Patient 3’UTR mutations regulate post-transcriptional gene expression and cellular oncogenicity in prostate cancer

Samantha Schuster

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Andrew Hsieh, Chair

Adam Geballe

Edith Wang

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Patient 3’UTR mutations regulate post-transcriptional gene expression and cellular oncogenicity in prostate cancer

Samantha Schuster

Chair of the Supervisory Committee:

Andrew Hsieh

Department of Genome Sciences

While cancer genomics has historically focused on the coding genome, the 3' untranslated region (3’UTR) represents a largely unexplored avenue for uncovering alternative genetic drivers of disease. The 3’UTR controls mRNA stability and translation through sequence and structure-based features that recruit *trans*-acting factors such as RNA-binding proteins (RBPs) and microRNAs (miRNAs). Though many studies have discovered ways in which RBPs, miRNAs, or other aspects of the 3’UTR are oncogenic, little is known about how somatic mutations to the 3’UTR may function in cancer. In this dissertation, I discuss how I built and utilized a pair of massively parallel reporter assays (MPRAs) to better understand the role of 3’UTR mutations in prostate cancer. Using these novel tools, I measured how thousands of 3’UTR mutations found in advanced stage prostate tumors affect multiple levels of post-transcriptional gene expression, uncovering hundreds of patient mutations that significantly affect either mRNA stability or translational efficiency. Additionally, I performed CRISPR-Cas9 base editing to demonstrate that introduction of oncogenic patient 3’UTR mutations into their

endogenous genomic context leads to increased cell growth and stress resistance. In parallel, I applied this same base editing technology to a patient mutation of interest in the 5’ untranslated region, proving that single-nucleotide changes to either UTR can have significant effects on mRNA translation. This work represents an unprecedented view of the extent to which disease-relevant UTR mutations affect mRNA stability, translation efficiency, and cancer, where use of cutting-edge technology brings us closer to understanding the full breadth of genetic drivers in cancer.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my family, especially my parents, for always supporting me in everything. For my dad wondering “how I came out so much smarter than him” to give me encouragement and boost my morale, and my mom for always being there for me no matter what I was pursuing. Also my brother, for making fun of me enough that I just had to go get a PhD to one-up his Master’s degree.

I want to acknowledge the teachers and mentors I have had throughout my academic career, from my high school Biology teachers who first opened the world of genetics to me and my English teachers who taught me how to write (turns out, very important in biology) to Dr. Hunter and Dr. Slamon who allowed me to first explore research within their labs in undergrad. Also thank you to my entire thesis committee for their insight and advice the past 5 years.

Thank you to Andrew, for taking me in as a new grad student right out of undergrad and replacing my nervousness and imposter syndrome with enthusiasm and support. Though I still don’t quite “Jump out of bed in the morning and scream F yeah!” (quote: Andrew Hsieh, 2017), each time we meet and talk science I get reinvigorated and excited about research. Thank you to the rest of the lab for creating a welcoming, close environment where we can do good science and also support each other as individuals and friends.

I wouldn’t have gotten this far in my PhD without the support of my friends, particularly Jeet Patel who supported and encouraged me through countless rough times and Jessica Gianopulos who I can always count on for a reenergizing chat over lunch. I’m so glad to have made friends throughout MCB and particularly our cohort, where we can share in our interests outside of science (love to our GBBO group Courtnee Clough and Jilliane Perkins) and support each other as we go through the stresses of graduate school together. A special shout out also to Alex Germanos because we spent a ridiculous amount of time together both living and

working together for 3 years and still came out close friends. Thank you to the MCB staff, Maia and Denise, for keeping our program running and the students feeling supported, this graduate experience wouldn't have been the same without you or in a different program.

And finally, a huge thank you to Pierce Eggan, for being perfect and loving and supportive, for taking me on great adventures to get away from it all and making a cozy home together so that every day is made better.

CHAPTER 1: INTRODUCTION

Portions of this chapter are adapted from: Schuster SL, Hsieh AC. The Untranslated Regions of mRNAs in Cancer. Trends Cancer. 2019 Apr;5(4):245-262. doi: 10.1016/j.trecan.2019.02.011. Epub 2019 Mar 22. PMID: 30961831; PMCID: PMC6465068.

1.1 Prostate cancer pathogenesis and genomics

1.1.1 The progression and treatment of prostate cancer

Prostate cancer (PCa) is the most commonly diagnosed cancer and the second-leading cause of cancer-related death in US men[1]. Despite prostate cancer being widespread, the prognosis for most men is favorable, with the overall 5-year survival rate being almost 97%. However, this risk is very dependent on staging, where the 73% of cases that are diagnosed at the localized stage have essentially a 100% 5-year survival rate, with treatment with surgery or localized radiation being typically curative. Unfortunately, these numbers drop to an only 32% 5-year survival rate in men diagnosed with metastatic disease[2]. Once patients have presented with metastatic disease, whether upon diagnosis or having recurred after localized treatment, systemic therapies are needed to control the disease. Overall, 20-30% of patients diagnosed with localized disease progress after surgery within 5 years, though this number can be over 50% for high-risk localized PCa patients[3–7].

Treatment-naïve prostate cancer is highly dependent on androgen signaling for growth, and therefore the first line of systemic therapy against advanced PCa is androgen deprivation therapy (ADT). Common forms of ADT include leuprolide, goserelin, buserelin, and others, all of which are gonadotropin-releasing hormone (GnRH) agonists. These GnRH activators increase the release of luteinizing hormone, which triggers androgen synthesis. While this initially causes a spike in androgen levels, negative feedback then inhibits the system, resulting in castrate levels of androgens, decreased cell proliferation, and prolonged patient survival. Other ADT options include the GnRH antagonist degarelix or physical castration, either of which can also

be used to decrease androgen levels in the body[8,9]. Though the initial results of ADT in patients are excellent, many tumors acquire resistance to such treatment, with 2-3 year durations of stable disease before progression. Prostate cancer that has progressed despite castrate levels of testosterone is termed castration-resistant PCa (CRPC).

Metastatic prostate cancer, particularly metastatic castration-resistant prostate cancer (mCRPC), is often considered incurable. Despite significant research into therapies for mCRPC, numerous clinical trials, and several FDA approvals, no treatment has been found that prolongs patient survival for more than a short period before resistance is again acquired and the patient eventually succumbs to the disease. These novel therapies include second-line hormonal therapies, PARP inhibitors, prostate-specific membrane antigen (PSMA)-based radionucleotides, and others[10]. Second-line hormonal therapies including abiraterone and enzalutamide target the androgen metabolism pathway in distinct ways from ADT, providing new ways to decrease androgen-based cell growth after ADT resistance. Abiraterone acetate is a CYP17 inhibitor, a key enzyme of the androgen synthesis pathway, thereby suppressing native testosterone production[11,12]. Enzalutamide blocks androgen signaling in several ways: it is a competitive inhibitor of testosterone binding to the androgen receptor (AR), inhibits AR translocation to the nucleus, and prevents AR from binding target DNA sites[13]. These AR-targeting therapies are effective in CRPC because even ADT-resistant tumors still often utilize the AR pathway to increase cell proliferation by upregulating AR expression. Therefore, targeting other nodes of this pathway can still be effective[14]. Both abiraterone and enzalutamide are now considered standard care for mCRPC, increasing patient survival by medians of 2-5 months depending on the trial context[10].

For patients harboring mutations in homologous recombination repair (HRR) pathway genes, such as ATM, BRCA1, BRCA2, and others, PARP inhibitors are effective gene-targeted therapies for advanced disease. These treatments work via synthetic lethality, where the decreased function of DNA repair upon HRR mutation sensitizes cells to drugs targeting other

DNA repair pathways. Two PARP inhibitors, rucaparib and olaparib, are now FDA approved for mCRPC treatment[15–17]. For patients without such precision medicine treatments available, targeted radionucleotides can be an option. One such treatment strategy delivers beta-particle radiation to cells of prostate origin by tethering Lutetium-177 to PSMA, a membrane protein highly expressed on prostate cancer cells[18]. Alternatively, radium-223 can be used to suppress metastases, as this alpha emitter selectively traffics to regions of increased bone growth, effectively treating the most common site of metastasis in prostate cancer[19]. Other treatments of advanced PCa include the chemotherapy docetaxel and the immunotherapy sipuleucel-T, both of which have been found to extend mCRPC patient survival[7,20]. Despite these many advancements in the treatment of mCRPC, aggressive prostate cancer tumors invariably acquire resistance to therapies, leading to the estimated ~35,000 deaths annually from PCa in the US[1]. The wide-ranging mechanisms of resistance to therapy suggest further need to understand the underlying genetic heterogeneity of prostate cancer.

1.1.2 Genomics of prostate cancer

Prostate cancer has a relatively low burden of mutations compared to many other cancers, with a median of approximately 1 mutation per megabase[21]. In primary prostate cancer, the most prevalent genomic aberrations are *ETS* gene family fusions, which are found in ~50% of cases[22,23]. The *TMPRSS2-ERG* fusion accounts for 40-50% of these, wherein the 5' end of the androgen responsive *TMPRSS2* gene is fused with the transcription factor *ERG*, driving increased oncogenic ERG expression[24]. Additional common mutations in localized PCa include point mutations in *SPOP* (8-11%) and *FOXA1* (2-3%), each of which are associated with higher AR transcriptional activity. Approximately 45% of localized PCa tumors contain mutations that may be targeted by available genetic therapies, with 19% of patients harboring mutations in DNA repair genes that make them candidates for PARP inhibitors and 25% of patients having alterations within the druggable RAS/PI3K pathway[22,25].

The most striking difference between the genomics of localized PCa and mCRPC is the novel amplification of *AR*, which is present in 60-80% of aggressive tumors[26,27]. This amplification, whether of *AR* itself, its promoter, or a distant enhancer, succeeds in driving *AR* expression to such high levels that androgen deprivation therapy is no longer effective[28]. Though most AR pathway mutations are of *AR* itself, an additional 10-20% of patients harbor mutations in FOXA1, SPOP, NCOR1/2, or ZBTB16, diversifying the mechanisms by which castration resistance can be acquired. Additionally, loss of tumor suppressors is more common in advanced PCa than localized, with ~50% of patients harboring *TP53* loss and 40% with *PTEN* loss.

It is also important to note that there are significant differences in prostate cancer genomics between patient populations with different ancestry. Men of African descent are less likely than European ancestry patients to harbor *ETS*-family mutations including the *TMPRSS2-ERG* fusion, while they display a higher percentage of mutations to *FOXA1, SPOP,* and DNA damage repair genes[29,30]. Chinese patients also have markedly higher incidence of *FOXA1* mutations than noted in white populations[31]. Diversifying the patient population in both genomic studies and clinical trials of prostate cancer will be an important part in understanding genomic heterogeneity as it stems from ancestry as well as across cancer progression.

While these genomic studies of both localized and advanced PCa give significant insight as to resistance mechanisms and possible therapeutic opportunities, currently established driver mutations are not found in most tumors. In localized disease, ~26% of tumors could not be assigned to a major mutational category, and in advanced PCa, about 10% of patients did not harbor any potentially actionable mutations. Importantly, with PARP inhibitors being the only approved precision medicine in prostate cancer, the number of tumors with treatable driver mutations is even lower. This failure to find major driver mutations in a subset of patients could be attributed to these tumors harboring less recurrent mutations that are nonetheless important in small numbers of patients. This idea has been bolstered by the discovery of the “long tail” of

oncogenic drivers in PCa, wherein detailed analysis of a very large cohort of patients allowed the SU2C International Prostate Cancer Dream Team to find many previously unknown potential driver mutations with incidences of <3%[32]. Intriguingly, this long tail of significant mutations may also include non-coding mutations, which are often missed by the use of exome sequencing versus whole-genome sequencing (WGS). For a group of 200 intermediate risk, localized tumors that underwent WGS, 70 such non-coding mutations were found with incidences over 2%, placing them around similar recurrence of FOXA1 and TP53 in this cohort[23]. Further genomic analysis beyond the exome with focus on non-coding mutations, epigenetics, and post-transcriptional regulation are likely the next frontiers of understanding advanced prostate cancer.

1.2 Regulatory functions of the 3'UTR in normal physiology and cancer

1.2.1 Introduction to the 3' untranslated region

Post-transcriptional processes account for approximately 60% of variation in protein expression[33], and as such are vital for our complete understanding of proteome diversity. The 3' untranslated region (3'UTR) is the domain downstream of the main coding sequence (CDS) of messenger RNA (mRNA) that controls critical post-transcriptional gene regulation processes. The 3'UTR contains a myriad of regulatory elements involved in mRNA processing, mRNA stability, and translation initiation. Moreover, emerging evidence suggests that the 3'UTR can be either directly mutated or co-opted in diseases such as cancer[34,35]. Nevertheless, most studies of cancer genomics have only focused on genetic aberrations of protein-coding regions of the genome[36,37]. This is largely because of the high cost of whole-genome sequencing in comparison to targeted exome sequencing, which under-captures UTRs. Though CDS-focused studies have undoubtedly increased our understanding of many cancers, these approaches have found actionable mutations in only 57% of tumors[36]. Therefore, the 3'UTR may harbor yet undiscovered drivers of cancer pathogenesis at the post-transcriptional level.

