

Figure S1. 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)** Significantly mutated regions (SMRs). Distribution of the nucleotide-bias (NB) scores (x axis) and p-values (in $-\log_{10}$ scale) (y axis) of the identified SMRs from the PAN505 dataset separated by region type.

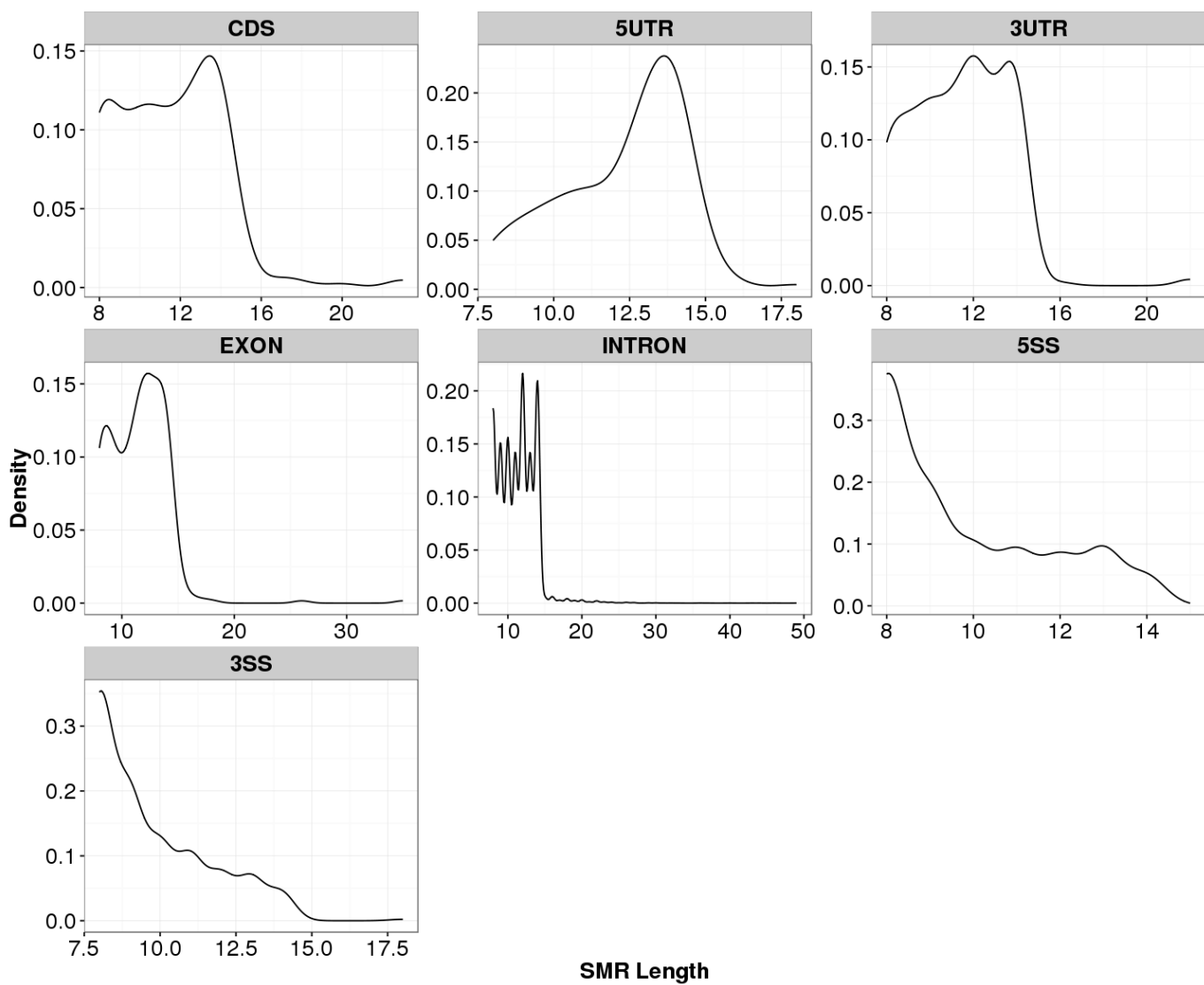


Figure S2. Length distribution of SMRs. Length distribution of the identified significantly mutated regions (SMRs) for the PAN505 dataset separately for each region type.

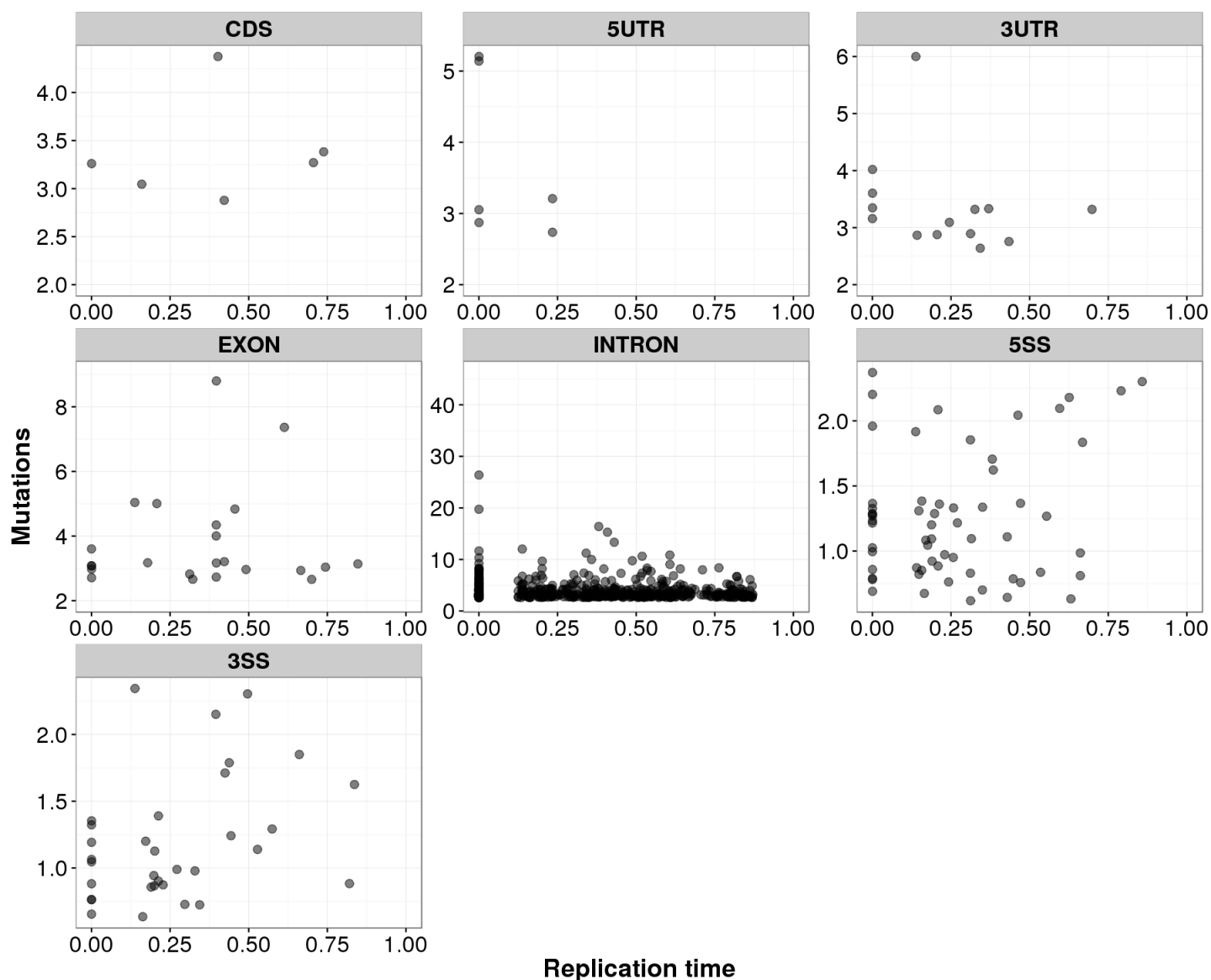


Figure S3. Comparison of mutation count with replication timing. Relation between replication timing (x-axis) and number of mutations (y-axis) in the SMRs from the PAN505 cohort falling in regions with replication timing data from (Lochovsky et al. 2015).

