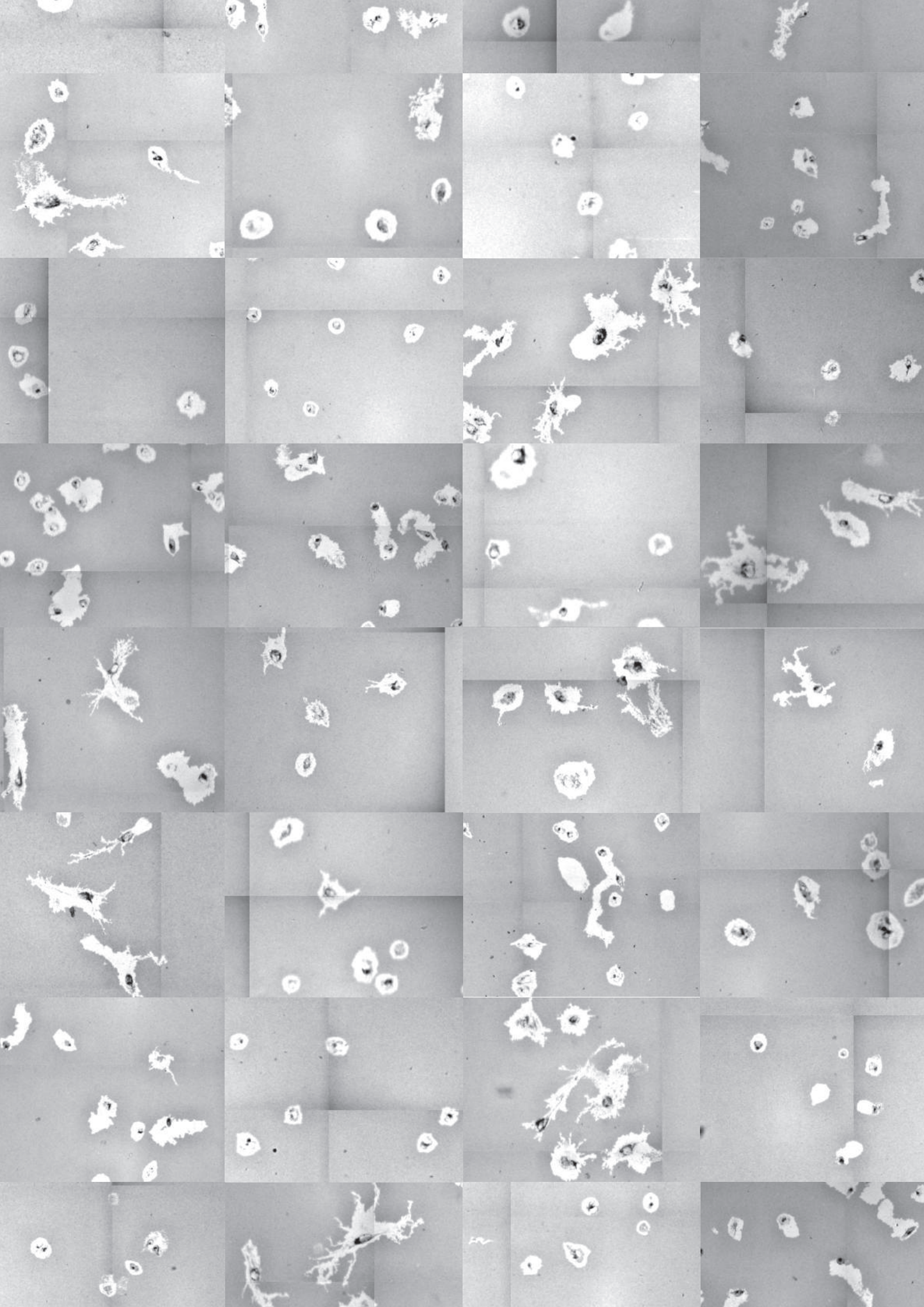
RNA splicing in breast cancer progression

Koedoot, E.



Chapter 5

Unraveling the spliceosomal control of breast cancer cell motility behavior

Esmee Koedoot, Marcel Smid, Peter Stoilov, Román González-Prieto, Alfred C.O. Vertegaal, John W.M. Martens, Sylvia E. Le Dévédec, Bob van de Water

Ready for submission

Highlights

- A targeted RNAi screen of 244 splicing factors uncovered ten splicing factors that act as common migratory modulators with limited effects on cell proliferation and survival
- Splicing factors SRSF7 and SMNDC1 are critical for RNA processing and regulate the expression of genes involved in cell-matrix interactions and cellular adhesions
- SRSF7 functions in a large complex together with previously identified splicing factors PRPF4B and BUD31, making this complex a candidate for drug development

◄ IN THE PICTURE

Migratory tracks of triple-negative breast cancer cells. Cells were seeded on a bead-coated surface on which they leave tracks when migrating. These migratory tracks were used as a measurement for cell motility.

◄ IN BEELD

Sporen van beweging van triple-negatieve borstkankercellen. Cellen werden uitgezaaid op een met korrels bedekt oppervlak waarop sporen achterblijven tijdens het bewegen. Deze sporen werden gebruikt als een meting voor de beweeglijkheid van cellen.

There is increasing evidence for a role of alternative splicing in the progression and metastasis of breast cancer. The role of the splicing machinery in the tumor cell migration of aggressive triple-negative breast cancer (TNBC) remains elusive. Here we systematically evaluated the role of RNA splicing in TNBC cell motility behavior. A targeted RNAi screen of all 244 major splicing factors in two highly motile TNBC cancer cell lines uncovered ten splicing factors, including CWC15, PHF5A, SF3A3, SMNDC1, SNRPA1, SNRPB2, SNRPD2, SNRPG, SRSF7 and U2AF1, that act as common migratory modulators in TNBC with limited effects on cell proliferation and survival. All of these factors were associated with aggressive breast cancer subtypes and/or metastasis-free survival in breast cancer patients. Two of the most prominent factors, SRSF7 and SMNDC1, are critical for RNA processing in association with regulating the expression of gene networks involved in cell-matrix adhesion signaling. SRSF7 resided in a large protein-protein interaction splicing complex together with PRPF4B and BUD31, and was affecting similar exon inclusion events, making this complex a candidate for drug development.

Introduction

Breast cancer is the most prevalent cancer in women and can be classified based on hormone receptor expression. In contrast to the luminal subtypes, the triple-negative breast cancer (TNBC) subtype lacks overexpression of the three hormone receptors (estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2)[39]. Although this subtype only accounts for approximately 15% of human breast cancer (BC) cases[14], it is disproportionally responsible for BC-related deaths having an overall 5 year relapse of 30-40% due to development of metastases[15,72,73,316–318]. Therefore, unravelling the underlying mechanisms that drive TNBC metastasis formation is essential to inhibit progression and reduce TNBC mortality.

Many TNBC cell lines demonstrate a mesenchymal phenotype associated with increased migratory behaviour that typically coincides with metastatic progression[24,39–42,319]. Cell migration is critical for many steps of the metastatic cascade, including local invasion, intravasation, extravasation, and colonization of secondary sites[27] and is controlled by modulators of the actin cytoskeleton as well as cell adhesion components[59,86]. On top, the expression, activity and post-translational modifications of these complex machineries are regulated by various signalling pathways and transcriptional programs that all together control the cell migratory behaviour[87,320,321]. In the past, we have performed several RNAi-screens to systematically unravel the global cell signalling landscape and identify genes involved in the regulation of TNBC cell migration[66] (Koedoot et al, unpublished). Interestingly, in these independent screening efforts, various signalling components that control the migratory and metastatic behaviour of TNBC cells are directly related to RNA splicing programs (e.g. SRPK1, PRPF4B and BUD31)[66] (Koedoot et al, unpublished). So far a systematic evaluation of the role of splicing factors in breast cancer cell motility and metastasis formation is lacking.

RNA splicing catalyses the removal of introns from the pre-mRNA is essential for mRNA maturation. Additionally, it provides an extra layer of gene control by selective inclusion of exons or introns (alternative splicing) leading to different transcripts derived from the same gene. In this way, alternative splicing already demonstrated to affect multiple pathways pivotal in cancer development and progression, such as apoptosis, metabolism, sustained proliferation, oncogenes, tumor suppressor genes and angiogenesis[67,133] (Koedoot et al, unpublished). The splicing reaction is catalysed by a highly dynamic complex consisting of five small nuclear RNAs and approximately 250 associated proteins or splicing factors[69]; their role in TNBC cell migratory behaviour as well as the underlying transcriptional modulation remains unclear.

Here, we performed an imaging based RNAi-screen targeting 244 splicing factors to unravel the spliceosomal control of TNBC migratory behaviour. We identified that the splicing factors SMNDC1 and SRSF7 modulate TNBC cell motility, alternative splicing and particularly genes related to migratory pathways.

Methods

Cell culture

MDA-MB-231 (ATCC-HTB-26) and Hs578T (ATTC-HTB-126) cell lines were cultured in RPMI-1640 medium (Gibco, Life Technologies, Carlsbad, CA, USA) supplemented with 10% fetal bovine serum (FBS), 100 U/ml penicillin (Invitrogen) and 100 µg/ml streptomycin (Invitrogen) (referred to as complete medium). Cells were maintained in a humidified incubator containing 5% CO_2.

Transient siRNA transfection

The spliceosome siRNA library containing 244 splicing factors was designed based on the study of Hegele et al[69] and purchased from Dharmacon. Knockdown was acquired by reversed transfection of 50nM siRNA by using INTERFERin (Polyplus, Illkirch) according to the manufacturer's guidelines. Medium was replaced by complete medium 18 hours after transfection. Cells were used for different assays 65 to 88 hours after transfection.

