A functional network of gastric-cancer-associated splicing events controlled by dysregulated splicing factors

Shanshan Cheng [1,2,3,†], **Debleena Ray** [3,†], **Raymond Teck Ho Lee**[3], **Kishore Babu Naripogu**[3], **Permeen Akhtar Bt Mohamed Yusoff**[3], **Pamela Bee Leng Goh**[3], **Yujing Liu**[2,3,4], **Yuka Suzuki**[2,3], **Kakoli Das**[3], **Hsiang Sui Chan**[5], **Wai Keong Wong**[6], **Weng Hoong Chan**[6], **Pierce Kah-Hoe Chow**[7,8,9], **Hock Soo Ong**[10], **Prema Raj**[11], **Khee Chee Soo**[7,9,12], **Patrick Tan**[3], **David M. Epstein**[3,*] and **Steven G. Rozen** [2,3,*]

[1]Department of Epidemiology and Biostatistics, Key Laboratory for Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, 13 Hangkong Rd, Wuhan, Hubei 430030, China, [2]Centre for Computational Biology, Duke–NUS Medical School, 8 College Rd, Singapore 169857, Singapore, [3]Cancer & Stem Cell Biology Programme, Duke–NUS Medical School, 8 College Rd, Singapore 169857, Singapore, [4]Singapore MIT Alliance, 4 Engineering Dr 3, Singapore 117576, Singapore, [5]Department of General Surgery, Gleneagles Medical Centre, 6A Napier Rd, Singapore 258500, Singapore, [6]Department of Upper Gastrointestinal & Bariatric Surgery, Singapore General Hospital, 1 Hospital Dr, Singapore 169608, Singapore, [7]Division of Surgical Oncology, National Cancer Center Singapore, 11 Hospital Dr, Singapore 169610, Singapore, [8]Department of HPB and Transplant, Singapore General Hospital, 1 Hospital Dr, Singapore 169608, Singapore, [9]Clinical, Academic & Faculty Affairs, Duke–NUS Medical School, 8 College Rd, Singapore 169857, Singapore, [10]Department of General Surgery, Singapore General Hospital, 1 Hospital Dr, Singapore 169608, Singapore, [11]General Surgery, Mount Elizabeth Medical Center, 3 Mount Elizabeth, Singapore 228510, Singapore and [12]Yong Loo Lin School of Medicine, National University of Singapore, 21 Lower Kent Ridge Rd, Singapore 119077, Singapore

ABSTRACT

Comprehensive understanding of aberrant splicing in gastric cancer is lacking. We RNA-sequenced 19 gastric tumor–normal pairs and identified 118 high-confidence tumor-associated (TA) alternative splicing events (ASEs) based on high-coverage sequencing and stringent filtering, and also identified 8 differentially expressed splicing factors (SFs). The TA ASEs occurred in genes primarily involved in cytoskeletal organization. We constructed a correlative network between TA ASE splicing ratios and SF expression, replicated it in independent gastric cancer data from The Cancer Genome Atlas and experimentally validated it by knockdown of the nodal SFs (*PTBP1*, *ESRP2* and *MBNL1*). Each SF knockdown drove splicing alterations in several corresponding TA ASEs and led to alterations in cellular migration consistent with the role of TA ASEs in cytoskeletal organization. We have therefore established a robust network of dysregulated splicing associated with tumor invasion in gastric cancer. Our work is a resource for identifying oncogenic splice forms, SFs and splicing-generated tumor antigens as biomarkers and therapeutic targets.

INTRODUCTION

Gastric cancer is the fourth most common cancer type, the third leading cause of cancer-related deaths worldwide and extremely common in East Asia (1,2). Gastric cancer is often diagnosed at advanced stages due to late emergence of symptoms, and it is associated with poor 5-year survival rates and limited treatment options (3). Gastric cancer also manifests substantial heterogeneity, and has been classified into subtypes based on histological features (4) and genotypes and gene expression profiles (5–7). Research on gastric carcinogenesis has primarily focused on acquired molecular aberrations, including chromosomal instability, microsatellite instability, altered epigenetic profiles and somatic muta-

tions, with the goal of finding key players that can potentially lead to improved therapy and prognosis (8).

One molecular aspect of gastric cancer that is less studied is alternative splicing. In general, it is estimated that 95% of multi-exon human genes undergo alternative splicing (9,10), and alternative splicing is a major mechanism that generates functional proteomic diversity. Splicing is modulated by both the spliceosome and trans-acting splicing factors (SFs) (11). The choice of alternative splice sites is regulated by these trans-acting SFs (12) and by the cis-regulatory exonic/intronic splicing enhancer or silencer sequences that they bind to (13).

Alternative splicing has important regulatory functions in cellular development and differentiation (14), is highly tissue specific (15) and has been linked to some diseases (16). Several lines of evidence indicate that tumors exploit alternative splicing to acquire growth and survival advantage (17). First, several cancer types harbor recurrent somatic hotspot mutations affecting SFs, notably *U2AF1* (18) and *SF3B1* (19–21). Second, several cancer types exhibit alteration of the expression of SFs. Examples include RBFOX2 downregulation in ovarian cancer (22), SRSF1 overexpression in breast cancer (23), and PRPF6 and PTBP1 overexpression in colorectal cancer (24,25). Third, the oncogenic effect of aberrant splicing of known tumor drivers has been delineated in several instances comprising several cancer types, showing that alternative splicing sometimes directly promotes tumor progression (17). Examples include *AR* (androgen receptor) (26), *BCL2L11* (also known as *BIM*) (27), *BRAF* (28), *VEGFA* (29), *BCL2L1* (30,31), *CD44* (32) and *FGFR2* (33).

To date, a few large-scale RNA-seq studies have described aberrant splicing repertoires of several cancer types: breast cancer (34), chronic lymphocytic leukemia (35), lung cancer (36), acute myeloid leukemia (37) and uveal melanoma (38,39). With respect to gastric cancer, we are aware of three studies that globally examined alternative splicing, and each study addressed a different question. One study focused on differences in alternative splicing between Epstein–Barr positive and negative gastric cancers in The Cancer Genome Atlas (TCGA) data, followed by focused, functional studies in an embryonic kidney cell line (6,40). Another study explored gastric-cancer-associated splicing within a larger multi-omics study, and then focused on functional studies of a specific ZAK kinase splice form (41). A third study examined the splicing aberrations in gastric carcinomas using the TCGA data, focusing on the prognostic value of a few splicing events in predicting patient survival (42). In addition to these studies that globally examined alternative splicing, other studies in gastric cancer focused on splicing in single specific genes, including *CD44*, *FGFR2*, *BIRC5*, *MST1R*, *CD82*, *WISP1*, *SPP1*, *WNT2B*, *CDH1*, *MUTYH*, *FHIT*, *MUC1*, *NUF2* and *TERT* [reviewed in (43)].

Given the potential impact of RNA-seq analysis of tumor-associated (TA) aberrant splicing and given the recent advances in understanding the role of splicing in the biology of other tumor types, we sought to use high-coverage RNA sequencing to systematically identify the repertoire of high-confidence TA alternative splicing events (ASEs) in primary human gastric cancer tissue and to experimentally identify their upstream regulators. Our aim was to construct a network of dysregulated splicing in gastric cancer and to use knockdown experiments to functionally validate it. This network illuminates the mechanisms of gastric tumorigenesis, and points to high-confidence TA, and in some cases oncogenic, splice forms that are common across gastric cancer subtypes. These splice forms can then be the subjects of further functional study, including those splice forms that might encode TA antigens that could be exploited in immunotherapy, or that might be used as biomarkers for distinguishing tumor from non-tumor tissue in the operating room.

MATERIALS AND METHODS

The tissue samples, the cell lines and the experimental methods are described in detail in Supplementary Information.

RNA-seq analysis for identifying TA ASEs and differentially expressed SFs

We used high-coverage (>200 million reads), stranded, RNA sequencing on rRNA-depleted input RNA in order to provide a basis for very high confidence identification of alternative splicing. Supplementary Information provides further details. We used rMATS V3.2.5 for exon-centric alternative splicing analysis of both the primary tissue samples and the gastric cell lines (44). rMATS quantifies the relative abundance of an ASE by a percent spliced in (PSI) value, and it does so for all major types of ASEs, including skipped exons, alternative 5′ splice sites, alternative 3′ splice sites, mutually exclusive exons and retained introns. The identification of ASEs was guided by GENCODE gene annotation v19 with the argument '-novelSS 1' to additionally allow detection of novel splice sites. Differential splicing was called by RNASeq-MATS.py with the options '-t paired -len 100 -analysis P -libType fr-firststrand -novelSS 1 -c 0.1'. Figure 1B shows PSIs in tumor and normal samples at an ASE in the *HLA-DMB* gene as examples. For matched tumor and normal tissue samples, we performed paired analysis. We sought to maximize the specificity of our analysis because we intended to identify high-confidence TA ASEs that are common across gastric cancer subtypes. In the future, these could be prioritized for functional analysis or might be exploited in immunotherapy or used as biomarkers for distinguishing tumor from non-tumor tissue in the operating room. We first selected ASEs with false discovery rates (FDRs) <0.05 based on the rMATS *P*-values and with median junction read counts ≥10 in both tumor and normal samples. To further maximize specificity, we inspected the sashimi plots of all remaining individual ASEs (Supplementary Information and Supplementary Figures S1–S4). We excluded ASEs with any of the following characteristics, as illustrated in the examples in Supplementary Figure S2: (i) ASEs with many intronic reads, because these likely arose from unprocessed transcripts and may obscure the signal arising from mature transcripts; and (ii) ASEs with complex transcript structures involving several combinations of multiple exons, for which rMATS is often unable to reliably estimate PSI (45).