Despite the critical role of the 3'UTR in controlling gene regulation, some studies argue that UTRs are not functionally important in cancer because regulatory region mutations rarely have effect sizes as large as protein-coding mutations[38–40]. However, many other studies have found evidence for the functional and clinical relevance of UTR mutations in cancer, both in genome-wide and gene-specific analyses. Whole-genome sequencing has identified areas of recurrent mutations across the UTRs of many cancer-related genes, arguing for extensive deregulation of UTR function in disease[34,35,41]. In addition to whole-genome studies, many targeted analyses have been performed of particular UTR regulatory elements, providing functional evidence for the significance of 3'UTR mutations[42–45]. In this section, I will explore the regulatory functions of the 3'UTR, how this region can be hijacked in cancer, and recent advances in technology for studying UTR elements.

1.2.2 Cis- *and* trans-*regulatory control of cancerous gene expression in the 3'UTR*

The 3'UTR primarily functions through a dynamic interplay between sequence and structural motifs, collectively called *cis*-regulatory elements, and RNA-binding protein (RBPs) and small RNAs called *trans*-acting factors (**Figure 1-1A**). Together, they can shape the cellular proteome by tuning the metabolism and translation of specific mRNAs. Importantly, deregulation of *cis*-regulatory elements can drive cancer pathogenesis by inducing oncogenic gene expression. Here I will highlight a series of sequence-based and structural *cis*-elements and their related *trans*-acting factors implicated in cancer.

AU-rich elements (AREs) are one of the most widely studied mRNA *cis*-regulatory elements. It has been shown that ARE-containing mRNAs are overexpressed in cancer and may play a role in mitotic progression[46]. Numerous RBPs have been observed to bind AREs, causing variable effects on mRNA stability or translation, with HuR (*ELAVL1*) and AUF1 (*HNRNPD*) being two of the most well-studied[47–49]. HuR is overexpressed in many cancers and associated with poor patient outcomes, and may have potential as a therapeutic target[50]. While

HuR stabilizes its target mRNAs[51,52], AUF1 usually acts through increasing mRNA decay, though its mechanism has proven to be variable across targets and systems, both increasing and decreasing either mRNA stability or translation depending on the context[53]. Yet another ARE-binding RBP is FXR1, which has been found to simultaneously stabilize the *cMYC* mRNA and increase its translation[54]. In this way, FXR1 is oncogenic, with high expression associated with poor patient prognosis and necessary for tumor growth and metastasis *in vivo*. This variability in both the *trans*-acting factor recruited to a given ARE site and the downstream effect of this interaction underscores the complexity of 3'UTR-mediated post-transcriptional gene regulation. Even given the presence of a known *cis*-regulatory element like an ARE within a 3'UTR sequence, it is difficult to determine how that element may affect mRNA dynamics.

Though AREs were one of the first discovered 3'UTR regulatory elements[55,56], many others have been described. For example, the guanine-rich element (GRE) promotes mRNA translation through interactions with the RBP CELF1. Interestingly, the GRE is enriched in the 3'UTRs of 10 epithelial-to-mesenchymal transition (EMT) related genes and CELF1 is upregulated in breast cancer tissues and promotes lung metastasis in xenograft models[57]. In addition to sequence-based motifs, RBPs can also bind specific structural elements (**Figure 1-1B**). For example, the BAT element is a structural motif consisting of a stem-loop with an asymmetrical bulge. This structural element represses translation of two EMT-related genes, *Dab2* and *ILEI*, via the binding of heterologous nuclear protein (hnRNP) E1. TGFβ signaling can phosphorylate hnRNP E1, releasing it from the stem-loop, increasing protein expression of these genes, and promoting EMT[58].

These sequence- and structure-specific motifs demonstrate the functional consequences of UTR-mediated deregulation in cancer. The following sections will highlight additional examples of well-studied and emerging 3'UTR *trans*-acting factors and *cis*-elements that can promote cancer phenotypes. Importantly, many fundamental questions about *cis*-regulatory elements still remain. For example, it is unknown how commonly these motifs are somatically

mutated in cancer. Moreover, the mechanisms of many elements and what drives their selectivity in disease are open questions. Lastly, some of these elements can be found within the same mRNAs, raising the intriguing potential of *cis*- on *cis*- interactions that will require novel experimental designs to unlock their regulatory logic.

1.2.3 Deregulation of 3'UTR miRNA binding sites in cancer

MicroRNAs are one of the most extensively studied 3'UTR-mediated post-transcriptional regulators. These non-protein-coding RNAs are 20-22 nucleotides long and generally cause decreases in protein expression through translation repression followed by mRNA decay (**Figure 1-1C**)[59,60]. miRNAs are produced from precursor RNAs that are processed by the Drosha-DGCR8 complex and cleaved by the protein Dicer. The mature miRNA is then incorporated into the miRNA-induced silencing complex (miRISC) to functionally target and regulate mRNAs[61]. miRNAs exert their control over the translation of specific genes through complementary base-pairing between the miRNA and target mRNAs. Though the vast majority of miRNAs target sequences in the 3'UTR of mRNAs, they can also bind to the 5'UTR[62] or within coding sequences[63]. Since miRNAs act through base-pairing of a relatively short sequence, each miRNA can have dozens to hundreds of targets throughout the genome, allowing a single miRNA to coordinately control functionally related pathways. miRNAs have been predicted to regulate the activity of >60% of all protein-coding genes in humans[64]. This abundance of target genes, along with their widespread conservation across eukaryotes, points to the ubiquitous importance of miRNAs in post-transcriptional gene regulation and hints at the numerous ways this tightly regulated control could go awry in cancer.

miRNAs have been extensively studied in the context of disease and cancer (reviewed in [61,65–68]). Expression of many miRNAs have been correlated with cancer development and prognosis in both oncogenic and tumor suppressive capacities. These include the classical "oncomiRs" miR-21[69] and the miR-17-92[70] family, which can induce or exacerbate tumor

formation in mouse models, respectively. One example of a tumor suppressive miRNA is miRNA-34a, which has been shown to impair progenitor cell function and metastatic potential in prostate cancer models[71]. Intriguingly, a small molecule therapeutic has been discovered that can increase miRNA-34a expression and inhibit tumor growth *in vivo*[72], suggesting this form of UTR-mediated regulation can provide opportunities for therapeutic development. It is important to note that in some cases, the same miRNA can have both oncogenic and tumor suppressive functions depending on cellular context, which reveals the intricate complexities of miRNA-mediated control of gene expression networks[73,74].

While much is known about the oncogenic and tumor suppressive functions of particular miRNAs, exceedingly little is known about how somatic mutations in miRNA binding sites of target gene 3'UTRs can impact tumorigenic gene expression and cancer phenotypes. Here I will highlight one known example of how genetic alterations within the 3'UTR can drive aberrant gene expression in cancer and speculate on the future expansion of this field. Targeted 3'UTR sequencing of genes previously associated with colorectal cancer (CRC) in a cohort of CRC patients uncovered a somatic mutation in the miR-136-5p target region of the *E2F1* 3'UTR[75]. E2F1 is an oncogenic transcription factor that regulates the cell cycle and is associated with poor prognosis in patients[76]. This novel 3'UTR mutation was associated with 4-fold increase in expression of *E2F1* in the patient. Moreover, it was shown that the mutated *E2F1* 3'UTR increased gene expression in reporter assays. It remains to be seen whether this mutation would produce oncogenic phenotypes *in vivo*; nonetheless, its discovery illustrates that non-coding mutations of specific 3'UTR miRNA binding sites may regulate expression of known oncogenes.

The recent increase in cancer whole-genome sequencing enhances the potential for discovering somatic 3'UTR mutations that impact miRNA binding. For example, one study analyzed 3'UTR somatic mutations across four cancer types to uncover alterations that disrupt miRNA targeting[45]. 3'UTR somatic mutations were discovered that altered miRNA target sites in

cancer-associated genes such as *TAL1*, *MITF*, *EPHA3*, *BMPR1B*, and *KDM5A*. However, these have yet to be functionally tested. Additionally, large-scale computational analysis of 3'UTR sequencing data has been coupled with experimentally-validated miRNA binding sites to create a database containing somatic mutations predicted *in silico* to affect miRNA binding[77]. Novel large-scale functional assays will be necessary to determine the causal effects of these mutations on RNA metabolism, protein expression, and cancer phenotypes.

1.2.4 Alternative 3'UTR isoforms in cancer

Alternative polyadenylation (APA) is a mechanism of gene regulation conserved across many eukaryotes that occurs through the differential usage of multiple polyadenylation sites within an mRNA to change the length of the 3'UTR on the mature transcript. The polyadenylation site (PAS) is marked by several upstream and downstream consensus sequences, including the core AAUAAA hexamer, that indicate where the transcript will be cleaved and polyadenylated. Importantly, the similarity of a sequence to the canonical motif determines the strength of each polyadenylation site[78]. Often, the PAS closer to the coding sequence (the proximal PAS) has weaker consensus sequences, and more 3' PASs (distal PASs) have stronger consensus sequences[79]. This allows for global regulation of 3'UTR length based on conditions that favor the use of strong versus weak consensus sequences. By changing the site at which precursor mRNA is cleaved and polyadenylated, alternative polyadenylation changes the length of a transcript's 3'UTR and consequently includes or excludes *cis*-regulatory elements that regulate gene expression. Since many regulatory elements within the 3'UTR destabilize mRNA or repress translation, longer 3'UTRs generally lower protein expression, whereas shorter 3'UTRs increase protein expression (**Figure 1D**). Illustratively, more than 50% of conserved microRNA binding sites are located downstream of proximal poly(A) sites and therefore are only active when the distal PAS is used over the proximal PAS[80,81].

Interestingly, cancer cells generally exhibit global 3'UTR shortening. One analysis across seven tumor types found 91% of genes that underwent differential APA between matched tumor and normal tissue expressed shorter transcripts in the context of cancer[82,83]. It is suggested that this systemic change helps proto-oncogenes escape regulation by suppressive elements in their 3'UTRs that normally maintain balanced expression and prevent malignancy. For example, APA in triple-negative breast cancer patients leads to 3'UTR shortening of the oncogenes *c-JUN* and *NRAS*[84]. This can result in the exclusion of inhibitory Pumilio complex recognition sites from the 3'UTR, which enables the persistent expression of these oncogenic transcripts. Interestingly, in this study, increased APA was associated with a higher incidence of lymph node invasion. In addition to increasing expression through transcript shortening, APA can repress expression by promoting longer isoforms of particular genes. An example of this in glioblastoma involves MGMT, which is a DNA repair enzyme associated with resistance to alkylating agents, such as temazolamide. Unlike normal brain tissue, gliomas utilize APA to express an isoform of *MGMT* with a longer 3'UTR that mediates lower mRNA levels of *MGMT* in part due to addition of miRNA binding sites[85]. This change in expression is associated with increased sensitivity to temazolamide *in vitro*. Though these examples highlight how of 3'UTR shortening increases gene expression, it is important to note that some 3'UTR elements are stabilizing[86] and therefore shortening of particular 3'UTRs may result in decreased expression.

Proximal versus distal poly(A) site usage and subsequent changes in 3'UTR length can be regulated by a variety of factors. In general, more abundant APA machinery enables the use of weaker poly(A) sites and shifts the balance towards usage of proximal, noncognate PAS and shorter 3'UTRs. In particular, upregulation of the polyadenylation factor CSTF64, an RNA-binding subunit of the cleavage stimulation factor (CSTF) complex, is associated with 3'UTR shortening in TCGA data across five tumor types[83]. Though upregulation of APA machinery like CSTF64 and CPSF3 promotes proximal PAS usage, other studies have discovered APA factors that promote distal PAS usage, including PABPN1, CFIm25, and CFIm68[87–90]. In glioblastoma,

3’UTR shortening is mediated by downregulation of the Cleavage Factor Im (CFIm) complex subunit CFIm25. Depletion of CFIm25 increases growth and invasive ability *in vitro*, as well as drastically increases the size of tumors in a subcutaneous glioblastoma mouse model[91]. These studies raise the important questions of how competing APA signals are integrated to coordinate the regulation of mRNA targets and induce cancer phenotypes.

In addition to alternative polyadenylation, differential 3’UTR splicing can be a mechanism of altering the 3’UTR isoform present in cells. While splicing of 3’UTR introns has not been widely studied, a newfound interest in these events has followed a recent study discovering their widespread importance in cancer[92]. This study of 3’UTR splicing across 10 cancer types determined that such events, similarly to APA, are upregulated in cancer, particularly in known oncogenes. They demonstrate that a particularly pervasive 3’UTR splicing variant in the *CTNNB1* 3’UTR contributes significantly to its oncogenicity in hepatocellular carcinoma and colon adenocarcinoma. As such, APA and 3’UTR splicing are two mechanistically different processes that each lead to the diversification of 3’UTRs possible for a single gene, thereby altering the regulatory elements present on the mRNA.

1.2.5 Dynamic UTR modifications in cancer

Post-transcriptional modifications of mRNA have been known to occur for decades, since the discovery of N^6-methyladenosine (m^6A), 5-methylcytidine (m^5C) and 2’O methylation in the 1970s[93–95]. Recent advances in whole-transcriptome, high-resolution techniques to map these modifications in response to changing cellular states have revealed their dynamic nature and invigorated the field[96–99]. Furthermore, the functional identification of some of the major regulators of these processes has opened the field to deeper study of epitranscriptomic changes and their importance as modulators of gene expression.

