Supporting Information

Ke et al. 10.1073/pnas.1714573115

SI Experimental Procedures

Data Used. The breast cancer PSI data file (PSI_download_BRCA.txt) and exon annotation file (TCGA_SpliceSeq_Gene_Structure.txt) were acquired from the website bioinformatics.mdanderson.org/TCGASpliceSeq/. The PSI file included seven types of splice events (alternate acceptors, alternate donors, ES, retained intron, alternate promoters, alternate terminators, and mutually exclusive exons), with splice events occurring in over 75% of samples retained. The expression file (BRCA.mRNAseq_RPKM.txt and BRCA.mRNAseq_raw_counts.txt) and clinical information file (BRCA.merged_only_clinical_clin_format.txt) for breast cancer tumor subtypes in the TCGA were downloaded from the November 2015 data archive (https://portal.gdc.cancer.gov/).

Breast Cancer Subtype Classification. The breast cancer samples were classified according to ER, PR, and HER2 status derived from the TCGA clinical dataset (BRCA.merged_only_clinical_clin_format.txt) (5): luminal A: ER^+, $PR^{+/-}$, $HER2^-$; luminal B: ER^+, $PR^{+/-}$, $HER2^+$; TNBC: ER^-, PR^-, $HER2^-$; HER2+: ER^-, PR^-, $HER2^+$. Other samples that did not follow the above criteria were excluded from subsequent analyses. Clinical subtype results are shown in Dataset S1.

Comparison of Splicing Profiles. Two methods were used to characterize the splicing profiles across breast cancer subtypes. First, splicing profiles for each subtype were directly compared against a pool of the other subtypes. To maintain balanced sets for comparison and avoid biases due to sample selection, we randomly selected 30 samples for a given subtype and a pool of 30 samples from the other three subtypes (TNBC vs. non-TNBC; luminal A vs. nonluminal A; luminal B vs. nonluminal B; HER2+ vs. non-HER2+). Splicing events with $|\Delta PSI| > 0.1$ and FDR corrected $P < 0.05$ (two-sided Mann–Whitney test) were identified as differential AS events, and the entire procedure was repeated 10,000 times. Then, Student's t test was applied to examine the significance of the number of differential AS events between any two comparisons (TNBC vs. non-TNBC; luminal A vs. nonluminal A; luminal B vs. nonluminal B; HER2+ vs. non-HER2+). To avoid possible bias due to the $|\Delta PSI|$ threshold, we also tried a series of $|\Delta PSI|$ values, including 0.05, 0.1, 0.15, and 0.2, which all showed similar results. Finally, splice events with a significant difference between any two breast cancer subtypes were identified. A heatmap of these 3,049 differential splice events and four breast cancer subtypes was draw with the R package pheatmap (version 1.0.8).

Overlap Analysis of Gene-Associated Events in TNBC-Specific Events. Overlap analysis of gene-associated events in TNBC-specific events was used to assess how much of the TNBC-specific splicing profile could be explained by each splicing factor. First, TNBC-specific splicing ES events were isolated by comparing splicing events from the TNBC subtype against those from the other three subtypes. The significant splicing events were determined based on $|\Delta PSI| > 0.1$ and FDR < 0.05 (Mann–Whitney test). Second, gene-associated ES events were determined by correlation analyses using gene-expression levels (RPKM) and PSI values across the breast cancer samples. Correlation coefficients (R value) and corresponding P values were calculated with the Pearson method, and gene-associated events were determined by $R > 0.2$ and $P < 0.0001$. Common events were considered as events that occurred (overlapped) in both gene-associated and TNBC-specific ES events. Finally, enrichment analyses were conducted by Fisher's exact test to determine whether gene-associated splicing events were enriched in TNBC-specific splicing events, in which all associated splicing events (7,806) from the 221 splicing factors were used as the background control, with 576 TNBC-specific splicing events.

Coexpression Analysis. The list of 221 known or putative splicing factors was obtained from a previous study (Dataset S4) (31). These 221 splicing factors and the breast cancer expression dataset with all samples were used for coexpression analyses. Correlation coefficients (R value) and corresponding P values between these genes were then calculated with the Pearson method. The correlation matrix ordered by hierarchical clustering was shown as a heatmap. Clustering distance was determined by the Euclidean method. A core set of genes was found to be coexpressed with R value 0.64 ± 0.11 (mean ± SD) and $P <$ 2.25e-12. The R functions cor and cor.test were used to calculate the correlation coefficients and P values, respectively. Pheatmap (version 1.0.8, R package) was used to draw the heatmap.

Physical Interaction Networks. To identify the significance of the 57 proteins (Fig. 1*C*), we constructed a protein–protein interaction network using the Cytoscape_3.2.0 (62) plugin GeneMANIA (63). Two gene products are considered linked if they are found to interact in a protein–protein interaction study. These interactions were collected from the biological general repository for interaction datasets BioGRID (64) and PathwayCommons (www.pathwaycommons.org/). The *Homo sapiens* (human) database was selected; the queried genes were the 57 genes in Fig. 1*C*, with no other related genes involved.

RNA-Seq Analysis. Raw sequence data were processed through the standard Illumina pipelines for base-calling and fastq file generation. Paired-end reads were mapped to the human genome primary assembly (GRCh37) (65), and the Ensembl human gene annotation for GRCh37 genebuild was used to improve accuracy of the mapping with STAR software (STAR_2.4.2a) (66). FeatureCounts (version 1.4.6-p5) (67) was used to assign sequence reads to genes. Mitochondrial genes, ribosomal genes, and genes possessing fewer than five raw reads in half the samples were removed. Differential expression analysis was performed with the Bioconductor edgeR package 1.6 (68), in which we used an overdispersed Poisson model with a common dispersion parameter combined with the exact test. Significant genes were determined by an adjusted P value of <0.05 based on the Benjamini–Hochberg multiple-testing correction and a log2-transformed fold change >2 or less than −2. For AS analysis, MISO software (version 0.5.4) (69) was used to analyze RNA-seq data and estimate the percentage of splicing isoforms (Ψ values, for percent spliced isoform). Significantly different AS events were determined by a Bayes factor >6 and $|\Delta PSI| > 0.2$.

Survival Analysis. Kaplan–Meier Plotter (kmplot.com/analysis/) was used to conduct survival analysis. Survival type was set to relapse-free survival (RFS). Intrinsic subtype was set to basal. Significantly different survival was based on a log-rank test, $P < 0.05$.

RNA Extraction, qPCR, and Semiquantitative RT-PCR. Total RNA was isolated using TRIzol reagent (Life Technologies) and reverse transcribed using a PrimeScript RT reagent kit with gDNA Eraser (TaKaRa) according to the manufacturer's instructions. The cDNA was then used for RT-PCR or qRT-PCR. RT-PCR was conducted using 2× Goldstar Best Master Mix (catalog no. 0655M; CWBIO), and qRT-PCR was performed using SYBR Green PCR Master

Mix (Applied Biosystems) on an ABI 7500 apparatus. The RT-PCR products were separated on a 3% agarose gel and scanned with a Tanon-3500 GIS (Tanon). The amount of each splicing isoform was measured with ImageQuant 5.2. Primers are listed in Table S1.

Immunohistochemistry and Immunofluorescence. Tissue chips of human breast cancers were purchased from several companies (AlenaBio catalog nos. BR8013, BR248a, T086, BR1009; Shanghai Outdo Biotech catalog nos. HBre-Duc150Sur-02, HBre-Duc046Sur-02). Antigen retrieval was performed by boiling the samples twice in 10 mM sodium citrate buffer for 20 min. Endogenous peroxidases were inactivated by incubation with 3% H_2O_2 for 15 min. After blocking with 10% goat serum and 0.1% Triton X-100/PBS for 2 h, the samples were incubated with primary antibody against TDP43 (1:150; Abcam) at 4 °C overnight, with HRP-linked secondary antibodies (Sigma) at 1:200 in 0.25% BSA in PBS (BSA/PBS) for 1 h, and then finally were developed with 3,3-diaminobenzidine (DAB). Following counterstaining with hematoxylin, the slides were examined with an inverted microscope (Nikon), with visible brown granules indicative of positive staining. The staining intensity and degree were reviewed based on judgment standards as follows: "+," no or faint staining; "++," moderate staining; "+++," strong staining.

For immunofluorescence, cells were fixed with 4% paraformaldehyde and blocked with 5% goat serum/PBS and 0.1% Triton X-100/PBS. Cells were then incubated with primary antibodies for 2 h and with secondary antibodies for 1 h at room temperature. The primary antibodies used for immunostaining were rabbit anti-TDP43 (Abcam), mouse anti-SRSF3 (Santa Cruz), rabbit anti-SRSF3 (Abcam), and mouse anti-BrdU (Life Technologies). The secondary antibodies used for immunostaining were fluorescein-labeled anti-rabbit (KPL), fluorescein-labeled anti-mouse (KPL), cy3 goat anti-mouse (Life Technologies), and cy3 goat anti-rabbit (Life Technologies). Samples were mounted and counterstained with DAPI (H-1200; Vector Laboratories) before visualization. BrdU was observed under an inverted microscope (Nikon), and colocalization analysis was performed by laser-scanning confocal microscopy.

