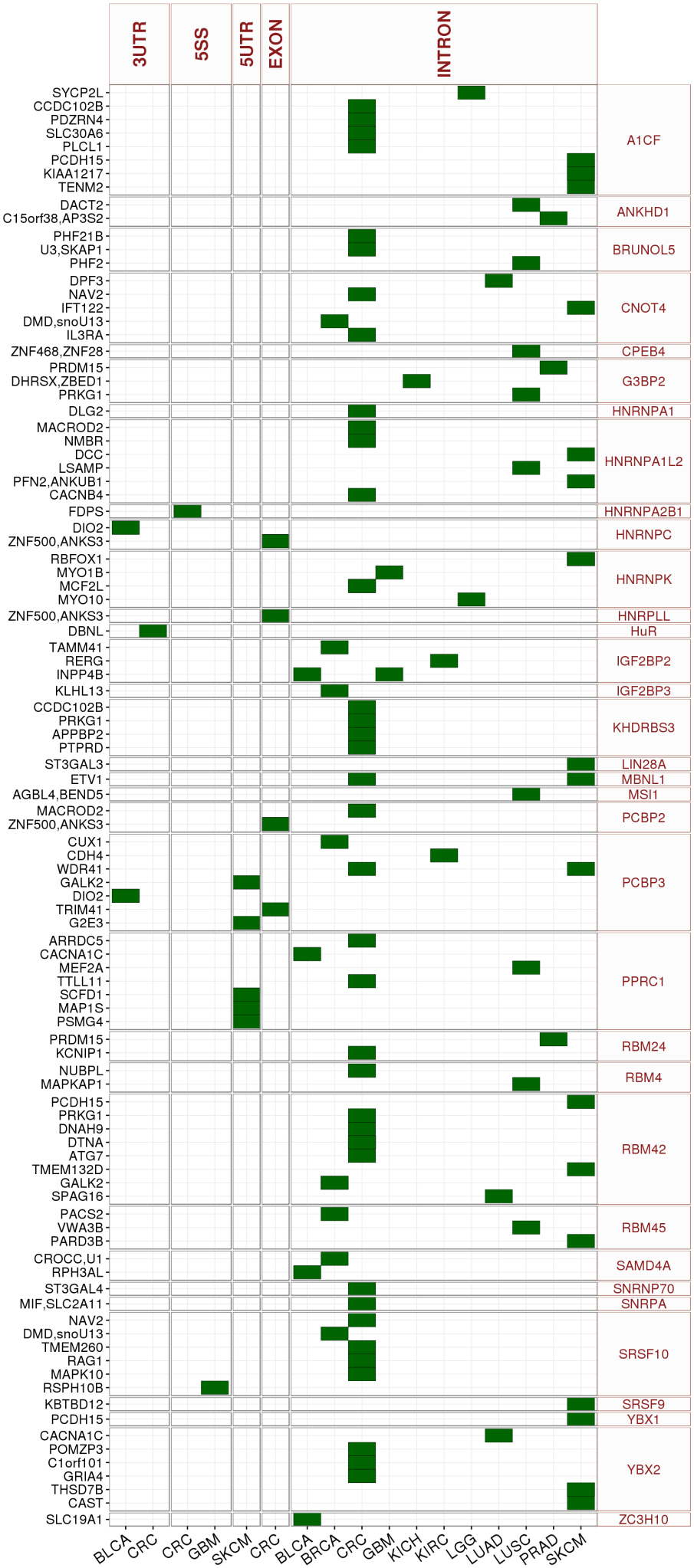


Figure S14. Significant changes in transcript expression associated to mutations in enriched motifs in SMRs. We show the genes (y axis) that have a significant change, separated by tumor type (x axis). Corresponding motifs are given on the right. The most numerous cases (PTBP1, PCBP1, RBM8A and ZNF638) are given in Figure S14.

