

Functional requirement of alternative splicing in epithelial-mesenchymal transition of pancreatic circulating tumor

Lixing Xiao,^{1,3} Haoxiu Sun,^{2,3} Rui Cheng,^{1,3} Rongrong Yang,¹ Xiyun Jin,¹ Zhaochun Xu,¹ Yideng Cai,¹ Yuexin Yang,¹ Fenglan Pang,¹ Guangfu Xue,¹ Pingping Wang,² Qinghua Jiang,^{1,2} and Huan Nie¹

¹School of Life Science and Technology, Harbin Institute of Technology, Harbin 150000, China; ²School for Interdisciplinary Medicine and Engineering, Harbin Medical University, Harbin 150076, China

Circulating tumor cells (CTCs) that undergo epithelial-to-mesenchymal transition (EMT) can provide valuable information regarding metastasis and potential therapies. However, current studies on the EMT overlook alternative splicing. Here, we used single-cell full-length transcriptome data and mRNA sequencing of CTCs to identify stage-specific alternative splicing of partial EMT and mesenchymal states during pancreatic cancer metastasis. We classified definitive tumor and normal epithelial cells via genetic aberrations and demonstrated dynamic changes in the epithelial-mesenchymal continuum in both epithelial cancer cells and CTCs. We provide the landscape of alternative splicing in CTCs at different stages of EMT, uncovering cell-type-specific splicing patterns and splicing events in cell surface proteins suitable for therapies. We show that *MBNL1* governs cell fate through alternative splicing independently of changes in gene expression and affects the splicing pattern during EMT. We found a high frequency of events that contained multiple premature termination codons and were enriched with C and G nucleotides in close proximity, which influence the likelihood of stop codon readthrough and expand the range of potential therapeutic targets. Our study provides insights into the EMT transcriptome's dynamic changes and identifies potential diagnostic and therapeutic targets in pancreatic cancer.

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is a highly aggressive and lethal malignancy because of the lack of early diagnosis and limited response to treatments.¹ Metastatic recurrence is a significant problem in patients with PDAC, with a 5-year survival rate of only 9% for fatal disease and metastatic disease emerging in nearly 98% of patients.² Metastatic spread is a highly inefficient and complex multistep process.³ The formation of tumor metastases first requires circulating tumor cells (CTCs) to colonize metastatic organs or tissues through the circulatory system.⁴ CTCs are vital components of liquid biopsies for the diagnosis of residual cancer, monitoring of therapy response, and prognosis of recurrence.⁵ Epithelial-to-mesenchymal transition (EMT) promotes the generation of CTCs in epithelial cancers by increasing tumor cell invasiveness and motility to promote tumor cell intravasation and facilitate tumor cell survival in the peripheral system.⁶

EMT is involved in cancer invasion and metastasis, which is the process by which polarized epithelial cells lose their characteristic intercellular connections and polarity and acquire characteristics of mesenchymal cells.⁷ Experimental and clinical studies have highlighted that CTC with phenotypic hallmarks of EMT plays a critical role in the metastatic and recurrence of solid malignancy.⁷ EMT is now well documented not to be a linear process, but rather a dynamic one with the existence of partial EMT (pEMT) states between the epithelial and mesenchymal poles.⁸ Cells transitioning in these pEMT states exhibit stem-cell-like properties and display high plasticity, which is critical for metastasis and drug resistance.⁹ RNA alternative splicing (AS) drives multiple malignant phenotypes, including aggressiveness, angiogenesis, and abnormal metabolism, it is not surprising that it simultaneously affects different aspects of EMT progression.^{10,11} In line with this notion, AS events have been reported to play a functional role in EMT. For instance, Brown et al.¹² showed that CD44 isoform switching is essential for EMT to occur and for cancer cells to metastasize. Reciprocally, the RNA-binding protein (RBP) AKAP8 suppresses tumor metastasis by antagonizing EMT-associated AS.¹³ Together, these findings imply a role for AS in EMT, but, until now, this phenomenon has not been assessed in PDAC.

AS is an essential mechanism that acts on the pre-mRNA transcripts of more than 90% of mammals, allowing cells to create different versions of transcripts and proteins by selectively including or excluding different exons during RNA processing.¹⁴ The parsing of unique

Received 28 July 2023; accepted 19 January 2024;
<https://doi.org/10.1016/j.omtn.2024.102129>.

³These authors contributed equally

Correspondence: Pingping Wang, School for Interdisciplinary Medicine and Engineering, Harbin Medical University, Harbin 150076, China.

E-mail: wangpingping@hrbmu.edu.cn

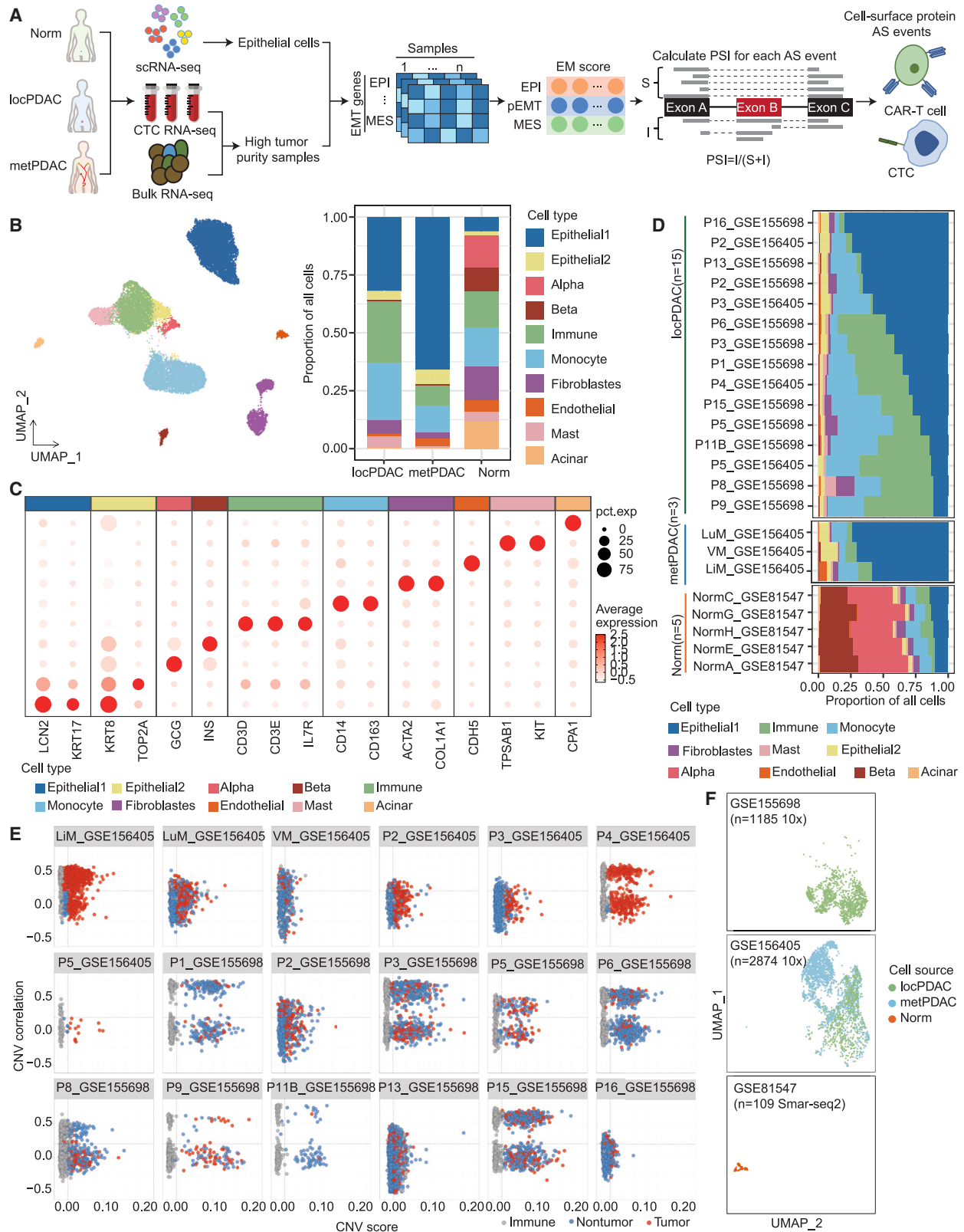
Correspondence: Qinghua Jiang, School of Life Science and Technology, Harbin Institute of Technology, Harbin 150000, China.

E-mail: qhjiang@hit.edu.cn

Correspondence: Huan Nie, School of Life Science and Technology, Harbin Institute of Technology, Harbin 150000, China.

E-mail: nh1212@hit.edu.cn





(legend on next page)

alterations of tumor-specific RNA splicing in PDAC can reveal the key elements involved in the metastatic changes, and these elements can be exploited for novel immunotherapeutic strategies.¹⁵ Progression and metastasis are also influenced by complex and dynamic features in tumor AS. However, efforts to understand PDAC progression and metastasis through alternative splicing analysis of EMT progression in CTCs have not been reported.

Here, we analyzed the epithelial, pEMT, and mesenchymal status of human PDAC CTCs to identify the full EMT continuum and create a dynamic landscape by profiling AS from the epithelial stage, through pEMT, to the mesenchymal stage. We demonstrated that the EMT program of CTCs exhibits significant heterogeneity. We found that EMT is accompanied by a significant AS modality switch, which is orchestrated mainly by the splicing regulator. These findings provide insight into RNA processing during EMT progression, enhance our understanding of the molecular mechanisms involved in PDAC metastasis, and offer a resource for future investigations of AS regulators critical for tumor progression.

