# Differential RNA editing between epithelial and mesenchymal tumors impacts mRNA abundance in immune response pathways

Tracey Chan[1], Ting Fu[2], Jae Hoon Bahn[3], Hyun-Ik Jun[3], Jae-Hyung Lee[3,4], Giovanni Quinones-Valdez[5], Chonghui Cheng[6], Xinshu Xiao[1,2,3,7,8,9*]

[1]Bioinformatics interdepartmental program, UCLA, Los Angeles, California, CA, USA
[2]Molecular, Cellular and Integrative Physiology Interdepartmental program, UCLA, Los Angeles, California, CA, USA
[3]Department of Integrative Biology and Physiology, UCLA, Los Angeles, California, CA, USA
[4]Department of Life and Nanopharmaceutical Sciences & Oral Microbiology, School of Dentistry, Kyung Hee University, Seoul, Korea
[5]Department of Bioengineering, UCLA, Los Angeles, California, CA, USA
[6] Lester & Sue Smith Breast Center & Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA
[7]Molecular Biology Institute, UCLA, Los Angeles, California, CA, USA
[8]Institute for Quantitative and Computational Sciences, UCLA, Los Angeles, California, CA, USA
[9]Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, California, CA, USA

*Correspondence: gxxiao@ucla.edu

**Abstract**

Recent studies revealed global shifts in RNA editing, the modification of RNA sequences, across many cancers. Besides a few sites implicated in tumorigenesis or metastasis, most tumor-associated sites, predominantly in noncoding regions, have unknown function. Here, we characterize editing profiles between epithelial (E) and mesenchymal (M) phenotypes in seven cancer types, as epithelial-mesenchymal transition (EMT) is a key paradigm for metastasis. We observe distinct editing patterns between E and M tumors and EMT induction upon loss of ADAR enzymes in cultured cells. E-M differential sites are highly enriched in genes involved in immune and viral processes, some of which regulate mRNA abundance of their respective genes. We identify a novel mechanism in which ILF3 preferentially stabilizes edited transcripts. Among editing-dependent ILF3 targets is the transcript encoding PKR, a crucial player in immune response. Our study demonstrates the broad impact of RNA editing in cancer and relevance of editing to cancer-related immune pathways.

**Introduction**

RNA editing, the modification of specific nucleotides in RNA sequences, expands diversity in proteins and gene regulatory mechanisms[1,2]. The most frequent type of RNA editing in human cells is A-to-I editing, which refers to the deamination of adenosine (A) to inosine (I) and is catalyzed by the Adenosine Deaminases Acting on RNA (ADAR) family of enzymes[3]. Three ADAR genes are encoded in the human genome, namely ADAR1, ADAR2 and ADAR3. Catalytically active ADAR1 and ADAR2 are widely expressed across tissues. In contrast, ADAR3 is exclusively expressed in certain brain regions and is catalytically inactive[4]. As inosine is recognized as guanosine (G) in translation and sequencing, A-to-I editing is also referred to as A-to-G editing. Though millions of editing events have been revealed across the human transcriptome, only a small proportion of editing events have been functionally characterized. The effects of most editing sites, primarily within non-coding regions, have yet to be discovered[5,6].

Increasing evidence has established the importance of RNA editing dysregulation in cancer. A number of studies have delineated mechanisms through which individual RNA editing sites, mostly causing recoding events (i.e., amino acid changes), promote or suppress tumor development[7–10]. Besides functioning in tumorigenesis, edited RNA transcripts can be translated into edited peptides, which may be recognized as cancer antigens and activate an anti-tumor immune response[11,12]. Furthermore, across various cancer types, genome-wide aberrations in RNA editing were observed and associated with clinical features[13–15]. Within each cancer type, editing levels generally increased or decreased in tumors, compared to matched normal samples. Editing levels of specific sites were correlated with tumor stage, subtype, and patient survival, and for a smaller subset of nonsynonymous sites, editing altered cell proliferation and drug sensitivity in cell line experiments[13]. As RNA editing has the potential to inform development of improved cancer diagnostics and patient-specific treatments, thorough investigation of the precise functions and regulatory mechanisms of the many cancer-type-specific RNA editing changes is needed[10].

In cancer progression, activation of epithelial-mesenchymal transition (EMT) facilitates metastasis by enabling tumor cells to gain an invasive phenotype, infiltrate the circulatory and lymphatic systems, and reach distant sites for colonization[16–18]. A few RNA editing sites have been associated with this process so far. Specifically, editing events in SLC22A3, FAK, COPA, RHOQ, and miR-200b were demonstrated to accelerate metastasis[12,19–23]. While miR-200b normally targets ZEB1 and ZEB2, which are key EMT-inducing transcription factors, editing alters its targets and enhances cell invasiveness and motility[23]. The SLC22A3 recoding event also promoted EMT, causing expression changes in EMT marker genes[19]. In contrast, a recoding event in GABRA3 inhibited metastasis and was present only in non-invasive cell lines and non-metastatic tumors[22]. These studies highlight the functional relevance of RNA editing in metastasis and the merit of a more comprehensive investigation.

Here, we present a global analysis and comparison of RNA editing profiles between epithelial (E) and mesenchymal (M) phenotypes of primary tumors across

3

multiple cancer types. Using RNA-seq data derived from bulk tumors and single cells, we observed distinct editing patterns between phenotypes, with editing differences often enriched among immune response pathway genes. Supported by experimental evidence, we show that differential editing sites affect host gene mRNA abundance and identify a novel mechanism of editing-dependent stabilization of mRNAs by ILF3. One of the target genes of ILF3 is EIF2AK2, coding for Protein Kinase R (PKR), a key player in immune and viral response.

**Results**

*Altered RNA editing profiles between epithelial and mesenchymal tumors*

EMT is known to be accompanied by substantial transcriptome remodeling[17,24–28]. Given the previously reported functional relevance of RNA editing in EMT[19,23,29], we hypothesized that epithelial and mesenchymal tumors possess different transcriptome-wide RNA editing profiles. Thus, we analyzed RNA-seq datasets of primary tumors from The Cancer Genome Atlas (TCGA). We focused on seven cancer types that have been previously studied in the context of EMT and have relatively large sample sizes available from TCGA (Fig. 1A). To classify tumors to epithelial (E) and mesenchymal (M) phenotypes, we utilized a well-established EMT scoring system, where scoring and categorization of tumors into these E and M phenotypes enabled systematic identification of cancer-specific differences in treatment response between phenotypes, as well as associations with survival[30]. Of all categorized tumors for each cancer type, we further refined the subset of tumors such that metadata were matched between the two groups (Supplementary Table 1).

Applying methods previously published from our lab[1,31,32], we quantified editing levels at over 4 million editing sites recorded in the REDIportal database[33]. We then identified sites that were differentially edited between E and M tumors in each cancer type. To control for false discoveries, we filtered out predicted differential editing sites that repeatedly exhibited differences in editing when phenotype labels were shuffled randomly. Principal components analysis on differential editing levels showed that E and M tumors could be well separated by the first two principal components of editing (Fig. 1A). These first two principal components did not appear to be confounded by sample metadata and suggest that most of the variation in editing is explained by the distinction of E and M phenotypes (Supplementary Fig. 1).

Based on the differential editing sites, most cancer types, including LUAD, LUSC, PRAD, KIRC and HNSC, demonstrated a hyperediting trend in the M phenotype (Fig. 1B). In contrast, two cancer types, BRCA and OV, had a trend of hypoediting in the M samples. The majority of differential editing sites in all cancer types were located in the 3' untranslated regions (UTRs) or introns (Fig. 1C). The above results suggest that distinct RNA editing profiles exist between E and M phenotypes.

*Editing patterns are shared among cancer types and distinct from differential expression*

Given dominant trends of hyperediting or hypoediting that distinguished E and M phenotypes in an individual cancer type, we asked whether genes with differential editing patterns were shared or distinct across cancer types. We examined the statistical significance of overlap in differentially edited genes between pairs of cancer types by Rank-rank Hypergeometric Overlap (RRHO). Extending Gene Set Enrichment Analysis (GSEA) to two dimensions, RRHO tests the significance of the intersection of gene lists, ranked by a metric of differential expression, across two genome-wide datasets[34]. We applied RRHO to RNA editing here by ranking genes according to the significance of tested editing differences between E and M. Beyond shared directionality, we found significant overlap in genes with editing changes among multiple cancer types (Fig. 2A). The significant overlap in genes based on differential editing suggests that editing changes in EMT may affect common pathways across cancer types.

It should be noted that differentially edited genes do not overlap with differentially expressed genes (Fig. 2B). This observation indicates that gene expression changes in EMT did not confound the RNA editing differences observed. Thus, altered editing potentially represent a distinct layer of molecular changes in EMT.

*Differential editing occurs in genes of immune relevance*

Next, we examined the gene ontologies enriched among genes with differential editing in EMT. In this analysis, background control genes were chosen randomly from those that did not have differential editing sites but had similar gene length and GC content as the differentially edited genes (Methods). Across multiple cancer types, differentially edited genes were enriched with viral-host interaction features, interferon (IFN) and other immune response pathways, metabolic processes and translational regulation (Fig. 2C).

The observation of immune-relevant categories is of particular interest. RNA editing has been described as a mechanism to label endogenous double-stranded RNAs and consequently prevent IFN induction[35–39]. However, the roles of editing events in genes directly associated with immune response such as those in the IFN response pathways have not been well characterized. Our observation indicates that RNA editing may directly affect immune response genes in EMT.

*Contribution of cell types to differential editing*

Given the observed enrichment of differential editing in immune-relevant genes, we asked whether our identified differential editing events primarily occur in cancer cells or in other cell types in the tumor microenvironment. To address this question, we analyzed single cell (sc) RNA-seq data from three non-small cell lung cancer (NSCLC) patients, each with three tumor samples from the tumor edge, core, and in-between[40]. Following quality control measures, we clustered the cells in two rounds and labeled cell

5

types based on marker genes to obtain T cells, B cells, myeloid cells, endothelial cells, fibroblasts, epithelial cells, mast cells, alveolar cells, and cancer cells (Supplementary Fig. 2A-C, Methods). Supporting the accuracy of this clustering, expression of marker genes was generally highest in their expected cell types when RPKM was calculated from pooled cells and when a signature gene expression matrix was predicted by CIBERSORTx[41] (Supplementary Fig. 2D).

Gauging the contribution of individual cell types to bulk tumor differential editing, we calculated the percent of differentially edited genes from the bulk tumor analysis that were expressed in each cell type (with cells of each type pooled together). Cancer cells expressed the highest proportion of genes that were differentially edited (Fig. 3A). As we would not expect cell types expressing few differentially edited genes to be strongly represented by bulk tumor differential editing, the editing differences observed among bulk tumors may be predominantly attributable to the cancer cells.