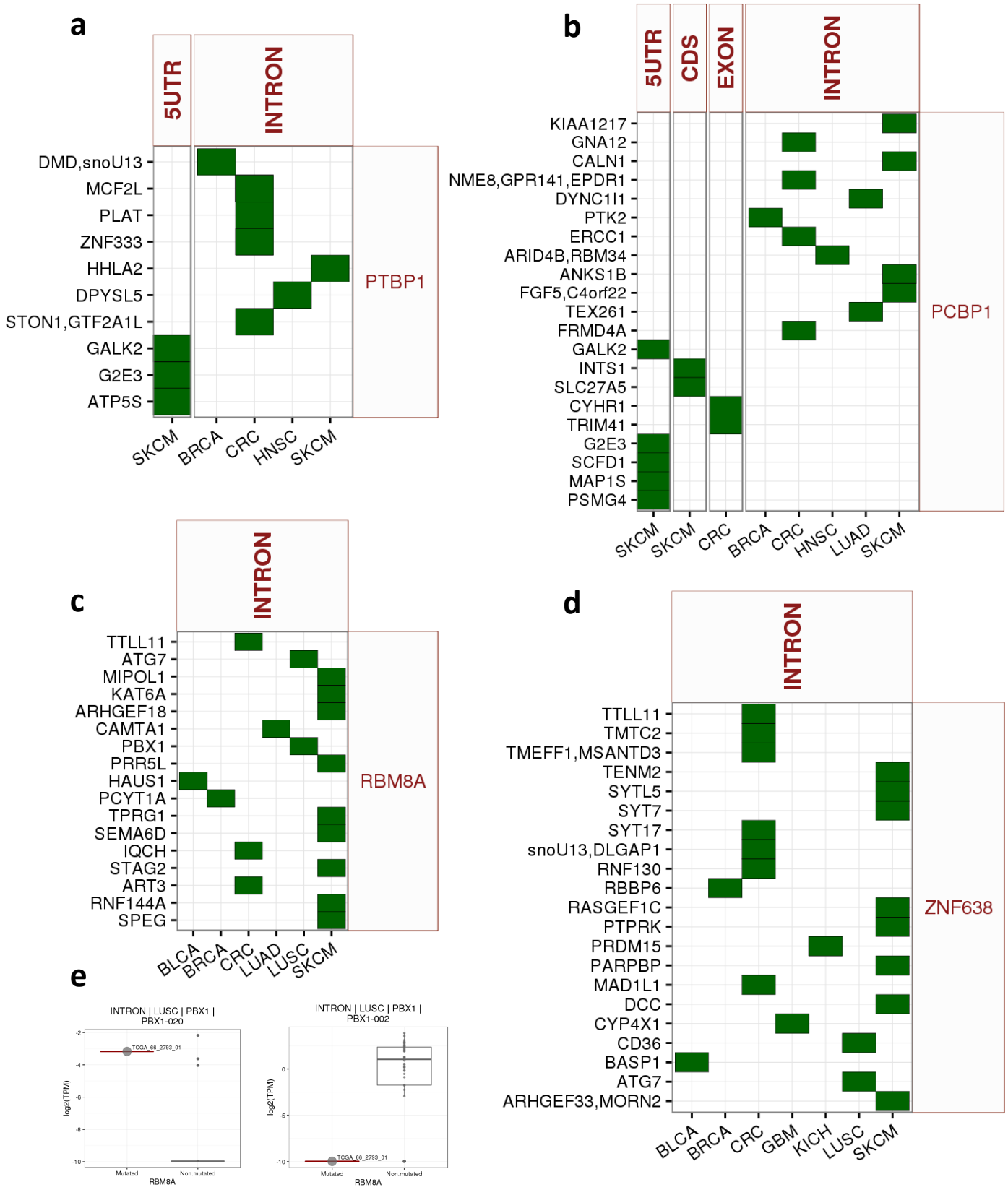


Figure S15. Significant changes in transcript expression associated to mutations in enriched motifs at intronic SMRs. We show the cases in the motifs PTBP1 **(a)**, PCBP1 **(b)**, RBM8A **(c)**, and ZNF638 **(d)**. We show the genes (y axis) that have a significant change, separated by tumor type (x axis). **(e)** Expression change in two transcripts from the cancer gene PBX1 associated to mutations in RBM8A motif in lung squamous cell carcinoma (LUSC). The two transcripts change in opposite directions.

