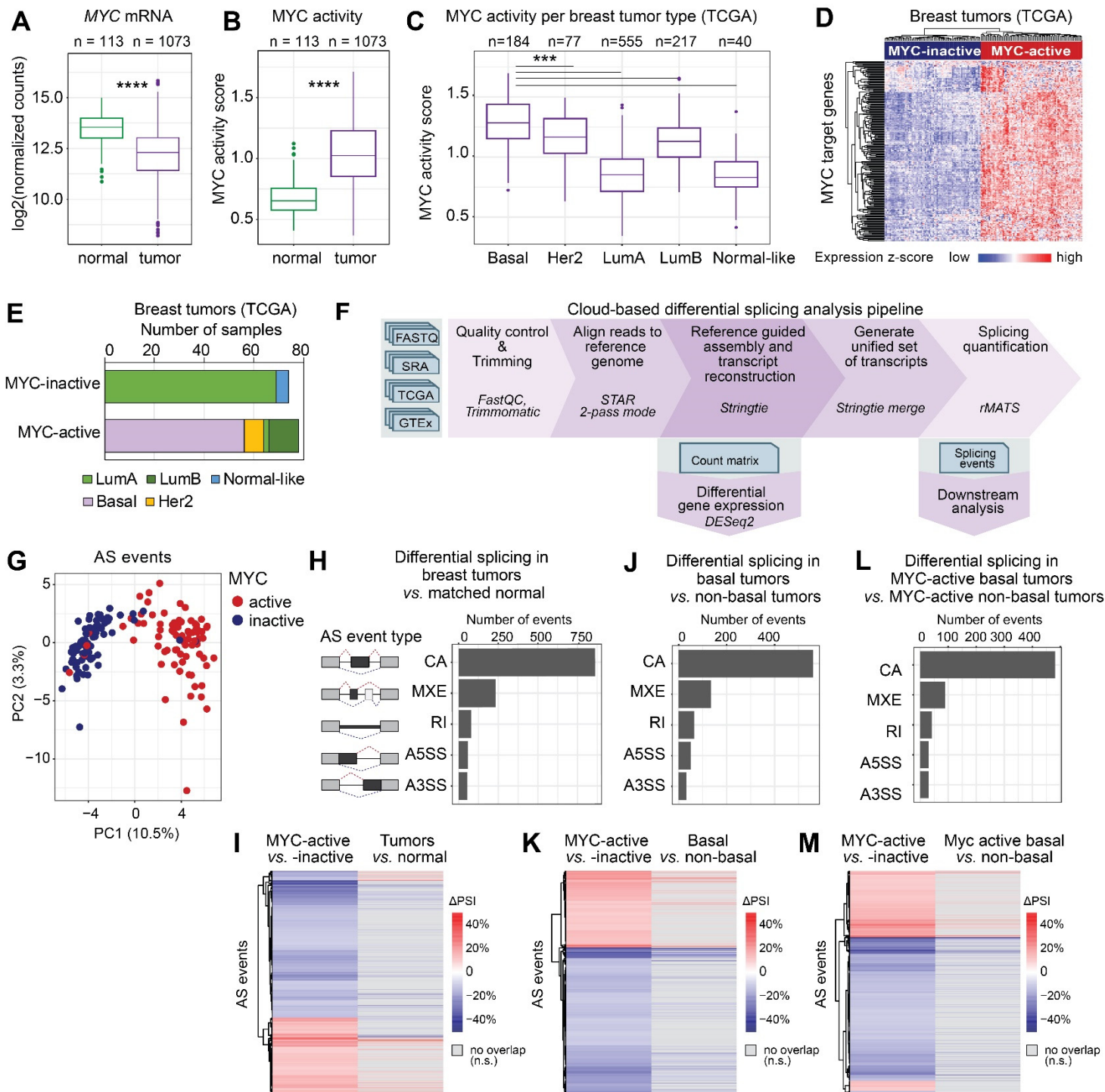


**Cell Reports, Volume 41**

**Supplemental information**

**MYC regulates a pan-cancer network  
of co-expressed oncogenic splicing factors**

**Laura Urbanski, Mattia Brugiolo, SungHee Park, Brittany L. Angarola, Nathan K. Leclair, Marina Yurieva, Phil Palmer, Sangram Keshari Sahu, and Olga Anczuków**



**Figure S1. AS signature in breast tumors classified based on MYC activity** (related to Figure 1).

**(A)** Expression of MYC mRNA transcripts in TCGA adjacent breast normal (n=113) and tumor samples (n=1,073) (median±interquartile range; t-test, \*\*\*\* $P < 0.0001$ ).

**(B)** MYC activity score, calculated based on expression of known MYC target genes defined in GSEA Hallmark MYC targets V1 list (see Methods), for TCGA adjacent breast normal (n=113) and tumor samples (n=1,073) (median±interquartile range; t-test, \*\*\*\* $P < 0.0001$ ).

**(C)** MYC activity score defined as in (B) for five breast tumors subtypes as defined by the PAM50 classification (median±interquartile range; ANOVA, \*\*\* $P < 0.001$ ).

**(D)** Expression of MYC target genes in MYC-active (n=78) or -inactive (n=74) TCGA breast tumors. Each column represents one tumor and each row represents the expression of one gene, normalized as z-score across all tumors.

