

Publication News Updates

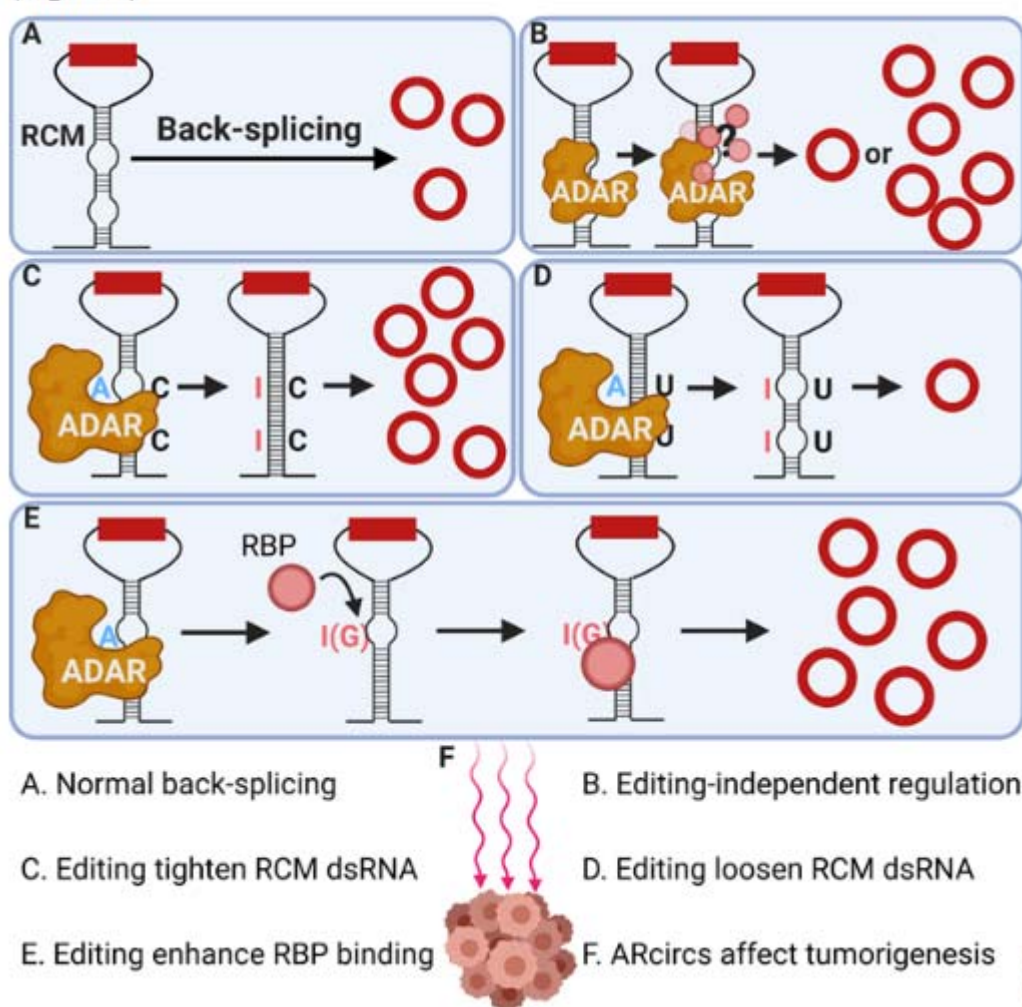
Source: A/Prof CHEN Leilei, Polly

Email: antel@nus.edu.sg

1. [Shen H, An Ö, Ren X, Song Y, Tang SJ, Ke X, Han J, Tay DJ, Ng VHE, Bellido Molias F, Pitcheshwar P, Leong KW, Tan KK, Yang H, Chen L. ADARs act as potent regulators of circular transcriptome in cancer. *Nature Communications*, 13\(1\):16 pages Number ARTN 1508 21 Mar 2022.](#) (You can use **Figure 1** shown below)

A short description: Previous studies have shown that aberrantly expressed circRNAs contribute to many diseases, including cancers, but precise mechanisms underlying the regulation of circRNA biogenesis in cancer cells remain unknown. In a novel step forward, Polly Chen and her team established the bidirectional regulatory role of Adenosine deaminases acting on RNA (ADARs) on circRNA. They found that ADARs-regulated circRNAs are ubiquitous in multiple cancers, and play critical roles in cancer development.

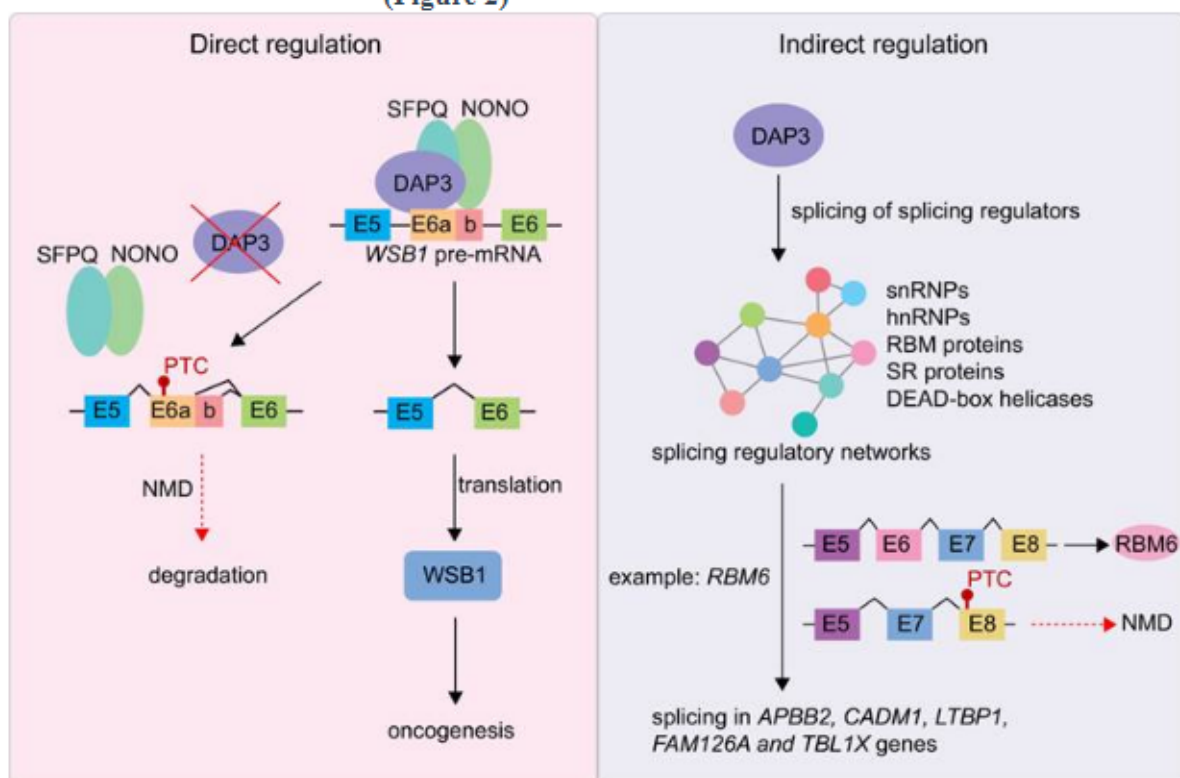
(Figure 1)



2. [Han J, An Ö, Ren X, Song Y, Tang SJ, Shen H, Ke X, Ng VHE, Tay DJ, Tan HQ, Kappei D, Yang H, Chen L. Multilayered Control of Splicing Regulatory Networks by DAP3 Leads to Widespread Alternative Splicing Changes in Cancer. *Nature Communications*, 13\(1\):1793 04 Apr 2022.](#) (You can use **Figure 2** shown below)

A short description: A recent study helmed by Polly Chen has yielded important insights into the splicing regulatory roles of a frequently overexpressed cancer-associated protein, DAP3, in cancer development. Results revealed that DAP3 coordinates splicing regulatory networks to modulate global alternative splicing in cancer via both RNA-protein and protein-protein interactions. With these findings, the team established that targeting DAP3-driven splicing events and blocking the splicing regulatory ability of DAP3 and/or specifically targeting DAP3-driven splicing events may hold great promise for cancer treatment.

(Figure 2)



ARTICLE



<https://doi.org/10.1038/s41467-022-29138-2>

OPEN

ADARs act as potent regulators of circular transcriptome in cancer

Haoqing Shen^{1,2}, Omer An¹, Xi Ren¹, Yangyang Song¹, Sze Jing Tang¹, Xin-Yu Ke^{1,2}, Jian Han¹, Daryl Jin Tai Tay¹, Vanessa Hui En Ng¹, Fernando Bellido Molias¹, Priyanka Pitcheshwar^{1,2}, Ka Wai Leong¹, Ker-Kan Tan^{3,4,5}, Henry Yang¹ & Leilei Chen^{1,2,✉}

Circular RNAs (circRNAs) are produced by head-to-tail back-splicing which is mainly facilitated by base-pairing of reverse complementary matches (RCMs) in circRNA flanking introns. Adenosine deaminases acting on RNA (ADARs) are known to bind double-stranded RNAs for adenosine to inosine (A-to-I) RNA editing. Here we characterize ADARs as potent regulators of circular transcriptome by identifying over a thousand of circRNAs regulated by ADARs in a bidirectional manner through and beyond their editing function. We find that editing can stabilize or destabilize secondary structures formed between RCMs via correcting A:C mismatches to I(G)-C pairs or creating I(G).U wobble pairs, respectively. We provide experimental evidence that editing also favors the binding of RNA-binding proteins such as PTBP1 to regulate back-splicing. These ADARs-regulated circRNAs which are ubiquitously expressed in multiple types of cancers, demonstrate high functional relevance to cancer. Our findings support a hitherto unappreciated bidirectional regulation of circular transcriptome by ADARs and highlight the complexity of cross-talk in RNA processing and its contributions to tumorigenesis.

¹Cancer Science Institute of Singapore, National University of Singapore, Singapore, Singapore. ²Department of Anatomy, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ³Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ⁴Division of Colorectal Surgery, University Surgical Cluster, National University Health System, Singapore, Singapore. ⁵NUS Centre for Cancer Research, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ✉email: polly_chen@nus.edu.sg

Unlike canonical linear RNAs, circular RNAs (circRNAs) are a type of RNA molecules with a covalently closed continuous loop structure. Since the circular form of RNA in the cytoplasm fraction of eukaryotic cells was first observed using electron microscope in 1979¹, circRNAs have been identified in different eukaryotes including plants, fungi, mice, and humans². However, in the following decades, because of their naturalness of low abundance and non-coding feature, the vast majority of circRNAs remained neglected. Only with recent advances in high-throughput sequencing, circRNAs have been characterized as ubiquitously expressed, biologically conserved and tissue-specific RNA molecules³. Diverse functions of circRNAs include competing with linear splicing, sponging microRNA (miRNA), interacting with RNA-binding proteins (RBPs), and producing small peptides³. Importantly, aberrantly expressed circRNAs have been found in many diseases such as neurological diseases, cardiovascular diseases, and cancers⁴. Since circRNAs are ubiquitous and functional, it is worth further investigation of precise mechanisms underlying the regulation of circRNA biogenesis in cells.

CircRNAs are generated by “back-splicing”, which is splicing between a downstream 5′ splice donor and an upstream 3′ splice acceptor⁵. This process requires spatial proximity of non-sequential splice sites, which is usually facilitated by RBPs which bind to flanking introns^{5,6} and/or base-pairing formed by reverse complementary matches (RCMs) in flanking introns such as inverted repeat Alu elements (IRAlus)^{7,8}. RBPs can also facilitate or disrupt the intra-intronic base-pairing. DEXH-box helicase 9 (DHH9) negatively regulates circRNA biogenesis by binding to and unwinding the base-paired IRAlus in flanking introns⁹. On the contrary, nuclear factor 90 (NF90) and its 110 kDa isoform NF110 bind to the base-pairs formed by flanking introns, leading to an increased circRNA production¹⁰. However, the role of other RBPs (particularly dsRNA-binding proteins) in regulating circRNA biogenesis remains largely unexplored.

Adenosine deaminases acting on RNA (ADARs) protein family, known to preferentially bind to dsRNAs formed by IRAlu elements¹¹, holds great potential as a potent circRNA regulator. Upon dsRNA binding, ADARs may catalyze adenosine to inosine (A-to-I) editing, the most prevalent type of RNA editing in eukaryotes¹², in their bound dsRNAs. Till now, a few studies have reported controversial findings about the effect of ADARs on circRNA biogenesis. It has been suggested that ADARs could suppress the generation of circRNAs by editing and “melting” the dsRNA^{8,13}. However, another study claimed that ADARs alone had no major effect on circRNA biogenesis, although double knockdown of ADAR1 and DHH9 repressed circRNA biogenesis to a greater extent than the single knockdown of each gene⁹. Besides, there is still a lack of experimental evidence supporting that ADARs-mediated editing can destabilize (and unwind) dsRNAs formed by IRAlu elements. The role of ADARs in circRNA biogenesis warrants a deeper investigation from the facts that: (1) ADARs preferentially edit A:C mismatches rather than A-U base pairs¹⁴, presumably resulting in a more stable secondary structure; (2) through editing, ADARs can strengthen or weaken binding of RBPs by altering RNA sequences of *cis*-elements and/or creating or destroying RBP binding motifs^{15–17}; and (3) independent of their editing function, ADARs can also block the access of other RBPs (e.g. U2AF65) to the latter’s original binding sites, contributing to changes in splicing¹⁵. These abovementioned facts challenge the common opinion that ADARs function as repressor of circRNA biogenesis, presumably dependent on their editing capability.

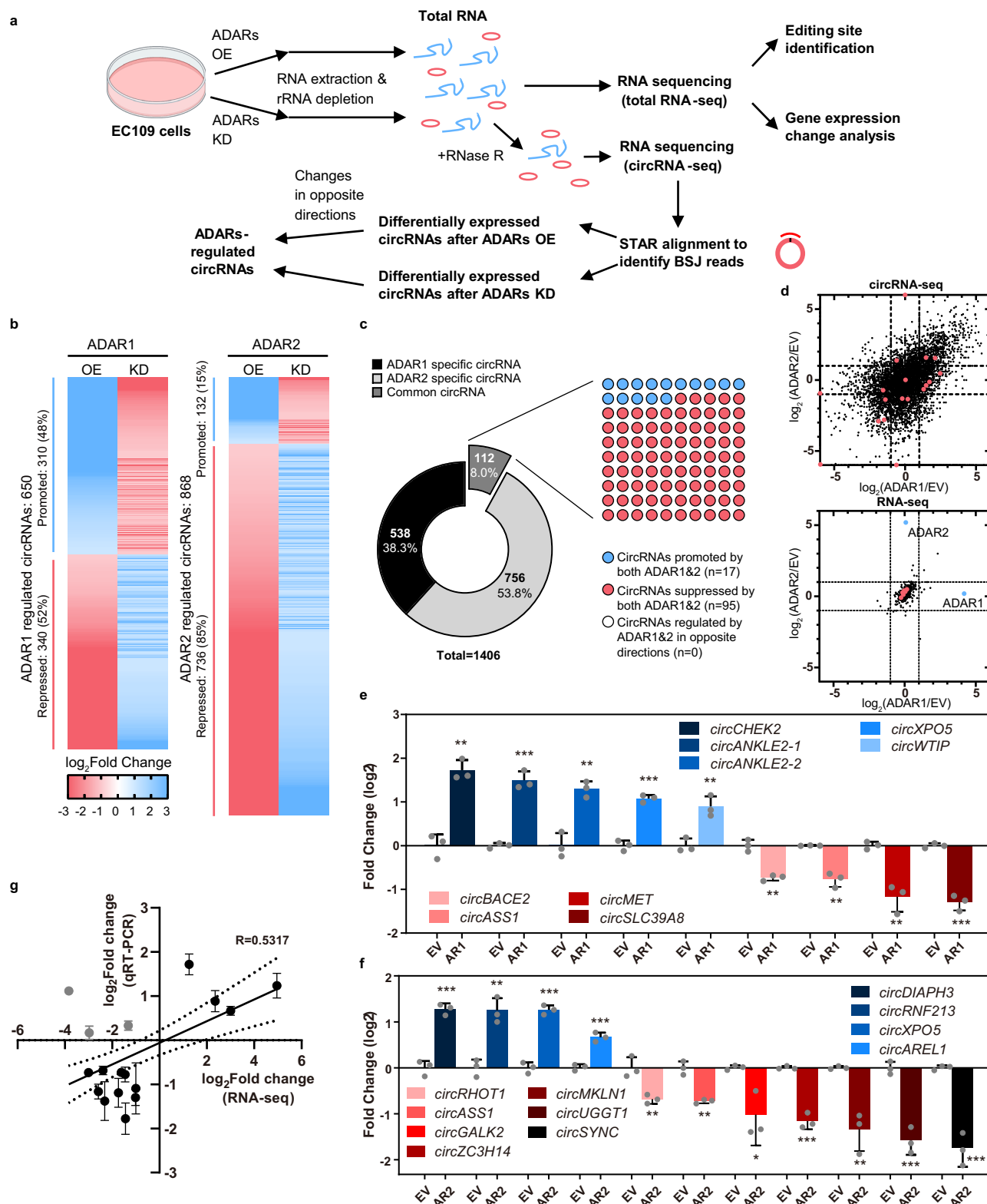
It is known that the differentially expressed ADARs and its resultant dysregulation of A-to-I RNA editome are implicated in multiple cancer types such as esophageal squamous cell

carcinoma (ESCC), hepatocellular carcinoma (HCC), colorectal cancer (CRC), breast cancer, and gastric cancer^{18–24}. Herein, we comprehensively define the regulatory role of ADARs in circRNAs and explore the functional relevance of target circRNAs to cancer. We uncover over a thousand circRNAs either promoted or repressed by ADAR1 and/or ADAR2 via editing-dependent or -independent mechanisms. Next, our mechanistic investigation deciphers editing-dependent mechanisms of action by which editing can stabilize or destabilize secondary structures formed between RCMs within the flanking introns via correcting A:C mismatches to I(G)-C pairs or creating I(G).U wobble pairs, respectively. We also find that editing facilitates the recruitment of RBPs such as PTBP1 to regulate back-splicing. Moreover, we show that these ADARs-regulated circRNAs (ARcircs) are not merely by-products of back-splicing, but indeed influence tumorigenesis. Our findings provide a previously undescribed bidirectional regulation of circular transcriptome by ADARs and highlight a complex crosstalk between RNA editing machinery and circRNA biogenesis and its implications in cancer.

Results

ADARs regulate circRNA biogenesis bidirectionally. To query the role of ADARs in regulating circRNAs, we modulated the expression level of ADAR1 or ADAR2 (ADAR1/2) by either forced expression or silencing in EC109 which is an esophageal squamous carcinoma cell line that has been frequently used for ADARs and A-to-I editing research^{23–25}. Untreated or RNase R-treated RNA samples were subsequently sent for total RNA sequencing (RNA-Seq) or circRNA sequencing (circRNA-Seq), respectively (Fig. 1a). We applied an in-house pipeline for circRNA detection and identified a total of 37,916 circRNAs. To ensure the reliability of our analysis, we compared the performance for circRNA identification between our pipeline and two commonly used benchmark methods CIRI2 and CIRCexplorer2^{26–28} and obtained a high percentage of overlapping circRNAs between our in-house pipeline and CIRI2 or CIRCexplorer2 (87% for CIRCexplorer2; 70% for CIRI2) (Supplementary Fig. 1a). With our stringent filter criteria (see Methods), a total of 650 and 868 circRNAs were identified as high-confidence ADAR1 or ADAR2-regulated circRNAs (ARcircs), respectively (Fig. 1b and Supplementary Data 1). Intriguingly, both ADAR1 and ADAR2 proteins were found to regulate circRNAs in both directions (Fig. 1b, c). Unlike ADAR1 which exerts its suppressive or promoting effect on approximately the same amount of circRNAs (promoting: 48% vs repressing: 52%), ADAR2 is most likely to be a potent repressor of circRNAs rather than an enhancer (promoting: 15% vs repressing: 85%) (Fig. 1b, c). Among approximately a hundred circRNAs regulated by both ADAR proteins (defined as common circRNAs), no circRNA was found to be regulated by ADAR1/2 in opposite directions (Fig. 1c). Of note, using the same filter criteria, 93% (1,313/1,406) of ARcircs identified by our pipeline were also found by CIRCexplorer2 and demonstrated the same pattern of changes upon modulation of ADAR1/2 expression (Supplementary Fig. 1b and Supplementary Data 1). To rule out the possibility that such effects might arise from the changes in linear mRNA expression, we analyzed expression changes in all detected circRNAs and their host gene transcripts upon modulation of ADAR1/2 expression and observed drastic expression changes in circRNAs, but not their corresponding linear mRNAs (Fig. 1d). Our finding is consistent with previous reports that ADARs has no major effect on global gene expression, even those undergoing A-to-I editing^{29–31}.

To confirm these findings, we randomly selected 21 candidate circRNAs for experimental validation (Supplementary Fig. 1c). Upon RNase R digestion, all linear forms underwent more than



50-fold reduction in their expression; while circRNAs demonstrated strong resistance to the digestion, indicating that these candidate circRNAs are truly circularized RNA molecules (Supplementary Fig. 1d). Sanger sequencing analyses of purified circRNA products further confirmed these back-splicing events (Supplementary Fig. 1e). We then went on to validate the regulatory effects of ADARs on these candidate ARcircs identified by circRNA-Seq. Expression change was successfully verified for

20 out of 23 candidate ARcircs (ADAR1-regulated: 9 out of 10; and ADAR2-regulated: 11 out of 13), but not for their host linear mRNAs (Fig. 1e–g and Supplementary Fig. 1f). Of note, 2 ARcircs *circASS1* and *circXPO5*, were confirmed to be common targets for both ADARs (Fig. 1e–g and Supplementary Fig. 1e). Moreover, these ARcircs were further validated in ADAR1/2-knockdown cells (Supplementary Fig. 1g, h). These data indicate that ADARs indeed function as potent bidirectional regulators of circular

Fig. 1 ADAR1 and ADAR2 regulate circRNA biogenesis bidirectionally. **a** Workflow for identification of ADARs-regulated circRNAs (ARcircs). **b** Heat maps indicating the fold change in expression of candidate circRNAs, upon modulation of ADAR1 or ADAR2 expression through lentivirus-based knockdown (KD) or overexpression (OE). A relative decrease in the KD or OE samples is indicated as blue, while an increase is indicated as red. **c** Doughnut chart depicts the percentage of circRNAs regulated specifically by either ADAR1 or ADAR2, or by both ADAR proteins. **d** 10 × 10 dot plot illustrates the number of common circRNAs which are regulated by ADAR1 and ADAR2 in either same or opposite direction. **e** Scatterplots displaying the fold changes in expression levels of all 37,916 detected circRNAs (upper) or their corresponding host genes (lower), upon overexpression of ADAR1/2 versus empty vector control in EC109 cells. A total of 20 randomly selected circRNAs from **(b)** and their corresponding host genes are indicated as red dots. Blue dots indicate ADAR1 and ADAR2. **e, f** Quantitative real-time PCR (qRT-PCR) validation of the indicated circRNAs. **g** Correlation between fold change calculated from circRNA-Seq and qRT-PCR validation data. Dash lines show 95% confidence interval. CircRNAs not validated are showed in grey color. **e–g** Data are presented as the mean ± S.D. of technical triplicates from a representative experiment of 2 independent experiments (unpaired, two-tailed Student's *t*-test; **P* < 0.05; ***P* < 0.01; ****P* < 0.001). Exact *P* values and source data are provided in Source Data file.

transcriptome, with ADAR2 appearing skewed towards a repressor. For those circRNAs regulated by both ADARs, they are most likely to be regulated in the same direction.

ADARs regulate circRNAs through or beyond their editing function. Base-pairing of reverse complementary matches (RCMs) residing in circRNA flanking introns facilitates circRNA production^{7,8}. To further dissect the mechanism underpinning the regulatory role of ADARs in circRNA biogenesis, we first identified 41,551 high-confidence A-to-I editing sites from the total RNA-Seq data and 1,043 ARcircs with ≥1 RCM locating in their flanking introns from the circRNA-Seq data (Fig. 2a and Supplementary Data 2). To ensure the specificity of identified RNA editing sites, we checked the proportion of each possible type of mismatches and found that A to G accounts for approximately 90% of all detected mismatches, consistent with previous studies reporting A-to-I editing as the most common type of RNA editing in humans^{32,33} (Supplementary Fig. 2a). Moreover, we analyzed the sequence preference for neighboring nucleotides surrounding editing sites (± 2 nt) and found “G” is preferred to be excluded at 5' neighbour but included at 3' neighbour of editing sites, as reported previously^{34,35} (Supplementary Fig. 2b). We then went on to analyze the distribution of these editing sites and RCMs across the flanking introns (Fig. 2a). Not surprisingly, RCMs are obviously enriched in the intronic region proximal to the back-splicing junctions (−500 nt ~ +500 nt; black dots, Fig. 2b), indicating that intronic matches between flanking introns are truly involved in back-splicing. A previous study reported that both RCMs and editing sites from RADAR (Rigorously Annotated Database of A-to-I RNA editing)³⁶ are preferentially distributed near the splice sites of circularized exons⁸. However, from our analysis, the locations of editing sites identified in either our own EC109 total RNA-seq data or the RADAR database are not enriched in the proximal back-splice junction (BSJ) region (Fig. 2b, red and blue lines). A recent study also suggested that such an enrichment was not observed in the flanking introns of circRNAs regulated by ADARs in the mouse bone marrow or liver tissue samples³⁷.