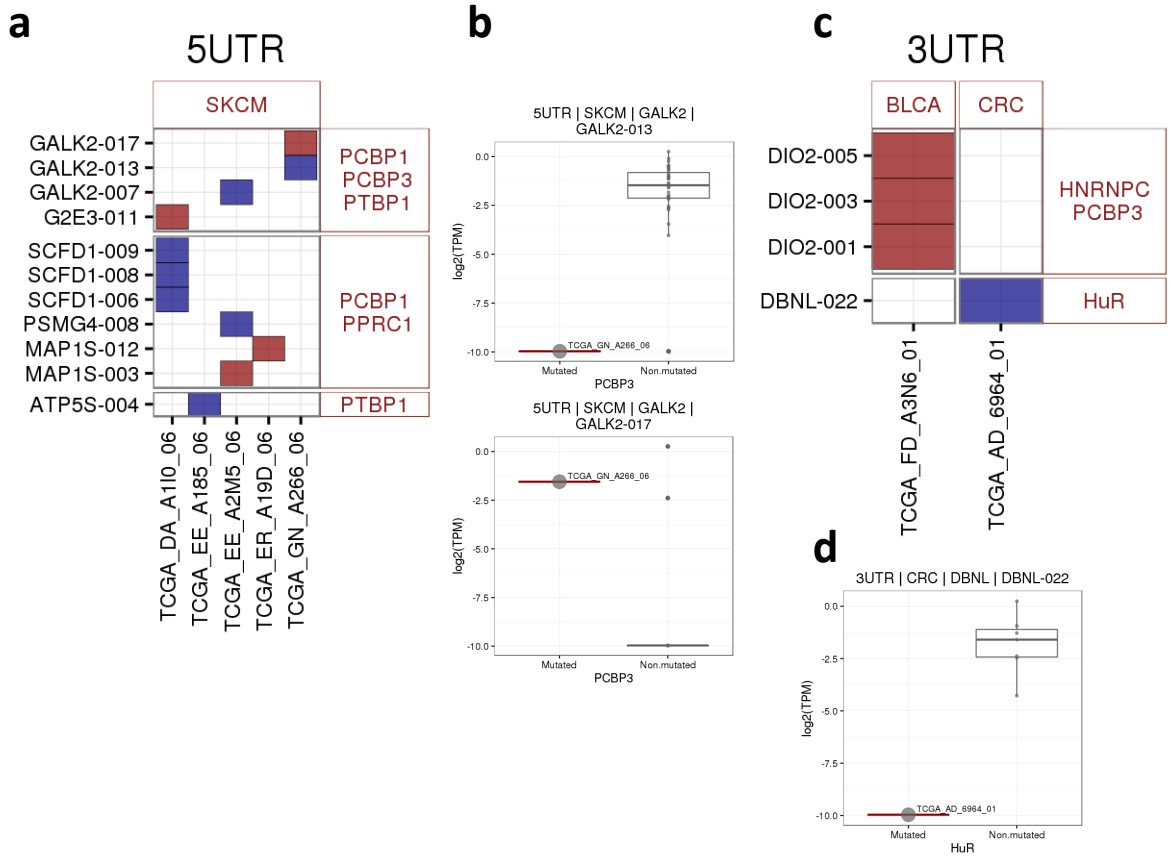


Figure S16. Significant changes in transcript expression associated with mutations in enriched motifs in UTRs. (a) Significant changes in transcript expression associated with mutations in 5UTR motifs. For each patient (x axis) we show the transcripts (y axis) that have a significant increase (red) or decrease (blue) in expression. We separate them according to tumor type (indicated above). If the mutation falls on multiple labeled overlapping motifs, e.g. PCBP1, PCBP3, PTBP1, we giveall the RBP labels. **(b)** We show the example of two transcripts of *GALK2* changing expression in opposite directions in association to a mutation in an SKCM patient on a PCBP3 motif. **(a)** Significant changes in transcript expression associated with mutations in 5UTR motifs. For each patient (x axis) we show the transcripts (y axis) that have a significant increase (red) or decrease (blue) in expression. We separate them according to tumor type (indicated above). **(b)** We show the example of a *DBNL* transcript changing expression in association to a mutation in an CRC patient on a HuR (ELAVL1) motif.

