Journal of Translational Medicine

RESEARCH

Open Access

CPSF4-mediated regulation of alternative splicing of HMG20B facilitates the progression of triple-negative breast cancer

Guansheng Zhong[1†], Qinyan Shen[2†], Xinli Zheng[3], Kun Yu[4], Hongjiang Lu[5], Bajin Wei[1], Haidong Cui[1], Zhijun Dai[1] and Weiyang Lou[1*]

Abstract

Background Aberrant alternative splicing (AS) contributes to tumor progression. A crucial component of AS is cleavage and polyadenylation specificity factor 4 (CPSF4). It remains unclear whether CPSF4 plays a role in triple-negative breast cancer (TNBC) progression through AS regulation. In this study, our objective is to investigate the prognostic value of CPSF4 and pinpoint pivotal AS events governed by CPSF4 specifically in TNBC.

Methods We examined the expression levels and prognostic implications of CPSF4 in patients diagnosed with TNBC through public databases. CPSF4-interacting transcripts, global transcriptome, and alternative splicing were captured through RNA immunoprecipitation sequencing (RIP-seq) and RNA sequencing (RNA-seq). The top 10 CPSF4-regulated alternative splicing events (ASEs) were validated using qRT-PCR. TNBC cells transfected with high mobility group 20B (HMG20B) siRNA were subjected to CCK-8 and transwell assays.

Results In TNBC, CPSF4 exhibited heightened expression levels and was correlated with unfavorable prognosis. Overexpression of CPSF4 significantly promoted colony formation and migration, whereas knockdown of CPSF4 had the opposite effect. Inhibition of CPSF4 altered the transcriptome profile of MDA-MB-231 cells. CPSF4-regulated numerous genes showed enrichment in cancer-related functional pathways, including mRNA processing, cell cycle, RNA transport, mRNA surveillance pathway, and apoptosis. CPSF4-regulated ASEs were highly validated by qRT-PCR. CPSF4 modulated selective splicing events by inhibiting alternative 3′ splice site events of HMG20B and promoted cell proliferation, migration, and invasion.

Conclusion CPSF4 promotes TNBC progression by regulating AS of HMG20B. These findings contribute to the development of more useful prognostic, diagnostic and potentially therapeutic biomarkers for TNBC.

Keywords Triple-negative breast cancer, Alternative splicing, RNA sequencing, CPSF4, Biomarker

†Guansheng Zhong and Qinyan Shen have contributed equally to this work.

*Correspondence:
Weiyang Lou
lou15968783311@sina.com
Full list of author information is available at the end of the article

Introduction

Triple-negative breast cancer (TNBC) represents a subtype of breast cancer [1]. TNBC accounts for 10% to 17% of all breast cancers and is associated with early metastasis, high recurrence rates, and resistance to therapy, attributed to its poorly differentiated and highly malignant biological features, resulting in poor prognosis [2–4]. Because TNBC patients do not derive advantages from hormonal therapy or HER2-targeted medications, chemotherapy stands as the primary modality for systemic treatment in these cases [5, 6]. Neoadjuvant chemotherapy (NACT) has emerged as a novel strategy preceding surgical tumor resection, aiming to render initially unresectable tumors amenable to surgery, allowing for breast-conserving surgery, improving long-term prognosis, and providing an opportunity for preoperative testing of novel agents [7, 8]. Thus, enhancing patient outcomes and survival necessitates advancements in TNBC biomarker screening, the development of predictive factors for treatment response, and novel therapeutic agents.

Alternative splicing (AS) is a post-transcriptional process generating alternative mRNA transcripts crucial for normal development and contributing to the proteomic complexity of the mammalian genome [9, 10]. Studies across the genome reveal that most human genes undergo AS [11, 12]. There is evidence indicating the involvement of AS in cellular differentiation, determining lineages, acquiring tissue identity, maintaining tissues, and developing organs [13, 14]. Moreover, the systematic and coordinated alteration of AS, orchestrated by functionally associated ribonucleic acid (RNA)-binding proteins (RBPs), could influence the process of carcinogenesis [15]. Dysregulation of alternative splicing, leading to changes in splice isoforms, is implicated in regulating numerous cancer phenotypes. These include but are not limited to cellular physiological processes, angiogenesis, aberrant energy metabolism, and evasion of immune responses [16, 17]. Aberrations in alternative splicing have been documented to impact various genes, including BRCA1 DNA repair associated (BRCA1), fibroblast growth factor receptor 1/2 (FGFR1/2), human epidermal growth factor receptor 2 (HER2), and differentiation mediation transcription factor 1 (DMTF1)/krüppel-like factor 6 (KLF6), across different subtypes of breast cancer. These abnormalities often result in distinct or even contradictory functionalities [18]. Cleavage and polyadenylation specificity factor 4 (CPSF4) is an important mRNA processing factor that plays a key role in the post-transcriptional regulation of gene expression [19]. In cancer research, the abnormal expression of CPSF4 has been closely associated with the occurrence and progression of various tumors [20]. Notably, CPSF4 is also believed to play a significant role in the process of AS, as it can influence the production of different splice isoforms of genes by regulating the activity of splicing factors and the selection of splice sites, thereby altering the phenotype of cancer cells [21]. CPSF4 plays a critical role in the maturation of mRNA 3' ends and polyadenylation [19]. Studies suggest CPSF4 involvement in the splicing of tumor protein p53 (TP53) mRNA [22]. However, the specific mechanisms by which CPSF4 operates in AS in TNBC remain unclear.

High mobility group 20B (HMG20B) is a high mobility group protein primarily located in the cell nucleus [23]. It contains a highly conserved HMG domain, which enables it to bind to DNA and regulate transcriptional activity. As a nuclear protein, HMG20B interacts with DNA and other proteins to modulate transcription factor activity and participate in cell cycle progression and proliferation control. Studies indicate that HMG20B complexes can mediate chromatin remodeling and repression of specific genes by altering chromatin structure, thus playing an essential role in neuronal differentiation [24]. Previous research has reported that HMG20B is upregulated in hepatocellular carcinoma and may serve as a novel prognostic marker [25]. Additionally, mutant forms of HMG20B in epithelial cancers can dominantly induce cytokinesis failure [26]. However, its functional mechanisms in TNBC remain unclear and warrant further investigation.

In this study, RNA sequencing (RNA-seq) was utilized to capture the transcriptome changes resulting from CPSF4 knockdown in TNBC cell lines, specifically MDA-MB-231. To comprehensively identify CPSF4-associated mRNAs, we conducted RNA immunoprecipitation sequencing (RIP-seq). Through integrated analysis, it was discovered that CPSF4 binding has the potential to influence a variety of alternative splicing events in MDA-MB-231 cells. Specifically, CPSF4-mediated regulation of HMG20BAS was found to facilitate the advancement of TNBC.

Materials and methods

Analysis platforms and databases

University of Alabama at Birmingham Cancer Data Analysis Portal (UALCAN; http://ualcan.path.uab.edu/) database is an open web platform for the analysis and visualization of cancer genomic data. This database specifically focuses on information extracted from The Cancer Genome Atlas (TCGA) and other public databases, providing an in-depth analysis of cancer-related gene expression [27]. Based on UALCAN database, we analyzed the expression levels of CPSF4 in breast cancer (BRCA) compared to normal tissue and among different BRCA subtypes. The Kaplan–Meier Plotter (http://kmplot.com)

database assesses survival rates associated with various genes across 21 cancer types [28].

Cell culture and transfection

The breast epithelial cells (MCF10A) and various human breast cancer cell lines (MDA-MB-231, MDA-MB-468, BT-549, and BT-20) were sourced from the American Type Culture Collection (ATCC, USA). The MCF10A cells were cultured in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F-12) (Thermo Fisher Scientific, Inc.). MDA-MB-231, MDA-MB-468, BT-549, and BT-20 cell lines were cultured in Roswell Park Memorial Institute Medium 1640 (RPMI 1640) medium (GIBCO, USA), supplemented with 10% fetal bovine serum (FBS) (GIBCO, USA). The cells were kept in a humidified incubator at 37 ℃ with a 5% CO_2 atmosphere. Sangon Biotech synthesized si-CPSF4 and si-HMG20B, which were subsequently transfected into the cells using Lipofectamine™ 2000 transfection reagent, following the guidelines provided by the manufacturer. The overexpression vector (CPSF4-pcDNA3.1–3*FLAG) (NM_006693.4, Youbio, China) was transfected according to the manufacturer's instructions. The small interfering RNA (siRNA) sequences are as follows: si-CPFS4#1: GGGCGCUGCUGUCUGUGAATT; si-CPFS4#2: GCC ACAUCAGUGGUGAGAATT; si-CPFS4#3: GGCCCU CGUGUAAAUUCAUTT; si-HMG20B#1: GAGAGA AGCAGCAGUACAU; si-HMG20B#2: GAGCUUCGG CGCUUGCGGA; si-HMG20B#3: GGACACAGGGCA GACGAAA; si-NC: UUCUCCGAACGUGUCACGUTT.

Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

Firstly, total RNA was isolated from the cells utilizing TRIzol reagent, in accordance with the guidelines provided by the manufacturer (Invitrogen; Thermo Fisher Scientific, Inc.). Subsequently, reverse transcription was conducted using PrimeScript RT Master Mix (Takara Biotechnology Co.). For quantitative PCR (qPCR) analysis, Takara Biotechnology TB green premix Ex Taq™ II (Catalog number: RR820A) was employed. The relative expression levels of CPSF4 or HMG20B were determined using the $2^{-\Delta\Delta Cq}$ method [29], with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) serving as the internal control. The sequences of primers used are the following: CPFS4: F: 5'-GTCCTCCTTGATCCAGTT AA-3'; R: 5'-CATCTGTTGGCGTAGTGT-3';HMG20B: F: 5'-CACCTTCGATGTTCCCATCT −3'; R: 5'-GAA GGCCACATTCATCTTCC −3'; GAPDH: F: 5'-GGT CGGAGTCAACGGATTTG-3'; R: 5'-GGAAGATGG TGATGGGATTTC-3'.