Antibodies

Mouse anti-Tubulin (T-9026, Sigma-Aldrich), mouse anti-FAK (610087, BD Biosciences), mouse anti-N-cadherin (610920, BD Biosciences), rabbit anti-p-FAKY397 (446-24ZG, Thermo Fisher), goat anti-MMP9 (sc-6840, Santa Cruz), mouse anti-fibronection (610077, BD Biosciences), mouse anti-pCas(p130) (P-27820, Transduction), mouse anti-Cas9 (14697P, Cell Signaling), rhodamine phalloidin (R415, Molecular Probes), mouse anti-GFP (11814460001, Roche), mouse anti-SRSF7 (sc-390126, Santa Cruz), mouse anti-SNRPG (sc-398741, Santa Cruz), rabbit anti-SNRPA1 (17368-1-AP, ProteinTech), rabbit anti-SMNDC1 (NBP1-47302, Novus Biologicals) and mouse anti-Vimentin (ab8069, Abcam) were all commercially purchased. Anti-

mouse and anti-rabbit horseradish peroxidase (HRP) conjugated secondary antibodies were purchased from Jackson ImmunoResearch.

Quantitative cell migration assays.

Phagokinetic track (PKT) assay: Cells were transfected in a 96-wells format (6,000 cells/well for MDA-MB-231 and 5,000 cells for Hs578T). Phagokinetic Track (PKT) assays were performed according the protocol described in Fokkelman et al.[65].

Random cell migration assay: Hs578T and MDA-MB-231 cells were transfected in a 96-wells format as previously described(6,000 cells/well for MDA-MB-231 and 5,000 cells for Hs578T). Glass-bottom 96-well plates (SensoPlate, Greiner Bio-One, Frickenhausen, Germany) were coated with 10µg/µl fibronectin in PBS for 1 hour at 37°C. 65 hours after transfection, knockdown cells were trypsinized, resuspended in complete medium and plated in these fibronectin coated plates. 69 hours after transfection, cells were stained with 100 ng/ml Hoechst 33342 for 1 hour to mark the nuclei, followed by replacing the Hoechst containing medium with complete medium to avoid phototoxicity. 72 hours after transfection, cells were imaged on a Nikon Eclipse Ti microscope, equipped with a 37°C incubation chamber with CO_2 flow, Perfect Focus System, an automated xy-stage and 20x objective (0.75 NA, 1.00WD). Images were captured with a DS-Qi1MC CCD camera every 12 minutes for at least 12 hours at 3 positions per well using NIS software (Nikon, Amsterdam, The Netherlands). Plates were then washed with PBS, fixed using 1% paraformaldehyde and 0.1% TritionX100 in PBS for 15 minutes at room temperature and stored in PBS at 4°C for later immunostaining. Images were converted to tiffs and analyzed by using CellProfiler[90] and R-scripts described in Wink et al[322].

Cell proliferation assay.

High content imaging nuclei counting: Cells were transfected in a 96-wells format as described above (6,000 cells/well for MDA-MB-231 and 5,000 cells for Hs578T). 88 hours after transfection, cells were fixed by adding 30µL 50% trichloroacetic acid directly to 100µl medium per well. Plates were put on the shaker for 1 hour at 4°C, washed five times with distilled water followed by 200 ng/ml Hoechst 33358 staining overnight at 4°C. Cells were imaged on a Nikon Eclipse Ti microscope with a 10x objective.

To cover the whole well surface 8x8 images were captured for each well using a DS-Qi1MC CCD camera and NIS software (Nikon, Amsterdam, The Netherlands). Images were converted to tiffs and nuclei were counted with CellProfiler[90].

Sulforhodamine B (SRB) assay: After nuclei counting, Sulforhodamine B (SRB) colorimetric assays were performed to determine the total amount of protein as a measure for proliferation. Water was removed and plates were air-dried. Cells were stained with 60µl 0.4% SRB in 1% acetic acid for 30 minutes at room temperature, washed 5 times with 1% acetic acid and air-dried in between the washing steps. Protein-bound SRB was dissolved in 200µl 10mM unbuffered Tris (pH>10) per well and absorbance was measured at 530 nm with a FLUOstar OPTIMA plate reader (BMG Labtech, Offenburg, Germany).

Immunofluorescent staining

Cells fixed in 1% paraformaldehyde and 0.1% TritonX100 were washed 3 times 10 minutes with 0.5% Bovine Serum Albumin (BSA) in PBS and stained with 30µl primary antibody per well overnight at 4°C. Cells were washed 3 times 10 minutes with 0.5% BSA in PBS, stained with corresponding secondary antibody and 200 ng/ml Hoechst 33358 for 1 hour at room temperature (30µl/well). Cells were washed once with 0.5% BSA in PBS, twice with PBS for 10 minutes and the plate was stored at 4°C in dark until imaging. Imaging was performed with a Nikon Eclipse Ti microscope with a 20x objective (0.75 NA, 1.00WD) and confocal imaging (lasers: 561, 488 and 408 nm).

Cell death assays

Cells were transfected in a 96-wells format as described previously (6,000 cells/well for MDA-MB-231 and 5,000 cells for Hs578T). 65 hours after transfection, cells were stained with 100 ng/ml Hoechst 33342 for 1 hour to mark the nuclei, followed by replacing the Hoechst containing medium with medium containing propidium iodide (PI) and Annexin V to stain for necrotic and apoptotic cells respectively. 72 hours after transfection, cells were imaged on a Nikon Eclipse Ti microscope with fitted with a 37°C/5%CO_2 incubation chamber, a perfect focus system, a 20x objective (0.75 NA, 1.00WD) and confocal imaging (lasers: 561, 488 and 408 nm). 4 images per well were acquired. After imaging, plates were fixed with 30µl 50% trichloroacetic acid for 1 hour at 4°C, washed 5 times with distilled water and stored for both nuclei counting and SRB assays (previously described). The fraction of PI and Annexin V positive cells was determined by using CellProfiler and R studio.

Next generation RNA sequencing

RNA was isolated 72 hours after knockdown using the RNeasy plus mini kit (Qiagen). DNA libraries were prepared with the TruSeq Stranded mRNA Library Prep Kit and sequenced according to the Illumina TruSeq v3 protocol on an Illumina HiSeq2500 sequencer. 20 million and 100 million 100 base pair paired-end reads were generated for the first (siKP, siSMNDC1, siSNRPA1, siSNRPG and siSRSF7) and second experiment (siKP, siSMNDC1 and siSRSF7), respectively. For both experiment, samples from 3 biological replicates were used. Reads were aligned against the human GRCh38 reference genome using using the HiSat2 aligner (version 2.2.0.4). Gene expression was quantified using the HTseq-count software (version 0.6.1) based on the ENSEMBL gene annotation for GRCH38 (release 84). Count data was normalized and log2 fold changes and adjusted P-values were calculated using the DESeq2 package[92]. DEGs were selected by effect size (log2 fold change bigger than 1 or smaller than -1) and adjusted p-value (smaller than 0.01) and used for over-representation analysis for KEGG pathways using ConsensusPathDB[94].

For the intron retention analysis, RNA-seq reads were aligned to the current human genome (GRCh38) using Hisat2[95]⁹⁵⁹⁵⁹⁵⁹⁵(Kim et al. 2015). Differential intron retention analysis was carried out in R using DexSeq packages[96,97]. Here, the differences of intron inclusion were calculated based the counts from the intron and the two adjacent exons. The sizes of the exons were limited to 100nt immediately adjacent to

the intron to reduce artifacts deriving from alternative promoters, alternative splice sites and alternative poly-adenylation sites. Deferentially retained introns were selected by effect size (relative change in inclusion more than 0.1) and statistical significance of the change (adjusted p-value less than 0.01). Alternative exon inclusion was analyzed using the rMATS package version 3.0.8[98]. Differentially spliced exons were selected by effect size (relative change in inclusion more than 0.1) and statistical significance of the change (FDR < 0.05). RNA sequencing data will be available in Sequence Read Archive upon publication.

RT-qPCR

RNA was collected 72 hours after knockdown using the RNeasy plus mini kit (Qiagen). cDNA was synthesized using the RevertAid H minus first strand cDNA synthesis kit (Thermo Fisher Scientific). RT-qPCR was performed using the SYBR Green PCR master mix (Thermo Fisher Scientific) on a 7500 Fast Real-Time PCR machine (Applied Biosystems/Thermo Fisher Scientific). Used primers are shown in Table 1. Relative gene expression was calculated after correction for GAPDH expression using the 2ΔΔCt method.

Table 1. ***Primers sequences used for RT-qPCR.***

Target	Forward primer	Reverse primer
GAPDH	5'-CTGGTAAAGTGGATATTGTTGCCAT-3'	5'-TGGAATCATATTGGAACATGTAAACC-3'
SMNDC1	5'-GCAACAAACCCATGTCAAAA-3'	5'-CAAAGCTTTCTTCTTTTTATATTCACG-3'
SNRPA1	5'-TCCGCAAGTCAGAGTACTGG-3'	5'-CCGTTTGCCCTTGAACATT-3'
SNRPG	5'-TCACCCTCCCGAGTTGAA-3'	5'-AAATCCCCGCAATATTCCTT-3'
SRSF7	5'-AAAAGGATCGAGGTATTTCCAAT-3'	5'-GCTGCTTCTTGGTCGTTGAA-3'
LAMA5	5'-CTTCGTGCAGCGGCTCCT-3'	5'-GCCTTCCGGGCGGTG-3'
FN1	5'-AACAAACACTAATGTTAATTGCCCA-3'	5'-TCGGGAATCTTCTCTGTCAGC-3'
FAK	5'-TGGGCGGAAAGAAATCCTGC-3'	5'-GGCTTGACACCCTCGTTGTA-3'
U1snRNA	5'-CCATGATCACGAAGGTGGTTT-3'	5'-ATGCAGTCGAGTTTCCCACAT-3'
U2snRNA	5'-TTCTCGGCCTTTTGGCTAAG-3'	5'-CTCCCTGCTCCAAAAATCCA-3'
U4snRNA	5'-GCCAATGAGGTTTATCCGAGG-3'	5'-TCAAAAATTGCCAATGCCG-3'
U6snRNA	5'-CTCGCTTCGGCAGCACA-3'	5'-AACGCTTCACGAATTTGCGT-3'
COL5A1	5'-GACAAGAAGTCCGAAGGGGC-3'	5'-TAGGAGAGCAGTTTCCCACG-3'
CPT1A	5'-GAACAGGTATCTACAGTCGGTGAGG-3'	5'-CTCACGTAATTTGTAGCCCACCAGG-3'
DHRS11	5'-CAACATCGAGGAGCTGGCTG-3'	5'-CTACACCGCTGTGCTGAGAA-3'
MRM1	5'-GAGGGCTCAGGTCTATCCCA-3'	5'-GAAGAAGAATTCCTGCAGCCAC-3'
TRIOBP	5'-TGAGGAGGCAGATGAGCTGG-3'	5'-CATCCTTGGTGTGGATCTGGAAG-3'
TTF2	5'-GTGATACTGCCCCAGCGTAA-3'	5'-CCAAACTCCAGTGCCACTCT-3'
FOS	5'-GGGGCAAGGTGGAACAGTTA-3'	5'-AGTTGGTCTGTCTCCGCTTG-3'
HMOX1	5'-CAGTCAGGCAGAGGGTGATAG-3'	5'-CCTGCAACTCCTCAAAGAGC-3'
NR4A2	5'-AACTGCACTTCGGCAGAGTT-3'	5'-GCCTGTCCAATCTCCTCCTTAG-3'
SAT1	5'-GCTGATCAAGGAGCTGGCTAA-3'	5'-TGCAACCAGGCAGTGGTAAA-3'

PCR gel electrophoresis

RNA samples were collected 72 hours after transfection using the RNeasy plus mini kit (Qiagen), followed by cDNA synthesis using the RevertAid H minus first strand cDNA synthesis

kit (Thermo Fisher Scientific). PCR was performed using the MyTaq Red Mix (Bioline). Used primers are shown in Table 2. Products were loaded on a 2% agarose gel and visualized with a transilluminator. Bands were quantified using ImageJ.