We selected an initial list of 73 SFs to investigate from a database of experimentally validated SFs, SpliceAid-F (46), and from a review of the literature (Supplementary Table
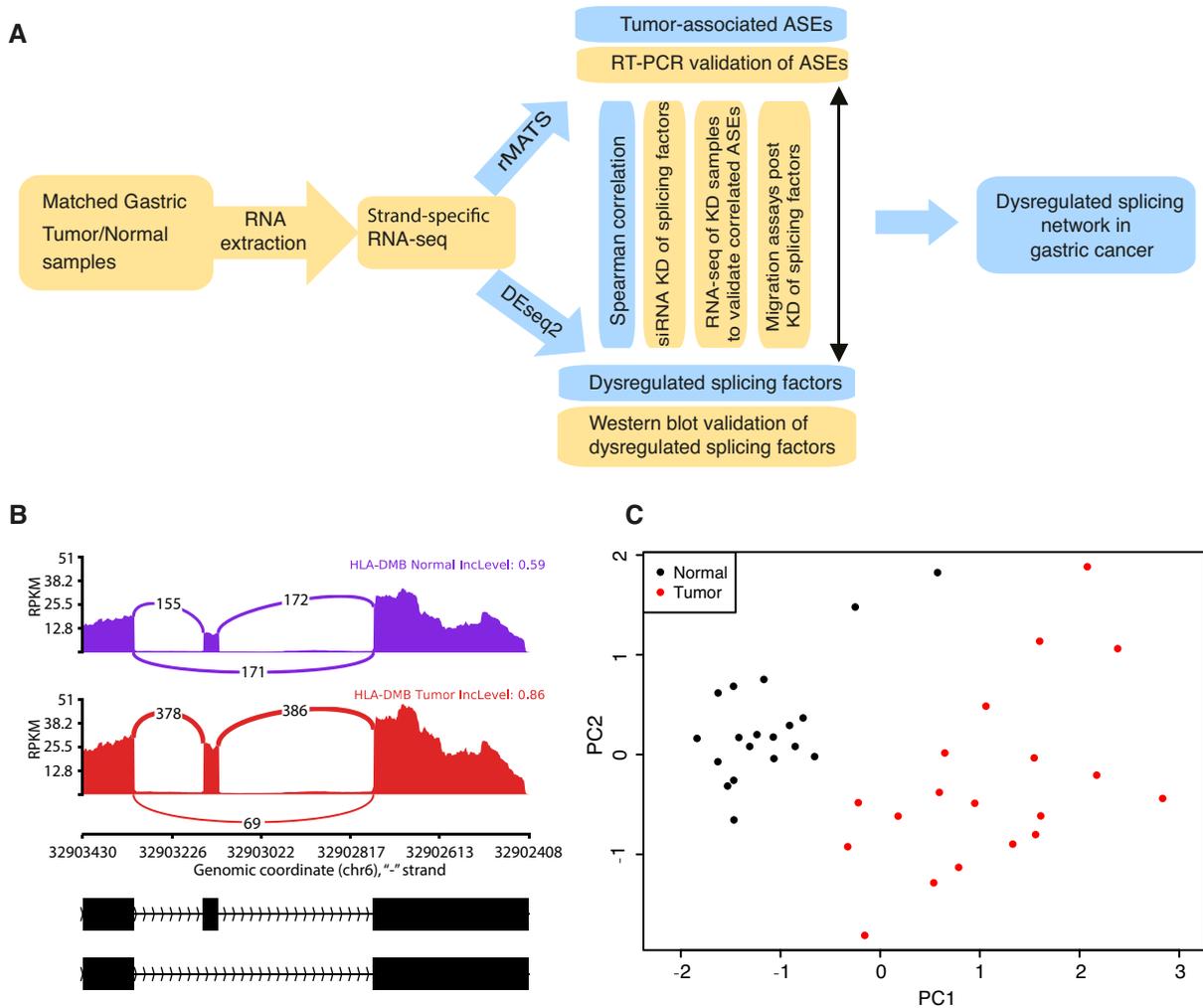
A

B

C

Figure 1. Experimental design and identification of TA ASEs. (**A**) Flowchart of study design. (**B**) Sashimi plot of an example of a differentially spliced ASE in tumor (red) and normal (purple) samples above a schematic representation of the alternative exon–intron structures (bottom panel in black). RPKM = reads per kilobase per million reads, averaged across samples within the same group. Numbered arcs within the sashimi plot, e.g. 155 in the upper left, indicate the average number of sequencing reads that span the indicated exon–exon junction. In this example, the middle exon was included with an average PSI of 0.59 in the normal sample and an average PSI of 0.86 in the tumor. (**C**) Principal component analysis over PSIs of 118 TA ASEs strongly separates tumors (red) from non-malignant samples (black).

S4). We detected changes in SF transcript levels in the 19 paired primary tumor and normal samples by first computing gene counts using HTSeq (47) and then analyzing differential gene expression using DESeq2 (48). To identify SFs that were differentially expressed at the transcript level, we initially required FDR ≤ 0.05 and absolute $\log_2$ fold change ≥ 0.5. We further included SFs that did not satisfy these criteria but were known from the literature to be associated with some cancer types. This yielded 22 candidate SFs with dysregulated transcript levels. We then assessed the 22 SFs' protein levels via western blot in several gastric cancer cell lines (Supplementary Table S2) to generate the final list of differentially expressed SFs.

Experimental assessment of SF regulation of ASEs

We used knockdown experiments to assess whether the SFs directly or indirectly controlled correlated ASEs. We

knocked down each SF with siRNAs targeting three different regions of the gene (Supplementary Table S5). In addition, three replicates of a non-targeting siRNA were used as negative controls for each cell line. For each knocked-down SF, we used rMATS to determine which ASEs had significant PSI changes in the direction predicted by the correlation network. We evaluated the ratio of the number, d, of such ASEs relative to the number, c, of TA ASEs that were both detectable in the cell line and significantly correlated with the SF in the network model. To compute the statistical significance of d/c, we calculated an empirical null distribution based on 10 000 random samples as follows. For each sample, we randomly selected c ASEs that were detectable in the cell line, but without regard to the statistical significance of their correlation to the SF in the network model, and then determined d' based on the knockdown data. The P-value was the fraction of the 10 000 random samples in which $d'/c \geq d/c$.

RESULTS

ASEs in gastric cancer occur in genes involved in cytoskeleton organization

We generated deep ($\sim$200 million reads per sample), strand-specific RNA-seq data from 19 gastric tumors and their adjacent normal tissues (Figure 1A, Supplementary Table S3). Of the 19 tumor samples, 11 were of intestinal subtype, while the others were either diffuse or of a mixed subtype. Most were late-stage samples and most were poorly differentiated (Supplementary Table S1).

Using rMATS (44), we identified a total of 238 545 ASEs across the 19 tumor–normal pairs. To focus on high-confidence differential splice form abundance between gastric tumors and corresponding normal tissues, we screened for ASEs that showed a change in PSI in either direction between tumor and normal samples with an rMATS FDR < 0.05. We then manually curated the ASEs and retained 118 ASEs in 100 genes that showed strong evidence of tumor association ('Materials and Methods' section, Supplementary Table S6, Supplementary Figures S1–S4). Henceforth, we refer to these as TA ASEs. Several known splicing variants affecting cancer-critical genes were rediscovered among the 118 TA ASEs, including ASEs in *CD44*, *KRAS*, *CTTN* and *TNC* (49–52). The list of TA ASEs we identified here had limited overlap with gastric-cancer-associated alternative splice forms identified in other studies (40,41) (9 and 13 overlapping genes, respectively). This is likely because the previous studies relied on difficult-to-estimate differences in abundances of entire transcript isoforms rather than on differences in PSI at ASEs. ASEs are spatially localized, and therefore, compared to the abundances of entire transcript isoforms, PSIs at ASEs are more amenable to estimation based on relatively short, next-generation sequencing reads. Principal component analysis on the PSIs of the 118 TA ASEs cleanly separated tumor from normal samples (Figure 1C). Thus, the 118 TA ASEs constitute an alternative splicing signature that distinguishes gastric tumors from adjacent non-malignant tissue based on PSIs.

In a panel of gastric cancer cell lines and using HFE-145 (a normal gastric cell line) and normal primary stomach RNA as controls, reverse transcriptase polymerase chain reaction (RT-PCR) experiments on 25 top-ranked TA ASEs confirmed that 19 had PSI differences in agreement with those observed in the primary tissues (Figure 2A, Supplementary Table S7, Supplementary Figure S5). These RT-PCR experiments confirmed the existence and directionality of PSI differences that we initially detected in the RNA-seq data. Furthermore, these experiments indicated that the splicing changes in primary tumors were recapitulated in the cell lines, which could then serve as model systems for experimental study of the role of splicing alteration in gastric cancer.