We next separated cancer cells to epithelial and mesenchymal cell clusters (Fig. 3B, Methods). Sampling epithelial cells to match mesenchymal cells in terms of cell number (200 cells) and metadata, we pooled cells within each phenotype together and detected RNA editing events (Supplementary Fig. 3). Although the scRNA-seq primarily sequences the 3' ends of mRNAs, a relatively small number of RNA editing events were still captured. We identified nine editing sites with significant differences between E and M (Fig. 3C). All nine differential sites exhibited higher editing levels in the M phenotype, which is consistent with the hyperediting trend in M observed in bulk LUAD and LUSC tumors (Fig. 1B). Notable differentially edited genes include RHOA, which is active in cell migration and is associated with metastasis in multiple cancer types[42–44], and ARL16, a reported negative regulator of RIG-I activity[45], consistent with the observed enrichment of immune-relevant genes that were differentially edited in bulk tumors. Overall, the findings from single cell data support the hypothesis that editing differences between bulk E and M tumors mainly reflect changes occurring in cancer cells.

*ADAR1 or ADAR2 knockdown induced EMT*

Given the differential editing profiles between E and M tumors, an important question is whether the editing changes are functionally relevant to EMT. To address this question, we first examined if changes in ADAR expression affect EMT. Using cell culture systems commonly employed in EMT studies, we carried out knockdown (KD) experiments of ADAR1 or ADAR2 in two cell lines, A549 and MCF10A, via siRNAs. Upon ADAR1 KD, A549 cells showed elongated spindle-like mesenchymal morphology (Fig. 4A). We also confirmed the loss of epithelial markers (E-cadherin and $\gamma$-Catenin) and gain of mesenchymal marker (Vimentin) in ADAR1 KD A549 cells (Fig. 4B). Similar results were observed upon ADAR2 KD in A549 cells (Fig. 4C-D) and reproducible in MCF10A cells (Fig. 4E-F). These findings suggest that loss of either catalytically active ADAR enabled EMT in the two cell lines. The phenotypic changes following ADAR2 KD are consistent with a previous report that ADAR2-deficiency can induce EMT in SW480 cells[29]. Together, these results indicate that knockdown of ADARs, which impairs RNA editing at genome scale, promotes EMT.

*Impact of RNA editing on mRNA abundance*

Given ADAR's primary role in RNA editing, we next asked how RNA editing may affect genes relevant to EMT, especially those related to immune response (Fig. 2C). Since a relatively large fraction of differential editing sites is located in 3' UTRs, we examined the hypothesis that these sites may affect mRNA abundance of their respective genes. Thus, we first calculated the correlation between editing levels and mRNA abundance for differentially edited sites observed in the E-M comparison. Using a regression model accounting for confounding factors including age, gender and race, we observed a total of 127 genes whose editing sites are significantly correlated with mRNA abundance (FDR<10%) in at least one type of cancer (Fig. 5A). In addition, among these genes, 77% (94 of 122 testable genes) demonstrated a significant correlation in at least one human tissue type based on a similar analysis of GTEx data, 78% (73/94 genes) of which showed the same direction of correlation between cancer and at least one GTEx tissue.

To further evaluate the regulatory role of RNA editing on mRNA abundance, we next examined the change in mRNA expression levels upon ADAR1 KD. We used ADAR1 KD RNA-Seq data from 5 cell lines: U87, HepG2, K562, HeLa and B cells[1,46,47], respectively. Out of the 127 edited genes identified above, 126 of them were detectable at an expression level of at least 1 FPKM (and edited) in at least one cell line (control or ADAR1 KD condition). Among them, 71% (89 genes, red dots, Fig. 5B) showed inverse correlation between ADAR1 KD and editing level coefficient in at least one cell line (Fig. 5B). These genes showed an enrichment of negative expression changes upon ADAR1 KD, indicating a likely stabilizing effect imposed by RNA editing (p = 2.7e-4, binomial test). Among expression-correlated editing sites in the 89 genes, 64% are located in 3' UTRs, a percentage that's significantly higher than that of E-M differential editing sites in general (p = 2.4e-4, Fig. 5C). We thus refer to the 89 genes as putative target genes whose expression is modulated by RNA editing (Supplementary Table 2).

Next, we experimentally validated the regulation of mRNA abundance by six editing sites within three genes: RNF24, RHOA, and MRPS16. We used a minigene reporter with bi-directional promoters for mCherry and eYFP[48] and cloned edited and unedited versions of each editing site and its surrounding 3' UTR region into the 3' UTR of mCherry. Using expression of eYFP as an internal control, we compared mCherry expression between cells carrying the edited and unedited versions for each editing site. All six editing sites induced significant expression differences in the direction consistent with the editing-expression correlations observed in primary tumors (Fig. 5D, Supplementary Table 3). While positive editing associations were dominant among predicted target genes, there also exist negative associations between editing and expression levels. We tested one example of the latter category (RHOA).

7

*ILF3 as an editing-dependent regulator of mRNA abundance*

Since mRNA stability is closely regulated by RNA-binding proteins (RBPs)[49–52], we next asked whether RBPs are involved in the modulation of mRNA abundance by RNA editing sites. To this end, we analyzed enhanced ultraviolet crosslinking and immunoprecipitation (eCLIP) datasets of 126 RBPs in two cell lines (HepG2 and K562) from ENCODE[46,53]. We asked whether RBP binding signals are enriched significantly closer to editing sites in the 89 potential target genes than expected by chance. This analysis identified ILF3 as a top protein with significantly short distances to the editing sites in both cell lines (Supplementary Fig. 4A). To validate this finding and test this relationship in a different cell type, we performed eCLIP-seq of ILF3 in A549 cells. The same observation was made via this data set (Fig. 6A). As observed in HepG2 and K562 cells, differential editing sites within predicted target genes were significantly closer to ILF3 binding regions in A549 cells than random gene-matched control sites. Furthermore, 75 (84%) of the 89 genes showed a significant correlation between their gene expression and the expression of ILF3 (FDR<10%), 37 of which had an absolute correlation coefficient of at least 0.2 (Fig. 6B). Importantly, the majority of the significant correlations were positive, consistent with the known roles of ILF3 in stabilizing its target mRNAs[54–56].

*Impact of ILF3 on immune-relevant genes*

ILF3 promotes an antiviral response through its binding to RNAs[57–59]. Given the fact that immune-relevant genes are differentially edited in E-M (Fig. 2C), we next asked whether ILF3 regulates the mRNA abundance of these EMT-associated differentially edited, immune-relevant genes. Among the 89 genes whose expression was affected by RNA editing, 20 genes fall into the immune or viral GO categories. Interestingly, the ILF3 binding sites were significantly closer to the differential editing sites of these 20 genes than differential sites in immune-related genes without editing-expression associations (Fig. 6C). Together, these results suggest that ILF3 binds close to the editing sites of immune-related genes.

Since we observed that differential editing between bulk E and M tumors mainly reflected changes occurring in cancer cells (Fig. 3A), we next asked whether the above regulatory relationship between ILF3 and immune-related genes also occurs in cancer cells. To this end, we analyzed gene expression of individual cell types identified in the NSCLC scRNA-seq dataset. Within each cell type, we correlated ILF3 expression with expression of the 20 immune-related target genes. In cancer cells, all 20 genes had expression levels positively correlated with ILF3 expression at 10% FDR (Fig. 6D). Though significant correlations were also observed in other cell types, only cancer cells showed correlation coefficients of at least 0.2 in magnitude. This result suggests that the mRNA stabilizing function of ILF3 is prominent in cancer cells, in line with our observation that E-M differential editing primarily occurs in cancer cells.

*PKR expression is affected by 3' UTR editing through ILF3 regulation*

Among the 20 immune-related genes putatively regulated by ILF3, the gene EIF2AK2, coding for Protein Kinase R (PKR), had most significant expression-editing correlation (Supplemental Table 2) and expression correlation with ILF3 (Fig. 6D). Activated by dsRNA, PKR suppresses translation and promotes apoptosis through its phosphorylation activity[60,61]. PKR also regulates various signaling pathways, such as NF-κB and p38 MAPK, in response to cellular stress[60]. Using the editing minigene reporter, we examined the effects of seven 3' UTR editing sites on EIF2AK2 mRNA abundance in A549 cells. Five of the seven editing sites showed significantly higher normalized mCherry expression compared to their unedited counterparts (Fig. 6E, Supplementary Fig. 4B), suggesting that the editing sites enhanced mRNA abundance, consistent with the positive editing-expression correlation in primary tumors.

Based on the eCLIP data, all five editing sites are located within ILF3 binding sites (Fig. 6F, Supplementary Fig. 4C-D). To test the hypothesis that ILF3 regulates PKR mRNA abundance in an editing-dependent manner, we generated ILF3 KD A549 cells (Fig. 6G). The edited and unedited reporters, demonstrating differential expression in control cells, no longer produced different expression levels upon ILF3 KD (Fig. 6H). These results support the hypothesis that ILF3 mediates regulation of mRNA expression levels by 3' UTR RNA editing for PKR.

## Discussion

As most cancer patient deaths are due to metastasis, thorough understanding of the molecular mechanisms underlying metastasis is crucial to developing effective preventative measures[62]. EMT plasticity is thought to underlie cell dissemination and metastatic formation in many cancer types[18]. Supported by studies on primary tumors and various model systems, features of EMT have been associated with metastasis[16,18,63,64]. For instance, higher expression of mesenchymal markers, with preserved epithelial markers in the absence of nearly all canonical EMT transcription factors, was detected in cells located at the leading edge of primary human HNSC tumors[64]. Furthermore, this partial EMT program was correlated with multiple metastatic characteristics, including abundance of lymph node metastases, lymphovascular invasion, and tumor grade[64]. While mutations are understood to drive primary tumorigenesis and are often found in reported oncogenes and tumor suppressor genes, the existence of recurrently mutated genes specific to metastasis is not clear[18]. Accordingly, mechanisms regulating cell invasiveness beyond genetic variation need to be more thoroughly investigated. Our study is the first to report a systematic characterization of RNA editing in EMT phenotypes across several cancer types. Through a combination of experimental and computational analyses, we observed many editing differences in EMT-relevant genes, especially those related to immune and viral response, with the potential of affecting mRNA abundance of these genes. We also show that higher expression levels of these edited transcripts may be due to stabilization by ILF3.

Located in noncoding regions, most editing sites have unknown function. To assess the contribution of differential editing to altered cell phenotypes in cancer, we focused on the capacity of editing to regulate host gene mRNA abundance. To our knowledge, very few studies have examined this question on the transcriptome-wide scale[65,66]. Previously, several studies demonstrated this regulatory role for a handful of editing sites through alteration of miRNA binding sequences or mRNA secondary structure or otherwise unknown mechanisms[6,20,67–72]. Expanding on these previous studies, we incorporated tissue-rich data from GTEx and ADAR KD expression changes from five cell lines to computationally support associations of editing with mRNA abundance. We also validated the effects of specific editing sites and explored the involvement of RBPs in this regulatory mechanism. It should be noted that we were able to detect associations between editing and mRNA abundance levels, even though differentially expressed genes did not significantly overlap differentially edited genes. These findings do not contradict each other because editing levels are relatively low. Consequently, inosine may affect mRNA abundance, but when present at low levels, may not necessarily lead to significant expression differences.