Table 2. ***Primer sequences used for PCR.***

Target	Forward primer	Reverse primer
DMTF1 A3SS	ACCTGAAGAAATTGAGAAGCTCAAGGA	TCAGTCGGCACCGATCTTTGAC
SPAG9 SE	TGGTGTCTTGTAATCGTTTGTGGGT	GCATGTGCCATTGAACAATAGGGT
NCOR2 SE	AGGATGAGGAGATGGAGGCGT	TCCGTGCACTCGCTCTTGA
RABGGTB SE	TCGCATCCTATGGCTCAAAGAAAG	AACAAAGCCAAAGTTGCCACC
CD44 A3SS	ATACCATCTCAGCAGGCTGGG	GTAAGCATCTCAGTGTCCTTTGCAT
PPP4R1 A3SS	ACCTCTCGCTGCTTCAGGA	TCACTTGCAGCATACTTGTCCAA
TMEM218 A5SS	ACAACAGAGTGGGAACTTTCTGTC	CTGGTAGAAGACACCAGGGCAT
CEP83 SE	ACTTCAGCTCCTGCTTGAAGAAC	GTTGTATTTGGGCTCTTAGCAATTCCA
CHFR SE	AGCAAGGCATGACACAAGAATCC	ACGACGAAAAGGACGCAGTCT
DLG1 SE	GTTGAGAAGAAAGAACGAGCCCGA	TCGCTGGCATTAGAAGTTACATGC
PARP6 SE	ACATTGTCAAACTACCTCTCAGCAG	TGCAGTTTGGTGTAGGATGCAT

CRISPR-Cas9 inducible knockdown

To obtain inducible Cas9 cell lines MDA-MB-231 cells were transduced with the lentiviral Edit-R inducible Cas9 plasmid (Dharmacon). Cells were selected using 2 µg/ml blasticidin and grown single cell after which a clone was selected that was fully Cas9 inducible; from now on called MDA-MB-231 ind-Cas9. sgRNAs were obtained from the human Sanger Arrayed Whole Genome Lentiviral CRISPR Library (Sigma-Aldrich) (sgSRSF7 #1: CACCGGATTTGTGGCTCCCGAGTGGT, sgSRSF7 #2: CACCGCAAAGGGACGTCGGGCAGGGT, sgSMNDC1 #1: CACCGAACTTGCAAGCGTCTCAGAGT, sgSMNDC1 #2: CACCGTGTTATGAAGCGGAGATTGGT). MDA-MB-231 inducible Cas9 cells were transduced with sgRNA containing lentivurs followed by puromycin selection. Western blot and live cell imaging assays were performed 72 hours after 1 µg/ml doxycycline treatment.

Generation BAC-GFP reporter cell lines

Human SRSF7 and SMNDC1 BAC clones were selected and GFP-tagged as previously described[323,324] and stably introduced in Hs578T and MDA-MB-231 cells using G-418 for selection. Selected cells were FACS sorted twice for GFP expression.

GFP pulldown and mass spectrometry

Pulldown and mass spectrometry were performed as described previously in Chapter 6 (Koedoot et al, Splicing factors control triple-negative breast cancer cell mitosis through SUN2 interaction and sororin intron retention). Samples were collected from 2x T175 3 days after plating.

Western blotting

Samples were collected 72 hours after transfection in RIPA lysis buffer (1% w/w deoxycholate, 50 mM Tris (pH 7.5), 0.15 M NaCl, 0.1% sodium dodecyl sulfate (SDS), 1% v/v NP-40, 2 mM EDTA, 1% v/v protease inhibitor cocktail (P8340, Sigma-Aldrich)). Proteins were separated by electrophoresis using SDS-PAGE gels, transferred to PVDF membranes (Merck Millipore), blocked in 5% w/v BSA and overnight incubated with the corresponding primary antibody at 4°C. Membranes were incubated with secondary antibody for 1 hour at room temperature, exposed to Pierce ECL western blotting substrate (Thermo Fisher Scientific) and visualized with the Amersham Imager 600 (GE Healthcare). At least 2 biological replicates were performed per experiment. Tubulin was used as a loading control.

Data analysis and statistics

Sample sizes were based on previously published similar experiments. When not indicated, all experiments were performed in biological triplicates. Data sets were compared with Student's t-test (two-tailed, equal variances) or one-way ANOVA (for comparison of more than 2 groups) using GraphPad Prism 6.0. Results were considered to be significant if p-value < 0.05.

Results

A systematic high throughput spliceosome RNAi screen for TNBC cell migration and proliferation

To discover the spliceosome components driving TNBC cell migration, we performed a high-content imaging SMARTpool (pool of 4 single siRNAs per target) RNAi screen targeting all 244 spliceosome components in two highly motile TNBC cancer cell lines (Hs578T and MDA-MB-231). TNBC cell migratory behaviour was quantitatively assessed using both the PhagoKinetic Track (PKT)[65] and live cell migration assays[64]; in a parallel screen the effects on proliferation were assessed using nuclei counting and a SulfoRhodamine B (SRB) assays (Fig. 1A). To select factors that were primarily involved in cell migration, splicing factors that were strongly inhibiting proliferation upon knockdown were excluded (Fig. 1B.i). The remaining splicing factors were clustered based on the fold change in cell migration speed upon knockdown (Fig. 1B.ii). We selected splicing factors that upon knockdown consistently reduced TNBC cell migration speed and reduced the overall net migratory track area and length of the major track axis in the PKT assay, two parameters that are critical in reflecting migration behavior (Fig. 1B-D, Suppl. Fig. 1-2). A significant positive correlation between the different quantitative migration parameters in both Hs578T and MDA-MB-231 cells was observed (Fig. 1E). Based on the effects on live cell migration and track phenotype, we selected 26 splicing factor hits affecting cell migration in these two TNBC cell lines; this group of selected hits was enriched for core splicing factors[136] (Suppl. Fig. 3A-B) with many factors part of U2 and Sm complexes (Suppl. Fig. 3C-D). The U2 complex is involved in splice site recognition and is attached to the RNA during the whole splicing reaction[325]. The seven Sm proteins form a heptameric ring and are involved in assembly of the snRNPs[325]; remarkably, six out of the seven Sm proteins were

