

Specific alternative splicing and polyadenylation facilitate metastasis mediated by CTC clusters

Wen Zhang (✉ zhangwen@cicams.ac.cn)

National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital

<https://orcid.org/0000-0003-3150-2751>

Quanyou Wu

National Cancer Center <https://orcid.org/0000-0002-3180-8606>

Guoliang Li

Zhenrong Yang

Defeng Kong

Zhaoru Gu

Duo Wan

Qi Zhang

Shujun Cheng

Kaitai Zhang

Cancer Institute (Hospital), Peking Union Medical College & Chinese Academy of Medical Sciences

<https://orcid.org/0000-0001-9560-9641>

Article

Keywords: CTC clusters, alternative splicing, alternative polyadenylation, mRNA therapies, antisense oligonucleotides

Posted Date: December 7th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1128296/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Circulating tumor cell (CTC) clusters possess a much higher capability to seed metastasis than single CTCs. However, the mechanism underlying this phenomenon is still elusive and no reports have investigated the role of posttranscriptional RNA regulation in CTC clusters. Here, we compared alternative splicing (AS) and alternative polyadenylation (APA) profiles between single CTCs and CTC clusters. 994 and 836 AS events were identified in single CTCs and CTC clusters, separately. About ~20% of AS events exhibited alterations between both cell types. The differential splicing of SRSF6 was a core event that caused AS profiles' disturbance and made CTC clusters more dangerous. Concerning APA, we identified global 3' UTRs lengthening in CTC clusters compared with single CTCs. This change was mainly regulated by 14 core APA factors, especially PPP1CA. The altered APA profiles boosted the cell cycle of CTC clusters and reflected that CTC clusters endured less oxidative stress. Our study investigated the posttranscriptional regulation mechanisms in CTC clusters, found that the perturbation of AS and APA contributed to the superiority of CTC clusters compared with single CTCs, and laid the foundation for developing antisense oligonucleotides that inhibit metastasis by reducing CTC clusters.

Key Points

- Establishing a pipeline for revealing AS and APA profiles based on single-cell RNA sequencing libraries.
- About ~20% of AS events exhibit alterations between CTC clusters and single CTCs.
- The CTC clusters-specific splicing of SRSF6 causes AS profiles' disturbance and make CTC clusters more dangerous.
- Identification of global 3' UTRs lengthening in CTC clusters compared with single CTCs for the first time, which boosts the cell cycle of CTC clusters.

Introduction

Metastasis, the spread of cells from the primary tumor to distant organs followed by their unlimited proliferation, is one of the most devastating aspects of malignancy[1]. It is widely known that metastasis is mainly mediated by circulating tumor cells (CTC), which transit through the bloodstream and seed in distant locations for subsequent progress[2]. There are at least two forms of CTC in the blood of cancer patients: single CTCs and CTC clusters[3, 4]. Compared with single CTCs, CTC clusters possess a much higher capability to seed metastasis. It is reported that CTC clusters have 23~50-fold increased metastatic potential and the presence of them is associated with a significantly poor prognosis, affirming the bright prospect of hampering CTC clusters to reduce cancer dissemination[5]. However, only a few studies tried to dissect the molecular biology of CTC clusters and no reports have aimed to reveal the posttranscriptional regulation mechanism in CTC clusters, especially on the aspect of alternative splicing (AS) and alternative polyadenylation (APA).

Alternative splicing is a mechanism in which almost all introns of pre-mRNAs are discarded and exons are selectively retained, thereby contributing to protein complexity in eukaryotes[6, 7]. Alternative polyadenylation is a process that produces isoforms with distinct 3'termini from the same gene, resulting in altered mRNA function, localization, stability, and translation efficiency[8]. The perturbation of AS and APA can lead to various types of diseases, including cancer[9]. It has been realized for a long time that modulating AS and APA by mRNA-based therapies is a promising approach to treat diseases. Recently, the FDA-approved mRNA vaccines for COVID-19 (BNT162b1, mRNA-1273, Givosiran, etc.) and dozens of mRNA-targeted drugs for refractory diseases[10], ushering in an upsurge of mRNA-based therapies and provided a solid rationale to target mRNA as an anticancer strategy[11]. Compared with traditional protein-based therapies, RNA-based therapies have potential advantages because they can theoretically target any gene (especially the previously undruggable targets), have more direct and simple development procedures, and are more stable at room temperature, a feature conducive to their distribution and storage[12]. Several RNA-based therapeutic platforms have been developed, including antisense oligonucleotides (ASOs), microRNA, small interfering RNA, and aptamers[13]. Among them, the application of ASOs to convert AS and APA is currently one of the most well-established and promising cancer treatment strategies[14]. To illustrate, as of January 1, 2021, a total of 229 clinical trials have studied 60 oligonucleotide drugs for the treatment of cancer, of which 195 trials have applied ASOs as interventions, only 7 trials for microRNA, 17 trials for siRNAs and 9 trials for aptamers. Besides, there are 15 phase 2/3 or phase 3 clinical trials investigating oligonucleotide therapeutics as cancer treatment, and they are all focused on ASOs[15]. Because the most common mechanism of ASOs is to modulate AS and APA[14], the sparse understanding of AS and APA profiles and dynamics in CTC clusters inhibits the development of such promising therapies that are expected to interfere with CTC clusters and significantly reduce cancer dissemination and metastasis.

In this study, we mainly focused on revealing the profile and dynamic perturbation of AS and APA in CTC clusters compared with single CTCs, investigating why these key posttranscriptional regulation mechanisms changed and whether these alterations contributed to the more malignant phenotype of CTC clusters. Through solving these questions, our study provided a genome-wide posttranscriptional regulation profile of single CTCs and CTC clusters, unraveled molecular characteristics that promote metastasis, and enabled the identification of ASOs targets for hampering CTC clusters.

Materials And Methods

Compilation of Single-cell RNA sequencing (scRNA-seq) data

To investigate and compare the AS and APA profiles in single CTCs and CTC clusters, we collected four scRNA-seq datasets, including SRP131647[16], SRP133387[17], SRP066632[18], SRP186111[19]. These datasets were produced according to the Smart-seq2 protocol[20], which is appropriate for AS quantification. Dataset SRP131647 and SRP133387 contain single CTCs and CTC clusters from matched breast cancer patients and xenografts. Dataset SRP066632 contains single CTCs derived from breast cancer patients, while SRP186111 includes CTC clusters from breast cancer patients and xenografts. We

downloaded raw sequencing data in FASTQ format from the NCBI SRA database. Dataset SRP131647 was used to draw main conclusions, while the other three datasets were obtained for validating results. In this study, we defined dataset SRP131647 as discovery set, SRP133387 as validation set 1, SRP066632 as validation set 2, and SRP186111 as validation set 3.

Raw data processing and gene/transcript expression quantification

FastQC (version 0.11.8) was applied to perform Initial quality checks on raw sequencing data, and Trim-Galore (version 0.5.0) was employed to remove adaptor sequences and low-quality reads. To remove potential contaminations in the xenograft models, we used STAR (version 2.7.0f)[21] to map the reads from datasets SRP131647, SRP133387, and SRP186111 to a combined human (hg38) and mouse (mm10) genome reference. For dataset SRP066632, we directly mapped trimmed reads to human genome (hg38) reference. Cells having uniquely aligned reads less than 1,000,000 or showing more than 5% of reads derived from mice were filtered out. Salmon (version 1.3.1)[22] was used to calculate the expression of genes and transcripts in each cell. To make the quality control more stringent, we further removed cells that met any of the following criteria: (1) the number of genes detected < 2000; (2) the percentage of mitochondrial genes > 20%; (3) the expression level of EPCAM < 10 and PTPRC > 10 TPM. Finally, 92 single CTCs and 82 CTC clusters from ten breast cancer patients and three xenografts were retained in dataset SRP131647. The other three datasets for validation contained a total of 202 single CTCs and CTC clusters, in which dataset SRP133387 preserved 66 single CTCs and 64 CTC clusters from matched donors, SRP066632 contained 41 single CTCs, and SRP186111 included 31 CTC clusters.

AS events quantification

AS events were detected and quantified by Outrigger (version 1.1.1), a module of Expedition[23], with default settings. Percent Spliced In (PSI) was used to measure AS events, which represents the exon-inclusion ratio. By default, Outrigger only calculates the PSI values of AS events with more than ten junction reads; otherwise, it outputs missing values. AS events identified in at least five cells of each cell type were kept for further analysis. IGV (version 2.5.3)[24] was used to draw Sashimi plots when visualizing exon coverages and splice junctions.