We next questioned if the editing capability of ADARs is indispensable for their regulation of circRNAs. To this end, the ADAR1/2 mutants depleted of either editing activity only (DeAD mutants)³⁸ or both RNA binding and editing capabilities (EAA mutants)^{39,40} were generated. Upon overexpression of each wildtype or mutant form, the ADAR1/2 EAA mutant was incapable of regulating all ARcircs (Fig. 2c, d and Supplementary Fig. 2c, d), suggesting that RNA binding ability is critical for ADARs to regulate ARcircs; unlike the EAA mutant, the DeAD mutant was able to modulate the expression of approximately half of ARcircs such as *circXPO5*, *circASS1*, and *circRNF213*, to a similar or less extent than the wildtype form (Fig. 2c, d and Supplementary Fig. 2c, d). These data suggested that ADARs can regulate circRNAs through their editing-dependent and/or

independent functions. To further interrogate whether such an editing-dependent/independent regulation of circRNA biogenesis can be observed in a transcriptome-wide manner, we overexpressed the ADAR1/2 DeAD mutant or the empty vector (EV) control in EC109 cells and performed circRNA-Seq to identify editing-dependent and -independent ARcircs. From this batch of circRNA-Seq, we could detect 76.3% (1,073/1,406) of ARcircs identified from our previous circRNA-Seq. Among these 1,073 ARcircs, 767 were identified as editing-dependent ARcircs and 306 as editing-independent ones regulated by ADAR1 and/or ADAR2 (Methods and Supplementary Data 3). All these findings strongly indicate that ADARs can bidirectionally regulate circRNAs via editing-dependent and/or independent mechanisms in a transcriptome-wide manner.

ADAR1 promotes *circCHEK2* biogenesis via its direct binding and editing of *circCHEK2* flanking introns. So far, there is a lack of experimental evidence about the mechanism underpinning the regulation of circRNA biogenesis via ADAR-mediated editing. *CircCHEK2*, an editing-dependent ARcirc generated by back-splicing between exon 3 and 9 of its host gene *CHEK2* (Fig. 3a), was chosen as an exemplary target for further study. We first analyzed the publicly available ADAR1 RNA immunoprecipitation sequencing (fRIP-Seq) dataset^{41,42} and found that ADAR1 binding peaks enriched in both flanking introns 2 and 9, especially the identified RCM pair with the highest BLAST score (Fig. 3a). A high probability of dsRNA formation between the predicted RCM was supported by secondary structure prediction using RNAfold⁴³. Further, by performing RNA immunoprecipitation (RIP) assay, we confirmed the association of ADAR1 with the dsRNA structure formed between the identified RCMs in vivo (Fig. 3b). We then provided experimental evidence that upon ADAR1 overexpression, three editing sites (sites #1, #2 and #3) within RCMs could be detected with editing frequencies ranging from 19% to 28.6% (Fig. 3c). Moreover, the site #1 is located in a previously reported ADAR1 binding motif⁴⁴ (Fig. 3c). All these findings suggest that ADAR1 indeed binds and edits the dsRNA structure formed between RCMs located in the flanking introns of *circCHEK2*.

To explore whether editing of RCM has an effect on *circCHEK2* expression, we generated a *circCHEK2* minigene containing the partial sequence of *CHEK2* gene, including the entire exons 3–9, part of exon 2, exon 10 and flanking introns 2 and 9 (Fig. 3d). We transfected the *circCHEK2* minigene together with the wildtype ADAR1 or DeAD mutant into cells. Like endogenous *circCHEK2*, exogenous *circCHEK2* derived from the minigene was also regulated by ADAR1 dependent on the latter's editing function (Fig. 3e). Increased editing was observed at the same editing sites within the flanking intronic sequence of the exogenous pre-mRNA (Fig. 3c). It is known that cellular machineries recognize inosine as guanosine (G), due to their high structural similarity. To further understand whether all three

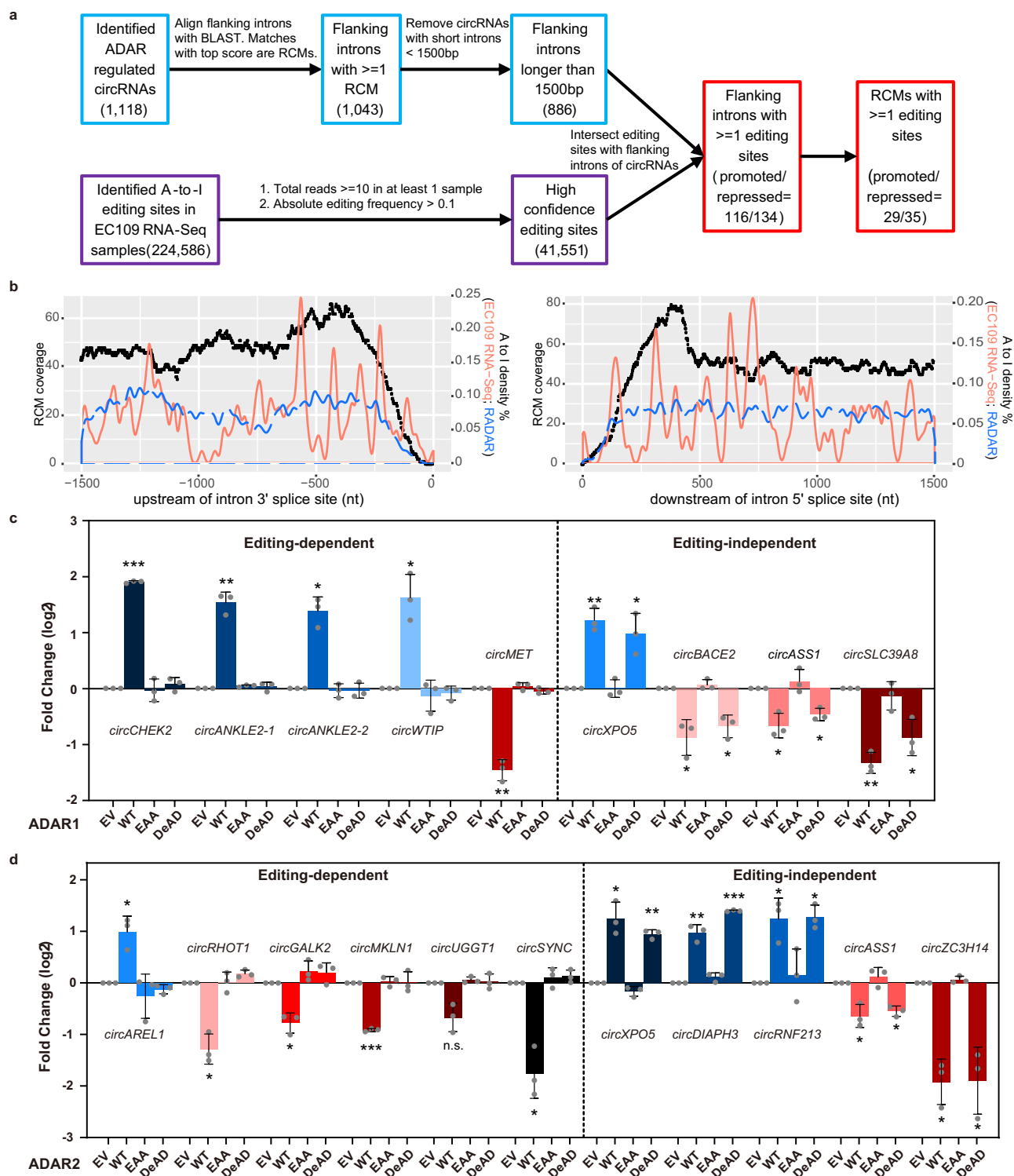
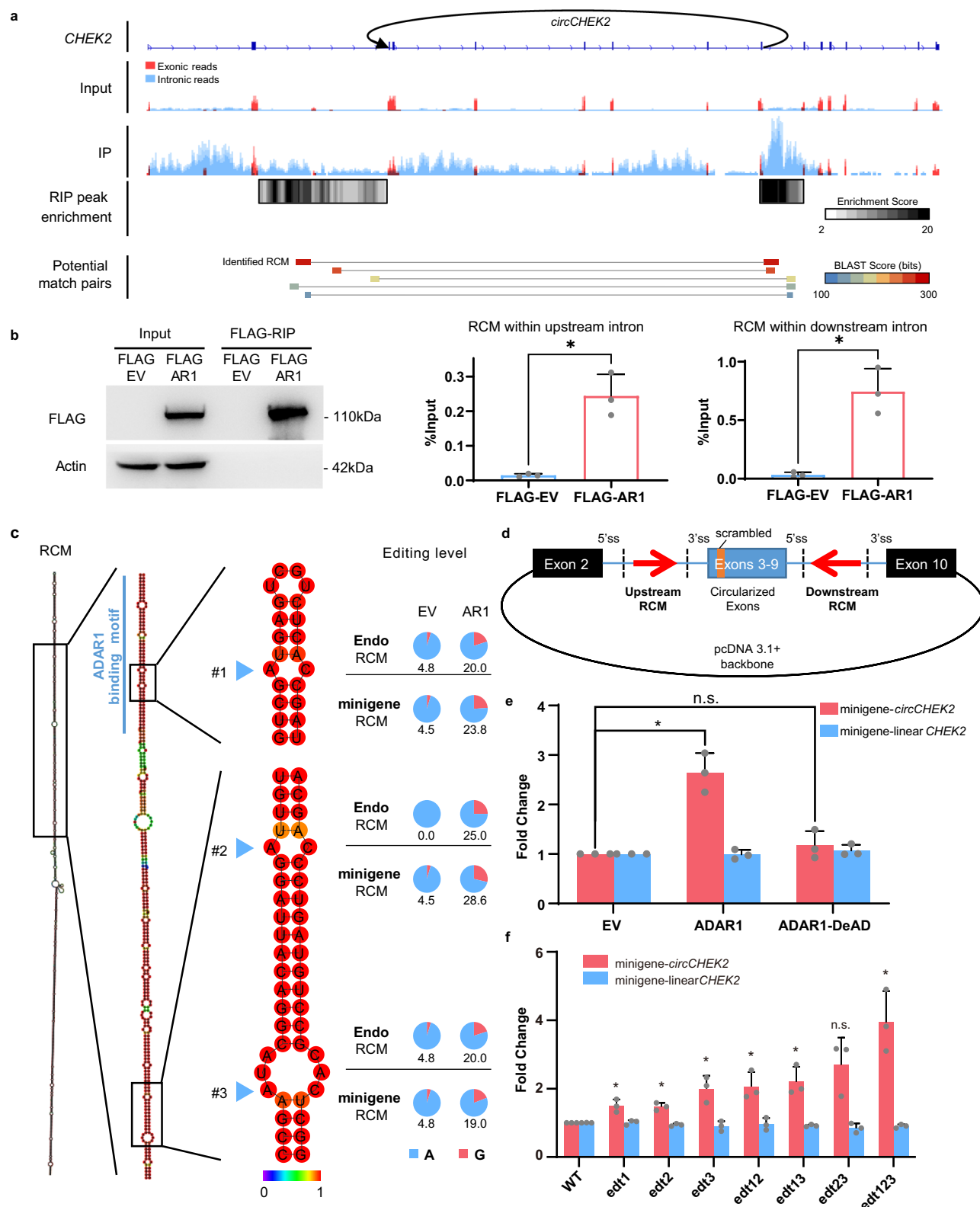


Fig. 2 ADAR1/2 regulate circRNA biogenesis via either editing-dependent or -independent mechanisms. **a** Workflow for identification of reverse complementary matches (RCMs) and A-to-I editing sites at flanking introns of ARcircs. **b** Distribution of RCMs (left Y-axis) and editing sites (right Y-axis) across the flanking intronic region spanning 1,500nt upstream (left panel) and 1,500nt downstream (right panel) of back-splicing junction site of ARcircs. Black dotted lines indicate the distribution of RCMs. Red or blue lines indicate the distribution of editing sites identified from our EC109 RNA-seq data or the RADAR database, respectively. **c, d** qRT-PCR analysis of expression change of the indicated ARcircs, upon overexpression of the wildtype (WT), EAA mutant, or DeAD mutant form of ADAR1 (**c**) or ADAR2 (**d**) versus empty vector (EV) control in EC109 cells. Each dot represents the mean value of technical triplicates from an independent experiment. Data are presented as the mean \pm S.D. of 3 biological replicates (paired, two-tailed Student's *t*-test; **P* < 0.05; ***P* < 0.01; ****P* < 0.001). Exact *P* values and source data are provided in Source Data file.



editing sites are involved in such regulation, we introduced an A-to-G mutation into the minigene at each editing site to mimic a fully edited site. Intriguingly, each single, double, or triple combination of these mutations led to increased expression of minigene- *circCHEK2* (Fig. 3f). Of note, these three editing sites demonstrated synergistic effect on promoting *circCHEK2* biogenesis (Fig. 3f). Collectively, ADAR1-mediated editing of RCMs can promote *circCHEK2* biogenesis.

A-to-I editing may alter circRNA production via stabilizing or destabilizing dsRNA formed between RCMs. As dsRNA between circRNA flanking introns is one of the key factors for circRNA biogenesis, we next asked whether ADAR1-mediated editing of RCMs alters the secondary structure. Based on the in silico secondary structure prediction, all three edited adenosines form A:C mismatches in the dsRNA, and editing at A:C mismatches which changes A:C to I(G)-C may enable a more perfect

Fig. 3 ADAR1 binds and edits *circCHEK2* RCM to promote *circCHEK2* biogenesis. **a** Genome browser tracks of *CHEK2* loci reveal ADAR1 binding peaks (top) from ADAR1-fRIPseq data and predicted RCM pairs within *circCHEK2* flanking introns (bottom). Black arrow: the circular junction site of *circCHEK2* in a 5'-3' direction. Reads mapped to exonic or intronic regions (GENCODE annotation) are colored in red or blue, respectively. Potential match pairs are indicated in different colors and the pair with the highest BLAST score is defined as "Identified RCM". **b** RIP-qPCR analysis of the association of ADAR1 protein to the *circCHEK2* RCM region in EC109 cells transfected with FLAG empty vector (FLAG EV) or FLAG-ADAR1 (FLAG AR1). WB and qPCR analyses of FLAG-RIP immunoprecipitates are shown in the left and right panels, respectively. **c** Secondary structure formed by RCMs of *circCHEK2*, as predicted by RNAfold (left). Location of a reported ADAR1 binding motif is indicated by blue line. Blue arrows indicate 3 editing sites identified within RCMs of *circCHEK2*. Base-pair probabilities are shown by a color spectrum (middle). Pie charts illustrating the editing frequency (indicated by red slice) of each editing site in the indicated samples (right). Editing frequency of each editing site was measured using TA cloning (see Methods). **d** Schematic diagram illustrating the structure of *circCHEK2* minigene. A 20-bp sequence of exon 3 was scrambled to distinguish minigene-produced transcripts from endogenous transcripts. **e** Fold change in expression of minigene-produced *circCHEK2* and linear *CHEK2*, upon overexpression of WT or DeAD ADAR1, compared to EV control. **f** Fold change in expression between *circCHEK2* and linear *CHEK2* derived from the WT or mutated minigenes carrying A-to-G mutation(s) at editing sites. Edt1, A-to-G mutation at site #1; edt12, A-to-G mutations at sites #1 and #2, and so forth. **b, e, f** Data are presented as the mean \pm S.D. of 3 biological replicates. Each dot represents the mean value of technical triplicates from an independent experiment. Data is presented as mean \pm S.D. of 3 biological replicates. Statistical significance is determined by paired, two-tailed Student's t-test (*, $P < 0.05$; n.s., not significant). Exact P values and source data are provided in Source Data file.

secondary structure, which potentially facilitates circRNA production (Fig. 4a). It has been known that tightly folded RNAs travel more rapidly than unfolded RNAs of the same length or molecular weight⁴⁵. To test our hypothesis, we generated RNA probes containing the *circCHEK2* RCM sequence with or without single, double, or triple A to G mutations at the three editing sites and performed native polyacrylamide gel electrophoresis (PAGE). As expected, probes with mutations at all three editing sites (edt123) migrated more rapidly on gel (Fig. 4b), suggesting that editing may enable RCMs within the flanking intronic sequence to form a more compact structure to stabilize the dsRNA.

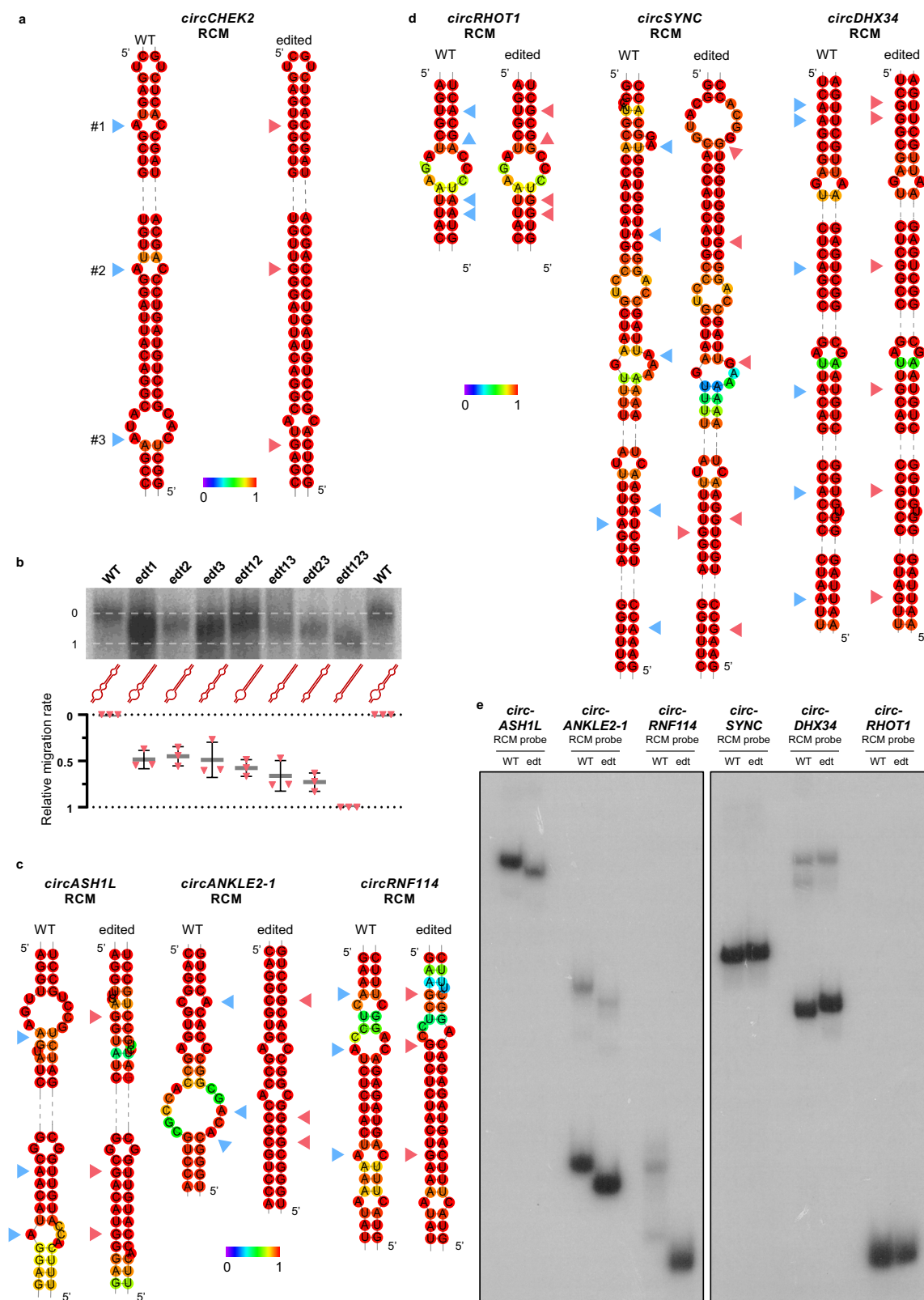
To obtain more experimental evidence supporting that editing of RCMs can alter dsRNA structure, we selected 6 additional editing-dependent ARcircs identified by circRNA-Seq (Supplementary Data 3), including 3 ADAR1/2-promoted circRNAs (*circASH1L*, *circANKLE2-1* and *circRNF114*) and 3 ADAR1/2-repressed circRNAs (*circSYNC*, *circDHX34* and *circRHOT1*). From our RNA-Seq data, all 6 ARcircs have editing sites within their RCMs demonstrating $\geq 10\%$ increase in editing frequency upon overexpression of the corresponding ADAR protein. Using the same strategy, we found that the majority of editing sites within RCMs of *circASH1L*, *circANKLE2-1* and *circRNF114* locate at A:C mismatches where editing was predicted to lead to more compact dsRNA structures (Fig. 4c); on the contrary, upon editing, RCMs of *circSYNC*, *circDHX34* and *circRHOT1* hypothetically form looser dsRNA structures via changing A-U base pairs to weaker G:U wobble base pairs or affecting the structures of neighboring regions (Fig. 4d). Intriguingly, native PAGE analysis showed that for *circASH1L*, *circANKLE2-1* and *circRNF114*, the edited RCM probes with A-to-G mutations at all editing sites (indicated by arrows, Fig. 4c, d) migrated faster in the gel than the unedited/wildtype probes (Fig. 4e, left panel); while for *circSYNC*, *circDHX34* and *circRHOT1*, the edited probes migrated slightly slower than the unedited/wildtype probes (Fig. 4e, right panel). Therefore, there could be a universal editing-dependent mechanism by which ADARs regulate circRNA biogenesis via editing-mediated change in the secondary structure formed by flanking introns.

A-to-I editing enhances PTBP1 binding to flanking introns of *circCHEK2* to promote its biogenesis. Other than causing structural changes, A-to-I editing of intronic sequence has been proved to be a regulator of splicing by creating or modifying auxiliary *cis*-acting elements for splicing factor binding^{15,17}. Since the canonical machinery of spliceosome also functions in circRNA biogenesis, we next asked whether editing could facilitate

the binding of splicing regulators and affect back-splicing via changing *cis*-acting elements. To this end, we predicted RBPs which demonstrate binding preference near *circCHEK2* editing sites using RBPmap⁴⁶. Two RBPs, TDP43 and PTBP1 with a respective binding motif near the editing site #1 and #2, are of particular interest (Fig. 5a). RNA pulldown assay was performed by incubating the whole cell lysates with the wildtype (WT) or triple mutant (edt123) RNA probe. Intriguingly, PTBP1 was found to bind more strongly to the edt123 than the WT probe, while TDP43 did not show any distinct binding preference between 2 probes (Fig. 5b and Supplementary Fig. 3a). We further performed PTBP1 RIP assay in EC109 cells with or without overexpression of ADAR1 and found that binding of PTBP1 to the *circCHEK2* RCM region was significantly enhanced upon overexpression of ADAR1 (Fig. 5c). Intriguingly, the proportion of edited RCM transcripts (shown by editing frequencies of all 3 editing sites) was increased in PTBP1 RIP products when compared to the 'Input' samples, particularly in ADAR1-overexpressing cells (Fig. 5d), further confirming that editing enhances PTBP1 binding to the *circCHEK2* RCM region.

Next, we determined the regulatory effect of PTBP1 and TDP43 on *circCHEK2* biogenesis. In the absence of PTBP1 knockdown, there was an approximately 5-fold higher expression of *circCHEK2* derived from the triple mutant (edt123) minigene than the WT counterpart; however, upon knockdown of PTBP1, the difference in the efficiency of *circCHEK2* production between edt123 and WT minigene was significantly attenuated (Fig. 5e and Supplementary Fig. 3b). However, such changes were not observed upon silencing of TDP43 (Fig. 5f and Supplementary Fig. 3c). Endogenously, silencing of PTBP1 also reduced the promoting effect on *circCHEK2* biogenesis caused by ADAR1 overexpression (Fig. 5g and Supplementary Fig. 3d). Previous study reported that PTBP1 could affect the translation of ADAR1 in glioma cells⁴⁷, which may serve as an additional regulatory mechanism of PTBP1 on *circCHEK2*. However, we did not observe any obvious reduction in ADAR1 protein level upon silencing of PTBP1 (Supplementary Fig. 3e). All these data suggested that besides editing-mediated change in the secondary structure formed by *circCHEK2* flanking introns, editing can also enhance PTBP1 binding to the flanking introns and promote *circCHEK2* biogenesis.

Editing can alter RBP binding sites in the flanking introns of circRNAs in a transcriptome-wide manner. Inspired by our observations, we next sought to investigate whether editing-mediated changes in binding sites of RBPs may serve as a general mechanism to regulate circRNA biogenesis. With the same



stringent filter (Fig. 2a), we identified 571 editing sites distributed within flanking introns of 92 editing-dependent ARcircs. We next retrieved the sequence surrounding editing sites (± 10 nt) and analysed RBP binding motifs before and after editing using RBPmap⁴⁶, followed by the calculation of the number of circRNAs which have altered RBP binding sites on flanking introns due to editing. We found that among 132 analysed RBPs with

annotated binding sites in RBPmap, 129 RBPs, including PTBP1 and those which have been shown to regulate circRNA biogenesis such as MBNL1⁵, FUS⁴⁸, SFPQ⁴⁹, HNRNPL⁵⁰, KHSRP⁵⁰, and QKI⁶, were found to have editing-mediated changes in their binding sites at flanking introns of more than 10 editing-dependent ARcircs (Fig. 5h and Supplementary Data 4), implying that altering RBP binding affinity is an important mechanism for

Fig. 4 A-to-I editing alters the dsRNA structure formed by RCMs within flanking introns of circRNAs. **a** Predicted secondary structures formed by *circCHEK2* RCM with or without A-to-I(G) editing by RNAfold. Partial RNA structures which contain editing sites and neighboring sequences are shown. Blue and red arrows indicate unedited/wildtype (WT) adenosines and edited/mutated (edited) sites, respectively. **b** Migration on native polyacrylamide gel of RNA probes containing the *circCHEK2* RCM sequence with or without A-to-I(G) editing at each editing site. Simplified secondary structure of each probe is shown. Calculation of the relative migration rate was discussed in Methods. Data are presented as the mean \pm S.D. of biological triplicates. **c, d** Predicted secondary structures formed by RCMs of ADAR1/2-promoted circRNAs (*circASH1L*, *circANKLE2-1* and *circRNF114*) **c** and ADARs-repressed circRNAs (*circRHOT1*, *circSYNC* and *circDHX34*) **d** by RNAfold, with or without editing. Partial RNA structures which contain editing sites and neighboring sequences are shown. Blue and red arrows indicate WT adenosines and edited sites, respectively. **e** Migration on native polyacrylamide gel of RNA probes containing the wildtype (WT) and edited (edt) partial RCM sequences of the indicated circRNAs. Representative result of $n = 2$. **a, c, d** Base-pair probabilities are shown by a color spectrum. Source data are provided in Source Data file.

editing to regulate circRNA biogenesis. Taken together, editing can not only alter the stability of secondary structure formed between RCMs, but also affect RBP binding to flanking intronic sequences, leading to changes in circRNA production.

