Cell Reports, Volume *23*

Supplemental Information

Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types

Michael Seiler, Shouyong Peng, Anant A. Agrawal, James Palacino, Teng Teng, Ping Zhu, Peter G. Smith, The Cancer Genome Atlas Research Network, Silvia Buonamici, and Lihua Yu

Supplemental Figures

Figure S1

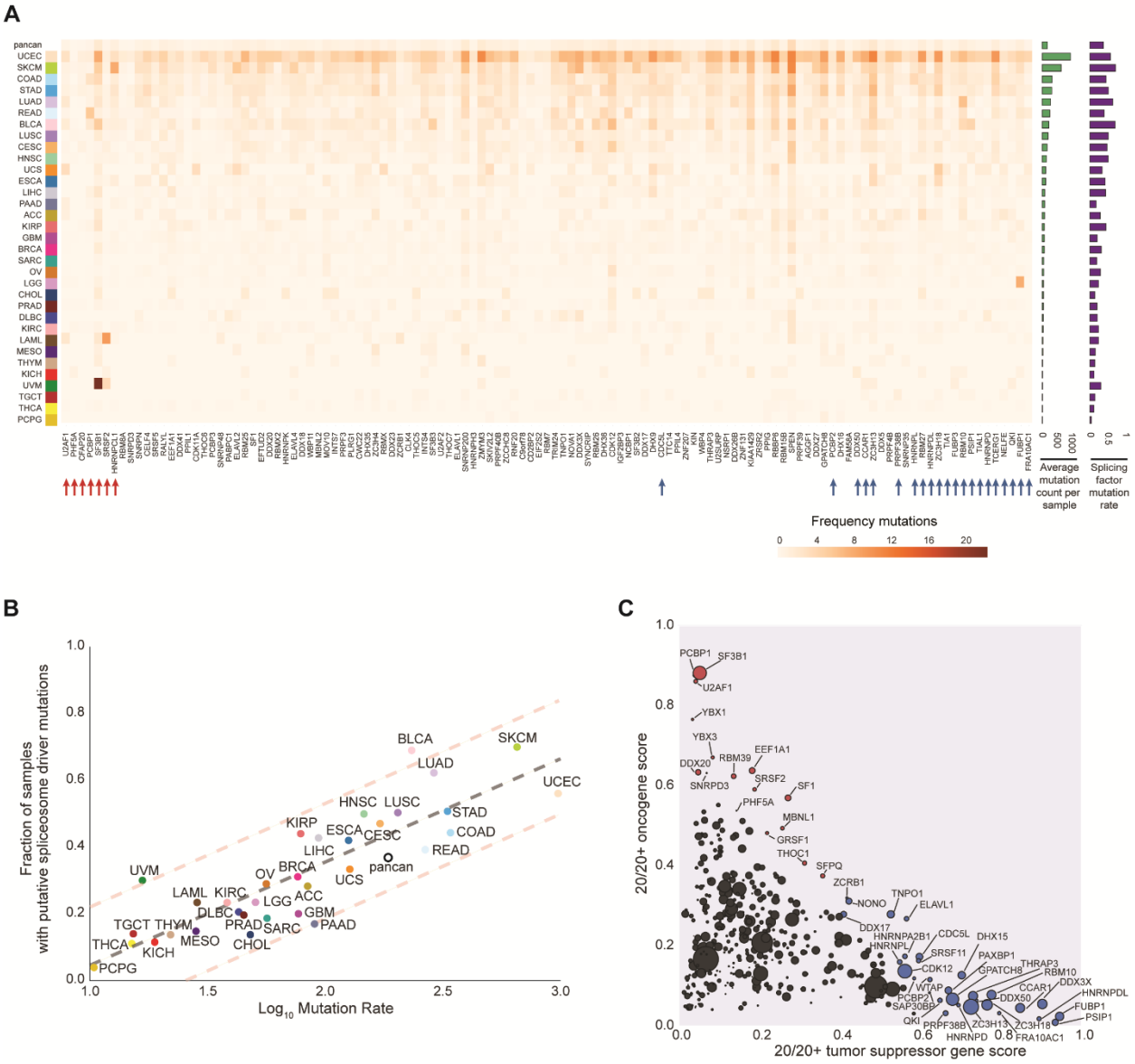

Figure S1. **Mutations of 119 splicing factor genes pan-TCGA. Related to Figure 1.**