APA events quantification

To quantify APA events, we use the coverage data (wig files generated by STAR) as input for the DaPars2 algorithm (version 2.0)[25, 26], which directly infers the APA usage by comparing standard RNA-seq data from multiple samples. The Percentage Distal Usage Index (PDUI) score was calculated by DaPars2, which quantifies the fraction of distal poly(A) site usage for the corresponding gene. PDUI score near 0 means that the gene tends to use proximal PAS, while genes favoring distal PAS usage were assigned PDUI scores near 1. Because DaPars2 would produce many missing values in the PDUI matrix due to the relatively low sequencing depth of scRNA-seq compared to standard bulk RNA-seq, we further imputed the missing value by scDaPars (version 0.0.1) [27], a tool designed explicitly for scRNA-seq data to tackle the sparsity by sharing APA information across similar cells.

APA events in primary breast tumor and breast cancer cell lines

APA profiles of primary breast tumor and breast cancer cell lines were obtained from Synapse (<https://www.synapse.org/>, syn7888354). As previously described [28], RNA-seq BAM files of 837 primary breast tumor samples, 105 paired nontumor samples, and 47 breast cancer cell lines were obtained from the National Cancer Institute's Genomic Data Commons (GDC) (<https://gdc.cancer.gov>). DaPars was further utilized to identify APA events and calculate PDUI values for each sample.

Modality assignment

The PSI/PDUI distribution of an AS/APA event in a group of cells can be summarized by the modality. Anchor (version 1.1.1), the other module of Expedition, was applied to assign modalities to AS/APA events. Since PSI/PDUI values are continuous between 0 and 1, Anchor fits the values to a Beta distribution. The probability density function $Pr(\alpha, \beta)$ of the Beta distribution is defined between (0, 1), with parameters $\alpha > 0$ and $\beta > 0$. Anchor assigns modalities to AS/APA events based on the goodness of fit of four models (excluded ($1 = \alpha < \beta$), included ($\alpha > \beta \geq 1$), bimodal ($\alpha = \beta < 1$), and middle ($\alpha = \beta > 1$)). Multimodal modality was used as the null model where parameters α and β equal 1. During the assignment process, Anchor first measures how well the AS/APA event fits the excluded and included models. If the fit is poor, test the middle and bimodal models. If still poorly fitted, then Anchor will assign multimodal modality to this event.

Investigating regulators of differential AS/APA events between single CTCs and CTC clusters

From SpliceAid2[29], we collected 68 reliable splicing factors (SFs), most of which belong to hnRNPs or SR protein families. Twenty-two core APA factors were obtained from one previous report[30]. Spearman correlation test was applied pair-wise to identify the interaction between SFs (core APA factors) and differential AS (APA) events, the regulatory role of differentially spliced genes participating in RNA splicing (mRNA 3'-end processing) on differential AS (APA) events. The pairs were significantly associated if the absolute value of Spearman correlation $|Rs| > 0.3$ and the adjusted P-value < 0.05 . Cytoscape (version 3.7.1)[31] was used to draw dysregulation networks.

Exploration of the functional meaning of differential AS/APA events between single CTCs and CTC clusters

Protein-coding genes affected by differential AS/APA events were obtained to perform GO enrichment analysis through the R package "clusterProfiler"[32]. To evaluate the effects of alternative exons on isoforms' coding ability, we retrieved protein-coding information from the annotation file GENCODE v27. To identify whether alternative exons inclusion would affect functional domains, we searched domains of each isoform through Pfam[33].

highly conserved miRNA sites in acquired and lost 3' UTRs

We retrieved highly conserved miRNA binding sites and their corresponded miRNA families from TargetScanHuman 7.2 [34]. This information, along with the genomic position of altered 3'UTRs predicted by DaPars2 was used to calculate the number of miRNA sites in acquired and lost 3' UTRs.

Effects of APA Events on Drug Sensitivity

We obtained the drug sensitivity data from the Genomics of Drug Sensitivity in Cancer (GDSC) data portal [35] to evaluate the effects of APA events on drug sensitivity. Spearman correlation analysis was conducted to measure the association between the IC50 from GDSC and the PDUI of APA events. The absolute value of correlation coefficients > 0.3 and $FDR < 0.05$ indicates a significant correlation between drug sensitivity and PDUI, which is equivalent or stricter than commonly used in drug sensitivity studies[36].

Cell cycle phase classification

CellCycleScoring function implemented in Seurat was utilized to identify the cell cycle status of each cell[37]. This function relies on 97 cell cycle marker genes, including 43 for S phase and 44 for G2/M phase[38]. CellCycleScoring assigns each cell a score according to its expression of S and G2/M phase markers. The cells were assigned to G1 if both S-score and G2/M-score were lower than average (i.e., negative). Otherwise, the cells were assigned to S or G2/M, depending on which score is larger.

Statistical Analysis

All statistical tests were two-sided and performed in R software. We applied the Benjamini-Hochberg procedure to adjust P-values and considered P-values less than 0.05 as statistical significance in analyses.

Results

The AS profiles in single CTCs and CTC clusters

To study the pattern of AS events in single CTCs and CTC clusters, we utilized a set of scRNA-seq data derived from matched breast cancer patients and xenografts with accession number SRP131647 in the NCBI SRA database. After stringent quality control (see methods), 92 single CTCs and 82 CTC clusters in this dataset were kept for downstream analysis. We also gathered three other datasets (validation sets) to confirm the results obtained from dataset SRP131647 (discovery set). The number of uniquely mapped reads of single CTCs and CTC clusters in all datasets were similar (Fig. 1A, Fig. S1A-C) and no obvious 3'-end bias was found (Fig. 1B, Fig. S1D-F), suggesting that it is reasonable to investigate AS events based on these datasets. In datasets SRP131647 and SRP133387, the AS profiles of cells derived from patients and xenografts were similar (Fig. 1C-D), justifying the strategy that comparing all single CTCs with all CTC clusters regardless of their origin. There was no significant difference in the number of AS events identified in single CTCs and CTC clusters, and these numbers were of the same order of magnitude across different datasets (Fig. 1E, Fig. S1G-I). In total, 994 high-confidence AS events were

identified in single CTCs, including 975 skipped exon (SE) events and 19 mutually exclusive exon (MXE) events (Fig. S1J). These AS events affected 778 genes, of which 769 were coding genes, 1 was antisense RNA, and 8 were processed transcripts annotated by GENCODE v27. Parallely, we identified 836 high-confidence AS events in CTC clusters, containing 820 SE and 16 MXE. These AS events regulated 670 genes, of which 662 were coding genes, 1 was antisense RNA, and 7 were processed transcripts (Fig. 1F). The low proportion of MXE in identified AS events and high fraction of protein-coding genes regulated by AS were concordant with other studies dissecting AS profiles of malignancies based on bulk RNA-seq data[39, 40]. Among all alternative exons in single CTCs and CTC clusters, about half had intact codons (the length of intact codons is a multiple of three) (Fig. 1G). The distribution of median PSI value of AS events identified in single CTCs and CTC clusters was similar and the heterogeneity across each cell was low in terms of PSI's distribution (Fig. 1H-J). Next, we classified AS events into five modalities according to PSI's distribution. An AS event was "excluded" if all PSIs of a group of cells were near 0, "included" if all PSIs were near 1, "middle" if all PSIs were near 0.5, and "bimodal" if some PSIs were close to 1 and others near 0. Otherwise, it was classified as "multimodal" (Fig. 1K). We found that in both single CTCs and CTC clusters, the number of included AS events was the most, and the number of excluded events was the second. In single CTCs, bimodal and multimodal AS events accounted for 26.7%, while these two modalities accounted for 16.5% in CTC clusters. No middle AS events were detected in both cell types. Although the whole modality pattern of AS events in single CTCs and CTC clusters were similar, the proportion of bimodal events in CTC clusters was much lower than that in single CTCs (Fig. 1L; Table S1). To validate the reliability of these modalities assigned to AS events, we gathered dataset SRP133387 as validation set 1, which contained 66 single CTCs and 64 CTC clusters from matched breast cancer patients and xenografts after strict quality control. Among shared AS events detected in single CTCs from both datasets, 89.3% (582/652) were assigned to the same modalities, while in CTC clusters, 95.3% (562/590) of modalities were the same (Fig. 1M-N). To further confirm the robustness of these modalities, two other datasets were collected. One was SRP066632 (validation set 2), which preserved 41 single CTCs from breast cancer patients. This dataset confirmed 65.1% (216/332) of modalities in single CTCs (Fig. S1K). The other was SRP186111 (validation set 3), which contained 31 CTC clusters from breast cancer patients and xenografts. This dataset confirmed 80.2% (303/378) of modalities assigned to AS events in CTC clusters (Fig. S1L). Altogether, these results revealed the AS profiles of single CTCs and CTC clusters and confirmed that the modalities assigned to AS events were reliable.

