SRRM1 coordinates an alternative splicing program that promotes expression of oncogenic protein isoforms.

Kamal Othman[1 2 3], Lauren Viola[1 2 3], Hira Fatima[1 2 3], Jessica Lapierre[1 3], Graham MacLeod[4], Craig D Simpson[5], Christopher Chu[1 3], Russell Beltran[1 3], Yangjing Zhang[1 2 3], Stephane Angers[4 6 7], Olivier Saulnier[8 *] C. Jane McGlade[1 2 3 *].

1-The Arthur and Sonia Labatt Brain Tumour Research Centre, 555 University Avenue, Toronto, ON, M5G 1X8, Canada.

2-Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, ON, M5G 2M9, Canada.

3-Program in Cell and Systems Biology, 555 University Avenue, Toronto, ON, M5G 1X8, Canada.

4-Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada.

5-SPARC BioCentre, The Hospital for Sick Children, 555 University Avenue, Toronto, ON, M5G 1X8, Canada.

6-Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON, M5S 3M2, Canada.

7-Department of Biochemistry, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, M5S 1A8, Canada.

8-Genomics and Development of Childhood Cancers Lab, Inserm U1330, Institut Curie, PSL University, SIREDO Oncology Center, Paris, France.

*Correspondence: Olivier.Saulnier@curie.fr (O. S.), jmcglade@sickkids.ca (C. J. M.).

Abstract:

The alternative splicing of the adapter protein NUMB is dysregulated in multiple cancer types, regulating its functional divergence towards either tumor suppression or oncogenesis in an isoform-dependant manner. Here we utilized a NUMB exon 9 (E9) splicing reporter in a genome-wide CRISPR screen to identify splicing regulators SRRM1 and SRSF11 that promote NUMB oncogenic splicing in colorectal, lung and breast cancer cell lines. We show that SRRM1 and SRSF11 share common protein interactors, RNA targets and effects on an oncogenic splicing program which favors the expression of pro-tumorigenic isoforms. In addition to NUMB E9, SRRM1 regulates cancer-associated splicing events in genes encoding signaling proteins, transcription factors and actin cytoskeleton regulators, many of which are also developmentally regulated, including CD44, MKNK2, ECT2, DIAPH1, KAT5, TCF7L2, FOXM1 and TBX3. Loss of SRRM1 in colorectal and lung cancer cells reduces cell proliferation, colony formation and invasion capabilities, as well as expression of tumour promoters Cyclin D1, Notum, and PRDX2. Our data indicate that SRRM1 regulation of alternative splicing represents a node to target multiple properties of malignant cells, with broad effects on cellular signaling, proliferation, EMT, apoptosis resistance and stemness.

Introduction:

Alternative splicing of pre-mRNA is a mechanism that contributes to proteomic diversity by generating mature mRNA transcripts that encode distinct protein products. The core of the splicing machinery is the spliceosome which consists of five snRNPs (small nuclear ribonucleoprotein particles; snRNP U1, U2, U4, U5 and U6) and hundreds of additional proteins that interact with cis elements at intron-exon boundaries to catalyze the splicing reaction [1]. Additional cis elements act as splicing enhancers or splicing suppressors through the recruitment of auxiliary splicing regulators such as the SR and HNRNP protein families [1-3]. Alternatively spliced exons frequently encode protein sequences within modular domains or intrinsically disordered regions that contain sites of post-translational modification or protein interaction motifs. Therefore, patterns of alternative splicing have important functional implications by re-wiring protein networks that impact cellular physiology [3] and contribute to both the transcriptomic and proteomic complexity that underlie cellular functioning and cell fate determination [2].

Mis-regulation of alternative splicing and the associated changes in protein isoform expression has been linked to cancer initiation, progression and drug resistance [4]. Aberrant splicing can arise through multiple mechanisms, including mutations that disrupt cis-acting splice sites, or through mutations or misexpression of trans-acting splicing factors that alter spliceosome activity [5]. Activation of oncogenic signaling pathways can also alter splicing factor activity to promote the reversal of developmentally regulated splicing patterns. This can result in the expression of tumorigenic protein isoforms and/or reduced expression of isoforms normally associated with cell differentiation and cell fate [4].

Alternative splicing of the NUMB gene is developmentally regulated and is frequently altered in cancer [6-11]. In mammals, NUMB has an essential role during development and a conserved function as an inhibitor of Notch-dependent activation of gene transcription [9,12,13]. Mammalian NUMB also functions to regulate activity of additional pathways including receptor tyrosine kinase, Hedgehog and TP53 signaling [14-16]. NUMB is alternatively spliced to generate four protein isoforms with distinct patterns of expression in different cell types [9,10]. The two largest isoforms (p71 and p72) arise due to inclusion of coding exon 9 (E9) that inserts 49 amino acids in the central region of the protein. The two smaller proteins (p65 and p66) result from E9 skipping. Expression of NUMB E9 is developmentally regulated with stem and progenitor cells primarily expressing E9 included isoforms and differentiated cells primarily expressing E9 skipped isoforms [6,10,17-19]. A number of tissue-specific splicing factors have been shown to alter NUMB E9 splicing during development and cell differentiation including Rbfox3 [18], SRSF family members [6,17,20], RBM4/5/10 [21] and PTBP1 [17,22].

Dysregulation of NUMB alternative splicing is broadly associated with multiple cancer types, and is characterized by increased expression of E9 included isoforms that impact cellular processes associated with tumor progression [7,9,11,15,23]. Despite being broadly implicated in multiple tumour types, the mechanisms that dysregulate NUMB E9 splicing in the context of cancer are less well understood. While mutations in the NUMB gene have not been identified, alterations in splicing factor activity have been implicated in lung cancer where loss of function mutations in RBM10, for example, were shown to increase E9 inclusion [7]. Other evidence suggests that oncogenic signaling, particularly through RAS, alters NUMB splicing at least partly through PTBP1 expression [17,22].

Understanding the mechanistic details by which splicing alterations arise and contribute to cancer provides an opportunity to develop strategies to therapeutically modify pathological splicing events and/or the downstream consequences. Here we use a genome-wide CRISPR screen based on the NUMB E9 splicing event to identify SRRM1 and SRSF11 as common factors that regulate cancer-associated splicing of NUMB exon 9 in colorectal, lung and breast cancer cells. Furthermore, we show that in addition to NUMB, SRRM1 regulates a set of tumor-associated splicing events and promotes expression of oncogenic protein isoforms associated with cell proliferation, actin cytoskeleton dynamics, signaling and stemness.

Results:

A genome wide CRIPSR screen identifies regulators of NUMB exon 9 splicing.

A mini-gene reporter in which alternative splicing of NUMB E9 can be monitored by expression of EGFP or mCherry fluorophores was constructed by cloning EGFP and mCherry coding sequences in tandem downstream of NUMB genomic region including exons 8-10 and corresponding introns. A frameshift mutation was inserted into E9 to allow for translation of EGFP or mCherry fusions in alternate frames depending on the inclusion or skipping of E9 (Fig 1A). The frame shift insertion of two nucleotides (E9+2) did not alter NUMB reporter splicing relative to the non-mutagenized version of the reporter (Supp Fig 1A/B). Reporter responsiveness was validated based on expression of EGFP and mCherry fusions at both the mRNA and protein level in response to transient transfection with known NUMB splicing regulators, constitutively active MEKK (MEKK-EE) or splicing factor SRSF1 (Supp Fig 1C/D).

An inducible DLD1 reporter cell line was generated (Fig 1B) and its suitability for a flow-based genome wide CRISPR screen was verified by transduction with lentiviral vectors expressing Cas9 and sgRNAs targeting known NUMB E9 splicing regulators SRSF1, SRSF3 or PTBP1 sgRNAs, as well as negative and positive controls targeting AAVS1 or mCherry respectively (Fig 1C). Reporter expression in transduced cells was monitored using fluorescence-activated cell sorting (FACS) and gated to capture the top 5% EGFP high/mCherry low and mCherry high/EGFP low populations. Knockout of PTBP1 or SRSF3 and SRSF1 resulted in a shift of the reporter population towards the EGFP high:mCherry low population or the EGFP low:mCherry high population respectively, confirming that the DLD1 reporter cell line was responsive to CRISPR-Cas9 deletion of splicing regulators (Fig 1C).

To identify factors that regulate NUMB E9 splicing, the DLD1 reporter cell line was transduced with the TKOv3 pooled sgRNA library targeting 18,053 protein-coding genes (4 guides per gene) and cultured for 9 days in puromycin selection. Cells within the top ~5% or ~20% of the EGFP high/mCherry low or EGFP low/mCherry high populations were captured by FACS and the sgRNA enrichment was analyzed via high throughput sequencing (Fig 1D). The data were collected from three independent biological replicates. The efficiency of sgRNA targeting and the effect of fitness on sgRNA enrichment scoring was tracked through the loss of sgRNA sequences in the unsorted cell populations at day 0 (T0, after selection) versus day 9 (T9, screen endpoint). Bayes factor (BF) scores [24,25] were determined using the BAGEL algorithm. Precision versus recall scores of core fitness DLD1 genes (Supp Fig 2A) [25] demonstrated efficient (AUC = 0.9704) sgRNA targeting considering the shorter screen duration (9 days). Bayes factor distribution of DLD1 non-essential versus fitness genes determined through BAGEL analysis (FDR < 0.05), yielded good separation indicating that CRISPR-Cas9 targeting was successful (Supp Fig 2B). In total, 515 unique hits were identified (false discovery rate FDR < 0.2, log fold enrichment LFC > 0.585) across both the 5% and 20% sorted populations (Supplementary Data 1).

The enriched biological process terms from Gene Ontology (GO) analysis of screen hits included those related to mRNA splicing/processing and histone modification (Fig 2A/B). Interestingly, other than the nucleoplasm which is the main site of splicing factor localization [26], cellular component GO enrichment analysis also identified the Spt-Ada-Gcn5 acetyltransferase (SAGA) complex which is an evolutionary conserved transcriptional coactivator that regulates gene expression through histone acetyltransferase and deubiquitylase activity (Supp Fig 3A). Among the identified genes, 47 are annotated in the Cancer Gene Census (CGC) according to the

Catalogue of Somatic Mutations in Cancer (COSMIC, https://cancer.sanger.ac.uk/census) including several known oncogenic driver genes such as TP53, PIK3CA, KRAS, and ARID1A. Several splicing factors previously shown to regulate NUMB E9 AS were also identified including SRSF3 [20] and ESRP1 [27].

High confidence hits from the top 5% EGFP high or mCherry high sorted populations representing genes whose knockout had the largest effect on the splicing of the E9 reporter (Fig 2C) and hits with a previously established role associated with alternative splicing (Fig 2D) were prioritized for validation. High confidence hits identified included the splicing factors serine and arginine repetitive matrix 1 (SRRM1), serine and arginine rich splicing factor 11 (SRSF11) and DEAD-box helicase 23 (DDX23) whose involvement in regulating NUMB E9 AS has not previously been reported. String analysis of the splicing associated hits from the 5% sorted populations (Fig 2E) highlighted potential interactions of multiple factors identified from the screen, including connections between SRRM1, SRSF11, and DDX23.

Splicing factors SRRM1, SRSF11 and DDX23 promote NUMB exon 9 inclusion.

To validate potential NUMB E9 splicing regulators, we examined the effect of siRNA knockdown of a selected subset of 33 hits on endogenous NUMB E9 splicing. Selected targets whose knockdown resulted in a greater than 0.25 Log2PSI change compared to non-targeted siRNA controls were further validated in biological triplicate (Supp fig 3C). Amongst these, siRNA mediated knockdown of SRRM1, SRSF11 and DDX23 demonstrated strong and consistent effects on endogenous NUMB E9 splicing (Fig 2F). Targeting of gene expression was confirmed in siRNA treated cells via RT-qPCR and immunoblot (Supp fig 3 D/E). The effect of SRRM1, SRSF11 and DDX23 knockdown on endogenous NUMB E9 splicing was further

examined in HEK293, MDA-MB-468 breast cancer cells and A549 lung cancer cells. SRRM1 knockdown with either siRNA or shRNA targeting resulted in greater E9 skipping across all cell lines tested (Fig 2F and Supp fig 3F/G), indicating a broad role for this factor in promoting inclusion of NUMB E9. Depletion of DDX23 and SRSF11 also resulted in E9 skipping across multiple cell lines tested but had inconsistent effects in HEK293 and MDA-MB-468 cells. These results identify SRRM1, DDX23 and SRSF11 as novel regulators of endogenous NUMB E9 splicing in multiple cellular contexts.

SRRM1 and SRSF11 have common interactions with SR proteins and the splicing machinery.

SRRM1 and SRSF11 are members of the SR family of splicing factors and their expression and activity have been implicated in cancer associated splicing events [28-31]. To further elucidate the mechanisms underlying splicing regulation by SRRM1 and SRSF11 we used mass spectrometry to identify proteins that co-precipitated with Flag tagged SRRM1 or Flag tagged SRSF11 from DLD1 or HEK293T cell lysates (Supplementary data 2). The SRRM1 and SRSF11 interactomes were enriched for RNA binding proteins (RBPs), splicing factors and proteins involved in several spliceosome complexes including the core spliceosome snRNPs U1, U2 and tri-snRNP (U4/U5-U6) as well as the EJC, TREX and PRP19 complexes (Fig 3A, Supp Fig 4A). GO term enrichment analysis revealed a high enrichment of genes associated with RNA splicing via the spliceosome. (Fig 3D-E, Supp Fig 4D-E). Of note, SRRM1 was identified in the SRSF11 interaction network. While SRSF11 was not identified in the SRRM1 interactome, both proteins share the known SRRM1 binding partner, SRRM2, as well as multiple members of the SR family of splicing factors and components of the U1, U2 and Tri-snRNP [32]. In agreement with

prior studies demonstrating SRRM1 role in mRNA export by the Exon Junction Complex (EJC) [33] shared SRRM1 and SRSF11 interactors (Fig 3F) included EJC proteins EIF4A3, RBM8A and RNPS1, as well as factors also associated with mRNA export such as MAGOH, MAGOHB and DDX39B. In addition, multiple m6a mRNA methylation factors were also identified as SRRM1 interactors including WTAP, RBM15, ALKBH5 and ZC3H13 as well as VIRMA (also known as KIAA1429) (Fig 3A, Supp Fig 4A), which was also a hit in the splicing reporter screen (Fig 2D), while SRSF11 bound the m6a readers IGF2BP2 and ELAVL1 [34].

Identification of SRRM1 and SRSF11 regulated splicing events.

To identify additional splicing events regulated by SRRM1 and SRSF11, we conducted bulk RNA-sequencing (RNA-seq) in DLD1 cells line following siRNA silencing of SRRM1 or SRSF11 (Fig 4A). Analysis of RNA-seq data demonstrated expected changes in NUMB E9 splicing (Supp Fig 5A/B). Principal component analysis of 3 biological replicates collected for each condition demonstrated high similarity (Supp Fig 5C). Exonic splicing events with a ΔPSI (Percent Spliced In) > 10% and a FDR < 0.05 compared to control samples were considered alternatively spliced exon (ASE/SE) events of interest (Fig 4B, Supplementary data 3). In total, SRRM1 and SRSF11 silencing resulted in the identification of 712 and 1214 differential ASE events respectively, of which 166 events were common to both SRRM1 and SRSF11 (Fig 4C). Directionality of changes in response to SRRM1 or SRSF11 silencing of common ASEs was ~94% concordant (156/166) with only 10 ASEs demonstrating opposite PSI shifts between the two splicing factors (Fig 4D/E, Supp Fig 5D/E/F). SRRM1 silencing resulted in both exon inclusion (406 events) and exon skipping (306 events), while SRSF11 silencing resulted in exon inclusion in 90 events and exon skipping in 60 events (Supp Fig 5D/E/F).

Comparison of differentially spliced genes (DSGs) with ASE events regulated by SRRM1 or SRSF11 revealed additional common targets shared between the two factors independent of exon location. These results suggest that SRRM1 and SRSF11 may regulate the splicing of different exons within the same gene (Supp Fig 5G). GO term analysis of the biological process enriched terms associated with the DSGs included DNA repair (Fig 4F/G) in both SRRM1 and SRSF11 knockdown cells, as well as microtubule cytoskeleton organization (SRRM1 silencing) and actin cytoskeleton organization (SRSF11 silencing) (Fig 4F/G).

Analysis of altered gene expression in response to SRRM1 and SRSF11 silencing revealed only 78 differentially expressed genes (DEGs) in SRRM1 depleted cells and 206 DEGs in SRSF11 depleted cells (fold change > 2, FDR < 0.05)(Supplementary data 3). Among the DEGs identified, 7 were shared between SRRM1 and SRSF11, 5 of which were altered similarly in both conditions (Supp Fig 5H). Comparison of the DSGs and DEGs revealed minimal overlap with only 3 genes common between the DSGs and DEGs with SRRM1 and SRSF11 silencing (Supp Fig 5I/J). This suggests that silencing of SRRM1 and SRSF11 primarily affects alternative splicing events, and secondarily alters the expression of a specific subset of genes.