Tumor heterogeneity may help explain the seemingly divergent roles of ADARs and their editing activity in this study. Our cell line experiments showed EMT induction upon KD of either ADAR1 or ADAR2. Loss of ADAR2 was also previously reported to trigger EMT by mediating the extracellular shuttling of EMT-repressing miR-200s[29]. Given these findings for both enzymatic ADARs, we might expect to observe hypoediting in the M phenotype (which we observed for BRCA and OV). Conversely, another study showed that ADAR1 knockdown diminished EMT phenotypes in oral squamous cell carcinoma cell lines with unknown mechanism[73]. As primary tumors are heterogenous not only across patients but also within an individual tumor, in terms of genetic variation among malignant cells and cell type composition, hyperediting or hypoediting trends in M-categorized tumors may reflect the diverse nature of primary tumors in addition to the pleiotropy of ADARs[74].

Considering tumor heterogeneity and the roles of stromal and immune cells in EMT, it is important to examine the contributions of different cell types to differential editing observed in the E-M comparisons. Our results using single-cell data supported that cancer cells are a main cell type underlying differential editing between E and M phenotypes in lung cancer, although contributions by other cell types cannot be excluded. Furthermore, cancer cells demonstrated the strongest expression correlation between ILF3 and immune-relevant differentially edited genes among all cell types considered in lung cancer. These findings suggest that RNA editing is likely an important aspect of transcriptome remodeling of cancer cells in EMT, at least in lung cancer. Single-cell analysis of RNA editing in other cancer types should be conducted in the future.

RNA editing is known to be important to innate immunity by preventing viral dsRNA sensors, such as MDA5 and RIG-I, from sensing host dsRNA[35,39,75]. In this study, we provided multiple lines of evidence to support that RNA editing differences in EMT may affect immune response genes directly, adding a new dimension to the

relationships between RNA editing and innate immunity. Interestingly, a major RBP that mediates this relationship is ILF3. ILF3 was identified as a PKR substrate and serves as a negative regulator of viral replication upon phosphorylation[76,77]. Upon viral infection and sensing viral dsRNA, PKR activates, suppresses translation, and promotes apoptosis of affected cells[61]. Importantly, this mechanism has been targeted in oncolytic virotherapy for cancer. Cancer cells that have low PKR expression are sensitive to oncolytic viruses[78–80]. Our study showed that ILF3 mediates the RNA editing-dependent regulation of PKR expression, suggesting that these two factors are intertwined in gene regulation. Additional studies on their interaction during viral infection or cancer treatment will be informative for therapeutic development. Previously, ADAR1 loss has been shown to render tumor cells sensitive to immunotherapy through enhanced inflammatory response[81,82]. Our findings on the regulation of immune response genes by RNA editing may add additional mechanisms in this process that will need further investigation.

The functional roles of RNA editing in cancer have been increasingly recognized in recent years. Highlighting the extensive editing differences between EMT phenotypes and their impact on mRNA abundance, especially for genes involved in the immune response, our work extends the basis for future studies on the contribution of editing to metastasis and patient outcomes.

**Methods**

*Plasmid construction*

For bi-directional reporters, full length or partial 3' UTR regions (1~2kb) of candidate genes were cloned from the genomic DNA extracted from HMLE or A549 cells. Edited versions of 3' UTR inserts were generated using overlap-extension PCR (Supplementary table 3). Edited and unedited versions of 3' UTR regions were then cloned into the pTRE-BI-red/yellow vector via ClaI and SalI-HF enzyme sites[48]. To obtain a lentiviral vector expressing ILF3 shRNA, oligos containing the target sequence (GGTCTTCCTAGAGCGTATAAA, TRCN0000329788) were ordered from Integrated DNA Technologies (IDT) and cloned into pLKO.1 via EcoRI and AgeI enzyme sites.

*Cell culture and transfection*

A549, Hela and HEK293T cells were maintained in DMEM with 10% FBS and Antibiotic-Antimycotic reagent (Gibco). MCF10A cells were maintained in DMEM/F12, supplemented with 5% Horse serum, 20ng/ml human EGF (PeproTech), 0.5mg/ml Hydrocortisone (Sigma), 100ng/ml Cholera Toxin (Sigma), 10ug/ml Insulin (Sigma), and Antibiotic-Antimycotic reagent (Gibco). For siRNA treatment, A549 or MCF10A cells were seeded at $1\times10^5$cells per well in 6-well plates. After 24 hours, siRNAs (Supplementary table 3) were introduced at the final concentration of 10nM~100nM using lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's protocol. Media were changed 24 hours post-transfection, and cells were harvested 72 hours post-transfection. For transfection of bi-directional reporters, Hela and HEK293T cells

were seeded in 12-well plates to reach 90% confluency by the time of transfection. A549 cells were seeded at $0.15 \times 10^5$ cells per well in 12-well plates 24 hours before transfection. Reporter plasmids were transfected at 200ng per 12-well with lipofectamine 3000 (Invitrogen), following the manufacturer's protocol. Cells were harvested 16 hours post-transfection.

*Western blot*

Cells were lysed with RIPA buffer containing protease inhibitor (EDTA-free, Thermo Fisher Scientific) at 4°C for 30 minutes. The whole cell lysates were then centrifuged at 12,000g, 4°C for 15 minutes. The supernatants were collected for protein concentration measurement using Bradford assay (Pierce™ Detergent Compatible Bradford Assay Kit, Thermo Fisher Scientific). Protein samples were prepared by mixing protein lysates with 4x SDS protein loading dye at 3:1 ratio. The mixture was boiled for 5 minutes. 10 ug of each protein samples were loaded on SDS-PAGE gels and transferred to nitrocellulose membranes for antibody incubations. Antibodies used were as follows: ADAR1 antibody (Santa Cruz Biotechnology, sc-73408, 1:200), ADAR2 antibody (Santa Cruz Biotechnology, sc-73409, 1:200), E-cadherin antibody (Cell Signaling Technology, #3195, 1:1000), γ-Catenin antibody (BD Transduction Laboratories, 610253, 1:8000), N-cadherin antibody (BD Transduction Laboratories, 610920, 1:500), Vimentin antibody (Cell Signaling Technology, 5741, 1:1000), NF90(ILF3) antibody (BETHYL Laboratories, A303-651A, 1:1000), β-actin-HRP antibody(Santa Cruz Biotechnology, sc-47778, 1:2000), goat anti-rabbit IgG-HRP(Santa Cruz Biotechnology, sc-2004, 1:2000), goat anti-mouse IgG-HRP(Santa Cruz Biotechnology, sc-2005, 1:2000). Membrane blots were incubated with SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific) and visualized under the imager (Syngene PXi).

*RNA isolation and real-time qPCR*

Cells were lysed using TRIzol (Thermo Fisher Scientific). Total RNA was isolated using Direct-zol RNA Miniprep Plus kit (Zymo Research) following the manufacturer's protocol. 2 ug of total RNA was used for cDNA synthesis with SuperScript IV (Thermo Fisher Scientific). The real-time qPCR reaction was assembled using the PowerUp™ SYBR® Green Master Mix (Thermo Fisher Scientific). The reaction was performed in the CFX96 Touch Real-Time PCR detection system (Bio-Rad) with the following settings: 50°C for 10 minutes, 95°C for 2 minutes, 95°C for 15 seconds, 60°C for 30 seconds, and with the last two steps repeated for 45 cycles. For bi-directional reporter assays, mCherry expression was normalized against eYFP expression within the same sample. ILF3 expression was normalized against the expression of internal control gene *TBP*. For qPCR validating the eCLIP peaks, the final libraries were diluted to the same concentration at 0.01ng/ul. 5ul of diluted libraries were used in each qPCR reaction. Around 80 bp upstream each EIF2AK2 editing site was amplified. The expression of each EIF2AK2 region was normalized against the expression of 18s.

*Categorization of tumors as epithelial and mesenchymal*

12

We downloaded fragments per kilobase million (FPKM) data of primary tumors from patients across seven cancer types from TCGA: BRCA, LUAD, LUSC, PRAD, OV, KIRC, and HNSC (GDC Data Portal; https://portal.gdc.cancer.gov/). To assess E and M phenotypes of the tumors of each cancer type, we quantified the enrichment of E and M gene sets by applying gene set variation analysis (GSVA)[83]. We obtained pan-cancer E and M gene sets from a 2014 publication by Tan and colleagues (Table S1A from their publication)[30]. Tumors with high E scores and low M scores were considered to have an E phenotype, while tumors with low E and high M scores were classified as M. Subsets of E and M tumors were selected for each cancer type to minimize confounding of E and M distinction by patient and sample metadata.

*Quantification and comparison of RNA editing levels*

We downloaded RNA-seq fastq files of categorized tumors from the GDC Legacy Archive (https://portal.gdc.cancer.gov/legacy-archive). We mapped reads to hg19 with HISAT2, using default parameters. Dense clusters of editing sites, or hyperedited regions, can lead to many mismatches in reads. Consequently, these reads may be left unmapped and hinder accurate detection of editing in these regions. To rescue reads that were originally unmapped due to high density of editing activity, we applied a hyperediting pipeline and combined the recovered reads with uniquely mapped reads for downstream analyses[32,84]. To analyze editing sites of high confidence, we downloaded the REDIportal database, comprising over 4 million editing sites identified across 55 tissues of 150 healthy humans from GTEx (http://srv00.recas.ba.infn.it/atlas/)[33]. We applied methods used in our previous studies to detect editing at REDIportal sites in the tumor samples. We filtered out editing sites found in dbSNP (version 147) and COSMIC (version 81), except for reported cancer-related editing sites[8,13,19,85–88], since editing sites have been shown to be mistakenly recorded as SNPs[89,90]. Within each sample, we also filtered out editing events that overlapped with sample-specific somatic mutations and copy number variants. Somatic variants were obtained from the publicly released MC3 MAF (https://gdc.cancer.gov/about-data/publications/mc3-2017), and copy number variants were obtained from copy number segment data downloaded from the GDC data portal.

Differential editing sites were defined as editing sites with significantly different editing levels between E and M phenotypes. To identify such sites, we used an adaptive coverage approach[32]. For an individual editing site, we determined the highest read coverage threshold that was satisfied in at least five samples of both phenotypes, among twenty, fifteen, and ten reads. If none of these thresholds was satisfied and fewer than ten samples in each phenotype had at least five reads covering the site, we did not test the site for differential editing. Using the highest coverage determined, we calculated the mean editing levels among samples of each phenotype separately. We then consecutively lowered the read coverage threshold by 5 reads and compared the new mean editing levels of each phenotype, when including additional samples, to the original high-coverage-only editing means. If the differences in mean editing levels were less than 0.03, we used the lower read coverage threshold to delineate which samples

to include for the differential test. Editing levels between E and M samples were compared by a Wilcoxon rank-sum test. Editing differences were considered significant if the Wilcoxon p-value < 0.05 and the magnitude of the difference > 0.05. To account for false positives, we shuffled phenotype labels and retested for significant differences for each differential editing site, 100 times. If a site showed significant differences for shuffled labels over ten times, it was filtered out and no longer considered a differential editing site.