A. Frequency of non-silent somatic mutation of 119 putative driver splicing factor genes are plotted across all tumor cohorts. Genes are sorted by %hotspot mutation high to low and %LoF mutation low to high from left to right. Cohorts are ordered by average mutation counts per sample per cohort from top to bottom (right green bar). The fractions of samples with non-silent mutations in any of the 119 putative driver genes in each cohort are shown as purple bars on the right. **B.** All TCGA cohorts are plotted based on the fraction of samples in that cohort with any non-silent mutation in a putative splicing factor

driver gene (y-axis) and the Log10 overall mutational burden per sample (x-axis). The linear regression line is plotted in gray, while the 95% confidence intervals are presented in red. **C.** Genes with significant (q<0.1) driver score using 20/20+ (Tokheim et al., 2016) are plotted by tumor suppressor gene (TSG) score (x-axis) vs. oncogene (OG) score (y-axis) across TCGA (pancan). Genes with greater OG score are colored red, while genes with greater TSG score are colored blue.

Figure S2

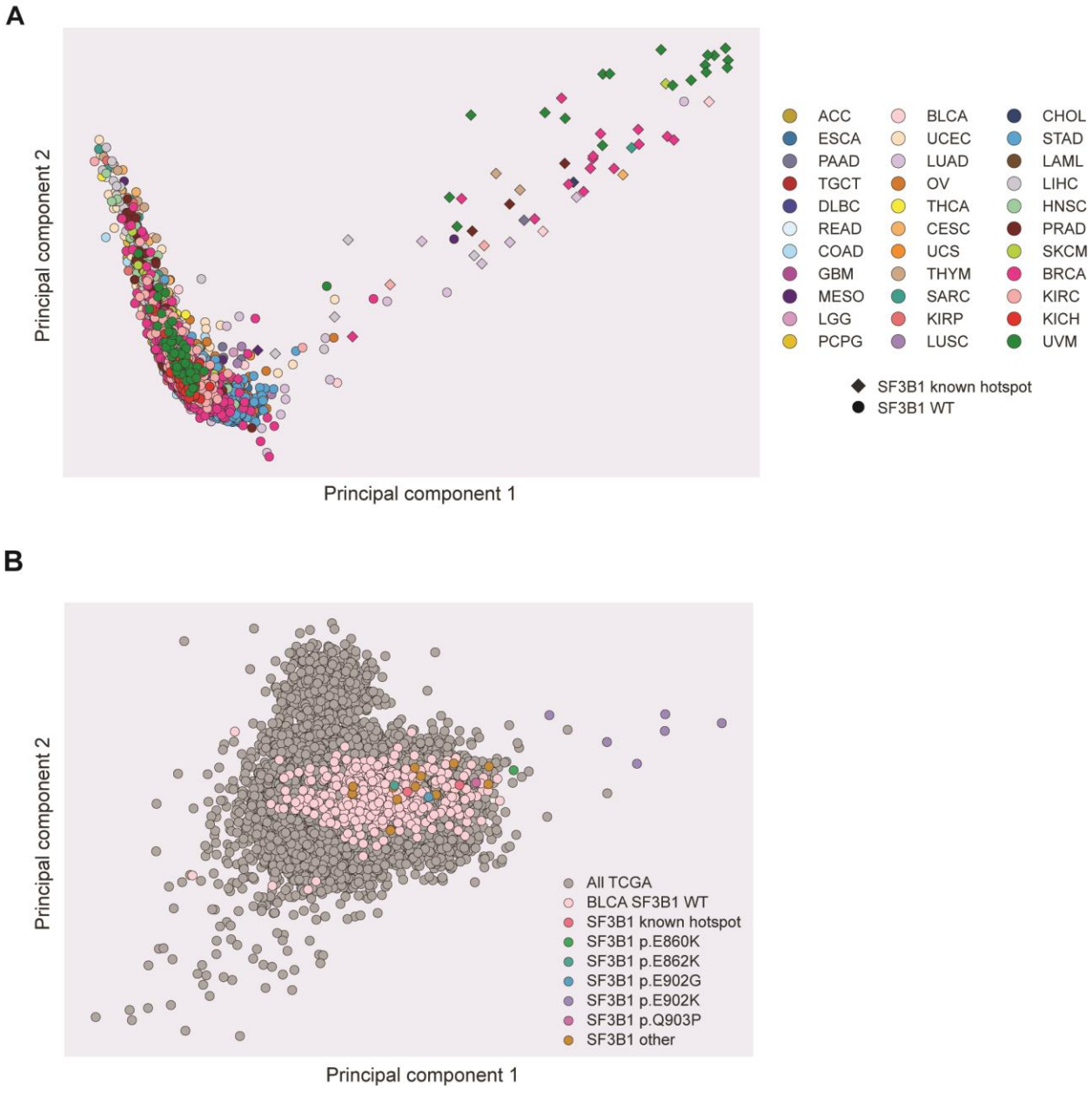

Figure S2. Splicing signatures from *SF3B1* HDs 4-8 hotspot mutations and p.E902K

stratify solid tumors across TCGA. Related to Figure 2.

A. PCA using the expression of aberrant 3' splice sites specific to known *SF3B1* mutations stratifies tumor samples from all solid tumor cohorts in TCGA. Samples are colored by cohort and diamond markers are used to represent samples with known *SF3B1* hotspot mutations, while circle markers represent all other samples. **B**. PCA using the expression of junctions upregulated by the presence of *SF3B1* p.E902K stratifies samples from all solid tumor cohorts. BLCA samples with *SF3B1* mutations are colored separately to show the *SF3B1* p.E902K mutation confers this splicing phenotype.