Enriched RBP motifs near the SRRM1 and SRSF11 regulated ASEs were identified using rMAPS2 (Supp Fig 6). The majority of the top 20 enriched motifs identified were common to both SRRM1 and SRSF11(Supp Table 1and 2). While SRRM1 does not have well defined target RNA binding motif, among the most highly enriched motifs identified in both the SRRM1 and SRSF11-regulated ASEs were the SRSF10 binding motifs A[AG]AG[AG][AG][AG] and AGAGA[AG][AG]. Of note these motifs are highly similar to the less well characterized SRSF11 binding motif AG[AG][AG][AG] providing further evidence that SRSF11 likely regulates several of the ASEs also regulated by SRRM1 (Fig 4C/D/H). RBPs corresponding to

19 of the most enriched motifs were also identified as SRRM1 and SRSF11 interactors (Fig 4H). These factors include members of the SR family of splicing factors (SRSF 1-3, 5-7, 9-10) as well as a core component of the EJC RBM8A, the mRNA processing factors PABPC4, PABPN1 and ZC3H1, the m6a reader IGF2BP2 and the oncogenic splicing factor TRA2B (Fig 4H). ASEs were also enriched for the epithelial splicing regulatory protein 1 (ESRP1) motif (Supp Fig 6A/C) in agreement with ESRP1 known role in NUMB E9 splicing [27].

SRRM1 and SRSF11 promote a set of cancer-associated splicing events.

In order to identify SRRM1 and/or SRSF11 regulated splicing events that may contribute to oncogenesis, we analyzed the list of differentially spliced genes (DSGs) for significant PSI changes in The Cancer Genome Atlas (TCGA) patient samples [35]. Of the 593 SRRM1-regulated DSGs identified, 388 exhibited significant splicing changes in COAD, 479 in BRCA, 413 in LUAD and 457 in LUSC patient samples (Supp Fig 7A). Similarly, of the 949 SRSF11 regulated DSGs, 642 exhibited significant PSI changes in COAD, 788 in BRCA, 633 in LUAD and 753 in LUSC patient samples (Supp Fig 8A). Next, we examined how specific shifts in exon PSI varied with SRRM1 silencing (Supp Fig 7B-E) or SRSF11 silencing (Supp Fig 8B-E) in comparison to exon-matched PSI shifts in COAD, LUAD and LUSC tumor and normal patient samples. We found that 40~41% of SRRM1 and 41~47% of SRSF11 exon-matched ASEs that exhibited reduced PSI with factor knockdown, had a higher PSI in tumors versus normal TCGA patient samples or vice versa (Supp Fig 7B-E, 8B-E).

Of the SRRM1 regulated splicing events that are also altered in COAD, BRCA, LUAD or LUSC patient samples, we identified a set of 22 events that are annotated in VastDB and predicted to impact protein isoform expression (Fig 4I). The set of splicing events impact genes

with known functions related to the hallmarks of cancer including proliferation, invasion and metastasis (Fig 5A). For example, the splicing of CD44 which has a well characterized role in tumour invasion and metastasis [36,37] has also previously been identified as an SRRM1 target [38]. We next examined a subset of these cancer associated splicing events in colon, lung and breast cancer cell lines with SRRM1 or SRSF11 siRNA knockdown (Supp Fig 9A/B). MKNK2 is a serine threonine kinase that undergoes alternative splicing of exon 14a to form the functionally distinct isoforms MKNK2A that acts as a tumor-suppressor by inducing p38-mediated cell death and MKNK2B which activates eIF4E without triggering the p38 MAPK stress response [39,40]. Analysis of TCGA RNA-seq data demonstrates that MKNK2 exon 14a PSI is significantly lower in COAD, LUAD and LUSC tumours (Supp Fig 9A). In line with these findings, we demonstrate that knockdown of SRRM1 or SRSF11 in DLD1, A549 and MDA-MB-468 cancer cell lines resulted in reduced MKNK2B isoform expression (Fig 5B and Supp Fig 10A).

SRRM1 knockdown also induced skipping of cassette exons in the T-box transcription factor (TBX3) as well as in the Forkhead box protein M1 (FOXM1) transcription factor. TBX3 is a transcriptional regulator linked to stem cell self-renewal and maintenance, and is a downstream effector of the Wnt signaling pathway [41]. Both SRRM1 and SRSF11 knockdown reduced TBX3 exon 2a inclusion in DLD1, HCT116, and A549 cell lines (Fig 5C, Supp Fig 10B). TBX3 exon 2a is located within the T-box DNA binding domain, and skipping of this exon has been shown to be important for differentiation of embryonic stem cells [42],while inclusion in cancer is associated with senescence inhibition [43]. In keeping with a tumour promoting role, exon 2a inclusion is significantly increased in BRCA, COAD, LUSC as well as BLCA and READ tumour types (Supp Fig 9A). Similarly, the Forkhead (FH) transcription factor, FOXM1 has been implicated in oncogenesis due to its elevated expression across several cancers as well as its

ability to promote proliferation, metastasis, stemness and drug resistance [44]. SRRM1 knockdown reduced the inclusion of FOXM1 exon 6 (also known as exon Va) in DLD1, HCT116 and A549 cancer cell lines (Fig 5C, Supp Fig 10C). FOXM1 exon 6 is located immediately adjacent to the FH DNA binding domain and contains a putative ERK1/2 phosphorylation site that renders the FOXM1c isoform MEK responsive [45]. Analysis of TCGA data showed that exon 6 inclusion was significant higher in COAD (Supp Fig 9A) and prior studies have shown increased inclusion of exon 6 in pancreatic and ovarian tumours [46,47].

SRRM1-regulated exon matched splicing events that demonstrated significant PSI shifts across TCGA samples also included the histone acetyltransferase KAT5 (also known as TIP60) and the splicing factor PTBP1. KAT5 can act as a tumor suppressor by acetylating and activating pro-apoptotic genes such as p53 [48,49]. The KAT5 exon 5 skipped isoform lacks the N-terminal domain required for chromatin binding and acts as a dominant negative inhibitor of the exon 5 included tumour suppressive isoform [50]. In patient samples, KAT5 exon 5 PSI values were significantly lower in COAD, LUAD, LUSC and several other tumour types (Supp Fig 9A). Consistent with these results, knockdown of SRRM1 resulted in increased KAT5 exon 5 inclusion in DLD1, HCT116 and A549 cell lines. Similarly, SRSF11 knockdown had the same effect of increased exon 5 inclusion in HCT116 and A549 cell lines (Fig 5C, Supp Fig 10D).

SRRM1 knockdown also downregulated exon 9 inclusion of the hnRNP family member PTBP1 in all cell lines tested (Fig 5C, Supp Fig 10E). Increased inclusion of exon 9 was observed in both LUAD and THCA patient samples compared to normal control (Supp Fig 9A). Whilst the role of PTBP1 exon 9 as it relates to oncogenic activity has not been elucidated, exon 9 is important for regulation of PTBP1 splicing activity during neural development [51]. Of note,

PTBP1 has been shown to promote oncogenic splicing of both NUMB and PKM2 in breast cancer [17,52], as well as promote the stemness in gastric cancer [53].

We also identified previously uncharacterized splicing events in the Rho GEF epithelial cell transforming sequence 2 (ECT2) and Rho effector DIAPH1 that are regulated by SRRM1[54-56]. Knockdown of SRRM1 resulted in reduced ECT2 exon 4 inclusion in DLD1, HCT116, A549 and MDA-MB-468 cancer cell lines (Supp Fig 10F, Supp Fig 11F). Similarly, knockdown of SRSF11 also significantly reduced ECT2 exon 4 inclusion in A549 and MDA-MB-468 cells (Supp Fig 10F, Supp Fig 11F). Though its functional role has not been characterized, ECT2 exon 4 PSI is significantly higher in both LUAD and LUSC cancer samples (Supp Fig 9A). SRRM1 knockdown also resulted in reduced DIAPH1 exon 2 inclusion across DLD1, HCT116, A549 and MDA-MB-468 cancer cell lines whilst SRSF11 knockdown resulted in reduced DIAPH1 exon 2 inclusion only in the DLD1 and MDA-MB-468 cancer cell lines (Supp Fig 10G, Supp Fig 11G). A significant increase in DIAPH1 exon 2 PSI is observed in COAD, LUSC, and HNSC tumours (Supp Fig. 9). Taken together, these results demonstrate that SRRM1 and SRSF11 promote the expression of cancer-associated isoforms of multiple genes linked to multiple hallmarks of cancer.

SRRM1 knockdown impairs cancer cell proliferation, colony formation and gene expression.

To characterize the phenotypic effects of SRRM1 knockdown, we generated inducible short hairpin RNA (shRNA) HCT116 and A549 cell lines (Supp Fig 12A/B). Knockdown of SRRM1 resulted in significantly reduced colony formation (Fig 6A) and proliferation (Supp Fig 12D/E). SRRM1 knockdown also resulted in reduced cell migration as well as invasive capacity

of A549 cell lines (Fig 6C/D). To assess how specific SRRM1 splicing events may contribute to the observed growth suppression in SRRM1 knockdown conditions, we considered the predicted downstream consequences of specific protein isoform expression. Silencing of SRRM1 results in altered splicing of MKNK2 and reduced expression of the MNK2B isoform which in turn activates phosphorylation of p38 and increases COX2 expression as part of the stress response pathway. Indeed, we observed enhanced phosphorylation of p38 and increased expression of p38 target gene COX2 in response to SRRM1 knockdown in DLD1, A549 and MDA-MB-468 cells compared to controls (Supp Fig 12F/G).

SRRM1 depletion also caused the altered expression of a small set of genes that are significantly enriched for E2F targets and G2M checkpoint associated genes (Supp Fig 5H,K; Supplementary data 3). In particular, SRRM1 knockdown in DLD1cells significantly reduced expression of genes with known functions in cell cycle progression (CCND1, RalGPS2), tumour progression and drug resistance (ALDHA1, PRDX2, PGK1), epithelial-mesenchymal transition (TMEM107, Lnc01816, DCST1-AS1), fatty acid metabolism (FADS2) and Wnt signaling (Notum, PRDX2, DDB2, CCND1, CDKN2B) (Fig 6E). Analysis of transcription factor target enrichment using ChEA3 [57] identified the SRRM1 splicing targets TBX3, FOXM1 and TCF7L2 amongst the most highly enriched for this differentially expressed gene set (Fig 6F). Analysis of selected target genes by RT-qPCR in DLD1, HCT116, A549 and MDA-MB-468 cells demonstrated reduced expression of PRDX2, DDB2, CCND1, NOTUM as well as METTL3 in response to knockdown of SRRM1 (Fig 6G). These data suggest that in addition to the oncogenic effects of directly regulating isoform expression of proteins such as NUMB, MKNK2 and CD44, SRRM1 can also promote a tumorigenic gene expression profile through altered splicing of transcription factors.

Discussion:

Detailed studies of the mechanisms and targets of dysregulated splicing can advance our understanding of tumourigenesis[4]. Here we used a CRISPR-Cas9 screen, based on a splicing event frequently dysregulated in cancer, to identify the SR protein family members, SRRM1 and SRSF11 as regulators of NUMB E9 splicing. We assessed the extent of functional overlap between SRRM1 and SRSF11 by comparing their shared interactomes and global effects on alternative splicing, as well as identify a set of coordinately regulated splicing events. SRRM1 functions as a splicing co-activator through association with other SR family members known to bind exonic splicing enhancers (ESEs). Indeed, SRRM1 and SRSF11 commonly regulated exons were enriched with binding motifs for an overlapping set of SR proteins including those that have been linked to cancer-associated splicing events including SRSF1, SRSF2, SRSF3, SRSF10 and TRA2B [4]. This set of proteins were also present in the SRRM1 and SRSF11 interactomes, along with SRRM2 and the U1 and U2 snRNP spliceosome components that together form a splicing co-activator complex [58].

The roles of SRSF11 and SRRM1 in cancer associated splicing are less well known. SRSF11 is a regulator of neuronal microexon splicing [59], and has also been implicated in CRC metastasis, by regulating the AS of HSPA12A and resultant increased N-cadherin expression [31]. Dysregulation of SRRM1 has been observed in several cancer contexts. In B-ALL, upregulation of SRRM1 is part of a high-risk gene signature and associated splicing program, that has a functional role in leukemia cell survival, proliferation and invasion [60]. Elevated SRRM1 expression has also been shown to contribute to tumor cell invasion in hepatocellular carcinoma [28], as well as increased metastasis and aggressiveness in prostate cancer [29]. Elevated plasma SRRM1 levels have been reported to distinguish high risk PCa patients from control individuals

suggesting potential utility as a biomarker [30]. In addition, SRRM1 was shown to promote expression of the oncogenic splice variant of CD44 in response to activated RAS signalling, suggesting that in addition to overexpression, oncogenic signaling may provide a mechanism to dysregulate SRRM1 activity [38]. Related to this, we conducted the screen that identified SRRM1 as a regulator of NUMB E9 splicing in DLD1 cells which express activated RAS^{G13D}. Coupled with the observation that activated RAS signaling also promotes NUMB E9 inclusion [17], these findings suggest a close link between SRRM1 regulated splicing events and activated RAS signalling in cancer.

In addition to promoting NUMB E9 inclusion, we identified an additional set of 24 SRRM1 and SRSF11 dependent cancer associated splicing events in genes that contribute to the cancer hallmarks. Indeed, the downregulation of SRRM1 resulted in reduced proliferation and colony formation of colon and lung cancer cell lines. We validated SRRM1 and SRSF11 regulation of specific splicing events including those in MNK2, TBX3, FOXM1, KAT5, ECT2 and DIAPH1, in DLD-1 and HCT116 colon cancer cells, as well as A549 lung cancer and MB-MDA468 breast cancer lines, supporting that notion that the observed splicing patterns may be common features of multiple tumour types. For example, among the SRRM1 and SRSF11 regulated ASEs that demonstrated oncogenic splicing patterns in COAD, LUAD and LUSC cancers was exon 14 in MKNK2, a serine threonine kinase that regulates translation through phosphorylation of eIF4E. MKNK2 undergoes alternative splicing of exon 14 to form functionally distinct isoforms MKNK2A and MKNK2B that act as a tumor-suppressor by phosphorylating p38 MAPK and inducing p38-mediated cell death, or activate eIF4E without triggering the p38 MAPK stress response respectively.

Whilst ECT2 and DIAPH1's role in oncogenesis and metastasis has been previously established [54,55,61,62], the potential functional impacts of the alternative splicing events we identified has not been characterized. In our analysis of data derived from TCGA, ECT2 exhibits significantly elevated exon 4 inclusion in tumor vs normal samples from LUAD and LUSC patients, whilst DIAPH1 exhibits significantly increased exon 2 inclusion in tumor vs normal samples from COAD and LUSC patients. Both ECT2 exon 4 and DIAPH1 exon 2 sequences are within regions of the protein with predicted functions in subcellular localization and protein-protein interactions, and further studies are required to determine protein isoform specific activities.

Depletion of SRRM1 also resulted in significant changes in expression of a small group of genes including the reduced expression of WNT pathway-associated genes such as NOTUM [63], PRDX2 [64,65], DDB2 [66] and CCND1 [67,68] and elevated expression of the tumor-suppressive WNT-inhibitor CDKN2B[69,70]. The effects of SRRM1 on WNT related gene expression may be a consequence of alternative splicing events we identified within the C-clamp DNA binding region of TCF7L2 (TCF4), which forms a complex with β-catenin important for WNT target gene expression [71-73]. Although we were unable to validate the specific splicing event regulated by SRRM1 due to the complexity of splicing in this region, these TCF7L2 isoforms are known to play differential roles sustaining oncogenic Wnt signaling [73].

We also found that SRRM1 regulates specific splicing events in the TBX3 and FOXM1 transcription factors, which both have roles in development, cancer, and are downstream effectors of WNT signaling [74,75]. In agreement with altered TBX3, FOXM1 and TCF7L2 transcriptional activation function, SRRM1 loss caused reduced expression of potential target genes with functions in proliferative and invasive capacity, EMT, as well as resistance to

apoptosis such as ALDH3A1[76], CHI3L1[77], FADS2 [78], METTL3 [34], PGK1 [79], and RFC4 [80]. As well, SRRM1 knockdown reduced expression of oncogenic lncRNAs UCA1 [81,82] and DCST1-AS1 [83] and elevated expression of tumor-suppressive TP53AIP1 [84]and ACTA2-AS1 [85].

A recurrent theme in oncogenic splicing is the reversion of alternative splicing programs back to embryonic patterns, resulting in the expression of protein isoforms that promote cellular phenotypes associated with tumor cells [86]. Our study suggests that SRRM1 promotes a set of exon inclusion events including NUMB E9, TBX3 exon 2a, ECT2 exon 4, and PTBP1 exon 9 [51,87,88] which are both elevated in tumor tissue [35] and in early development [89]. Indeed SRRM2, the main co-factor of SRRM1 has been shown to be vital in determining cell fate in early development and maintenance of cellular pluripotency [90]. As well, TBX3 exon2a inclusion for example, has been shown to be important for maintenance of ES cell self renewal [42]. Of note, multiple components of the core SAGA transcriptional co-activator complex, were identified as hits in our CRISPR screen, including SUPT7L, TAF5L, TAF6L, SUPT20H, and TADA1 [91], which regulates expression of RNA processing factors that activate an alternative splicing program during cellular reprogramming that regulates pluripotency [92].

Overall, this study has expanded the repertoire of known factors that regulate NUMB E9 splicing and further identifies the capacity of SRRM1 and SRSF11 to promote oncogenesis by coordinately regulating a program of oncogenic splicing events. This work also highlights SRRM1 as a potential therapeutic target, by linking its function to proliferation, EMT, apoptosis resistance, stemness and the regulation of WNT signaling.

