

# UC Santa Cruz

## UC Santa Cruz Electronic Theses and Dissertations

### Title

Investigating Functional Roles of Driver Mutations in the Context of Co-Occurring Mutations and Environmental Stress

### Permalink

<https://escholarship.org/uc/item/4n60n83x>

### Author

Liang, Cindy

### Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

SANTA CRUZ

**Investigating functional roles of driver mutations in the context of  
co-occurring mutations and environmental stress**

A dissertation submitted in partial satisfaction of the requirements for the  
degree of

DOCTOR OF PHILOSOPHY

in

MOLECULAR, CELL AND DEVELOPMENTAL BIOLOGY

by

**Cindy E Liang**

June 2024

The Dissertation of Cindy  
Liang is approved:

---

Professor Angela N. Brooks,  
chair

---

Professor Joshua Arribere

---

Professor Olena Vaske

---

Peter Biehl  
Vice Provost and Dean of Graduate Studies

Copyright © by  
Cindy E. Liang  
2024

## Contents

Figure List.....	v
Abstract.....	vi
Acknowledgements.....	viii
Chapter 1 - Introduction.....	1
1.1 -Lung cancer is a complex pathology influenced by genetics and the environment.....	1
1.2 -Profiling genetic interactions of multiple mutations and aberrant splicing as a mechanism to understand lung cancer.....	2
1.3 - U2AF1 is a splicing factor recurrently mutated in lung adenocarcinoma.....	4
1.4 - Aberrant splicing is associated with increased viability in stress.....	5
1.5 - Use of new sequencing technologies to understand the cancer epi-transcriptome.....	7
1.6 - References .....	9
Chapter 2 - <i>U2AF1</i> <sup>S34F</sup> enhances tumorigenic potential by exhibiting synergy with <i>KRAS</i> mutation and altering response to environmental perturbations.....	13
2.1 - Introduction.....	13

2.2 - Results.....	15
2.3 - Discussion.....	28
2.4 - Methods.....	31
2.5 - Figures.....	44
2.6 - Acknowledgements.....	51
2.7 - Data access.....	52
2.8 - References.....	53
Chapter 3: direct RNA sequencing reveals altered modification landscape from <i>U2AF1<sup>S34F</sup></i> and cigarette smoke concentrate.....	59
3.1 - Introduction.....	59
3.2 - Results.....	61
3.3 - Discussion.....	64
3.4 - Methods.....	65
3.5 - Figures .....	72
3.6 - Acknowledgements.....	76
3.7 - References.....	76

## Figure List

### Chapter 2

Figure 1. *KRAS*<sup>G12V</sup> suppresses the effect of *U2AF1*<sup>S34F</sup> on the transcriptome while altering gene expression in oncogenic pathways.

Figure 2. Co-occurring *U2AF1*<sup>S34F</sup> and *KRAS*<sup>G12V</sup> produces unique splicing events distributions, while *U2AF1*<sup>S34F</sup> alone increases splicing in stress granule protein genes.

Figure 3. Co-occurring *U2AF1*<sup>S34F</sup> and *KRAS*<sup>G12V</sup> mutations increase oncogenic potential and proliferation.

Figure 4. Altered splicing in stress granule genes in *U2AF1*<sup>S34F</sup> HBEC3kts is associated with enhanced stress response.

### Chapter 3.

Figure 1. EpiNano and xPore predict differential modification in *U2AF1*<sup>S34F</sup> cells.

Figure 2. In untreated HBEC3kts, *VAMP8* modification in *U2AF1*<sup>WT</sup> lines is associated with altered protein abundance.

Figure 3. xPore predicts a global increase of m6A modifications caused by *U2AF1*<sup>S34F</sup> and cigarette smoke exposure.

## Abstract

### Investigating phenotypic and transcriptomic alterations of *U2AF1<sup>S34F</sup>* to genetic and environmental stress

by

Cindy E Liang

*U2AF1<sup>S34F</sup>*, a somatic splicing factor mutation, is frequently recurrent in human neoplasias such as lung adenocarcinoma (ADC). Although *U2AF1<sup>S34F</sup>* has been shown to occur early in tumor lineages, the mutation, alone, is insufficient for producing tumors. However, lung ADC patients with *U2AF1<sup>S34F</sup>* frequently have co-occurring *KRAS* mutations and smoking histories. We hypothesized that *U2AF1<sup>S34F</sup>* interacts with oncogenic *KRAS* and environmental stress to promote tumor-forming potential. To elucidate interaction of *U2AF1<sup>S34F</sup>* with a co-occurring mutation, we generated human bronchial epithelial cells (HBEC3kts) with *U2AF1<sup>S34F</sup>* or with co-occurring *U2AF1<sup>S34F</sup>* and *KRAS<sup>G12V</sup>*. From analyzing short-read transcriptome sequences, we found synergistic effects of co-occurring mutations on gene expression in cell cycle and inflammatory pathways associated with increased tumors in mouse xenografts, anchorage-independent growth, proliferation, and altered cytokine production. Interestingly, HBEC3kts harboring only *U2AF1<sup>S34F</sup>* display increased splicing in stress granule protein genes and increased viability in cigarette

smoke concentrate. Our results suggest that *U2AF1<sup>S34F</sup>* may prime cells for transformation by allowing precancerous cells to survive longer when environmental stress is present, permitting *U2AF1<sup>S34F</sup>* cells to accumulate transforming mutations, such as *KRAS<sup>G12V</sup>*

Next, I sought to further investigate the impact of *U2AF1<sup>S34F</sup>* and environmental stress response by profiling the mRNA modification landscape of *U2AF1<sup>S34F</sup>* and *U2AF1<sup>WT</sup>* HBEC3kts exposed to cigarette smoke concentrate (CSC) using Nanopore direct RNA sequencing (dRNA-seq). Preliminary results show that RNA modifications in autophagy gene *VAMP8* were associated with altered protein expression levels. We also show that CSC and the presence of *U2AF1<sup>S34F</sup>* both increase the number of RNA modifications in the transcriptome, with the highest number of modifications occurring in CSC treated *U2AF1<sup>S34F</sup>* HBEC3kts. We hypothesize that *U2AF1<sup>S34F</sup>* and CSC modify the RNA modification landscape in a synergistic way to increase oncogenic potential.

## **Acknowledgements**

For their mentorship and scientific expertise, I thank my advisor Dr. Angela Brooks, Dr. Eva Hrabeta-Robinson, Dr. Namrita Dhillon, and my thesis committee, Dr. Joshua Arribere and Dr. Olena Vaske. I would like to especially acknowledge Dr. Hrabeta-Robinson for the conceptualization of and experiments performed during the early stages of the U2AF1 KRAS project, several of which are included in chapter 2 of this thesis. For the cell lines used in this work, I thank Drs. Dennis Fei and Harold Varmus. For their collaboration and support, I thank my past and present colleagues in the Brooks lab, Dr. Cameron Soulette, Dr. Alexis Thornton, Dr. Alison Tang, Dr. Brandon Saint-John, Dr. Megan Durham, Dr. Jon Akutagawa, Dr. Amit Behera, Dr. Robert Shelansky, Colette Felton, Dennis Mulligan, and Gali Bai. Thank you to the UCSC Nanopore Group, including Vikas Peddu for performing pilot HBEC3kt direct RNA modification analysis on m6aNet and xPore, Marcus Viscardi for discussions on troubleshooting Nanopore experiments, and past Nanopore group member Dr. Miten Jain for his guidance on early direct RNA library preparation protocols. I would like to thank Dr. Shaheen Sikandar and Isobel Fetter for collaborating with us for the mouse xenograft experiments and providing insightful feedback on the manuscript. I would also like to acknowledge the UCSC Chemical Screening Center and Dr. Beverley Rabbitts, for providing expertise and mentorship in our immunostaining and cell imaging experiments. This work was made possible with the support of UCSC staff, faculty and facilities, with special acknowledgements to Teel Lopez, Carrie Niblett, Kaitlyn Murphy, and Grace Kistler-Fair.

## Chapter 1 - Introduction

### **1.1 Lung cancer is a complex pathology influenced by genetics and the environment**

Lung cancer is the leading cause of cancer deaths and most frequently diagnosed cancer ([Bray et al. 2024](#)). Although lung cancer is widely known to be linked with smoking history ([Weber et al. 2023](#)), around 25% of lung cancer cases worldwide occur in patients who have never smoked ([Sun et al. 2007](#)), implying that lung cancer incidence occurs from a complex interaction of genetic and environmental factors.

We focus on lung ADC, as it is the most prevalent lung cancer subtype ([Zhang et al. 2023](#)). Genomic sequencing of lung ADC patients stratified by smoking history has shown that genetics and the environment are inextricably linked. Patients with smoking histories possess different mutational signatures compared to patients without ([Imielinski et al. 2012](#); [Alexandrov et al. 2018](#)). Although we are still understanding the ways in which environmental exposure to cigarette smoke can alter the genome, it is thought that the compounds in cigarette smoke make chemical adducts on DNA, which induces mutations. However, the field has yet to fully profile how these mutations can change the biology of the cell.

There are two hurdles of lung cancer treatment that can be addressed with advances in precision medicine. First, early stages of lung cancer are asymptomatic, so patients

are often diagnosed at a later stage, where treatment options are more invasive. For instance, if surgery is not an option because of the location of the tumor or metastasis, chemotherapy and radiation are often used. Second, patients with preexisting mutations may be ineligible for specific therapies. For instance, patients with *EGFR* mutations are ineligible for certain immunotherapies and require alternative drugs ([Lisberg et al. 2018](#)). These factors lead us to the utility of using next-generation sequencing (NGS) in cancer medicine. First, sequencing has the potential to identify cancer-specific signatures that can serve as prognostic markers at earlier stages of disease. Second, information like biological pathways impacted by differing mutational statuses gained from sequencing can inform effective treatments for patients.

### **1.2 - Profiling genetic interactions of multiple mutations and aberrant splicing as a mechanism to understand lung cancer**

The current state of precision medicine for cancer takes advantage of NGS to profile mutations in patients. In NGS, DNA or RNA is fragmented and amplified before sequencing ([Qin 2019](#)). The resulting nucleotide reads can then be analyzed computationally for the presence of mutations. RNA reads can be additionally analyzed for gene expression levels and splicing of transcripts. However, there are limitations to how NGS is currently applied clinically. Current approaches for using NGS to determine eligibility of lung cancer patients for certain treatments focus on

looking for alterations in panels of genes ([Lindeman et al. 2018](#); [Simarro et al. 2023](#)). However, sequencing analysis has shown that multiple mutations can interact with each other in lung cancer, leading to enrichment in oncogenic gene sets such as those involved in epithelial-to-mesenchymal transition ([Li et al. 2019](#)). Studies like this indicate a need for a holistic approach that accounts for the impact of co-occurring mutations on biological pathways that impact cancer incidence.

During the processing of genetic information, DNA is transcribed into unprocessed pre-mRNA in the nucleus ([Wahl, Will, and Luhrmann 2009](#)). Then, pre-mRNA is processed by a complex of proteins and RNA called the spliceosome to create mRNA in a process called splicing. The mature RNA is then exported to the cytoplasm and translated into proteins. Several categories of splicing events have been profiled, illustrating the complexity of transcripts produced by this process ([Blencowe 2006](#)). The most common splicing event in mammals are cassette exon events, where an exon flanked by introns is constitutively removed. The identity of exons at the beginning and ends of transcripts can also be modulated by splicing, such as in alternative first and last exon events respectively. Additionally, unspliced introns can remain in the processed mRNA in intron retention events ([Grabski et al. 2020](#)).

Since spliced transcripts dictate what proteins are produced in the cell, they are often implicated in pathologies such as cancer. Splicing factors (SF), spliceosome proteins that regulate this process, are recurrently mutated in cancer and are associated with

oncogenic phenotypes such as reduced immune infiltration, increased cell proliferation, and reduced survival in cancer patients ([Seiler et al. 2018](#); [Li et al. 2017](#)). Mutations in different SF genes within tumor types often lead to deregulation of the same cancer pathways. Furthermore, many common SF mutations are mutually exclusive within their tumor cohort, suggesting a functional convergence in these potential driver mutations ([Seiler et al. 2018](#)). As *U2AF1<sup>S34F</sup>* is the most significantly recurrent splicing factor mutation in lung ADC ([Imielinski et al. 2012](#); [Brooks et al. 2014](#)), studying its role in transformation should not only elucidate the impact of its aberrant isoforms on lung cancer, but also provide broader insights on how these aberrant isoforms may impact cancer progression.

### **1.3 U2AF1 is a splicing factor recurrently mutated in lung adenocarcinoma**

U2AF1 is a subunit of the U2 Auxiliary Factor complex ([Zamore and Green 1989](#)). This splicing factor recognizes and binds to AG nucleotides at the 3' splice site to facilitate spliceosome assembly ([Wu et al 1999](#)). Comparisons made between lung ADC samples wild-type (WT) and S34F mutant for *U2AF1* have revealed numerous genes significantly differentially spliced, making *U2AF1<sup>S34F</sup>* a strong candidate for understanding how SF mutations in lung ADC affect oncogenicity ([Brooks et al. 2014](#)). Furthermore, *U2AF1<sup>S34F</sup>* has been found to be a truncal mutation in lung ADC primary samples ([Esfahani et al. 2019](#)), so understanding its functional impact on oncogenic potential can reveal information about early-stage cancer.

Critical barriers are present to understanding how *U2AF1<sup>S34F</sup>* pushes the cell towards a more oncogenic fate. First, the introduction of *U2AF1<sup>S34F</sup>* alone is insufficient to transform noncancerous cells and causes no change to growth phenotype when expressed in immortalized HBEC3kts (Fei et al., 2016). Second, the function of aberrant splicing resulting from SF mutations are largely unknown.

In the second chapter of my work, I explore *U2AF1<sup>S34F</sup>*'s functional impacts on preneoplastic potential by investigating the effect of co-occurring *U2AF1<sup>S34F</sup>* and *KRAS<sup>G12V</sup>* mutations on splicing and gene expression, and how these transcriptional changes impact biological function in the cell.

#### **1.4 Aberrant splicing is associated with increased viability in stress**

Alternative splicing changes have been shown to increase survival growth in the presence of environmental stressors. For instance, in human cancer cells, exogenous expression of a long isoform of *RPS24* caused by aberrant cassette exon inclusion is associated with increased viability in hypoxic media (Erin et al. 2024).

*U2AF1<sup>S34F</sup>* has also been shown to confer resistance to environmental stress. Existing work on *U2AF1<sup>S34F</sup>* present alone in HBEC3kts has shown that it confers increased viability after treatment with ionizing radiation (Palangat et al. 2019). Another group studying acute myeloid leukemia cells found that *U2AF1<sup>S34F</sup>* confers altered splicing

and binding in stress granule genes (Biancon et al. 2022). This transcriptome alteration was associated with increased stress granule formation and viability in sodium arsenite, a chemical agent of oxidative stress and DNA damage (Ruiz-Ramos et al. 2009).

Cigarette smoke is a source of many kinds of cellular stress. It can create reactive oxygen species (ROS), which induce DNA damage and inflammation (Caliri, Tommasi, and Besaratinia 2021). Although work has been done to understand the impact of cigarette smoke on RNA expression, the intersection between splicing factor mutations and cigarette smoke exposure has yet to be explored.

Another category of transcriptional alteration that has been shown to alter stress response is RNA modifications. One of the most common eukaryotic mRNA modifications are N6-methyladenosine (m6A) modifications (Yue, Liu, and He 2015). m6A modifications occur at RRACH consensus motifs, with R representing A or G nucleotides and H representing A, C, or U, and regulate important developmental processes such as differentiation and proliferation by targeting transcripts for decay (Wang et al., 2014). Overexpression of the m6A-writer protein METTL3 helps mouse epithelial cells survive oxidative stress caused by the drug colistin (Wang et al. 2019), indicating a functional role of RNA modifications on stress response. However, a gap in the field lies in understanding the relationship between RNA modifications, splicing factor mutations, and cigarette smoke exposure.

## **1.5 Use of new sequencing technologies to understand the cancer**

### **epi-transcriptome**

The invention of new sequencing techniques allows us to profile the transcriptome in new ways. Older short-read sequencing methods generate reads using a library preparation process where RNA is fragmented, then amplified using PCR ([Qin 2019](#)). There are two limitations to this method. First, the assembly of full-length isoforms from fragmented transcripts is difficult when there are many repetitive sequences ([Hardwick et al. 2019](#)). Second, methods which utilize PCR are prone to PCR bias, a phenomenon where genic loci with higher abundance of certain nucleotides are artificially over- or under-represented in the library due to differences in amplification efficiency ([Aird et al. 2011](#)). Third-generation sequencing platforms such as Nanopore and PacBio allow for the sequencing of longer stretches of RNA, which permits the assembly of isoforms that cannot be resolved with short-read sequencing alone. For instance, analysis of *U2AF1<sup>S34F</sup>* HBEC3kts mRNA reads generated by Nanopore cDNA sequencing revealed multiple complex isoforms that were not captured by short-read sequencing ([Soulette et al. 2023](#)). In a recent evaluation of long-read sequencing technologies, of which I am a co-author, it was revealed that longer, more accurate sequences generated by long-read technologies resulted in more accurate transcripts, again highlighting the utility of long-read sequencing ([Pardo-Palacios et al. 2024](#)).

Nanopore dRNA-seq in particular is a PCR-free method that sequences native RNA molecules by passing nucleotide molecules through a pore ([Workman et al. 2019](#)). Changes in current caused by different molecules passing through are then recorded and assigned nucleotides. This feature allows for the identification of chemical modifications on the RNA. For instance, analysis on dRNA-seq data has been shown to identify the presence of N6-methyladenosine (m6A) modifications ([Lorenz et al., 2020](#)).

In the third chapter of my work, I profile the transcriptomes of *U2AF1<sup>S34F</sup>* and *U2AF1<sup>WT</sup>* cells exposed to cigarette smoke concentrate (CSC) using dRNA-seq to profile the impact of cigarette smoke and *U2AF1<sup>S34F</sup>* on the transcriptome and stress response.