*Identification of differentially expressed genes*

HTSeq-Count data were downloaded from the GDC data portal. We identified genes with significantly different mRNA expression levels between E and M tumors of each cancer type, using limma-voom[91]. Metadata significantly correlated with the top two principal components of expression were included as covariates in the linear models. Expression differences were considered significant if log2-fold change was at least 1 and adjusted p-value was less than 0.05.

*Rank-rank hypergeometric overlap*

To measure the similarity in patterns of editing changes across cancer types, we ranked genes based on differential editing between E and M phenotypes for each cancer type. More specifically, the ranking metric was the statistical significance of the differential editing test (-log10(Wilcoxon p-value)), multiplied by the sign of the editing difference (mean of M editing levels – mean of E editing levels). Accordingly, genes at the top of the ranked list had the highest increases in editing in M, while genes at the bottom had the largest decreases in editing in M. For each gene with multiple editing sites tested, the site with the most significant change in editing levels was used to represent the gene. We used the RRHO package within Bioconductor to test for significance of overlap between ranked gene lists, with a step size of 30 genes between each rank[92]. To make RRHO maps comparable across pairs of cancer types, we scaled the log-transformed p-values to account for different lengths of gene lists and then applied the Benjamini-Yekutieli correction for multiple testing[34].

We also ran RRHO between gene rankings by differential editing and differential gene expression for each cancer type. To order genes based on differential gene expression, genes were ranked according to the signed statistical significance of differential expression tests (signed by the direction of expression change in M). As a result, genes at the top of the list were more highly expressed in M and genes at the bottom, more lowly expressed in M.

*Gene ontology enrichment analysis*

To evaluate whether an individual GO term was enriched in differential editing in one cancer type, we compared the occurrence of the term among query genes – genes containing differential editing sites – to its occurrences within 10,000 sets of control genes. In each set, one control gene was randomly selected for each query gene

14

among non-differentially edited genes that matched the query gene based on gene length and GC content (within 10%). Query genes that did not have at least ten matched control genes were excluded. We calculated the p-value of the term's enrichment among query genes from the normal distribution fit to occurrences of the term among control gene sets.

Likewise, we tested the occurrence of each GO term represented among differentially expressed genes to its occurrences among 10,000 sets of non-differentially expressed control genes, randomly selected to match the differentially expressed query genes for gene length and GC content.

*scRNA-seq dataset analysis*

We downloaded fastq files from 15 tumor samples of five NSCLC patients (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6149) and ran CellRanger (version 3.0.2) to map reads and obtain count matrices. We excluded the tumor samples from three LUSC patients exhibiting low percentages of valid barcodes and mapped reads. For the remaining samples, we loaded the filtered feature-barcode matrices from CellRanger and merged the datasets into a single Seurat object with the R package Seurat[93] (version 3.0.2). Next, we filtered out cells that did not meet the following criteria: 101-6000 expressed genes, over 200 UMIs, and less than 10% UMIs corresponding to the mitochondrial genome. Following normalization by sctransform[94] (version 0.2.0), we performed dimensional reduction with PCA. Based on an elbow plot, we decided to consider the first ten PCs for downstream clustering and TSNE embedding. To assign cell identity labels to clusters, we matched differentially expressed genes of clusters to reported marker genes. One cluster had differentially expressed markers of multiple cell types, so we subclustered its cells. To assess the accuracy of our final labeling of nine cell types, we examined expression of marker genes across the cell types in two approaches. In one approach, we used CIBERSORTx (https://cibersortx.stanford.edu/) to generate a gene expression signature matrix, which is a matrix of expression signatures characterizing cell types. To create this matrix from expression profiles of single cells labeled by cell type, CIBERSORTx identified differentially expressed genes. In the second approach, we pooled reads of each cell type together and calculated FPKM. These FPKM values calculated from pooled cells were also used to correlate ILF3 expression with expression of editing-correlated genes.

To identify cancer cells with E and M phenotypes, we subclustered the tumor cells. To this end, we first ran sctransform and PCA on only the tumor cells. Using the first twelve PCs, we clustered the cells and performed non-linear dimension reduction by UMAP. As a cluster of 200 M cells was identified, we sampled 200 E cells with similar numbers of features, numbers of UMIs, and percentages of reads mapped to the mitochondrial genome. For each phenotype, we compiled reads of cells together and detected editing levels at REDIportal sites. For each testable editing site, E and M editing levels were compared by a Fisher's Exact test. An editing site was considered

differential if the difference in editing levels was at least 0.05 and the Fisher's Exact p-value < 0.05.

*Regression analysis*

For each differential editing site, association between editing level and host gene mRNA abundance was tested by fitting a linear model of log-transformed gene FPKM against editing level and potentially confounding covariates (using the lm function in R). For associations in GTEx data, we included age, gender, and race as covariates. For associations in TCGA data, we included metadata that were significantly correlated with the top two principal components of expression, as in the differential expression analysis.

*eCLIP-seq generation*

The eCLIP experiment was performed following a published protocol[53]. The antibody used for this experiment is: ILF3/NF90 antibody (Bethyl Laboratories, A303-651A). $10^7$ cells of A549 were ultraviolet crosslinked at 254 nm (800 mJ cm$^{-2}$). Crosslinked sample was then used for each eCLIP experiment, which was performed as described[53]. For controls, we sampled 2% of the pre-immunoprecipitation (post lysis and fragmentation) sample and prepared libraries identically to the ILF3 eCLIP (including the membrane size selection step). These libraries served as "size-matched input" (SMInput) to minimize nonspecific background signal in the identical size range on the membrane as well as any inherent biases in ligations, PCR with reverse transcription (RT-PCR), gel migration and transfer steps.

*eCLIP-seq peak calling and distance analysis*

We obtained eCLIP peak data for 96 RBPs in K562, 83 RBPs in HepG2, and ILF3 in A549 cells, as described previously[46]. Briefly, after demultiplexing and trimming adapters, we aligned reads in multiple rounds with STAR. First, reads aligning to rRNA sequences were discarded, and then the unmapped reads were aligned to Alu sequences, permitting a maximum of 100 alignments for an individual read. In the final alignment step, the remaining unmapped reads were uniquely aligned to the hg19 genome. Then read enrichment within a sliding window, considering both genome and Alu-aligned reads, was tested for significance by a Poisson model in order to call eCLIP peaks[46,74].

To assess the proximity of a single RBP's binding to differential editing sites compared to random controls, we calculated the distance from each differential editing site or control to the closest eCLIP peak in the same gene. Control sites consisted of adenosines within genes containing differential editing sites[32]. We then calculated the area under the curve (AUC) of the cumulative distribution of distances from differential editing sites to the closest eCLIP peaks. Given our interest in close binding, we considered distances up to 10,000 bases only for AUC calculation. Similarly, we calculated the AUC of the distribution of closest distances between eCLIP peaks and

16

controls, for each of 10,000 sets of random controls. We computed the p-value of the AUC for differential editing sites from the normal distribution fit to the AUC values of control sets[32].

## Acknowledgements

## References

1.  Bahn, J. H. *et al.* Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res.* **22**, 142–150 (2012).
2.  Tan, M. H. *et al.* Dynamic landscape and regulation of RNA editing in mammals. *Nat. Publ. Gr.* **550**, 249–254 (2017).
3.  Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.* **17**, 83–96 (2016).
4.  Cho, D. S. C. *et al.* Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. *J. Biol. Chem.* **278**, 17093–17102 (2003).
5.  Walkley, C. R. & Li, J. B. Rewriting the transcriptome: adenosine-to- inosine RNA editing by ADARs. *Genome Biol.* **18**, 1–13 (2017).
6.  Brümmer, A., Yang, Y., Chan, T. W. & Xiao, X. Structure-mediated modulation of mRNA abundance by A-to-I editing. *Nat. Commun.* **8**, 1–12 (2017).
7.  Chen, L. *et al.* Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. *Nat. Med.* **19**, 209–216 (2013).
8.  Xu, L.-D., Öhman, M., Xu, L.-D. & Öhman, M. ADAR1 Editing and its Role in Cancer. *Genes (Basel).* **10**, (2019).
9.  Kung, C.-P., Maggi, L. B. & Weber, J. D. The Role of RNA Editing in Cancer Development and Metabolic Disorders. *Front. Endocrinol. (Lausanne).* **9**, 762 (2018).
10. Xu, X., Wang, Y. & Liang, H. The role of A-to-I RNA editing in cancer development. *Curr. Opin. Genet. Dev.* **48**, 51–56 (2018).
11. Zhang, M. *et al.* RNA editing derived epitopes function as cancer antigens to elicit

immune responses. *Nat. Commun.* **9**, 3919 (2018).

12. Peng, X. *et al.* A-to-I RNA Editing Contributes to Proteomic Diversity in Cancer. *Cancer Cell* 1–12 (2018). doi:10.1016/j.ccell.2018.03.026

13. Han, L. *et al.* The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. *Cancer Cell* **28**, 515–528 (2015).

14. Paz-Yaacov, N. *et al.* Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors. *Cell Rep.* **13**, 267–276 (2015).

15. Fumagalli, D. *et al.* Principles Governing A-to-I RNA Editing in the Breast Cancer Transcriptome. *Cell Rep.* **13**, 277–289 (2015).

16. Heerboth, S. *et al.* EMT and tumor metastasis. *Clin. Transl. Med.* **4**, 6 (2015).

17. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial–mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **15**, 178–196 (2014).

18. Dongre, A. & Weinberg, R. A. New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. *Nat. Rev. Mol. Cell Biol.* **20**, 69–84 (2019).

19. Fu, L. *et al.* RNA editing of SLC22A3 drives early tumor invasion and metastasis in familial esophageal cancer. *PNAS* **114**, 4631–4640 (2017).

20. Amin, E. M. *et al.* The RNA-editing enzyme ADAR promotes lung adenocarcinoma migration and invasion by stabilizing FAK. *Sci. Signal.* **10**, (2017).

21. Han, S.-W. *et al.* RNA editing in RHOQ promotes invasion potential in colorectal cancer. *J. Exp. Med.* **211**, 613–21 (2014).

22. Gumireddy, K. *et al.* The mRNA-edited form of GABRA3 suppresses GABRA3-mediated Akt activation and breast cancer metastasis. *Nat. Commun.* **7**, 10715 (2016).

23. Wang, Y. *et al.* Systematic characterization of A-to-I RNA editing hotspots in microRNAs across human cancers. *Genome Res.* **27**, 1112–1125 (2017).