Materials & Methods:

Construction of human Numb exon 9 dual fluorescent splicing reporters.

A dual fluorescent vector was constructed by cloning EGFP from a N1-EGFP vector into the N-terminal of a N2-mCherry vector such that EGFP and mCherry proteins are in alternate reading frames. To generate the splicing reporter, we utilized the human Numb genomic sequence chr14: 72778081-72952456 (NCBI build 36.1/hg18) obtained in a pBACCe3.6 vector (Empire Genomics, Buffalo, NY, USA; cat. #RP11-76J16) as previously described by our lab [17]. The human genomic region encompassing Numb exons 8/9, the initial fragment of exon 10, and introns 8-9/9-10 (5302 nucleotides total) were amplified by PCR and subcloned into the dual fluorescent vector.

To allow for expression of EGFP or mCherry fusions depending on the inclusion or skipping of exon 9 (E9), a frameshift mutation inserting two additional nucleotides was introduced into Numb E9 using site directed mutagenesis (TAGCACCCGTA**tt**GCAATGCCTGT; inserted nucleotides are bolded and underlined). To ensure that the frameshift mutation did not alter the splicing of the E9 reporter, we examined the effect of insertion of one (E9+1) or two (E9+2) additional nucleotides on E9 splicing when inserted into a 2HA-Numb E9 reporter [17] (Supp Fig 1 A, B). Insertion of two nucleotides (E9+2) did not alter Numb reporter splicing relative to the non-mutagenized version of the reporter, and was thus cloned into a pcDNA5/FRT/TRE-based plasmid compatible for insertion in Flp-In T-REx cells (Thermofisher: Cat: K650001). The pcDNA5 vectors were then used to generate stable dual fluorescent Flp-In HEK293 and DLD1 Numb E9 reporter cell lines.

Generation of stable Flp-In-DLD1/Flp-In-HEK reporter cell lines.

Tetracycline/doxycyline-inducible stable cell lines were generated by transfecting 1 μg of pcDNA5/FRT/TRE-based Numb E9 reporter plasmid and 500 ng of plasmid encoding pOG44 recombinase in HEK293 Flp-In or DLD1 Flp-In lines using Lipofectamine 3000 (Life Technologies) according to the manufacturer's instructions. Cell lines with stably integrated constructs were selected for and maintained with 15 μg/mL blasticidin and 200 μg/mL hygromycin. Transgene expression was induced by addition of 10 μg/mL tetracycline.

Cell culture.

Flp-In DLD1 colorectal adenocarcinoma cells were generously donated by the Gingras Lab. DLD1 cell lines were grown in RPMI 1640x medium (Wisent Bioproducts: Cat: 350-045-CL). HCT116 colorectal carcinoma cell line was grown in McCoy's 5A medium (WISENT, Cat: 317-010-CL). A549 and HEK293 cell lines were grown in DMEM (WISENT, Cat: 319-005-CL). MDA-MB-468 human breast cancer cell line was grown in Leibovitz L-15 medium (Gibco, Cat: 21083-027). All media were supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. Cell lines were authenticated based on STR profiling using services from The Centre for Applied Genomics in SickKids.

DNA transfection.

Cells were grown to 70-80% confluence in 6-well plates (Nunc) and were transfected with 1 μg of plasmid DNA 18-20 hours after seeding. Cells were transfected with plasmids expressing known Numb E9 splicing regulators previously established by our group [17] (SRSF1, PTBP1, constitutive active form of MEKK, MEKK-EE) or an empty control vector

(pcDNA3.1+) using Lipofectamine 2000 (Invitrogen) and Opti-MEM reduced serum medium (Life Technologies) according to the manufacturer's directions. Cells were maintained in serum-supplemented medium during the transfection process. 48-72 hours post-transfection, cells were lysed in the appropriate lysis buffer for either RNA or protein extraction and validation of Numb reporter splicing via immunoblots/RT-qPCR.

Immunoblots.

Cells were washed twice with PBS and lysed with 1X Laemmli Sample Buffer (63mM TrisHCl pH6.8, 2% SDS, 10% Glycerol). Cell lysates were collected into microfuge tubes, boiled for ten minutes, mixed for 10 minutes and centrifuged for 10 minutes at maximum speed. A BCA assay (Pierce) was used to determine the protein concentration in each lysate. A mixture of 2-mercaptoethanol (final concentration 5%) and bromophenol blue (final concentration 0.001%) was added to each lysate and 50 μg of protein were loaded onto 8-12 % gels for SDS-PAGE. Proteins were then transferred to PVDF membranes by wet transfer and blocked with 1% fish gelatin (Millipore Sigma, #G7765) in TBS. Membranes were incubated in primary antibody (listed below) overnight. The following day, membranes were washed three times in TBST for 10 minutes each, then incubated with HRP conjugated secondary antibodies for 1 hour followed by another set of TBST washes before development with ECL substrates. BioRad Gel Doc Imager was used to capture the signal after addition of ECL substrates. Membranes were stained with Coomassie blue stain for quantification of total protein expression. BioRad's Image Lab software was used to quantify band signals.

- Rabbit anti-EGFP (Invitrogen, Ref: A-11122) at 1:1000
- Rabbit anti-mCherry (Rockland, Ref: 600-401-p16) at 1:1000

- Rabbit anti-Gaussia luciferase (Thermofisher, Cat: PA1-181) at 1:1000
- Rabbit anti-Red Firefly luciferase (Thermofisher, Cat: PA1-179) at 1:1000
- Rabbit anti-SRRM1 (Invitrogen, Ref: PA5-62100) at 1:1000
- Mouse anti-Flag M2 (Sigma-Aldrich, Ref: F1804-200UG) at 1:1000
- Rabbit anti-phospho-p38 (Cell Signaling Technology, Ref: D3F9) at 1:1000
- Mouse anti-p38 (Santa Cruz, Ref: sc-7972) at 1:1000
- Mouse anti-GAPDH (Invitrogen, Ref: MA5-15738) at 1:10000

Splicing factor KO FACs test screen.

Five lentiCRISPRv2 vectors expressing guides targeting mCherry, SRSF1, SRSF3 and PTBP1 were obtained from Sigma Aldrich (gRNA sequence shown in Table 1). The lentiCRISPRv2 vector targeting AAVS1 (adeno-associated virus integration site 1) used as a negative control was generously provided by the Angers lab. The lentiCRISPRv2 vectors were used to examine whether knockdown of known Numb E9 splicing regulators was capable of influencing reporter EGFP/mCherry expression as examined by FACS. DLD1 Flp-In cells expressing the Numb E9 splicing reporter were transduced at an MOI of 0.3 in the presence of 8 mg/mL polybrene. 24 hours after infection, the medium was replaced with fresh medium containing puromycin (2.0 mg/mL) and cells were incubated for an additional 72 hours before passaging. 12 days post infection, cells were harvested and fixed in ice cold PBS buffer containing 1% PFA for 30 minutes and then washed twice with PBS. Cells were passed through a nylon mesh with a pore size of 40 mm to eliminate large aggregates and then sorted based on the relative EGFP:mCherry expression as indicated in (Fig 1C) using the Flow Cytometry core facility at the Peter Gilgan Centre for Research and Learning. Gates were generated to capture the top 5% mCherry high:EGFP low and EGFP high:mCherry low populations using the

negative control sg-AAVS1 cells and were maintained for all the other targeted genes to examine shifts in the dual fluorescent population in response to their KO.

Table 1: gRNA sequences used in splicing factor KO FACs test screen. Related to Figure 1C.

gRNA name	Sequence (PAM)
mCherry	AGTAGTCGGGGATGTCGGCG
PTBP1	GGGCCACGCACCGCCTGGTT
SRSF1	ATCGACCTCAAGAATCGCCG
RBFOX2	TTATCATGGATTCCCAGCGC
SRSF3	TTACCTACGAGACCTAGAGA
AAVS1	ATCCTGTCCCTAGTGGCCC

Virus library production and MOI determination.

For virus library production, 9 million HEK293T cells were seeded on 15 cm plates and transfected 24 hours later with a mix of 14 μg lentiviral pLCKO vector containing the library, 16μg packaging vector psPAX2, 1.56μg envelope vector pMD2.G, 93.6 μL X-treme Gene transfection reagent (Roche) and 1.4 mL Opti-MEM medium (Life Technologies). 24 hours post-transfection the medium was changed to serum-free, high BSA growth medium (DMEM, 1.1 g/100mL BSA, 1% penicillin/streptomycin). The virus-containing medium was harvested 48 and 72 hours after transfection, centrifuged at 1,500 rpm for 5 minutes and then frozen. DLD1 dual-fluorescent reporter cells were infected with a titration of the lentiviral gRNA library along with polybrene (8 μg/mL). After 24 hours the medium was replaced with medium containing puromycin (2 μg/mL) and cells were incubated for an additional 48 hours. Multiplicity of infection (MOI) of the titrated virus was determined 72 hours post-infection by comparing the percent survival of infected cells to a non-infection control.

Genome-wide CRISPR-Cas9 screen.

100 million DLD1 Flp-In cells expressing the dual fluorescence splicing reporter were infected with the lentiviral TKOv3 library (70,948 sgRNAs) at an MOI of 0.3 in the presence of 8 mg/mL polybrene to achieve a library coverage of ~300 cells/gRNA. 24 hours after infection, the medium was replaced with fresh medium containing Puromycin (2.0 mg/mL) and cells were incubated for an additional 72 before collecting a T0 unsorted cell population for technical analysis of the CRISPR screen. After selection, cells were split such that each sgRNA would be represented by an average of 300 cells in the population and passaged every three days. For cell sorting, 30 million cells were seeded and the following day 1μg/mL of tetracycline was added to induce dual fluorescent reporter expression. 24 hours later the cells were harvested and fixed with 1% PFA in ice cold PBS for 30 minutes and resuspended in sorting buffer (Hanks Balanced Salt Solution, 25mM HEPES pH 7.0, 1mM EDTA, 1% albumin) at a concentration of 10 million cells per millilitre. Cells were passed through a nylon mesh filter with a pore size of 40 mm to eliminate large aggregates. Filtered cells were sorted based on the relative EGFP:mCherry expression using Flow Cytometry core facilities at the Peter Gilgan Centre for Research and Learning using a Beckman Coulter MoFlo Astrios Cell Sorter. For each replicate, 900,000-1,100,00 cells with the highest or lowest 5% EGFP:mCherry ratio were collected. Additionally, 30 million cells were collected prior to sorting as a reference point for comparing enrichment in the sorted populations. The same procedure was followed for sorting of cells with the 20% highest or lowest EGFP:mCherry ratio.

Genomic DNA was extracted from the cell pellets of unsorted samples using the QIAamp Blood Maxi Kit (QIAGEN) while gDNA from sorted populations was purified with the Qiagen Midi Kit as per the manufacturer's recommendations. gDNA was precipitated using ethanol and

sodium chloride then resuspended in sterile water. gRNA inserts were amplified via PCR and prepared for sequencing using primers harboring Illumina TruSeq adapters with i5 and i7 barcodes. The resulting libraries were sequenced on an Illumina HiSeq2500 at the Donnelly Sequencing Centre. Each read was completed with standard primers for dual indexing with Rapid Run V1 reagents. The first 20 cycles of sequencing were "dark cycles", or base additions without imaging. The actual 26 base pair reads began after the dark cycles and contained two index reads with the i7 sequence read first, and the i5 sequence read second.

CRISPR screen analysis

Sequencing reads were mapped to TKOv3 library sgRNAs using the count function in the MAGeCK software (v0.5.8). Gene and sgRNA level enrichment in the mCherry high:EGFP low and EGFP high:mCherry low cell populations were calculated using the test function in MAGeCK in comparison to unsorted cells with control gRNAs targeting LacZ, EGFP and luciferase excluded during analysis. Genes over-represented at an FDR < 0.2, and a fold enrichment of at least 1.5x (Log fold enrichment > 0.585) with either gating regimen (5% or 20% highest or lowest mCherry:EGFP ratio) were considered as hits.

Gene ontology biological processes, molecular function and cellular component term enrichment analysis of all unique hits from all sorted populations (FDR < 0.2, LFC > 0.585 with gene expression > 1.0 TPM in DLD1 cells according to Depmap CCLE https://depmap.org/portal/ccle/) were uploaded to the online Database for Annotation, Visualization, and Integrated Discovery (DAVID, https://david.ncifcrf.gov/home.jsp), and clustered using an EASE of 1.0 and medium stringency. STRING interaction network of genes shown in (Fig 2D) identifying a potential regulatory splicing complex operating on Numb E9

alternative splicing was conducted using string database (https://string-db.org/). Only curated databases and experimentally determined evidence were considered. Nodes were colored based on biological process and molecular function go terms related to mRNA splicing, via spliceosome, RNA splicing, RNA binding and mRNA processing.

siRNA transfections.

DLD1, HCT116, A549, MDA-MB-468 and HEK293 cell lines were transfected with 25 nM of On-TARGETplus Smartpool siRNA (a mixture of four siRNA provided as a single reagent; Horizon Discovery) using Dharmafect 2 transfection reagent (Horizon Discovery) and Opti-MEM (Life Technologies) as per the manufacturer's instructions. A non-targeting siRNA pool was used as control. Transfected cells were harvested 48 hours or 72 hours post transfection for RNA collection using Qiagen's RNeasy RNA extraction kit. RNA samples were used for RT-qPCR and RT-PCR experiments.

RT-qPCR.

1μg of each RNA sample was reverse transcribed using the High-capacity cDNA Reverse Transcription kit (Applied Biosystems) and a prepared mixture of random oligomers and oligo-dT primers following the manufacturer's instructions. cDNA samples were diluted as required and 2 μL of dilute cDNA was added to 18 μL of the appropriate real-time reaction master mix in a 96-well clear PCR plate. Real-time reaction master mixes were prepared with PowerUp SYBR Green Master Mix (Applied Biosystems) with 12.5 μM of each forward and reverse primer according to the manufacturer's protocol. Reactions were set up in triplicate. RT-qPCR reactions were run on an Applied Biosystems Quantstudio3 thermal cycler, using cycling conditions

described by the manufacturer, and an annealing temperature of 60ºC. Reactions were normalized to GAPDH expression. A list of RT-qPCR primers are described in (Table 2).

Semi-quantitative PCR

1 µg of RNA was reverse-transcribed using the High-capacity cDNA Reverse Transcription kit (Applied Biosystems) and a prepared mixture of random oligomers and oligo-dT primers following the manufacturer's instructions. 1/20th of the cDNA reaction volume was used for PCR amplification with 12.5 µM of isoform-specific primers and 0.25 µL of Taq DNA polymerase (Roche) in a final reaction volume of 50 µL. PCR reactions were run on a BioRad thermocycler and had an initial denaturing step at 94ºC for 2 minutes, followed by multiple cycles of denaturing, annealing and 72ºC elongation, before a final elongation step for seven minutes at 72ºC. PCR products were then resolved on 1-3% agarose gels pre-stained with RedSafe dye (Chembio). A BioRad gel doc was used to capture images, and BioRad's Image Lab software was used to quantify band signal intensity. A list of primers used for semi-quantitative PCR can be found in (Table 3).

Table 2: List of primers used for quantitative PCR.

Target	Forward Sequence (5'→3')	Reverse Sequence (5'→3')
Numb +E9 Isoform	CCCCTGATGCTGCTAACAAG	CCTGGAAGAGACCTGGAGAG
Total Numb	ATGCAAGATGCCAAGAAAGC	AGGAGAGGTGGGAGAGGATG
GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
SRRM1	CAGCTGGAAGTGAAGAATCCAG	CCCGCGATGTTTTCTTGTGC
DDX23	GCGCATGGGTTGTGAGATTG	AACCACATAGGTACAGCGGC
SRSF11	GACTAATGTCTCCCCGAGCG	AAAGGCGAATCATCCGGCG

COX2	CCGAGGTGTATGTATGAGTGT	CTGTGTTTGGAGTGGGTTTC
PRDX2	CACCTGGCTTGGATCAACACC	CAGCACGCCGTAATCCTCAG
CCND1	GCTGCGAAGTGGAAACCATC	CCTCCTTCTGCACACATTTGAA
DDB2	ACCTCCGAGATTGTATTACGCC	TCACATCTTCTGCTAGGACCG
METTL3	CTATCTCCTGGCACTCGCAAGA	GCTTGAACCGTGCAACCACATC
NOTUM	TACCTGAAGGAGTCCAGGGG	CGTTCCACCAGTAGGGGTTC

Table 3: List of primers used for semi-quantitative PCR.