## 1.6 References

- Aird, Daniel, Michael G. Ross, Wei-Sheng Chen, Maxwell Danielsson, Timothy Fennell, Carsten Russ, David B. Jaffe, Chad Nusbaum, and Andreas Gnirke. 2011a. “Analyzing and Minimizing PCR Amplification Bias in Illumina Sequencing Libraries.” *Genome Biology* 12 (2): R18.
- Alexandrov, Ludmil B., Young Seok Ju, Kerstin Haase, Peter Van Loo, Iñigo Martincorena, Serena Nik-Zainal, Yasushi Totoki, et al. 2016. “Mutational Signatures Associated with Tobacco Smoking in Human Cancer.” *Science* 354 (6312): 618–22.
- Biancon, Giulia, Poorval Joshi, Joshua T. Zimmer, Torben Hunck, Yimeng Gao, Mark D. Lessard, Edward Courchaine, et al. 2022. “Precision Analysis of Mutant U2AF1 Activity Reveals Deployment of Stress Granules in Myeloid Malignancies.” *Molecular Cell* 82 (6): 1107–22.e7.
- Blencowe, Benjamin J. 2006. “Alternative Splicing: New Insights from Global Analyses.” *Cell* 126 (1): 37–47.
- Bray, Freddie, Mathieu Laversanne, Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Isabelle Soerjomataram, and Ahmedin Jemal. 2024. “Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.” *CA: A Cancer Journal for Clinicians*, April. <https://doi.org/10.3322/caac.21834>.
- Brooks, Angela N., Peter S. Choi, Luc de Waal, Tanaz Sharifnia, Marcin Imielinski, Gordon Saksena, Chandra Sekhar Pedamallu, et al. 2014. “A Pan-Cancer Analysis of Transcriptome Changes Associated with Somatic Mutations in U2AF1 Reveals Commonly Altered Splicing Events.” *PloS One* 9 (1): e87361.
- Caliri, Andrew W., Stella Tommasi, and Ahmad Besaratinia. 2021. “Relationships among Smoking, Oxidative Stress, Inflammation, Macromolecular Damage, and Cancer.” *Mutation Research-Reviews in Mutation Research* 787 (January): 108365.
- Esfahani, Mohammad S., Luke J. Lee, Young-Jun Jeon, Ryan A. Flynn, Henning Stehr, Angela B. Hui, Noriko Ishisoko, et al. 2019. “Functional Significance of U2AF1 S34F Mutations in Lung Adenocarcinomas.” *Nature Communications* 10 (1): 5712.
- Fei, Dennis Liang, Hayley Motowski, Rakesh Chatrikhi, Sameer Prasad, Jovian Yu, Shaojian Gao, Clara L. Kielkopf, Robert K. Bradley, and Harold Varmus. 2016. “Wild-Type U2AF1 Antagonizes the Splicing Program Characteristic of

U2AF1-Mutant Tumors and Is Required for Cell Survival.” *PLoS Genetics* 12 (10): e1006384.

Grabski, David F., Lucile Broseus, Bandana Kumari, David Rekosh, Marie-Louise Hammarskjold, and William Ritchie. 2021. “Intron Retention and Its Impact on Gene Expression and Protein Diversity: A Review and a Practical Guide.” *Wiley Interdisciplinary Reviews. RNA* 12 (1): e1631.

Hardwick, Simon A., Anoushka Joglekar, Paul Flicek, Adam Frankish, and Hagen U. Tilgner. 2019. “Getting the Entire Message: Progress in Isoform Sequencing.” *Frontiers in Genetics* 10 (August): 709.

Imielinski, Marcin, Alice H. Berger, Peter S. Hammerman, Bryan Hernandez, Trevor J. Pugh, Eran Hodis, Jeonghee Cho, et al. 2012. “Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing.” *Cell* 150 (6): 1107–20.

Kerry, Jenna, Erin J. Specker, Morgan Mizzone, Andrea Brumwell, Leslie Fell, Jenna Goodbrand, Michael N. Rosen, and James Uniacke. 2024. “Autophagy-Dependent Alternative Splicing of Ribosomal Protein S24 Produces a More Stable Isoform That Aids in Hypoxic Cell Survival.” *FEBS Letters* 598 (5): 503–20.

Li, Yafang, Xiangjun Xiao, Yohan Bossé, Olga Gorlova, Ivan Gorlov, Younghun Han, Jinyoung Byun, et al. 2019. “Genetic Interaction Analysis among Oncogenesis-Related Genes Revealed Novel Genes and Networks in Lung Cancer Development.” *Oncotarget* 10 (19): 1760–74.

Li, Yongsheng, Nidhi Sahni, Rita Panca, Daniel J. McGrail, Juan Xu, Xu Hua, Jasmin Coulombe-Huntington, et al. 2017. “Revealing the Determinants of Widespread Alternative Splicing Perturbation in Cancer.” *Cell Reports* 21 (3): 798–812.

Lindeman, Neal I., Philip T. Cagle, Dara L. Aisner, Maria E. Arcila, Mary Beth Beasley, Eric H. Bernicker, Carol Colasacco, et al. 2018. “Updated Molecular Testing Guideline for the Selection of Lung Cancer Patients for Treatment With Targeted Tyrosine Kinase Inhibitors: Guideline From the College of American Pathologists, the International Association for the Study of Lung Cancer, and the Association for Molecular Pathology.” *Archives of Pathology & Laboratory Medicine* 142 (3): 321–46.

Lisberg, A., A. Cummings, J. W. Goldman, K. Bornazyan, N. Reese, T. Wang, P. Coluzzi, et al. 2018. “A Phase II Study of Pembrolizumab in EGFR-Mutant, PD-L1+, Tyrosine Kinase Inhibitor Naïve Patients With Advanced NSCLC.” *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer* 13 (8): 1138–45.

- Lorenz, Daniel A., Shashank Sathe, Jaclyn M. Einstein, and Gene W. Yeo. 2020. "Direct RNA Sequencing Enables m6A Detection in Endogenous Transcript Isoforms at Base-Specific Resolution." *RNA* 26 (1): 19–28.
- Palangat, Murali, Dimitrios G. Anastasakis, Dennis Liang Fei, Katherine E. Lindblad, Robert Bradley, Christopher S. Hourigan, Markus Hafner, and Daniel R. Larson. 2019. "The Splicing Factor U2AF1 Contributes to Cancer Progression through a Noncanonical Role in Translation Regulation." *Genes & Development* 33 (9-10): 482–97.
- Pardo-Palacios, Francisco J., Dingjie Wang, Fairlie Reese, Mark Diekhans, Silvia Carbonell-Sala, Brian Williams, Jane E. Loveland, et al. 2023. "Systematic Assessment of Long-Read RNA-Seq Methods for Transcript Identification and Quantification." *bioRxiv : The Preprint Server for Biology*, July. <https://doi.org/10.1101/2023.07.25.550582>.
- Qin, Dahui. 2019. "Next-Generation Sequencing and Its Clinical Application." *Cancer Biology & Medicine* 16 (1): 4–10.
- Ruiz-Ramos, Ruben, Lizbeth Lopez-Carrillo, Alfonso D. Rios-Perez, Andrea De Vizcaya-Ruiz, and Mariano E. Cebrian. 2009. "Sodium Arsenite Induces ROS Generation, DNA Oxidative Damage, HO-1 and c-Myc Proteins, NF-kappaB Activation and Cell Proliferation in Human Breast Cancer MCF-7 Cells." *Mutation Research* 674 (1-2): 109–15.
- Seiler, Michael, Shouyong Peng, Anant A. Agrawal, James Palacino, Teng Teng, Ping Zhu, Peter G. Smith, Cancer Genome Atlas Research Network, Silvia Buonamici, and Lihua Yu. 2018. "Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types." *Cell Reports* 23 (1): 282–96.e4.
- Simarro, Javier, Gema Pérez-Simó, Nuria Mancheño, Emilio Ansotegui, Carlos Francisco Muñoz-Núñez, José Gómez-Codina, Óscar Juan, and Sarai Palanca. 2023. "Impact of Molecular Testing Using Next-Generation Sequencing in the Clinical Management of Patients with Non-Small Cell Lung Cancer in a Public Healthcare Hospital." *Cancers* 15 (6). <https://doi.org/10.3390/cancers15061705>.
- Soulette, Cameron M., Eva Hrabeta-Robinson, Carlos Arevalo, Colette Felton, Alison D. Tang, Maximillian G. Marin, and Angela N. Brooks. 2023. "Full-Length Transcript Alterations in Human Bronchial Epithelial Cells with U2AF1 S34F Mutations." *Life Science Alliance* 6 (10). <https://doi.org/10.26508/lsa.202000641>.
- Sun, Sophie, Joan H. Schiller, and Adi F. Gazdar. 2007. "Lung Cancer in Never Smokers--a Different Disease." *Nature Reviews. Cancer* 7 (10): 778–90.

- Wahl, Markus C., Cindy L. Will, and Reinhard Lührmann. 2009. "The Spliceosome: Design Principles of a Dynamic RNP Machine." *Cell* 136 (4): 701–18.
- Wang, Jian, Muhammad Ishfaq, Liang Xu, Chunli Xia, Chunli Chen, and Jichang Li. 2019. "METTL3/m6A/miRNA-873-5p Attenuated Oxidative Stress and Apoptosis in Colistin-Induced Kidney Injury by Modulating Keap1/Nrf2 Pathway." *Frontiers in Pharmacology* 10 (May): 517.
- Wang, Xiao, Zhike Lu, Adrian Gomez, Gary C. Hon, Yanan Yue, Dali Han, Ye Fu, et al. 2014. "N6-Methyladenosine-Dependent Regulation of Messenger RNA Stability." *Nature* 505 (7481): 117–20.
- Workman, Rachael E., Alison D. Tang, Paul S. Tang, Miten Jain, John R. Tyson, Roham Razaghi, Philip C. Zuzarte, et al. 2019. "Nanopore Native RNA Sequencing of a Human poly(A) Transcriptome." *Nature Methods* 16 (12): 1297–1305.
- Wéber, András, Eileen Morgan, Jerome Vignat, Mathieu Laversanne, Margherita Pizzato, Harriet Rungay, Deependra Singh, et al. 2023. "Lung Cancer Mortality in the Wake of the Changing Smoking Epidemic: A Descriptive Study of the Global Burden in 2020 and 2040." *BMJ Open* 13 (5): e065303.
- Wu, S., C. M. Romfo, T. W. Nilsen, and M. R. Green. 1999. "Functional Recognition of the 3' Splice Site AG by the Splicing Factor U2AF35." *Nature* 402 (6763): 832–35.
- Yue, Yanan, Jianzhao Liu, and Chuan He. 2015. "RNA N6-Methyladenosine Methylation in Post-Transcriptional Gene Expression Regulation." *Genes & Development* 29 (13): 1343–55.
- Zamore, P. D., and M. R. Green. 1989. "Identification, Purification, and Biochemical Characterization of U2 Small Nuclear Ribonucleoprotein Auxiliary Factor." *Proceedings of the National Academy of Sciences of the United States of America* 86 (23): 9243–47.
- Zhang, Yanting, Salvatore Vaccarella, Eileen Morgan, Mengmeng Li, Jaione Etxeberria, Eric Chokunonga, Shyam Shunker Manraj, Bakarou Kamate, Abidemi Omonisi, and Freddie Bray. 2023. "Global Variations in Lung Cancer Incidence by Histological Subtype in 2020: A Population-Based Study." *The Lancet Oncology* 24 (11): 1206–18.

## **Chapter 2 - *U2AF1*<sup>S34F</sup> enhances tumorigenic potential by exhibiting synergy with *KRAS* mutation and altering response to environmental perturbations**

### **2.1 - Introduction**

Splicing factor mutations play important roles in cancer, leading to global dysregulation of RNA splicing in protein-coding genes ([Imielinski et al., 2012](#)). Although work still remains to fully characterize the functional consequences of dysregulated splicing, aberrant splicing caused by splicing factor mutations have been shown to create isoforms which result in abnormal gene expression ([Tang et al. 2020](#); [Soulette et al. 2023](#)). Here, we focus on *U2AF1*<sup>S34F</sup>. *U2AF1* is among the most significantly mutated genes in lung ADC ([Imielinski et al., 2012](#)). Of the *U2AF1* mutations, *U2AF1*<sup>S34F</sup> occurs the most frequently in lung ADC ([Brooks et al, 2014](#)), where it has been observed to co-occur with known cancer-driver mutations like those in *KRAS* ([Imielinski et al., 2012](#)). In lung ADC, *U2AF1*<sup>S34F</sup> has also been found to be a basal mutation ([Esfahani et al., 2019](#)), indicating that it may potentiate the accumulation of further genetic perturbation. Additionally, *U2AF1*<sup>S34F</sup> has been shown to alter protein translation through mRNA binding ([Palangat et al. 2019](#)). As such, understanding its impact on oncogenesis in the context of co-occurring mutations may pave the way for earlier diagnostic and treatment options for patients.

*U2AF1* is a subunit of the U2 Auxiliary Factor complex ([Zamore and Green 1989](#)). Wild-type *U2AF1* (*U2AF1*<sup>WT</sup>) facilitates spliceosome assembly by recognizing and

binding to the 3' splice site (Wu et al, 1999). U2AF1 can also directly bind to mRNA to repress protein translation (Palangat et al. 2019). In *U2AF1<sup>S34</sup>* cells, an amino acid substitution in the second zinc finger alters 3' splice site choice ([Fei et al, 2016](#), Palangat et al. 2019). Other impacts of this mutation include altered binding to mRNA leading to translational dysregulation, altered polyadenylation of transcripts, increased R loop formation, and reduced NMD activity (Palangat et al., 2019; [Park et al, 2016](#); [Chen et al, 2018](#); [Nguyen et al, 2018](#); [Cheruiyot et al, 2021](#) ). Multiple functional consequences of this mutation have been reported, including increased survival advantage in cells exposed to ionizing radiation, altered inflammatory cytokine secretion, and increased stress granule production (Palangat et al. 2019; Biancon et al., 2022). Despite these phenotypes, *U2AF1<sup>S34F</sup>* by itself is insufficient for human lung cell lines to form tumors in mouse xenograft experiments (Fei et al., 2016). Confoundingly, *U2AF1<sup>S34F</sup>* is associated with poorer prognosis in cancer patients, but the mutation decreases proliferation in cancer cell lines ([Zhu et al., 2021](#)).

Mutations in *KRAS* associated with lung cancer are also insufficient to independently cause *in vivo* transformation in HBEC3kts ([Sato et al. 2013](#)). However, mutations in *KRAS* have recently been reported to alter splicing ([Lo et al., 2022](#)) through downregulating the phosphorylation of splicing factors. Although it has been hypothesized that *U2AF1<sup>S34F</sup>* confers tumorigenic potential independent of that conferred by the driver mutations it co-occurs with ([Imielinski et al., 2012](#)), work still

remains to fully understand what that is. Additionally, the cooperativity of splicing perturbations caused by co-occurring  $U2AF1^{S34F}$  and  $KRAS^{G12V}$  in a pre-cancerous model has yet to be studied.

Here, we introduced  $KRAS^{G12V}$  to HBEC3kt lines with  $U2AF1^{S34F}$ . We pair short-read mRNA sequencing with *in vivo* and *in vitro* assays to assess the impact of transcriptome alterations caused by co-occurring  $U2AF1^{S34F}$  and  $KRAS^{G12V}$  on preneoplastic potential and compare these perturbations to those caused by  $U2AF1^{S34F}$  or  $KRAS^{G12V}$  alone. Our results reveal synergistic effects of co-occurring  $U2AF1^{S34F}$  and  $KRAS^{G12V}$  in gene expression and splicing, which translated to enhanced oncogenic potential. Additionally, we discovered increased splicing in stress granule genes conferred by  $U2AF1^{S34F}$  alone, which translated to enhanced resistance to environmental stress. We propose a model where  $U2AF1^{S34F}$  enhances preneoplastic potential by allowing cells to survive stress and synergize with the transcriptomic effects of subsequent accumulated mutations to create oncogenic phenotypes.

## 2.2 - Results

$KRAS^{G12V}$  suppresses the effect of  $U2AF1^{S34F}$  on the transcriptome while altering gene expression in oncogenic pathways

To understand the role of *U2AF1*<sup>S34F</sup> in cancer, we first analyzed currently available sequencing data from lung ADC primary samples with *U2AF1* mutations on cBioportal to identify co-occurring mutations in known lung ADC driver genes. We find that *U2AF1*<sup>S34F</sup> significantly co-occurs with *KRAS* mutations (Fig 1A), with *KRAS* mutations at the *G12* locus being the most common. From this analysis, we identified *KRAS*<sup>G12X</sup> as a candidate mutation to study in the context of *U2AF1*<sup>S34F</sup>.

We obtained two parental isogenic HBEC3kt clones that were either wild-type or mutant for *U2AF1* (Fei et al., 2016): one cell line was homozygous *U2AF1*<sup>WT</sup>, and the other cell line was heterozygous for *U2AF1*<sup>S34F</sup> at its endogenous locus (Sup 1A). Homozygous *U2AF1*<sup>S34F</sup> mutation is lethal, so was not a consideration (Wadugu et al. 2021). A *KRAS*<sup>G12V</sup> pLenti6\_V5 plasmid (Sato, *et al* Mol Cancer Res (2013) 11 (6): 638–650) has previously been used to identify genetic perturbations required to accomplish *in vivo* transformation of HBEC3kt (Sato et al. 2013), which we obtained to study the impact of co-occurring *U2AF1*<sup>S34F</sup> and *KRAS*<sup>G12V</sup> mutations on preneoplastic potential. We exogenously overexpressed *KRAS*<sup>G12V</sup> in *U2AF1*<sup>WT</sup> and *U2AF1*<sup>S34F</sup> HBEC3kt cells using this construct. As a transduction control, we also introduced *LacZ* using the same plasmid backbone. A total of 4 cell lines were generated per parental HBEC3kt clone: (1) *U2AF1*<sup>WT</sup> + *LacZ* (2) *U2AF1*<sup>S34F</sup> + *LacZ*, (3) *U2AF1*<sup>WT</sup> + *KRAS*<sup>G12V</sup>, (4) *U2AF1*<sup>S34F</sup> + *KRAS*<sup>G12V</sup>. Lentivirus is known to unpredictably integrate into the host genome, and other groups' use of this *KRAS*<sup>G12V</sup> vector reported variation in *KRAS*<sup>G12V</sup> expression (Sato et al. 2013). When we

validated our cell lines for *KRAS*<sup>G12V</sup> integration via immunoblot and gene expression analysis, we observed that *KRAS* overexpression in the was inconsistent across the cell lines (Sup 1B, Sup 2B-C).