The most abundant and well-studied mRNA modification is m^6A[93–95], which is enriched in 3’UTRs[96], though also present in 5’UTRs[100]. This modification has been extensively reviewed in

[101,102]. In short, the m^{6}A machinery includes the “writer” methyltransferase complex composed of two core catalytic subunits METTL3[103] and METTL14[104–107] and several non-catalytic subunits[108], which are variably required to add methyl marks to the nitrogen-6 position of adenosine. These are countered by two AlkB family dioxygenases capable of erasing the m^{6}A mark, FTO[109] and ALKBH5[110]. The m^{6}A mark is read by at least five YTH-domain family proteins: YTHDF1-3, which are primarily cytoplasmic, and YTHDC1 and YTHDC2, which are nuclear[111–114], that mediate a variety of effects on gene expression. For example, YTHDF2 can either promote cap-independent translation[100] or mRNA degradation[113,115] by binding 5’UTR or 3’UTR m^{6}A sites, respectively. The readers YTHDF1[112] and YTHDF3[116] promote mRNA translation initiation, and YTHDC1[117] can drive alternative splicing. m^{6}A marks have also been shown to have roles in alternative polyadenylation[118] and changing mRNA structure and subsequent affinity for RNA-binding proteins[119,120] (**Figure 1-1E**).

As modifications that fine-tune mRNA metabolism and translation, particularly in response to cellular stresses that often exist in the tumor microenvironment[100], m^{6}A marks can play major roles in deregulating cancerous gene expression. Overexpression of the methyltransferases METTL3 and METTL14 have been implicated in maintaining stemness in cancer cell populations. Knockdown or mutation of either of these genes in human leukemia cells decreases cell growth, induces apoptosis, and delays the development of leukemia when transplanted into mice[121,122]. Additionally, high METTL3 levels in glioblastoma enhance SOX2 m^{6}A modification, mRNA stability, and stemness[123]. Interestingly, m^{6}A demethylation machinery has similarly been linked to maintenance of cancer stem cells. High expression of ALKBH5 correlates with poor patient prognosis in glioblastoma[124] and increased FTO expression has been found in a subset of AMLs[125]. Knockdown of these demethylases in models of their respective tumor types decreases both cell proliferation and stem-cell features like colony formation *in vitro* and decreases tumorigenicity *in vivo*. In breast cancer, ALKBH5 has been found to have a role in the response to hypoxia in the tumor environment. In particular, exposure

to hypoxia increases HIF1α/HIF2α-dependent expression of ALKBH5, leading to NANOG 3'UTR demethylation, increased NANOG transcript stabilization, and breast cancer stem cell enrichment[126]. These results illustrate the potential clinical importance of epitranscriptomic gene regulation, where these complexes may represent a new approach to target therapy-resistant cancer stem cells. However, the fact that opposing machinery can drive similar phenotypes warrants additional studies regarding the specificity of each "writer" and "eraser" in cancer.

There are other post-transcriptional mRNA modifications of the UTR that are likely deregulated in cancer, though they are considerably less well studied. These include N^1-methyladenosine (m^1A) and pseudouridine, both of which were first discovered on tRNA and rRNA, but more recently found and mapped on eukaryotic mRNA[97,127,128]. The functions of these RNA modifications on mRNA are not entirely clear, though m^1A marks positively correlate with protein levels and there is evidence they can re-structure 5'UTRs to influence translation[129]. Pseudouridines stabilize tRNA and rRNA structure, and as such may also influence mRNA structure-related gene expression. Though the *in vivo* relevance is unclear, *in vitro* modification of mRNA uridine to pseudouridine has been shown to enhance translation by decreasing PKR activation and eIF2α phosphorylation[130]. Both m^1A and pseudouridine modifications change in response to cellular stress such as heat shock and nutrient deprivation[97,128,131], indicating that these are dynamic marks on mRNA that could contribute to cancer pathogenesis. The advent of high-resolution mapping techniques coupled with genomic screening tools may provide a means to delineate their function in normal and disease states[96,97,99,127,128,131–133].

1.2.6 Downstream open reading frames may represent novel mechanisms for translation regulation

While upstream open reading frames (uORFs) placed within 5'UTRs have been studied for decades and are widely known to regulate the expression of main coding sequences[134,135], downstream ORFs are significantly newer to the field. These ORFs have been noted in

ribosome profiling experiments[136–138], but hadn't been studied in detail until recently[139]. While dORFs are not usually conserved between species, their presence consistently increases the translation of the main ORF encoded on the same mRNA (**Figure 1-1F**). This upregulation of translation seems to be dependent on the number of dORFs present on an mRNA, but not by their peptide sequence or length. Interestingly, two dORFs have been found to regulate cell proliferation[140], opening questions as to how such elements function to regulate either normal cellular homeostasis or may be co-opted in disease. The mechanism of how dORFs regulate the translation of their upstream counterparts and whether this is a conserved phenomenon remain open questions.

1.3 Tools for investigation of the 3'UTR

1.3.1 Studying UTRs in cancer on a whole-genome scale

The importance of UTRs as post-transcriptional regulatory regions in cancer has recently been expanding in earnest. Even so, research into these areas still lags far behind our knowledge and understanding of protein-coding regions, and there are still many unanswered questions for the field. Though attention in cancer genomics has been shifting to whole-genome sequencing, most studies of non-coding regulatory regions still primarily focus on promoter elements, neglecting the importance of UTRs and post-transcriptional regulation. This stems from the challenge posed by our incomplete understanding of UTR-mediated mechanisms of gene regulation as compared to the CDS or promoters. Whereas assigning function to CDS mutations is relatively straightforward using the knowledge of how the mutation changes protein biochemistry, the path between a non-coding mutation and its potential effects is often murky. To determine the relationship between a UTR mutation and its function, we must answer: Does it affect a known *cis*-element or structural motif? Does it functionally alter post-transcriptional regulation? Does it work combinatorially with any other regulatory element or mutation? These

questions could be addressed by adapting analyses currently performed for promoter and coding regions for UTR studies.

Several analyses in recent years have used published whole-genome sequencing data to identify recurrently mutated non-coding regions in cancer and uncovered or confirmed promoter and enhancer elements with potential relevance to cancer[35,38,141,142,149]. These computational pipelines, some of which include searching for promoter transcription factor binding sites or comparison with RNA-seq to identify functional transcriptional differences, could be adapted to analyze mutations of known UTR *cis*-elements. Other new technologies, such as mRNA structural analysis through sequencing-based platforms[143–146], measurement of translation initiation and efficiency through ribosome profiling[147], or analysis of RBP binding to mRNA[148] could also be combined with this mutational analysis to correlate functional consequences with cancer-specific UTR mutations. The PCAWG dataset has recently been used to achieve a pan-cancer view of non-coding mutations, wherein they discover recurrent mutations in both the 5'UTR and 3'UTR associated with significant changes in oncogenic expression[149]. Such pan-cancer analyses are important resources for the field. One recent study building on this dataset used an algorithm originally designed to study synonymous CDS mutations to detect 3'UTR cancer mutations predicted to disrupt protein-RNA interactions and thereby regulate gene expression within the PCAWG dataset[150]. They found that high burdens of these disruptive 3'UTR SNVs are more prevalent in cancer than normal tissue and associated with poor patient prognosis; however, very limited functional experiments were done to test how these mutations were working. Despite these intriguing findings and the many instances of UTR-mediated cancer oncogenicity noted above, both the PCAWG study and others note that there are few non-coding driver mutations in cancer with effect sizes or recurrency similar to known protein-coding mutations[40]. Genome-wide mutational analysis of the UTRs combined with functional correlates are a great addition to our understanding of UTR biology in cancer; however, direct functional analysis and validation of UTR elements and such mutations is still

lacking. Moreover, it is unknown how the majority of these mutations contribute to disease outcomes.

1.3.2 Functional genomics of UTR biology

Recently developed massively parallel reporter assays (MPRAs) for 3'UTRs, which can measure the effect of UTR elements on a variety of functional readouts of gene expression from transcript stability to translational efficiency, may help bridge the gap between regulatory cancer genomics and functional consequence (**Table 1-1**). Such MPRAs are powerful tools in which a library of reporter plasmids is created combining thousands of different 3'UTRs with a single control coding sequence, then introduced into cells and reporter expression measured. In this way, the effect of individual UTRs or UTR elements can be assessed by differences in reporter gene expression. Through the use of different expression readouts, such as RNAseq, polysome profiling, or flow cytometry, MPRAs can answer questions about how UTRs can affect specific aspects of gene regulation (RNA expression, translation efficiency, and protein expression, respectively). Though many of these UTR-based MPRAs have been published, most have focused on delineating regulatory dynamics in normal settings, measuring the effects of randomized sequences, normal UTRs, or known regulatory elements. To our knowledge, the only MPRAs conducted to date studying disease-relevant UTR regulation are MPRAu[151], out of the Sabeti lab, and both PLUMAGE[152] and my own study out of the Hsieh lab. MPRAu assessed the effect of thousands of 3'UTR mutations with significance in genome-wide association studies, evolutionary adaptation, or cancer, finding over 2,000 UTR variants that altered gene expression. One limitation of this study is that the readout was steady-state RNA-seq, and therefore it is unknown how any UTR variants affect mRNA translation or mRNA stability versus transcription. Though MPRAs are useful to causally link UTR elements with their effect on gene expression, the use of exogenous reporter plasmids and fragmented UTR sequences means that complex effects between the 5'UTR, CDS, and 3'UTR are lost. Additionally, these methods

cannot be used to explore cellular effects of UTR manipulation, such as changes in cell state or growth.

In order to explore the endogenous and cellular effects of UTR regulation, research must turn to cutting-edge CRISPR tools. New technologies such as CRISPR base editing and prime editing make functional studies relating endogenous *cis*-elements and cancer-related outcomes possible[153]. It is now possible to orthogonally validate reporter assay results using genomic CRISPR editing, creating a new gold standard in the field[152,154]. Additionally, CRISPR editing screens provide exciting new avenues for high-throughput exploration of the endogenous cellular effects of single point mutations. Multiple such screens have been performed using base editing or earlier homologous-directed repair-based CRISPR systems to tile across important oncogenes such as BRCA1, BRCA2, and JAK1, assigning functional meaning to previous variants of unknown significance in these genes[155–158]. Though similar to earlier deep mutational scanning approaches, these CRISPR editing screens are now able to assess the effect of mutations located within their endogenous genomic loci instead of from exogenously introduced plasmid or viral systems, thus increasing their applicability to actual disease states. Two tandem *Cell* publications in 2021 used cytosine base editing to go even further, using sgRNA libraries covering thousands of variants across the genome[159,160]. Hanna *et al*. assessed the function of over 52,000 ClinVar variants in 3,584 genes to determine how these mutations affected both normal cell growth and resistance to stress conditions. Cuella-Martin *et al*. investigated how mutations across 86 genes involved in DNA damage response affect cellular response to DNA damaging agents, discovering novel mutations of potential clinical significance. These base editing screens exemplify the power of this technology to assign functional significance to unstudied mutations and differentiate between passenger and driver mutations in cancer. It remains to be seen how our view of non-coding mutations may expand and change when such tools are adapted beyond the coding region into promoters, enhancers, UTRs, and other regulatory regions of the genome. In conjunction with clinical data, these assays may decode a

new layer of driving genetic events of poorly understood or incurable cancers. As such, the UTRs represent a new frontier in post-transcriptional gene regulation in health and disease with great potential for the discovery of new biology and therapeutic opportunities.