Cell Culture. All cell lines were obtained from ATCC and were maintained under standard culture conditions (37 °C, 5% CO_2) in culture medium with 1% penicillin/streptomycin solution. The MDA-MB231 cells were cultured in DMEM/F12 with 10% FBS. The HEK293T cells were cultured in DMEM with 10% FBS. The 4T1 and HCC1806 cells were cultured in RPMI-1640 medium with 10% FBS. The MCF10A cells were cultured in DMEM/F12 supplemented with 5% horse serum, 20 ng/mL EGF, 10 μg/mL insulin, 0.5 μg/mL hydrocortisone, 100 ng/mL cholera toxin, and antibiotics. The 3D culture of MCF10A cells was performed as described previously. In brief, growth factor-reduced Matrigel (catalog no. 354230; BD Biosciences) was layered into eight-well glass chamber slides (catalog no. C7057-1PAK; Sigma) to form a reconstituted basement membrane. Then 5,000 MCF10A cells per well were seeded on top of the layer in medium containing 2% Matrigel and 5 ng/mL EGF. Cells were cultured in a 5% CO_2 humidified incubator at 37 °C for 8–10 d.

Knockdown and Overexpression. The shRNAs were inserted into a cloning vector (pLKO) and were transfected with psPAX2 and pMD2.G (4:3:1) into 293T cells to produce lentiviral particles. The shRNA virus was then infected into target cells, with sh-scramble (catalog no.1864; Addgene) or sh-TRC (catalog no.10879; Addgene) as controls. The MDA-MB231 and HCC1806 cells were collected for RNA extraction and incubated with the virus for 3 d. The knockdown efficiency was verified by real-time qPCR and Western blot analysis. The sh-RNA sequences are listed in Table S1.

For overexpression, full-length TDP43 (catalog no. 27470; Addgene), SRSF3 (catalog no. 46736; Addgene), and PAR3b (catalog no. 19388; Addgene) vectors were subcloned into the pCDH lentiviral expression vector. For NUMB overexpression, cDNA from normal human breast tissue was used as a template to amplify NUMB into the pCDH overexpression vector. The truncated or mutated vectors related to these proteins were amplified from plasmids containing the full-length related genes. Sequencing was performed to verify all plasmid constructs to exclude mutations. Plasmids were transiently transfected into 293T cells with Lipofectamine (Invitrogen) according to the manufacturer's manual. Cells were collected for further analyses or experiments 48 h later. For stably expressed proteins, psPAX2 and pMD2.G were cotransfected into 293T cells to produce lentiviral particles, which were then used to infect MDA-MB231 or HCC1806 cells. Stable cell lines were selected with 2 μg/mL puromycin 72 h after infection and were maintained in culture for 10–15 d.

Cell Proliferation, Cell Cycle, and Cell Apoptosis Assays. Cell viability was measured by MTS assay using the CellTiter 96 AQueous One Solution Reagent (catalog no. G3581; Promega) according to the manufacturer's protocols. For cell cycle analysis, cells were fixed with 70% ethanol at 4 °C overnight, washed with PBS, and then incubated in PBS with 1 mg/mL RNase A, 100 mg/mL PI, and 0.6% Nonidet P-40 at 37 °C for 30 min. Cell-cycle distribution was determined by flow cytometry (LSRFortessa; BD). An apoptosis assay was performed using an FITC Annexin V apoptosis detection kit I (556547; BD Biosciences). After flow cytometry, data were analyzed with FlowJo software.

Wound-Healing Assay. The MDA-MB231 cells were seeded into six-well plates in confluent monolayers. Scratch wounds were created using a sterilized tip in serum-free medium. Images were captured at 0 and 36 h after wounding. Averages ± SEM for three independent measurements are shown.

Colony-Formation Assay. The MDA-MB231 cells were seeded into six-well plates (5,000 cells per well) and incubated at 37 °C for 2 wk. Colonies were stained with 0.1% crystal violet. The number of colonies ± SEM for three independent measurements are shown.

Cell Migration and Invasion Assays. For cell migration, the MDA-MB231 or HCC1806 cells were starved for 15 h in DMEM/F12 containing 0.1% BSA. Cells (1×10^5) in serum-free medium were placed in the top chamber of the Transwell (Corning). The cells in the top chamber were then transferred to the bottom chamber, which contained complete medium. Cells were incubated at 37 °C for 8–12 h for MDA-MB231 and for 24–48 h for HCC1806. For visualization, cells in the top well were removed by wiping the top of the membrane with cotton swabs. The membrane filter was then stained with 0.5% crystal violet and photographed. For MDA-MB231 cell invasion, the top Transwell chamber was first coated with Matrigel (BD Biosciences); all other procedures were the same as those applied for migration, although with a longer invasion time (12–24 h).

Co-IP and Western Blotting. Cells were lysed in cell lysis buffer with protease inhibitor mixture (no. B14001; BioTools) for 30 min. For DNase/RNase treatment, cell extracts were digested with 50 μg/mL DNase (no. 11284932001; Roche) and RNase A (no. EN0531; Thermo Scientific) for 15 min at 37 °C before incubation with beads. The M2 beads (Sigma) or protein A agarose beads (Santa Cruz) were incubated with whole-cell extracts at 4 °C for 8–12 h. The beads were washed with cell lysis buffer three to five times. Finally, the beads were boiled in 5 × SDS for 10 min. The eluents were analyzed by Western blotting. Protein lysates were separated

by SDS/PAGE, transferred onto PVDF membranes, probed with primary antibodies at 4 °C overnight and with HRP-linked secondary antibodies (Sigma) at room temperature for 1 h, and then detected with chemiluminescent HRP substrate (Millipore). The antibodies used for immunoblotting were as follows: PARP (no. 9542; CST), TDP43 (ab109535; Abcam), SRSF3 (ab198291; Abcam), SRSF7 (ab137247; Abcam), vimentin (D21H3; CST), E-cadherin (24E10; CST), FLAG (no. 14793; CST), fibronectin (F7387; Sigma), N-cadherin (ab76057; Abcam), PRR3 (ab64646; Abcam), α-tubulin (T5168; Sigma), GAPDH (sc-25778; Santa Cruz), P21 (ab109199; Abcam), P27 (610241; BD), P57 (ab75974; Abcam), SRSF1 (32-4500; Life Technologies), cleaved caspase-3 (9661; CST), and hnRNP A1 (ab5832; Abcam).

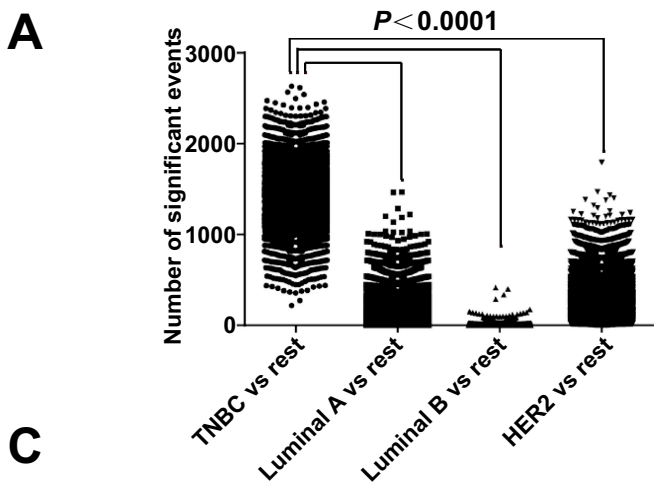
RIP and RIP-Seq. The RIP kit (Millipore catalog no. 17-701) was used per the manufacturer's instructions. MDA-MB231 cells overexpressing Flag-SRSF3 or flag-TDP43 were collected and lysed in RIP lysis buffer. The cells were then immunoprecipitated with Flag (1:20; CST) or IgG (5 μg; Millipore) antibodies against the RNA-binding protein of interest at 4 °C for 8 h with protein A/G magnetic beads. The protein A/G magnetic beads were then washed with RIP buffer five times to discard unbound material. RNA was purified for downstream applications, including RT-PCR, real-time qPCR, and RIP-seq. The PCR-related primers are listed in Table S1. For RIP-seq, RNA from three independent RIP assays was mixed for deep sequencing. ASPeak software was used to identify protein-associated transcripts. The transcripts were regarded as significantly binding to SRSF3 or TDP43 if they fulfilled the following conditions: FPKM >10 in the RIP or input datasets, ratio of log2 FPKMrip/FPKMinput greater than or equal to −0.1, and detected by ASPeak.