Using these four cell lines, we performed short-read RNA sequencing and cell-based assays to understand how *U2AF1*<sup>S34F</sup> and *KRAS*<sup>G12V</sup> co-occurrence alters the transcriptome and biology of HBEC3kt (Fig 1B). We performed short-read Illumina sequencing of polyA-selected RNA on the cell lines. Then, we performed differential gene expression analysis and gene set enrichment analysis (Fig 1C). A gene set uniquely downregulated in *U2AF1*<sup>S34F</sup> + *KRAS*<sup>G12V</sup> HBEC3kts is the KRAS Signaling Down gene set, corresponding to genes downregulated when KRAS is active ([Liberzon et al. 2015](#)). In contrast, the *U2AF1*<sup>WT</sup> + *KRAS*<sup>G12V</sup> cell line had a positive enrichment for the KRAS Signaling Up gene set, corresponding to genes upregulated when KRAS is active, and no significant enrichment in the KRAS Signaling Down gene set. From this result, we propose that *U2AF1*<sup>S34F</sup> in the presence of oncogenic *KRAS* alters the KRAS signaling pathway.

We also observed that individually, *U2AF1*<sup>S34F</sup> and *KRAS*<sup>G12V</sup> produced opposite enrichment patterns to each other in KRAS signaling, coagulation, and inflammatory pathway gene sets (KRAS Signaling Up, Coagulation, IL6 JAK STAT3 Signaling, Inflammatory Response, TNFA Signaling Via NFKB, Interferon Alpha Response, Interferon Gamma Response). We hypothesize that *KRAS*<sup>G12V</sup> suppresses

*U2AF1*<sup>S34F</sup>-specific gene expression signatures in the *U2AF1*<sup>S34F</sup> + *KRAS*<sup>G12V</sup> cell line. To understand this observation, we quantified the ratio of *U2AF1*<sup>S34F</sup> mRNA in cells with and without *KRAS*<sup>G12V</sup> using our short-read RNA-seq data. A subset of lung ADC primary samples have been reported to have “quasi-WT” status, which represents tumors with low S34F:WT mRNA ratios, but unchanged absolute *U2AF1*<sup>S34F</sup> or total *U2AF1* mRNA levels (Fei et al. 2016). These quasi-WT S34F:WT mRNA ratios range from 0.27-0.31. We found that the *U2AF1*<sup>S34F</sup> mRNA fraction in *U2AF1*<sup>S34F</sup> + *KRAS*<sup>G12V</sup> cells falls within the range of 0.26-0.37 (Fig 1D). This suggests that the presence of *KRAS*<sup>G12V</sup> may suppress *U2AF1*<sup>S34F</sup> expression.

We then examined the mutational status of typical-S34F and quasi-WT samples from the TCGA lung ADC cohort studied by Fei et al. Consistent with our hypothesis that mutant *KRAS* suppresses *U2AF1*<sup>S34F</sup> expression signature, we found that quasi-WT patients carry a higher proportion of *KRAS* mutations (3/4 patients) than typical-S34F patients (5/9 patients) (Fig 1E). Together, these results support the hypothesis that *KRAS*<sup>G12V</sup> suppresses the *U2AF1*<sup>S34F</sup> transcriptomic signature.

We also found variation in the *U2AF1*<sup>S34F</sup> mRNA ratios between isogenic clones (Supp 1D). The *U2AF1*<sup>S34F</sup> mRNA fraction of clone 1 *U2AF1*<sup>S34F</sup> + *KRAS*<sup>G12V</sup> ranged from 0.31-0.37, while the *U2AF1*<sup>S34F</sup> mRNA fraction of clone 2 *U2AF1*<sup>S34F</sup> + *KRAS*<sup>G12V</sup> ranged from 0.26-0.28. Although clone 1 ranged higher in *U2AF1*<sup>S34F</sup>

mRNA ratio than clone 2, it was still well within the typical-S34F range reported by Fei et al. (0.43 or above).

Co-occurring *U2AF1*<sup>S34E</sup> and *KRAS*<sup>G12V</sup> produces unique splicing events distributions, while *U2AF1*<sup>S34E</sup> alone increases splicing in stress granule protein genes

As *U2AF1*<sup>S34F</sup> and oncogenic *KRAS* are both known to alter splicing (Graubert et al., 2011; Okeyo-Uwuor et al., 2015; Fei et al., 2016; Yip et al., 2017; Palangat et al., 2019; Esfahani et al., 2019; Soulette et al., 2023; Lo et al., 2022), we next sought to understand the effect of co-occurring *U2AF1*<sup>S34F</sup> and *KRAS*<sup>G12V</sup> on differential splicing. First, we examined the mutations' global effects on splicing.

We detected and quantified splicing events from short-read data using JuncBASE (Brooks et al., 2011). Interestingly, we discovered that splicing was greatly attenuated in clone 1, compared to clone 2 (Supp 3, A-F). This finding was consistent with our gene expression results, which showed variations in gene expression between clones of similar genotypes (Supp 2A). Here, we show splicing results from clone 2. As expected, *U2AF1*<sup>S34F</sup> + *LacZ* HBEC3kts exhibited the most changes in differentially spliced genes, compared to *U2AF1*<sup>WT</sup> + *LacZ* (Fig 2A). *U2AF1*<sup>WT</sup> + *KRAS*<sup>G12V</sup> HBEC3kts displayed the lowest amount of differentially spliced genes (Fig 2B), while *U2AF1*<sup>S34F</sup> + *KRAS*<sup>G12V</sup> HBEC3kts displayed an intermediate number (Fig 2C).

Consistent with the effect that  $KRAS^{G12V}$  has on expression of  $U2AF1^{S34F}$ , these results suggest that  $KRAS^{G12V}$  suppresses the splicing changes mediated by  $U2AF1^{S34F}$ .

We next compared the categories of splicing events that were significantly different (adjusted  $p < 0.25$  and  $\Delta\text{PSI} \geq 10$ ) between  $U2AF1^{S34F} + LacZ$  and  $U2AF1^{WT} + LacZ$ , and  $U2AF1^{S34F} + KRAS^{G12V}$  and  $U2AF1^{WT} + LacZ$  in clone 2. JuncBASE categorizes splicing events in eight different categories: cassette exon, mutually exclusive exon, coordinate cassette exons, alternative 5' splice site, alternative 3' splice site, alternative first exon, alternative last exon, and retained intron (Fig 2D). We found that co-occurring  $U2AF1^{S34F}$  and  $KRAS^{G12V}$  mutations had a similar fraction of cassette exon events, a splicing event type characteristic of  $U2AF1^{S34F}$  ([Graubert et al., 2011](#); [Okeyo-Uwuor et al., 2015](#); [Fei et al., 2016](#); [Yip et al., 2017](#); [Palangat et al., 2019](#); [Esfahani et al., 2019](#); [Soulette et al., 2023](#)). In contrast, the fraction of alternative first exon events was increased and the fraction of alternative last exon events was decreased in this cell line. Interestingly, the alternative first exon events in  $U2AF1^{S34F} + KRAS^{G12V}$  HBEC3kts appeared to be at an intermediate fraction between those in  $U2AF1^{S34F} + LacZ$  and  $U2AF1^{WT} + KRAS^{G12V}$  cells. Similar to gene expression enrichment patterns, we hypothesize that the effects of co-occurring  $U2AF1^{S34F}$  and  $KRAS^{G12V}$  mutations on splicing antagonize with each other to create intermediate splicing events proportions.

To examine potential biological pathways impacted by the differentially spliced genes, we performed GSEA on differentially spliced genes. Although a less stringent threshold was used for creating the ranked gene list for this analysis (adjusted  $p < 0.25$ ), too few splicing events passed the threshold for significance for GSEA analysis for clone 1 ( $n = 75$  genes pass adjusted  $p$  value  $< 0.25$  for clone 1  $U2AF1^{S34F} + LacZ$ , compared to  $n = 1455$  genes for clone 2). Because clone 1 appeared to have a dearth of splicing events to analyze and different global gene expression patterns (Supp 1A), we continued remaining analyses with clone 2. In contrast to the differential gene expression GSEA results, far fewer Hallmark gene sets were significantly differentially enriched amongst our genotypes. Only one gene set, p53 Pathway, was found to be significantly enriched, and was only present in the  $U2AF1^{S34F} + KRAS^{G12V}$  comparison. This finding highlights the non-overlapping roles of gene expression and splicing on transcripts belonging to certain pathways in the cell.

A recently appreciated role of  $U2AF1^{S34F}$  is its ability to confer resistance to the effects of stress.  $U2AF1^{S34F}$  alone has been shown to increase cell proliferation following ionizing radiation exposure (Palangat et al. 2019).  $U2AF1^{S34F}$  has also recently been found to confer altered splicing and binding to stress granule gene sets in an MDS cell line (Biancon et al. 2022). Stress granules are RNA-protein condensates that may help cells survive stress and can help cancer cells resist chemotherapy (Chen et al. 2018). In the MDS line, altered splicing in stress granule

genes was associated with enhanced viability under sodium arsenite, a chemical agent of oxidative stress (Biancon et al. 2022).

We hypothesized *U2AF1<sup>S34F</sup>* may also be altering stress response through aberrant splicing in our cell lines. Although we did not find significant enrichment in stress granule gene sets after overlapping transcripts differentially spliced and differentially bound by U2AF1, we did observe increased splicing in stress granule protein genes in *U2AF1<sup>S34F</sup> + LacZ*, compared to other genotypes (Fig 3E). Interestingly, gene expression changes in this gene set did not reveal as strong of a difference among the genotypes, revealing the utility of integrating multiple kinds of RNA sequence analysis for a more complete view of potentially altered pathways (Supp 3A).

#### Co-occurring *U2AF1<sup>S34F</sup>* and *KRAS<sup>G12V</sup>* mutations increase oncogenic potential and proliferation

Following this transcriptomic profiling, we sought to understand the functional consequences of differentially expressed gene sets. We first sought to explore gene sets with similar enrichment patterns in both *U2AF1<sup>S34F</sup> + LacZ* and *U2AF1<sup>S34F</sup> + KRAS<sup>G12V</sup>* HBEC3kts, as they indicated *U2AF1<sup>S34F</sup>*-specific effects which persisted when *KRAS<sup>G12V</sup>* was present. One category that fit this criteria was the inflammatory pathway gene sets (Complement, IL2 STAT5 Signaling, IL6 JAK STAT3 Signaling,

Inflammatory Response, TNFA Signaling Via NFkB, Interferon Alpha Response, and Interferon Gamma Response). Oncogenic Ras has been found to increase production of cytokines such as IL-6 in multiple cell types (Ancrile et al. 2007; Liu et al. 2021). To probe how these pathways are altered in our *U2AF1<sup>S34F</sup> + KRAS<sup>G12V</sup>* cell line, we measured inflammatory cytokine production in our HBEC3kt genotypes from clone 2. For most cytokines tested, we observe that *U2AF1<sup>WT</sup> + KRAS<sup>G12V</sup>* HBEC3kts secrete the highest levels of inflammatory cytokines. Co-occurrence of *U2AF1<sup>S34F</sup>* with *KRAS<sup>G12V</sup>* suppresses the levels of secreted cytokines IL-1 $\beta$ , IL-6, IL-8, TNF $\alpha$ , GM-CSF, and IFN $\gamma$  (Figure 3A, Table S1). High levels of IL-1 $\beta$ , TNF $\alpha$ , GM-CSF, and IFN $\gamma$  have been shown to promote antitumor activity in animal models (Lee et al. 2019; Hoving et al. 2006; Kim et al. 2019; Gerber et al. 2013). We hypothesized that *U2AF1<sup>S34F</sup>* creates a microenvironment conducive to tumor growth by bringing cytokine secretion down to an intermediate level in *KRAS<sup>G12V</sup>*-mutant cells. We note that our cytokine results are inconsistent with previous work done on *U2AF1<sup>S34F</sup>* HBEC3kts, which showed that *U2AF1<sup>S34F</sup>* increases the secretion of cytokines such as IL-8 (Palangat et al. 2019). However, the clonal background of the cells used in the aforementioned study was not reported, and it is possible that different steady-state cytokine secretion levels may be present in HBEC3kts from different isogenic clones.

HBEC3kts with *U2AF1<sup>S34F</sup>* also exhibited expression in gene sets related to cell cycle progression (E2F Targets, G2M checkpoint). To understand how these gene expression differences translate to altered phenotype, we next asked how *U2AF1<sup>S34F</sup>*

and  $KRAS^{G12V}$  mutations affect proliferative potential. We stained HBEC3kts with EdU and phospho histone-H3 (PHH3) (Figures 3B-C, Table S3) to measure the proportion of cells undergoing S-phase and M-phase, respectively (Flomerfelt et al. 2016; Kim et al. 2017). Previous models with  $U2AF1^{S34F}$  have found that  $U2AF1^{S34F}$  by itself suppresses growth phenotypes such as proliferation and colony-forming potential (Fei et al. 2019; Zhu et al. 2021). Consistent with these findings, we observed lower normalized EdU and PHH3 intensity in  $U2AF1^{S34F} + LacZ$  HBEC3kts compared to  $U2AF1^{WT} + LacZ$ . However, when  $KRAS^{G12V}$  and  $U2AF1^{S34F}$  co-occur, we observe increased proliferation compared to  $U2AF1^{S34F}$  by itself. Notably, M-phase staining in  $U2AF1^{S34F} + KRAS^{G12V}$  HBEC3kts was elevated to above  $U2AF1^{WT} + LacZ$  levels (Figure 3C), indicating that  $KRAS^{G12V}$  confers increased mitosis in  $U2AF1^{S34F}$ -mutant cells.

Mutant  $KRAS$ , including  $KRAS^{G12V}$  is known to induce oncogenic phenotypes, such as anchorage-independent growth (Muñoz-Maldonado et al. 2019). Due to the enhanced proliferation in  $U2AF1^{S34F} + KRAS^{G12V}$  cells and the unique negative enrichment score in the KRAS Signaling Down gene set observed in this line, we hypothesized that co-occurring  $U2AF1^{S34F}$  and  $KRAS^{G12V}$  would alter anchorage-independent growth as well. We cultured clone 2 HBEC3kts of the four genotypes on low attachment plates and measured viability. Consistent with the xenograft results,  $U2AF1^{S34F} + KRAS^{G12V}$  HBEC3kts survive anchorage-independent growth conditions better than other genotypes over 10 days in low-attachment conditions (Figure 3D, Table S1).

Finally, we sought to understand how  $U2AF1^{S34F}$  and  $KRAS^{G12V}$  co-occurrence in HBEC3kts impacts the ability of cells to form tumors *in vivo*. Cells from the four genotypes from clone 2 were injected into immunodeficient (NSG) mice. We find that  $U2AF1^{S34F} + KRAS^{G12V}$  HBEC3kts formed more tumors than HBEC3kts with either mutation alone (Figure 3E). This suggests that co-occurring  $U2AF1^{S34F}$  and  $KRAS^{G12V}$  mutations synergize to transform HBEC3kts cells *in vivo*.

Similar to the heterogeneity observed in  $KRAS$  gene expression, we also observed phenotypic heterogeneity between our isogenic clones. When we asked how the tumor formation differed between cells from clone 1 and clone 2, we found that clone 2  $U2AF1^{S34F} + KRAS^{G12V}$  HBEC3kts formed tumors more frequently than clone 1 (Figure S3 B). In contrast, the one tumor formed by a  $U2AF1^{WT} + KRAS^{G12V}$  was from clone 1. Similarly, when we examined viability in low-attachment conditions, we observed that the  $U2AF1^{S34F} + KRAS^{G12V}$  cell line from clone 2 was more viable in low attachment conditions at earlier timepoints than clone 1 (Figure S3 C, Table S1). Importantly, no tumors were formed with  $U2AF1^{S34F}$ , suggesting that  $U2AF1^{S34F}$  alone is insufficient for *in vivo* transformation. These results are consistent with the hypothesis that inconsistent  $KRAS^{G12V}$  integration in clone 1 impacts oncogenic potential unpredictably.

We also assayed for other cancer hallmarks in clone 2, such as the long-term ability to survive and proliferate into colonies, which is a marker of cancer stemness and can be assessed with a clonogenicity or colony-forming assay (Esquer et al. 2020).

Interestingly, co-occurring *KRAS*<sup>G12V</sup> and *U2AF1*<sup>S34F</sup> suppressed colony-forming potential (Figure S3 D, Table S1). These findings are consistent with previous work performed on *U2AF1*<sup>S34F</sup>-mutant cancer cell lines (Zhu et al. 2021). We also performed a wound-healing assay to assess the invasive potential of *U2AF1*<sup>S34F</sup> + *KRAS*<sup>G12V</sup> HBEC3kts (Figure S3 E, Table S1) and observed that *U2AF1*<sup>S34F</sup> decreases invasive potential in *KRAS*<sup>G12V</sup> background. Our results indicate that enhanced proliferation conferred by *U2AF1*<sup>S34F</sup> + *KRAS*<sup>G12V</sup> may work in concert with pathways outside of stemness and invasion to confer oncogenic potential.

#### Altered splicing in stress granule genes in *U2AF1*<sup>S34F</sup> HBEC3kts correlates with enhanced stress response

Stress granules are often induced by an agent of oxidative stress (Biancon et al. 2022, Lian and Gallouzi et al. 2009 ). Oxidative stress is relevant to cancer formation because it can produce mutations by causing DNA damage (Poetsch 2020). First, we examined lung ADC primary samples with and without mutations in *U2AF1* for overall mutation counts and found that the presence of *U2AF1* mutation increases mutational burden (Fig 4A). In lung ADC patients, a common source of oxidative stress is exposure to cigarette smoke. To understand the connection between oxidative

stress response and lung ADC, we next analyzed splicing alterations in patient data from TCGA and found greater numbers of splicing alterations in lung ADC patients with smoking histories, compared to never-smokers (Fig 4B).

Previous studies on *U2AF1<sup>S34F</sup>* have observed an increase in viability after exposing cells to oxidative stress like radiation and sodium arsenite, compared to *U2AF1<sup>WT</sup>* cells (Palangat et al. 2019; Biancon et al. 2022). We followed up on this line of inquiry by measuring how *U2AF1<sup>S34F</sup>* alone impacts viability in cigarette smoke concentrate (CSC) (Fig 4C). We treated *U2AF1<sup>S34F</sup>* or *U2AF1<sup>WT</sup>* HBEC3kts with CSC and measured viability after three days. *U2AF1<sup>S34F</sup>* HBEC3kts displayed higher viability than *U2AF1<sup>WT</sup>* HBEC3kts at all concentrations tested.