Target	**Forward Sequence (5'→3')**	**Reverse Sequence (5'→3')**
Endogenous E9 Numb isoforms	GAAGGGGAGGCAGAGAGC	GATGCTGGCTGGGAGATG
HA tagged Numb reporter	TCCAGTGTGGTGGAATTCTG	GATGCTGGCTGGGAGATG
b-actin	AAGATCAAGATCATTGCTCCTC	GGGTGTAACGCAACTAAGTC
Dual fluorescent Numb reporter	GAAGGGGAGGCAGAGAGC	GTTTACGTCGCCGTCCAG
KAT5-E5	CTAAGAACGGACTTCCTGGG	GTGAAACCACCTCCACCTTC
TBX3-E2a	GACCCCGAAATGCCAAAGAGG	GGGGCTGGTATTTGTGCATGG
ECT2-E4	GTGATATTGGTTCAAGAAGCTGG	CAAATTCTTCCACTGACTCCATC
DIAPH1-E2	GACCCGGGACAAGAAGAAG	ATGACTGTGCTGTGGGATCA
MKNK2B	GCTGCGACCTGTGGAGCCTGGG	GAGGAGGAAGTGACTGTCCCAC
Total MKNK2	CCGGACGCCAAGAAGAGG	CTTGACGGCGTACTCCTGG
FOXM1-E6	CGCCACAACCTTTCCCTG	GTAGCAGTGGCTTCATCTTCC
PTBP1-E9	CGTGACTACACACGCCCAGAC	GGTTGAGGTTGCTGACCAGC

Generation of flag-tagged SRRM1 and SRSF11 DLD1 and HEK293 cell lines

1 µg of pcDNA5/FRT/TRE-based plasmid encoding FLAG-tagged SRRM1 or SRSF11 was co-transfected with 500 ng of pOG44 recombinase plasmid in DLD1 or HEK293 Flp-In lines with Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. Parental Flp-In lines were maintained with media containing 15 µg/mL blasticidin and 100 µg/mL zeocin. To select for cell lines with stably integrated constructs, cells were grown in media containing 15 µg/mL blasticidin and 200 µg/mL hygromycin. Transgene expression was activated by supplementing media with 10 µg/mL of tetracycline.

FLAG Affinity Purification coupled with Mass Spectrometry (AP-MS) sample preparation

Flag-tagged SRRM1 or SRSF11 DLD1 or HEK293 cell lines were seeded onto two 15 cm plates and induced for 24 hours with 1 µg/mL tetracycline. Cell pellets were collected and lysed in ice cold TAP lysis buffer containing 50 mM HEPES-NaOH (pH 8.0), 100 mM NaCl, 2 mM EDTA, and 10% glycerol with freshly added 0.1% NP-40, 1 mM DTT, 1 mM PMSF, and a single COMPLETE protease inhibitor tablet (Roche Applied Science) at a 1:6 pellet weight-to-volume ratio. Resuspended pellets were frozen and thawed by placing the pellets on dry ice for five minutes and thawed in a 37°C water bath until a small ice pellet remained. Samples were quickly moved back onto ice and sonicated with three 10 second bursts with a 2 second rest at an amplitude of 35%. To solubilize chromatin and reduce the detection of interactions mediated by RNA or DNA, 1 µL of benzonase nuclease (Sigma-Aldrich, E8263, 250U) was added to each lysate and incubated for 30 minutes with rotation at 4°C. Lysates were cleared by centrifugation at 20,000 rcf for 20 minutes at 4°C then transferred to tubes containing 25 µL of 50% magnetic anti-FLAG M2 beads slurry (Sigma-Aldrich, M8823) pre-washed with lysis buffer. FLAG

immunoprecipitation was performed for 3 hours at 4°C with rotation. After incubation, beads were pelleted by centrifugation at 1,000 rpm for 1 minute and magnetized to aspirate the unbound lysate. The beads were then de-magnetized and washed with 1 mL of lysis buffer and the total volume (with beads) were transferred to a new tube. Beads were washed once more with 1 mL of lysis buffer then washed with 50 mM ammonium bicarbonate (ABC) at pH 8. All wash steps were performed on ice using cold lysis buffer and ABC.

All sample preparation, LC-MS/MS data collection, and MS data searches were carried out at the SPARC BioCentre at The Hospital for Sick Children (Toronto, ON, Canada). Samples were reduced with DTT (10 mM, 60°C, 1 hour), alkylated with iodoacetamide (20 mM, room temperature, 45 minutes, dark), and digested overnight at 37°C with trypsin (2 µg; Pierce). Peptides were dried by vacuum centrifugation, desalted on C18 ziptips (Millipore) using a DigestPro MSi (Intavis Bioanalytical Instruments), and dried again by vacuum centrifugation before resuspension in Buffer A (2% acetonitrile, 0.1% formic acid). Samples were analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS) using an EASY-nanoLC 1200 system with a 1 h analysis and an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific). The LC portion of the analysis consisted of a 18 minute linear gradient running 3-20% of Buffer A to Buffer B (0.1% FA, 80% acetonitrile), followed by a 31 minute linear gradient running 20-35% of Buffer A to Buffer B, a 2 minute ramp to 100% Buffer B and 9 minute hold at 100% Buffer B, all at a flow rate of 250 nL/minute. Samples were loaded into a 75 µm x 2 cm Acclaim PepMap 100 Pre-column followed by a 75 µm x 50 cm PepMax RSLC EASY-Spray analytical column filled with 2 µM C_{18} beads (Thermo Fisher Scientific). MS1 acquisition resolution was set to 120000 with an automatic gain control (AGC) target value of 4×10^5 and maximum ion injection time (IT) of 50 ms for a scan range of *m/z* 375-1500, with

dynamic exclusion set to 10 s. Isolation for MS2 scans was performed in the quadrupole with an isolation window of *m/z* 0.7. MS2 scans were performed in the ion trap with maximum ion IT of 10 ms, AGC target value of 1 x 10^4, and higher-energy collisional dissociation (HCD) activation with an NCE of 30.

Mass spectrometry data analysis

MS raw files were analyzed using PEAKS Studio software (Bioinformatics Solutions Inc.) and Proteome Discoverer (version 2.5.0.400) and fragment lists searched against the mouse UniProt Reference database (Uniprot_UP000000589 downloaded Sept. 15, 2020; 21966 entries). For both search algorithms, the parent and fragment mass tolerances were set to 10 ppm and 0.6 Da respectively, and only complete tryptic peptides with a maximum of three missed cleavages were accepted. Carbamidomethylation of cysteine was specified as a fixed modification. Deamidation of asparagine and glutamine, oxidation of methionine, acetylation of the protein N-terminus, and phosphorylation of serine, threonine, and tyrosine residues were specified as variable modifications. Membership in spliceosomal sub-complexes of prey proteins identified in AP-MS experiments was analyzed using human protein annotations in the Spliceosome Database ([93]). In (Fig 3A, Supp Fig 4A), only prey with a saint score > 0.95 for at least one of the baits are shown. The network map was generated using Cytoscape. Gene ontology biological processes term enrichment analysis of all unique SRRM1 or SRSF11 protein interactors (BFDR < 0.05) in DLD1 cells or (Saint score > 0.95) in HEK293 were uploaded to the online Database for Annotation, Visualization, and Integrated Discovery (DAVID, https://david.ncifcrf.gov/home.jsp), and clustered using an EASE of 1.0 and medium stringency. Data are available for download at MassIVE: MSV000100469

Bulk RNA-seq data analysis

Raw reads were mapped to the human genome using the hg19 reference genome with the STAR aligner (version 2.7.11a) [94]. PCR duplicates and low-mapping quality reads were removed using Picard tools and SAMtools. Differential expression analysis was performed using the R package DESeq2 (v.1.42) [95]. Only genes with a FDR adjusted P value < 0.05 and foldchange > 2 were considered statistically significant and graphed as a volcano plot using ggplot2 packages on R software. Differential splicing analysis was performed using rMATS-turbo (version 4.2.0) [96]. Splicing events supported by at least 15 unique reads with a ΔPSI > 10% and an FDR of < 0.05 were considered statistically significant. Statistical analysis was performed and plots were generated using R environment version 4.3.1. Sashimi plots were generated using ggsashimi.py script [97] with the parameter –min-coverage 5 and –aggr mean. Motif enrichment analysis was performed using rMAPS [98]. Significantly differently expressed genes with SRRM1 knockdown were uploaded to ChEA3 online tool (https://maayanlab.cloud/chea3/) [57] to predict transcription factor regulation. The transcription factors predicted to regulate these genes were cross referenced to the list of differentially spliced genes to identify specific transcription factors that may mediate the change in gene expression observed with SRRM1 knockdown (supplementary data).

Gene ontology biological processes term enrichment analysis of all unique genes that had alternatively spliced exons (ASEs) regulated by SRRM1 or SRSF11 were uploaded to the online Database for Annotation, Visualization, and Integrated Discovery (DAVID, https://david.ncifcrf.gov/home.jsp) and were clustered using an EASE of 1.0 and medium stringency. Gene set enrichment analysis (GSEA) of all differentially expressed genes in response to SRRM1 knockdown was conducted using GSEA software from the Broad institute

using a pre-ranked list (SIGN(logFC)*-log10(pvalue)), maximum term size (500), minimum term size (15) and 2000 permutations using MSigDB gmt file (h.all.v2024.1.Hs.symbols.gmt) obtained from (https://www.gsea-msigdb.org/gsea/index.jsp) [99,100]. RNAseq data are available at DOI: 10.5281/zenodo.18154912.

TCGA alternative splicing analysis

Our study utilized alternative splicing analysis that was conducted by [35] on RNA-sequencing (RNA-seq) samples from TCGA database. BRCA, COAD, LUAD and LUSC PSI values were extracted from Kahles et al., supplementary tables for all alternative splicing events that involved genes identified to be differentially spliced from our own DLD1 bulk RNA-seq analysis. A two-sided unpaired wilcoxon test of the difference of the PSI value between tumor and normal samples in BRCA, COAD, LUAD and LUSC was conducted. Specific alternative splicing events that demonstrated significant differences between tumor and normal samples (p-value < 0.05) were matched to the specific alternative splicing events from our bulk RNA-seq data analysis using the specific coordinates of the upstream, spliced and downstream exons using the GRCh73/hg19 human genome assembly. These steps were conducted using R environment version 4.4.1. GraphPad Prism software was used to graph the difference between PSI in tumor and normal samples (BRCA/COAD/LUAD/LUSC PSIΔ) versus the difference between PSI in the NTC and SRRM1/SRSF11 knockdown conditions (SRRM1si/SRSF11si PSIΔ) to determine similarity and differences between TCGA PSI changes and DLD1 RNA-seq PSI changes (Supp Fig 7/9).

For TCGA PSI analysis of MKNK2 E14a, ECT2 E4, DIAPH1 E2, KAT5 E5 and TBX3 E2a (Supp Fig 9A), the hg19 chromosomal coordinates of each splicing event used for extraction

is described in (Table 4). PSI values and sample barcodes were imported into R Studio version 4.41. Sample barcodes were broken down to identify the cancer (i.e., BRCA, COAD, LUAD etc.) and sample type (i.e., normal solid tissue, primary solid tumor, metastatic). The PSI values for each splicing event were compared between primary solid tumor samples and normal solid tissue samples within each cancer type. Only cancer types with five or more paired normal tissue samples were included in the analysis. A Wilcoxon rank-sum test was performed to determine if the difference in PSI for each splicing event between normal and tumor samples were statistically significant for each cancer type. Figures were created using GraphPad Prism.

Table 4: List of chromosomal coordinates used in TCGA PSI analysis.

Target	Exon Size	Chromosomal Coordinates
ECT2 Exon 4	93 base pairs	chr3:172473273:172473365
DIAPH1 Exon 2	27 base pairs	chr5:140967791:140967817
MKNK2 Exon 14a	226 base pairs	chr19:2039630:2039855
TBX3 Exon 2a	60 base pairs	chr12:115117718:115117777
KAT5 Exon 5	156 base pairs	chr11:65480819:65480974
FOXM1 Exon 6	45 base pairs	chr12:2974521:2974565
PTBP1 Exon 9	78 base pairs	chr19:805492:805569

Wound Healing Migration and Invasion Assays

For HCT116 wound healing assay, 80,000 cells (shNTC, shSRRM1 611 and shSRRM1 960) were seeded in 96-well incucyte imagelock plates (Sartorius) coated with 100 ug/mL of matrigel (Corning BD Cat. No. 354234). HCT116 cells were seeded in normal media. For A549 wound healing assays, cells were induced with 10 ug/mL of tetracycline for 72 hours prior to seeding. On day of seeding, 70,000 cells (shNTC, shSRRM1 611, and shSRRM1 960) were seeded in 96-well incucyte imagelock plates coated with matrigel. A549 cells were seeded in media supplemented with tetracycline. After 24 hours, the cell monolayer (at 90–100% of confluence) was wounded by scraping with the Incucyte® Wound Maker 96-Tool (Sartorius).

Cells were washed with medium to remove floating cells and replaced with medium supplemented with 10% FBS, 1% penicillin/streptomycin and 10 μg/ml tetracycline. Relative wound density was imaged and measured with an Incucyte® Live-Cell imager (Sartorius) every 4 hours post-wounding. The wound width was evaluated in a minimum of four wells using the Incucyte® Cell Migration Scratch Wound Analysis software. For the invasion assay, Matrigel (4 mg/mL, Corning BD Cat. No. 354234) was added after wounding and washing following Sartorius Incucyte guidelines. The mean of three independent experiments from shNTC, shSRRM1 611 and shSRRM1 960 cell lines were analyzed using GraphPad Prism by conducting a two-way ANOVA comparing the NTC to SRRM1 knockdown condition to determine the statistical significance of SRRM1 knockdown on cell migration and invasion.

Proliferation assays

For the HCT116 proliferation assay, 2,500 cells (shNTC, shSRRM1 611 shSRRM1 and 960 HCT116) were seeded in 96-well incucyte imagelock plates (Sartorius). After 24 hours, cells (at ~10% of confluence) had their medium replaced with fresh medium supplemented with 10% FBS, 1% penicillin/streptomycin and 10 μg/mL tetracycline. For the A549 proliferation assay, 2,500 cells (shNTC, shSRRM1 611, shSRRM1 960) were seeded in 96-well incucyte imagelock plates in medium supplemented with tetracycline. Phase area confluence was imaged and measured with an Incucyte® Live-Cell imager (Sartorius) every 6 hours. Cell confluence was evaluated in a minimum of eight wells using the Incucyte® AI confluence and basic fluorescence analysis software following Sartorius Incucyte guidelines. The mean of three independent experiments from shNTC, shSRRM1 611 and shSRRM1 960 cell lines were analyzed using

GraphPad Prism by conducting a two-way ANOVA comparing the NTC to SRRM1 knockdown condition to determine the statistical significance of SRRM1 knockdown on cell proliferation.

Colony formation assays.

675, 1250 and 2500 cells (shNTC, shSRRM1 611, shSRRM1 960 HCT116) were seeded in 6-well plates and cultured in standard conditions. Fresh medium was replaced every week. After 14 days, cells were fixed and stained with crystal violet (0.1%w/v). Plates were scanned to generate representative images. The crystal violet stain was then solubilized using 1% SDS and the signal was acquired by measuring absorbance at 590 nm. Statistical significance was determined using paired t-test between shNTC and shSRRM1 biological triplicates.

Acknowledgements: The authors thank the following: SPARC molecular analysis, the Sickkids Flow cytometry facility, Donnelly Sequencing Centre and Thomas Gonatopoulos-Pournatzis and Benjamen Blencowe for technical advice, Anne Claude Gingras for providing Flp-In DLD1 cell line. This work was supported by funding from the Canadian Institutes of Health Research (Project Grant FRN 106507) and the Canadian Cancer Society (Challenge Grant - 707475) and Natural Sciences and Engineering Council of Canada (RGPIN-2019-06485) to CJM. LV is a recipient of a Canada Graduate Scholarship – Masters. HF is a recipient of an Ontario Graduate Scholarship. CC and RB received an USRA awards from NSERC.

Author Contributions: KO performed experiments, acquired and analyzed data, drafted and critically reviewed the manuscript. LV, HF, JL, CDS, CC, RB AZ performed experiments, acquired and analyzed data. GM, SA analyzed and interpreted data. OS analyzed and interpreted

data and helped draft the manuscript. CJM supervised KO, LV, HF, CC, RB, AZ, analyzed and interpreted data, drafted and revised the manuscript.

Competing Interest Statement: The authors declare no conflict of interest.

References:

1 Smith, C. W. & Valcarcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci* **25**, 381-388, doi:10.1016/s0968-0004(00)01604-2 (2000).

2 Marasco, L. E. & Kornblihtt, A. R. The physiology of alternative splicing. *Nat Rev Mol Cell Biol* **24**, 242-254, doi:10.1038/s41580-022-00545-z (2023).

3 Ule, J. & Blencowe, B. J. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol Cell* **76**, 329-345, doi:10.1016/j.molcel.2019.09.017 (2019).

4 Bradley, R. K. & Anczukow, O. RNA splicing dysregulation and the hallmarks of cancer. *Nat Rev Cancer* **23**, 135-155, doi:10.1038/s41568-022-00541-7 (2023).

5 Rogalska, M. E., Vivori, C. & Valcarcel, J. Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. *Nat Rev Genet* **24**, 251-269, doi:10.1038/s41576-022-00556-8 (2023).

6 Li, Y. *et al.* A splicing factor switch controls hematopoietic lineage specification of pluripotent stem cells. *EMBO Rep* **22**, e50535, doi:10.15252/embr.202050535 (2021).

7 Bechara, E. G., Sebestyen, E., Bernardis, I., Eyras, E. & Valcarcel, J. RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol Cell* **52**, 720-733, doi:10.1016/j.molcel.2013.11.010 (2013).

8 Dooley, C. M., James, J., Jane McGlade, C. & Ahmad, I. Involvement of numb in vertebrate retinal development: evidence for multiple roles of numb in neural differentiation and maturation. *J Neurobiol* **54**, 313-325, doi:10.1002/neu.10176 (2003).