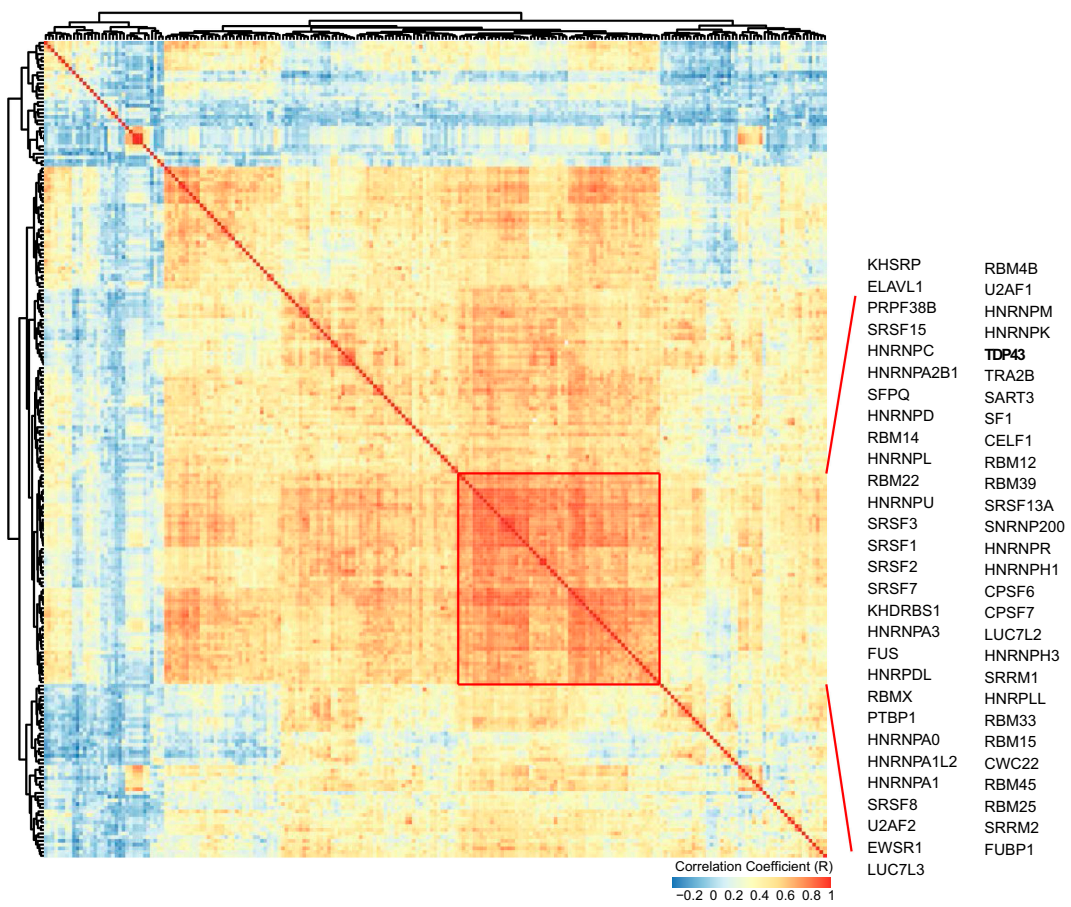
Splicing Reporter Constructs. The wild-type SYNGAP1 minigene construct was created by inserting the SYNGAP1 exon 13–15 regions separated by introns 13–14 (corresponding to hg19 chr5:4889604–4891365) into the mammalian expression vector (no. 17448; Addgene). The TDP43- and SRSF3-binding sites (CCTCAAC and TGTGTG, respectively) were mutated by introducing site-directed mutagenesis (CGTTTAC and AACGAA) using Phusion High Fidelity polymerase (no. M0491; New England Biolabs) and primers containing the corresponding mutations. All primers used in this study are listed in Table S1.

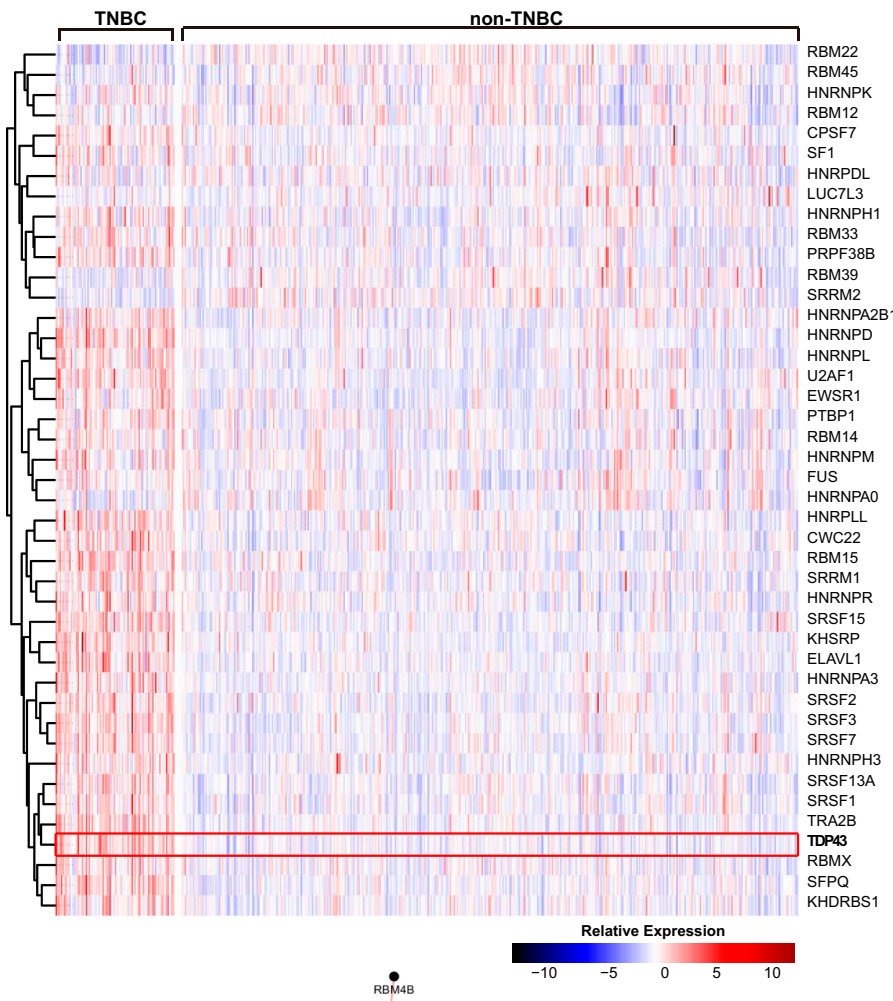
Xenograft Assay. The MDA-MB231 cells transduced with sh-SRSF3 or sh-TDP43 were resuspended (1×10^6) in PBS/Matrigel mixture and then were injected orthotopically into 8-wk-old female nude mice. Each experimental group contained six mice. Tumor size was measured using a caliper every week. At the end of 8 wk, all mice were killed by cervical dislocation, and tumors were collected and weighed.

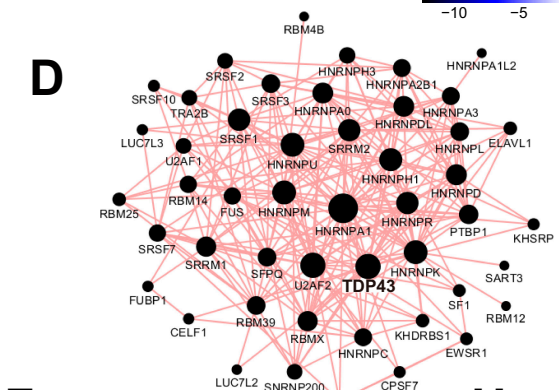
Lung Metastasis Assay. Stable luciferase-expressing 4T1 or MDA-MB231 cells in PBS were injected into the tail veins of mice together with cells expressing a neutral hairpin as the control. The 4T1-luc cells were injected into 8-wk-old BALB/C mice ($n = 8$). The MDA-MB231-luc cells were injected into 8-wk-old NOD/SCID mice ($n = 5$). For bioluminescence imaging, mice were injected in the abdomen with 100 mg/g of D-luciferin in PBS. Five minutes after injection, mice were anesthetized by anesthetic ketamine, and bioluminescence was imaged with a CCD camera once a week. Four weeks (4T1-luc cells) or 8 wk (MDA-MB231-luc cells) later, all mice were killed by cervical dislocation for lung tissue sampling, and data were analyzed.

LC-MS/MS. Bands of PAR3 were excised, destained with 50 mM ammonium bicarbonate in 30% acetonitrile, and dehydrated with acetonitrile. Samples were then reduced by 10 mM DTT at 56 °C for 30 min and alkylated with 55 mM iodoacetamide at room temperate for 10 min. Samples were then subjected to in-gel digestion with trypsin (enzyme:protein ratio 1:50) at 37 °C overnight. Peptides were extracted with 60% acetonitrile and 1% TFA and were lyophilized for further analysis. Reverse-phase HPLC (RP-HPLC) separation was achieved on the Easy nLC system (Thermo Fisher Scientific) using a self-packed column (75 μm × 150 mm; 3 μm ReproSil-Pur C18 beads, 120 Å; Dr. Maisch GmbH) at a flow rate of 300 nL/min. The mobile phase A of RP-HPLC was 0.1% formic acid in water, and B was 0.1% formic acid in acetonitrile. The peptides were eluted using a gradient (2–90% mobile phase B) over a 60-min period into a Q Exactive mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in data-dependent mode with each full MS scan (m/z 350–2,000) followed by MS/MS for the 15 most intense ions with the following parameters: precursor ion charge equal to or greater than +2, 2-Da precursor ion isolation window, and 27 normalized collision energies of higher energy collisional dissociation (HCD). Dynamic exclusion was set for 30 s. The full MS and subsequent MS/MS analyses were scanned in the Orbitrap analyzer with $R = 70{,}000$ and $R = 17{,}500$, respectively. The MS data were analyzed using MaxQuant software (www.coxdocs.org/doku.php?id=maxquant:start, version 1.6.0.1). Proteins were identified by searching MS and MS/MS data of peptides against the UniProt human proteome database (November 2017). Trypsin/P was selected as the digestive enzyme with two potential missed cleavages. The search included variable modifications of methionine oxidation and N-terminal acetylation and fixed modification of cysteine carbamidomethylation. Peptides with a minimum of seven amino acids were used for analysis. For peptide and protein identification, the FDR was set to 0.01.