ADARs-mediated circRNA regulation exists in multiple cancer types. ADAR1 and ADAR2 are ubiquitously expressed in many tissue types¹¹. We wondered if ADARs function as potent regulators of circular transcriptome in multiple cancer types. To address this, we selected five validated ARcircs and detected their expression changes upon overexpression of the wildtype or mutant form of ADAR1/2 in MB231 (breast cancer cell line), MKN28 (gastric tubular adenocarcinoma cell line), SNU398 (hepatocellular carcinoma cell line), and HCT15 (colorectal cancer cell line). Intriguingly, we observed the same pattern of editing-dependent or independent regulation of ARcircs in these cell lines as EC109 cells (Fig. 6a–d and Supplementary Fig. 4a, b). We then investigated the expression pattern of *circCHEK2* and the association between expression levels of *ADAR1* and *circCHEK2* in 17 matched pairs of primary HCC and non-tumor (NT) liver samples as well as 20 matched pairs of primary colorectal cancer (CRC) and NT colon samples. We found that 41% (7 out of 17) and 60% (12 out of 20) of HCC and CRC patients demonstrated a ≥ 2 -fold increase in *circCHEK2* expression in tumors compared to their NT samples, respectively (Fig. 6e, f, upper panels). Next, both HCC and CRC patients were stratified into two groups: ADAR1-down and ADAR1-up, based on the decreased or increased expression of ADAR1 in tumors compared to their matched NT samples, respectively (Fig. 6e, f, lower panels). We found that in the ADAR1-down or ADAR1-up group of HCC patients, 3 out of 6 (50%) or 6 out of 11 (54.5%) showed ≥ 2 -fold decrease or increase in *circCHEK2* expression in tumors, respectively (Fig. 6e). Likewise, in the ADAR1-down or ADAR1-up group of CRC patients, 4 out of 6 (67%) or 8 out of 14 (57%) showed ≥ 2 -fold decrease or increase in *circCHEK2* expression in tumors, respectively (Fig. 6f). These findings suggested that ADARs-mediated circRNA regulation is most likely present in multiple cancer types.

Impacts of ARcircs on tumorigenesis. To investigate the potential involvement of ARcircs in tumorigenesis, we utilized CasRX (also known as RfxCas13d) system^{51,52} and designed guide RNAs (gRNAs) against the back-splicing junction sequence of each ARcirc for a specific and efficient knockdown without affecting their host genes expression (Fig. 7a). In 2 different types of cancer cell lines EC109 and SNU398, knockdown of *circCHEK2*, *circGALK2*, and *circSLC39A8* significantly reduced the tumorigenic ability of cells, as manifested by decreased frequencies of focus formation and colony formation in soft agar, suggesting that these ARcircs have a cancer-promoting role (Fig. 7b–e). We further provided in vivo evidence that *circCHEK2* knockdown in EC109 and SNU398 cells led to a significant

reduction in tumor growth rate than the control counterparts (Fig. 7f, g). All these data suggested that these ARcircs are of functional relevance to multiple types of cancers.

Discussion

Although several previous studies reported that RNA editing enzymes ADARs function as repressors of circRNA biogenesis^{8,13} or have no major regulatory effect on circRNAs⁹, our study demonstrates that ADARs are potent regulators of circular transcriptome and they can regulate over a thousand of circRNAs in both directions through and beyond their editing functions. However, it remains unknown what mechanisms determine the direction of circRNA regulation by ADARs. One key factor is the position of the edited adenosine within a dsRNA. Previously proposed model suggested that ADARs destabilize the secondary structure through altering A-U base pairs located in the dsRNA stem, leading to repression of circRNA biogenesis. Here, we provided experimental evidence that adenosines at A:C mismatches, which gain editing preference than those at A-U pairs¹⁴, can stabilize the dsRNA structure formed between flanking introns, promoting circRNA biogenesis. It has been known that canonical splicing signals and spliceosomal machinery are required for back-splicing, and editing at *cis*-acting elements (e.g., branch point site, splicing enhancers/silencers) can result in changes of splicing pattern^{15–17}. Therefore, the location of editing sites within the host gene transcript may influence circRNA expression. In this work, we demonstrate that ADAR1-mediated A-to-I editing can enhance binding of splicing factor PTB1 to the flanking intron of *circCHEK2*, rendering increased expression of *circCHEK2*. Although PTB1 is well documented as a pyrimidine-rich sequence binding protein, a previous study also showed that guanosine containing triplets contribute to PTB1 binding⁵³. This explains our observation that A-to-I (G) substitutions within *circCHEK2* RCMs could enhance PTB1 binding. Moreover, we provided a large-scale prediction of editing-mediated changes on RBP binding motifs on flanking introns of ARcircs and found that upon editing, most analyzed RBPs have altered binding sites in flanking introns of more than 10 editing-dependent ARcircs, further suggesting that editing may regulate circRNA biogenesis through affecting RBP binding in a transcriptome-wide manner. One should note that there may be other editing-dependent and -independent mechanisms underpinning the regulation of circRNAs by ADARs, such as altering the circRNA turnover and splicing.

RNA editing, alternative splicing, polyadenylation, and back-splicing are crucial RNA processing steps that expand transcriptome diversity. As each step heavily involves base-pairing (e.g., dsRNA formation of IRAlu elements), it is not surprising that these processes undergo extensive crosstalk. These dsRNAs recruit RBPs, dramatically increasing the complexity of RNA processing network. One example is DHX9, which regulates RNA editing as a binding partner of ADARs and also suppresses circRNA biogenesis via unwinding dsRNAs formed by IRAlu

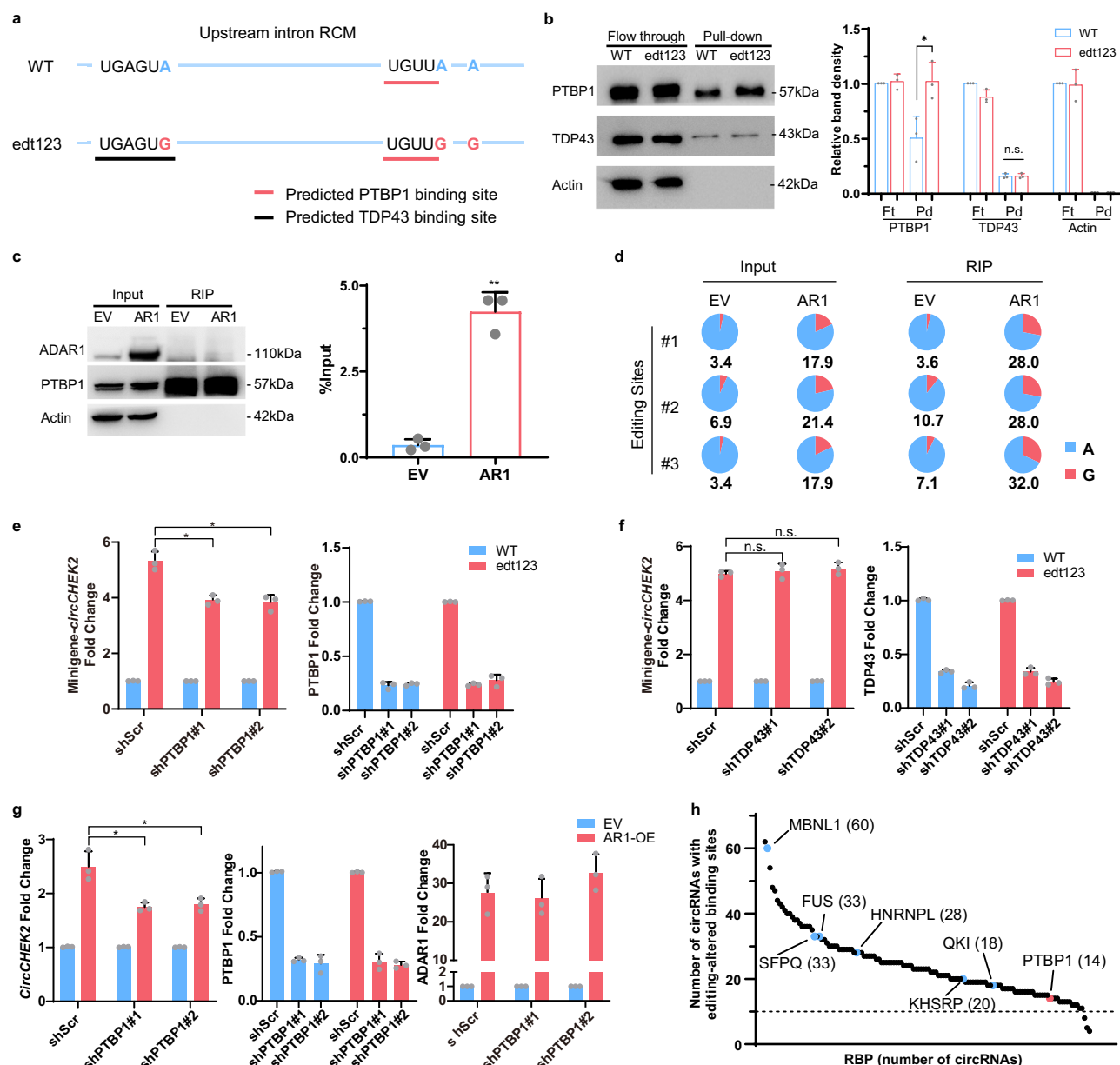


Fig. 5 A-to-I editing enhances PTBP1 binding to intron and promotes *circCHEK2* biogenesis. **a** In silico prediction of PTBP1 (red line) and TDP43 (black line) binding to unedited (WT) or triple mutated (edt123) RCM sequence of *circCHEK2* using RBPmap. **b** WB analysis of RNA pull-down products showing the binding affinity of PTBP1 and TDP43 to the WT or edt123 RNA probes. **c** RIP-qPCR analysis of the binding of PTBP1 protein to the *circCHEK2* RCM region in EC109 cells transfected with ADAR1 or empty vector control (EV). WB and qPCR analyses of PTBP1 RIP immunoprecipitates are shown in the left and right panels, respectively. **d** Pie charts illustrating the editing frequency (indicated by red slice) of each editing site (#1, #2, #3) in the indicated Input or RIP samples. Editing frequency of each editing site was measured using TA cloning (see Methods). **e, f** Left panels: Fold change in expression of *circCHEK2* produced by minigenes with or without A-to-G mutations at three editing sites, upon knockdown of PTBP1 **c** or TDP43 **d**. Right panels: qPCR analysis showing the knockdown efficiency of PTBP1 and TDP43. **g** Left panel: Fold change in expression of endogenous *circCHEK2* with or without lentivirus-mediated overexpression of ADAR1 in EC109 cells, upon knockdown of PTBP1. Middle panel: qPCR analysis showing the knockdown efficiency of PTBP1 in the indicated cells. Right panel: qPCR analysis illustrating the efficiency of ADAR1 overexpression in the indicated cells. **b, c, e, f, g** Each dot represents the mean of technical triplicates. Data are presented as mean \pm S.D. of 3 biological replicates (paired, two-tailed Student's *t*-test. n.s., not significant; **P* < 0.05; ***P* < 0.01). **h** Number of editing-dependent ARcircs of which flanking introns have editing-mediated changes in the binding sites of each RBP. Black dots indicate RBPs included in this analysis and the number in the bracket denotes the number of circRNAs with altered binding motifs of the corresponding RBP due to editing in flanking introns. Those which have been previously reported to regulate circRNA biogenesis are highlighted in blue. PTBP1 is highlighted in red. Exact *P* values and source data are provided in Source Data file.

elements^{9,54}. Interestingly, co-depletion of ADAR1 and DHX9 leads to synergistic effect on circRNA production. This implies the possibility that DHX9 plays as a regulator of circRNA biogenesis by tuning the editing frequency. Besides DHX9, recent studies on other non-ADAR editing regulators^{24,55} indicate an

additional layer of editing-dependent regulation of circRNA biogenesis.

Depending on their binding sites along RNA transcripts, ADARs can protect mRNA from degradation, regulate precursor microRNA processing and alter splicing pattern⁵⁶. Herein, we

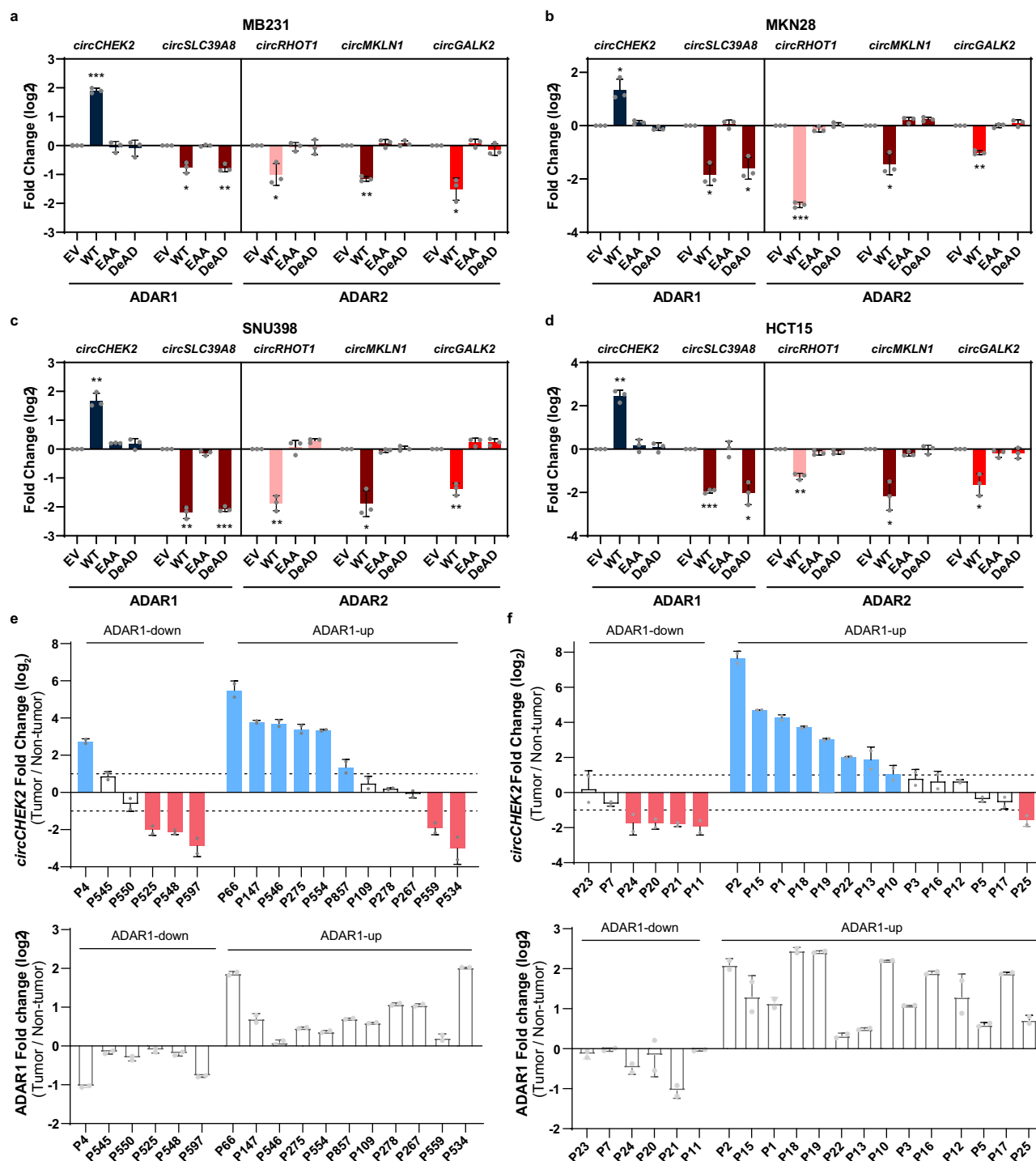


Fig. 6 ADARs-mediated circRNA regulation exists in multiple cancer types. **a–d** Fold change in expression of 5 validated ARcircs upon overexpression of WT, EAA, or DeAD form of ADAR1 or ADAR2, compared to the EV control, in MB231 **a**, MKN28 **b**, SNU398 **c**, and HCT15 **d**. Each dot represents the mean value of technical triplicates. Data are presented as the mean \pm S.D. of 3 biological replicates. Statistical significance was calculated by paired, two-tailed Student's *t*-test; *, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001. **e, f** Fold change in expression levels of *circCHEK2* (upper panels) and *ADAR1* (lower panels) between 17 primary HCC tumors **e** and 20 primary CRC tumors **f** and their matched NT liver and colon samples. Upper panels, cases demonstrating ≥ 2 fold higher or lower *circCHEK2* expression than their matched NT samples are shown by blue or red bars, respectively. Lower panels, patients were stratified into 2 groups: 'ADAR1-up' and 'ADAR1-down', according to the pattern of change in *ADAR1* expression between HCC or CRC and their matched NT samples. Each dot represents the mean value of technical triplicates. Data are presented as the mean \pm S.D. of 2 independent experiments. Exact *P* values and source data are provided in Source Data file.

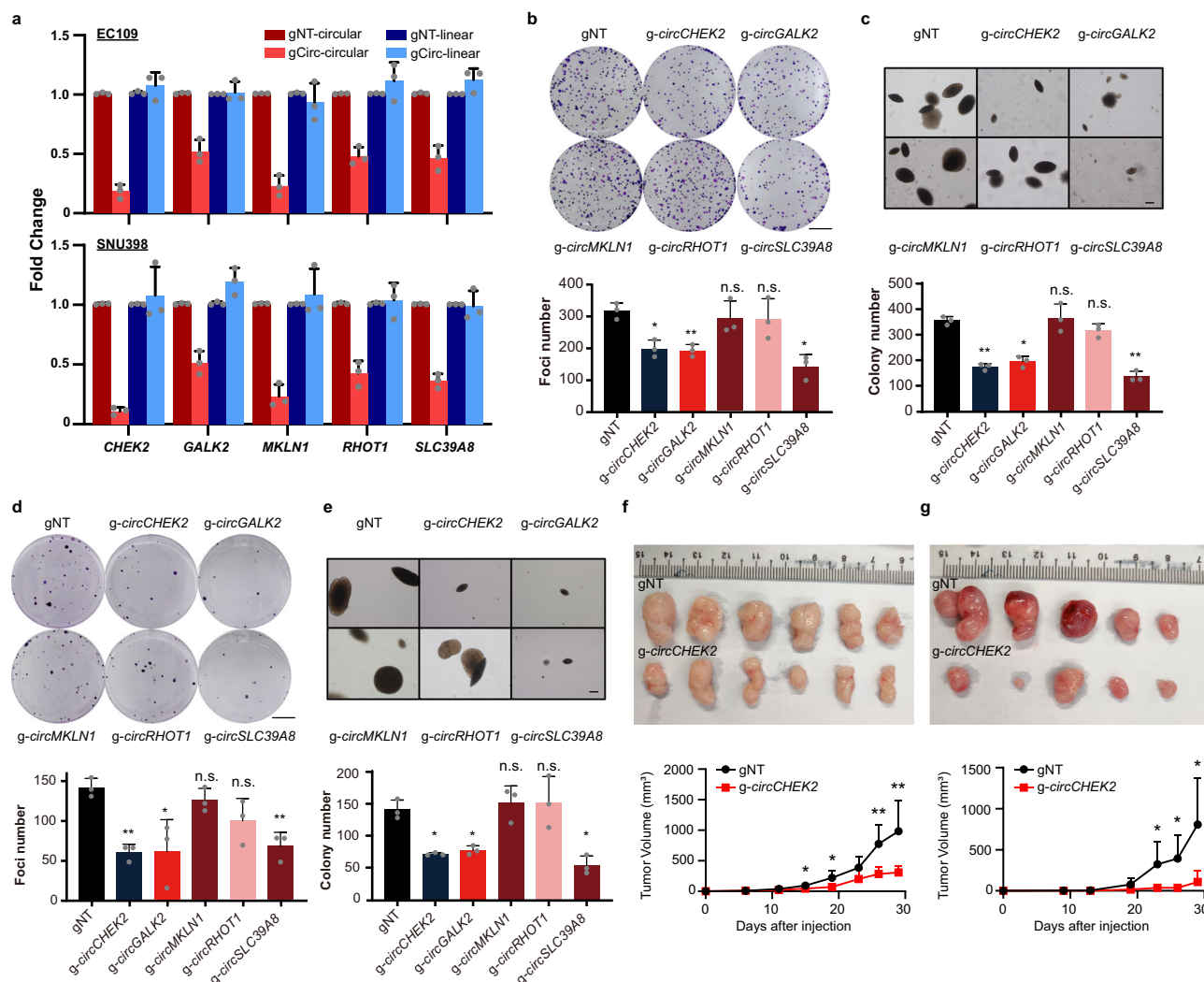


Fig. 7 ADARs-regulated circRNAs affect tumorigenesis. **a** Fold change in expression of circRNAs (circular) and their corresponding host genes (linear) in EC109 (upper panel) and SNU398 cells (lower panel), upon CasRX and guide RNA-mediated knockdown of circRNAs (gCirc) compared to non-targeting control (gNT). gNT-circular, circRNA expression of indicated gene upon treatment of CasRX and non-targeting guide RNA, and so forth. **b, c** Quantification of foci (**b**; scale bar, 1 cm) and colonies formed in soft agar (**c**; scale bar, 200 μm) by EC109 cells with or without circRNAs knockdown. **d, e** Quantification of foci (**d**; scale bar, 1 cm) and colonies formed in soft agar (**e**; scale bar, 200 μm) by SNU398 cells with or without circRNAs knockdown. **f, g** Images and growth curves of xenograft tumors derived from EC109 (n = 6) (**f**) and SNU398 cells (n = 5) (**g**). Data are presented as the mean ± S.D. of tumor volumes. Statistical significance was calculated by unpaired, two-tailed Student's *t*-test; *, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001. Exact *P* values and source data are provided in Source Data file.

mapped the identified RNA editing sites by RNA-Seq as well as RCMs of ARcircs to the region near the back-splicing sites. Our results implied the importance of base-pairing of flanking introns in circRNA biogenesis, consistent with previous studies^{7,8}. However, we found that unlike RNA binding, RNA editing is not always required for ADARs-mediated regulation of circRNAs. Nevertheless, the precise editing-independent mechanisms remain further investigations.

Circular RNAs were long considered as by-products of aberrant splicing without biological functions. Only until recently, plenty of circRNAs were found to play critical roles in multiple aspects of cellular and physiological functions, and the dysregulated circRNAs have been implicated in tumorigenesis^{3,57}. However, how ADARs-mediated changes in circRNA production contribute to cancer remain elusive. Herein, we provided extensive evidence supporting that the editing-dependent and/or independent regulation of circRNA expression by ADAR proteins is present in multiple types of

cancer cell lines and more importantly, the association between expression levels of ADARs and *circCHEK2* could also be found in HCC and CRC patient samples. We further showed these ARcircs are not merely by-products but indeed affect tumorigenesis, which poses an additional important function of ADARs in cancer. Of note, in our study, *circSLC39A8* exhibits an oncogenic role; however, the biogenesis of *circSLC39A8* could be repressed upon overexpression of ADAR1 which is largely characterized as an oncogene⁵⁶. In fact, the tumor promoting effect of ADAR1 is most likely arising from functional changes of multiple target genes via ADAR1's editing-dependent and independent functions. The effect of ADAR1 or ADAR2 on tumorigenesis may differ depending on their target genes and/or cell or tissue types. For example, ADAR1-mediated protein-recoding editing of antizyme inhibitor 1 (AZIN1) promotes hepatocarcinogenesis²¹; however, ADAR1-mediated editing can also suppress tumorigenesis by recoding Gamma-aminobutyric acid receptor subunit alpha-3 (GABRA3) in breast cancer⁵⁸. Likewise,

each ARcirc may have its distinct cancer-related functions and thus, one should note that contribution of ADARs-regulated circRNA biogenesis to cancer is unlikely to be attributed to one single ARcirc.

In sum, by identifying more than a thousand circRNAs regulated by ADARs, we uncover that ADARs could regulate circRNAs in both direction via editing-dependent and editing-independent mechanisms. We provide experimental evidence that ADAR1/2 can edit RCM of ARcirc, altering the secondary structure formed between RCMs within the flanking introns and enhanced binding of RBP to the site of action. Moreover, ADARs-mediated circRNA regulation is most likely to be present in the same manner across different types of cancer cells, including breast, esophagus, liver, stomach and colon. We show that these ARcircs were not merely by-products of back-splicing, but functional molecules influencing tumorigenesis. These findings improve our understanding of the interaction between ADARs and circRNA biogenesis and its biological importance, particularly in the context of cancer.

Methods

Ethical statement. Our research complies with all relevant ethical regulations. All animal experiments were approved by the Institutional Animal Care and Use Committees of National University of Singapore (NUS; Singapore) with the protocol numbers R16-1644 and R20-1586. All human tissue samples used in this study were approved by the committees for ethics review at Sun Yat-Sen University, the National University of Singapore, and the National University Hospital, Singapore. Written informed consent for all patients were provided for the use of their clinical specimens for medical research.

Cell lines. EC109 cells were kindly provided by Professor TSAO, George Sai Wah (Director, Faculty Core Facility, Li Ka Shing Faculty of Medicine); SNU-398, HCT15 and MB231 cells were purchased from the American Type Culture Collection (ATCC); MKN28 cells were obtained from Japanese Collection of Research Bioresources Cell Bank. MB231 cells were maintained in Dulbecco's Modified Eagle Medium (Biowest) while the rest cell lines were maintained in Roswell Park Memorial Institute (RPMI) medium (Biowest), all supplemented with 10% fetal bovine serum (FBS; Thermo Scientific) and incubated at 37 °C with 5% CO₂.

Identification of circRNAs by circRNA sequencing. The expression level of ADAR1 or ADAR2 was modulated (forced overexpression or silencing) using a lenti-viral system in EC109 cells. RNA extraction was performed using RNeasy Mini kit (Qiagen) and subjected to rRNA depletion and RNase R treatment to digest all the linear RNAs. Samples were purified with Beckman RNAClean beads, the retrieved RNA was fragmented using divalent cations at an elevated temperature. Libraries were prepared using the TruSeq library preparation protocol (Illumina) using a modified protocol. Fragments were purified with Beckman AMPure beads and resolved in EB buffer for end repair and adding A at the 3' end. Y-adaptor was added afterwards. The product was then amplified to construct the cDNA library and sequenced on an Illumina HiSeq4000 or NovaSeq6000 instrument.