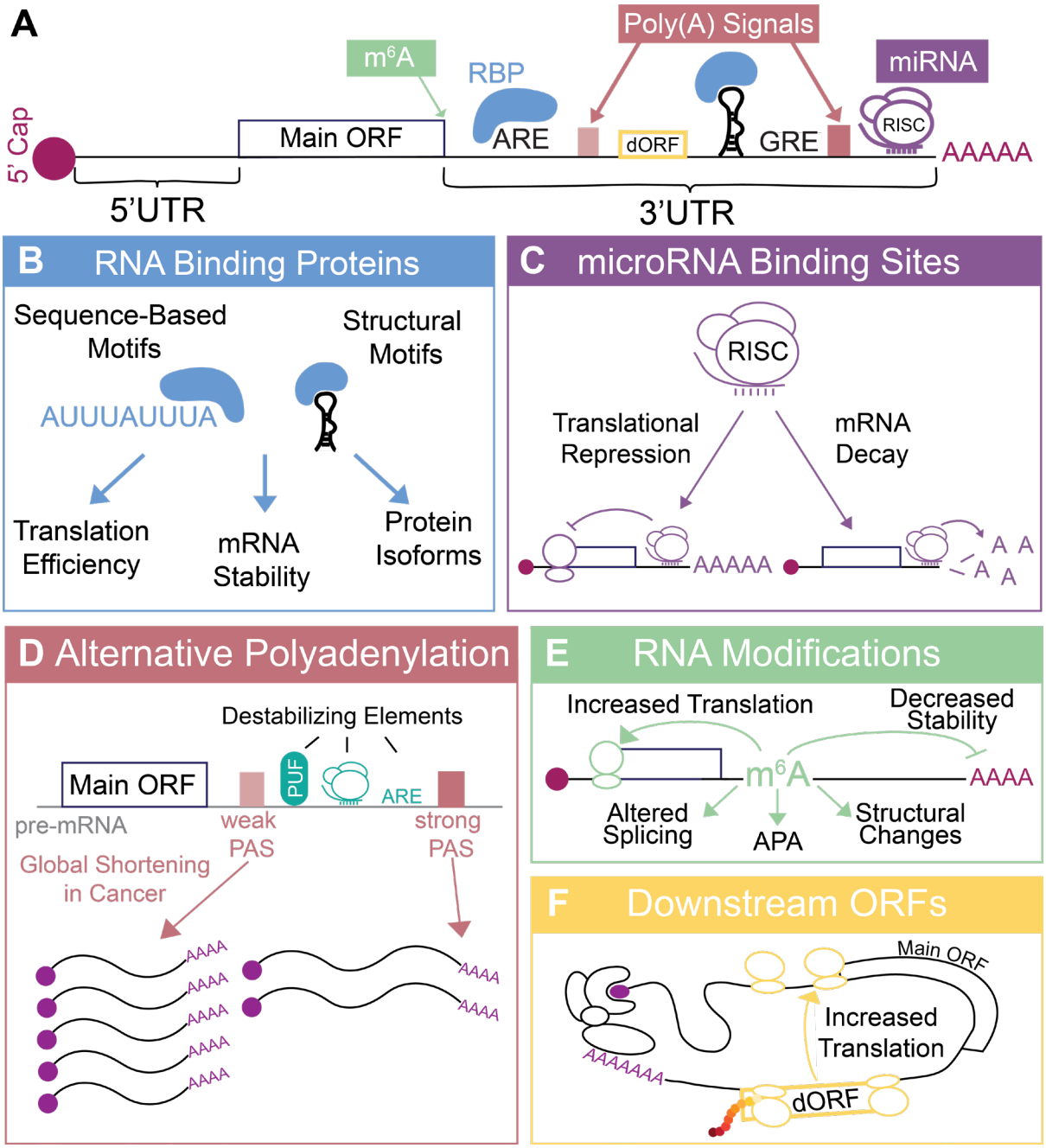

Figure 1-1: mRNA regulation by 3'UTR elements

(A) Schematic of an mRNA highlighting various regulatory elements within the 3'UTR: m^{6}A modification, a downstream ORF (dORF), two alternative polyadenylation (polyA) signals, miRNA binding site occupied by the RNA-indued silencing complex (RISC), and multiple sequence and structural elements bound by RNA-binding proteins (RBPs). **(B)** RNA binding proteins can bind to sequence-based or structural motifs to affect a variety of processes. **(C)** MicroRNAs (miRNAs) integrated with the RISC cause translational repression and/or mRNA decay. **(D)** Multiple polyadenylation signals (PAS) in the same 3'UTR can give rise to differentially cleaved and polyadenylated transcripts, affecting downstream gene expression. Often, use of proximal PASs increases gene expression versus distal PASs. **(E)** RNA modifications such as m^{6}A can mediate a variety of downstream effects. **(F)** Downstream ORFs increase the translation of the preceding main ORF in a translation-dependent manner.

Table 1-1: Massively parallel reporter assays designed for investigation of the 3'UTR.

Sequence Length	Sequence Source	Measurement	Experimental Design	Citation
200mer	Random combinations of known 3'UTR *cis*-elements	Transcript abundance, translation initiation & efficiency	RNA-seq & ribosome profiling of plasmid library in HeLa cells	[161]
110mer	3'UTRs of ~7,000 zebrafish transcripts	mRNA degradation	RNA-seq over time course of mRNA libraries in zebrafish embryos	[162]
8mer	Random sequence inserted into endogenous 3'UTR	Protein expression, translation efficiency	FACS + DNA-seq of genomically-integrated dual fluorescent reporters in 293 cells	[163]
34mer	3'UTR sequences conserved across vertebrates	Protein expression, translation efficiency	FACS + DNA-seq of genomically-integrated dual fluorescent reporters in 293 cells	[164]
200mer + 160mer	Deep mutational scanning of active 3'UTR ARE + full coverage of *CXCL2* + highly conserved sequences	mRNA stability, protein expression, translation efficiency	RNA-seq over time course & FACS + DNA-seq of plasmid library transfected into human epithelial cells	[165]
25-45mer	>3 million unique: PASs from human genome + other randomized seqs	Alternative polyadenylation	RNA-seq from 293T cells	[166]
70mer	Protein-occupied sites	Transcript abundance	RNA-seq normalized to plasmid DNA-seq of transduced primary mouse T cells	[167]
160mer	ARE or CDE containing 3'UTR segments	mRNA Stability + steady-state abundance	RNA-seq time-course of dox-treated Jurkat and Beas2B cells transduced with tet-responsive plasmid	[168]
162mer	miRNA binding sequences in various native contexts with inserted mutations	Protein abundance	FACS + DNA-seq in stably integrated K562 cells	[169]
162mer	Native and mutated PASs	RNA expression + alternative polyadenylation	RNA-seq from transfected K562 cells, normalized to plasmid DNA-seq	[170]
100mer	Wildtype vs mutant 3'UTRs with mutations from GWAS, ClinVar, and evolutionarily selected sequences		RNA-seq normalized to plasmid DNA-seq from transfected HEK293, HepG2, CM12878, SK-N-SH, K562, and HMEC cells	[151]

CHAPTER 1 REFERENCES

1. Siegel, R.L., Miller, K.D., Fuchs, H.E., and Jemal, A. (2022). Cancer statistics, 2022. CA. Cancer J. Clin. *72*, 7–33. 10.3322/CAAC.21708.
2. Institute, N.C. SEER Cancer Stat Facts: Prostate Cancer. https://seer.cancer.gov/statfacts/html/prost.html.
3. Berge, V., Berg, R.E., Hoff, J.R., Wessel, N., Svindland, A., Karlsen, S.J., and Eri, L.M. (2012). Five-year progression-free survival in 577 patients operated on with laparoscopic radical prostatectomy for localized prostate cancer. https://doi.org/10.3109/00365599.2011.604790 *46*, 8–13. 10.3109/00365599.2011.604790.
4. Sundi, D., Wang, V.M., Pierorazio, P.M., Han, M., Bivalacqua, T.J., Ball, M.W., Antonarakis, E.S., Partin, A.W., Schaeffer, E.M., and Ross, A.E. (2013). Very-high-risk localized prostate cancer: definition and outcomes. Prostate Cancer Prostatic Dis. 2014 171 *17*, 57–63. 10.1038/pcan.2013.46.
5. Buyyounouski, M.K., Pickles, T., Kestin, L.L., Allison, R., and Williams, S.G. (2012). Validating the interval to biochemical failure for the identification of potentially lethal prostate cancer. J. Clin. Oncol. *30*, 1857–1863. 10.1200/JCO.2011.35.1924.
6. Studer, U.E., Whelan, P., Albrecht, W., Casselman, J., De Reijke, T., Hauri, D., Loidl, W., Isorna, S., Sundaram, S.K., Debois, M., et al. (2006). Immediate or deferred androgen deprivation for patients with prostate cancer not suitable for local treatment with curative intent: European organisation for research and treatment of cancer (EORTC) trial 30891. J. Clin. Oncol. *24*, 1868–1876. 10.1200/JCO.2005.04.7423.
7. Kantoff, P.W., Higano, C.S., Shore, N.D., Berger, E.R., Small, E.J., Penson, D.F., Redfern, C.H., Ferrari, A.C., Dreicer, R., Sims, R.B., et al. (2010). Sipuleucel-T Immunotherapy for Castration-Resistant Prostate Cancer. N. Engl. J. Med. *363*, 411–422. 10.1056/NEJMOA1001294/SUPPL_FILE/NEJMOA1001294_DISCLOSURES.PDF.
8. Harris, W.P., Mostaghel, E.A., Nelson, P.S., and Montgomery, B. (2009). Androgen deprivation therapy: progress in understanding mechanisms of resistance and optimizing androgen depletion. Nat. Clin. Pract. Urol. 2009 62 *6*, 76–85. 10.1038/NCPURO1296.
9. Sharifi, N., Gulley, J.L., and Dahut, W.L. (2005). Androgen Deprivation Therapy for Prostate Cancer. JAMA *294*, 238–244. 10.1001/JAMA.294.2.238.
10. Swami, U., McFarland, T.R., Nussenzveig, R., and Agarwal, N. (2020). Advanced Prostate Cancer: Treatment Advances and Future Directions. Trends in Cancer *6*, 702–715. 10.1016/J.TRECAN.2020.04.010.
11. Logothetis, C.J., Efstathiou, E., Manuguid, F., and Kirkpatrick, P. (2011). Abiraterone acetate. Nat. Rev. Drug Discov. *10*, 573–574. 10.1038/NRD3516.
12. Attard, G., Belldegrun, A.S., and De Bono, J.S. (2005). Selective blockade of androgenic steroid synthesis by novel lyase inhibitors as a therapeutic strategy for treating metastatic prostate cancer. BJU Int. *96*, 1241–1246. 10.1111/J.1464-410X.2005.05821.X.
13. Hoffman-Censits, J., and Kelly, W.K. (2013). Enzalutamide: A novel antiandrogen for patients with castrate-resistant prostate cancer. Clin. Cancer Res. *19*, 1335–1339. 10.1158/1078-0432.CCR-12-2910/85299/AM/ENZALUTAMIDE-A-NOVEL-ANTI-ANDROGEN-FOR-PATIENTS.
14. Chen, C.D., Welsbie, D.S., Tran, C., Baek, S.H., Chen, R., Vessella, R., Rosenfeld, M.G., and Sawyers, C.L. (2003). Molecular determinants of resistance to antiandrogen therapy. Nat. Med. 2004 101 *10*, 33–39. 10.1038/nm972.
15. de Bono, J., Mateo, J., Fizazi, K., Saad, F., Shore, N., Sandhu, S., Chi, K.N., Sartor, O., Agarwal, N., Olmos, D., et al. (2020). Olaparib for Metastatic Castration-Resistant Prostate Cancer. N. Engl. J. Med. *382*, 2091–2102. 10.1056/NEJMOA1911440/SUPPL_FILE/NEJMOA1911440_DATA-SHARING.PDF.

16. Abida, W., Patnaik, A., Campbell, D., Shapiro, J., Bryce, A.H., McDermott, R., Sautois, B., Vogelzang, N.J., Bambury, R.M., Voog, E., et al. (2020). Rucaparib in Men With Metastatic Castration-Resistant Prostate Cancer Harboring a BRCA1 or BRCA2 Gene Alteration. J. Clin. Oncol. *38*, 3763–3772. 10.1200/JCO.20.01035.
17. Abida, W., Campbell, D., Patnaik, A., Shapiro, J.D., Sautois, B., Vogelzang, N.J., Voog, E.G., Bryce, A.H., McDermott, R., Ricci, F., et al. (2020). Non-BRCA DNA Damage Repair Gene Alterations and Response to the PARP Inhibitor Rucaparib in Metastatic Castration-Resistant Prostate Cancer: Analysis From the Phase II TRITON2 Study. Clin. Cancer Res. *26*, 2487–2496. 10.1158/1078-0432.CCR-20-0394.
18. Sartor, O., de Bono, J., Chi, K.N., Fizazi, K., Herrmann, K., Rahbar, K., Tagawa, S.T., Nordquist, L.T., Vaishampayan, N., El-Haddad, G., et al. (2021). Lutetium-177–PSMA-617 for Metastatic Castration-Resistant Prostate Cancer. N. Engl. J. Med. *385*, 1091–1103. 10.1056/NEJMOA2107322/SUPPL_FILE/NEJMOA2107322_DATA-SHARING.PDF.
19. Parker, C., Nilsson, S., Heinrich, D., Helle, S.I., O'Sullivan, J.M., Fosså, S.D., Chodacki, A., Wiechno, P., Logue, J., Seke, M., et al. (2013). Alpha Emitter Radium-223 and Survival in Metastatic Prostate Cancer. N. Engl. J. Med. *369*, 213–223. 10.1056/NEJMOA1213755/SUPPL_FILE/NEJMOA1213755_DISCLOSURES.PDF.
20. Tannock, I.F., de Wit, R., Berry, W.R., Horti, J., Pluzanska, A., Chi, K.N., Oudard, S., Théodore, C., James, N.D., Turesson, I., et al. (2004). Docetaxel plus Prednisone or Mitoxantrone plus Prednisone for Advanced Prostate Cancer. https://doi.org/10.1056/NEJMoa040720 *351*, 1502–1512. 10.1056/NEJMOA040720.
21. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A. V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al. (2013). Signatures of mutational processes in human cancer. Nat. 2013 5007463 *500*, 415–421. 10.1038/nature12477.
22. Abeshouse, A., Ahn, J., Akbani, R., Ally, A., Amin, S., Andry, C.D., Annala, M., Aprikian, A., Armenia, J., Arora, A., et al. (2015). The Molecular Taxonomy of Primary Prostate Cancer. Cell *163*, 1011–1025. 10.1016/j.cell.2015.10.025.
23. Fraser, M., Sabelnykova, V.Y., Yamaguchi, T.N., Heisler, L.E., Livingstone, J., Huang, V., Shiah, Y.-J., Yousif, F., Lin, X., Masella, A.P., et al. (2017). Genomic hallmarks of localized, non-indolent prostate cancer. Nature *541*, 359–364. 10.1038/nature20788.
24. Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science (80-.). *310*, 644–648. 10.1126/SCIENCE.1117679/SUPPL_FILE/TOMLINS.SOM.PDF.
25. Castel, P., Toska, E., Engelman, J.A., and Scaltriti, M. (2021). The present and future of PI3K inhibitors for cancer therapy. Nat. Cancer 2021 26 *2*, 587–597. 10.1038/s43018-021-00218-4.
26. Robinson, D., Van Allen, E.M., Wu, Y.M., Schultz, N., Lonigro, R.J., Mosquera, J.M., Montgomery, B., Taplin, M.E., Pritchard, C.C., Attard, G., et al. (2015). Integrative clinical genomics of advanced prostate cancer. Cell *161*, 1215–1228. 10.1016/j.cell.2015.05.001.
27. Quigley, D.A., Dang, H.X., Zhao, S.G., Maher, C.A., Small, E.J., and Feng, F.Y. (2018). Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. Cell *174*, 758–769. 10.1016/j.cell.2018.06.039.
28. Viswanathan, S.R., Ha, G., Hoff, A.M., Wala, J.A., Carrot-Zhang, J., Whelan, C.W., Haradhvala, N.J., Freeman, S.S., Reed, S.C., Rhoades, J., et al. (2018). Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. Cell *174*, 433-447.e19. 10.1016/J.CELL.2018.05.036.
29. Lowder, D., Rizwan, K., McColl, C., Paparella, A., Ittmann, M., Mitsiades, N., and Kaochar, S. (2022). Racial disparities in prostate cancer: A complex interplay between socioeconomic inequities and genomics. Cancer Lett. *531*, 71–82. 10.1016/J.CANLET.2022.01.028.