RESULTS

Single-cell resolution identified epithelial cells from PDAC tumor and normal samples

We sought to assess the EMT status of CTCs during metastasis, reasoning that this may unveil the diverse regulation of metastasis in pancreatic cancer and reveal potential therapeutic targets. To comprehensively understand the dynamics of cancer spread from the primary site through circulation to metastatic sites, we collected data on localized PDAC (locPDAC), metastatic PDAC (metPDAC), and healthy samples (Table S1). We explored intrinsic characteristics of locPDAC and metPDAC through a comparative analysis between epithelial cells and CTCs (Figure 1A). We first needed to obtain pure cancer epithelial cells along with normal pancreatic duct cells. Single-cell RNA sequencing (RNA-seq) data were obtained from a total of 33 samples, which included 19 localized cases and 3 metastases. These metastases stemmed from the liver, lung, and surgical vaginal apex in human PDAC. Additionally, we obtained 11 normal samples. A standard single-cell analysis process was performed to identify the major cell types in PDAC datasets. To identify cell types on the basis of the similarity of their transcriptomic signatures between cells, we performed clustering analysis using uniform manifold approximation and projection (UMAP). Data integration showed segregation into two main cell clusters (Figures S1A–S1C). To further understand whether this resulted from a batch effect, we analyzed the cellular composition of these two groups. Our findings show that cluster 1 predominantly comprises endothelial cells and fibroblasts, whereas these cell types are rare in cluster 2 (Figures S1D–S1H).

This suggests that the observed disparity is more likely attributable to cell type than to a batch effect. After quality-control filtering, we cataloged 35,545 high-quality single cells into 10 distinct cell clusters annotated with canonical marker gene expression (Figures 1B and 1C). We noticed there were two different clusters of epithelial cells and conducted further analysis on these distinct cell types. As epithelial 2 cluster shows high expression of the cell-cycle-associated *TOP2A* gene, we investigated the cell cycle status of individual cells within this cluster in greater detail. We found that not all cells in this cluster were in a proliferating state. Hence, our categorization of this cluster is not solely driven by cell cycle effects (Figure S1I). The pie chart of proportions (Smartseq2 vs. 10x) for these two different epithelial cell clusters shows that the ratio of Smartseq2 to 10x in these two cell types is similar, indicating that their differences are not likely due to the two different protocols (Figure S1J). The bar plot, however, reveals significant differences in the number of duct cells between these two cell types, with epithelial 1 showing a higher yield of cells from normal tissues (Figure S1K). We also performed a Gene Ontology (GO) enrichment analysis on the differentially expressed genes (DEGs) within these two epithelial cell types. The results suggested that epithelial 1 is correlated with peptidase activity, while epithelial 2 appears to be engaged with inflammation-related biological processes (Figures S1L–S1M). Therefore, we retained only epithelial 1 cells for subsequent analysis.

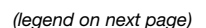
The yield of the epithelial transcriptome varied by histological subtype. To ensure accuracy in subsequent analyses, we removed any samples with an epithelial percentage less than 10% (Figure 1D). To separate defined tumor cells from a potentially non-malignant population, we obtained epithelial cells from each tumor tissue, using genetic aberrations by inferring copy number variations (CNVs) from gene expression data (Figures S1N–S1P; materials and methods).¹⁶ We calculated the correlation of each cell's CNV segment expression with the average CNV segment expression of the top 5% samples on the basis of the CNV score. We identified 4,059 malignant epithelial cells, defined by a correlation or CNV score greater than 0.2 (Figures 1E, S1Q, and S1R). Epithelial cancer cells in tumor tissues were removed because of their ambiguous identities arising from the lack of CNV. In the subsequent tumor cell analysis, we retained the inferred malignant tumor cells and epithelial cells from the normal sample (Figure 1F).

Presentation of EMT gene signatures reveals dynamic changes in CTCs during the metastasis process

Having obtained cancer cells and normal epithelial cells of PDAC, we ventured forth to perform a deeper characterization of EMT dynamics of epithelial cells from the primary site, through the circulation to the metastasis site. Recent studies have demonstrated that a

Figure 1. Cell types identified in PDAC on the basis of single-cell RNA sequencing

(A) Conceptual overview of study design and samples used. (B) UMAP plot of all single-cell transcriptomes color coded by main cell type (left) and bar plot showing proportions of cell types in pooled with localized PDAC (locPDAC), metastatic PDAC (metPDAC), and normal samples (right). (C) Bubble plot showing highly expressed marker genes in each cell type, with cell types in rows and genes in columns. The size of each bubble represents the percent of cells expressing marker and color represents the level of average expression. (D) Quantification of main cell types per patient. (E) Clarification of cancer cells on the basis of CNV inference. The two-dimensional plot between the mean squares and correlation of the CNV signal for each sample. Red and blue colors represent single cells defined as malignant and non-malignant, respectively. Gray represents immune cells. (F) UMAP of all epithelial single-cell transcriptomes color coded by the cell source.



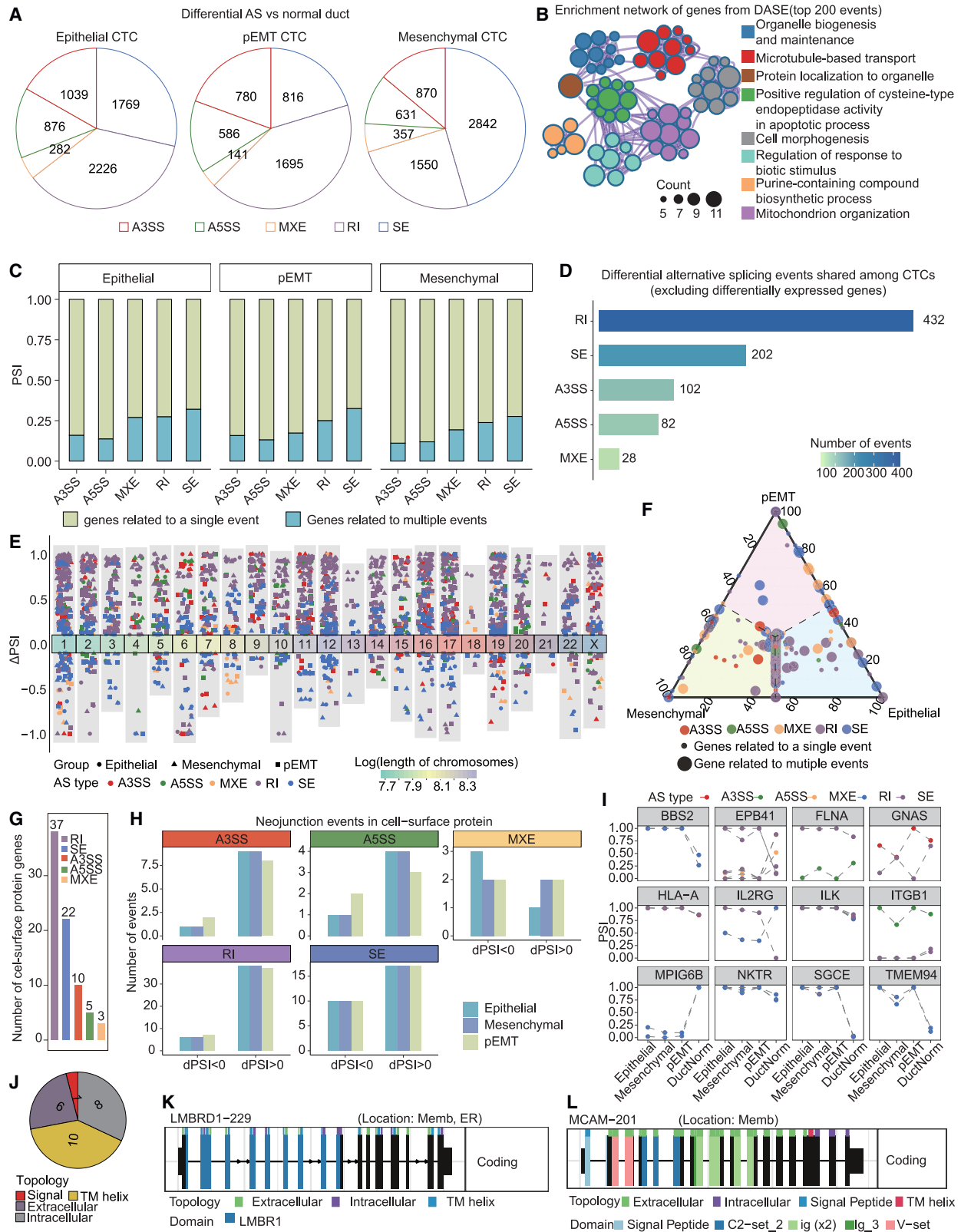
purity of 50% can be achieved with the CTC-chip using the peripheral blood of patients with metastatic cancer (lung, prostate, breast, colon, and pancreatic cancers).¹⁷ We assessed the tumor purity of samples and inferred the immune cell composition within the CTC samples. We grouped B cells, T cells, natural killer (NK) cells, plasma cells, eosinophils, neutrophils, and monocytes collectively as total white blood cells (WBCs). The results indicated that the WBC content of CTCs in high tumor purity was significantly lower compared with that in normal peripheral blood mononuclear cells (PBMCs) (Figures S2A and S2B; Table S1; materials and methods). To identify the EMT gene preferentially expressed in a particular cell cluster using CTCs from pancreatic cancer patients, we next performed principal-component analysis (PCA) using 30 CTCs with relatively high purity and 15 normal PBMC samples. The results showed that CTCs were separated from normal blood cells of healthy donors, and the first and second principal components explained 55% and 19% of the variability (74% in total; Figure 2A). This suggests that the expression profiles of CTCs were similar between localized and metastatic tumors. To investigate the dynamic changes in epithelial and mesenchymal markers of CTCs during cancer metastasis, we compared EMT genes (identified by Tan et al.¹⁸) in CTCs and epithelial cancer cells (materials and methods). We separately screened EMT genes that were differentially expressed in epithelial cancer cells and CTCs (Figures S2C and S2D; Table S2). We found 28 differentially expressed EMT genes in duct PDAC, consisting of 11 epithelial genes and 17 mesenchymal genes (Figure 2B). Most of the differentially expressed EMT genes demonstrated an upregulation trend, with more mesenchymal genes being upregulated than epithelial genes. Moreover, out of all the differentially expressed EMT signatures, 10 epithelial genes, and 8 mesenchymal genes were found to be differentially expressed in both epithelial duct samples and CTCs (Figures 2C and S2E). The result showed opposite expression patterns of the same gene in CTCs and epithelial cells: *TACSTD2*, *EHF*, and *OR7E14P* in epithelial genes and *VCAM1*, *DDR2*, *SFRP1*, and *ECM2* in mesenchymal genes. This implies EMT dynamics in cancer epithelial cells. We observed that differentially expressed EMT genes were shared across multiple biological processes. Specifically, the differentially expressed EMT genes were significantly enriched in terms associated with the activation of the EMT process during metastasis. This includes activities related to cadherin binding, integrin binding, and functioning as extracellular matrix structural constituents (Figure 2D).