The regulation and functional consequences of disturbed AS profiles in CTC clusters

We next asked how many AS events changed modalities and found that only ~20% (129/786) of AS events shared between single CTCs and CTC clusters exhibited alterations (Fig. 2A). This finding agreed with previous studies reporting that CTC clusters and single CTCs exhibited similar molecular profiles[5, 17]. All these differential events were implicated in bimodal or multimodal modality, and no events changed from excluded (included) modality to included (excluded) modality. As cells transition from single CTCs to CTC clusters, the change from bimodal events to excluded or included events was the most prominent (Fig. 2B). It has been accepted that splicing factors (SFs) are the primary regulator of AS events[41]. Thus, we wondered whether the expression change of SFs would be the major mechanism

resulting in these differential AS events between single CTCs and CTC clusters. From SpliceAid2[29], We collected 68 SFs (Table S2) and paired them with differential AS events to construct a dysregulation network. To our surprise, among all SFs investigated, only 11 regulated differential AS events of eight genes (Fig. 2C), suggesting that the expression change of SFs could not explain the differential AS events between single CTCs and CTC clusters. We next performed Gene Ontology (GO) enrichment analysis on differential AS events and found that the most significant GO term enriched was “RNA splicing” (Fig. 2D). This result indicated that differential AS events-related genes that participate in “RNA splicing” may be the key mechanism regulating widespread differential AS events in CTC clusters. To test this hypothesis, we paired genes enriched in “RNA splicing” with differential AS events to construct another dysregulation network. Sixteen genes were correlated with 47 differential AS events in this network, and SRSF6 was the core factor that regulated the greatest number of events (Fig. 2E). To further confirm that the differential splicing of SRSF6 would affect more widespread changes of AS profiles in CTC clusters, we conducted a similar analysis on the other three datasets. As a result, SRSF6 was also a core factor in the dysregulation network (Fig. S2). These results suggested that differentially spliced genes that participate in “RNA splicing” were more important than the expression of SFs in mediating differential AS events. Among these 16 differentially spliced genes, the AS events of at least 12 would lead to the change of protein-coding abilities or the functional domains (Fig. S3), highlighting the significance of maintaining their dominant isoforms and reasoning why the differential splicing of these genes would lead to more widespread changes of AS profiles in CTC clusters. For instance, the alternative splicing of SRSF6 led to two isoforms: SRSF6-201 and SRSF6-202. SRSF6-201 is a protein-coding isoform while SRSF6-202 suffers nonsense-mediated decay, a process eliminating mRNA transcripts that contain premature stop codons. When the alternative exon was spliced out, the isoforms expressed by SRSF6 changed from SRSF6-202 to SRSF6-201 and thus gained the function of regulating AS (Fig. 2F). The modality of SRSF6 changed from bimodal to excluded when single CTCs transitioned to CTC clusters (Fig. 2G), and this change was confirmed in dataset SRP133387 (Table S1). Thus, SRSF6 played its correct splice-regulation role in all CTC clusters, while in some single CTCs, SRSF6 suffered nonsense-mediated decay and loss function. Because it is widely known that SRSF6 promotes proliferation and metastasis of many types of malignancy, including breast cancer[42-45], our results indicated that the gain of function of SRSF6 through alternative splicing in all CTC clusters resulted in the perturbation of AS profiles and contributed to the more malignant phenotype of CTC clusters compared with single CTCs.

CTC clusters showed global 3' UTRs lengthening compared with single CTCs

We noticed that differential AS events-related genes were also significantly enriched in “RNA export from nucleus” and “mRNA 3'-end processing” (Fig. 2D). This led us to ask whether the transition of APA profiles was noticeable between single CTCs and CTC clusters. We identified a total of 1686 APA events in single CTCs and CTC clusters. Compared with primary breast tumor and breast cancer cell lines, the CTCs preferred short 3' UTRs (Fig. 3A). Similar to the distribution of PSI, the distribution of PDUI value in each CTC is also homogeneous (Fig. 3B-C), as well as in primary breast tumors, tumor-adjacent tissues, and breast cancer cell lines (Fig. S4A-C). Interestingly, the APA events identified in breast cancer cell lines and primary breast tumors were very alike, while about 60% APA events detected in single CTCs and CTC

clusters were CTC-specific (Fig. 3D). In terms of the APA profiles in CTCs, although some genes were affected by multiple APA events, most genes experienced one APA event (Fig. 3E). In contrast to the modality profile of AS events where included modality accounted for the most, the modality profile of APA events was characteristic of excluded modality, also suggesting that single CTCs and CTC clusters favored short 3' UTRs (Fig. 3F; Table S3). This result was consistent with other studies reporting that shortening 3' UTRs is a hallmark of tumor cells[28, 46, 47]. To confirm the reliability of modalities assigned to APA events, we conducted a similar analysis on dataset SRP133387 (validation set 1) and compared the results derived from both datasets. For single CTCs, 85.4% (1440/1686) of APA events had the same modalities, while for CTC clusters, 86.7% (1462/1686) of modalities were the same (Fig. 3G-H). To further validate the reliability of these modalities, two other datasets were also analyzed. Dataset SRP066632 (validation set 2) confirmed 61.3% (1034/1686) of modalities in single CTCs (Fig. S4D), while dataset SRP186111 confirmed 70.6% (1191/1686) of modalities in CTC clusters (Fig. S4E). Altogether, these results unraveled the APA profiles of single CTCs and CTC clusters and confirmed that the modalities assigned to APA events were reliable. We next compared APA profiles between single CTCs and CTC clusters and found that CTC clusters exhibited global lengthening of 3' UTRs (Fig. 3I-J). Validation set 1 also confirmed this tendency (Fig. S4F). We found two APA events with longer 3' UTR in CTC clusters were associated with drug sensitivity. Longer 3'UTR of C7orf50 was associated with lower IC50 of Palbociclib while longer 3'UTR of CBR3-AS1 was related to higher IC50 of Doxorubicin (Fig. 3K-N), suggesting these APA events may affect the drug sensitivity of CTC clusters. Besides, we found 3' UTR of MRPL40 was longer in CTC clusters and longer 3' UTR of MRPL40 is associated with poor survival of breast cancer patients (Fig. 3O-P), indicating this APA event may play a crucial role in CTC clusters.

PPP1CA mediated the widespread perturbation of APA profiles in CTC clusters

We further asked how many AS events changed modalities and identified 223 differential APA events between single CTCs and CTC clusters. The most notable changes of modalities were excluded to multimodal and multimodal to included, including 46 and 55 differential APA events, respectively (Fig. 4A). This result matched the finding that CTC clusters tended to use long 3' UTRs. Among those differential APA events with longer 3' UTRs in CTC clusters, 52.8% of transcripts (94/178) gained at least one highly conserved miRNA binding site and miR-325-3p binds to the greatest number of miRNA binding sites gained in CTC clusters. While for differential APA events with shorter 3' UTRs in CTC clusters, 58.5% (24/41) lost at least one highly conserved miRNA binding site (Fig. 4B-C). Some researchers suggested that 3' UTR lengthening would mediate cellular senescence[48]. To check whether the 3' UTR lengthening in CTC clusters reflected a more senescent phenotype than single CTCs, we compared the 3' UTR lengthening genes in CTC clusters with senescence-related 3' UTR lengthening genes reported by those researchers. AS a result, among 195 differential APA events-related genes, 157 showed larger median PDUI values in CTC clusters, in which only 5 genes belonged to senescence-related 3' UTR lengthening genes, indicating that the 3' UTR lengthening in CTC clusters was not associated with cellular senescence (Fig. 4D; Table S4). Besides, there was almost no overlap between differential AS events-related genes, differential APA events-related genes, and differential expression genes, suggesting that gene expression levels, AS and APA regulated CTC clusters at almost completely different layers (Fig. 4E). We next

wondered how extensive the differential AS events-related genes enriched in “mRNA 3'-end processing” regulated the differential APA events and found that only two of these genes were correlated with differential APA events of 10 genes, in which SRSF6 regulated APA of 9 genes (Fig. 4F). This result further highlighted the significance of differential splicing of SRSF6 on mediating the perturbation of posttranscriptional regulation approaches in CTC clusters. Other studies have also reported that SRSF6 exerted an impact on PAS usage[49, 50]. However, this dysregulation network cannot explain why 223 APA events had modality changes between single CTCs and CTC clusters. Previous reports have indicated that alterations in core APA factors' expression or activities can influence PAS usage in eukaryotes, such as CSTF2[51] and NUDT21[52]. Thus, we investigated how extensive these core APA factors regulated differential APA events and which APA factor(s) played essential roles in CTC clusters. To address these issues, we evaluated the relationship between 22 core APA factors (Table S5) with differential APA events and revealed that 14 core APA factors regulated events of 81 genes (Fig. 4G). Among these core factors, the regulatory ability of PPP1CA was the strongest, indicating that this factor largely determined the differential APA events between single CTCs and CTC clusters (Fig. 4H). Results from the other three datasets confirmed that PPP1CA regulated the greatest number of differential APA events (Fig. S5A-B). Thus, in contrast to the regulation of differential AS events where the differential AS events-related genes enriched in “RNA splicing” mediated more perturbation of AS profiles than the expression of SFs, the expression of core APA factors was the main mechanism for regulating differential APA events instead of the differential AS events-related genes enriched in “mRNA 3'-end processing”. In addition to being a core APA factor, PPP1CA was also implicated in promoting multiple types of malignancy, including breast cancer[53-55], and it was upregulated in CTC clusters compared to single CTCs (Fig. 4I). Besides, those genes regulated by PPP1CA were significantly enriched in “cellular response to oxidative stress” and “cell cycle G2/M phase transition”, suggesting an important role of PPP1CA in promoting a more malignant phenotype of CTC clusters through changing the APA profiles (Fig. 4H).