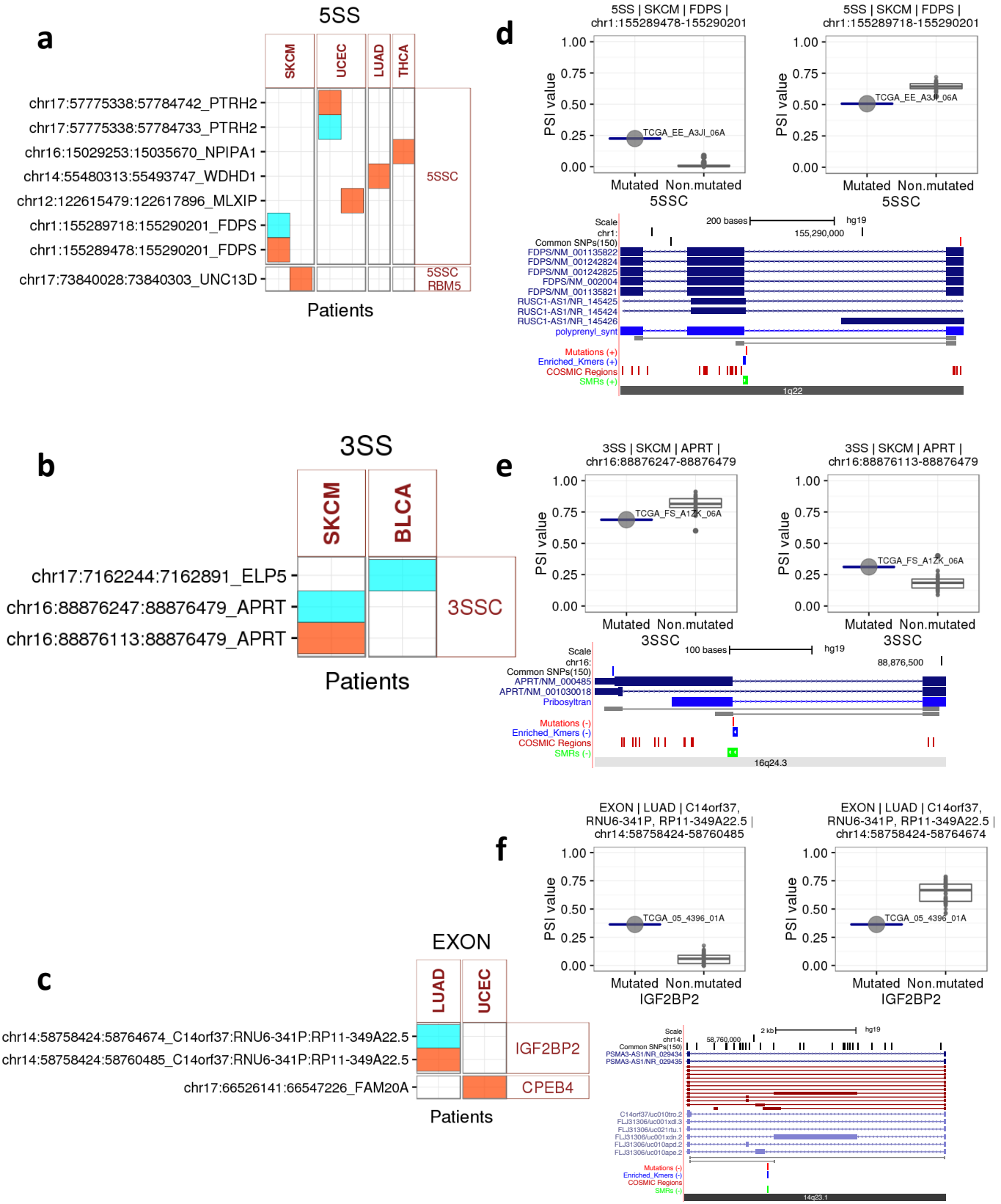


Figure S17. Significant changes in exon-exon junctions. Significant changes in junction inclusion associated to mutations in enriched motifs in 5SS **(a)**, 3SS **(b)** and EXON **(c)** SMRs. We show the junctions (y axis) that have a significant increase (orange) or decrease (cyan) in inclusion. We separate them according to tumor type (indicated above). On the right side we give the motifs in which the mutation occurred. We give the examples of a significant change in *FDPS* **(d).** The boxplots show the PSI values (y axis) of the two changing junctions separated by samples with mutations and without mutations, indicated in the x axis. We also show the UCSC screenshot of the junctions changing in *FDPS*, indicating the position of the SMR (green), the enriched motif (blue), the mutation (red) and the two junctions (gray). We show also the examples of *APRT* (3 splice-site motif) **(e)**, and *C14orf 37* (IGF2BP2 motif on an EXON SMR) **(f)**.