To evaluate the epithelial stability of PDAC epithelial ductal cells and CTCs, we calculated the epithelial score (EPI) and mesenchymal score (MES) for each sample on the basis of the differentially expressed EMT genes (materials and methods). Our results show an increase in both the EPI and MESs in the CTCs, with the CTC samples tending to have high MESs and low EPIs, which is not significant in ductal samples (Figures 2E and S2F). To verify the universality of this phenomenon, we carried out the same analysis process with these differentially expressed EMT genes in two other datasets (one is the CTC dataset, and the other is the PDAC dataset from The Cancer Genome Atlas [TCGA]). The results were consistent with our analysis, showing an increase in both the EPI and MES in the CTCs, with the CTC samples tending to have higher MESs and lower EPIs, and in the TCGA datasets, while the EPIs remained unchanged, the range of MESs changed greatly, implying that the degree of mesenchymal differentiation in tumor tissues is heterogeneous (Figure S2G). These findings confirm that entry into circulation is a complex event involving alterations in EMT signatures that are crucial for the distant spread of cancer cells.

To measure the overall EMT dynamics from the epithelium to the mesenchyme, we generated an EM score for each sample on the basis of these differentially expressed EMT genes (materials and methods). The distribution of EM scores for both CTC and epithelial duct samples shows that scores for CTC are concentrated primarily between 0 and 0.7, whereas epithelial duct samples predominantly score near -1 and 1 (Figure 2F). Then, we subdivided them into epithelial, pEMT, and mesenchymal phenotypes on the basis of the cutoff at the intersection of the epithelial duct and CTC density curve (Figure 2G; Table S3). Normal PBMCs predominantly exhibit a pEMT phenotype, while all normal ducts present an epithelial phenotype (Figures 2H and S2H). This suggests that samples in epithelial phenotype are less mobile and more structured. We also found that the transition from pEMT to a fully mesenchymal phenotype was accompanied by an increase in the epithelial duct and a decrease in CTCs (Figure S2I). In CTC samples, we found that locPDAC accounted for a slightly higher proportion than metPDAC in the full mesenchymal phenotype (Figure 2I). In addition, we observed that locPDAC epithelial duct samples contributed to 25% of both the epithelial and mesenchymal phenotypes, whereas metPDAC epithelial duct samples predominantly exhibited the mesenchymal phenotype (Figure 2J). The result suggests a distinction in the behavior between

Figure 2. EMT phenotypes of mRNA across different PDAC patient groups

(A) Principal-component analysis between CTCs and normal PBMC samples. (B) Differentially expressed EMT gene counts in CTCs and duct PDAC compared with normal samples. (C) Heatmap showing the log₂ fold change values for differentially expressed epithelial (top) and mesenchymal genes (bottom) shared in CTCs and duct PDAC. (D) Cnet plot representing the top 7 Gene Ontology enriched terms of differentially expressed EMT genes. Enriched terms with high similarity were clustered and rendered as a network. Each gene node is colored according to its EMT type. The node size of a term represents the number of enriched genes and is colored by differentially expressed epithelial genes (EPI) and mesenchymal genes (MES) in CTCs and ductal PDAC, respectively. (E) Analysis of the correlation between epithelial score (E score) and mesenchymal score (M score). Points are colored by different groups. (F) Distributions of EM score within the duct and CTC. The distribution of the EM score from the CTC is colored in blue, while the distribution of the EM score from the duct is colored in green. The black dashed lines represent the intersection of the CTC and duct EM score distribution curves. The red dots indicate the corresponding EM score at the intersection point. (G) Quantification of different patient groups in epithelial, partial, and mesenchymal phenotypes. (H) Boxplot showing EM scores in different CTC and duct groups. (I and J) Distribution of EMT status in CTC and duct groups. (K) Overview diagram showing the dynamic changes of the EMT stage from the primary site, through the circulation to the metastatic site. The color and shape of each point are determined by the group type and display the EM score of the relevant sample.



(legend on next page)

localized and metastatic forms of PDAC and implies some level of dynamic balance between these cell types during the process of EMT (Figure 2K). The EMT score measures the potential for metastasis and could also be connected to disease development. To determine whether this was a common occurrence in pancreatic cancer CTCs, we validated this hypothesis on three external datasets including CTC and duct tumor samples. The results are consistent with our analysis that CTC samples predominantly display pEMT and mesenchymal phenotype, duct normal samples are largely epithelial phenotype, and duct tumor samples vary across different EMT stages (Figure S2J). We analyzed the overall survival of TCGA patients, and a higher EM score was associated with poorer overall survival (Figure S2K). In addition, we also compared the correlation between EM score and the expression of shared EMT genes with opposite expression patterns, and we found that the expression of Mesenchymal gene *VCAM1* and epithelial gene *TACSTD2* were significantly correlated with EM score (Figure S2L). Evidence has shown that CTCs can cluster with neutrophils, enhancing proliferation and promoting metastasis. Neutrophils interact with CTCs through binding to highly expressed adhesion molecules on tumor cells, such as *VCAM1*.¹⁹ *TACSTD2* overexpression is related to increased growth and metastasis of pancreatic carcinoma.²⁰ These results suggest that the acquisition of mesenchymal features by epithelial cells is closely associated with the migration of CTCs, leading to cancer progression.

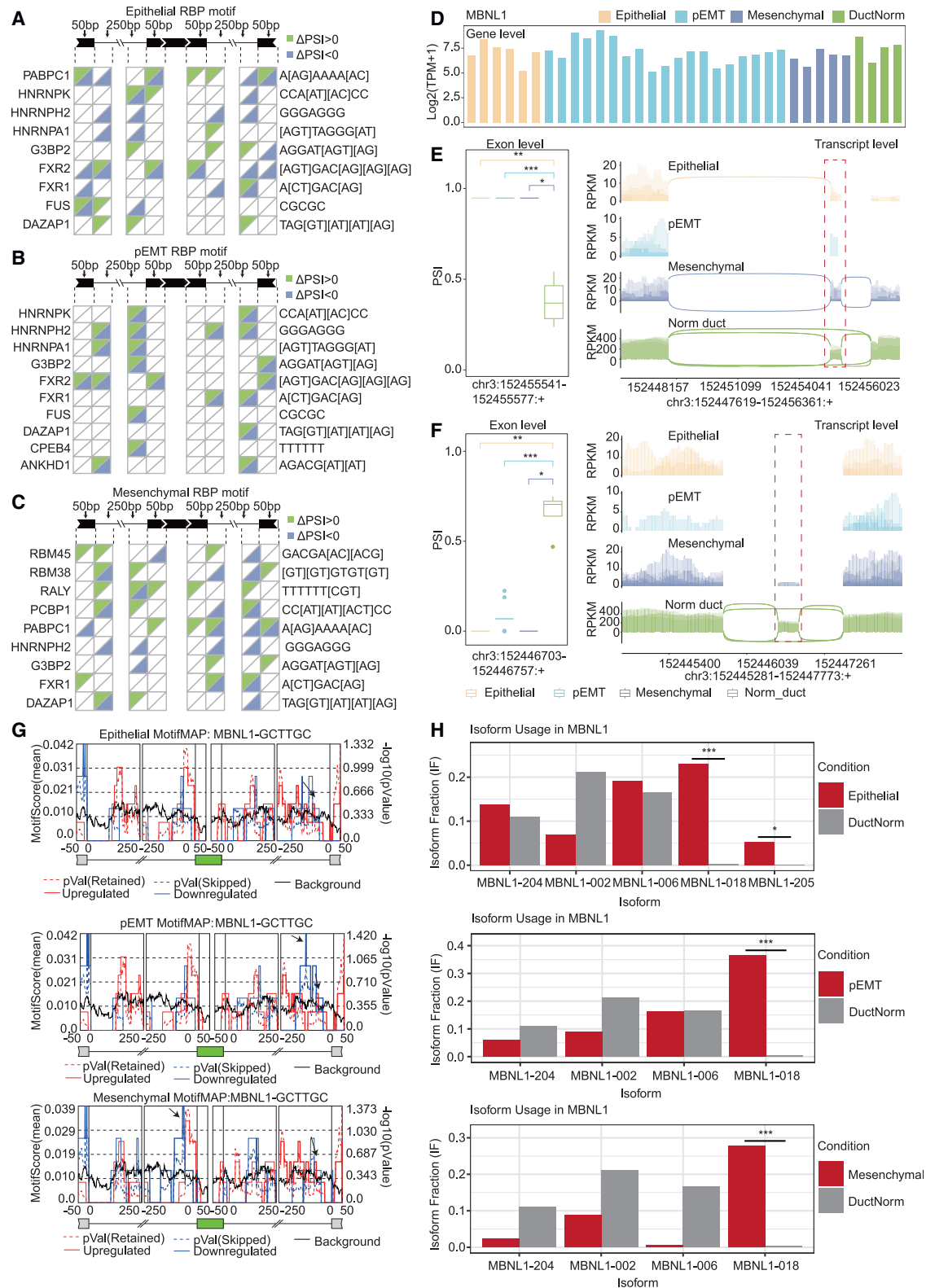
A range of AS events occur in CTCs during EMT