Western blot analysis

The protein samples were initially loaded onto 10% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-*PAGE*) gels and subsequently transferred onto polyvinylidene fluoride (PVDF) membranes (ISEQ00010, Millipore). Following this, the membranes underwent blocking with 5% skim milk in buffer solution for one hour. They were then subjected to overnight incubation with primary antibodies at 4℃, followed by a subsequent one-hour incubation with horseradish peroxidase-conjugated secondary antibodies at room temperature. Membranes were visualized using chemiluminescence. Subsequently, Western blot bands were quantified using Image J software. The primary antibodies used were CPSF4 antibody (15023-1-AP, Proteintech) and GAPDH antibody (60004-1-IG, Proteintech). The secondary antibody employed was HRP-conjugated Affinipure Goat Anti-Rabbit IgG(H + L) (SA00001-2, Proteintech).

Colony formation assay

A single-cell suspension was seeded into 6 cm culture dishes at a density of 800 cells per well. After 14 days, the cells were fixed with 4% paraformaldehyde and stained with 1% crystal violet. The number of visible colonies was then counted.

RNA extraction and library preparation

MDA-MB-231 cell samples were flash-frozen in liquid nitrogen and ground into a fine powder to facilitate subsequent RNA extraction. Total RNA extraction was conducted utilizing TRIzol (Life Technologies, Carlsbad, CA, USA), followed by additional purification through two rounds of phenol–chloroform treatment. To prepare the sequencing libraries, 1 μg of total RNA per sample was utilized in conjunction with the VAHTS Stranded mRNA-seq Library Prep Kit (Vazyme, Nanjing, China). Library preparation procedures were performed in accordance with the instructions. Subsequently, paired-end sequencing analysis (150 nt) was carried out using the Illumina HiSeq X Ten system.

Data preprocessing and alignment

Raw reads containing more than 2-N bases were discarded. Adapter sequences and low-quality bases in the raw sequencing reads were trimmed using the FASTX toolkit [30]. Reads shorter than 16 nucleotides were filtered out from the dataset. The remaining clean reads were aligned to the human genome (GRCH38) utilizing TopHat2, with allowance for up to four mismatches. Only reads that mapped uniquely were retained for gene read counting and subsequent calculation of fragments per kilobase per million mapped reads (FPKM).

Differential expression analysis

We conducted alignment of the quality-filtered reads against the GRCH38 using TopHat2, with a tolerance for up to 4 mismatches [30]. Subsequently, reads that mapped uniquely were utilized for gene read counting and the computation of FPKM. Subsequently, gene expression levels were quantified using FPKM values. Differential expression genes (DEGs) were identified from RNA-seq data using the R software package edgeR, which detects differentially expressed genes [31]. Genes with $P<0.05$ and fold change (FC) ≥ 2 or ≤ 0.5 were considered significantly differentially expressed between the siCPSF4 and control groups.

RIP-seq data analysis

The read mapping and quality filtering methods used for CPSF4 RIP-seq data mirrored those applied to RNA-seq data. Upon aligning uniquely mapped reads to the genome sequence, we utilized the random IP method to identify binding sites (peaks) of CPSF4 on transcripts. Peaks with a $P<0.05$ were selected as CPSF4 binding sites. Subsequently, we used the HOMER software to extract peak sequences and detect enriched motifs.

Alternative splicing analysis

We utilized the ABLas pipeline to define and quantify alternative splicing events (ASEs) and regulated alternative splicing events (RASEs) induced by siCPSF4 [32, 33]. These events were comprehensively categorized into types based on splice junction read counts. Statistical analysis was performed using Fisher's exact test for sample comparisons to ascertain statistical significance, with a significance threshold set at $P \leq 0.05$ for the detection of RASEs.

Functional annotation of DEGs

In order to evaluate the functional categories of DEGs, we employed the KOBAS 2.0 server to determine both Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [34]. Enriched terms were identified utilizing the hypergeometric test, with subsequent control of the Benjamini–Hochberg False Discovery Rate (FDR) procedure.

Validation of ASEs

To validate the authenticity of ASEs in cells, a subset of RASEs was quantified via qRT-PCR, with normalization against the reference gene GAPDH. For the quantitative analysis of two distinct splicing isoforms associated with specific ASEs using qRT-PCR, two pairs of primers were meticulously designed to selectively amplify these isoforms subsequent to the initial synthesis of the first-strand complementary deoxyribonucleic acid (cDNA) utilizing random primers. The altered ratio of AS events in qRT-PCR was calculated using the formula: AS transcripts level /Model transcripts level [35].

Cell proliferation assay

Cells (2×10^3/well) were seeded into 96-well plates and subsequently incubated at 37 ℃ with 5% CO_2. Cell Counting Kit-8 (CCK-8) solution (Dojindo Molecular Technologies, Inc.) was used for determination. The results were assessed utilizing microplate reader (Bio-Rad Laboratories, Inc.).

Cell migration and invasion assay

For migration assays, cells were seeded into the upper chambers of transwell inserts (Corning, Inc.) at a density of 5×10^4 cells, while for invasion assays, cells were seeded at a density of 1×10^5 cells. The transwell inserts were either left uncoated or pre-coated with Matrigel (BD Biosciences). Cells were cultured at 37 ℃ with 5% CO_2 for 24 h, and then the non-migrated cells were removed. Subsequently, they were stained with 1% crystal violet for one hour and enumerated using inverted microscope (Leica Microsystems GmbH).

Statistical analysis

Data analysis utilized GraphPad Prism V9.0 (Dotmatics) and SPSS 19.0 (IBM Corp.). Group differences were assessed via unpaired Student's t-test ($P<0.05$). Results represent mean ± standard deviation (SD) from three independent experiments.

Result

CPSF4 overexpression in TNBC is negatively correlated with patient prognosis

To understand the role of CPSF4 in breast cancer, we analyzed its expression levels in the UALCAN database and breast cancer cell lines. Results revealed higher CPSF4 expression in 1097 tumor samples compared to 114 normal tissue samples (Fig. 1A). Furthermore, the expression of CPSF4 was found to be increased in TNBC in comparison to both normal breast tissue and other subtypes of breast cancer (Fig. 1B). Subsequently, CPSF4 levels were assessed in breast cancer cell lines, with higher expression observed in MDA-MB-231 and MDA-MB-468 cells, chosen for subsequent experiments (Fig. 1C). Furthermore, we examined the relationship between CPSF4 expression and patient survival rates in breast cancer. Elevated CPSF4 mRNA levels correlated with shorter distant metastasis-free survival (DMFS) in breast cancer patients, indicating a potential association between high CPSF4 expression and adverse prognosis (Fig. 1D).

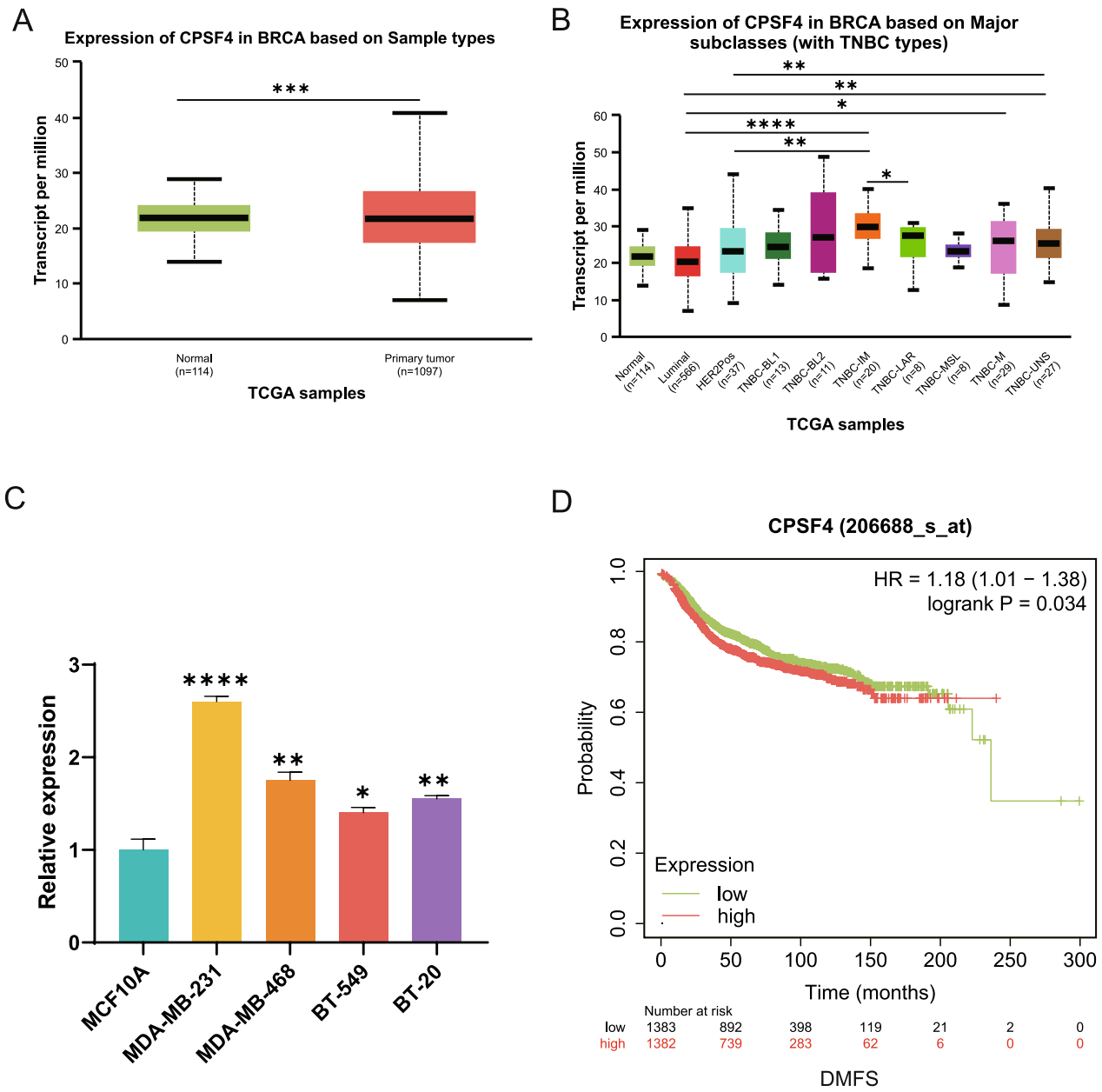


Fig. 1 Upregulation of Cleavage and Polyadenylation Specificity Factor 4 (CPSF4) expression in Triple-Negative Breast Cancer (TNBC) correlates with poor patient prognosis. Expression analysis of CPSF4 in breast cancer (**A**) and different breast cancer subtypes (**B**) in the University of Alabama at Birmingham Cancer Data Analysis Portal (UALCAN) database. **C** Quantitative reverse transcription polymerase chain reaction (qRT-PCR) detection of CPSF4 expression levels in breast cancer cell lines. **D** Analysis of the prognostic value of different CPSF4 expression levels in breast cancer using the Kaplan–Meier Plotter database. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$