9 Dho, S. E., Othman, K., Zhang, Y. & McGlade, C. J. NUMB alternative splicing and isoform-specific functions in development and disease. *J Biol Chem* **301**, 108215, doi:10.1016/j.jbc.2025.108215 (2025).

10 Dho, S. E., French, M. B., Woods, S. A. & McGlade, C. J. Characterization of four mammalian numb protein isoforms. Identification of cytoplasmic and membrane-associated variants of the phosphotyrosine binding domain. *J Biol Chem* **274**, 33097-33104, doi:10.1074/jbc.274.46.33097 (1999).

11 Misquitta-Ali, C. M. *et al.* Global profiling and molecular characterization of alternative splicing events misregulated in lung cancer. *Mol Cell Biol* **31**, 138-150, doi:10.1128/MCB.00709-10 (2011).

12 McGill, M. A. & McGlade, C. J. Mammalian numb proteins promote Notch1 receptor ubiquitination and degradation of the Notch1 intracellular domain. *J Biol Chem* **278**, 23196-23203, doi:10.1074/jbc.M302827200 (2003).

13 McGill, M. A., Dho, S. E., Weinmaster, G. & McGlade, C. J. Numb regulates post-endocytic trafficking and degradation of Notch1. *J Biol Chem* **284**, 26427-26438, doi:10.1074/jbc.M109.014845 (2009).

14 Gulino, A., Di Marcotullio, L. & Screpanti, I. The multiple functions of Numb. *Exp Cell Res* **316**, 900-906, doi:10.1016/j.yexcr.2009.11.017 (2010).

15 Colaluca, I. N. *et al.* NUMB controls p53 tumour suppressor activity. *Nature* **451**, 76-80, doi:10.1038/nature06412 (2008).

16 Liu, X. *et al.* Numb positively regulates Hedgehog signaling at the ciliary pocket. *Nat Commun* **15**, 3365, doi:10.1038/s41467-024-47244-1 (2024).

17 Rajendran, D., Zhang, Y., Berry, D. M. & McGlade, C. J. Regulation of Numb isoform expression by activated ERK signaling. *Oncogene* **35**, 5202-5213, doi:10.1038/onc.2016.69 (2016).

18 Kim, K. K., Nam, J., Mukouyama, Y. S. & Kawamoto, S. Rbfox3-regulated alternative splicing of Numb promotes neuronal differentiation during development. *J Cell Biol* **200**, 443-458, doi:10.1083/jcb.201206146 (2013).
19 Zhang, Y. & McGlade, C. J. Pan tissue analysis of human Numb alternative splicing patterns. *BioRxiv* **doi: https://doi.org/10.1101/2023.09.26.559629** (2024).
20 Ke, H. *et al.* Loss of TDP43 inhibits progression of triple-negative breast cancer in coordination with SRSF3. *Proc Natl Acad Sci U S A* **115**, E3426-E3435, doi:10.1073/pnas.1714573115 (2018).
21 Tarn, W. Y. *et al.* RBM4 promotes neuronal differentiation and neurite outgrowth by modulating Numb isoform expression. *Mol Biol Cell* **27**, 1676-1683, doi:10.1091/mbc.E15-11-0798 (2016).
22 Hollander, D. *et al.* A network-based analysis of colon cancer splicing changes reveals a tumorigenesis-favoring regulatory pathway emanating from ELK1. *Genome Res* **26**, 541-553, doi:10.1101/gr.193169.115 (2016).
23 Zhang, Y. *et al.* Numb exon 9 inclusion regulates Integrinbeta5 surface expression and promotes breast cancer metastasis. *Oncogene* **41**, 2079-2094, doi:10.1038/s41388-022-02225-w (2022).
24 Hart, T. *et al.* High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**, 1515-1526, doi:10.1016/j.cell.2015.11.015 (2015).
25 Hart, T. & Moffat, J. BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics* **17**, 164, doi:10.1186/s12859-016-1015-8 (2016).
26 Hochberg-Laufer, H. *et al.* Availability of splicing factors in the nucleoplasm can regulate the release of mRNA from the gene after transcription. *PLoS Genet* **15**, e1008459, doi:10.1371/journal.pgen.1008459 (2019).
27 Xu, T. *et al.* Alternative splicing downstream of EMT enhances phenotypic plasticity and malignant behavior in colon cancer. *Elife* **11**, doi:10.7554/eLife.82006 (2022).
28 Song, X. & Ma, J. SRRM1 promotes the proliferation, migration, and invasion of hepatocellular carcinoma cells by regulating the JAK/STAT signaling pathway. *Tissue Cell* **79**, 101954, doi:10.1016/j.tice.2022.101954 (2022).
29 Jimenez-Vacas, J. M. *et al.* Dysregulation of the splicing machinery is directly associated to aggressiveness of prostate cancer. *EBioMedicine* **51**, 102547, doi:10.1016/j.ebiom.2019.11.008 (2020).
30 Montero-Hidalgo, A. J. *et al.* Clinical value of circulating splicing factors in prostate cancer: SRRM1 as a novel predictive biomarker and therapeutic target. *Mol Ther Oncol* **32**, 200910, doi:10.1016/j.omton.2024.200910 (2024).
31 Pan, Y. J. *et al.* Alternative splicing of HSPA12A pre-RNA by SRSF11 contributes to metastasis potential of colorectal cancer. *Clin Transl Med* **12**, e1113, doi:10.1002/ctm2.1113 (2022).
32 Eldridge, A. G., Li, Y., Sharp, P. A. & Blencowe, B. J. The SRm160/300 splicing coactivator is required for exon-enhancer function. *Proc Natl Acad Sci U S A* **96**, 6125-6130, doi:10.1073/pnas.96.11.6125 (1999).
33 Le Hir, H., Izaurralde, E., Maquat, L. E. & Moore, M. J. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J* **19**, 6860-6869, doi:10.1093/emboj/19.24.6860 (2000).

34 Sun, T., Wu, R. & Ming, L. The role of m6A RNA methylation in cancer. *Biomed Pharmacother* **112**, 108613, doi:10.1016/j.biopha.2019.108613 (2019).
35 Kahles, A. *et al.* Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell* **34**, 211-224 e216, doi:10.1016/j.ccell.2018.07.001 (2018).
36 Miletti-Gonzalez, K. E. *et al.* The CD44 receptor interacts with P-glycoprotein to promote cell migration and invasion in cancer. *Cancer Res* **65**, 6660-6667, doi:10.1158/0008-5472.CAN-04-3478 (2005).
37 Cho, S. H. *et al.* CD44 enhances the epithelial-mesenchymal transition in association with colon cancer invasion. *Int J Oncol* **41**, 211-218, doi:10.3892/ijo.2012.1453 (2012).
38 Cheng, C. & Sharp, P. A. Regulation of CD44 alternative splicing by SRm160 and its potential role in tumor cell invasion. *Mol Cell Biol* **26**, 362-370, doi:10.1128/MCB.26.1.362-370.2006 (2006).
39 Mogilevsky, M. *et al.* Modulation of MKNK2 alternative splicing by splice-switching oligonucleotides as a novel approach for glioblastoma treatment. *Nucleic Acids Res* **46**, 11396-11404, doi:10.1093/nar/gky921 (2018).
40 Maimon, A. *et al.* Mnk2 alternative splicing modulates the p38-MAPK pathway and impacts Ras-induced transformation. *Cell Rep* **7**, 501-513, doi:10.1016/j.celrep.2014.03.041 (2014).
41 Zimmerli, D. *et al.* TBX3 acts as tissue-specific component of the Wnt/beta-catenin transcriptional complex. *Elife* **9**, doi:10.7554/eLife.58123 (2020).
42 Wang, X. *et al.* hnRNPLL controls pluripotency exit of embryonic stem cells by modulating alternative splicing of Tbx3 and Bptf. *EMBO J* **40**, e104729, doi:10.15252/embj.2020104729 (2021).
43 Fan, W., Huang, X., Chen, C., Gray, J. & Huang, T. TBX3 and its isoform TBX3+2a are functionally distinctive in inhibition of senescence and are overexpressed in a subset of breast cancer cell lines. *Cancer Res* **64**, 5132-5139, doi:10.1158/0008-5472.CAN-04-0615 (2004).
44 Kalathil, D., John, S. & Nair, A. S. FOXM1 and Cancer: Faulty Cellular Signaling Derails Homeostasis. *Front Oncol* **10**, 626836, doi:10.3389/fonc.2020.626836 (2020).
45 Ma, R. Y. *et al.* Raf/MEK/MAPK signaling stimulates the nuclear translocation and transactivating activity of FOXM1c. *J Cell Sci* **118**, 795-806, doi:10.1242/jcs.01657 (2005).
46 Huang, C., Du, J. & Xie, K. FOXM1 and its oncogenic signaling in pancreatic cancer pathogenesis. *Biochim Biophys Acta* **1845**, 104-116, doi:10.1016/j.bbcan.2014.01.002 (2014).
47 Zhao, C. *et al.* Splicing Factor DDX23, Transcriptionally Activated by E2F1, Promotes Ovarian Cancer Progression by Regulating FOXM1. *Front Oncol* **11**, 749144, doi:10.3389/fonc.2021.749144 (2021).
48 ME, L. L. *et al.* New p53 related genes in human tumors: significant downregulation in colon and lung carcinomas. *Oncol Rep* **16**, 603-608, doi:10.3892/or.16.3.603 (2006).
49 Gorrini, C. *et al.* Tip60 is a haplo-insufficient tumour suppressor required for an oncogene-induced DNA damage response. *Nature* **448**, 1063-1067, doi:10.1038/nature06055 (2007).
50 Hamard, P. J. *et al.* PRMT5 Regulates DNA Repair by Controlling the Alternative Splicing of Histone-Modifying Enzymes. *Cell Rep* **24**, 2643-2657, doi:10.1016/j.celrep.2018.08.002 (2018).

51 Gueroussov, S. *et al.* An alternative splicing event amplifies evolutionary differences between vertebrates. *Science* **349**, 868-873, doi:10.1126/science.aaa8381 (2015).
52 He, X. *et al.* Involvement of polypyrimidine tract-binding protein (PTBP1) in maintaining breast cancer cell growth and malignant properties. *Oncogenesis* **3**, e84, doi:10.1038/oncsis.2013.47 (2014).
53 Ni, T. *et al.* PTBP1 drives c-Myc-dependent gastric cancer progression and stemness. *Br J Cancer* **128**, 1005-1018, doi:10.1038/s41416-022-02118-5 (2023).
54 Fields, A. P. & Justilien, V. The guanine nucleotide exchange factor (GEF) Ect2 is an oncogene in human cancer. *Adv Enzyme Regul* **50**, 190-200, doi:10.1016/j.advenzreg.2009.10.010 (2010).
55 Cook, D. R. *et al.* Aberrant Expression and Subcellular Localization of ECT2 Drives Colorectal Cancer Progression and Growth. *Cancer Res* **82**, 90-104, doi:10.1158/0008-5472.CAN-20-4218 (2022).
56 Wang, H. K., Liang, J. F., Zheng, H. X. & Xiao, H. Expression and prognostic significance of ECT2 in invasive breast cancer. *J Clin Pathol* **71**, 442-445, doi:10.1136/jclinpath-2017-204569 (2018).
57 Keenan, A. B. *et al.* ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res* **47**, W212-W224, doi:10.1093/nar/gkz446 (2019).
58 Blencowe, B. J. *et al.* The SRm160/300 splicing coactivator subunits. *RNA* **6**, 111-120, doi:10.1017/s1355838200991982 (2000).
59 Gonatopoulos-Pournatzis, T. *et al.* Genome-wide CRISPR-Cas9 Interrogation of Splicing Networks Reveals a Mechanism for Recognition of Autism-Misregulated Neuronal Microexons. *Mol Cell* **72**, 510-524 e512, doi:10.1016/j.molcel.2018.10.008 (2018).
60 Closa, A. *et al.* A convergent malignant phenotype in B-cell acute lymphoblastic leukemia involving the splicing factor SRRM1. *NAR Cancer* **4**, zcac041, doi:10.1093/narcan/zcac041 (2022).
61 Lin, Y. N. *et al.* Expression of DIAPH1 is up-regulated in colorectal cancer and its down-regulation strongly reduces the metastatic capacity of colon carcinoma cells. *Int J Cancer* **134**, 1571-1582, doi:10.1002/ijc.28486 (2014).
62 Narumiya, S., Tanji, M. & Ishizaki, T. Rho signaling, ROCK and mDia1, in transformation, metastasis and invasion. *Cancer Metastasis Rev* **28**, 65-76, doi:10.1007/s10555-008-9170-7 (2009).
63 Kakugawa, S. *et al.* Notum deacylates Wnt proteins to suppress signalling activity. *Nature* **519**, 187-192, doi:10.1038/nature14259 (2015).
64 Feng, A. L. *et al.* PRDX2 plays an oncogenic role in esophageal squamous cell carcinoma via Wnt/beta-catenin and AKT pathways. *Clin Transl Oncol* **22**, 1838-1848, doi:10.1007/s12094-020-02323-9 (2020).
65 Wang, W. *et al.* PRDX2 promotes the proliferation of colorectal cancer cells by increasing the ubiquitinated degradation of p53. *Cell Death Dis* **12**, 605, doi:10.1038/s41419-021-03888-1 (2021).
66 Gilson, P. *et al.* Emerging Roles of DDB2 in Cancer. *Int J Mol Sci* **20**, doi:10.3390/ijms20205168 (2019).
67 Rohrs, S. *et al.* Chronological expression of Wnt target genes Ccnd1, Myc, Cdkn1a, Tfrc, Plf1 and Ramp3. *Cell Biol Int* **33**, 501-508, doi:10.1016/j.cellbi.2009.01.016 (2009).
68 Sommer, G. *et al.* The RNA-binding protein La contributes to cell proliferation and CCND1 expression. *Oncogene* **30**, 434-444, doi:10.1038/onc.2010.425 (2011).

69 Krimpenfort, P. *et al.* A natural WNT signaling variant potently synergizes with Cdkn2ab loss in skin carcinogenesis. *Nat Commun* **10**, 1425, doi:10.1038/s41467-019-09321-8 (2019).
70 Li, B., Li, A., You, Z., Xu, J. & Zhu, S. Epigenetic silencing of CDKN1A and CDKN2B by SNHG1 promotes the cell cycle, migration and epithelial-mesenchymal transition progression of hepatocellular carcinoma. *Cell Death Dis* **11**, 823, doi:10.1038/s41419-020-03031-6 (2020).
71 Nusse, R. & Clevers, H. Wnt/beta-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* **169**, 985-999, doi:10.1016/j.cell.2017.05.016 (2017).
72 Mao, C. D. & Byers, S. W. Cell-context dependent TCF/LEF expression and function: alternative tales of repression, de-repression and activation potentials. *Crit Rev Eukaryot Gene Expr* **21**, 207-236, doi:10.1615/critreveukargeneexpr.v21.i3.10 (2011).
73 Weise, A. *et al.* Alternative splicing of Tcf7l2 transcripts generates protein variants with differential promoter-binding and transcriptional activation properties at Wnt/beta-catenin targets. *Nucleic Acids Res* **38**, 1964-1981, doi:10.1093/nar/gkp1197 (2010).
74 Chen, Y. *et al.* Wnt-induced deubiquitination FoxM1 ensures nucleus beta-catenin transactivation. *EMBO J* **35**, 668-684, doi:10.15252/embj.201592810 (2016).
75 Jauregi-Miguel, A. *et al.* The developmental factor TBX3 engages with the Wnt/beta-catenin transcriptional complex in colorectal cancer to regulate metastasis genes. *Proc Natl Acad Sci U S A* **122**, e2419691122, doi:10.1073/pnas.2419691122 (2025).
76 Terzuoli, E. *et al.* ALDH3A1 Overexpression in Melanoma and Lung Tumors Drives Cancer Stem Cell Expansion, Impairing Immune Surveillance through Enhanced PD-L1 Output. *Cancers (Basel)* **11**, doi:10.3390/cancers11121963 (2019).
77 Qiu, Q. C. *et al.* CHI3L1 promotes tumor progression by activating TGF-beta signaling pathway in hepatocellular carcinoma. *Sci Rep* **8**, 15029, doi:10.1038/s41598-018-33239-8 (2018).
78 Li, Y. L., Tian, H., Jiang, J., Zhang, Y. & Qi, X. W. Multifaceted regulation and functions of fatty acid desaturase 2 in human cancers. *Am J Cancer Res* **10**, 4098-4111 (2020).
79 He, Y. *et al.* PGK1-mediated cancer progression and drug resistance. *Am J Cancer Res* **9**, 2280-2302 (2019).
80 Liu, L. *et al.* An RFC4/Notch1 signaling feedback loop promotes NSCLC metastasis and stemness. *Nat Commun* **12**, 2693, doi:10.1038/s41467-021-22971-x (2021).
81 Han, Y. *et al.* UCA1, a long non-coding RNA up-regulated in colorectal cancer influences cell proliferation, apoptosis and cell cycle distribution. *Pathology* **46**, 396-401, doi:10.1097/PAT.0000000000000125 (2014).
82 Wang, C. J. *et al.* The lncRNA UCA1 promotes proliferation, migration, immune escape and inhibits apoptosis in gastric cancer by sponging anti-tumor miRNAs. *Mol Cancer* **18**, 115, doi:10.1186/s12943-019-1032-0 (2019).
83 Tang, L. *et al.* Long Noncoding RNA DCST1-AS1 Promotes Cell Proliferation and Metastasis in Triple-negative Breast Cancer by Forming a Positive Regulatory Loop with miR-873-5p and MYC. *J Cancer* **11**, 311-323, doi:10.7150/jca.33982 (2020).
84 Liang, Y., Wang, S. & Liu, J. Overexpression of Tumor Protein p53-regulated Apoptosis-inducing Protein 1 Regulates Proliferation and Apoptosis of Breast Cancer Cells through the PI3K/Akt Pathway. *J Breast Cancer* **22**, 172-184, doi:10.4048/jbc.2019.22.e21 (2019).