24. Harvey, S. E. *et al.* Coregulation of alternative splicing by hnRNPM and ESRP1 during EMT. *RNA* **24**, 1326–1338 (2018).

25. Warzecha, C. C. & Carstens, R. P. Complex changes in alternative pre-mRNA splicing play a central role in the epithelial-to-mesenchymal transition (EMT). *Semin. Cancer Biol.* **22**, 417–427 (2012).

26. Brown, R. L. *et al.* CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *J. Clin. Invest.* **121**, 1064–1074 (2011).

27. Xu, Y. *et al.* Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. *Genes Dev.* **28**, 1191–1203 (2014).

28. Hu, X. *et al.* The RNA-binding protein AKAP8 suppresses tumor metastasis by antagonizing EMT-associated alternative splicing. *Nat. Commun.* **11**, 486 (2020).

29. Shelton, P. M. *et al.* The Secretion of miR-200s by a PKCζ/ADAR2 Signaling Axis Promotes Liver Metastasis in Colorectal Cancer. *Cell Rep.* **23**, 1178–1191 (2018).

30. Tan, T. Z. *et al.* Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol. Med.* **6**, 1279–93 (2014).

31. Lee, J., Ang, J. K. & Xiao, X. Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. *RNA*

**19**, 725–732 (2013).

32. Tran, S. S. *et al.* Widespread RNA editing dysregulation in brains from autistic individuals. *Nat. Neurosci.* **22**, 25–36 (2019).

33. Picardi, E., D'Erchia, A. M., Lo Giudice, C. & Pesole, G. REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.* **45**, D750–D757 (2017).

34. Plaisier, S. B., Taschereau, R., Wong, J. A. & Graeber, T. G. Rank-rank hypergeometric overlap: Identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res.* **38**, 1–17 (2010).

35. Liddicoat, B. J. *et al.* RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349**, 1115–20 (2015).

36. George, C. X., Ramaswami, G., Li, J. B. & Samuel, C. E. Editing of Cellular Self-RNAs by Adenosine Deaminase ADAR1 Suppresses Innate Immune Stress Responses. *J. Biol. Chem.* **291**, 6158–68 (2016).

37. Chung, H. *et al.* Human ADAR1 Prevents Endogenous RNA from Triggering Translational Shutdown. *Cell* **172**, 811-824.e14 (2018).

38. Savva, Y. A., Rezaei, A., Laurent, G. S. & Reenan, R. A. Reprogramming, Circular Reasoning and Self versus Non-self: One-Stop Shopping with RNA Editing. *Front. Genet.* **7**, 1–8 (2016).

39. Mannion, N. M. *et al.* The RNA-Editing Enzyme ADAR1 Controls Innate Immune Responses to RNA. *Cell Rep.* **9**, 1482–1494 (2014).

40. Lambrechts, D. *et al.* Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **24**, 1277–1289 (2018).

41. Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).

42. Zaoui, K. *et al.* Ran promotes membrane targeting and stabilization of RhoA to orchestrate ovarian cancer cell invasion. *Nat. Commun.* **10**, 2666 (2019).

43. Ridley, A. Rho GTPases and cell migration. *J. Cell Sci.* 2713–2722 (2001).

44. Yu, X. *et al.* CXCL12/CXCR4 promotes inflammation-driven colorectal cancer progression through activation of RhoA signaling by sponging miR-133a-3p. *J. Exp. Clin. Cancer Res.* **38**, 32 (2019).

45. Yang, Y.-K. *et al.* ARF-like protein 16 (ARL16) inhibits RIG-I by binding with its C-terminal domain in a GTP-dependent manner. *J. Biol. Chem.* **286**, 10568–80 (2011).

46. Quinones-Valdez, G. *et al.* Regulation of RNA editing by RNA-binding proteins in human cells. *Commun. Biol.* **2**, 19 (2019).

47. Wang, I. X. *et al.* ADAR Regulates RNA Editing, Transcript Stability, and Gene Expression. *Cell Rep.* **5**, 849–860 (2013).

48. Yang, E.-W. *et al.* Allele-specific binding of RNA-binding proteins reveals functional genetic variants in the RNA. *Nat. Commun.* **10**, 1338 (2019).

49. Glisovic, T., Bachorik, J. L., Yong, J. & Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters* **582**, 1977–1986 (2008).

50. Mayr, C. What are 3′ utrs doing? *Cold Spring Harb. Perspect. Biol.* **11**, (2019).

51. Dassi, E. Handshakes and Fights: The Regulatory Interplay of RNA-Binding Proteins. *Front. Mol. Biosci.* **4**, 67 (2017).

52. Turner, M. & DÍaz-Muñoz, M. D. RNA-binding proteins control gene expression

and cell fate in the immune system. *Nature Immunology* **19**, 120–129 (2018).

53. 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).

54. Shim, J., Lim, H., R.Yates, J. & Karin, M. Nuclear Export of NF90 Is Required for Interleukin-2 mRNA Stabilization. *Mol. Cell* **10**, 1331–1344 (2002).

55. Vumbaca, F., Phoenix, K. N., Rodriguez-Pinto, D., Han, D. K. & Claffey, K. P. Double-stranded RNA-binding protein regulates vascular endothelial growth factor mRNA stability, translation, and breast cancer angiogenesis. *Mol. Cell. Biol.* **28**, 772–83 (2008).

56. Kuwano, Y. *et al.* MKP-1 mRNA stabilization and translational control by RNA-binding proteins HuR and NF90. *Mol. Cell. Biol.* **28**, 4562–75 (2008).

57. Harashima, A., Guettouche, T. & Barber, G. N. Phosphorylation of the NFAR proteins by the dsRNA-dependent protein kinase PKR constitutes a novel mechanism of translational regulation and cellular defense. *Genes Dev.* **24**, 2640–53 (2010).

58. Castella, S., Bernard, R., Corno, M., Fradin, A. & Larcher, J.-C. Ilf3 and NF90 functions in RNA biology. *WIREs RNA* **6**, 243–256 (2015).

59. Li, X. *et al.* Coordinated circRNA Biogenesis and Function with NF90/NF110 in Viral Infection. *Mol. Cell* **67**, 214-227.e7 (2017).

60. Garcia, M. A. *et al.* Impact of Protein Kinase PKR in Cell Biology: from Antiviral to Antiproliferative Action. *Microbiol. Mol. Biol. Rev.* **70**, 1032–1060 (2006).

61. Gal-Ben-Ari, S., Barrera, I., Ehrlich, M. & Rosenblum, K. PKR: A kinase to remember. *Frontiers in Molecular Neuroscience* **11**, (2019).

62. Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging Biological Principles of Metastasis. *Cell* **168**, 670–691 (2017).

63. Aiello, N. M. *et al.* EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. *Dev. Cell* **45**, 681–695 (2018).

64. Puram, S. V *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611–1624 (2017).

65. Sharpnack, M. F. *et al.* Global Transcriptome Analysis of RNA Abundance Regulation by ADAR in Lung Adenocarcinoma. *EBioMedicine* **27**, 167–175 (2018).

66. Gu, T., Fu, A. Q., Bolt, M. J. & White, K. P. Clinical Relevance of Noncoding Adenosine-to-Inosine RNA Editing in Multiple Human Cancers. *JCO Clin. cancer informatics* **3**, 1–8 (2019).

67. Borchert, G. M. *et al.* Adenosine deamination in human transcripts generates novel microRNA binding sites. *Hum. Mol. Genet.* **18**, 4801–4807 (2009).

68. Wang, Q. *et al.* ADAR1 regulates ARHGAP26 gene expression through RNA editing by disrupting miR-30b-3p and miR-573 binding. *RNA* **19**, 1525–1536 (2013).

69. Zhang, L., Yang, C., Varelas, X. & Monti, S. Altered RNA editing in 3′ UTR perturbs microRNA-mediated regulation of oncogenes and tumor-suppressors. *Sci. Rep.* 1–13 (2016). doi:10.1038/srep23226

70. Ma, C. *et al.* ADAR1 promotes robust hypoxia signaling via distinct regulation of multiple HIF-1α-inhibiting factors. *EMBO Rep.* **20**, (2019).

71. Sagredo, E. A. *et al.* ADAR1-mediated RNA-editing of 3′UTRs in breast cancer. *Biol. Res.* **51**, 36 (2018).

72. Stellos, K. *et al.* Adenosine-to-inosine RNA editing controls cathepsin S expression in atherosclerosis by enabling HuR-mediated post-transcriptional regulation. *Nat. Med.* **22**, (2016).

73. Liu, X. *et al.* ADAR1 promotes the epithelial-to-mesenchymal transition and stem-like cell phenotype of oral cancer by facilitating oncogenic microRNA maturation. *J. Exp. Clin. Cancer Res.* **38**, 315 (2019).

74. Bahn, J. H. *et al.* Genomic analysis of ADAR1 binding and its involvement in multiple RNA processing pathways. *Nat. Commun.* **6**, 6355 (2015).

75. Pestal, K. *et al.* Isoforms of RNA-Editing Enzyme ADAR1 Independently Control Nucleic Acid Sensor MDA5-Driven Autoimmunity and Multi-organ Development. *Immunity* **43**, 933–944 (2015).

76. Saunders, L. R. *et al.* Characterization of Two Evolutionarily Conserved, Alternatively Spliced Nuclear Phosphoproteins, NFAR-1 and -2, that Function in mRNA Processing and Interact with the Double-stranded RNA-dependent Protein Kinase, PKR. *J. Biol. Chem.* **276**, 32300–32312 (2001).

77. Harashima, A., Guettouche, T. & Barber, G. N. Phosphorylation of the NFAR proteins by the dsRNA-dependent protein kinase PKR constitutes a novel mechanism of translational regulation and cellular defense. *Genes Dev.* **24**, 2640–53 (2010).

78. Strong, J. E. The molecular basis of viral oncolysis: usurpation of the Ras signaling pathway by reovirus. *EMBO J.* **17**, 3351–3362 (1998).

79. Stojdl, D. F. *et al.* Exploiting tumor-specific defects in the interferon pathway with a previously unknown oncolytic virus. *Nat. Med.* **6**, 821–825 (2000).

80. Danziger, O., Shai, B., Sabo, Y., Bacharach, E. & Ehrlich, M. Combined genetic and epigenetic interferences with interferon signaling expose prostate cancer cells to viral infection. *Oncotarget* **7**, 52115–52134 (2016).

81. Ishizuka, J. J. *et al.* Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. *Nature* 1 (2018). doi:10.1038/s41586-018-0768-9

82. Liu, H. *et al.* Tumor-derived IFN triggers chronic pathway agonism and sensitivity to ADAR loss. *Nat. Med.* **25**, 95–102 (2019).

83. Hanzelmann, S., Castelo, R. & Guinney, J. GSVA : gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* **14**, (2013).

84. Porath, H. T., Carmi, S. & Levanon, E. Y. A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat. Commun.* **5**, 4726 (2014).