30. Kohaar, I., Zhang, X., Tan, S.H., Nousome, D., Babcock, K., Ravindranath, L., Sukumar, G., Mcgrath-Martinez, E., Rosenberger, J., Alba, C., et al. (2022). Germline mutation landscape of DNA damage repair genes in African Americans with prostate cancer highlights potentially targetable RAD genes. Nat. Commun. 2022 131 *13*, 1–9. 10.1038/s41467-022-28945-x.
31. Li, J., Xu, C., Lee, H.J., Ren, S., Zi, X., Zhang, Z., Wang, H., Yu, Y., Yang, C., Gao, X., et al. (2020). A genomic and epigenomic atlas of prostate cancer in Asian populations. Nature *580*, 93–99. 10.1038/S41586-020-2135-X.
32. Armenia, J., Wankowicz, S.A.M., Liu, D., Gao, J., Kundra, R., Reznik, E., Chatila, W.K., Chakravarty, D., Han, G.C., Coleman, I., et al. (2018). The long tail of oncogenic drivers in prostate cancer. Nat. Genet. 2018 505 *50*, 645–651. 10.1038/s41588-018-0078-z.
33. Schwanhüusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342. 10.1038/nature10098.
34. Mularoni, L., Sabarinathan, R., Deu-Pons, J., Gonzalez-Perez, A., and López-Bigas, N. (2016). OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. Genome Biol. *17*, 128. 10.1186/s13059-016-0994-0.
35. Weinhold, N., Jacobsen, A., Schultz, N., Sander, C., and Lee, W. (2014). Genome-wide analysis of noncoding regulatory mutations in cancer. Nat. Genet. *46*, 1160–1165. 10.1038/ng.3101.
36. Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371-385.e18. 10.1016/j.cell.2018.02.060.
37. Network, T.C.G.A.R. (2014). Comprehensive molecular characterization of urothelial bladder carcinoma. Nature *507*, 315–322. 10.1038/nature12965.
38. Hornshøj, H., Nielsen, M.M., Sinnott-Armstrong, N.A., Świtnicki, M.P., Juul, M., Madsen, T., Sallari, R., Kellis, M., Ørntoft, T., Hobolth, A., et al. (2018). Pan-cancer screen for mutations in non-coding elements with conservation and cancer specificity reveals correlations with expression and survival /631/67/69 /631/114 article. npj Genomic Med. *3*, 1. 10.1038/s41525-017-0040-5.
39. Fredriksson, N.J., Ny, L., Nilsson, J.A., and Larsson, E. (2014). Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. Nat. Genet. *46*, 1258–1263. 10.1038/ng.3141.
40. Elliott, K., and Larsson, E. (2021). Non-coding driver mutations in human cancer. Nat. Rev. Cancer 2021 218 *21*, 500–509. 10.1038/s41568-021-00371-z.
41. Puente, X.S., Beà, S., Valdés-Mas, R., Villamor, N., Gutiérrez-Abril, J., Martín-Subero, J.I., Munar, M., Rubio-Pérez, C., Jares, P., Aymerich, M., et al. (2015). Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature *526*, 519–524. 10.1038/nature14666.
42. Zeraati, M., Moye, A.L., Wong, J.W.H., Perera, D., Cowley, M.J., Christ, D.U., Bryan, T.M., and Dinger, M.E. (2017). Cancer-associated noncoding mutations affect RNA G-quadruplex-mediated regulation of gene expression. Sci. Rep. *7*, 708. 10.1038/s41598-017-00739-y.
43. Schulz, J., Mah, N., Neuenschwander, M., Kischka, T., Ratei, R., Schlag, P.M., Castaños-Vélez, E., Fichtner, I., Tunn, P.-U., Denkert, C., et al. (2018). Loss-of-function uORF mutations in human malignancies. Sci. Rep. *8*, 2395. 10.1038/s41598-018-19201-8.
44. Wethmar, K., Schulz, J., Muro, E.M., Talyan, S., Andrade-Navarro, M.A., and Leutz, A. (2016). Comprehensive translational control of tyrosine kinase expression by upstream open reading frames. Oncogene *35*, 1736–1742. 10.1038/onc.2015.233.
45. Ziebarth, J.D., Bhattacharya, A., and Cui, Y. (2012). Integrative Analysis of Somatic Mutations Altering MicroRNA Targeting in Cancer Genomes. PLoS One *7*, e47137. 10.1371/journal.pone.0047137.
46. Hitti, E., Bakheet, T., Al-Souhibani, N., Moghrabi, W., Al-Yahya, S., Al-Ghamdi, M., Al-Saif, M., Shoukri, M.M., Lánczky, A., Grépin, R., et al. (2016). Systematic Analysis of AU-Rich Element

Expression in Cancer Reveals Common Functional Clusters Regulated by Key RNA-Binding Proteins. Cancer Res. *76*, 4068–4080. 10.1158/0008-5472.CAN-15-3110.
47. Chen, C.Y.A., and Shyu, A. Bin (1995). AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem. Sci. *20*, 465–470. 10.1016/S0968-0004(00)89102-1.
48. García-Mauriño, S.M., Rivero-Rodríguez, F., Velázquez-Cruz, A., Hernández-Vellisca, M., Díaz-Quintana, A., De la Rosa, M.A., and Díaz-Moreno, I. (2017). RNA binding protein regulation and cross-talk in the control of AU-rich mRNA Fate. Front. Mol. Biosci. *4*. 10.3389/fmolb.2017.00071.
49. Barreau, C., Paillard, L., and Osborne, H.B. (2006). AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res. *33*, 7138–7150. 10.1093/NAR/GKI1012.
50. Wu, X., and Xu, L. (2022). The RNA-binding protein HuR in human cancer: A friend or foe? Adv. Drug Deliv. Rev. *184*. 10.1016/j.addr.2022.114179.
51. Brennan, C.M., and Steitz, J.A. (2001). HuR and mRNA stability. Cell. Mol. Life Sci. *58*, 266–277. 10.1007/PL00000854.
52. Peng, S.S.Y., Chen, C.Y.A., Xu, N., and Shyu, A. Bin (1998). RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. EMBO J. *17*, 3461–3470. 10.1093/emboj/17.12.3461.
53. Moore, A.E., Chenette, D.M., Larkin, L.C., and Schneider, R.J. (2014). Physiological networks and disease functions of RNA-binding protein AUF1. Wiley Interdiscip. Rev. RNA *5*, 549–564. 10.1002/WRNA.1230.
54. George, J., Li, Y., Kadamberi, I.P., Parashar, D., Tsaih, S.W., Gupta, P., Geethadevi, A., Chen, C., Ghosh, C., Sun, Y., et al. (2021). RNA-binding protein FXR1 drives cMYC translation by recruiting eIF4F complex to the translation start site. Cell Rep. *37*. 10.1016/J.CELREP.2021.109934.
55. Shyu, A.B., Greenberg, M.E., and Belasco, J.G. (1989). The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways. Genes Dev. *3*, 60–72. 10.1101/GAD.3.1.60.
56. Caput, D., Beutler, B., Hartog, K., Thayer, R., Brown-Shimer, S., and Cerami, A. (1986). Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators. Proc. Natl. Acad. Sci. U. S. A. *83*, 1670–1674. 10.1073/PNAS.83.6.1670.
57. Chaudhury, A., Cheema, S., Fachini, J.M., Kongchan, N., Lu, G., Simon, L.M., Wang, T., Mao, S., Rosen, D.G., Ittmann, M.M., et al. (2016). CELF1 is a central node in post-transcriptional regulatory programmes underlying EMT. Nat. Commun. *7*, 13362. 10.1038/ncomms13362.
58. Chaudhury, A., Hussey, G.S., Ray, P.S., Jin, G., Fox, P.L., and Howe, P.H. (2010). TGF-β-mediated phosphorylation of hnRNP E1 induces EMT via transcript-selective translational induction of Dab2 and ILEI. Nat. Cell Biol. *12*, 286–293. 10.1038/ncb2029.
59. Djuranovic, S., Nahvi, A., Green, R., Wilczynska, A., Spriggs, R. V., Robinson, S.W., Godfrey, J.D., Willis, A.E., and Bushell, M. (2012). miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. Science *336*, 237–240. 10.1126/science.1215691.
60. Meijer, H.A., Kong, Y.W., Lu, W.T., Wilczynska, A., Spriggs, R. V, Robinson, S.W., Godfrey, J.D., Willis, A.E., and Bushell, M. (2013). Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. Science *340*, 82–85. 10.1126/science.1231197.
61. Peng, Y., and Croce, C.M. (2016). The role of MicroRNAs in human cancer. Signal Transduct. Target. Ther. *1*, 15004. 10.1038/sigtrans.2015.4.
62. Zhou, H., and Rigoutsos, I. (2014). MiR-103a-3p targets the 5' UTR of GPRC5A in pancreatic cells. RNA *20*, 1431–1439. 10.1261/rna.045757.114.
63. Hausser, J., Syed, A.P., Bilen, B., and Zavolan, M. (2013). Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. Genome Res. *23*, 604–615. 10.1101/gr.139758.112.
64. Friedman, R.C., Farh, K.K.H., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. *19*, 92–105. 10.1101/gr.082701.108.

65. Rupaimoole, R., and Slack, F.J. (2017). MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nat. Rev. Drug Discov. *16*, 203–222. 10.1038/nrd.2016.246.
66. Di Leva, G., Garofalo, M., and Croce, C.M. (2014). MicroRNAs in Cancer. Annu. Rev. Pathol. Mech. Dis *9*, 287–314. 10.1146/annurev-pathol-012513-104715.
67. Jansson, M.D., and Lund, A.H. (2012). MicroRNA and cancer. Mol. Oncol. *6*, 590–610. 10.1016/J.MOLONC.2012.09.006.
68. Hayes, J., Peruzzi, P.P., and Lawler, S. (2014). MicroRNAs in cancer: biomarkers, functions and therapy. Trends Mol. Med. *20*, 460–469. 10.1016/j.molmed.2014.06.005.
69. Medina, P.P., Nolde, M., and Slack, F.J. (2010). OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. Nature *467*, 86–90. 10.1038/nature09284.
70. He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J., et al. (2005). A microRNA polycistron as a potential human oncogene. Nature *435*, 828–833. 10.1038/nature03552.
71. Liu, C., Kelnar, K., Liu, B., Chen, X., Calhoun-Davis, T., Li, H., Patrawala, L., Yan, H., Jeter, C., Honorio, S., et al. (2011). The microRNA miR-34a inhibits prostate cancer stem cells and metastasis by directly repressing CD44. Nat. Med. *17*, 211–215. 10.1038/nm.2284.
72. Xiao, Z., Han Li, C., Chan, S.L., Xu, F., Feng, L., Wang, Y., Jiang, J.D., Sung, J.J.Y., Cheng, C.H.K., and Chen, Y. (2014). A small-molecule modulator of the tumor-suppressor mir34a inhibits the growth of hepatocellular carcinoma. Cancer Res. *74*, 6236–6247. 10.1158/0008-5472.CAN-14-0855.
73. Babar, I.A., Cheng, C.J., Booth, C.J., Liang, X., Weidhaas, J.B., Saltzman, W.M., and Slack, F.J. (2012). Nanoparticle-based therapy in an in vivo microRNA-155 (miR-155)-dependent mouse model of lymphoma. Proc. Natl. Acad. Sci. *109*, E1695–E1704. 10.1073/pnas.1201516109.
74. Levati, L., Pagani, E., Romani, S., Castiglia, D., Piccinni, E., Covaciu, C., Caporaso, P., Bondanza, S., Antonetti, F.R., Bonmassar, E., et al. (2011). MicroRNA-155 targets the SKI gene in human melanoma cell lines. Pigment Cell Melanoma Res. *24*, 538–550. 10.1111/j.1755-148X.2011.00857.x.
75. Lopes-Ramos, C.M., Barros, B.P., Koyama, F.C., Carpinetti, P.A., Pezuk, J., Doimo, N.T.S., Habr-Gama, A., Perez, R.O., and Parmigiani, R.B. (2017). E2F1 somatic mutation within miRNA target site impairs gene regulation in colorectal cancer. PLoS One *12*, e0181153. 10.1371/journal.pone.0181153.
76. Engelmann, D., and Pützer, B.M. (2012). The dark side of E2F1: in transit beyond apoptosis. Cancer Res. *72*, 571–575. 10.1158/0008-5472.CAN-11-2575.
77. Bhattacharya, A., Ziebarth, J.D., and Cui, Y. (2013). SomamiR: A database for somatic mutations impacting microRNA function in cancer. Nucleic Acids Res. *41*, D977–D982. 10.1093/nar/gks1138.
78. Cheng, Y., Miura, R.M., and Tian, B. (2006). Prediction of mRNA polyadenylation sites by support vector machine. Bioinformatics *22*, 2320–2325. 10.1093/bioinformatics/btl394.
79. Tian, B., Hu, J., Zhang, H., and Lutz, C.S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res. *33*, 201–212. 10.1093/nar/gki158.
80. Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., and Burge, C.B. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science *320*, 1643–1647. 10.1126/science.1155390.
81. Ji, Z., Lee, J.Y., Pan, Z., Jiang, B., and Tian, B. (2009). Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc. Natl. Acad. Sci. U. S. A. *106*, 7028–7033. 10.1073/pnas.0900028106.
82. Mayr, C., and Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell *138*, 673–684. 10.1016/j.cell.2009.06.016.