85 Pan, Q., Huang, Y., Wang, Y., Li, D. & Lei, C. LncRNA ACTA2-AS1 suppress colon adenocarcinoma progression by sponging miR-4428 upregulation BCL2L11. *Cancer Cell Int* **21**, 203, doi:10.1186/s12935-021-01769-3 (2021).
86 Bonnal, S. C., Lopez-Oreja, I. & Valcarcel, J. Roles and mechanisms of alternative splicing in cancer - implications for care. *Nat Rev Clin Oncol* **17**, 457-474, doi:10.1038/s41571-020-0350-x (2020).
87 Dong, L., Lyu, X., Faleti, O. D. & He, M. L. The special stemness functions of Tbx3 in stem cells and cancer development. *Semin Cancer Biol* **57**, 105-110, doi:10.1016/j.semcancer.2018.09.010 (2019).
88 Cook, D. R. *et al.* The ect2 rho Guanine nucleotide exchange factor is essential for early mouse development and normal cell cytokinesis and migration. *Genes Cancer* **2**, 932-942, doi:10.1177/1947601912437035 (2011).
89 Mazin, P. V., Khaitovich, P., Cardoso-Moreira, M. & Kaessmann, H. Alternative splicing during mammalian organ development. *Nat Genet* **53**, 925-934, doi:10.1038/s41588-021-00851-w (2021).
90 Carvalho, S. *et al.* SRRM2 splicing factor modulates cell fate in early development. *Biol Open* **13**, doi:10.1242/bio.060415 (2024).
91 Chen, Y. C. & Dent, S. Y. R. Conservation and diversity of the eukaryotic SAGA coactivator complex across kingdoms. *Epigenetics Chromatin* **14**, 26, doi:10.1186/s13072-021-00402-x (2021).
92 Hirsch, C. L. *et al.* Myc and SAGA rewire an alternative splicing network during early somatic cell reprogramming. *Genes Dev* **29**, 803-816, doi:10.1101/gad.255109.114 (2015).
93 Cvitkovic, I. & Jurica, M. S. Spliceosome database: a tool for tracking components of the spliceosome. *Nucleic Acids Res* **41**, D132-141, doi:10.1093/nar/gks999 (2013).
94 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
95 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
96 Wang, Y. *et al.* rMATS-turbo: an efficient and flexible computational tool for alternative splicing analysis of large-scale RNA-seq data. *Nat Protoc* **19**, 1083-1104, doi:10.1038/s41596-023-00944-2 (2024).
97 Garrido-Martin, D., Palumbo, E., Guigo, R. & Breschi, A. ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. *PLoS Comput Biol* **14**, e1006360, doi:10.1371/journal.pcbi.1006360 (2018).
98 Park, J. W., Jung, S., Rouchka, E. C., Tseng, Y. T. & Xing, Y. rMAPS: RNA map analysis and plotting server for alternative exon regulation. *Nucleic Acids Res* **44**, W333-338, doi:10.1093/nar/gkw410 (2016).
99 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
100 Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273, doi:10.1038/ng1180 (2003).

Figures:

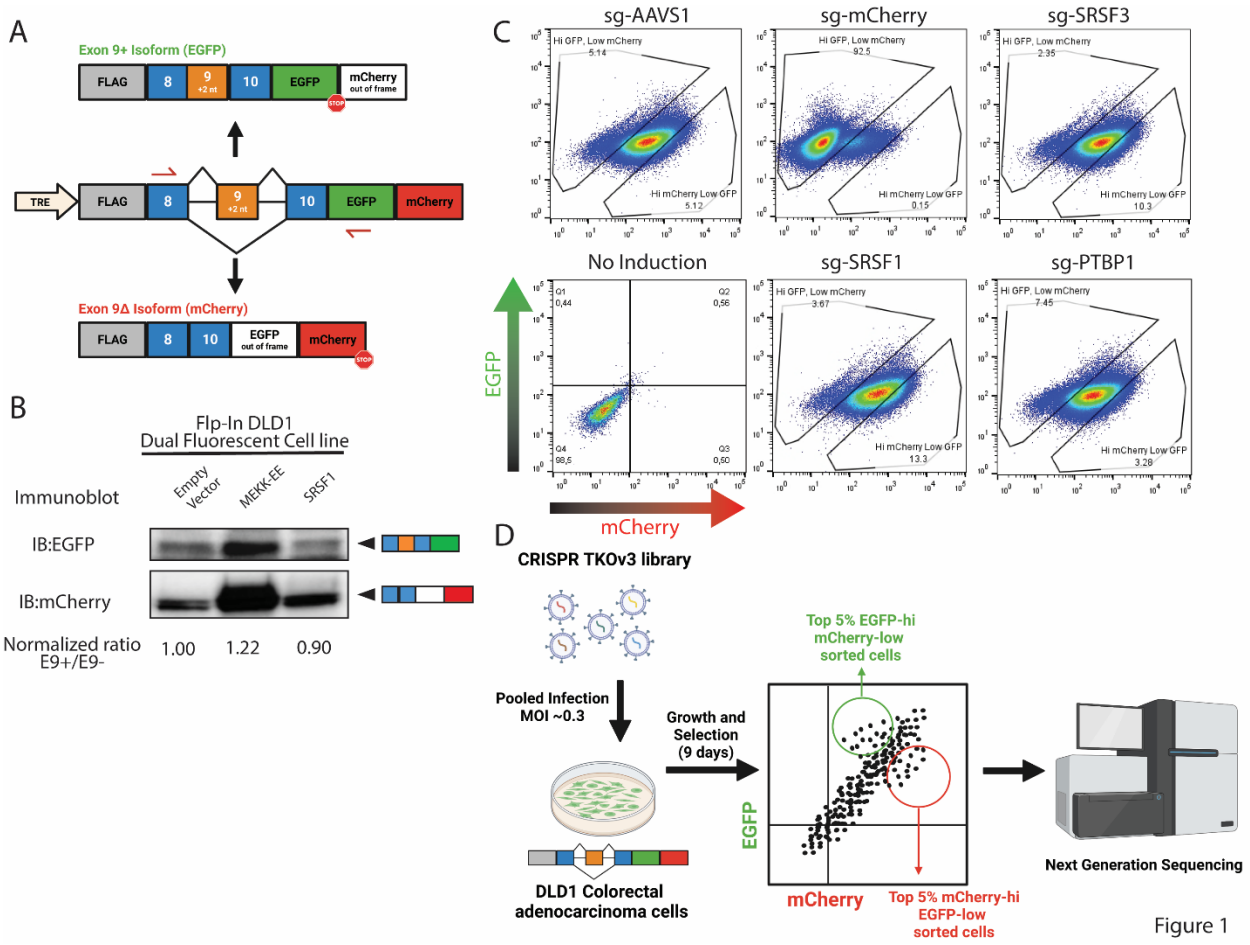


Figure 1: Dual fluorescent Numb splicing reporter based CRISPR knockout screen.

(A) Schematic diagram of the dual fluorescent Numb E9 Splicing reporter. The reporter includes Numb exon 8, E9 (including a +2 bp frameshift) with E9 flanking introns (2934-nt upstream and 1987-nt downstream) and 88bps of exon 10. Mutations were introduced into exon such that the inclusion or skipping of E9 results in EGFP or mCherry expression respectively.

(B) Immunoblots of lysates from E9 reporter expressing Flp-in DLD1 cells, transfected with MEKK-EE, SRSF1 or empty vector and probed with either EGFP a or mCherry antibodies. Numbers under blots denote ratio of relative band intensity quantifications utilizing ImageJ imaging software of the inclusion (EGFP)/exclusion (mCherry) product normalized to empty vector control.

(C) FACS analysis of Flp-in DLD1 dual fluorescent reporter cell line following 24hr induction and transduced with lentiviral expression cassettes for Cas9 and sgRNAs targeting known RBP regulators of Numb E9. Gates were generated to capture the Top 5% EGFP high:mCherry low and EGFP low:mCherry high populations on the negative control sg-AAVS1 cells and were maintained for all the other targeted gene conditions.

(D) Overview of genome-wide CRISPR knockout screen using the TKOv3 sgRNA library to identify novel pathways and factors regulating Numb E9 splicing. Dual fluorescent flp-in DLD1 Numb E9 reporter cell lines were transduced with a human CRISPR lentiviral library that has 70,948 guides (4 guides/gene) targeting 18,053 protein-coding genes at ~0.3 MOI and subjected to puromycin selection for 9 days before PFA fixation and fluorescent activated cell sorting (FACs). Genomic DNA from the top 5%/20% EGFP high:mCherry low and mCherry-high:EGFP-low populations sorted populations was collected and sequenced via NGS.

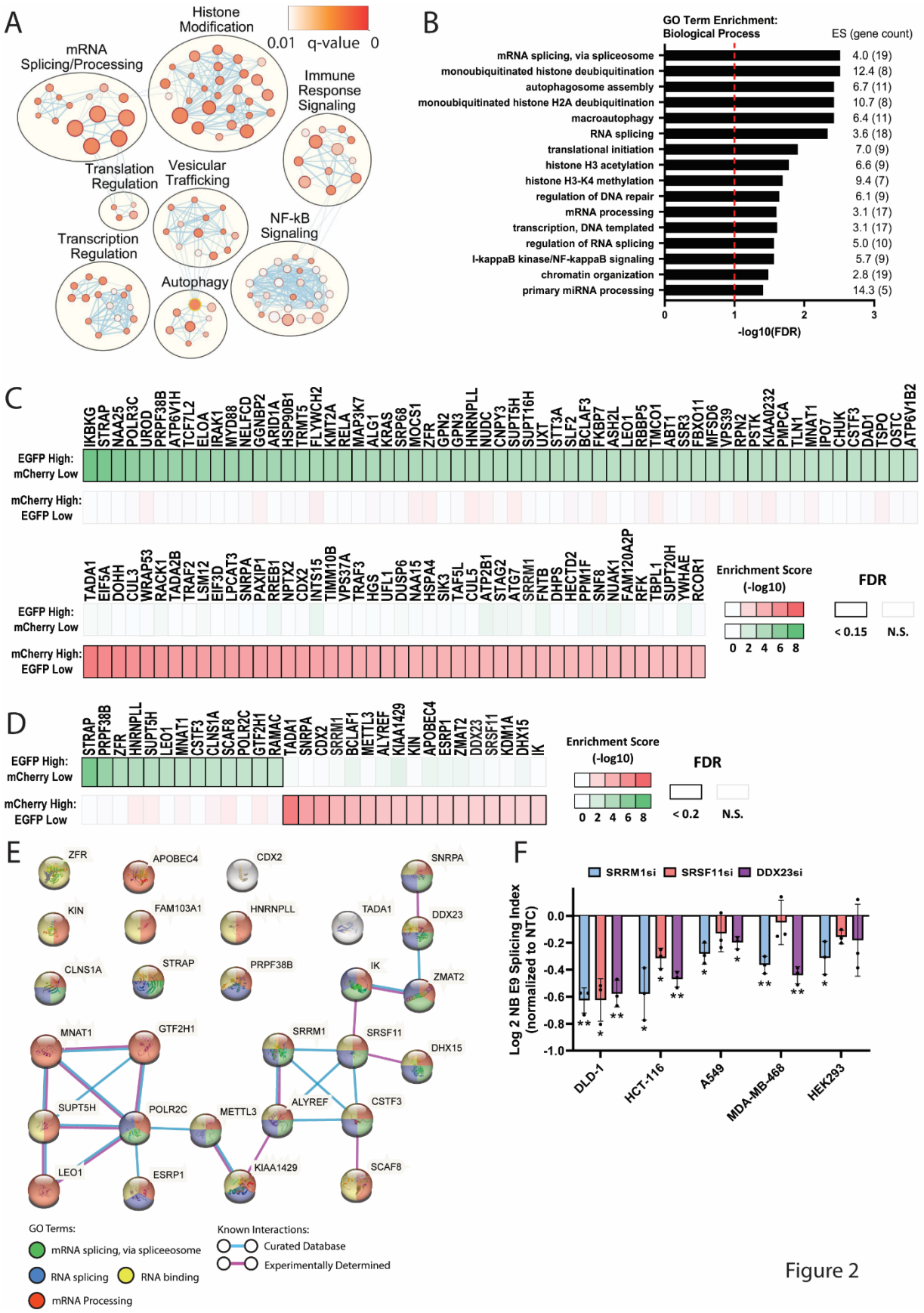

A
Histone Modification
0.01 q-value 0
mRNA Splicing/Processing
Immune Response Signaling
Translation Regulation
Vesicular Trafficking
NF-kB Signaling
Transcription Regulation
Autophagy
B
GO Term Enrichment: Biological Process
ES (gene count)
mRNA splicing, via spliceosome 4.0 (19)
monoubiquitinated histone deubiquitination 12.4 (8)
autophagosome assembly 6.7 (11)
monoubiquitinated histone H2A deubiquitination 10.7 (8)
macroautophagy 6.4 (11)
RNA splicing 3.6 (18)
translational initiation 7.0 (9)
histone H3 acetylation 6.6 (9)
histone H3-K4 methylation 9.4 (7)
regulation of DNA repair 6.1 (9)
mRNA processing 3.1 (17)
transcription, DNA templated 3.1 (17)
regulation of RNA splicing 5.0 (10)
I-kappaB kinase/NF-kappaB signaling 5.7 (9)
chromatin organization 2.8 (19)
primary miRNA processing 14.3 (5)
0 1 2 3
-log10(FDR)
C
IKBKG STRAP NAA25 POLR3C UROD PRPF38B ATP6V1H TCF7L2 ELOA IRAK1 MYD88 NELFCD GGNBP2 ARID1A HSP90B1 TRMT5 FLYWCH2 KMT2A RELA MAP3K7 ALG1 KRAS SRP68 MOCS1 ZFR GPN2 GPN3 HNRNPLL NUDC CNPY3 SUPT5H SUPT16H UXT STT3A SLF2 BCLAF3 FKBP7 ASH2L LEO1 RBBP5 TMCO1 ABT1 SSR3 FBXO11 MFSD6 VPS39 RPN2 PSTK KIAA0232 PMPCA TLN1 MNAT1 IPO7 CHUK CSTF3 DAD1 TSPO OSTC ATP6V1B2
EGFP High: mCherry Low
mCherry High: EGFP Low
TADA1 EIF5A DOHH CUL3 WRAP53 RACK1 TADA2B TRAF2 LSM12 EIF3D LPCAT3 SNRPA PAXIP1 RREB1 NPTX2 CDX2 INTS15 TIMM10B VPS37A TRAF3 HGS UFL1 DUSP6 NAA15 HSPA4 SIK3 TAF5L CUL5 ATP2B1 STAG2 ATG7 SRRM1 FNTB DHPS HECTD2 PPM1F SNF8 NUAK1 FAM120A2P RFK TBPL1 SUPT20H YWHAE RCOR1
Enrichment Score (-log10)
0 2 4 6 8
FDR
< 0.15 N.S.
D
STRAP PRPF38B ZFR HNRNPLL SUPT5H LEO1 MNAT1 CSTF3 CLNS1A SCAF8 POLR2C GTF2H1 RAMAC TADA1 SNRPA CDX2 SRRM1 BCLAF1 METTL3 ALYREF KIAA1429 KIN APOBEC4 ESRP1 ZMAT2 DDX23 SRSF11 KDM1A DHX15 IK
EGFP High: mCherry Low
mCherry High: EGFP Low
Enrichment Score (-log10)
0 2 4 6 8
FDR
< 0.2 N.S.
E
ZFR APOBEC4 CDX2 SNRPA
KIN FAM103A1 HNRNPLL TADA1 DDX23
CLNS1A STRAP PRPF38B IK ZMAT2
MNAT1 GTF2H1 SRRM1 SRSF11 DHX15
SUPT5H POLR2C METTL3 ALYREF CSTF3
LEO1 ESRP1 KIAA1429 SCAF8
GO Terms:
mRNA splicing, via spliceeosome
RNA splicing
RNA binding
mRNA Processing
Known Interactions:
Curated Database
Experimentally Determined
F
SRRM1si SRSF11si DDX23si
Log 2 NB E9 Splicing Index (normalized to NTC)
0.0 -0.2 -0.4 -0.6 -0.8 -1.0
DLD-1 HCT-116 A549 MDA-MB-468 HEK293

Figure 2

Figure 2: Identification of Numb exon 9 splicing reporter regulators.