85. Dong, X. *et al.* CDK13 RNA Over-Editing mediated by ADAR1 associates with poor prognosis of hepatocellular carcinoma patients. *Cell. Physiol. Biochem.* **47**, 2602–2612 (2018).

86. Chen, Y. Bin *et al.* ADAR2 functions as a tumor suppressor via editing IGFBP7 in esophageal squamous cell carcinoma. *Int. J. Oncol.* **50**, 622–630 (2017).

87. Jiang, Q. *et al.* Hyper-Editing of Cell-Cycle Regulatory and Tumor Suppressor RNA Promotes Malignant Progenitor Propagation. *Cancer Cell* **35**, 81-94.e7 (2019).

88. Beghini, A. *et al.* RNA hyperediting and alternative splicing of hematopoietic cell phosphatase (PTPN6) gene in acute myeloid leukemia. *Hum. Mol. Genet.* **9**,

2297–2304 (2000).

89.  Levanon, E. Y. *et al.* Evolutionarily conserved human targets of adenosine to inosine RNA editing. *Nucleic Acids Res.* **33**, 1162–1168 (2005).

90.  Eisenberg, E. *et al.* Identification of RNA editing sites in the SNP database. *Nucleic Acids Res.* **33**, 4612–4617 (2005).

91.  Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, 1–17 (2014).

92.  Rosenblatt, J. & Stein, J. RRHO: Test overlap using the Rank-Rank Hypergeometric test. R package version 1.26.0. (2014). doi:10.18129/B9.bioc.RRHO

93.  Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1–15 (2019).