CPSF4 promotes cell proliferation and migration in TNBC

To elucidate the function of CPSF4, we investigated its role in TNBC cell lines. We established CPSF4 overexpression and knockdown models in MDA-MB-231 and MDA-MB-468 cell lines. Subsequently, the overexpression efficiency and sh-CPSF4 silencing efficiency of CPSF4 in MDA-MB-231 and MDA-MB-468 cell lines were analyzed by qRT-PCR and Western Blot (Supplementary Fig. 1A-1F). Colony formation assay results demonstrated that CPSF4 knockdown inhibited cell colony formation (Fig. 2A, B), while CPSF4 overexpression had the opposite effect (Fig. 2C). Transwell assay results showed that CPSF4

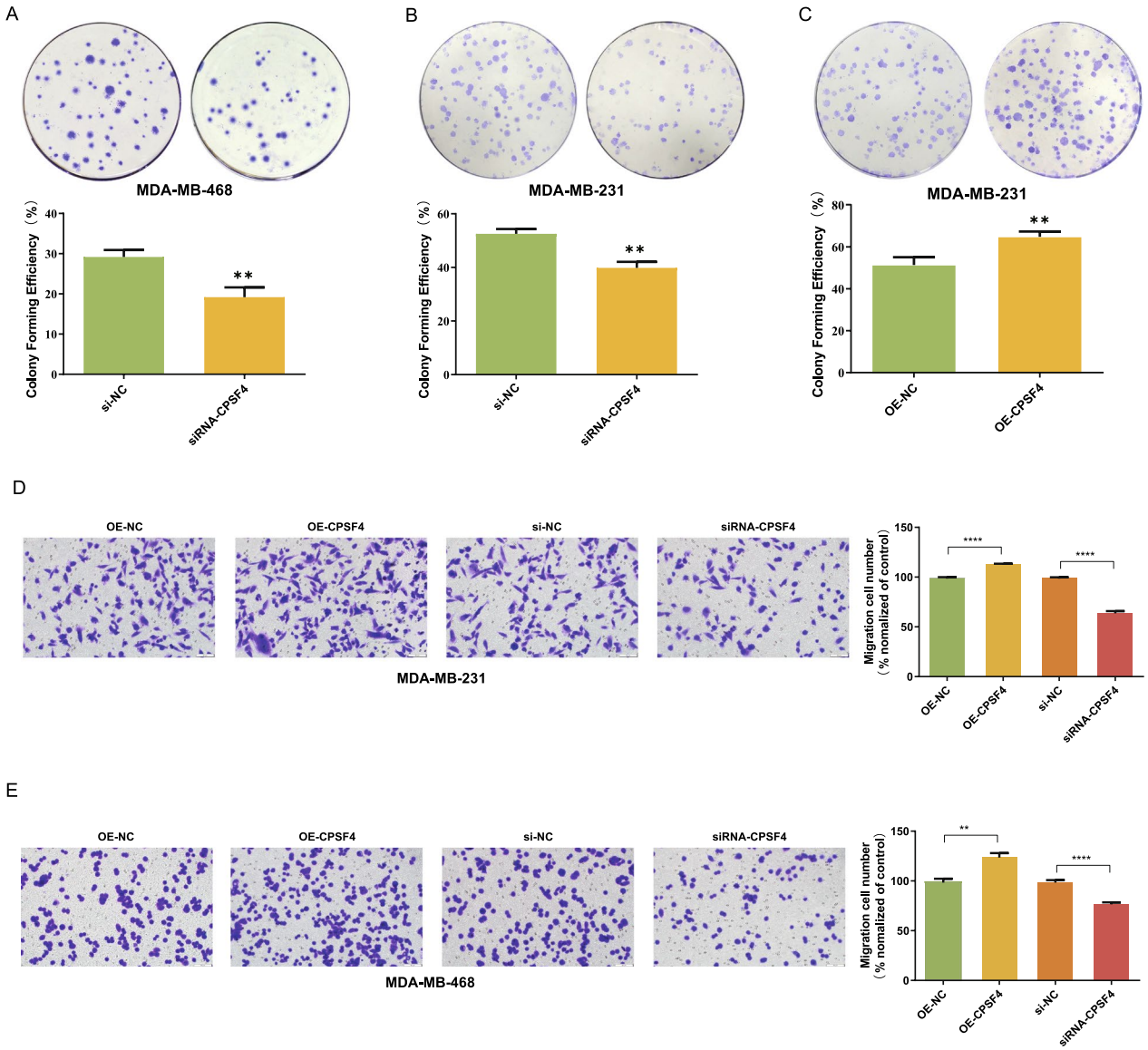


Fig. 2 CPSF4 Promotes Cell Proliferation and Migration in TNBC. **A**, **B** Cell proliferation of MDA-MB-231 and MDA-MB-468 cells transfected with siRNA-NC or si-CPSF4 detected by colony formation assay. **C** Cell proliferation of MDA-MB-231 cells transfected with overexpression negative control (OE-NC) or OE-CPSF4 detected by colony formation assay. **D** Transwell assay showing the migration of MDA-MB-231 cells with CPSF4 overexpression or knockdown. **E** Transwell assay showing the migration of MDA-MB-468 cells with CPSF4 overexpression or knockdown. $^{**}P < 0.01$, $^{****}P < 0.0001$

overexpression promoted cell migration, whereas CPSF4 knockdown inhibited cell migration (Fig. 2D, E).

CPSF4-mediated transcriptional differences in TNBC

To elucidate the mechanism by which CPSF4 regulates AS in TNBC, RNA-seq was employed to profile the transcriptomes of si-CPSF4 and control cells. For both si-CPSF4 and control cells, a total of six RNA-seq libraries were prepared and subsequently subjected to sequencing. Each experimental condition was replicated three times biologically. FPKM values further confirmed significant CPSF4 knockdown in MDA-MB-231 cells (Fig. 3A). The Pearson correlation coefficients between the si-CPSF4 and control groups were notably high, exceeding 0.992, suggesting a considerable similarity in gene expression across the majority of genes. Despite this, unsupervised hierarchical clustering unveiled a clear demarcation between the si-CPSF4 and control samples. Notably, the three biological replicates clustered together, as illustrated in Fig. 3B. Differential expression