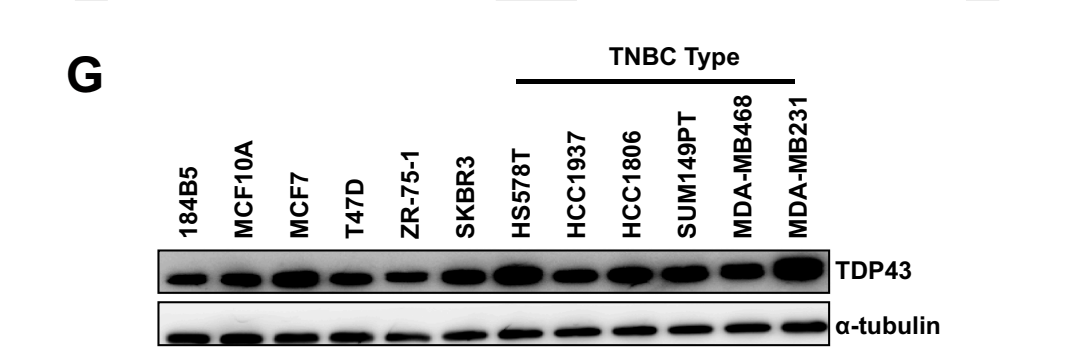


Fig. S1. Related to Fig. 1. Unique alternative splicing profile and high expression of TDP43 with poor prognosis in TNBC. (*A*) The number of different splicing events in a given subtype vs. those in a pool of the other three subtypes using the resampling method. The criteria used to identify significantly different splicing events between groups of samples were $|\Delta PSI| > 0.1$, Mann–Whitney test followed by FDR correction, $P < 0.05$. TNBC vs. rest represents the comparison between 30 randomly resampled TNBC and 30 randomly resampled non-TNBC samples (luminal A, luminal B, and HER2+ subtypes) 10,000 times. Luminal A vs. rest, luminal B vs. rest, and HER2+ vs. rest subtypes represent comparisons between luminal A and nonluminal A, between luminal B and nonluminal B, between and HER2+ and non-HER2+, respectively. *P* values were calculated by Student's *t* test. (*B*) Heatmap showing pairwise gene coexpression matrix for 221 splicing factors with 57 splicing factors grouped into a single cluster. (*C*) Heatmap showing the expression of 43 differentially expressed splicing factors in TNBC and non-TNBC samples. (*D*) Protein–protein interaction network of the 57 splicing factors using Cytoscape_3.2.0. (*E*) Positional distribution of the TDP43-binding motif (the sequence shown above the graphs) within TNBC-specific AS events using web server rMAPS. Motif enrichment scores (*Left*, full lines) and $-\log_{10}$ (*P* value) (*Right*, dotted lines) were plotted against the genomic positions of TNBC-specific AS events. Red and blue traces show up-regulated and down-regulated splicing events in TNBC samples, respectively. Significant differences were determined by comparison with the background set (black trace). *P* values were calculated by Wilcoxon's rank sum test. Arrows indicate peaks of enrichment for the TDP43-binding motif in TNBC-specific AS events. (*F*) Quantification of the percentage of positively stained TDP43 protein nuclei in grouped samples. (*G*) Protein expression levels of TDP43 in various breast cancer cell lines and immortalized mammary epithelial cell lines. (*H*) Kaplan–Meier plots showing survival analysis of breast cancer patients with luminal A, luminal B, and HER2 subtypes, which were stratified by TDP43 expression level. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by one-way ANOVA.

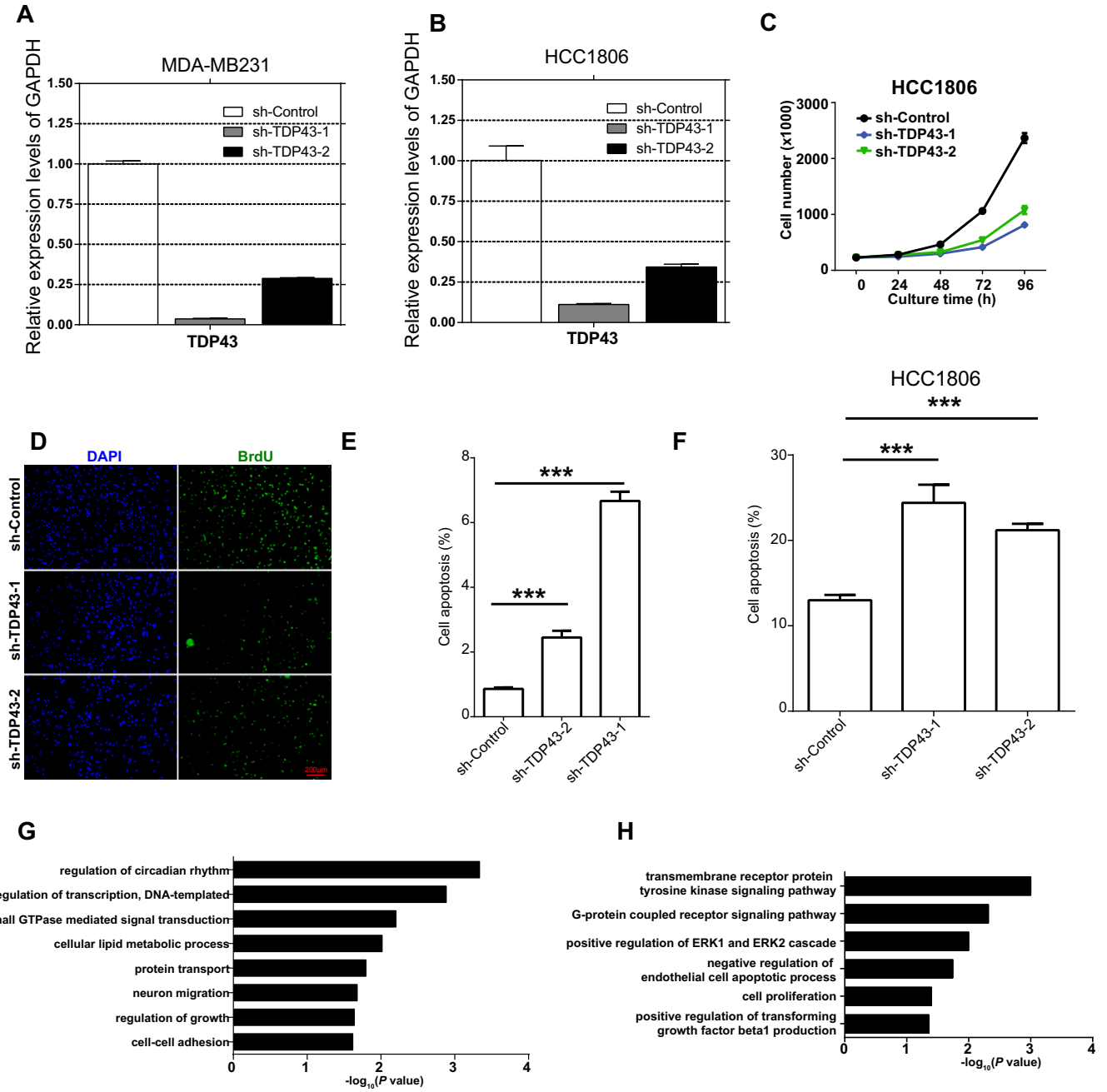


Fig. S2. Related to Fig. 2. Knockdown of TDP43 inhibits breast cancer progression. (*A* and *B*) Knockdown efficiencies determined by qRT-PCR analysis in MDA-MB231 (*A*) and HCC1806 (*B*) cells. (*C*) HCC1806 cell growth is inhibited upon knockdown of TDP43 as shown in the cell-growth curve. (*D*) BrdU staining upon TDP43 knockdown by two independent shRNAs. (*E* and *F*) Knockdown of TDP43-induced cell apoptosis by annexin-V/PI staining in MDA-MB231 (*E*) and HCC1806 (*F*) cells. (*G*) GO analyses of sh-TDP43–regulated AS targets. (*H*) GO analyses of sh-TDP43–regulated gene expression. ***$P < 0.001$ by one-way ANOVA.