High-confidence circRNAs were identified by using an established in-house bioinformatics pipeline⁵⁹. First, raw reads were mapped to the reference human genome (*hg19*) by STAR (v2.5.2a)⁶⁰ with the chimeric junction reads option on (--chimSegmentMin 20). The gtf file from the GENCODE⁶¹ (*gencode.v27lift37.annotation.gtf*) was used for the gene and junction annotations. The expression of the circRNAs were quantified as read counts that map to the junctions, resulting in a total of 37,916 circRNAs having ≥ 1 read in at least 1 out of the 3 samples. To identify differentially expressed circRNAs in ADAR1/2 overexpressed and silenced samples compared to the control, the below criteria were applied:

1. Total reads in both (EV + ADAR OE) and (Scr + ADAR KD) ≥ 10 ,
2. Fold change in circRNA expression (ADAR versus EV) ≤ 0.5 (suppressing) or ≥ 2 (promoting), and
3. Fold change in circRNA expression (shADAR versus shScr) ≤ 0.5 (promoting) or ≥ 2 (suppressing).

As a result, we identified 1,406 circRNAs which are potentially regulated by ADARs (Supplementary Data 1).

To conduct a reliable comparison of the performance for circular RNA identification between our in-house pipeline and the other two commonly used benchmark methods- CIRI2 (v2.0.6) and CIRCexplorer2 (v2.3.0)^{26–28}, we considered annotated circRNAs wherever possible and required the junction positions to be identical (chr-start-end).

To identify editing-dependent ARcircs using circRNA-Seq datasets of the DeAD mutant or EV control-overexpressing EC109 cells (DeAD or EV), the following criteria were applied:

- (1) Total reads in (EV + DeAD) ≥ 10 ;
- (2) The resultant list of circRNAs was overlapped with ARcircs presented in Fig. 1b and Supplementary Data 1;
- (3) We obtained a total of 1,073 ARcircs fulfilling (1) and (2). We next defined “editing-dependent ARcircs” using the following filter criteria: there is no or minor change ($0.8 < \text{fold change} < 1.25$) in expression between the EV and DeAD-overexpressing cells or the pattern of change in expression upon overexpression of DeAD mutants is opposite to that of the corresponding wildtype ADAR1/2 when compared to the EV control. The remaining ARcircs which are 1) regulated by the wildtype or DeAD form of ADAR1 or/and ADAR2 in the same direction and 2) demonstrate ≥ 1.25 -fold change in expression upon overexpression of DeAD mutant versus EV control are defined as “editing-independent ARcircs” (Supplementary Data 3).

Identification of high-confidence A-to-I editing events from the total RNA sequencing data. A bioinformatics pipeline adapted from a previously published method⁶² was used to identify RNA editing events from total RNA-Seq data by using CSI NGS Portal (<https://csibioinfo.nus.edu.sg/csingsportal>)⁵⁹. For each sample, raw reads were mapped to the reference human genome (*hg19*) with a splicing junction database generated from transcript annotations derived from UCSC, RefSeq, Ensembl and GENCODE⁶¹ by using Burrows–Wheeler Aligner with default parameters (*bwa mem*, v0.7.17-r1188)⁶³. To retain high quality data, PCR duplicates were removed (*samtools markdup -r*, v1.9)⁶⁴ and the reads with mapping quality score < 20 were discarded. Junction-mapped reads were then converted back to the genomic-based coordinates. An in-house perl script was utilized to call the variants from samtools pileup data and the sites with at least two supporting reads were retained. The candidate events were filtered by removing the single nucleotide polymorphisms (SNPs) reported in different cohorts (1000 Genomes Project⁶⁵, NHLBI GO Exome Sequencing Project, and dbSNP v138⁶⁶) and excluding the sites within the first six bases of the reads caused by imperfect priming of random hexamer during cDNA synthesis. For the sites not located in Alu elements, the candidates within the four bases of a splice junction on the intronic side, and those residing in the homopolymeric regions and in the simple repeats were all removed. Candidate variants located in the reads that map to the non-unique regions of the genome by using BLAST-like alignment tool⁶⁷ were also excluded. At last, only A-to-G editing sites based on the strand information from the strand-specific RNA-Seq data were considered for all the downstream analyses. The genomic regions of the editing variants and the associated genes were annotated by using ANNOVAR (v2018)⁶⁸ with the *refGene* table. As reported by us previously¹⁵, ADARs-mediated global editing changes has been confirmed by analysing the same RNA-Seq dataset of ADARs OE or KD cells.

To identify high-confidence editing events, the editing sites were required to be supported by ≥ 10 reads in ≥ 1 sample, and ADAR1/2-overexpressing samples to result in more than 10% change in the editing level compared to the control. This resulted in 41,151 high-confidence editing sites from the RNA-Seq of our EC109 cell lines used in further analyses. To analyze the sequence preference for the neighbour nucleotide surrounding editing sites, the sequence context of these editing sites was extracted using “bedtools getfasta”, i.e. editing site plus 2 neighbour nucleotides on either side in a strand-specific manner. Then the nucleotide frequencies were converted to a position probability matrix and the sequence logo was plotted by using the “seqLogo” package (v1.56.0).

In silico prediction of RCMs within flanking intronic sequences. To identify RCMs, we adopted a previously published method⁸. We used circRNAs exhibiting > 2 or < 0.5 -fold change upon ADAR1 or ADAR2 overexpression with total reads in EV and ADAR1/2 more than 50 in circRNA-seq for RCM identification, which results in 1,118 circRNAs in total (Supplementary Data 2). First, a BLAST⁶⁹ alignment was performed for each intron pair flanking circRNA junctions to identify all the potential candidates. The RCM with the top BLAST score for each circRNA was retained yielding 1,043 circRNAs with at least 1 RCM. The circRNAs with short flanking introns (< 1500 bp) were further removed leaving 886 circRNAs as the final list. The intronic region spanning 1500 bp upstream and 1500 bp downstream of the circRNA back-splicing junctions was considered to plot the distribution of RCM coverage, which is defined as the sum of the number of top-scoring RCMs at each base across all 886 circRNAs.

To investigate if there is an enrichment of RNA editing events within RCMs as previously reported⁸, 41,151 high-confidence RNA editing sites identified from the RNA-Seq data were overlapped with the flanking introns of 886 circRNAs identified from the circRNA-Seq. The density plots showing the distribution of editing sites were generated within the same region as the plot illustrating the distribution of the RCM coverage mentioned above, but we used the density distribution rather than coverage since each editing site indicates a single nucleotide variation whereas RCMs are regions of variable length. In addition, we have also analyzed the distribution of 2,576,459 A-to-I RNA editing sites downloaded from RADAR database³⁶ within the same regions.

Plasmid constructions. Minigene fragments were amplified from human placenta genomic DNA (Sigma) (for intronic sequences) or EC109 cDNA (for exonic sequences) using PrimeSTAR Max DNA polymerase (Clontech) with overlapped

primers and ligated into one piece of DNA, followed by ligation into pcDNA3.1+ vector. KAPA HiFi polymerase (KAPA Biosystems) was used to introduce point mutations into minigene using primers with corresponding mutation(s).

Overexpression plasmids were obtained by cloning coding sequences of protein, which were amplified by PrimeSTAR Max DNA polymerase (Clontech), into pLenti6 vector. ADARs-targeting short hairpin RNAs (shRNAs) were designed using RNAi Platform (Broad Institute) and were cloned into pLKO.1_puro plasmid using AgeI and EcoRI restriction sites.

CasRX system (pXR001: EF1a-CasRx-2A-EGFP and pXR003: CasRx gRNA cloning backbone) was a gift from Patrick Hsu (pXR001: Addgene plasmid #109049, <http://n2t.net/addgene:109049>, RRID:Addgene_109049; pXR003: Addgene plasmid #109053, <http://n2t.net/addgene:109053>, RRID:Addgene_109053)^{51,52}. Guide RNA (gRNA) sequences were designed using sequence of circRNAs around BSJ with a length of 21 bp and cloned into pXR003 plasmid using BbsI restriction sites.

Plasmids transfection. A total of 2 µg plasmids (protein overexpression construct, shRNA plasmids or minigene plasmids) or a mixture of 1 µg pXR001 and 1 µg pXR003 plasmids were transfected into cells a well of 12-well plate using Lipofectamine 2000 (Invitrogen) with a ratio of 1:2 (DNA:reagent).

RNA extraction and RT-qPCR. RNA was extracted using RNeasy mini kit (Qiagen) with on column treatment of DNaseI. cDNA was synthesized using Advantage RT-for-PCR kit (Clontech) with random hexamer primers and subsequently qPCR was performed with GoTaq DNA polymerase (Promega). Fold change was calculated by $2^{-\Delta\Delta C_t}$, $\Delta C_t = C_{t\text{target}} - C_{t\text{actin}}$, $\Delta\Delta C_t = \Delta C_{t\text{sample}} - \text{average}\Delta C_{t\text{control}}$. Primers used in RT-qPCR are listed in Supplementary Data 5.

Western blot. Cells were lysed with RIPA buffer (Sigma) supplemented with 1x cComplete EDTA-free protease inhibitor cocktail (Roche) and concentrations of total protein were quantified using Bradford assay (Biorad). 10% SDS-PAGE were used to separate proteins, followed by transferred onto polyvinylidene difluoride membranes (Millipore) and incubated with primary antibodies (1:1,000 dilution) overnight at 4 °C and secondary antibodies (1:10,000 dilution) at room temperature for 1 h. Enhanced chemiluminescence (GE Healthcare) was used to visualize the blots. Primary antibodies used in this study are as listed: anti-PTBP1 (Abcam, ab133734), anti-TDP43 (Proteintech, 10782-2-AP), anti-ADAR1 (Abcam, ab88574), anti-FLAG-HRP (Sigma, A8592), anti-β actin HRP (Santa Cruz Biotechnology, sc-47778HRP), anti-mouse IgG HRP-linked (Cell Signaling Technology Cat# 7076, RRID:AB_330924), anti-rabbit IgG HRP-linked (Cell Signaling Technology Cat#7074, RRID:AB_2099233). ImageJ (1.51J8) was used to measure band density of blots.

RNA immunoprecipitation (RIP). A 10-cm dish of EC109 cells was transfected with 10 µg of FLAG, FLAG-ADAR1, plenti6, or plenti6-ADAR1 plasmid individually. After 48 h culturing, cells were collected and lysed in buffer containing 50 mM Tris, pH7.5, 150 mM NaCl, 1 mM EDTA and 1% Triton X-100 supplemented with cComplete protease inhibitor (Roche) and SUPERase-In RNase Inhibitor (Invitrogen). For FLAG-RIP, lysates were then incubated with anti-FLAG M2 magnetic beads (Sigma) overnight at 4 °C with rotation followed by six times of washing with 1× TBS buffer (0.5 M Tris, 1.5 M NaCl). For PTBP1 RIP, the lysates were pre-cleared using 50 µL protein A-agarose suspension (Roche) at 4 °C for overnight. A total of 2 µL anti-PTBP1 antibody was then added into the lysate and incubate at 4 °C for 1 h, followed by overnight incubation at 4 °C with the addition of 50 µL protein A-agarose suspension. The beads were then washed with washing buffer [150 mM NaCl, 0.04 U/µL SUPERase-In RNase Inhibitor (Invitrogen)] for 6 times with each time for 10 min at 4 °C. 10% of beads was used for protein elution while the rest was subjected to RNA extraction using RNeasy miniprep kit (Qiagen). Extracted RNA was reverse-transcribed using Advantage RT-for PCR kit (Clontech) with random hexamer and subsequently qPCR was performed. Input indicates 1% of the total cell lysate. $\%input = 2^{-\Delta C_t} \times 100\%$; $\Delta C_t = C_{tRIP} - C_{tinput}$ [dilution factor]. Sequences of primers are listed in Supplementary Data 5.

Analysis of editing frequency by TA cloning. The region containing editing site(s) was amplified using PCR method, followed by purification using PCR product purification kit (Qiagen). Purified PCR products were then ligated into pGEM-T easy vector (Promega) using T4 quick ligase (Promega). A total of 20–28 individual plasmids were sent for Sanger sequencing. The number of unedited ‘A’ or edited ‘G’ clones was counted, followed by the calculation of the percentage of edited clones by ‘G/(A + G)’. The percentage of edited clones (a readout of ‘editing frequency’) was determined and shown by pie chart (represented by red slice).

In vitro transcription. PCR was used to generate DNA template for in vitro transcription with a primer pair containing T7 promoter sequences (5'-CGAAATTAATACGACTCATATAGG at forward) and sequence of interest. DNA template was subjected to in vitro transcription with RiboMAX™ Large Scale RNA Production Systems (Promega) according to the manufacturer's protocol. Synthesized RNA probes were then purified by RNeasy mini kit (Qiagen).

Native PAGE analysis. Each RNA probe (50 pmol) was first dephosphorylated using rSAP (New England Biolabs), followed by ³²P labelling with γ-³²P-ATP (Perkin Elmer) and T4 PNK (New England Biolabs). Labelled probes were then purified by G25 column (GE healthcare). 0.5 pmol labelled RNA probe were incubated at 95 °C for 5 min and then gradually cool down to help form secondary structure. Probes were then loaded on 4% or 8% native polyacrylamide gel, followed by gel drying and gel exposure to BioMax® MS film (Carestream Kodak). Sequences of probes are listed in Supplementary Data 5. For *circCHEK2*, position of probe with lowest migration rate was labelled as 0, while position of probe with the highest migration rate was labelled as 1. The related migration rate of each probe was measured by: (Distance between the probe with 0)/(Distance between 1 and 0).

Whole cell extraction. Whole cell extraction was performed with the kit (Active motif) according to the manufacturer's protocol.

RNA pull-down assay. RNA probe was generated as discussed above but only with an addition of 3'-aptamer at reverse primer. For each reaction, 50 µL Dynabeads MyOne C1 (Invitrogen) was used to incubate with 25 µg RNA probe in 300 µL binding buffer (100 mM NaCl, 10 mM MgCl₂, 50 mM Hepes, 0.5% Igeal CA-630, and pH 7.4) for 30 min at 4 °C with rotation for probe binding to the beads. Beads were then washed washing buffer (250 mM NaCl, 10 mM MgCl₂, 50 mM Hepes, 0.5% Igeal CA-630, and pH 7.4), for 10 min at 4 °C for three times. 1 mg whole cell extract was supplemented with 4 µL 10 mg/ml yeast tRNA (Invitrogen) and SUPERase In (Invitrogen). The mixture was added to the beads and topped up to 300 µL with washing buffer, followed by incubation for 30 min at 4 °C with rotation. After washing three times, beads were subjected to 2× Laemmli buffer (Sigma) at 95 °C to elute proteins followed by western blot (WB). Sequences of probes are listed in Supplementary Data 5.

RBPmap analysis. The ±10nt sequences surrounding editing sites were retrieved. The sequence with A or G at each editing site was used as input for RBP motif analysis using RBPmap⁴⁶. Briefly, for each sequence, if the RBP binding affinity (Z-score) is changed because of an A-to-G mutation, the RBP binding motif is affected by the editing site. The number of circRNAs which have altered RBP binding sites on flanking introns due to editing was calculated using our in-house script.

Foci formation assay. A total of 1,000 cells were seeded in each well of 6-well plates after transfection and incubated at 37 °C for 7–9 days. Plates were stained with crystal violet solution (0.1% crystal violet, 20% methanol in PBS) to visualize colonies. Colonies were calculated using OpenCFU⁷⁰. Triplicate independent experiments were conducted with technical triplicates.

Soft agar assay. A total of 2,000 cells (EC109) or 5,000 cells (SNU398) were seeded into 0.4% low-melting agarose (Lonza Rockland) in each well of 6-well plates with 0.6% low-melting agarose at the bottom. Plates were incubated at 37 °C for 2 weeks and stained with crystal violet solution (0.05% crystal violet, 20% methanol in PBS) for visualization. Colonies were calculated using OpenCFU⁷⁰. Triplicate independent experiments were conducted with technical triplicates.

In vivo tumorigenicity assay. Four-six-weeks-old male and female NOD scid gamma (NSG) mice (The Jackson Laboratory, RRID:IMSR_JAX:005557) were maintained in pathogen-free (SPF) facility in NUS Comparative Medicine Department. Less than 5 mice with same sex were housed in a cage at 20–25 °C and 50% humidity with a 12 h light/dark cycle. A total of 6 female mice (EC109) or 4 female + 1 male mice (SNU398) were used to subcutaneously inject with one million (EC109) or two million (SNU398) cells into the right and left flanks. Tumor growth was monitored by measuring the length and width at indicated day points. Tumor volume was determined by the formula: $0.5 \times \text{length} \times \text{width}^2$. All animal experiments were approved by and performed in accordance with the Institutional Animal Care and Use Committees of National University of Singapore (NUS; Singapore).

Human tissues. A total of 17 matched pairs of primary HCC and adjacent non-tumor (NT) tissues were obtained from the Sun Yat-Sen University Cancer Centre (Guangzhou, China), between 2002 and 2007. A total of 20 matched pairs of primary CRC and adjacent NT colon tissues were obtained from the National University Hospital, Singapore.

Quantification and statistical analysis. All quantitative data represent the mean ± SD. Statistical significance was assessed with paired or unpaired two-tailed Student's *t*-tests using Prism 8 (GraphPad software). For all figures: n.s., not significant; *, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The circRNA-Seq data generated in this study have been deposited in the Gene Expression Omnibus (GEO) under accession code [GSE164681](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164681). The EC109 RNA-Seq data have been published previously and is also available at GEO under accession [GSE131658](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131658)¹⁵. Human genome reference hg19 was obtained from GENCODE. A-to-I editing sites from RADAR database were obtained from <http://RNAedit.com>. Information on SNPs was obtained from 1000 Genomes Project (<https://www.internationalgenome.org/>), NHLBI GO Exome Sequencing Project, and dbSNP v138. The data supporting the findings of this study are available from the corresponding authors upon reasonable request. Source data for the figures and supplementary figures are provided as a Source Data file. Source data are provided with this paper.

Code availability

The codes used in the data analysis are available in Supplementary Software.

Received: 9 April 2021; Accepted: 15 February 2022;

Published online: 21 March 2022

References

- Hsu, M. T. & Coca-Prados, M. Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature* **280**, 339–340 (1979).
- Chen, L., Huang, C., Wang, X. & Shan, G. Circular RNAs in eukaryotic cells. *Curr. Genomics* **16**, 312–318 (2015).
- Kristensen, L. S. et al. The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* **20**, 675–691 (2019).
- Lee, E. C. S. et al. The roles of circular RNAs in human development and diseases. *Biomed. Pharmacother.* **111**, 198–208 (2019).
- Ashwal-Fluss, R. et al. circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* **56**, 55–66 (2014).
- Conn, S. J. et al. The RNA binding protein quaking regulates formation of circRNAs. *Cell* **160**, 1125–1134 (2015).
- Zhang, X. O. et al. Complementary sequence-mediated exon circularization. *Cell* **159**, 134–147 (2014).
- Ivanov, A. et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.* **10**, 170–177 (2015).
- Aktas, T. et al. DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature* **544**, 115–119 (2017).
- Li, X. et al. Coordinated circRNA Biogenesis and Function with NF90/NF110 in Viral Infection. *Mol. Cell* **67**, 214–227 e7 (2017).
- Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev. Biochem.* **79**, 321–349 (2010).
- Zinshteyn, B. & Nishikura, K. Adenosine-to-inosine RNA editing. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **1**, 202–209 (2009).
- Shi, L. et al. Circular RNA expression is suppressed by androgen receptor (AR)-regulated adenosine deaminase that acts on RNA (ADAR1) in human hepatocellular carcinoma. *Cell Death Dis.* **8**, e3171 (2017).
- Wong, S. K., Sato, S. & Lazinski, D. W. Substrate recognition by ADAR1 and ADAR2. *RNA* **7**, 846–858 (2001).
- Tang, S. J. et al. Cis- and trans-regulations of pre-mRNA splicing by RNA editing enzymes influence cancer development. *Nat. Commun.* **11**, 799 (2020).
- Hsiao, Y. E. et al. RNA editing in nascent RNA affects pre-mRNA splicing. *Genome Res.* **28**, 812–823 (2018).
- Chen, Y. T. et al. Tumor-associated intronic editing of HNRPLL generates a novel splicing variant linked to cell proliferation. *J. Biol. Chem.* **293**, 10158–10171 (2018).
- Fumagalli, D. et al. Principles governing A-to-I RNA editing in the breast cancer transcriptome. *Cell Rep.* **13**, 277–289 (2015).
- Chan, T. H. et al. ADAR-mediated RNA editing predicts progression and prognosis of gastric cancer. *Gastroenterology* **151**, 637–650.e10 (2016).
- Shigeyasu, K. et al. AZIN1 RNA editing confers cancer stemness and enhances oncogenic potential in colorectal cancer. *JCI Insight* **3** (2018).
- Chen, L. et al. Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. *Nat. Med.* **19**, 209–216 (2013).
- Song, Y. et al. RNA editing mediates the functional switch of COPA in a novel mechanism of hepatocarcinogenesis. *J. Hepatol.* **74**, 135–147 (2021).
- Qin, Y. R. et al. Adenosine-to-inosine RNA editing mediated by ADARs in esophageal squamous cell carcinoma. *Cancer Res.* **74**, 840–851 (2014).
- Han, J. et al. Suppression of adenosine-to-inosine (A-to-I) RNA editome by death associated protein 3 (DAP3) promotes cancer progression. *Sci. Adv.* **6**, eaba5136 (2020).
- Fu, L. et al. RNA editing of SLC22A3 drives early tumor invasion and metastasis in familial esophageal cancer. *Proc. Natl Acad. Sci. USA* **114**, E4631–E4640 (2017).
- Chen, L. et al. The bioinformatics toolbox for circRNA discovery and analysis. *Brief. Bioinform.* **22**, 1706–1728 (2021).
- Zhang, X. O. et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* **26**, 1277–1287 (2016).
- Gao, Y., Zhang, J. & Zhao, F. Circular RNA identification based on multiple seed matching. *Brief. Bioinform.* **19**, 803–810 (2018).
- Liscovitch, N., Bazak, L., Levanon, E. Y. & Chechik, G. Positive correlation between ADAR expression and its targets suggests a complex regulation mediated by RNA editing in the human brain. *RNA Biol.* **11**, 1447–1456 (2014).
- Chung, H. et al. Human ADAR1 prevents endogenous RNA from triggering translational shutdown. *Cell* **172**, 811–824 e14 (2018).
- Yang, C. C. et al. ADAR1-mediated 3' UTR editing and expression control of antiapoptosis genes fine-tunes cellular apoptosis response. *Cell Death Dis.* **8**, e2833 (2017).
- Athanasias, A., Rich, A. & Maas, S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.* **2**, e391 (2004).
- Li, J. B. et al. Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324**, 1210–1213 (2009).
- Wang, Y., Park, S. & Beal, P. A. Selective recognition of RNA substrates by ADAR deaminase domains. *Biochemistry* **57**, 1640–1651 (2018).
- Lehmann, K. A. & Bass, B. L. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **39**, 12875–12884 (2000).
- Ramaswami, G. & Li, J. B. RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.* **42**, D109–D113 (2014).
- Kapoor, U. et al. ADAR-deficiency perturbs the global splicing landscape in mouse tissues. *Genome Res.* **30**, 1107–1118 (2020).
- Lai, F., Drakas, R. & Nishikura, K. Mutagenic analysis of double-stranded RNA adenosine deaminase, a candidate enzyme for RNA editing of glutamate-gated ion channel transcripts. *J. Biol. Chem.* **270**, 17098–17105 (1995).
- Valente, L. & Nishikura, K. RNA binding-independent dimerization of adenosine deaminases acting on RNA and dominant negative effects of nonfunctional subunits on dimer functions. *J. Biol. Chem.* **282**, 16054–16061 (2007).
- Qi, L. et al. An RNA editing/dsRNA binding-independent gene regulatory mechanism of ADARs and its clinical implication in cancer. *Nucleic Acids Res.* **45**, 10436–10451 (2017).
- Davis, C. A. et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).
- Wu, C.-S. et al. Widespread interaction between ADAR1 and transcriptional byproducts. *bioRxiv*, 870782 (2019).
- Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
- Bahn, J. H. et al. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res.* **22**, 142–150 (2012).
- Woodson, S. A. & Koculi, E. Analysis of RNA folding by native polyacrylamide gel electrophoresis. *Methods Enzymol.* **469**, 189–208 (2009).
- Paz, I., Kost, I., Ares, M. Jr., Cline, M. & Mandel-Gutfreund, Y. RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.* **42**, W361–W367 (2014).
- Yang, B. et al. PTBP1 induces ADAR1 p110 isoform expression through IRES-like dependent translation control and influences cell proliferation in gliomas. *Cell Mol. Life Sci.* **72**, 4383–4397 (2015).
- Errichelli, L. et al. FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nat. Commun.* **8**, 14741 (2017).
- Stagsted, L. V. W., O'Leary, E. T., Ebbesen, K. K. & Hansen, T. B. The RNA-binding protein SFPQ preserves long-intron splicing and regulates circRNA biogenesis in mammals. *eLife* **10**, e63088 (2021).
- Fei, T. et al. Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. *Proc. Natl Acad. Sci. USA* **114**, E5207–E5215 (2017).
- Konermann, S. et al. Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173**, 665–676 e14 (2018).
- Li, S. et al. Screening for functional circular RNAs using the CRISPR-Cas13 system. *Nat. Methods* (2020).
- Han, A. et al. De novo prediction of PTBP1 binding and splicing targets reveals unexpected features of its RNA recognition and function. *PLoS Comput. Biol.* **10**, e1003442 (2014).
- Hong, H. et al. Bidirectional regulation of adenosine-to-inosine (A-to-I) RNA editing by DEAH box helicase 9 (DHX9) in cancer. *Nucleic Acids Res.* **46**, 7953–7969 (2018).