**(E)** Classification of MYC-active and -inactive tumors from TCGA shown per breast tumor subtype.

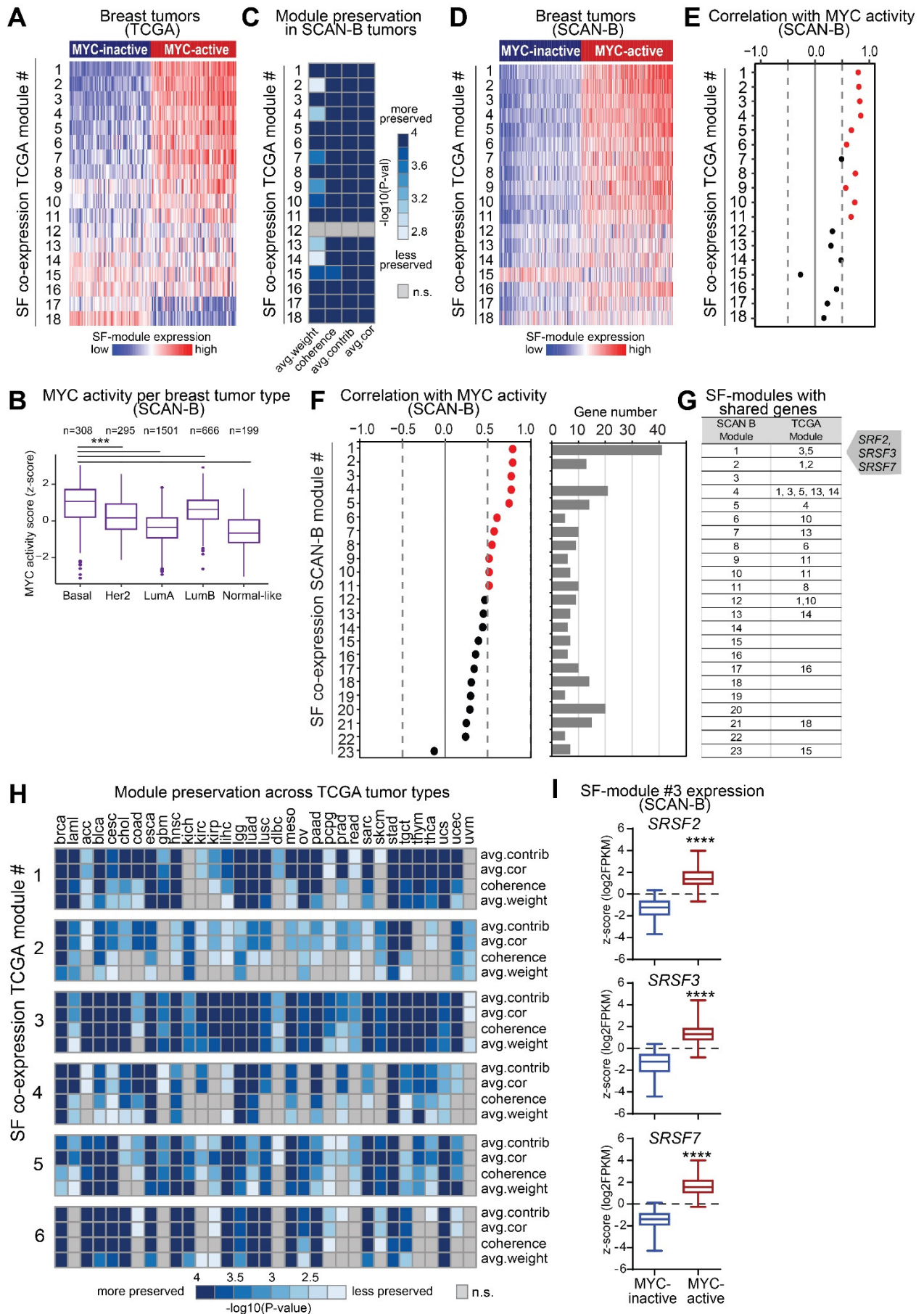
**(F)** Schematic of the cloud-based differential AS analysis pipeline.

**(G)** Principal component analysis of top 5,000 variable AS events between MYC-active (n=78) and -inactive (n=74) TCGA breast tumors.

**(H, I)** Significant AS events detected in matched TCGA breast tumors (n=108) and adjacent normal tissues (n=108) ( $\Delta\text{PSI}>|10\%$ ,  $\text{FDR}<0.05$ ), shown per AS event type (H). Overlap between tumor-associated AS events and MYC-active AS events (I). See also [Table S1D](#).

**(J,K)** Significant AS events detected in basal (n=70) versus non-basal TCGA breast tumors (n=70) ( $\Delta\text{PSI}>|10\%$ ,  $\text{FDR}<0.05$ ), shown per AS event type (J). Overlaps between basal-associated and MYC-associated AS events from TCGA breast tumors are shown (K). See also [Table S1E](#).

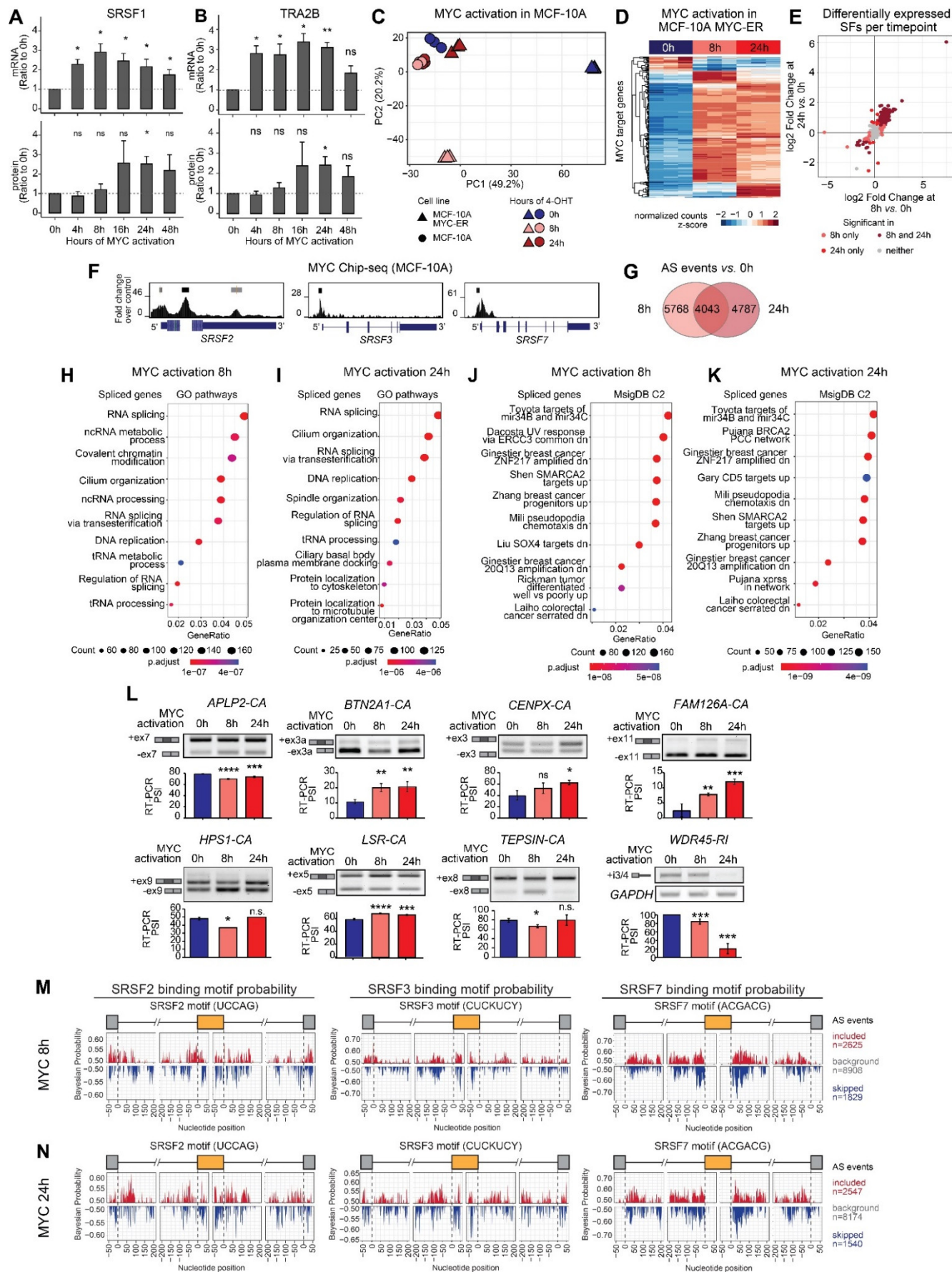
**(L,M)** Significant AS events detected in basal MYC-active (n=57) and non-basal MYC-active (n=23) TCGA breast tumors ( $\Delta\text{PSI}>|10\%$ ,  $\text{FDR}<0.05$ ), shown per AS event type (J). Overlaps between MYC active basal-associated and MYC-associated AS events from TCGA breast tumors are shown (M). See also [Table S1F](#).