Differential APA events boosted the cell cycle of CTC clusters and reflected that CTC clusters endured less oxidative stress

To further investigate the functional consequence of the altered APA profiles, we performed GO enrichment analysis on differential APA events between single CTCs and CTC clusters. Interestingly, these genes were enriched in “antigen processing and presentation of peptide antigen via MHC class I”, “regulation of cell cycle G2/M phase transition”, “cellular response to oxidative stress”, and other critical signaling pathways such as IFN, TNF, NF-kappaB pathways (Fig. 5A). These GO terms referred to the biological processes that ensure the survival of CTCs in the circulating blood, where they suffered high oxidative stress, immune cells surveillance, and anti-tumor chemokines. These results suggested that differential APA events may contribute to the survival and subsequent development of CTC clusters. Oxidative stress is a critical impediment to the survival and metastasis of CTCs[56]. Previous studies have reported that oxidative stress led to the overall 3' UTRs shortening, and APA was a crucial mechanism for cells to stabilize mRNA in response to stress[57]. Thus, the global lengthening of 3' UTRs reflected that CTC clusters suffered less oxidative stress than single CTCs during transportation in the

blood. In addition, the differential APA events enriched in “cellular response to oxidative stress” further suggested that the oxidative stress was different between single CTCs and CTC clusters.

The result that differential APA events were enriched in “regulation of cell cycle G2/M phase transition” led us to ask whether there was a difference in the cell cycle status between single CTCs and CTC clusters. As a result, we found that compared to single CTCs, more CTC clusters were in the G2/M phase (Fig. 5B). This result was further confirmed by the three other datasets (Fig. 5C-D; Table S6). Among the ten genes enriched in “regulation of cell cycle G2/M phase transition”, four showed an evident difference in terms of median delta PDU ($|\text{delta PDU}| > 0.1$) between single CTCs and CTC clusters, in which the 3' UTRs of H2AFY, SSNA1, CDK1, PSMD2 were longer in CTC clusters (Fig. 5E). Besides, in the G2/M phase, the 3' UTRs of H2AFY, SSNA1, CDK1, PSMD2 were significantly longer (Fig. 5F-I) while of other six genes were not (Fig. S5C-I), indicating a regulatory effect of the APA events involved in these four genes on cell cycle status. These results altogether suggested that through differential APA of these G2/M phase regulatory genes, CTC clusters tended to go into the G2/M phase, thus were more proliferative as indicated by previous studies[17].

Discussions

To our knowledge, this study is the first to focus on revealing the regulation and functional consequence of AS and APA profiles in single CTCs and CTC clusters. Through comprehensive analysis of scRNA-seq datasets produced according to the Smart-seq2 protocol, we identified 994 and 836 high-confidence AS events in single CTCs and CTC clusters, respectively. The number of events detected in this study was lower than other studies based on bulk tumor RNA-seq data, where the total number of SE and MXE events was more than 3,000[39, 40]. This inconsistency is partly because of the relatively low sequencing depth of scRNA-seq compared with bulk RNA-seq and the biological difference of single cell from bulk cells. Since this study is based on scRNA-seq data, it is better to categorize the distribution of AS and APA events into modalities to detect the difference between single CTCs and CTC clusters. Methods for detecting difference applied by many other studies on bulk RNA-seq datasets assume that a unimodal distribution within a condition well characterizes AS and APA events. But the AS and APA events are not always unimodal within a group of cells, and thus the efficiency of these methods is insufficient[58]. We found that the number of included and excluded AS events accounted for the most, and no middle modality was identified in single CTCs and CTC clusters. These results strongly agreed with other studies suggesting that either one or the other isoform is dominant in a given cell, with few cells producing both isoforms[59-61].

We next identified little difference in AS profiles between single CTCs and CTC clusters. Previous studies have illustrated that both cell types exhibited similar gene expression profiles and CpG methylation profiles[5, 17], here we extended the knowledge and suggested that the AS profiles were also similar between single CTCs and CTC clusters. The molecular profiles of CTCs clustered based on the patients of origin or the specific xenograft models[17] justified finding differences within matched samples in this study. We investigated the upstream regulation of differential AS events and found that the self-splicing

of SFs (instead of the expression alteration of SFs) was the primary mechanism resulting in the perturbation of AS profiles in CTC clusters. This result explained why AS profiles changed and suggested correct directions to design AS-targeted therapies to interfere with CTC clusters. For example, we in this study concluded that the gain of function of SRSF6 through AS in all CTC clusters contributed to the more malignant phenotype of this group of cells. Thus, we could try to design ASOs switching the splicing of SRSF6 to attack CTC clusters in the future.

We next investigated the APA profiles and found that CTC clusters exhibited global lengthening of 3' UTRs. This result highlighted the apparent difference of APA profiles between single CTCs and CTC clusters. We suggested that the global 3' UTR lengthening reflected less oxidative stress suffered by CTC clusters in the circulating blood. Many other researchers proposed that because of the physical structure of CTC clusters, it is reasonable to suspect that CTC clusters are subject to less oxidative stress exerted by circulating blood[62, 63]. Here, we provided molecular evidence supporting this hypothesis. By investigating the upstream regulatory mechanism of differential APA events, we found that PPP1CA played a crucial role in promoting CTC clusters through changing the APA profiles. This result suggested that targeting PPP1CA may be a promising approach to interfere with CTC clusters.

Finally, we found that CTC clusters were more advanced in cell cycle status than single CTCs, and this was mediated through differential APA of G2/M phase regulatory genes. This result revealed and highlighted the importance of APA on CTC clusters for the first time and proposed a novel molecular mechanism mediating the more proliferative phenotype of CTC clusters as indicated by previous studies[17].

Our study indeed holds some limitations. First, due to the feature of scRNA-seq data and the limitation of algorithms detecting AS events, we only investigated two types of AS events among seven (SE, MXE, A3, A5, AL, AF, RI) in this study. Almost all studies aimed at dissecting AS profiles based on short-read scRNA-seq data were subject to this restriction[64-66]. Second, the number of single CTCs and CTC clusters analyzed was relatively small, and a larger sample size may be more convincing. Third, although we interrogated the functional consequences of differential AS and APA events, most isoforms' specific and clear functions are still largely unknown. Uncovering the precise functions of each isoform will help to more thoroughly assess the impact of AS and APA profile alterations. However, no high-throughput approaches have been developed to distinguish essential isoforms from isoforms that have little effect in metastasis. Thus, in the future, time-consuming wet experiments are required to explore the function of abundant isoforms generated by AS and APA.

In summary, our research provides comprehensive insights into the AS and APA profiles of single CTCs and CTC clusters, reveals specific mechanisms that lead to the perturbation of these posttranscriptional regulation approaches, and highlights the association between phenotypes of CTCs (such as cell cycle status) and posttranscriptional regulation dynamics. This study uncovers the posttranscriptional diversity of single CTCs and CTC clusters for the first time and is expected to direct and stimulate translational research aiming to develop ASOs for attacking CTC clusters.

Conclusions

Our study comprehensively revealed the dynamics, regulation, and functional consequence of AS and APA profiles in CTC clusters for the first time, found that the perturbation of AS and APA contributed to the superiority of CTC clusters, and laid the foundation for developing ASOs that interfere with CTC clusters.

Abbreviations

CTC: circulating tumor cell, AS: alternative splicing, APA: alternative polyadenylation, ASOs: antisense oligonucleotides, scRNA-seq: Single-cell RNA sequencing, PSI: Percent Spliced In, PDUI: Percentage Distal Usage Index, SFs: splicing factors, SE: skipped exon, MXE: mutually exclusive exon, GO: Gene Ontology

Declarations

Ethical Approval and Consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

scRNA-seq data are available under accession no. SRP131647, SRP133387, SRP066632, SRP186111 in the NCBI SRA database.

Competing interests

The authors have declared that no conflict of interest exists.

Funding

This research was supported by National Key Research and Development Project grant nos. 2019YFC1315700 and 2017YFC1308702, CAMS Initiative for Innovative Medicine grant no. 2017-I2M-1-005, and National Natural Science Foundation of China grant no. 81472013. Funders only provided funding and had no role in the study design, data collection, data analysis, interpretation, and writing of the report.

Authors' contributions

WZ, KZ, and SC was the overall principal investigators who conceived the study, obtained financial support, supervised the entire study, and reviewed the manuscript; QW was responsible for the study

design, performed the data analyses, interpreted the results and wrote the initial manuscript; GL, ZY, and DK oversaw statistical analyses, interpreted the results; ZG and DW performed the sequencing analyses; QZ reviewed the manuscript. All authors approved the final report for publication.