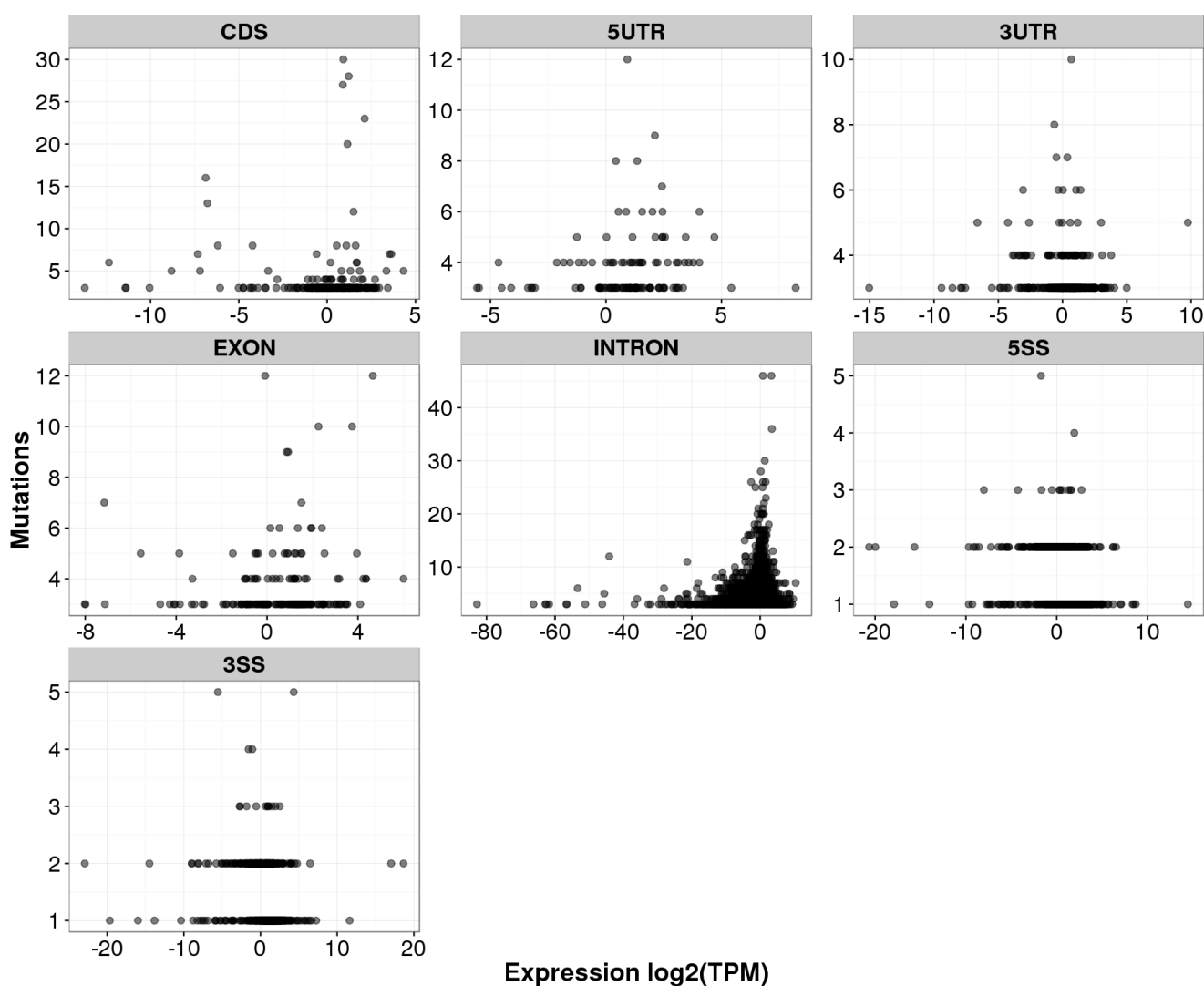


Figure S4. Comparison between expression and mutation count. For every SMR in the PAN505 cohort, we compared the average expression of the transcripts, in $\log_{10}(\text{TPM})$ units (x axis), with the mutation count (y axis), within the same patients harboring the mutations in the SMR. For each transcript 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.

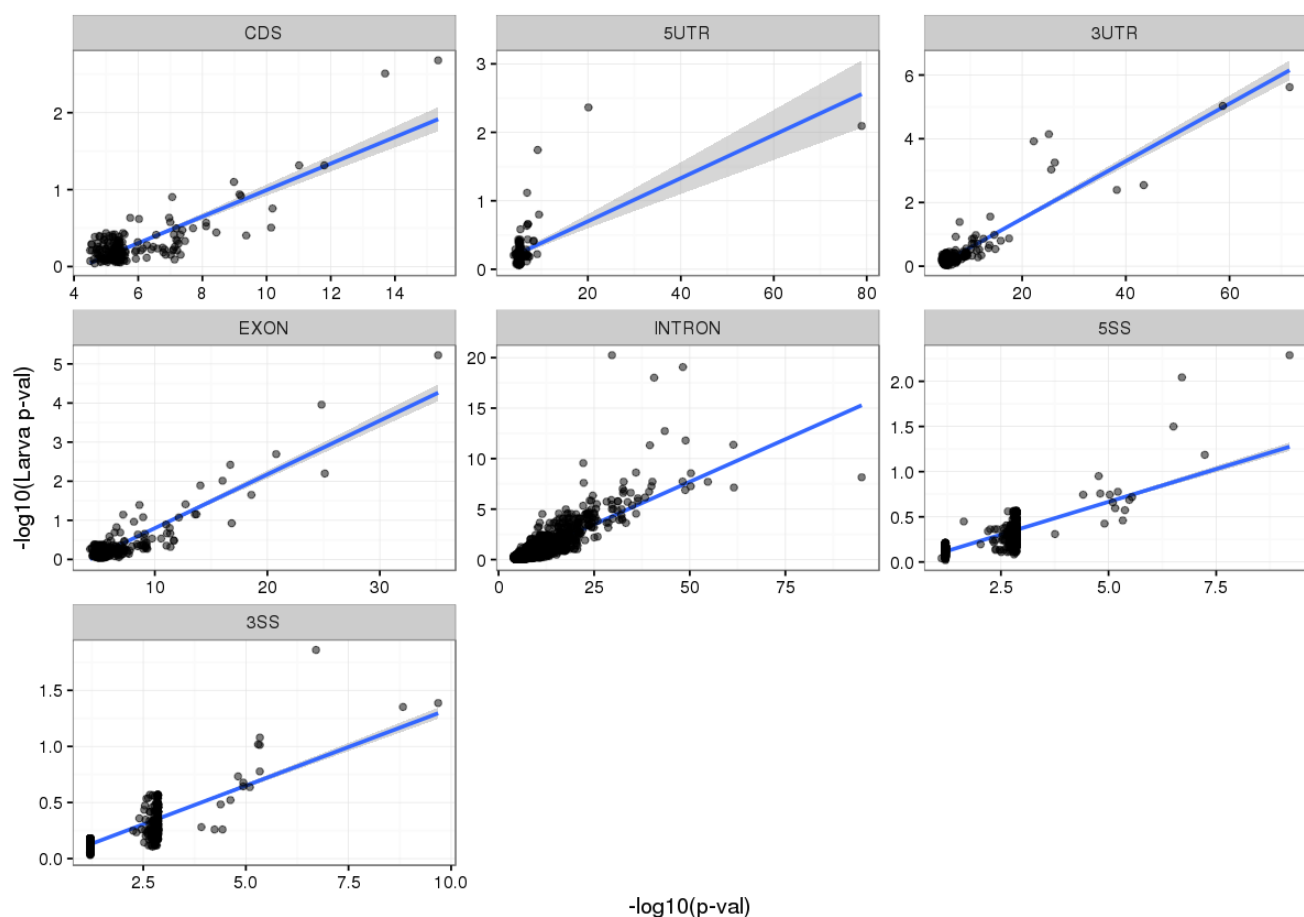


Figure S5. Comparison with LARVA. For each of our predicted SMRs, we compare in $-\log_{10}$ scale the p-value we provided with the p-value provided by LARVA (Lochovsky et al. 2015) using a model that accounts for over-dispersion of the mutation rate and replication timing (p.bbd.cor, see Methods). The Pearson's correlation values 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).