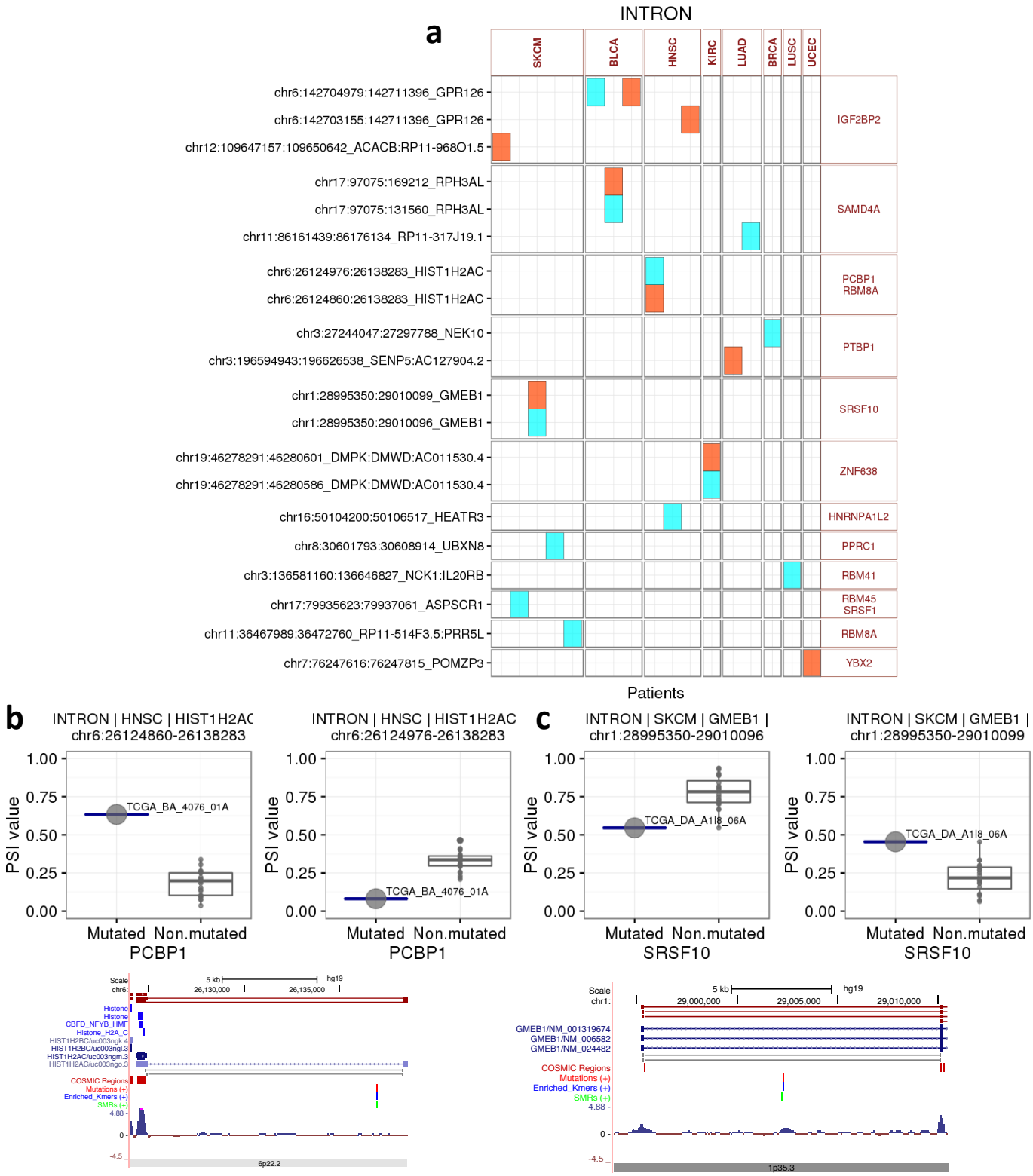


Figure S18. Significant changes in exon-exon junctions. (a) Significant changes in junction inclusion associated to mutations in enriched motifs in INTRON SMRs. We show the junctions (y axis) that have a significant increase (orange) or decrease (cyan) in inclusion. We separate them according to tumor type (indicated above). On the right side we give the motifs in which the mutation occurred. Significant junction changes in *HIST1H2AC* **(b)** associated to mutations in melanoma (SKCM) at intronic PCBP1 motifs, and in *GMEB1* **(c)** associated to mutations in SKCM at intronic SRSF10 motifs. The boxplots show the PSI values (y axis) of the two changing junctions separated by samples with mutations and without mutations, indicated in the x axis. The UCSC screenshots show the changing junctions as well as the SMR (green), motif (blue) and mutations (red).