**Figure S2.** SFs co-expression in MYC active breast tumors and across cancer types (related to Figures 2 and 3).

- (A)** Relative expression of each SF-module across MYC-active (n=78) and -inactive (n=74) TCGA breast tumors. Each column represents one tumor and each row represents the relative expression of one SF-module, normalized as a Z-score across all tumors.
- (B)** MYC activity scores for SCAN-B breast tumors (n=2,969), shown per tumor subtype as defined by the PAM50 classification (median±interquartile range; ANOVA, \*\*\* $P<0.001$ ).
- (C)** Preservation of TCGA-derived SF-modules in the SCAN-B breast cancer patient cohort (n=2,969), represented as  $-\log_{10}(P\text{-values})$  for four module preservation statistics (avg. weight - average edge weight; coherence - module coherence; avg. contrib. - average node contribution; avg. cor. - average correlation coefficient) using NetRep (Ritchie et al. 2016).
- (D)** Relative expression of each TCGA-derived SF-module in MYC-active (n=212) vs. -inactive (n=188) SCAN-B breast tumors. Each column represents one tumor and each row represents the relative expression of each module, normalized across all tumors.
- (E)** Correlation of TCGA-derived SF-module expression with MYC activity in SCAN-B breast tumors. SF-modules with correlation coefficient  $r>0.5$  are shown in red.
- (F)** Correlation of SCAN-B-derived SF-module expression with MYC activity in SCAN-B breast tumors. SF-modules with correlation coefficient  $r>0.5$  are shown in red. The total number of genes in each SF-module is shown on the right.
- (G)** SF-modules from SCAN-B which share genes with SF-modules from TCGA are indicated. The SCAN-B SF-module which contains *SRSF2*, *SRSF3* and *SRSF7* genes is highlighted.
- (H)** Preservation of six SF-modules across 33 TCGA tumor types, represented as  $-\log_{10}(P\text{-values})$  for four module preservation statistics (avg. weight - average edge weight; coherence - module coherence; avg. contrib. - average node contribution; avg. cor. - average correlation coefficient) using NetRep (Ritchie et al. 2016). See tumor names in [Table S2E](#).
- (I)** Expression of pan-cancer SF-module hub genes *SRSF2*, *SRSF3*, and *SRSF7* in MYC-active (n=212) vs. -inactive (n=188) SCAN-B breast tumors (median±interquartile range; t-test, \*\*\*\* $P<0.0001$ ).





**Figure S3. MYC activation leads to gene expression and AS changes in human mammary epithelial cells (related to Figure 3).**

**(A-B)** mRNA and protein expression of SFs SRSF1 (A) and TRA2 $\beta$  (B) in MCF-10A MYC-ER cells after MYC activation, normalized to 0h. Transcripts levels are measured by Q-PCR, normalized to GAPDH (n=3; mean  $\pm$ sd; t-test, \* $P$ <0.05, \*\* $P$ <0.01). Protein levels are measured by Western blot, normalized to  $\beta$ -catenin (n=3; mean  $\pm$ sd; t-test, \* $P$ <0.05, \*\* $P$ <0.01).

**(C)** Principal component analysis based on gene expression counts for MCF-10A MYC-ER at indicated time points following MYC activation, and in control non-inducible MCF-10A cells.

**(D)** Normalized gene expression (z-score) of 200 MYC target genes from GSEA MSigDB Hallmark MYC targets v1 in MCF-10A MYC-ER at indicated time points following MYC activation. Each column represents one sample and each row represents one gene.

**(E)** Expression of SF in MCF-10A MYC-ER at 8h and 24h. Color indicates whether the SF was significantly differentially expressed in neither dataset, both datasets, or either 8h or 24h only. Each point represents one SF. (neither = 133, both = 121, only 8h = 43, only 24h = 18)

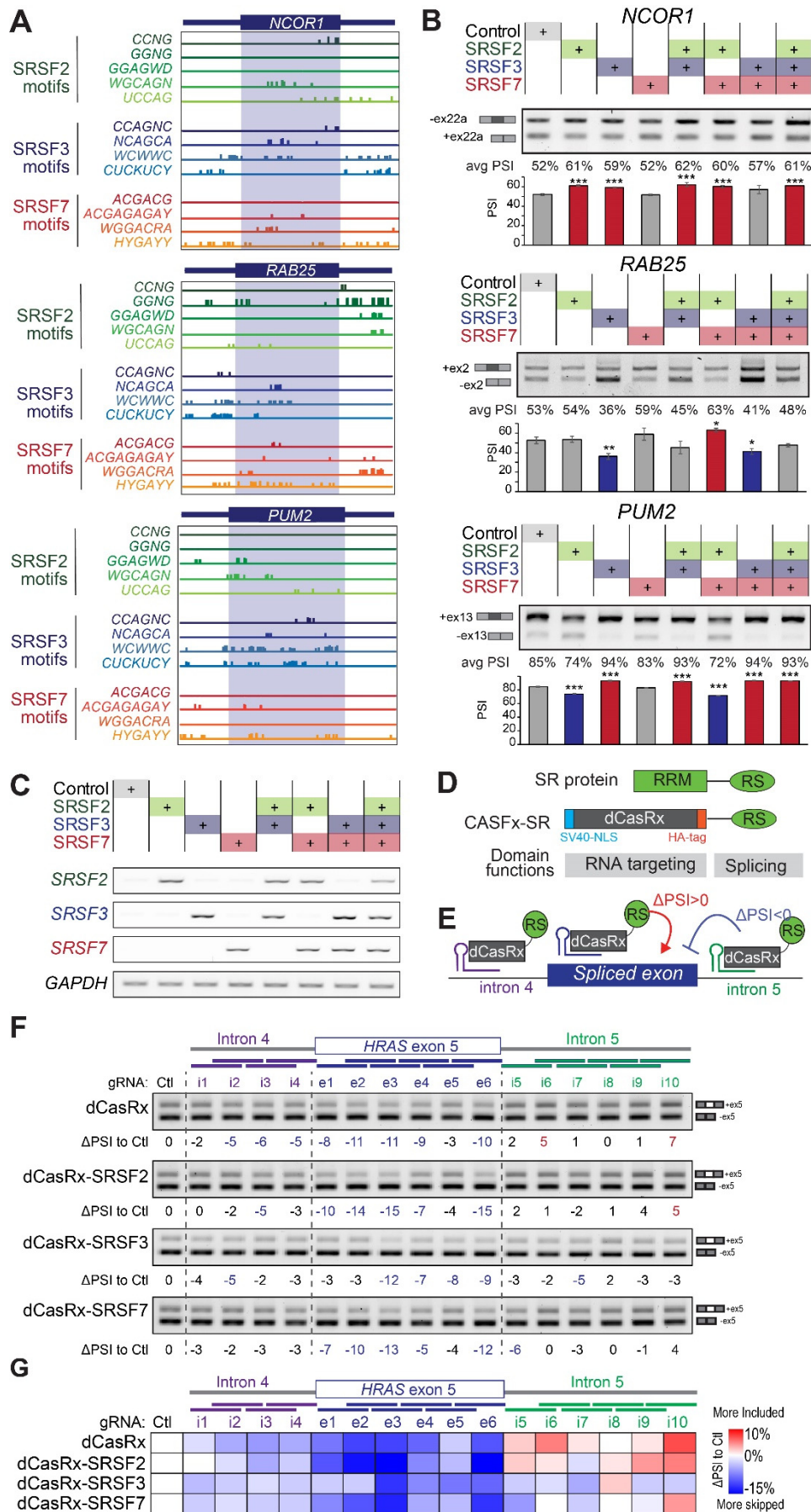
**(F)** MYC binding to *SRSF2*, *SRSF3*, and *SRSF7* genomic region as detected by MYC ChIP seq experiments in MCF-10A from the ENCODE database (ENCODE ENCSR000DOS). Significant peaks are shown by rectangles above the tracks.