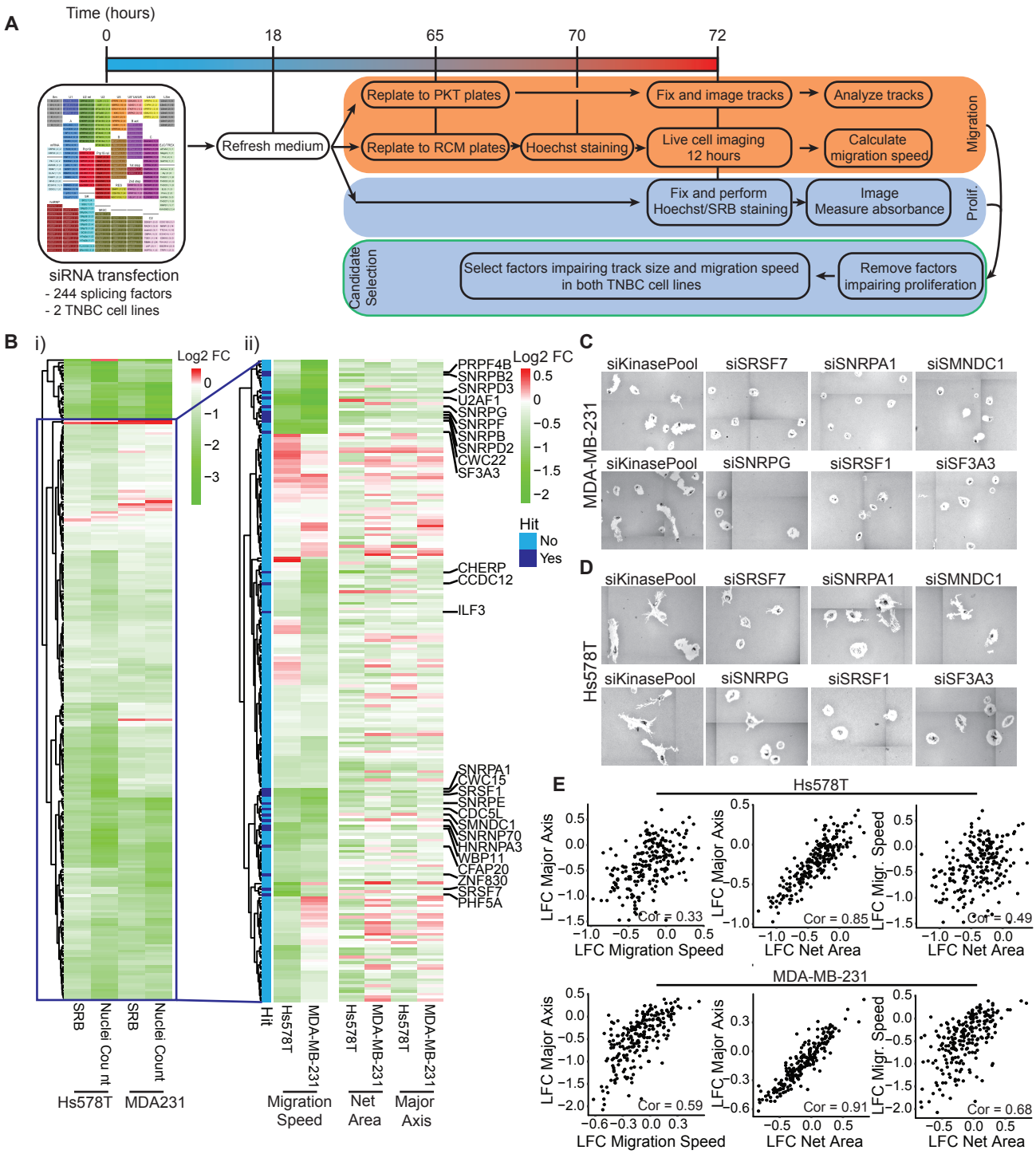


Figure 1. ***Spliceosome RNAi screen identifies novel splicing regulators of tumor cell migration.*** (A) Overview of the time line of the primary screen and candidate selection. (B) Hierarchical clustering (Euclidean distance, complete linkage) of log2 fold change of splicing factor knockdown on proliferation in Hs578T and MDA-MB-231 (i). Splicing factors strongly inhibiting proliferation in both cell lines (upper cluster) were excluded from further analysis. Hierarchical clustering (Euclidean distance, complete linkage) of log2 fold change of splicing factor knockdown effects on live cell migration in Hs578T and MDA-MB-231 cells (ii). Splicing factors that upon depletion strongly inhibited live cell migration, net area and major axis were selected for further validation (dark blue labeling). Example track phenotypes of hit knockdown in MDA-MB-231 (C) and Hs578T (D). (E) Pearson correlation between different migration parameters used in the primary screen.

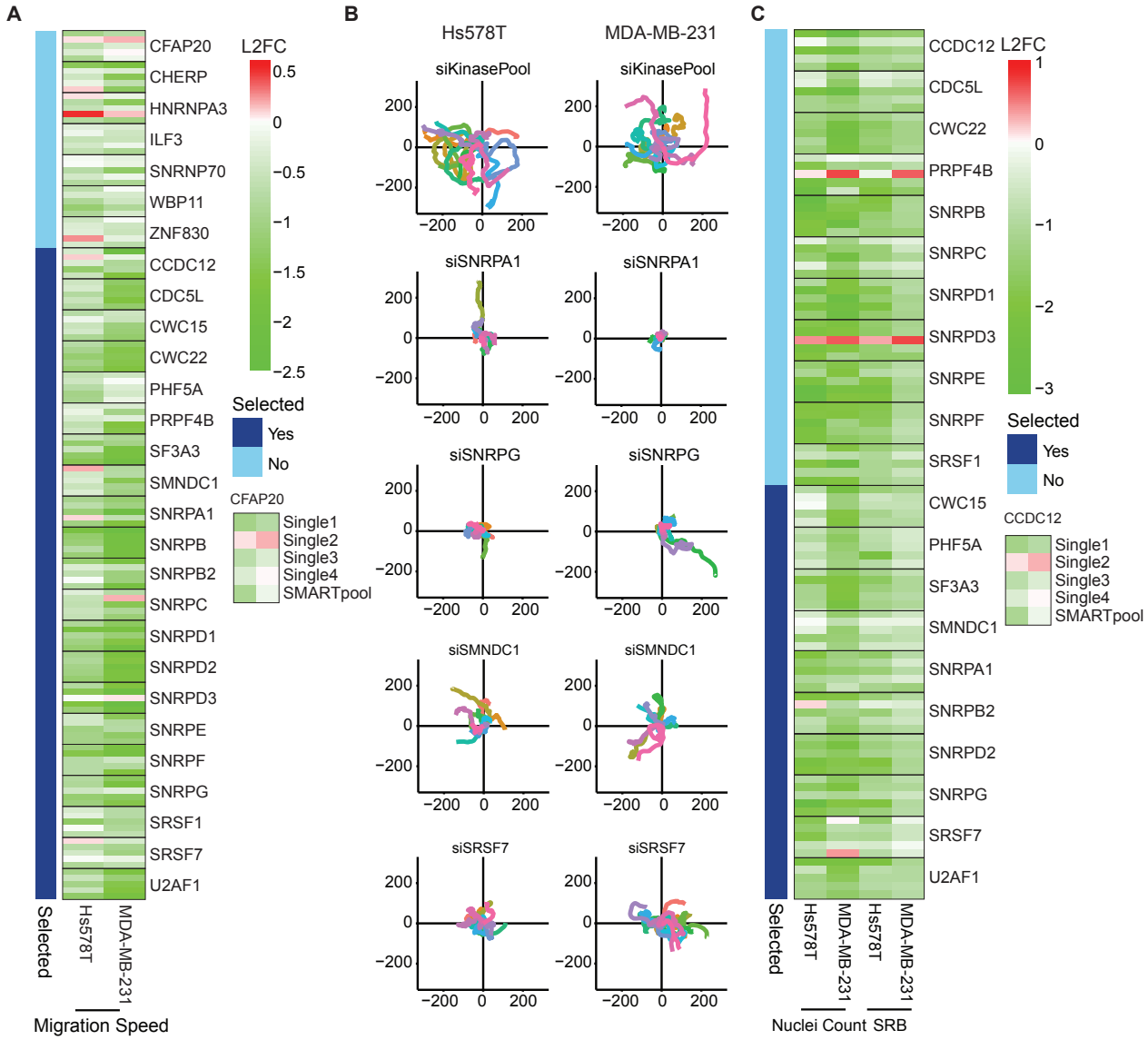


Figure 2. ***Candidate validation in a secondary screen.*** (A) Hierarchical clustering (Euclidean distance, complete linkage) of log2 fold change of live cell migration speed 72 hours after single and SMARTpool siRNA knockdown. (B) Single cell trajectories of cell migration upon SMARTpool knockdown of SNRPA1, SNRPG, SMNDC1 and SRSF7 in Hs578T and MDA-MB-231 cell lines. (C) Hierarchical clustering (Euclidean distance, complete linkage) of log2 fold change of proliferateion 72 hours after knockdown of single and SMARTpool siRNAs of validated candidates.

identified as a hit in our primary screen. To validate the selected primary hits, the migration and proliferation screens were repeated using SMARTpool as well as 4 single siRNA sequences.

From this validation screen we selected splicing factors that showed consistent inhibition of live cell migration in both TNBC cell lines with the SMARTpool and at least 2 single siRNAs (Fig. 2A-B, Suppl. Fig. 4). To ensure the selection of candidates that specifically affect cell migration and not cell proliferation, for further studies we defined 10 splicing factors that upon knockdown demonstrated minimal effects on cell proliferation (Fig. 2C). These selective modulators of cell migration included CWC15, PHF5A, SF3A3, SMNDC1, SNRPA1, SNRPB2, SNRPD2, SNRPG,