55. Quinones-Valdez, G. et al. Regulation of RNA editing by RNA-binding proteins in human cells. *Commun. Biol.* **2**, 19 (2019).
56. Xu, L. D. & Öhman, M. ADAR1 Editing and its Role in Cancer. *Genes (Basel)* **10** (2018).
57. Vo, J. N. et al. The landscape of circular RNA in cancer. *Cell* **176**, 869–881 e13 (2019).
58. Tian, N. et al. A structural determinant required for RNA editing. *Nucleic Acids Res.* **39**, 5669–5681 (2011).
59. An, O. et al. CSI NGS portal: an online platform for automated NGS data analysis and sharing. *Int. J. Mol. Sci.* **21** (2020).
60. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
61. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
62. Ramaswami, G. et al. Identifying RNA editing sites using RNA sequencing data alone. *Nat. Methods* **10**, 128–132 (2013).
63. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
64. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
65. Genomes-Project, C. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
66. Sherry, S. T. et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
67. Kent, W. J. BLAT-the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
68. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
69. Camacho, C. et al. BLAST + : architecture and applications. *BMC Bioinforma.* **10**, 421 (2009).
70. Geissmann, Q. OpenCFU, a new free and open-source software to count cell colonies and other circular objects. *PLoS ONE* **8**, e54072 (2013).

Acknowledgements

We thank and acknowledge Prof. Xin-Yuan Guan (The University of Hong Kong, Hong Kong, China) for providing RNA samples of HCC cases. L. C. was supported by National Research Foundation Singapore; Singapore Ministry of Education under its Research Centres of Excellence initiative; Singapore Ministry of Education's Tier 2 Grants [MOE2018-T2-1-005 and MOE2019-T2-2-008]; NMRC Clinician Scientist-Individual Research Grant (CS-IRG, project ID: MOH-000214); and Singapore Ministry of Education's Tier 3 Grants [MOE2014-T3-1-006].

Author contributions

L.C. conceived and supervised the study. L.C. and H.S. designed and performed the experiments. H.Y., O.A. and X.R. conducted all the bioinformatics analyses. Y.S. helped in conducting primary tumor sample related experiments. K.T. provided the matched pairs of CRC and NT samples. Y.S. and X.K. assisted in conducting mouse-related experiments. X. K. helped with data analysis. S.J.T., J.H., X.K., D.J.T.T., V.H.E.N., F.B.M., P.P., and K.W.L. provided insightful suggestions and experimental materials. H.S. and O.A. wrote the manuscript. L.C. edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-29138-2>.

Correspondence and requests for materials should be addressed to Leilei Chen.

Peer review information *Nature Communications* thanks Trees-Juen Chuang, Leng Han and the other, anonymous, reviewer for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.












© The Author(s) 2022

ARTICLE


<https://doi.org/10.1038/s41467-022-29400-7>

OPEN

Multilayered control of splicing regulatory networks by DAP3 leads to widespread alternative splicing changes in cancer

Jian Han ^{1✉}, Omer An ¹, Xi Ren¹, Yangyang Song ¹, Sze Jing Tang ¹, Haoqing Shen ¹, Xinyu Ke¹, Vanessa Hui En Ng ¹, Daryl Jin Tai Tay ¹, Hui Qing Tan ², Dennis Kappei ^{1,3,4}, Henry Yang ¹ & Leilei Chen ^{1,4,5✉}

The dynamic regulation of alternative splicing requires coordinated participation of multiple RNA binding proteins (RBPs). Aberrant splicing caused by dysregulation of splicing regulatory RBPs is implicated in numerous cancers. Here, we reveal a frequently overexpressed cancer-associated protein, DAP3, as a splicing regulatory RBP in cancer. Mechanistically, DAP3 coordinates splicing regulatory networks, not only via mediating the formation of ribonucleoprotein complexes to induce substrate-specific splicing changes, but also via modulating splicing of numerous splicing factors to cause indirect effect on splicing. A pan-cancer analysis of alternative splicing across 33 TCGA cancer types identified DAP3-modulated mis-splicing events in multiple cancers, and some of which predict poor prognosis. Functional investigation of non-productive splicing of *WSB1* provides evidence for establishing a causal relationship between DAP3-modulated mis-splicing and tumorigenesis. Together, our work provides critical mechanistic insights into the splicing regulatory roles of DAP3 in cancer development.

¹Cancer Science Institute of Singapore, National University of Singapore, Singapore 117599, Singapore. ²Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117549, Singapore. ³Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117596, Singapore. ⁴NUS Centre for Cancer Research, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117594, Singapore. ⁵Department of Anatomy, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117594, Singapore.

✉email: csihj@nus.edu.sg; polly_chen@nus.edu.sg

RNA-binding proteins (RBPs) participate in every aspect of RNA processing and regulation, from RNA transcription, splicing, cleavage, and polyadenylation to RNA modifications, degradation, transportation, and translation. Each of these processes involves multiple RBPs which form regulatory networks to execute the dynamic control of the transcriptome and proteome complexity. The spatiotemporal regulation of RNA processing by RBPs is vital for normal development and physiology, so that any disruption in RNA processing may lead to human diseases¹. Aberrant RNA splicing is frequently observed in almost all types of cancers and each cancer hallmark may be affected by aberrant splicing². Alternative splicing in cancer cells can switch a gene from a tumor-suppressive or non-oncogenic isoform to an oncogenic isoform. For instance, mutually exclusive alternative splicing of the *PKM* pre-mRNA generates either an exon 9 inclusion isoform *PKM1*, or an exon 10 inclusion isoform *PKM2*. Through upregulation of heterogeneous nuclear ribonucleoproteins (hnRNPs) that repress *PKM* exon 9 inclusion, cancer cells express high level of the *PKM2* isoform to maintain aerobic glycolysis³. Alternative splicing is also a prevalent mechanism to regulate gene expression by altering the stability and degradation of mRNA transcripts. Inclusion of “poison exons” that contain premature termination codons can trigger nonsense-mediated decay (NMD), a translation-dependent RNA surveillance process that degrades mRNAs⁴. For example, mutation of *SRSF2* alters its RNA-binding recognition and thereby promotes mis-splicing and NMD of *EZH2*, results in impaired hematopoietic differentiation in myelodysplastic syndromes (MDS) development⁵.

Next-generation sequencing reveals that aberrant splicing often occurs at a genome-wide scale in cancer cells, with numerous genes alternatively spliced to drive cancer initiation and progression. Although *cis*-acting mutations can lead to deficient pre-mRNA splicing by inactivating a splice site within the pre-mRNA, most cancer-driven or associated splicing events are resulted from altered expression and/or function of splicing factors and regulators due to *trans*-acting mutations or other genomic and epigenomic mechanisms. For example, genomic amplification of a proto-oncogene and splicing factor *SRSF1* is observed in many cancer types, which leads to a malignant transformation of rodent fibroblasts to sarcomas by inducing an oncogenic isoform of *RPS6KB1*⁶. In addition to identifying changes in splicing regulatory *cis*-elements in RNA, it is therefore critical to delineate the precise mechanisms underpinning misregulation of *trans*-acting splicing regulators that bind to RNA and their functional relevance to cancer.

RBPs can form complex and interlaced RNA processing regulatory networks via direct binding to RNA or protein–protein interactions with other RBPs. Different RBPs may share a high degree of sequence recognition similarity and therefore work in a cooperative or competitive mode^{7,8}. Many RBPs, and particularly splicing factors, auto- or cross-regulate expression of themselves or other RBPs through alternative splicing coupled with NMD (AS-NMD) respectively, increasing the complexity of splicing regulatory networks^{9,10}. In this study, we identify death-associated protein 3 (DAP3) as an RNA splicing regulator and uncovered its multilayered control of cancer transcriptome. DAP3 binds extensively to endogenous RNAs in vivo and demonstrates an RNA-binding preference for coding sequences (CDS) with a significant enrichment in splicing-associated motifs such as the 5' splice site consensus sequence and SR-protein hexamer motif. DAP3 can not only facilitate the association of splicing factor proline and glutamine rich (SFPQ) and Non-POU domain containing octamer binding (NONO) with target RNAs for splicing modulation, but also modulate the expression of numerous splicing factors via AS-NMD, leading to global changes in splicing. Notably, such widespread splicing changes modulated by DAP3 can be observed in multiple cancer types and are of

clinical relevance and prognostic values. By further investigations of functional importance of DAP3-driven splicing events in cancer using in vitro and in vivo models, our study provides critical mechanistic insights into the role of DAP3 in cancer development as a critical regulator of RNA splicing.

Results

RNA-binding landscape of DAP3 indicates its potential role in RNA splicing. It has previously been shown that DAP3 interacts with RNA editing enzymes adenosine deaminases acting on RNA (ADARs) and functions as a potent repressor of adenosine-to-inosine (A-to-I) RNA editing in cancer cells¹¹. Surprisingly, however, DAP3 does not directly bind to the regions proximal to editing sites¹¹ and the RNA-binding landscape of DAP3 remains largely unclear so far. Herein, we performed a comprehensive transcriptome-wide analysis of RNA-binding sites of DAP3 using our eCLIP-Seq data¹¹ (GEO accession number: GSE144318) and found that DAP3 binds extensively to endogenous RNAs in vivo with 9699 genes showed DAP3-bound peaks in both biological duplicates (Supplementary data 1). No presence of four selected RBPs including SFPQ, NONO, U2AF35, and U2AF65 in the DAP3 eCLIP elutes confirmed that our eCLIP experiment could specifically pull down DAP3-bound RNAs (Supplementary Information; Supplementary Fig. 1a, b). Further analysis showed that ~70% of DAP3-bound peaks were mapped to exons (37,624 out of 53,596 peaks and 45,575 out of 64,555 peaks in biological duplicates DAP3-1 and DAP3-2, respectively) (Fig. 1a; Supplementary data 1). Of note, DAP3 demonstrated a strong RNA-binding preference for CDS appearing slightly skewed towards 5' end of CDS (Fig. 1b). Moreover, although DAP3 binds predominantly to protein-coding RNAs (~80% of eCLIP peaks), ~6% of eCLIP peaks was mapped to non-coding RNAs (ncRNAs), such as two well-known ncRNAs *NEAT1* and *MALAT1* (Fig. 1c). We also observed that DAP3 can bind to multiple regions of one gene, such as 5'UTR, first intron and CDS of *ARHGEF16* gene and different sites within *TARDBP* 3'UTR (Fig. 1c). We also conducted a comparative analysis of gene structure between DAP3-bound and unbound genes and did not observe any significant difference in the number of exons or isoforms between two groups of genes (Supplementary Fig. 2a, b). Further, the de novo motif analysis of DAP3-binding sites by HOMER¹² showed the top three enriched DAP3 RNA-binding motifs: GAAGAAGAU, C(A/U)(A/U)C repeats, and AGGUAAGU (Fig. 1d). The most enriched motif GAAGAAGAU contains a representative GAAGAA hexamer, which is an exonic splicing enhancer (ESE) essential for constitutive and alternative splicing¹³. The presence of AGGUAA GU motif, the 5' splice site consensus sequence in vertebrates¹⁴, indicates DAP3 could bind to the exon–intron splice junctions (Supplementary Fig. 3a, b). We next selected two target transcripts *SCYL1* and *ATAD3A* for the RNA electrophoretic mobility shift assay (REMSA), due to the presence of top ranked DAP3 RNA-binding motifs in their eCLIP peaks (Supplementary Fig. 3c). We confirmed the direct binding of DAP3 to these target RNAs in vitro (Fig. 1e).

Gene ontology (GO) enrichment analysis revealed that DAP3 tends to bind and affect genes associated with RNA processing (e.g., splicing and degradation), DNA metabolic process and mitotic cell cycle regulation (Fig. 1f). All these observations indicate a potential role of DAP3 in modulating RNA splicing.

DAP3 modulates widespread alternative splicing. To investigate the role of DAP3 in RNA splicing modulation, we quantified splicing changes caused by DAP3 depletion in two esophageal squamous cell carcinoma (ESCC) cell lines EC109 and KYSE180, from previously published total RNA-Seq data¹¹ using rMATS

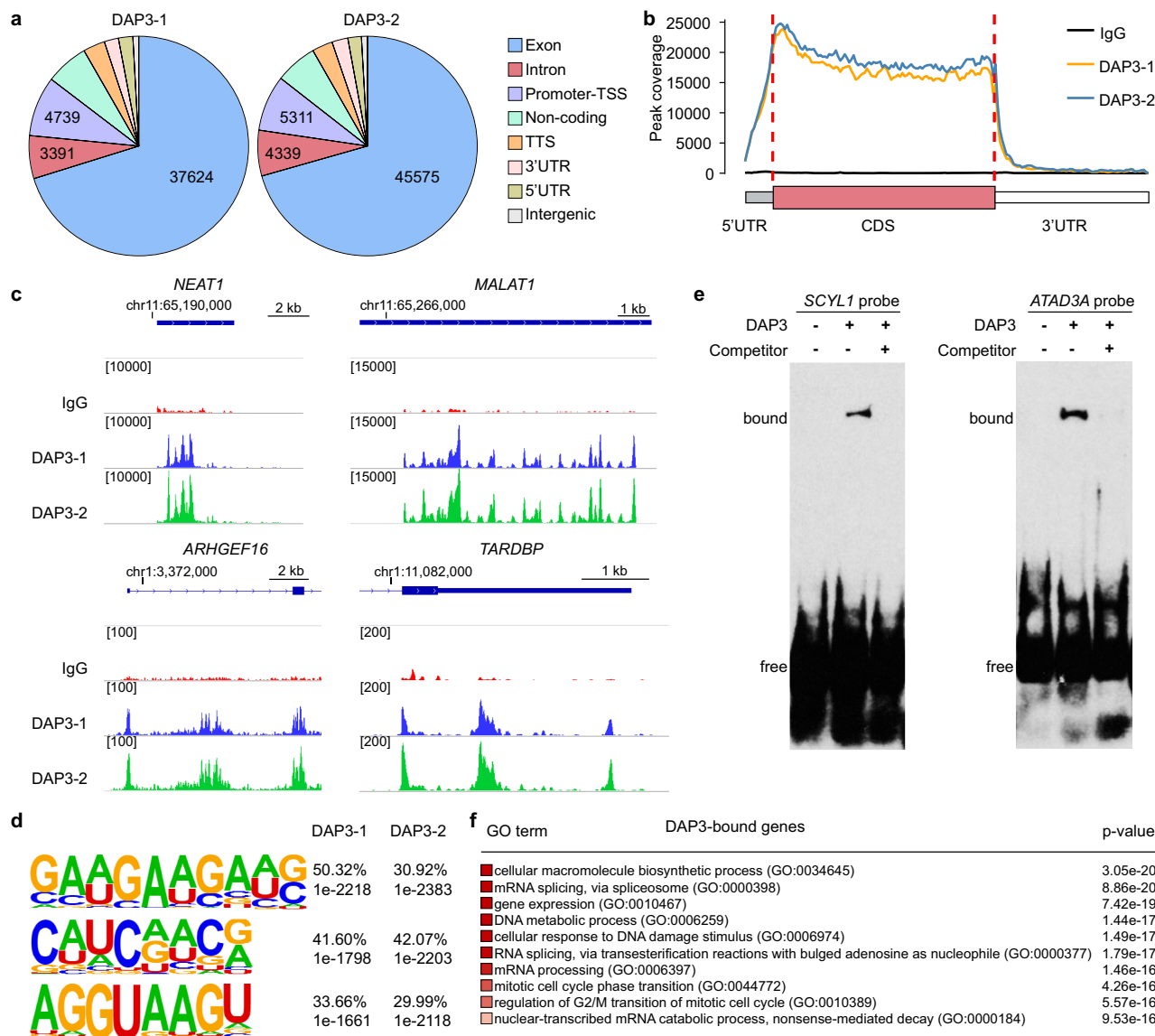


Fig. 1 RNA-binding landscape of DAP3 by eCLIP-Seq. **a** The distribution of the DAP3 eCLIP peaks in the human genome from biological duplicates (DAP3-1 and DAP3-2). TSS transcription start site, TTS transcription termination site, UTR untranslated region. **b** The binned DAP3 eCLIP peak coverage across all expressed genes in EC109 cells. The 5'UTR, CDS, and 3'UTR of each gene are split into 13, 100, and 70 bins, respectively. CDS coding sequence. **c** Integrated genome viewer (IGV) browser tracks of the DAP3 eCLIP peaks spanning the genomic loci of *NEAT1* and *MALAT1*, 5' region of *ARHGEF16*, and 3' region of *TARDBP* gene. Detailed information about all significantly enriched eCLIP peaks can be found in Supplementary data 1. **d** Top three most significantly enriched de novo sequence motifs in the DAP3-binding peaks using HOMER¹². The percentage of peaks containing the discovered motifs and the *p*-values of the motifs calculated by binomial test against the random genomic background were shown. **e** REMSA analysis of the binding of DAP3 to two RNA probes that were generated from DAP3-binding sequences within *SCYL1* and *ATAD3A* transcripts. A 100-fold molar excess of nonbiotinylated RNA probes was used as competitor. **f** Gene ontology analysis of DAP3-bound genes. The top 10 most significantly enriched biological processes are shown. The significance of enrichment for GO sets were evaluated by the WebGestalt³⁹ portal default hypergeometric test. Source data are provided in Source Data file.

pipeline¹⁵. Compared to the scramble controls, alternative splicing events in *DAP3*-depleted cells [$|\Delta$ percent spliced-in (Δ PSI)| $\geq 10\%$, false discovery rate (FDR) < 0.05 , and splice junction read coverage ≥ 20] were defined as DAP3-modulated splicing events. Intriguingly, we identified 7400 and 11,820 DAP3-modulated splicing events in 3262 and 4582 genes in EC109 and KYSE180 cells, respectively (Fig. 2a; Supplementary data 2 and 3). Approximately half of the DAP3-modulated splicing events are skipped exon (SE), followed by mutually exclusive exon (MXE; $\sim 20\%$), alternative 5' splice site (A5SS; $\sim 10\%$), alternative 3' splice site (A3SS; $\sim 10\%$), and intron retention (IR; $\sim 10\%$) in both cell lines (Fig. 2a; Supplementary data 2 and 3).

DAP3 was inclined to repress exon skipping (58% and 71% of DAP3-repressed vs. 42% and 29% DAP3-promoted SE events in EC109 and KYSE180, respectively), intron retention (72% and 77% vs. 28% and 23%), and usage of distal A5SS (65% and 58% vs. 35% and 42%) and A3SS (65% and 65% vs. 35% and 35%) events (Fig. 2b). Based on the observation that half of the DAP3-modulated splicing events belongs to SE, we further analyzed whether there is a preferential regulation of constitutively included, excluded or alternatively spliced exons by DAP3. SE events identified in the EC109 and KYSE180 scramble control cells (108,399 and 110,895 in EC109 and KYSE180, respectively) were stratified into groups based on their PSI values (Supplementary

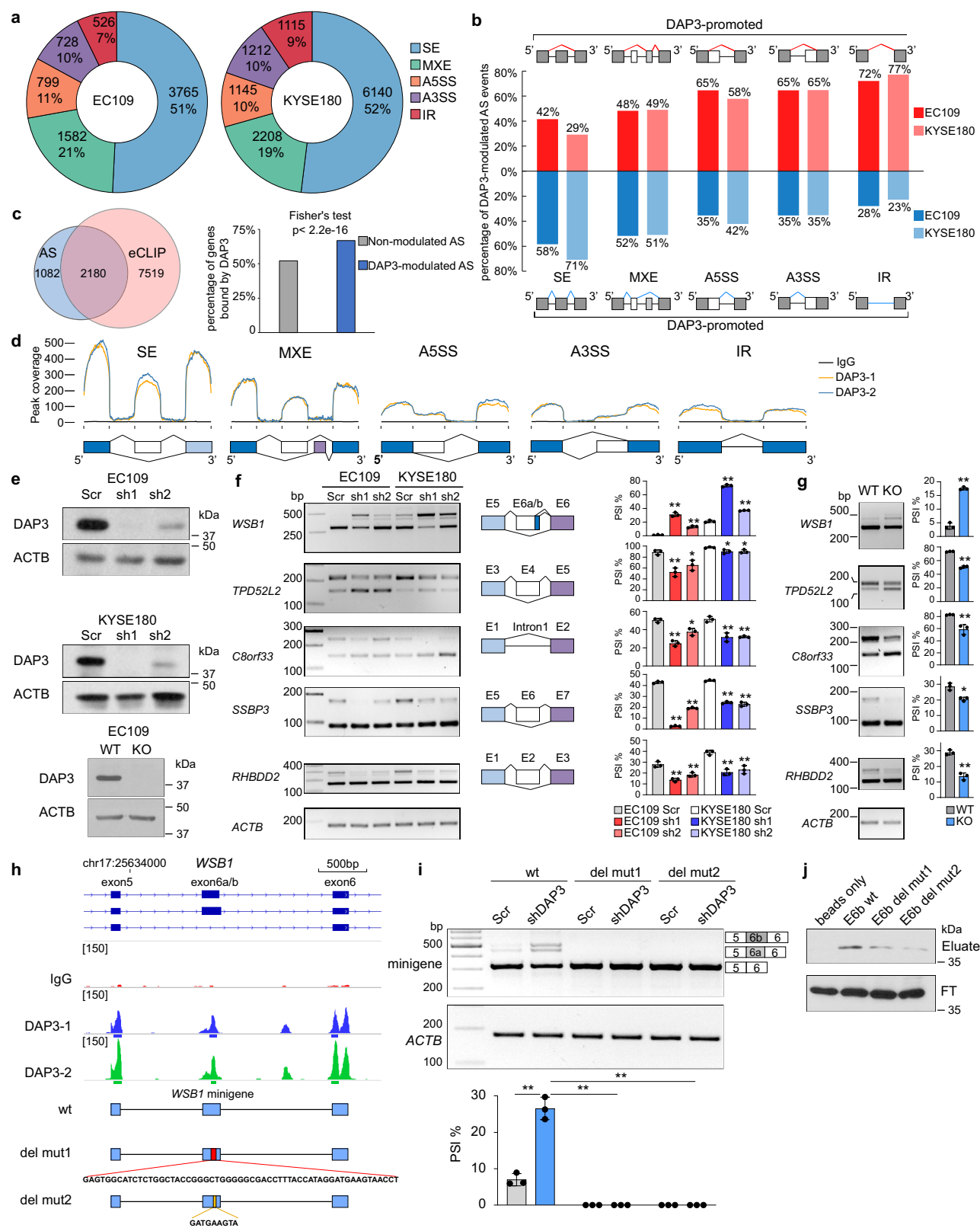


Fig. 4a). We found that less than 1% of SE events in “PSI = 0%” (i.e., constitutively excluded) group and less than 3% of SE events in “PSI = 100%” (i.e., constitutively included) group were modulated by DAP3; while 15–25% of SE events detected in the remaining groups with basal PSI values ranging from 20 to 80% were modulated by DAP3. These findings indicate that compared to constitutively included or excluded exons, those with basal PSI

values ranging from 20 to 80% are preferentially affected by DAP3.

A combined analysis of the DAP3 eCLIP-Seq and RNA-Seq datasets of EC109 cells showed that 67% (2180/3262) or 52% (5121/9828) of genes with or without DAP3-modulated alternative splicing were bound by DAP3, respectively, demonstrating a significant enrichment of DAP3-binding peaks in genes undergoing

Fig. 2 DAP3 depletion leads to widespread splicing changes in cancer cells. **a** Pie charts showing the distribution of each type of significantly altered splicing events in DAP3-depleted EC109 and KYSE180 cells compared to the scramble controls detected by rMATS¹⁵. SE skipped exon, MXE mutually exclusive exons, A5SS alternative 5' splice site, A3SS alternative 3' splice site, IR intron retention. **b** Percentage of DAP3-modulated splicing events demonstrating increased or decreased PSI upon DAP3 depletion in EC109 and KYSE180 cells. **c** Venn diagram showing the numbers of genes which underwent DAP3-modulated splicing and genes containing DAP3 eCLIP peaks in EC109 cells. Bar chart showing the percentage of genes with or without DAP3-modulated splicing that are bound by DAP3 (two-sided Fisher's Test). **d** The binned DAP3 eCLIP peak coverage across the splice junctions of five types of DAP3-modulated splicing events in EC109 cells. **e** Western blot analysis of DAP3 protein expression in the indicated DAP3 knockdown (KD, sh1, and sh2), knockout (KO) and their control [scramble (Scr) and wildtype (WT)] samples. **f, g** Semiquantitative RT-PCR analyses of five randomly selected DAP3-modulated splicing events. **h** IGV browser tracks of the DAP3 eCLIP peaks spanning exon 5–6 and intervening introns of *WSB1* gene. Significant peaks are marked by blue and green bars. Schematic diagram illustrates the genomic fragments inserted into the wildtype (wt) and mutant *WSB1* splicing minigenes. del mut1, lacking a 58 bp DAP3-binding sequence in exon E6a/b; del mut2, lacking a 9 bp DAP3-binding motif in exon E6a/b. **i** Semiquantitative RT-PCR analyses of splicing changes of exogenous *WSB1* transcripts derived from the indicated minigenes in Scr control and DAP3-KD cells. **j** RNA pulldown assay detecting the binding of DAP3 to *WSB1* exon E6b wt, del mut1, and del mut2 RNA probes. WB analysis of DAP3 proteins in RNA pulldown (eluate) products and flow-through (FT) fractions. **f, g, i** Data are represented as mean \pm s.d. of $n = 3$ biologically independent samples. Statistical significance is determined by unpaired, two-tailed Student's *t*-test (* $p < 0.05$, ** $p < 0.01$). Exact *p*-values and source data are provided in Source Data file.

DAP3-modulated splicing (Fisher's test, $p < 2.2e-16$; Fig. 2c). Our further randomization analysis confirmed that such a binding enrichment was not due to the difference in expression level or sample size between these two groups of genes (Supplementary Information; Supplementary Fig. 4c and d). Based on these observations, binding of DAP3 to its target RNA transcripts may be required for at least a subset of DAP3-modulated alternative splicing events. To further understand whether the binding affinity of DAP3 to RNA targets is associated with the strength of its splicing regulation, we divided DAP3-bound genes based on the fold enrichment of their eCLIP peaks into four groups (≥ 4 , ≥ 8 , ≥ 12 , and ≥ 16 fold) and examined the percentage of bound genes with or without DAP3-modulated splicing among all DAP3-bound genes. We found that the proportion of bound genes undergoing DAP3-modulated splicing remains within the range of 20–25% among different groups (Supplementary Fig. 4e), indicating an increased binding affinity of DAP3 does not potentiate its splicing regulation.