**(G)** Overlap of AS events detected MCF-10A MYC-ER cells after MYC activation at 8h or 24h, compared to 0h. Events were considered overlapping if they were detected in both datasets and changing in the same direction.

**(H-K)** Gene ontology analysis using GO biological processes enrichment and MSigDB C2 chemical and genetic perturbations signature for spliced genes at 8h (H,J) and 24h (I,K) after MYC activation.

**(L)** Validation of selected of MYC-regulated AS events in MCF-10A MYC-ER cells at 8h or 24h vs. 0h after MYC activation by RT-PCR using primers that amplify included and skipped events. A representative gel is shown, along with isoform structures. AS quantification as PSI from RT-PCR are shown (n=3; mean $\pm$ sd; t-test, \* $P$ <0.05, \*\* $P$ <0.01, \*\*\* $P$ <0.001, \*\*\*\* $P$ <0.0001, n.s. – not significant). Gene name and AS event types are indicated.

**(M,N)** Probability of the SRSF2, SRSF3, or SRSF7 binding motifs for significant CA exons included ( $\Delta$ PSI $\geq$ 10%) or skipped ( $\Delta$ PSI $\leq$ -10%) in MCF-10A MYC.ER cells at 8h (M) or 24h (N) after MYC activation vs. 0h. Randomly sampled non-significant AS events serve as background. The number of included, skipped, or background CA exons is indicated for each condition. Motif densities are quantified using RBPmap and Bayesian probabilities are calculated by subtracting background motif density from motif density in significantly included or skipped CA events. Probabilities are plotted for each nucleotide position as an aggregate of all AS events mapped across a meta-transcript with CA exon (yellow) and upstream/downstream exons as well as surrounding introns.



**Figure S4. Binding motifs and regulation by pan-cancer SF-module hub genes (related to Figure 4).**

**(A)** Position of predicted binding motifs for SRSF2, SRSF3, or SRSF7 in the spliced sequences (blue box) and surrounding introns (100bp) of validated MYC-regulated AS events in MCF-10A. Consensus motifs sequences derived from *in vitro* or



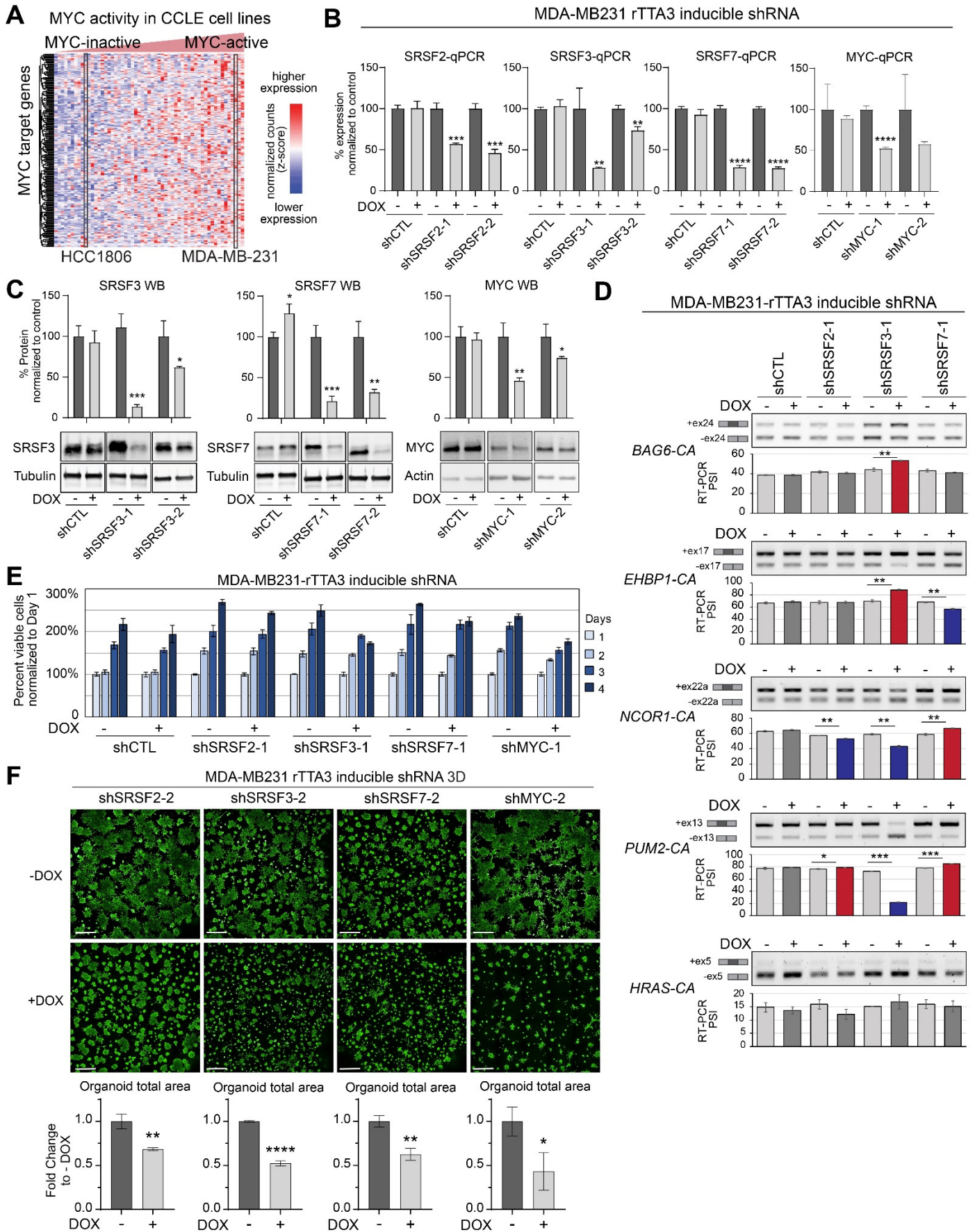
*in vivo* studies are indicated (Ajiro et al., 2016; Dominguez et al., 2018; Kim et al., 2015; Konigs et al., 2020; Paz et al., 2014; Ray et al., 2013).