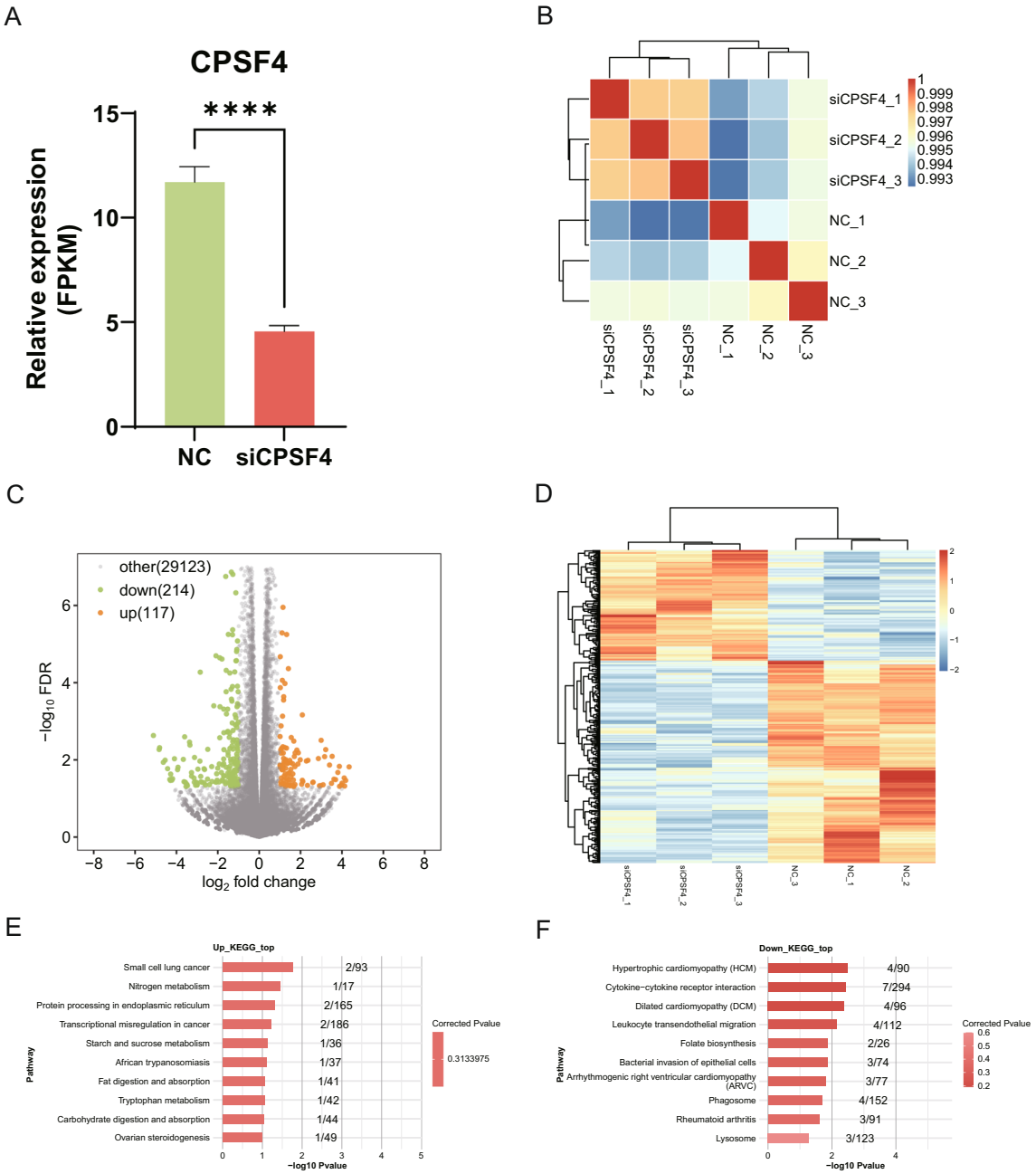


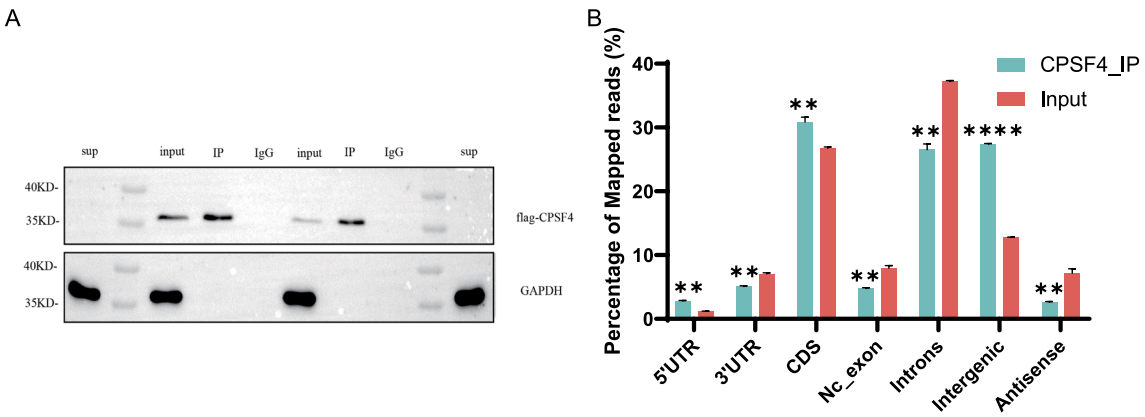
Fig. 3 RNA sequencing (RNA-seq) analysis of transcriptome profiles regulated by CPSF4 in TNBC cells. **A** Evaluation of knockdown efficiency of si-CPSF4 in MDA-MB-231 cells by qRT-PCR. **B** The heatmap displays the hierarchical clustering of the Pearson correlation matrix for transcriptional expression levels in control and si-CPSF4 samples. **C** A volcano plot of differentially expressed genes associated with CPSF4, with upregulated and downregulated genes shown in orange and green, respectively. **D** Hierarchical clustering of differentially expressed genes (DEGs) in control and si-CPSF4 samples. **E** Top 10 kyoto encyclopedia of genes and genomes (KEGG) pathways enriched in upregulated genes. **F** Top 10 KEGG pathways enriched in downregulated genes. ****$P < 0.0001$

analysis identified 331 DEGs in RNA-seq data, with 117 upregulated and 214 downregulated in the CPSF4-overexpressing group (Fig. 3C). Furthermore, hierarchical clustering of normalized FPKM values of DEGs showed clear separation between si-CPSF4 and control samples, with high consistency among the three replicate datasets (Fig. 3D). The KEGG enrichment analysis conducted on DEGs regulated by CPSF4 unveiled enrichment in several pathways. Specifically, the KEGG pathways associated with genes upregulated by CPSF4 include ovarian

steroidogenesis, carbohydrate digestion and absorption, and tryptophan metabolism (Fig. 3E). Meanwhile, the genes downregulated by CPSF4 are enriched in aspects related to cytokine-cytokine receptor interaction, folate biosynthesis, and lysosomes (Fig. 3F). These results highlight the significant impact of CPSF4 knockdown on the transcriptional expression of a gene set and provide insights into the potential roles of these DEGs.

CPSF4 binding characteristics revealed by RIP-seq analysis

Studies have indicated the pivotal role of CPSF4 in the regulation of AS [19, 22]. To better understand the role of CPSF4 in TNBC progression, RIP-seq analysis was conducted in MDA-MB-231 cells to confirm the transcripts interacting with CPSF4. In two replicates, a substantial amount of CPSF4 was found in the immunoprecipitated fraction in MDA-MB-231 cells (Fig. 4A). After trimming adapters and filtering out low-quality reads, sequencing of cDNA libraries derived from RNA input and CPSF4 products produced a total of 41072240 and 47433886 reads, respectively, from two CPSF4 immunoprecipitation replicates. Similarly, two RNA input replicates yielded 51864506 and 61267184 reads, respectively. Upon mapping unique CPSF4 reads, notable enrichment was observed in the 3' untranslated region (3'UTR), coding sequence (CDS) and 5' untranslated region (5'UTR) regions of the reference genome (Fig. 4B). Additionally, the Homer software detected significantly enriched motifs within CPSF4 binding peaks, with "GAAGAA G" consistently identified among the top 10 motifs in both independent experiments (Fig. 4C). In summary, RIP-seq analysis demonstrated distinct RNA binding characteristics of CPSF4 in TNBC, shedding light on its potential regulatory roles in the disease progression.



C

CPSF4 IP1

Rank	Motif	P-value	log P-pvalue	% of Targets	% of Background
1	CUACG	1e-216	-4.980e+02	69.26%	45.99%
2	GACGAGGA	1e-177	-4.077e+02	69.89%	48.94%
3	GAAGAAG	1e-174	-4.020e+02	67.66%	46.74%
4	CAUCGAG	1e-171	-3.960e+02	48.67%	28.64%
5	GAGCCGGA	1e-161	-3.718e+02	67.64%	47.54%
6	GUACCG	1e-156	-3.605e+02	80.51%	62.19%
7	UGGCGGA	1e-139	-3.208e+02	60.22%	41.45%
8	CAAGGUGA	1e-135	-3.128e+02	69.39%	51.11%
9	CGGAUGA	1e-135	-3.114e+02	82.78%	66.35%
10	CGGCG	1e-107	-2.479e+02	94.88%	84.53%

CPSF4 IP2

Rank	Motif	P-value	log P-pvalue	% of Targets	% of Background
1	GAAGAAG	1e-281	-6.474e+02	77.66%	54.48%
2	CGACGA	1e-243	-5.596e+02	73.01%	51.01%
3	CAUCGA	1e-199	-4.604e+02	75.54%	55.98%
4	GAGCGGGA	1e-194	-4.469e+02	72.13%	52.51%
5	GAACG	1e-185	-4.273e+02	79.77%	61.63%
6	GGUGAG	1e-170	-3.924e+02	40.60%	23.45%
7	GCGGCGG	1e-161	-3.726e+02	64.35%	46.10%
8	GACGGACG	1e-161	-3.719e+02	82.81%	66.60%
9	GAGGA	1e-148	-3.425e+02	61.07%	43.53%
10	UCGCCA	1e-131	-3.025e+02	62.12%	45.66%

Fig. 4 CPSF4 binding spectrum and binding motifs revealed by RNA immunoprecipitation sequencing (RIP-seq) analysis. **A** Western blot analysis of CPSF4 protein in MDA-MB-231 cells. **B** Distribution of reads across the entire reference genome. **C** Motif analysis using HOMER software showing the top 10 preferred binding motifs of CPSF4. $^{**}P < 0.01$, $^{****}P < 0.0001$

Identification and functional analysis of CPSF4 binding genes

To validate reliable CPSF4 binding genes, we employed the methods of ABlife and Piranha to call CPSF4 binding peaks. The computation of overlapping peaks was conducted through ABlife software. Across two RIP-seq sample replicates, a total of 3953 peak clusters showed consistent overlapping (Fig. 5A). Following this, we performed GO and KEGG enrichment analyses to delve deeper into the potential biological functions associated with the genes bound by CPSF4. The enrichment analysis of biological processes in GO terms unveiled associations between CPSF4-bound genes and several processes. Notably, these processes included negative regulation of transcription by RNA polymerase II, translation and apoptotic processes (Fig. 5B). Similarly, the enriched KEGG pathways encompassed various pathways, notably including Focal adhesion, Endometrial cancer, and Cellular senescence (Fig. 5C). To investigate the potential direct binding of CPSF4 to DEGs as RBP, and its subsequent impact on their expression, we conducted an integrated analysis of RNA-seq and RIP-seq data. We found that among the 331 DEGs, only 3 genes tended to directly interact with CPSF4 (Fig. 5D). The majority of DEGs did not exhibit regulation via direct binding to CPSF4.