Gene set enrichment analysis

We investigated enrichment for high-confidence TA ASEs in gene sets in the MSigDB 'GO gene sets' and 'curated gene sets' collections (53). To avoid detecting enrichment for alternatively spliced genes in general, as opposed to genes with TA alternative splicing, we used the set of genes harboring detected ASEs, and not the set of all genes, as the background set. We found that TA ASEs prominently affect genes involved in cytoskeleton organization and cell motility (adjusted *P*-values by hypergeometric tests between 7.23×10^{-6} and 4.91×10^{-2}, Figure 2B, Supplementary Figure S6, figures plotted using the Bioconductor package 'clusterProfiler' (54)). In the 'GO gene sets' collection, 'actin filament based process', 'actin binding' and 'cell leading edge', which are all essential functions in cellular movement, showed the most significant enrichment. In the 'curated gene sets' collection, 'KEGG_ECM_RECEPTOR_INTERACTION', which contains genes involved in adhesion, migration, differentiation, proliferation and apoptosis, and 'KEGG_TIGHT_JUNCTION', which contains genes that mediate cell adhesion, showed significant enrichment. In addition, the set 'COULOUARN_TEMPORAL_TGFB 1_SIGNATURE_UP', which contains genes associated with a more invasive phenotype in experimental systems, showed significant enrichment. These findings regarding gene set enrichment are consistent with previous reports in other cancer types of TA ASEs in cytoskeletal genes such as *TPM1*, *CTTN*, *CALD1* and *ENAH* (50,55–58).

We also draw attention to two additional enriched gene sets in the 'curated gene sets' collection (Supplementary Figure S6). 'DUTERTRE_ESTRADIOL_RESPONSE_24HR_DN' contains genes that respond to estradiol treatment. Thus, estrogen may play a role in gastric cancer. This finding is consistent with reports of the association of estrogen receptors with gastric cancer, in terms of both expression dysregulation and isoform switches (59,60). Another highly enriched gene set in the 'curated gene sets' collection, 'DANG_REGULATED_BY_MYC_DN', contains genes downregulated by the MYC oncogene. This is consistent with reports that MYC regulates several SFs, including PTBP1 (61), which was a key dysregulated SF identified in the current study.

Because some previous studies (40,42) reported larger sets of TA ASEs based on relaxed filtering criteria, we analyzed gene set enrichment for 1595 genes harboring 2104 TA ASEs identified using relaxed filtering criteria in our data (Supplementary Table S11). While the themes noted earlier remained, and a few more gene sets were enriched, FDRs were higher, suggesting that using the stringent criteria for identifying TA ASEs is more likely to pinpoint relevant biology (Supplementary Figure S10).

Expression of several SFs is dysregulated in gastric cancer

To understand the mechanisms regulating splicing aberration in the 19 tumor–normal pairs, we examined the data for mutations at splice sites in TA ASEs and looked for recurrent mutations or expression level differences in the 73 experimentally confirmed SFs listed in the SpliceAid-F database (46). In addition, we examined the expression level differences of several regulators of SFs, specifically SF kinases and phosphatases, including CLK1, CLK2, CLK3, CLK4, SRPK1, SRPK2, SRPK3, DYRK1A, DYRK2, PIM1, PIM2, PRPF4, PPP1CA, PPP1CB, PPP1CC, PP2CA, PPP2CB, MAPK1, MAPK3,

A

RPS24

AFMID
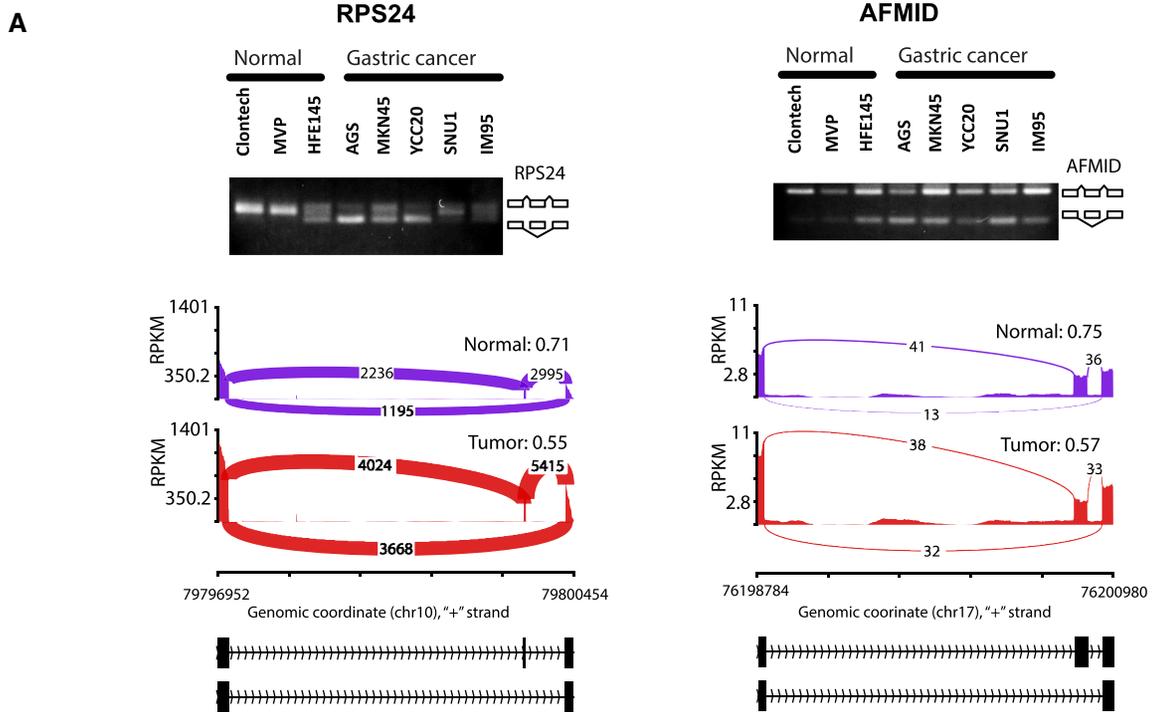
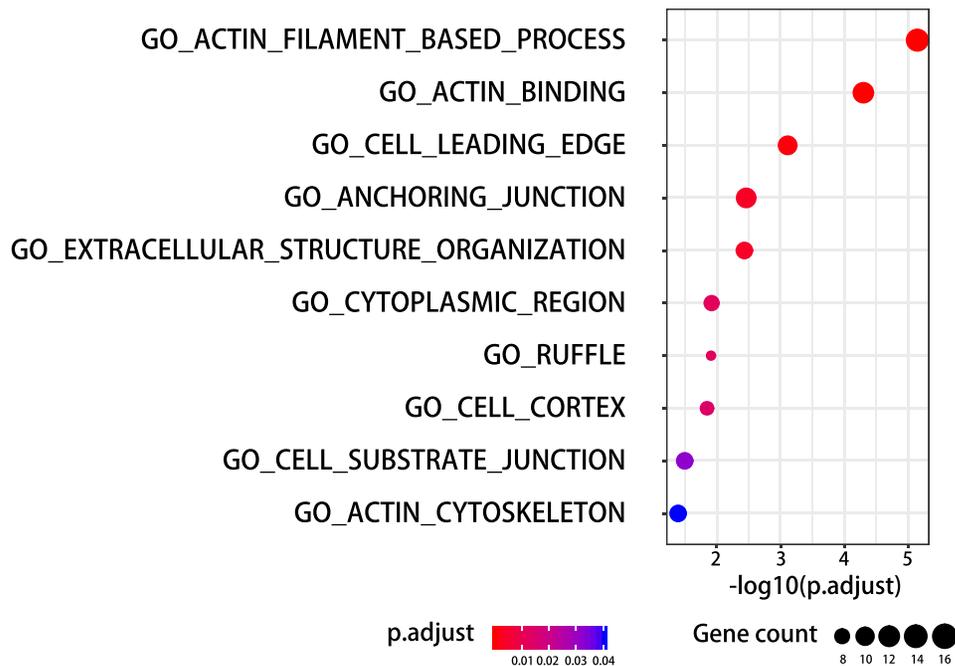**B** Enrichment of TA ASEs in MSigDB C5 (Gene Ontology) gene sets

Figure 2. Confirmation of TA ASEs. (**A**) Top panels show gel electrophoresis images of RT-PCR confirmation of example TA ASEs above sashimi plots and diagrams of the ASEs' exon–intron structures, as in Figure 1B. (**B**) Top enriched Gene Ontology (GO) terms (C5 of MSigDB) in genes affected by TA ASEs. Dot size indicates the number of genes harboring TA ASEs in that GO term; dot color indicates the significance of enrichment (hypergeometric test *P*-value adjusted by the Benjamini–Hochberg method). Figure plotted using the Bioconductor package 'clusterProfiler' (54).

