UNIVERSITY OF CALIFORNIA

Los Angeles

Computational Analysis of RNA Editing in Human Cancer

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Bioinformatics

by

Tracey Chan

2021

ABSTRACT OF THE DISSERTATION

Computational Analysis of RNA Editing in Human Cancer

by

Tracey Chan

Doctor of Philosophy in Bioinformatics

University of California, Los Angeles, 2021

Professor Xinshu Xiao, Chair

Discoveries of several oncogenic and tumor-suppressive RNA editing sites have revealed critical roles of editing in cancer and led to the detection of large-scale aberrations in tumors depending on cancer type. Yet, how these abnormal editing events arise, their contributions to tumor development and spread, and the cancer editomes in individual cell types are generally unknown. In this dissertation, we investigated the functional consequences of altered editing in tumors from the level of bulk tissues to single cells.

Expanding upon examples of individual editing sites that promote tumor metastasis, we probed the global editing differences between epithelial and mesenchymal phenotypes in multiple cancer types. Supported by experimental validations, differential editing sites were found to affect mRNA abundance of immune response genes. Furthermore, we identified a novel mechanism of editing-dependent stabilization involving ILF3.

Next, we analyzed RNA editing profiles of single cells and individual cell types in lung cancer. Cancer cells were distinctly hyperedited compared to other cell types in the tumor microenvironment. Gene ontology enrichment analyses further suggested cell-type specificity of differential editing. As we observed that cancer-specific editing correlated with features of immune suppression and overall survival, increased editing levels in cancer cells may support tumor progression by repressing innate immune responses.

Prompted by the apparently diverse contributions of editing to tumor immunity, we sought to identify dsRNA editing candidates that may indicate response to immune checkpoint blockade treatment in melanoma patients. Notably, this analysis revealed the strong candidate dsRNAs that were also correlated with interferon stimulation signatures in lung cancer. Underlying the observed dsRNA associations with treatment response and survival is likely enhanced innate immune signaling activated by unedited dsRNAs.

The dissertation of Tracey Chan is approved.

Thomas G Graeber

Matteo Pellegrini

Sriram Sankararaman

Xinshu Xiao, Committee Chair

University of California, Los Angeles

2021

To the Lord who is my Rock and my Redeemer

and my family

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Professor Xinshu Xiao for her guidance and mentorship. Her passion, visionary thinking, and work ethic have been inspirational. I appreciate her willingness to always make time to discuss projects and any challenges or questions, despite her many responsibilities. I am also grateful for her care and efforts in helping us develop as scientists.

I would like to thank all other members of my committee: Professor Thomas Graeber, Professor Matteo Pellegrini, and Professor Sriram Sankararaman. I appreciate the insightful comments and questions they all provided during our meetings. I would like to thank Professor Graeber and his group also for sharing cancer genomic resources, specifically gene signatures of chronic inflammation.

Next, I would like to thank current and former members of the Xiao lab, as they have made the lab such a pleasant and collaborative work environment. I am thankful for their help, support, and contributions to this work. I have learned a lot from them and have enjoyed getting to know them from eating lunch to exploring LA together. Thanks especially to Ting Fu, with whom I had the privilege of working on multiple projects.

I would also like to thank my family and friends for their generous support throughout the years. I give special thanks to my parents, whose personal sacrifices and encouragement enabled me to pursue a PhD. I also am grateful to my sister and

her husband for always being there for me and giving me advice. Thanks to Justin Leong, Juhyun Kim, Christina Lau, Karen Wong, and Grace Tang for also cheering me on through my PhD journey. I thank the pastors and other members of Immanuel Bible Church in North Hills, California, especially May and Frank Chiu, Mimi Chang, Lulu and Ed Kung, and Joy and Jesse Liang for their ministries and accountability.

This work was partially supported by NIH-NCI T32LM012424 (Biomedical Big Data Training Grant at UCLA, 2016-2018).

Chapter 2 is a version of the published open access article: Chan TW*, Fu T*, Bahn JH, Jun HI, Lee JH, Quinones-Valdez G, Cheng C, Xiao X. RNA editing in cancer impacts mRNA abundance in immune response pathways. Genome Biol 21, 268 (2020). https://doi.org/10.1186/s13059-020-02171-4. (*Contributed equally). T.W.C., T.F., C.C., and X.X. designed the study with inputs from all other authors. T.W.C., J.H.L., and G.Q.V. conducted the bioinformatics analyses. T.F., J.H.B., and H.I.J. conducted the molecular and cellular experiments. All authors contributed to the writing of the paper. X.X. was the principal investigator.

Chapter 3 is currently being prepared as a manuscript for submission: Chan TW, Choudhury M, Xiao X. “The landscape of RNA editing in single cells of lung cancer.” T.W.C. and X.X. designed the study and wrote the paper. T.W.C performed all described analyses except for dsRNA prediction. M.C. predicted dsRNAs from editing-enriched regions. X.X. was the principal investigator.

Chapter 4 is currently being prepared as a manuscript for submission: Chan TW, Choudhury M, Xiao X. “Candidate RNA editing events relevant to immunotherapy response in melanoma.” T.W.C. and X.X. designed the study and wrote the paper. T.W.C performed all described analyses except for MDA5 eCLIP-seq mapping and peak calling. M.C. mapped the MDA5 eCLIP-seq reads and identified eCLIP peaks. X.X. was the principal investigator. I would also like to acknowledge Jae Hoon Bahn, who generated the eCLIP-seq.

VITA

2011-2015	B.S., Bioengineering California Institute of Technology
2015-2021	Graduate Student Researcher, Bioinformatics Interdepartmental Program, University of California, Los Angeles
2020	Teaching Assistant, MCDB 187AL: Research Immersion Lab in Genomic Biology, University of California, Los Angeles

Awards

2016-2018	Big Biomedical Data Training Program Grant, University of California, Los Angeles
2017, 2018	Big Biomedical Data Training Program Challenge Award, University of California, Los Angeles
2017	Poster Award, Institute for Quantitative and Computational Biosciences, University of California, Los Angeles
2014	Robert and Delpha Noland Summer Internship, California Institute of Technology
2012, 2013	Summer Undergraduate Research Fellowship, California Institute of Technology

Publications

Chan TW, Choudhury M, Xiao X. The landscape of RNA editing in single cells of lung cancer. In preparation.

Chan TW, Choudhury M, Xiao X. Candidate RNA editing events relevant to immunotherapy response in melanoma. In preparation.

Fu T, Amoah K, **Chan TW**, Bahn JH, Lee JH, Cheung R, Kosuri S, Xiao X. Massively Parallel Screen for 3'UTR Variants Regulating mRNA Abundance. In preparation.

Quinones-Valdez G, Fu T, **Chan TW**, Xiao X. scAllele: a versatile tool for the detection and analysis of variants in scRNA-seq. In preparation.

Wang H, **Chan TW**, Vashisht AA, Drew BG, Calkin AC, Harris TE, Wohlschlegel JA, Xiao X, Reue K. Lipin 1 modulates mRNA splicing during fasting adaption in liver. To appear in JCI Insight, September 2021.

Chan TW*, Fu T*, Bahn JH, Jun HI, Lee JH, Quinones-Valdez G, Cheng C, Xiao X. RNA editing in cancer impacts mRNA abundance in immune response pathways. Genome Biol 21, 268 (2020). https://doi.org/10.1186/s13059-020-02171-4 (*Contributed equally)

Urtecho G, Insigne KD, Tripp AD, Brinck M, Lubock NB, Kim H, **Chan TW**, Kosuri S. Genome-wide Functional Characterization of Escherichia coli Promoters and Regulatory Elements Responsible for their Function. bioRxiv, 2020. 10.1101/2020.01.04.894907

Brümmer A, Yang Y, **Chan TW**, Xiao X. Structure-mediated modulation of mRNA abundance by A-to-I editing. Nat Commun 8, 1255 (2017). https://doi.org/10.1038/s41467-017-01459-7

Chapter 1: Introduction

Post-transcriptional modifications of individual RNA nucleotides broaden the proteome and regulate gene expression[1,2]. The most abundant form of RNA editing in the human transcriptome is the deamination of adenosine (A) to inosine (I), which is called A-to-I editing. Two members of the adenosine deaminases acting on RNA (ADAR) family of proteins, ADAR1 and ADAR2, catalyze this process[3]. Since ribosomes and other RNA processing machinery read inosines as guanosines (Gs), A-to-I editing events are interpreted as A-to-G mismatches to the genome[2]. In protein-coding regions, edited sites are RNA mutations that can cause changes in amino acid sequence and ultimately protein function. As the majority of editing sites are located in noncoding regions, their functions are still largely unknown. Effects of certain noncoding editing sites include alternative splicing patterns, altered miRNA sequences or target accessibility, and inosine-specific mRNA degradation[2–5].

Besides the functions reported for specific editing sites, modifying endogenous dsRNAs to avoid activating innate immune responses is considered an essential role of RNA editing[2]. Due to the high abundance and similarity of Alu repetitive elements in the human genome, a pair of inverted Alu repeats may reside in the same RNA transcript and form a dsRNA structure similar to the structures of many viral RNAs[2,6]. As a result, cytosolic dsRNA sensors may recognize these Alu-formed dsRNAs similarly as viral dsRNAs and erroneously induce production of interferons (IFNs), suspension of translation, and inhibition of cell growth[7]. Since ADARs target dsRNAs, RNA editing

events impair base pairing and likely remodel these RNA structures, thus preventing recognition by dsRNA sensors and providing a mechanism of tolerance to self RNAs[2,7–10]. This role is thought to explain the embryonic lethality and inflammation of both ADAR1-null mice and ADAR1-editing-deficient mice[7,11,12]. Furthermore, loss of editing may mediate onset of Aicardi-Goutières syndrome (AGS), a genetic disease with clinical presentation similar to infection, caused by ADAR1 loss-of-function mutations[13–15].

Acting as oncogenic drivers or tumor suppressors, RNA editing events in both protein-coding and noncoding regions may play important roles in the pathogenesis of various cancer types and consequently may serve as clinical targets for therapy[16–19]. Demonstrated in multiple cancer types, an AZIN1 recoding site inhibits degradation of oncogenic ODC and cyclin D1, thereby facilitating cell proliferation and metastasis[16,20,21]. Other tumorigenic editing events include miR-200b editing, which prevents degradation of transcription activators ZEB1 and ZEB2 in epithelial-mesenchymal transition (EMT) and enables novel targeting of tumor suppressor LIFR[22]. Inversely, recoding sites in Gabra3 and PODXL were shown to suppress the function of the wildtype genes of enhancing cell invasiveness in breast and gastric cancers, respectively[23,24]. In addition, recoding sites in AZIN1, GRIA2, and COG3 were found to affect cancer cell sensitivity to drugs in cell line experiments, suggesting the therapeutic potential of editing events[25].

On a transcriptome-wide scale, distinct editing patterns were observed in tumors across multiple cancer types[25–27]. For an individual cancer type, tumor editing levels

were globally increased or decreased compared to normal samples. The consequences of these widespread editing changes have not yet been determined. Also unknown is the representation of individual cells and cell types in editing alterations, considering the heterogenous nature of tumors and their microenvironment. So far, studies have reported functions of ADAR1 and editing in the development of several immune cell types[7,8,28]. However, RNA editing in immune cells within tumors has not been characterized in previous studies to our knowledge. Further understanding of the functions and regulatory mechanisms of altered editing in cancer will guide development of novel editing-based tumor diagnostic and treatment measures. Focusing on the relevance of EMT to tumor metastasis, we systematically investigated the potential functions of editing differences between epithelial (E) and mesenchymal (M) phenotypes in Chapter 2. In Chapters 2 and 3, we examined cell type contributions to editing differences between E and M tumors and between tumors and normal samples, using single cell RNA sequencing datasets.

Given that the innate immune system is a first line of defense against tumors, altered RNA editing may help maintain tumor cell proliferation by regulating the innate immune response[8]. On the other hand, translated recoding events may trigger adaptive immune responses specific to these tumor-associated edited peptides[29,30]. We explored the contribution of cancer-specific RNA editing to immunity in non-small cell lung cancer (NSCLC) tumors in Chapter 3.

Despite improved clinical outcomes that have been achieved with immune checkpoint blockade (ICB) in certain cancer types, only a subset of patients successfully respond to treatment[31,32]. Currently, determining optimal predictors of response is a challenge[32,33]. In a mouse model for melanoma, ADAR1 deficiency was shown to promote ICB response through increased innate immune signaling[34]. Supported by other reports of ADAR1 dependency in certain cancer cells[35,36] and the association of editing with patient survival, RNA editing may be indicative of tumor sensitivity to ICB. Chapter 4 reveals dsRNA editing events associated with ICB response and survival in melanoma patients.

Chapter 2: RNA editing in cancer impacts mRNA abundance in immune response pathways

2.1 Abstract

RNA editing generates modifications to the RNA sequences, thereby increasing protein diversity and shaping various layers of gene regulation. Recent studies have revealed global shifts in editing levels across many cancer types, as well as a few specific mechanisms implicating individual sites in tumorigenesis or metastasis. However, most tumor-associated sites, predominantly in noncoding regions, have unknown functional relevance. Here, we carry out integrative analysis of RNA editing profiles between epithelial and mesenchymal tumors, since epithelial-mesenchymal transition is a key paradigm for metastasis. We identify distinct editing patterns between epithelial and mesenchymal tumors in seven cancer types using TCGA data, an observation further supported by single-cell RNA sequencing data and ADAR perturbation experiments in cell culture. Through computational analyses and experimental validations, we show that differential editing sites between epithelial and mesenchymal phenotypes function by regulating mRNA abundance of their respective genes. Our analysis of RNA-binding proteins reveals ILF3 as a potential regulator of this process, supported by experimental validations. Consistent with the known roles*

* The work appearing in this chapter is published: Chan, T.W., Fu, T., Bahn, J.H. et al. RNA editing in cancer impacts mRNA abundance in immune response pathways. Genome Biol 21, 268 (2020). https://doi.org/10.1186/s13059-020-02171-4

of ILF3 in immune response, epithelial-mesenchymal differential editing sites are enriched in genes involved in immune and viral processes. The strongest target of editing dependent ILF3 regulation is the transcript encoding PKR, a crucial player in immune and viral response. Our study reports widespread differences in RNA editing between epithelial and mesenchymal tumors and a novel mechanism of editing-dependent regulation of mRNA abundance. It reveals the broad impact of RNA editing in cancer and its relevance to cancer-related immune pathways.

2.2 **Introduction**

RNA editing, the modification of specific nucleotides in RNA sequences, expands diversity in proteins and gene regulatory mechanisms[37,38]. 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[39]. 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[5,40].

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[16,17,20,41]. 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[29,30]. Furthermore, across various cancer types, genome-wide aberrations in RNA editing were observed and associated with clinical features[25–27]. 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[25]. 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[17].

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[42–44]. 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[22,23,30,45–47]. 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[22]. The SLC22A3 recoding event also promoted EMT, causing expression changes in EMT marker genes[45]. In contrast, a recoding event in GABRA3 inhibited metastasis and was present only in non-invasive cell lines and non-metastatic tumors[23]. 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 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.

2.3 Results

2.3.1 Altered RNA editing profiles between epithelial and mesenchymal tumors

EMT is known to be accompanied by substantial transcriptome remodeling[43,74–78]. Given the previously reported functional relevance of RNA editing in EMT[22,45,79], 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 into 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[51]. 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 our previously published methods [37,53,80], we quantified editing levels at over 4 million editing sites recorded in the REDIportal database[54]. 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.

2.3.2 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[65]. We applied RRHO to RNA editing here by ranking genes according to the significance of tested editing differences between E and M and the direction of editing differences (Methods). In addition to shared directionality of global editing trends, we found significant overlap in genes with editing changes among multiple cancer types (Fig. 2A). Within pairs of cancer types, most significant overlaps were enriched at the bottom left or top right corners, where genes were hyperedited or hypoedited in both

cancer types, respectively. These significant overlaps in genes based on differential editing suggest 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 represents a distinct layer of molecular changes in EMT.

2.3.3 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, Supplementary Fig. 2).

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[9,10,12,81,82]. 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.

2.3.4 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[83]. Following quality control measures, we clustered the cells in two rounds and labeled cell types based on marker genes to obtain T cells, B cells, myeloid cells, endothelial cells (EC), fibroblasts (Fibro), epithelial cells (Epi), mast cells, alveolar cells, and cancer cells (Supplementary Fig. 3A-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[84] (Supplementary Fig. 3D).

To gauge the contribution of individual cell types to bulk tumor differential editing, we examined gene expression and editing profiles of each cell type. Specifically, we pooled cells of each type and calculated the percent of differentially edited genes from the bulk tumor analysis that were expressed in each cell type. Cancer cells expressed the highest proportion of genes that were differentially edited (Fig. 3A). We then

measured the extent of editing in each cell type by calculating the percent of bulk tumor differential editing sites that were edited. Consistent with the expression analysis, the highest proportion of differential sites were edited in cancer cells (Fig. 3B). Therefore, the editing differences observed among bulk tumors may be mainly attributable to the cancer cells.

We next separated cancer cells to epithelial and mesenchymal cell clusters (Fig. 3C, 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. 4). 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. 3D). 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). Two sites overlapped with differentially edited sites in LUAD or LUSC and both had hyperediting in M cells, consistent with the direction in bulk tumors (Supplementary Fig. 5). This small overlap likely reflects the low coverage on editing sites in the single cell data, and/or the possibility that more differential editing sites, which were not identified in our study due to limits in power, exist in the bulk tumors.

Notable differentially edited genes include RHOA, which is active in cell migration and is associated with metastasis in multiple cancer types[85–87], and ARL16, a reported negative regulator of RIG-I activity[88], 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.

2.3.5 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 γ-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[79]. Together, these results indicate that knockdown of ADARs promotes EMT.

As expected, ADAR KD induced significant editing changes measured by RNA-seq in A549 cells (Supplementary Fig. 6A-B), with ADAR1 KD affecting a large number of editing sites but ADAR2 having fewer targets. A minority of ADAR2-responding sites

had increased editing upon ADAR2 KD, reflecting the likely compensation by ADAR1. The reverse, compensation of ADAR1 loss by ADAR2, was not observed. Among the lung cancer E-M differential editing sites that were testable in the above A549 RNA-seq data, the vast majority responded to KD of either ADAR or double KD (Supplementary Fig. 6C). These results confirm the impairment of RNA editing at genome scale upon the loss of ADARs.

We next examined mRNA expression of ADARs in the bulk E and M tumors across cancer types. In several cancer types with a hyperediting trend in M, higher mRNA expression of ADAR1 or ADAR2 likely contributed to increased editing levels in M tumors (Supplementary Fig. 7). However, ADAR expression was not consistent with RNA editing differences for some cancer types. Thus, although ADAR KD caused EMT in cell culture models, ADAR expression alone may not sufficiently explain the global editing trends observed in bulk tumors.

2.3.6 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[37,72,89], 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).

2.3.7 ILF3 as an editing-dependent regulator of mRNA abundance

Since mRNA stability is closely regulated by RNA-binding proteins (RBPs)[90–93], 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[71,72]. 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. 8A). 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[94–96].

2.3.8 Impact of ILF3 on immune-relevant genes

ILF3 promotes an antiviral response through its binding to RNAs[97–99]. 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-B), 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.

2.3.9 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[100,101]. PKR also regulates various signaling pathways, such as NF-κB and p38 MAPK, in response to cellular stress[100]. Using the editing minigene reporter, we examined the individual effects of seven 3' UTR editing sites on PKR 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. 8B). To assess the collective impact of multiple RNA editing sites on PKR mRNA abundance, we measured endogenous PKR expression in A549

cells upon ADAR1 or ADAR2 KD. We first confirmed that the 3' UTR editing sites in PKR were edited endogenously in A549 cells. Importantly, these editing sites are mainly regulated by ADAR1 instead of ADAR2 (Supplementary Fig. 8C). Upon ADAR1 KD, PKR expression level was significantly reduced by about 40% (Fig. 6F). In contrast, PKR expression did not change upon ADAR2 KD, as expected. These results suggest that the editing sites enhanced PKR mRNA abundance, consistent with the positive editing-expression correlation in primary tumors.

Based on the eCLIP data, the five editing sites that individually promoted PKR mRNA abundance are located within ILF3 binding sites (Fig. 6G, Supplementary Fig. 8D-E). To test the hypothesis that ILF3 regulates PKR mRNA abundance in an editing-dependent manner, we generated ILF3 KD A549 cells (Fig. 7A). The edited and unedited reporters, demonstrating differential expression in control cells, no longer produced different expression levels upon ILF3 KD (Fig. 7B). Together, our data suggest that ILF3 promotes PKR mRNA expression in an editing-dependent manner by binding to the PKR mRNA.

2.3.10 ILF3 knockdown induced EMT in A549 cells

Since ILF3 was found to stabilize transcripts that were differentially edited between E and M tumors, we next asked if ILF3 regulates the EMT process. We carried out ILF3 KD experiments via two different siRNAs in A549 cells. Upon ILF3 KD, cell morphology changed from tightly connected, round cells towards more dispersed,

spindle-shaped cells (Fig. 7C), consistent with expected EMT phenotypes. Additionally, we observed reduced expression of the epithelial marker E-cadherin along with increased expression of the mesenchymal marker N-cadherin in the ILF3 KD cells (Fig. 7D, E for protein and RNA levels, respectively). Thus, these data show that ILF3 deficiency induces EMT in A549 cells, supporting a significant role of ILF3 in regulating EMT.

2.4 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[102]. EMT plasticity is thought to underlie cell dissemination and metastatic formation in many cancer types[44]. Supported by studies on primary tumors and various model systems, features of EMT have been associated with metastasis[42,44,103,104]. 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[104]. Furthermore, this partial EMT program was correlated with multiple metastatic characteristics, including abundance of lymph node metastases, lymphovascular invasion, and tumor grade[104]. 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[44]. 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[105,106]. 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[5,46,107–112]. 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.

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.

Our cell line experiments showed EMT induction upon KD of either ADAR1 or ADAR2 in lung and breast cell lines. In contrast, we observed hyperediting in M tumors of most cancer types. The seemingly opposite trends may reflect the complexity of tumor biology that is not effectively recapitulated by cell culture models. Although the cell culture models can support the likely importance of RNA editing in EMT, the exact mechanisms and related regulation can only be investigated using *in vivo* models in the future. In addition, we did not observe large differences in ADAR expression levels that are consistent with observed editing differences between E and M tumors for all cancer

types. Other proteins that directly or indirectly affect ADAR function likely contribute to the regulation of E-M RNA editing differences, which remains to be investigated.

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[9,12,113]. 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[97,114]. Upon viral infection and sensing of viral dsRNA, PKR activates, suppresses translation, and promotes apoptosis of affected cells[101]. Importantly, this mechanism has been targeted in oncolytic virotherapy for cancer. Cancer cells that have low PKR expression are sensitive to oncolytic viruses[115–117]. Our study showed that ILF3 mediates the RNA editing-dependent regulation of PKR expression. We also observed that ILF3 KD induced EMT in A549 cells. These data reveal novel insights into the reciprocal regulation between PKR and ILF3 and their potential contributions to EMT. Additional studies on their interaction during viral infection or cancer treatment will also be informative for therapeutic development. Previously, ADAR1 loss has been shown to render tumor cells sensitive to immunotherapy through enhanced inflammatory response[34,36]. 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.

2.5 Methods

2.5.1 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.

2.5.2 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 1×10^5 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 $0.15x10^5$ 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.

2.5.3 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).

2.5.4 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). Primers used for qPCR are listed in Supplementary Table 3. 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.

2.5.5 Quantification of RNA editing levels by Sanger sequencing

Regions of interest were amplified from cDNA using Thermo Scientific™ DreamTaq™ Green PCR Master Mix (2X). Primers used for PCR are listed in Supplementary Table 3. The amplicons were gel extracted and premixed with the reverse primer for Sanger sequencing. The peak signals of A and G nucleotides were measured by 4Peaks for editing level calculation (G/(A+G)).

2.5.6 Categorization of tumors as epithelial and mesenchymal

We downloaded fragments per kilobase million (FPKM) data of primary tumors from patients across seven cancer types in TCGA: BRCA, LUAD, LUSC, PRAD, OV, KIRC, and HNSC, from the Genomic Data Commons (GDC) Data Portal[49]. 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)[50]. We obtained pan-

cancer E and M gene sets from a 2014 publication by Tan and colleagues (Table S1A from their publication)[51]. 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.

2.5.7 Quantification and comparison of RNA editing levels in TCGA tumors

We downloaded RNA-seq fastq files of categorized tumors from the GDC 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[52,53]. 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 [54,55]. 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[25,41,45,56–59], since editing sites have been shown to be mistakenly recorded as SNPs[60,61]. 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[62], 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[53]. 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.

2.5.8 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[63]. 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.

2.5.9 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 in R to test for significance of overlap between ranked gene lists, with a step size of 30 genes between each rank[64].

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.

To make RRHO maps comparable across cancer types and across overlaps based on differential editing and differential expression, 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[65].

2.5.10 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 for each query gene was randomly selected 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. We repeated this assessment of GO term enrichment separately for lists of differential hyperedited and hypoedited genes in each cancer type.

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.

2.5.11 scRNA-seq dataset analysis

We downloaded fastq files from 15 tumor samples of five NSCLC patients[66] 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[67] (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[68] (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[69] 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 RPKM. These RPKM 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 cancer cells. To this end, we first ran sctransform and PCA on only the cancer 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.

2.5.12 RNA-seq generation for ADAR KD A549 cells

A549 cells were seeded at 1×10^5 cells per well in 6-well plates 24 hours before siRNA transfection. siRNAs (Supplementary Table 3) were introduced at the final concentration of 22nM using lipofectamine RNAiMAX (Invitrogen), according to the manufacturer's protocol. For individual KD of ADAR1 or ADAR2, 11nM siRNA of ADAR1 or ADAR2 were mixed with 11nM control siRNAs. For double KD of ADAR1 and ADAR2, 11nM siRNA of ADAR1 and 11nM siRNA of ADAR2 were mixed. Media were changed 24 hours post-transfection. The transfected cells were harvested 48 hours post-transfection. Total RNA was extracted for RNA-seq library generation for three biological replicates of each condition. RNA sequencing libraries were generated using NEBNext Ultra II Directional RNA library Prep kit and NEBNext multiplex oligos for Illumina according to the manufacturer's instructions (New England Biolabs, E7760S). Library concentrations were measured by Qubit fluorometric assay (Life Technologies), and libraries were sequenced on an Illumina HiSeq-4000 with 150-bp paired-end reads.

2.5.13 A549 ADAR KD RNA-seq analysis

Following mapping of RNA-seq reads with HISAT2 and a hyperediting pipeline[53], we detected editing events at REDIportal sites as we did for the TCGA tumor samples. We then removed dbSNP variants while retaining previously reported cancer editing sites. To identify differential editing sites between each ADAR KD condition and control or between each individual ADAR KD and double KD, we used REDIT-LLR on sites that were edited in the control condition (editing level ≥ 0.05)[70]. A site was considered

differentially edited if the difference in mean editing levels between conditions was at least 0.05 and REDIT-LLR p-value < 0.05.

2.5.14 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.

2.5.15 eCLIP-seq generation

Following a published protocol[71], we performed an eCLIP experiment comprising three libraries from two ILF3-immunoprecipitated biological replicates and one control. The antibody used for this experiment is: ILF3/NF90 antibody (Bethyl Laboratories, A303-651A). For each sample, 10M A549 cells were ultraviolet (UV) crosslinked at 254 nm (800 mJ cm^{-2}). We then performed cell lysis, RNA fragmentation, immunoprecipitation, adapter ligation, and other library preparation steps on UV crosslinked samples, as described[71]. For the size-matched input control (SMInput), we prepared a library from sampling 2% of one pre-immunoprecipitation

UV crosslinked sample. This control is used to normalize binding signal, given biases that may be introduced through various experimental steps.

2.5.16 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[72]. 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[72,73].

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[53]. 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

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[53].

2.6 Acknowledgements

We thank members of the Xiao and Cheng laboratories for helpful discussions and comments on this work. The results published here are in part based upon data generated by The Cancer Genome Atlas managed by the NCI and NHGRI. We thank the GTEx consortium for generating the RNA-seq data. We thank the ENCODE Project Consortium (specifically the groups of Dr. Gene Yeo and Dr. Brenton Graveley) for generating the eCLIP-seq and RNA-seq data sets used in this study. We appreciate the helpful discussions with Dr. Eric Van Nostrand and Dr. Gene Yeo on the ILF3 eCLIP-seq experiments.

This work was supported in part by grants from the National Institutes of Health (U01HG009417, R01AG056476 to X.X. and R35GM131876 to C.C.) and the Jonsson Comprehensive Cancer Center at UCLA. C.C. is a CPRIT Scholar in Cancer Research (RR160009). T.W.C. was supported by the NIH-NCI National Cancer Institute T32LM012424.