CPSF4-mediated ASEs in TNBC

To delineate the involvement of CPSF4 in the regulation of AS, we leveraged uniquely mapped read counts obtained from transcriptome sequencing data. This enabled us to investigate AS events that are dependent on CPSF4. In addition to the known ASEs annotated in the reference files, we discovered a multitude of novel AS events. These events were classified into ten distinct AS types. Among the prominent AS types were ES, A5SS, and A3SS. (Fig. 6A). Subsequently, we extracted significantly differentially expressed RASEs between siCPSF4 and control samples, detecting 1983 RASEs upon CPSF4 silencing. These high-confidence RASEs were further classified, with the majority comprising ES (280 events), A5SS (396 events), and A3SS (509 events) (Fig. 6B). These findings indicate that CPSF4 comprehensively regulates ASEs in TNBC. By means of pathway mapping and statistical analysis, we identified a total of 1380 genes that contribute to these regulated

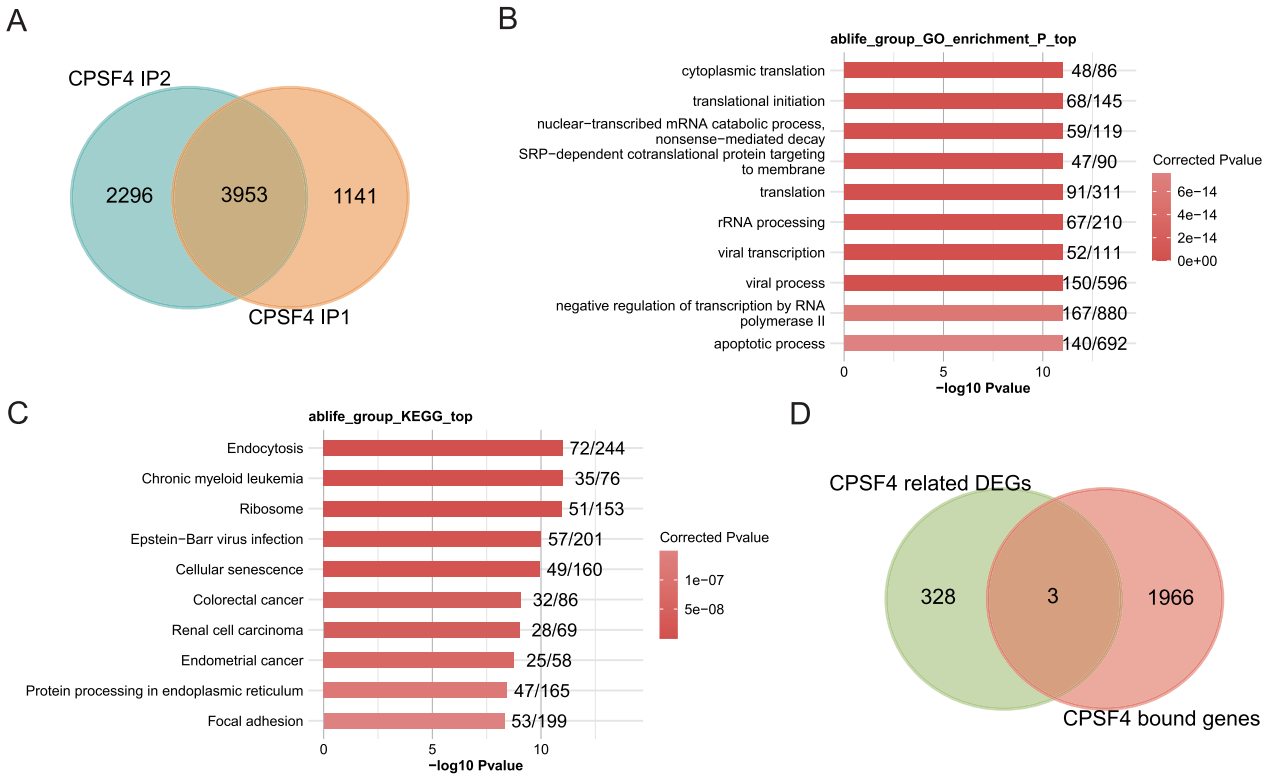


Fig. 5 DEGs Regulated by CPSF4 Not Directly Bound to CPSF4. **A** Venn diagram showing the overlap between peaks in two replicate sequences of RIP-seq samples. **B** Gene ontology (GO) enrichment analysis showing the top 10 biological processes associated with genes bound to CPSF4. **C** KEGG enrichment analysis displaying the top 10 pathways associated with genes bound to CPSF4. **D** Venn diagram illustrating the overlap between RIP-seq and DEGs

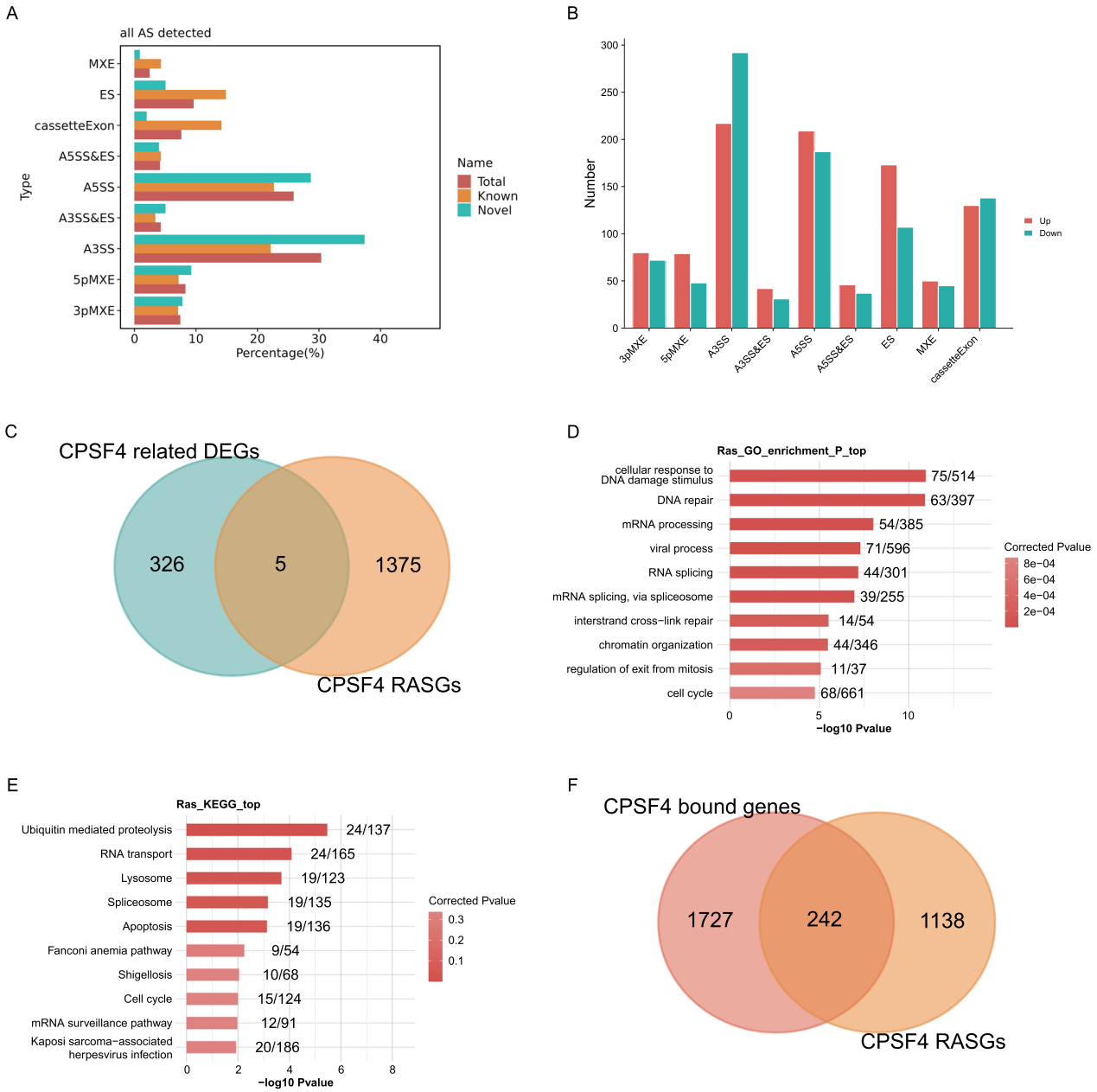


Fig. 6 CPSF4 Modulates Alternative Splicing Events (ASEs) in MDA-MB-231 cells. **A** Classification of all detected ASEs. **B** Classification of ASEs regulated by CPSF4. **C** Overlap between CPSF4-regulated DEGs and regulated alternative splicing genes (RASGs). **D** Top 10 biological processes associated with CPSF4-regulated alternative splicing genes as determined by GO enrichment analysis. **E** Top 10 pathways associated with CPSF4-regulated alternative splicing genes identified through KEGG analysis. **F** Venn diagram depicting the overlap between binding genes and RASGs

RASEs, hence termed CPSF4-regulated RASGs. While distinct alternative splicing genes were observed between si-CPSF4 and control samples, only transcripts of 5 RASGs were significantly regulated (Fig. 6C). Thus, changes in ASEs cannot solely be ascribed to transcriptional upregulation or downregulation. Further analysis was conducted to explore the potential functions of the CPSF4-regulated RASGs. GO term analysis, specifically focusing on biological processes, indicated a significant enrichment of CPSF4-regulated genes in mRNA processing, RNA splicing, mRNA splicing via spliceosome and cell cycle processes(Fig. 6D). In a similar vein, the enriched KEGG pathways were linked to RNA transport, spliceosome, mRNA surveillance pathway

and apoptosis (Fig. 6E). We conducted an extensive examination of genes that overlap between the binding genes and RASG. Among the variable splicing genes regulated by CPSF4, 242 of them were found to intersect with CPSF4-binding genes (Fig. 6F). In summary, CPSF4 influences the expression of genes in TNBC through AS regulation.