AS is an important method of regulating gene expression that expands the cell proteome. After demonstrating the significance of the EMT program in the tumor cycle, our next step was to identify the splicing regulators responsible for it at the exon level. Instead of using blood samples as a control, normal epithelial cells were extracted from bulk RNA-seq to investigate the intrinsic differences between CTCs and epithelial cells (materials and methods). We first analyzed the gene coverage of both the CTC RNA-seq and bulk RNA-seq data. This comparison confirms that the gene coverage across these two datasets is sufficiently similar to permit valid comparative analysis (Figure S3A). To quantify the contribution of AS to dynamic transcript accumulation during the emergence of EMT, we calculated the use of alternative exons using the rMATS statistical model. This model estimates the abundance of each exon and assigns it a “percentage splicing in” (PSI) value (Ψ) to calculate the use of alternative exons. We identified five types of AS events, including exon skipping, mutually exclusive exons, alternative 5'

splice sites, alternative 3' splice sites, and intron retention. On the basis of the location of each AS event, we obtained the intersection of events in CTC samples and normal epithelial ducts (Figure S3B). We then identified differential AS events in epithelial, pEMT, and mesenchymal CTCs, respectively, compared with normal epithelial duct samples (adjusted $p < 0.05$ and delta PSI > 0.1 ; Table S4). The relative proportions of differential AS event types remained stable when comparing epithelial, pEMT, and mesenchymal types (Figures 3A, S3C, and S3D). Notably, some of the differential AS events identified in our results have been reported as indeed existing. For example, the SE splicing event associated with the RBP gene *MBNL1*, located at chr3:152445281-152445539+, was experimentally verified by Chu et al.²¹ Similarly, the SE splicing event related to cell surface protein *FLNA* at chrX:154352552-154352675- in exon 40 was also validated by Oda et al.²² GO analysis of these events revealed their association with signaling pathways related to protein metabolism (e.g., organelle biogenesis and maintenance, protein localization) and tumor growth factor (e.g., apoptotic signaling pathway, establishment of organelle localization), as shown in Figure 3B. On the basis of our calculation of the number of events related to those genes, we discovered that on average, 20% of genes generated multiple isoforms, indicating transcriptional and post-transcriptional diversity within CTCs in different EMT states (Figure 3C). We then conducted a detailed analysis to explore the composition of those alternative spliced genes. To explore the relationship between differential AS events and DEGs, we analyzed the DEGs in epithelial, pEMT, and mesenchymal phenotype compared with normal epithelial duct samples. We found that a few of the genes were both differentially spliced and expressed, whereas the majority of genes were alternatively spliced without expression changes (Figure S3E). In subsequent analyses, we only retained events whose related gene showed no significant changes in expression between CTCs and normal epithelial ducts (Figure 3D). To gain insights into the regulation of AS and functional relationships among genes, we explored the chromosomal distribution of these events. Our analysis revealed that relatively long chromosomes, such as chr1, chr2, and chr3, had a larger number of splicing events compared with other chromosomes in A5SS and MXE, suggesting that these chromosomal regions display increased transcriptional and post-transcriptional activity (Figures 3E and S3F). We went on to compare the delta PSI of differential AS events among epithelial cells, pEMT, and stromal cells. We found those AS events were different across the EMT stage. These differences were predominantly driven by strong relative enrichment of epithelial

Figure 3. Differential alternative splicing events during EMT in CTCs

(A) The distribution of AS types in EMT-specific AS events, comparing epithelia, pEMT, and mesenchymal CTCs. (B) Enrichment network representing the top 8 significantly enriched terms of genes related to differential AS events ($p < 0.05$). Enriched terms with high similarity were clustered and rendered as a network, while each node represents an enriched term and is colored according to its cluster. Node size indicates the number of enriched genes. (C) Bar graph showing the proportion of genes associated with a single event or multiple events. (D) The count of differential AS events without differences at gene level among epithelial, pEMT, and mesenchymal CTC. (E) Distribution of delta PSI of shared SE events across chromosomes. (F) The ternary phase diagram reveals variations in epithelial, pEMT, and stromal AS events, analyzed using delta PSI. Each point, colored per AS type, signifies genes: small dots for those associated with a unique splicing event, and larger ones for genes linked to multiple events. (G) The count of differential AS events related to cell surface proteins. (H) The frequencies of exon use in differential AS events related to cell surface proteins. (I) Cell surface protein genes associated with multiple differential AS events in epithelial, pEMT, mesenchymal, and normal duct samples. (J) The pie chart shows the topology present in ORF transcripts affected by differential AS events related to cell surface protein genes. (K and L) Examples showing topology and domains in ORF transcripts, such as in LMBRD1-229 (K) and MCAM-201 (L).



(legend on next page)

and mesenchymal, compared with their low abundances in pEMT. The relative abundances of RI events increased progressively from mesenchymal to epithelial (Figure 3F). In addition, we found that multiple differential AS events were significantly associated with survival outcomes (Figure S3G). Together, the differential AS events clearly show the necessity of discriminating AS-specific roles for a full understanding of the transcriptional landscape of EMT.

Cell surface proteins that are specifically altered splicing in tumor cells are ideal targets to minimize off-target effects and are important markers for the detection and isolation of CTCs. Our analysis of splicing data revealed a total of 89 aberrant splicing events associated with 77 cell surface proteins among the differential AS events (Figure 3G). To identify retained exons in these cell surface protein-related AS events, we used a threshold of $\Delta\text{PSI} > 0.1$, while skipped exons were defined as those with $\Delta\text{PSI} < -0.1$. The result showed that there exhibited a higher number of retained exons, which could potentially impact isoform length (Figure 3H). Among them, we found 12 cell surface protein genes associated with multiple AS events, which might increase the number of inferred neoantigens in metPDAC (Figure 3I). On the basis of the location of these cell surface protein gene-related events, we found that they affected 124 open reading frame (ORF) regions. These regions are associated with 22 different transcripts, which could potentially affect the reading frame and may serve as suitable therapeutic targets (Table S5). To further screen for PDAC-specific targets in cancer-specific AS events, we then analyzed the topology of those 22 related transcripts. We identified 25 distinct topologies, 6 of which were associated with the extracellular domain (Figure 3J). Furthermore, transcripts contain structures with multiple topologies that may influence cellular function, possibly leading to changes in phenotype or even disease conditions (Figures 3K and 3L). However, we believe that this is a cautious estimate as junction-spanning polypeptides derived from AS are inadequately represented because of trypsin's cleavage properties. Consequently, all of the RNA-level candidates we have identified might be considered appropriate for additional validation and can be developed further as targets for immune therapy. Together, these data indicate a specific and sustained requirement for diverse transcripts of cell surface protein during all stages of EMT.

AS of RBP orchestrates the EMT-specific AS events

After highlighting the significance of particular AS events in EMT, our subsequent objective was to pinpoint the splicing regulators accountable for them. Our focus was on skipped exon events, which are one of the

most common types of AS. We used the online web server rMAPS2 (RNA Map Analysis and Plotting Server 2) to identify discriminative *cis*-regulatory sequence motifs enriched in the cassette exons that are preferentially included in epithelial, pEMTs, and mesenchymal phenotype, as well as motifs in their flanking introns (materials and methods).²³ These *cis*-regulatory motifs corresponded to predicted RBPs. To investigate the association between AS and RBPs, we analyzed the correlations between the gene expression levels of RBPs and the PSI values of exon-skipped events. We sorted the significantly correlated RBPs according to the absolute value of correlation and took the top 10 RBPs for downstream analysis (correlation R is greater than 0.6, and the p value is less than 0.05; Figures S4A–S4C). We observed that the well-conserved binding motifs of these RBPs were significantly over-represented in the transcript sequences of flanking the skipped/retained exons compared with non-differentially spliced exons (Figures 4A–4C). Most RBPs were common among epithelial, pEMT, and mesenchymal, such as *DAZAP1*, *FXR1*, *G3BP2*, and *HNRNPH2* (Figure S4D). It is worth noting that when the same RBPs bind to the same RNA target sequence but possess distinct biochemical properties, it may result in varying splicing regulatory outcomes.²⁴ As an example, in the case of pEMT-specific skipping events ($\Delta\text{PSI} > 0.1$), we identified the *HNRNPH2* motif occurring separately in the intron sequence located 250 bp upstream or downstream of the target exon sequence (Figure 4B). In contrast, for epithelial and mesenchymal skipping events ($\Delta\text{PSI} > 0.1$), this motif was absent in the intron sequence found 250 bp downstream of the target exon sequence (Figures 4A and 4C). The results suggest that splicing factors have a common regulatory effect on AS patterns and that RBPs have potential position-dependent effects on exon splicing in both the process of pEMT and mesenchymal phenotype.