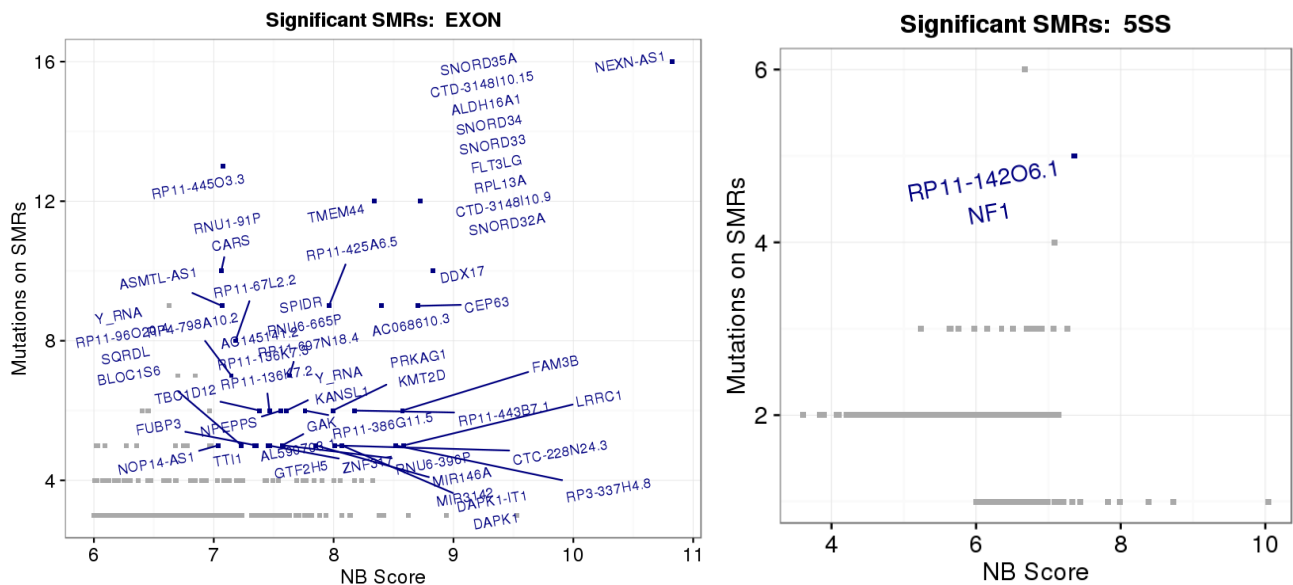


Figure S6. EXON and 5SS SMRs. For EXON and 5SS SMRs we give the nucleotide bias (NB) score (x axis) and the number of mutations (y axis). We show the gene name for the SMRs with NB-score > 6 and with 5 or more mutations.

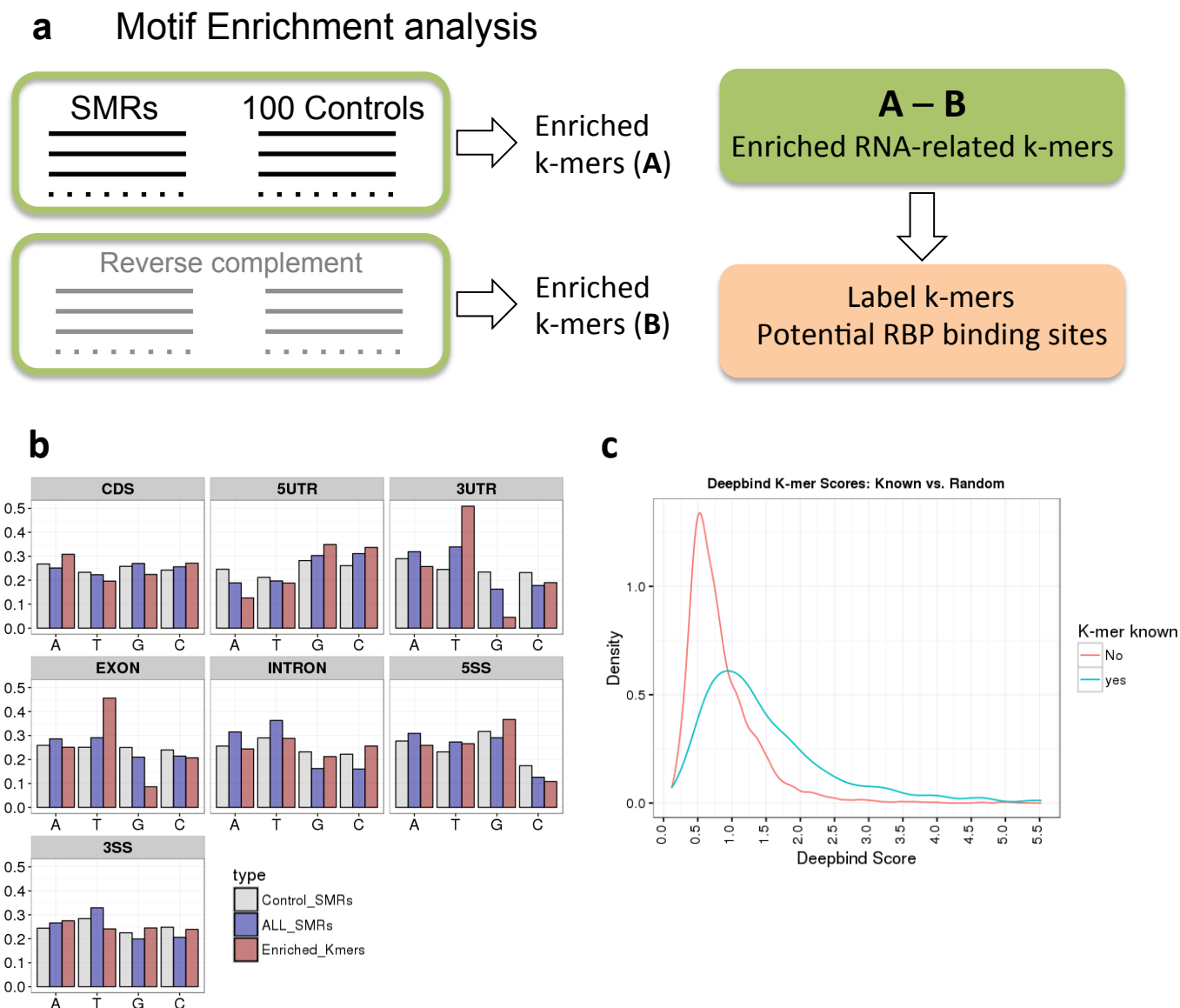
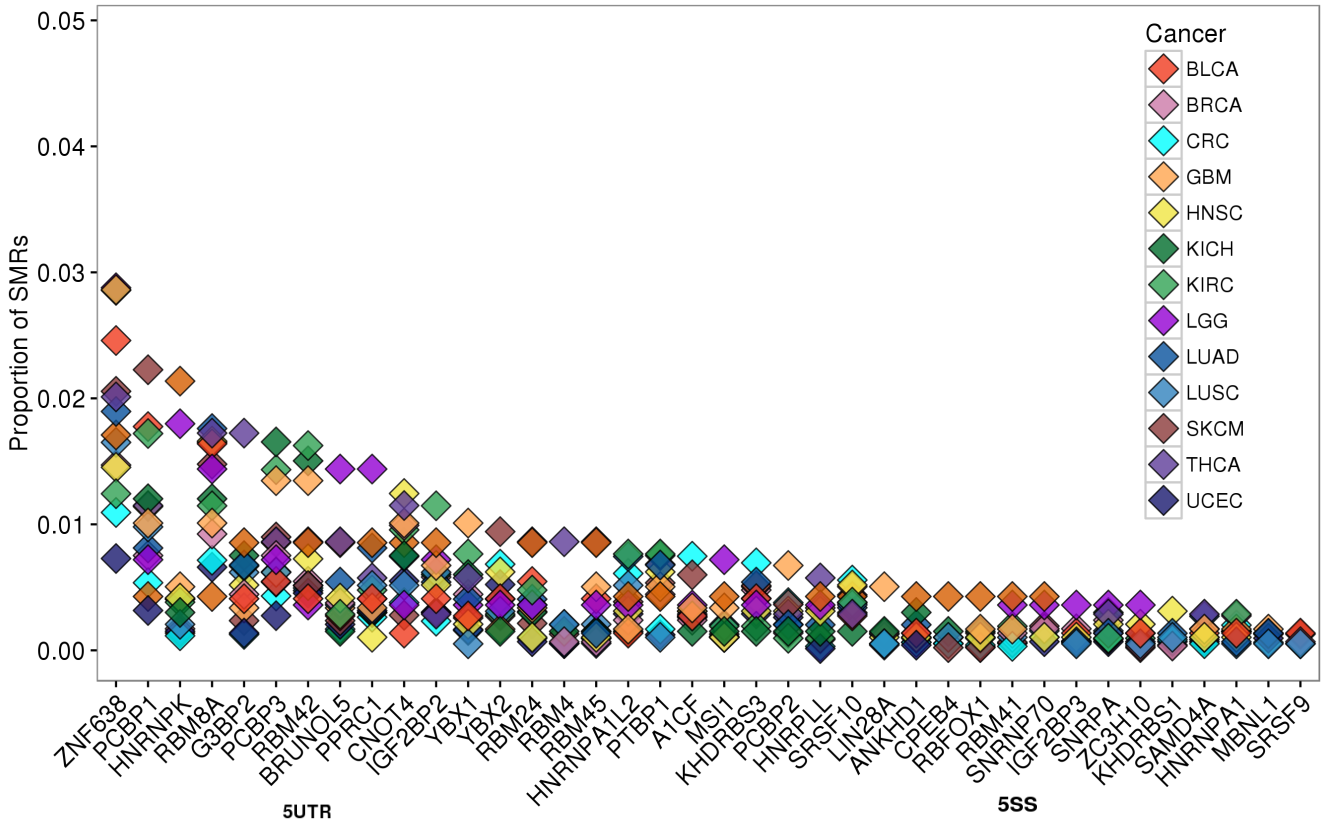
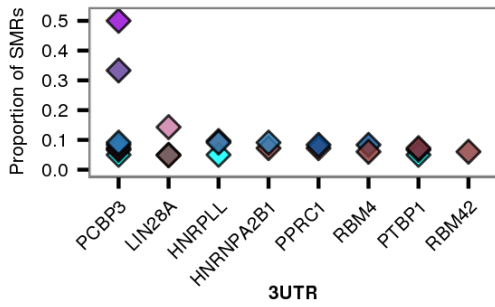