Together our results lead us to a model of oncogenic transformation. *U2AF1<sup>S34F</sup>* has been reported to be a truncal mutation in lung cancer and MDS ([Esfahani et al, 2019](#), [Montgomery et al, 2021](#)). We propose that *U2AF1<sup>S34F</sup>*, when present in precancerous cells, allows cells to better survive an initial onslaught of oxidative stress. The surviving cells are then more likely to persist and accumulate further mutations like *KRAS<sup>G12V</sup>*, which act synergistically with *U2AF1<sup>S34F</sup>* to alter splicing and gene expression to increase oncogenic potential (Fig 4D).

### 2.3 - Discussion

Despite its status as a recurrent mutation, the role of  $U2AF1^{S34F}$  role in lung cancer has been difficult to understand, since the mutation confers anti-proliferative and anti-invasive phenotypes when present alone in model systems. This aspect limits the ability of researchers to identify the functional role of  $U2AF1^{S34F}$  in lung cancer, and limits the use of  $U2AF1^{S34F}$  as a prognostic marker for lung ADC patients. To gain a better understanding of this mutation, we examined the role of  $U2AF1^{S34F}$  in early cancer formation in two directions: how  $U2AF1^{S34F}$  may synergize with other cancer drivers like  $KRAS^{G12V}$ , and how  $U2AF1^{S34F}$  by itself can impact stress response in the cell.

We performed transcriptome sequencing on  $U2AF1^{S34F} + KRAS^{G12V}$  HBEC3kts and revealed synergistic effects on gene expression in cancer-relevant pathways. Notably, the KRAS Signaling Down hallmark pathway becomes down regulated only in  $U2AF1^{S34F} + KRAS^{G12V}$  HBEC3kts.  $U2AF1^{S34F} + KRAS^{G12V}$  HBEC3kts also shared a suppression in inflammatory pathway gene expression and enrichment in cell cycle gene set expression observed in HBEC3kts with only  $U2AF1^{S34F}$ . Together, these gene set enrichment patterns correlated with increased *in vivo* tumor formation for  $U2AF1^{S34F} + KRAS^{G12V}$  HBEC3kts. The HBEC3kt line has previously been used to study the number of genetic alterations required to transform noncancerous

immortalized cells. Interestingly, sh-p53, overexpression of *KRAS*<sup>G12V</sup>, and overexpression of c-MYC were required to fully transform immortalized HBEC3kts (Sato et al. 2013), whereas we observe full transformation with only *KRAS*<sup>G12V</sup> and *U2AF1*<sup>S34F</sup>

Despite shared gene set enrichment patterns between *U2AF1*<sup>S34F</sup> + *KRAS*<sup>G12V</sup> and *U2AF1*<sup>S34F</sup> + *LacZ* HBEC3kts, validation of these pathways revealed different ways in which these phenotypes manifest. For instance, although cell cycle gene sets were both enriched in *U2AF1*-mutant cell lines regardless of *KRAS* status, *U2AF1*<sup>S34F</sup> + *LacZ* cells exhibited reduced proliferation, whereas the co-occurrence of *U2AF1* and *KRAS* mutations increased proliferation. In contrast, other gene set enrichment patterns translated to more consistent phenotypes. For instance, a suppression in inflammatory gene enrichment translated to a suppression of inflammatory cytokines for all HBEC3kt lines.

We also examined splicing-level changes in the transcriptome caused by *U2AF1* and *KRAS* mutation. Interestingly, we found little overlap in gene set enrichment between differentially expressed and differentially spliced genes, highlighting the use of applying multiple kinds of RNA-seq analysis to assess the synergistic impact of mutations. In our case, increased alteration in stress granule protein genes unique to *U2AF1*<sup>S34F</sup> + *LacZ* HBEC3kts was found only in the splicing analysis and not in the differential gene expression results. We also observed synergy in the global usage of

splicing event types in cells with co-occurring *U2AF1* and *KRAS* mutations. Namely, *U2AF1* and *KRAS* co-occurrence increases the proportion of alternative first exon usage compared to *U2AF1*<sup>S34F</sup> + *LacZ* HBEC3kts, while cassette exon event proportion remains the same.

When we followed up on this splicing analysis by quantifying cellular response to stress, we found that *U2AF1*<sup>S34F</sup> confers resistance to exposure to cigarette smoke concentrate. Our work leads us to a model of *U2AF1*<sup>S34F</sup> conferring oncogenic potential that is dependent on the presence of environmental stress. When stress is present, we propose that *U2AF1*<sup>S34F</sup> confers survival advantage over *U2AF1*<sup>WT</sup> cells, which allows for continued proliferation and accumulation of stronger oncogenic drivers like *KRAS*<sup>G12V</sup>, which synergize with *U2AF1*<sup>S34F</sup> to increase oncogenic potential.

In analyzing sequence data from our replicates, we observed variations between isogenic clones that appeared to impact oncogenic potential. First, the expression of oncogenic *KRAS* was inconsistent in clone 1. Others who have worked with HBEC3kts transfected with *KRAS*<sup>G12V</sup> have noted that *KRAS* expression levels correlate with increased oncogenic potential (Sato et al. 2013). Second, we observed that clone 1 has considerably fewer significant splicing events compared to clone 2, even in lines without *KRAS*<sup>G12V</sup>. This finding demonstrates the utility of considering

both aberrant splicing and gene expression when making predictions regarding oncogenic potential.

More work is left to be done to understand the role of splicing in stress response. In yeast, intron retention (IR) has been linked to fitness advantage in the presence of environmental stressors like starvation ([Lukačičin et al. 2022](#)). Although we did not find evidence of altered intron retention in our short-read analysis, previous work from this lab has shown *U2AF1<sup>S34F</sup>* increasing IR from long-read data (Soulette et al. 2023), leaving an interesting avenue to pursue as long-read sequencing technologies improve.

## 2.4 - Methods

### EXPERIMENTAL MODEL DETAILS

#### Mouse models

*NOD.Cg-Prkdc<sup>scid</sup>Il2rg<sup>tm1Wjl</sup>/SzJ* (NSG) (stock #005557) mice were purchased from The Jackson Laboratory and bred at the University of California Santa Cruz (UCSC). All mice used for this study were maintained at the UCSC Animal Facility in accordance with the guidelines set forth by UCSC and the Institutional Animal Care and Use Committee (Protocol number SIKAS2010).

## METHOD DETAILS

Cell lines: Host HBEC3kt cell lines homozygous for wildtype *U2AF1* (*U2AF1* WT) and heterozygous with one copy of *U2AF1*<sup>S34F</sup> at the endogenous locus were obtained as a gift from the laboratory of Harold Varmus (Cancer Biology Section, Cancer Genetics Branch, National Human Genome Research Institute, Bethesda, United States of America and Department of Medicine, Meyer Cancer Center, Weill Cornell Medicine, New York, United States Of America) and maintained as described by Fei et al, PLoS Genet 12(10): e1006384. These cell lines were used for lentiviral transduction and blasticidin selection to generate a stable expression of *KRAS*<sup>G12V</sup> or *LacZ* using plasmids obtained as gifts from the laboratory of John D Minna (Hamon Center for Therapeutic Oncology Research, The University of Texas Southwestern Medical Center) and used as described in (Sato, *et al* Mol Cancer Res (2013) 11 (6): 638–650).

A total of 8 cell lines were created: (1) *U2AF1*<sup>WT</sup> + *LacZ* clone 1, (2) *U2AF1*<sup>WT</sup> + *LacZ* clone 2, (3) *U2AF1*<sup>S34F</sup> + *LacZ* clone 1, (4) *U2AF1*<sup>S34F</sup> + *LacZ* clone 2, (5) *U2AF1*<sup>WT</sup> + *KRAS*<sup>G12V</sup> clone 1, (6) *U2AF1*<sup>WT</sup> + *KRAS*<sup>G12V</sup> clone 2, (7) *U2AF1*<sup>S34F</sup> + *KRAS*<sup>G12V</sup> clone 1, (8) *U2AF1*<sup>S34F</sup> + *KRAS*<sup>G12V</sup> clone 2 (Supp 1A). These cell lines were tested for mycoplasma (IDEXX).

Mouse xenograft of HBEC3kts: Clone 1 and clone 2 HBEC3kt lines were cultured as previously described. Cells were allowed to recover from cold storage in liquid nitrogen after seeding for one passage in a T-25 flask. Cells were passaged to a 10cm plate, then to a final 15cm plate, and allowed to grow to 80% confluency. At 80% confluency, the media in the 15cm plates were aspirated and cells were washed twice with DPBS. To suspend cells for injection, cells were trypsinized with standard protocols ([Ramirez et al. 2004](#)) and live cell counts were assessed by Trypan Blue staining. For each cell line, 9 million cells were resuspended in Keratinocyte SFM media containing 40% Matrigel and subcutaneously injected into the fourth abdominal fat pads on both sides of male NSG mice. 2-5 million cells were injected at each site in 100 uL media + Matrigel (5 million in 1st xenograft experiment, 2 million in 2nd). Mice were monitored every week for tumor growth. All mice were euthanized if tumor growth reached end point (1500 mm<sup>3</sup>), the tumors were ulcerated, or mice showed signs of distress. Tumor size was measured using digital calipers. A total of 4 female and 14 male mice were used.

RNA extraction of HBEC3kts: For RNA sequencing of HBEC3ktlines, cells were allowed to recover from cold storage in liquid nitrogen after seeding for one passage in a T-25 flask. Cells were passaged at 70-80% confluency to ensure they remained in log-phase growth into a 10cm plate. Once cells in the 10cm plates reached 70-80% confluency, they were washed twice in ice cold DPBS then collected in Tri-reagent

for storage at -80°C until the bulk RNA was extracted using Direct-Zol RNA Miniprep Kit (Cat#R2050, Zymo Research).

Short-read RNA-Seq of  $U2AF1^{WT} + LacZ$ ,  $U2AF1^{S34E} + LacZ$ ,  $U2AF1^{WT} + KRAS^{G12V}$ , and  $U2AF1^{S34E} + KRAS^{G12V}$  HBEC3kts: For Illumina sequencing, n=3 10cm plates per HBEC3kt genotype of both clones, for a combined n=6 per genotype, were cultured for RNA extraction as described above. Concentrations of purified RNA in nuclease-free water were determined by Nanodrop-2000 Spectrophotometer and Qubit RNA BR Assay (ThermoFisher Scientific). RINe numbers ranging from 7.8-10 were determined by TapeStation 4150 RNA ScreenTape Analysis (Agilent Technologies) before sending RNA to UC Davis DNA Technologies and Expression Analysis Core Laboratory for poly-A strand specific library preparation to obtain 60 million paired end read pairs by NovaSeq S4 (PE150) sequencing.

Viability assay: HBEC3kts of differing genotypes were seeded in a 96 well-plate in triplicate and grown in supplemented KSM. At multiple time points, (0, 4, and 6 days), cells were rinsed twice with DPBS, CellTiter-Glo (Promega) reagent was added, and transferred to white opaque 96 well-plates to measure luminescence. Luminescence at each timepoint was quantified using the VarioSkan platereader and normalized to the average relative luminescence units (RLU) of the 0 day timepoint.

Western blot analysis: Cell lines were cultured to 85% confluency in 10cm plates.

After preparation of protein lysates in 1ml of RIPA buffer supplemented with protease inhibitor cocktail (manufacturer's #5892970001, Roche Molecular Systems, Inc, USA) proteins were denatured using standard denaturation techniques in beta mercaptoethanol laemmli buffer, and 15ug of denatured protein lysate was separated on a 4-15% Mini-Protean TGX Precast Protein Gel (Cat#4561086, Bio-Rad Laboratories, Inc. USA). After transfer to 0.2 um PVDF membrane using TransBlot Turbo Transfer system ( Cat# 1704272, BioRad Laboratories Inc., USA), membranes were incubated shaking at room temperature in 5% milk block in 1x PBST followed by incubation in KRAS<sup>G12V</sup> primary antibody at 1:250 dilution (Cat#14412) and B actin conjugated to HRP at 1:500 dilution (Cat#sc-47778 HRP) in milk block overnight on an orbital shaker at 4°C. The next day, blots were washed in PBST and incubated with secondary HRP-conjugated antibody (Cat#7074) at 1:1000 dilution at room temperature for 1h. After washing in PBST, bands were detected using WesternSure PREMIUM Chemiluminescent substrate (Cat# 926-95000, Li-COR Biosciences, USA) and visualized on a C-Digit Blot Scanner (Li-COR Biosciences, USA).

Secreted cytokine analysis: Growth triplicates of each cell line were seeded in 6 well plates and cultured with standard protocols described above to 85% confluency. Conditioned media (3 ml) above the cells was collected and cell debris spun out at 3000 x g for 10 mins at 4°C and supernatant was stored in -80°C before sending to

Eve Technologies (Calgary, Canada) for Human High Sensitivity T-Cell Discovery Array 14-plex (HDHSTC14) Elisa assay. Data was plotted and significance was calculated with a Mann-Whitney test on GraphPad Prism.

Proliferation assays: Cell staining was performed at UCSC's Chemical Screening Center, using the BioTek EL406 with peri/syringe/wash modules for automated washing and dispensing of reagents. Cells were cultured as previously described in optical-bottom black opaque 96 well-plates (Corning 3904). The plate was taken to the Chemical Screening Center and incubated with EdU for 1 hour at 37°C and 5% CO<sub>2</sub>. Following EdU incorporation, cells were fixed with 5% formaldehyde (Fisher, F79-500) in basal media (ATCC, PCS-300-030) for 30 minutes at 37°C and 5% CO<sub>2</sub>. Cells were blocked with 2% BSA in PBST for 20-60min in the dark at room temperature. Following blocking, click reagent (15ml 100mM Tris pH7.4, 0.6ml 100mM CuSO<sub>4</sub>, 155.5ul 200mg/ml Na Ascorbate, 15.5ul 10mg/ml Rhodamine-Azide) was added to the cells to visualize EdU incorporation and cells were incubated in the dark for 1h at room temperature. Following azide incorporation, cells were stained with Hoechst (2.5uL in 2% BSA) to visualize nuclei and incubated in the dark for 2h at room temperature. Cells were then incubated with a primary antibody for Phospho-Histone H3 (Ser10) Recombinant Rabbit Monoclonal Antibody (9H12L10) (Invitrogen# 701258) at 1:5000 dilution in PBS and BSA, followed by incubation in a chicken anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647 secondary antibody (Invitrogen A-21443) at 1:1000 dilution.

Immunofluorescent imaging: Imaging and quantification of immunofluorescent signal was performed with the Perkin Elmer Opera Phenix Plus and Harmony bioinformatics software at UCSC's Chemical Screening Center. Plotting and statistical analysis was performed on Python.v3.7.7 and Jupyter notebook v6.3.0. Significance was calculated using a Kruskal-Wallis test with Dunn's multiple comparisons test.

Growth in Low Attachment (GILA): Cells were grown to 85% confluency on regular tissue culture-treated 6-well plates, harvested by trypsinization, filtered over a nylon 70um mesh and seeded in triplicate in KSFM media at 2500 cells per well in Ultra-low attachment 96 well plates (Manufacturers # 3474, Corning, USA) and time points were collected for viability assays over an 8-day period. An early time point was collected at the time of seeding and used for normalization. Viability was assayed using CellTiterGlo according to manufacturer's instructions (Cat#G7570, Promega, USA) and luminescence was measured on a VarioScan LUX plate reader (ThermoFisher, USA). Significance was calculated using a Kruskal-Wallis test with Dunn's multiple comparisons test using GraphPad Prism.

Clonogenicity assay: Colony formation was assessed by seeding the cells in triplicate at 200 cells per 10cm plate and cultured under normal conditions except that media

was changes only twice over a 10 day period. Cells were fixed in 100% methanol for 20 mins and stained in 0.5% Crystal Violet in 25% methanol for 5 mins before drying and photographing. Colonies of approximately 2mm or larger were counted in 4 separate quadrants of each plate. Significance was calculated using a Kruskal-Wallis test with Dunn's multiple comparisons test using GraphPad Prism.

Wound healing assay: HBEC3kts were seeded in 6-well plates. A 200uL pipettor and filter tip was used to create the wound in a confluent monolayer of cells. The wound was imaged at 0 and 3 hours. The number of cells that had migrated into the wound between the two time points was counted. Significance was calculated using a Mann-Whitney test using GraphPad Prism.

Cigarette smoke treatment: CSC was obtained from Murty Pharmaceuticals (Cat#nc1560725). HBEC3kts were seeded in a 96 well-plate and grown to 50% confluency. They were then treated with 0, 15, 60, and 120ug/mL CSC for three days. Following treatment, the cells were washed twice with DPBS and assayed using CelTiterGlo as described above. Luminescence was normalized to the 0ug/mL control. Significance was calculated using a Mann-Whitney test using GraphPad Prism.

Nutrient depletion assay: Cells were seeded in a 96 well-plate and grown to 80% confluency. Supplemented media was removed and replaced with basal medium

(ATCC, PCS-300-030) for each timepoint. After treatment, cells were stained for EdU as described above.

RNA-Seq Data Analysis: Raw sequencing reads in fastq files were aligned to a version of the human genome hg38 that has a region of repeats masked to make the U2AF1 locus alignable ([Miller et al., 2022](#)), using STAR.v2.7.3a (Dobin et al. 2013) with the parameters `--outSAMtype BAM SortedByCoordinate --twopassMode Basic --quantMode GeneCounts --bamRemoveDuplicatesType UniqueIdentical` and the Gencode v33 primary assembly gtf file. Aligned BAM files were indexed with Samtools.v1.10 (Danecek et al. 2021). Mapped reads in BAM files were counted with HTSeq.v0.12.4 (Anders et al. 2015) for all the annotated genes in gencode.v33.primary\_assembly.annotation.gtf with `-stranded = reverse` and `nonunique=none` parameters.

*U2AF1*<sup>S34F</sup> mRNA ratio: Aligned reads from clone 1 and clone 2 HBEC3kts were loaded onto the Integrative Genomics Viewer (Robinson et al. 2011) (IGV) at the *U2AF1*<sup>S34F</sup> mutational locus. The fraction of A (mutant) nucleotides at this locus obtained from IGV was plotted. Significance was calculated with a Mann-Whitney test on GraphPad Prism.

Differential expression analysis: Differential expression analysis was performed with the DESeq2 v1.40.2 on R v4.3.1 (Love et al. 2014) on aligned RNA sequences from

clone 1 and clone 2 of our HBEC3kt lines. Gene counts were normalized and a LRT calculation was performed to account for batch differences between samples from clone 1 and clone 2. statistical analysis was performed on expression differences in the following pairwise comparisons:  $U2AF1^{S34F} + LacZ$  vs.  $U2AF1^{WT} + LacZ$ ,  $U2AF1^{WT} + KRAS^{G12V}$  vs.  $U2AF1^{WT} + LacZ$ , and  $U2AF1^{S34F} + KRAS^{G12V}$  vs.  $U2AF1^{WT} + LacZ$ . The log<sub>2</sub> fold change (FC) values from each comparison was filtered for adjusted p-value (padj) < 0.05, and the filtered log<sub>2</sub> FC values along with gene names were exported as a .RNK file for gene set enrichment analysis using a custom Python script that is available upon request.