Figure S3

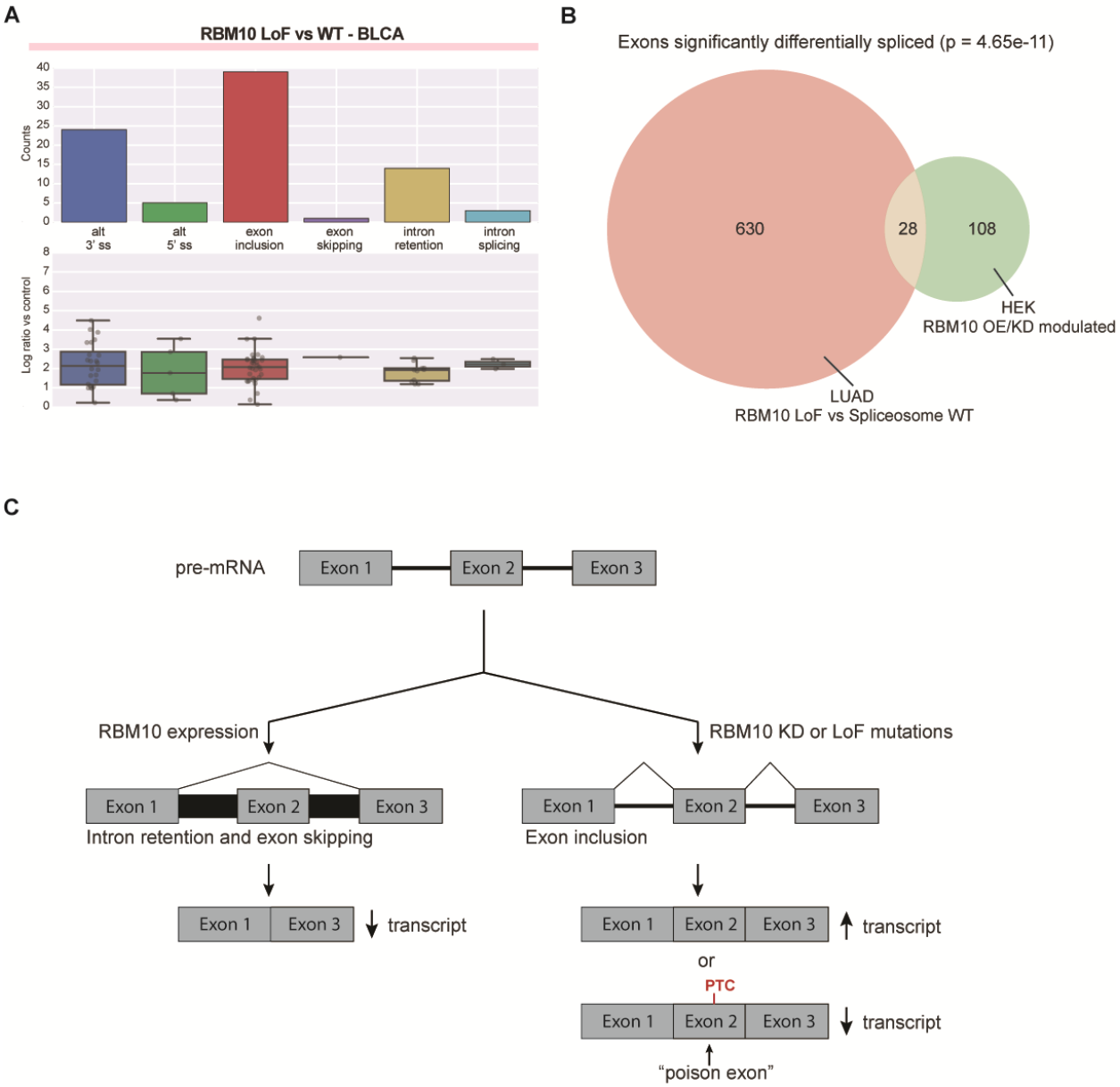

Figure S3. *RBM10* LoF mutation associated differential splicing in BLCA and overlap of *RBM10* LoF modulated exons with published data. Related to Figure 4.

A. Differential splicing events associated with *RBM10* LoF mutations (n=11) compared to splicing factor gene WT (n=40) are identified using RNAseq in BLCA. Events with corrected q-value < 0.05 were considered significant. Below each splicing event count, the PSI log2 fold change of each individual event is detailed in a boxplot. **B.** Exons that are significantly upregulated upon *RBM10* loss in LUAD overlap with exons that are significantly downregulated upon *RBM10* over-expression (OE) and upregulated upon knockdown (KD) in HEK293 cells (Wang et al., 2013). **C**. Overview of spliced gene expression changes because of *RBM10* LoF mutations.