Figure S7. Motif enrichment in SMRs. (a) Association with RNA processing motifs is assessed by identifying enriched motifs that are specific of the gene strands. (b) For each region type we give the proportion of A, C, G and T (y axis) for found enriched k-mers in SMRs (k=6), all SMRs, and control regions built by sampling sequences of same length and GC content as the SMRs from the same region type (Methods). (c) 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.cnice.es/>) (Giudice et al. 2016)

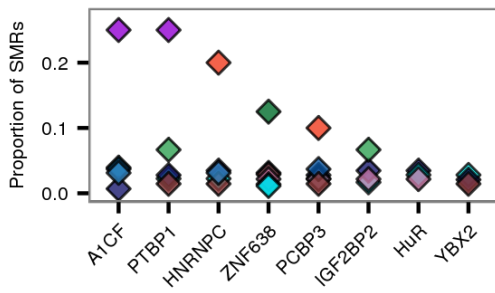
INTRON



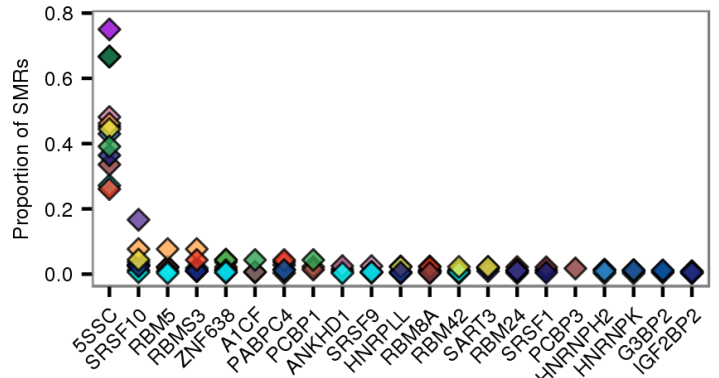
5UTR



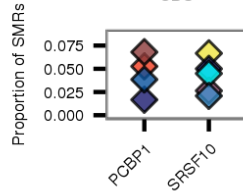
3UTR



5SS



CDS



3SS

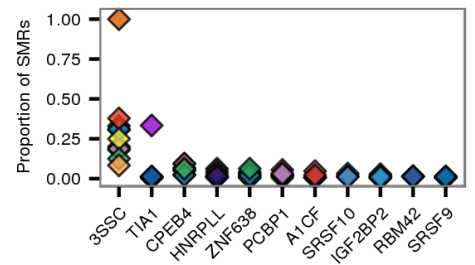


Figure S8. Proportion of SMRs with enriched motifs. For the region types INTRON, 5SS, 3SS, tUTR and 3UTR, and for each cancer type (coded by color), we give the proportion (y axis) of SMRs that contain at least one mutated RBP motif (x axis).