MAPK7-14, AKT1, PTK6, FYN, SRC, ABL1, PRKACA, PRKACB, TOP1, FASTK and AURKA (62,63). We did not find any splice site mutations, or any recurrently mutated SFs (Supplementary Table S4). We did find the Aurora kinase A (AURKA) to be upregulated in gastric tumors and the protein kinase CAMP-activated catalytic subunit beta (PRKACB) to be downregulated in gastric tumors. While AURKA was previously reported to regulate splicing of two apoptotic genes, BCL-X and MCL1, through phosphorylation of SRSF1 (64), we did not observe significantly differential splicing of these two genes in our data. PRKACB, which acts on SRSF1 and SRSF7, was reported to regulate alternative splicing of the gene, TAU (65,66), although the latter was not found to be differentially spliced in our data either. Therefore, we focused on differential expression of SFs as possible drivers of PSI differences between tumor and normal tissue. The mRNA levels of 22 of the 73 SFs differed between tumor and normal samples [DESeq2 FDR < 0.05 and abs($\log_2$(fold change)) > 0.5, see 'Materials and Methods' section]. We then assessed expression of these 22 SFs plus MBNL1 (because of previous reports indicating its role in cancer) in a panel of 8 gastric cancer and 2 non-malignant gastric epithelium cell lines by western blot analysis. SF proteins SF3B3, PTBP1, SQSTM1, HNRNPF, HNRNPL and HNRNPK were upregulated in gastric cancer cell lines, while MBNL1 was downregulated (Supplementary Figure S11). ESRP2 mRNA and protein were absent from both of the non-malignant cell lines that we studied. However, ESRP2 mRNAs were consistently downregulated in tumors in our 19 tumor–normal pairs and in the TCGA gastric cancer data. We therefore proceeded on the hypothesis that ESRP2 downregulation contributes to gastric carcinogenesis.

PTBP1, *ESRP2* and *MBNL1* play key regulatory roles in the networks of gastric-cancer-associated splicing

To better understand the roles of these eight SFs (the six upregulated SFs and the two downregulated SFs, MBNL1 and ESRP2) in gastric cancer, we examined associations between the SFs and their potential ASE targets. We first computed Spearman's correlations between the mRNA levels of the dysregulated SFs and the PSI values of the 118 high-confidence TA ASEs (Supplementary Table S8). Some dysregulated SFs were correlated to more TA ASEs than others, with *ESRP2*, *SF3B3*, *PTBP1* and *MBNL1* correlated to the most TA ASEs (Supplementary Figure S7). In all, 59% of the TA ASEs were significantly correlated (Benjamini–Hochberg FDR < 0.05, Supplementary Table S8) to ≥1 dysregulated SF (Figure 3A). The average absolute correlation coefficient over significantly correlated SF–ASE pairs was 0.64 (95% confidence interval = 0.42–0.79). Statistical significance of the correlation coefficients was estimated by permutation tests that randomly permuted the association of PSIs with the ASEs before computing the ASE–SF correlations. We did this separately for the tumor and for the matched non-malignant tissue to remove a possible indirect association due to the fact that both the ASEs and SFs were TA.

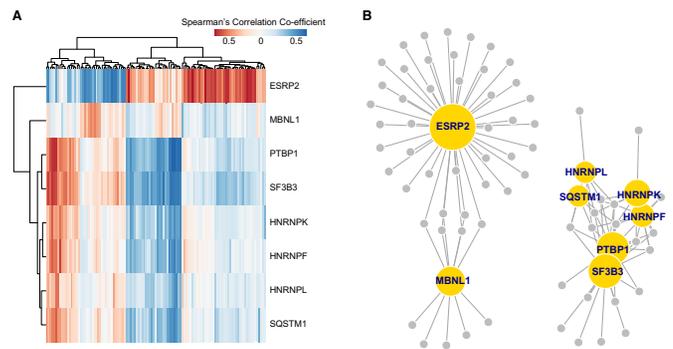To visualize the association between dysregulated SFs and TA ASEs, we computed a hub-and-spoke network

Figure 3. Expression of the dysregulated SFs and the PSIs of TA ASEs is highly correlated. (**A**) Heatmap showing Spearman's correlation coefficients between the DESeq2-normalized log counts of the 8 dysregulated SFs and the PSIs of the 118 TA ASEs (data in Supplementary Table S8). Red indicates high positive correlation, while blue indicates high negative correlation. (**B**) A hub-and-spoke network model showing correlations (spokes) between the eight dysregulated SFs (hubs in yellow) and their potential target TA ASEs (gray). The model features three regulatory clusters.

model of aberrant splicing events (Figure 3B), which depicted highly correlated ASEs and SFs (those with Benjamini–Hochberg FDR < 0.05 based on the permutation test *P*-values, Supplementary Table S8). Hierarchical clustering of the eight SFs based on their co-expression with the ASEs and assessed by silhouette indices indicated that the optimal number of clusters was 3 (Figure 3A, Supplementary Information, Supplementary Table S12). Cluster I consisted of *PTBP1*, *SF3B3*, *HNRNPK*, *HNRNPF*, *HNRNPL* and *SQSTM1*, Cluster II consisted of *ESRP2* and Cluster III consisted of *MBNL1*. This model suggested that the dysregulated SFs in gastric cancer regulate more than half of the identified TA ASEs either directly or indirectly.

We experimentally assessed the possible regulatory roles of the SFs by using RNAi to knock down one representative SF from each of the three clusters in a gastric cell line, followed by RNA-seq analysis of the consequences. For the knockdown target in Cluster I, we selected PTBP1, which was correlated with the second largest number of ASEs in that cluster (Supplementary Figure S7). (The SF correlated with the largest number of ASEs was SF3B3, which cannot be knocked down because it is required for cell viability.) We studied PTBP1 in the AGS cell line, a gastric adenocarcinoma cell line that expresses high levels of this gene. After knockdown of PTBP1, 4 out of the 11 *PTBP1*-correlated ASEs that were detectable in AGS showed changes in splicing in the expected direction (*P* = 0.246, see 'Experimental assessment of SF regulation of ASEs', Figure 4A). Because TA ASEs in Cluster I are correlated to multiple SFs (Figure 3B), it is possible that the seven ASEs that were correlated to PTBP1 but showed no effect upon PTPB1 knockdown might be regulated by SFs in Cluster I other than PTBP1.

For Cluster II, centered on the SF ESRP2, we also used AGS cells, which show high ESRP2 expression. Of 34 detectable TA ASEs correlated to ESRP2, 9 showed expected changes in splicing after ESRP2 knockdown (Figure 4B, *P* < 0.0001).

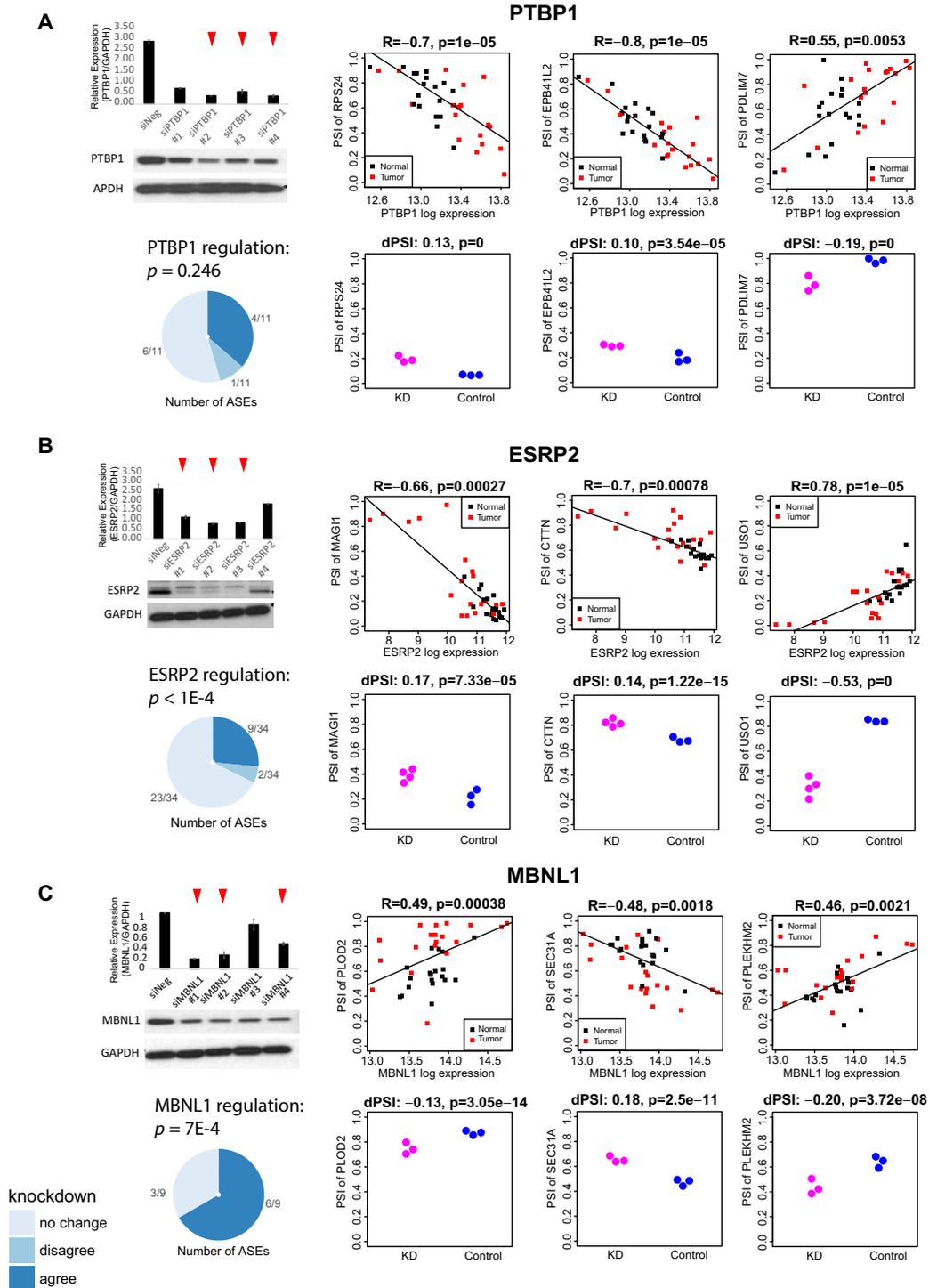For Cluster III, centered on MBNL1, we used HFE-145 cells, a non-neoplastic gastric epithelial cell line with high