Identification and validation of CPSF4-mediated ASE

To identify potential regulators binding to mRNA based on RIP-seq data and RASGs, we pinpointed all peaks that were common to both bindings. Signals were identified by descending peak height, revealing coiled-coil-helix-coiled-coil-helix domain containing 3 (CHCHD3), epsin 1 (EPN1), HMG20B, high mobility group nucleosomal binding domain 2 (HMGN2), myeloid leukemia factor 2 (MLF2), NMDA receptor synaptonuclear signaling and neuronal migration factor (NSMF), 5'-nucleotidase, cytosolic IIIB (NT5C3B), platelet activating factor acetylhydrolase 1b catalytic subunit 3 (PAFAH1B3), pleckstrin homology domain containing J1 (PLEKHJ1), and ubiquitin conjugating enzyme E2 A (UBE2A). Splicing ratio changes of 10 CPSF4-bound RASGs were analyzed in RNAseq data. EPN1, HMG20B, and PAFAH1B3 exhibited A3SS, CHCHD3, HMGN2, and NT5C3B showed A5SS, MLF2, NSMF, and PLEKHJ1 presented A5SS, and A3SS&ES for A3SS&ES (Fig. 7A). We examined the dependency of these 10 ASEs on CPSF4 in the sequencing samples of MDA-MB-231 cells. The results showed that 10 ASEs underwent differential splicing between the high and low CPSF4 groups, with the AS ratios for CHCHD3, NSMF, PAFAH1B3 and UBE2A decreasing, while the AS ratios for EPN1, HMG20B, HMGN2, MLF2, NT5C3B and PLEKHJ1 increased (Fig. 7B). These findings are consistent with our RNA-seq data results. We validated these 10 ASEs, obtaining results consistent with those from our RNA-seq data (Fig. 7B). Overall, our findings demonstrate that

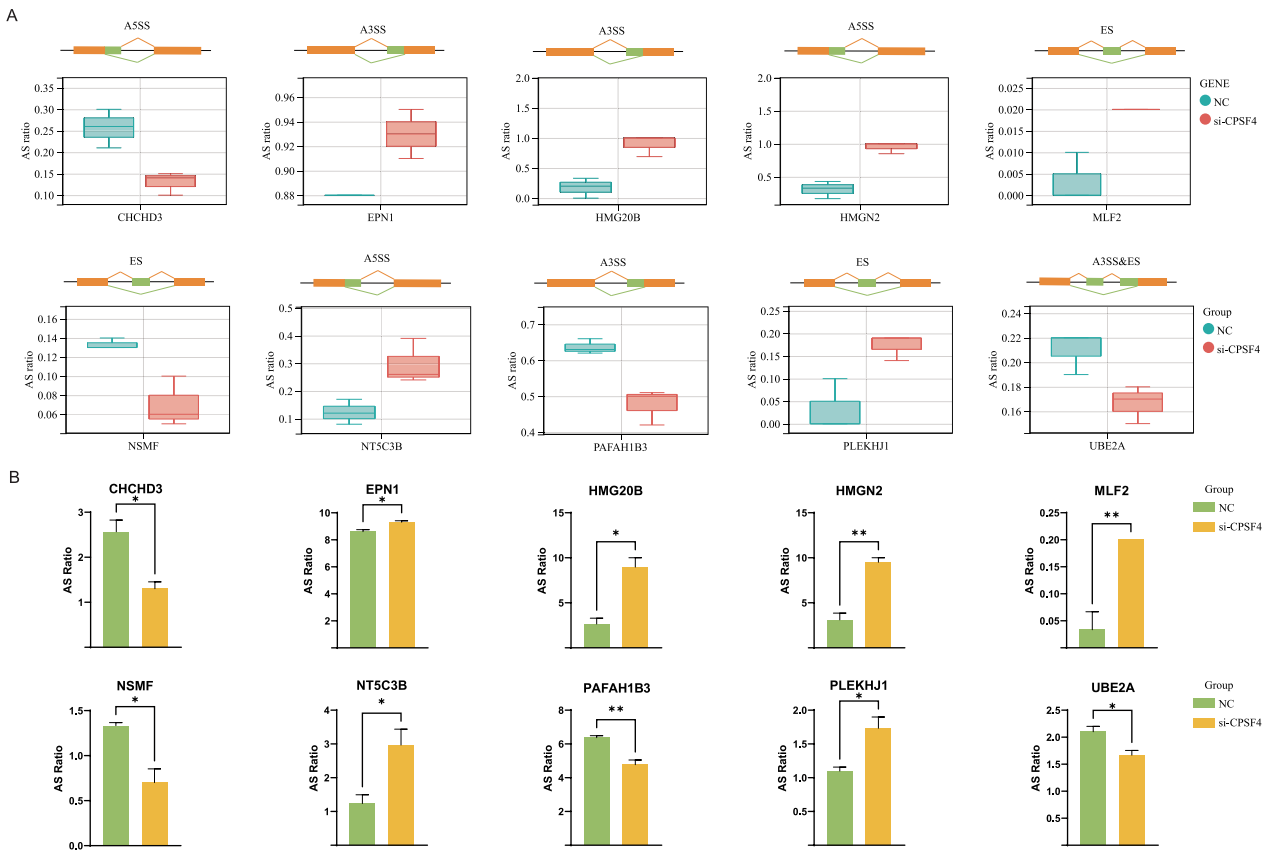


Fig. 7 Identification and validation of ASEs mediated by CPSF4 in TNBC cells. **A** Proportional changes of ASEs for coiled-coil-helix-coiled-coil-helix domain containing 3 (CHCHD3), epsin 1 (EPN1), high mobility group 20B (HMG20B), high mobility group nucleosomal binding domain 2 (HMGN2), myeloid leukemia factor 2 (MLF2), NMDA receptor synaptonuclear signaling and neuronal migration factor (NSMF), 5'-nucleotidase, cytosolic IIIB (NT5C3B), platelet activating factor acetylhydrolase 1b catalytic subunit 3 (PAFAH1B3), pleckstrin homology domain containing J1 (PLEKHJ1) and ubiquitin conjugating enzyme E2 A (UBE2A) in RNA-seq data. **B** Validation of ASEs for CHCHD3, EPN1, HMG20B, HMGN2, MLF2, NSMF, NT5C3B, PAFAH1B3, PLEKHJ1, and UBE2A in MDA-MB-231 cells after CPSF4 knockdown using qRT-PCR. *$P < 0.05$, **$P < 0.01$

CPSF4 exerts regulatory control over a wide range of ASEs through its functional binding to RNA in TNBC.

CPSF4 regulates HMG20B AS and further promotes TNBC progression

To explore the function of HMG20B in TNBC, we utilized siRNA to downregulate HMG20B (Fig. 8A). As anticipated, CCK-8 assays revealed significant suppression of MDA-MB-231 cell viability upon HMG20B knockdown (Fig. 8B). Concurrently, transwell assays indicated a significant decrease in the migratory and invasive abilities of cells in the HMG20B knockdown group compared to the control group (Fig. 8C–E). In addition, analysis of the read distribution of the RIP-seq data for HMG20B showed that there were multiple binding sites for CPSF4 in the intronic region of HMG20B (Fig. 8F). Therefore, these results suggest that CPSF4 promotes the AS alternative splicing of HMG20B.

Discussion

TNBC remains a challenging malignancy lacking validated therapeutic targets, leaving conventional treatments such as surgery, radiotherapy, and chemotherapy as the mainstay of therapy [22]. Hence, novel therapeutic targets and strategies must be developed for its management. Recent advancements in high-throughput sequencing technologies and computational tools have led to a growing focus on the analysis of tumor-associated alternative splicing (AS) events. Li et al. demonstrated that the splicing factor SNRPB contributes to the progression of ovarian cancer by regulating the abnormal skipping of exons in POLA1 and BRCA2 [36]. Xu et al. revealed that PUF60 promotes both cell cycle advancement and the progression of lung cancer by modulating the alternative splicing of CDC25C [37]. Similarly, Xie et al. revealed the inhibitory effect of NONO on bladder cancer lymphatic metastasis through selective splicing of SETMAR [38]. However, the significance of AS genes in TNBC remains elusive, particularly in the context of TNBC immunotherapy. Through RNA-seq analysis, we detected a multitude of ASEs controlled by CPSF4, suggesting its implication in pivotal pathways linked to the pathogenesis of TNBC.

CPSF4 was widely recognized for its role in pre-mRNA 3′-end processing [39]. Clinical studies have linked CPSF4 upregulation with poor prognosis in hepatocellular carcinoma [40]. In lung cancer, CPSF4 promotes tumor initiation phenotypes via the VEGF/NRP2/TAZ signaling pathway, serving as a prognostic marker and therapeutic target [20]. Additionally, CPSF4 transcriptionally regulates MDM4 to promote metastasis in TNBC [21]. However, previous investigations into CPSF4 in TNBC have primarily focused on its transcriptional regulation. Further exploration of CPSF4's role in regulating AS genes and its impact on TNBC malignancy, prognosis, and the tumor microenvironment is warranted. In our investigation, utilizing RIP-seq and RNA-seq datasets, we compared expression profiles between control and CPSF4-knockdown cells. Our results revealed CPSF4's capability to influence the AS patterns of genes involved in diverse pathways. These findings emphasize the substantial regulatory function of CPSF4 in AS patterns. Therefore, CPSF4 might impact the progression of TNBC and could represent a promising therapeutic target in future interventions.

The RIP-seq technique provides a systematic method for identifying RNA transcripts that are bound by RBPs [41]. In this investigation, RIP-seq analysis unveiled the RNA-binding motif of CPSF4 to be GAAGAAG. DEG and AS analyses revealed that only five genes showed both differential expression and alternative splicing. This suggests that CPSF4 governs separate gene sets through both transcriptional and post-transcriptional mechanisms. Previous reports have linked splicing events to increased cancer cell proliferation and invasion [42, 43]. Our findings revealed that genes undergoing AS are enriched in pathways strongly linked to tumorigenesis. These pathways include mRNA processing, cell cycle regulation, RNA transport, the mRNA surveillance pathway, and apoptosis. Therefore, focusing on specific ASEs linked to CPSF4 could serve as a viable approach to decrease TNBC cell proliferation and prevent metastasis.