Further analysis of the eCLIP peak coverage for all five types of DAP3-modulated splicing events indicated a strong tendency of binding to exons rather than introns (Fig. 2d). The HOMER¹² motif enrichment analysis of sequences from DAP3-modulated cassette exons identified a top ranked AGGUAAGU motif (Supplementary Fig. 4f), which matches with the third enriched motif from DAP3-bound sequences (Fig. 1d) and the second enriched motif identified from sequences of DAP3-binding peaks located within the region from the upstream to downstream constitutive exon of DAP3-modulated splicing event (Supplementary Fig. 4g).

Next, a total of 20 DAP3-modulated splicing events were experimentally validated in DAP3-knockdown (KD) cells (Figs. 2e, f, 4b, and 5g). Five of which were chosen for further verification in DAP3-knockout (KO) cells (Fig. 2e and g). To provide experimental evidence that DAP3 binding to target RNA is required for its splicing regulation, we selected one of the DAP3 target genes *WSB1* for further investigation. We first constructed a wildtype (wt) minigene consisting of *WSB1* exons 5–6 and intervening introns. Based on the identified DAP3 eCLIP peaks on *WSB1* gene, we generated two deletion mutant minigenes by deleting a 58 bp DAP3-binding sequence on *WSB1* exon E6a/b (del mut1) or a 9 bp GATGAAGTA DAP3-binding motif (del mut2) (Fig. 2h). We found depletion of DAP3 led to the inclusion of previously unannotated exon E6a and E6b in the wt minigene (Fig. 2i), consistent with the splicing change of endogenous *WSB1* upon DAP3 depletion (Fig. 2f). However, such splicing changes were not observed in *WSB1* transcripts derived from del mut1 or del mut2 minigene upon DAP3 knockdown (Fig. 2i). Next, RNA pulldown assays using RNA probe consisting of the *WSB1* wt exon E6b (E6b wt), the 58nt DAP3-binding sequence-depleted

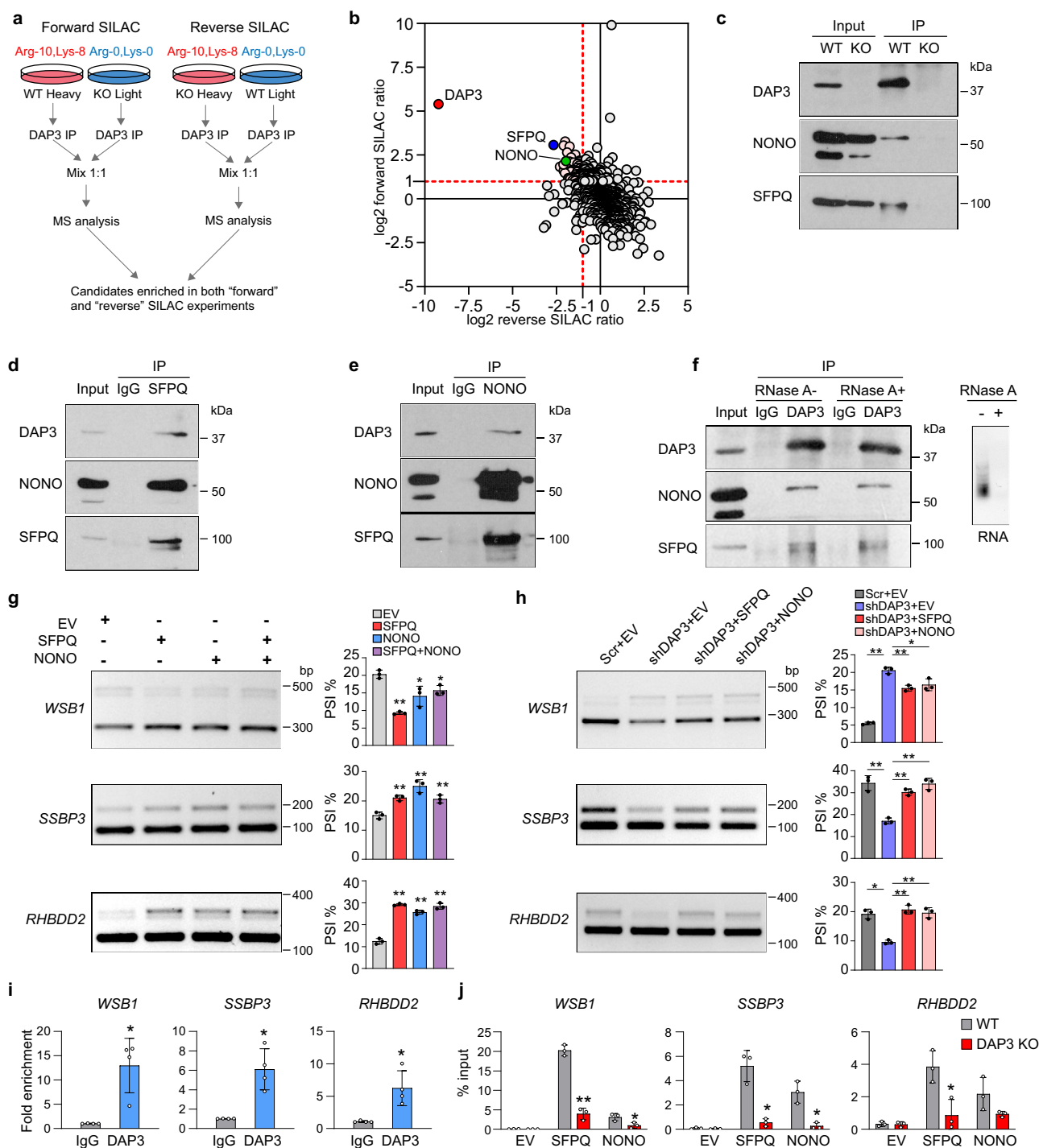
exon E6b (E6b del mut1), or the 9nt GATGAAGTA binding motif-depleted exon E6b (E6b del mut2) further confirmed the deleted sequences were required for DAP3 binding (Fig. 2j). These findings suggest that for genes undergoing DAP3-modulated splicing such as *WSB1*, the binding of DAP3 to its target RNA is required for its splicing regulation.

Moreover, given that one-third (1082/3262) of genes with DAP3-modulated alternative splicing events was not bound by DAP3, it is possible that DAP3 may also modulate splicing through RNA-binding-independent mechanisms. Because A-to-I RNA editing could potentially affect splicing¹⁶, we further intersected DAP3-regulated editing sites with DAP3-modulated splicing events and found only 1–2% of splicing events having editing sites detected in the region from the upstream to downstream constitutive exon of an alternatively spliced event (Supplementary Fig. 4h and i). Although this does not exclude the possibility that DAP3-regulated RNA editing in nascent RNAs could impact pre-mRNA splicing, editing-mediated alternative splicing does not stand as a dominant mechanism of DAP3-mediated splicing regulation.

GO analysis suggested that genes undergoing DAP3-modulated alternative splicing were functionally enriched in pathways associated with DNA replication, mitotic cell cycle phase transition, and RNA processing (Supplementary Fig. 5). Together, these findings provided solid evidence that DAP3 is a critical regulator of widespread alternative RNA splicing and dramatically reshapes transcriptome in cancer cells by altering thousands of alternative splicing events.

DAP3 facilitates the association of splicing regulators SFPQ and NONO with target RNAs for splicing modulation.

To dissect the mechanism by which DAP3 functions as an RBP to modulate splicing, we first explored whether DAP3 interacts with other RBPs, such as splicing factors, to facilitate or block their binding to target RNAs. We performed immunoprecipitation coupled to mass spectrometry (IP-MS) in combination with SILAC (stable isotope labelling with amino acids in cell culture) to detect and quantify DAP3-interacting proteins (Fig. 3a). We identified two well-known RNA splicing regulators SFPQ and NONO among the top DAP3-interacting proteins in both forward and reverse SILAC experiments (Fig. 3b; Supplementary data 4). Reciprocal co-IP analysis further confirmed that DAP3 interacted with SFPQ and NONO (Fig. 3c–e), and their interactions were not abolished by RNase A treatment prior to the co-IP experiments (Fig. 3f). Of note, by examining three exemplary target transcripts *WSB1*, *SSBP3*, and *RHBDD2*, we noticed that opposite to the effect of DAP3 depletion, overexpression of SFPQ



and NONO alone or together significantly promoted skipping of *WSB1* exon E6a/E6b, whereas repressed skipping of *SSBP3* exon 6 and *RHBDD2* exon 2 (Fig. 3g; Supplementary Fig. 6a). Moreover, introduction of SFPQ and NONO into *DAP3*-depleted cells reversed splicing changes of *WSB1*, *SSBP3*, and *RHBDD2* caused by *DAP3*-KD (Fig. 3h; Supplementary Fig. 6b). These observations prompted us to examine whether *DAP3* might form a splicing regulatory complex with SFPQ and NONO to interact with mRNA transcripts and modulate RNA splicing cooperatively. To this end, after experimental verification of the association of *DAP3* with *WSB1*, *SSBP3*, and *RHBDD2* (Fig. 3i; Supplementary Fig. 7), we found that in the absence of *DAP3*, the association of SFPQ and NONO to these transcripts were

significantly reduced (Fig. 3j; Supplementary Fig. 6c). Altogether, we demonstrated that *DAP3* complexes with the splicing regulators SFPQ and NONO in an RNA-independent manner and facilitates their association with target RNA transcripts for splicing modulation.

DAP3 modulates splicing of numerous splicing factors. As described above, approximately one-third of transcripts which undergo *DAP3*-modulated alternative splicing do not have *DAP3*-binding sites detected by eCLIP-Seq (Fig. 2c). We next explored whether *DAP3* could modulate alternative splicing via other mechanisms independent of RNA binding. We first looked

Fig. 3 DAP3 facilitates the binding of splicing factors SFPQ and NONO to target RNAs. **a** Flowchart of the immunoprecipitation coupled to mass spectrometry (IP-MS) in combination with SILAC (stable isotope labelling with amino acids in cell culture) to detect DAP3 interactors. **b** Scatterplot of DAP3-immunoprecipitated proteins retrieved from both forward and reverse SILAC-based IP-MS. DAP3 and several top interactors (e.g., SFPQ and NONO) are highlighted. **c** Co-IP analysis of protein extracts from the WT and DAP3-KO EC109 cells. IP was performed with a DAP3 antibody, followed by western blot analysis of DAP3-pulldown products using DAP3, SFPQ, and NONO antibodies. **d, e** Co-IP analysis of protein extracts from EC109 cells. IP was performed with a SFPQ (**d**) or NONO (**e**) antibody, followed by western blot analysis of SFPQ or NONO-pulldown products using the indicated antibodies. IgG antibody was used as negative control. **f** Co-IP analysis of protein extracts from EC109 cells. RNase A treated (RNase A+) or untreated (RNase A-) protein lysates were used for IP using DAP3 antibody, followed by western blot analysis of DAP3-pulldown products using the indicated antibodies. IgG antibody was used as negative control. Agarose gel demonstrating the successful digestion of total RNA. **g** Semiquantitative RT-PCR analyses of the indicated splicing events upon overexpression of SFPQ or NONO alone or together in EC109 cells. **h** Semiquantitative RT-PCR analyses of the indicated splicing events upon overexpression of SFPQ or NONO in DAP3-KD EC109 cells. **i** RIP-qPCR analysis of the association between DAP3 protein and the indicated mRNA transcripts (*WSB1*, *SSBP3*, and *RHBDD2*). Data are represented as mean \pm s.d. of $n = 4$ biologically independent samples. **j** RIP-qPCR analysis of the binding of SFPQ or NONO to the indicated mRNA transcripts in DAP3-KO and WT EC109 cells that were transfected with Flag-tagged SFPQ or NONO, respectively. RIP was conducted using anti-Flag M2 beads. **g, h, j** Data are represented as mean \pm s.d. of $n = 3$ biologically independent samples. **g-j** Statistical significance is determined by unpaired, two-tailed Student's *t*-test (* $p < 0.05$, ** $p < 0.01$). Exact *p*-values and source data are provided in Source Data file.

into over 300 genes which are involved in mRNA processing and have their splicing modulated by DAP3 (Supplementary Fig. 5), including genes encoding the spliceosomal small nuclear ribonucleoproteins (*SNRPA*, *SNRPA1*, *SNRPB*, *SNRPE*, *SNRPG*, and *SNRNP27*), RNA-binding motif proteins (*RBM3*, *RBM4*, *RBM5*, *RBM6*, *RBM7*, *RBM10*, *RBM15*, *RBM23*, *RBM28*, and *RBM39*), Serine/Arginine-rich (SR) splicing factors (*SRSF1*, *SRSF3*, *SRSF5*, *SRSF7*, *SRSF10*, and *SRSF11*), heterogeneous nuclear ribonucleoproteins (*HNRNPC*, *HNRNPD*, *HNRNPK*, *HNRNPH1*, and *HNRNPH2*), and DEAD-box helicases (*DDX5*, *DDX19B*, *DDX23*, *DDX39A*, *DDX42*, and *DDX46*), and found that many of which form functional protein association networks associated with various steps of RNA processing such as splicing and polyadenylation (Fig. 4a). We then went on to show that depletion of DAP3 truly promoted skipping of *RBM6* exon 6, *NSRPI* exon 2, *AKAP17A* exon 5b, *FMR1* exon 12, and *HNRNPH1* exon 4, as well as the usage of a distal 3'SS of *SNRPB* exon 7 and a distal 5'SS of *PRPF4B* exon 12 and *TIA1* exon 8; while repressed skipping of *RBM4* exon 3 and the usage of a distal 3'SS of *TIAL1* exon 3 (Fig. 4b). It is well-known that splicing factors can autoregulate their own expression or cross-regulate other splicing factors through alternative splicing coupled nonsense-mediated decay (AS-NMD)^{9,10}. Hypothetically, among 10 selected DAP3-modulated splicing factors, DAP3 can trigger the NMD of *RBM6*, *HNRNPH1*, *AKAP17A*, and *TIA1* by introducing a premature termination codon (PTC). As expected, cycloheximide (CHX) treatment or *UPF1* knockdown significantly restored the expression of these PTC-containing isoforms (Fig. 4c and d). On the other hand, there was no obvious NMD observed in the remaining six splicing factors (Supplementary Fig. 8), suggesting that these alternatively spliced isoforms are translated into protein variants that might have distinct functions or activities.

To further investigate whether the DAP3-triggered NMD of splicing factors indeed contributes to DAP3-modulated splicing, *RBM6*, a known cancer-related splicing regulator¹⁷ was selected for further investigation. Skipping of *RBM6* exon 6 causes an open reading frameshift and introduces a PTC on exon 8 (Fig. 5a). DAP3 was found to bind to *RBM6* exon 6 and flanking exons and introns (Fig. 5b). Consistent with the observation that DAP3 depletion promoted NMD of *RBM6* mRNA transcript (Fig. 4c and d), the mRNA level of the canonical exon 6-included *RBM6* isoform was significantly reduced in DAP3-KD cells (Fig. 5c). A reduction in protein expression of *RBM6* was also confirmed upon DAP3 depletion (Fig. 5d). By performing RNA-Seq analysis of *RBM6* KD cells, we identified a total of 94 alternative splicing events, which were modulated by *RBM6* and DAP3 (defined as “co-modulated” targets) (Fig. 5e and f; Supplementary data 5).

Notably, DAP3-KD-induced splicing changes in five validated co-modulated targets could be rescued by the restoration of *RBM6* expression (Fig. 5g and h; Supplementary Fig. 9). Altogether, our results indicated that DAP3 could modulate genome-wide changes in splicing indirectly via affecting the expression of splicing factors/regulators.

Clinical relevance of DAP3-modulated mis-splicing in cancers.

It has been reported that DAP3 is overexpressed in a broad range of cancer types and appears to have strong oncogenic effect¹¹. We next conducted a pan-cancer analysis of alternative splicing across 33 cancer types using the TCGA Spliceseq dataset¹⁸ to study whether DAP3-modulated splicing events are frequently dysregulated in different cancer types and evaluate their clinical significance. Of 20 experimentally validated DAP3-modulated splicing events (Figs. 2f, 4b, and 5g), 18 of them (90%) were detected in almost all 33 TCGA cancer types with varying PSI values (Supplementary Fig. 10). To examine whether these DAP3-modulated splicing events are dysregulated in cancers, we compared the PSI values of these splicing events between tumors and non-tumor (NT) samples in several representative cancer types, which demonstrate significantly higher expression of DAP3 in tumors as reported previously¹¹. We found that DAP3-modulated splicing events (e.g., *WSB1*, *SNRPB*, *TIAL1*, *TBL1X*, *SSBP3*, and *CADM1*) were significantly dysregulated in tumors compared to NT samples in multiple cancer types, such as esophageal carcinoma (ESCA), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), glioblastoma multiforme (GBM), liver hepatocellular carcinoma (LIHC), and stomach adenocarcinoma (STAD) (Fig. 6a). Moreover, lower PSI values of *WSB1*, *TBL1X*, and *SNRPB* gene in tumors were significantly correlated with the shorter overall survival (OS) time of patients with ESCA and LIHC, whereas the higher PSI values of *TIAL1* and *CADM1* in tumors predicted poorer prognosis in patients with LIHC and COAD (Fig. 6b). It is known that different cancer types display diverse alternative splicing landscapes¹⁹. We further conducted a comparative analysis of DAP3-modulated alternative splicing using the RNA-Seq datasets of DAP3-depleted ESCC cells and 13 matched pairs of tumors and NT samples from the TCGA ESCA. Same as DAP3-depleted ESCC cells, SE is the major type of differentially regulated splicing type in ESCA tumors, followed by MXE, A5SS, A3SS and IR (Figs. 2a; 6c). Among 3811 differentially spliced events (across 1606 genes) detected in tumors (FDR < 0.05, $|\Delta\text{PSI}| \geq 10\%$), 14% (535/3,811) of which were identified as “DAP3-modulated splicing events” (Fig. 6c–e; Supplementary data 6). For example, DAP3-modulated splicing of *WSB1*, *SNRPB*, *TIAL1*, and *TBL1X* gene were significantly altered in tumors compared to their matched NT samples (Fig. 6f). All these observations

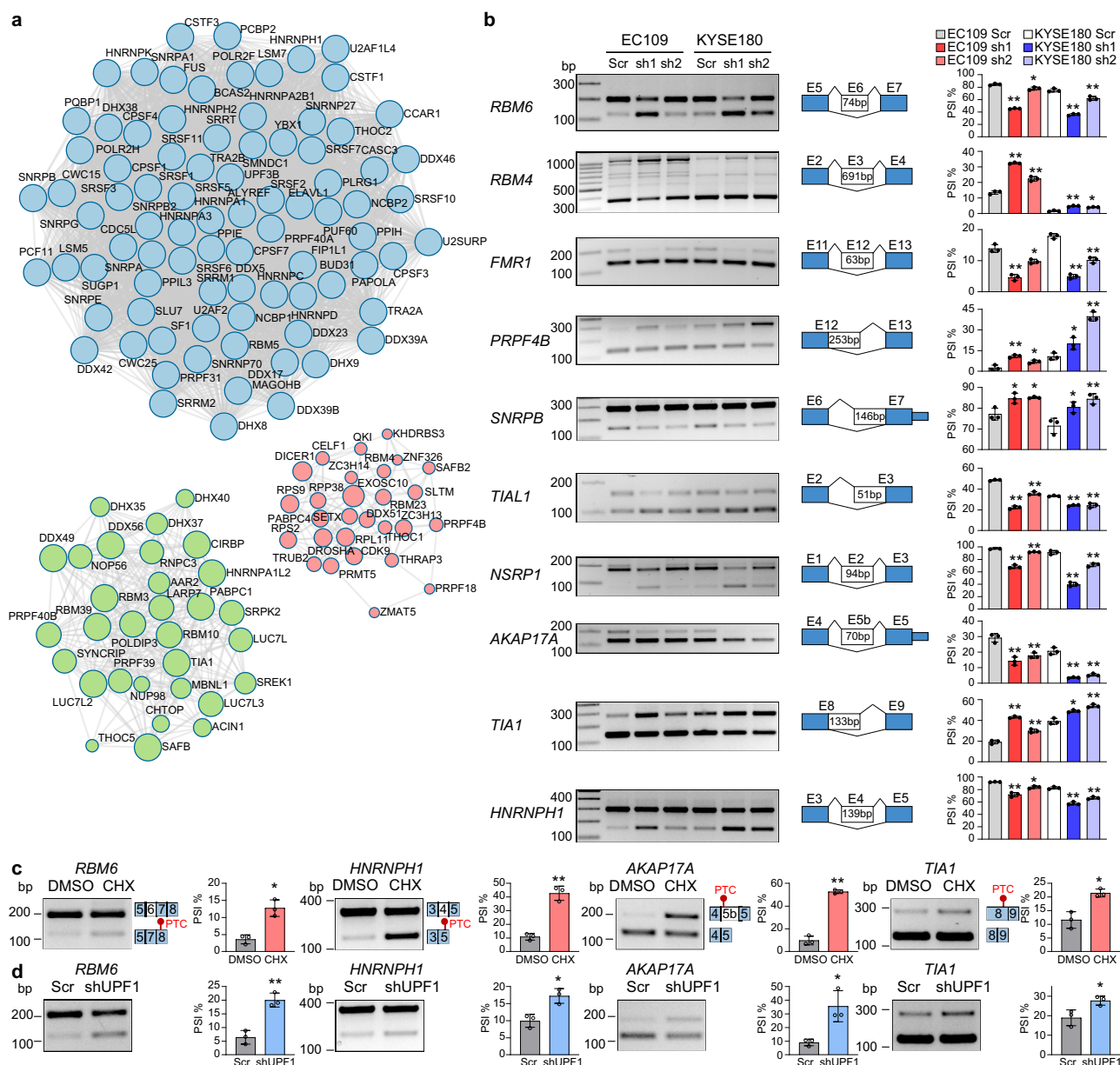


Fig. 4 DAP3 orchestrates splicing regulatory networks by modulating alternative splicing of splicing factors. **a** Functional protein association networks of genes that are involved in mRNA processing and have their splicing modulated by DAP3 were identified in Metascape⁴⁵(<http://metascape.org>). **b** Semiquantitative RT-PCR analyses of splicing changes of 10 representative splicing factors *RBM6*, *RBM4*, *FMR1*, *PRPF4B*, *SNRNPB*, *TIAL1*, *NSRNP1*, *AKAP17A*, *TIA1*, and *HNRNPH1* upon DAP3 depletion in EC109 and KYSE180 cells. **c**, **d** Semiquantitative RT-PCR analyses of splicing changes of *RBM6*, *HNRNPH1*, *AKAP17A*, and *TIA1* upon inhibition of NMD in EC109 cells. Cells were treated with CHX or vehicle control (DMSO) for 6 h (**c**) or transfected shUPF1 or the scramble control (Scr) for 48 h (**d**). **b–d** Data are represented as mean \pm s.d. of $n = 3$ biologically independent samples. Statistical significance is determined by unpaired, two-tailed Student's *t*-test (* $p < 0.05$, ** $p < 0.01$). Exact *p*-values and source data are provided in Source Data file.

suggest that prevalent dysregulations of DAP3-modulated splicing events are clinically relevant to multiple cancer types.