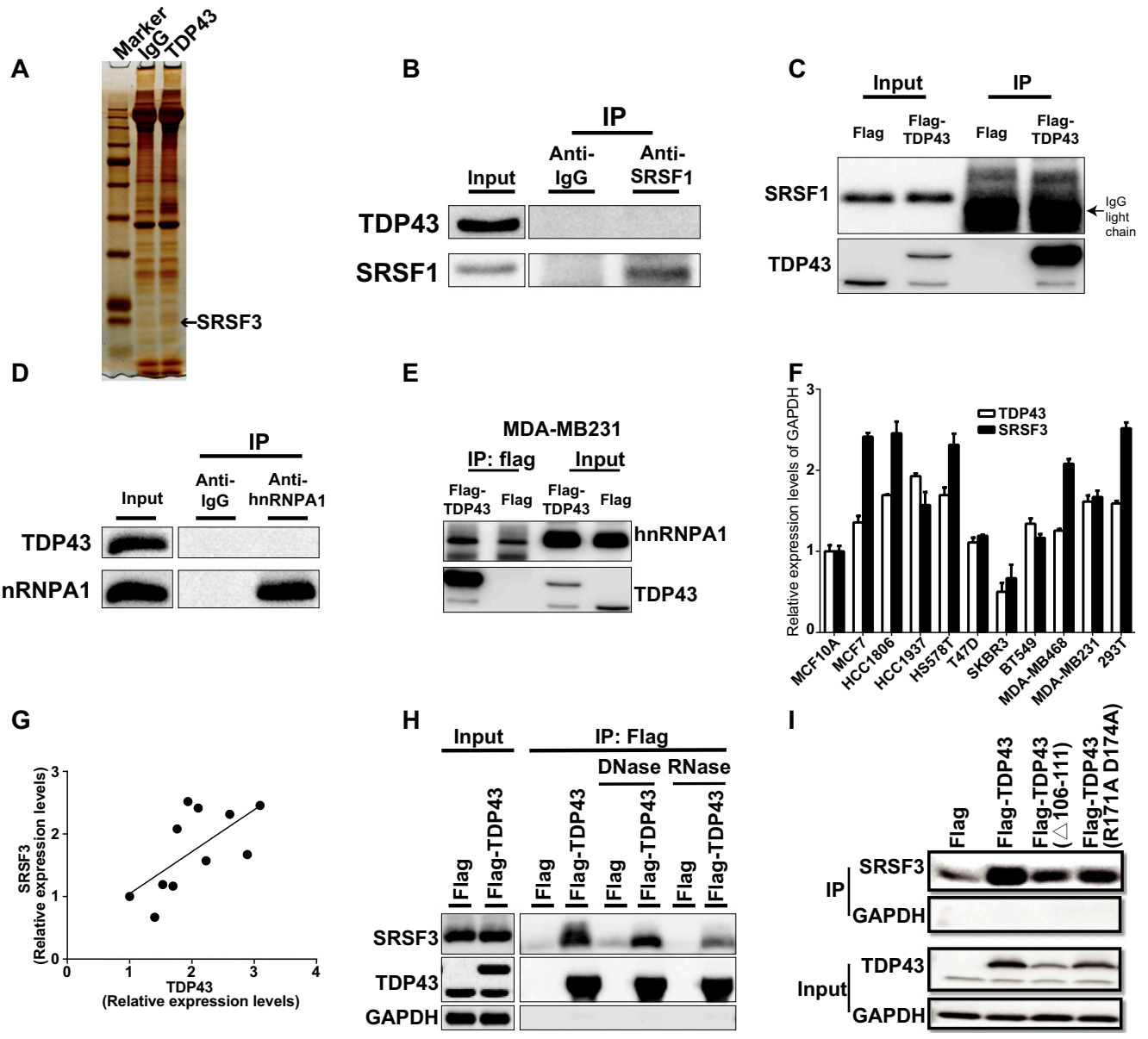


Fig. S3. Related to Fig. 5. (*A*) Immunopurification and MS analysis in MDA-MB231 cells. TDP43-associated proteins were pulled down with an anti-TDP43 antibody, separated with SDS/PAGE, and visualized with silver staining. (*B–D*) Endogenous immunopurification with antibodies against the indicated proteins in MDA-MB231 cells was followed with immunoblotting using antibodies against the indicated proteins. (*E*) Flag-TDP43 pulled down with hnRNPA1. (*F*) The expression of TDP43 or SRSF3 mRNAs was analyzed by qRT-PCR in breast cancer cell lines. (*G*) The relative level of SRSF3 was plotted against that of TDP43. (*H*) Cellular extracts from 293T cells stably expressing FLAG (control) or Flag-TDP43 were immunopurified with anti-Flag affinity beads in the presence of RNase or DNase. Immunocomplexes were then immunoblotted using antibodies against the indicated proteins. (*I*) 293T cells stably expressing FLAG (control), Flag-TDP43, the Flag-TDP43 Δ106–111 deletion mutation, or the Flag-TDP43 R171A D174A mutation were immunopurified with anti-Flag affinity beads. Indicated protein levels were detected by Western blotting.

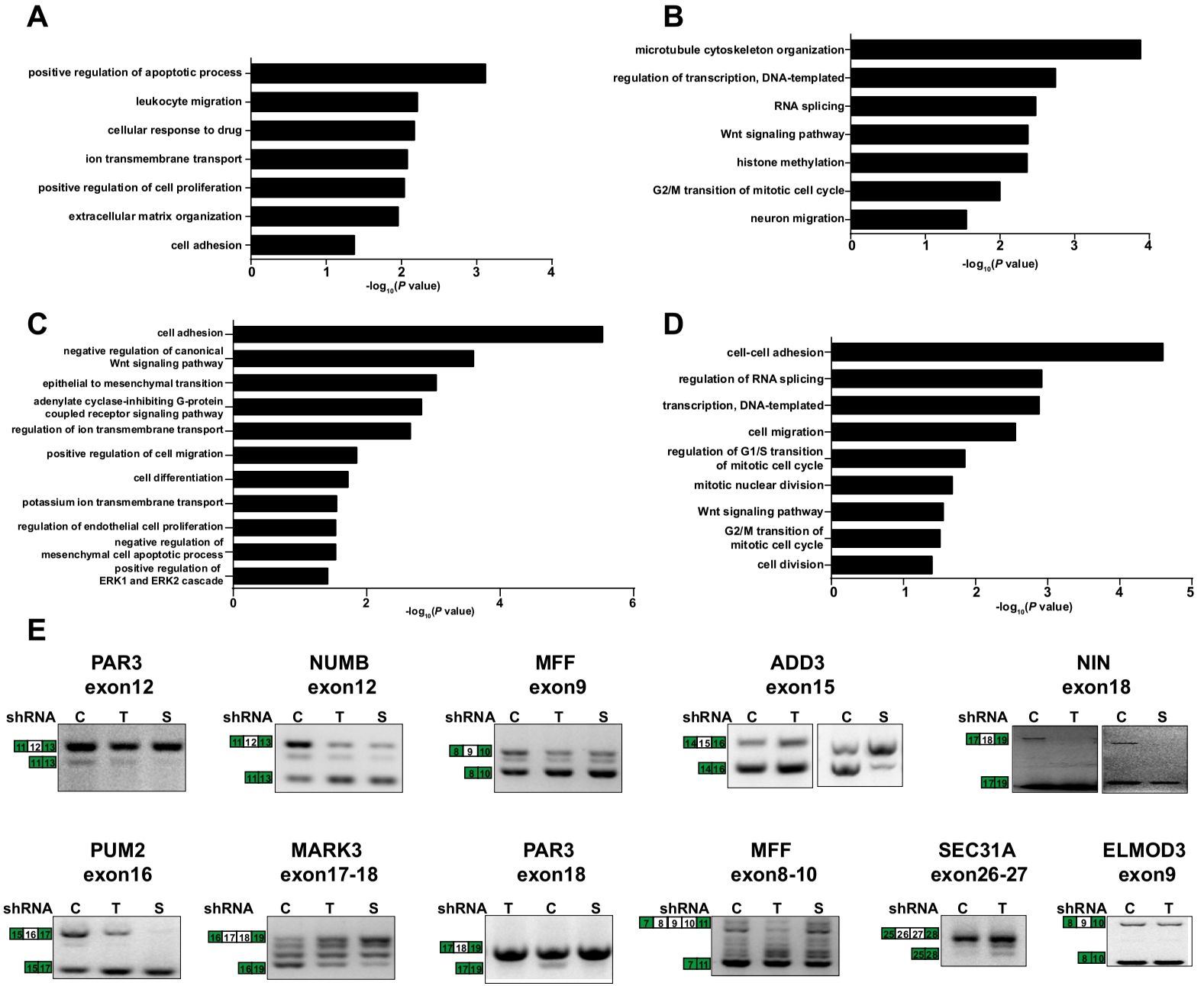


Fig. S4. Related to Fig. 6. (*A–D*) GO analysis of differentially expressed genes (*A* and *C*) and AS targets (*B* and *D*) upon SRSF3 knockdown (*A* and *B*) and TDP43/SRSF3 knockdown (*C* and *D*), respectively. (*E*) RT-PCR validation of TDP43- and SRSF3-downstream candidate splicing events. C, sh-Control; S, sh-SRSF3, T, sh-TDP43.