Notably, the relative expression level of *MBNL1* was not significantly different between pEMT PDAC, mesenchymal PDAC, and normal epithelial cells (Figure 4D). However, *MBNL1* generated distinct transcripts through its unique AS mechanism (Figures 4E and 4F). Moreover, *MBNL1* influences the recruitment of other splicing factors and the spliceosome, leading to its direct regulation of AS. Therefore, we have selected *MBNL1* for further investigation to understand its role in epithelial, pEMT, and mesenchymal emergence. To investigate the effect of *MBNL1* on AS changes between CTC and normal epithelial cells, we examined the positional distribution of *MBNL1*-binding motifs. Our analysis revealed a significant overrepresentation of the *MBNL1* binding motif in the flanking retained exons of CTCs compared with non-differentially skipped exons. Additionally, we

Figure 4. RBP regulated exon inclusion in CTCs during EMT

(A–C) The differential enrichment of RBPs between exon skipped ($\Delta\text{PSI} < 0$) and exon inclusion ($\Delta\text{PSI} > 0$). The motif sequences of the corresponding RBP are labeled on the right. Cell colors represent the smallest p value less than 0.05 for each enriched region, while blanks represent non-significant enrichment. (D) Bar graphs show expression levels of *MBNL1* across epithelial, pEMT, mesenchymal, and normal duct samples. (E and F) The boxplot on the left represents the PSI of *MBNL1*. The genomic location of the alternative exon was labeled at the bottom. Sashimi plots on the right show the alternative spliced exons according to genomic locations. Junction reads are plotted as arcs whose width is determined by the number of reads aligned to the junction spanning the exons connected by the arc. (G) Positional distribution of *MBNL1*-binding motifs of epithelial (top), pEMT (middle), mesenchymal (bottom) CTC in SE events. Motif enrichment scores (solid line) and p values (dashed line) were plotted according to AS event positions. Arrows indicate peaks of enrichment for exons. (H) Bar plot showing isoform use of *MBNL1*. Data were analyzed using the Mann-Whitney two-sided test. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, and **** $p \leq 0.0001$.

observed significant differences in the positioning of *MBNL1* binding sites within CTC regions undergoing AS changes (Figure 4G). In addition, we further analyzed the differential transcript use of *MBNL1*. Although there were no evident alterations in the expression levels of *MBNL1*, we noticed considerable variation in the use of the *MBNL1*-018 transcript isoform across epithelial, pEMT, and mesenchymal CTCs. On the other hand, significant changes in the *MBNL1*-205 transcript were only seen in epithelial CTCs (Figure 4H). Collectively, these findings suggest that as EMT evolves step by step, it is accompanied by stage-specific variations in transcript diversity. These findings suggest that *MBNL1* can differentially regulate the transition among epithelial, pEMT, and mesenchymal phenotypes through AS, independent of gene expression, potentially affecting CTC remodeling and differentiation into pEMT and mesenchymal subtypes.

Stop codon readthrough affects the diversity of therapeutic targets in CTCs

To provide a comprehensive understanding of the complexity of a given AS event, we analyzed changes in AS event modality relative to the stepwise progression from normal epithelial cells to pEMT and finally to mesenchymal. Our analysis began by assigning each AS event to one of three modalities: (1) included, wherein exons are entirely preserved ($\Psi > 0.95$); (2) excluded, wherein exons are fully skipped ($\Psi < 0.05$); and (3) middle, wherein an exon may be included or skipped in RNA molecules transcribed from the gene ($0.05 < \Psi < 0.95$). Our results reveal that nearly 90% of AS events exhibited modality switches from the norm to epithelial phenotype, while 28.2% of AS events showed modality changes from epithelial to pEMT, and 35.6% of them changed from pEMT to mesenchymal phenotype (Figures 5A–5C). We observed that among the modality change AS events, the majority were attributed to the middle mode. The middle mode represents an exon that skips over a percentage of transcripts while including the remaining portions. As various combinations of exons in a gene can generate multiple protein isoforms with unique functions or properties, the incorporation of these exons may play a critical role in determining the commitment of cells toward pEMT fate (Figure 5D). We found that *CYB5R3*-related splicing events have different patterns in different EMT stages (Figure 5E). *CYB5R3* is an important component in maintaining glucose use, where it interacts with glucose to stabilize glucokinase.²⁵ We found that the transcript of *CYB5R3* has an extracellular topology domain, which may play a crucial role in processes like signal transduction. At the same time, compared with normal duct cells, we found that *CYB5R3*-214 showed significant isoform use differences during EMT (Figure 5E). Our findings reveal a clear pattern that identifying the CTC-specific isoform may lead to the discovery of specific therapeutic targets.

Exon skipping and retention not only alter transcript length but also introduce premature termination codons (PTCs) within gene coding sequences.²⁶ This can activate the evolutionarily conserved nonsense-mediated decay machinery, which guards against the negative consequences of mRNAs harboring PTCs.²⁷ To comprehensively compare the impact of stop codon identity on these splicing events, we con-

ducted a systematic analysis by tallying the occurrences of each possible PTC, including TAA, TAG, and TGA (Figure 5F). Importantly, our analysis revealed a high frequency of single events harboring multiple PTCs, with longer exons exhibiting an increased number of PTCs (Figures 5G and 5H). Given the known influence of C's and G's in promoting efficient translation termination compared with A's and U's, we investigated the nucleotide use within 40 nt upstream and 60 nt downstream of the stop codon in protein-coding transcripts (Figure 5I). Our investigation revealed that several patterns in this region demonstrate a repeated three-nucleotide periodicity. Notably, we observed a significant enrichment of C and G nucleotides near the PTC. However, as the distance from the PTC increases, the frequency of C and G nucleotides decreases. Taken together, it appears that evolution has intricately refined PTC to enhance translation termination efficiency. Targeting crucial stop codons for readthrough is a promising approach that may be achieved without disrupting general translation termination. This trait could potentially provide therapeutic benefits when dealing with crucial genes identified to have PTCs.

DISCUSSION

In this study, we provide quantitation of the epithelial to EMT continuum of CTCs in PDAC, covering localized and metastatic sites. Our findings reveal that CTCs, especially those obtained from metastatic samples, can destabilize the epithelium by preferentially expressing mesenchymal genes. This could lead to broad perturbations in the metastatic process through multiple pathways. Additionally, a genome-wide exon-based analysis of CTCs showed considerable variation between different EMT stages, allowing for investigation into the identification of cell-type-specific isoforms, AS events, and an unknown regulatory mechanism of EMT formation. We identified 77 splicing events in cell surface proteins that were specifically expressed in epithelial, pEMT, and mesenchymal states, respectively, which may represent potential targets for therapeutic intervention. These results suggest that abnormal AS is prevalent in EMT progression, with a complex mode switch from epithelial to pEMT and then to mesenchymal. Furthermore, we discovered that *MBNL1* significantly regulates frequent differential AS switches in the epithelial, pEMT, and mesenchymal stages through AS without appreciable changes at the gene expression level. Finally, upon examination of the epithelial, pEMT, and mesenchymal termination codons, we observed a high frequency of single events harboring multiple PTCs. This finding could potentially lead to transcript diversity and contribute to the identification of useful markers for diagnosis and prognosis.

A previous study discovered that the EMT program is highly plastic and demonstrated that epithelial stabilization leads to enhanced collective migration of epithelial cancer cells, as evidenced by the use of single-cell RNA-seq and genetic mouse models.²⁸ Epithelial cells undergo a process known as EMT, during which they become more mobile and invasive by losing their normal cell-cell adhesion and polarity. This allows them to break off from the main tumor and enter the bloodstream as CTCs.^{29,30} We explored epithelial stability during tumor metastasis to identify disease-associated cell-type-specific EMT

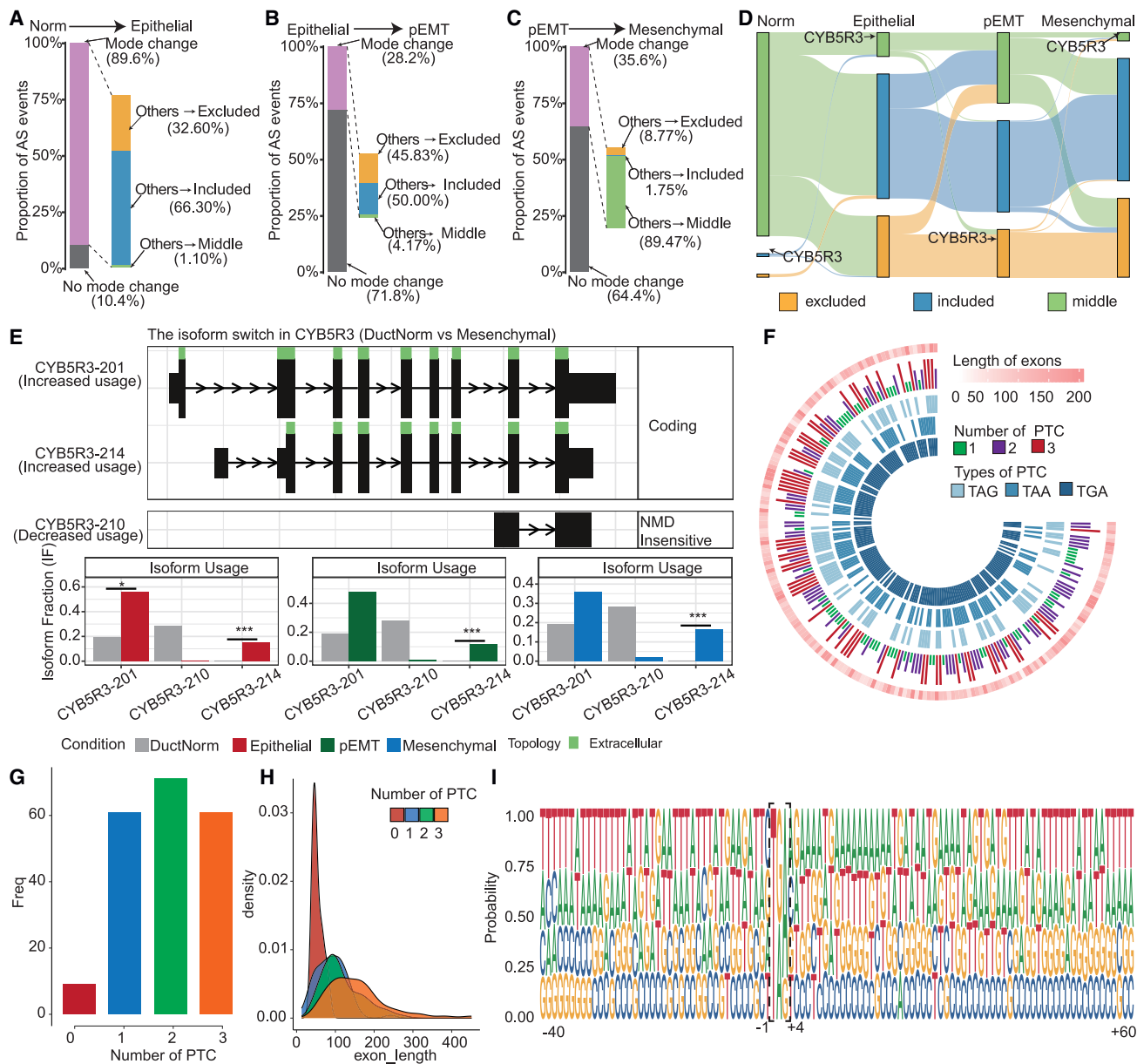


Figure 5. Genes have alternative splice variants containing different stop codons