83. Xia, Z., Donehower, L.A., Cooper, T.A., Neilson, J.R., Wheeler, D.A., Wagner, E.J., and Li, W. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. Nat. Commun. *5*, 5274. 10.1038/ncomms6274.
84. Miles, W.O., Lembo, A., Volorio, A., Brachtel, E., Tian, B., Sgroi, D., Provero, P., and Dyson, N. (2016). Alternative Polyadenylation in Triple-Negative Breast Tumors Allows NRAS and c-JUN to Bypass PUMILIO Posttranscriptional Regulation. Cancer Res. *76*, 7231–7241. 10.1158/0008-5472.CAN-16-0844.
85. Kreth, S., Limbeck, E., Hinske, L.C., Schütz, S. V., Thon, N., Hoefig, K., Egensperger, R., and Kreth, F.W. (2013). In human glioblastomas transcript elongation by alternative polyadenylation and miRNA targeting is a potent mechanism of MGMT silencing. Acta Neuropathol. *125*, 671–681. 10.1007/s00401-013-1081-1.
86. Zybura-Broda, K., Wolder-Gontarek, M., Ambrozek-Latecka, M., Choros, A., Bogusz, A., Wilemska-Dziaduszycka, J., and Rylski, M. (2018). HuR (Elavl1) and HuB (Elavl2) Stabilize Matrix Metalloproteinase-9 mRNA During Seizure-Induced Mmp-9 Expression in Neurons. Front. Neurosci. *12*, 224. 10.3389/fnins.2018.00224.
87. Li, W., You, B., Hoque, M., Zheng, D., Luo, W., Ji, Z., Park, J.Y., Gunderson, S.I., Kalsotra, A., Manley, J.L., et al. (2015). Systematic Profiling of Poly(A)+ Transcripts Modulated by Core 3' End Processing and Splicing Factors Reveals Regulatory Rules of Alternative Cleavage and Polyadenylation. PLOS Genet. *11*, e1005166. 10.1371/journal.pgen.1005166.
88. Gruber, A.R., Martin, G., Keller, W., and Zavolan, M. (2012). Cleavage factor Im is a key regulator of 3' UTR length. RNA Biol. *9*, 1405–1412. 10.4161/rna.22570.
89. Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kühn, U., Menzies, F.M., Vrielink, J.A.F.O., Bos, A.J., Drost, J., Rooijers, K., et al. (2012). The Poly(A)-Binding Protein Nuclear 1 Suppresses Alternative Cleavage and Polyadenylation Sites. Cell *149*, 538–553. 10.1016/j.cell.2012.03.022.
90. de Klerk, E., Venema, A., Anvar, S.Y., Goeman, J.J., Hu, O., Trollet, C., Dickson, G., den Dunnen, J.T., van der Maarel, S.M., Raz, V., et al. (2012). Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. Nucleic Acids Res. *40*, 9089–9101. 10.1093/nar/gks655.
91. Masamha, C.P., Xia, Z., Yang, J., Albrecht, T.R., Li, M., Shyu, A.-B., Li, W., and Wagner, E.J. (2014). CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature *510*, 412–416. 10.1038/nature13261.
92. Chan, J.J., Zhang, B., Chew, X.H., Salhi, A., Kwok, Z.H., Lim, C.Y., Desi, N., Subramaniam, N., Siemens, A., Kinanti, T., et al. (2022). Pan-cancer pervasive upregulation of 3′ UTR splicing drives tumourigenesis. Nat. Cell Biol. 2022 246 *24*, 928–939. 10.1038/s41556-022-00913-z.
93. Desrosiers, R., Friderici, K., and Rottman, F. (1974). Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. Proc. Natl. Acad. Sci. U. S. A. *71*, 3971–3975.
94. Rottman, F., Shatkin, A.J., and Perry, R.P. (1974). Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing. Cell *3*, 197–199.
95. Dubin, D.T., and Taylor, R.H. (1975). The methylation state of poly A-containing messenger RNA from cultured hamster cells. Nucleic Acids Res. *2*, 1653–1668.
96. Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and near Stop Codons. Cell *149*, 1635–1646. 10.1016/J.CELL.2012.05.003.
97. Dominissini, D., Nachtergaele, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M.S., Dai, Q., Di Segni, A., Salmon-Divon, M., Clark, W.C., et al. (2016). The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. Nature *530*, 441–446. 10.1038/nature16998.
98. Dai, Q., Moshitch-Moshkovitz, S., Han, D., Kol, N., Amariglio, N., Rechavi, G., Dominissini, D., and He, C. (2017). Nm-seq maps 2′-O-methylation sites in human mRNA with base precision. Nat. Methods *14*, 695–698. 10.1038/nmeth.4294.

99. Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206. 10.1038/nature11112.
100. Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S.-B., and Jaffrey, S.R. (2015). 5′ UTR m6A Promotes Cap-Independent Translation. Cell *163*, 999–1010. 10.1016/j.cell.2015.10.012.
101. Nachtergaele, S., and He, C. (2017). The emerging biology of RNA post-transcriptional modifications. RNA Biol. *14*, 156–163. 10.1080/15476286.2016.1267096.
102. Roundtree, I.A., Evans, M.E., Pan, T., and He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. Cell *169*, 1187–1200. 10.1016/j.cell.2017.05.045.
103. Bokar, J.A., Shambaugh, M.E., Polayes, D., Matera, A.G., and Rottman, F.M. (1997). Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA *3*, 1233–1247.
104. Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3–METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. *10*, 93–95. 10.1038/nchembio.1432.
105. Ping, X.-L., Sun, B.-F., Wang, L., Xiao, W., Yang, X., Wang, W.-J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y.-S., et al. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. Cell Res. *24*, 177–189. 10.1038/cr.2014.3.
106. Schwartz, S., Mumbach, M.R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G.G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D., et al. (2014). Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5’ sites. Cell Rep. *8*, 284–296. 10.1016/j.celrep.2014.05.048.
107. Wang, Y., Li, Y., Toth, J.I., Petroski, M.D., Zhang, Z., and Zhao, J.C. (2014). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat. Cell Biol. *16*, 191–198. 10.1038/ncb2902.
108. Deng, X., Su, R., Weng, H., Huang, H., Li, Z., and Chen, J. (2018). RNA N6-methyladenosine modification in cancers: current status and perspectives. Cell Res. *28*, 507–517. 10.1038/s41422-018-0034-6.
109. Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.-G., et al. (2011). N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat. Chem. Biol. *7*, 885–887. 10.1038/nchembio.687.
110. Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.-M., Li, C.J., Vågbø, C.B., Shi, Y., Wang, W.-L., Song, S.-H., et al. (2013). ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. Mol. Cell *49*, 18–29. 10.1016/j.molcel.2012.10.015.
111. Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R., and Qian, S.-B. (2015). Dynamic m(6)A mRNA methylation directs translational control of heat shock response. Nature *526*, 591–594. 10.1038/nature15377.
112. Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell *161*, 1388–1399. 10.1016/j.cell.2015.05.014.
113. Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120. 10.1038/nature12730.
114. Xu, C., Wang, X., Liu, K., Roundtree, I.A., Tempel, W., Li, Y., Lu, Z., He, C., and Min, J. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. Nat. Chem. Biol. *10*, 927–929. 10.1038/nchembio.1654.
115. Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J., and Wu, L. (2016). YTHDF2 destabilizes

m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. Nat. Commun. *7*, 12626. 10.1038/ncomms12626.
116. Li, A., Chen, Y.-S., Ping, X.-L., Yang, X., Xiao, W., Yang, Y., Sun, H.-Y., Zhu, Q., Baidya, P., Wang, X., et al. (2017). Cytoplasmic m6A reader YTHDF3 promotes mRNA translation. Cell Res. *27*, 444–447. 10.1038/cr.2017.10.
117. Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., Sun, H.-Y., Li, A., Ping, X.-L., Lai, W.-Y., et al. (2016). Nuclear m 6 A Reader YTHDC1 Regulates mRNA Splicing. Mol. Cell *61*, 507–519. 10.1016/j.molcel.2016.01.012.
118. Ke, S., Alemu, E.A., Mertens, C., Gantman, E.C., Fak, J.J., Mele, A., Haripal, B., Zucker-Scharff, I., Moore, M.J., Park, C.Y., et al. (2015). A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. Genes Dev. *29*, 2037–2053. 10.1101/gad.269415.115.
119. Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015). N6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. Nature *518*, 560–564. 10.1038/nature14234.
120. Liu, N., Zhou, K.I., Parisien, M., Dai, Q., Diatchenko, L., and Pan, T. (2017). N 6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. Nucleic Acids Res. *45*, 6051–6063. 10.1093/nar/gkx141.
121. Weng, H., Huang, H., Wu, H., Qin, X., Zhao, B.S., Dong, L., Shi, H., Skibbe, J., Shen, C., Hu, C., et al. (2018). METTL14 Inhibits Hematopoietic Stem/Progenitor Differentiation and Promotes Leukemogenesis via mRNA m6A Modification. Cell Stem Cell *22*, 191-205.e9. 10.1016/J.STEM.2017.11.016.
122. Vu, L.P., Pickering, B.F., Cheng, Y., Zaccara, S., Nguyen, D., Minuesa, G., Chou, T., Chow, A., Saletore, Y., MacKay, M., et al. (2017). The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. Nat. Med. *23*, 1369. 10.1038/nm.4416.
123. Visvanathan, A., Patil, V., Arora, A., Hegde, A.S., Arivazhagan, A., Santosh, V., and Somasundaram, K. (2018). Essential role of METTL3-mediated m 6 A modification in glioma stem-like cells maintenance and radioresistance. Oncogene *37*, 522–533. 10.1038/onc.2017.351.
124. Zhang, S., Zhao, B.S., Zhou, A., Lin, K., Zheng, S., Lu, Z., Chen, Y., Sulman, E.P., Xie, K., Bögler, O., et al. (2017). m 6 A Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma Stem-like Cells by Sustaining FOXM1 Expression and Cell Proliferation Program. Cancer Cell *31*, 591-606.e6. 10.1016/j.ccell.2017.02.013.
125. Li, Z., Weng, H., Su, R., Weng, X., Zuo, Z., Li, C., Huang, H., Nachtergaele, S., Dong, L., Hu, C., et al. (2017). FTO Plays an Oncogenic Role in Acute Myeloid Leukemia as a N6-Methyladenosine RNA Demethylase. Cancer Cell *31*, 127–141. 10.1016/J.CCELL.2016.11.017.
126. Zhang, C., Samanta, D., Lu, H., Bullen, J.W., Zhang, H., Chen, I., He, X., and Semenza, G.L. (2016). Hypoxia induces the breast cancer stem cell phenotype by HIF-dependent and ALKBH5-mediated m 6 A-demethylation of NANOG mRNA. Proc. Natl. Acad. Sci. *113*, E2047–E2056. 10.1073/pnas.1602883113.
127. Li, X., Xiong, X., Wang, K., Wang, L., Shu, X., Ma, S., and Yi, C. (2016). Transcriptome-wide mapping reveals reversible and dynamic N1-methyladenosine methylome. Nat. Chem. Biol. *12*, 311–316. 10.1038/nchembio.2040.
128. Carlile, T.M., Rojas-Duran, M.F., Zinshteyn, B., Shin, H., Bartoli, K.M., and Gilbert, W. V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature *515*, 143–146. 10.1038/nature13802.
129. Zhou, H., Kimsey, I.J., Nikolova, E.N., Sathyamoorthy, B., Grazioli, G., McSally, J., Bai, T., Wunderlich, C.H., Kreutz, C., Andricioaei, I., et al. (2016). m1A and m1G disrupt A-RNA structure through the intrinsic instability of Hoogsteen base pairs. Nat. Struct. Mol. Biol. *23*, 803–810. 10.1038/nsmb.3270.