2.7 Figures

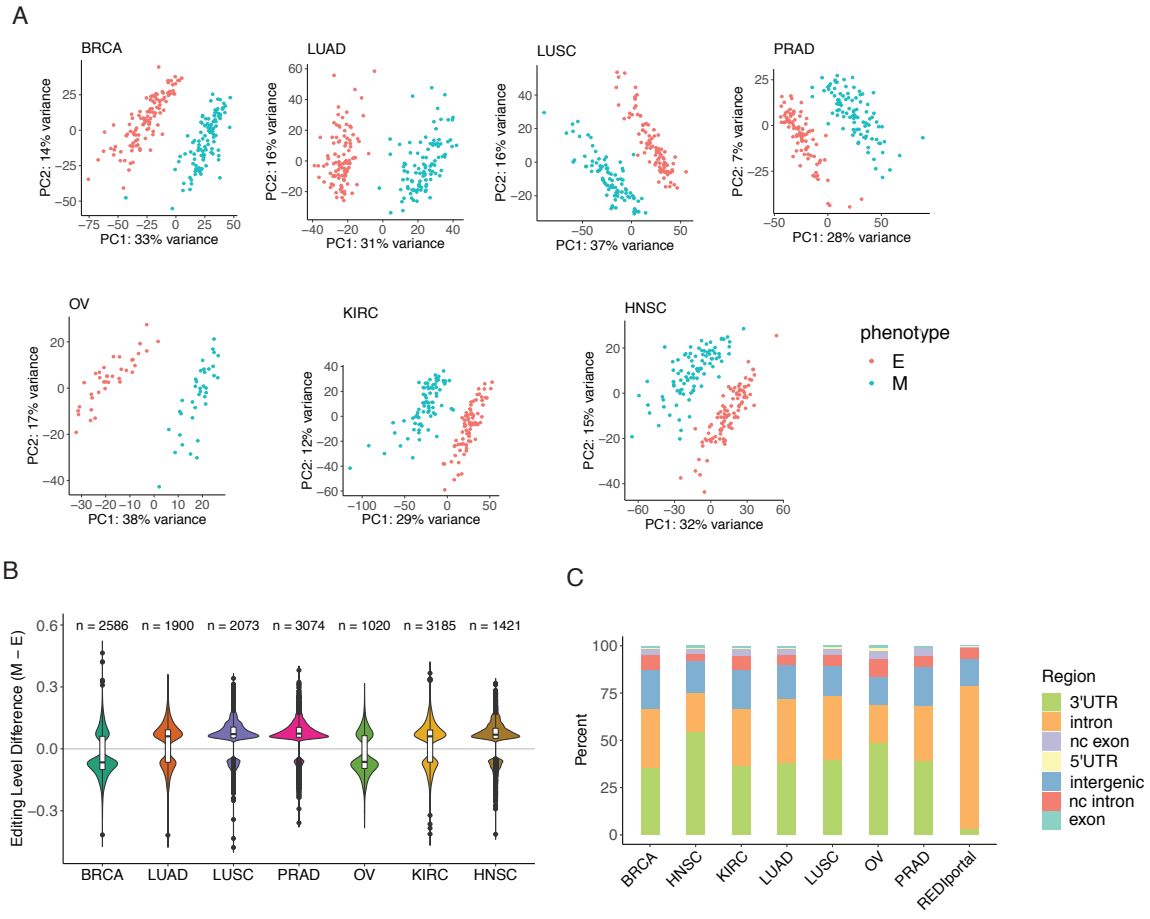


Figure 2.1 Overview of differential editing in cancer EMT

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

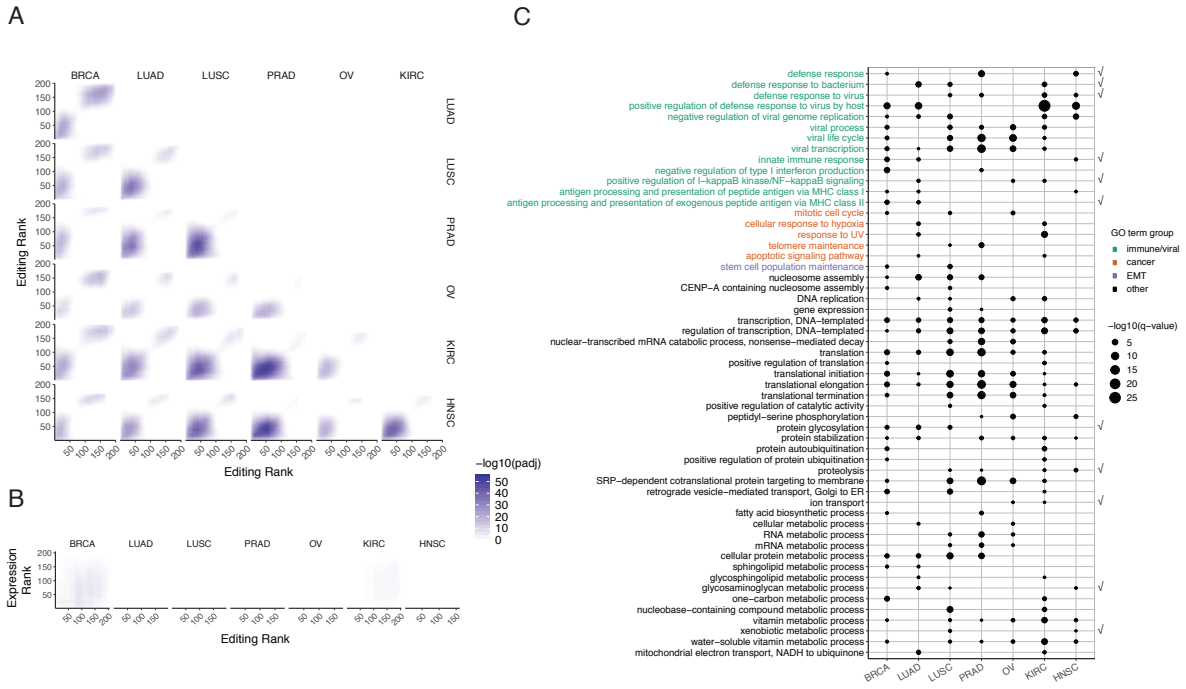


Figure 2.2 Differential editing patterns are shared among cancer types yet distinct from differential gene expression.

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.

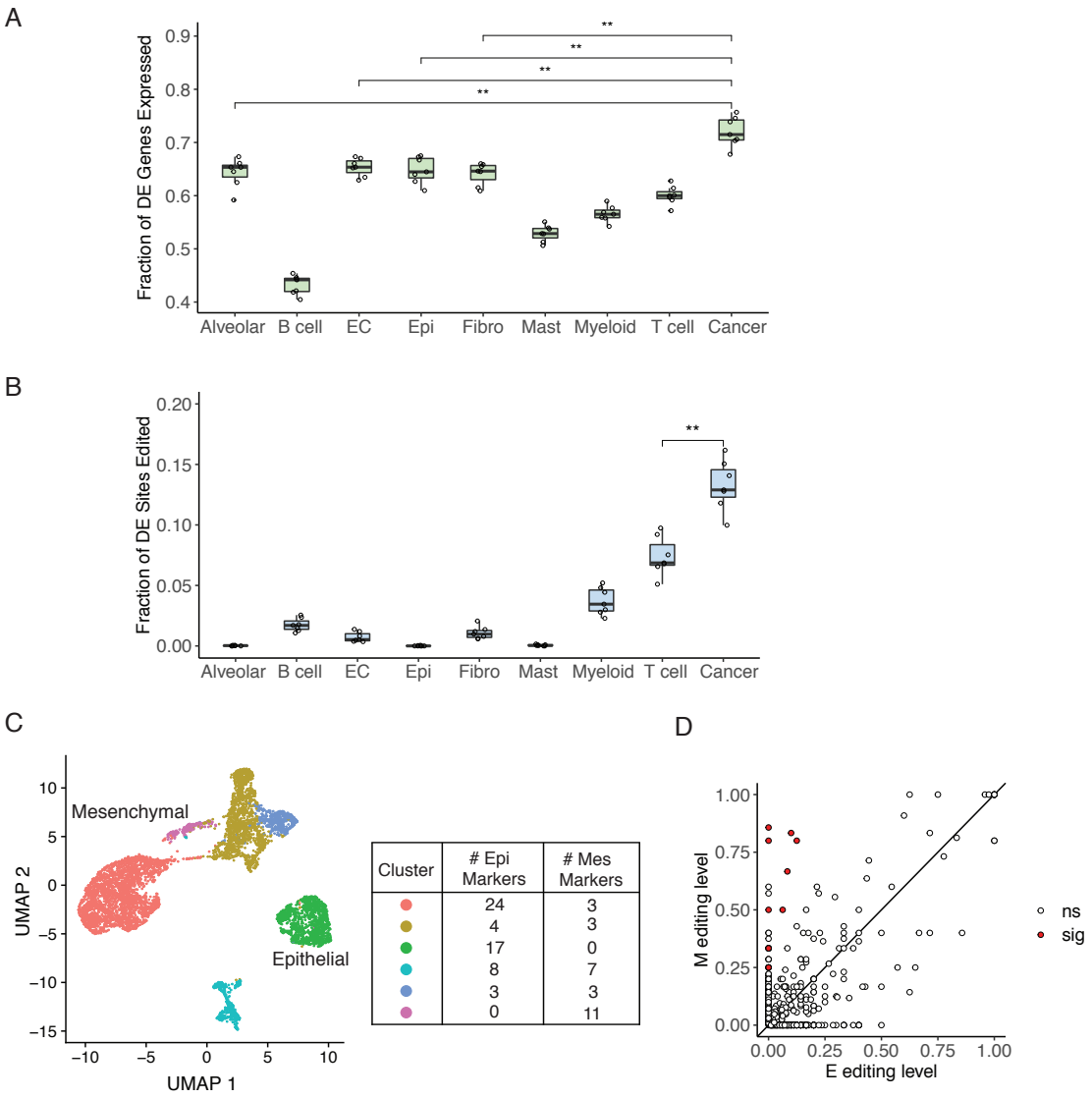


Figure 2.3 Contribution of cell types to differential editing.

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. EC stands for endothelial cells. **B** Proportion of differential editing sites from bulk tumor analysis that were edited in individual cell types. A site was considered as edited in a cell type if the site was covered by at least 5 reads and editing was supported by at least 2 reads. Each point represents the proportion of sites from one cancer type. Proportions for top cell types were compared by Mann Whitney U test, with p value significance shown. **p ≤ 0.01. **C** 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. **D** 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.

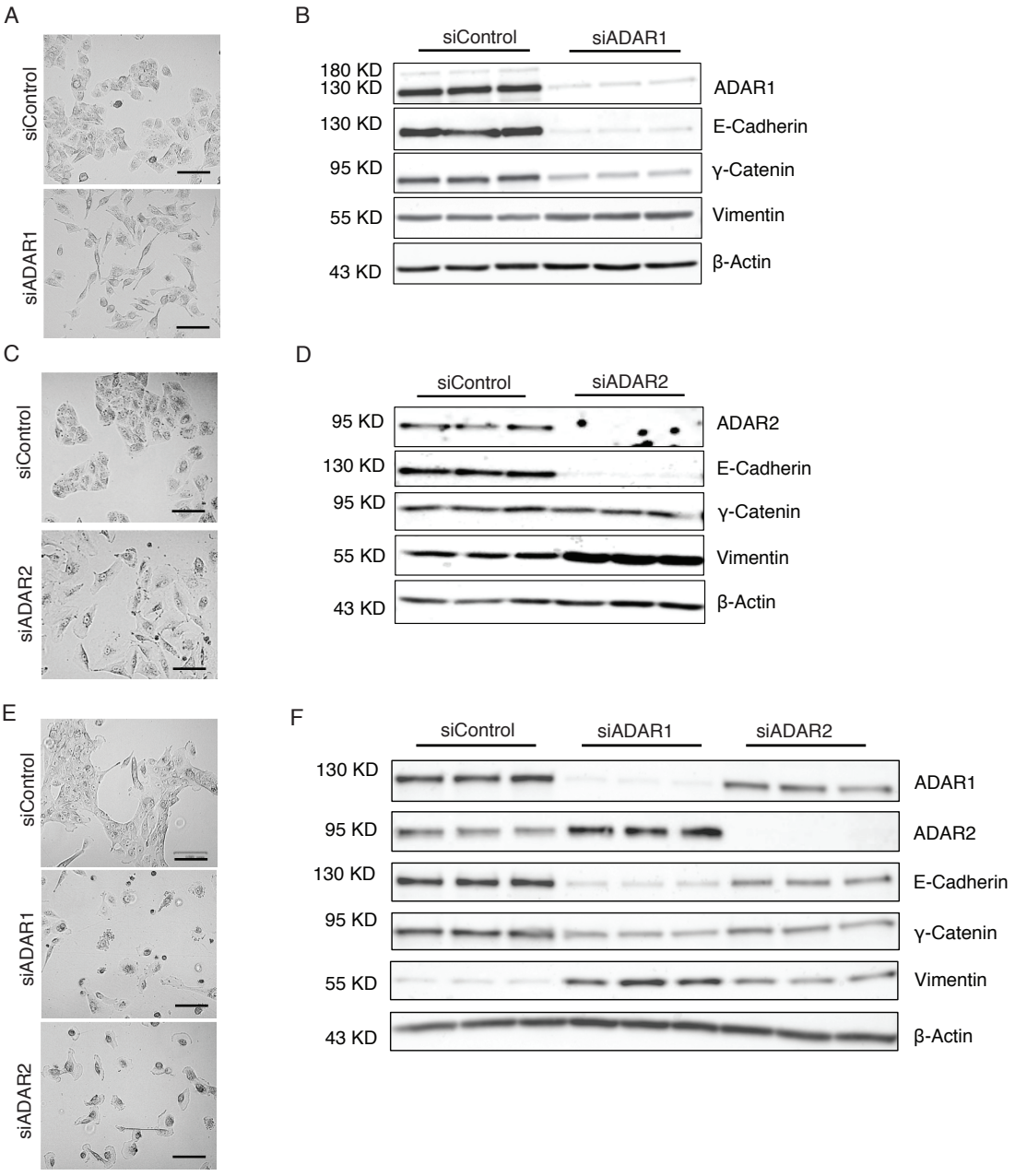


Figure 2.4 ADAR1 or ADAR2 knockdown induced EMT.

A Images of A549 cells transfected with siRNAs for ADAR1 knockdown (KD) (siADAR1) or control siRNAs (siControl). Scale bars, 100 μm. **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 100 nM siRNA for 72 h. Three biological

replicates were used in each condition. **C** Images of A549 cells transfected with siRNAs for ADAR2 KD (siADAR2) or control siRNAs (siControl). Scale bars, 100 μm. **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 11 nM siRNA for 72 h. Three biological replicates were used in each condition. **E** Images of MCF10A cells with ADAR1 or ADAR2 KD or control siRNAs. Scale bars, 100 μm. **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 11 nM siRNA for 72 h. Three biological replicates were used in each condition.

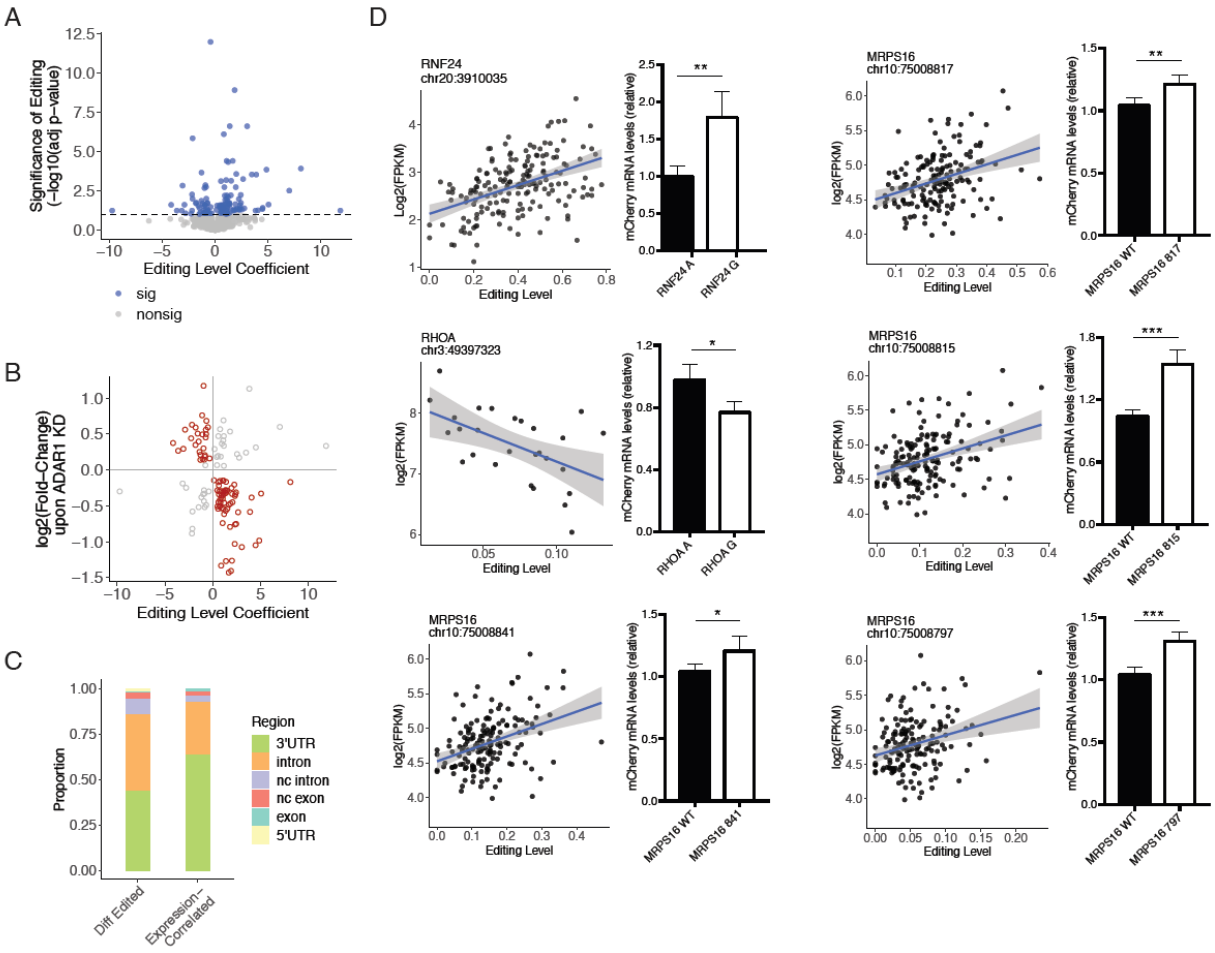


Figure 2.5 Effects of editing on mRNA abundance.

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. All minigenes were tested in Hela cells with five biological replicates. Normalized expression values (mean ± SD) 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).

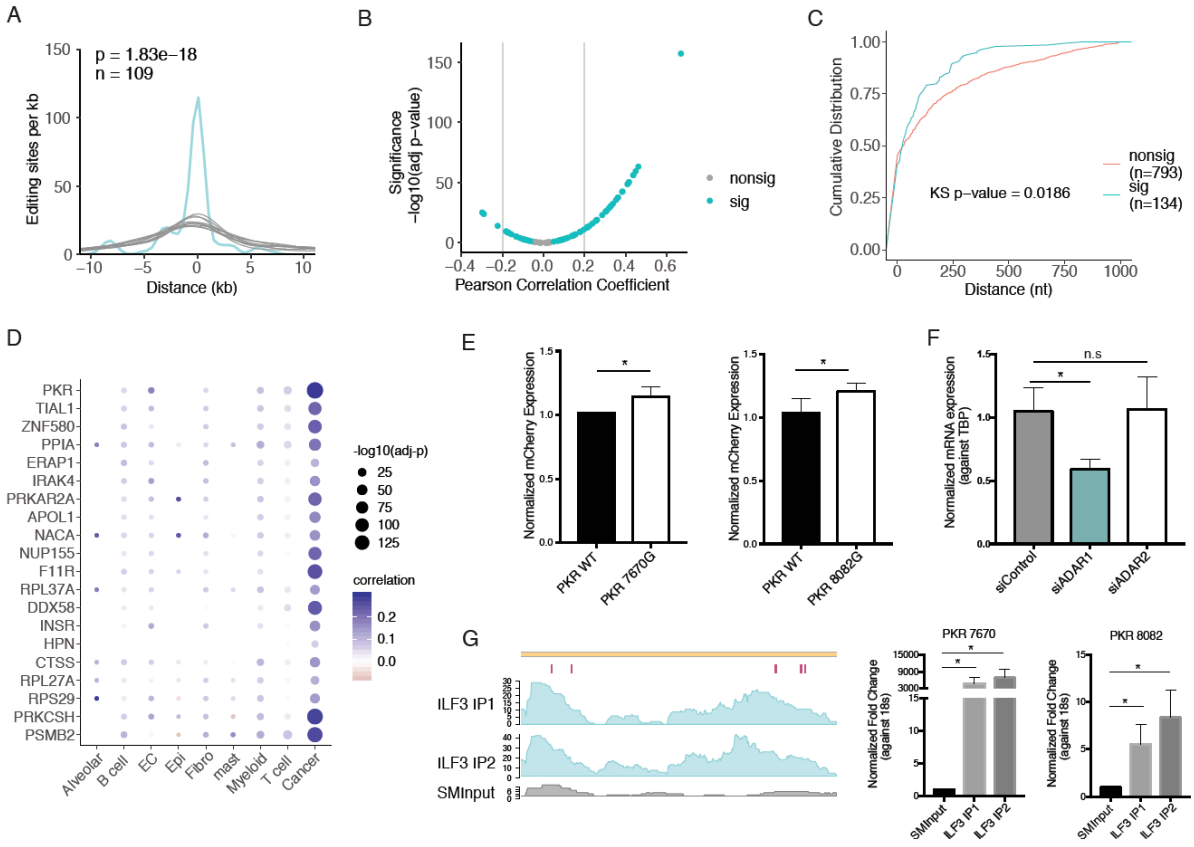


Figure 2.6 ILF3 binds closely to the differential editing sites in editing-expression-correlated genes.

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 the 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 (mean ± SD) for nonedited or edited versions of sites in the 3′UTR of PKR in A549 cells. Five biological replicates were performed. p value calculated by two-sided t-test (same below), *p < 0.05. **F** Normalized mRNA expression (mean ± SD) of endogenous PKR in siControl, siADAR1, and siADAR2 A549 cells. Three biological replicates were performed. *p < 0.05. n.s., not significant. **G** 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 (left). Right: Validation of PKR eCLIP signal overlapping two editing sites. PKR expression (mean ± SD) was measured by qRT-PCR in the IP or SMInput samples and normalized against the expression of 18s rRNA, *p < 0.05. (n = 3)

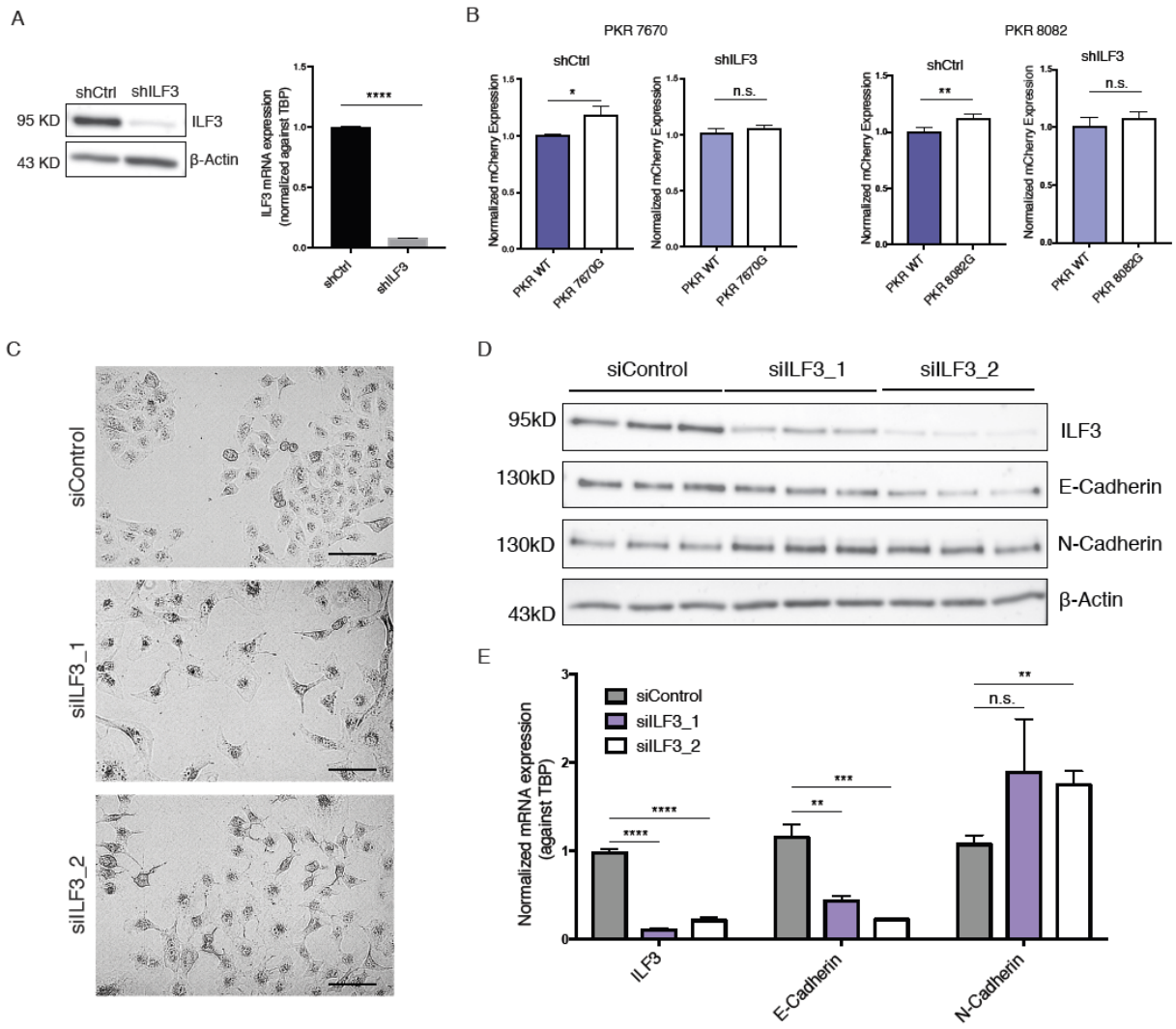


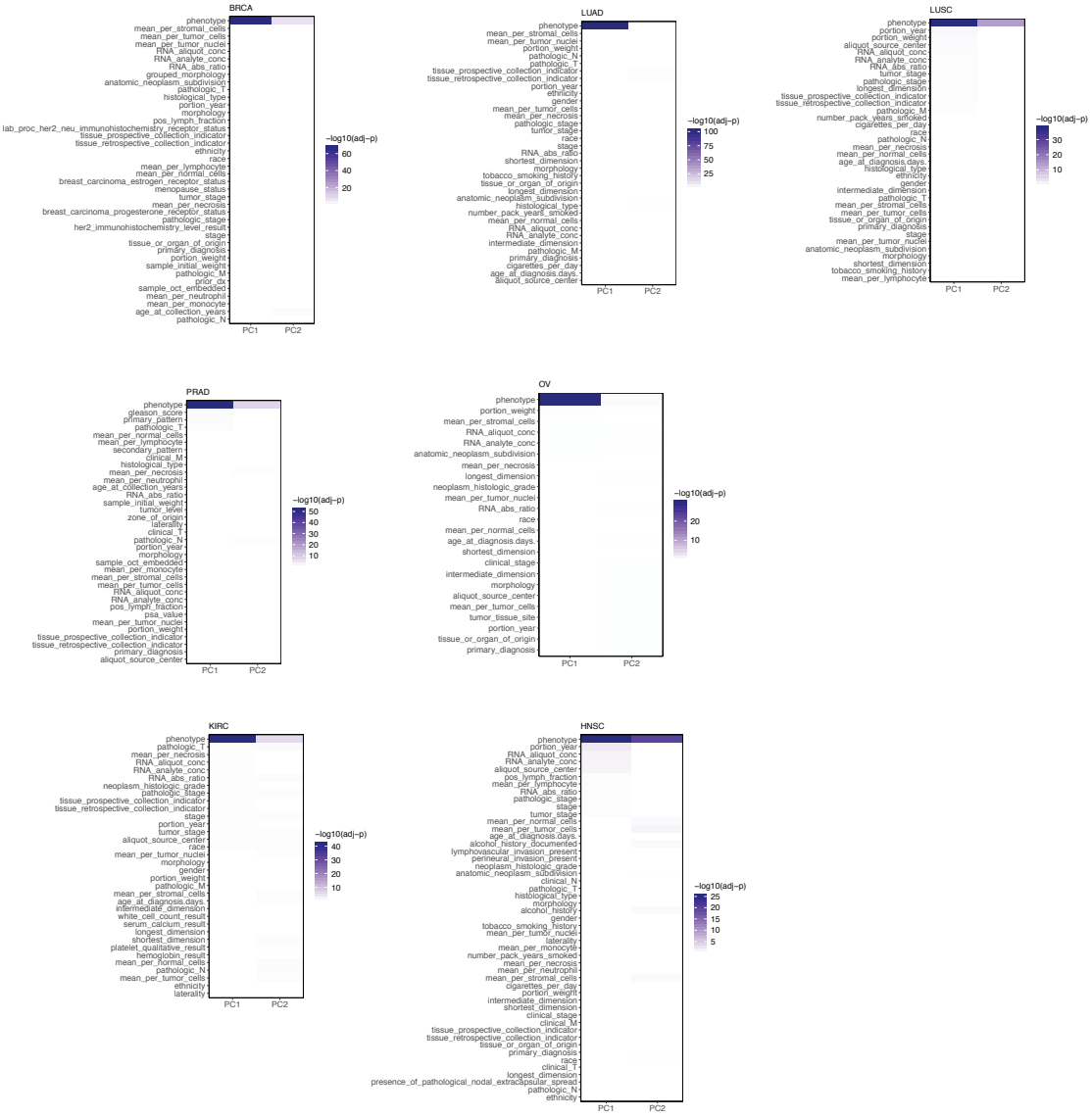
Figure 2.7 ILF3 regulates PKR mRNA abundance and EMT in A549 cells.

A Western blot confirming shRNAmediated ILF3 KD in A549 cells (left). ILF3 mRNA levels (mean ± SD) were quantified in A549 shCtrl and ILF3 KD cells by qRT-PCR (right). ILF3 mRNA expression was normalized against gene TBP mRNA expression. Three biological replicates were performed. p value calculated via t-test, ****$p < 0.0001$.