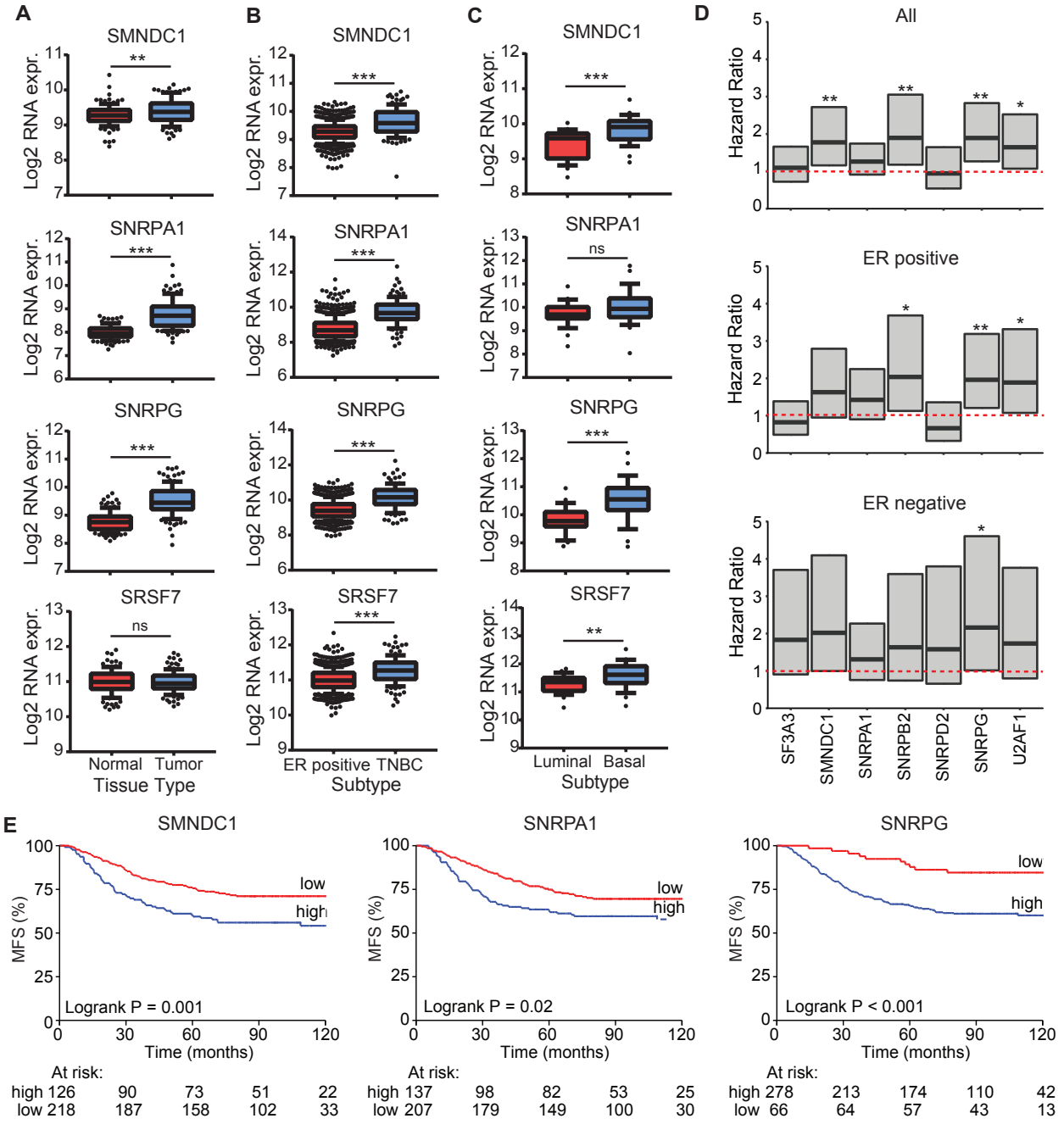
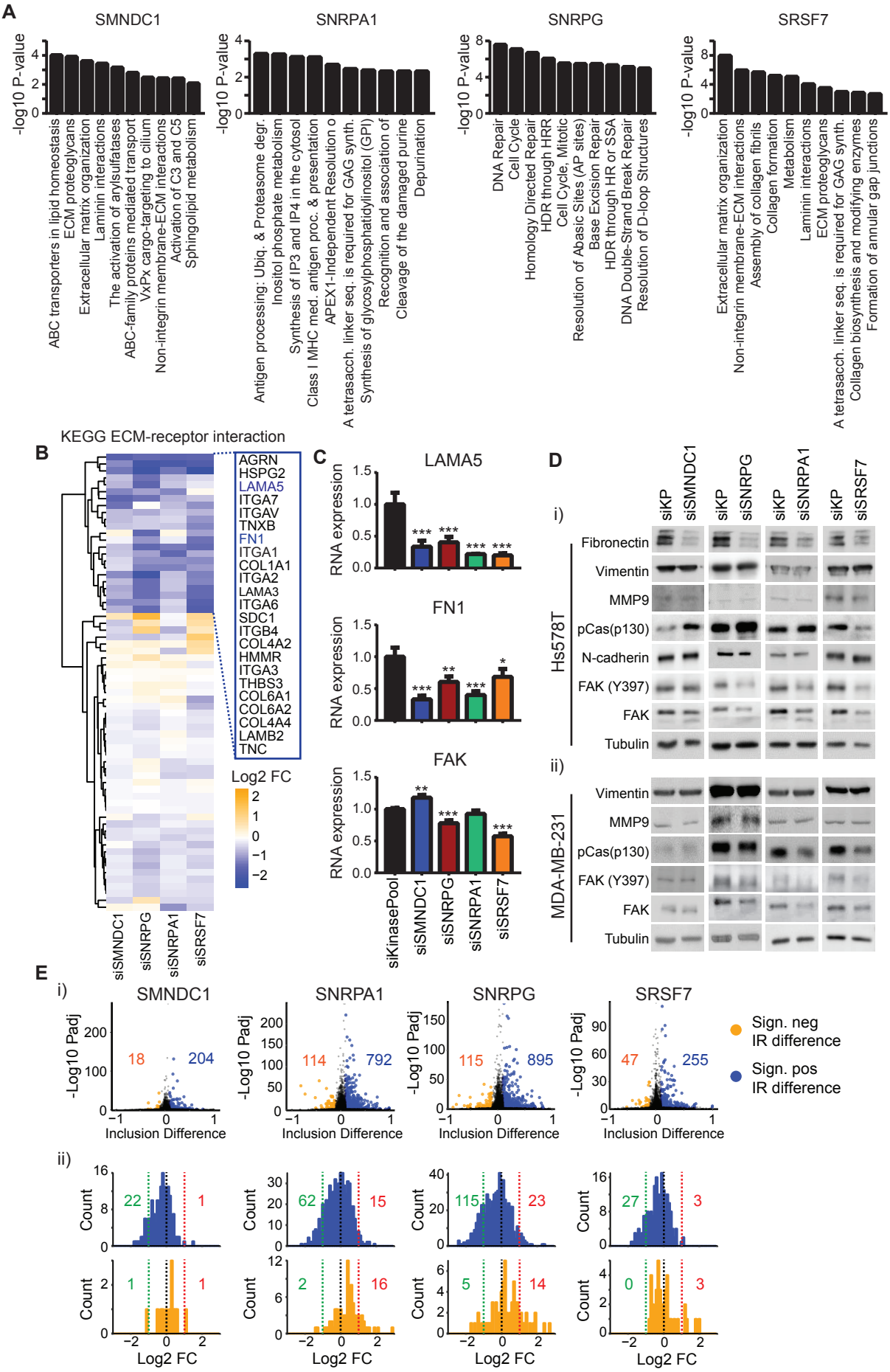


Figure 3. ***Candidate splicing factors are associated with more aggressive tumor subtype and metastasis formation.*** (A) Log2 RNA expression levels of splicing factors in normal and tumor breast cancer patient tissue (n=114), (B) ER positive (n=807) and TNBC (n=116) patient tumors and (C) luminal (n=25) and basal (n=27) breast cancer cell lines, (D) Hazard ratio and 95% confidence interval of candidates for metasis-free survival in all, ER positive and ER negative tumor subtypes. (E) Kaplan Meier curves of expression of SMNDC1, SNRPA1 and SNRPG of metastasis-free survival in BC patients, optimal split. * P < 0.05, ** P < 0.01, *** P < 0.001.

SRSF7 and U2AF1. The RNA splicing reaction is catalyzed by 5 small nuclear RNAs (snRNAs). First, to assess the role of 10 selected candidates in RNA splicing in TNBC cell lines, we assessed the snRNA expression upon candidate knockdown (Suppl. Fig. 5). All of the 10 selected hits significantly affected the expression of at least one of the snRNAs, indicative for


A
SMNDC1
SNRPA1
SNRPG
SRSF7
-log10 P-value
ABC transporters in lipid homeostasis
ECM proteoglycans
Extracellular matrix organization
Laminin interactions
The activation of arylsulfatases
ABC-family proteins mediated transport
VxPx cargo-targeting to cilium
Non-integrin membrane-ECM interactions
Activation of C3 and C5
Sphingolipid metabolism
Antigen processing: Ubiq. & Proteasome degr.
Inositol phosphate metabolism
Synthesis of IP3 and IP4 in the cytosol
Class I MHC med. antigen proc. & presentation
APEX1-Independent Resolution o
A tetrasacch. linker seq. is required for GAG synth.
Synthesis of glycosylphosphatidylinositol (GPI)
Recognition and association of
Cleavage of the damaged purine
Depurination
DNA Repair
Cell Cycle
Homology Directed Repair
HDR through HRR
Cell Cycle, Mitotic
Resolution of Abasic Sites (AP sites)
Base Excision Repair
HDR through HR or SSA
DNA Double-Strand Break Repair
Resolution of D-loop Structures
Extracellular matrix organization
Non-integrin membrane-ECM interactions
Assembly of collagen fibrils
Collagen formation
Metabolism
Laminin interactions
ECM proteoglycans
A tetrasacch. linker seq. is required for GAG synth.
Collagen biosynthesis and modifying enzymes
Formation of annular gap junctions
B
KEGG ECM-receptor interaction
AGRN
HSPG2
LAMA5
ITGA7
ITGAV
TNXB
FN1
ITGA1
COL1A1
ITGA2
LAMA3
ITGA6
SDC1
ITGB4
COL4A2
HMMR
ITGA3
THBS3
COL6A1
COL6A2
COL4A4
LAMB2
TNC
Log2 FC
siSMNDC1
siSNRPG
siSNRPA1
siSRSF7
C
LAMA5
FN1
FAK
RNA expression
siKinasePool
siSMNDC1
siSNRPG
siSNRPA1
siSRSF7
D
siKP
siSMNDC1
siSNRPG
siSNRPA1
siSRSF7
i)
Hs578T
Fibronectin
Vimentin
MMP9
pCas(p130)
N-cadherin
FAK (Y397)
FAK
Tubulin
ii)
MDA-MB-231
Vimentin
MMP9
pCas(p130)
FAK (Y397)
FAK
Tubulin
E
i)
SMNDC1
SNRPA1
SNRPG
SRSF7
-Log10 Padj
Inclusion Difference
18
204
114
792
115
895
47
255
Sign. neg IR difference
Sign. pos IR difference
ii)
Count
Log2 FC
22
1
1
1
62
15
2
16
115
23
5
14
27
3
0
3

◄ **Figure 4.** ***SMNDC1, SNRPA1, SNRPG and SRSF7 knockdown results in decreased expression of migratory signaling pathways.*** (A) Reactome pathway over-expression analysis of significantly downregulated genes (DEGs) of candidate knockdown in MDA-MB-231 cells. (B) Heatmap (Euclidean distance, complete linkage) of log2 fold change expression levels of genes involved in the KEGG ECM-receptor interaction pathway 72 hours after candidate splicing factor knockdown. (C) Examples of log2 fold change in RNA expression levels of genes involved in the ECM-receptor interaction pathway (LAMA5 and FN1) or migration 72 hours after knockdown of candidate splicing factors. (D) Protein levels of genes cell migration or EMT 72 hours after splicing factor knockdown in Hs578T (i) or MDA-MB-231 (ii). (E). Effects of SMNDC1, SNRPA, SNRPG and SRSF7 knockdown on intron retention in MDA-MB-231 cells 72 hours after transfection. (i) Inclusion difference between splicing factor knockdown and control knockdown. Significant positive and negative events are labeled in blue and orange, respectively. (ii) Histogram of total gene expression of genes showing significant positive (blue) or negative (orange) intron inclusion events upon splicing factor knockdown.

their role in the regulation of splicing. Interestingly, the specific snRNAs affected were highly dependent on the splicing factor knockdown, suggesting that the ten candidates use different mechanisms to regulate the splicing reaction.