130. Anderson, B.R., Muramatsu, H., Nallagatla, S.R., Bevilacqua, P.C., Sansing, L.H., Weissman, D., and Karikó, K. (2010). Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. Nucleic Acids Res. *38*, 5884–5892. 10.1093/nar/gkq347.
131. Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., León-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S., et al. (2014). Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. Cell *159*, 148–162. 10.1016/J.CELL.2014.08.028.
132. Kuppers, D., Arora, S., Lim, Y., Lim, A., Carter, L., Corrin, P., Plasier, C., Basom, R., Delrow, J., Wang, S., et al. (2018). N6-methyladenosine mRNA marking promotes selective translation of regulons required for human erythropoiesis. bioRxiv, 457648. 10.1101/457648.
133. Li, X., Zhu, P., Ma, S., Song, J., Bai, J., Sun, F., and Yi, C. (2015). Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nat. Chem. Biol. *11*, 592–597. 10.1038/nchembio.1836.
134. Hinnebusch, A.G., Ivanov, I.P., and Sonenberg, N. (2016). Translational control by 5'-untranslated regions of eukaryotic mRNAs. Science *352*, 1413–1416. 10.1126/science.aad9868.
135. Morris, D.R., and Geballe, A.P. (2000). Upstream Open Reading Frames as Regulators of mRNA Translation. Mol. Cell. Biol. *20*, 8635–8642. 10.1128/MCB.20.23.8635-8642.2000.
136. Mackowiak, S.D., Zauber, H., Bielow, C., Thiel, D., Kutz, K., Calviello, L., Mastrobuoni, G., Rajewsky, N., Kempa, S., Selbach, M., et al. (2015). Extensive identification and analysis of conserved small ORFs in animals. Genome Biol. 2015 161 *16*, 1–21. 10.1186/S13059-015-0742-X.
137. Ji, Z., Song, R., Regev, A., and Struhl, K. (2015). Many lncRNAs, 5′UTRs, and pseudogenes are translated and some are likely to express functional proteins. Elife *4*. 10.7554/ELIFE.08890.
138. Bazzini, A.A., Johnstone, T.G., Christiano, R., MacKowiak, S.D., Obermayer, B., Fleming, E.S., Vejnar, C.E., Lee, M.T., Rajewsky, N., Walther, T.C., et al. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J. *33*, 981–993. 10.1002/EMBJ.201488411.
139. Wu, Q., Wright, M., Gogol, M.M., Bradford, W.D., Zhang, N., and Bazzini, A.A. (2020). Translation of small downstream ORFs enhances translation of canonical main open reading frames. EMBO J. *39*. 10.15252/EMBJ.2020104763.
140. Chen, J., Brunner, A.D., Cogan, J.Z., Nuñez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D., et al. (2020). Pervasive functional translation of noncanonical human open reading frames. Science (80-.). *367*, 140–146. 10.1126/science.aav5912.
141. Melton, C., Reuter, J.A., Spacek, D. V, and Snyder, M. (2015). Recurrent somatic mutations in regulatory regions of human cancer genomes. Nat. Genet. *47*, 710–716. 10.1038/ng.3332.
142. Rheinbay, E., Parasuraman, P., Grimsby, J., Tiao, G., Engreitz, J.M., Kim, J., Lawrence, M.S., Taylor-Weiner, A., Rodriguez-Cuevas, S., Rosenberg, M., et al. (2017). Recurrent and functional regulatory mutations in breast cancer. Nature *547*, 55–60. 10.1038/nature22992.
143. Watters, K.E., and Lucks, J.B. (2016). Mapping RNA Structure In Vitro with SHAPE Chemistry and Next-Generation Sequencing (SHAPE-Seq). In Methods in molecular biology (Clifton, N.J.), pp. 135–162. 10.1007/978-1-4939-6433-8_9.
144. Wan, Y., Qu, K., Ouyang, Z., and Chang, H.Y. (2016). Genome-Wide Probing of RNA Structures In Vitro Using Nucleases and Deep Sequencing. In Methods in molecular biology (Clifton, N.J.), pp. 141–160. 10.1007/978-1-4939-3079-1_9.
145. Ritchey, L.E., Su, Z., Tang, Y., Tack, D.C., Assmann, S.M., and Bevilacqua, P.C. (2017). Structure-seq2: sensitive and accurate genome-wide profiling of RNA structure in vivo. Nucleic Acids Res. *45*, e135. 10.1093/nar/gkx533.
146. Smola, M.J., Rice, G.M., Busan, S., Siegfried, N.A., and Weeks, K.M. (2015). Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile

and accurate RNA structure analysis. Nat. Protoc. *10*, 1643–1669. 10.1038/nprot.2015.103.
147. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science (80-.). *324*, 218–223. 10.1126/science.1168978.
148. Garzia, A., Morozov, P., Sajek, M., Meyer, C., and Tuschl, T. (2018). PAR-CLIP for Discovering Target Sites of RNA-Binding Proteins. In Methods in molecular biology (Clifton, N.J.), pp. 55–75. 10.1007/978-1-4939-7540-2_5.
149. Rheinbay, E., Nielsen, M.M., Abascal, F., Wala, J.A., Shapira, O., Tiao, G., Hornshøj, H., Hess, J.M., Juul, R.I., Lin, Z., et al. (2020). Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nat. 2020 5787793 *578*, 102–111. 10.1038/s41586-020-1965-x.
150. Wei, W., Gao, W., Li, Q., Liu, Y., Chen, H., Cui, Y., Sun, Z., and Liu, Z. (2022). Comprehensive characterization of posttranscriptional impairment-related 3′-UTR mutations in 2413 whole genomes of cancer patients. npj Genomic Med. 2022 71 *7*, 1–12. 10.1038/s41525-022-00305-0.
151. Griesemer, D., Xue, J.R., Reilly, S.K., Ulirsch, J.C., Kukreja, K., Davis, J.R., Kanai, M., Yang, D.K., Butts, J.C., Guney, M.H., et al. (2021). Genome-wide functional screen of 3′UTR variants uncovers causal variants for human disease and evolution. Cell, 1–14. 10.1016/j.cell.2021.08.025.
152. Lim, Y., Arora, S., Schuster, S.L., Corey, L., Fitzgibbon, M., Wladyka, C.L., Wu, X., Coleman, I.M., Delrow, J.J., Corey, E., et al. (2021). Multiplexed functional genomic analysis of 5' untranslated region mutations across the spectrum of prostate cancer. Nat. Commun. *12*. 10.1038/S41467-021-24445-6.
153. Anzalone, A. V., Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. Nat. Biotechnol. *38*, 824–844. 10.1038/S41587-020-0561-9.
154. Xu, Y., Poggio, M., Jin, H.Y., Shi, Z., Forester, C.M., Wang, Y., Stumpf, C.R., Xue, L., Devericks, E., So, L., et al. (2019). Translation control of the immune checkpoint in cancer and its therapeutic targeting. Nat. Med. *25*, 301–311. 10.1038/s41591-018-0321-2.
155. Findlay, G.M., Daza, R.M., Martin, B., Zhang, M.D., Leith, A.P., Gasperini, M., Janizek, J.D., Huang, X., Starita, L.M., and Shendure, J. (2018). Accurate classification of BRCA1 variants with saturation genome editing. Nat. 2018 5627726 *562*, 217–222. 10.1038/s41586-018-0461-z.
156. Coelho, M.A., Karakoc, E., Bhosle, S., Gonçalves, E., Burgold, T., Cooper, S., Cattaneo, C.M., Veninga, V., Consonni, S., Dinçer, C., et al. (2022). Base editing screens map mutations affecting IFNγ signalling in cancer. bioRxiv, 2022.03.29.486051. 10.1101/2022.03.29.486051.
157. Kweon, J., Jang, A.H., Shin, H.R., See, J.E., Lee, W., Lee, J.W., Chang, S., Kim, K., and Kim, Y. (2019). A CRISPR-based base-editing screen for the functional assessment of BRCA1 variants. Oncogene 2019 391 *39*, 30–35. 10.1038/s41388-019-0968-2.
158. Huang, C., Li, G., Wu, J., Liang, J., and Wang, X. (2021). Identification of pathogenic variants in cancer genes using base editing screens with editing efficiency correction. Genome Biol. *22*, 1–25. 10.1186/S13059-021-02305-2/FIGURES/6.
159. Hanna, R.E., Hegde, M., Fagre, C.R., DeWeirdt, P.C., Sangree, A.K., Szegletes, Z., Griffith, A., Feeley, M.N., Sanson, K.R., Baidi, Y., et al. (2021). Massively parallel assessment of human variants with base editor screens. Cell *184*, 1064-1080.e20. 10.1016/J.CELL.2021.01.012.
160. Cuella-Martin, R., Hayward, S.B., Fan, X., Chen, X., Huang, J.W., Taglialatela, A., Leuzzi, G., Zhao, J., Rabadan, R., Lu, C., et al. (2021). Functional interrogation of DNA damage response variants with base editing screens. Cell *184*, 1081-1097.e19. 10.1016/J.CELL.2021.01.041.
161. Cottrell, K.A., Chaudhari, H.G., Cohen, B.A., and Djuranovic, S. (2018). PTRE-seq reveals mechanism and interactions of RNA binding proteins and miRNAs. Nat. Commun. *9*, 301. 10.1038/s41467-017-02745-0.
162. Rabani, M., Pieper, L., Chew, G.-L., and Schier, A.F. (2017). A Massively Parallel Reporter Assay of 3′ UTR Sequences Identifies In Vivo Rules for mRNA Degradation. Mol. Cell *68*, 1083-1094.e5.

10.1016/J.MOLCEL.2017.11.014.
163. Wissink, E.M., Fogarty, E.A., and Grimson, A. (2016). High-throughput discovery of post-transcriptional cis-regulatory elements. BMC Genomics *17*, 177. 10.1186/s12864-016-2479-7.
164. Oikonomou, P., Goodarzi, H., and Tavazoie, S. (2014). Systematic Identification of Regulatory Elements in Conserved 3′ UTRs of Human Transcripts. Cell Rep. *7*, 281–292. 10.1016/J.CELREP.2014.03.001.
165. Zhao, W., Pollack, J.L., Blagev, D.P., Zaitlen, N., McManus, M.T., and Erle, D.J. (2014). Massively parallel functional annotation of 3′ untranslated regions. Nat. Biotechnol. *32*, 387–391. 10.1038/nbt.2851.
166. Bogard, N., Linder, J., Rosenberg, A.B., and Seelig, G. (2019). A Deep Neural Network for Predicting and Engineering Alternative Polyadenylation. Cell *178*, 91-106.e23. 10.1016/J.CELL.2019.04.046.
167. Litterman, A.J., Kageyama, R., Le Tonqueze, O., Zhao, W., Gagnon, J.D., Goodarzi, H., Erle, D.J., and Ansel, K.M. (2019). A massively parallel 3′ UTR reporter assay reveals relationships between nucleotide content, sequence conservation, and mRNA destabilization. Genome Res. *29*, 896–906. 10.1101/gr.242552.118.
168. Siegel, D.A., Le Tonqueze, O., Biton, A., Zaitlen, N., and Erle, D.J. (2022). Massively parallel analysis of human 3′ UTRs reveals that AU-rich element length and registration predict mRNA destabilization. G3 Genes|Genomes|Genetics *12*. 10.1093/G3JOURNAL/JKAB404.
169. Vainberg Slutskin, I., Weingarten-Gabbay, S., Nir, R., Weinberger, A., and Segal, E. (2018). Unraveling the determinants of microRNA mediated regulation using a massively parallel reporter assay. Nat. Commun. *9*, 529. 10.1038/s41467-018-02980-z.
170. Slutskin, I.V., Weinberger, A., and Segal, E. (2019). Sequence determinants of polyadenylation-mediated regulation. Genome Res. *29*, 1635–1647. 10.1101/GR.247312.118.