B Normalized mCherry expression (mean ± SD) for nonedited or edited versions of sites in the 3′ UTR of PKR in shCtrl or ILF3 KD A549 cells. Five biological replicates were

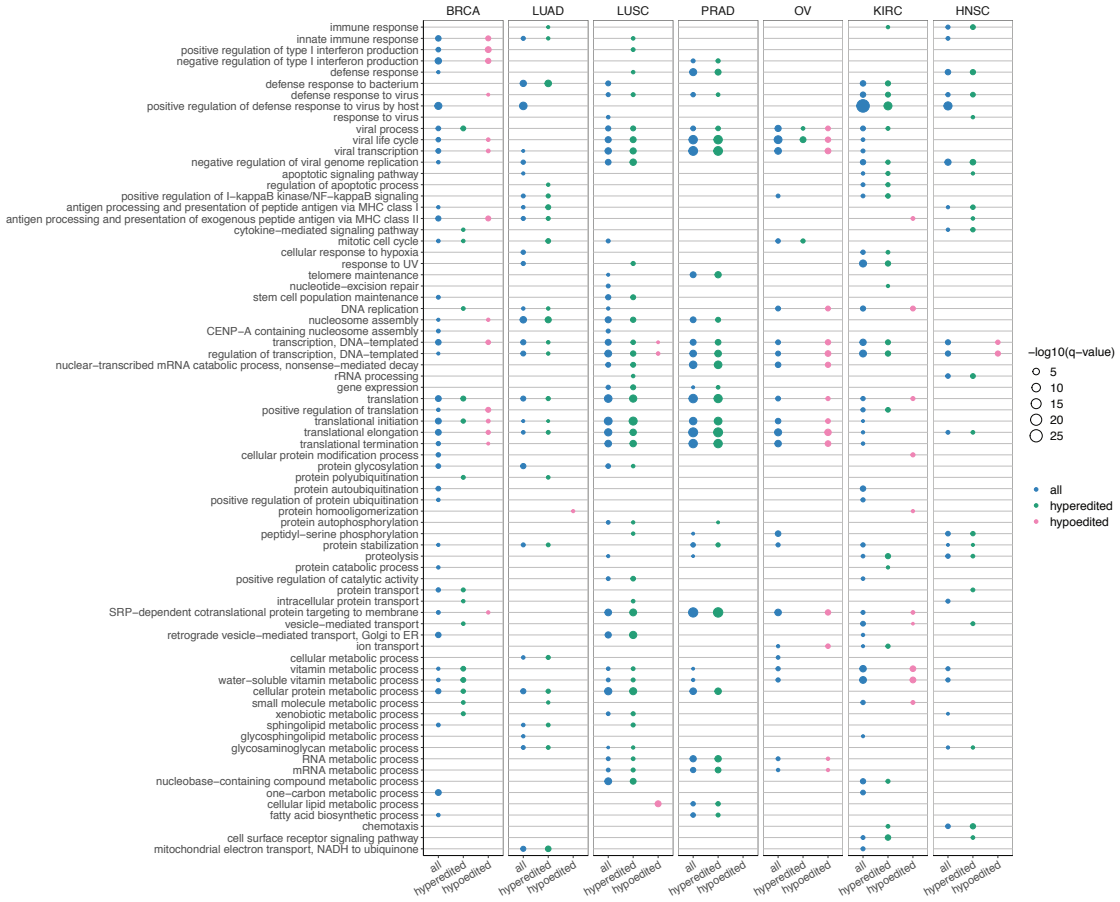
performed. Normalized expression values were compared between edited and nonedited versions by two-sided t-test. *p < 0.05, **p < 0.01, n.s., not significant. **C** Images of A549 cells transfected with siRNAs targeting ILF3 (two different siRNAs were used to KD ILF3, siILF3_1, and siILF3_2) or control siRNAs (siControl). Scale bars: 100 μm. **D** Western blot detecting protein levels of ILF3, E-Cadherin, N-Cadherin, and internal control β-Actin in the siControl, siILF3_1, and siILF3_2 A549 cells. Three biological replicates were carried out for each experiment. **E** Normalized mRNA expression levels (mean ± SD) for ILF3, E-Cadherin, and N-Cadherin in the siControl, siILF3_1, and siILF3_2 A549 cells. Three biological replicates were carried out for each experiment. The expression values were compared between siILF3 and siControl via t-test. **p < 0.01, ***p < 0.001, ****p < 0.0001, n.s., not significant

2.8 Supplementary Figures



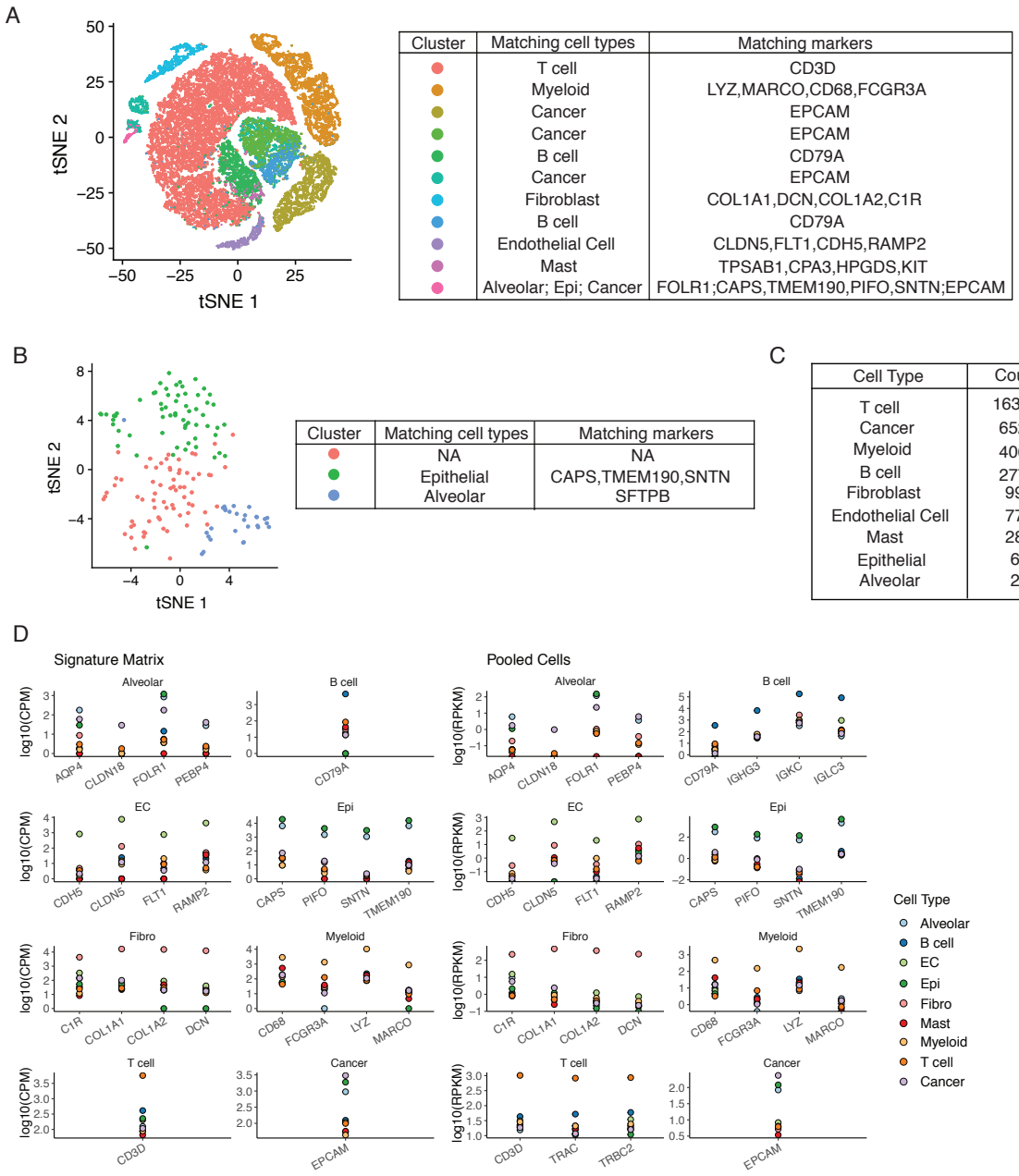
Supplementary Figure 2.1 Differential editing not confounded by metadata.

Heatmaps of significance (log10-transformed adjusted p-values) of correlations between the top two principal components and E/M phenotype among metadata fields in each cancer type. Darker color indicates smaller p-value and stronger association.



Supplementary Figure 2.2 Gene ontology enrichment among differentially edited genes.

Significance of enrichment of gene ontology (GO) terms among all differentially edited genes (blue), only hyperedited genes (green) or only hypoedited genes (pink) of each cancer type. Point size represents the statistical significance of enrichment (log10-transformed adjusted p-value). Terms significantly enriched in at least two cancer types are shown. For cancer types with a global hyperediting trend in M tumors, GO enrichment among hyperedited genes is similar to that among all differentially edited genes. Likewise, for cancer types with a hypoediting trend (BRCA and OV), enrichment among ypoedited genes is similar to that among all differentially edited genes.



Cluster	Matching cell types	Matching markers
●	T cell	CD3D
●	Myeloid	LYZ,MARCO,CD68,FCGR3A
●	Cancer	EPCAM
●	Cancer	EPCAM
●	B cell	CD79A
●	Cancer	EPCAM
●	Fibroblast	COL1A1,DCN,COL1A2,C1R
●	B cell	CD79A
●	Endothelial Cell	CLDN5,FLT1,CDH5,RAMP2
●	Mast	TPSAB1,CPA3,HPGDS,KIT
●	Alveolar; Epi; Cancer	FOLR1;CAPS,TMEM190,PIFO,SNTN;EPCAM

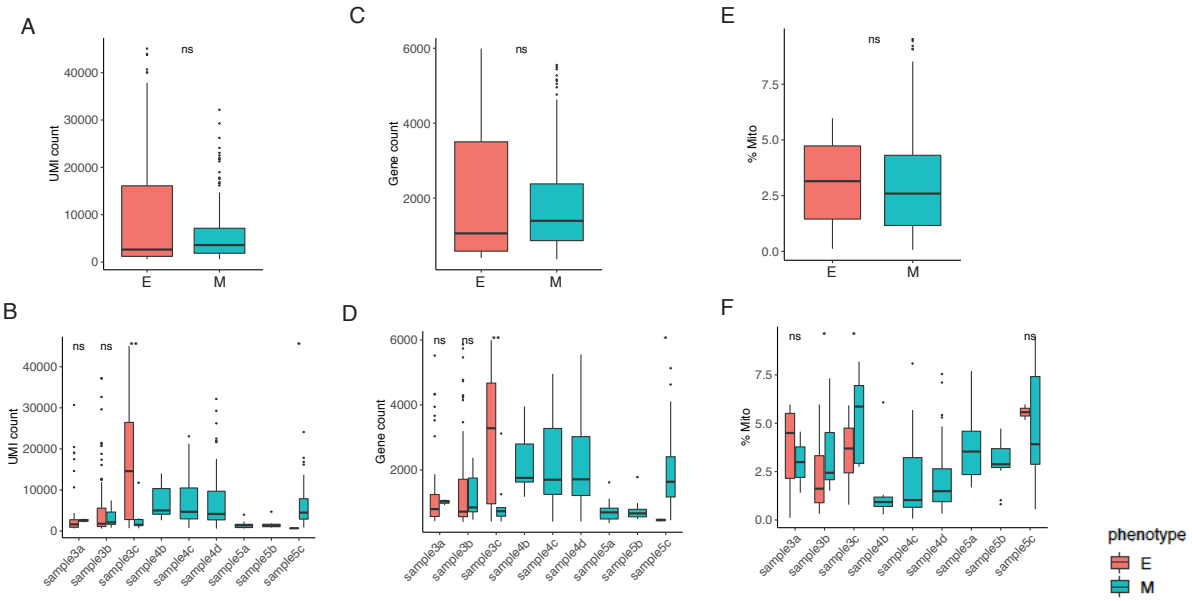
Cluster	Matching cell types	Matching markers
●	NA	NA
●	Epithelial	CAPS,TMEM190,SNTN
●	Alveolar	SFTPB

Cell Type	Count
T cell	16354
Cancer	6526
Myeloid	4063
B cell	2770
Fibroblast	998
Endothelial Cell	771
Mast	285
Epithelial	62
Alveolar	29

Supplementary Figure 2.3 Clustering of single cells from three lung cancer tumors.

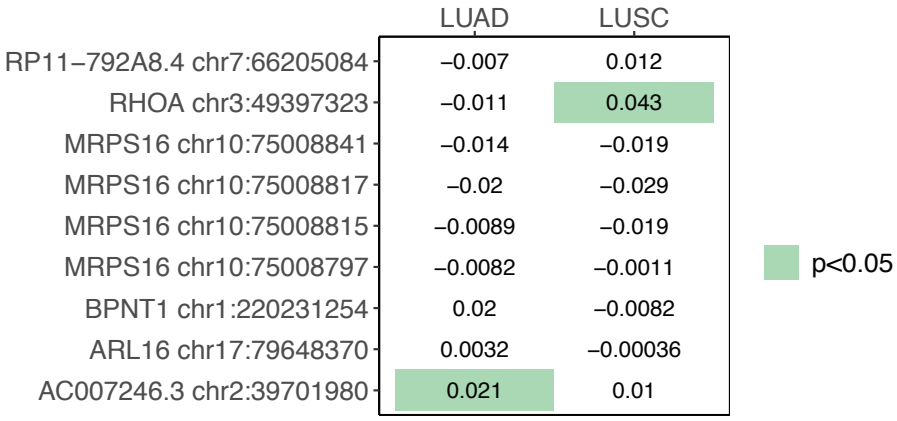
A TSNE projection of cells based on expression profiles, with color indicating cluster identity (left). Cell types were assigned to clusters by matching differentially expressed genes of clusters to known cell type markers (right). **B** TSNE projection of only cells

from cluster 10 to further refine cell type assignment (left). Similar to A, cell types were labeled using differentially expressed genes that matched cell type markers (right). **C** Counts of cells for each cell type after 2 rounds of clustering and cell type assignment (A and B). **D** Log2-transformed expression values of marker genes across cell types. Signature matrix on the left indicates expression values assigned for each cell type by CIBERSORTx. On the right, Pooled Cells indicate that expression values were calculated from pooling reads from cells of the same type together.



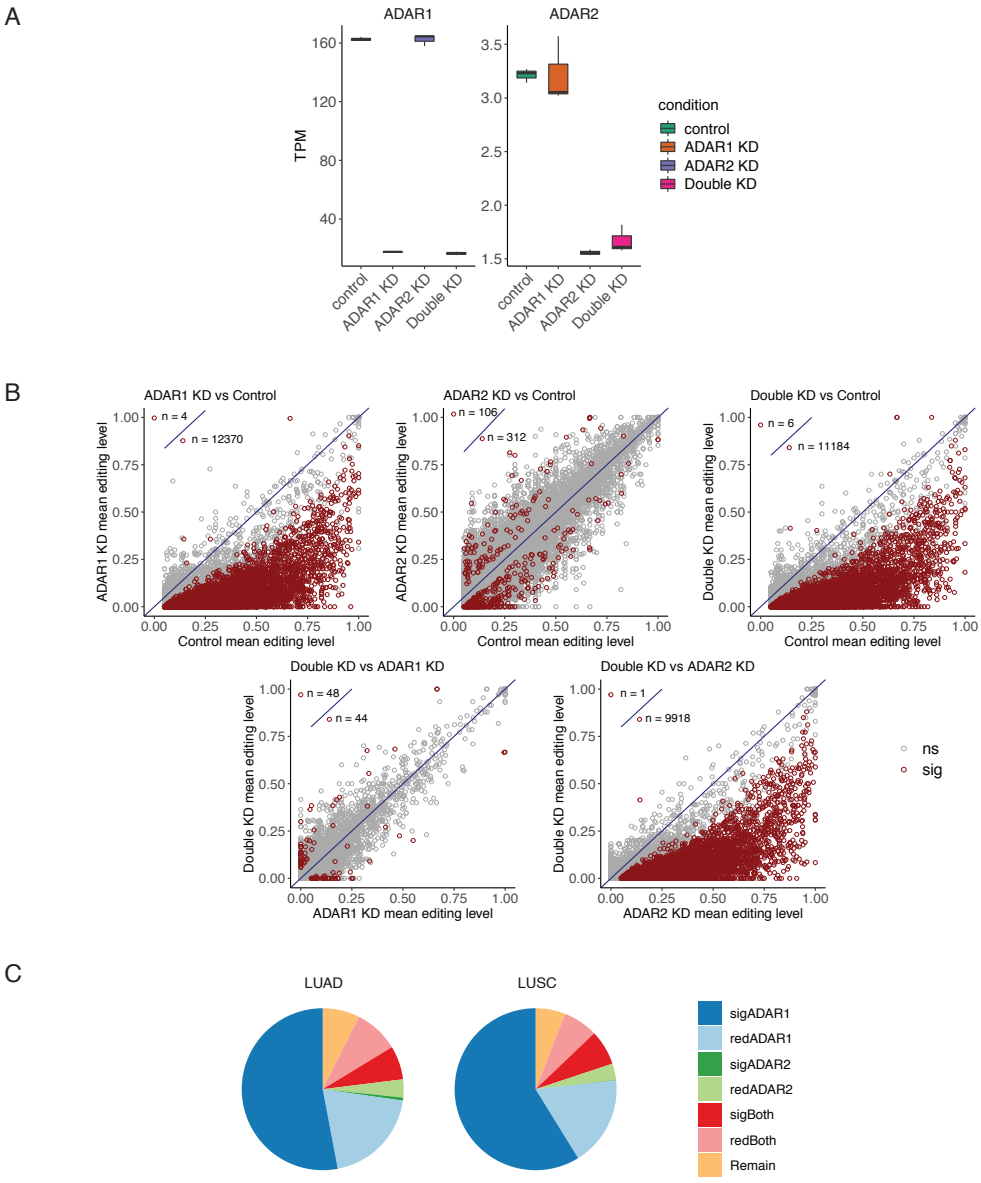
Supplementary Figure 2.4 E and M assignment of single cells not confounded by metadata.

Comparison between E and M cells altogether (top) and within each tumor sample (bottom) of metadata fields: UMI count (A-B), gene count (C-D), and percent of reads mapping to the mitochondrial genome (E-F). Metadata values were compared by Mann Whitney U tests, and significance of p-values are shown. ns: $p > 0.05$, * $p \leq 0.05$, ** $p \leq 0.01$.



Supplementary Figure 2.5 LUAD and LUSC tumor editing differences of differential

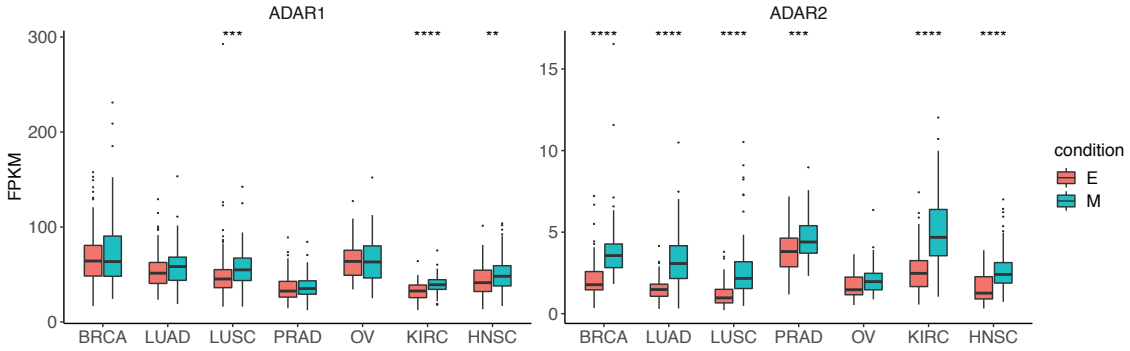
For each editing site, the difference in mean editing levels between M and E tumors (M - E) in each cancer type is listed. Green highlight indicates Wilcoxon p-value < 0.05.



Supplementary Figure 2.6 Altered editing upon knockdown of ADAR1, ADAR2, or both.

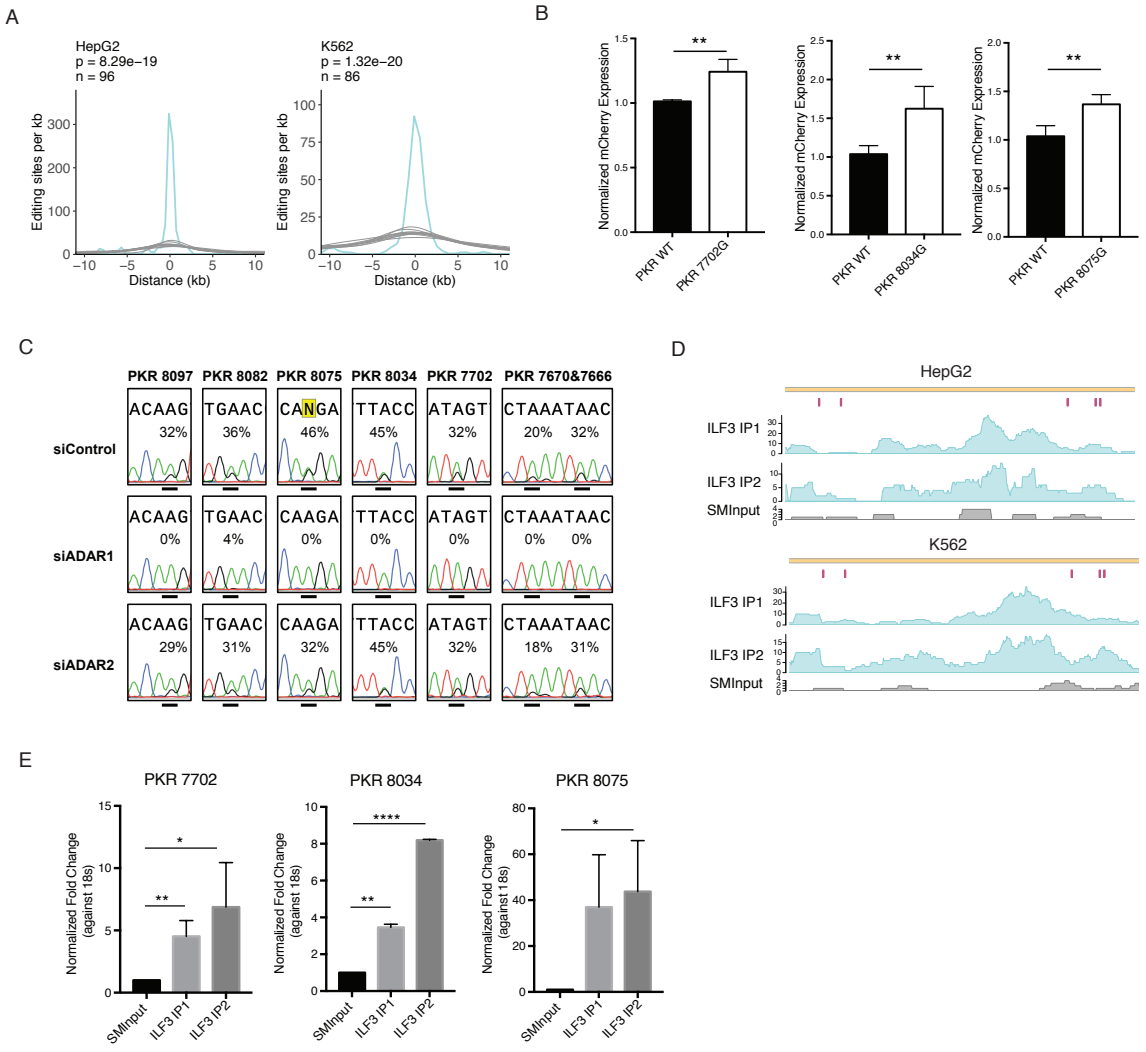
A Distributions of mRNA expression of ADAR1 and ADAR2 under ADAR KD and control conditions. Expression levels were quantified as transcripts per million (TPM).

B Mean editing levels of testable sites in five comparisons between ADAR KD conditions or control experiment. Sites with significant editing differences between conditions are colored red, while gray represents nondifferential sites. Y=x line shown in blue. **C** Proportions of lung cancer E-M differential sites that were also differential in ADAR KD conditions (compared to controls). sigADAR1: sites that were differential only in ADAR1 KD. sigADAR2: sites that were differential only in ADAR2 KD. sigBoth: sites that were differential in both ADAR1 KD and ADAR2 KD, or in double KD. The prefix 'red' indicates reduced editing level by at least 0.05 upon KD from control, but did not pass the statistical significance requirement. 'Remain': editing sites that were not significantly different or reduced across any comparison.



Supplementary Figure 2.7 Expression of ADARs in E and M tumors.

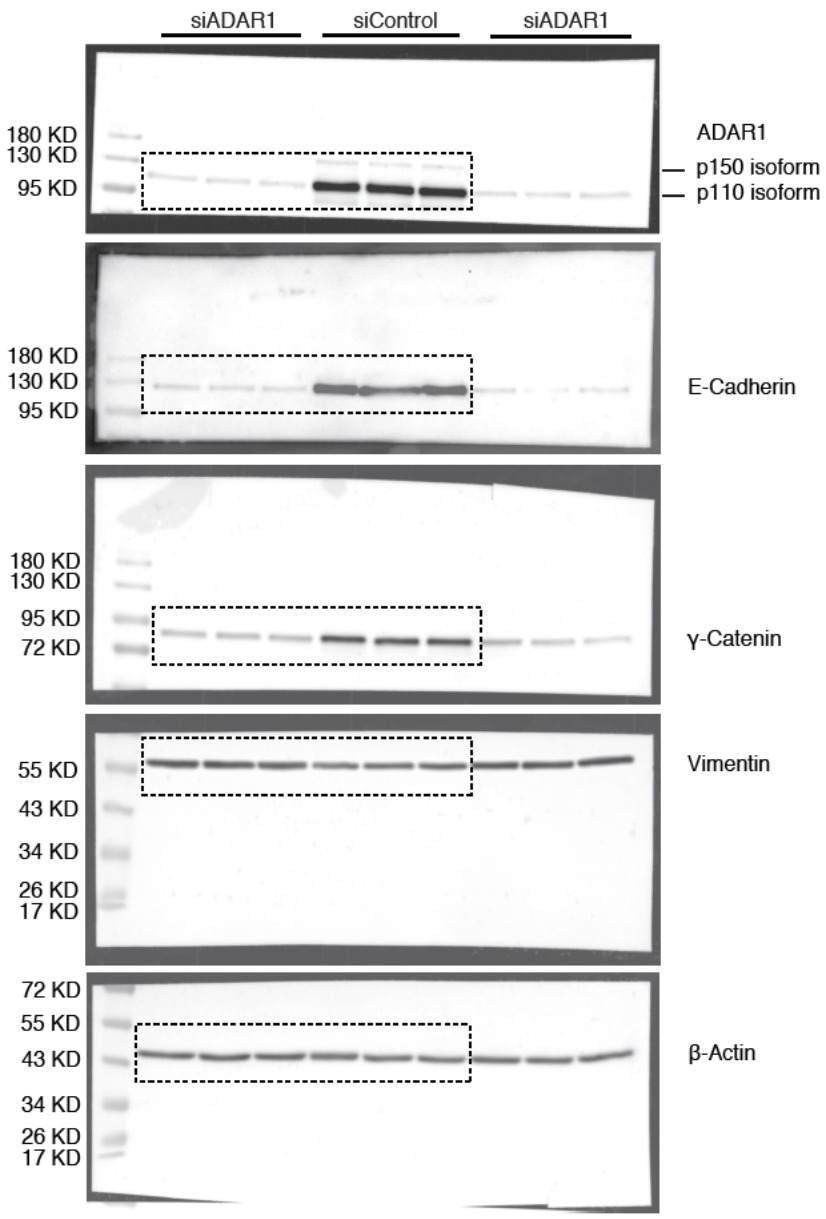
Distributions of mRNA expression of ADAR1 (left) and ADAR2 (right) in E and M tumors across cancer types. Expression values, measured as Fragments Per Kilobase per Million mapped reads (FPKM), were compared by Mann Whitney U tests, and significance of p-values are shown. ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$.



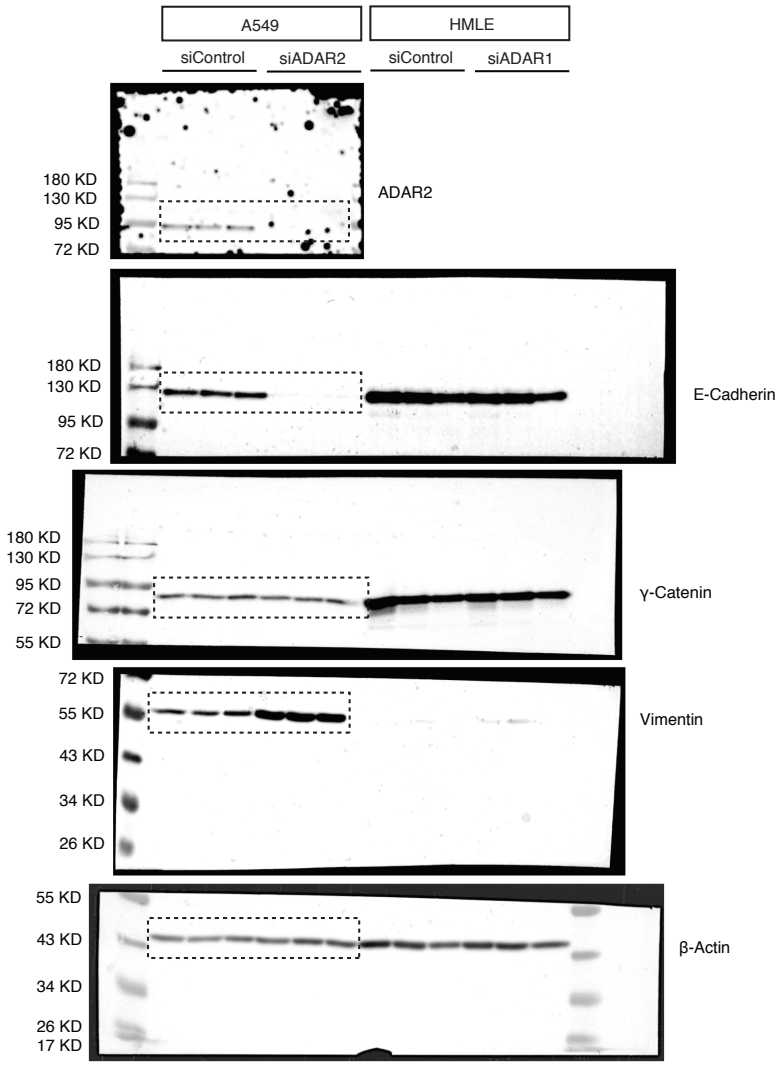
Supplementary Figure 2.8 ILF3 binds closely to the differential editing sites in editing-expression correlated genes.