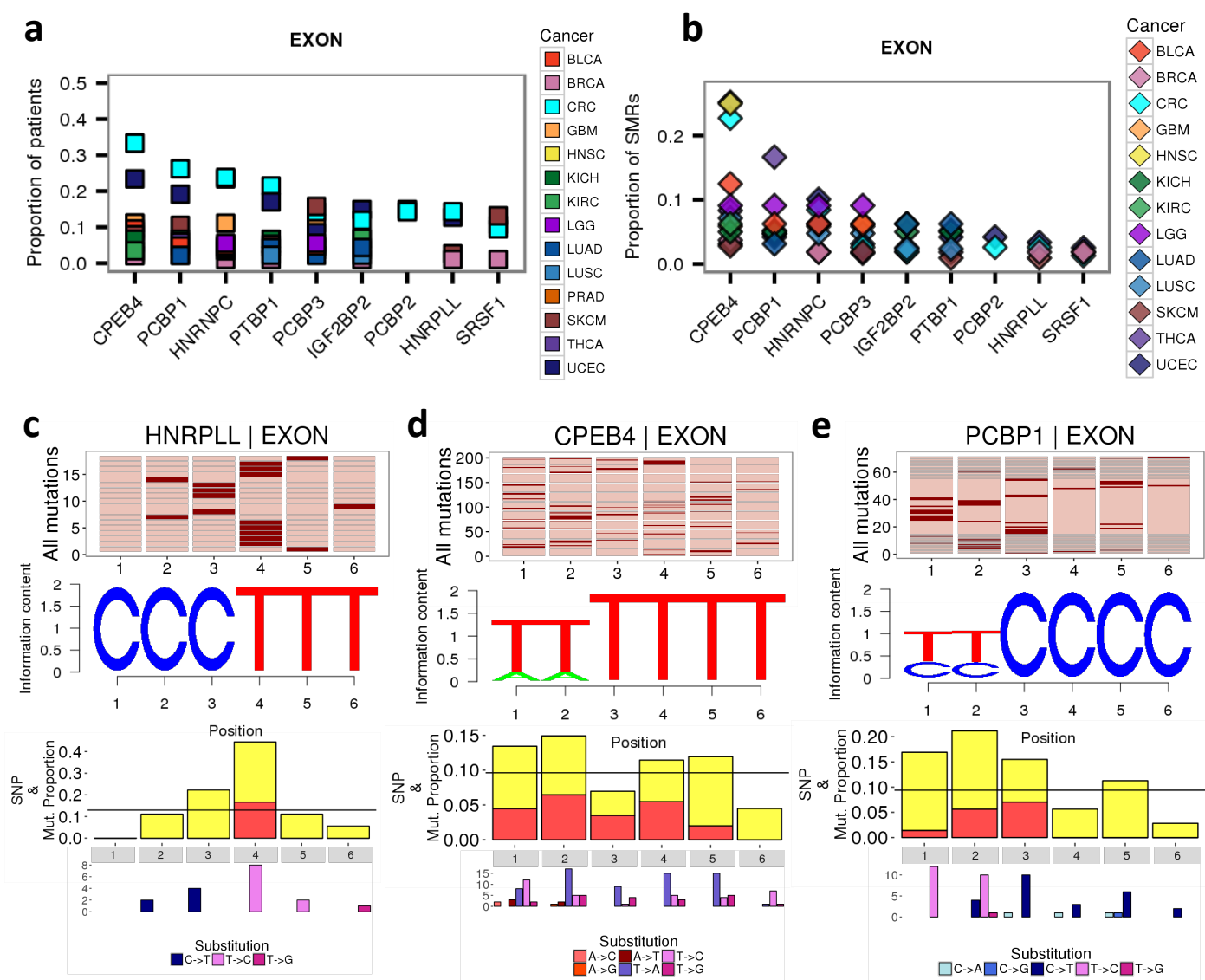


Figure S9. Mutations on RBP motifs on EXON SMRs. For the EXON (exons in long-non coding RNA genes) SMRs and for each cancer type (coded by color), we give in the y axis the proportion patients (a) and of SMRs (b) that contain at least one mutated RBP motif (x axis). We show the mutation patterns on HNRPLL (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.

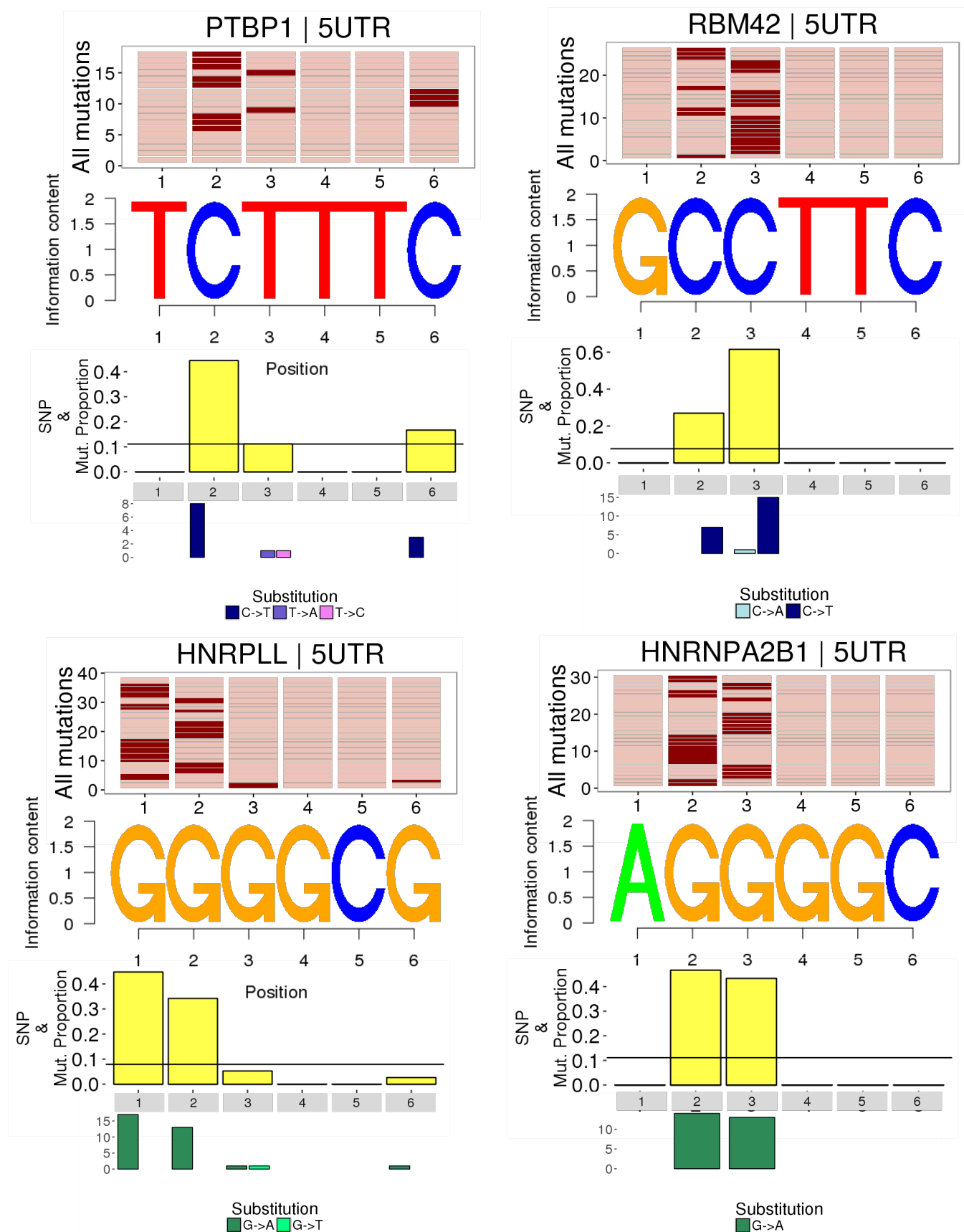


Figure S10. Mutation patterns on motifs in 5UTRs. We show the mutation patterns on PTBP1, RBM42, HNRPLL, HNRNPA2B1 motifs in 5UTR SMRs. In the upper panels we indicate in red the positions covered by the motif and in dark red the position of the mutations. The barplots below show the proportion of somatic 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.