**(B)** MYC-regulated AS events in HEK293 cells transfected with the coding sequence for either one, two, or all three SR proteins, or control plasmids, is measured by RT-PCR using primers that both amplify included and skipped isoforms. A representative gel is shown, along with isoform structures. AS quantification as PSI from RT-PCR and standard deviation are shown (n=3; mean±sd; t-test, \* $P$ <0.05, \*\* $P$ <0.001, \*\*\* $P$ <0.0001).

**(C)** Expression of HA-tagged SR proteins in HEK293 cells transfected with the coding sequence for either one, two, or all three SR proteins, or control plasmids is measured by RT-PCR with HA-tag and gene specific primers and *GAPDH* reference control.

**(D,E)** Domain structure of CASFx-SR compared to SR protein (D), and gRNA-guided AS modulation principle (E).

**(F,G)** *HRAS* exon 5 inclusion measured by RT-PCR using primers that amplify included and skipped isoforms in HEK293T cells transfected with dCasRx or indicated CASFx-SRs and gRNAs (F). Change in PSI ( $\Delta$ PSI) is measured relative to control (ctl) non-targeting gRNA. Positional effects are visualized as a  $\Delta$ PSI heatmap (G).



**Figure S5.** Knockdown of pan-cancer SF-module hub genes leads to changes in MYC-regulated AS events in MYC-active breast cancer cells (related to Figure 5).

**(A)** Breast cancer lines (n=57) from the Cancer Cell Line Encyclopedia (CCLE) are ranked by MYC activity based on MYC target gene expression. Rows represent expression of genes (normalized counts z-score) and columns represent cell lines. Highlighted cell lines are HCC1806 and MDA-MB231, which were chosen respectively as MYC-inactive and MYC-active breast cancer cell lines for validation.

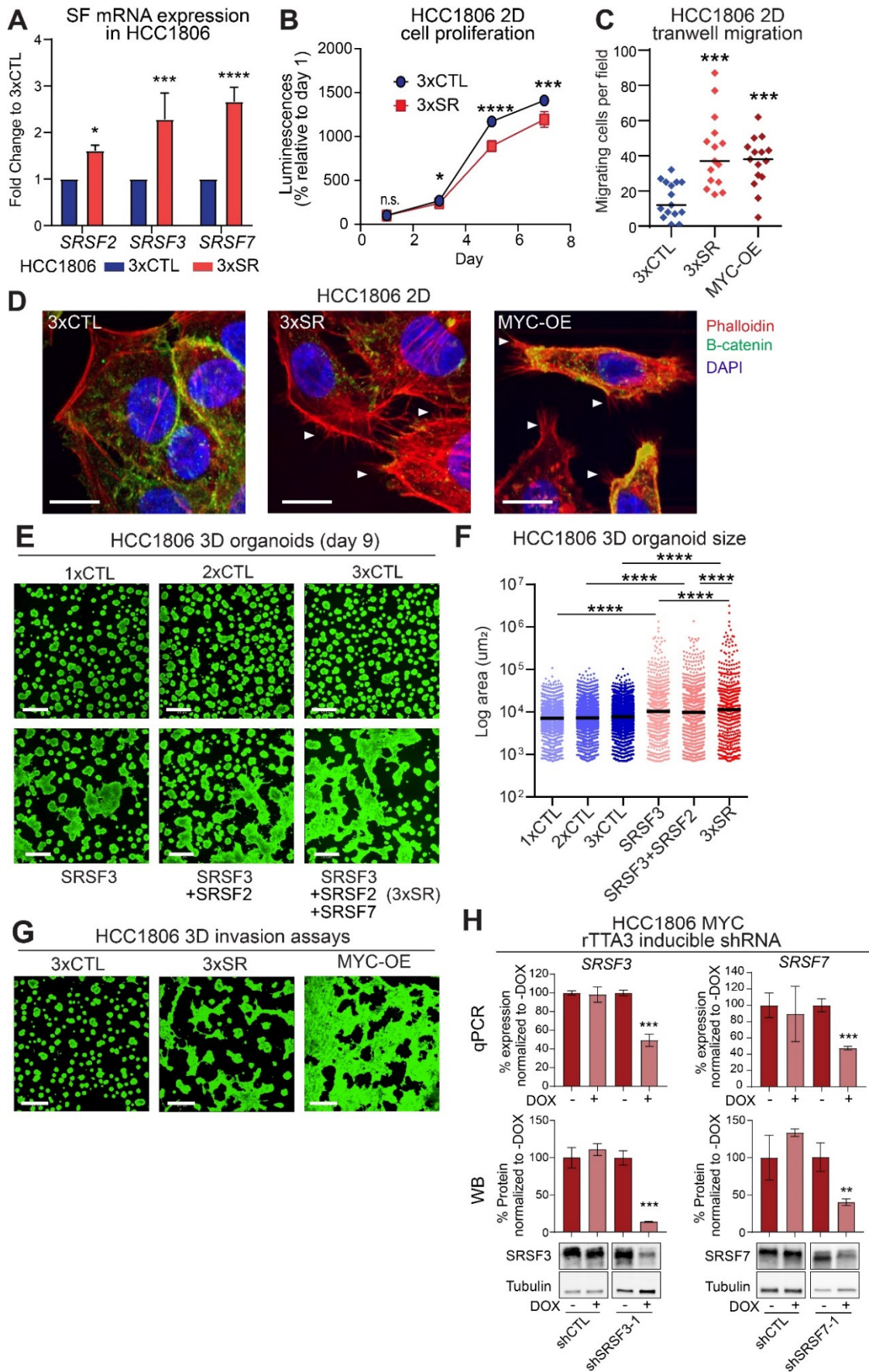
**(B)** Expression of pan-cancer SF-module hub genes *SRSF2*, *SRSF3*, *SRSF7*, or *MYC*, in MDA-MB231-rTTA3 cell lines expressing DOX-inducible shRNAs targeting *SRSF2*, *SRSF3*, *SRSF7*, *MYC*, or control (CTL). Percent expression in +DOX is normalized to the corresponding -DOX cell line (n=3; mean±sd; t-test, \* $P < 0.01$ , \*\* $P < 0.001$ , \*\*\* $P < 0.0001$ ).