Differential splicing analysis on HBEC3kts: Aligned .BAM files were trimmed to remove nonstandard chromosomes with samtools v.1.13. Splice junction usage was counted using juncBASE (Brooks et al. 2011) using the parameters outlined in the .sh file provided in the resources table. JuncBASE count files were statistically analyzed with the compareSampleSets.py module, using the following commands: python /JuncBASE/compareSampleSets.py --in\_prefix /path/to/juncbase\_tables/ --all\_psi\_output PSI.txt --mt\_correction BH --which\_test t-test --thresh 10 --delta\_thresh 5.0 --sample\_set1 wt\_rep1,wt\_rep2,wt\_rep3 --sample\_set2 mut\_rep1,mu\_rep2,mu\_rep3 --as\_only python /JuncBASE/compareSampleSets.py --in\_prefix /path/to/juncbase\_tables/ --all\_psi\_output output.txt --mt\_correction BH --which\_test t-test --thresh 10 --delta\_thresh 10.0 --sample\_set1 wt\_rep1,wt\_rep2,wt\_rep3 --sample\_set2 mut\_rep1,mu\_rep2,mu\_rep3. Then,

redundant splicing events were filtered out using the JuncBASE script `makeNonRedundantAS.py`. We also filtered out junction-only alternative acceptor and alternative donor events, as these events have less read support than other categories. Additionally, we excluded intron retention events that consisted of novel junctions. To generate the .RNK file for splicing changes, we took the absolute value of the deltaPSI values produced by `compareSampleSets.py` and filtered them for  $\text{padj} < 0.25$ . Unlike differential gene expression, multiple splicing events are possible for a given gene. To convert our results into a .RNK file readable by GSEA, we handled duplicate deltaPSI entries by keeping the entry with the highest  $\text{abs}(\text{deltaPSI})$  value. The deltaPSI values and gene names were then exported as a .RNK file for GSEA.

To compare splicing event type distributions between the genotypes, splicing events were filtered for  $\text{padj} < 0.25$  and  $\text{abs}(\text{deltaPSI}) > 10$ . Statistical differences between splicing event distributions were calculated using a Fisher's exact test and row-wise Fisher's exact test with R's `rstatix` library.

Stress granule protein gene analysis: Significant ( $\text{padj} < 0.25$ ) splicing events were overlapped with genes from *U2AF1<sup>S34F</sup>* HBEC3kt samples that have differential U2AF1 binding (Palangat et al. 2019). The absolute deltaPSI for each genotype comparison was then plotted as a heatmap. NA values corresponding to genes without significant change in splicing from certain pairwise comparisons did not pass the filters were replaced with 0 for plotting.

GSEA: gene set enrichment analysis was performed by inputting .RNK files generated from differential expression or splicing analysis into GSEAPreranked on the GSEA v.4.3.2 software (Subramanian et al. 2005). The “Collapse/Remap to gene symbols” option was set to “No\_Collapse” and default settings were used for the remaining options. The positive and negative GSEA output tables for each gene set were combined, and the normalized enrichment scores (NES) were filtered for FDR q-value < 0.25 and nominal p-value < 0.05. Filtered NES and the identities of their corresponding gene sets were plotted in heatmaps using Python.v3.7.7 and Jupyter notebook v6.3.0. NA values corresponding to gene sets where NES from certain pairwise comparisons did not pass the filters were replaced with 0 for plotting.

#### Lung ADC primary sample RNA-seq analysis

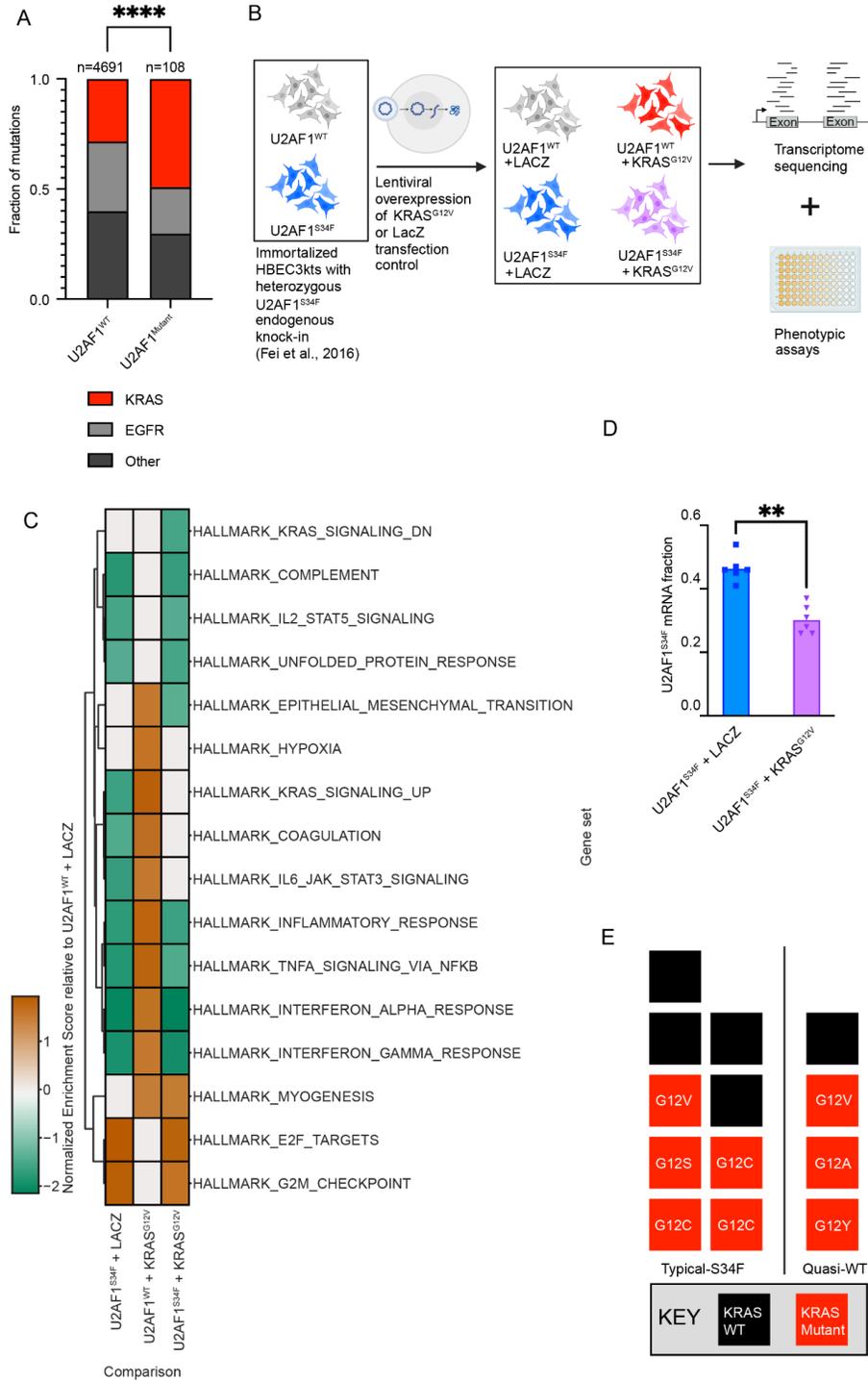
*U2AF1*<sup>S34F</sup> and *KRAS*<sup>G12V</sup> co-occurrence: Lung ADC patient mutational statuses was obtained from cBioPortal (Cerami et al. 2012). Overlapping studies as well as the TSP Nature, 2008 were excluded from analysis. Co-occurrence p value was obtained from cBioPortal’s Mutual Exclusivity analysis and the mutational status of patients were plotted with GraphPad Prism.

*U2AF1*<sup>S34F</sup> and smoking history splicing alteration status: Lung ADC primary sample RNA-seq data was obtained from TCGA and smoking status for the patients was obtained from Campbell et al. 2016. Splicing alteration status was obtained by running juncBASE to compare lung ADC against matched normal tissues.

"jcn\_only", novel intron retention events, and events where more than 25% of the samples were missing data were excluded. Splicing event PSI medians and interquartile range (IQR) were computed from samples with a smoking designation from Campbell et. al. A splicing event was considered altered in an individual sample if the PSI - median was more than 1.5xIQR for that event, and the delta PSI was more than 10% from the median. P-values were calculated using a Wilcoxon ranksum unpaired test.

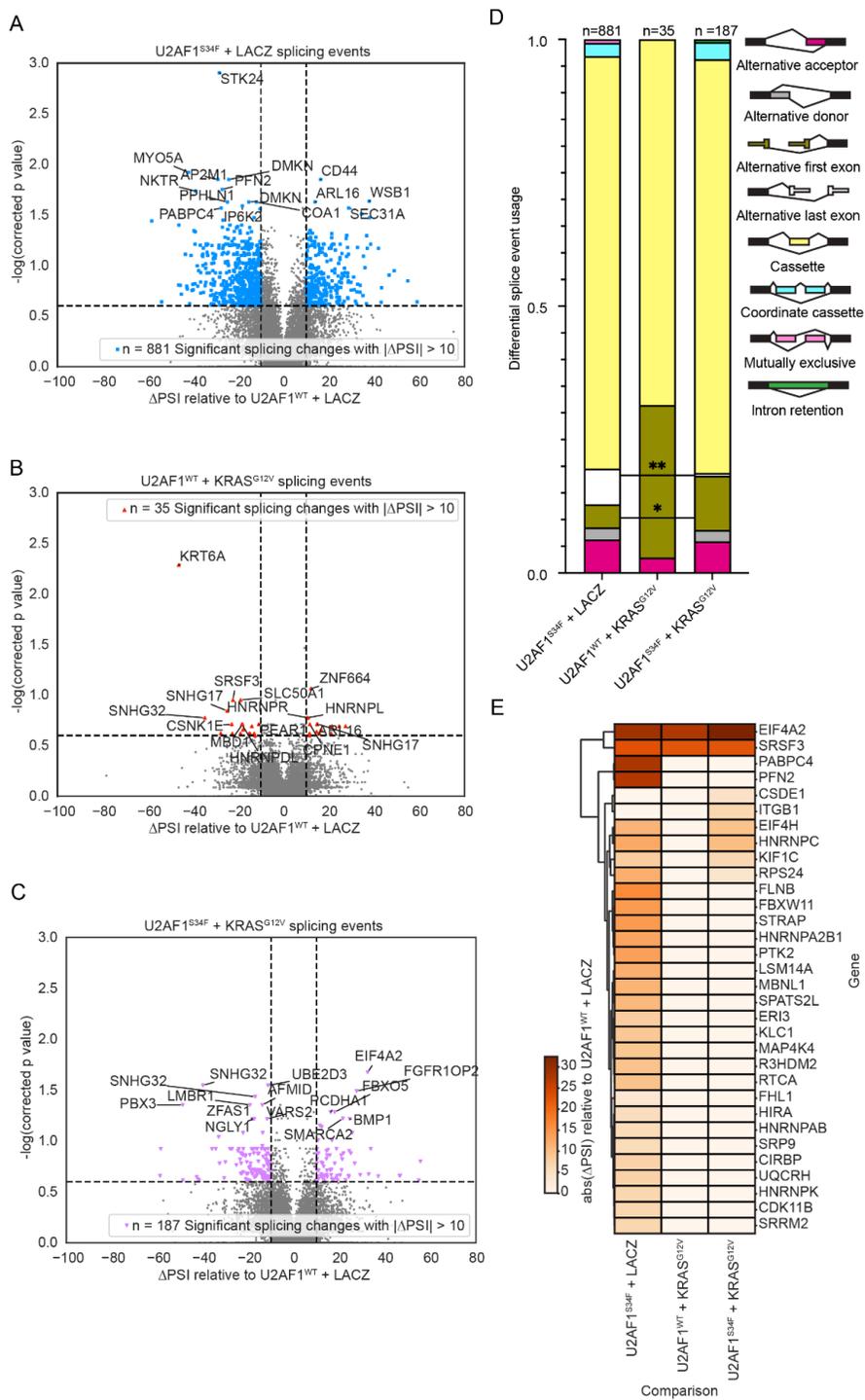
## 2.5 - Figures

FIGURE 1



**Figure 1** *KRAS*<sup>G12V</sup> suppresses the effect of *U2AF1*<sup>S34F</sup> on the transcriptome while altering gene expression in oncogenic pathways. **(A)** Distribution of *KRAS*, *EGFR*, and other mutations in lung ADC patients with and without mutations in *U2AF1*. **(B)** Experimental pipeline for study. Illumina RNA sequencing was performed on HBEC3kt lines with *U2AF1*<sup>S34F</sup> alone, *KRAS*<sup>G12V</sup> alone, co-occurring *U2AF1*<sup>S34F</sup> and *KRAS*<sup>G12V</sup>, and a wild-type control. Phenotypic assays for oncogenic phenotypes were also performed. **(C)** Heatmap of gene enrichment scores for gene sets differentially expressed between each genotype and the wild-type control. **(D)** *U2AF1*<sup>S34F</sup> mRNA fraction in HBEC3kt lines with differing mutational backgrounds. Bars represent mean *U2AF1*<sup>S34F</sup> mRNA fraction. **(E)** Distribution of *KRAS* mutations in lung ADC patients observed to display quasi-WT or typical-S34F expression patterns. Each box represents a single patient. \*\* P ≤ 0.01, \*\*\*\* P ≤ 0.0001. See also Figures S1-S2, Table S1, Table S2.

FIGURE 2



**Figure 2 Co-occurring *U2AF1*<sup>S34F</sup> and *KRAS*<sup>G12V</sup> produces unique splicing events distributions, while *U2AF1*<sup>S34F</sup> alone increases splicing in stress granule protein genes. (A-C)** Volcano plots of differentially spliced genes compared to the wild-type control in *U2AF1*<sup>S34F</sup> + *LacZ*, *U2AF1*<sup>WT</sup> + *KRAS*<sup>G12V</sup>, and *U2AF1*<sup>S34F</sup> + *KRAS*<sup>G12V</sup> HBEC3kts. Colored dots and numbers displayed in inset represent genes with significant splicing changes (adjusted  $p < 0.25$ ) of magnitude greater than 10% ( $|\Delta\text{PSI}| \geq 10$ ). The top 15 most significantly differentially spliced genes were labeled. **(D)** Distribution of splicing event types as categorized by JuncBASE **(E)** Heatmap of splicing changes in stress granule protein genes. \*  $P \leq 0.05$ , \*\*  $P \leq 0.01$ . See also Figure S3A, Table S1, Data S1.

FIGURE 3

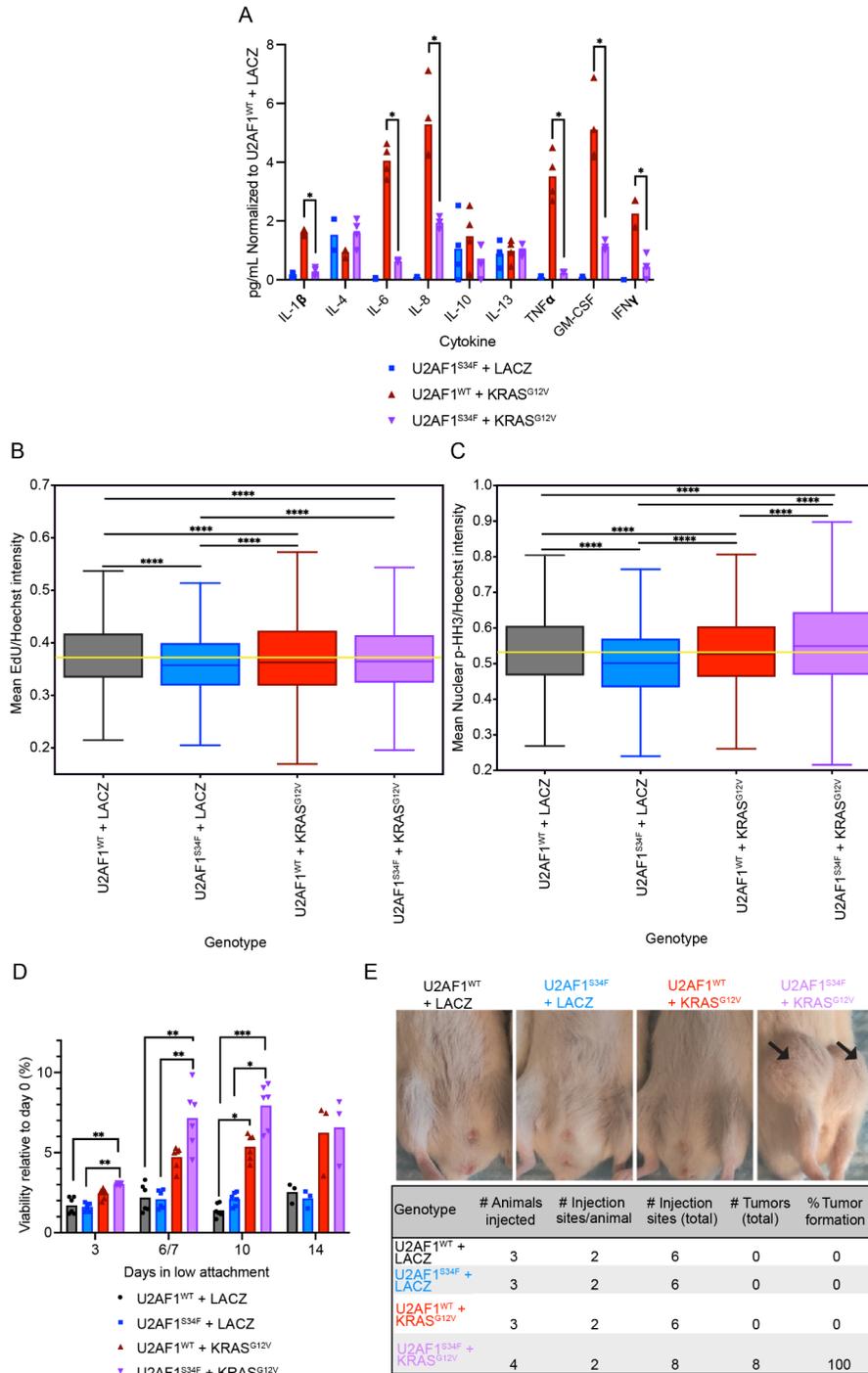
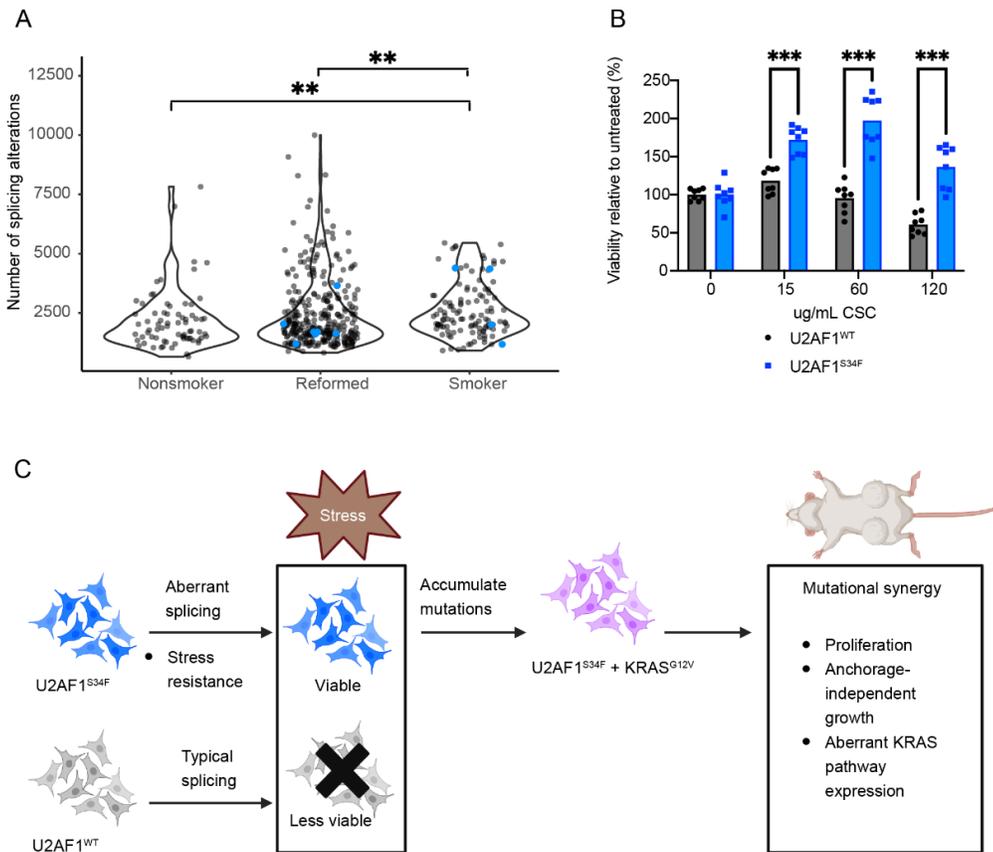