Acknowledgements

Not applicable

Authors' information

WZ is the PI of the Department of Immunology in National Cancer Center, China, engaged in the field of mRNA vaccines and therapies. KZ is a professor in the Department of Etiology and Carcinogenesis of National Cancer Center, China, focused on researching circulating tumor cells. SC is an academican of the Chinese Academy of Engineering, who is interested in the molecular characteristics of cancers.

References

- 1 Fidler IJ. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature reviews Cancer* 2003; 3: 453-458.
- 2 Keller L, Pantel K. Unravelling tumour heterogeneity by single-cell profiling of circulating tumour cells. *Nature reviews Cancer* 2019; 19: 553-567.
- 3 Fidler IJ. The relationship of embolic homogeneity, number, size and viability to the incidence of experimental metastasis. *European journal of cancer* 1973; 9: 223-227.
- 4 Liotta LA, Saidel MG, Kleinerman J. The significance of hematogenous tumor cell clumps in the metastatic process. *Cancer research* 1976; 36: 889-894.
- 5 Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, Spencer JA *et al*. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* 2014; 158: 1110-1122.
- 6 Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics* 2008; 40: 1413-1415.

- 7 Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008; 456: 470-476.
- 8 Gruber AJ, Zavolan M. Alternative cleavage and polyadenylation in health and disease. *Nature reviews Genetics* 2019; 20: 599-614.
- 9 Chabot B, Shkreta L. Defective control of pre-messenger RNA splicing in human disease. *The Journal of cell biology* 2016; 212: 13-27.
- 10 Wang F, Zuroske T, Watts JK. RNA therapeutics on the rise. *Nature reviews Drug discovery* 2020; 19: 441-442.
- 11 Khan P, Siddiqui JA, Lakshmanan I, Ganti AK, Salgia R, Jain M *et al.* RNA-based therapies: A cog in the wheel of lung cancer defense. *Molecular cancer* 2021; 20: 54.
- 12 Katzmann JL, Packard CJ, Chapman MJ, Katzmann I, Laufs U. Targeting RNA With Antisense Oligonucleotides and Small Interfering RNA: JACC State-of-the-Art Review. *Journal of the American College of Cardiology* 2020; 76: 563-579.
- 13 Kulkarni JA, Witzigmann D, Thomson SB, Chen S, Leavitt BR, Cullis PR *et al.* The current landscape of nucleic acid therapeutics. *Nature nanotechnology* 2021; 16: 630-643.
- 14 Desterro J, Bak-Gordon P, Carmo-Fonseca M. Targeting mRNA processing as an anticancer strategy. *Nature reviews Drug discovery* 2020; 19: 112-129.
- 15 Xiong H, Veedu RN, Diermeier SD. Recent Advances in Oligonucleotide Therapeutics in Oncology. *International journal of molecular sciences* 2021; 22.

- 16 Szczerba BM, Castro-Giner F, Vetter M, Krol I, Gkoutela S, Landin J *et al.* Neutrophils escort circulating tumour cells to enable cell cycle progression. *Nature* 2019; 566: 553-557.
- 17 Gkoutela S, Castro-Giner F, Szczerba BM, Vetter M, Landin J, Scherrer R *et al.* Circulating Tumor Cell Clustering Shapes DNA Methylation to Enable Metastasis Seeding. *Cell* 2019; 176: 98-112.e114.
- 18 Jordan NV, Bardia A, Wittner BS, Benes C, Ligorio M, Zheng Y *et al.* HER2 expression identifies dynamic functional states within circulating breast cancer cells. *Nature* 2016; 537: 102-106.
- 19 Donato C, Kunz L, Castro-Giner F, Paasinen-Sohns A, Strittmatter K, Szczerba BM *et al.* Hypoxia Triggers the Intravasation of Clustered Circulating Tumor Cells. *Cell reports* 2020; 32: 108105.
- 20 Picelli S, Björklund Å K, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature methods* 2013; 10: 1096-1098.
- 21 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* 2013; 29: 15-21.
- 22 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* 2017; 14: 417-419.
- 23 Song Y, Botvinnik OB, Lovci MT, Kakaradov B, Liu P, Xu JL *et al.* Single-Cell Alternative Splicing Analysis with Expedition Reveals Splicing Dynamics during Neuron Differentiation. *Molecular cell* 2017; 67: 148-161.e145.
- 24 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G *et al.* Integrative genomics viewer. *Nature biotechnology* 2011; 29: 24-26.

- 25 Feng X, Li L, Wagner EJ, Li W. TC3A: The Cancer 3' UTR Atlas. *Nucleic acids research* 2018; 46: D1027-d1030.
- 26 Li L, Huang KL, Gao Y, Cui Y, Wang G, Elrod ND *et al.* An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nature genetics* 2021; 53: 994-1005.
- 27 Gao Y, Li L, Amos CI, Li W. Analysis of alternative polyadenylation from single-cell RNA-seq using scDaPars reveals cell subpopulations invisible to gene expression. *Genome research* 2021.
- 28 Xiang Y, Ye Y, Lou Y, Yang Y, Cai C, Zhang Z *et al.* Comprehensive Characterization of Alternative Polyadenylation in Human Cancer. *Journal of the National Cancer Institute* 2018; 110: 379-389.
- 29 Piva F, Giulietti M, Burini AB, Principato G. SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. *Human mutation* 2012; 33: 81-85.
- 30 Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, 3rd *et al.* Molecular architecture of the human pre-mRNA 3' processing complex. *Molecular cell* 2009; 33: 365-376.
- 31 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 2003; 13: 2498-2504.
- 32 Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology* 2012; 16: 284-287.
- 33 Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C *et al.* The Pfam protein families database. *Nucleic acids research* 2012; 40: D290-301.

- 34 Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 2015; 4.
- 35 Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic acids research* 2013; 41: D955-961.
- 36 Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S *et al.* Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nature chemical biology* 2016; 12: 109-116.
- 37 Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nature biotechnology* 2015; 33: 495-502.
- 38 Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* 2016; 539: 309-313.
- 39 Li S, Hu Z, Zhao Y, Huang S, He X. Transcriptome-Wide Analysis Reveals the Landscape of Aberrant Alternative Splicing Events in Liver Cancer. *Hepatology (Baltimore, Md)* 2019; 69: 359-375.
- 40 Wu Q, Zhang Y, An H, Sun W, Wang R, Liu M *et al.* The landscape and biological relevance of aberrant alternative splicing events in esophageal squamous cell carcinoma. *Oncogene* 2021; 40: 4184-4197.
- 41 Lee Y, Rio DC. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annual review of biochemistry* 2015; 84: 291-323.

- 42 Cohen-Eliav M, Golan-Gerstl R, Siegfried Z, Andersen CL, Thorsen K, Ørntoft TF *et al.* The splicing factor SRSF6 is amplified and is an oncoprotein in lung and colon cancers. *The Journal of pathology* 2013; 229: 630-639.
- 43 Jensen MA, Wilkinson JE, Krainer AR. Splicing factor SRSF6 promotes hyperplasia of sensitized skin. *Nature structural & molecular biology* 2014; 21: 189-197.
- 44 Park S, Brugiolo M, Akerman M, Das S, Urbanski L, Geier A *et al.* Differential Functions of Splicing Factors in Mammary Transformation and Breast Cancer Metastasis. *Cell reports* 2019; 29: 2672-2688.e2677.
- 45 Wan L, Yu W, Shen E, Sun W, Liu Y, Kong J *et al.* SRSF6-regulated alternative splicing that promotes tumour progression offers a therapy target for colorectal cancer. *Gut* 2019; 68: 118-129.
- 46 Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 2009; 138: 673-684.
- 47 Xia Z, Donehower LA, Cooper TA, Neilson JR, Wheeler DA, Wagner EJ *et al.* Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nature communications* 2014; 5: 5274.
- 48 Chen M, Lyu G, Han M, Nie H, Shen T, Chen W *et al.* 3' UTR lengthening as a novel mechanism in regulating cellular senescence. *Genome research* 2018; 28: 285-294.
- 49 Müller-McNicoll M, Botti V, de Jesus Domingues AM, Brandl H, Schwich OD, Steiner MC *et al.* SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes & development* 2016; 30: 553-566.