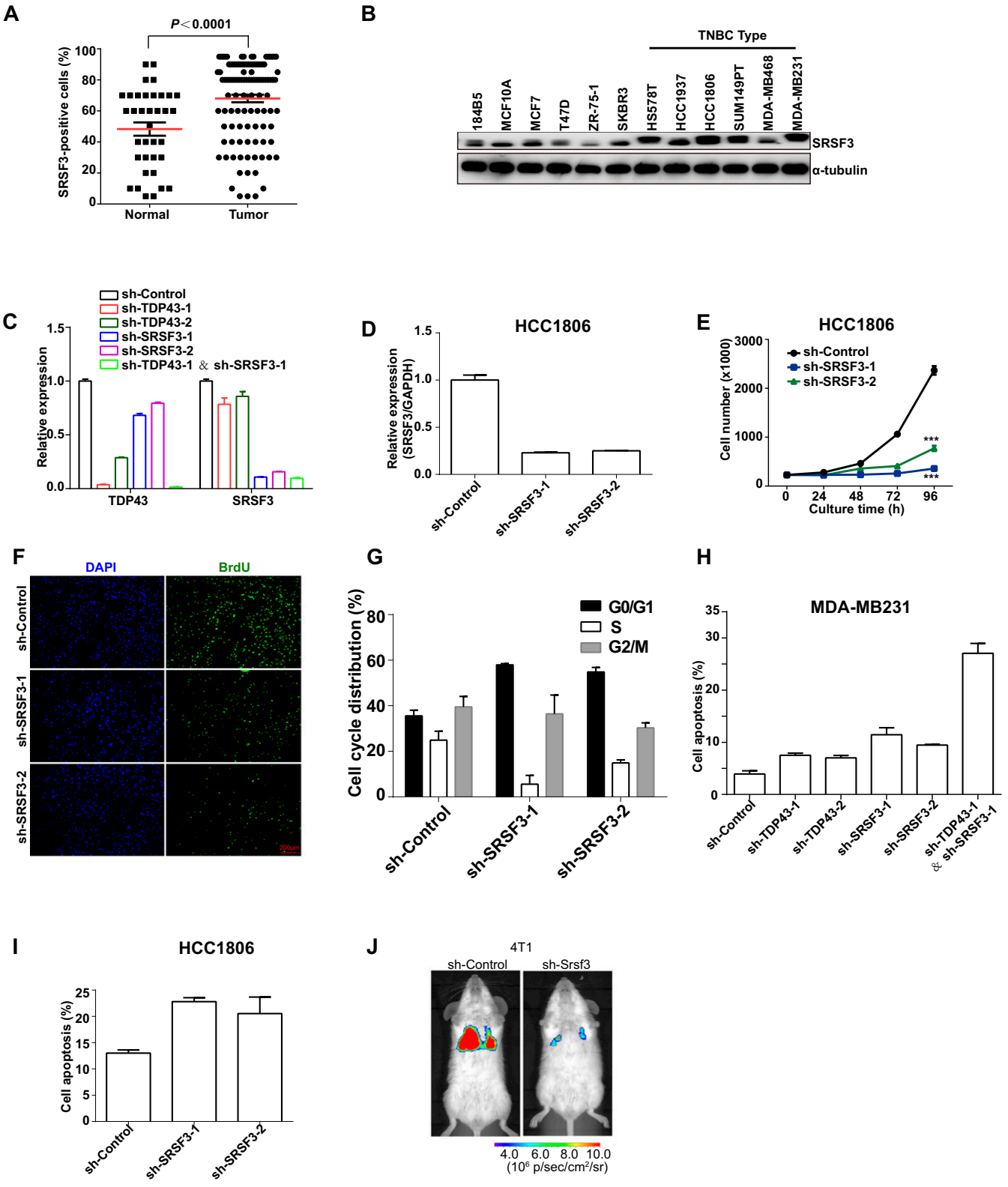


Fig. S5. Related to Fig. 7. (*A*) Immunohistochemical staining of SRSF3 in tissue chips with normal breast tissue and breast carcinomas. (*B*) Protein expression levels of SRSF3 in various breast cancer cell lines and immortalized mammary epithelial cells lines. (*C* and *D*) Knockdown efficiencies were determined by qRT-PCR analysis in MDA-MB231 (*C*) and HCC1806 (*D*) cells. (*E*) HCC1806 cell growth was inhibited upon knockdown of SRSF3. (*F*) BrdU staining upon SRSF3 knockdown by two independent shRNAs. (*G*) Cell-cycle distribution was analyzed upon knockdown of TDP43. (*H* and *I*) Cell apoptosis increased upon TDP43, SRSF3, or TDP43 and SRSF3 knockdown by shRNAs by annexin-V/PI in MDA-MB231 (*H*) and HCC1806 (*I*) cells. (*J*) Fluorescence density of metastasis ability following tail-veil injection of 4T1 cells expressing luciferase ($n = 9$). ***$P < 0.001$ by one-way ANOVA.

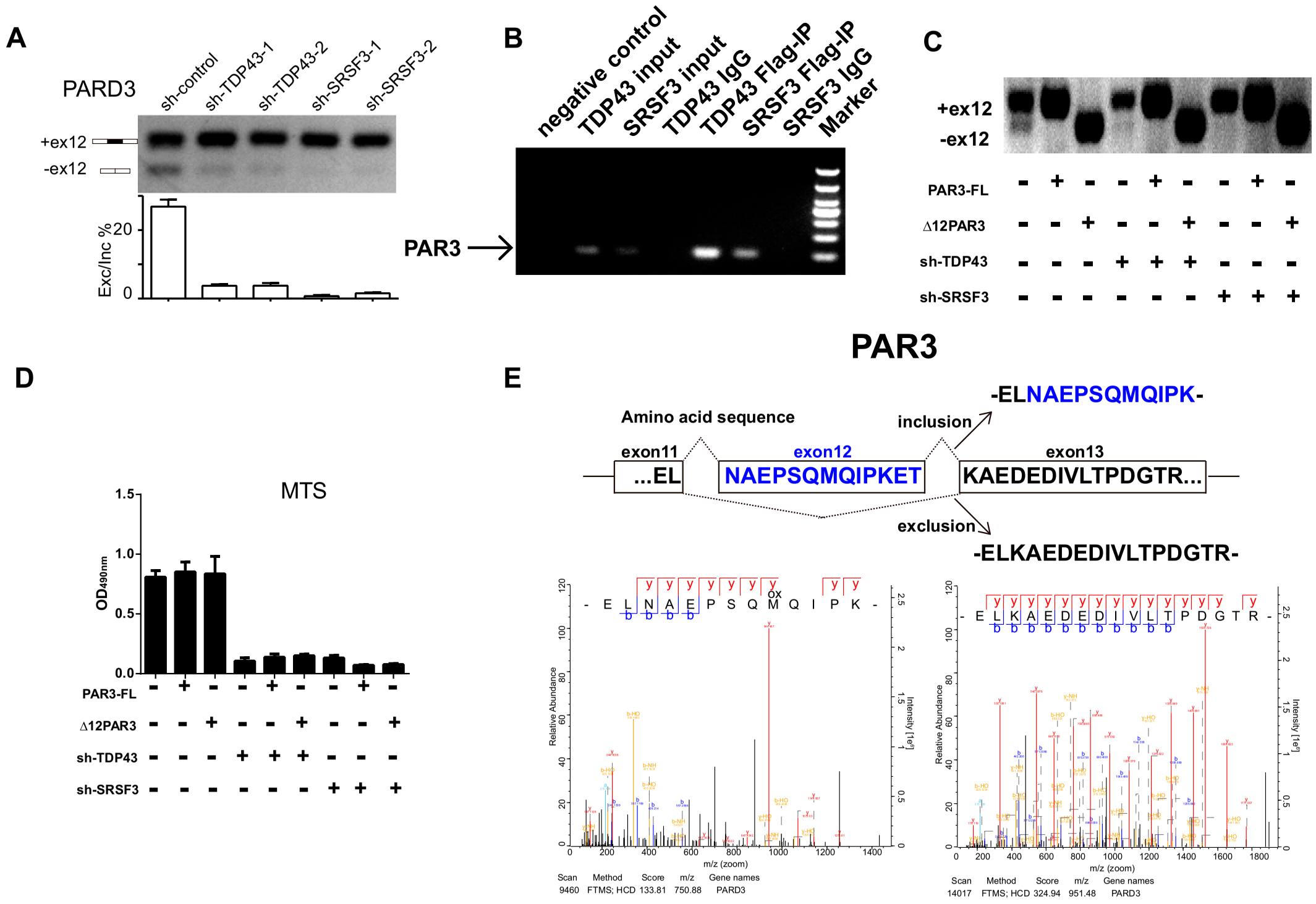


Fig. S6. Related to Fig. 8. (*A*) RT-PCR analysis of changes in *PAR3* exon 12 skipping with knockdown of TDP43 or SRSF3 using primers e and f (locations shown in Fig. 8*A*). (*B*) RIP RT-PCR results. Binding of PAR3 pre-mRNAs with TDP43 or SRSF3 was detected by the RIP assay in cells exogenously expressing FLAG-TDP43, FLAG-SRSF3, or vector control. (*C*) RT-PCR analysis of splicing of *PAR3* exon12 in MDA-MB231 cells expressing the indicated constructs. (*D*) The MTS assay determined the viability of MDA-MB231 cells expressing the indicated constructs. (*E*, *Upper*) Schematic representation of different amino acid sequences based on PAR3 exon 12 inclusion and exclusion. (*Lower*) MS peaks of two peptides, ELNAEPSQMQIPK and ELKAEDEDIVLTPDGTR, which represent PAR3 exon 12 inclusion and exclusion, respectively.