DAP3 increases WSB1 expression via repressing AS-NMD of WSB1 to promote tumorigenesis. We next looked for experimental evidence that supports a functional link of DAP3-modulated mis-splicing to tumorigenesis. Here, we were particularly interested in one of the DAP3 target genes, *WSB1* (WD repeat and SOCS box-containing protein 1) which is an E3-ubiquitin ligase that promotes ATM ubiquitination and degradation to drive tumorigenic progression²⁰, with the following reasons including (1) differentially spliced-in tumors of multiple cancer types (Fig. 6a), (2)

lower PSI value in tumors predicts poor prognosis (Fig. 6b), and (3) hypothetically, its expression can be affected by DAP3 via AS-NMD (Fig. 7a). As shown in Fig. 2f and g, inclusion of a non-canonical exon E6a or E6b by usage of cryptic splicing sites at intron 5 in the *WSB1* mRNA transcript was promoted by DAP3-KD or KO. While exon E6b has an additional 12-nucleotide sequence at 3' end than exon E6a, inclusion of either E6a or E6b introduces a PTC that may lead to selective degradation by NMD (Fig. 7a). Suppression of NMD by either CHX treatment or *UPF1* knockdown significantly restored the expression of E6a/E6b-included *WSB1* isoform, confirming that inclusion of exon E6a or E6b truly introduces a PTC-containing isoform and triggers NMD (Fig. 7b). Both KD and KO

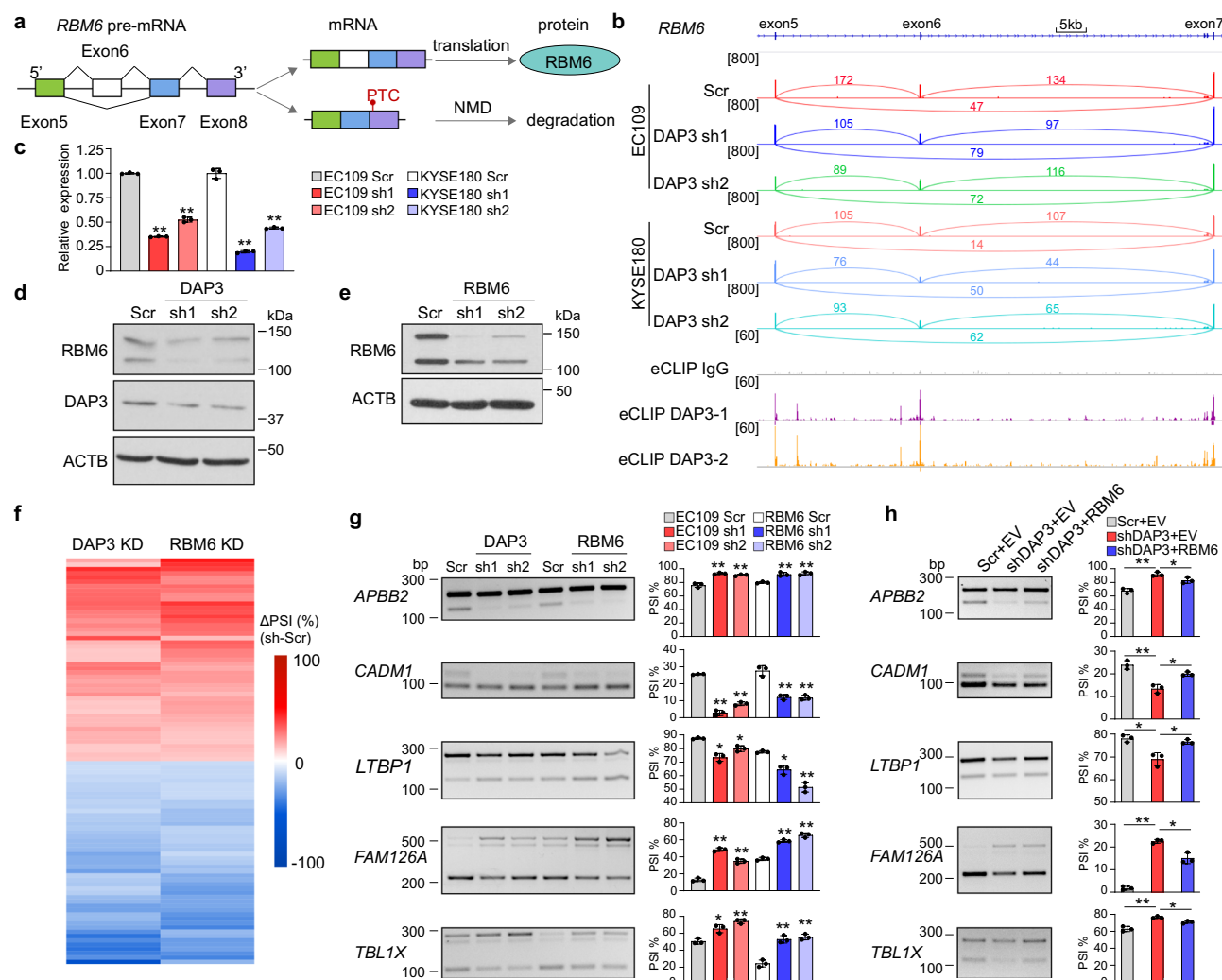


Fig. 5 DAP3 positively regulates RBM6 expression via repressing AS-NMD of RBM6. **a** Schematic diagram depicting skipping of RBM6 exon 6 causes a shift in the open reading frame, resulting in the introduction of a PTC into RBM6 transcript and possible NMD. **b** Visualization of RNA-Seq data of DAP3-depleted EC109 and KYSE180 cells and DAP3 eCLIP-seq peaks spanning the RBM6 gene locus using IGV. Significant peaks are marked by purple and orange bars. **c** qRT-PCR analysis of expression of the RBM6 exon 6-included isoform after DAP3 depletion in EC109 and KYSE180 cells. ACTB was used as a housekeeping gene internal control. Data are represented as mean \pm s.d. of technical triplicates. **d** Western blot analysis of RBM6 protein expression after DAP3 depletion in EC109 cells. **e** Western blot analysis of RBM6 protein expression after RBM6 knockdown in EC109 cells. **f** Heatmap showing the co-modulated splicing events upon knockdown of DAP3 and RBM6. **g** Semiquantitative RT-PCR analyses of five randomly selected splicing events co-modulated by DAP3 and RBM6. Data are represented as mean \pm s.d. of $n = 3$ biologically independent samples. **h** Semiquantitative RT-PCR analyses of the indicated splicing events after re-expressing RBM6 in DAP3-depleted cells. Data are represented as mean \pm s.d. of $n = 3$ biologically independent samples. **c, g, h** statistical significance is determined by unpaired, two-tailed Student's *t*-test (* $p < 0.05$, ** $p < 0.01$). Exact *p*-values and source data are provided in Source Data file.

of DAP3 significantly decreased the expression of WSB1 at mRNA and protein levels through repressing the skipping of E6a/E6b (Fig. 7c–e). Conversely, overexpression of DAP3 increased WSB1 protein expression (Fig. 7f). Restoring DAP3 expression in KO cells could rescue WSB1 expression (Supplementary Fig. 11a). As a downstream target of WSB1²⁰, the protein expression of ATM was negatively regulated by DAP3-mediated change in WSB1 protein expression (Fig. 7d–f).

Previously we have shown DAP3-promoted tumorigenesis¹¹, but whether DAP3-modulated splicing contributes to its oncogenic function is not clear. As shown above, DAP3 depletion caused AS-NMD of WSB1 to suppress WSB1 expression. Therefore, we hypothesized that DAP3-modulated AS-NMD of WSB1 might contribute to the oncogenic function of DAP3. To this end, we examined the changes in tumorigenic ability of WT or DAP3-KO cells after restoring WSB1 expression by performing in vitro

foci formation and anchorage independent soft agar assays and in vivo xenograft assay (Fig. 7h–k). Overexpression of WSB1 in WT EC109 cells promoted tumorigenesis in vitro and in vivo when compared to WT cells expressing empty vector control (Fig. 7h–k), indicating the oncogenic role of WSB1 in cancer cells. While KO of DAP3 significantly repressed the tumorigenicity, reintroduction of WSB1 gene into DAP3-KO cells significantly attenuated DAP3-KO-induced suppression of tumorigenicity both in vitro and in vivo (Fig. 7h–k). Similarly, restoring WSB1 expression in the DAP3-KD cells also partially rescued the reduced tumorigenicity (Supplementary Fig. 11b–e). These data indicate that DAP3 depletion represses tumorigenesis at least partially through promoting AS-NMD of WSB1 gene. In sum, our functional investigation of nonproductive splicing of WSB1 supports a causal relationship between DAP3-modulated mis-splicing and tumorigenesis.

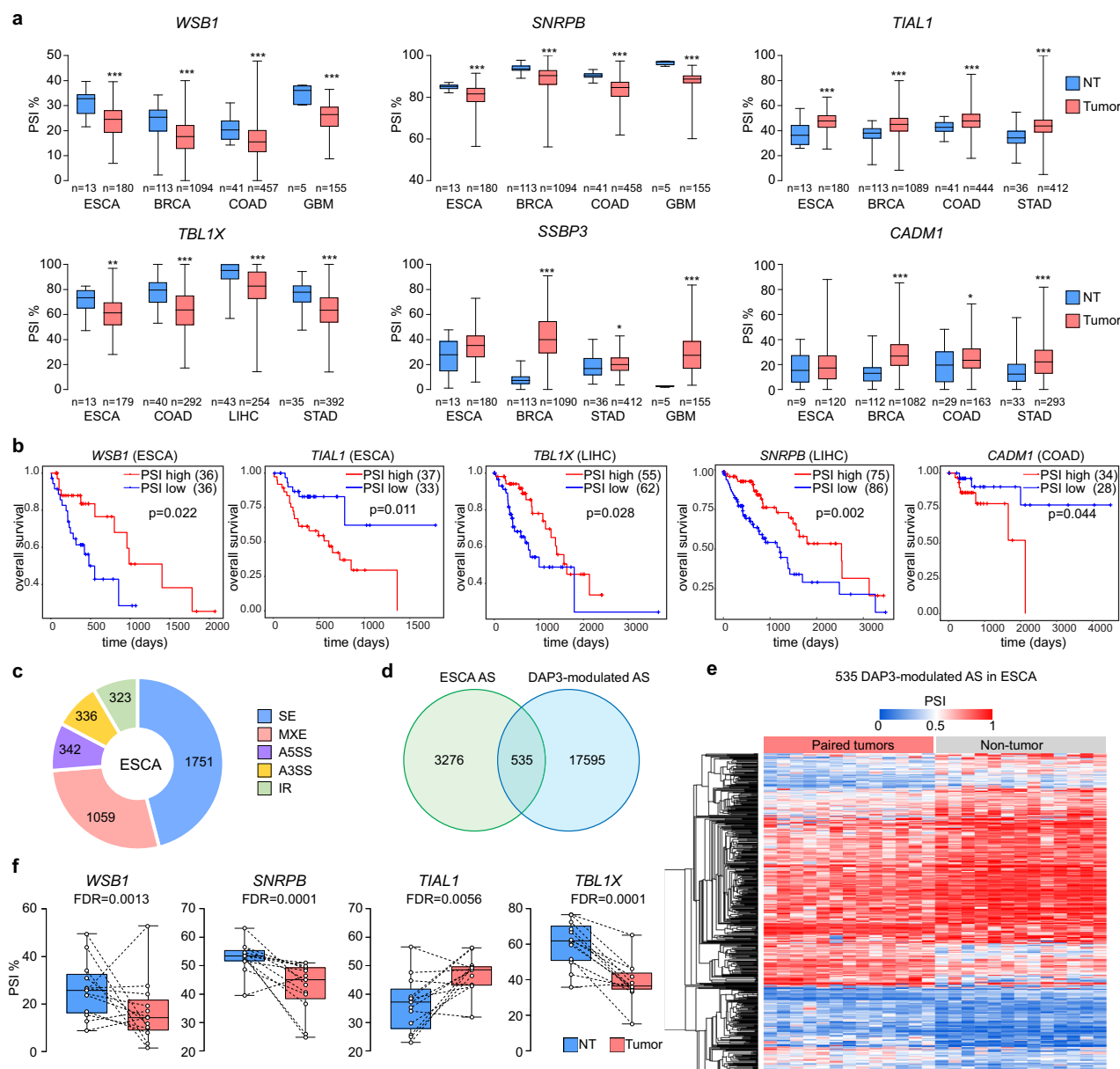


Fig. 6 Clinical relevance of DAP3-modulated mis-splicing in multiple cancer types. **a** Boxplots showing the PSI values of six experimentally validated DAP3-modulated splicing events in tumors and their matched NT samples of six representative TCGA cancer types including esophageal carcinoma (ESCA), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), glioblastoma multiforme (GBM), liver hepatocellular carcinoma (LIHC), and stomach adenocarcinoma (STAD). Statistical significance was determined by two-tailed nonparametric Mann-Whitney test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). **b** Kaplan-Meier OS plots comparing patients demonstrating the top 25th percentile of PSI values of the indicated DAP3-modulated splicing events ("PSI high" group) and those with the bottom 25th percentile of PSI values ("PSI low" group). Statistical significance was determined by log-rank test. **c** Pie chart showing the distribution of each type of differentially spliced events in 13 matched pairs of ESCA tumors and their NT samples detected by rMATS analysis¹⁵. **d** Venn diagram showing the number of significantly altered spliced events identified by RNA-Seq analysis of both DAP3-depleted ESCC cells (EC109 and KYSE180) and the TCGA ESCA. **e** Heatmap showing 535 significantly altered DAP3-modulated splicing events detected in both ESCC cell lines and ESCA tumors as described in **d**. The heatmap was generated using hierarchical clustering of tumors and NT samples (average linkage and Euclidean distance used). The color spectrum indicates PSI values. **f** Boxplots showing the PSI values of four representative DAP3-modulated splicing events in 13 matched pairs of ESCA tumors and NT tissues. **a, f** The box extends from the 25th to 75th percentiles. The line in the middle of the box is plotted at the median. The whiskers indicate min to max. Exact p -values and source data are provided in Source Data file.

Discussion

It is known that high level of fidelity required for splicing needs additional action of a complex interplay of RBPs that bind adjacent to splicing sites and promote recruitment of the spliceosome or outcompete spliceosomal components for binding to target RNAs²¹. Even modest changes in the abundance or activity of individual RBPs or core spliceosomal proteins can result in

aberrations or mistakes in splicing, which may be deleterious to cells and may result in cell death or cellular transformation^{8,22}. In this study, through application of eCLIP-Seq, RNA-Seq, and proteomics analyses, we characterize DAP3 as a widespread alternative splicing regulatory RBP which modulates thousands of splicing events and dissect its associated regulatory mechanisms and functional relevance to cancer. DAP3, which is overexpressed

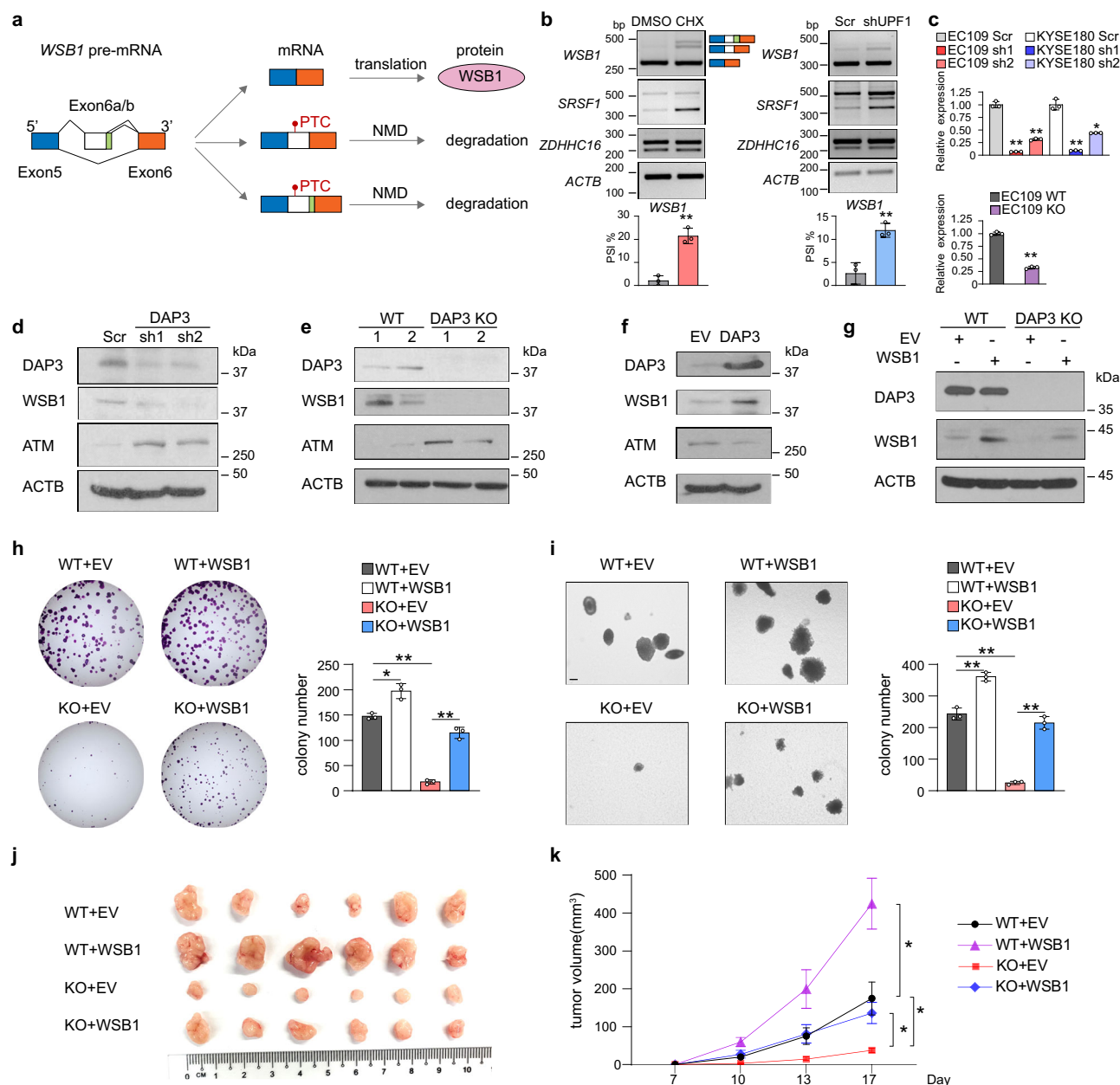


Fig. 7 DAP3 increases WSB1 expression via suppressing AS-NMD of WSB1 to promote tumorigenesis. **a** Schematic diagram illustrating the inclusion of non-canonical exon E6a and E6b introduces a PTC to *WSB1* mRNA transcript, which may result in NMD. **b** Semiquantitative RT-PCR analysis of the *WSB1* splicing upon inhibition of NMD. Left panel: EC109 cells were treated with CHX or vehicle control (DMSO) for 6 h. Right panel: EC109 cells were transfected with shUPF1 or the scramble control (Scr) for 48 h. *SRSF1* or *ZDHHC16* serves as a positive or negative control, respectively, to ensure successful inhibition of NMD. *ACTB* was used as a housekeeping gene internal control. Data are represented as mean \pm s.d. of $n = 3$ biologically independent samples. **c** qRT-PCR analysis of the change in expression of *WSB1* canonical isoform (E6a/6b-skipped) upon *DAP3*-KD (upper panel) or KO (lower panel). *ACTB* was used as a housekeeping gene internal control. Data are represented as mean \pm s.d. of technical triplicates. **d-f** Western blot analyses of DAP3, WSB1, ATM, and ACTB protein expression in **d** DAP3-KD EC109 cells, **e** DAP3-KO EC109 cells ("1" and "2" indicate two different WT or KO clones), and **f** EC109 cells with DAP3 overexpression. EV empty vector control. EV empty vector control. **g** Western blot analysis of DAP3 and WSB1 protein expression in DAP3-KO or WT EC109 cells that were overexpressed with the empty vector control or *WSB1* construct. **h, i** Quantification of foci formation (**h**) or soft agar colony formation (**i**) induced by the indicated stable cells. Scale bar: 200 μ m. Data are represented as mean \pm s.d. of $n = 3$ biologically independent wells. **j** Xenograft tumors derived from the indicated stable cell lines at end point ($n = 6$ mice per group). **k** Growth curve of tumors derived from the indicated cells in mice, over a 17-day observation period. Data are presented as the mean \pm s.e.m. Statistical significance is determined by unpaired, two-tailed Student's *t*-test ($*p < 0.05$). Exact *p*-values and source data are provided in Source Data file.

in multiple cancer types, has been proven to be a cancer-promoting gene¹¹. DAP3 has been reported to interact with ADAR proteins and repress A-to-I RNA editing; however, it does not bind proximal to the editing site within target RNA¹¹. Here, our eCLIP-Seq analysis revealed a widespread binding of DAP3 to RNAs transcribed from 9699 genes (approximately one-third of the total number of genes in the human genome). Many of these bound genes are critical regulators of mRNA processing including splicing and polyadenylation, gene expression, and mitotic cell cycle regulation. Of note, we found that DAP3 has an RNA-binding preference for exonic sequences with two identified top enriched binding motifs that are splicing regulatory sequence motifs, GAAGAA and AGGUAAGU. The purine-rich GAAGAA hexamer, appearing to be one of the strongest ESEs critical for constitutive and alternative splicing¹³, has also been found as an internal exonic binding motif of SR proteins, such as SRSF1^{23,24}, transformer 2 alpha homolog (TRA2A), and transformer 2 beta homolog (TRA2B)²⁵. The RNA-binding preference of DAP3 towards the GAAGAA hexamer may partially explain why DAP3 tends to promote exon inclusion in both cancer cell lines. The other motif AGGUAAGU is a consensus sequence at the 5' splice site, which is complementary to the nucleotides 4–11 of U1RNA¹⁴. In concordance with this finding, we observed that DAP3 could modulate approximately a thousand A5SS events with a slight preference for the proximal 5' splice site. However, such regulations may also be controlled via other *cis*- and/or *trans*-acting mechanisms such as the strength of 5' splice sites and/or involvement of other splicing factors/regulators, which bind to the 5' splice sites.

Although we found a significant enrichment of DAP3-binding peaks in genes undergoing DAP3-modulated splicing and provided experimental evidence supporting the binding of DAP3 to its target gene *WSB1* is required for DAP3-mediated splicing change in *WSB1*, it is unlikely that extensive splicing changes mediated by DAP3 are simply through its binding to splicing consensus sequence or regulatory elements. Moreover, we also showed that the binding affinity of DAP3 to RNA targets is not associated with the strength of its splicing regulation. In this

study, we delineated two distinct mechanisms of how DAP3 functions in splicing regulation (Fig. 8). First, DAP3 directly binds to target RNA and mediates the recruitment of splicing factors such as SFPQ and NONO to the binding sites via protein–protein interaction independent of their interactions with RNA. SFPQ and NONO belong to the multifunctional Drosophila behavior/human splicing (DBHS) family of proteins with highly conserved N-terminal RNA recognition motifs (RRMs). Although they are not essential components for spliceosome assembly, numerous studies have identified them as spliceosome-associated proteins and play an important role in alternative splicing^{26–30}. Using a proteomic approach, we identified SFPQ and NONO as DAP3 interactors. A previous study has demonstrated a SFPQ interacting RBP, Dido3, is required to recruit SFPQ to its target exon for efficient alternative splicing³¹. Similarly, we observed that loss of DAP3 could significantly compromise the association of SFPQ and NONO with their target RNA transcripts, indicating DAP3 functions as an important mediator to facilitate binding of SFPQ and NONO to RNA. Overexpression of SFPQ and NONO could readily rescue splicing changes of several exemplary DAP3 target transcripts *WSB1*, *SSBP3*, and *RHBDD2* in DAP3-depleted cells. Another mechanism of DAP3-modulated splicing involves indirect modulation of splicing by fine-tuning the splicing pattern of hundreds of splicing factors, as supported by our eCLIP-Seq and RNA-Seq data. These splicing factors include key components of splicing machinery, such as spliceosomal small nuclear ribonucleoproteins and heterogeneous nuclear ribonucleoproteins, as well as important splicing regulators, such as SR splicing factors, RBM proteins, and DEAD-box helicases. These splicing factors form dynamic splicing regulatory networks, which control splicing of tens of thousands of target genes. DAP3 alters the expression of these splicing factors via nonproductive splicing or isoform switching, thereby coordinately modulating splicing of multiple genes via their respective splicing factors. This is exemplified by DAP3-mediated AS-NMD of *RBM6* gene in this study. Depletion of DAP3 induces skipping of *RBM6* exon 6, causing a frameshift in its coding sequence and triggers NMD, indicating a positive

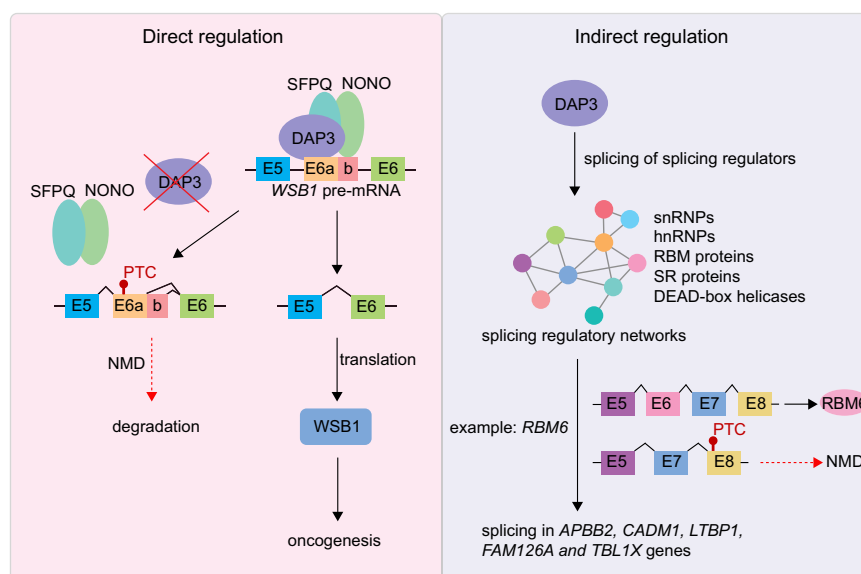


Fig. 8 Multilayered control of splicing regulatory networks by DAP3 leads to widespread alternative splicing changes in cancer. There are two distinct mechanisms of how DAP3 exerts its splicing regulatory functions. First, DAP3 directly binds to target RNA and mediates the recruitment of splicing factors, such as SFPQ and NONO, to the binding sites. Another mechanism involves indirect modulation of splicing through fine-tuning the splicing pattern of hundreds of splicing factors by DAP3. When DAP3 is overexpressed in many cancer types, global splicing changes can be observed and contribute to tumorigenesis.

regulation of DAP3 on RBM6 expression. Similar to DAP3, RBM6 also preferentially binds to exonic sequences; however, it has a distinct binding preference for sequences with CUCUGAA motif¹⁷. DAP3 cross-regulates RBM6, which in turn leads to a cascade of secondary splicing changes, expanding the repertoire of DAP3-affected splicing events. However, we also observed approximately a hundred splicing events that are discordant in DAP3-depleted and RBM6-depleted cells. Because splicing regulation often involves numerous RBPs and we showed DAP3-modulated splicing of hundreds of RBPs, it is possible that the modulation of these discordant splicing events involves other DAP3-modulated RBPs other than RBM6 alone. Therefore, the overall effects of these discordant splicing events after DAP3 depletion were not dominated by RBM6 downregulation.

The precise coordination and regulation of RNA processing by RBPs is essential for maintaining cellular homeostasis, which may otherwise result in various diseases including cancer. Our previous study demonstrated that DAP3 is overexpressed in multiple cancer types¹¹. In this study, we demonstrate that many DAP3-modulated splicing events were significantly altered in patients' tumors and such changes demonstrate prognostic values in multiple cancer types. Here we conducted detailed functional analyses of the AS-NMD of *WSB1* gene, whose alternative splicing could account for the oncogenic phenotypic changes driven by DAP3. *WSB1* contains seven WD40 repeats and a SOCS box at the C-terminus and functions as an E3-ubiquitin ligase in ubiquitination and proteasomal degradation³². It regulates the metastatic potential of renal carcinoma, osteosarcoma, and hormone receptor negative breast cancer by modulating pVHL, RhoGDI2, metalloproteinase (MMP) activity, vascular endothelial growth factor (VEGF) secretion^{33–35}. More importantly, it ubiquitinates ATM for degradation to overcome oncogene-induced senescence, thereby playing an important role in tumor initiation²⁰. Besides transcriptional regulation by oncogenic transcription factors, such as c-Myc, HIF-1, and CREB-ATF^{36,37}, our study provides another mechanism of *WSB1* expression regulation in cancer cells, which is via DAP3-modulated alternative splicing. DAP3 suppresses the nonproductive splicing of *WSB1*, which is prone to NMD. Since DAP3 is widely overexpressed in cancer cells, such DAP3-driven stabilization of *WSB1* mRNA transcripts may be a key step in tumor initiation.