94.  Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
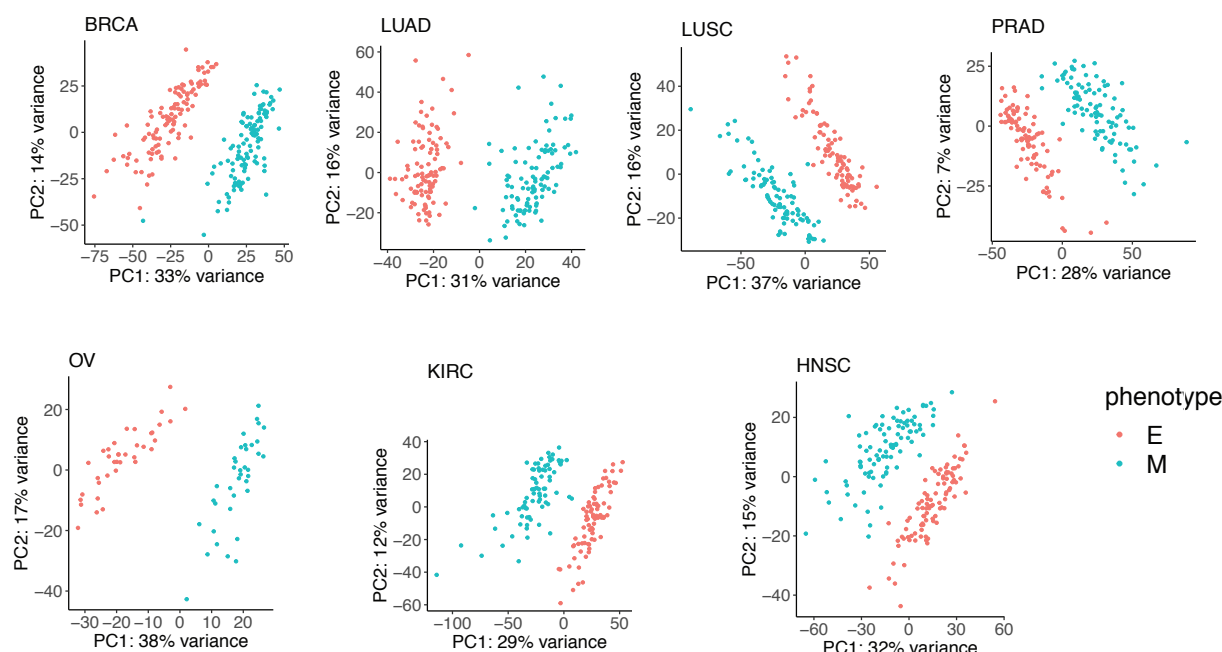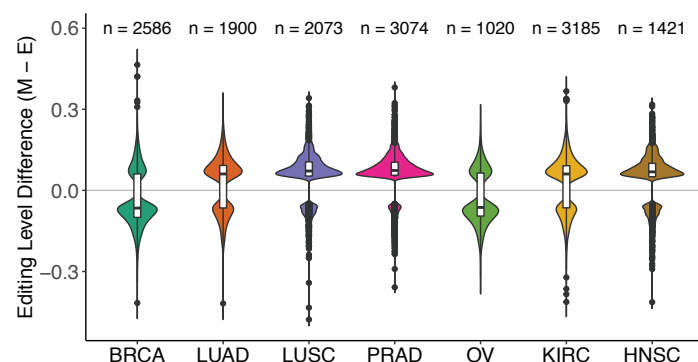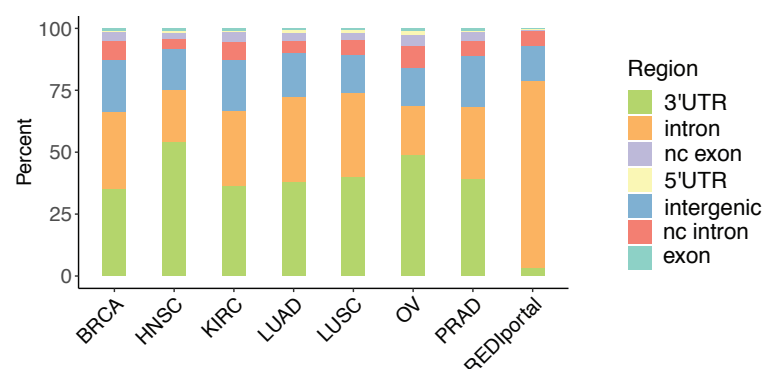
**Fig. 1 Overview of Differential Editing in Cancer EMT.** The following cancer types were studied: breast invasive carcinoma (BRCA), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), prostate adenocarcinoma (PRAD), ovarian serous cystadenocarcinoma (OV), kidney renal clear cell carcinoma (KIRC), head and neck squamous cell carcinoma (HNSC). **A.** First two principal components of differential editing profiles separate tumor samples into epithelial (E) and mesenchymal (M) phenotypes across cancer types. **B.** Distributions of differences in mean editing levels between E and M tumors in each cancer type. The number of differential editing sites is listed on top of each distribution. **C.** Differential editing sites are mostly found in 3' UTR and intronic regions in all cancer types, with higher proportions of 3'UTR sites compared to that of all editing sites from the REDIportal database.
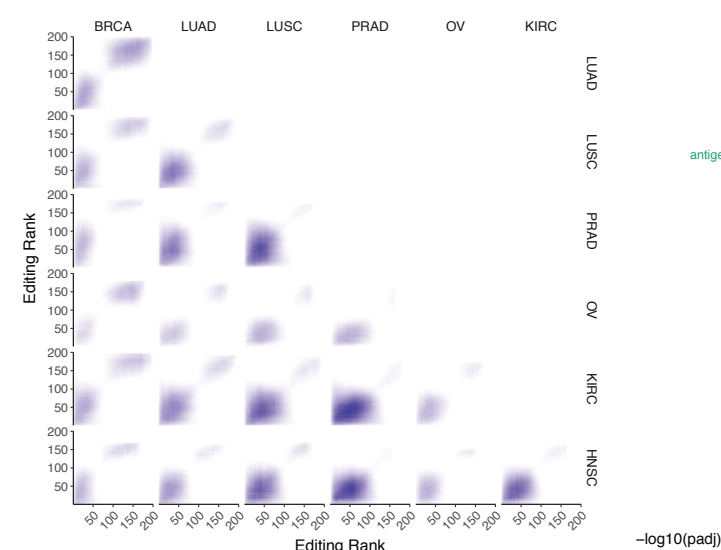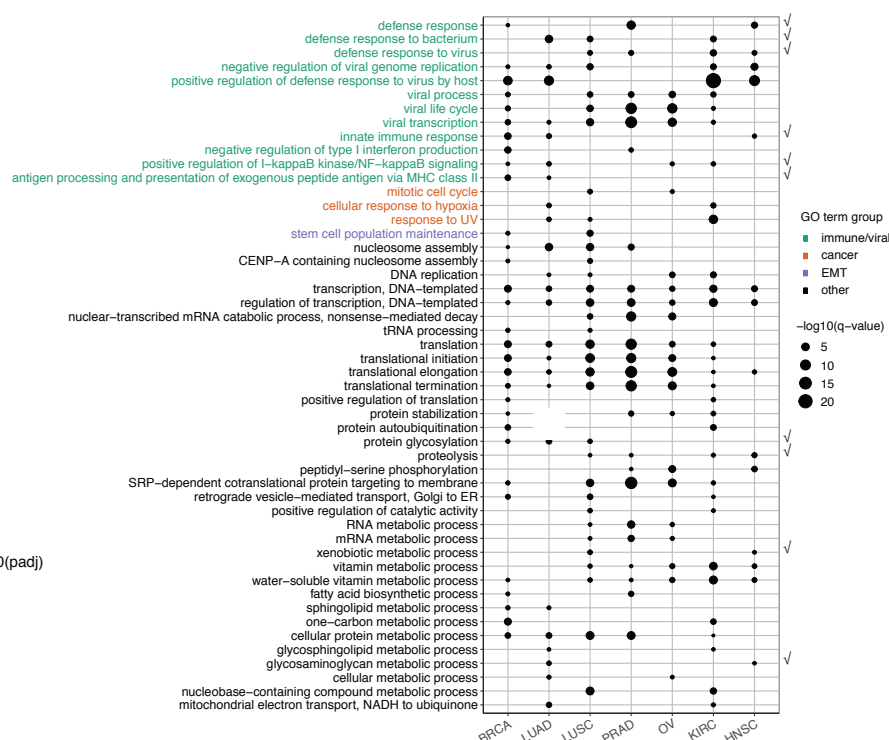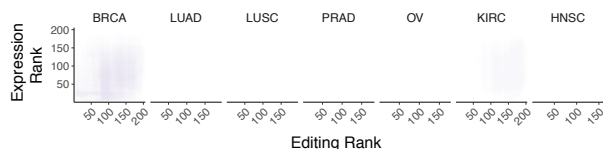
**Fig. 2 Differential Editing Patterns Are Shared Among Cancer Types Yet Distinct from Differential Gene Expression. A.** Rank-rank hypergeometric overlap (RRHO) map of RNA editing across pairs of cancer types. Each heatmap (for two cancer types) represents the matrix of log10-transformed adjusted p-values that indicate the extent of overlap in two gene lists at each possible pair of ranks. For an individual cancer type, genes were ranked by the signed significance of RNA editing differences (M-E). Genes with higher editing in the M phenotype are at lower ranks, while those with higher editing levels in E tumors are at higher ranks. Higher pixel darkness indicates stronger enrichment of overlapping genes within the rank thresholds given by the x and y coordinates. The step size between ranks was 30 genes. **B.** RRHO map of editing and gene expression within each cancer type. Each heatmap contains log10-transformed adjusted p-values of hypergeometric overlap between genes ranked by editing differences (x-axis) and genes ranked by expression differences (y-axis) in a single cancer type. Similar to ranking genes by differential editing, genes were ranked by the signed significance of expression differences, such that genes at lower ranks have higher expression in M tumors, while genes at higher ranks have higher expression in the E phenotype. The step size between ranks was 30 genes. **C.** Significance of enrichment of gene ontology (GO) terms in differentially edited genes of each cancer type represented by point size (log10-transformed adjusted p-value). Terms significantly enriched in at least two cancer types are shown. Check mark on the right indicates terms that were also significantly enriched in differentially expressed genes in at least two cancer types. Text color indicates category of biological relevance.
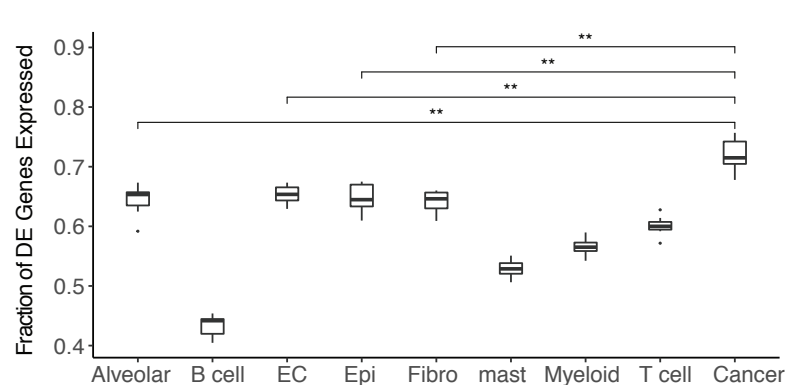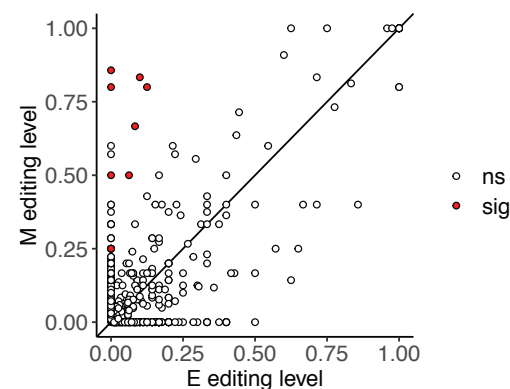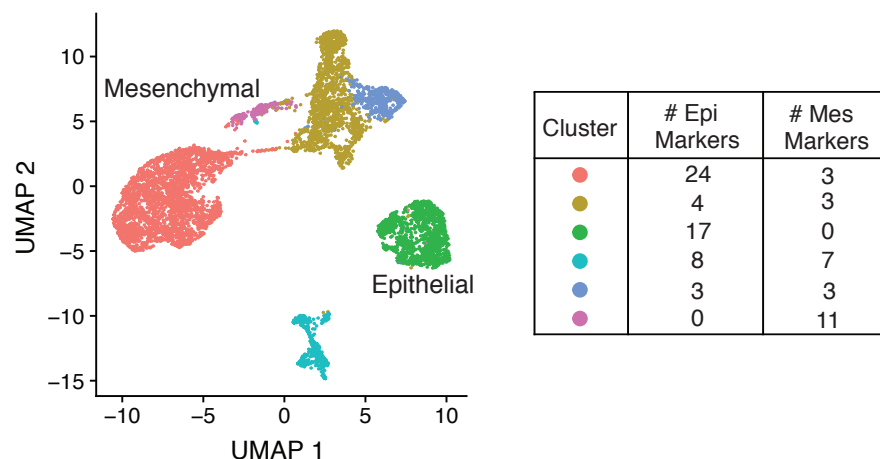
**Fig. 3 Contribution of Cell Types to Differential Editing. A.** Proportions of differentially edited (DE) genes from bulk tumor analysis that were expressed in cell types identified in lung cancer single cell RNA-seq data. Each point represents the proportion of genes from one cancer type. A gene was considered as expressed in a cell type if its expression ≥ 1 RPKM. RPKM values were calculated within each cell type by pooling reads of the same cell type together. Proportions were compared for top cell types by Mann Whitney U test, with significance of p-values shown. **p ≤ 0.01. **B.** UMAP projection of 6526 tumor cells based on expression profiles, colored by cluster assignment (scatterplot, left). By differential expression of epithelial or mesenchymal markers (table, right), Green and Purple clusters were labeled as epithelial and mesenchymal, respectively. **C.** Scatterplot of editing levels of pooled E and M cells, with y = x line. Editing sites exhibiting significant differences between E and M were labeled in red. Differences were considered significant if the difference between editing levels ≥ 0.05 and Fisher's Exact p-value < 0.05.
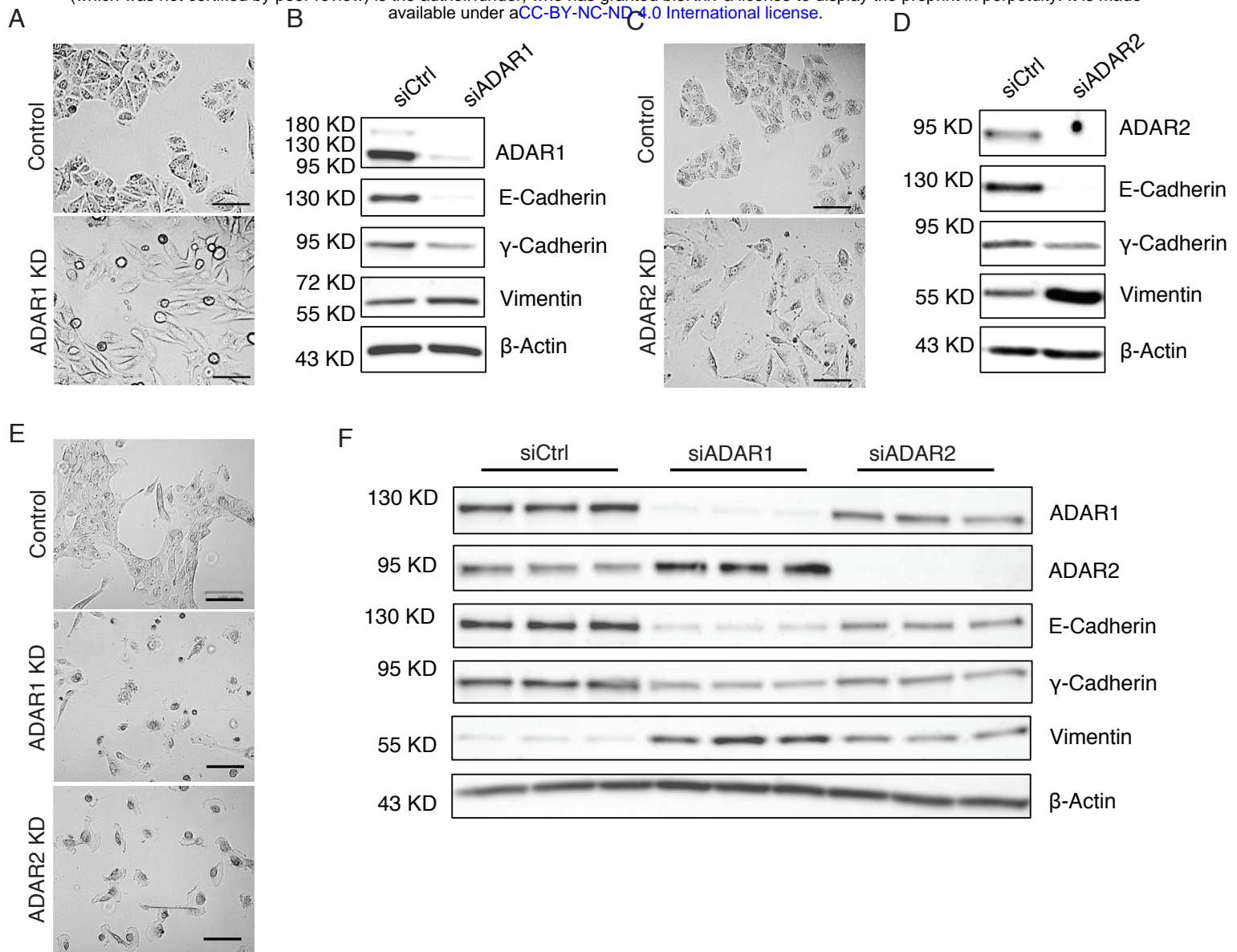
**Fig. 4 ADAR1 or ADAR2 knockdown induced EMT. A.** Images of A549 cells transfected with siRNAs for ADAR1 knockdown (KD) (siADAR1) or control siRNAs (siCtrl). Scale bars, 100um. **B.** Loss of epithelial markers (E-cadherin and γ-Catenin) and induction of mesenchymal marker (Vimentin) in A549 cells upon ADAR1 KD. Cells were treated with 100nM siRNA for 72 hours. **C.** Images of A549 cells transfected with siRNAs for ADAR2 KD (siADAR2) or control siRNAs (siCtrl). Scale bars, 100um. **D.** Loss of epithelial markers (E-cadherin and γ-Catenin) and induction of mesenchymal marker (Vimentin) in A549 cells upon ADAR2 KD. Cells were treated with 11nM siRNA for 72 hours. **E.** Images of MCF10A cells with ADAR1 or ADAR2 KD or control siRNAs. Scale bars, 100um. **F.** Loss of epithelial markers (E-cadherin and γ-Catenin) and induction of mesenchymal markers (Vimentin) in MCF10A cells upon ADAR1 KD or ADAR2 KD. Cells were treated with 11nM siRNA for 72 hours. Three biological replicates were used in each condition.
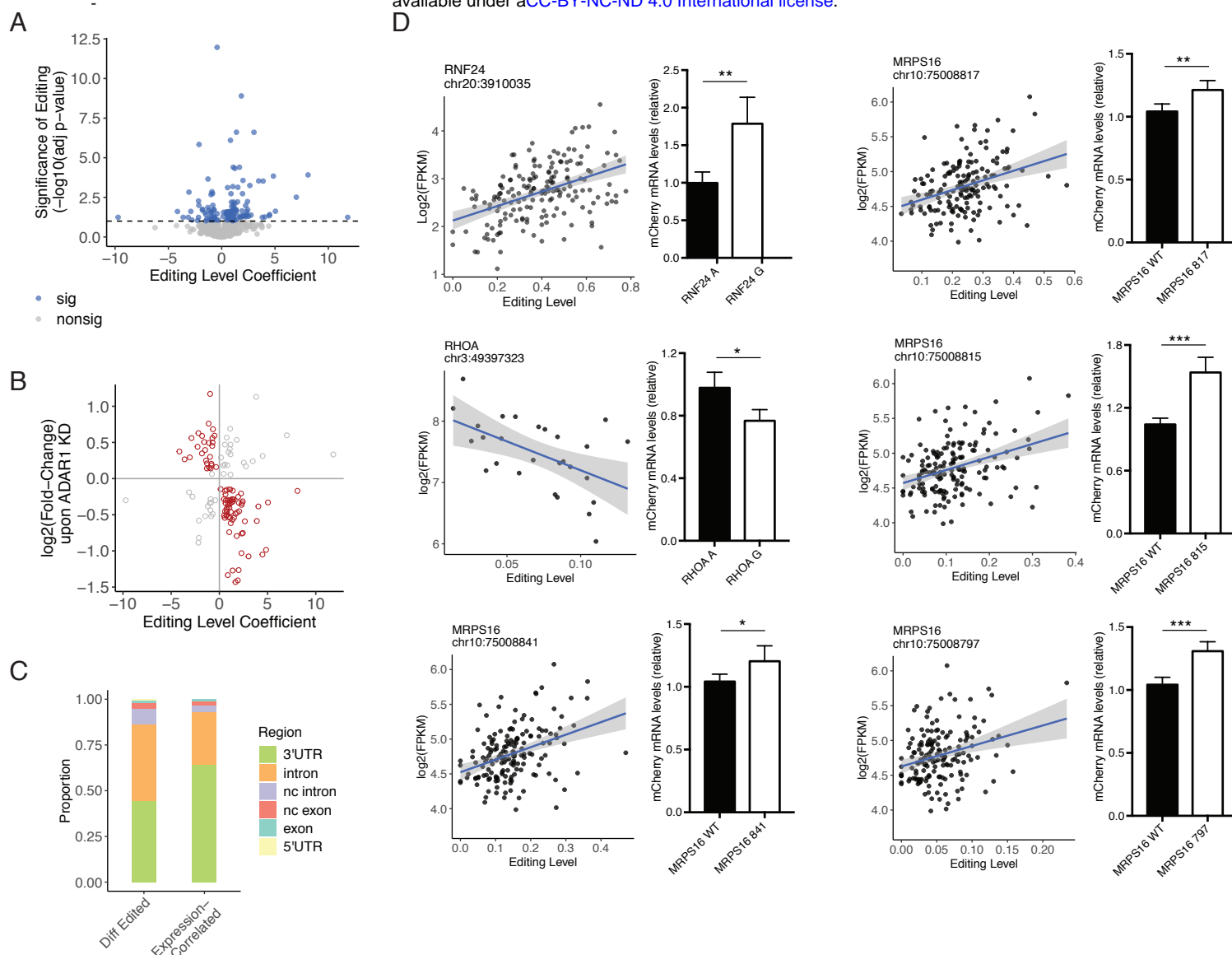