A Histogram of distances between differential editing sites in editing-correlated genes and the closest ILF3 eCLIP peaks in HepG2 and K562 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 for each cell line. 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

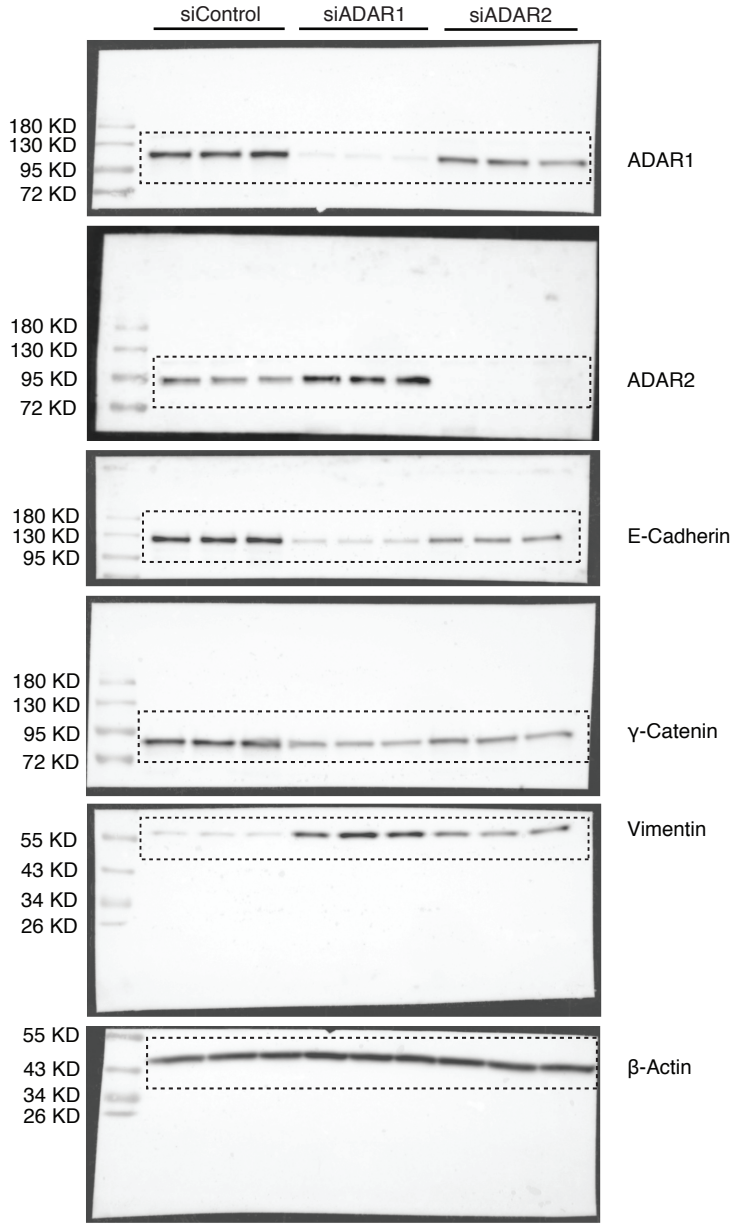
values of 10,000 sets of random gene-matched A's. **B** Normalized mCherry expression for nonedited or edited versions of sites in the 3'UTR of PKR in A549 cells. Five biological replicates were performed. Normalized expression values were compared between edited and nonedited versions by two-sided t-test. **p<0.01. **C** Editing levels of PKR 3'UTR editing sites in siControl, siADAR1 and siADAR2 A549 cells measured by Sanger sequencing. The peak signals of A and G nucleotides were measured by 4Peaks for editing level calculation (G/(A+G)). The editing level of each editing site (underlined) is shown in the graph. **D** Read coverage of ILF3 eCLIP-seq in HepG2 and K562 cells for two biological replicates (ILF3 IP1 and ILF3 IP2, turquoise) and size-matched input (SMInput, gray) in each cell line. The five validated 3' UTR editing sites affecting PKR mRNA abundance in A549 cells are labeled in magenta.
E Validation of PKR eCLIP signal overlapping three 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 (other than two replicates for 8034). P-value calculated by t-test. *p<0.05, **p<0.01, ****p<0.0001.



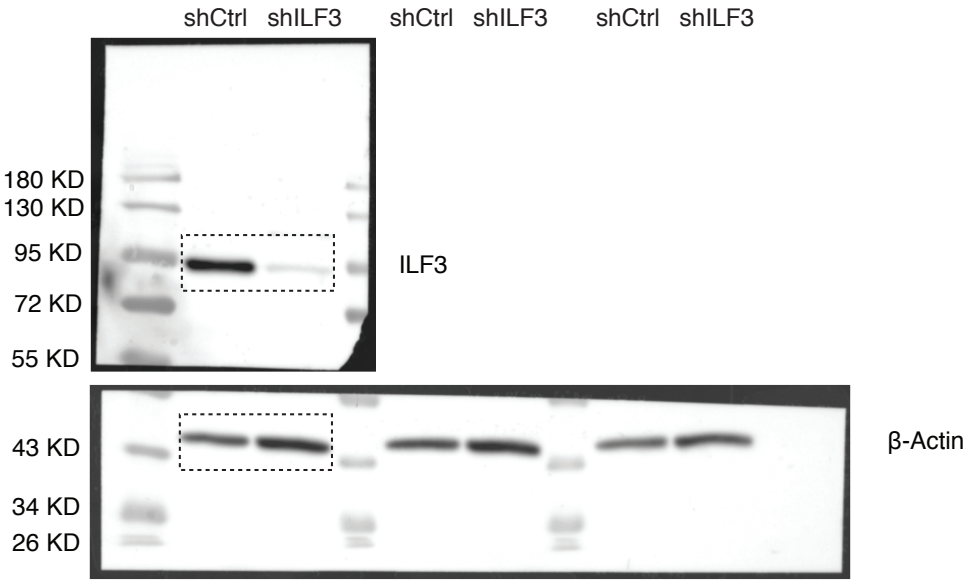
Supplementary Figure 2.9 Uncropped western blot images for Figure 4A.



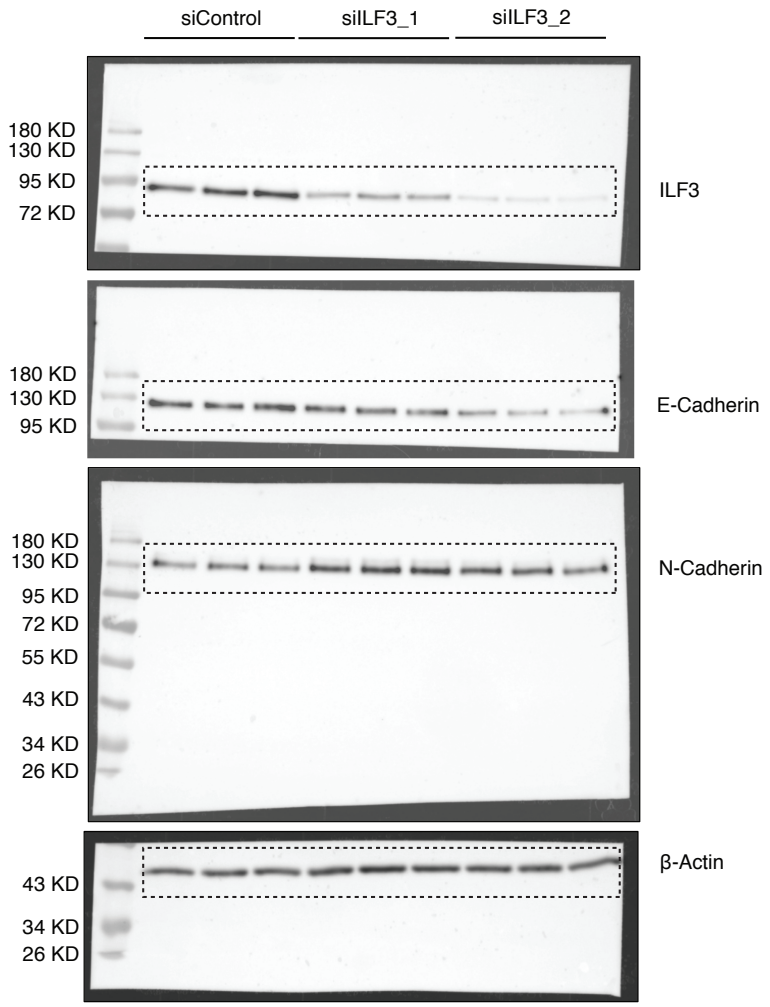
Supplementary Figure 2.10 Uncropped western blot images for Figure 4B.



Supplementary Figure 2.11 Uncropped western blot images for Figure 4C.



Supplementary Figure 2.12 Uncropped western blot images for Figure 7A.



Supplementary Figure 2.13 Uncropped western blot images for Figure 7D.

2.9 Supplementary Tables

Primary_Site	Cancer_Type	Abbreviation	E	M
Breast	Breast Invasive Carcinoma	BRCA	110	110
Lung	Lung Adenocarcinoma	LUAD	100	100
Lung	Lung Squamous Cell Carcinoma	LUSC	93	93
Prostate	Prostate Adenocarcinoma	PRAD	97	97
Ovary	Ovarian Serous Cystadenocarcinoma	OV	37	37
Kidney	Kidney Renal Clear Cell Carcinoma	KIRC	78	78
Head and Neck	Head and Neck Squamous Cell Carcinoma	HNSC	94	94

Table S 2.1 Primary tumor samples used in this study.

Cancer types and the corresponding numbers of categorized E and M tumor samples analyzed in this study.

gene	rna_region	editing_site	editlevel_est	adj_edit_pvalue	consis_cell	consis_logfoldchange
ACOX1	3UTR	chr17:73940077	0.675590107	0.057558713	K562	-0.3
ALDH6A1	3UTR	chr14:74526975	2.642817316	0.000566336	HepG2,K562	-0.18,-0.59
APOL1	3UTR	chr22:36662697	5.043197874	0.023870074	Hela_cyto	-0.33
APOL6	3UTR	chr22:36057354	1.325809518	0.019982168	U87	-0.284
APOOL	3UTR	chrX:84346877	0.881841353	0.024752325	U87,K562	-0.3,-0.39
ARHGAP29	intron	chr1:94634050	2.330463483	0.03925702	HepG2	-0.76
ARPIN	3UTR	chr15:90444654	0.92754555	0.022276919	U87	-0.535
ARRDC1	intron	chr9:140504867	-1.806802676	0.019982168	Hela_w,Hela_cyto	0.347,0.229
ASB16-AS1	noncodingIntron	chr17:42255926	0.596184402	0.074699173	HepG2	-0.17
C19orf48	intron	chr19:51306695	2.354947542	0.025715263	U87	-0.339
C19orf71	intron	chr19:3539995	2.291666376	0.046870966	K562,Hela_w,Hela_cyto	-0.15,-0.461,-1.03
CCNYL1	3UTR	chr2:208619479	1.084244552	3.98E-05	U87	-0.307
CEP104	3UTR	chr1:3729872	-0.807298896	0.084946682	Hela_cyto	0.183
CERS2	noncodingIntron	chr1:150933884	2.051475252	0.044422316	Hela_cyto	-0.451
CINP	intron	chr14:102813733	0.72775171	0.03712208	U87,HepG2,K562,Hela_w,Hela_cyto	-0.211,-0.2,-0.2,-0.511,-0.387
CPM	intron	chr12:69243317	3.885443958	0.000297212	U87,Hela_w,Hela_cyto	-0.257,-0.22,-0.586
CTSS	3UTR	chr1:150704508	4.848763834	0.000143564	U87	-0.985
CXorf56	3UTR	chrX:118673269	0.492247834	0.085175369	U87,K562,Hela_w	-0.265,-0.51,-0.216
CYP20A1	3UTR	chr2:204170105	-0.32963639	0.085175369	Bcell	0.16
DDX58	3UTR	chr9:32456380	1.228575069	0.007953617	U87,HepG2	-0.218,-0.3
EIF2AK2	3UTR	chr2:37328097	1.382670474	2.46E-07	U87	-0.167
EMC1	3UTR	chr1:19542609	1.373657011	0.022662878	U87,Hela_w,Hela_cyto	-0.534,-0.148,-1.27
ERAP1	3UTR	chr5:96110544	1.356393229	0.033976694	U87	-0.286
F11R	3UTR	chr1:160967890	1.02385868	0.094957352	HepG2,K562,Hela_cyto	-0.21,-0.32,-0.741
FAM129A	3UTR	chr1:184761312	-1.577706747	0.028593802	Hela_w	0.506
FAM20B	3UTR	chr1:179042544	1.059798513	0.05628566	HepG2	-0.15
FCF1	3UTR	chr14:75202544	0.791115041	0.03925702	U87,K562	-0.458,-0.25
FGD5-AS1	noncodingExon	chr3:14986104	-0.664122639	0.0638645	Hela_w,Hela_cyto	0.473,0.683
FLNA	intron	chrX:153579411	1.999356154	0.000270927	U87,Hela_w,Hela_cyto	-0.274,-0.301,-1.27
FOXRED2	3UTR	chr22:36884057	2.434689666	0.018654036	U87,Hela_w,Hela_cyto	-0.753,-0.152,-0.19
GFOD2	3UTR	chr16:67716119	-1.381194392	0.072767736	Hela_w,Hela_cyto	0.25,0.21
GGCX	3UTR	chr2:85773390	1.080678262	0.057558713	HepG2,K562,Hela_cyto	-0.35,-0.27,-0.232
H2AFV	intron	chr7:44872505	0.47460002	0.03897698	U87	-0.285
H6PD	3UTR	chr1:9328120	1.348878853	0.018654036	HepG2	-0.57
HPN	intron	chr19:35538724	-1.792221114	0.01799247	HepG2	0.59
HSPB11	intron	chr1:54387885	0.737673592	0.025715263	U87,K562	-0.22,-0.42
IGFBP7	exon	chr4:57976234	3.027681172	2.48E-07	Hela_w,Hela_cyto	-0.17,-1.08
INSR	intron	chr19:7146479	-1.472096252	0.038996066	U87,HepG2	0.395,0.19
IRAK4	3UTR	chr12:44181739	-1.084182315	0.085721376	Hela_w,Hela_cyto	0.436,0.494
KNOP1	3UTR	chr16:19714150	1.382159872	0.077693483	U87,HepG2,K562	-0.267,-0.32,-0.23
LPP	3UTR	chr3:188598857	2.422937272	0.005835479	U87,HepG2	-0.314,-0.21
MDM4	3UTR	chr1:204526595	1.113953173	0.022276919	K562	-0.54
METTL7A	3UTR	chr12:51324639	-1.111493276	0.057558713	HepG2,Hela_w	0.76,0.509
MFSD12	intron	chr19:3540230	1.90702215	0.022276919	HepG2,Hela_w,Hela_cyto	-0.27,-1.06,-1.41
MREG	3UTR	chr2:216808321	-2.891316786	0.024460223	U87,HepG2,Hela_w,Hela_cyto	0.286,0.28,0.56,0.509
MRTO4	3UTR	chr1:19586458	0.585692697	0.025465398	U87,K562,Hela_w	-0.452,-0.27,-0.168
MYO19	intron	chr17:34853684	1.518375774	0.000735249	U87,HepG2	-0.541,-0.26
NACA	intron	chr12:57125246	1.121990286	0.085084237	U87,K562	-0.15,-0.15
NBPF10	3UTR	chr1:145369556	0.127875665	0.051780731	U87	-0.145
NOP14	3UTR	chr4:2940462	1.239264708	4.72E-05	U87	-0.33
NPLOC4	3UTR	chr17:79530071	2.025722247	0.049162216	U87,HepG2,Hela_cyto	-0.336,-0.22,-0.141
NUP155	3UTR	chr5:37291446	1.183902296	0.081405156	U87,HepG2,K562	-0.298,-0.18,-0.46
PAICS	3UTR	chr4:57326875	-2.171471286	0.005405145	Hela_w,Hela_cyto	0.614,0.629
PCCB	intron	chr3:136050150	4.046404225	0.055133335	U87,HepG2,K562	-0.385,-0.36,-0.35
PDLIM5	intron	chr4:95542079	-4.144319345	0.024752325	Hela_w,Hela_cyto	0.374,0.241
PINK1-AS	noncodingExon	chr1:20976109	1.182220784	0.088652742	Hela_w,Hela_cyto	-0.307,-0.616
PLBD2	3UTR	chr12:113827916	1.666217711	3.98E-05	U87,Hela_cyto	-1.43,-0.735
PPIA	3UTR	chr7:44841857	-1.281380018	0.000220889	HepG2,Hela_cyto	0.21,0.181
PRKAR2A	3UTR	chr3:48787856	-0.736600952	0.018612605	Hela_w,Hela_cyto	0.362,0.452
PRKCSH	intron	chr19:11561241	-0.71592712	0.01985595	U87,HepG2,K562	0.314,0.56,0.24
PSMB2	3UTR	chr1:36067752	1.044145565	0.000659271	U87,HepG2,K562,Hela_w,Hela_cyto	-0.325,-0.19,-0.24,-0.403,-0.263
PXMP4	3UTR	chr20:32293069	-0.70072022	0.058943984	Hela_w,Hela_cyto	0.498,0.446
RBBP9	3UTR	chr20:18468339	0.915381444	0.006388843	U87,Bcell	-0.358,-0.188
RBM8A	3UTR	chr1:145513094	0.636371761	0.04105897	U87	-0.217
RDH13	intron	chr19:55550840	-0.902703462	0.054168949	HepG2	0.14
RNF24	3UTR	chr20:3910035	1.843212681	1.26E-09	U87,HepG2	-0.469,-0.4
RPL27A	3UTR	chr11:8708835	2.246128853	0.010482013	U87,Hela_w	-0.177,-0.206
RPL37A	3UTR	chr2:217366903	0.876470214	0.046870966	Hela_w	-0.295
RPL7L1	3UTR	chr6:42856051	-0.826267054	0.010067513	Hela_w,Hela_cyto	0.184,0.205
RPS29	3UTR	chr14:50039685	-2.260993657	0.091869511	HepG2,Bcell	0.44,0.166
SERBP1	3UTR	chr1:67874859	-0.506580781	0.085175369	Hela_w,Hela_cyto	0.594,0.545
SERINC2	intron	chr1:31895247	4.465208772	0.045831535	HepG2,Hela_w,Hela_cyto	-1.05,-0.685,-0.803
SLC25A16	intron	chr10:70279509	-3.60902466	0.057558713	Hela_cyto	0.266
SLC35F5	noncodingIntron	chr2:114465433	0.751160766	0.057338066	HepG2,K562,Hela_cyto	-0.46,-0.49,-0.503
SMPD4	intron	chr2:130915581	-1.292145797	0.013733024	HepG2	0.14
TIAL1	3UTR	chr10:121331855	-3.068906487	0.001484171	Hela_w,Hela_cyto	0.294,0.236
TIMM50	intron	chr19:39982388	-1.057991191	0.02951378	HepG2	0.15
TMEM120B	3UTR	chr12:122216342	0.64131591	0.049727225	HepG2,K562,Hela_w	-0.34,-0.27,-0.229
TMEM59	3UTR	chr1:54496708	1.802146018	0.048523878	HepG2,K562,Hela_cyto	-0.19,-0.29,-0.794
TTC9C	intron	chr11:62507302	0.553395068	0.000767773	Hela_w,Hela_cyto	-0.542,-0.369
TXNDC15	3UTR	chr5:134236752	0.872473286	0.001864405	HepG2,Hela_w,Hela_cyto	-0.44,-0.681,-1.33
UBA1	intron	chrX:47067783	1.955430639	0.001864405	U87,Hela_cyto	-0.19,-0.252
VPS41	3UTR	chr7:38764332	0.754541392	0.081323723	HepG2,Hela_cyto	-0.15,-0.154
ZDHHC20	3UTR	chr13:21949165	1.493075791	0.006085932	Hela_cyto	-0.288
ZNF432	3UTR	chr19:52535007	8.096825832	0.000121773	HepG2,K562	-0.17,-0.14
ZNF552	intron	chr19:58321472	2.305572887	0.052306474	U87	-0.363
ZNF580	intron	chr19:56149705	-0.96540403	0.081405156	U87,K562,Hela_cyto	0.226,0.3,0.29
ZNF587B	3UTR	chr19:58357228	-0.957656315	0.024752325	Hela_w,Hela_cyto	0.872,1.17
ZSWIM1	3UTR	chr20:44513183	1.686372763	0.05628566	U87,Hela_w	-0.485,-0.249

Table S 2.2 List of editing sites predicted to regulate host gene mRNA abundance.

Editing-expression associations (editlevel_est represents editing level regression coefficient and adj_edit_pvalue is the adjusted p-value of the coefficient) were supported by consistent expression changes upon ADAR KD in at least one cell line.

Cloning for minigenes			
RNF24 Fw	GTACCATCGATAGTCTCATGTGGATATGCCTG		
RNF24 Rv	ACGCGTCGACGTTGGGGTAATTTCTGTTGTC		
RNF24 G R	GTAACAAGACCCCGTCTCAACAACAAC		
RNF24 G F	GTTGTTGTTGAGACGGGGTCTTGTTAC		
RhoA_ClaI_F	GTACCATCGATAACCTTGCTGCAAGCACAG		
RhoA_SalI_R	ACGCGTCGACGGATACAGGAAGTTTAGAAAACTGCCTTTATTC		
RhoA_G_F	GTTGGTAACTTTTGTGAATTGGGCTGTAACTAC		
RhoA_G_R	GTAGTTACAGCCCAATTCACAAAAGTTACCAAC		
MRPS16_ClaI_F	GTACCATCGATATGAGCTGACTTTAGTGAGCATAG		
MRPS16_SalI_R	ACGCGTCGACGGAAAATTGAAATCGCACACTGAAATATC		
MRPS16_841_R	CAACACCACAGCCAGCCAATTTTTTAAG		
MRPS16_841_F	CTTAAAAAATTGGCTGGCTGTGGTGTTG		
MRPS16_817_R	CGAGTAGCTGGAACtACGGGTGAG		
MRPS16_817_F	CTCACCCGTaGTTCCAGCTACTCG		
MRPS16_815_R	CGAGTAGCTGGAACCATGGGTGAG		
MRPS16_815_F	CTCACCCATGGTTCCAGCTACTCG		
MRPS16_797_R	ACTgCCTCAGCCTCCCGAGTAGCTGGA		
MRPS16_797_F	TCCAGCTACTCGGGAGGCTGAGGcAGT		
PKR_ClaI_F	GTACCATCGATTGTTACATCATTGCACTTGTAAGTAC		
PKR_SalI_R	ACGCGTCGACAATGTCTAGCATGGGCAAATC		
PKR_8097_F	CAAGTAAATACAGGTCTCAGTCAGATG		
PKR_8097_R	CATCTGACTGAGACCTGTATTTACTTG		
PKR_8082_F	CAGTCAGATGGACCCCAAGAGCCAC		
PKR_8082_R	TGGCTCTTGGGGTCCATCTGACTGAG		
PKR_8075_F	GATGAACCCCAGGAGCCACATG		
PKR_8075_R	CATGTGGCTCCTGGGGTTCATC		
PKR_8034_F	TCTCACACTTTTGCCTGTTACATGG		
PKR_8034_R	CCATGTAACAGGCAAAAGTGTGAGA		
PKR_7702_F	GAATCACAGTTGATGGTTATATGGTGAC		
PKR_7702_R	GTCACCATATAACCATCAACTGTGATTC		
PKR_7670_F	GTGGCTTAAATTCTGAATAACTAGAAACTG		
PKR_7670_R	CAGTTTCTAGTTATTCAGAATTTAAGCCAC		
PKR_7666_F	GGCTTAAATTCTAAATGACTAGAAACTGTATAATAG		
PKR_7666_R	GCCTATTATACAGTTTCTAGTCATTTAGAATTTAAG		
PKR_p2_seq	AGGAGTTGGCAACTAATTG		
Tre-F-seq	ACTACACCATCGTGGAACAG		
Tre-R-seq	GATTATGATCCTCTGGAG		
Generating ILF3 KD contructs			
ILF3_sh_F	CCGGGGTCTTCCTAGAGCGTCTAAACTCGAGTTTAGACGCTCTAGGAAGACCTTTTTG		
ILF3_sh_R	AATTCAAAAAGGTCTTCCTAGAGCGTCTAAACTCGAGTTTAGACGCTCTAGGAAGACC		
qPCR primers			
mCherry qPCR F	ACTACGACGCTGAGGTCAA		
mCherry randomR	CGTTCGTACTGTTCCACGATG		
eYFP qPCR F	AAGATCCGCCACAACATCGA		
eYFP randomR	ACTCCAGCAGGACCATGTG		
qTBP-Fw	CAGCAACTTCCTCAATTCCTTG		
qTBP-Rv	GCTGTTTAACTTCGCTTCCG		
NF90(ILF3) qPCR F	AACCATGGAGGCTACATGAAT		
NF90(ILF3) qPCR R	CGCTCTAGGAAGACCCAAAATC		
18S_F	CTCTTAGCTGAGTGTCCCGC		
18S_R	CTGATCGTCTTCGAACCTCC		
7670 qPCR F	CAGGTCCAAATCAAATTAACCCCATAAG		
7670 qPCR R	CTGTATAATAGGCAAAACTGTGAGGC		
7702 qPCR F	TTGCCTCACAGTTTTGCCTATTATAC		
7702 qPCR R	GTTGATAGTTATATGGTGACATTAGTGGC		
8034 qPCR F	AGATGTACAGTCGCCCCAC		
8034 qPCR R	TTGTCTCACACTTTTACCTGTTACATG		
8075 qPCR F	GCCCCACTACTGGCTTAAATTC		
8075 qPCR R	GAGCCACATGTATTTGAGGGGTAC		
8082 qPCR F	CTGAAAACCATGTAACAGGTAAAAGTG		
8082 qPCR R	CAGTCAGATGAACCCCAAGAG		
PKR qPCR F	CCTGTCCTCTGGTTCTTTTGCT		
PKR qPCR R	GATGATTCAGAAGCGAGTGTGC		
Ecadherin qPCR F	GCCCTTGGAGCCGCAG		
Ecadherin qPCR R	TCAAAATTCACTCTGCCCAGGA		
ZO-1 qPCR F	CAACATACAGTGACGCTTCACA		
ZO-1 qPCR R	CACTATTGACGTTTCCCCACTC		
Ncadherin qPCR F	TCAGGCGTCTGTAGAGGCTT		
Ncadherin qPCR R	ATGCACATCCTTCGATAAGACTG		
Vimentin qPCR F	CGAGGAGAGCAGGATTTCTC		
Vimentin qPCR R	GGTATCAACCAGAGGGAGTGA		
MMP9 qPCR F	TTCTACGGCCACTACTGTGCCT		
MMP9 qPCR R	AATCGCCAGTACTTCCCATCCT		
siRNAs		Vendor	
ADAR1	CGCAGAGUUCCUCACCUGUAUU	Thermo Scienctific Dharmacon	
ADAR2	GCCUGGUUUGCAGUACACTT	Ambion cat# AM51331 ADARB1 Silencer validated siRNA	
Non-targeting siRNA #2	N/A	Thermo Scienctific Dharmacon	
DsiRNAs			
Name	sequence 1	sequence 2	Vendor
hs.Ri.ILF3.13.1	rGrGrArArCrUrCrUrArUrCrArCrArArUrUrUrGrArArArAGA	rUrCrUrUrUrUrCrArArArUrUrGrUrGrArUrArGrArGrUrUrCrCrUrU	IDT
hs.Ri.ILF3.13.3	rGrCrArArArGrCrArUrUrCrUrUrCrCrGrUrUrUrArUrCrCAA	rUrUrGrGrArUrArArArCrGrGrArArGrArArUrGrCrUrUrUrGrCrCrA	IDT
Negative Control DsiRNA	N/A	N/A	IDT
Checking endogenous PKR editing levels			
PKR_editing_F	GGGATTAAGGAAAGGTAAGCATCAAAG		
PKR_editing_R	CAGGTCCAAATCAAATTAACCCCATAAG		

Table S 2.3 List of primers and siRNAs used in this study.

Chapter 3: The landscape of RNA editing in single cells of lung cancer

3.1 Introduction

Among the various types of RNA editing in the human transcriptome, the most prevalent is A-to-I editing, i.e., the deamination of adenosine (A) to inosine (I)[3]. Proteins of the adenosine deaminases acting on RNA (ADAR) family, specifically ADAR1 and ADAR2, catalyze this conversion. Facilitated by advances in genomic technologies, recent studies have highlighted the multi-faceted roles of ADAR activity in cancer[16,17,118,119].

Specific editing sites have been discovered to transform cancer cell behavior, as well as impact the anti-tumor immune response. Edited at higher levels in multiple cancer types, a recoding site in AZIN1 increases cell growth and invasion by heightening the protein's affinity to antizyme and consequently preventing antizyme-dependent degradation of two oncoproteins[16,20,21]. In other examples, translation of mRNAs containing recoding events may generate tumor-associated edited peptides, which may prompt anti-tumor T cell responses specific to these editing-derived antigens[29,30]. Besides protein sequence alteration[118], editing can affect transcript stability of tumorigenic genes[16,120]. For example, a FAK-stabilizing intronic site enables a migratory and invasive phenotype in lung adenocarcinoma (LUAD) cells[46]. In addition,

editing may regulate mRNA degradation by altering miRNA biogenesis, targeting, or binding[5,16,118,121].