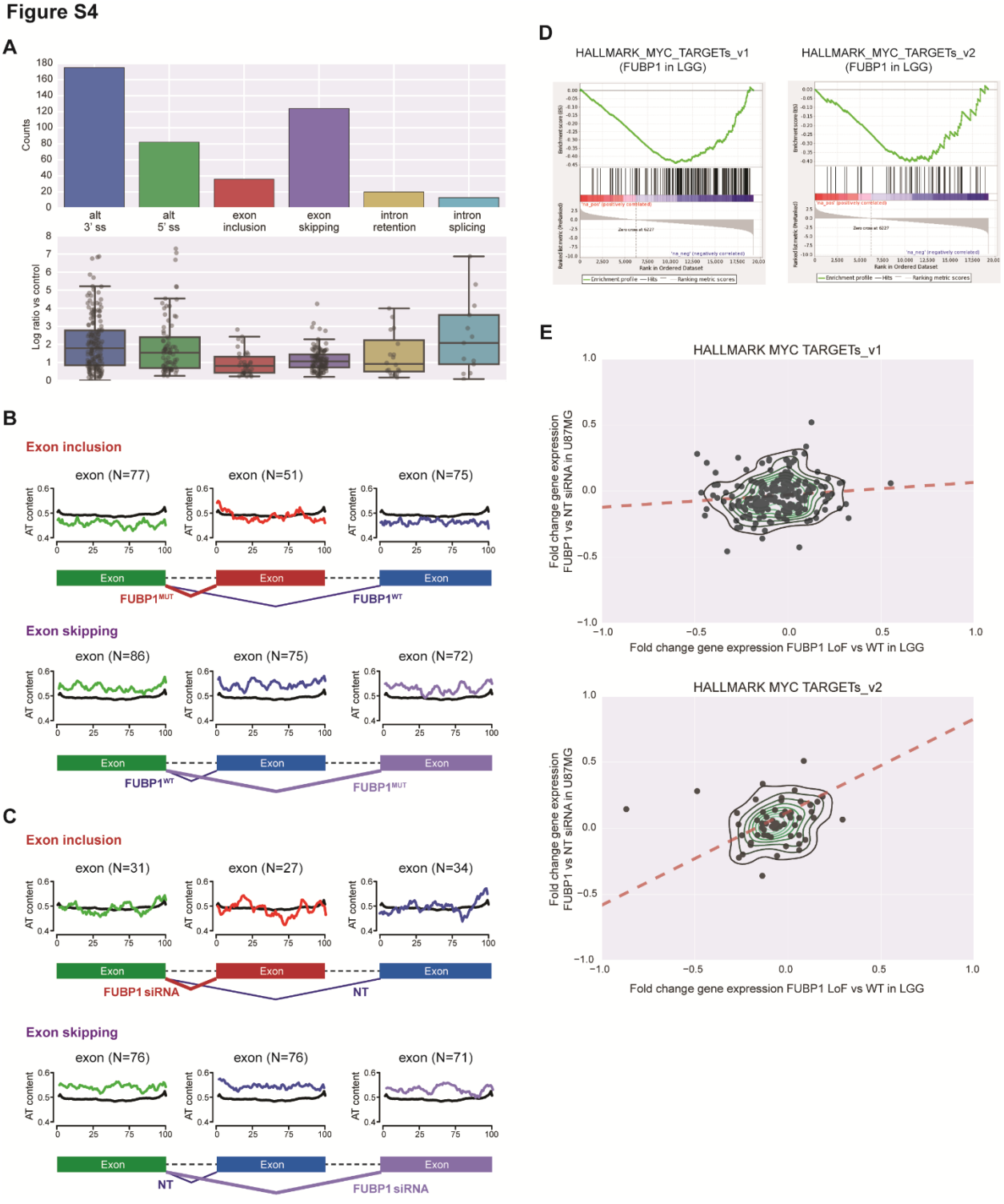

Figure S4. *FUBP1* LoF-induced cassette exon events are associated with increased AT content, and *FUBP1* LoF mutations are associated with down-regulation of *myc* target genes. Related to Figure 5.

A. Differential splicing events associated with *FUBP1* siRNA knock down in U87MG

cells (n=3) vs. non-targeting (NT) siRNA (n=3). Events with corrected q-value < 0.05 were considered significant. Below each splicing event count the PSI log2 fold change of each individual event is detailed in a boxplot. **B.** AT content analysis identified AT content bias associated with alternatively spliced exons in LGG samples with *FUBP1* LoF mutations. Each exon is normalized to length of 100. The reference AT content for 10,000 randomly selected exons are plotted in black. Exons preferentially included (MUT exon of exon inclusion event, colored red) in *FUBP1* LoF mutant samples have higher AT content towards the 5' end of exons (left) relative to exons preferentially spliced in WT (WT exon of exon inclusion event, colored in blue). All three exons associated with exon skipping events in *FUBP1* LoF mutant samples have higher AT contents than background ($p < 0.00015$ in all three comparisons, student's t-test). AT content profiling was restricted to exons of length >= 60 and <= 240. **C.** AT content analysis of alternatively spliced exons identified in U87MG cells *FUBP1* siRNA knock down. All three exons associated with exon skipping events in *FUBP1* siRNA knock-down samples have higher AT contents than background ($p < 0.00019$ in all three comparisons, student's t-test). AT content profiling was restricted to exons of length >= 60 and <= 240. **D.** Gene set enrichment analysis (GSEA) using MSigDB hallmark gene sets identified significant downregulation of *MYC* target genes. **E.** Fold changes of *MYC* target genes from MSigDB hallmark gene sets are correlated between LGG *FUBP1* LoF vs *FUBP1* WT and U87MG siRNA knock down vs NT siRNA comparisons ($p = 0.12$ and $p = 5.9e-4$ for *MYC* targets v1 and v2, respectively).