**(C)** Protein expression of *SRSF3*, *SRSF7*, or *MYC*, in MDA-MB231-rTTA3 cell lines expressing DOX-inducible shRNAs targeting *SRSF3*, *SRSF7*, *MYC*, or control (CTL) is measured by western blot using indicated antibodies, normalized to Tubulin or Actin loading control (n=3; mean±sd; t-test, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ).

**(D)** RT-PCR validations of selected MYC-regulated AS events in MDA-MB231-rTTA3 cell lines expressing DOX-inducible shRNAs targeting *SRSF2*, *SRSF3*, *SRSF7*, or control (CTL), with primers amplifying both included and skipped AS events. A representative gel is shown, along with isoform structures. AS quantification as PSI from RT-PCR are shown (n=3; mean±sd; t-test, \* $P < 0.05$ , \*\* $P < 0.001$ , \*\*\* $P < 0.0001$ ). Gene name and AS event types are indicated.

**(E)** CellTiter Glo proliferation assay for MDA-MB231-rTTA3 cell lines expressing DOX-inducible shRNAs targeting *SRSF2*, *SRSF3*, *SRSF7*, *MYC*, or control (CTL) normalized to day 1 for each condition (n=4; mean±sd).

**(F)** Representative images of 3D-grown MDA-MB231-rTTA3 cells expressing DOX-inducible shRNA targeting *SRSF2*, *SRSF3*, *SRSF7*, *MYC*, or control (CTL) at day 9 stained with calcein (scale bar: 1mm). For each condition, >30 fluorescent confocal z-stack images are combined as a maximal projection image. Total organoid area is quantified and plotted compared to -DOX control for each condition (n=3, 25 fields per replicate; mean±sd; t-test, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.0001$ , n.s.- not significant).



**Figure S6.** Expression of pan-cancer SF-module hub genes controls cell invasion in a MYC-inactive breast cancer cells (related to Figure 5).

**(A)** Expression of pan-cancer SF-module hub genes *SRSF2*, *SRSF3*, and *SRSF7* in 3xSR and 3xCTL HCC1806 cells (n=3; mean±sd; t-test, \* $P$ <0.05, \*\* $P$ <0.01, \*\*\* $P$ <0.001).

**(B)** CellTiter Glo proliferation assay for 3xSR and 3xCTL HCC1806 cells (n=5; t-test, \* $P$ <0.05, \*\* $P$ <0.01, \*\*\* $P$ <0.001, \*\*\*\* $P$ <0.0001, n.s. – not significant).

**(C)** Cell migration of 2D-grown 3xCTL, 3xSR and MYC-OE HCC1806 cells in transwell assays (n=3, 5 fields per replicate; median; t-test, \*\*\* $P$ <0.001).

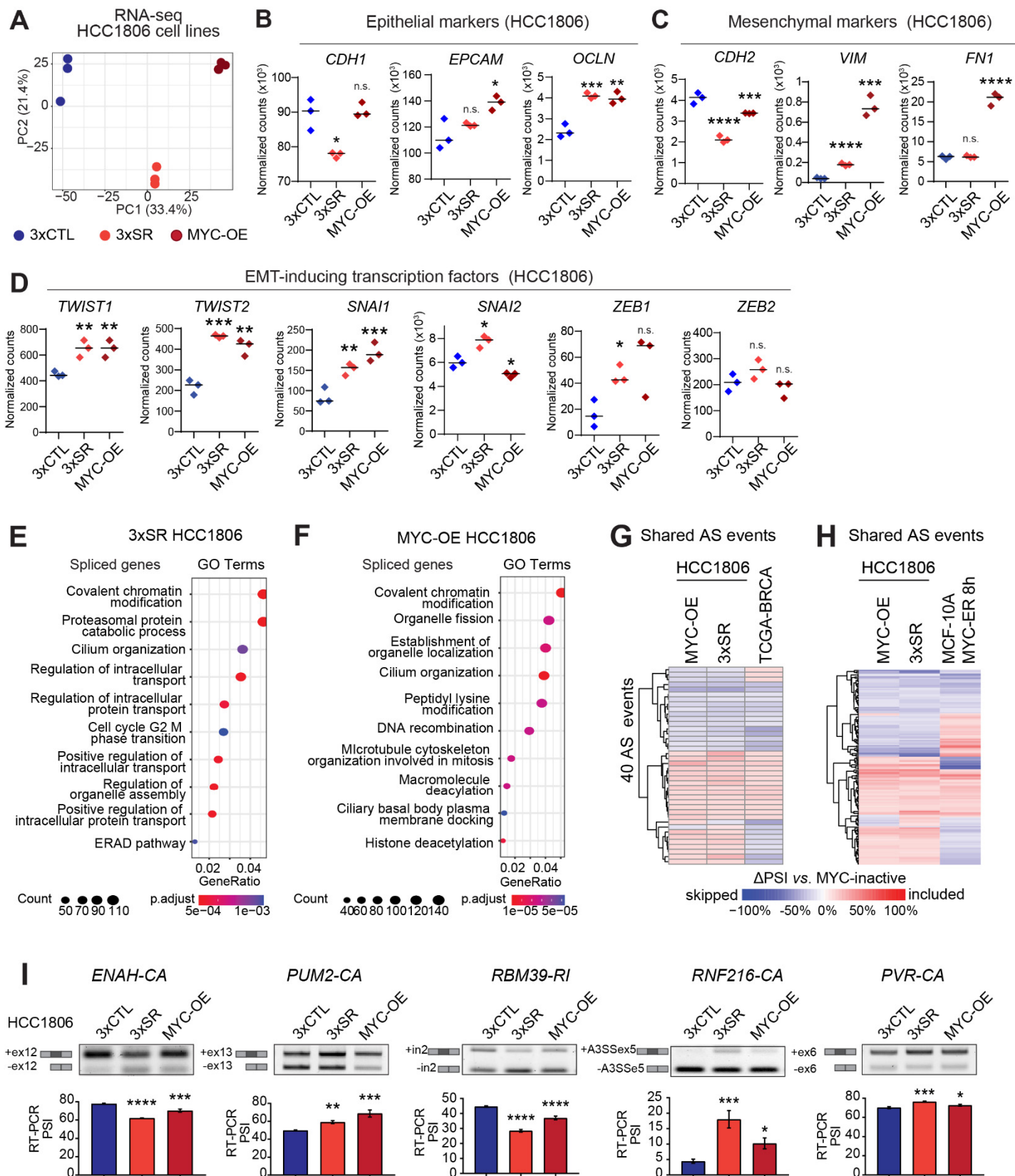
**(D)** Representative images of 2D-grown 3xCTL, 3xSR, and MYC-OE HCC1806 cells stained with phalloidin (red),  $\beta$ -catenin (green), and DAPI (blue) (scale bar 15  $\mu$ m). Actin-rich filopodia are indicated by arrowheads.