(A) Enrichment map of GO biological processes and reactome terms from the CRISPR screen hits (All unique hits from 5% and 20% sorted populations FDR <0.2, LFC >0.585). Only genes with an expression >1.0 TPM in DLD1 cells were considered. Node size represents the number of genes belonging to individual terms; color of node represents significance of the gene set (only q value < 0.01 gene sets are shown), edge thickness denotes shared interactors between terms.

(B) Biological process GO term enrichment analysis for CRISPR screen hits. The same hit list as in **(A)** was uploaded to the online Database for Annotation, Visualization and Integrated Discovery, DAVID, and clustered using an EASE of 1.0 and medium stringency. The top enriched terms are arranged in order of significance (-log10(FDR) with Enrichment score (ES) values and total hit gene count for each term in between brackets denoted on the right. Only terms with an FDR score < 0.05 are shown. The red dotted line indicates an FDR of 0.1.

(C) Relative effects of identified CRISPR screen hits detected by sgRNA enrichment in the 5% sorted cell populations. Genes identified in the 5% EGFP high/mCherry low sorted population (green) and the 5% mCherry high/EGFP low population (red) from DLD1 splicing reporter expressing cells with an FDR < 0.15, a log fold enrichment (LFC) > 0.585 and gene expression >1.0 TPM in DLD1 cells are shown.

(D) List of CRISPR screen hits identified with a previously established link to splicing regulation with an FDR <0.2, LFC > 0.585 and TPM >1 in DLD1 cells from the 5% sorted populations.

(E) STRING interaction network (https://string-db.org/) of genes shown in **(D)** identifying a potential regulatory splicing complex operating on Numb E9 alternative splicing

(F) Examination of endogenous Numb E9 splicing index (log2 fold change expression of the Numb+E9 transcript relative to total Numb transcript expression, measured via RT-qPCR.) in DLD1, HCT-116, A549, MDA-MB-468 and HEK293 cells following transfection with siRNAs targeting SRRM1, SRSF11 or DDX23. (n = 3, mean Log2SI ± s.e.m.) All results depicted are relative to non-target control siRNA and significant differences with a p-values < 0.05 indicated with * and p-value < 0.01 with ** according to paired t-test statistical analysis.

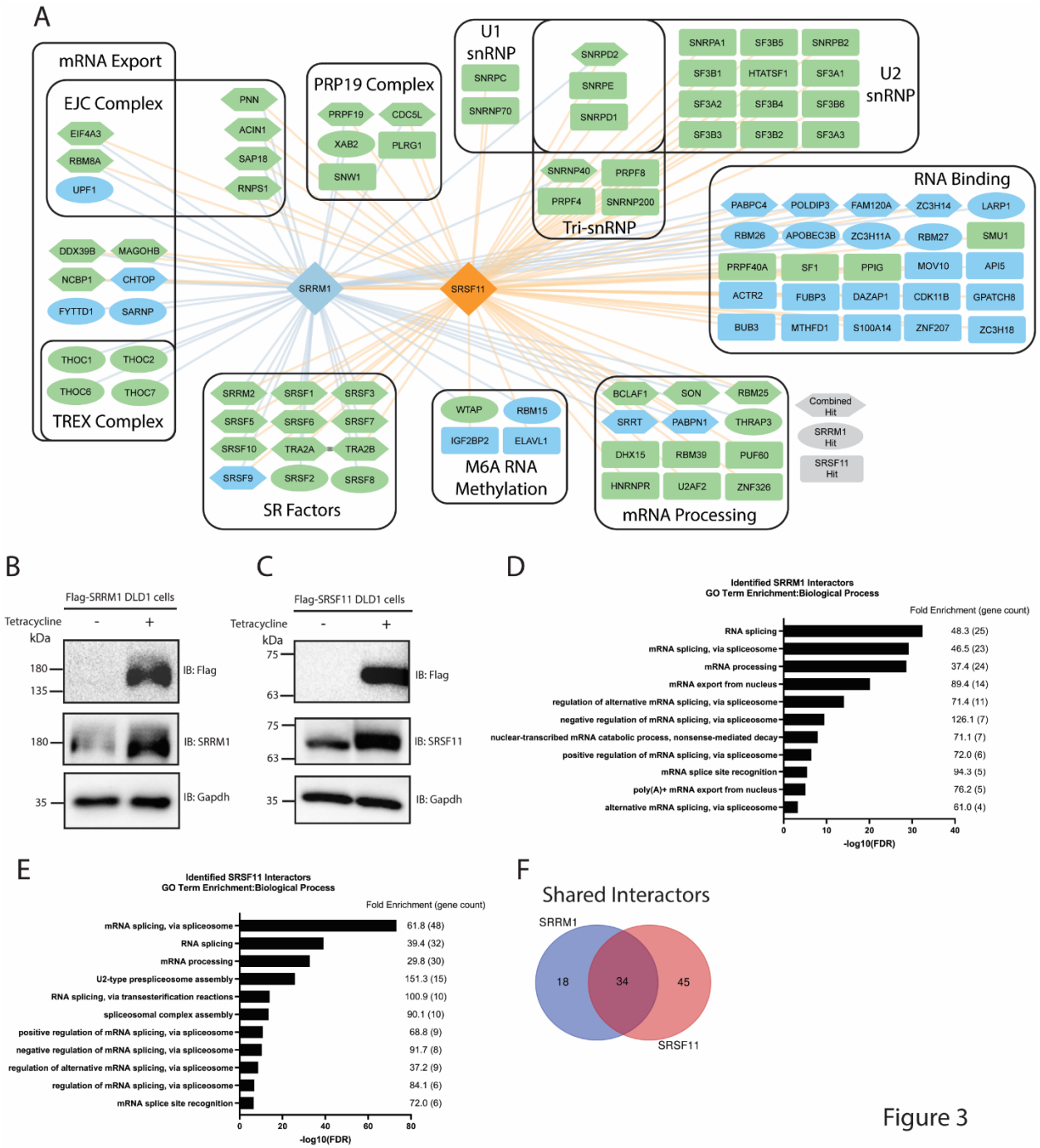

A
mRNA Export
EJC Complex
EIF4A3
RBM8A
UPF1
PNN
ACIN1
SAP18
RNPS1
DDX39B
MAGOHB
NCBP1
CHTOP
FYTTD1
SARNP
THOC1
THOC2
THOC6
THOC7
TREX Complex
PRP19 Complex
PRPF19
CDC5L
XAB2
PLRG1
SNW1
U1 snRNP
SNRPC
SNRNP70
SNRPD2
SNRPE
SNRPD1
SNRNP40
PRPF8
PRPF4
SNRNP200
Tri-snRNP
SNRPA1
SF3B5
SNRPB2
SF3B1
HTATSF1
SF3A1
SF3A2
SF3B4
SF3B6
SF3B3
SF3B2
SF3A3
U2 snRNP
RNA Binding
PABPC4
POLDIP3
FAM120A
ZC3H14
LARP1
RBM26
APOBEC3B
ZC3H11A
RBM27
SMU1
PRPF40A
SF1
PPIG
MOV10
API5
ACTR2
FUBP3
DAZAP1
CDK11B
GPATCH8
BUB3
MTHFD1
S100A14
ZNF207
ZC3H18
SRRM1
SRSF11
SRRM2
SRSF1
SRSF3
SRSF5
SRSF6
SRSF7
SRSF10
TRA2A
TRA2B
SRSF9
SRSF2
SRSF8
SR Factors
WTAP
RBM15
IGF2BP2
ELAVL1
M6A RNA Methylation
BCLAF1
SON
RBM25
SRRT
PABPN1
THRAP3
DHX15
RBM39
PUF60
HNRNPR
U2AF2
ZNF326
mRNA Processing
Combined Hit
SRRM1 Hit
SRSF11 Hit
B
Flag-SRRM1 DLD1 cells
Tetracycline - +
kDa
180
135
IB: Flag
180
IB: SRRM1
35
IB: Gapdh
C
Flag-SRSF11 DLD1 cells
Tetracycline - +
kDa
75
63
IB: Flag
75
IB: SRSF11
63
35
IB: Gapdh
D
Identified SRRM1 Interactors
GO Term Enrichment:Biological Process
Fold Enrichment (gene count)
RNA splicing 48.3 (25)
mRNA splicing, via spliceosome 46.5 (23)
mRNA processing 37.4 (24)
mRNA export from nucleus 89.4 (14)
regulation of alternative mRNA splicing, via spliceosome 71.4 (11)
negative regulation of mRNA splicing, via spliceosome 126.1 (7)
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay 71.1 (7)
positive regulation of mRNA splicing, via spliceosome 72.0 (6)
mRNA splice site recognition 94.3 (5)
poly(A)+ mRNA export from nucleus 76.2 (5)
alternative mRNA splicing, via spliceosome 61.0 (4)
0 10 20 30 40
-log10(FDR)
E
Identified SRSF11 Interactors
GO Term Enrichment:Biological Process
Fold Enrichment (gene count)
mRNA splicing, via spliceosome 61.8 (48)
RNA splicing 39.4 (32)
mRNA processing 29.8 (30)
U2-type prespliceosome assembly 151.3 (15)
RNA splicing, via transesterification reactions 100.9 (10)
spliceosomal complex assembly 90.1 (10)
positive regulation of mRNA splicing, via spliceosome 68.8 (9)
negative regulation of mRNA splicing, via spliceosome 91.7 (8)
regulation of alternative mRNA splicing, via spliceosome 37.2 (9)
regulation of mRNA splicing, via spliceosome 84.1 (6)
mRNA splice site recognition 72.0 (6)
0 20 40 60 80
-log10(FDR)
F
Shared Interactors
SRRM1
18
34
45
SRSF11

Figure 3

Figure 3: Investigating SRRM1/SRSF11 interactome using AP-MS in DLD1 colorectal adenocarcinoma cells.

(A) Protein-Protein interactors of SRRM1 (blue edges) and SRSF11 (orange edges) identified via affinity-purification mass spectrometry in DLD1 cell lines. Interactors with a BFDR ≤ 0.05 are depicted. Unique prey are depicted in ovals for SRRM1 and rectangles for SRSF11 whilst shared prey between the two baits are depicted in hexagons. Green Nodes denote hits that possessed a GOBP term that belongs to RNA splicing and/or mRNA splicing, via spliceosome according to DAVID analysis (https://davidbioinformatics.nih.gov).

(B) Immunoblots demonstrating expression of Flag tagged SRRM1 DLD1 cells ± tetracycline (10ug/ml) induction blotted with either anti-FLAG or anti-SRRM1 antibodies with corresponding anti-Gapdh blots used as a loading control.

(C) Immunoblots demonstrating expression of Flag tagged SRSF11 DLD1 cells ± tetracycline (10ug/ml) blotted with anti-FLAG or anti-SRSF11 antibodies with corresponding anti-Gapdh blots used as a loading control.

(D) Biological process GO term enrichment analysis for SRRM1 protein interactors analyzed through DAVID. The top enriched terms are arranged in order of significance (-log10 (FDR)) with fold enrichment values and total hit gene count for each term in between brackets denoted on the right.

(E) Same as **(D)** for SRSF11 protein interactors.

(F) Venn diagram depicting the number of shared interactors between SRRM1 and SRSF11 in DLD1 cells.

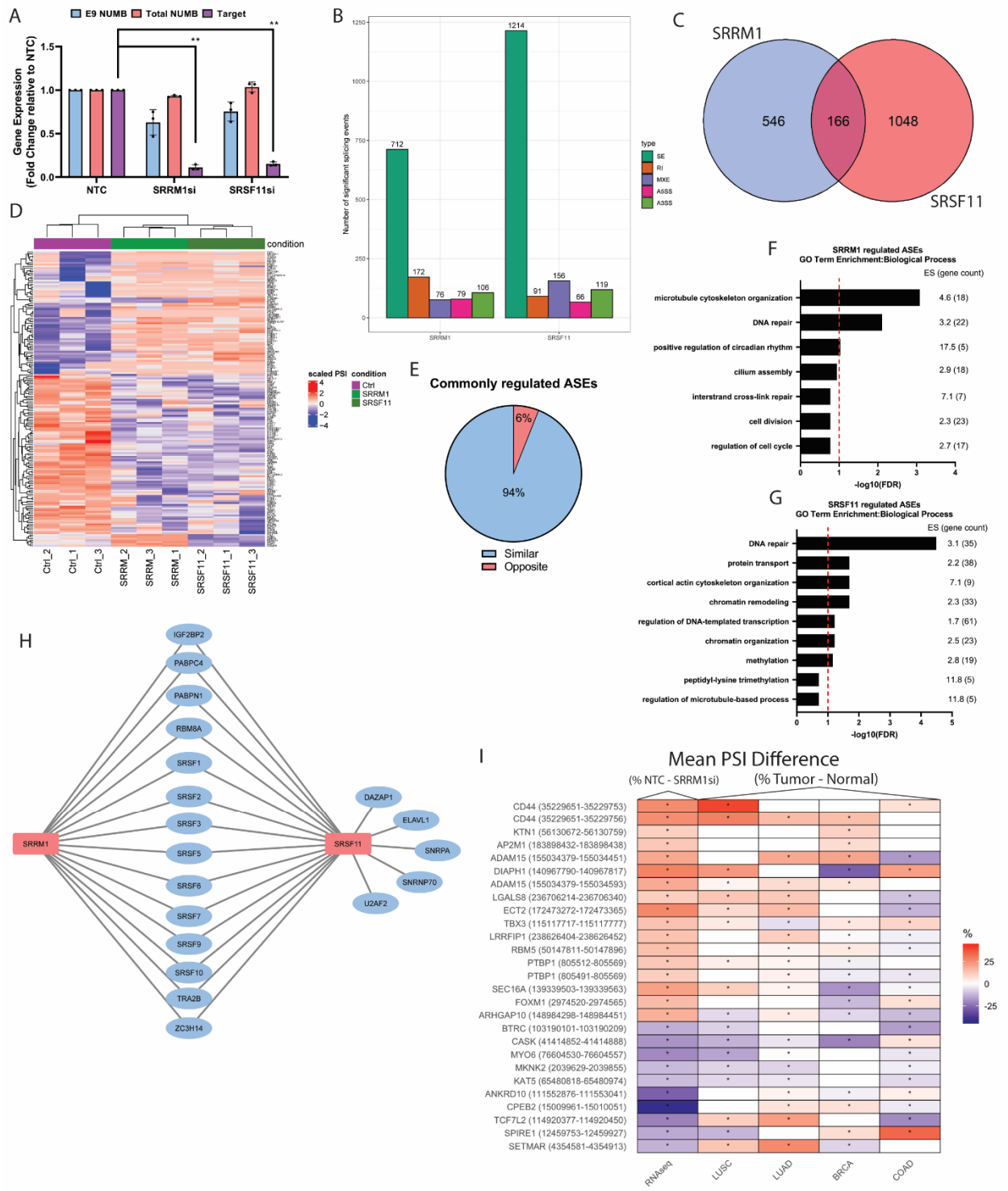

A
E9 NUMB
Total NUMB
Target
Gene Expression (Fold Change relative to NTC)
NTC
SRRM1si
SRSF11si
B
Number of significant splicing events
712
172
76
79
106
1214
91
156
66
119
SRRM1
SRSF11
type
SE
RI
MXE
A5SS
A3SS
C
SRRM1
546
166
1048
SRSF11
D
condition
scaled PSI
Ctrl
SRRM1
SRSF11
Ctrl_2
Ctrl_1
Ctrl_3
SRRM_2
SRRM_3
SRRM_1
SRSF11_2
SRSF11_1
SRSF11_3
E
Commonly regulated ASEs
6%
94%
Similar
Opposite
F
SRRM1 regulated ASEs
GO Term Enrichment:Biological Process
ES (gene count)
microtubule cytoskeleton organization 4.6 (18)
DNA repair 3.2 (22)
positive regulation of circadian rhythm 17.5 (5)
cilium assembly 2.9 (18)
interstrand cross-link repair 7.1 (7)
cell division 2.3 (23)
regulation of cell cycle 2.7 (17)
-log10(FDR)
G
SRSF11 regulated ASEs
GO Term Enrichment:Biological Process
ES (gene count)
DNA repair 3.1 (35)
protein transport 2.2 (38)
cortical actin cytoskeleton organization 7.1 (9)
chromatin remodeling 2.3 (33)
regulation of DNA-templated transcription 1.7 (61)
chromatin organization 2.5 (23)
methylation 2.8 (19)
peptidyl-lysine trimethylation 11.8 (5)
regulation of microtubule-based process 11.8 (5)
-log10(FDR)
H
IGF2BP2
PABPC4
PABPN1
RBM8A
SRSF1
SRSF2
SRSF3
SRSF5
SRSF6
SRSF7
SRSF9
SRSF10
TRA2B
ZC3H14
SRRM1
SRSF11
DAZAP1
ELAVL1
SNRPA
SNRNP70
U2AF2
I
Mean PSI Difference
(% NTC - SRRM1si)
(% Tumor - Normal)
CD44 (35229651-35229753)
CD44 (35229651-35229756)
KTN1 (56130672-56130759)
AP2M1 (183898432-183898438)
ADAM15 (155034379-155034451)
DIAPH1 (140967790-140967817)
ADAM15 (155034379-155034593)
LGALS8 (236706214-236706340)
ECT2 (172473272-172473365)
TBX3 (115117717-115117777)
LRRFIP1 (238626404-238626452)
RBM5 (50147811-50147896)
PTBP1 (805512-805569)
PTBP1 (805491-805569)
SEC16A (139339503-139339563)
FOXM1 (2974520-2974565)
ARHGAP10 (148984298-148984451)
BTRC (103190101-103190209)
CASK (41414852-41414888)
MYO6 (76604530-76604557)
MKNK2 (2039629-2039855)
KAT5 (65480818-65480974)
ANKRD10 (111552876-111553041)
CPEB2 (15009961-15010051)
TCF7L2 (114920377-114920450)
SPIRE1 (12459753-12459927)
SETMAR (4354581-4354913)
RNAseq
LUSC
LUAD
BRCA
COAD
%
25
0
-25

Figure 4

Figure 4: SRRM1 and SRSF11 knockdown results in altered downstream splicing.