(A–C) Bar plot showing the proportion of AS events with (modality change) or without (no modality change) modality changes between neighboring developmental stages during EMT. (D) Sankey diagram showing the AS modality transition from norm to epithelial, then to pEMT, and finally to mesenchymal. (E) Isoforms related to *CYB5R3*. Top: two analyzed isoforms for the *CYB5R3* gene: PFAM domains are indicated by black. Extracellular topology colored in green. UCSC gene transcript IDs are shown for each isoform. Bottom: the average isoform fraction in epithelial (left), pEMT (middle), and mesenchymal (right), compared with the norm duct. Data were analyzed using the Mann-Whitney two-sided test. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, and **** $p \leq 0.0001$. (F) Circos plot of PTC (TAA, TAG, and TGA) signals in overall shared SE AS events. (G) Bar plot showing the frequency of SE AS events with different types of PTCs. (H) Distribution of different types of PTCs in shared SE AS events. (I) The frequencies of each nucleotide are plotted for all positions 40 nt upstream to 60 nt downstream of the stop codon. Nucleotides are plotted in order of increasing frequencies.

gene signatures. The analysis of the molecular features that distinguish CTCs is an outstanding tool to provide insights into the understanding of crucial aspects of the metastatic process and could foster use in liquid biopsy-based personalized cancer treatment. On the basis of the EMT signature, our results showed that the EMT program of

CTCs is highly heterogeneous, with many cells present in pEMT and mesenchymal states, whereas epithelial cancer cells acquire a more stable epithelial phenotype away from pEMT and mesenchymal states. Our results are consistent with previously reported that the EMT program of cancer cells exhibits molecular alterations, including

decreased expression of epithelial markers (E-cadherin, ZO-1, claudins, and occludins) and increased expression of mesenchymal markers (vimentin, N-cadherin, fibroblast-specific protein1, and fibronectin).^{31,32} Notably, we observed that both normal epithelial cells and healthy donor samples from blood fluctuated in a small range of EM scores, whereas epithelial cancer cells and CTCs demonstrated distinct polarization. These results suggest that epithelial cells can undergo EMT with broad effects. In addition, pEMT-phenotype CTCs were more similar to health donors, suggesting that this pEMT intermediate cell state may therefore be more stable than the other transition states and provided evidence of a metastatic advantage for pEMT CTCs, which was supported by previous studies.^{33,34} Overall, our analysis of EMT signals of epithelial cells and CTCs establishes the importance of epithelial cells losing their intercellular adhesion and acquiring mesenchymal in supporting metastatic dissemination and highlights that pEMT-phenotype CTCs and mesenchymal CTCs had different metastases capacity.

AS is a common process in which transcribed mRNAs generate distinct transcript isoforms, leading to protein products that possess unique functional properties.³⁵ However, the current understanding of how to monitor disease progression from CTCs to metastasis is primarily derived from the observations at the gene expression level.³⁶ Ignoring transcriptional isoforms in transcriptome analyses can lead to incomplete information about the complex process of EMT. Furthermore, it is common for previous studies of CTCs from PDAC patients to use purified healthy donor volunteer blood as a control, while the cells in purified blood from a healthy donor include red blood cells, WBCs (such as lymphocytes and neutrophils), and platelets.^{4,36–38} Nonetheless, one main concern is that CTCs are rare cells that have detached from the localized tumor and entered the bloodstream, and they may provide a closer comparison with the epithelial cells of the localized tumor compared with other circulating cells in the bloodstream, including their genetic and molecular profile, as well as their morphology and behavior. Therefore, using healthy donor blood as a control might potentially mask important differences in EMT changes between CTCs and healthy individuals. Here, we describe AS changes in epithelial-type CTCs, pEMT-type CTCs, and mesenchymal-type CTCs using normal epithelial samples as controls to gain a deeper understanding of the unique characteristics of each CTC subtype.

We identified multiple splicing events in cell surface proteins that are suitable targets for future research. It has been reported that abnormal splicing events provided the corresponding putative AS-derived neo-epitopes of 16 cancer types.³⁶ Lin et al.³⁶ illustrated that many highly tumor-specific AS events are putative targets for TCR T cell approaches in the GBM. *MBNL1* is well known to regulate RNA AS, localization, and decay and drives tumor dedifferentiation as well, implicating their activity to some extent in the relapse and distant metastasis.^{39–41} We found that gene *MBNL1*, with the absence of appreciable changes in the gene expression, produced significantly different AS events in CTC. Notably, we found a large subset of epithelial-, pEMT-, and mesenchymal-specific AS events were pre-

dicted to be under the control of *MBNL1*. Genome-wide transcriptomic analysis of the *MBNL1* gene has revealed a potential close relationship between *MBNL1* and breast cancer metastatic colonization.⁴² Together, our findings provide further evidence that specific AS isoforms might have key regulatory roles independently of expression alterations and are a necessary first step for understanding the isoform-specific functions of key proteins.

The premature stop codons may lead to mRNA destruction through nonsense-mediated mRNA decay or the production of a truncated protein.⁴² Using the modality strategy, we explored the distribution and variation of AS events across different cell populations to reveal exon inclusion bias during EMT. We discovered a high frequency of single events harboring multiple PTCs that might support models of translation initiation inhibition. This highlights the significance of the stop codon context in regulating translation termination. However, achieving specificity for PTC readthrough poses a significant challenge for nonsense suppression therapies, requiring a deeper understanding of translation termination and stop codon readthrough in varied sequence contexts.^{43,44} Despite this challenge, such readthrough events can be beneficial in therapy when a critical gene contains a PTC.

In summary, this study suggests that the EMT program at the exon level within the CTC is heterogeneous and contributes to the functional annotation of the splicing factor by indicating the necessity for studying alternatively spliced isoforms during transcriptional analysis. These findings have allowed us to explore previously unidentified transcriptional processes involved in tumor metastasis and offer insights into potential therapeutic interventions.

MATERIALS AND METHODS

Data acquisition

We downloaded 3 PDAC single-cell RNA-seq (scRNA-seq) datasets (GSE156405, GSE81547, and GSE155698), 3 CTC RNA-seq datasets (GSE144561, GSE40174, and GSE114704), and 2 bulk RNA-seq datasets (GSE57973 and GSE193268) from the Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>). The corresponding RNA-seq data were obtained from the TCGA data portal (<https://tcga-data.nci.nih.gov/tcga/>). Datasets, including GEO: GSE156405, GSE81547, GSE155698, GSE144561, GSE193268, and GSE57973, were used for analysis. Datasets, including GEO: GSE40174 and GSE114704 and TCGA samples were used for validation purposes (Table S1).

Single-cell data quality pre-processing and initial cell type discovery

Quality control and downstream analysis were performed using Seurat version 3.2.3.⁴⁵ Each scRNA-seq dataset contained a single-cell transcriptome count matrix. Samples were analyzed after excluding low-quality cells (<1,000 genes/cell and >20% mitochondrial genes). The harmony algorithm was used to integrate single cells from different platforms and tissues, followed by the construction of a K-nearest neighbors (KNN) graph using the FindNeighbors function

in Seurat. Finally, cell clustering and UMAP visualization were performed using the FindClusters and RunUMAP functions, respectively. The annotations of cell identity on each cluster were defined by the expression of known marker genes.

Inference of CNV from scRNA-seq data

To distinguish cancer cells from normal ones, chromosomal gene expression was perturbed to infer CNV aberrations. For the accuracy of cancer cell inference, we have compared the results of two methods: inferCNV and CopyKat.^{46,47} First, we separately used CopyKat and inferCNV to compute the signal for copy number profiles from the gene expression matrix. To compare the results of these two methods, we calculated the CNV scores of all cells using the method from previous studies.⁴⁸ Because of the algorithmic differences between the two methods, we scaled the CNV score. We then arranged the cells in ascending order and defined the high and low of the CNV score with 25% as the cutoff value. We found that in the results of CopyKat, there is a higher proportion of high scores in the epithelial cells, while in immune cells, there is a higher proportion of low scores. On the other hand, the results from InferCNV present similar proportions for both types of cells. Therefore, in subsequent analyses, we adopted the predictions of CopyKat. Second, we excluded cells that were not predicted and the diploid cells that were predicted to be normal epithelial or immune cells. Third, we calculated the mean square and the correlation of CNV per cell with the average of the top 5% of cells from the aneuploid single cells. This allowed us to detect cancer cells and enhance the precision of our findings. Finally, cells specifically with a correlation or CNV score greater than 0.2, are categorized as cancer cells. Considering that P5_GES156405 contained very few malignant epithelial cells, this sample was removed from subsequent analyses.