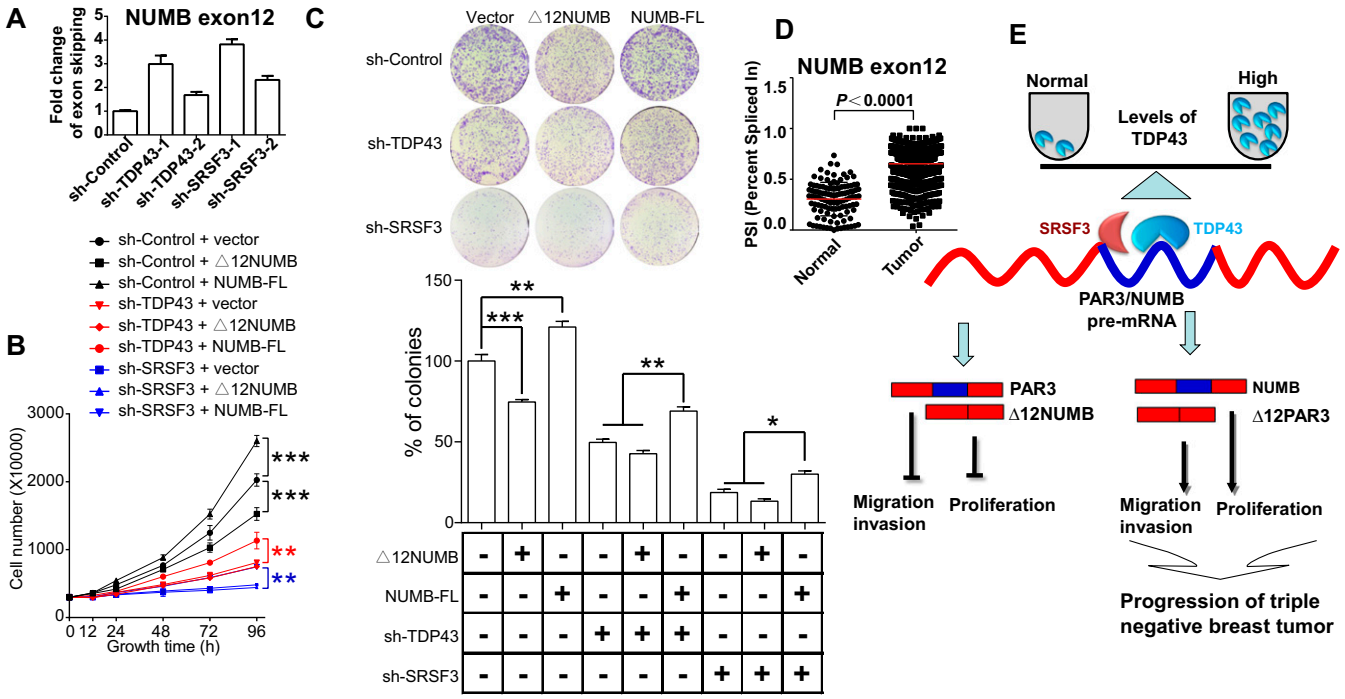


Fig. S7. Antiproliferation effects of reduced TDP43 and SRSF3 are mediated by control of *NUMB* exon 12 splicing. (*A*) qPCR analysis of changes in *NUMB* exon 12 skipping with knockdown of TDP43 or SRSF3. Data shown represent three independent experiments. (*B*) Cell-growth curve showing cell proliferation in MDA-MB231 cells. Two *NUMB* isoforms were transduced into MDA-MB231 cells stably expressing sh-TDP43 or sh-SRSF3. Data shown represent three independent experiments. (*C*) MDA-MB231 cells expressing the indicated vectors were analyzed by the colony-formation assay. (*D*) PSIs of *NUMB* exon 12 in normal ($n = 113$) and breast cancer ($n = 1,085$) tissue in the TCGA. (*E*) Working model of TDP43 in TNBC progression. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by one-way ANOVA.

Table S1. Primers and shRNA sequences used in this paper

Primer names	Sequence
shRNA	
sh-homo-TDP43-1-F	CCGGGCTCTAATTCTGGTGCAGCAACTCGAGTTGCTGCACCAGAATTAGAGCTTTTTG
sh-homo-TDP43-1-R	AATTCAAAAAGCTCTAATTCTGGTGCAGCAACTCGAGTTGCTGCACCAGAATTAGAGC
sh-homo-TDP43-2-F	CCGGCCTAATTCTAAGCAAAGCCAACTCGAGTTGGCTTTGCTTAGAATTAGGTTTTTG
sh-homo-TDP43-2-R	AATTCAAAAACCTAATTCTAAGCAAAGCCAACTCGAGTTGGCTTTGCTTAGAATTAGG
sh-homo-SRSF3-1-F	CCGGCAGTGACACAAAGGTGTAATTCTCGAGAATTACACCTTTGTGTCACTGTTTTTG
sh-homo-SRSF3-1-R	AATTCAAAAACAGTGACACAAAGGTGTAATTCTCGAGAATTACACCTTTGTGTCACTG
sh-homo-SRSF3-2-F	CCGGGCTAGATGGAAGAACACTATGCTCGAGCATAGTGTTCTTCCATCTAGCTTTTTG
sh-homo-SRSF3-2-R	AATTCAAAAAGCTAGATGGAAGAACACTATGCTCGAGCATAGTGTTCTTCCATCTAGC
sh-mus-srsf3-1f	CCGGATCATAAGAGGCACGTGATATCTCGAGATATCACGTGCCTCTTATGATTTTTTG
sh-mus-srsf3-1r	AATTCAAAAAATCATAAGAGGCACGTGATATCTCGAGATATCACGTGCCTCTTATGAT
sh-mus-tdp43-1f	CCGGGGAGAGGATTTGATCATTAAACTCGAGTTTAATGATCAAATCCTCTCCTTTTTG
sh-mus-tdp43-1r	AATTCAAAAAGGAGAGGATTTGATCATTAAACTCGAGTTTAATGATCAAATCCTCTCC
qRT-PCR	
homo-TDP43-F	CCTTGCGTTCATAGCGTTGATAC
homo-TDP43-R	TGCCATAGGAATACTGTCTACATGC
homo-SRSF3-1F	TGGCAACAAGACGGAATTGGA
homo-SRSF3-1R	CAAAGCCGGGTGGGTTTCTA
mus-Tdp43-F	GAGTGCCTCTGTGCATTTGA
mus-Tdp43-R	TTCCCATTGATATGCTCTGCT
mus-Tdp43-F	GACCACTCAGAAGTGTGTGGG
mus-Tdp43-R	TCCTCAAATTCGACGAAAGCAAA
ABI1-F	AAACCGCCAAGTCCTCCCAT
ABI1-R	TGCCTGGTCATTGTGCCATAC
MBNL1-F	GCCCAATACCAGGTCAACCA
MBNL1-R	TGTTAAAGACTGCGGTGGCA
SYNGAP1-1F	AACCCCAACATCCAAAGGCA
SYNGAP1-1R	GGTTCGAAACACTGCTGCTG
MBNL2-1F	AGGCCAAAATCAAAGCTGCG
MBNL2-1R	GTGAGAGCCTGCTGGTAGTG
SYNGAP1-MUT-1R	AGTCCTTCATGGCTCGAGGCGTTTACAGGTGAGGGGCTC
SYNGAP1-MUT-1F	GAGCCCCTCACCTGTAAACGCCTCGAGCCATGAAGGACT
SYNGAP1-MUT-2F	CTTCTGCTAACGAACCCCCTTCCCTTCTG
SYNGAP1-MUT-2R	AGGGGGTTCGTTAGCAGAAGGTAAGGTCAG
P-SYNGAP1-1F	GACGAGCTGTACAAGACGCGTGAAGCCCTCCTGAAGC
P-SYNGAP1-1R	TCCAGAGGTTGATTGTCGACCTAAGGGTTCTGGAAGGAG
P-SYNGAP1-2F	GGCATGGACGAGCTGTAC
P-SYNGAP1-2R	GGCTTTTCCTTGGTTGGGG
P-SYNGAP1-3F	TAACAGCCTCACCCTTCCAG
P-SYNGAP1-3R	TGGACAAATGAAAGGAGAGGC
P-SYNGAP1-4F	TGACACCATACCCATCCCAC
P-SYNGAP1-4R	GAGGGTTCTGGAAGGAGAGG
PARD3-1F	TGGATTTAGGTAGTTCTCCCAGC
PARD3-1R	CCGTGGACGATGGAAAGGAA
PARD3-2F	GCAGAGATGTGGGTCCTTCC
PARD3-2R	TGTCGATGGCAGCTCTGAAG
PARD3-3F (primer e)	TTTCGCTGTTGAGAAGCACC
PARD3-3R (primer f)	CTGCGTGGTTCTCTTTTGACC
PARD3-4F	GGGGACGGCCACATGAAAG
PARD3-4R	TTCCAAGCGATGCACCTGTAT
NUMB forward	CTCCCTGTGCTCACAGATCA
NUMB reverse	CGGACGCTCTTAGACACCTC
NUMBex10-11F	ACCTGAGGACCCCTTCTCAT
NUMBex10-11R	TCAGTGCCATTAGCTTGGAA
NUMBex11-12F	AATTGCAGCCACATGTTCG
NUMBex11-12R	GTCGGCCTCAGAGGGAGTA
NUMBEx10-12F	GAAGTAGAAGGGGAGGCAGA
NUMBEx10-12R	CCCACTCGGTCCCTTGGAA
NUMBEx10F	TAGAAGGGGAGGCAGAGAGC
NUMBEx10R	TGTCACTGGTTTGGTCATCG
PUM2-F	TGCTATAGGCTCAGCCCTCA
PUM2-R	CTGCGGCCAGAAGGCATAATA
FNIP1-F	GCTCATGCTCAGCAAAGTGTTT

Table S1. Cont.