Globally, altered editing profiles were evident in tumors of many cancer types[25–27]. The functional implications of most of these editing changes, especially those located in noncoding regions, are not clear. One crucial role of editing in normal cells is to modify endogenous double-stranded RNAs (dsRNAs), likely altering their secondary structures, which consequently prevents self-activation of innate immune response pathways[7,10,122–124]. Considering this editing-mediated regulation of immunity, one proposed consequence of increased editing levels in cancer cells may be repressed interferon production and sustained cell growth[41]. However, in certain cancer types, global editing levels were lower in tumors than in matched normal samples[25]. Furthermore, these observations were obtained from bulk tumors. As tumors are highly heterogeneous with respect to cancer cells and their microenvironment, whether editing aberrations occur in different cell types within tumors and how editing varies across single cancer cells are unknown.

Demonstrating both oncogenic and tumor-suppressive capacities, RNA editing has profound potential for clinical application. In mouse models, combining ADAR deficiency with immune checkpoint blockade (ICB) or DNA methyltransferase inhibitor (DNMTi) therapy improved treatment efficacy through induced interferon (IFN) signaling[34,125]. For particular cancer cells with pre-existing expression of interferon stimulated genes (ISGs), ADAR loss alone caused cell lethality[35,36]. On the level of site-

specific editing, recoding events individually affected drug sensitivity of two cell lines[25]. Deeper understanding of the functions and regulation of editing will aid development of more effective cancer therapies, management of treatment resistance, and classification of treatment-responsive patients.

In this study, we characterized RNA editing in single cells of tumor and normal biopsies from lung cancer patients. We also investigated tumor-associated editing differences across individual cell types and the relationships between editing in cancer cells and features of immune response. These analyses enabled a first global view of the RNA editing landscape in distinct cell types of lung cancer and the potential association of RNA editing with tumor immunity and patient survival.

3.2 Results

3.2.1 **Characterization of RNA editing in single cells**

To examine RNA editing profiles of different cell types in lung cancer, we analyzed single-cell (sc) RNA-seq data from tumors (n = 46) and tumor adjacent tissues (TATs, n = 3) of 30 non-small cell lung cancer (NSCLC) patients[126]. After filtering by quality control metrics, we clustered cells in multiple rounds based on normalized gene expression profiles and assigned cell type labels according to their expression of marker genes (Fig. S1, Methods). Briefly, we labeled clusters initially as immune or non-immune and then subclustered these broad types separately. We annotated immune

subtypes as B cells, T cells, mast cells, macrophages, neutrophils, and plasmacytoid dendritic cells (pDCs). From non-immune clusters, we obtained epithelial cells, endothelial cells, fibroblasts, hepatocytes, and melanocytes. Cancer cells were distinguished from non-malignant epithelial cells by comparing copy-number variation (CNV) estimates to those of reference fibroblasts and endothelial cells[130] (Fig. S2A-B). After removing PCR duplicates, single cells had about half a million uniquely mapped reads on average (Fig. S2C).

We identified editing sites in each cell using our published methods[37,53,80] and requiring their presence in the REDIportal v2 database[132]. On average, an individual editing site was found to be edited (with ≥1 edited read and ≥5 reads in total coverage) in 14 cells and covered (with ≥5 total reads) in 314 cells (Fig. 1A, Fig. S3A). In an individual cell, a mean of 672 sites were edited and 15,037 were covered (Fig. 1B, Fig. S3B). Of the total 1,096,361 sites found to be edited in one or more cells, 50,576 sites were edited in at least 50 cells and located primarily in introns and 3'UTRs, within Alu regions (Fig. 1C-D). To estimate the general editing level of a single cell, we calculated the mean of editing levels of all sites in the cell. Parallel to ADAR1 expression levels, overall editing levels were highest in cancer cells (Fig. S3C-D).

Next, we examined the editing levels of two recoding sites (in AZIN1 and CCNI respectively) known to have higher editing levels in LUAD than normal controls[136]. Both sites have been shown to have cancer-specific functions[16,20,21,29]. Consistent with these reported findings, we observed highest recoding rates in cancer cells at both sites (Fig.

1E). These results support the validity of the single cell RNA editing profiles derived here.

The discovery of targetable oncogenic driver mutations and development of corresponding inhibitors have helped improve survival of lung cancer patients following treatment[137]. Interestingly, we found that editing profiles were distinct among cancer cells grouped by oncogenic driver mutation (Fig. 1F). Furthermore, editing levels across driver mutations were not well correlated with ADAR expression, suggesting the presence of other regulatory mechanisms that may be linked to individual driver mutations (Fig. S3E).

3.2.2 **Tumor-associated editing is distinct among cell types**

Previous studies detected numerous editing changes in bulk LUAD tumors compared to matched tumor-adjacent normal samples[25,26], a finding we reproduced by analyzing the LUAD and control samples in the TCGA project (Fig. S4, Methods). Considering the heterogeneity of the tumor microenvironment observed across patients[83,138], we asked whether tumor-associated editing aberrations were cell-type-specific. To address this question, we first compared editing frequencies of tumor-increased sites from our bulk LUAD analysis across cell types in the scRNA-seq data. We found that these sites were most frequently edited in cancer cells compared to other cell types (Fig. 2A). This observation suggests that heightened editing in bulk tumors

likely reflect altered editing in cancer cells more often than in other cell types present in tumors.

To further investigate the specificity of tumor-associated editing in individual cell types, we applied REDIT LLR[70] to test for editing differences between tumor and normal conditions of each cell type. Using the eight samples from the three patients with TATs, we pooled cells of the same cell type within a sample together. For non-epithelial cell types, we compared editing levels of pooled cells from tumor samples to those from normal samples. Considering cancer cells as the tumor condition of epithelial cells in LUAD, we compared editing levels of pooled cancer cells to pooled non-malignant epithelial cells. Strikingly, cancer cells alone displayed a dominant trend of increased editing compared to non-malignant epithelial cells (Fig. 2B). In contrast, editing levels were generally reduced in the other cell types in tumors (Fig. 2C).

We next sought to determine whether the types of genes harboring tumor-associated editing differ by cell type. To this end, we performed a gene ontology (GO) enrichment analysis on differentially edited genes in each cell type. Enrichment was assessed against background genes with gene length and expression levels comparable to differentially edited genes, as described previously[53,120]. While several categories, such as cell proliferation and apoptotic process, were identified across multiple cell types, certain pathways appeared to be enriched in only specific cell types (Fig. 2D-E). For instance, enrichment of regulation of transforming growth factor beta receptor signaling pathway and cytoskeleton organization were exclusive to cancer

cells. Moreover, the inflammatory pathways of tumor necrosis factor-mediated signaling pathway and interleukin-1-mediated signaling pathway were significantly enriched in differential editing of T cells only. Considering these findings together, tumor-associated editing changes appear distinct across cell types not only in direction but also in gene categories.

3.2.3 **Cancer editing associated with immune suppression**

Helping to prevent autoimmunity, ADAR marks endogenous dsRNAs by RNA editing so that these dsRNAs do not unnecessarily activate cytosolic sensors, such as MDA5 and PKR, and their signaling pathways, which would lead to IFN production, translational repression, and growth arrest[10,40,139]. Consistent with this role of ADAR in normal tissues, certain cancer cells were found to be vulnerable to ADAR loss through stimulation of ISGs and growth inhibition in both in vitro and in vivo models[34–36]. Cancer cells may upregulate RNA editing to exploit this mechanism of innate immune suppression.

To explore this model in human tumors, we examined the contribution of RNA editing to ISG expression in cancer cells. For multiple sets of ISGs, we quantified their overall expression as the mean expression of genes in the set. The following sets of ISGs were included, the Hallmark IFN-alpha and IFN-gamma response gene sets, 60 genes suppressed by IU-dsRNA during poly(IC) transfection[140] (ISGdsRNA), 38 cancer-specific genes associated with resistance to ICB[141] (ISG RS), 38 genes associated with

ADAR dependence[36] (ISGliu), and 30 genes with prolonged IFN-induced expression[142,143] (ISGchronic). We observed that overall editing levels were negatively associated with ISG expression across single cancer cells for several ISG sets, consistent with the model of editing-mediated suppression of IFNs in cancer cells (Fig. 3A). Furthermore, this relationship seemed dependent on the clinically identified oncogenic driver mutation, as cancer cells with the BRAF V600E driver mutation exhibited the strongest negative correlations (Fig. 3B). Grouping cancer cells by treatment timepoint, we also observed most evident negative associations in the progressive disease state (Fig. S5).

As ISGs regulate activation and recruitment of different immune cells[144], we asked whether editing levels in cancer cells are also linked to the infiltration of immune cell types in tumors. For this analysis, we subclustered T cells and macrophages of lung tumors separately (Fig. S6A-B) and calculated the proportions of cell types within each tumor, using our annotations of single cells. To measure a cancer-specific editing index in each tumor, we calculated the mean of editing levels over pooled cancer cells in the tumor. While a negative association between cancer editing and cell proportion was statistically significant only for natural killer (NK) cells, most of the other immune subpopulations also exhibited the same trend (Fig. 3C). The association of reduced immune infiltration and higher editing levels in cancer cells may be partly explained by suppressed interferon signaling that we observed (Fig. 3A-B).

In bulk LUAD tumors from TCGA, we also observed a significant negative correlation between mean editing levels over tumor-increased sites and NK infiltration, as estimated by quanTIseq[134] (Fig. 3D). In contrast, tumor-increased editing levels were positively associated with proportions of CD8+ T cells and M1 macrophages. In agreement with this observation, overall editing levels were highest in tumors of the C2 immune subtype[145], characterized by the highest M1 and CD8+ T cell signatures among all immune subtypes identified across TCGA cancer types (Fig. S7).

3.2.4 **Potential dsRNAs edited to avoid IFN induction**

We next asked which editing substrates are most relevant to ISG signatures in lung cancer. To this end, we searched for pertinent candidates among our in-house set of dsRNAs predicted from editing-enriched regions[146,147]. First, we calculated an editing index for each dsRNA in each cancer cell. We then correlated individual dsRNA editing index levels with mean expression of the ISGdsrna set in cancer cells. This analysis uncovered a number of dsRNAs whose editing levels strongly correlate with ISG expression negatively (Fig. 4A-B). Interestingly, among these top associations were dsRNAs from genes with previously reported oncogenic functions. For example, CTSB was found to promote cancer cell invasiveness, and higher expression of CTSB was correlated with worse survival in multiple cancer types[148–150]. Similarly, AHR and MACC1 were associated with metastasis through enhanced cell motility[151,152].

To further investigate the impact of top dsRNAs, we examined their clinical relevance. For several dsRNAs with the strongest negative ISG correlations, higher editing levels were associated with worse overall survival in TCGA LUAD patients (Fig. 4C). These associations are consistent with the role of RNA editing to suppress dsRNA recognition, avoid innate immune activation, and in effect support tumor growth. Though previously reported in certain cancer types, we did not observe a significant survival association for CTSB gene expression (Fig. S8), highlighting the significant association of RNA editing and survival, unconfounded by gene expression of CTSB.

As associations between overall editing and ISG expression varied by driver mutation, we tested whether CTSB dsRNA editing correlates with the ISGdsrna signature for individual driver mutations separately. While all driver mutations exhibited a trend of negative association, editing levels of the CTSB dsRNA were most strongly associated with ISG expression for cancer cells with EGFR del19 (Fig. S9).

3.2.5 **Interplay between RNA editing and tumor mutation burden**

Another factor linked to cancer immunity is tumor mutation burden (TMB), which has shown potential for predicting response to immune checkpoint blockade (ICB) in certain cancer types, including NSCLC[153–156]. Though higher TMB is thought to help elicit a greater anti-tumor immune response and is associated with better clinical outcomes during ICB treatment, TMB is negatively correlated with survival following targeted EGFR therapy[157,158]. We investigated how RNA editing in individual cell types

relates to somatic mutation burden in cancer cells. We found that overall editing levels in cancer cells were strongly associated with estimated nonsynonymous TMB (Fig. 5A). Assessing this apparent relationship in bulk tumors, we correlated mean tumor-increased editing levels and TMB in the LUAD data from TCGA. Consistent with the observation in the scRNA-seq data, higher cancer-specific editing levels corresponded to higher mutation burden (Fig. 5B).

To further examine the relationship between editing and DNA mutations, we tested the correlations between editing of individual genes and TMB in bulk tumors. Nearly all significant associations between gene editing levels and TMB were positive (Fig. 5C). Additionally, significantly correlated genes were enriched for certain cancer-relevant GO categories: apoptotic process and cellular response to DNA damage stimulus (Fig. 5D). These observations suggest that global editing is correlated with TMB in lung cancer, and gene-specific editing may contribute to the load of RNA mutations of tumors.

Since TMB is associated with cytolytic activity in certain cancer types like LUAD[159], higher TMB may prompt IFN-γ secretion by cytotoxic T cells[160,161] through increased neoantigen load. As a result, IFN-γ may induce ADAR p150 expression[162] in cancer cells. Exploring this possible mechanism to explain the observed association between RNA editing and TMB, we tested whether ADAR expression levels are correlated with TMB in both bulk tumors and pooled cancer cells from the scRNA-seq data. We found that ADAR expression was weakly associated with TMB (Fig. S10).

Thus, IFN-stimulated ADAR expression does not clearly mediate the relationship between RNA editing and TMB.

We next considered the effects of both RNA editing and TMB on survival of LUAD patients from TCGA. To do so, we fit a Cox proportional hazards model with TMB, tumor-increased mean editing level, age, and gender. Of all the covariates included in the model, only mean editing was significantly associated with overall survival (Fig. 5E), suggesting that increased RNA editing is an important factor in LUAD prognosis.

3.3 Discussion

We conducted the first global study of RNA editing in single cells from cancer. Single cell analysis enabled us to examine the variability of RNA editing due to tumor heterogeneity and dissect tumor-associated editing changes by cell type in lung cancer patients. We report distinct shifts in editing according to cell type, with regards to both direction and enriched biological processes. We have also shown that editing levels in cancer cells correlate with interferon stimulation signatures, tumor infiltration, tumor mutation load and patient survival.

As a mechanism to avoid mistakenly inducing an innate immune response and halting cell growth, ADAR-catalyzed inosines mark endogenous dsRNAs as self

RNAs[7,10,123,124]. The resulting ADAR dependence in certain cancer cell lines, patient-derived xenografts, and mouse models indicates that ADAR acts as an immune checkpoint in cancer with important clinical implications[34–36,125]. However, beyond previous negative associations between hyperediting and general tumor inflammation based on combined TCGA cancer types[34], this regulatory mechanism has not been demonstrated extensively yet in human tumors. Furthermore, the most critical dsRNAs in these cancer types have not been previously characterized. Our findings provide additional support for the hypothesis that RNA editing suppresses the innate immune response in NSCLC. We also identified potential dsRNAs whose editing levels likely affect interferon stimulation in lung cancer cells and are associated with survival. Notably, the most highly associated dsRNAs overlapped genes with functions in tumorigenesis and metastasis, such as CTSB, MACC1 and AHR[148,151,152]. In addition, MACC1 and AHR were reported to play a role in the expression of PD-L1, an immune checkpoint[163,164], in gastric and lung cancers, respectively[165,166]. Thus, these genes may suppress anti-tumor immune response via multiple mechanisms, which will need to be validated in future studies.

For certain lung cancer patients whose tumors develop resistance to targeted therapy, immunotherapy may be a suitable option as the next line of treatment. Despite limited clinical data, patients with BRAF V600E alteration appear to have a higher response rate to ICB than patients with other driver mutations in EGFR or ALK[167]. Consistently, we observed strongest negative correlations between editing and ISG signatures in cancer cells with BRAF V600E, suggesting that ADAR loss could cause

IFN production, growth arrest, and possibly improved response to immunotherapy. Consequently, patients with this particular driver mutation may be potential candidates for a combination therapy of ICB and ADAR inhibition. Further investigation of how driver mutations and RNA editing affect innate and adaptive immunity will inform biomarker identification and patient selection for ICB combined with ADAR silencing.

Our results suggest that NK cells may have close relevance to RNA editing-mediated immune response in tumors. NK cells are cytotoxic innate immune cells that can prevent tumor progression. Their antitumor activity can be activated inherently without requiring specific antigen presentation. Thus, NK cells are increasingly recognized as potential candidates to facilitate cancer immunotherapy[168]. We observed that the editing levels of cancer cells negatively correlated with the number of NK cells in the tumor. This negative correlation, together with the observation of heightened editing in cancer cells, points to the possibility that hyperediting in cancer cells contributes to their evasion from NK cell destruction. Future studies need to be carried out to substantiate such a causal relationship. In addition, whether RNA editing may affect the cytotoxicity or IFN production of NK cells needs to be investigated.

Although TMB is a biomarker for ICB response in NSCLC [156,169], higher TMB was previously associated with poor survival following targeted therapy and chemotherapy[157,158]. We, along with others[136], have observed an association between TMB and RNA editing in bulk tumors of LUAD. Why these features are positively correlated is not yet clear. Importantly, we show that this relationship is specific to

cancer cells of tumors and may involve editing of specific genes in DNA damage response. In addition, we found that tumor-increased RNA editing was strongly associated with survival while TMB was not. In summary, these data highlight the prognostic value of RNA editing and motivate further evaluation of editing metrics as predictors for response to cancer therapies.

3.4 Methods

3.4.1 **Single cell RNA-seq data processing and cell type assignment**

We downloaded scRNA-seq fastq files from BioProject PRJNA591860[126]. After checking data quality with FastQC and aligning reads to hg19 with STAR[127], we ran HTSeq count to obtain gene-level expression counts with --mode=intersection-nonempty. Using the R package Seurat[128], we combined these count data into a Seurat object and retained cells based on the following quality control metrics: sequencing depth (> 50,000 reads), features (> 500 genes), and mitochondrial read content (< 50%). For these remaining cells, expression profiles were normalized by sctransform[68] before linear dimensional reduction with principal components analysis (PCA). We used the top 20 PCs to cluster cells (with cluster resolution = 0.09) and run non-linear dimensional reduction by UMAP. Based on expression of broad compartment marker genes for epithelial (EPCAM), endothelial (CLDN5), immune (PTPRC), and stromal (COL1A2) cells[129], we labeled clusters initially as immune and non-immune cells.

Next, we subclustered non-immune cells after running sctransform and PCA on only non-immune cells (20 PCs, cluster resolution = 0.12). Each cluster was assigned a cell type based on its expression of known cell type marker genes[126]. Alveolar cells were grouped with epithelial cells for downstream analyses. To classify cancer cells apart from non-malignant epithelial cells, we applied inferCNV[130] to estimate large-scale CNVs in epithelial cells with fibroblasts and endothelial cells as reference and spike-in controls. Specifically, amplification and deletion regions were approximated from comparing gene expression profiles of input cells to those of labeled reference cells (1000 fibroblasts and 500 endothelial cells). Input cells included all epithelial cells, as well as the spike-ins: 1000 fibroblasts and 500 endothelial cells.

Immune cells were subclustered with the same procedure, using 24 PCs and cluster resolution = 0.07. Including only lung biopsies, we also further subclustered T cells (20 PCs, resolution = 0.18) and macrophages (11 PCs and resolution = 0.2) separately in the same manner.

3.4.2 **Bulk RNA-seq data processing**

We downloaded RNA-seq fastq files for tumor and normal samples of the TCGA Lung Adenocarcinoma (LUAD) project from the Genomic Data Commons (GDC) Legacy Archive[49]. Reads were first mapped to hg19 with HISAT2[131]. To account for hyperedited regions that could cause many mismatches to the reference genome in reads, we ran a hyperediting pipeline[52,53] on initially unmapped reads. Reads rescued

from this pipeline were combined with uniquely mapped reads from the first round of alignment.

3.4.3 **Identification of RNA editing sites**

Using our previously published methods[37,53,80,120], we detected editing events within each single cell at sites recorded in the REDIportal v2[132] database. If an editing site overlapped a variant listed in dbSNP (version 147) or COSMIC (version 81) and was not previously reported as a cancer-associated editing site, we excluded the site from downstream analyses.

Similarly, we quantified editing levels at REDIportal sites in bulk tumor and matched normal tissue samples from TCGA LUAD. In addition to removing sites within dbSNP and COSMIC databases, we filtered out editing events overlapping sample-specific somatic mutations from the MC3 MAF[133] and CNV data from the GDC data portal.

3.4.4 **Differential editing by cell type**

To compare editing levels between tumor and normal conditions in a single cell type, we first pooled counts of edited and unedited reads of single cells of that cell type within each sample. We used only tumor samples from the three patients with matched normal samples. Treating pooled cells of each sample as a replicate (5 tumors and 3

normals) for each cell type, we ran REDIT-LLR[70] to test for tumor versus normal editing differences. A site was considered testable if it was covered by at least 5 reads in at least one pooled sample in each condition. Significance of editing differences was determined by REDIT FDR-adjusted p-value < 0.05 and difference in mean editing levels ≥ 0.05.

3.4.5 **Differential editing in bulk tumors**

We used REDIT-regression[70] to identify sites that were differentially edited between bulk tumors and normal samples from TCGA LUAD. For each testable editing site, the following covariates were considered in the regression model: gender, race, age, and sample type (tumor or normal). An editing site was defined as differential if the sample type FDR-adjusted p-value < 0.05 and the difference in mean editing levels ≥ 0.05.

3.4.6 **Quantification of tumor mutation burden**

We downloaded the list of prioritized somatic mutations provided for individual cancer cells (Table S3)[126]. For each tumor sample, tumor mutation burden was estimated as the number of unique nonsynonymous mutations from this list across cancer cells in that sample. This mutation list was also used to filter editing events in cancer cells before correlating TMB and editing levels.

For each bulk TCGA LUAD tumor, we calculated tumor mutation burden as the number of unique somatic nonsynonymous mutations from the MC3 MAF.

3.4.7 **Immune cell infiltration estimates**

We downloaded quanTIseq[134] estimates of immune cell proportions in TCGA tumors from TIMER2.0 (http://timer.cistrome.org)[135]. The quanTIseq method was chosen due to the interpretability of its output as cell fractions and the inclusion of more immune subtypes.

3.5 Figures

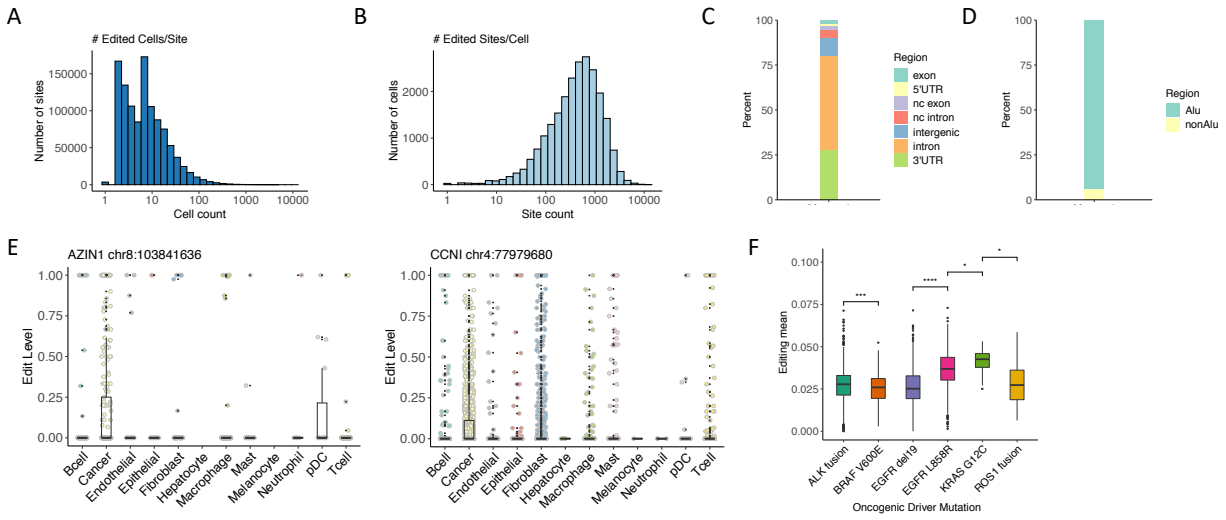


Figure 3.1 Overview of editing events detected in single cells.

A Histogram of the number of single cells edited per editing site, on a base 10 logarithm-transformed scale. A site was considered edited in a cell if the site was covered by at least five reads and editing was supported by at least one read. **B** Histogram of the number of sites edited per cell, on a base 10 logarithm-transformed scale. Editing criteria were the same as in A. **C** Percentages of where common editing sites were located show that sites were found mainly in intronic and 3'UTR regions of genes. Common editing sites comprised sites edited in at least 50 cells. **D** Most common editing sites were in Alus. **E** Distributions of single cell editing levels at each of two cancer-associated recoding sites across cell types. For the AZIN1 recoding site, $p < 0.05$ for pairwise comparisons of cancer vs the following cell types: T cells, macrophages, B cells, fibroblasts, mast cells. For the CCNI recoding site, $p < 0.001$ for pairwise comparisons of cancer vs the following cell types: pDCs, T cells, macrophages,

B cells, epithelial cells, endothelial cells, fibroblasts, mast cells. Editing levels were compared for each pair of cell types by Mann Whitney U test, when sufficient observations were present. **F** Overall editing levels of cancer cells grouped by oncogenic driver mutation. Overall editing level was calculated as the mean over editing levels of all sites for each cell. Editing levels were compared among driver mutation groups by Mann Whitney U test, with p value significance shown. *p <= 0.05. ***p <= 0.001. ****p <= 0.0001.

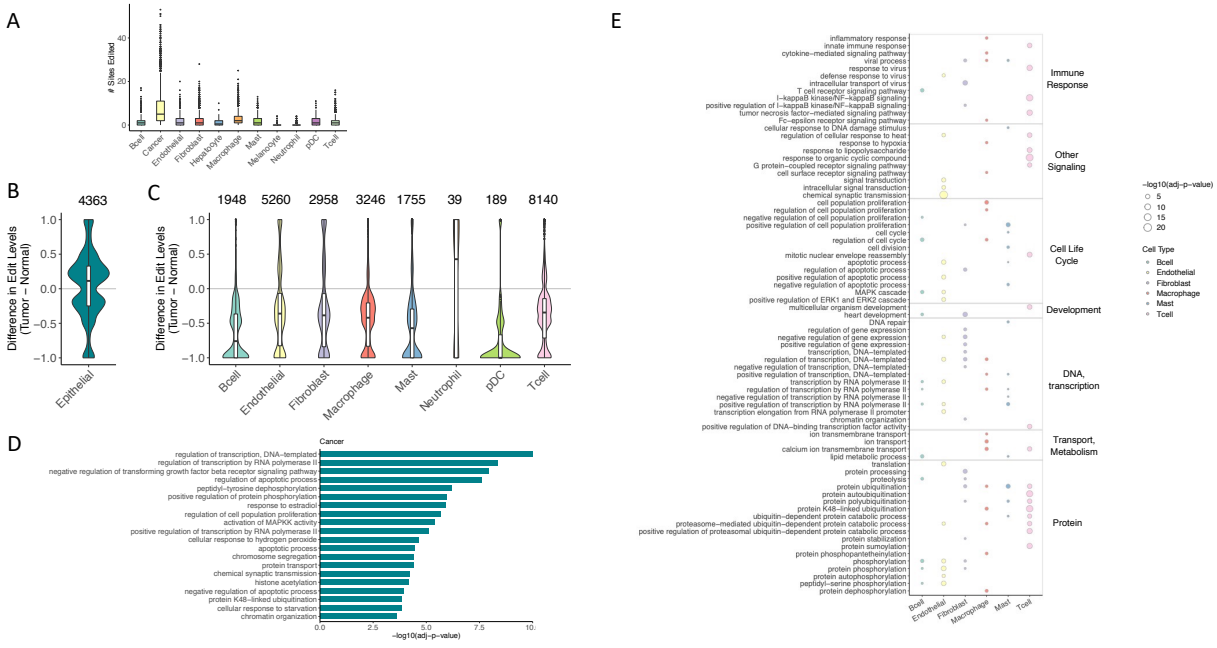


Figure 3.2 Tumor-increased editing is specific to cancer cells.

A Counts of edited sites in single cells, considering only the significantly tumor-increased editing sites from bulk tumors. A site was considered edited in a single cell if the site was covered by at least five reads and editing was supported by at least one read. p <= 0.0001 when comparing cancer to any other cell type by Mann Whitney U

test. **B** Distribution of differences in mean editing levels between pooled cancer cells and pooled non-malignant epithelial cells. Only sites with significant differences (mean difference >= 0.05, REDIT LLR adjusted p-value < 0.05) are included. The number of significantly different editing sites is labeled on top. **C** Similar to **B** but for differential editing sites between pooled cells of tumor samples and pooled cells of normal samples for each non-epithelial cell type. **D** Top gene ontology (GO) enrichment in genes containing differential editing sites between cancer cells and non-cancerous epithelial cells, compared to background genes without differential sites but matched according to gene length and expression. Only the 20 most significantly enriched terms, each with a minimum of 10 corresponding genes, are shown. **E** Top GO enrichment among tumor vs normal differential edited genes in non-epithelial cell types. As in D, only the top 20 terms are included for each cell type, and background genes were chosen in the same manner. Larger circle size represents higher statistical significance of enrichment. Color of circle fill corresponds to cell type. Terms are grouped by broader categories labeled on the right.