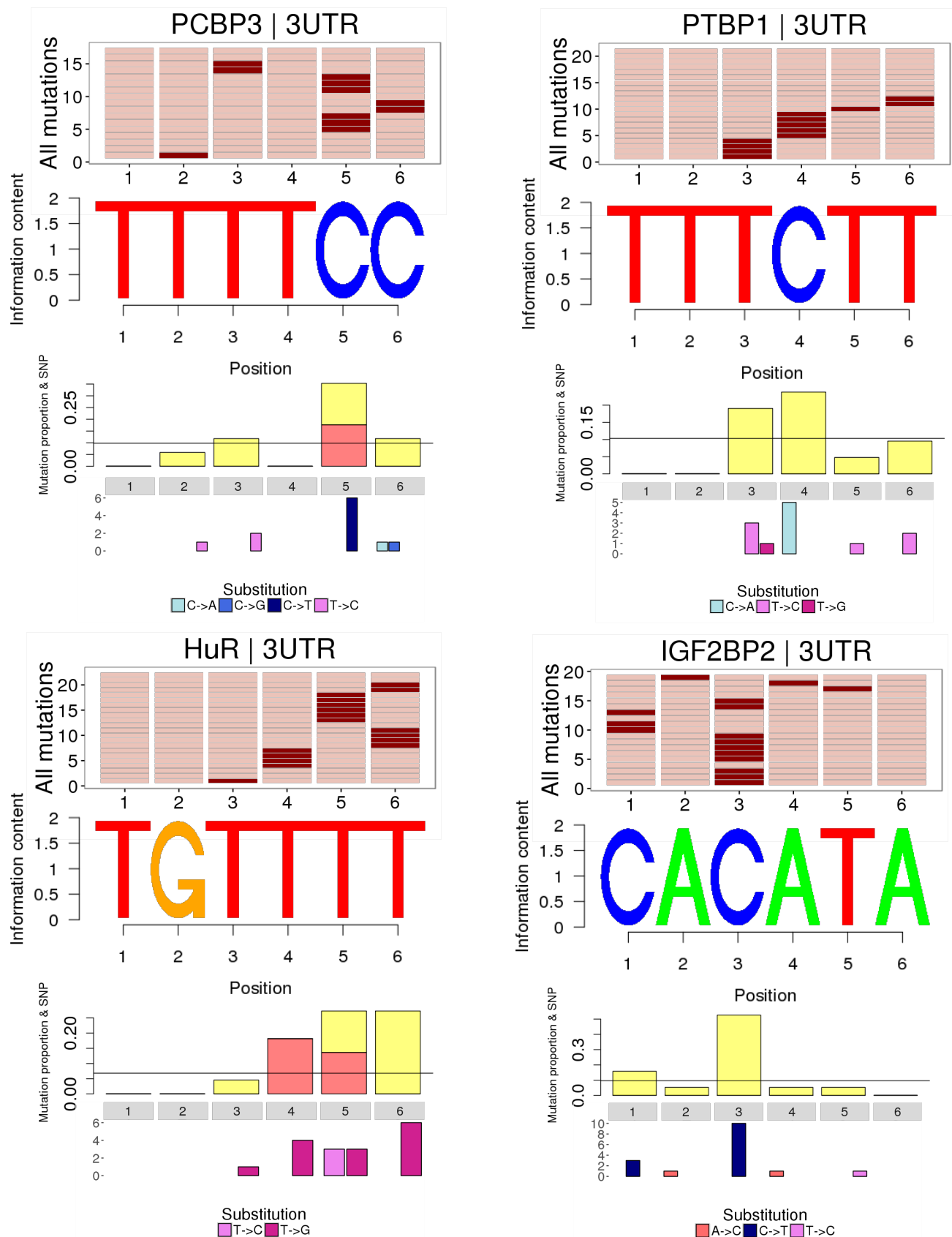


Figure S11. Mutation patterns on motifs in 3UTRs. We show the mutation patterns on PCBP3, PTBP1, HuR (ELAVL1), IGF2BP2 in 3UTR SMRs. In the upper panels we indicate in red the positions covered by the motif and in dark red the position of the mutations. The barplots below show the proportion of somatic 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.

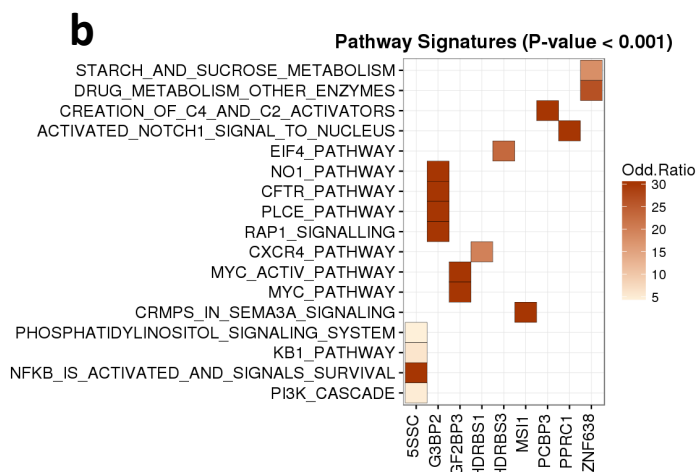
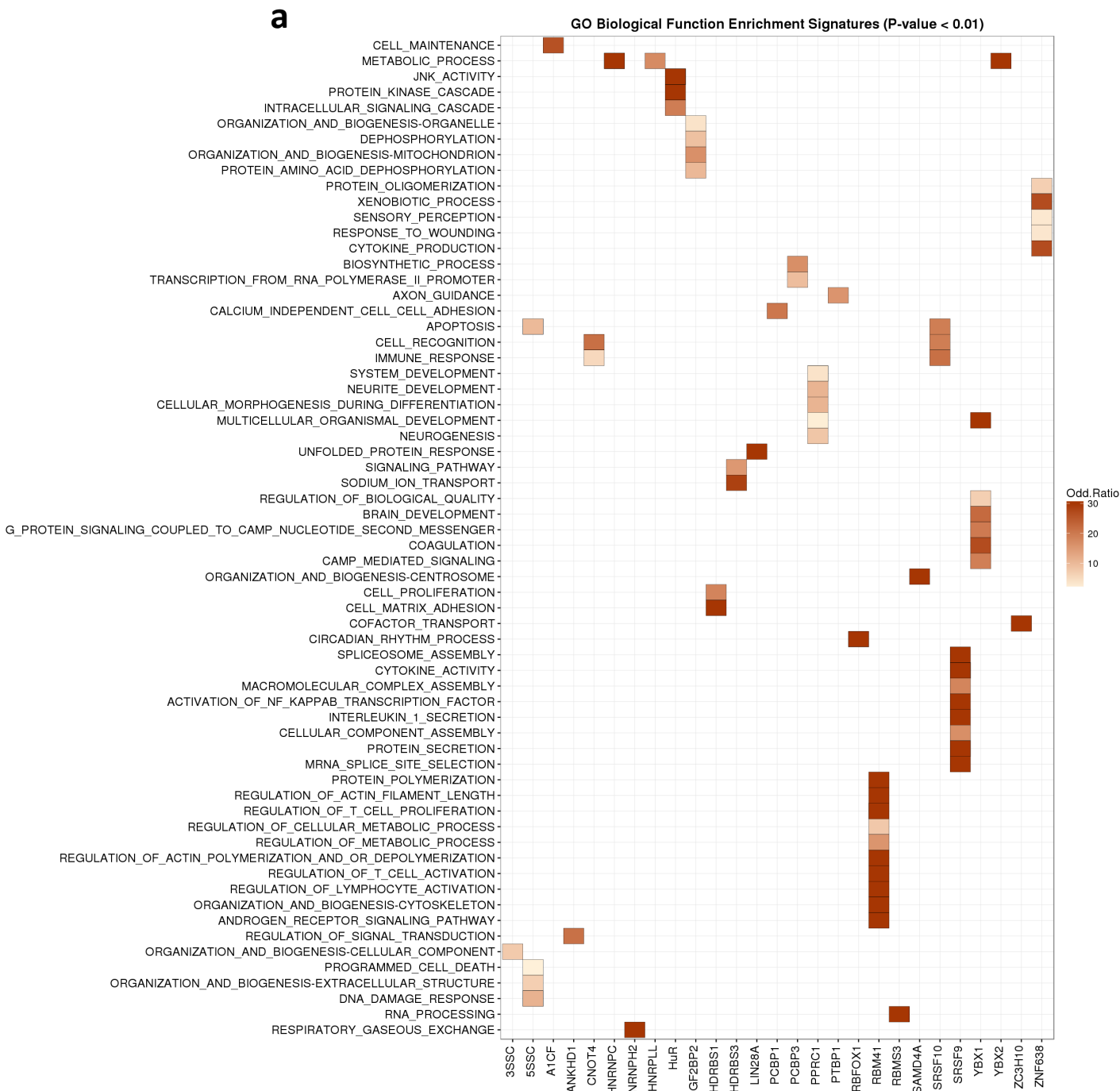


Figure S12. Gene set enrichment analysis. (a) **GO Biological function enrichment.** Enriched categories are given in the y axis related to the genes with SMRs with enriched motifs (x axis). The color scale is related to the odds-ratio of the Fisher's exact test. (b) **Pathway enrichment analysis.** 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 S13. 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.