**(E,F)** Representative images (E) of 3D-grown 1xCTL and *SRSF3*-OE, 2xCTL and *SRSF3*+*SRSF2*-OE, as well as 3xCTL and 3xSR (*SRSF3*+*SRSF2*+*SRSF7*) HCC1806 organoids at day 9 stained with calcein (scale bar 500  $\mu$ m). For each condition, >20 fluorescent confocal z-stacks are combined as a maximal projection image. Average area of 3D grown HCC1806 structures at day 9 are quantified (n=3, 15-25 fields; median; One-way ANOVA, \*\*\*\* $P$ <0.0001, not shown are n.s.) (F).

**(G)** Representative images of 3D-grown 3xCTL, 3xSR, and MYC-OE HCC1806 organoids in a collagen matrix at day 9 stained with calcein (scale bar 500  $\mu$ m). For each condition, >20 fluorescent confocal z-stacks are combined as a maximal projection image.

**(H)** RNA and protein expression of *SRSF3* and *SRSF7* in HCC1806-rTTA3 MYC-OE cells expressing DOX-inducible shRNA targeting *SRSF3*, *SRSF7*, or control (CTL) is measured by qPCR and Western blot respectively using indicated antibodies, normalized to Tubulin loading control. Percent expression in +DOX is normalized to the corresponding -DOX cell line (n=3; mean±sd; t-test, \*\* $P$ <0.01, \*\*\* $P$ <0.001).





**Figure S7. Overexpression of pan-cancer SF-module hub genes leads to the expression of EMT marker genes and changes in AS in MYC-inactive breast cancer cells (related to Figure 6).**

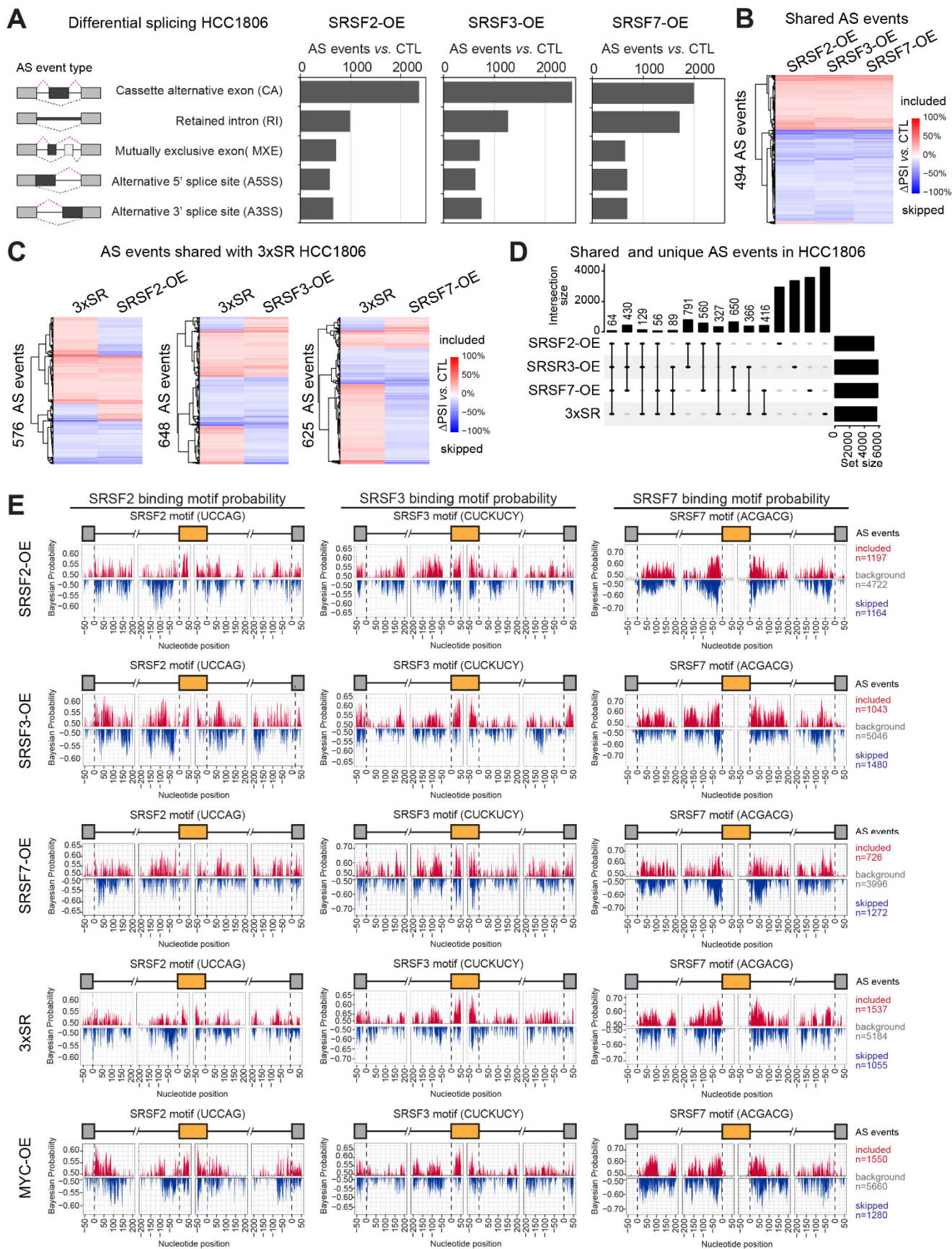
**(A)** Principal component analysis based on gene counts from RNA-seq for 3xCTL, 3xSR, and MYC-OE HCC1806 cells (n=3 per condition). PC1 and PC2 represent 54% of the variance between samples.

**(B-D)** Gene expression of epithelial (B) and mesenchymal cell markers (C), as well as EMT-inducing transcription factors (D) in 3xCTL, 3xSR, and MYC-OE HCC1806 cells profiled by RNA-seq as normalized counts (n=3; median; t-test, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.005$ , \*\*\*\* $P < 0.0005$ , n.s. – not significant).

**(E-F)** Gene ontology analysis using GO biological processes enrichment for spliced genes in 3xSR (E) or MYC-OE (F) vs. 3xCTL HCC1806 cells.

**(G-H)** Overlapping significant AS events between MYC-active 3xSR HCC1806, MYC-OE HCC1806, and TCGA breast tumors (G), and MCF-10A MYC-ER 8h (H).

**(I)** RT-PCR validations of selected AS events in 3xSR, MYC-OE, and 3xCTL HCC1806 cells with primers amplifying both included and skipped events. A representative gel is shown, along with isoform structures. AS quantification as PSI from RT-PCR are shown (n=3; mean±sd; t-test, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ , n.s. – not significant). Gene name and AS event types are indicated.