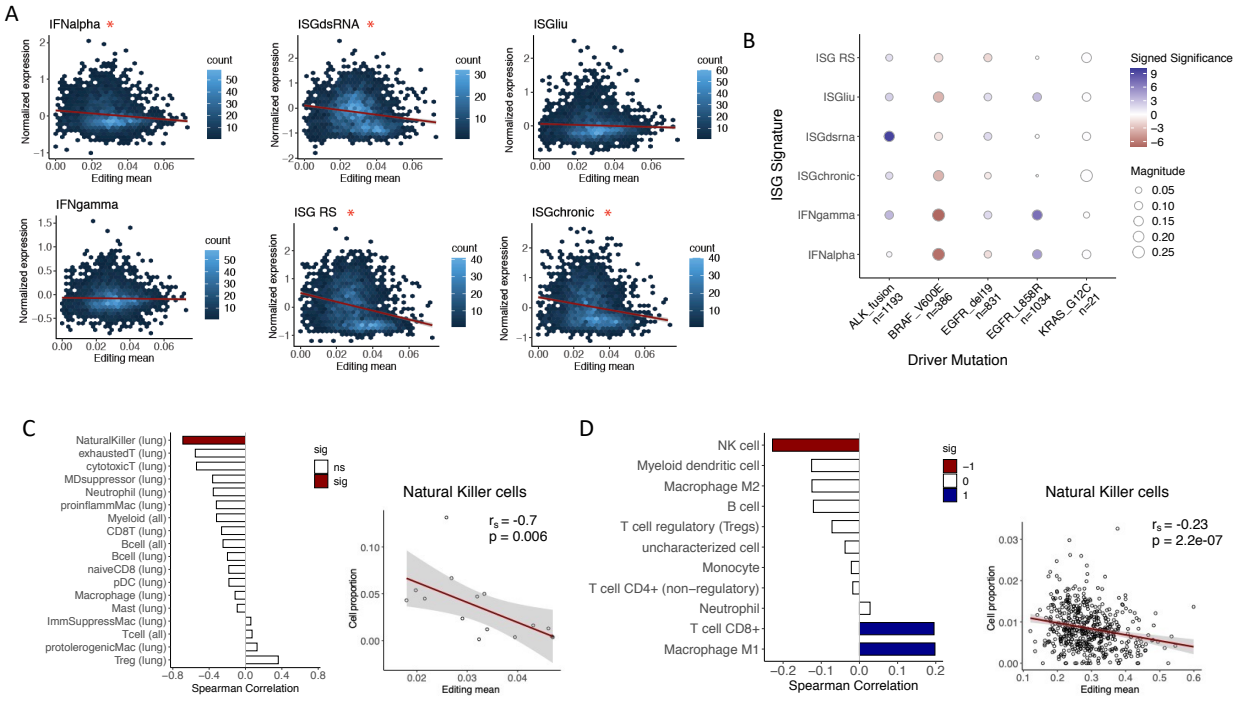


Figure 3.3 Editing in cancer cells associated with immune suppression.

A Hexagonal 2-dimensional histograms of mean editing level and mean normalized ISG expression in single cancer cells for multiple ISG signatures. Mean editing level was calculated over editing levels of all sites for each cell. Red asterisk indicates a significant negative spearman correlation, with FDR-adjusted p-value < 0.05.

B Spearman correlations between mean editing levels and mean expression of multiple ISG signatures across single cancer cells, grouped by oncogenic driver mutation. The number of cancer cells with each driver mutation is listed under each mutation label. The size of each circle indicates the magnitude of the Spearman correlation coefficient, and the color intensity corresponds to significance of the adjusted p-value. Blue represents positive correlations and red, negative ones. **C** Bar plot on the left shows Spearman correlations between cancer editing levels and infiltration of different immune cell types. For each tumor, single cancer cells were pooled, and overall cancer editing

level was calculated as the mean editing level over all sites in the pooled cancer cells. Red color (sig) indicates significant correlation with $p < 0.05$. Nonsignificant correlations are shown in white (ns). For each cell type, in parentheses, lung signifies that only lung biopsies were included. In contrast, ”all” signifies all samples were included. On the right is the scatterplot of cancer editing and infiltration of Natural Killer cells, with Spearman correlation coefficient and p-value listed. **D** Bar plot of Spearman correlations between tumor-increased editing and estimated infiltration of different immune cell types in bulk TCGA LUAD tumors. Tumor-increased editing was calculated as the mean editing level over sites with significantly higher editing levels in tumors than in matched normal samples. Colored bars indicate significance by FDR-adjusted $p < 0.05$, with positive correlations in blue (1) and negative ones in red (-1). Nonsignificant correlations are shown in white (0). Scatterplot of tumor-increased editing and quanTIseq-estimated proportion of Natural Killer cells with listed Spearman correlation coefficient and p-value on the right.

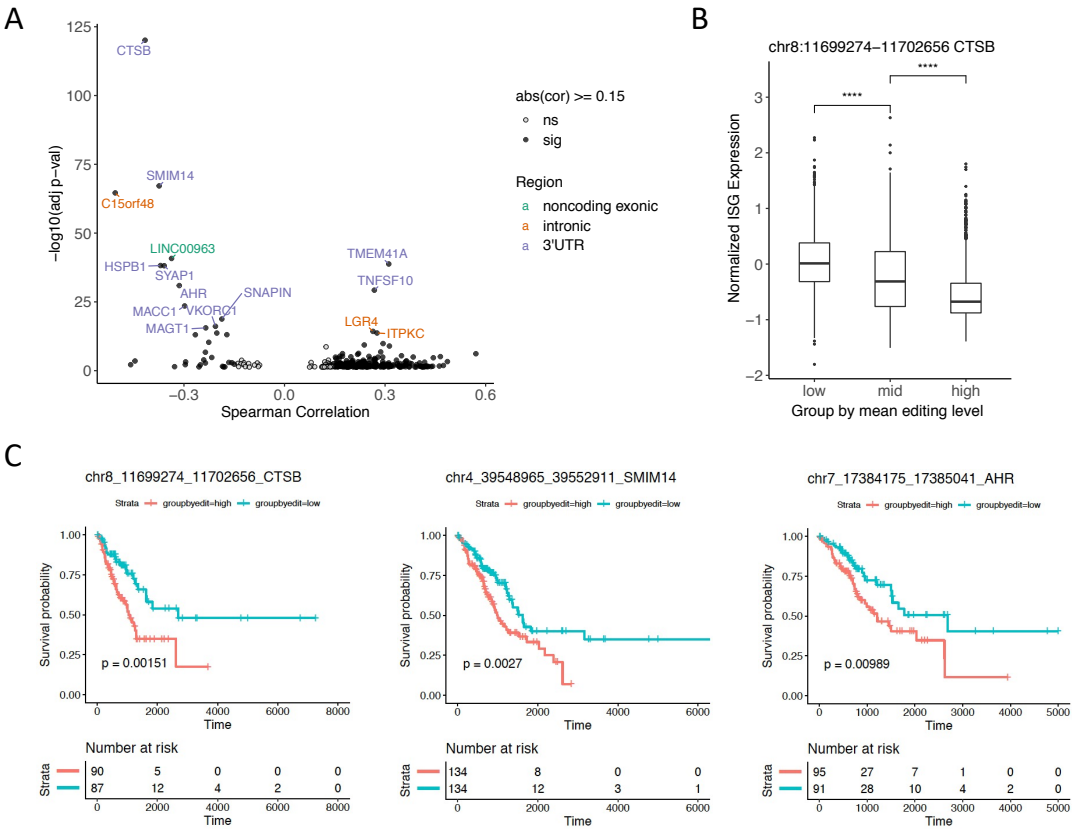


Figure 3.4 Candidate editing substrates contributing to ISG signatures in cancer cells.

A Scatterplot of Spearman correlation coefficient and its statistical significance (-log10(FDR-adjusted p-value)) for associations between editing levels of individual dsRNAs and mean ISG expression levels across single cancer cells. Only dsRNAs passing FDR of 5% are included in the plot. Black-filled points correspond to dsRNAs with correlation coefficients of at least 0.15 in magnitude (sig), and dsRNAs that have correlation coefficients lower than 0.15 have gray-filled points (ns). The top 15 dsRNAs are labeled by the names of their overlapping genes, which are colored by the specific regions the dsRNAs overlap. If a dsRNA overlapped multiple regions of a gene, we used the following prioritization to assign a single region to the dsRNA: 3' UTR > exonic

> 5’ UTR > intronic. **B** Example of editing-ISG association for a dsRNA in CTSB. Distributions of mean ISG expression are shown for cancer cells grouped by tertile editing levels of the labeled CTSB dsRNA. ISG expression was compared across dsRNA editing groups by Mann Whitney U test, with p value significance labeled. ****p <= 0.0001. **C** Kaplan Meier plots showing worse survival for patients with higher editing levels of 3 dsRNAs (labeled on top). For one dsRNA, patients in the high group (red curve) had editing levels among the top third across all patients with a mean editing level for that dsRNA. Patients in the low group (blue curve) had editing levels among the lowest third.

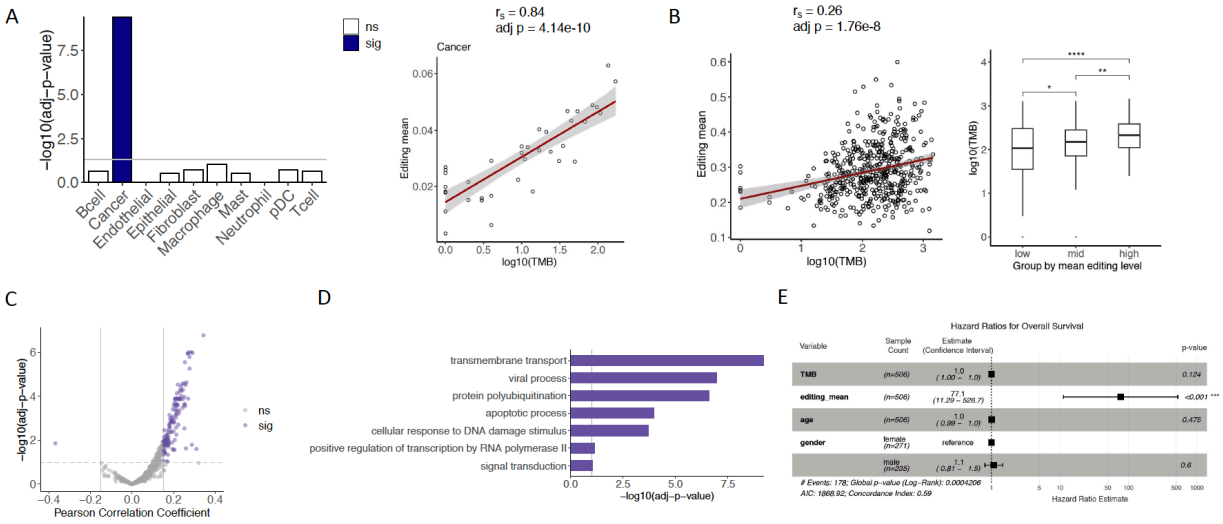
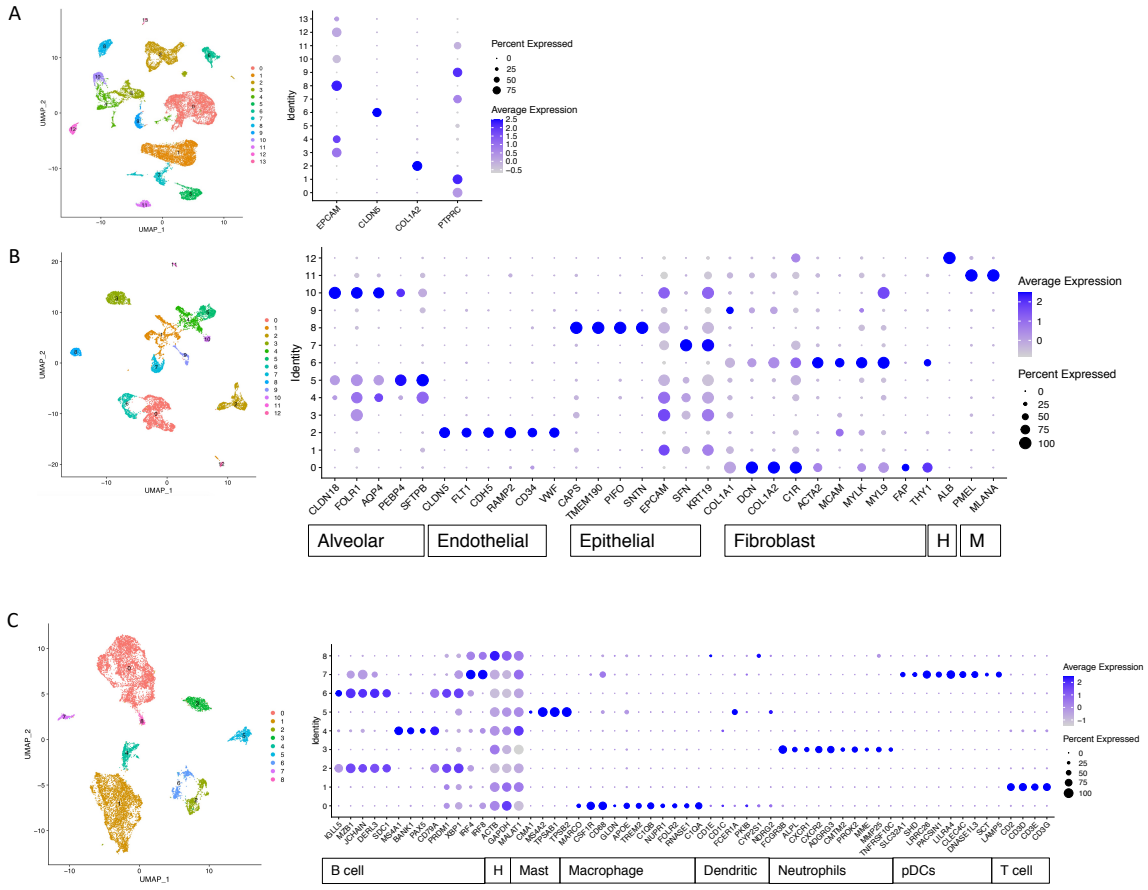


Figure 3.5 Relationship between RNA editing and tumor mutation burden specific to cancer cells.

A Left bar plot shows statistical significance of Spearman correlation between mean editing levels of each cell type and estimated TMB across tumors. For each tumor, single cells were pooled for each cell type, and mean editing level was calculated over

all sites in the pooled cells. Blue-filled bar indicates significant correlation by FDR 5% (sig). Correlations that were not significant are shown as white bars (ns). Scatterplot shows positive correlation between TMB and mean editing of pooled cancer cells, with labeled Spearman correlation coefficient and adjusted p-value. **B** Scatterplot of TMB and mean editing in bulk TCGA LUAD tumors. On the right, distributions of TMB across tumors grouped by mean editing level tertiles. Editing mean was calculated as the mean editing level of tumor-increased sites (with significantly higher editing levels in tumors than in matched normal samples). **C** Scatterplot of Pearson correlation coefficient and statistical significance of associations between gene editing levels and TMB in TCGA LUAD. Purple indicates significance by passing FDR 10% and minimum correlation coefficient of 0.15 (sig). Nonsignificant correlations are shown in gray (ns).
D Gene ontology terms enriched among genes with editing levels significantly correlated with TMB (from C), compared to background non-correlated genes with similar gene length and expression levels. Bar length represents statistical significance of enrichment, and vertical gray line indicates threshold of FDR 10%.
E Forest plot showing hazard ratio estimates and p-values from fitting a Cox regression model with TMB, editing mean, age, and gender as covariates. Editing mean was calculated as the mean editing level of tumor-increased sites (with significantly higher editing levels in tumors than in matched normal samples).

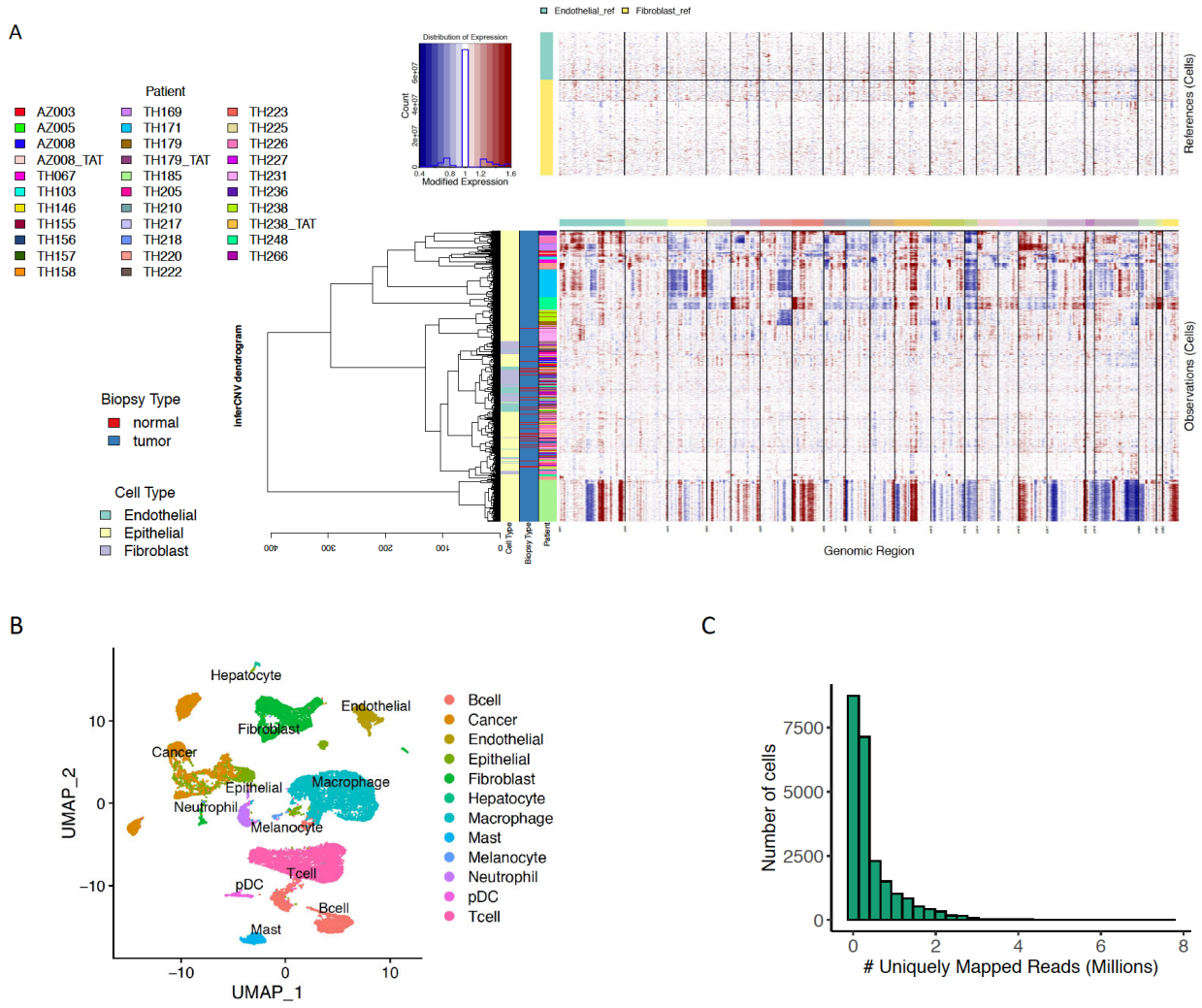
3.6 Supplementary Figures



Supplementary Figure 3.1 Clustering of single cells.

A UMAP projection of cells based on normalized gene expression, with cells colored by cluster identity (left). Right bubble plot shows cluster expression of broad compartment markers (epithelial, endothelial, immune, stromal), with darker blue indicating higher expression and point size representing percent of cells in a cluster that express a marker. **B** Sub-clustering of non-immune cells based on normalized gene expression. Similar to A, UMAP projection of non-immune cells is shown next to marker gene expression profiles of specific non-immune cell types. H and M stand for hepatocyte and

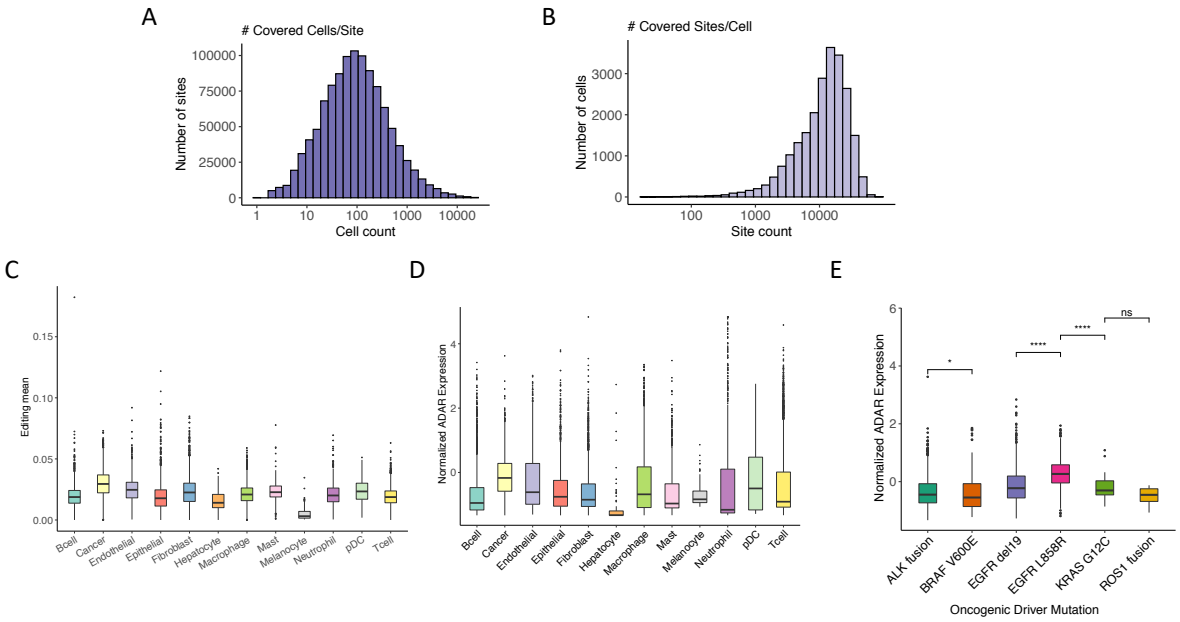
melanocyte, respectively. **C** Sub-clustering of immune cells based on normalized gene expression. Similar to B, UMAP projection of immune cells is shown next to marker gene expression profiles of specific immune cell types, as well as immune house-keeping genes (H).



Supplementary Figure 3.2 Labeling cancer cells.

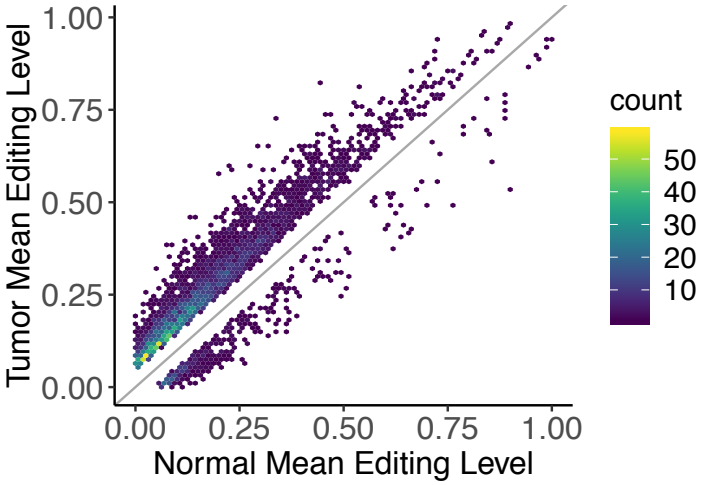
A Heatmaps of genome-wide CNV signals of epithelial cells and reference controls, fibroblasts and endothelial cells, estimated by inferCNV. Red regions indicate relative

amplifications, while blue regions represent deletions. The top heatmap includes only reference endothelial cells and fibroblasts. The heatmap below includes epithelial cells, as well as spike-in fibroblasts and endothelial cells. This heatmap Is labeled with cell type, biopsy type, and patient information for individual cells. Control spike-ins of fibroblasts and endothelial cells show low CNV estimates, as expected. **B** Final cell type assignment of single cells following cancer cell annotation based on inferCNV results. **C** Histogram of the number of uniquely mapped reads in single cells after removing PCR duplicates.



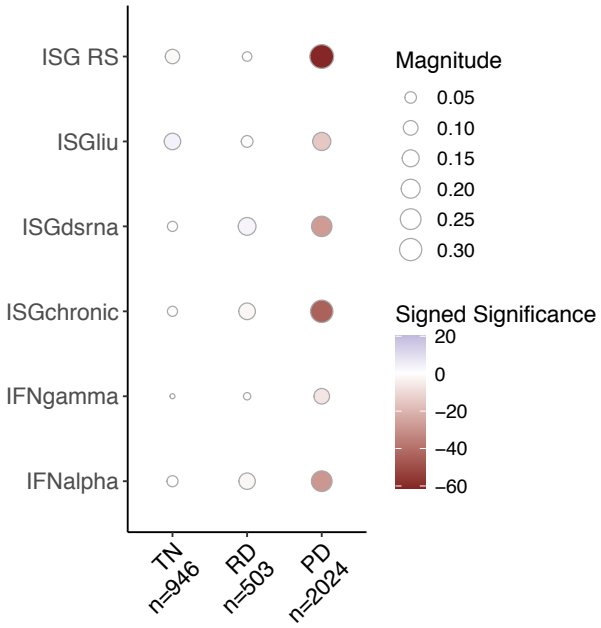
Supplementary Figure 3.3 Coverage, mean editing, and ADAR expression in single cells.

A Histogram of the number of single cells covered for each editing site, on a base 10 logarithm-transformed scale. A site was defined as covered if total coverage was at least five reads. **B** Histogram of the number of sites covered (with at least five reads) in each cell, on a base 10 logarithm-transformed scale. **C** Distributions of overall editing levels in single cells across cell types. Overall editing level was calculated as the mean over editing levels of all sites for each cell. By Mann Whitney U test, p <= 0.0001 for all pairwise comparisons of cancer vs another cell type. **D** Distributions of sctransform-normalized ADAR expression in single cells across cell types. By Mann Whitney U test, p <= 0.0001 for all pairwise comparisons of cancer vs another cell type. **E** Normalized ADAR expression levels were compared among driver mutation groups by Mann Whitney U test, with p value significance shown. *p <= 0.05. ****p <= 0.0001. ns p>0.05.



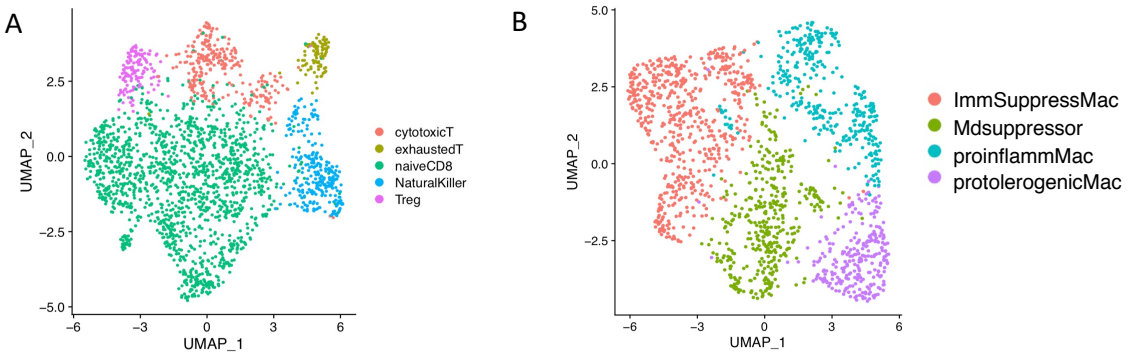
Supplementary Figure 3.4 Hyperediting in bulk LUAD tumors.

2-dimensional histogram of mean editing level over tumors and mean editing level over normal samples for 4380 differential editing sites. Significance of differential editing was determined by REDIT regression with FDR 5% and a minimum difference in mean editing levels of 5%.



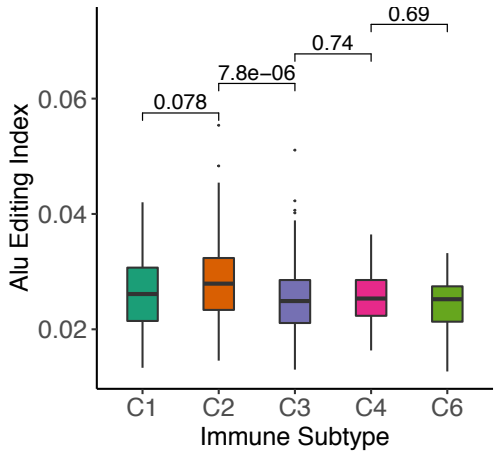
Supplementary Figure 3.5 Editing association with ISG signatures by treatment timepoint.

Spearman correlations between mean editing levels and mean expression of ISG signatures across single cancer cells grouped by treatment timepoint. The number of cancer cells within each category is labeled. The size of each circle indicates the magnitude of the Spearman correlation coefficient, and the color intensity corresponds to significance of the adjusted p-value. Blue represents positive correlations and red, negative ones. TN stands for treatment naïve, RD stands for residual disease, and PD stands for progression.



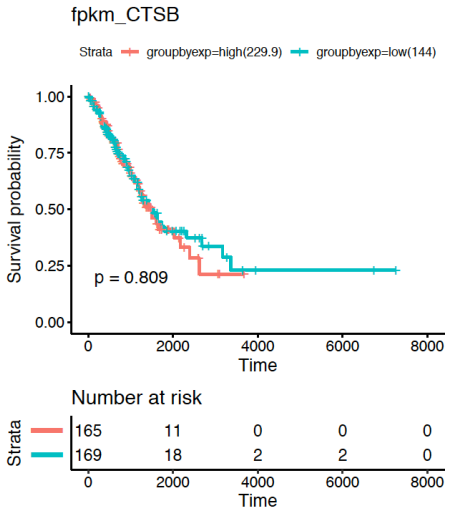
Supplementary Figure 3.6 Clustering of T cells and Macrophages.

A UMAP projection of T cells cells based on normalized gene expression profiles. Cells are colored by subtype of T cells determined by marker gene expression. **B** UMAP projection of macrophages, similar to A. Color indicates subtype of Macrophages based on marker gene expression.