Bulk RNA-seq quality pre-processing

The CTC datasets read were aligned to the GRCh38 human reference genome from the University of California, Santa Cruz (<http://genome.ucsc.edu>), using the STAR version 2.7.4a aligner with default settings (Alex Dobin; <https://github.com/alexdobin/STAR>).⁴⁹ We used ESTIMATE to evaluate tumor purity.⁵⁰ The formula for calculating ESTIMATE-based tumor purity was developed as follows: Tumor_purity = $\cos(0.6049872018 + 0.0001467884 \times \text{ESTIMATEScore})$. We retained 30 CTC samples with a tumor purity greater than 50%, and 15 normal samples with a purity less than 50%. To further check the tumor purity of the sample, we analyzed the immune cell composition in the selected CTC samples, as inferred by the CIBERSORT algorithm.⁵¹

Differential expression analysis

To perform sample-level differentially expressed EMT genes in single-cell PDAC samples and CTC samples in parallel, we used a pseudo-bulk approach for single-cell data. We use the functions from the SingleCellExperiment package to extract the counts and metadata of single-cell data.⁵² To calibrate the pseudo-bulk data, we added two bulk RNA-seq datasets as references, namely, GSE193268, which comprises 3 liver metastasis samples, and GSE57973, which includes 3 normal duct samples. We applied ComBat-seq to eliminate any batch effects and used DESeq2 to perform differential expression analysis in

pseudo-batch samples and CTC samples separately.^{53,54} An average $\log_2$ fold change (FC) value greater than 1 was defined as upregulation, while a value less than -1 was defined as down-regulation. Only genes with adjusted p values less than 0.05 and $|\log_2 \text{FC}| \geq 1$ were considered significantly differentially expressed.

Generation of EMT scores

All scRNA-seq samples are analyzed using the average expression of each gene. To create the EMT expression matrix, we selected specific genes on the basis of the average expression of the 315 Thiery gene signature. The different expressed EMT genes in CTC and epithelium were used to define the cancer cell EMT stage. Epithelial and mesenchymal scores were obtained from singscore method for the epithelial duct samples and CTC samples respectively.⁵⁵ An “EM score” was calculated as follows:

$$\text{EM score} = \tanh \left(\frac{\text{EPI Score}}{2 * (\text{Mes Score})} \right)$$

where EPI is the epithelial score and MES is the mesenchymal score. These EM scores were then used to organize all samples across the EMT spectrum from epithelial to mesenchymal with the cutoff at the intersection of the Epi duct and CTC density curve (-0.05 and 0.76), where epithelial has an EM score >0.76 , pEMT has an EM score ≤ 0.76 and ≥ -0.05 , and mesenchymal has an EM score <-0.05 .

AS analysis

PSI (Ψ) of alternatively spliced events were quantified with filtered and sorted BAM files using rMATS (version 4.1.2) with default parameters. We analyzed AS of epithelial CTCs, pEMT CTCs, and mesenchymal CTCs compared with normal epithelial duct samples. The normal epithelial duct samples were obtained from the bulk RNA-seq dataset (GSE57973). First, we calculated AS events in epithelial CTCs, pEMT CTCs, mesenchymal CTCs, and normal epithelial duct samples, respectively, and removed events with an NA percentage greater than 85%. Then, we identified the intersection of these events in epithelial CTCs, pEMT CTCs, mesenchymal CTCs, and normal duct samples, on the basis of the position of each AS event. We included only the intersected events, filtering out the others. Finally, missing values were imputed using the mean value, and a Wilcoxon test was used to compute differences in AS events across epithelial CTCs, pEMT CTCs, and mesenchymal CTCs, in comparison with normal duct samples. The p values were corrected using the Benjamini & Hochberg method. Differential splicing events were identified at a threshold of $\text{abs}(\Delta\text{PSI}) > 0.1$ and adjusted p values less than 0.05. We used IsoformSwitchAnalyzeR to analyze changes in isoform use and the dominant ORF of the isoforms.⁵⁶

Motif enrichment analysis

To analyze motif enrichment of alternative spliced exons, we used rMAPS2 with default parameters and considered all non-significant SE events as background. The results of motif enrichment analysis for epithelial-, pEMT-, and mesenchymal-CTC ES events compared

with background ES events. We identified significantly enriched RBPs on the basis of the following criterion: the minimum p value in at least one of the four regions needed to be less than 0.05.

Statistical analysis

The statistical analyses were conducted using R (version 4.1.3). Survival package for survival analysis. Survival analysis was performed using the Survival package. The clusterProfiler package was used to conduct GO and pathway enrichment analysis.⁵⁷ For visualization, ggplot2 (version 3.3.6) and ggseqlogo (version 0.1) packages were used. Spearman's rank correlation analysis was implemented for non-normally distributed data, while Pearson correlation was used for continuous variables that met normal distribution criteria.

DATA AND CODE AVAILABILITY

The codes for data processing can be found at <https://github.com/JiangBioLab/PancreaticCancerAS>.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.omtn.2024.102129>.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (grants T2325009, 62032007, and 32270789), the Major Program of the National Natural Science Foundation of China (grant T2293752), the National Key R&D Program of China (grant 2022ZD0117700), and the Special Science and Technology Innovation Project of Xiong'an New Area in China (grant 2022XAGG0117).

AUTHOR CONTRIBUTIONS

Q.J., H.N., and P.W. conceived and designed the study. L.X., H.N., P.W., H.S., and R.C. performed the research. All authors received and verified the data. L.X. and R.C. performed the data analysis and interpretation. Q.J., L.X., and R.C. drafted the manuscript. H.X., R.Y., X.J., and Z.X. contributed to the critical revision of the manuscript. Y.C., Y.Y., F.P., and G.X. participated in paper writing and editing. All authors critically reviewed and approved the final version of the manuscript for submission.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Connor, A.A., and Gallinger, S. (2022). Pancreatic cancer evolution and heterogeneity: integrating omics and clinical data. *Nat. Rev. Cancer* 22, 131–142.
- Tian, C., Huang, Y., Clauser, K.R., Rickelt, S., Lau, A.N., Carr, S.A., Vander Heiden, M.G., and Hynes, R.O. (2021). Suppression of pancreatic ductal adenocarcinoma growth and metastasis by fibrillar collagens produced selectively by tumor cells. *Nat. Commun.* 12, 2328.
- Mehlen, P., and Puisieux, A. (2006). Metastasis: a question of life or death. *Nat. Rev. Cancer* 6, 449–458.
- Lin, D., Shen, L., Luo, M., Zhang, K., Li, J., Yang, Q., Zhu, F., Zhou, D., Zheng, S., Chen, Y., and Zhou, J. (2021). Circulating tumor cells: biology and clinical significance. *Signal Transduct. Targeted Ther.* 6, 404.
- Nikanjam, M., Kato, S., and Kurzrock, R. (2022). Liquid biopsy: current technology and clinical applications. *J. Hematol. Oncol.* 15, 131.
- Qi, L.N., Xiang, B.D., Wu, F.X., Ye, J.Z., Zhong, J.H., Wang, Y.Y., Chen, Y.Y., Chen, Z.S., Ma, L., Chen, J., et al. (2018). Circulating Tumor Cells Undergoing EMT Provide a Metric for Diagnosis and Prognosis of Patients with Hepatocellular Carcinoma. *Cancer Res.* 78, 4731–4744.
- Brabletz, S., Schuhwerk, H., Brabletz, T., and Stemmler, M.P. (2021). Dynamic EMT: a multi-tool for tumor progression. *EMBO J.* 40, e108647.
- Lüönd, F., Sugiyama, N., Bill, R., Bornes, L., Hager, C., Tang, F., Santacrose, N., Beisel, C., Ivanek, R., Bürglin, T., et al. (2021). Distinct contributions of partial and full EMT to breast cancer malignancy. *Dev. Cell* 56, 3203–3221.e11.
- Zhang, N., Ng, A.S., Cai, S., Li, Q., Yang, L., and Kerr, D. (2021). Novel therapeutic strategies: targeting epithelial-mesenchymal transition in colorectal cancer. *Lancet Oncol.* 22, e358–e368.
- Wright, C.J., Smith, C.W.J., and Jiggins, C.D. (2022). Alternative splicing as a source of phenotypic diversity. *Nat. Rev. Genet.* 23, 697–710.
- Biamonti, G., Bonomi, S., Gallo, S., and Ghigna, C. (2012). Making alternative splicing decisions during epithelial-to-mesenchymal transition (EMT). *Cell. Mol. Life Sci.* 69, 2515–2526.
- Brown, R.L., Reinke, L.M., Damerow, M.S., Perez, D., Chodosh, L.A., Yang, J., and Cheng, C. (2011). CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *J. Clin. Invest.* 121, 1064–1074.
- Hu, X., Harvey, S.E., Zheng, R., Lyu, J., Grzeskowiak, C.L., Powell, E., Piwnicka-Worms, H., Scott, K.L., and Cheng, C. (2020). The RNA-binding protein AKAP8 suppresses tumor metastasis by antagonizing EMT-associated alternative splicing. *Nat. Commun.* 11, 486.
- Wang, Z.W., Pan, J.J., Hu, J.F., Zhang, J.Q., Huang, L., Huang, Y., Liao, C.Y., Yang, C., Chen, Z.W., Wang, Y.D., et al. (2022). SRSF3-mediated regulation of N6-methyladenosine modification-related lncRNA ANRIL splicing promotes resistance of pancreatic cancer to gemcitabine. *Cell Rep.* 39, 110813.
- Bradley, R.K., and Anczuków, O. (2023). RNA splicing dysregulation and the hallmarks of cancer. *Nat. Rev. Cancer* 23, 135–155.
- Kim, N., Kim, H.K., Lee, K., Hong, Y., Cho, J.H., Choi, J.W., Lee, J.I., Suh, Y.L., Ku, B.M., Eum, H.H., et al. (2020). Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* 11, 2285.
- Bankó, P., Lee, S.Y., Nagygyörgy, V., Zrínyi, M., Chae, C.H., Cho, D.H., and Telekes, A. (2019). Technologies for circulating tumor cell separation from whole blood. *J. Hematol. Oncol.* 12, 48.
- Tan, T.Z., Miow, Q.H., Miki, Y., Noda, T., Mori, S., Huang, R.Y.J., and Thiery, J.P. (2014). Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol. Med.* 6, 1279–1293.
- Park, J., Wysocki, R.W., Amoozgar, Z., Maiorino, L., Fein, M.R., Jorns, J., Schott, A.F., Kinugasa-Katayama, Y., Lee, Y., Won, N.H., et al. (2016). Cancer cells induce metastasis-supporting neutrophil extracellular DNA traps. *Sci. Transl. Med.* 8, 361ra138.
- Chen, W., Li, M., Younis, M.H., Barnhart, T.E., Jiang, D., Sun, T., Lang, J.M., Engle, J.W., Zhou, M., and Cai, W. (2022). ImmunoPET of trophoblast cell-surface antigen 2 (Trop-2) expression in pancreatic cancer. *Eur. J. Nucl. Med. Mol. Imag.* 49, 861–870.
- Chu, Y., Hu, J., Liang, H., Kanchwala, M., Xing, C., Beebe, W., Bowman, C.B., Gong, X., Corey, D.R., and Mootha, V.V. (2020). Analyzing pre-symptomatic tissue to gain insights into the molecular and mechanistic origins of late-onset degenerative trinucleotide repeat disease. *Nucleic Acids Res.* 48, 6740–6758.
- Oda, H., Sato, T., Kunishima, S., Nakagawa, K., Izawa, K., Hiejima, E., Kawai, T., Yasumi, T., Doi, H., Katamura, K., et al. (2016). Exon skipping causes atypical phenotypes associated with a loss-of-function mutation in FLNA by restoring its protein function. *Eur. J. Hum. Genet.* 24, 408–414.
- Hwang, J.Y., Jung, S., Kook, T.L., Rouchka, E.C., Bok, J., and Park, J.W. (2020). rMAPS2: an update of the RNA map analysis and plotting server for alternative splicing regulation. *Nucleic Acids Res.* 48, W300–W306.