Figure S5

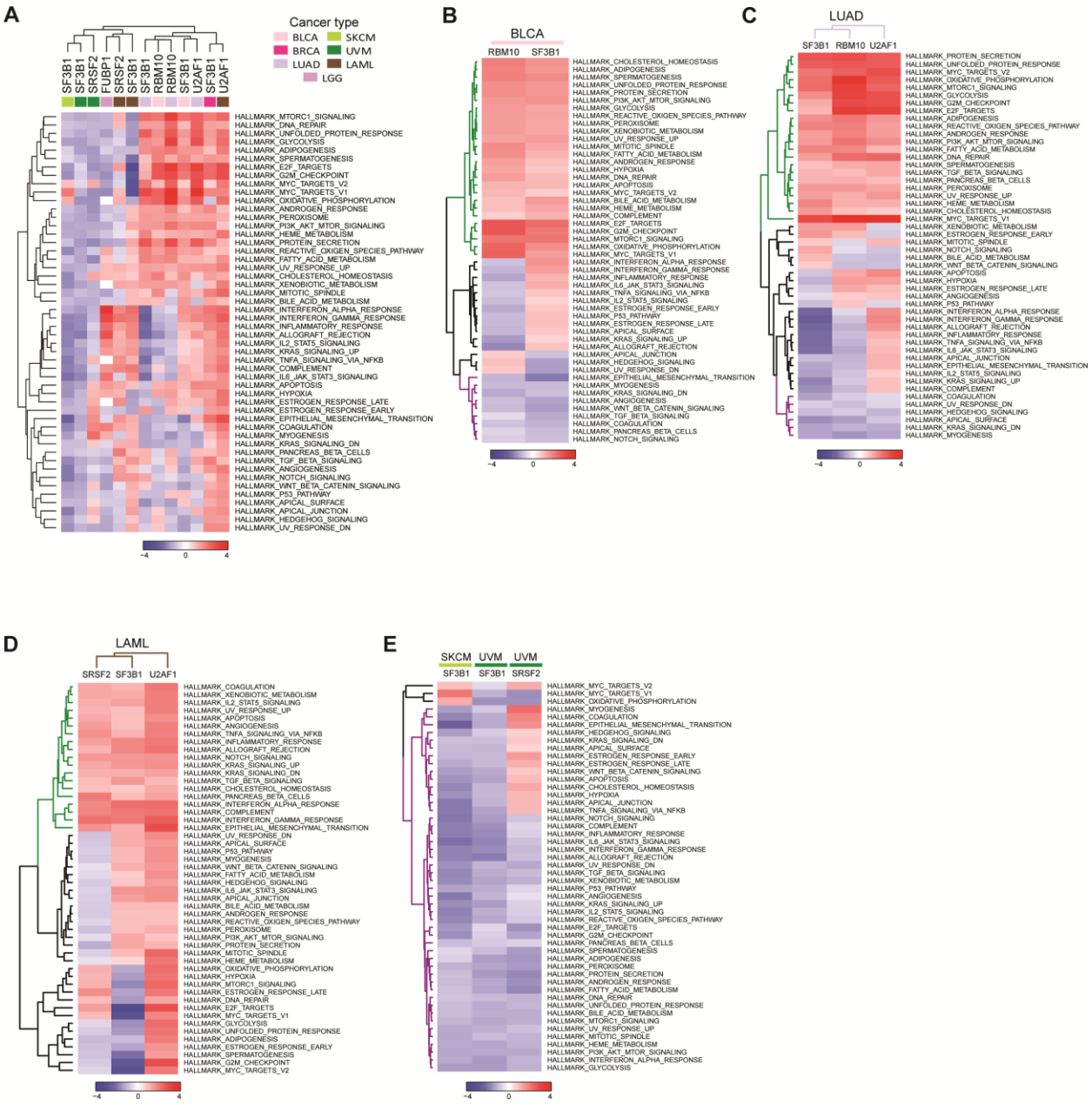

Figure S5. Gene set enrichment analysis using hallmark gene sets identified common pathways within each cohort. Related to Figure 6.

A. Clustering analysis of normalized enrichment scores (NESs) of hallmark gene sets suggest that cancer pathway modulation associated with splicing factor gene mutations across different diseases is lineage-specific. **B.** Clustering analysis of NESs for splicing factor gene mutations found in BLCA. Commonly upregulated pathways are marked in green, whereas commonly downregulated pathways are marked in purple. **C.** Clustering analysis of NESs for splicing factor gene mutations found in LUAD. Commonly

upregulated pathways are marked in green, whereas commonly downregulated pathways are marked in purple. **D**. Clustering analysis of NESs for splicing factor gene mutations found in LAML. Commonly upregulated pathways are marked in green. **E**. Clustering analysis of NESs for splicing factor gene mutations found in UVM and SKCM. Commonly downregulated pathways are marked in purple.