Figure 3 Co-occurring *U2AF1*<sup>S34F</sup> and *KRAS*<sup>G12V</sup> mutations increase oncogenic

**potential and proliferation. (A)** Secreted cytokine measurements of each genotype, normalized to *U2AF1<sup>WT</sup> + LACZ* levels. Bars represent mean normalized cytokine concentration. **(B)** EdU assay measuring S phase, normalized by cell density. **(C)** Phospho Histone H3 assay for number of cells undergoing M phase, normalized by cell density. Yellow lines correspond to the median of the wild-type control. The middle line in the body of each boxplot represent medians of each genotype, box limits represent quartiles, and whiskers represent the range of the most extreme, non-outlier data points. **(D)** Viability in low-attachment vessel for each HBEC3kt genotype. Relative viability is calculated by dividing the viability for each genotype at a certain time point, by the genotype's viability at day 0. Bars represent mean viability. **(E)** Top, representative injection site images of mouse xenografts. Bottom, tumor formation quantification for each HBEC3kt genotype injected. \*  $P \leq 0.05$ , \*\*  $P \leq 0.01$ , \*\*\*  $P \leq 0.001$ , \*\*\*\*  $P \leq 0.0001$ . See also Figure S3B-E, Table S1, Table S3.

FIGURE 4



**Figure 4 Altered splicing in stress granule genes in *U2AF1*<sup>S34F</sup> HBEC3kts is associated with enhanced stress response. (A) Splicing alteration distribution in lung ADC patients with smoking histories. Blue dots represent patients with *U2AF1* mutations. (B) Viability in cigarette smoke concentrate (CSC). Concentrations are in ug/mL CSC. Bars show mean viability of each cell line. (C) Working model for *U2AF1*<sup>S34F</sup>'s role in priming cells for oncogenic transformation. \*\* P ≤ 0.01, \*\*\* P ≤ 0.001. See also Data S2, Table S1.**

## **2.6 - Acknowledgements**

The short-read Illumina sequencing was carried by the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. Technical support was provided by Dr. Beverley Rabbitts, UCSC Chemical Screening Center RRID SCR\_021114. Purchase of the Perkin Elmer Opera Phenix Plus and BioTek EL406 used in this research was made possible through the National Institutes of Health Grant 1S10OD028730-01A1, awarded to the UCSC Chemical Screening Center RRID SCR\_021114. Funding for this work came from TRDRP grant T29KT0401, CRCC grant C23CR5710.

We would like to thank Drs. Christina M. Towers, Judith Campisi, and Birgit Schilling for providing critical advice and feedback on early stages of this study.

### **Author contributions**

A.N.B. supervised this study. E.H.R. and A.N.B. conceived of this study.

E.H.R. performed HBEC3kt transfections, cell culture, RNA extraction, Westerns, viability in low attachment assays, proliferation assays, clonogenicity and wound healing assays, collected conditioned media for cytokine measurements, analyzed cBioPortal lung ADC patient data, and generated plots on GraphPad Prism.

C.E.L. performed cell culture, RNA extraction, viability in cigarette smoke assay,

proliferation assays, and analyzed cBioPortal lung ADC patient data. C.E.L. aligned reads, ran HTSeq, DESeq2, performed post-processing of JuncBASE tables, ran GSEA, generated heatmaps of gene sets, performed statistical analyses, and generated plots on GraphPad Prism and Python.

C.A. ran JuncBASE and generated alternative splicing tables for HBEC3kt reads.

A.B. Aligned reads and ran HTSeq.

C.M.S. stratified lung ADC patient sample data based on smoking history and performed JuncBASE analysis of this data.

I.J.F performed the in vivo xenografts studies under supervision of S.S.S. I.J.F and S.S.S provided critical feedback on the manuscript.

C.E.L. wrote and edited original draft. C.E.L., E.H.R., A.B., and I.J.F. wrote methods.

I.J.F, S.S.S, E.H.R., and A.N.B. provided critical feedback on the manuscript.

## **2.6 - Data Access**

Cell lines generated from this study are available upon request. Sequence data have been submitted to GEO under GSE267349. All code has been deposited to GitHub and will be made publically available at <https://github.com/cindyeliang/u2af1-kras> at

publication.

## 2.8 References

Ancrile, Brooke, Kian-Huat Lim, and Christopher M. Counter. 2007. “Oncogenic Ras-Induced Secretion of IL6 Is Required for Tumorigenesis.” *Genes & Development* 21 (14): 1714–19.

Anders, Simon, Paul Theodor Pyl, and Wolfgang Huber. 2015. “HTSeq—a Python Framework to Work with High-Throughput Sequencing Data.” *Bioinformatics* 31 (2): 166–69. <https://doi.org/10.1093/bioinformatics/btu638>.

Biancon, Giulia, Poorval Joshi, Joshua T. Zimmer, Torben Hunck, Yimeng Gao, Mark D. Lessard, Edward Courchaine, et al. 2022. “Precision Analysis of Mutant U2AF1 Activity Reveals Deployment of Stress Granules in Myeloid Malignancies.” *Molecular Cell* 82 (6): 1107–22.e7.

Brooks, Angela N., Peter S. Choi, Luc de Waal, Tanaz Sharifnia, Marcin Imielinski, Gordon Saksena, Chandra Sekhar Pdamallu, et al. 2014. “A Pan-Cancer Analysis of Transcriptome Changes Associated with Somatic Mutations in U2AF1 Reveals Commonly Altered Splicing Events.” *PloS One* 9 (1): e87361.

Brooks, Angela N., Li Yang, Michael O. Duff, Kasper D. Hansen, Jung W. Park, Sandrine Dudoit, Steven E. Brenner, and Brenton R. Graveley. 2011. “Conservation of an RNA Regulatory Map between *Drosophila* and Mammals.” *Genome Research* 21 (2): 193–202.

Campbell, Joshua D., Anton Alexandrov, Jaegil Kim, Jeremiah Wala, Alice H. Berger, Chandra Sekhar Pdamallu, Sachet A. Shukla, et al. 2016. “Distinct Patterns of Somatic Genome Alterations in Lung Adenocarcinomas and Squamous Cell Carcinomas.” *Nature Genetics* 48 (6): 607–16. <https://doi.org/10.1038/ng.3564>.

Cerami, Ethan, Jianjiong Gao, Ugur Dogrusoz, Benjamin E. Gross, Selcuk Onur Sumer, Bülent Arman Aksoy, Anders Jacobsen, et al. 2012. “The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data.” *Cancer Discovery* 2 (5): 401–4. <https://doi.org/10.1158/2159-8290.CD-12-0095>.

Chen, Hsiao-Yun, Liang-Ting Lin, Mong-Lien Wang, Kun-Ling Tsai, Pin-I Huang, Yi-Ping Yang, Yi-Yen Lee, et al. 2018. “Musashi-1 Promotes Chemoresistant Granule Formation by PKR/eIF2 $\alpha$  Signalling Cascade in Refractory Glioblastoma.” *Biochimica et Biophysica Acta, Molecular Basis of Disease* 1864 (5 Pt A): 1850–61.

Chen, Liang, Jia-Yu Chen, Yi-Jou Huang, Ying Gu, Jinsong Qiu, Hao Qian, Changwei Shao, et al. 2018. “The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing Factor Mutations.” *Molecular Cell* 69 (3): 412–25.e6.

Cheruiyot, Abigael, Shan Li, Sridhar Nonavinkere Srivatsan, Tanzir Ahmed, Yuhao Chen, Delphine S. Lemacon, Ying Li, et al. 2021. “Nonsense-Mediated RNA Decay Is a Unique Vulnerability of Cancer Cells Harboring SF3B1 or U2AF1 Mutations.” *Cancer Research* 81 (17): 4499–4513.

Danecek, Petr, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard, Andrew Whitwham, et al. 2021. “Twelve Years of SAMtools and BCFtools.” *GigaScience* 10 (2). <https://doi.org/10.1093/gigascience/giab008>.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. “STAR: Ultrafast Universal RNA-Seq Aligner.” *Bioinformatics* 29 (1): 15–21.

Esfahani, Mohammad S., Luke J. Lee, Young-Jun Jeon, Ryan A. Flynn, Henning Stehr, Angela B. Hui, Noriko Ishisoko, et al. 2019. “Functional Significance of U2AF1 S34F Mutations in Lung Adenocarcinomas.” *Nature Communications* 10 (1): 5712.

Esquer, Hector, Qiong Zhou, Adedoyin D. Abraham, and Daniel V. LaBarbera. 2020. “Advanced High-Content-Screening Applications of Clonogenicity in Cancer.” *SLAS Discovery : Advancing Life Sciences R & D* 25 (7): 734–43.

Fei, Dennis Liang, Hayley Motowski, Rakesh Chatrikhi, Sameer Prasad, Jovian Yu, Shaojian Gao, Clara L. Kielkopf, Robert K. Bradley, and Harold Varmus. 2016. “Wild-Type U2AF1 Antagonizes the Splicing Program Characteristic of U2AF1-Mutant Tumors and Is Required for Cell Survival.” *PLoS Genetics* 12 (10): e1006384.

Gerber, Scott A., Abigail L. Sedlacek, Kyle R. Cron, Shawn P. Murphy, John G. Frelinger, and Edith M. Lord. 2013. “IFN- $\gamma$  Mediates the Antitumor Effects of Radiation Therapy in a Murine Colon Tumor.” *The American Journal of Pathology* 182 (6): 2345–54. <https://doi.org/10.1016/j.ajpath.2013.02.041>.

Flomerfelt, Francis A., and Ronald E. Gress. 2016. “Analysis of Cell Proliferation and Homeostasis Using EdU Labeling.” *Methods in Molecular Biology* 1323: 211–20.

Graubert, Timothy A., Dong Shen, Li Ding, Theresa Okeyo-Owuor, Cara L. Lunn, Jin

Shao, Kilannin Krysiak, et al. 2011. “Recurrent Mutations in the U2AF1 Splicing Factor in Myelodysplastic Syndromes.” *Nature Genetics* 44 (1): 53–57.

Hoving, Saske, Ann L. B. Seynhaeve, Sandra T. van Tiel, Gisela aan de Wiel-Ambagtsheer, Ernst A. de Bruijn, Alexander M. M. Eggermont, and Timo L. M. ten Hagen. 2006. “Early Destruction of Tumor Vasculature in Tumor Necrosis Factor-Alpha-Based Isolated Limb Perfusion Is Responsible for Tumor Response.” *Anti-Cancer Drugs* 17 (8): 949–59.  
<https://doi.org/10.1097/01.cad.0000224450.54447.b3>.

Imielinski, Marcin, Alice H. Berger, Peter S. Hammerman, Bryan Hernandez, Trevor J. Pugh, Eran Hodis, Jeonghee Cho, et al. 2012. “Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing.” *Cell* 150 (6): 1107–20.

Kim, Ji-Ye, Hyang Sook Jeong, Taek Chung, Moonsik Kim, Ji Hee Lee, Woo Hee Jung, and Ja Seung Koo. 2017. “The Value of Phosphohistone H3 as a Proliferation Marker for Evaluating Invasive Breast Cancers: A Comparative Study with Ki67.” *Oncotarget* 8 (39): 65064–76.

Lee, Ping-Hsien, Tori N. Yamamoto, Devikala Gurusamy, Madhusudhanan Sukumar, Zhiya Yu, Jane Hu-Li, Takeshi Kawabe, et al. 2019. “Host Conditioning with IL-1 $\beta$  Improves the Antitumor Function of Adoptively Transferred T Cells.” *The Journal of Experimental Medicine* 216 (11): 2619–34. <https://doi.org/10.1084/jem.20181218>.

Lian, Xian Jin, and Imed-Eddine Gallouzi. 2009. “Oxidative Stress Increases the Number of Stress Granules in Senescent Cells and Triggers a Rapid Decrease in p21waf1/cip1 Translation.” *The Journal of Biological Chemistry* 284 (13): 8877–87.

Liberzon, Arthur, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo. 2015. “The Molecular Signatures Database (MSigDB) Hallmark Gene Set Collection.” *Cell Systems* 1 (6): 417–25.

Liu, Huashan, Zhenxing Liang, Chi Zhou, Ziwei Zeng, Fengwei Wang, Tuo Hu, Xiaowen He, Xiaojian Wu, Xianrui Wu, and Ping Lan. 2021. “Mutant KRAS Triggers Functional Reprogramming of Tumor-Associated Macrophages in Colorectal Cancer.” *Signal Transduction and Targeted Therapy* 6 (1): 144.

Lo, April, Maria McSharry, and Alice H. Berger. 2022. “Oncogenic KRAS Alters Splicing Factor Phosphorylation and Alternative Splicing in Lung Cancer.” *BMC Cancer* 22 (1): 1315.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550. <https://doi.org/10.1186/s13059-014-0550-8>.

- Lukačičin, Martin, Adriana Espinosa-Cantú, and Tobias Bollenbach. 2022. “Intron-Mediated Induction of Phenotypic Heterogeneity.” *Nature* 605 (7908): 113–18.
- Miller, Christopher A., Jason R. Walker, Travis L. Jensen, William F. Hooper, Robert S. Fulton, Jeffrey S. Painter, Mikkael A. Sekeres, et al. 2022. “Failure to Detect Mutations in U2AF1 due to Changes in the GRCh38 Reference Sequence.” *The Journal of Molecular Diagnostics: JMD* 24 (3): 219–23. <https://doi.org/10.1016/j.jmoldx.2021.10.013>.
- Montgomery, Nathan D., Jonathan Galeotti, Steven M. Johnson, Leah Commander, Eric T. Weimer, Pranil K. Chandra, Tariq Nazir, Thomas B. Alexander, Joshua F. Zeidner, and Matthew C. Foster. 2021. “Bilineal Evolution of a U2AF1-Mutated Clone Associated with Acquisition of Distinct Secondary Mutations.” *Blood Advances* 5 (24): 5612–16.
- Muñoz-Maldonado, Carmen, Yitzhak Zimmer, and Michaela Medová. 2019. “A Comparative Analysis of Individual RAS Mutations in Cancer Biology.” *Frontiers in Oncology* 9 (October): 1088. <https://doi.org/10.3389/fonc.2019.01088>.
- Nguyen, Hai Dang, Wan Yee Leong, Weiling Li, Pavankumar N. G. Reddy, Jack D. Sullivan, Matthew J. Walter, Lee Zou, and Timothy A. Graubert. 2018. “Spliceosome Mutations Induce R Loop-Associated Sensitivity to ATR Inhibition in Myelodysplastic Syndromes.” *Cancer Research* 78 (18): 5363–74.
- Okeyo-Owuor, T., B. S. White, R. Chatrikhi, D. R. Mohan, S. Kim, M. Griffith, L. Ding, et al. 2015. “U2AF1 Mutations Alter Sequence Specificity of Pre-mRNA Binding and Splicing.” *Leukemia* 29 (4): 909–17.
- Palangat, Murali, Dimitrios G. Anastasakis, Dennis Liang Fei, Katherine E. Lindblad, Robert Bradley, Christopher S. Hourigan, Markus Hafner, and Daniel R. Larson. 2019. “The Splicing Factor U2AF1 Contributes to Cancer Progression through a Noncanonical Role in Translation Regulation.” *Genes & Development* 33 (9-10): 482–97.
- Park, Sung Mi, Jianhong Ou, Lynn Chamberlain, Tessa M. Simone, Huan Yang, Ching-Man Virbasius, Abdullah M. Ali, et al. 2016. “U2AF35(S34F) Promotes Transformation by Directing Aberrant ATG7 Pre-mRNA 3’ End Formation.” *Molecular Cell* 62 (4): 479–90.
- Poetsch, Anna R. 2020. “The Genomics of Oxidative DNA Damage, Repair, and Resulting Mutagenesis.” *Computational and Structural Biotechnology Journal* 18 (January): 207–19.