24. Fu, X.D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* *15*, 689–701.
25. Watanabe, H., Du, W., Son, J., Sui, L., Asahara, S.I., Kurland, I.J., Kuo, T., Kitamoto, T., Miyachi, Y., De Cabo, R., et al. (2023). Cyb5r3-based mechanism and reversal of secondary failure to sulfonylurea in diabetes. *Sci. Transl. Med.* *15*, eabq4126.
26. Jung, H., Lee, K.S., and Choi, J.K. (2021). Comprehensive characterisation of intronic mis-splicing mutations in human cancers. *Oncogene* *40*, 1347–1361.
27. Lindeboom, R.G.H., Supek, F., and Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* *48*, 1112–1118.
28. Carstens, J.L., Yang, S., Correa De Sampaio, P., Zheng, X., Barua, S., Mcandrews, K.M., Rao, A., Burks, J.K., Rhim, A.D., and Kalluri, R. (2021). Stabilized epithelial phenotype of cancer cells in primary tumors leads to increased colonization of liver metastasis in pancreatic cancer. *Cell Rep.* *35*, 108990.
29. Castro-Giner, F., and Aceto, N. (2020). Tracking cancer progression: from circulating tumor cells to metastasis. *Genome Med.* *12*, 31.
30. Tao, J., Zhu, L., Yakoub, M., Reißfelder, C., Loges, S., and Schölch, S. (2022). Cell-Cell Interactions Drive Metastasis of Circulating Tumor Microemboli. *Cancer Res.* *82*, 2661–2671.
31. Mittal, V. (2018). Epithelial Mesenchymal Transition in Tumor Metastasis. *Annu. Rev. Pathol.* *13*, 395–412.
32. Díaz-López, A., Díaz-Martín, J., Moreno-Bueno, G., Cuevas, E.P., Santos, V., Olmeda, D., Portillo, F., Palacios, J., and Cano, A. (2015). Zeb1 and Snail1 engage miR-200f transcriptional and epigenetic regulation during EMT. *Int. J. Cancer* *136*, E62–E73.
33. Pal, A., Barrett, T.F., Paolini, R., Parikh, A., and Puram, S.V. (2021). Partial EMT in head and neck cancer biology: a spectrum instead of a switch. *Oncogene* *40*, 5049–5065.
34. Ruscetti, M., Quach, B., Dadashian, E.L., Mulholland, D.J., and Wu, H. (2015). Tracking and Functional Characterization of Epithelial-Mesenchymal Transition and Mesenchymal Tumor Cells during Prostate Cancer Metastasis. *Cancer Res.* *75*, 2749–2759.
35. Wang, F., Tan, P., Zhang, P., Ren, Y., Zhou, J., Li, Y., Hou, S., Li, S., Zhang, L., Ma, Y., et al. (2022). Single-cell architecture and functional requirement of alternative splicing during hematopoietic stem cell formation. *Sci. Adv.* *8*, eabg5369.
36. Lin, E., Cao, T., Nagrath, S., and King, M.R. (2018). Circulating Tumor Cells: Diagnostic and Therapeutic Applications. *Annu. Rev. Biomed. Eng.* *20*, 329–352.
37. Holcomb, J.B., Tilley, B.C., Baraniuk, S., Fox, E.E., Wade, C.E., Podbielski, J.M., Del Junco, D.J., Brasel, K.J., Bulger, E.M., Callcut, R.A., et al. (2015). Transfusion of plasma, platelets, and red blood cells in a 1:1:1 vs a 1:1:2 ratio and mortality in patients with severe trauma: the PROPPR randomized clinical trial. *JAMA* *313*, 471–482.
38. Yu, M., Bardia, A., Wittner, B.S., Stott, S.L., Smas, M.E., Ting, D.T., Isakoff, S.J., Ciciliano, J.C., Wells, M.N., Shah, A.M., et al. (2013). Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* *339*, 580–584.
39. Wang, E.T., Cody, N.A.L., Jog, S., Biancolella, M., Wang, T.T., Treacy, D.J., Luo, S., Schroth, G.P., Housman, D.E., Reddy, S., et al. (2012). Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* *150*, 710–724.
40. Ray, D., Yun, Y.C., Idris, M., Cheng, S., Boot, A., Iain, T.B.H., Rozen, S.G., Tan, P., and Epstein, D.M. (2020). A tumor-associated splice-isoform of MAP2K7 drives dedifferentiation in MBNL1-low cancers via JNK activation. *Proc. Natl. Acad. Sci. USA* *117*, 16391–16400.
41. Batra, R., Charizanis, K., Manchanda, M., Mohan, A., Li, M., Finn, D.J., Goodwin, M., Zhang, C., Sobczak, K., Thornton, C.A., and Swanson, M.S. (2014). Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Mol. Cell* *56*, 311–322.
42. Fish, L., Pencheva, N., Goodarzi, H., Tran, H., Yoshida, M., and Tavazoie, S.F. (2016). Muscleblind-like 1 suppresses breast cancer metastatic colonization and stabilizes metastasis suppressor transcripts. *Genes Dev.* *30*, 386–398.
43. Wangen, J.R., and Green, R. (2020). Stop codon context influences genome-wide stimulation of termination codon readthrough by aminoglycosides. *Elife* *9*, e52611.
44. Anzalone, A.V., Zairis, S., Lin, A.J., Rabadan, R., and Cornish, V.W. (2019). Interrogation of Eukaryotic Stop Codon Readthrough Signals by in Vitro RNA Selection. *Biochemistry* *58*, 1167–1178.
45. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* *184*, 3573–3587.e29.
46. Timothy Tickle, I.T., Georgescu, C., Brown, M., and Haas, B. (2019). inferCNV of the Trinity CTAT Project.
47. Gao, R., Bai, S., Henderson, Y.C., Lin, Y., Schalck, A., Yan, Y., Kumar, T., Hu, M., Sei, E., Davis, A., et al. (2021). Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.* *39*, 599–608.
48. Peng, J., Sun, B.F., Chen, C.Y., Zhou, J.Y., Chen, Y.S., Chen, H., Liu, L., Huang, D., Jiang, J., Cui, G.S., et al. (2019). Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* *29*, 725–738.
49. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15–21.
50. Yoshihara, K., Shahmoradgol, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* *4*, 2612.
51. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* *12*, 453–457.
52. Amezcua, R.A., Lun, A.T.L., Becht, E., Carey, V.J., Carpp, L.N., Geistlinger, L., Marini, F., Rue-Albrecht, K., Risso, D., Pagès, H., et al. (2020). Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* *17*, 137–145.
53. Zhang, Y., Parmigiani, G., and Johnson, W.E. (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom. Bioinform.* *2*, lqaa078.
54. 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.
55. Foroutan, M., Bhuva, D.D., Lyu, R., Horan, K., Cursors, J., and Davis, M.J. (2018). Single sample scoring of molecular phenotypes. *BMC Bioinf.* *19*, 404.
56. Vitting-Seerup, K., and Sandelin, A. (2017). The Landscape of Isoform Switches in Human Cancers. *Mol. Cancer Res.* *15*, 1206–1220.
57. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS A J. Integr. Biol.* *16*, 284–287.