Figure 4. siRNA knockdown of individual SFs confirms predicted regulatory targets. Each of the panels (**A**), (**B**) and (**C**) corresponds to knockdown of one SF, as indicated. The bar plots (top left of each panel) show the quantitative RT-PCR expression of the respective SF post knockdown by four individual siRNAs (data in Supplementary Table S14). Red triangles indicate the most effective knockdowns, which were subsequently characterized by RNA-seq. Below each bar plot is the western blot confirming protein-level knockdown of the SF. 'siNeg' indicates data from a non-targeting control siRNA. The pie chart (bottom left of each panel) summarizes the validation results of predicted targets of the SF via RNA-seq experiments; dark blue portion indicates the correlated events showing significant differential splicing in a direction expected from the correlations in RNA-seq from the initial 19 tumor–normal pairs. On the right of each of panels (A), (B) and C are plots showing three example ASEs that were directly or indirectly regulated by the SF. In each panel, the top row plots relationships between the PSIs of 3 TA ASEs and the SF mRNA level across all samples in the initial 19 tumor–normal pairs; *R* indicates the strength of the correlation; *p* indicates the statistical significance of the correlation. The bottom row of plots in each panel shows the knockdown-driven splicing changes in the TA ASEs; dPSI = average change in PSI; KD = knockdown; the *P*-value, reported by the rMATS software, reflects the statistical significance of splicing changes due to the knockdown. For ESRP2, there was a technical replicate of one of the siRNA knockdowns, so there were a total of four knockdown experiments.

MBNL1 expression. Of nine *MBNL1*-correlated ASEs detectable in the knockdown samples, six showed expected changes in splicing (Figure 4C, $P = 0.0007$). Taken together, these results confirm that many of the correlations reflect direct or indirect regulation of the TA ASE by a dysregulated SF (Figure 4, Supplementary Table S9). We therefore have developed a novel methodology for constructing a splicing network model that captures hypotheses regarding regulatory relationships between SFs and ASEs and that are amenable to experimental assessment of biological relevance.

Knocked-down SFs modulate the same biology as their predicted TA ASE targets

Having experimentally established a regulatory relationship between dysregulated SFs and many of their network-model-predicted TA ASE targets, we hypothesized that the SFs and their target TA ASEs modulate the same biological functions. GO analysis showed that the genes affected by TA ASEs are involved in cytoskeletal organization (Figure 2B). We therefore experimentally investigated whether the SFs regulating these TA ASEs also affect this GO category. Cytoskeletal organization refers to organization and distribution of actin and actin-binding proteins that play key roles in focal adhesion formation and in cell migration and invasion. These in turn correlate to metastatic potential (53). To assess the role of the SFs in cell migration, we measured AGS cells' trans-well migration after separate knockdowns of PTBP1, ESRP2 and MBNL1 (three different siRNAs for each target, Figure 5A and B). These three SFs appeared to be dominant regulatory nodes in the gastric-cancer-associated splicing network model (Figure 3B). As expected, knockdown of ESRP2 and MBNL1 led to increased migration in all three knockdown replicates as compared to controls, while loss of PTBP1 led to a reduction in migration in two out of three replicates. We further knocked down each of the SFs with its three targeting siRNAs pooled, and observed the expected significant changes in migration (Supplementary Figure S8). Notably, the knockdown of these three SFs did not affect cell proliferation (Supplementary Figure S9B), thereby showing that the changes observed in cell migration are independent of altered cell growth. Furthermore, 24 h after knockdown of MBNL1 and ESRP2 some cells showed a more mesenchymal morphology compared to control cells (Supplementary Figure S9A). Taken together, our data indicate that dysregulated expression of key SFs (MBNL1, ESRP2 and PTBP1) alters cell migration, possibly via changes in alternative splicing. We therefore concluded that the key SFs and TA ASEs identified in the gastric cancer splicing network (Figure 3B) may play important roles in tumor metastasis.

Confirmation of correlations between SF levels and ASE PSIs in a second dataset

As an independent validation of the findings in our primary gastric tumor dataset, we repeated our analysis in the TCGA gastric cancer samples, comprising 323 tumors and 31 non-malignant gastric epithelium samples (6). As in the initial dataset of 19 tumor–normal pairs, principal component analysis over the TCGA PSI values of the 118 TA

ASEs strongly separated tumor and normal samples (Figure 5C). The TCGA data also confirmed the direction of gene expression changes in the SFs at the hubs of the network model. Most importantly, 45 out of 48 significant positive correlations between the level of an SF transcript and the PSI of a TA ASE in the initial data were positive in the TCGA data, and all 60 negative correlations in the initial data were negative in the TCGA data ($P < 2.2 \times 10^{-16}$, Fisher's exact test, two-sided, Figure 5D, Supplementary Table S8). Thus, the dysregulated splicing network reported here represents a general mechanism underlying gastric cancer progression.

DISCUSSION

This study has identified a comprehensive profile of TA ASEs and their potential regulators in gastric cancer. We combined computational and experimental approaches to build a hub-and-spoke network model of TA ASEs potentially driven by dysregulation of several SFs (*PTBP1*, *SF3B3*, *HNRNPF*, *HNRNPK*, *HNRNPL*, *SQSTM1*, *ESRP2* and *MBNL1*, Figure 3B). Knockdown of three hub SFs (PTBP1, MBNL1 and ESRP2) established causal relationships between these SFs and TA ASEs within the network model (Figure 4). GO analysis indicated that genes harboring gastric TA ASEs are involved in cytoskeletal organization, which is associated with invasion and migration. Indeed, knockdown of the hub SFs (PTBP1, MBNL1 and ESRP2) led to the expected changes in cell migration (Figure 5A and B). Thus, these SFs function in migration and invasion potentially via regulation of the predicted target TA ASEs. Recapitulation of the gastric-cancer-associated splicing network model (Figure 3B) in the TCGA gastric cancer RNA-seq data demonstrates the robustness of our study (Figure 5C and D).

The importance of the three hub SFs (PTBP1, ESRP2 and MBNL1) and a handful of their target ASEs was previously known in cancer biology. *PTBP1* transcripts are known to be upregulated in gastric tumors (40). In a colon cancer cell line, knockdown of PTBP1 led to reduction in some cancer-associated splice forms (67). In breast, colon, renal and gastric cell lines, upregulation of PTBP1 increased proliferation or migration and invasion (67–71). CTTN (cortactin), a splicing target of PTBP1, regulates interactions between components of adherens junctions and plays a critical role in cytoskeletal organization (72). One splice form of this gene promotes invasion and migration in colon cancer cell lines (71). It has been proposed that PTBP1 contributes to the Warburg effect via regulation of the alternative splicing of the pyruvate kinase muscle (*PKM*) gene (68,70). However, neither the 19 tumor–normal pairs nor the TCGA data showed significant elevation of the proposed oncogenic *PKM* splice form.