- 50 Schwich OD, Blümel N, Keller M, Wegener M, Setty ST, Brunstein ME *et al.* SRSF3 and SRSF7 modulate 3'UTR length through suppression or activation of proximal polyadenylation sites and regulation of CFIm levels. *Genome biology* 2021; 22: 82.
- 51 Takagaki Y, Seipelt RL, Peterson ML, Manley JL. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* 1996; 87: 941-952.
- 52 Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu AB *et al.* CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* 2014; 510: 412-416.
- 53 Chen M, Wan L, Zhang J, Zhang J, Mendez L, Clohessy JG *et al.* Deregulated PP1 α phosphatase activity towards MAPK activation is antagonized by a tumor suppressive failsafe mechanism. *Nature communications* 2018; 9: 159.
- 54 Sun H, Ou B, Zhao S, Liu X, Song L, Liu X *et al.* USP11 promotes growth and metastasis of colorectal cancer via PPP1CA-mediated activation of ERK/MAPK signaling pathway. *EBioMedicine* 2019; 48: 236-247.
- 55 Saidy B, Kotecha S, Butler A, Rakha EA, Ellis IO, Green AR *et al.* PP1, PKA and DARPP-32 in breast cancer: A retrospective assessment of protein and mRNA expression. *Journal of cellular and molecular medicine* 2021; 25: 5015-5024.
- 56 Micalizzi DS, Maheswaran S, Haber DA. A conduit to metastasis: circulating tumor cell biology. *Genes & development* 2017; 31: 1827-1840.
- 57 Zheng D, Wang R, Ding Q, Wang T, Xie B, Wei L *et al.* Cellular stress alters 3'UTR landscape through alternative polyadenylation and isoform-specific degradation. *Nature communications* 2018; 9: 2268.

- 58 Korthauer KD, Chu LF, Newton MA, Li Y, Thomson J, Stewart R *et al.* A statistical approach for identifying differential distributions in single-cell RNA-seq experiments. *Genome biology* 2016; 17: 222.
- 59 Marinov GK, Williams BA, McCue K, Schroth GP, Gertz J, Myers RM *et al.* From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome research* 2014; 24: 496-510.
- 60 Westoby J, Herrera MS, Ferguson-Smith AC, Hemberg M. Simulation-based benchmarking of isoform quantification in single-cell RNA-seq. *Genome biology* 2018; 19: 191.
- 61 Liu W, Zhang X. Single-cell alternative splicing analysis reveals dominance of single transcript variant. *Genomics* 2020; 112: 2418-2425.
- 62 Giuliano M, Shaikh A, Lo HC, Arpino G, De Placido S, Zhang XH *et al.* Perspective on Circulating Tumor Cell Clusters: Why It Takes a Village to Metastasize. *Cancer research* 2018; 78: 845-852.
- 63 Rodrigues P, Vanharanta S. Circulating Tumor Cells: Come Together, Right Now, Over Metastasis. *Cancer discovery* 2019; 9: 22-24.
- 64 Falcão AM, van Bruggen D, Marques S, Meijer M, Jäkel S, Agirre E *et al.* Disease-specific oligodendrocyte lineage cells arise in multiple sclerosis. *Nature medicine* 2018; 24: 1837-1844.
- 65 Linker SM, Urban L, Clark SJ, Chhatriwala M, Amatya S, McCarthy DJ *et al.* Combined single-cell profiling of expression and DNA methylation reveals splicing regulation and heterogeneity. *Genome biology* 2019; 20: 30.
- 66 Muñoz JF, Delorey T, Ford CB, Li BY, Thompson DA, Rao RP *et al.* Coordinated host-pathogen transcriptional dynamics revealed using sorted subpopulations and single macrophages infected with *Candida albicans*. *Nature communications* 2019; 10: 1607.

Figures

Figure 1

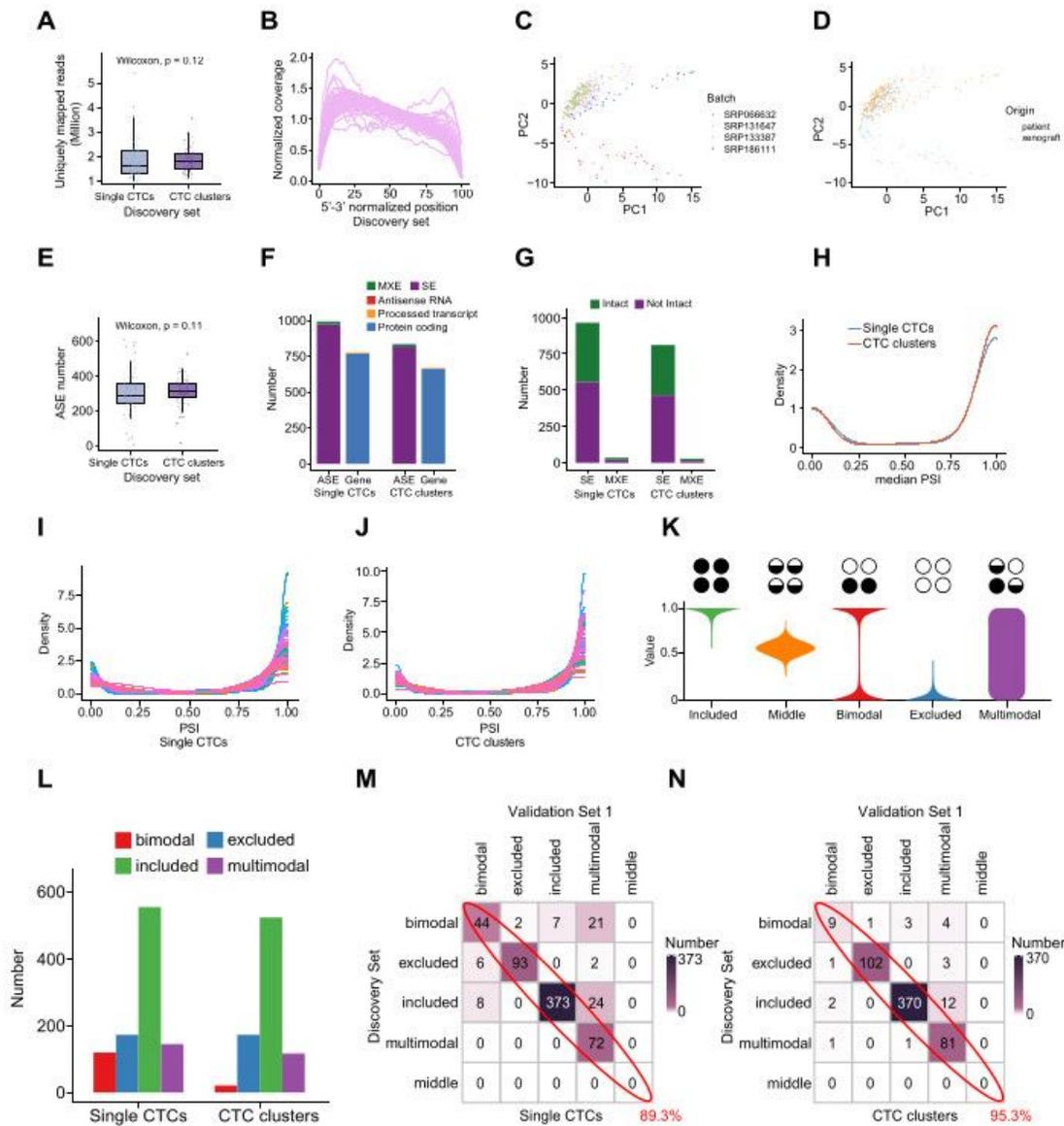


Figure 1

Overview of AS profiles in single CTCs and CTC clusters. (A) Boxplot showing uniquely mapped reads for single CTCs and CTC clusters from discovery set (SRP131647). (B) The normalized coverage of reads from scRNA-seq libraries from discovery set. (C-D) Principal component analysis of all cells in 4 datasets

based on AS profiles. (E) Boxplot showing the number of AS events detected for each cell from discovery set. (F) Bar plot showing the number of detected AS events and their related genes in single CTCs and CTC clusters. (G) The proportion of the alternative exons with intact codons. (H) Distribution of the median value of PSI in single CTCs and CTC clusters. (I-J) Distribution of PSI values of each cell in single CTCs group (I) and CTC clusters group (J). (K) A pattern diagram of the five modalities. (L) Statistics on the modalities assigned to AS events in single CTCs and CTC clusters. (M) Heatmap showing the consistency of modalities assigned to AS events in single CTCs from discovery set and validation set 1 (SRP133387). The darker the color of the principal diagonal, the stronger the consistency. (N) The consistency of modalities assigned to AS events in CTC clusters from discovery set and validation set 1.