This study identified the top 10 AS genes associated with CPSF4 (CHCHD3, EPN1, HMG20B, HMGN2, MLF2, NSMF, NT5C3B, PAFAH1B3, PLEKHJ1, and UBE2A). We validated the AS ratios of these cancer-related genes regulated by CPSF4 through qRT-PCR, consistent with the CPSF4-RASEs identified by RNA-seq. Among these genes, CHCHD3 is crucial for maintaining mitochondrial integrity and function [44]. HMGN2, a nuclear protein that binds to DNA, enhances the stability of nucleosome assembly and enhances gene transcription. It exerts inhibitory effects on the growth and migration of osteosarcoma cells [45]. MLF2, a member of the Myeloid Leukemia Factor family, serves as an oncogene in breast cancer and chronic myeloid leukemia [46, 47]. PAFAH1B3, one of the catalytic subunits of PAFAH, is essential for the growth, metastasis, angiogenesis, and development of drug resistance in cancer [48]. HMG20B, belonging to the high mobility group (HMG) family, is a non-sequence-specific DNA-binding protein. It interacts with the tumor suppressor protein BRCA2 DNA repair associated (BRCA2) and modulates cell division during cytokinesis [49]. In our investigation, the depletion of HMG20B led to notable reductions in cell proliferation, migration, and invasion. These results suggest that

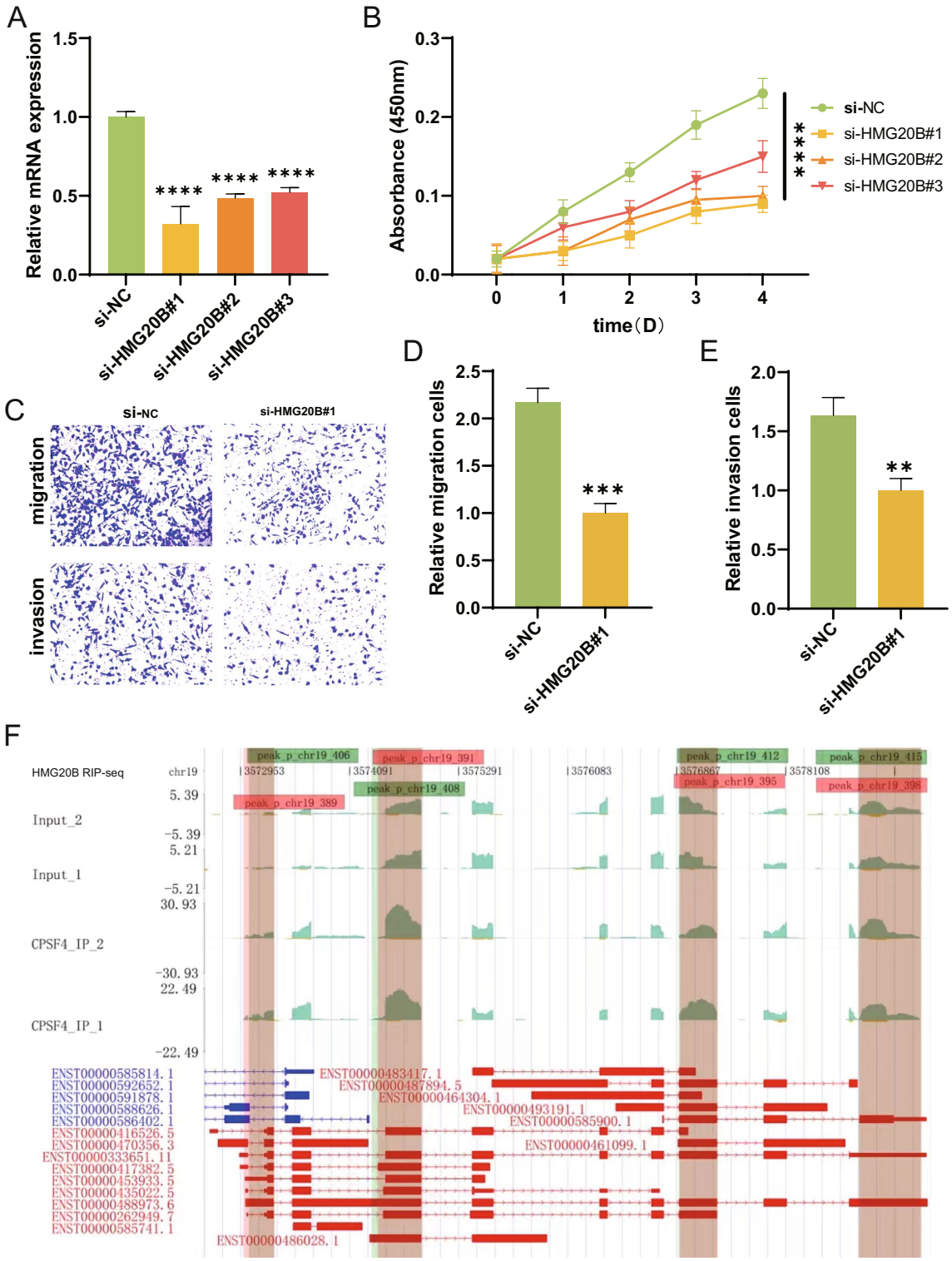


Fig. 8 Suppression of cell proliferation, migration, and invasion by HMG20B knockdown. **A** Relative expression levels of HMG20B in cells after HMG20B knockdown detected by qRT-PCR. **B** Viability of MDA-MB-231 cells transfected with siRNA negative control (si-NC) or siRNA targeting HMG20B (si-HMG20B#1/#2/#3) assessed by Cell Counting Kit-8 (CCK-8) assay. **C**–**E** Migration and invasion of MDA-MB-231 cells transfected with siRNA-NC or siRNA-HMG20B#1 detected by Transwell assay. **F** IGV-sashimi plots showed HMG20B peak reads distribution in RIP-seq. ****$P < 0.0001$

CPSF4 might facilitate the progression of TNBC by modulating the alternative splicing of HMG20B. However, the precise mechanisms still require further elucidation.

This study has several limitations. Although we provided RNA expression data for CPSF4 in tumor cells, the protein expression of CPSF4 has not been validated in clinical samples. Additionally, while the CPSF4-mediated transcriptional differences, CPSF4-binding RIP-seq analysis, and alternative splicing analysis were all derived from sequencing data, we plan to perform experimental validation in future studies. Furthermore, although we have validated the impact of HMG20B on the properties of cancer cells, additional experimental studies are needed to investigate the interaction between CPSF4 and HMG20B, as well as their influence on cancer cell characteristics. While we have clarified the regulatory function of CPSF4 in the alternative splicing of HMG20B, further inhibition experiments targeting specific splice variants of HMG20B are warranted. Moreover, additional studies are required to explore the underlying mechanisms concerning the regulatory role of CPSF4 in various alternative splicing events, including those involving CPSF4 and HMG20B.

Conclusion

In conclusion, our investigation into the interplay of CPSF4 with the transcriptome and alternative splicing in MDA-MB-231 cells, utilizing RIP-seq and RNA-seq, has unveiled novel findings. Particularly noteworthy is our discovery that CPSF4 modulates HMG20B alternative splicing, thereby fostering TNBC proliferation. These findings not only deepen our understanding of TNBC mechanisms but also present avenues for the identification of novel diagnostic markers and therapeutic targets.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-024-06004-x.

Supplementary material 1: Figure 1 Assay of transfection efficiency. (A) Evaluation of knockdown efficiency of small interfering RNA (siRNA) targeting CPSF4 (si-CPSF4) in MDA-MB-231 and MDA-MB-468 cells by qRT-PCR. (B) Evaluation of knockdown efficiency of si-CPSF4 in MDA-MB-468 cells by qRT-PCR. (C) Evaluation of overexpression efficiency of overexpression of CPSF4 (OE-CPSF4) in MDA-MB-231 cells by qRT-PCR. (D) Evaluation of overexpression efficiency of OE-CPSF4 in MDA-MB-468 cells cells by qRT-PCR. (E) Validation of si-CPSF4 knockdown efficiency in MDA-MB-231 cells by Western blot. (F) Validation of si-CPSF4 knockdown efficiency in MDA-MB-468 cells by western blot. (G) Validation of OE-CPSF4 overexpression efficiency in MDA-MB-231 cells by western blot. (H) Evaluation of overexpression efficiency of in MDA-MB-468 cells by Western blot.**$P < 0.01$, ***$P < 0.001$,****$P < 0.0001$.

Acknowledgements
We thank Ruixing Biotechnology Co., Ltd (Wuhan) for the RNA-seq and RIP-seq service, and the subsequent bioinformatics analysis.

Author contributions
GZ and WL designed the study. GZ, ZD and WL performed the data processing and analysis. QS, KY and HJL conducted several in vitro experiments. GZ, QS, and WL wrote the manuscript. GZ, QS, XZ, KY, HL, BW, HC, ZD and WL revised the manuscript. All authors read and approved the final manuscript.

Funding
This work was financially supported by the Grants LHDMZ24H300001 from Zhejiang Provincial Natural Science Foundation of China.

Availability of data and materials
The data analyzed was available on the request for the corresponding author.

Declarations

Competing interests
The authors declare there is no financial interest.

Author details
[1]Department of Breast Surgery, College of Medicine, The First Affiliated Hospital, Zhejiang University, Hangzhou 310000, Zhejiang, China. [2]Department of Surgical Oncology, Affiliated Dongyang Hospital of Wenzhou Medical University, Dongyang 322100, Zhejiang, China. [3]Department of Eye, Ear, Nose and Throat, The 903 Hospital of the Joint Logistics Support Force of the Chinese People's Liberation Army, Hangzhou 310000, Zhejiang, China. [4]Department of Head, Neck & Thyroid Surgery, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, Hangzhou 310000, Zhejiang, China. [5]Department of Radiology, The 903 Hospital of the Joint Logistics Support Force of the Chinese People's Liberation Army, Hangzhou 310000, Zhejiang, China.