In sum, our findings demonstrate that DAP3 coordinates splicing regulatory networks to modulate global alternative splicing in cancer via both RNA–protein and protein–protein interactions. Targeting DAP3-driven splicing events and blocking the splicing regulatory ability of DAP3 and/or specifically targeting DAP3-driven splicing events may hold great promise for cancer treatment.

Methods

Cell culture. EC109 and KYSE180 cells were cultured in HyClone RPMI 1640 medium (Thermo Fisher Scientific) supplemented with 10% FBS at 37 °C in a humidified incubator containing with 5% CO₂. HEK293T cells were cultured in DMEM high glucose medium (Biowest) supplemented with 10% FBS at 37 °C in a humidified incubator containing with 5% CO₂.

Generation of stable knockdown (KD) and overexpression cells. The DAP3-KD stable EC109 and KYSE180 cell lines were established using lentiviral transduction, followed by puromycin selection. The pLKO-DAP3-sh1 (5'GCTTATCCAGCTATAC GATAT3'), pLKO-DAP3-sh2 (5'ATCCTGGTTTCCAACTATAAC3'), pLKO-RBM6-sh1 (5'GACTGGTCTTCAGATACAAAT3') and pLKO-RBM6-sh2 (5'ACGGAACAC AAGTAGACTTAA3') constructs were used for the lentivirus packaging. EC109 cells stably expressing Flag-tagged *WSB1* were established by transduction of packaged lentiviral CSII-CMV-Flag-*WSB1* constructs, followed by the puromycin selection.

Generation of DAP3-knockout (KO) cells using CRISPR/Cas9 system. DAP3 sgRNAs were designed using the CRISPR design tool from The Massachusetts Institute of Technology (MIT) (<http://crispr.mit.edu>). The DAP3 sgRNA

(ATAGCTCTCGGACTCTCAAC) targeting exon 3 of DAP3 was cloned into the pX330A vector. EC109 cells were transfected with either empty vector or vector expressing DAP3 sgRNA and split into single cell. Clones grew from single cell were lysed by DirectPCR Lysis Reagent (Viagen Biotech) followed by the detection of indels in each single clone by T7EI assays. PCR products were TA cloned and Sanger sequenced to confirm the biallelic KO of DAP3 in each individual clone. Western blot analysis was performed to confirm DAP3-KO at protein level.

eCLIP-Seq data analysis. The eCLIP experiment was performed as previously described³⁸. Briefly, 20 million of EC109 cells were UV crosslinked, fragmented and immunoprecipitated using a DAP3 antibody (Abcam, ab2637) or control IgG (Invitrogen, 02-6202). Next, the bound protein-RNA products were subjected to gel electrophoresis and membrane transfer. Bound RNAs on the membrane corresponding to the protein size of DAP3 and 70 kDa above were extracted and further processed with adaptor ligation. The cDNA library was prepared by reverse transcription and sequenced by paired-end 100 bp sequencing performed on the Illumina HiSeq 4000 platform. The sequencing data were processed as previously described and clusters identified in IP samples were compared against paired size-matched input to obtain significantly enriched peaks using a Fisher's Exact test (or Yates' Chi-Square test if all observed and expected values were above 5), with *p*-values reported not corrected for multiple hypothesis testing^{7,38}. Peaks with fold enrichment (four fold) and significance (*p*-value <0.001) in immunoprecipitation versus paired size-matched input sample were defined as significant binding peaks. The eCLIP-seq data were first utilized in our previous study¹¹ (GEO accession number: GSE144318); however, the comprehensive analysis of eCLIP-seq data were provided only in the present study. The gene ontology pathway enrichment analysis was done using WebGestalt³⁹ portal (<http://www.webgestalt.org/>). The significance of enrichment for GO sets were evaluated by the pipeline default hypergeometric test.

Coverage plots of eCLIP peaks across transcripts and splice junctions. To understand the distribution of DAP3-binding peaks on transcripts, a previous method was adapted to generate the coverage plot for the eCLIP-Seq peaks⁴⁰. To increase the reliability of this analysis, we only considered “expressed genes” which have TPM (transcripts per million) ≥ 1 in each of 10 RNA-Seq datasets of EC109 cells generated by our team previously, resulting in 8170 such genes. The gene annotations for reference human genome (hg19) were downloaded from UCSC Table Browser⁴¹ (“knownGene” table), and only the longest isoform was considered as the representative transcript per gene. Next, each gene was split into 183 bins excluding the introns (13 for 5'UTR, 100 for CDS, and 70 for 3'UTR) proportional to the median 5'UTR, CDS, and 3'UTR lengths of the representative transcripts of expressed genes in EC109 cells. Finally, by using the significant peaks from the DAP3 eCLIP-Seq output, the cumulative peak coverage for each bin was calculated for these genes in a strand-specific manner with “bedtools coverage”⁴² and used to generate the final plot.

Similarly, the coverage plots for splice junctions were generated as described above. For each type of splicing events (SE, A5SS, A3SS, MXE, IR) identified by rMATS, the schematic representation of exons, introns and splice junctions involving the AS event is depicted as Fig. 2d. Based on the above schematic, each of the exonic and intronic regions were split into 100 bins, and the coverage of significant AS events in each bin was calculated in a strand-specific manner by using “bedtools coverage”. In case multiple events of the same type were present in the same gene, each was regarded as a different event as long as the upstream or downstream boundaries were different, otherwise the longest inclusive form of the alternatively spliced exon was considered. The cumulative coverage of the significant peaks from DAP3 eCLIP-Seq output was then plotted for each type of AS events.

Differential gene expression and alternative splicing analysis of RNA-Seq

data. Gene expression profiling and differential expression analysis were performed by using CSI NGS Portal⁴³ (<https://csibioinfo.nus.edu.sg/csingsportal>). Strand-specific RNA-Seq was performed as previously described¹¹. Briefly, clean reads were aligned to the reference human genome (hg19) by using STAR⁴⁴ with default parameters. The differential alternative splicing (AS) events between DAP3-KD duplicates and the scrambled control were identified by using rMATS¹⁵ for five major types (SE, A5SS, A3SS, MXE, IR). The FDR was calculated using the default parameters based on the Benjamini–Hochberg approach. In the rMATS output, only significant AS events were further considered defined as: sum(IJC_KD1, SJC_KD1) ≥ 20; sum(IJC_KD2, SJC_KD2) ≥ 20; sum(IJC_Scr, SJC_Scr) ≥ 20; abs(InclLevelDifference) ≥ 0.1; FDR < 0.05. The GO pathway enrichment analysis of DAP3-modulated alternatively spliced gene was done using WebGestalt³⁹ portal (<http://www.webgestalt.org/>). The protein–protein interaction network analysis of DAP3-modulated alternatively spliced RBPs was done using Metascape⁴⁵ portal (<http://metascape.org>).

RNA-Seq datasets of 13 matched pairs of ESCA tumor and NT tissues from the TCGA were used for transcriptome-wide splicing analysis using the same rMATS pipeline. Differentially spliced events in tumors relative to NT samples were identified based on the following filter criteria: |ΔPSI (tumor vs normal)| ≥ 0.1 and FDR < 0.05. The heatmap of differentially spliced events was generated using

hierarchical clustering (average linkage and Euclidean distance used) on Morpheus (<https://software.broadinstitute.org/morpheus>).

Motif enrichment analysis. Motif enrichment analyses of DAP3 eCLIP peak sequences, sequences of DAP3-modulated cassette exons and sequences of DAP3 eCLIP peaks located within the region from the upstream to downstream constitutive exon of DAP3-modulated splicing event were performed using “findMotifsGenome.pl” function in HOMER¹² for “de novo” motif enrichment with the flag “-rna”. The background is randomly selected sequences from the reference genome, corrected for sequence content (GC content and other bi-nucleotide composition, <http://homer.ucsd.edu/homer/motif/> and <http://homer.ucsd.edu/homer/motif/rnaMotifs.html> for details). Statistics was derived from binomial test against the random genomic background using the HOMER default setting.

TCGA SpliceSeq analysis of DAP3-modulated splicing events. Data from the TCGA SpliceSeq analysis¹⁸ of 33 cancer types were downloaded from <http://bioinformatics.mdanderson.org/TCGASpliceSeq>. Out of the 20 validated DAP3-modulated splicing events, 18 were manually mapped into the splicing events identified in TCGA SpliceSeq dataset according to the alternative splice site locations. The PSI values of each splicing event in tumor and NT samples were compared and statistical significance was determined by two-sided nonparametric Mann–Whitney test. For the OS analysis, we segregated the TCGA patients based on the PSI values of each DAP3-modulated splicing event. The patients with the top 25th percentile of PSI values were defined as “PSI high” group, while patients demonstrating the bottom 25th percentile of PSI as the “PSI low” group. The OS benefit was estimated by the log-rank test and presented by a Kaplan–Meier plot.

RNA purification and semiquantitative RT-PCR. Total RNA was purified using RNeasy Mini Kit (Qiagen) with RNase-Free DNase Set (Qiagen) to digest contaminating DNA. Reverse transcription was performed using SensiFAST™ cDNA synthesis kit (Bioline) and PCR was performed using exTEN 2× PCR Master Mix (Axil Scientific). RT-PCR products were subjected to electrophoresis in 2–3% agarose gel with ethidium bromide and visualized. Molecular weight markers (unit: bp) are labelled in gel images. Band intensity were measured by ImageJ for PSI calculation. Primer sequences are listed in Supplementary data 7.

Splicing minigene assay. The genomic DNA fragment spanning from exon 5 to exon 6 of WSB1 gene was cloned into pcDNA3.1(+) plasmid and deletion mutants were generated by mutagenesis cloning using PrimeSTAR Max DNA Polymerase (Takara). Minigene plasmids were transfected into cells using lipofectamine 2000. After 48 h total RNA was purified using RNeasy Mini Kit (Qiagen) with RNase-Free DNase Set (Qiagen) to digest contaminating DNA. Reverse transcription was performed using SensiFAST™ cDNA synthesis kit (Bioline) and PCR was performed using pcDNA3.1(+) forward and reverse primers by exTEN 2× PCR Master Mix (Axil Scientific). Primer sequences are listed in Supplementary data 7.

RNA pulldown assay. RNA probe was generated by RiboMAX™ Large Scale RNA Production Systems (Promega) using DNA template containing 5′-T7 promoter, WSB1 exon E6b wt or deletion mutant sequences, and 3′-aptamer. A total of 10 µg RNA probe was incubated with 20 µl Dynabeads MyOne C1 (Invitrogen) in 300 µl binding buffer (100 mM NaCl, 10 mM MgCl₂, 50 mM Hepes, pH 7.4, and 0.5% Igeal CA-630) for 30 min at 4 °C with rotation followed by three washes with washing buffer (250 mM NaCl, 10 mM MgCl₂, 50 mM Hepes, pH 7.4, and 0.5% Igeal CA-630). For each reaction, 1 mg whole-cell extract was diluted in 300 µl washing buffer and supplemented with 2 µl 10 mg/ml yeast tRNA (Invitrogen) and SUPERase-In™ RNase Inhibitor (Invitrogen). RNA immobilized beads were incubated with protein mixtures for 30 min at 4 °C with rotation. After three washes, bound proteins were eluted in 2× Laemmli buffer (Sigma) at 95 °C and analyzed by western blot. Primer sequences are listed in Supplementary data 7.

Co-immunoprecipitation (Co-IP). For the pulldown of DAP3, SFPQ and NONO protein, EC109 cells were lysed with prechilled lysis buffer (50 mM Tris-HCl, pH 7.5; 150 mM NaCl; 1% Nonidet P40; 0.5% sodium deoxycholate; 1× EDTA-free cComplete protease inhibitor (Roche)). The lysates were precleared with Dynabeads™ protein G (Invitrogen) at 4 °C overnight. The precleared lysates with incubated with anti-DAP3 (abcam, ab2637), anti-SFPQ (Santa Cruz, sc-271796) and anti-NONO (Santa Cruz, sc-166702) antibodies for 4 h at 4 °C and subsequently with Dynabeads™ protein G at 4 °C overnight. The Dynabeads™ protein G (Invitrogen) with bound proteins were washed with 150 mM NaCl with 1× EDTA-free cComplete protease inhibitor for six times and boiled with 2× protein loading buffer for 10 mins at 95 °C to elute bound proteins. Western blot analysis was performed to detect co-IP products. For the RNase A treatment prior to the immunoprecipitation, the total lysates were incubated with 0.1 µg/µl RNase A (Thermo Fisher Scientific) at 37 °C for 10 min.

Western blot analysis. Protein lysates were denatured and separated on SDS-PAGE gels, transferred onto polyvinylidene difluoride membranes, and immunoblotted with a primary antibody at 4 °C overnight, followed by incubation with a secondary antibody at room temperature for 1 h. The following antibodies are used

in this study: anti-DAP3 (1:1000, Abcam, ab2637), anti-β-actin (1:5000, Santa Cruz, sc-47778), anti-SFPQ (1:1000, Santa Cruz, sc-271796) and anti-NONO (1:1000, Santa Cruz, sc-166702), anti-WSB1 (1:1000, Novus, NBP2-82049), anti-ATM (1:1000, Santa Cruz, sc-377293), and anti-RBM6 (1:1000, Santa Cruz, sc-376201). β-Actin (ACTB) was used as a loading control. Molecular weight markers (unit: kDa) are labelled in blots.

Immunoprecipitation coupled to mass spectrometry (IP-MS) in combination with SILAC (stable isotope labelling by amino acids in cell culture). The DAP3-KO EC109 cells were cultured in “light” (84 mg/ml Arg-0 and 146 mg/ml Lys-0) media, and the WT cells were culture in “heavy” (84 mg/ml Arg-10 and 146 mg/ml Lys-8) media in the “forward” experiment for five consecutive passages. In the “reverse” experiment, the light and heavy labels were swapped. SILAC incorporation rates were >98% in both WT and KO samples. IP was performed using anti-DAP3 (abcam, ab2637). In the “forward” or “reverse” experiment, the IP products of the WT-heavy and KO-light or WT-light and KO-heavy samples were mixed at 1:1 ratio, respectively. The MS data acquisition and analysis were performed as described previously⁴⁶. The MS data are provided in Supplementary data 4.

RNA electrophoretic mobility shift assay (REMSA). The RNA probes were generated by in vitro transcription using RiboMAX™ Large Scale RNA Production Systems (Promega) and biotin-labelled using Biotin 3′ End DNA Labeling Kit (Thermo Fisher Scientific). Recombinant protein was prepared as described¹⁶. REMSA was performed using LightShift® Chemiluminescent RNA EMSA Kit, according to the manufacturer’s protocol (Thermo Fisher Scientific). Briefly, biotinylated RNA probes were heated for 5 min at 80 °C and placed on ice immediately to release secondary structure. The biotinylated RNA probes (0.5 pmol) were incubated with 40 ng of Flag-DAP3 proteins in the binding buffer containing 10 mM HEPES (pH 7.3), 20 mM KCl, 1 mM MgCl₂, 1 mM DTT, 100 ng/µl tRNA, and 0.2U/µl SUPERase-In™ RNase Inhibitor (Invitrogen) at room temperature for 30 min. In the RNA competition assay, a 100-fold molar excess of unlabelled RNA probes were preincubated with the reaction mixture at room temperature for 5 min before adding biotinylated probes. Samples were subjected to electrophoresis on a 5% native acrylamide gel, transferred to Amersham Hybond-NX (GE Healthcare) membrane, and detected by chemiluminescence. The probe sequences (motifs highlighted in red) are provided in Supplementary data 7.

RNA immunoprecipitation (RIP)-quantitative PCR (qPCR) analysis. Cells were lysed by prechilled lysis buffer (50 mM Tris (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1% Triton, 1×complete protease inhibitor (Roche), and 0.1U/µl SUPERase-In™ RNase Inhibitor (Invitrogen)) and incubated overnight with M2 magnetic beads at 4 °C. Then the M2 magnetic beads were washed six times with TBS buffer (50 mM Tris (pH 7.5), 150 mM NaCl, and 0.02U/µl SUPERase-In™ RNase Inhibitor (Invitrogen)). Bound proteins were eluted with 2× protein loading buffer after boiling at 95 °C for 10 min. Western blot was performed to examine pulldown efficiency. RNAs bound to the M2 magnetic beads were eluted with buffer RLT and purified with RNeasy Mini Kit (Qiagen). cDNA was synthesized using SensiFAST™ cDNA synthesis kit (Bioline) and qPCR was performed using GoTaq® qPCR Master Mix (Promega). Enrichment of pulled down RNAs were normalized to the input RNA expression levels. Primer sequences are listed in Supplementary data 7.

Foci and soft agar colony formation assay. For foci formation assay, EC109 cells were seeded at a density of 1×10^3 per well in six-well plates. Medium was replaced every 3 days. Visible colonies in each well were stained with crystal violet solution (0.1% crystal violet; 25% methanol) and quantified. A representative image of a stained well for each treatment was shown.

For soft agar assay, EC109 (1×10^3 per well) cells resuspended in 0.4% low-melting agarose were seeded on top of 0.6% low-melting agarose in six-well plates and incubated for 2 weeks. Visible colonies were stained with crystal violet solution (0.005% crystal violet; 25% methanol) and quantified. A representative image viewed under microscope for each treatment was shown.

In vivo tumorigenicity assay. NOD scid gamma (NSG) mice (The Jackson Laboratory, RRID:IMSR_JAX:005557) were maintained in pathogen-free (SPF) facility in NUS Comparative Medicine Department. Less than five mice with same sex were housed in a cage at 20–25 °C and 50% humidity with a 12 h light/dark cycle. For in vivo tumorigenicity assay, 0.5×10^6 EC109 cells were subcutaneously injected into the left or right flank of 4- to 6-week-old NOD scid gamma (NSG) mice (Fig. 7j; $n = 3$ males and $n = 3$ females for each group; Supplementary Fig. 11e; $n = 2$ males and $n = 3$ females for each group). Tumor growth was monitored, and tumor length (L) and width (W) measured at indicated time points. Tumor volume was calculated by the formula $V = 0.5 \times L \times W \times W$. All animal experiments were approved by and performed in accordance with the Institutional Animal Care and Use Committees (IACUC) of National University of Singapore. All tumors were harvested before or on the day of reaching the IACUC approved tumor size (15 mm at the largest diameter).

Statistics and reproducibility. Bioinformatic statistics used default settings for each individual pipeline and portal. All other statistical analyses were performed using Graphpad Prism v9.2.0 or Microsoft Excel 2019. All tests used in the study were two-sided and the exact *p*-values are provided in source data file. The number of replicates were provided in the figures and legends. Western blot and REMSA experiments were performed at least twice with similar results and representative data are shown.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The datasets generated and used in this study are available in the GEO repository, Accession ID: [GSE123020](https://www.ncbi.nlm.nih.gov/geo/) (RNA-Seq), [GSE172078](https://www.ncbi.nlm.nih.gov/geo/) (RNA-Seq), and [GSE144318](https://www.ncbi.nlm.nih.gov/geo/) (eCLIP-Seq). Databank URL: <http://www.ncbi.nlm.nih.gov/geo/>. All the other data are available within the article and its Supplementary Information. Source data are provided with this paper.

Code availability

The bioinformatics pipelines for RNA-Seq, eCLIP-Seq, and rMATS are available online at the CSI NGS Portal⁴³ (<https://csibioinfo.nus.edu.sg/csingsportal>). Bioinformatics code for downstream analysis is available upon reasonable request.

Received: 15 July 2021; Accepted: 2 March 2022;

Published online: 04 April 2022

References

- Gebauer, F., Schwarzl, T., Valcárcel, J. & Hentze, M. W. RNA-binding proteins in human genetic disease. *Nat. Rev. Genet.* **22**, 185–198 (2021).
- Sveen, A., Kilpinen, S., Ruusulehto, A., Lothe, R. & Skotheim, R. I. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* **35**, 2413–2427 (2016).
- David, C. J., Chen, M., Assanah, M., Canoll, P. & Manley, J. L. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* **463**, 364–368 (2010).
- Maquat, L. E. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat. Rev. Mol. Cell Biol.* **5**, 89–99 (2004).
- Kim, E. et al. SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell* **27**, 617–630 (2015).
- Karni, R. et al. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.* **14**, 185–193 (2007).
- Van Nostrand, E. L. et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).
- Ule, J. & Blencowe, B. J. Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol. Cell* **76**, 329–345 (2019).
- Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**, 926–929 (2007).
- Ni, J. Z. et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21**, 708–718 (2007).
- Han, J. et al. Suppression of adenosine-to-inosine (A-to-I) RNA editome by death associated protein 3 (DAP3) promotes cancer progression. *Sci. Adv.* **6**, eaba5136 (2020).
- Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Fairbrother, W. G., Yeh, R.-F., Sharp, P. A. & Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).
- Jacob, M. & Gallinaro, H. The 5' splice site: phylogenetic evaluation and variable geometry of association with U1RNA. *Nucleic Acids Res.* **17**, 2159–2180 (1989).
- Shen, S. et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl Acad. Sci. USA* **111**, E5593–E5601 (2014).
- Tang, S. J. et al. Cis- and trans-regulations of pre-mRNA splicing by RNA editing enzymes influence cancer development. *Nat. Commun.* **11**, 1–17 (2020).
- Bechara, E. G., Sebestyén, E., Bernardis, I., Eyra, E. & Valcárcel, J. RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol. Cell* **52**, 720–733 (2013).
- Ryan, M. et al. TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. *Nucleic Acids Res.* **44**, D1018–D1022 (2016).
- Kahles, A. et al. Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* **34**, 211–224. e216 (2018).
- Kim, J. J. et al. WSB1 overcomes oncogene-induced senescence by targeting ATM for degradation. *Cell Res.* **27**, 274–293 (2017).
- Cherry, S. & Lynch, K. W. Alternative splicing and cancer: insights, opportunities, and challenges from an expanding view of the transcriptome. *Genes Dev.* **34**, 1005–1016 (2020).
- Fu, X.-D. & Ares, M. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* **15**, 689–701 (2014).
- Sanford, J. R. et al. Splicing factor SRSF1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* **19**, 381–394 (2009).
- Wang, X. et al. Predicting sequence and structural specificities of RNA binding regions recognized by splicing factor SRSF1. *BMC Genomics* **12**, 1–11 (2011).
- Tacke, R., Tohyama, M., Ogawa, S. & Manley, J. L. Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* **93**, 139–148 (1998).
- Knott, G. J., Bond, C. S. & Fox, A. H. The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold. *Nucleic Acids Res.* **44**, 3989–4004 (2016).
- Gozani, O., Patton, J. G. & Reed, R. A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. *EMBO J.* **13**, 3356–3367 (1994).
- Patton, J. G., Porro, E., Galceran, J., Tempst, P. & Nadal-Ginard, B. Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev.* **7**, 393–406 (1993).
- Zhang, W.-J. & Wu, J. Y. Functional properties of p54, a novel SR protein active in constitutive and alternative splicing. *Mol. Cell Biol.* **16**, 5400–5408 (1996).
- Kameoka, S., Duque, P. & Konarska, M. M. p54nrb associates with the 5' splice site within large transcription/splicing complexes. *EMBO J.* **23**, 1782–1791 (2004).
- Mora Gallardo, C. et al. Dido3-dependent SFPQ recruitment maintains efficiency in mammalian alternative splicing. *Nucleic Acids Res.* **47**, 5381–5394 (2019).
- Choi, D. W. et al. Ubiquitination and degradation of homeodomain-interacting protein kinase 2 by WD40 repeat/SOCS box protein WSB-1. *J. Biol. Chem.* **283**, 4682–4689 (2008).
- Poujade, F.-A. et al. WSB-1 regulates the metastatic potential of hormone receptor negative breast cancer. *Br. J. Cancer* **118**, 1229–1237 (2018).
- Cao, J. et al. Hypoxia-induced WSB1 promotes the metastatic potential of osteosarcoma cells. *Cancer Res.* **75**, 4839–4851 (2015).
- Kim, J. J. et al. WSB1 promotes tumor metastasis by inducing pVHL degradation. *Genes Dev.* **29**, 2244–2257 (2015).
- Rhodes, D. R. & Chinnaiyan, A. M. Integrative analysis of the cancer transcriptome. *Nat. Genet.* **37**, S31–S37 (2005).
- Benita, Y. et al. An integrative genomics approach identifies Hypoxia Inducible Factor-1 (HIF-1)-target genes that form the core response to hypoxia. *Nucleic Acids Res.* **37**, 4587–4602 (2009).
- Van Nostrand, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).
- Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **45**, W130–W137 (2017).
- Van Nostrand, E. L. et al. Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol.* **21**, 90 (2020).
- Karolchik, D. et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**, D493–D496 (2004).
- Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
- An, O. et al. CSI NGS Portal: An Online Platform for Automated NGS Data Analysis and Sharing. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms21113828> (2020).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1–10 (2019).
- Roelofs, P. A. et al. Characterization of the mechanism by which the RB/E2F pathway controls expression of the cancer genomic DNA deaminase APOBEC3B. *elife* **9**, e61287 (2020).

Acknowledgements

This project was supported by National Research Foundation Singapore; Singapore Ministry of Education under its Research Centres of Excellence initiative; Singapore Ministry of Education's Tier 2 Grants (MOE2018-T2-1-005 and MOE2019-T2-2-008 to L.C.); Singapore Ministry of Education's Tier 3 Grants (MOE2014-T3-1-006 to L.C.); Singapore Ministry of Health's National Medical Research Council (NMRC) Clinician Scientist-Individual Research Grant (CS-IRG, project ID: MOH-000214 to L.C.); NMRC

Open Fund Young Individual Research Grant (OFYIRG19nov-0008:MOH-000537 to J.H.). We thank and acknowledge Prof Gene W. Yeo (Department of Physiology, National University of Singapore, Singapore; Department of Cellular and Molecular Medicine, University of California San Diego, USA) for providing technical supports to the eCLIP-Seq experiment.

Author contributions

J.H. performed the experiments and wrote the manuscript. O.A., X.R., and H.Y. conducted the bioinformatics analyses. Y.S., S.J.T., H.S., X.K., V.H.E.N., and D.J.T.T. provided insightful suggestions and experimental materials. H.Q.T. contributed to the eCLIP-Seq experiments. D.K. performed and analyzed SILAC-IP-MS experiments. L.C. supervised the study and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-29400-7>.

Correspondence and requests for materials should be addressed to Jian Han or Leilei Chen.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022