Figure 2

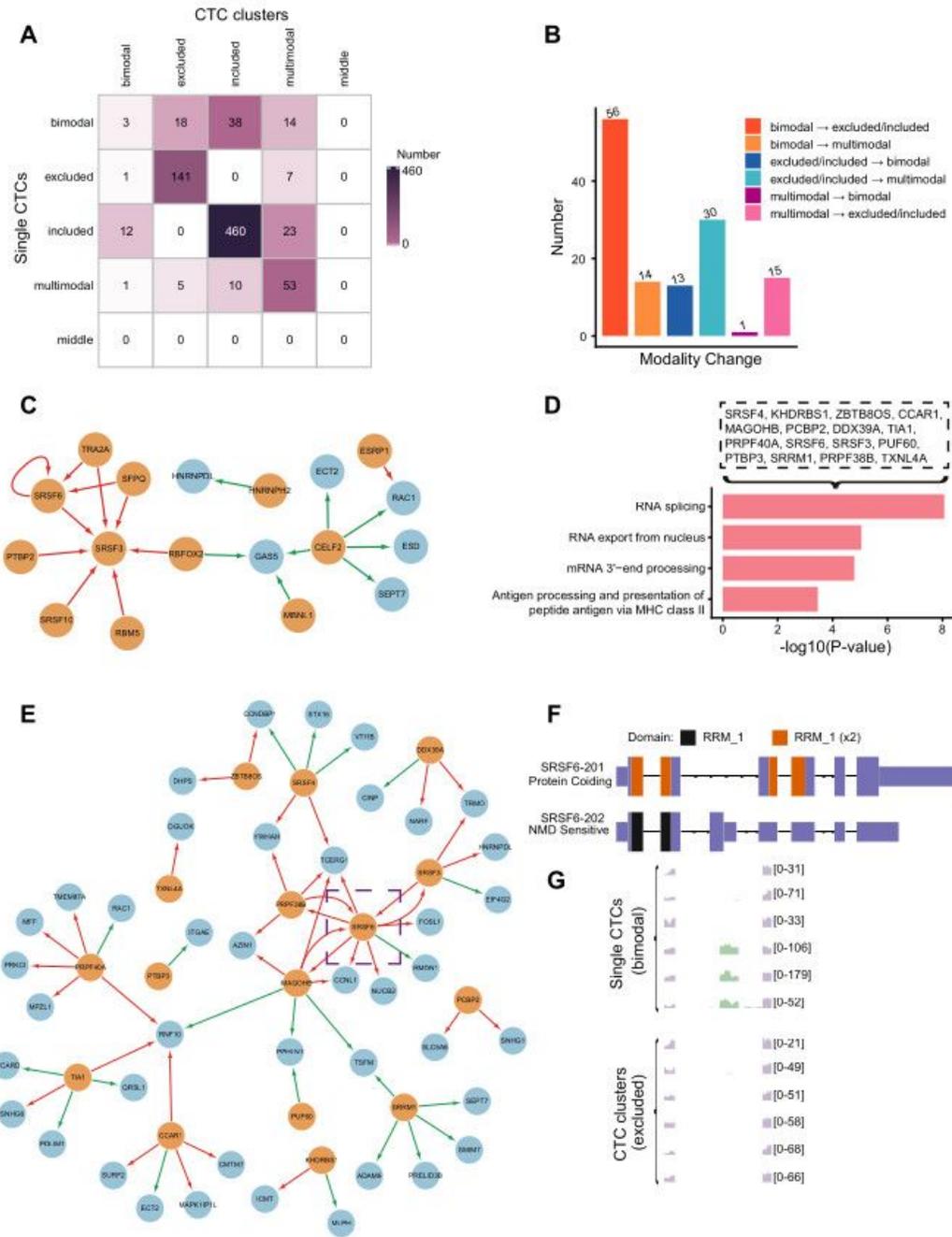


Figure 2

The differential AS events between single CTCs and CTC clusters. (A) Heatmap showing the AS events that changed modalities between single CTCs and CTC clusters. (B) Summary of the pattern of differential AS events. (C) Network illustrating the differential AS events regulated by the expression of SFs. Orange circles indicate SFs, while blue circles indicate differential AS events-related genes. Red lines mean that the expression of SFs was positively correlated with the PSI value of AS events. Green lines

mean negative correlations. (D) GO enrichment analysis of differential AS events-related genes. (E) Network illustrating the regulatory effects of differentially spliced genes that participate in “RNA splicing” on differential AS events. (F) The schematic plot of two transcripts of SRSF6, in which the lines indicate introns, the higher boxes indicate exons in the open reading frame (ORF), and the lower boxes indicate the 5' untranslated region (5'UTR) and 3' untranslated region (3'UTR). (G) Sashimi plot exhibiting the coverage of SRSF6’s alternative exon (green color) in single CTCs and CTC clusters.

Figure 3

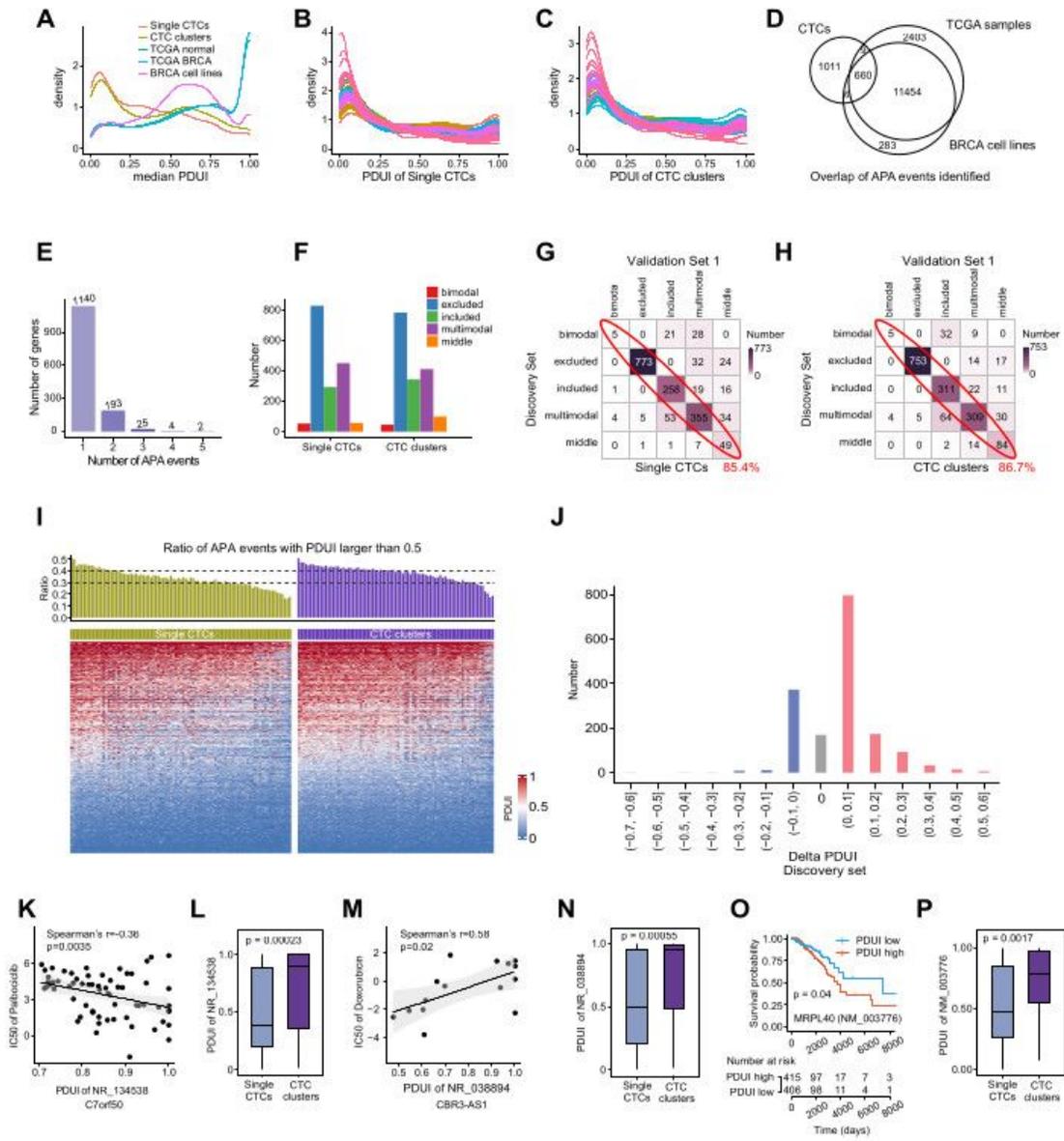


Figure 3

Overview of APA profiles in single CTCs and CTC clusters. (A) Distribution of median PDUI values in single CTCs, CTC clusters, TCGA normal breast tissues, TCGA breast cancers, and breast cancer cell lines. (B-C) Distribution of PDUI values of each cell in single CTCs (B) and CTC clusters (C). (D) Overlap of detected APA events in circulating tumor cells, TCGA samples, and breast cancer cell lines. (E) Statistics on the number of APA events implicated in each gene. (F) The distribution of modalities assigned to APA events in single CTCs and CTC clusters. (G) Heatmap showing the consistency of modalities assigned to APA events in single CTCs from discovery set and validation set 1. (H) The consistency of modalities assigned to APA events in CTC clusters from discovery set and validation set 1. (I-J) Heatmap and Bar plot showing global 3' UTRs lengthening in CTC clusters. (K) The PDUI value of C7orf50 was negatively correlated with the IC50 value of Palbociclib. (L) The PDUI value of C7orf50 was significantly larger in CTC clusters. (M) The PDUI value of CBR3-AS1 was positively correlated with the IC50 value of Doxorubicin. (N) The PDUI value of CBR3-AS1 was significantly larger in CTC clusters. (O) The Kaplan-Meier survival curve suggested that high PDUI of MRPL40 was associated with poor prognosis of breast cancer patients. (P) The PDUI value of MRPL40 was significantly larger in CTC clusters.