(A) Representative RT-qPCR of SRRM1 and SRSF11 silencing in DLD1 colorectal adenocarcinoma cell lines submitted for RNA-seq relative to non-target control. Error bars represent SEM (n =3 biological replicates for each condition, p-val < 0.001** according to paired student t-test).

(B) Number of SRRM1 (left) or SRSF11 (right) dependent splicing events identified by rMATs following silencing treatment via siRNA. SE: alternatively spliced exon, IR: intron retention, MXE: mutually exclusive exons, A5SS: alternative 5' splice sites, A3SS: alternative 3' splice sites.

(C) Venn diagram depicting overlap in SRRM1 and SRSF11 regulated alternatively spliced exon (ASE) events.

(D) Heatmap of normalized PSI values of ASE events regulated by both SRRM1 and SRSF11 in siSRRM1, siSRSF11 and siCTRL samples.

(E) The proportion of common target ASEs following SRRM1 or SRSF11 silencing shown in (C) (n=166) that are "similar" (blue) or "opposite" (orange) depending on whether delta PSI values vary in the same or opposite direction relative to control PSI values.

(F) Biological process GO term enrichment analysis for all genes that demonstrated dysregulated alternatively spliced exons (ASEs) regulated by SRRM1 analyzed through DAVID. The top enriched terms are arranged in order of significance (-log10(FDR) with Enrichment score (ES) values and total hit gene count for each term in between brackets denoted on the right. The red dotted line indicates an FDR of 0.1.

(G) Same as (**F**) for the list of genes that demonstrated dysregulated ASEs regulated by SRSF11.

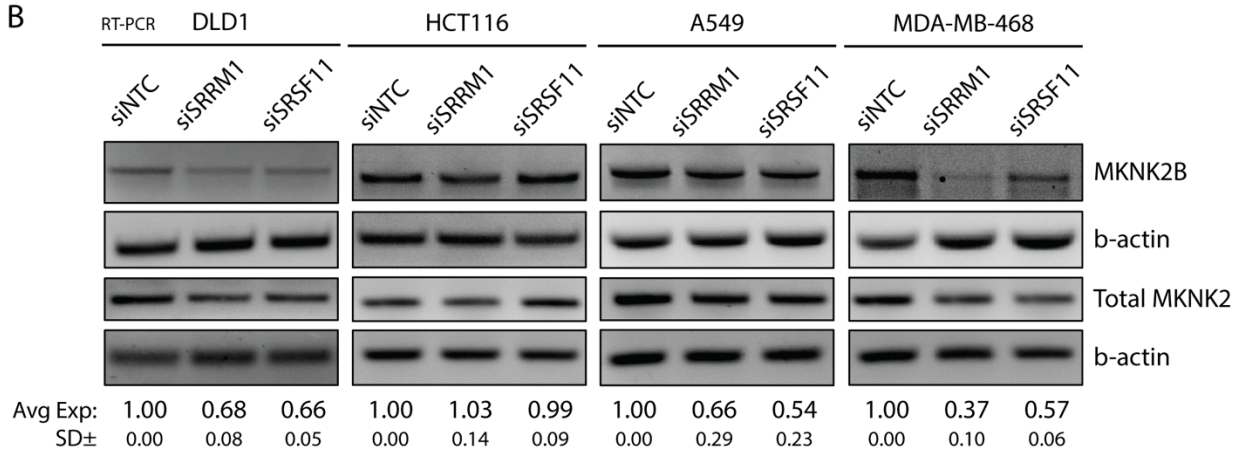
(H) Depicts proteins that were identified as protein-protein interactors of SRRM1 and/or SRSF11 through APMS with binding motifs enriched in the differentially spliced events downstream of SRRM1 and SRSF11 knockdown.

(I) Heatmap depicting splicing events that demonstrated high correlation in mean PSI difference (%NTC-SRRM1si) in the DLD1 RNAseq splicing analysis compared to mean PSI difference (%Tumor – Normal) of LUSC, LUAD, BRCA and COAD TCGA patient samples. Only values that demonstrated a significant change in tumor vs normal patient samples are depicted.

A

Cancer Hallmarks \ Spliced Gene	ADAM15	ANKRD10	AP2M1	ARHGAP10	BTRC	CASK	CD44	CPEB2	DIAPH1	ECT2	FOXM1	KAT5	KTN1	LGALS8	LRRFIP1	MKNK2	MYO6	NUMB	PTBP1	RBM5	SEC16A	SETMAR	SPIRE1	TBX3	TCF7L2
Sustaining Proliferative Signaling	■	■	■	■		■	■	■		■	■	■	■		■	■	■	■	■	■	■	■		■	■
Resisting Cell Death			■		■		■		■		■	■	■			■		■	■	■				■	■
Activating Invasion & Metastasis	■	■		■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■		■	■	■	■
Enabling Replicative Immortality			■				■							■		■		■	■			■		■	
Tumor Promoting Inflammation					■										■	■		■							
Genome Instability & Mutation									■			■				■									

B



C

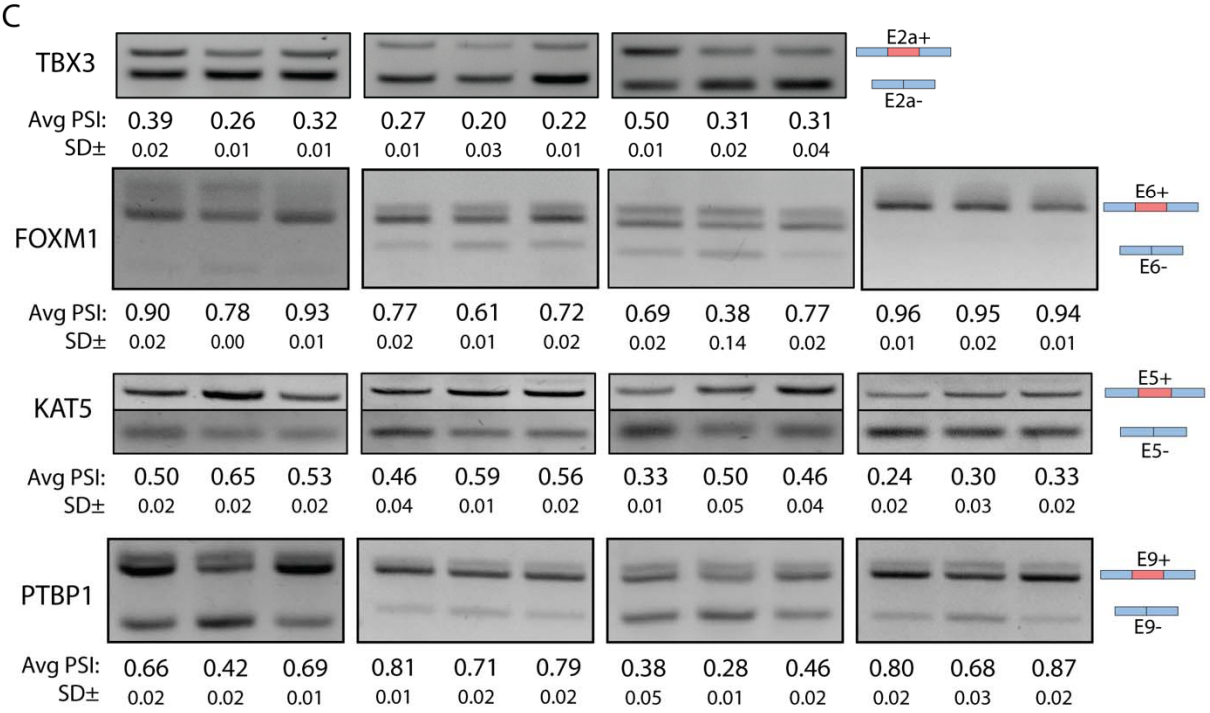


Figure 5

Figure 5: SRRM1 and SRSF11 promote expression of oncogenic protein isoforms in multiple cancer types.

(A) Cancer hallmarks associated with genes that demonstrate dysregulated splicing upon SRRM1 knockdown identified through RNA-seq rMATs analysis

(B) Representative semi-quantitative RT-PCR blots demonstrating effects of SRRM1 or SRSF11 knockdown via siRNA treatment on Total MKNK2 and MKNK2B isoform expression relative to β-actin compared with non-target control (NTC) treatment in DLD1, HCT116, A549 and MDA-MB-468 cancer cell lines. Average MKNK2B expression normalized to total MKNK2 expression ± SD of 3 biological replicates in each cell line is denoted below.

(C) Representative semi-quantitative RT-PCR blots demonstrating effects of SRRM1 or SRSF11 knockdown via siRNA treatment on shifts in TBX3 exon 2a, FOXM1 exon 6, KAT5 exon 5 and PTBP1 exon 9 isoform expression compared with non-target control (NTC) treatment in DLD1, HCT116, A549 and MDA-MB-468 cancer cell lines. Average percent spliced in (PSI) ± SD of 3 biological replicates in each cell line is denoted below.

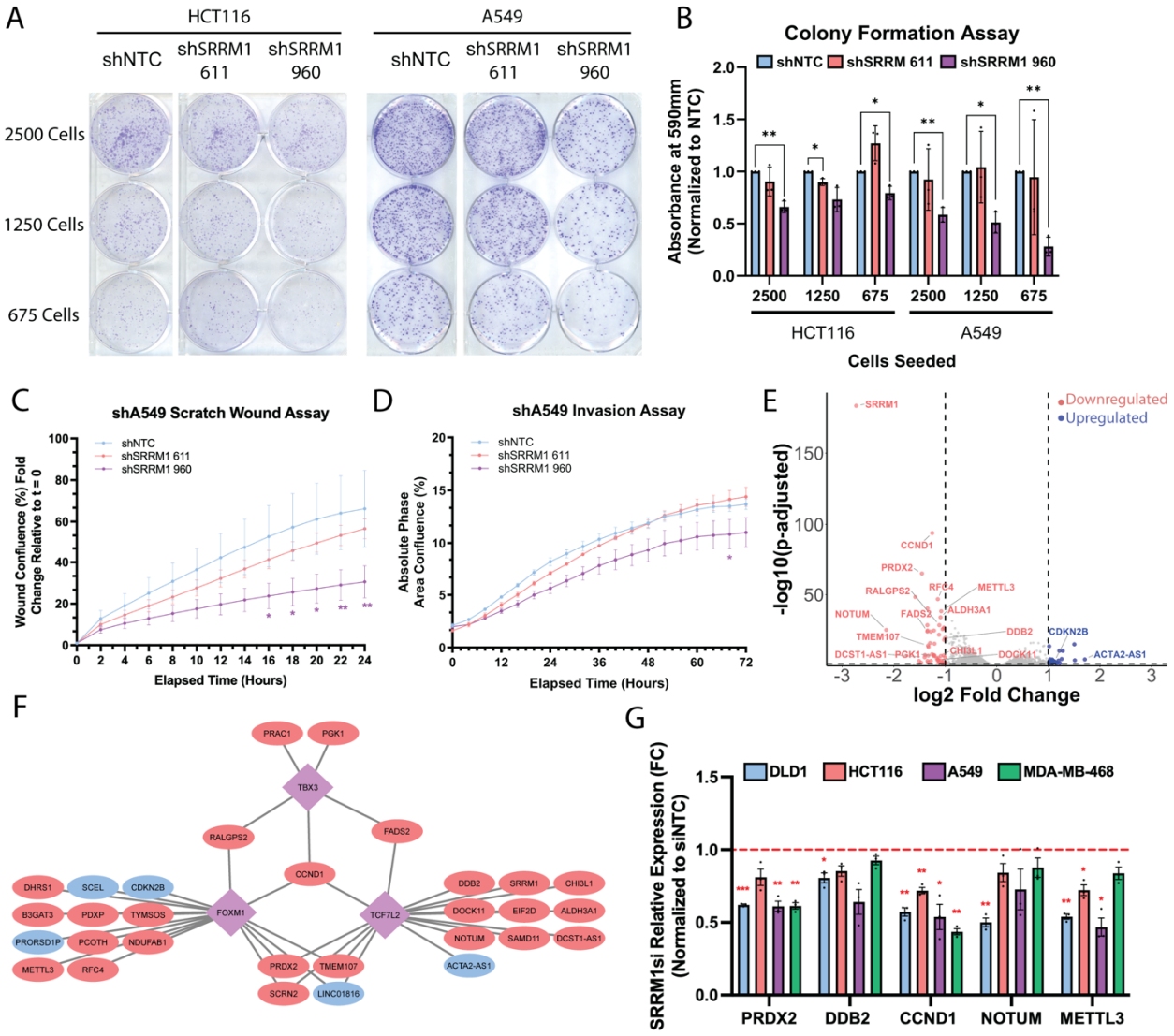

A
HCT116
A549
shNTC
shSRRM1 611
shSRRM1 960
2500 Cells
1250 Cells
675 Cells
B
Colony Formation Assay
shNTC
shSRRM 611
shSRRM1 960
Absorbance at 590mm (Normalized to NTC)
2500
1250
675
HCT116
A549
Cells Seeded
C
shA549 Scratch Wound Assay
Wound Confluence (%) Fold Change Relative to t = 0
Elapsed Time (Hours)
D
shA549 Invasion Assay
Absolute Phase Area Confluence (%)
Elapsed Time (Hours)
E
Downregulated
Upregulated
-log10(p-adjusted)
log2 Fold Change
SRRM1
CCND1
PRDX2
RALGPS2
RFC4
METTL3
NOTUM
FADS2
ALDH3A1
TMEM107
DDB2
CDKN2B
DCST1-AS1
PGK1
CHI3L1
DOCK11
ACTA2-AS1
F
PRAC1
PGK1
TBX3
RALGPS2
FADS2
CCND1
FOXM1
TCF7L2
DHRS1
SCEL
CDKN2B
B3GAT3
PDXP
TYMSOS
PRORSD1P
PCOTH
NDUFAB1
METTL3
RFC4
DDB2
SRRM1
CHI3L1
DOCK11
EIF2D
ALDH3A1
NOTUM
SAMD11
DCST1-AS1
ACTA2-AS1
PRDX2
TMEM107
SCRN2
LINC01816
G
DLD1
HCT116
A549
MDA-MB-468
SRRM1si Relative Expression (FC) (Normalized to siNTC)
PRDX2
DDB2
CCND1
NOTUM
METTL3

Figure 6

Figure 6 SRRM1 knockdown impairs cell growth and colony formation and alters gene expression.

- **(A)** Representative images from colony formation assays conducted on tetracycline-inducible HCT116 and A549 SRRM1 short hairpin cell lines (shSRRM1 611/960) relative to a non-target control line (NTC) seeded at either 2500, 1250 and 675 cells per well and cultured for 14 days (HCT116) or 10 days (A549) under tetracycline induction before crystal violet staining and imaging.
- **(B)** Quantification of colony formation assays **(C)** across 3 biological replicates normalized to NTC control. Significant differences are denoted with a p-values < 0.05 = *, p-values < 0.01 = ** according to a paired t-test.
- **(C)** Quantification of scratch wound migration assays conducted on A549 short hairpin cell lines measuring wound confluence using an Incucyte live cell imaging system. The wound confluence at each time point was normalized to the initial wound confluence. Cells were allowed to form a monolayer before scratching a wound and imaging migration across 24 hours. Significance was determine through a two-way ANOVA analysis between shNTC and shSRRM1 lines.
- **(D)** Same as **(C)** but cells were embedded in 8mg/ml Matrigel for an Invasion assay and imaged across 72 hours.
- **(E)** Volcano plot demonstrating -log10(FDR adjusted p-value) vs Log2fold change of gene expression in SRRM1 knockdown cells identified by RNA-seq Deseq2 analysis. Only genes that exhibited p-adjusted > 0.05 and Log2Fold change > 1 or < -1 were considered hits. Labeled downregulated oncogenic genes and upregulated tumor suppressive genes and potential targets of alternative spliced transcription factors (FOXM1, TBX3, TCF7L2) are shown.
- **(F)** Predicted targets of FOXM1, TBX3 and TCF7L2 transcription factors predicted through ChEA3 (https://maayanlab.cloud/chea3/), which exhibited significant expression changes in response to SRRM1 knockdown. Red nodes denote reduced expression, and blue nodes denote increased expression in response to SRRM1si treatment.
- **(G)** Real-time quantitative PCR measuring PRDX2, DDB2, CCND1, NOTUM and METTL3 gene expression in DLD1, HCT116, A549 and MDA-MB-468 cells treated with SRRM1 siRNA relative to non-target control treatment across 3 biological replicates. Relative fold change expression values were normalized to NTC control (red dotted line FC =1.0). Significant differences with a p-values < 0.05 = *, p-values < 0.01 = **, p-values < 0.001 = ***.