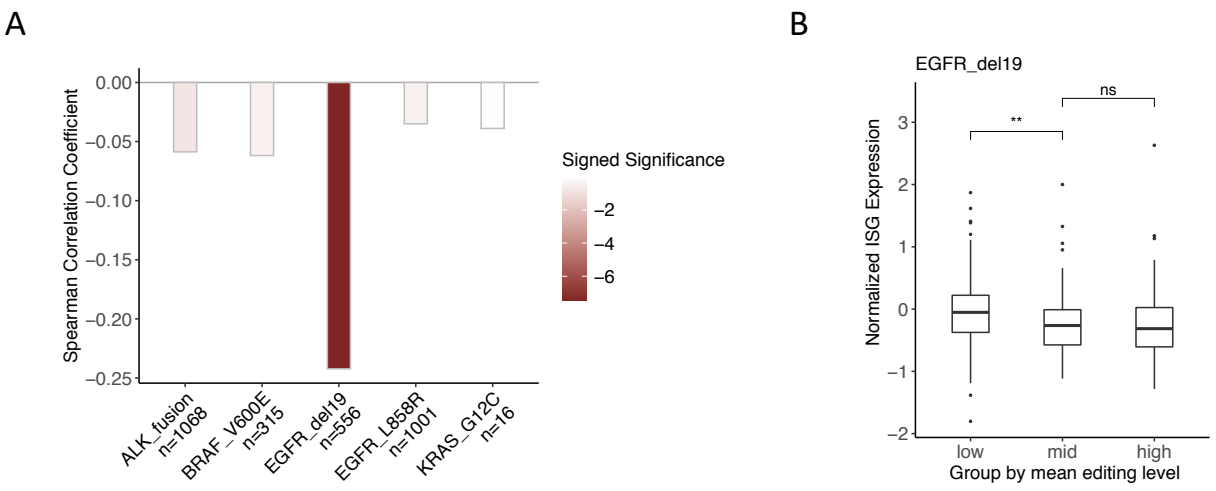
Supplementary Figure 3.7 Association of editing with immune subtype in TCGA LUAD tumors.

Alu editing index was compared between pairs of immune subtypes by Mann Whitney U test, with p-values labeled.



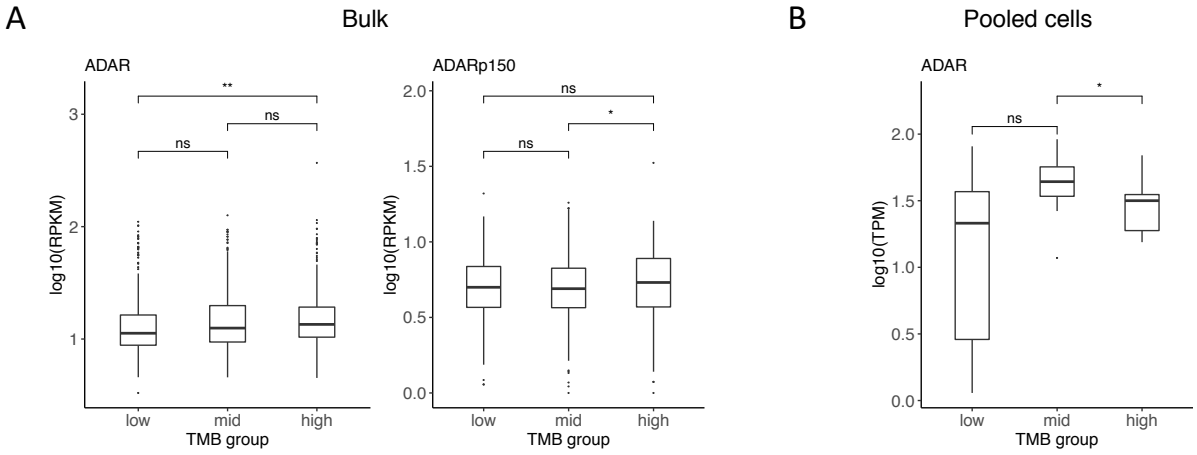
Supplementary Figure 3.8 CTSB expression not associated with survival.

Kaplan Meier plot showing similar survival for patients with relatively high and low CTSB expression. Patients in the high group (red curve) had CTSB FPKM among the top third across all patients, and patients in the low group (blue curve) had CTSB FPKM levels among the lowest third of all patients.



Supplementary Figure 3.9 Association of CTSB dsRNA editing and ISG expression is dependent on driver mutation in cancer cells.

A Barplot of Spearman correlation coefficient for associations between CTSB dsRNA editing levels and mean ISG expression levels across single cancer cells grouped by driver mutation. Color intensity represents statistical significance (-log10(FDR-adjusted p-value)) and red color indicates negative correlation. **B** Distributions of mean ISG expression are shown for cancer cells with EGFR del19 driver mutation. Cells were grouped by tertile editing levels of the CTSB dsRNA. ISG expression was compared across dsRNA editing groups by Mann Whitney U test, with p value significance labeled. **p <= 0.01. ns p > 0.05.



Supplementary Figure 3.10 ADAR Expression and TMB not strongly associated

A Distributions of ADAR and ADAR p150 expression in TCGA LUAD tumors, grouped by TMB tertiles. ADAR p150 expression was estimated based on the single exon unique to the ADAR p150 (not included in the ADAR p110 isoform). Expression was compared between groups by Mann Whitney U test, with significance of p-values labeled. **p <= 0.01. *p<=0.05. ns p > 0.05. **B** Distributions of ADAR expression in pooled cancer cells, grouped by TMB tertiles. For each tumor, single cancer cells were pooled, and ADAR TPM was calculated. Expression was compared between groups by Mann Whitney U test, with significance of p-values labeled. *p<=0.05. ns p > 0.05.

Chapter 4: Candidate RNA editing events relevant to immunotherapy response in melanoma

4.1 Introduction

In Chapter 3, our study of RNA editing in single cells of lung cancer demonstrated the significance of editing to tumor immunity and overall survival, with potential application to immune checkpoint blockade (ICB) treatment. Implying an immunosuppressive role in tumors, editing in cancer cells was negatively associated with cancer interferon-stimulated gene (ISG) expression and natural killer cell infiltration. Furthermore, the strongest negative correlation between editing and ISGs was observed in cancer cells with the driver mutation BRAF V600E, which was associated with higher rates of ICB response than several other driver mutations in NSCLC[167]. The BRAF V600E mutation was also previously observed to correlate with PD-1 treatment response in melanoma patients[170]. Thus, determining specific editing signatures that are most relevant to ICB response is warranted.

Various biomarkers have been considered so far in attempts to identify patients who would respond to ICB therapies[32,33]. For treatments inhibiting the interaction between PD-L1 and its receptor PD-1, the predictive value of PD-L1 expression has been assessed. Despite the direct role of PD-L1 as a target, certain patients with tumors considered to be PD-L1-negative may still respond to treatment[32]. Another biomarker is

tumor mutation burden (TMB), which is thought to correlate with immunogenic neoantigen load. Still, TMB does not appear to robustly predict ICB response either, especially for cancer types with generally low mutation loads[33,155]. Other proposed biomarkers that have not yet been FDA-approved include quantification of tumor lymphocyte infiltration or expression of gene signatures, such as IFN-γ signaling and T cell exhaustion[33,169,171,172]. Given variable performance of current biomarkers and the lack of standardized thresholds, discovery and integration of novel biomarkers are needed to realize effective stratification of responsive and treatment-resistant patients.

Labeling endogenous dsRNAs, RNA editing mediates the role of ADAR as an immune checkpoint in cancer. Upon loss of ADAR and depletion of most RNA editing events, unmodified dsRNAs likely activate cytosolic dsRNA sensors, resulting in excessive IFN production and translational inhibition in certain cancer cells[10,35,36]. These ADAR-dependent innate immune signaling pathways were shown to promote response to ICB and DNA methyltransferase inhibitor (DNMTi) therapies in mouse models[34,125]. However, since loss of ADAR1 is poorly tolerated *in vivo*, ADAR1 inhibition therapy faces significant challenges in toxicity and side effects[173]. Thus, deciphering the specific dsRNA targets of ADAR1 in cancer could provide critical insights in both therapeutic design and the mechanistic basis of this pathway.

In this Chapter, we present the first effort, to our knowledge, in search of dsRNA editing events relevant to ICB response in cancer. Specifically, we leverage the existence of multiple RNA-seq data sets derived from melanoma patients undergoing

ICB treatment. In the future, a similar approach could be applied once more data sets become available, followed by experimental characterizations of the predicted dsRNA substrates.

4.2 Results

4.2.1 Unedited dsRNA level is associated with ICB outcomes

ADAR loss caused cell lethality and promoted response to ICB in tumor models[34–36]. As dsRNA sensor activation and IFN upregulation mediated this ADAR dependency, we hypothesized that a metric of RNA editing or its absence on dsRNAs could be a biomarker for ICB response. To test this hypothesis, we used four datasets of melanoma patients with recorded response to ICB and pre-treatment RNA-seq data (Table S1). For each RNA-seq sample, we counted the number of dsRNAs with a minimum unedited expression level (Methods). We found that higher abundance of unedited dsRNAs was correlated with longer overall survival in the MGSP and TCGA SKCM patients (Fig. 1A). Furthermore, ICB responders in the MGSP dataset had larger numbers of unedited dsRNAs (Fig. 1B). Consistent with the model that higher levels of unedited dsRNAs induce IFNs and cancer cell growth arrest, these findings support the potential use of unedited dsRNA level as a biomarker for patient response to ICB.

4.2.2 **Clinical relevance of CTSB dsRNA**

Next, we investigated which dsRNAs are the most relevant to survival and response to ICB. To do so, we examined dsRNAs that are most often represented in the unedited dsRNA count. A dsRNA in the 3'UTR of CTSB was most frequently counted as unedited in MGSP patients and was also among the top dsRNAs in TCGA SKCM. Based on the secondary structure predicted by RNAfold[177], this region likely forms long stem structures, which are the preferred targets of MDA5[9,178] (Fig. S1A). In addition, the CTSB dsRNA contains multiple MDA5 eCLIP-peaks identified in HeLa cells, suggesting MDA5 binds to this dsRNA (Fig. S1B). Editing events in the CTSB 3'UTR may disrupt the dsRNA structure and prevent MDA5 activation.

Given the prevalence and potential MDA5 recognition of the CTSB dsRNA, we tested its prognostic value in the context of ICB treatment. In patients from both MGSP and TCGA SKCM datasets, higher unedited expression of the CTSB dsRNA was associated with longer survival (Fig. 2A). Higher gene expression of CTSB was also correlated with improved survival in these two datasets (Fig. 2B). Interestingly, the gene expression association appeared limited to patients treated with immunotherapy, as no significance was seen in SKCM patients who received other types of treatment (Fig. 2C). Indeed, in a previous study where all TCGA cancer types, including SKCM, were analyzed collectively, higher CTSB expression was reported to correlate with worse survival[149], which is opposite to the survival correlation observed here for ICB-treated melanoma patients. These data indicate that the involvement of CTSB in cancer

processes may be multifaceted: while the function of CTSB as a cysteine cathepsin may be oncogenic[179], the presence of unedited dsRNAs from the CTSB 3'UTR may activate innate immune signaling and improve clinical outcomes following ICB treatment in melanoma patients.

4.2.3 **Differentially edited Alus in response to ICB**

As Alus are prominent editing substrates in humans, we next considered Alus in our search of editing events associated with response to ICB. For this analysis, we first calculated the editing index for each individual Alu in each tumor sample. We then identified Alus with significantly different mean editing levels between responders (R) and non-responders (NR) to ICB in each dataset separately. We used two criteria to define significance: difference in mean editing index ≥ 0.05 and Mann Whitney U test p-value < 0.05. We found four differentially edited Alus that were shared by at least two datasets, with the same direction of editing difference. For most of these shared Alus, lower editing levels were observed in patients who responded to ICB (Fig. 3), which is consistent with the hypothesis that RNA editing deficiency induces IFNs and promotes ICB response.

Examining the neighboring region of these differential Alus revealed close proximity of inverted Alus in the same transcripts (Fig. S2A). These pairs of inverted Alu repeats likely form dsRNA structures that when unmodified may potentially activate an innate immune response[6]. Accordingly, regions containing these inverted Alu pairs were

predicted to form dsRNAs by RNAfold (Fig. S2B). Furthermore, for three differentially edited Alus, MDA5 eCLIP peaks overlapped the surrounding regions containing the inverted repeats (Fig. S2A). Thus, these Alu pairs may form dsRNAs that are substrates for MDA5 binding and activation.

Interestingly, shared differentially edited Alus were within genes involved in immune response. For example, the gene IFNGR2 encodes a component of the receptor for IFN-γ, which has multifaceted roles in tumor response to ICB[141,180]. In addition, PILRB is expressed on the surface of various immune cell types and was found to affect cytokine production in macrophages[181,182]. Altered editing levels within the dsRNA-forming transcripts of these genes may add another layer of regulation to anti-tumor immunity during ICB treatment.

4.3 Discussion

We described the first pre-treatment collection of editing events in dsRNAs associated with clinical outcomes of melanoma patients treated with ICB. In addition, we provided evidence that MDA5, a key sensor of long dsRNAs in innate immunity[9] and mediator of ADAR dependence in a mouse model for melanoma[34], likely targets these dsRNAs. As additional clinical and genomic data become available, the predictive power of these potential dsRNA biomarkers in ICB response may be further assessed.

Recent studies have reported that ADAR gene expression and general measures of RNA editing, namely the number of total editing events or Alu editing index, were not different between ICB responders[183] and non-responders or between samples before treatment and after progression[184]. These findings highlight the need to determine a refined set of editing events most relevant to ICB response. Comparing cell lines from patients before treatment to those after relapse, a previous study identified an Alu editing signature specific to relapsed tumors[184], but the possibility that editing differences were a consequence of treatment was not addressed. Here, we focused on editing features in pre-treatment tumors with the potential of future biomarkers for ICB response.

One of the most prominent unedited dsRNAs that we found in multiple melanoma datasets was the dsRNA in the CTSB 3'UTR, which was also the top dsRNA with editing levels negatively correlated with ISG expression in NSCLC from Chapter 3. The association of higher unedited expression level of this CTSB dsRNA with longer survival supports the hypothesized function of unedited dsRNAs to induce innate immune signaling pathways that would enhance tumor sensitivity to ICB treatment. Consistently, we had observed that lower editing levels of this dsRNA were correlated with improved survival in lung cancer patients. Together, these data suggest that editing of the CTSB dsRNA is an important factor contributing to ISG signatures and tumor growth. Though CTSB has been reported to enhance the metastatic potential of cancer cells[148,179], the activity of unedited dsRNAs from CTSB transcripts may establish a dominant anti-tumor role for CTSB during ICB treatment. This mechanism of CTSB dsRNA editing in

promoting patient survival and response to ICB will need to be further investigated in future studies.

Dependent on the source cell type, IFN-γ signaling has been associated with both response and resistance to ICB[141,172,185]. While IFNs trigger MHC-1 expression and function in the maturation of dendritic cells and T cells[180], IFN-γ signaling in cancer cells induces PD-L1 expression and inhibits cytotoxicity of T cells and Natural Killer cells[141]. Among the four Alus that were differentially edited between ICB responders and non-responders in multiple datasets, only the Alu in the IFNGR2 gene, a critical receptor of the IFN-γ pathway, had higher editing levels in responders. This association may suggest the existence of other roles of altered editing in IFNGR2 to support ICB response, which needs future exploration. For the remaining Alus, editing levels were significantly decreased in responders, indicating that they likely influence IFN activation and ICB response in a similar manner to the CTSB dsRNA.

In summary, this study demonstrates the potential value of dsRNA- and editing-based biomarkers to predict patient response to immunotherapy in melanoma. In the future, additional patient cohorts need to be analyzed to validate and extend these findings. This type of analysis may also lead to improved understanding of the fundamental pathways underlying cancer immunity and immunotherapy response.

4.4 Methods

4.4.1 **RNA-seq data preprocessing and identification of RNA editing sites**

We downloaded RNA-seq fastq files from four sources: 1) Gene Expression Omnibus (GEO) accession GSE78220[174], 2) GEO accession GSE91061[175], 3) dbGaP accession phs000452.v3.p1[176], and 4) GDC Legacy Archive for TCGA Skin Cutaneous Melanoma (SKCM) tumors[49]. We aligned reads to hg19 using HISAT2[131] with default parameters except that soft-clipping was disabled (--no-softclip). To obtain gene expression counts for calculation of Fragments Per Kilobase Million (FPKM), we ran HTSeq count.

To consider regions of hyperediting, we input unmapped reads into a hyperediting pipeline[52,53]. For each sample, the resulting rescued reads were combined with the original uniquely mapped reads. We next used our published methods to detect editing events at sites documented in the REDIportal database[54,55]. As carried out in a previous study, we then filtered out variants listed in dbSNP version 147 and COSMIC version 81, except for editing sites relevant to cancer[120].

4.4.2 **Differential editing of Alus**

We used the UCSC RepeatMasker to determine which editing sites overlapped individual Alu repeats. An editing index was then calculated for each Alu in each tumor

as the sum of edited reads across all sites within the Alu divided by the sum of all reads covering sites in the Alu. The difference in editing index between responders and non-responders for each Alu in one melanoma dataset was tested by Mann Whitney U test. Alu elements with p-value < 0.05 and a difference in mean editing levels ≥ 0.05 were considered significantly different in ICB response.

4.4.3 **Measuring the level of unedited dsRNAs**

We predicted dsRNA structures based on editing-enriched regions, as previously described[147]. For an individual dsRNA located in a 3'UTR, we quantified expression level (as FPKM) and mean editing level. Mean editing level of a dsRNA was calculated as the mean of editing levels of all sites within the dsRNA. We next obtained the unedited level of the dsRNA by subtracting mean editing level from 1. Then, we calculated unedited expression level of the dsRNA as the product of the dsRNA FPKM and unedited level. We considered a dsRNA as unedited if its unedited expression level was at least 20. The number of unedited dsRNAs was counted for each tumor sample.

4.4.4 **MDA5 eCLIP data**

eCLIP experiments of MDA5 in HeLa cells were carried out in-house (Bahn JH, unpublished data) using a similar protocol as reported for ADAR1 in our previous work[73]. Two biological replicates and a size-matched control were obtained, with about

20M pairs of reads obtained for each replicate. Read mapping and eCLIP peak calling were carried out in the same way as previously described[73].

4.4.5 Analysis of the relevance of RNA editing to survival and immunotherapy response

We first used the survfit function of the R package survival to obtain Kaplan-Meier survival curves for high and low groups of unedited dsRNA level. To define high and low groups, the tertiles of unedited dsRNA level were first calculated for each melanoma dataset. In one dataset, patients with unedited dsRNA counts within the top third across all patients were assigned to the high group, while patients in the low group had unedited dsRNA counts among the bottom third of all patients. The survdiff function was then used to compare high and low group survival curves by the log rank test. Survival curves were visualized by the ggsurvplot function of the R package survminer. Similarly, the same functions in R were used to test the association between survival and CTSB dsRNA unedited expression level or CTSB expression.

To test the association between ICB response and the number of unedited dsRNAs in one melanoma cohort, we fit a linear regression model of unedited dsRNA count on patient response (responder or non-responder) using the function lm in R.

We used bedtools closest to find inverted Alu repeats or MDA5 eCLIP peaks near differentially edited Alus and the CTSB dsRNA.

4.5 Figures

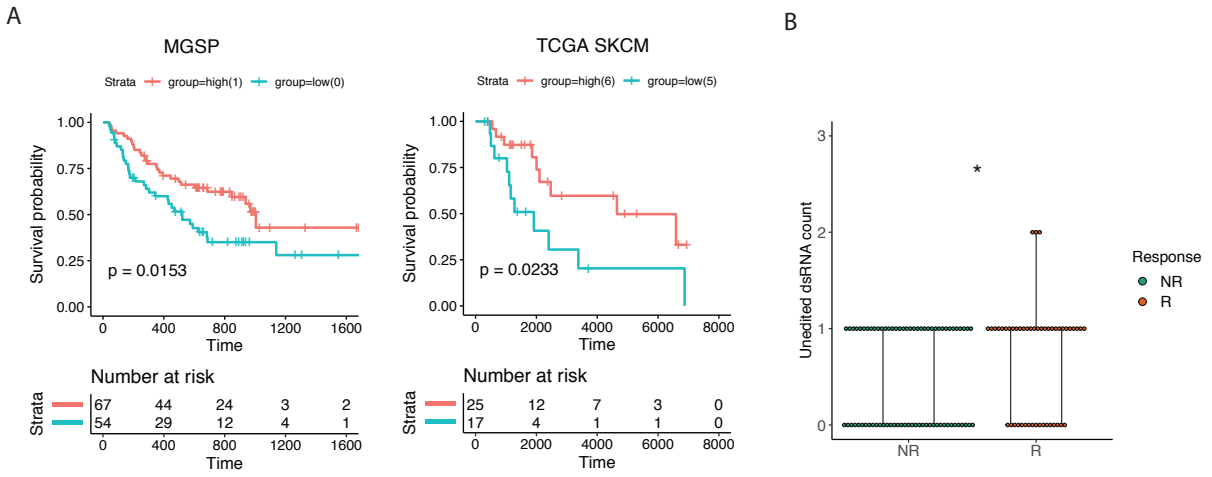


Figure 4.1 Number of unedited dsRNAs is associated with survival after immunotherapy

A Kaplan-Meier plots showing better survival for patients with higher numbers of unedited dsRNAs present in two datasets (labeled on top). The number of unedited dsRNAs in a sample was counted as the number of 3'UTR-overlapping dsRNAs that each had an unedited expression level above a cutoff of 20. For each dataset, patients of high (red curve) and low (blue curve) groups were determined by the tertiles of counts of unedited dsRNAs across all patients in the dataset. Log-rank p-value between high and low groups is labeled on each plot. **B** Distributions of unedited dsRNA counts for responders (R, orange) and non-responders (NR, green) from the MGSP dataset. Unedited dsRNAs were counted in the same way as in A. Association between response and unedited dsRNA count was tested by linear association. *p<0.05.

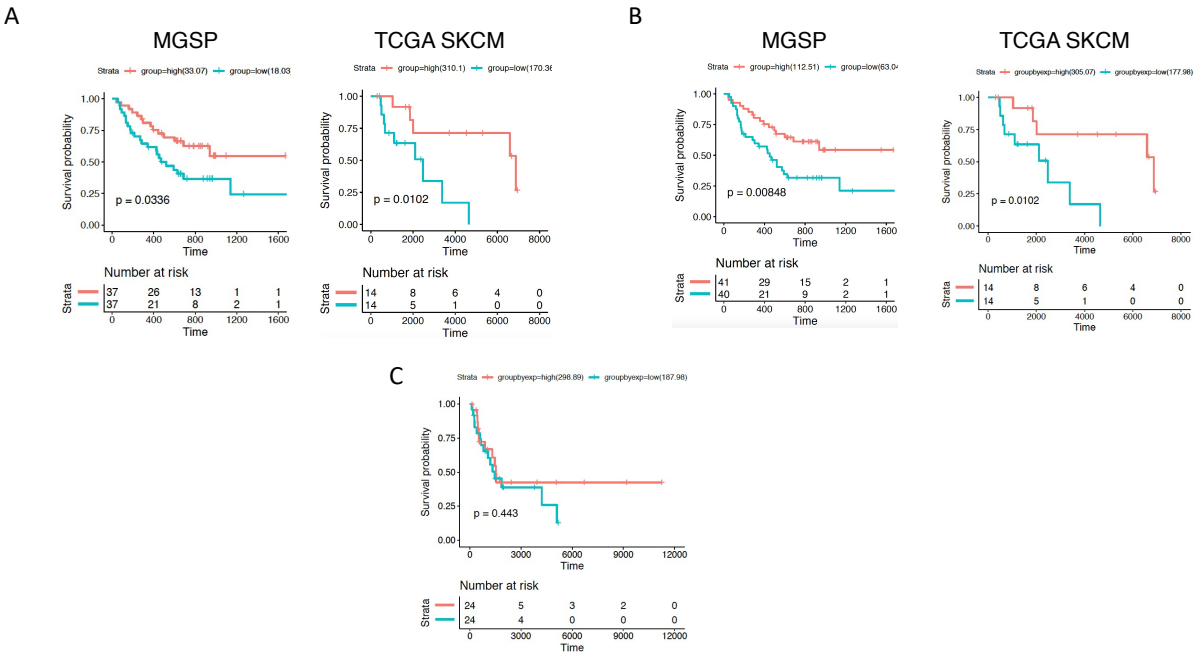


Figure 4.2 Survival association of CTSB dsRNA.

A Kaplan-Meier plots showing better survival for patients with higher unedited expression level of the CTSB dsRNA in two datasets (labeled on top). Patients in the high group (red curve) had unedited expression levels among the top third across all testable patients. Patients in the low group (blue curve) had unedited expression levels among the lowest third. Patients with recorded response to immunotherapy were included for each dataset. Log-rank p-value between high and low groups is labeled on each plot. **B** Similar to A except patients are grouped here by CTSB expression (FPKM) tertile levels. Patients with high CTSB expression (red) have longer survival than those with low expression (blue). **C** Kaplan-Meier plot showing CTSB expression is not associated with overall survival in patients not treated with immunotherapy in TCGA SKCM. Patients were spilt by tertiles of CTSB FPKM levels, with high group (red) indicating top third of all patients by CTSB expression and low group (blue) indicating bottom third.

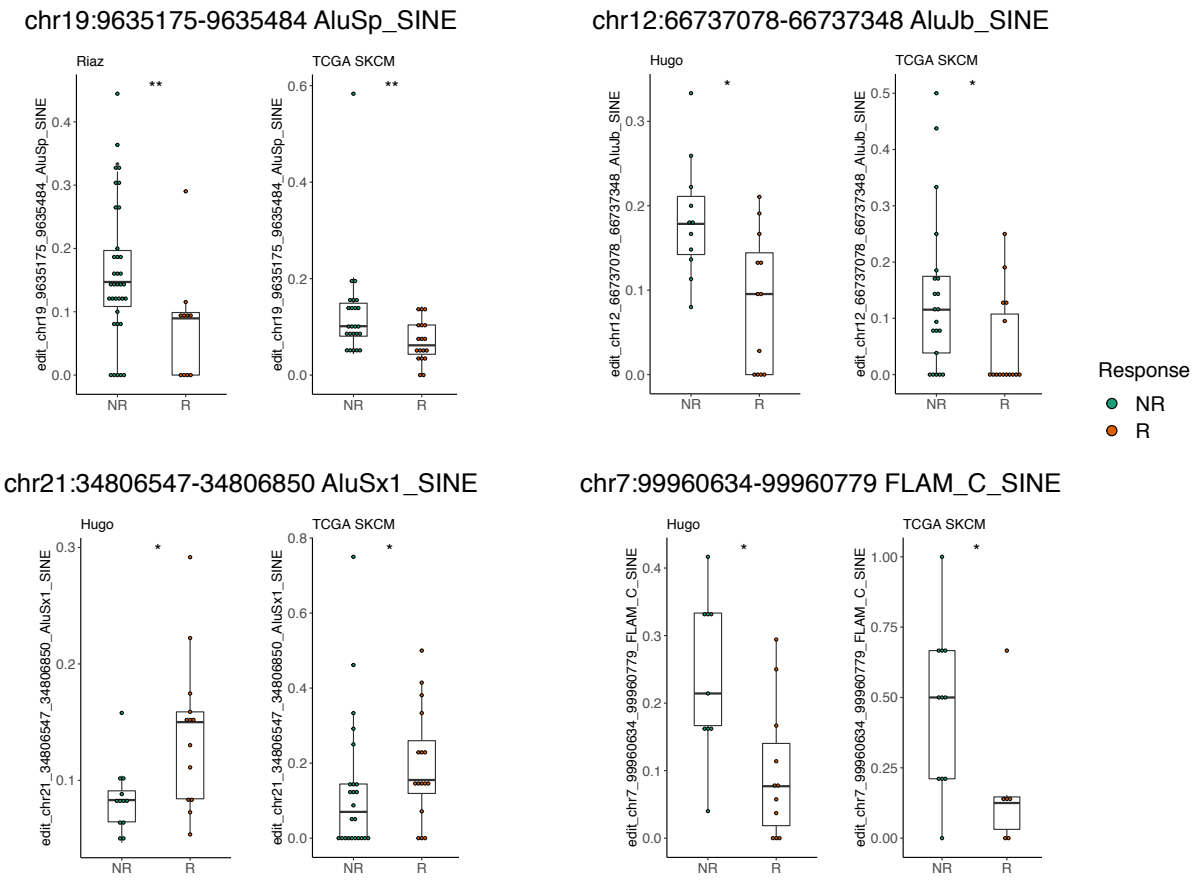
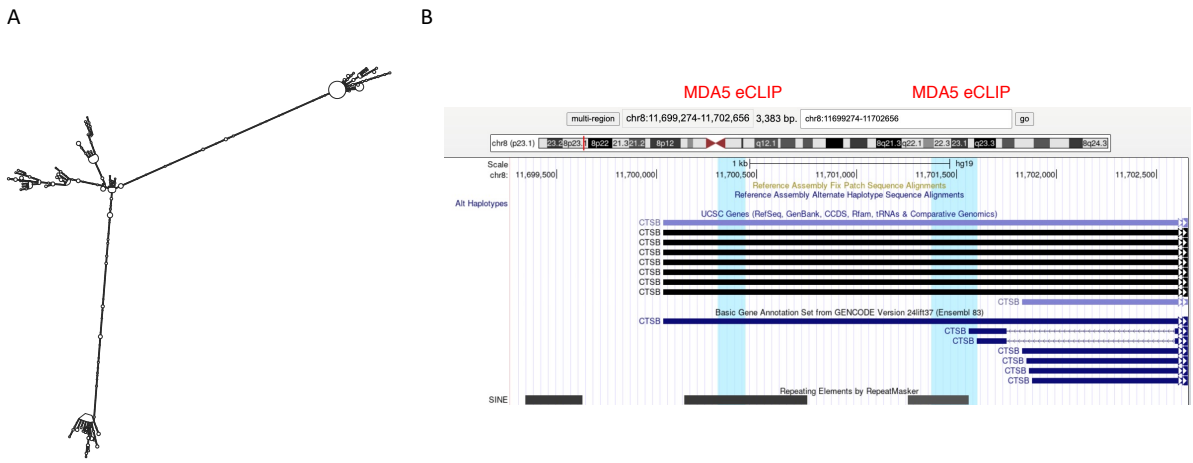


Figure 4.3 Differentially edited Alus in immunotherapy response.