**Fig. 5 Effects of Editing on mRNA Abundance. A.** Scatterplot of coefficient estimate and statistical significance (log10-transformed adjusted p-value) of editing level as a predictor of host mRNA expression in linear regression, accounting for potential confounding variables. For genes with multiple editing sites associated with expression, the most significantly associated site was used. Dashed line indicates significance threshold based on 10% false discovery rate (FDR). **B.** Scatterplot of editing level coefficient estimate from multiple linear regression models used in **A** and log2-transformed fold change of the corresponding gene observed in ADAR1 KD cells. Red points indicate expression changes in the direction consistent with the sign of the editing association, in contrast to the gray points. **C.** Editing sites associated with host expression (Expression-Correlated) are more often found in 3' UTR regions, compared to all differential editing sites (Diff Edited, not including intergenic sites). **D.** Validation of six editing sites affecting host mRNA abundance. For each site, a scatterplot of editing level and log2-transformed mRNA expression in the TCGA data is shown. On the right of each scatterplot is mCherry expression, normalized by eYFP expression, of minigenes with A or G, corresponding to nonedited or edited versions of the sites in the 3' UTR of each gene. Normalized expression values were compared between edited and nonedited versions by two-sided t-test. *p<0.05, **p<0.01, ***p<0.001. Note that RHOA and MRPS16 editing sites were identified as differential sites in the single cell RNA-seq analysis (Fig. 3C).
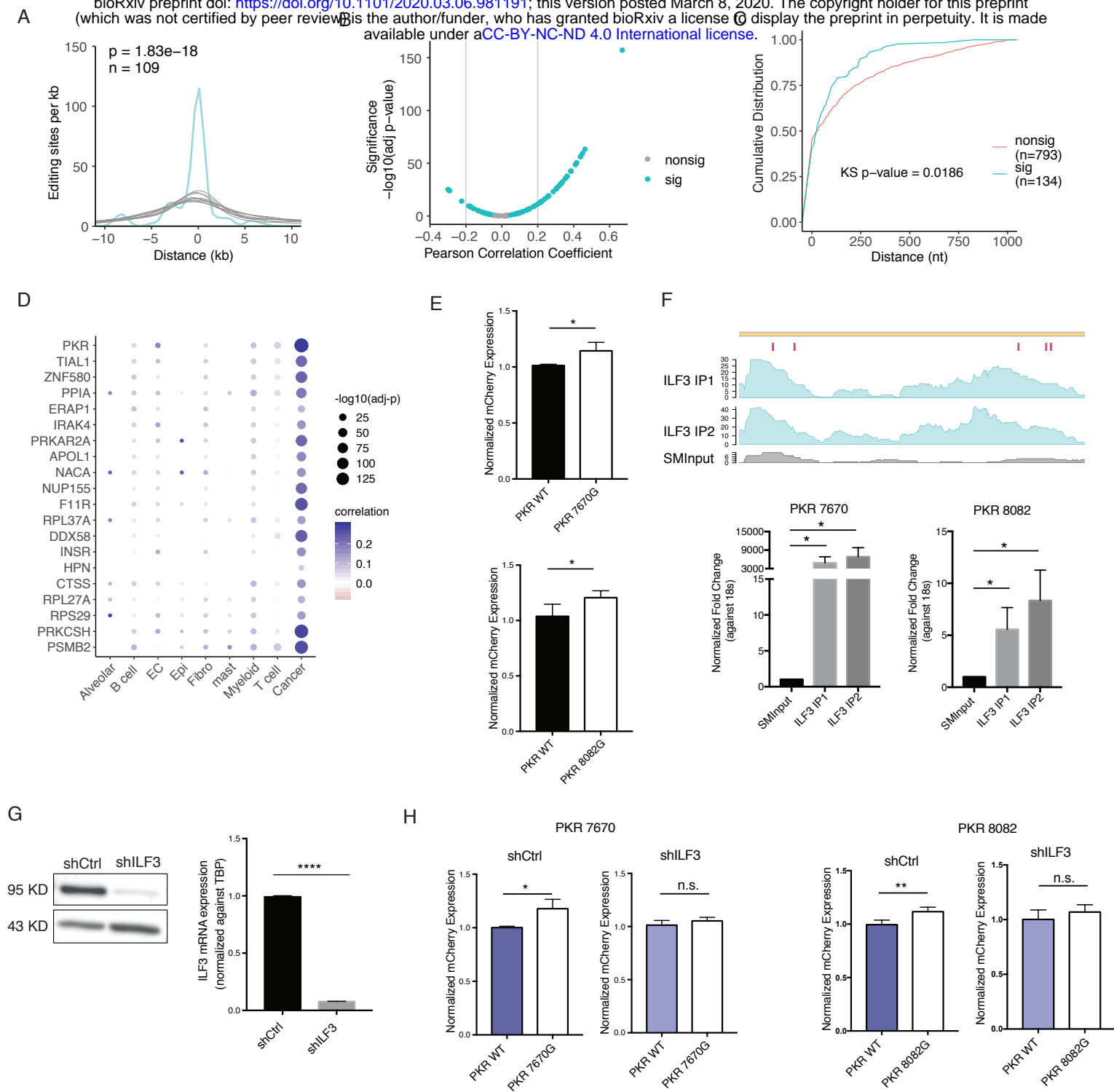
**Fig. 6 ILF3 Regulates mRNA Abundance through RNA Editing. A.** Histogram of distances between differential editing sites in editing-correlated genes and the closest ILF3 eCLIP peaks in A549 cells (turquoise), up to 10 kb. Gray curves represent distances for 10 sets of randomly picked A's in the same genes as differential editing sites. Number of differential editing sites is given by n. P-value was calculated by comparing the area under the curve (AUC) of the distance distribution for differential editing sites to a normal distribution fit to the AUC values of 10,000 sets of random gene-matched A's. **B.** Scatterplot of Pearson correlation coefficient and significance (log10-transformed adjusted p-value) of correlation between ILF3 mRNA expression and mRNA expression of editing-correlated genes. Genes passing 10% FDR are labeled as significant (sig, turquoise), others as nonsig. **C.** Cumulative distributions of distances between ILF3 eCLIP peaks and differential editing sites within editing-expression-associated genes (sig) or differential editing sites in genes without editing-expression associations (nonsig), up to 1 kb. Only genes associated with immune and viral related GO terms were included. P-value calculated by Kolmogorov-Smirnov test. **D.** For each cell type in the lung cancer scRNA-seq dataset, ILF3 mRNA expression was correlated with mRNA expression of editing-expression-correlated genes (identified in the TCGA data) by Pearson correlation. Genes associated with any immune or viral related GO term are shown. The size of each point indicates significance of correlation and color corresponds to values of the correlation coefficient. **E.** Normalized mCherry expression for nonedited or edited versions of sites in the 3' UTR of PKR in A549 cells. Normalized expression values were compared between edited and nonedited versions by two-sided t-test. *p<0.05. **F.** Read coverage of ILF3 eCLIP-seq in A549 cells for two biological replicates (ILF3 IP1 and ILF3 IP2, turquoise) and size-matched input (SMInput, gray). The five validated 3' UTR editing sites affecting PKR mRNA abundance in A549 cells are labeled in magenta (top). Bottom: Validation of PKR eCLIP signal overlapping two editing sites. PKR expression was measured by qRT-PCR in the IP or SMInput samples and normalized against the expression of 18s rRNA. Three technical replicates were performed. P-value calculated by t-test. *p<0.05. **G.** Western Blot confirming shRNA-mediated ILF3 KD in A549 cells (left). ILF3 mRNA levels were quantified in A549 shCtrl and ILF3 KD cells by qRT-PCR (right). ILF3 mRNA expression were normalized against gene TBP mRNA expression. P-value calculated via t-test. ****p<0.0001. **H.** Normalized mCherry expression for nonedited or edited versions of sites in the 3'UTR of PKR in shCtrl or ILF3 KD A549 cells. Normalized expression values were compared between edited and nonedited versions by two-sided t-test. *p<0.05, **p<0.01.