Splicing modulators of cell migration are associated with aggressive BC subtypes and metastasis-free survival in breast cancer patients

To further relate our splicing factor hits to breast cancer development, progression and metastasis formation in patients, we compared the expression levels of these factors between normal breast tissue and matched tumor tissue using RNA sequencing data derived from The Cancer Genome Atlas (TCGA) (Fig. 3A, Suppl. Fig. 6A). Except for SRSF7, U2AF1, SF3A3 and CWC15, all selected splicing factors were significantly higher expressed in tumor tissue. Furthermore, comparing different breast cancer subtypes, all 10 splicing factors were significantly higher expressed in the more aggressive TNBC compared to the ER positive subtype (Fig. 3B, Suppl. Fig. 6B). Moreover, all candidates except SNRPA1 and U2AF1 demonstrated elevated RNA levels in the basal breast cancer cell lines compared to the luminal cell lines (Fig. 3C, Suppl. Fig. 6C). Since splicing factor SF3B1 is frequently mutated in multiple cancer types[171,326,327] and has also been demonstrated to be a breast cancer driver gene[76], we investigated mutations, deletions and amplifications of the 10 splicing factor hits in breast cancer tumor tissue. Mutations and deletions were rarely identified in these patients (Suppl. Fig. 7). However, gene amplifications were observed with a higher frequency and, interestingly, for 5 of these factors significantly enriched in TNBC tumors compared to ER positive tumors (Suppl. Fig. 7). Finally, the association of the gene expression of the 10 splicing factors with BC distant metastasis-free survival in cohort of 344 untreated lymph node negative breast cancer patients was determined. Of the 7 splicing factors represented in this dataset, 4 factors (SMNDC1, SNRPB2, SNRPG and U2AF1) demonstrated significantly increased hazard ratios for metastasis formation (Fig. 3D), which was associated with a significant shorter metastasis free survival for SMNDC1, SNRPA1 and SNRPG (Fig. 3E). Altogether, we showed that the majority of our 10 selected splicing factor candidates are higher expressed in more aggressive BC tumor

subtypes and associated with an increased chance of metastasis formation in breast cancer patients.

SRSF7, SNRPG, SNRPA1 and SMNDC1 modulate expression of cell-matrix adhesion pathways

Given the role of splicing factors in regulating transcript expression and maturation, we anticipated that the depletion of splicing factors that modulate TNBC cell migration would affect transcriptional programs that control cell motility. Based on our cell migration screen and proliferation/viability data (Suppl. Fig. 8) as well as the clinical association with metastasis-free survival and BC subtypes, we further focused on four splicing factor candidates: SRSF7, SNRPG, SNRPA1 and SMNDC1. To investigate the transcriptional programs that are dependent on these four splicing factors, we first depleted the expression of these individual factors in MDA-MB-231 cells (Suppl. Fig. 9A) followed by next generation sequencing (NGS) in MDA-MB-231 cells. For these NGS experiments, we pooled the two single siRNAs that demonstrated the strongest knockdown (Suppl. Fig. 9B, red labeling) for each of the four splicing factors. NGS showed a knockdown efficiency greater than 80% for all four individual splicing factors, which was confirmed by RT-qPCR (Suppl. Fig. 9C). Candidate splicing factor knockdown resulted in a variable number of differentially expressed genes (DEGs, log 2 fold change > 1 or < -1 and adjusted P-value < 0.05), ranging from ~350 for SMNDC1 depletion to more than ~1750 for SNRPG depletion (Suppl. Fig. 10A). Importantly, for a selected gene set effects of splicing factor depletion observed by NGS in MDA-MB-231 were confirmed by RT-qPCR in both MDA-MB-231 and Hs578T cell lines (Suppl. Fig. 9D), providing confidence on the overall NGS procedure and the consistent effect of splicing factor knockdown in these TNBC cell lines. Depletion of SRSF7, SNRPG, SNRPA1 and SMNDC1 mainly resulted in downregulated genes (Suppl. Fig. 10A). 51 genes were commonly downregulated upon depletion of all 4 splicing factors, while 176 genes were downregulated in at least 3 splicing factor knockdown conditions (Suppl. Fig. 10B). Only limited self-regulation of the spliceosome was observed: depletion of SMNDC1, SNRPA1, SNRPG or SRSF7 showed minor effects on the expression level of other splicing factors, further indicating that the observed effect on TNBC cell migration are specific for the depletion of these individual factors (Suppl. Fig. 11). To unravel the pathways differently regulated upon candidate splicing factor knockdown, we performed over-representation analyses on the downregulated gene sets using Reactome and KEGG databases (Fig. 4A and Suppl. Fig. 12A). Knockdown of all four splicing factors resulted in downregulation of pathways related to cell migration, such as extracellular matrix (ECM)-receptor interaction, ECM organization and laminin interactions and involved downregulation of various extracellular matrix components and integrin cell-matrix adhesion receptors (Fig. 4A-C, Suppl. Fig. 12B-C). These effects were also confirmed at protein level: splicing factor knockdown resulted in reduced levels of focal adhesion kinase (FAK), p-FAK and fibronectin (Fig. 4D) and were reflected in the morphological changes of the F-actin network and focal adhesions of the cells (Suppl. Fig. 13). Despite the altered cellular phenotype, depletion of these four candidate splicing factors did not induce a bona-fide mesenchymal-to-epithelial

transition; no difference in levels of N-cadherin, vimentin and MMP9 was observed (Fig. 4D). Next to the commonly affected migratory pathways, splicing factor specific effects were also observed: SMNDC1 knockdown led to downregulation of ABC-transporters; SNRPA1 affected inositol phosphatase and immune signaling; SNRPG mediated cell cycle and DNA damage related pathways; and SRSF7 modulated amino acid metabolism (Fig. 4A, Suppl. Fig 12A). Although no exon-specific alternative splicing events could be identified, depletion of all 4 splicing factors did increase intron retention (Fig. 4E, top panel). Increased intron retention was mainly related to downregulation of total gene expression, while decreased intron retention often resulted in significant upregulation of the affected gene (Fig. 4E, bottom panels). Altogether, we demonstrated that splicing factors SNRPG, SNRPA1, SNRPG and SMNDC1 regulate gene expression via intron retention and as such control TNBC migratory pathways.

SRSF7 and SMNDC1 reside in splicing complexes

Our NGS results demonstrated that knockdown of SNRPG and SNRPA1 also reduced expression of genes involved in cell cycle and DNA damage repair, whereas SRSF7 and SMNDC1 knockdown were primarily affecting cell migration related pathways. Therefore, to unravel the role of splicing factors in cell migration specifically, we next focused on the SRSF7 and SMNDC1. First, we validated the prominent role of SRSF7 and SMNDC1 in TNBC cell migration: CRISPR/Cas9-mediated knockout of SRSF7 and SMNDC1 reduced the migration speed of MDA-MB-231 cells (Suppl. Fig. 14A-B). Moreover, inducible overexpression of SRSF7 and SMNDC1 caused a more scattered cellular phenotype of basal A SUM149PT cells that are otherwise organized in clusters (Suppl. Fig. 14C-D). Given that SRSF7 and SMNDC1 cooperate in large protein complexes to mediate RNA processing, we next uncovered the protein-protein interaction (PPI) complexes in which SMNDC1 and SRSF7 are functioning. We first established bacterial artificial chromosome (BAC) green fluorescent protein (GFP) MDA-MB-231 and Hs578T cell lines that express the SMNDC1-GFP and SRSF7-GFP fusion products. These BAC-GFP cell lines contain a genomic copy of the gene of interest that next to the introns and exons bears at least 10 kB flanking DNA with regulatory sequences including the promoter region, thereby ensuring endogenous regulation of the target gene. As expected for splicing factors, SMNDC1 and SRSF7 resided in the nucleus in both MDA-MB-231 and Hs578T cell lines (Fig. 5A-B). Depletion of SMNDC1 or SRSF7 reduced the GFP intensity (Fig. 5A-B and Suppl. Fig. 15), supporting that our BAC-GFP cell lines contain the correct fusion gene. To identify the functional complexes of SMNDC1 and SRSF7, we performed GFP pulldowns in these BAC-GFP reporter cell lines, followed by mass spectrometry. Remarkably, the SMNDC1 PPI complex consists of only 10 proteins in MDA-MB-231 and 9 proteins in Hs578T, of which 2 proteins were overlapping: SMNDC1 and SNRPA (Fig. 5C). In contrast SRSF7 resided in a bigger complex, containing 185 and 154 proteins Hs578T and MDA-MB-231 cells, respectively, with 130 overlapping proteins (Fig. 5C). Interestingly, the majority of genes present in the SMNDC1 complexes were also part of the SRSF7 complexes (Fig. 5D), suggesting that these factors are partly functioning in the same larger PPI complex. Such a putative overlapping function was confirmed by the 80 commonly down-regulated genes upon SRSF7 or SMNDC1 knockdown (Suppl. Fig. 10B). As expected, pathway over-representation analysis of the