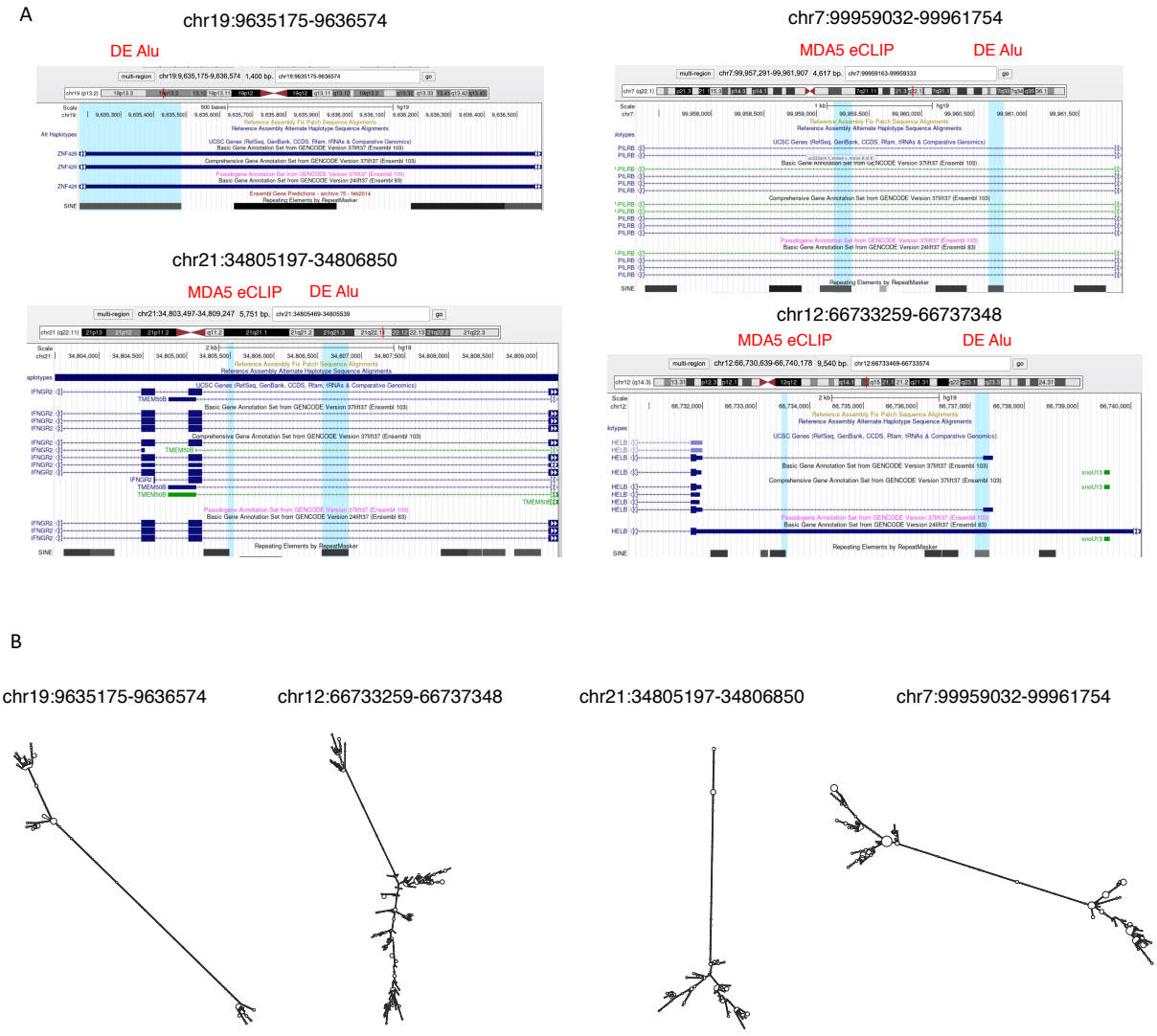
Four Alus with significantly different mean editing levels between responders (R, orange) and non-responders (NR, green) to immune checkpoint blockade. Each Alu shown was differentially edited in two datasets, with the same direction of editing difference between datasets. Editing levels were compared between responders and non-responders by Mann Whitney U test, with p-value significance shown. *p <= 0.05. **p <= 0.01.

4.6 Supplementary Figures



Supplementary Figure 4.1 CTSB dsRNA is a potential MDA5 target.

A Predicted secondary structure of CTSB dsRNA by RNAfold. **B** CTSB dsRNA region in UCSC genome browser, in which blue highlighted regions correspond to MDA5 eCLIP peaks. Bottom track shows Alu (SINE) repeats present.



Supplementary Figure 4.2 Potential dsRNAs containing differentially edited Alus

A Regions containing Alus differentially edited in ICB response shown in UCSC browser. Blue highlighted regions correspond to MDA5 eCLIP peaks or differentially edited Alus (DE Alu), as labeled on top. Bottom SINE track shows inverted pairs of Alus present in these regions. **B** Predicted secondary structure of extended regions containing differentially edited Alus and their inverted repeats by RNAfold.

4.7 Supplementary Tables

Dataset source	Treatment	# Patients with measured response	# RNA-seq samples analyzed
Hugo et al, Cell, 2016	anti-PD1	38	26
Liu et al, Nat Med, 2019 (MGSP)	anti-PD1	144	110
Riaz et al, Cell, 2017	anti-PD1	73	51
TCGA SKCM	anti-CTL4 and other	43	43

Table S 4.1 Melanoma datasets analyzed in this study.

RNA-seq samples from pre-treatment tumors only were analyzed.

Chapter 5: Concluding Remarks

Similar to somatic mutations, RNA editing events can drive the uncontrolled growth and invasion of cancer cells. This transformative potential has been demonstrated for a small number of editing sites. Functions of the majority of tumor-associated editing sites, primarily in noncoding regions, are not well understood. In this work, we carried out the first global investigations of RNA editing in EMT phenotypes and in single cells of cancer. We also demonstrated the role of editing to suppress innate immunity in lung tumors and highlighted the first compilation of candidate dsRNA editing events with potential predictive value in ICB response of melanoma patients.

In Chapter 2, we identified RNA editing differences between epithelial and mesenchymal tumors of various cancer types in TCGA. We found a consistent pattern of altered editing in lung cancer scRNA-seq data. Furthermore, we showed that EMT was induced upon knockdown of either ADAR1 or ADAR2. Enriched in immune response pathways, differential editing may regulate the mRNA abundance of these genes. Following a systematic assessment of RBP binding with ENCODE eCLIP data, we described a novel mechanism in which ILF3 mediates the effects of differential editing on mRNA expression of immune genes, such as PKR.

In Chapter 3, we examined RNA editing events detectable in single cells of lung cancer biopsies and compared editing profiles of tumors to those of normal samples in a cell-type-specific manner. These analyses indicated that tumor-increased editing was

unique to cancer cells and that certain pathways are differentially edited in specific cell types only. Additionally, we found that cancer-specific editing was negatively associated with cancer ISG expression and Natural Killer cell infiltration, extending previous findings of immunosuppressive editing to NSCLC tumors. This role of editing may partly explain the association we observed between higher editing levels and worse survival in NSCLC.

In Chapter 4, we reported dsRNA editing events associated with survival and response of melanoma patients treated with ICB. We showed that these dsRNAs of interest are likely MDA5 substrates, suggesting potential for MDA5 activation by their unedited forms.

Our analyses provided deeper insights into the various functions of RNA editing in cancer, especially regarding tumor immunity. From altered mRNA abundance of immune response genes to suppressed IFN signaling, editing facilitates multiple feedback mechanisms to promote or resist anti-tumor immune responses. A more complete understanding of these roles of editing will support development of novel avenues for cancer treatments, as well as the refinement of existing ones, like ICB combination therapies. During ICB response, higher abundance of unedited dsRNAs may induce IFNs and in effect, increase recruitment of immune cells to tumors and support tumor cell killing. Future studies should validate their clinical relevance and determine the precise functional mechanisms of dsRNA editing in ICB response. It is not yet clear how MDA5, PKR, and other dsRNA sensors distinguish edited dsRNAs

from unedited dsRNAs, though altered dsRNA structures are hypothesized. In the future, RNA editing may also be integrated with other biomarkers to more accurately predict ICB response and avoid unwarranted risk of immune-related adverse events for likely non-responsive patients.

Our findings also motivate further investigations of tumor-associated editing changes in single cells and in individual cell types. We used a pseudo-bulk approach to determine editing differences between tumor and normal within each cell type in Chapter 3. For more precise identification of differential editing across single cells, novel methods should be developed to account for variation among single cells and hierarchical structure. In Chapter 3, we also observed that certain immune pathways were uniquely enriched in T cell differentially edited genes. Previous literature has reported involvement of ADAR and RNA editing in T cell development[8,28], but cancer-specific functions of editing in T cells are unknown. Whether these differential editing events affect T cell phenotypes of exhaustion or cytotoxicity should be assessed in future studies.

References

1. Gott, J. M. & Emeson, R. B. Functions and mechanisms of RNA editing. *Annual Review of Genetics* **34**, 499–531 (2000).
2. Eisenberg, E. & Levanon, E. Y. A-to-I RNA editing - Immune protector and transcriptome diversifier. *Nature Reviews Genetics* **19**, 473–490 (2018).
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. Scadden, A. D. J. & Smith, C. W. J. Specific cleavage of hyper-edited dsRNAs. *EMBO J.* **20**, 4243–4252 (2001).
5. 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).
6. Schaffer, A. A. & Levanon, E. Y. Alu a-to-i RNA editing: Millions of sites and many open questions. in *Methods in Molecular Biology* **2181**, 149–162 (Humana Press Inc., 2021).
7. Ahmad, S., Mu, X. & Hur, S. The role of rna editing in the immune response. in *Methods in Molecular Biology* **2181**, 287–307 (Humana Press Inc., 2021).
8. Nakahama, T. & Kawahara, Y. Adenosine-to-inosine RNA editing in the immune system: friend or foe? *Cellular and Molecular Life Sciences* **77**, 2931–2948 (2020).
9. Liddicoat, B. J. *et al.* RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science (80-.).* **349**, 1115–1120 (2015).
10. Chung, H. *et al.* Human ADAR1 Prevents Endogenous RNA from Triggering

Translational Shutdown. *Cell* **172**, 811-824.e14 (2018).

11. Liddicoat, B. J., Chalk, A. M. & Walkley, C. R. ADAR1, inosine and the immune sensing system: Distinguishing self from non-self. *Wiley Interdisciplinary Reviews: RNA* **7**, 157–172 (2016).
12. Mannion, N. M. *et al.* The RNA-Editing Enzyme ADAR1 Controls Innate Immune Responses to RNA. *Cell Rep.* **9**, 1482–1494 (2014).
13. Rice, G. I. *et al.* Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type i interferon signature. *Nat. Genet.* **44**, 1243–1248 (2012).
14. Frassinelli, L., Galardi, S., Ciafrè, S. A. & Michienzi, A. RNA editing in interferonopathies. in *Methods in Molecular Biology* **2181**, 269–286 (Humana, New York, NY, 2021).
15. Guo, X. *et al.* Aicardi-Goutières syndrome-associated mutation at ADAR1 gene locus activates innate immune response in mouse brain. *J. Neuroinflammation* **18**, 1–16 (2021).
16. 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).
17. 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).
18. Kurkowiak, M. *et al.* The effects of RNA editing in cancer tissue at different stages in carcinogenesis. *RNA Biology* (2021). doi:10.1080/15476286.2021.1877024
19. Nakano, M., Fukami, T., Gotoh, S. & Nakajima, M. A-to-I RNA editing up-regulates human dihydrofolate reductase in breast cancer. *J. Biol. Chem.* **292**,

4873–4884 (2017).

20. Chen, L. *et al.* Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. *Nat. Med.* **19**, 209–216 (2013).
21. Hu, X. *et al.* RNA editing of AZIN1 induces the malignant progression of non-small-cell lung cancers. *Tumor Biol.* **39**, (2017).
22. Wang, Y. *et al.* Systematic characterization of A-to-I RNA editing hotspots in microRNAs across human cancers. *Genome Res.* **27**, 1112–1125 (2017).
23. Gumireddy, K. *et al.* The mRNA-edited form of GABRA3 suppresses GABRA3-mediated Akt activation and breast cancer metastasis. *Nat. Commun.* **7**, 10715 (2016).
24. Chan, T. H. M. *et al.* ADAR-Mediated RNA Editing Predicts Progression and Prognosis of Gastric Cancer. *Gastroenterology* **151**, 637–650 (2016).
25. 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).
26. Paz-Yaacov, N. *et al.* Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors. *Cell Rep.* **13**, 267–276 (2015).
27. Fumagalli, D. *et al.* Principles Governing A-to-I RNA Editing in the Breast Cancer Transcriptome. *Cell Rep.* **13**, 277–289 (2015).
28. Danan-Gotthold, M., Guyon, C., Giraud, M., Levanon, E. Y. & Abramson, J. Extensive RNA editing and splicing increase immune self-representation diversity in medullary thymic epithelial cells. *Genome Biol.* **17**, 1–13 (2016).
29. Zhang, M. *et al.* RNA editing derived epitopes function as cancer antigens to elicit immune responses. *Nat. Commun.* **9**, 3919 (2018).

30. Peng, X. *et al.* A-to-I RNA Editing Contributes to Proteomic Diversity in Cancer. *Cancer Cell* **33**, 1–12 (2018).

31. Rausch, M. P. & Hastings, K. T. Immune Checkpoint Inhibitors in the Treatment of Melanoma: From Basic Science to Clinical Application. in *Cutaneous Melanoma: Etiology and Therapy* 121–142 (Codon Publications, 2017). doi:10.15586/codon.cutaneousmelanoma.2017.ch9

32. Ganesan, S. & Mehnert, J. Biomarkers for Response to Immune Checkpoint Blockade. *Annual Review of Cancer Biology* **4**, 331–351 (2020).

33. Lei, Y., Li, X., Huang, Q., Zheng, X. & Liu, M. Progress and Challenges of Predictive Biomarkers for Immune Checkpoint Blockade. *Frontiers in Oncology* **11**, 609 (2021).

34. Ishizuka, J. J. *et al.* Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. *Nature* **565**, 43–48 (2019).

35. Gannon, H. S. *et al.* Identification of ADAR1 adenosine deaminase dependency in a subset of cancer cells. *Nat. Commun.* **9**, 5450 (2018).

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

37. Bahn, J. H. *et al.* Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res.* **22**, 142–150 (2012).

38. Tan, M. H. *et al.* Dynamic landscape and regulation of RNA editing in mammals. *Nat. Publ. Gr.* **550**, 249–254 (2017).

39. Chen, C. X. *et al.* A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains.

RNA **6**, 755–767 (2000).

40. Walkley, C. R. & Li, J. B. Rewriting the transcriptome: adenosine-to- inosine RNA editing by ADARs. *Genome Biol.* **18**, 1–13 (2017).
41. Xu, L.-D. & Öhman, M. ADAR1 Editing and its Role in Cancer. *Genes (Basel).* **10**, (2019).
42. Heerboth, S. *et al.* EMT and tumor metastasis. *Clin. Transl. Med.* **4**, 6 (2015).
43. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial–mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **15**, 178–196 (2014).
44. 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).
45. Fu, L. *et al.* RNA editing of SLC22A3 drives early tumor invasion and metastasis in familial esophageal cancer. *PNAS* **114**, 4631–4640 (2017).
46. Amin, E. M. *et al.* The RNA-editing enzyme ADAR promotes lung adenocarcinoma migration and invasion by stabilizing FAK. *Sci. Signal.* **10**, (2017).
47. Han, S.-W. *et al.* RNA editing in RHOQ promotes invasion potential in colorectal cancer. *J. Exp. Med.* **211**, 613–21 (2014).
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. Genomic Data Commons. Available at: https://portal.gdc.cancer.gov/. (Accessed: 10th March 2020)
50. Hanzelmann, S., Castelo, R. & Guinney, J. GSVA : gene set variation analysis for

microarray and RNA-Seq data. *BMC Bioinformatics* **14**, (2013).

51. 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).
52. 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).
53. Tran, S. S. *et al.* Widespread RNA editing dysregulation in brains from autistic individuals. *Nat. Neurosci.* **22**, 25–36 (2019).
54. 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).
55. REDIportal. Available at: http://srv00.recas.ba.infn.it/atlas/. (Accessed: 10th March 2020)
56. 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).
57. 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).
58. 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).
59. 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).

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

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

62. Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines | NCI Genomic Data Commons. Available at: https://gdc.cancer.gov/about-data/publications/mc3-2017. (Accessed: 10th March 2020)

63. 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).

64. 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

65. 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).

66. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, Bassez A, Decaluwé H, Pircher A, Van den Eynde K, Weynand B, Verbeken E, De Leyn P, Liston A, Vansteenkiste J, Carmeliet P, Aerts S, T. B. E-MTAB-6149 - Single cell sequencing of lung carcinoma. *ArrayExpress* (2018). Available at: https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6149/. (Accessed: 10th

March 2020)

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

68. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

69. CIBERSORTx. Available at: https://cibersortx.stanford.edu/. (Accessed: 10th March 2020)

70. Tran, S. S., Zhou, Q. & Xiao, X. Statistical inference of differential RNA-editing sites from RNA-sequencing data by hierarchical modeling. *Bioinformatics* **36**, 2796–2804 (2020).

71. 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).

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

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

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

75. 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).

76. 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).

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

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

79. 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).

80. 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).

81. 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).

82. 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).

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

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

85. Zaoui, K. *et al.* Ran promotes membrane targeting and stabilization of RhoA to

orchestrate ovarian cancer cell invasion. *Nat. Commun.* **10**, 2666 (2019).

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

87. 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).

88. 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).

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

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

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

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

93. 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).

94. 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).

95. 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).

96. 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).

97. 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).

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

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

100. 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).

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

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

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

104. 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).

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

106. 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).

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

108. 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).

109. 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

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

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

112. 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).

113. 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).

114. 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).

115. Strong, J. E., Coffey, M. C., Tang, D., Sabinin, P. & Lee, P. W. K. The molecular basis of viral oncolysis: Usurpation of the Ras signaling pathway by reovirus. *EMBO J.* **17**, 3351–3362 (1998).
116. 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).
117. 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).
118. Erdmann, E. A., Mahapatra, A., Mukherjee, P., Yang, B. & Hundley, H. A. To protect and modify double-stranded RNA–the critical roles of ADARs in development, immunity and oncogenesis. *Critical Reviews in Biochemistry and Molecular Biology* **56**, 54–87 (2021).
119. Song, B., Shiromoto, Y., Minakuchi, M. & Nishikura, K. The role of RNA editing enzyme ADAR1 in human disease. *Wiley Interdisciplinary Reviews: RNA* (2021). doi:10.1002/wrna.1665
120. Chan, T. W. *et al.* RNA editing in cancer impacts mRNA abundance in immune response pathways. *Genome Biol.* **21**, 268 (2020).
121. Marceca, G. P. *et al.* Detecting and characterizing a-to-i microrna editing in cancer. *Cancers* **13**, 1699 (2021).
122. Lamers, M. M., van den Hoogen, B. G. & Haagmans, B. L. ADAR1: 'Editor-in-Chief' of Cytoplasmic Innate Immunity. *Frontiers in immunology* **10**, 1763 (2019).

123. Ahmad, S. *et al.* Breaching Self-Tolerance to Alu Duplex RNA Underlies MDA5-Mediated Inflammation. *Cell* **172**, 797-810.e13 (2018).

124. Hartner, J. C., Walkley, C. R., Lu, J. & Orkin, S. H. ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling. *Nat. Immunol.* **10**, 109–115 (2009).

125. Mehdipour, P. *et al.* Epigenetic therapy induces transcription of inverted SINEs and ADAR1 dependency. *Nature* **588**, 169–173 (2020).

126. Maynard, A. *et al.* Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. *Cell* **182**, 1232-1251.e22 (2020).

127. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

128. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

129. Travaglini, K. J. *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).

130. inferCNV of the Trinity CTAT Project. Available at: https://github.com/broadinstitute/inferCNV. (Accessed: 23rd July 2021)

131. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

132. Mansi, L. *et al.* REDIportal: Millions of novel A-to-I RNA editing events from thousands of RNAseq experiments. *Nucleic Acids Res.* **49**, D1012–D1019 (2021).

133. Ellrott, K. *et al.* Scalable Open Science Approach for Mutation Calling of Tumor

Exomes Using Multiple Genomic Article Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. 271–281 (2018). doi:10.1016/j.cels.2018.03.002

134. Finotello, F. *et al.* Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* **11**, 34 (2019).

135. Li, T. *et al.* TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48**, W509–W514 (2020).

136. Wang, C. *et al.* Identification of A-to-I RNA editing profiles and their clinical relevance in lung adenocarcinoma. *Sci. China Life Sci.* 1–14 (2021). doi:10.1007/s11427-020-1928-0

137. Dong, J., Li, B., Lin, D., Zhou, Q. & Huang, D. Advances in Targeted Therapy and Immunotherapy for Non-small Cell Lung Cancer Based on Accurate Molecular Typing. *Frontiers in Pharmacology* **10**, 230 (2019).

138. Wu, F. *et al.* Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat. Commun.* **12**, 1–11 (2021).

139. Liddicoat, B. J. *et al.* RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science (80-.).* **349**, 1115–1120 (2015).

140. Vitali, P. & Scadden, A. D. J. Double-stranded RNAs containing multiple IU pairs are sufficient to suppress interferon induction and apoptosis. *Nat. Struct. Mol. Biol.* **17**, 1043–1050 (2010).

141. Benci, J. L. *et al.* Opposing Functions of Interferon Coordinate Adaptive and Innate Immune Responses to Cancer Immune Checkpoint Blockade. *Cell* **178**,

933-948.e14 (2019).

142. Cheon, H. *et al.* IFNβ-dependent increases in STAT1, STAT2, and IRF9 mediate resistance to viruses and DNA damage. *EMBO J.* **32**, 2751–2763 (2013).

143. Cheon, H. J. & Stark, G. R. Unphosphorylated STAT1 prolongs the expression of interferon-induced immune regulatory genes. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9373–9378 (2009).

144. Budhwani, M., Mazzieri, R. & Dolcetti, R. Plasticity of Type I Interferon-Mediated Responses in Cancer Therapy: From Anti-tumor Immunity to Resistance. *Front. Oncol.* **8**, 322 (2018).

145. Thorsson, V. *et al.* The Immune Landscape of Cancer. *Immunity* **48**, 812-830.e14 (2018).

146. Blango, M. G. & Bass, B. L. Identification of the long, edited dsRNAome of LPS-stimulated immune cells. *Genome Res.* **26**, 852–862 (2016).

147. Reich, D. P. & Bass, B. L. Mapping the dsrna world. *Cold Spring Harb. Perspect. Biol.* **11**, a035352 (2019).

148. Wu, J. S. *et al.* Cathepsin B defines leader cells during the collective invasion of salivary adenoid cystic carcinoma. *Int. J. Oncol.* **54**, 1233–1244 (2019).

149. Chen, C. H. *et al.* Study of Cathepsin B inhibition in VEGFR TKI treated human renal cell carcinoma xenografts. *Oncogenesis* **8**, 1–18 (2019).

150. Gong, F. *et al.* Cathepsin B as a potential prognostic and therapeutic marker for human lung squamous cell carcinoma. *Mol. Cancer* **12**, (2013).

151. Sadik, A. *et al.* IL4I1 Is a Metabolic Immune Checkpoint that Activates the AHR and Promotes Tumor Progression. *Cell* **182**, 1252-1270.e34 (2020).

152. Stein, U. *et al.* MACC1, a newly identified key regulator of HGF-MET signaling, predicts colon cancer metastasis. *Nat. Med.* **15**, 59–67 (2009).

153. Alborelli, I. *et al.* Tumor mutational burden assessed by targeted NGS predicts clinical benefit from immune checkpoint inhibitors in non-small cell lung cancer. *J. Pathol.* **250**, 19–29 (2020).

154. Wu, Y. *et al.* The Predictive Value of Tumor Mutation Burden on Efficacy of Immune Checkpoint Inhibitors in Cancers: A Systematic Review and Meta-Analysis. *Frontiers in Oncology* **9**, 1161 (2019).

155. McGrail, D. J. *et al.* High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann. Oncol.* **32**, 661–672 (2021).

156. Rizvi, N. A. *et al.* Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science (80-.).* **348**, 124–128 (2015).

157. Offin, M. *et al.* Tumor mutation burden and efficacy of EGFR-tyrosine kinase inhibitors in patients with EGFR-mutant lung cancers. *Clin. Cancer Res.* **25**, 1063–1069 (2019).

158. Lin, C. *et al.* Tumor Mutation Burden Correlates With Efficacy of Chemotherapy/Targeted Therapy in Advanced Non–Small Cell Lung Cancer. *Front. Oncol.* **10**, 480 (2020).

159. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).

160. Bhat, P., Leggatt, G., Waterhouse, N. & Frazer, I. H. Interferon-γ derived from

cytotoxic lymphocytes directly enhances their motility and cytotoxicity. *Cell Death Dis.* **8**, e2836 (2017).

161. Castro, F., Cardoso, A. P., Gonçalves, R. M., Serre, K. & Oliveira, M. J. Interferon-gamma at the crossroads of tumor immune surveillance or evasion. *Frontiers in Immunology* **9**, 847 (2018).

162. Patterson, J. B., Thomis, D. C., Hans, S. L. & Samuel, C. E. Mechanism of Interferon Action: Double-Stranded RNA-Specific Adenosine Deaminase from Human Cells Is Inducible by Alpha and Gamma Interferons. *Virology* **210**, 508–511 (1995).

163. Diskin, B. *et al.* PD-L1 engagement on T cells promotes self-tolerance and suppression of neighboring macrophages and effector T cells in cancer. *Nat. Immunol. 2020 214* **21**, 442–454 (2020).

164. Doroshow, D. B. *et al.* PD-L1 as a biomarker of response to immune-checkpoint inhibitors. *Nature Reviews Clinical Oncology* **18**, 345–362 (2021).

165. Tong, G. *et al.* MACC1 regulates PDL1 expression and tumor immunity through the c-Met/AKT/mTOR pathway in gastric cancer cells. *Cancer Med.* **8**, 7044–7054 (2019).

166. Wang, G. Z. *et al.* The Aryl hydrocarbon receptor mediates tobacco-induced PD-L1 expression and is associated with response to immunotherapy. *Nat. Commun.* **10**, 1125 (2019).

167. Addeo, A. *et al.* Immunotherapy in non-small cell lung cancer harbouring driver mutations. *Cancer Treatment Reviews* **96**, 102179 (2021).

168. St-Pierre, F., Bhatia, S. & Chandra, S. Harnessing natural killer cells in cancer

immunotherapy: A review of mechanisms and novel therapies. *Cancers* **13**, 1988 (2021).

169. Wang, Y. *et al.* FDA-Approved and Emerging Next Generation Predictive Biomarkers for Immune Checkpoint Inhibitors in Cancer Patients. *Frontiers in Oncology* **11**, 2115 (2021).
170. Kitano, S., Nakayama, T. & Yamashita, M. Biomarkers for immune checkpoint inhibitors in melanoma. *Frontiers in Oncology* **8**, 270 (2018).
171. Twomey, J. D. & Zhang, B. Cancer Immunotherapy Update: FDA-Approved Checkpoint Inhibitors and Companion Diagnostics. *AAPS Journal* **23**, 1–11 (2021).
172. Grasso, C. S. *et al.* Conserved Interferon-γ Signaling Drives Clinical Response to Immune Checkpoint Blockade Therapy in Melanoma. *Cancer Cell* **38**, 500-515.e3 (2020).
173. Heraud-Farlow, J. E. & Walkley, C. R. What do editors do? Understanding the physiological functions of A-to-I RNA editing by adenosine deaminase acting on RNAs. *Open Biol.* **10**, 200085 (2020).
174. Hugo, W. *et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **165**, 35–44 (2016).
175. Riaz, N. *et al.* Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell* **171**, 934-949.e15 (2017).
176. Liu, D. *et al.* Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nat. Med.* **25**, 1916–1927 (2019).

177. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).

178. Kato, H. *et al.* Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5. *J. Exp. Med.* **205**, 1601–1610 (2008).

179. Gondi, C. S. & Rao, J. S. Cathepsin B as a cancer target. *Expert Opinion on Therapeutic Targets* **17**, 281–291 (2013).

180. Jorgovanovic, D., Song, M., Wang, L. & Zhang, Y. Roles of IFN-γin tumor progression and regression: A review. *Biomarker Research* **8**, 1–16 (2020).

181. Banerjee, A. *et al.* Modulation of paired immunoglobulin-like type 2 receptor signaling alters the host response to Staphylococcus aureus-induced pneumonia. *Infect. Immun.* **78**, 1353–1363 (2010).

182. Rothlin, C. V & Ghosh, S. Lifting the innate immune barriers to antitumor immunity. *J. Immunother. Cancer* **8**, e000695 (2020).

183. Siddiqui, J. & Miles, W. O. RNA Editing Signatures Predict Response to Immunotherapies in Melanoma Patients. *bioRxiv* 2020.08.12.248393 (2020). doi:10.1101/2020.08.12.248393

184. Tusup, M. *et al.* Evaluation of the interplay between the adar editome and immunotherapy in melanoma. *Non-coding RNA* **7**, 1–11 (2021).

185. Castro, F., Cardoso, A. P., Gonçalves, R. M., Serre, K. & Oliveira, M. J. Interferon-gamma at the crossroads of tumor immune surveillance or evasion. *Frontiers in Immunology* **9**, 1 (2018).