Ramirez, Ruben D., Shelley Sheridan, Luc Girard, Mitsuo Sato, Young Kim, Jon Pollack, Michael Peyton, et al. 2004. "Immortalization of Human Bronchial Epithelial Cells in the Absence of Viral Oncoproteins." *Cancer Research* 64 (24): 9027–34. <https://doi.org/10.1158/0008-5472.CAN-04-3703>.

Robinson, James T., Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, and Jill P. Mesirov. 2011. "Integrative Genomics Viewer." *Nature Biotechnology* 29 (1): 24–26. <https://doi.org/10.1038/nbt.1754>.

Sato, Mitsuo, Jill E. Larsen, Woonchang Lee, Han Sun, David S. Shames, Maithili P. Dalvi, Ruben D. Ramirez, et al. 2013. "Human Lung Epithelial Cells Progressed to Malignancy through Specific Oncogenic Manipulations." *Molecular Cancer Research: MCR* 11 (6): 638–50.

Soulette, Cameron M., Eva Hrabeta-Robinson, Carlos Arevalo, Colette Felton, Alison D. Tang, Maximillian G. Marin, and Angela N. Brooks. 2023. "Full-Length Transcript Alterations in Human Bronchial Epithelial Cells with U2AF1 S34F Mutations." *Life Science Alliance* 6 (10). <https://doi.org/10.26508/lsa.202000641>.

Su, Yang, Tianxiang Zhang, Jinsen Lu, Lei Qian, Yang Fei, Li Zhang, Song Fan, et al. 2023. "Identification and Validation of the Prognostic Panel in Clear Cell Renal Cell Carcinoma Based on Resting Mast Cells for Prediction of Distant Metastasis and Immunotherapy Response." *Cells* 12 (1). <https://doi.org/10.3390/cells12010180>.

Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, et al. 2005. "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles." *Proceedings of the National Academy of Sciences of the United States of America* 102 (43): 15545–50. <https://doi.org/10.1073/pnas.0506580102>.

Tang, Alison D., Cameron M. Soulette, Marijke J. van Baren, Kevyn Hart, Eva Hrabeta-Robinson, Catherine J. Wu, and Angela N. Brooks. 2020. "Full-Length Transcript Characterization of SF3B1 Mutation in Chronic Lymphocytic Leukemia Reveals Downregulation of Retained Introns." *Nature Communications* 11 (1): 1438.

Wadugu, Brian A., Sridhar Nonavinkere Srivatsan, Amanda Heard, Michael O. Alberti, Matthew Ndonwi, Jie Liu, Sarah Grieb, et al. 2021. "U2af1 Is a Haplo-Essential Gene Required for Hematopoietic Cancer Cell Survival in Mice." *The Journal of Clinical Investigation* 131 (21). <https://doi.org/10.1172/JCI141401>.

Wu, S., C. M. Romfo, T. W. Nilsen, and M. R. Green. 1999. "Functional Recognition of the 3' Splice Site AG by the Splicing Factor U2AF35." *Nature* 402 (6763): 832–35.

Yip, Bon Ham, Violetta Steeples, Emmanouela Repapi, Richard N. Armstrong, Miriam Llorian, Swagata Roy, Jacqueline Shaw, et al. 2017. “The U2AF1S34F Mutation Induces Lineage-Specific Splicing Alterations in Myelodysplastic Syndromes.” *The Journal of Clinical Investigation* 127 (6): 2206–21.

Zamore, P. D., and M. R. Green. 1989. “Identification, Purification, and Biochemical Characterization of U2 Small Nuclear Ribonucleoprotein Auxiliary Factor.” *Proceedings of the National Academy of Sciences of the United States of America* 86 (23): 9243–47.

Zhu, Yuqian, Dandan Song, Juan Guo, Jiacheng Jin, Ying Tao, Zheng Zhang, Feng Xu, et al. 2021. “U2AF1 Mutation Promotes Tumorigenicity through Facilitating Autophagy Flux Mediated by FOXO3a Activation in Myelodysplastic Syndromes.” *Cell Death & Disease* 12 (7): 655.

## **CHAPTER 3 - direct RNA sequencing reveals altered modification landscape from U2AF1<sup>S34F</sup> and cigarette smoke concentrate**

### **3.1 INTRO**

RNA modifications are chemical groups that are covalently added to RNA nucleotides and processed by a group of proteins known as modification reader, writer, and erasers ([Wilkinson, Cui, and He, 2021](#)). Although the field is still profiling the functional roles of these modifications in biology, aberrant in m6A reader and writer protein expression have been found to promote oncogenic growth of lung cancer cells ([Wanna-udom et al. 2020](#); [Shen et al. 2020](#)). Additionally, m6A eraser protein FTO has been shown to mediate autophagy in HeLa cells ([Jin et al. 2018](#)).

Autophagy is a process where cells break down damaged organelles for reuse ([Parzych and Klionsky 2014](#)). During autophagy, portions of the cell are enveloped by a double membrane to form an autophagosome. Autophagosomes then fuse with lysosomes with the help of proteins, such as VAMP8, which allows for the digestion of the autophagosome's components ([Wang and Diao 2022](#)). Autophagy is known to play different roles in oncogenesis depending on the stage of the tumor ([Towers,](#)

[Wodetzki, and Thorburn 2020](#)). In established tumors, enhanced autophagy is thought to promote tumor growth, as it helps tumors meet high metabolic demands. In contrast, in precancerous cells, autophagy is thought to play a protective role by removing cell components damaged by ROS. Studying the RNA modification landscape in the context of cancer and environmental stressors therefore has potential to lead to new insights on lung cancer formation.

Since the advent of third-generation sequencing technologies such as those offered by the Oxford Nanopore Technologies (ONT) platform, several computational tools have been developed to analyze third-generation sequence data and predict the presence of RNA modifications. RNA modification detection tools fall in two general categories, based on their method of modification identification ([Zhong et al. 2023](#)). The tools m6Anet and xPore fall into the category that makes predictions based on differences in electric current signals made by molecules passing through the Nanopore ([Hendra et al. 2022](#); [Pratanwanich et al. 2021](#)). In contrast, other tools such as EpiNano use errors in base-calling to predict the presence of modifications ([Liu et al. 2019](#)).

A recent evaluation of RNA modification detection tool performance has revealed interesting guidelines for analysis ([Zhong et al. 2023](#)). First, it is recommended to combine results from tools that use different methods to predict RNA modifications, because each method comes with its own drawbacks. For instance, tools like EpiNano

require a matched control sample with minimal modifications for confident predictions, because direct RNA sequencing has a high false positive rate for defining base-calling errors ([Liu et al. 2019](#)). However, these tools were better able to detect m6As at GGACT, GGACA, GGACC, and AGACT motifs. Detection capabilities of tools also differed at different m6A motifs, with EpiNano, m6Anet, and xPore having differing recall and precision rates at different RRACH sites. It is thus recommended to combine results from multiple RNA modification detection tools that utilize different detection methods for an integrated analysis.

Here, I show preliminary results of RNA modification predictions of HBEC3kt cells with *U2AF1<sup>WT</sup>* or *U2AF1<sup>S34F</sup>*, grown with and without the presence of CSC, from EpiNano, and xPore. EpiNano was chosen because it represents a tool that utilizes the base calling error method and performed well in the evaluation ([Zhong et al. 2023](#)). Although xPore was evaluated to have high precision at the cost of low recall, it predicts differences based on current signals in a modification-agnostic manner, which allows for potential detection of non-m6A modifications ([Pratanwanich et al. 2021](#)).

### **3.2 Results**

RNA mod calling tools EpiNano and xPore predict altered RNA modification in *U2AF1<sup>S34E</sup>*-mutant cells

We first performed preliminary RNA modification predictions on untreated HBEC3kts with and without *U2AF1<sup>S34F</sup>*. EpiNano predicted a higher number of transcriptomic positions that carried modifications in *U2AF1<sup>S34F</sup>* HBEC3kts compared to *U2AF1<sup>WT</sup>* (Fig 1A). Next, we used xPore, which predicts the presence of any modifications based on current signal differences. Differences in reference used for alignment for EpiNano and xPore made it difficult to find overlap between the modified transcripts predicted by each method. However, xPore in particular detected two autophagy genes that were significantly differentially modified in *U2AF1<sup>S34F</sup>* compared to wild-type: *VAMP8* and *GABARAPL2* (Fig 1B). *VAMP8* was estimated to be less modified in *U2AF1<sup>S34F</sup>* HBEC3kts, while *GABARAPL2* was estimated to be more modified.

Differential modification in *VAMP8* associated with protein change, but not splicing and gene expression differences

We followed up on investigating modification in *VAMP8* by viewing direct RNA reads from three biological replicates of untreated *U2AF1<sup>WT</sup>* and *U2AF1<sup>S34F</sup>* HBEC3kts on IGV (Fig 2A, bottom 6 tracks). Because RNA modifications often cause errors in base-calling, potential sites of modification can be detected via base-calling errors which show up as colored nucleotide alignments on IGV. As a negative control for base-calling errors caused by RNA modifications, we also

viewed the same loci on Nanopore cDNA reads (top 6 tracks). We observed a locus on *VAMP8* near the 5' UTR that appeared to have modifications in *U2AF1<sup>WT</sup>* but not *U2AF1<sup>S34F</sup>* samples. We were unable to find a locus in *GABARAPL2* that corresponded to the xPore prediction.

Next, we asked whether potential differential modification in *VAMP8* corresponded with differences in splicing or gene expression. We examined splicing by viewing three replicates of *U2AF1<sup>WT</sup>* and *U2AF1<sup>S34F</sup>* Nanopore cDNA read coverage at the entire *VAMP8* locus (Fig 2B) and saw no changes in exon coverage across the transcript between the genotypes, indicating no changes in splicing. When we quantified normalized gene counts of *VAMP8* in *U2AF1<sup>WT</sup> + LacZ* and *U2AF1<sup>S34F</sup> + LacZ* cells, we also found no significant difference in gene expression (Fig 2C). However, when we interrogated relative protein abundance of *VAMP8* in *U2AF1<sup>WT</sup> + LacZ* and *U2AF1<sup>S34F</sup> + LacZ* cells, we detected lower *VAMP8* protein levels (Fig 2D). Together, our results suggest that differential RNA modification by itself may have a functional impact on protein expression in the cell.

xPore predicts a global increase of m6A modifications caused by *U2AF1<sup>S34F</sup>* and cigarette smoke exposure

Next, we examined preliminary dRNA-seq data on HBEC3kts exposed to DMSO or CSC, using xPore. When we examined global m6A levels, we observed that

*U2AF1<sup>S34F</sup>* by itself increased overall levels of modification (Fig 3C). The addition of CSC also increased modification levels in WT cells (Fig3B). Methylation levels were further enhanced by the treatment of *U2AF1<sup>S34F</sup>* cells with CSC (Fig 3D). Contrary to previous literature on m6A loci, most of our predicted modifications were not in the terminal exons of the transcript ([Ke et al. 2015](#)).

### 3.3 Discussion

Although the results are preliminary, this set of experiments and analyses demonstrate the utility of using diverse sequencing approaches and analyses to understand biology. Direct RNA modifications in *VAMP8* was associated with difference in protein levels when gene expression between cell lines with similar mutational profiles was unaltered, and cDNA coverage did not reveal evidence of altered splicing. Analyzing sequences from multiple kinds of RNA-sequencing library preparation techniques has the potential to uncover a more complete transcriptome. More work is needed to understand our results, especially the prediction of m6A modifications occurring primarily at non-terminal exons in DMSO- and CSC-treated samples. Although one hypothesis is that cigarette smoke treatment and *U2AF1<sup>S34F</sup>* mutation may be modifying RRACH motifs at noncanonical locations via a novel mechanism, it will be crucial to repeat these analyses on dRNA-seq results from more replicates.

dRNA-seq is still a developing technology. dRNA-seq from other cell lines produced

with the R9.4.1 MinION flow cells and the SQK-RNA002 kit used in this chapter have been evaluated by the Long-read RNA-Seq Genome Annotation Assessment Project to have poorer read depth and higher error rates than Nanopore cDNA sequencing methods, which led to the lowest number of annotated transcripts reported in the evaluation ([Pardo-Palacios et al. 2024](#)). Although a new dRNA-seq kit and flow cell has since been released by ONT, the RNA modification prediction tools used in this chapter have yet to be updated to process data produced using the new technology. In the interim, future steps for analyzing dRNA-seq results created with the R9.4.1 flow cells include benchmarking RNA modification predictions using short read sequence analysis of matched samples, orthogonal methods of RNA modification detection such as m6A-IP-seq, and comparison of predicted mod loci with known RRACH motifs.

### **3.4 Methods**

#### Cell culture and library preparation

For RNA modification analysis of untreated cells, clone 1 HBEC3kts wild-type or mutant for *U2AF1*<sup>S34F</sup> were cultured and total RNA was extracted as described in chapter 2. For RNA modification analysis of treated cells, HBEC3kts wild-type or mutant for *U2AF1*<sup>S34F</sup> were cultured until 50% confluency in standard media. Media was replaced and 15ug/mL CSC from Murty Pharmaceuticals (Cat#nc1560725) or an equal volume of sterile DMSO was added to the culture dish. Following three days of

incubation, cells were washed twice with warm DPBS and RNA was extracted as described in chapter 2.

### Direct RNA library preparation and sequencing

Total RNA was poly-A selected using NEXTflex Poly(A) Beads 2.0 (Cat#NOVA-512993) following the manufacturer's protocols. Library preparation of direct RNA samples were performed using the SQK-RNA002 Nanopore kit, with SuperScript 4 Reverse Transcriptase (Cat#18090010) and RNAClean XP beads (Cat#A63987). Kit instructions were followed, with the following addition: when eluting library from the wash buffer, beads were pelleted on a magnet in a 37°C incubator. RNA was sequenced on an R9.4.1 flow cell using the MinION using the following MinKnow parameters: 72 hour run length, 200bp minimum read length, adaptive sampling off, basecalling off, active channel selection on, reserve pores on, and 3 hours between pore scans.

### cDNA library preparation and sequencing

Total RNA was extracted from HBEC3kts as described above. Library preparation was performed using the SQK-PCS110 kit following kit instructions, R9.4.1 flow cell using the MinION. Reads were base-called using guppy4.2.2 with the following parameters: `~/bin/ont-guppy-4.2.2/bin/guppy_basecaller -r -c dna_r9.4.1_450bps_hac.cfg --qscore_filtering --min_qscore 7.0 -x cuda:0`. Alignment was performed using minimap2 v.2.17 the following parameters:

```
/private/home/rshelans/bin/minimap2_2.17/minimap2 -ax splice
/private/groups/brookslab/reference_sequence/GRCh38.u2af1_fix.v1.2020_04_01.fa
--junc-bed
/private/groups/brookslab/reference_annotations/gencode.v38.p13.annotation.bed
/private/groups/brookslab/abehera/Nanopore/HBEC3kt_U2AF1/fastq_guppy4.2.2_gp
u/guppy4.2.2_basecalled/all.WT1-3.fastq.gz > 20230215_ontcdna_hbecwt1_r3.sam.
```

### RNA-seq data processing and RNA modification prediction.

For the data shown in this chapter, one replicate of each condition was analyzed (untreated  $U2AF1^{WT}$  and  $U2AF1^{S34F}$ , DMSO-treated  $U2AF1^{WT}$  and  $U2AF1^{S34F}$ , CSC-treated  $U2AF1^{WT}$  and  $U2AF1^{S34F}$ ). For EpiNano analysis, basecalling was performed on raw reads using guppy v. 3.1.5 with the following parameters: -c rna\_r9.4.1\_70bps\_hac.cfg --qscore\_filtering yes --min\_qscore 7 --reverse\_sequence yes --u\_substitution yes. Basecalled files were then aligned using minimap2 v.2.17 using the following parameters: --junc-bed

```
/private/groups/brookslab/reference_annotations/gencode.v38.p13.annotation.bed
--secondary=no -ax splice -uf -k14
/private/groups/brookslab/reference_sequence/GRCh38.u2af1_fix.v1.2020_04_01.fa.
```

Then, aligned SAM files were converted to BAM files and indexed using samtools.

The BAM file was converted into a bed12 for processing with FLAIR using the

bam2Bed12.py script provided by FLAIR ([Tang et al. 2020](#)).

FLAIR was then used to generate a reference transcriptome for processing the HBEC3kt sequences to in EpiNano. First, the FLAIR correct module was used with these parameters: `flair.py correct -g`

```
/private/groups/brookslab/reference_sequence/GRCh38.u2af1_fix.v1.2020_04_01.fa  
-j 2mut2laczSJfiltered.bed -f
```

```
/private/groups/brookslab/reference_annotations/gencode.v38.annotation.gtf. Then,  
the corrected alignments were collapsed using the cat command. The combined files  
were used to generate the HBEC3kt FLAIR transcriptome using flair.py collapse and  
the following parameters: -g
```

```
/private/groups/brookslab/reference_sequence/GRCh38.u2af1_fix.v1.2020_04_01.fa  
-r
```

```
/scratch/celiang/guppy315/mut1r1fastq/pass/mut1r1hbec_guppy315_combined.fastq  
/scratch/celiang/guppy315/wt1r1fastq/pass/wt1r1hbec_guppy315_combined.fastq -q  
2022_01_05_hbecRep1_combined_corrected.bed -f
```

```
/private/groups/brookslab/reference_annotations/gencode.v38.annotation.gtf -p
```

```
/private/home/celiang/LRGASP/promoter.gencode.v27.20.bed -o
```

```
2022_01_05-hbecR1_combined-collapse --generate_map --temp_dir
```

```
/scratch/celiang/flair_temp/. Following the generation of the HBEC3kt reference  
transcriptome, base-called files were realigned to the HBEC3kt reference for
```

processing with EpiNano using the parameters:

```
/usr/local/bin/minimap2-2.14_x64-linux/minimap2 -t 4 -ax splice -uf -k14
```

```
2022_01_05-hbecR1_combined-collapse.isoforms.fa
```

```
mut1r1hbec_guppy315_combined.fastq >
```

2023\_2\_21\_flairome\_drna\_hbec\_mut1\_r1.sam. SAMs were converted to BAMs and

indexed with samtools, and a sequence dictionary is made for EpiNano using this

command: java -jar /private/groups/brookslab/bin/picard-tools-1.140/picard.jar

```
CreateSequenceDictionary R=2022_01_05-hbecR1_combined-collapse.isoforms.fa
```

```
O=2022_01_05-hbecR1_combined-collapse.isoforms.fa.dict
```

EpiNano Variants was then run with the following parameters: python3

```
EpiNano/Epinano_Variants.py -R
```

```
2022_01_05-hbecR1_combined-collapse.isoforms.fa -b
```

```
2023_2_21_sorted-flairome_drna_hbec_wt1_r1.bam -s sam2tsv.jar -T t -n 6.
```

Subsequently, EpiNano SumErr was run using: python3

```
/usr/local/bin/EpiNano/misc/Epinano_sumErr.py --file
```

```
2023_2_21_sorted-flairome_drna_hbec_wt1_r1.plus_strand.per.site.csv --out
```

```
2023_2_22wt1r1.sum_err.csv --kmer 0. The resulting CSVs of predicted
```

modifications were then used to plot modification overlap in gene IDs using Jupyter notebook.