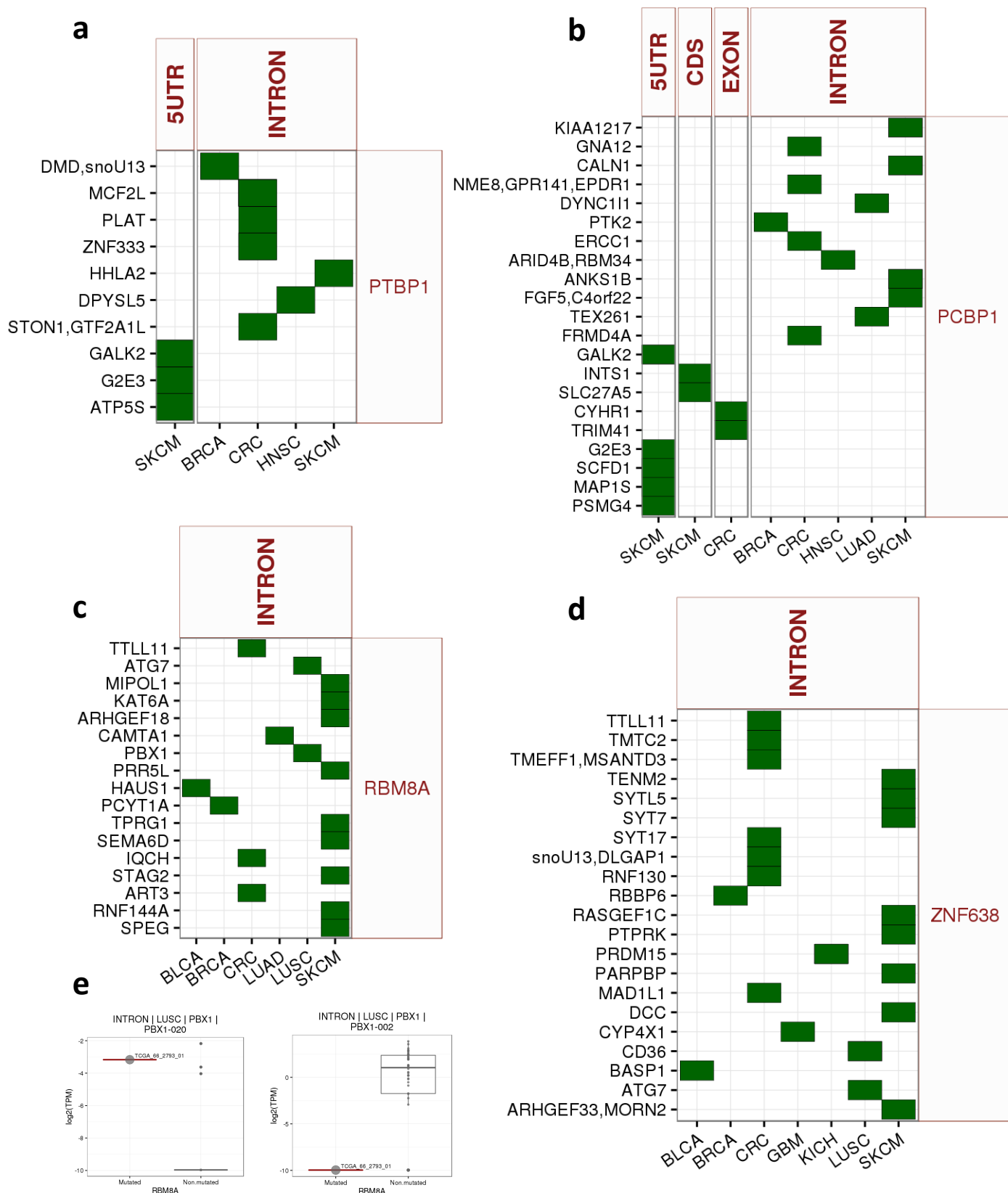


Figure S14. Significant changes in transcript expression associated to mutations in enriched motifs at intronic SMRs. We show the cases in the motifs PTBP1 (a), PCBP1 (b), RBM8A (c), and ZNF638 (d). We show the genes (y axis) that have a significant change, separated by tumor type (x axis). (e) Expression 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.

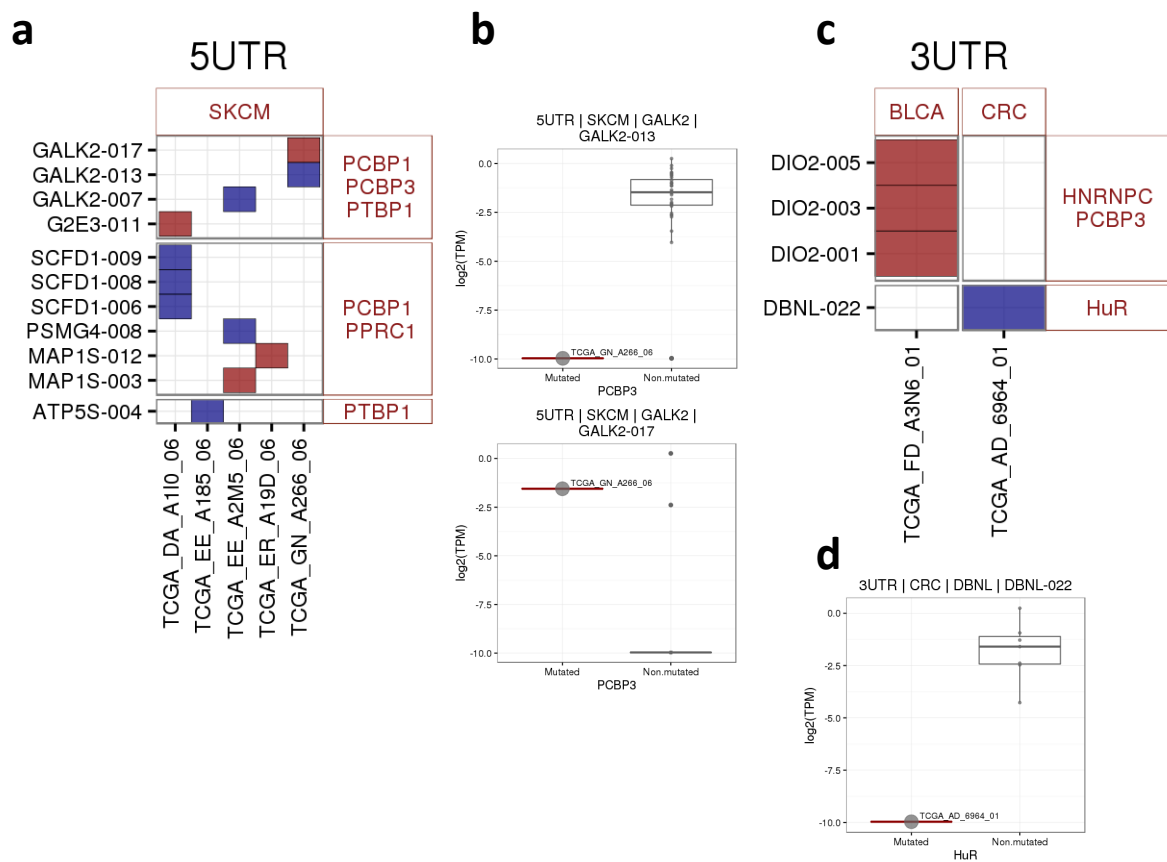


Figure S15. Significant changes in transcript expression associated with mutations in enriched motifs in UTRs. **(a)** Significant changes in transcript expression associated with mutations in 5UTR motifs. For each patient (x axis) we show the transcripts (y axis) that have a significant increase (red) 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 give all 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.