ESRP2 and its homolog ESRP1 are important in epithelial–mesenchymal transition (EMT) and EMT-driven tumor invasiveness (73). Knockdown of ESRP2 led to increased proliferation and cell migration in renal (74) and in head-and-neck cancer cell lines (75), which is concordant with our observations in gastric cancer cell lines. ESRP2 regulates the splicing of oncogenic transcripts of CD44 and ENAH (76), which we rediscovered in our data.
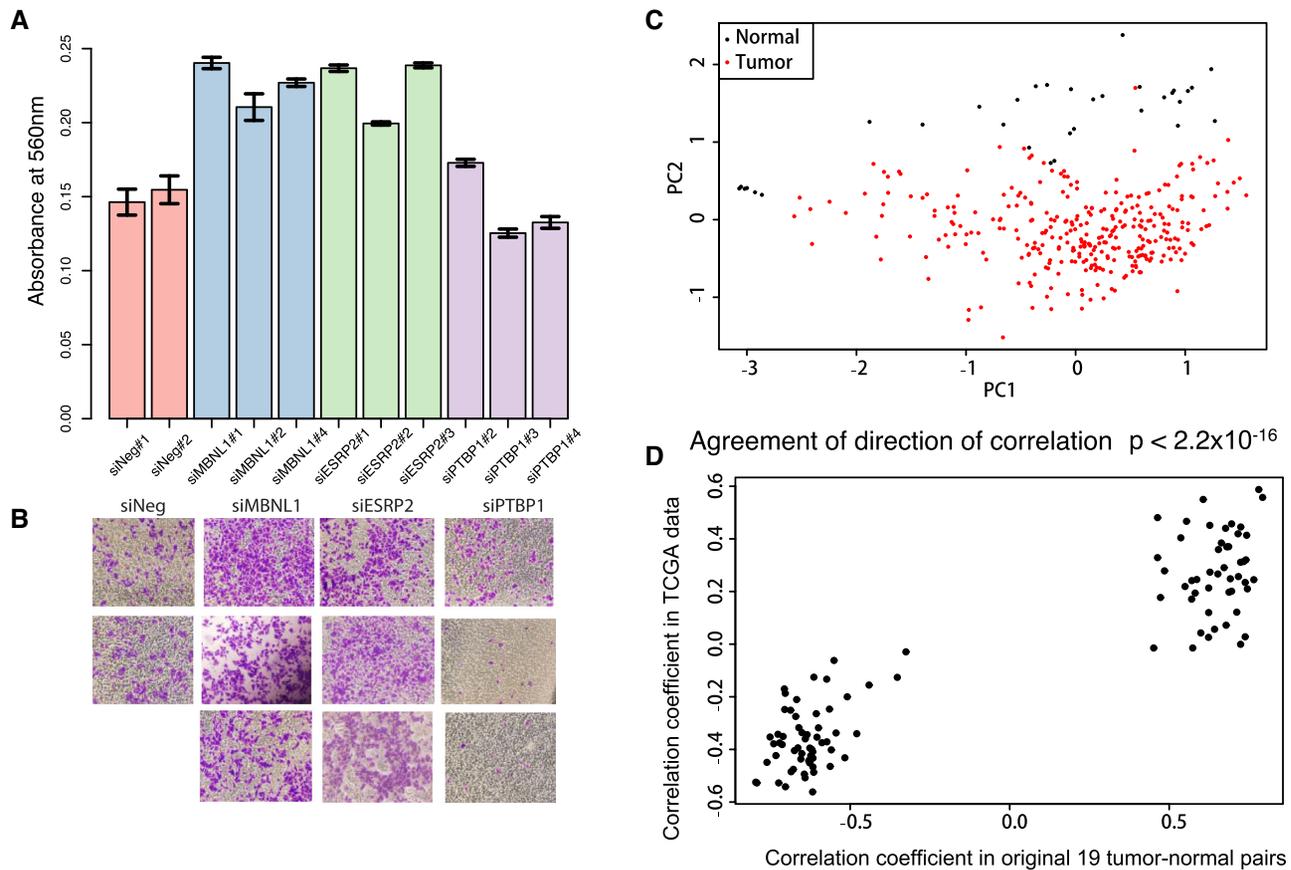
Figure 5. Changes in cell migration after SF knockdown and recapitulation of our computational analysis in TCGA data. (**A**) Bar plot of absorbance at 560 nm, a measure of the number of stained cells after migration, showing increased migration after knockdown of MBNL1 and ESRP2 and decreased migration after knockdown of PTBP1 (see also Supplementary Table S13). (**B**) Representative images of migrated AGS cells after knockdown with a control (siNeg) and/or with siPTBP1, siMBNL1 and siESRP2, showing more migration after knockdown with siMBNL1 and siESRP2 than control cells, and less migration after knockdown with siPTBP1. (**C**) Principal component analysis of the original 118 TA ASEs based on their PSIs in the TCGA data. This plot, like the analogous plot based on PSIs in the original 19 tumor–normal pairs, separates tumor from non-malignant samples. (**D**) Scatter plot of (i) the PSI-to-SF correlation coefficients in the original 19 tumor–normal pairs versus (ii) the corresponding PSI-to-SF correlations in the TCGA data. Each dot represents a single TA ASE–SF pair; only pairs with significant correlations in the initial set of 19 patients are shown; 45 out of 48 significant positive correlations in the initial data were positive in the TCGA data, and all 60 negative correlations were negative in the TCGA data ($P < 2.2 \times 10^{-16}$, Fisher's exact test, two-sided, Supplementary Table S8).

MBNL1 is the least studied SF of the three that we focused on in this study. RNA-seq data from 11 solid tumors (not including gastric cancer) showed reduced levels of *MBNL1* transcripts in several cancer types (77). In breast cancer, MBNL1 suppresses metastasis (64). Furthermore, knockdown of *MBNL1* led to splicing changes in several TA ASEs in multiple cancer types (77).

Taken together, all the above data are concordant with our findings that PTBP1 behaves as an oncogene, that ESRP2 and MBNL1 behave as tumor suppressors and that all three are important regulators of TA ASEs. Beyond re-discovering previously known oncogenic splice forms, our work has substantially expanded knowledge of the repertoire of potentially oncogenic TA ASEs regulated by these SFs.

Both the pathways enriched for the TA ASEs identified in the current study and the changes in migration after knockdown of key SFs suggest that aberrant splicing in gastric cancer promotes invasiveness, an essential hallmark of all cancers (78). Concordant with our findings in gastric can-cer, studies in other cancer types observed splicing changes that contribute to invasiveness. For example, an EGFR variant lacking exon 4 was more highly expressed in metastasized ovarian cancers than in primary tumors (79). The CD44v6 splice form seen in our data also enhanced the survival and proliferation of melanoma cells in brain metastases (80). In breast and colon cancers, SRSF1 promotes cell motility by switching the tyrosine kinase RON to a constitutively active splice form (81). Our study further suggests that global fine-tuning of an alternative splicing program promotes invasiveness in gastric cancer.

A limitation of the current work is the small sample size analyzed. Because of this, we focused on splicing aberrations common to all gastric cancer subtypes together, rather than examining splicing aberration separately for each subtype. However, as a consequence, the TA ASEs reported in this study represent a curated set of high-confidence splicing events common to all gastric cancer subtypes, which were derived from stringent filtering criteria and which were supported by strong evidence in the form of replication in

the TCGA data and in the form of the SF knockdown experiments. In the future, individual high-confidence TA ASEs could be studied experimentally to determine their specific roles in gastric oncogenesis. These high-confidence TA ASEs may be useful as biomarkers that can distinguish tumor from non-tumor tissue, particularly in the operating room where delineating clean surgical margins that demonstrate removal of tumor tissue is critical to achieving positive clinical outcomes (82). Furthermore, we are currently assessing the potential of these TA ASEs as a source of cancer antigens that can be exploited in immunotherapy, as proposed by Epstein *et al.* (83).

Overall, the results presented here provide a broad view of the alternative splicing landscape in gastric cancer. We have identified upstream regulators of these splicing alterations and have shown experimentally that three of these SFs regulate cell migration in gastric cell lines, potentially via their splicing targets. We speculate that the dysregulated splicing network plays a critical role in migration and invasion in gastric cancer. The splicing network model (Figure 3B) is a resource for future studies seeking to identify novel splice-activated oncogenes as biomarkers or as therapeutic targets.

DATA AVAILABILITY

The fastq files for the primary tissue samples and the cell lines have been deposited at the European Genome-phenome Archive at the European Bioinformatics Institute (http://www.ebi.ac.uk/ega/) under accession number EGAS00001002256.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

ACKNOWLEDGEMENTS

We thank Deblina Chatterjee, Tara Davis and Ernesto Guccione for discussions on study design and Niharika Sharma for technical support.