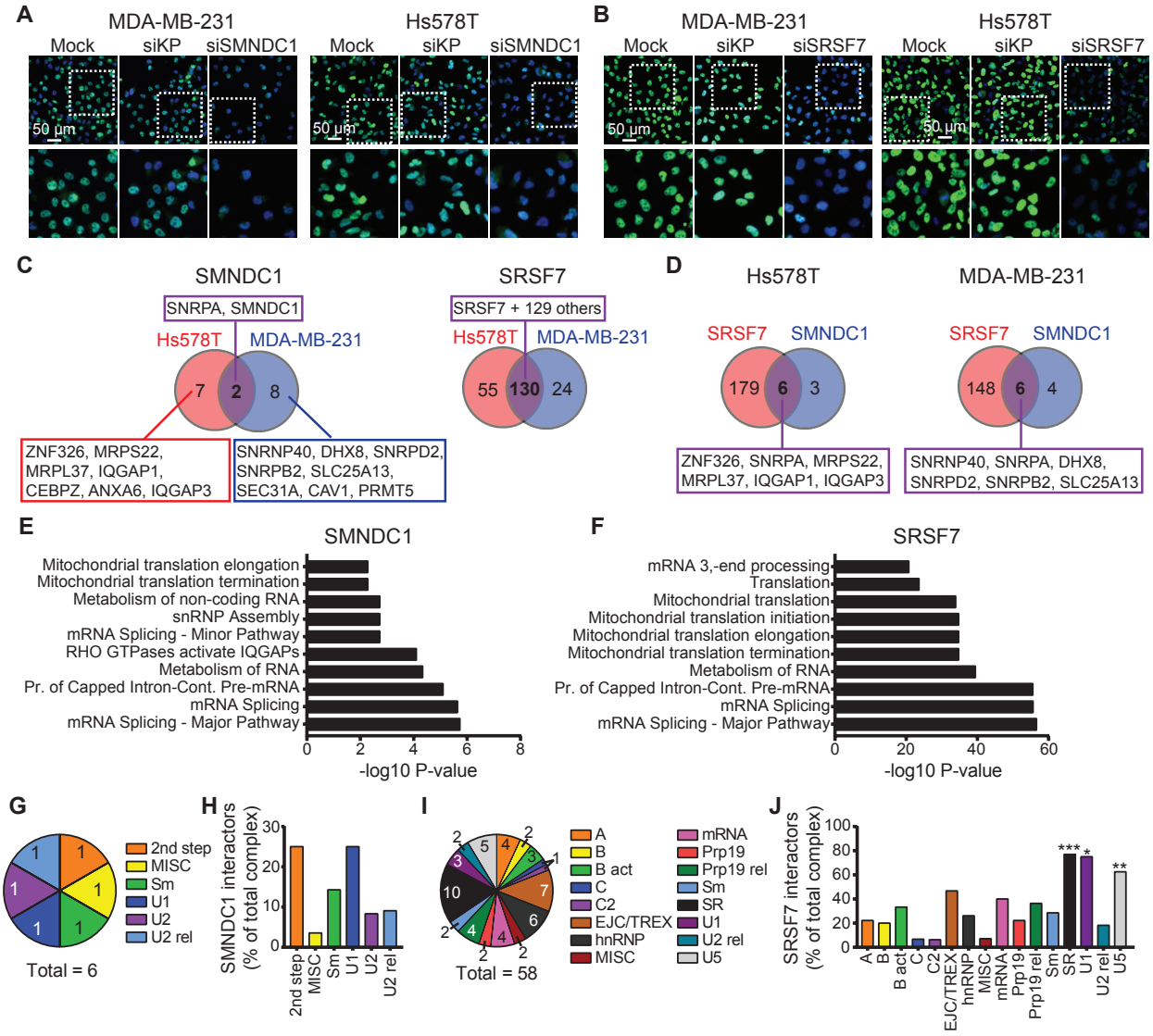


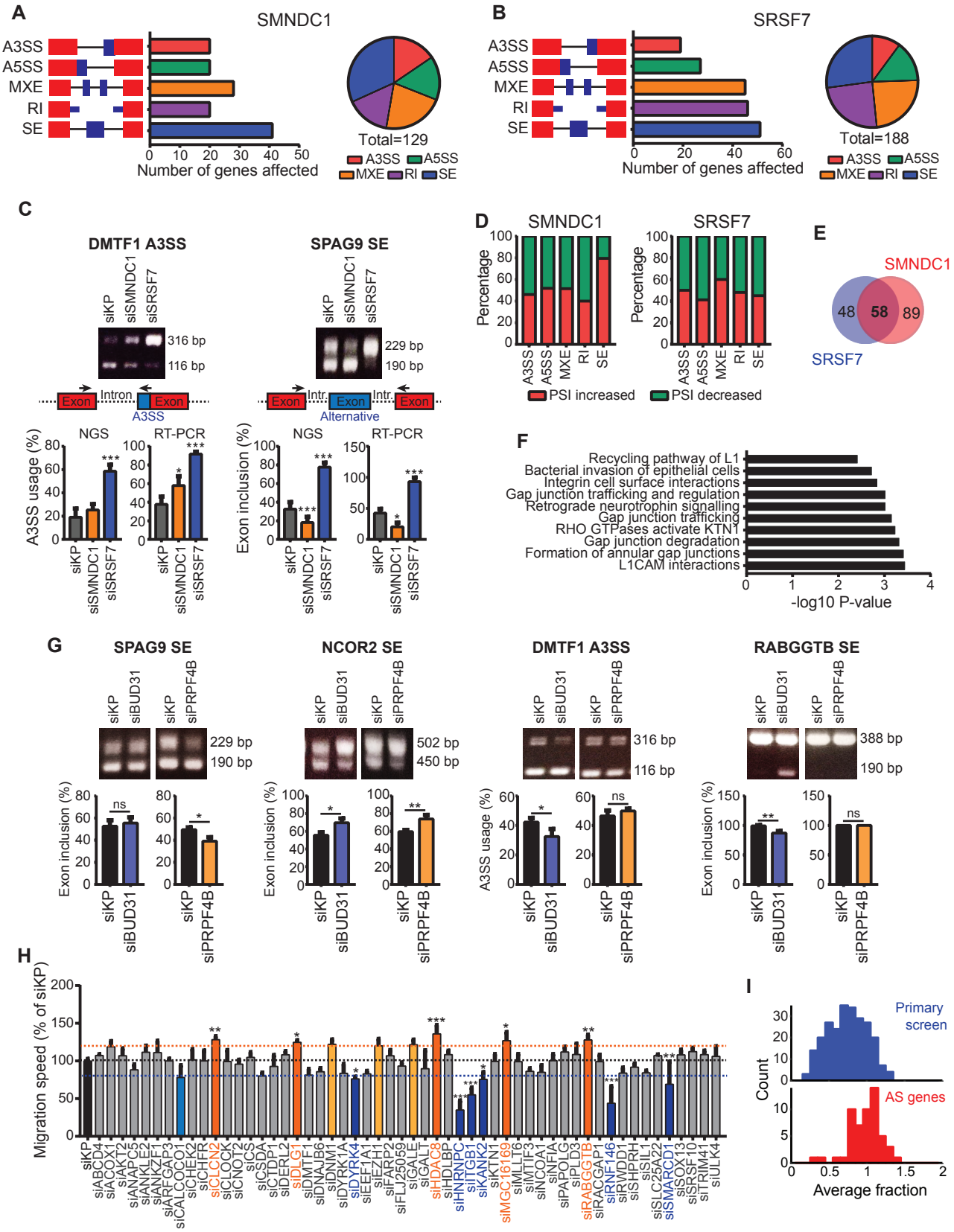
Figure 5. ***SMNDC1 and SRSF7 mainly interact with other splicing factors.*** (A) Validation of SMNDC1 BAC-GFP reporter cell lines. Scale bar = 50µm. (B) Validation of SRSF7 BAC-GFP reporter cell lines. Scale bar = 50µm. (C) Overlap of SMNDC1 (left) and SRSF7 (right) interactors in Hs578T (red) and MDA-MB-231 (blue) cell lines. (D) Overlap of the SRSF7 and SMNDC1 complexes in Hs578T (left) and MDA-MB-231 (right) cell lines. (E) Reactome over-representation analysis of proteins in complex with SMNDC1 in at least one of the cell lines. (F) Reactome over-representation analysis of the proteins in complex with SRSF7 in both MDA-MB-231 and Hs578T cell lines. (G) Distribution of the proteins interacting with SMNDC1 in at least one of the cell lines over the different spliceosome subcomplexes. (H) Enrichment of the spliceosome subcomplexes in the proteins interacting with SMNDC1 in at least one of the cell lines. (I) Distribution of the proteins interacting with SRSF7 in both Hs578T and MDA-MB-231 over the different spliceosome subcomplexes. (J) Enrichment of spliceosome subcomplexes in the proteins interacting with SRSF7 in both Hs578T and MDA-MB-231.

SMNDC1 and SRSF7 interactomes demonstrated strong enrichment of splicing and (mitochondrial) translation related pathways (Fig. 5E-F, Suppl. Fig. 16-17), confirming that the main function of these proteins regulation of RNA splicing. As expected due to the low number of interactors, splicing factors in the SMNDC1 complex were not enriched for a specific

spliceosome sub-complex (Fig. 5G-H). SRSF7 is a member of the family of serine/arginine (SR)-rich proteins that can interact with both RNA and other proteins via an RNA recognition motif (RRM) and a SR-rich domain, respectively. Activation of SR proteins is mediated by SR-rich protein-specific kinases (SRPK) mediated phosphorylation and SR proteins can regulate splicing with or without RNA binding[328]. According to our mass spectrometry analysis, SRSF7 mainly interacts with the SR proteins and the U1 and U5 spliceosome complexes in our TNBC cell lines (Fig. 5I-J). In this way, SRSF7 can connect the U1 snRNP at the 5' splice site with the U2AF complex at the 3' splice site and facilitate the recruitment of the U4/U5/U6 snRNP during the first steps of the splicing reaction.

SRSF7 knockdown can affect migration by modulating alternative exon usage

Our pulldown experiments demonstrated that both SRSF7 and SMNDC1 reside in a large PPI complex involved in (alternative) splicing. To get further insight in alternative splicing events that are mediated by SRSF7 and SMNDC1 we again performed next generation sequencing and increased the sequencing depth to 100 million paired end reads per sample. We confirmed the intron retention events identified in our first NGS analysis for both SMNDC1 and SRFS7 depletion (Suppl. Fig. 18). Also significant alternative exon usage (rMATS algorithm, $p < 0.01$, absolute inclusion difference > 0.1) was observed after both SMNDC1 and SRSF7 depletion (Fig. 6A-B): 129 (siSMNDC1) and 188 (siSRSF7) genes were affected by different alternative splicing events (alternative 3' splice site (A3SS) usage, alternative 5' splice site (A5SS) usage, mutually exclusive exon (MXE) usage, intron retention (IR) events and exon skipping (SE)). The reduced number of intron retention events observed in this analysis is related to the DEXseq2 method that was designed to detect intron retention events. In contrast, the rMATS algorithm only includes reads covering exons or exon junctions, while excluding reads only covering introns. Therefore, the rMATS algorithm is efficient for detection of alternative exon usage, but less sensitive for capturing intron retention. In general, SMNDC1 and SRSF7 knockdown affected alternative splicing of target genes via various types of events, with a slight enrichment for exon skipping (Fig. 6A-B). 11 out of 14 alternative splicing events could be validated by RT-PCR (Fig. 6C, Suppl. Fig. 19). While SRSF7 knockdown demonstrated equal numbers of increased and decreased exon skipping events, SMNDC1 knockdown mainly resulted in increased exon skipping (Fig. 6D). 58 genes were significantly affected by alternative splicing after both SMNDC1 and SRSF7 depletion (Fig. 6E); these genes were enriched for pathways involved in cell migration such as gap junctions and Rho GTPases (Fig. 6F). As described above, in our pulldown analysis we identified interactions between SRSF7 and our previously identified TNBC migration modulators, PRPF4B and BUD31 (Koedoot et al, unpublished). We confirmed that disruption of the functionality of the large PPI complex after PRPF4B or BUD31 depletion also resulted in similar alternative splicing events (Fig. 6G). Since the events were not consistently affected by depletion of all these factors in the same complex the alternative splicing pattern is likely splicing factor specific. Next, to identify the downstream components of these splicing factors that could be the critical components to affect cell migration behavior, we performed knockdown of the majority of the overlapping downstream splicing targets of SRSF7 and SMNDC1 in MDA-MB-231 cells followed by quantification of the cell migration speed. We