**Figure S8. Overexpression of pan-cancer SF-module hub genes SRSF2, SRSF3, or SRSF7 leads to changes in MYC-regulated AS events** (related to Figure 6).

(A) Number of significant AS events detected in SRSF2-OE, SRSF3-OE, or SRSF7-OE vs. CTL HCC1806 cells ( $n=3$  per condition;  $|\Delta$ PSI $\geq$ 10%; FDR $<$ 0.05) shown per AS event type.

**(B)** Overlapping significant AS events in SRSF2-OE, SRSF3-OE, and SRSF7-OE HCC1806 cells. Each row represents one AS event.

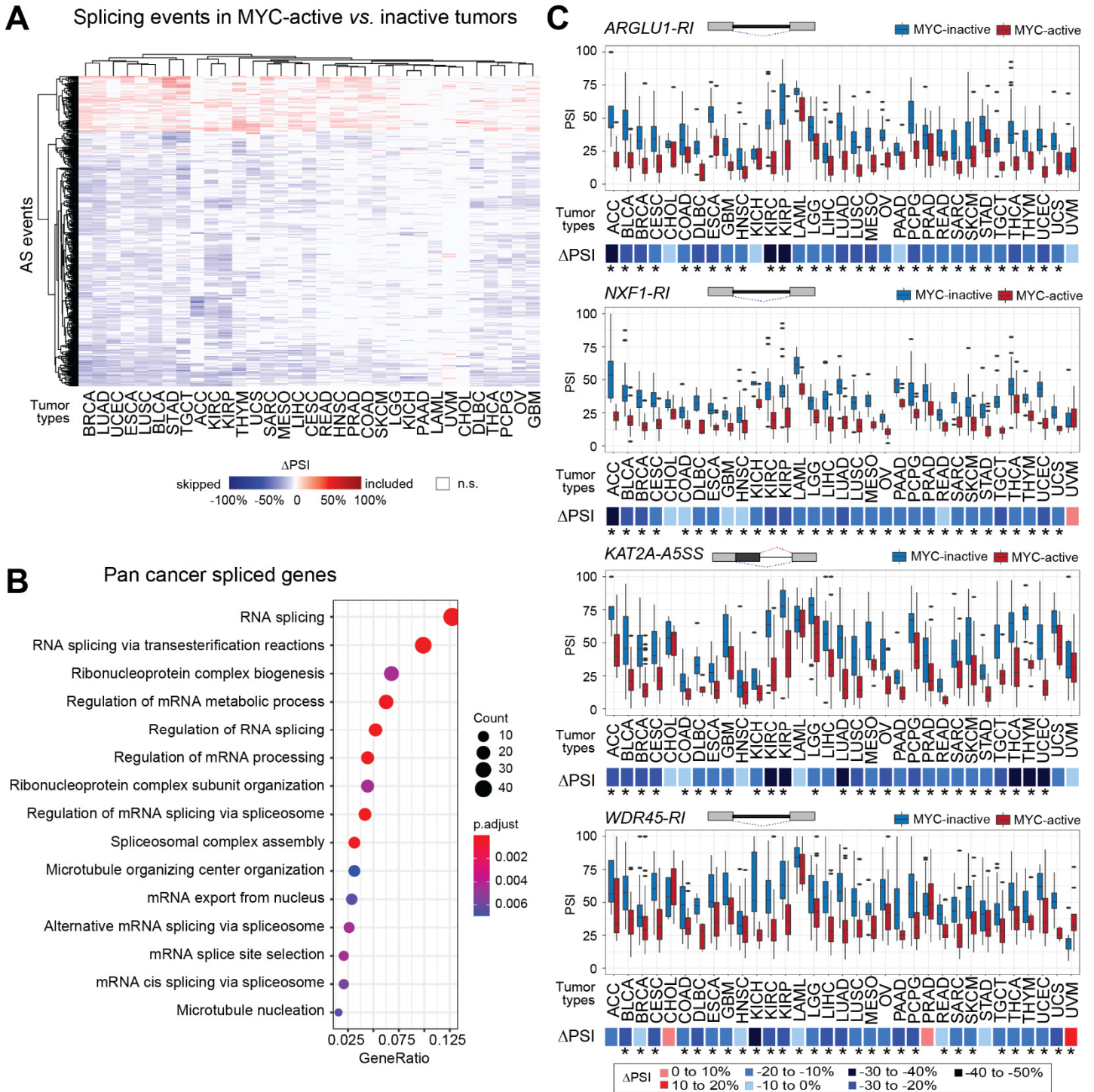
**(C)** Overlapping significant AS events in 3xSR vs. SRSF2-OE, or SRSF3-OE, or SRSF7-OE HCC1806 cells. Each row represents one AS event.

**(D)** Number of overlapping significant AS events in SRSF2-OE, SRSF3-OE, SRSF7-OE, 3xSR, or MYC-OE HCC1806 cells.

**(E)** Probability of the SRSF2, SRSF3, or SRSF7 binding motifs for significant CA exons included ( $\Delta\text{PSI} \geq 10\%$ ) or skipped ( $\Delta\text{PSI} \leq -10\%$ ) in SRSF2-OE, SRSF3-OE, SRSF7-OE, 3xSR, or MYC-OE HCC1806 compared to their respective controls. Randomly sampled non-significant CA events serve as background. The number of included, skipped, or background CA exons is indicated for each condition. Motif densities are quantified using RBPmap and Bayesian probabilities are calculated by subtracting background motif density from motif density in significantly included or skipped CA events. Probabilities are plotted for each nucleotide position as an aggregate of all CA events mapped across a meta-transcript with CA exon (yellow) and upstream/downstream exons as well as surrounding introns.

See also [Table S4](#).





**Figure S9. MYC-regulated AS events shared across tumor types** (related to Figure 7).

**(A)** Hierarchical clustering of all the pan-cancer AS events differentially spliced in MYC-active vs. -inactive tumors detected in  $\geq 25\%$  of TCGA tumors, all subtypes combined (see tumor names in Table S2E). Each column represents one tumor type and each row represents the differential percent-spliced-in ( $\Delta$ PSI) value of a specific AS event in MYC-active vs. -inactive, shown per tumor type. See also Table S5.

**(B)** Gene ontology analysis using GO biological processes for MYC-regulated AS events spliced in  $\geq 25\%$  of TCGA tumor types.

**(C)** Percent spliced in (PSI) values of pan cancer MYC-regulated AS event detected in MYC-active (red) vs. -inactive (blue) tumors, shown per tumor type (median $\pm$ interquartile range).  $\Delta$ PSI (MYC-active vs. -inactive tumors) are shown as a heatmap below per tumor type. Significant  $\Delta$ PSI changes (FDR $<$ 0.05) are indicated by an asterisk.