Received: 19 August 2024 Accepted: 17 December 2024
Published online: 27 December 2024

References

1. Anders CK, Carey LA, Burstein HJ. ER/PR negative, HER2-negative (triple-negative) breast cancer. UpToDate: UpToDate. 2021.
2. Kumar P, Aggarwal R. An overview of triple-negative breast cancer. Arch Gynecol Obstet. 2016;293:247–69.
3. Bai X, Ni J, Beretov J, Graham P, Li Y. Triple-negative breast cancer therapeutic resistance: where is the Achilles' heel? Cancer Lett. 2021;497:100–11.
4. Lee A, Djamgoz MB. Triple negative breast cancer: emerging therapeutic modalities and novel combination therapies. Cancer Treat Rev. 2018;62:110–22.
5. Mai N, Abuhadra N, Jhaveri K. Molecularly targeted therapies for triple negative breast cancer: history, advances, and future directions. Clin Breast Cancer. 2023. https://doi.org/10.1016/j.clbc.2023.05.012.
6. Hong R, Xu B. Breast cancer: an up-to-date review and future perspectives. Cancer Commun. 2022;42(10):913–36.
7. Shao Z, Chaudhri S, Guo M, Zhang L, Rea D. Neoadjuvant chemotherapy in triple negative breast cancer: an observational study. Oncol Res. 2016;23(6):291.
8. Gamucci T, Pizzuti L, Sperduti I, Mentuccia L, Vaccaro A, Moscetti L, et al. Neoadjuvant chemotherapy in triple-negative breast cancer: a multicentric retrospective observational study in real-life setting. J Cell Physiol. 2018;233(3):2313–23.
9. Wang Y, Liu J, Huang B, Xu YM, Li J, Huang LF, et al. Mechanism of alternative splicing and its regulation. Biomed Rep. 2015;3(2):152–8.
10. de Klerk E, AC't Hoen P. Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. Trends Genet. 2015;31(3):128–39.
11. Ramani AK, Calarco JA, Pan Q, Mavandadi S, Wang Y, Nelson AC, et al. Genome-wide analysis of alternative splicing in Caenorhabditis elegans. Genome Res. 2011;21(2):342–8.
12. Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human tissues. Genome Biol. 2004;5:1–15.

13. Gallego-Paez LM, Bordone MC, Leote AC, Saraiva-Agostinho N, Ascensão-Ferreira M, Barbosa-Morais N. Alternative splicing: the pledge, the turn, and the prestige: the key role of alternative splicing in human biological systems. Hum Genet. 2017;136:1015–42.
14. Furlanis E, Scheiffele P. Regulation of neuronal differentiation, function, and plasticity by alternative splicing. Annu Rev Cell Dev Biol. 2018;34:451–69.
15. Tao Y, Zhang Q, Wang H, Yang X, Mu H. Alternative splicing and related RNA binding proteins in human health and disease. Signal Transduct Target Ther. 2024;9(1):26.
16. Liu Y, Li Y, Du C, Kuang S, Zhou X, Zhang J, et al. Underlying mechanisms of epithelial splicing regulatory proteins in cancer progression. J Mol Med. 2022;100(11):1539–56.
17. Biamonti G, Infantino L, Gaglio D, Amato A. An intricate connection between alternative splicing and phenotypic plasticity in development and cancer. Cells. 2019;9(1):34.
18. Du J-X, Luo Y-H, Zhang S-J, Wang B, Chen C, Zhu G-Q, et al. Splicing factor SRSF1 promotes breast cancer progression via oncogenic splice switching of PTPMT1. J Exp Clin Cancer Res. 2021;40(1):171.
19. Yuemaierabola A, Guo J, Sun L, Yeerkenbieke B, Liu F, Ye D, et al. Comprehensive analysis of CPSF4-related alternative splice genes in hepatocellular carcinoma. J Cancer Res Clin Oncol. 2023;149(15):13955–71.
20. Song Y, Sun K, Gong L, Shi L, Qin T, Wang S, et al. CPSF4 promotes tumor-initiating phenotype by enhancing VEGF/NRP2/TAZ signaling in lung cancer. Med Oncol. 2022;40(1):62.
21. Lee K, Zheng Q, Lu Q, Xu F, Qin G, Zhai Q, et al. CPSF4 promotes triple negative breast cancer metastasis by upregulating MDM4. Signal Transduct Target Ther. 2021;6(1):184.
22. Dubois J, Traversier A, Julien T, Padey B, Lina B, Bourdon J-C, et al. The nonstructural NS1 protein of influenza viruses modulates TP53 splicing through host factor CPSF4. J Virol. 2019. https://doi.org/10.1128/jvi.02168-18.
23. Maiques-Diaz A, Nicosia L, Basma NJ, Romero-Camarero I, Camera F, Spencer GJ, et al. HMG20B stabilizes association of LSD1 with GFI1 on chromatin to confer transcription repression and leukemia cell differentiation block. Oncogene. 2022;41(44):4841–54.
24. Herchenröther A, Gossen S, Friedrich T, Reim A, Daus N, Diegmüller F, et al. The H2A.Z and NuRD associated protein HMG20A controls early head and heart developmental transcription programs. Nat Commun. 2023;14(1):472.
25. Aldughaim MMAMS, Al-yhya N, Wadaan MA, Al-ghamdi S, Habibullah MM, Alothaid H, Khan MF. The Tim8-Tim13 complex has multiple substrate binding sites and binds cooperatively to Tim23. J Mol Biol. 2022;34(2):101812.
26. Lee M, Venkitaraman AR. A cancer-associated mutation inactivates a region of the high-mobility group protein HMG20b essential for cytokinesis. Cell Cycle. 2014;13(16):2554–63.
27. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BV, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia. 2017;19(8):649–58.
28. Jia B, Zhao X, Wang Y, Wang J, Wang Y, Yang Y. Prognostic roles of MAGE family members in breast cancer based on KM-Plotter Data. Oncol Lett. 2019;18(4):3501–16.
29. Bustin SA, Mueller R. Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. Clin Sci. 2005;109(4):365–79.
30. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:1–13.
31. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
32. Jin L, Li G, Yu D, Huang W, Cheng C, Liao S, et al. Transcriptome analysis reveals the complexity of alternative splicing regulation in the fungus Verticillium dahliae. BMC Genomics. 2017;18:1–14.
33. Xia H, Chen D, Wu Q, Wu G, Zhou Y, Zhang Y, et al. 2017 CELF1 preferentially binds to exon-intron boundary and regulates alternative splicing in HeLa cells. Biochim et Biophys Acta (BBA) Gene Regul Mech. 1860;9:911–21.
34. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 20: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011;39:W316–22.
35. Song Q, Yi F, Zhang Y, Jun Li DK, Wei Y, Yu H, et al. CRKL regulates alternative splicing of cancer-related genes in cervical cancer samples and HeLa cell. BMC Cancer. 2019;19(1):499.
36. Li Y, Chen Z, Peng J, Yuan C, Yan S, Yang N, et al. The splicing factor SNRPB promotes ovarian cancer progression through regulating aberrant exon skipping of POLA1 and BRCA2. Oncogene. 2023;42(31):2386–401.
37. Xu N, Ren Y, Bao Y, Shen X, Kang J, Wang N, et al. PUF60 promotes cell cycle and lung cancer progression by regulating alternative splicing of CDC25C. Cell Rep. 2023;42(9):113041.
38. Xie R, Chen X, Cheng L, Huang M, Zhou Q, Zhang J, et al. NONO inhibits lymphatic metastasis of bladder cancer via alternative splicing of SETMAR. Mol Ther. 2021;29(1):291–307.
39. Misra A, Green MR. From polyadenylation to splicing: dual role for mRNA 3'end formation factors. RNA Biol. 2016;13(3):259–64.
40. Wang X, Dong J, Li X, Cheng Z, Zhu Q. CPSF4 regulates circRNA formation and microRNA mediated gene silencing in hepatocellular carcinoma. Oncogene. 2021;40(25):4338–51.
41. Nicholson CO, Friedersdorf MB, Bisogno LS, Keene JD. DO-RIP-seq to quantify RNA binding sites transcriptome-wide. Methods. 2017;118:16–23.
42. Wang Y, Chen D, Qian H, Tsai YS, Shao S, Liu Q, et al. The splicing factor RBM4 controls apoptosis, proliferation, and migration to suppress tumor progression. Cancer Cell. 2014;26(3):374–89.
43. Tien JF, Mazloomian A, Cheng S-WG, Hughes CS, Chow CC, Canapi LT, et al. CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion. Nucleic Acids Res. 2017;45(11):6698–716.
44. Darshi M, Mendiola VL, Mackey MR, Murphy AN, Koller A, Perkins GA, et al. ChChd3, an inner mitochondrial membrane protein, is essential for maintaining crista integrity and mitochondrial function. J Biol Chem. 2011;286(4):2918–32.
45. Liang G, Xu E, Yang C, Zhang C, Sheng X, Zhou X. Nucleosome-binding protein HMGN2 exhibits antitumor activity in human SaO2 and U2-OS osteosarcoma cell lines. Oncol Rep. 2015;33(3):1300–6.
46. Dave B, Granados-Principal S, Zhu R, Benz S, Rabizadeh S, Soon-Shiong P, et al. Targeting RPL39 and MLF2 reduces tumor initiation and metastasis in breast cancer by inhibiting nitric oxide synthase signaling. Proc Natl Acad Sci. 2014;111(24):8838–43.
47. Yang J, Cao D, Zhang Y, Ou R, Yin Z, Liu Y, et al. The role of phosphorylation of MLF2 at serine 24 in BCR-ABL leukemogenesis. Cancer Gene Ther. 2020;27(1):98–107.
48. Yuan Y, Jiang X, Tang L, Wang J, Duan L. Comprehensive analysis of the prognostic and immunological role of PAFAH1B in pan-cancer. Front Mol Biosci. 2021;8:799497.
49. Lee M, Daniels M, Garnett M, Venkitaraman A. A mitotic function for the high-mobility group protein HMG20b regulated by its interaction with the BRC repeats of the BRCA2 tumor suppressor. Oncogene. 2011;30(30):3360–9.

Publisher's Note