FUNDING

REFERENCES

1. Ferlay,J., Soerjomataram,I., Dikshit,R., Eser,S., Mathers,C., Rebelo,M., Parkin,D.M., Forman,D. and Bray,F. (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer*, **136**, E359–E386.
2. Torre,L.A., Bray,F., Siegel,R.L., Ferlay,J., Lortet-Tieulent,J. and Jemal,A. (2015) Global cancer statistics, 2012. *CA Cancer J. Clin.*, **65**, 87–108.
3. Cervantes,A., Roda,D., Tarazona,N., Rosello,S. and Perez-Fidalgo,J.A. (2013) Current questions for the treatment of advanced gastric cancer. *Cancer Treat. Rev.*, **39**, 60–67.
4. Lauren,P. (1965) The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol. Microbiol. Scand.*, **64**, 31–49.
5. Lei,Z., Tan,I.B., Das,K., Deng,N., Zouridis,H., Pattison,S., Chua,C., Feng,Z., Guan,Y.K., Ooi,C.H. *et al.* (2013) Identification of molecular subtypes of gastric cancer with different responses to PI3-kinase inhibitors and 5-fluorouracil. *Gastroenterology*, **145**, 554–565.
6. The Cancer Genome Atlas Research Network. (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*, **513**, 202–209.
7. Cristescu,R., Lee,J., Nebozhyn,M., Kim,K.M., Ting,J.C., Wong,S.S., Liu,J., Yue,Y.G., Wang,J., Yu,K. *et al.* (2015) Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat. Med.*, **21**, 449–456.
8. McLean,M.H. and El-Omar,E.M. (2014) Genetics of gastric cancer. *Nat Rev. Gastroenterol. Hepatol.*, **11**, 664–674.
9. Pan,Q., Shai,O., Lee,L.J., Frey,B.J. and Blencowe,B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, **40**, 1413–1415.
10. Wang,E.T., Sandberg,R., Luo,S., Khrebtukova,I., Zhang,L., Mayr,C., Kingsmore,S.F., Schroth,G.P. and Burge,C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
11. Wahl,M.C. and Luhrmann,R. (2015) SnapShot: spliceosome dynamics I. *Cell*, **161**, 1474e1.
12. Kornblihtt,A.R., Schor,I.E., Allo,M., Dujardin,G., Petrillo,E. and Munoz,M.J. (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.*, **14**, 153–165.
13. Wang,Z. and Burge,C.B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, **14**, 802–813.
14. Kalsotra,A. and Cooper,T.A. (2011) Functional consequences of developmentally regulated alternative splicing. *Nat. Rev. Genet.*, **12**, 715–729.
15. Yeo,G., Holste,D., Kreiman,G. and Burge,C.B. (2004) Variation in alternative splicing across human tissues. *Genome Biol.*, **5**, R74.
16. Garcia-Blanco,M.A., Baraniak,A.P. and Lasda,E.L. (2004) Alternative splicing in disease and therapy. *Nat. Biotechnol.*, **22**, 535–546.
17. Sveen,A., Kilpinen,S., Ruusulehto,A., Lothe,R.A. and Skotheim,R.I. (2016) Aberrant RNA splicing in cancer: expression changes and driver mutations of splicing factor genes. *Oncogene*, **35**, 2413–2427.
18. Graubert,T.A., Shen,D., Ding,L., Okeyo-Owuor,T., Lunn,C.L., Shao,J., Krysiak,K., Harris,C.C., Koboldt,D.C., Larson,D.E. *et al.* (2011) Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat. Genet.*, **44**, 53–57.
19. Papaemmanuil,E., Cazzola,M., Boultwood,J., Malcovati,L., Vyas,P., Bowen,D., Pellagatti,A., Wainscoat,J.S., Hellstrom-Lindberg,E., Gambacorti-Passerini,C. *et al.* (2011) Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N. Engl. J. Med.*, **365**, 1384–1395.
20. Wang,L., Lawrence,M.S., Wan,Y., Stojanov,P., Sougnez,C., Stevenson,K., Werner,L., Sivachenko,A., DeLuca,D.S., Zhang,L. *et al.* (2011) SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.*, **365**, 2497–2506.
21. Dvinge,H., Kim,E., Abdel-Wahab,O. and Bradley,R.K. (2016) RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer*, **16**, 413–430.

22. Venables,J.P., Klinck,R., Koh,C., Gervais-Bird,J., Bramard,A., Inkel,L., Durand,M., Couture,S., Froehlich,U., Lapointe,E. *et al.* (2009) Cancer-associated regulation of alternative splicing. *Nat. Struct. Mol. Biol.*, **16**, 670–676.

23. Anczukow,O., Akerman,M., Clery,A., Wu,J., Shen,C., Shirole,N.H., Raimer,A., Sun,S., Jensen,M.A., Hua,Y. *et al.* (2015) SRSF1-regulated alternative splicing in breast cancer. *Mol. Cell*, **60**, 105–117.

24. Adler,A.S., McCleland,M.L., Yee,S., Yaylaoglu,M., Hussain,S., Cosino,E., Quinones,G., Modrusan,Z., Seshagiri,S., Torres,E. *et al.* (2014) An integrative analysis of colon cancer identifies an essential function for PRPF6 in tumor growth. *Genes Dev.*, **28**, 1068–1084.

25. Takahashi,H., Nishimura,J., Kagawa,Y., Kano,Y., Takahashi,Y., Wu,X., Hiraki,M., Hamabe,A., Konno,M., Haraguchi,N. *et al.* (2015) Significance of polypyrimidine tract-binding protein 1 expression in colorectal cancer. *Mol. Cancer Ther.*, **14**, 1705–1716.

26. Nakazawa,M., Antonarakis,E.S. and Luo,J. (2014) Androgen receptor splice variants in the era of enzalutamide and abiraterone. *Horm. Cancer*, **5**, 265–273.

27. Ng,K.P., Hillmer,A.M., Chuah,C.T., Juan,W.C., Ko,T.K., Teo,A.S., Ariyaratne,P.N., Takahashi,N., Sawada,K., Fei,Y. *et al.* (2012) A common BIM deletion polymorphism mediates intrinsic resistance and inferior responses to tyrosine kinase inhibitors in cancer. *Nat. Med.*, **18**, 521–528.

28. Poulikakos,P.I., Persaud,Y., Janakiraman,M., Kong,X., Ng,C., Moriceau,G., Shi,H., Atefi,M., Titz,B., Gabay,M.T. *et al.* (2011) RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). *Nature*, **480**, 387–390.

29. Ladomery,M.R., Harper,S.J. and Bates,D.O. (2007) Alternative splicing in angiogenesis: the vascular endothelial growth factor paradigm. *Cancer Lett.*, **249**, 133–142.

30. Boise,L.H., Gonzalez-Garcia,M., Postema,C.E., Ding,L., Lindsten,T., Turka,L.A., Mao,X., Nunez,G. and Thompson,C.B. (1993) bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell*, **74**, 597–608.

31. Bauman,J.A., Li,S.D., Yang,A., Huang,L. and Kole,R. (2010) Anti-tumor activity of splice-switching oligonucleotides. *Nucleic Acids Res.*, **38**, 8348–8356.

32. Venables,J.P. (2004) Aberrant and alternative splicing in cancer. *Cancer Res.*, **64**, 7647–7654.

33. Cha,J.Y., Lambert,Q.T., Reuther,G.W. and Der,C.J. (2008) Involvement of fibroblast growth factor receptor 2 isoform switching in mammary oncogenesis. *Mol. Cancer Res.*, **6**, 435–445.

34. Eswaran,J., Horvath,A., Godbole,S., Reddy,S.D., Mudvari,P., Ohshiro,K., Cyanam,D., Nair,S., Fuqua,S.A., Polyak,K. *et al.* (2013) RNA sequencing of cancer reveals novel splicing alterations. *Sci. Rep.*, **3**, 1689.

35. Ferreira,P.G., Jares,P., Rico,D., Gomez-Lopez,G., Martinez-Trillos,A., Villamor,N., Ecker,S., Gonzalez-Perez,A., Knowles,D.G., Monlong,J. *et al.* (2014) Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Res.*, **24**, 212–226.

36. Liu,J., Lee,W., Jiang,Z., Chen,Z., Jhunjhunwala,S., Haverty,P.M., Gnad,F., Guan,Y., Gilbert,H.N., Stinson,J. *et al.* (2012) Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. *Genome Res.*, **22**, 2315–2327.

37. Brooks,A.N., Choi,P.S., de Waal,L., Sharifnia,T., Imielinski,M., Saksena,G., Pedamallu,C.S., Sivachenko,A., Rosenberg,M., Chmielecki,J. *et al.* (2014) A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events. *PLoS One*, **9**, e87361.

38. DeBoever,C., Ghia,E.M., Shepard,P.J., Rassenti,L., Barrett,C.L., Jepsen,K., Jamieson,C.H., Carson,D., Kipps,T.J. and Frazer,K.A. (2015) Transcriptome sequencing reveals potential mechanism of cryptic 3′ splice site selection in SF3B1-mutated cancers. *PLoS Comput. Biol.*, **11**, e1004105.

39. Alsafadi,S., Houy,A., Battistella,A., Popova,T., Wassef,M., Henry,E., Tirode,F., Constantinou,A., Piperno-Neumann,S., Roman-Roman,S. *et al.* (2016) Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat. Commun.*, **7**, 10615.

40. Armero,V.E.S., Tremblay,M.P., Allaire,A., Boudreault,S., Martenon-Brodeur,C., Duval,C., Durand,M., Lapointe,E.,

Thibault,P., Tremblay-Letourneau,M. *et al.* (2017) Transcriptome-wide analysis of alternative RNA splicing events in Epstein–Barr virus-associated gastric carcinomas. *PLoS One*, **12**, e0176880.

41. Liu,J., McCleland,M., Stawiski,E.W., Gnad,F., Mayba,O., Haverty,P.M., Durinck,S., Chen,Y.J., Klijn,C., Jhunjhunwala,S. *et al.* (2014) Integrated exome and transcriptome sequencing reveals ZAK isoform usage in gastric cancer. *Nat. Commun.*, **5**, 3830.

42. Shi,Y., Chen,Z., Gao,J., Wu,S., Gao,H. and Feng,G. (2018) Transcriptome-wide analysis of alternative mRNA splicing signature in the diagnosis and prognosis of stomach adenocarcinoma. *Oncol. Rep.*, **40**, 2014–2022.

43. Miura,K., Fujibuchi,W. and Sasaki,I. (2011) Alternative pre-mRNA splicing in digestive tract malignancy. *Cancer Sci.*, **102**, 309–316.

44. Shen,S., Park,J.W., Lu,Z.X., Lin,L., Henry,M.D., Wu,Y.N., Zhou,Q. and Xing,Y. (2014) rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E5593–E5601.

45. Liu,R., Loraine,A.E. and Dickerson,J.A. (2014) Comparisons of computational methods for differential alternative splicing detection using RNA-seq in plant systems. *BMC Bioinformatics*, **15**, 364.

46. Giulietti,M., Piva,F., D'Antonio,M., D'Onorio De Meo,P., Paoletti,D., Castrignano,T., D'Erchia,A.M., Picardi,E., Zambelli,F., Principato,G. *et al.* (2013) SpliceAid-F: a database of human splicing factors and their RNA-binding sites. *Nucleic Acids Res.*, **41**, D125–D131.

47. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

48. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.

49. Wang,W., Dong,L.P., Zhang,N. and Zhao,C.H. (2014) Role of cancer stem cell marker CD44 in gastric cancer: a meta-analysis. *Int. J. Clin. Exp. Med.*, **7**, 5059–5066.

50. Tsai,F.D., Lopes,M.S., Zhou,M., Court,H., Ponce,O., Fiordalisi,J.J., Gierut,J.J., Cox,A.D., Haigis,K.M. and Philips,M.R. (2015) K-Ras4A splice variant is widely expressed in cancer and uses a hybrid membrane-targeting motif. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 779–784.

51. Braeutigam,C., Rago,L., Rolke,A., Waldmeier,L., Christofori,G. and Winter,J. (2014) The RNA-binding protein Rbfox2: an essential regulator of EMT-driven alternative splicing and a mediator of cellular invasion. *Oncogene*, **33**, 1082–1092.

52. Tsunoda,T., Inada,H., Kalembeyi,I., Imanaka-Yoshida,K., Sakakibara,M., Okada,R., Katsuta,K., Sakakura,T., Majima,Y. and Yoshida,T. (2003) Involvement of large tenascin-C splice variants in breast cancer progression. *Am. J. Pathol.*, **162**, 1857–1867.

53. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 15545–15550.

54. Yu,G., Wang,L.G., Han,Y. and He,Q.Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, **16**, 284–287.

55. Gardina,P.J., Clark,T.A., Shimada,B., Staples,M.K., Yang,Q., Veitch,J., Schweitzer,A., Awad,T., Sugnet,C., Dee,S. *et al.* (2006) Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics*, **7**, 325.

56. Thorsen,K., Sorensen,K.D., Brems-Eskildsen,A.S., Modin,C., Gaustadnes,M., Hein,A.M., Kruhoffer,M., Laurberg,S., Borre,M., Wang,K. *et al.* (2008) Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. *Mol. Cell. Proteomics*, **7**, 1214–1224.

57. Abubaker,J., Bavi,P., Al-Haqawi,W., Sultana,M., Al-Harbi,S., Al-Sanea,N., Abduljabbar,A., Ashari,L.H., Alhomoud,S., Al-Dayel,F. *et al.* (2009) Prognostic significance of alterations in KRAS isoforms KRAS-4A/4B and KRAS mutations in colorectal carcinoma. *J. Pathol.*, **219**, 435–445.

58. Shapiro,I.M., Cheng,A.W., Flytzanis,N.C., Balsamo,M., Condeelis,J.S., Oktay,M.H., Burge,C.B. and Gertler,F.B. (2011) An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet.*, **7**, e1002218.

59. Tokunaga,A., Nishi,K., Matsukura,N., Tanaka,N., Onda,M., Shirota,A., Asano,G. and Hayashi,K. (1986) Estrogen and progesterone receptors in gastric cancer. *Cancer*, **57**, 1376–1379.

60. Ur Rahman,M.S. and Cao,J. (2016) Estrogen receptors in gastric cancer: advances and perspectives. *World J. Gastroenterol.*, **22**, 2475–2482.

61. David,C.J., Chen,M., Assanah,M., Canoll,P. and Manley,J.L. (2010) HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature*, **463**, 364–368.

62. Bates,D.O., Morris,J.C., Oltean,S. and Donaldson,L.F. (2017) Pharmacology of modulators of alternative splicing. *Pharmacol. Rev.*, **69**, 63–79.

63. Naro,C. and Sette,C. (2013) Phosphorylation-mediated regulation of alternative splicing in cancer. *Int. J. Cell Biol.*, **2013**, 151839.

64. Moore,M.J., Wang,Q., Kennedy,C.J. and Silver,P.A. (2010) An alternative splicing network links cell-cycle control to apoptosis. *Cell*, **142**, 625–636.

65. Shi,J., Qian,W., Yin,X., Iqbal,K., Grundke-Iqbal,I., Gu,X., Ding,F., Gong,C.X. and Liu,F. (2011) Cyclic AMP-dependent protein kinase regulates the alternative splicing of tau exon 10: a mechanism involved in tau pathology of Alzheimer disease. *J. Biol. Chem.*, **286**, 14639–14648.

66. Gu,J., Shi,J., Wu,S., Jin,N., Qian,W., Zhou,J., Iqbal,I.G., Iqbal,K., Gong,C.X. and Liu,F. (2012) Cyclic AMP-dependent protein kinase regulates 9G8-mediated alternative splicing of tau exon 10. *FEBS Lett.*, **586**, 2239–2244.

67. Hollander,D., Donyo,M., Atias,N., Mekahel,K., Melamed,Z., Yannai,S., Lev-Maor,G., Shilo,A., Schwartz,S., Barshack,I. *et al.* (2016) A network-based analysis of colon cancer splicing changes reveals a tumorigenesis-favoring regulatory pathway emanating from ELK1. *Genome Res.*, **26**, 541–553.

68. Sugiyama,T., Taniguchi,K., Matsuhashi,N., Tajirika,T., Futamura,M., Takai,T., Akao,Y. and Yoshida,K. (2016) MiR-133b inhibits growth of human gastric cancer cells by silencing pyruvate kinase muscle-splicer polypyrimidine tract-binding protein 1. *Cancer Sci.*, **107**, 1767–1775.

69. Jiang,J., Chen,X., Liu,H., Shao,J., Xie,R., Gu,P. and Duan,C. (2017) Polypyrimidine tract-binding protein 1 promotes proliferation, migration and invasion in clear-cell renal cell carcinoma by regulating alternative splicing of PKM. *Am. J. Cancer Res.*, **7**, 245–259.

70. He,X., Arslan,A.D., Ho,T.T., Yuan,C., Stampfer,M.R. and Beck,W.T. (2014) Involvement of polypyrimidine tract-binding protein (PTBP1) in maintaining breast cancer cell growth and malignant properties. *Oncogenesis*, **3**, e84.

71. Wang,Z.N., Liu,D., Yin,B., Ju,W.Y., Qiu,H.Z., Xiao,Y., Chen,Y.J., Peng,X.Z. and Lu,C.M. (2017) High expression of PTBP1 promote invasion of colorectal cancer by alternative splicing of cortactin. *Oncotarget*, **8**, 36185–36202.

72. MacGrath,S.M. and Koleske,A.J. (2012) Cortactin in cell migration and cancer at a glance. *J. Cell Sci.*, **125**, 1621–1626.

73. De Craene,B. and Berx,G. (2013) Regulatory networks defining EMT during cancer initiation and progression. *Nat. Rev. Cancer*, **13**, 97–110.

74. Mizutani,A., Koinuma,D., Seimiya,H. and Miyazono,K. (2016) The Arkadia-ESRP2 axis suppresses tumor progression: analyses in clear-cell renal cell carcinoma. *Oncogene*, **35**, 3514–3523.

75. Ishii,H., Saitoh,M., Sakamoto,K., Kondo,T., Katoh,R., Tanaka,S., Motizuki,M., Masuyama,K. and Miyazawa,K. (2014) Epithelial splicing regulatory proteins 1 (ESRP1) and 2 (ESRP2) suppress cancer cell motility via different mechanisms. *J. Biol. Chem.*, **289**, 27386–27399.

76. Warzecha,C.C., Sato,T.K., Nabet,B., Hogenesch,J.B. and Carstens,R.P. (2009) ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol. Cell*, **33**, 591–601.

77. Sebestyen,E., Singh,B., Minana,B., Pages,A., Mateo,F., Pujana,M.A., Valcarcel,J. and Eyras,E. (2016) Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res.*, **26**, 732–744.

78. Hanahan,D. and Weinberg,R.A. (2000) The hallmarks of cancer. *Cell*, **100**, 57–70.

79. Wang,H., Zhou,M., Shi,B., Zhang,Q., Jiang,H., Sun,Y., Liu,J., Zhou,K., Yao,M., Gu,J. *et al.* (2011) Identification of an exon 4-deletion variant of epidermal growth factor receptor with increased metastasis-promoting capacity. *Neoplasia*, **13**, 461–71.

80. Marzese,D.M., Liu,M., Huynh,J.L., Hirose,H., Donovan,N.C., Huynh,K.T., Kiyohara,E., Chong,K., Cheng,D., Tanaka,R. *et al.* (2015) Brain metastasis is predetermined in early stages of cutaneous melanoma by CD44v6 expression through epigenetic regulation of the spliceosome. *Pigment Cell Melanoma Res*, **28**, 82–93.

81. Ghigna,C., Giordano,S., Shen,H., Benvenuto,F., Castiglioni,F., Comoglio,P.M., Green,M.R., Riva,S. and Biamonti,G. (2005) Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Mol. Cell*, **20**, 881–90.

82. Epstein,D.M., Rozen,S.G. and Cheng,S. (2019) Method of detecting cancer tissue. Singapore Patent Application No. 10201904540S,21 May 2019.

83. Epstein,D.M., Wong,S.C. and Lee,R.T.H. (2019) Method and system for identifying and validating shared candidate antigens and shared antigen-specific T lymphocyte pairs. Singapore Patent Application No. 10201911055U, 22 November 2019.