Primer names	Sequence
FNIP1-R	GGTTAGGGCCACAGCTTTCA
SYNE1-F	TTCGGGCGTGTGGAAAGATA
SYNE1-R	GTGGGAGTTTTGCTGCGAAT
ABI1-2-F	TCTGTCCCTCCTCCTTCTGGA
ABI1-2-R	GTGGAGTCAACTGAGGCATAGG
PDGFA-F	GGCCAAGGTGGAATACGTCA
PDGFA-R	GGAGGAGAACAAAGACCGCA
CLSTN1-F	CGAGCCCTTCTCTGTGACTG
CLSTN1-R	GTTTCCCGGAACGGAGAGTT
PFKM-F	TGTCCACCAGATGACGACTG
PFKM-R	CCACTCTTAGATACCGGGGG
ST3GAL1-F	CTCCCCGCTATACCAACACC
ST3GAL1-R	CATCGGTAGGTGTCGTCCTC
GAPDH-F	AGCCACATCGCTCAGACAC
GAPDH-R	GCCCAATACGACCAAATCC
homo-β-actin F	GAGCACAGAGCCTCGCCTTT
homo-β-actin R	ATCCTTCTGACCCATGCCCA
PARD3-5F	TGATCGGATACAGCGTCTGAGG
PARD3-5R	AGAGCTGTACTGGCGCTGG
PARD3-6F	TCAAGCCGTTTTAGCCCTGA
PARD3-6R	GGTGTGCGCTATGAGGTAGT
MFF-1F	CGCTGACCTGGAACAAGGAT
MFF-1R	ATATGGGGAGGACAGACCCG
MFF-2F	TGTTCGCCAAAATGGACAGC
MFF-2R	AGGTCCTTCCTCCAGACTTTC
ADD3-F	ACCACAATCTCAGTTGCTTGC
ADD3-R	TCAGGTGACAGGACTTCTTCG
NIN-F	AGAGCTCTTGGAAAAGCACCA
NIN-R	TGCAGTTTTTCCTGGTTTTGAGA
MARK3-F	AGTATCAGTAGTGCAGCCACC
MARK3-R	GGGATCGAGGCTTTGCTTCT
SEC31A-F	CCAGGTCATATGCACACCCA
SEC31A-R	TGTTCAAAGCTGGAGGGTCA
ELMOD3-F	GACTCGGGCTTTGTCCTCAA
ELMOD3-R	AGATCAGCCATACGCCTTCG
P-tdp43-1f	GACTTCTAGAGCCACCATGTCTGAATATATTCGGG
P-tdp43-1r	GACTGGATCCCTACATTCCCCAGCCAGA
P-tdp43-2f	GACTAAGCTTGCCACCATGTCTGAATATATTCGGGT
P-tdp43-2r	GACTGGGCCCCCCATTCCCCAGCCAGAAGACT
P-tdp43-3f	GACTGAGCTCATGTCTGAATATATTCGGGT
P-tdp43-3r	GACTCTCGAGCTACATTCCCCAGCCAGA
P-tdp43-4r	GACTGGATCCTTATGATCTTCCGGAACGCGTGAGCT
P-SRSF3-1F	AAGACTCATCGGAGCGTGTG
P-SRSF3-1R	CAAGAGACAAACAGATTTGAACAGC
P-SRSF3-2F	TCGGAGCGTGTGGATTTGAG
P-SRSF3-2R	AAATGCACCAAGCTTATGAAGTCTC
P-SRSF3-3F	GGTCAAATGAAAGGAAAGCTAGCGACTACAAAGACCATG
P-SRSF3-3R	CATGGTCTTTGTAGTCGCTAGCTTTCCTTTCATTTGACC
P-SRSF3-4R	TGGGTTGCGGCCGCAGTCGACCTACTTGTCATCGTCATC
P-SRSF3–SR-R1	GGGAGGGAGAGGGGCGGATCCCTATTTTTCACCATTCGA
P-SRSF3–RRM-F2	ATTATAAAGATCATGATATCGATTACAAGGATGACGATG
P-SRSF3–RRM-F1	ACAAGGATGACGATGACAAGGGCAGAAGTAGAAATCGTG
P-TDP43–RRM1-F2	AGCAGTCCAGAAAACATCCTGCAAACTTCCTAATTCTAA
P-TDP43–RRM1-R1	TTAGAATTAGGAAGTTTGCAGGATGTTTTCTGGACTGCT
P-TDP43–RRM2-F2	TTTGAGAAGCAGAAAAGTGTCCAATGCCGAACCTAAGCA
P-TDP43–RRM2-R1	TGCTTAGGTTCGGCATTGGACACTTTTCTGCTTCTCAAA
P-TDP43–C term-R	GCTGGGTTGCGGCCGCAGTCGACCTAACTTCTTTCTAAC
P-TDP43–c term-F1	GGGGGAGGGAGAGGGGCGGATCCCTAACTTCTTTCTAAC
P-pard3-F1	GAAGACACCGACTCTAGAGCCACCATGAAAGTGACCGTG
P-pard3-F2	AGAACGAAGAATTTC
P-pard3-R1	CTCAATCCCACTGTAGAG
P-pard3-R2	GGGGGGAGGGAGAGGGGCGGATCCTCAGGAATAGAAGGG

Table S1. Cont.

Primer names	Sequence
P-pard3-F3	CTTCCACCCAAGGGAACTGAAAGCAGAAGATGAGGATAT
P-pard3-R3	ATATCCTCATCTTCTGCTTTCAGTTCCCTTGGGTGGAAG
P-pard3-F4	AAAAAGCATGGATTTAGGTAGTTCTCCCAGCAGAGATGT
P-pard3-R4	ACATCTCTGCTGGGAGAACTACCTAAATCCATGCTTTTT
P-NUMB743-1F	ACCTCTCCTACTTCTGATGCCACGACCTCTCTGGA
P-NUMB743-1R	TCCAGAGAGGTCGTGGCATCAGAAGTAGGAGAGGT
P-NUMB743-2F	gactACCGGTGCCACCATGAACAAATTACGGCAAAG
P-NUMB743-2R	gactACGCGTTTAAAGTTCAATTTCAAACGTC
P-NUMB743-3F	CACAATCTCCTACCTTCCAAGCTAATGGCACTGACTCAG
P-NUMB743-3R	CTGAGTCAGTGCCATTAGCTTGGAAGGTAGGAGATTGTG
P-NUMB743-4F	GTGTAGATGATGGCAGGTTGGCCTCAGCAGACAGGCATA
P-NUMB743-4R	TATGCCTGTCTGCTGAGGCCAACCTGCCATCATCTACAC

Table S2. Immunohistochemistry of paired breast cancer tissue arrays

	Breast cancer		Corresponding para-carcinoma tissue	
Sample no.	Stain degree	Positive cells, %	Stain degree	Positive cells, %
1	+++	90	+	5
2	+	40	+	3
3	+	70	+	10
4	+	70	+	30
5	++	70	+	5
6	+	10	+	3
7	+++	80	+/++	20
8	+++	80	+	10
9	+	60	+	10
10	+	30	+	20
11	+	70	+/++	70
12	+++	80	+/++	60
13	+++	95	+/++	50
14	+	40	+	3
15	+	60	+	1
16	+	70	0/+	1
17	++	85	+	30
18	+++	80	+	30
19	++	20	+	5
20	+/++	70	+/++	40
21	+++	70	+/++	15
22	+++	50	+	20
23	++	85	+	5
24	+	30	+	60
25	+	90	+	70
26	++	75	+	20
27	+/+ +	50	+	40
28	++	80	++/+++	80
29	++	60	++	60
30	++	40	+	3
31	+	60	+	5
32	++	85	+/++	30
33	++	70	+/++	30
34	++	70	+	10
35	+	5	0/+	1
36	++	80	+	10
37	+++	70	+/++	30
38	+++	80	++	70
39	+	60	+	40
40	+++	80	+/++	60
41	++	60	+	5
42	++	80	+	10
43	+++	85	+	10

Hyperplasia of mammary tissue is in bold red. Blue bold indicates higher expression in para-carcinoma tissue than in cancer tissue. +, no or faint staining; ++, moderate staining; +++, strong staining

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (XLSX)
Dataset S3 (XLSX)
Dataset S4 (XLSX)