Figure 4

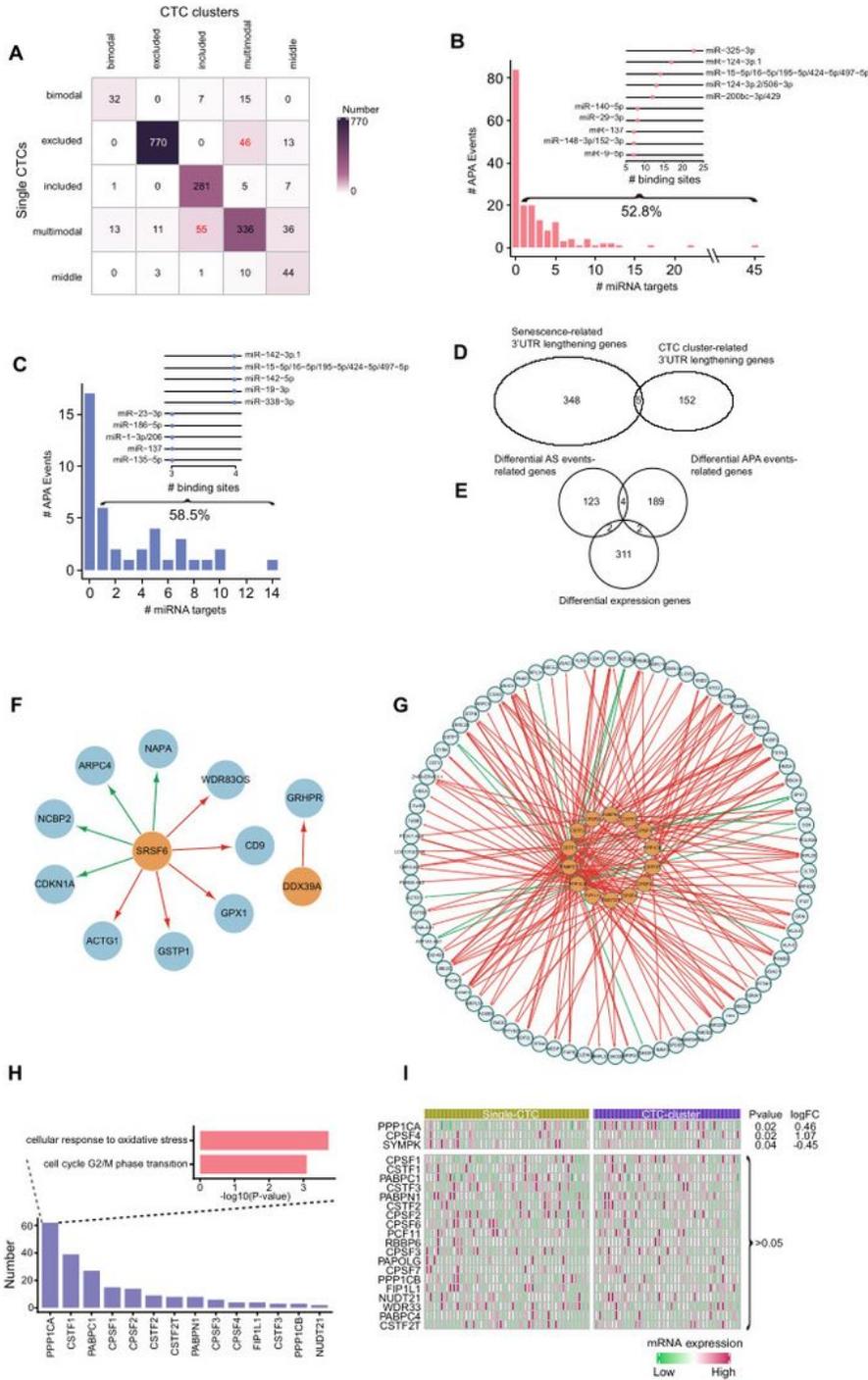


Figure 4

PPP1CA mediated the widespread perturbation of APA profiles in CTC clusters. (A) Heatmap showing APA events that changed modalities between single CTCs and CTC clusters. (B) Bars showing the number of APA events that gain highly conserved miRNA binding sites because of 3' UTR lengthening. Dots showing the number of binding sites each miRNA binds. (C) Number of APA events that lose highly conserved miRNA binding sites because of 3' UTR shortening. (D) The overlap between senescence-

related 3'UTR lengthening genes and CTC cluster-related 3'UTR lengthening genes. (E) Venn diagram showing little overlap between differential APA genes, differential AS genes, and differential expression genes. (F) Network illustrating the regulatory effects of differentially spliced genes that participate in "mRNA 3'-end processing" on differential APA events. (G) Network illustrating the differential APA events regulated by the expression of core APA factors. (H) Purple bars showing the number of differential APA events regulated by each core APA factors. Red bars showing the GO terms enriched by PPP1CA-regulated genes. (I) The expression of 22 core APA factors in single CTCs and CTC clusters.

Figure 5

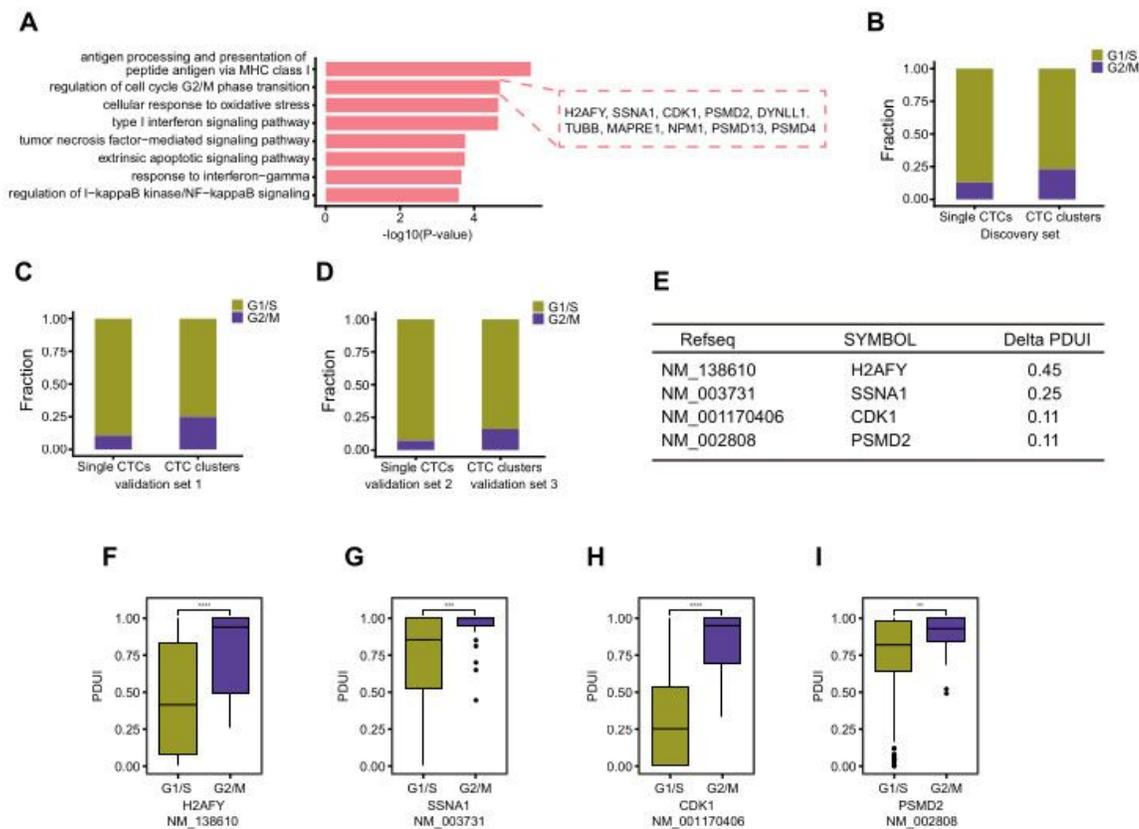


Figure 5

The functional meaning of differential APA events. (A) GO terms enriched by differential APA events-related genes. (B-D) Bar plot showing the cell cycle status of single CTCs and CTC clusters derived from discovery set (B), validation set 1 (C), validation set 2 and validation set 3 (D). (E) The delta median PDUI of four APA events between single CTCs and CTC clusters. Delta PDUI > 0 means that the corresponding APA event favors long 3'UTR in CTC clusters. (F-I) PDUI values of the four genes involved in "regulation of cell cycle G2/M phase transition" were connected with the cell cycle status.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)
- [SupplementaryFigures.docx](#)