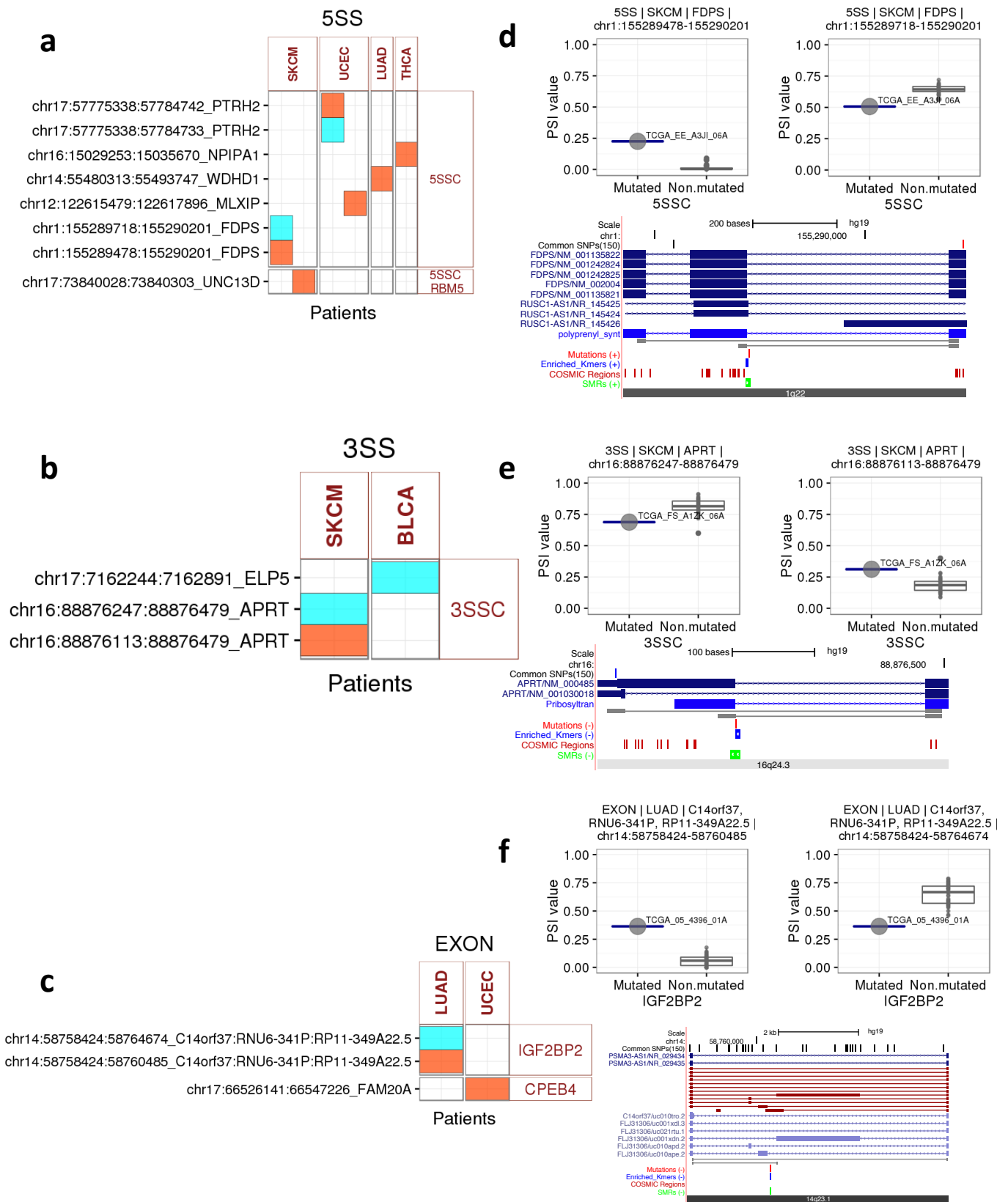


Figure S16. Significant changes in exon-exon junctions. Significant changes in junction inclusion associated to mutations in enriched motifs in 5SS (**a**), 3SS (**b**) and EXON (**c**) SMRs. We show the junctions (y axis) that have a significant increase (orange) or decrease (cyan) in inclusion. We 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**).

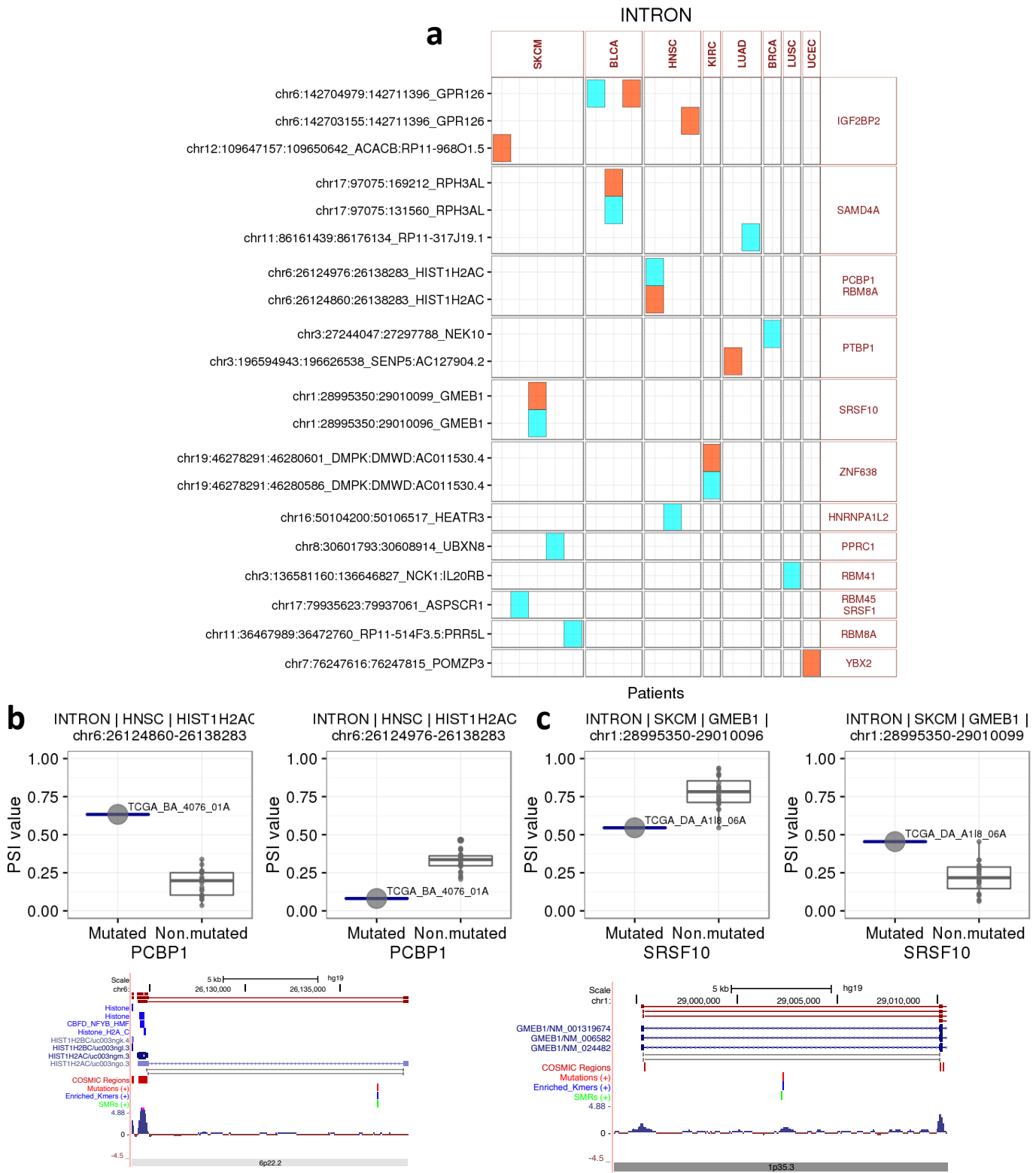


Figure S17. Significant changes in exon-exon junctions. (a) Significant changes in junction inclusion associated to mutations in enriched motifs in INTRON SMRs. We show the junctions (y axis) that have a significant increase (orange) or decrease (cyan) in inclusion. We separate them according to tumor 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).