A
SMNDC1
A3SS
A5SS
MXE
RI
SE
Number of genes affected
Total=129
B
SRSF7
Total=188
C
DMTF1 A3SS
SPAG9 SE
siKP
siSMNDC1
siSRSF7
316 bp
116 bp
229 bp
190 bp
Exon
Intron
A3SS
Alternative
NGS
RT-PCR
A3SS usage (%)
Exon inclusion (%)
D
Percentage
PSI increased
PSI decreased
E
SMNDC1
SRSF7
48
58
89
F
Recycling pathway of L1
Bacterial invasion of epithelial cells
Integrin cell surface interactions
Gap junction trafficking and regulation
Retrograde neurotrophin signalling
Gap junction trafficking
RHO GTPases activate KTN1
Gap junction degradation
Formation of annular gap junctions
L1CAM interactions
-log10 P-value
G
SPAG9 SE
NCOR2 SE
DMTF1 A3SS
RABGGTB SE
siBUD31
siPRPF4B
502 bp
450 bp
388 bp
ns
H
Migration speed (% of siKP)
I
Primary screen
AS genes
Count
Average fraction

◄ **Figure 6.** ***SMNDC1 and SRSF7 regulate alternative exon usage.*** (A) Number of genes affected by alternative splicing events 72 hours after knockdown of SMNDC1 in MDA-MB-231 cells. A3SS = alternative 3' splice site usage, A5SS = alternative 5' splice site usage, MXE = mutually exclusive exon, RI = retained intron, SE = skipped exon. (B) Same as in A, but now for SRSF7 knockdown. (C) Validation of alternative 3' splice site usage of DMTF1 and SPAG9 skipped exon by RT-PCR. (D) Percentage spliced in for the different significant alternative splicing events detected upon SMNDC1 (left) and SRSF7 (right) depletion. (E) Overlap of genes affected by alternative splicing after SMNDC1 and SRSF7 knockdown in MDA-MB-231 cells. (F) Reactome over-representation analysis of the 58 overlap genes showing significant alternative splicing events 72 hours after knockdown of both SRSF7 and SMNDC1. (G) RT-PCR of alternative splicing events validated upon SRSF7 knockdown, now for BUD31 and PRPF4B knockdown in MDA-MB-231 cells. (H) Effect of knockdown of alternatively spliced genes on random cell migration in MDA-MB-231. Orange: at least 20% increase in migration speed, blue: at least 20% decrease in migration speed. (I) Histogram of the average fraction of migration speed compared to control knockdown for splicing factors in the primary screen (top panel, blue) and genes alternatively spliced upon SRSF7 or SMNDC1 knockdown (red, bottom panel, genes in Fig. 6H).

identified 7 genes that upon depletion significantly inhibited the migratory behavior of MDA-MB-231 cells, including HNRNPC, ITG1B, RNF146, SMARCD1, KANK2, DYRK4 and CALCOCO1 (Fig. 6H). Interestingly, 8 genes significantly increased the migration speed upon depletion including HDAC8, MGC16169, DLG1, CLCN2 and RABGGTB. Since an increase in cell migration was not observed during our primary splicing screening efforts (Fig. 6I), we suggest that knockdown of SRSF7 and SMNDC1 could control TNBC motility by regulating the alternative splicing of positive but also negative regulators of cell migration.

Discussion

In the present study, we performed an imaging-based RNAi screen to uncover the role of RNA splicing in breast cancer cell motility. Through a systematic RNAi screening strategy we identified ten spliceosomal components that selective modulate migration in MDA-MB-231 and Hs578T cell lines without affecting cell proliferation and survival. Of these ten factors SMNDC1, SNRPA1, SNRPG and SRSF7 were associated with metastasis-free survival in human breast cancer patients. RNA sequencing revealed that these factors commonly modulate the expression of genes involved in focal adhesion signalling and ECM-receptor interactions. We further uncovered the protein-protein interaction complex where these factors reside and demonstrated that SRSF7 associates with factors of the SR, U1 and U5 spliceosomal complexes. Our deep RNA sequencing data indicate that the splicing factors SRSF7 and SMNDC1 are critical in the control of alternative splicing events of a large number of genes, some of which are directly controlling TNBC cell migration. Overall our data provide evidence that several splicing factors that are clinically associated with poor disease outcome and limited metastasis-free survival are critical for modulating TNBC cell migratory behaviour through modulation of RNA processing of genes that are essential in cell-matrix adhesion signalling.

Recent studies demonstrated a role for alternative splicing in essential pathways involved in (breast) cancer development and progression, including tumor cell migratory and invasive behavior[67,133] (Koedoot et al, unpublished). In our previous screening efforts, we identified an important role for splicing factors SPRK1, PRPF4B and BUD31 in TNBC cell migration and metastasis formation[66] (Koedoot et al, unpublished). Here we discovered ten splicing factors (U2AF1, SRSF7, SNRPG, SNRPD2, SNRPA1, SNRPB2, SMNDC1, SF3A3, PHF5A and CWC15), that upon knockdown strongly decreased cell migration speed in MDA-MB-231 and Hs578T cell lines. SRPK1, PRPF4B and BUD31 were also verified in our current screens, but did not meet the strict selection criteria used to select these 10 new candidates. Some of these new splicing factor candidates have previously been associated with cancer progression. For example, SNRPB2 expression is controlled by invasion-promoting MT1-MMP that is known for its direct link to tumorigenesis and metastasis[329]. SRSF7 depletion induced apoptosis and growth arrest in lung and colon cancer cells[330,331]. Furthermore, SF3A3 was identified as p53 inhibitor in lung cancer cells[332]. Lastly, U2AF1 is mutated in 50% of the patients with myelodysplastic syndromes (MDS) and tumors bearing this mutation are more susceptible to spliceosome inhibition[333]. So far, PHF5A, is the only splicing factor candidate from our screen that has been linked to BC migration before[231]. Therefore our current work provides novel functions of a panel of spliceosomal factors that likely contribute to TNBC cell migratory behaviour and possibly metastasis formation. The latter is supported by the fact that for the majority of our identified splicing factors high expression in breast cancer patients is associated with poor breast cancer clinical outcome, including poor metastasis-free survival and high expression in the more aggressive TNBC subtype.

Next generation sequencing revealed that depletion of SMNDC1, SNRPA1, SNRPG and SRSF7 all resulted in decreased expression of genes involved in focal adhesions and ECM interactions. This effect is likely directly affecting the motility behavior and associated cell phenotypes. We anticipate that for a large part the altered gene expression is related to perturbed RNA processing, since depletion of all 4 splicing factors resulted in increased intron retention. While in the past intron retention events were often overlooked or classified as noise, recent studies show a prominent role for intron retention in controlling gene expression during development but also in disease[334,335]. Differential intron retention patterns are commonly identified across multiple cancer types[336] and, in accordance with our study, commonly result in lower protein levels due to nuclear retention, degradation or reduced translation efficiency[334]. Next to alternative intron usage and decreased total gene expression, SMNDC1 and SRSF7 also affected alternative exon usage of negative and positive regulators of cell migration, such as RABGGTB and ITGB1, respectively. Future studies should elucidate the role of these alternative splicing events on breast cancer cell migration and directly link these events to metastasis formation.

To gain more insight on the overall network involved in SMNDC1 and SRSF7 splicing modulation, in the present study we also uncovered the protein-protein interaction network of SMNDC1 and SRSF7 in both Hs578T and MDA-MB-231 TNBC cells. While SMNDC1 was only part of a small complex (9 and 10 proteins in Hs578T and MDA-MB-231 respectively), SRSF7

(in)directly interacted with many different proteins (130 overlapping in Hs578T and MDA-MB-231) that were enriched for splicing factors part of the SR, U1 and U5 complex. SRSF7 is known as a SR-protein that are known for various functions in the splicing reaction: 1) interact with U2AF bound at the 3' splice site and U1 at the 5' splice site while bound the exonic splicing enhancer, 2) enable connection of the two splice sites by interacting with U1 and U2AF and attracting the U4/U5/U6 tri-snRNP and 3) repress splicing by binding to the intron close to splice site, thereby inhibiting spliceosome assembly by steric hindrance[328]. Remarkably, SRSF7 also interacted with previously identified TNBC motility modulators SRPK1, BUD31 and PRPF4B[66] (Koedoot et al, unpublished). Depletion of these three different components that resided in the same large protein complex resulted in partly similar effects on alternative exon usage, suggesting that alternative splicing events are splicing factor specific. Many components of the SRFS7 complex were part of our current screen, yet did not affect TNBC cell migration. This allows spliceosome specific pharmacological modulation of splicing events that are determining particular biological programs. Although the application of splicing factor SRPK1, SRPK2 and SF3B1 inhibitors for cancer treatment looks promising[286,337], inhibitors for other splicing factors are not yet available. Since our study and others[66,124,127,147] (Koedoot et al, unpublished) demonstrated a prominent role for various other splicing factors in breast cancer progression, it would be worthy to develop inhibitors for these factors as well and exploit their use in cancer treatment.

Acknowledgements

This work was supported by the ERC Advanced grant Triple-BC (grant no. 322737) and the Dutch Cancer Society project (grant nr 2011-5124).

Supplemental data availability

Supplemental Figures 1-19 are available upon request using the following link: https://doi.org/10.17026/dans-xs9-hf76.