Xpore analysis was performed using a NextFlow pipeline which can be accessed here: [https://github.com/vpeddu/diffmod\\_analysis/tree/main](https://github.com/vpeddu/diffmod_analysis/tree/main), using the following parameters: `nextflow run vpeddu/diffmod_analysis \ --dataprep \ --transcriptome gencode.v39.transcripts.fa \ --transcriptome_gtf gencode.v39.annotation.sorted.gtf \ --input_csv input_template.csv \ -with-singularity ubuntu:18.04 \ -resume`. For untreated HBEC3kts, differentially modified transcripts were plotted after filtering for xPore events with  $p < 0.05$  and plotted with R's Enhanced Volcano Plot package. For m6A modification prediction, the annotated last exon of each transcript was obtained from the supplied GTF file. Then, a table of RRACH motifs were created from xPore transcripts. Modifications that were predicted to take place at one of these motifs were then plotted and annotated for whether they were present in the last exons of the transcript.

#### IGV and gene expression comparisons of *VAMP8*

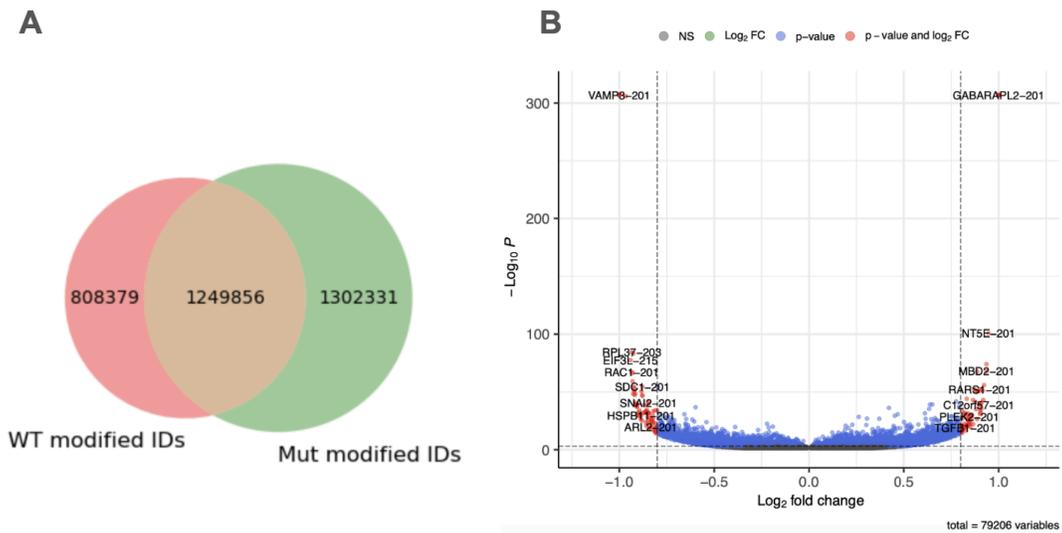
IGV was used to visualize three biological replicates of clone 1 *U2AF1*<sup>S34F</sup> and *U2AF1*<sup>WT</sup> HBEC3kts sequenced with the Nanopore cDNA kit and three biological replicates of clone 1 *U2AF1*<sup>S34F</sup> and *U2AF1*<sup>WT</sup> HBEC3kts sequenced with the Nanopore direct RNA kit. Normalized gene expression counts were obtained from DESeq2 analysis of clone 1 *U2AF1*<sup>S34F</sup> + *LacZ* and *U2AF1*<sup>WT</sup> + *LacZ* HBEC3kt Illumina reads as described in chapter 2.

### Western blot

Western blot was performed on two technical replicates of one biological replicate of clone 1 *U2AF1<sup>S34F</sup> + LacZ* and *U2AF1<sup>WT</sup> + LacZ* HBEC3kts, as described in chapter 2. 30ug of total protein was used as input for SDS-PAGE. Then, proteins were transferred to a PVDF membrane using a semi-dry transfer apparatus for 30 minutes. The blot was then incubated nonfat 5% milk/TBST block overnight, before addition of VAMP8 primary antibody (Cat#13060S) at 1:1000 dilution for overnight incubation. The blot was incubated in a secondary HRP antibody (Cat##7074) at 1:1000 dilution for one hour. For the loading control, blots were incubated in B actin conjugated to HRP (Cat#sc-47778 HRP) for one hour. Bands were visualized with ECL (Cat#926-95000) and imaged using the C-Digit Imager (Li-Cor).

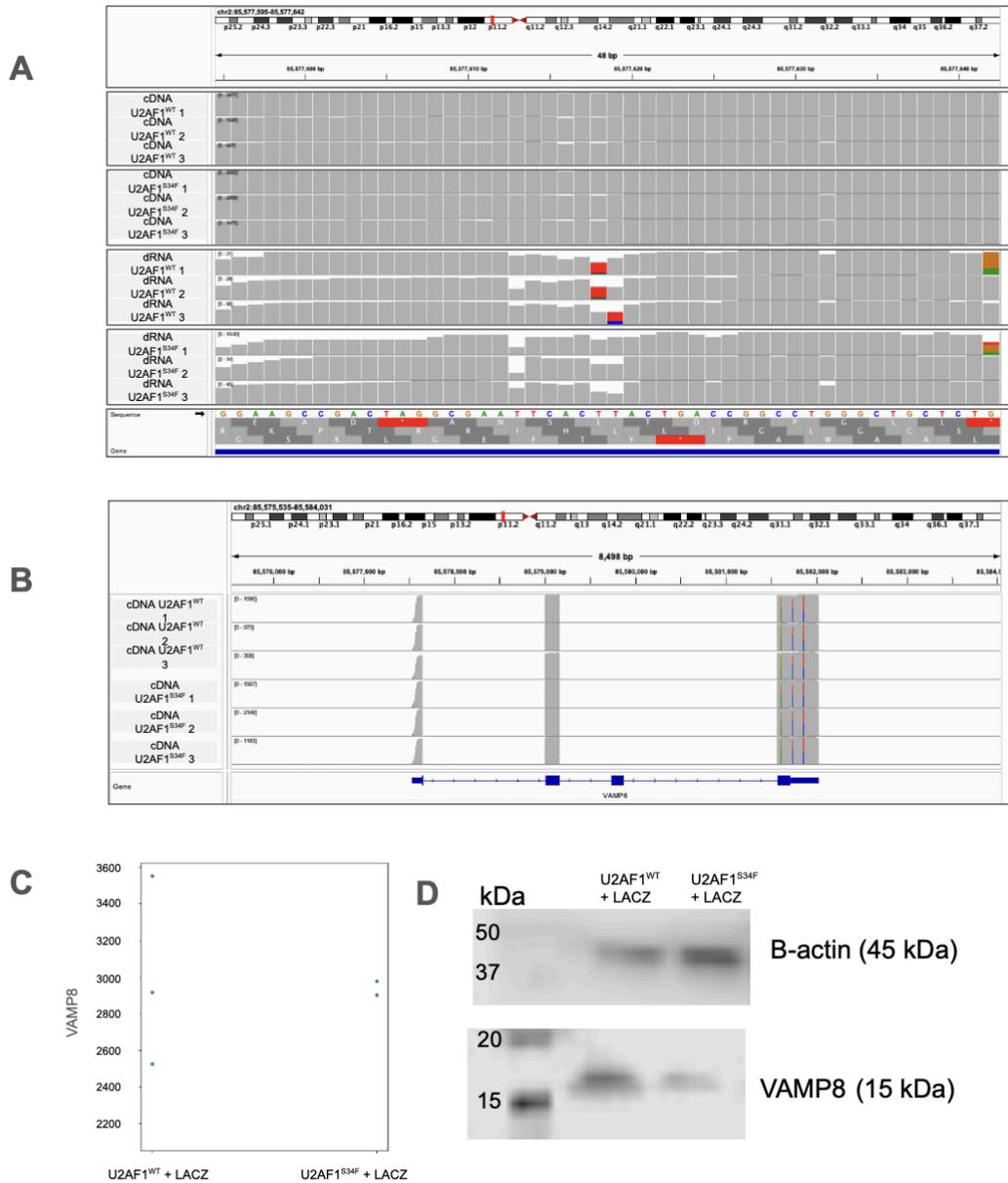
### 3.5 - Figures

FIGURE 1



**Fig 1 | EpiNano and xPore predict differential modification in  $U2AF1^{S34F}$  cells. A,** EpiNano predictions of m6A modification events in  $U2AF1^{WT}$  and  $U2AF1^{S34F}$  HBEC3kts without treatment. Red indicates transcript IDs that are modified in only  $U2AF1^{WT}$ , green indicates IDs that are modified in only  $U2AF1^{S34F}$ , and orange indicates the IDs modified in both  $U2AF1^{WT}$  and  $U2AF1^{S34F}$  transcripts. **B,** Global differential modification predictions between untreated  $U2AF1^{WT}$  and  $U2AF1^{S34F}$  HBEC3kts predicted by xPore. Each dot represent gene predicted to be modified by xPore. Y-axis negative log of the xPore p-value. X-axis is the Log<sub>2</sub> fold change of the modification.

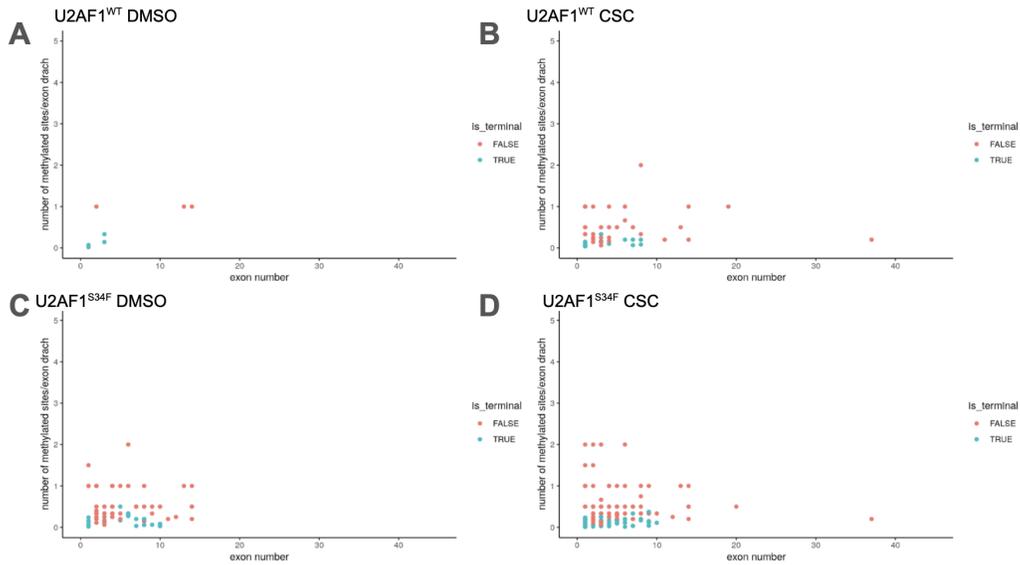
FIGURE 2



**Fig 2 | In untreated HBEC3kts, *VAMP8* modification in *U2AF1*<sup>WT</sup> lines is associated with altered protein abundance. A, Locus of potential *VAMP8* modification location detected in direct RNA sequences from *U2AF1*<sup>WT</sup> cells. First**

two sets of tracks show Nanopore cDNA sequence coverage in  $U2AF1^{WT}$  and  $U2AF1^{S34F}$  cells. Second two sets of tracks show direct RNA sequence coverage from the same genotypes. The last panel shows the nucleotide sequence of the locus **B**, Isoform-level view of *VAMP8* coverage in  $U2AF1^{WT}$  (top 3 tracks) and  $U2AF1^{S34F}$  (bottom 3 tracks) Nanopore cDNA reads. **C**, Normalized gene expression in  $U2AF1^{WT} + LacZ$  and  $U2AF1^{S34F} + LacZ$  cell lines (Wilcoxon ranksum pvalue=0.83). **D**, Western blot of VAMP8 in  $U2AF1^{WT} + LacZ$  and  $U2AF1^{S34F} + LacZ$  cell lines.

FIGURE 3



**Fig 3 | xPore predicts a global increase of m6A modifications caused by *U2AF1*<sup>S34F</sup> and cigarette smoke exposure.** Each dot represent a modified locus predicted by xPore with a modification probability > 75%. The exon number represents the number of exon in the transcript the modification is detected in. Blue indicates the modification is in a terminal exon, while Red indicates the modification is in a nonterminal exon. **A.** Predicted m6A modifications in *U2AF1*<sup>WT</sup> treated with DMSO control. **B.** Predicted m6A modifications in *U2AF1*<sup>WT</sup> treated with CSC. **C.** Predicted m6A modifications in *U2AF1*<sup>S34F</sup> treated with DMSO. **D.** Predicted m6A modifications in *U2AF1*<sup>S34F</sup> treated with CSC.

### 3.6 - Acknowledgements

Cindy E Liang performed cell culture, cigarette smoke treatment, direct RNA library prep and sequencing, Westerns, and EpiNano analysis. Vikas Peddu performed xPore analysis.

### 3.7 - References

Hendra, Christopher, Ploy N. Pratanwanich, Yuk Kei Wan, W. S. Sho Goh, Alexandre Thiery, and Jonathan Göke. 2022. “Detection of m6A from Direct RNA Sequencing Using a Multiple Instance Learning Framework.” *Nature Methods* 19 (12): 1590–98.

Jin, Shouheng, Xiya Zhang, Yanyan Miao, Puping Liang, Kaiyu Zhu, Yuanchu She, Yaoxing Wu, et al. 2018. “m6A RNA Modification Controls Autophagy through Upregulating ULK1 Protein Abundance.” *Cell Research* 28 (9): 955–57.

Ke, Shengdong, Endalkachew A. Alemu, Claudia Mertens, Emily Conn Gantman, John J. Fak, Aldo Mele, Bhagwattie Haripal, et al. 2015. “A Majority of m6A Residues Are in the Last Exons, Allowing the Potential for 3' UTR Regulation.” *Genes & Development* 29 (19): 2037–53.

Liu, Huanle, Oguzhan Begik, Morghan C. Lucas, Jose Miguel Ramirez, Christopher E. Mason, David Wiener, Schraga Schwartz, John S. Mattick, Martin A. Smith, and Eva Maria Novoa. 2019. “Accurate Detection of m6A RNA Modifications in Native RNA Sequences.” *Nature Communications* 10 (1): 4079.

Pardo-Palacios, Francisco J., Dingjie Wang, Fairlie Reese, Mark Diekhans, Silvia Carbonell-Sala, Brian Williams, Jane E. Loveland, et al. 2023. “Systematic Assessment of Long-Read RNA-Seq Methods for Transcript Identification and Quantification.” *bioRxiv : The Preprint Server for Biology*, July.

<https://doi.org/10.1101/2023.07.25.550582>.

Parzych, Katherine R., and Daniel J. Klionsky. 2014. “An Overview of Autophagy: Morphology, Mechanism, and Regulation.” *Antioxidants & Redox Signaling* 20 (3): 460–73.

Pratanwanich, Ploy N., Fei Yao, Ying Chen, Casslynn W. Q. Koh, Yuk Kei Wan, Christopher Hendra, Polly Poon, et al. 2021. “Identification of Differential RNA Modifications from Nanopore Direct RNA Sequencing with xPore.” *Nature Biotechnology* 39 (11): 1394–1402.

Sheng, Hao, Zhen Li, Shixin Su, Wenjing Sun, Xiaoya Zhang, Leilei Li, Jing Li, et al.

2020. “YTH Domain Family 2 Promotes Lung Cancer Cell Growth by Facilitating 6-Phosphogluconate Dehydrogenase mRNA Translation.” *Carcinogenesis* 41 (5): 541–50.

Tang, Alison D., Cameron M. Soulette, Marijke J. van Baren, Kevyn Hart, Eva Hrabeta-Robinson, Catherine J. Wu, and Angela N. Brooks. 2020. “Full-Length Transcript Characterization of SF3B1 Mutation in Chronic Lymphocytic Leukemia Reveals Downregulation of Retained Introns.” *Nature Communications* 11 (1): 1438.

Towers, Christina G., Darya Wodetzki, and Andrew Thorburn. 2020. “Autophagy and Cancer: Modulation of Cell Death Pathways and Cancer Cell Adaptations.” *The Journal of Cell Biology* 219 (1). <https://doi.org/10.1083/jcb.201909033>.

Wang, Lei, and Jiajie Diao. 2022. “VAMP8 Phosphorylation Regulates Lysosome Dynamics during Autophagy.” *Autophagy Reports* 1 (1): 79–82.

Wanna-Udom, Sasithorn, Minoru Terashima, Hanbing Lyu, Akihiko Ishimura, Takahisa Takino, Matomo Sakari, Toshifumi Tsukahara, and Takeshi Suzuki. 2020. “The m6A Methyltransferase METTL3 Contributes to Transforming Growth Factor-Beta-Induced Epithelial-Mesenchymal Transition of Lung Cancer Cells through the Regulation of JUNB.” *Biochemical and Biophysical Research Communications* 524 (1): 150–55.

Wilkinson, Emma, Yan-Hong Cui, and Yu-Ying He. 2021. "Context-Dependent Roles of RNA Modifications in Stress Responses and Diseases." *International Journal of Molecular Sciences* 22 (4). <https://doi.org/10.3390/ijms22041949>.

Zhong, Zhen-Dong, Ying-Yuan Xie, Hong-Xuan Chen, Ye-Lin Lan, Xue-Hong Liu, Jing-Yun Ji, Fu Wu, et al. 2023. "Systematic Comparison of Tools Used for m6A Mapping from Nanopore Direct RNA Sequencing." *Nature Communications* 14 (1): 1906.