CHAPTER 2

Multi-level functional genomics reveals molecular and cellular oncogenicity of patient-based 3' untranslated region mutations

This chapter presents work currently in revision at *Cell Reports*

Samantha L. Schuster[1,2], Sonali Arora[2], Cynthia L. Wladyka[2], Lukas Corey[2], Bethany L. Stackhouse[2], Lori Kollath[3], Eva Corey[3], Lawrence D. True[4], Dave Young[2], Patrick J. Paddison[2], Andrew C. Hsieh*[1,2,5]

Affiliations:

[1]Molecular and Cellular Biology Graduate Program; University of Washington; Seattle, WA, 98195; USA

[2]Division of Human Biology; Fred Hutchinson Cancer Center; Seattle, WA, 98109; USA

[3]Department of Urology; University of Washington; Seattle, WA, 98195; USA

[4]Department of Laboratory Medicine and Pathology; University of Washington; Seattle, WA, 98195; USA

[5]Departments of Medicine and Genome Sciences; University of Washington; Seattle, WA, 98195; USA

ABSTRACT

3’ untranslated region (3’UTR) somatic mutations represent a largely unexplored avenue of alternative oncogenic gene dysregulation. To determine the significance of 3’UTR mutations in disease, we identify 3’UTR somatic variants across 185 advanced prostate tumors, discovering 14,497 single-nucleotide mutations enriched in oncogenic pathways and 3’UTR regulatory elements. By developing two complementary massively parallel reporter assays, we measure how thousands of patient-based mutations affect mRNA translation and stability and identify hundreds of functional variants that allow us to define determinants of mutation significance. We demonstrate the clinical relevance of these mutations, observing that CRISPR-Cas9 endogenous editing of distinct variants increases cellular stress resistance and that patients harboring oncogenic 3’UTR mutations display particularly poor prognosis. This work represents an unprecedented view of the extent to which disease-relevant 3’UTR mutations affect mRNA stability, translation, and cancer progression, uncovering principles of regulatory functionality and potential therapeutic targets in previously unexplored regulatory regions.

Keywords: 3’UTR, untranslated region, mRNA stability, protein translation, prostate cancer, somatic mutation, massively parallel reporter assay, ZWILCH, IGF1R

INTRODUCTION

Metastatic, castration-resistant prostate cancer (mCRPC) is an advanced form of prostate cancer with a high mortality rate due to a current lack of curative treatment options[1]. While much is already known about how coding sequence (CDS) mutations, mRNA expression changes, and genomic structural variations affect prostate cancer, these findings do not explain all aspects of disease pathogenesis, leaving significant gaps in our ability to treat patients[2–7]. An alternative view of cancer gene regulation may give new insights as to how oncogenic pathways can be regulated in the absence of transcriptional changes. Importantly, only ~40% of protein expression can be explained by differences in mRNA expression, leaving much to be explored at the post-transcriptional level of oncogenic gene expression[8].

3' untranslated region (3'UTR) somatic mutations represent a largely unexplored avenue for understanding aberrant gene regulation in cancer. The 3'UTR of a gene is immediately downstream of the CDS and transcribed with it, but not translated. This region is a major mediator of mRNA stability and translation through sequence- and structure-based elements that recruit *trans*-acting factors. Many of these 3'UTR-interacting factors, such as RNA-binding proteins (RBPs) and microRNAs (miRNAs), are known to be dysregulated in cancer and contribute to pathogenesis[9]. For example, binding of the RBP EBP1 to the androgen receptor (AR) 3'UTR promotes its decay and translational repression and thereby acts as a tumor suppressor in prostate cancer[10–12]. Additionally, the microRNA *miR-21* is oncogenic in prostate cancer, with AR-mediated overexpression of this miRNA increasing both *in vitro* and *in vivo* growth[13,14]. Though these *trans*-acting factors demonstrate that the 3'UTR can be an important hub of cancer gene dysregulation, it is unknown how mutations to the 3'UTR itself affect post-transcriptional dynamics in cancer.

Select studies have established the importance of individual 3'UTR mutations in causing pathogenic changes in post-transcriptional gene regulation. For example, a 1-2 nucleotide deletion in the *EGFR* 3'UTR was found to increase *EGFR* mRNA stability independent of any

other EGFR dysregulation in colorectal tumors, thereby increasing cell sensitivity to EGFR inhibition[15]. Another study found a T→C single-nucleotide variant (SNV), also in the *EGFR* 3'UTR, that abolished a *miR-103a-3p* binding site, thereby increasing *EGFR* expression and derepressing cell growth[16]. These studies demonstrate how 3'UTR mutations can lead to clinically actionable overexpression patterns similar to more well-known amplifications or protein-coding mutations; however, such functional 3'UTR mutation studies are currently limited to individual variants. Massively parallel reporter assays (MPRAs) are powerful tools that combine reporter plasmid libraries with next-generation sequencing, enabling systematic analysis of gene expression changes caused by thousands of sequences of interest simultaneously[17]. They have recently been used to study basic principles of 3'UTR-mediated gene regulation, using random or evolutionarily-conserved sequences to identify patterns of function within this region[18–23]. However, the impact patient-based 3'UTR mutations have on post-transcriptional gene expression has not been explored in this high-throughput manner.

In this study, we have combined 3'UTR somatic mutation calling of 185 mCRPC patient tumors with a new dual MPRA technology to investigate how thousands of 3'UTR single-nucleotide mutations affect mRNA translation and stability. We find hundreds of these patient-derived mutations that cause significant changes to post-transcriptional gene regulation and with this knowledge determine specific sequence features associated with mutation function. CRISPR-Cas9 base editing allows us to further explore patient mutations within their endogenous genomic contexts, demonstrating that single-nucleotide changes in *ZWILCH* and *IGF1R* 3'UTRs are sufficient to drive pro-proliferative protein expression and cellular growth. We further explore patient outcome data, finding that oncogenic 3'UTR mutations that change mRNA translation or stability are associated with faster progression to advanced disease. This comprehensive study establishes the functional significance and clinical relevance of patient-based 3'UTR mutations, demonstrating their ability to drive molecular and cellular oncogenicity.

RESULTS

mCRPC patients harbor thousands of 3'UTR mutations in cancer-related genes and regulatory elements

To study how 3'UTR mutations affect post-transcriptional gene expression in prostate cancer, we first built a UTR-centric sequencing database including whole-genome sequencing (WGS) and UTR-sequencing of 185 mCRPC tumors. This dataset consists of in-house UTR-specific sequencing of matched normal and metastatic tissue for 79 patients from the University of Washington (UW) Tissue Acquisition Necropsy (TAN) program[24,25] and 5 samples from the UW LuCaP patient-derived xenograft (PDX) models[17]. Additionally, we obtained WGS of matched normal and metastatic tissue for 101 mCRPC patients from the Stand Up to Cancer (SU2C) West Coast Dream Team project[3] (**Figure 2-1A**). In addition to UTR-sequencing, exome sequencing was available for 47 of the UW patients, enabling us to compare the 3'UTR mutational landscape to that of the 5'UTR and coding sequence across 148 of our patients. To call a high confidence set of somatic mutations within each of these genomic regions, we implemented a combination of four mutation callers (Mutect2[26], Strelka2[27], MuSE[28], and Varscan2[29]) over both the UTRs and protein-coding space, using read cutoffs according to the sequencing depth of each dataset (**Figure 2-S1A**). For all further analysis, we retained only mutations that had been found in at least two callers (**Figure 2-S1B**), amounting to 14,497 3'UTR somatic mutations across 7,647 genes in these 185 mCRPC patient samples. Of this finalized set of 3'UTR mutations, 67.6% were identified by all four mutation callers, demonstrating a high level of agreement.

To further establish the robustness of these 3'UTR mutations, we compared the tumor mutational burden (TMB) and mutational signatures between the UTR and CDS regions in these patients. We find the overall TMB of the 3'UTR to be slightly lower than that of the CDS in mCRPC patients, but equivalent to the 5'UTR and within the range of previously published

prostate cancer TMB[30] (**Figure 2-S1C**). The 3'UTR mutations largely follow expected patterns of base substitutions, similar to that of the CDS, displaying a strong bias towards C→T/G→A transition mutations (**Figure 2-S1D**). To expand on such base change patterns, the COSMIC project has defined mutational signatures across cancer types in terms of base substitutions and their surrounding trinucleotide context[31]. In our dataset, both 3'UTR and CDS mutations follow established mutational signature patterns, with most patients exhibiting COSMIC signature 1, known to be found in almost all cancers, and many following signatures 5 & 6, which are specifically associated with prostate cancer (**Figures 2-S1E**). Additionally, we find enrichment of signatures 3, 8, and 16 specifically in 3'UTR mutations. Signature 3 is associated with failure to repair DNA double-stranded breaks, while signatures 8 and 16 have unknown etiology. The differences in mutational signatures between the CDS and 3'UTR potentially demonstrate the differing mutational constraints and selective pressures upon these genomic regions. Overall, the 3'UTR mutations called in our dataset follow expected patterns for prostate cancer, while uncovering alterations unique to this region.

These 3'UTR mutations greatly expand the space of potentially oncogenic mutations in mCRPC tumors, as many 3'UTR mutations fall into genes without CDS mutations, but with known relevance to cancer (**Figures 2-1B and 2-1C**). In fact, we observe significant enrichment of 3'UTR mutations in cancer-related gene sets and pathways, including breast, gastric, and prostate cancer signatures, suggesting that these mutations may be an important source of aberrant oncogenic expression (**Figure 2-1C**). We identify several well-studied cancer-associated genes with highly mutated 3'UTRs, including *FLRT2*, a cell adhesion protein identified as tumor suppressive in several cancers, whose 3'UTR is mutated in over 10% of our patients[32–34] (**Figure 2-1D**). *LPP* is another highly mutated cell adhesion protein with roles in regulating both metastasis and treatment resistance[35–38], and *FOXA1* and *SOX5*, each harboring 3'UTR mutations in 9 tumors (4.8%), are both established regulators of prostate

cancer[39–42]. The enrichment of 3'UTR mutations in established cancer-related genes leads us to hypothesize that these mutations are an alternative method of oncogenic dysregulation.

In addition to analyzing the genes targeted by 3'UTR mutations, we endeavored to delineate patterns of mutation across individual 3'UTR sequences. We observe that mutations are found equally distributed along the length of the 3'UTR, not enriched near the CDS nor the poly-A tail (**Figure 2-1E**). However, we find over 700 hotspot locations where multiple 3'UTR mutations are found within 50 bases of each other, with over 250 of these being within 5 bases of one another (**Figures 2-1F and 2-1G**). These mutation hotspots suggest repeated targeting of functional sequence motifs within the 3'UTR. Therefore, we set out to determine whether each 3'UTR mutation in our dataset could change known RBP motifs, miRNA seed sequences, the poly-A signal (PAS) AAUAAA[43], or the m^{6}A RRACH motif[44]. We then compared these results to an *in silico* expected distribution of 14,497 mutations randomly distributed across the 3'UTRome, keeping base changes and trinucleotide context of the mutations constant. We observe that patient 3'UTR mutations alter RBP, miRNA, and PAS motifs significantly more than expected by chance, while they are excluded from m^{6}A motifs (**Figure 2-1H**). Given the prevalence of mutations in cancer genes, hotspot areas, and regulatory motifs, we hypothesize that these mutations may be functional in altering 3'UTR-mediated gene expression, and therefore significant in cancer. The landscape of 3'UTR mutations in advanced prostate cancer tumors uncovers a wealth of potential functional mutations in cancer-related genes and functional *cis*-regulatory elements.

Patient-based 3'UTR mutations functionally impact gene-specific translation efficiency

To functionally assess how patient mutations change 3'UTR-mediated aspects of post-transcriptional gene regulation, we first built a polysome profiling-based massively parallel reporter assay (MPRA) able to measure how thousands of mutations change translation efficiency simultaneously (**Figure 2-2A**). Two hundred base-pair sequences of the 3'UTR

around each of 6,892 mutations from recurrently mutated genes were inserted downstream of the luciferase coding sequence in a modified pGL4 plasmid backbone, pLuc2CP-noARE. Sequencing of this plasmid library confirmed adequate and even representation of 3'UTR inserts (**Figure 2-S2A**).

The resultant plasmid library was transfected into PC3 cells, chosen to represent the mCRPC cellular environment similar to that of patient samples. Polysome profiling was performed on the transfected cells to fractionate mRNA by the number of attached ribosomes, resulting in distinct monosome-, low polysome-, and high polysome-associated pools of mRNA. Six biological replicates of the polysome-associated mRNA pools, plus total mRNA and plasmid DNA extracted from each sample, were sequenced. Translation efficiency (TE) was calculated based on the ratio of total polysome- or high polysome-associated mRNA to total mRNA for each 3'UTR insert, and wildtype and mutant pairs of 3'UTR inserts were analyzed for significant TE changes (FDR<0.10) caused by each mutation using the xtail package in R.

We achieved an average sequencing depth of 315 reads per insert in each sample (**Figure 2-S2B**), including even sequencing across replicates and sample types (**Figure 2-S2C**). We also detect high correlation between biological replicates in each sample group (**Figure 2-S2D**). Internal control sequences from published 3'UTR MPRAs have been included to validate our methods. Firstly, we found several "activator" or "repressor" sequences from Oikonomou *et al.* that significantly (FDR<0.10) changed TE relative to the average TE across all MPRA 3'UTR inserts, and 4/5 of these behave as originally published[20] (**Figure 2-S2E**). Also, addition of the Pumilio binding element to a blank sequence shows expected results agreeing with the PTRE-seq MPRA, causing a dose-dependent decrease in RNA expression while TE remains stable[19] (**Figure 2-S2F**). These quality control checks indicate robustness across our assay.

We discover 180 3'UTR point mutations that significantly change translation efficiency, with an over 1000-fold dynamic range (**Figure 2-2B**). As expected, the two methods of calculating TE, either using total or high polysome-associated mRNA, agree well, with total

polysomes uncovering a larger pool of functional mutations (**Figure 2-2C**). To ensure that these MPRA observations are significant, we validated several top hits using individual dual luciferase assays, observing expected changes in protein expression in accordance with those observed by polysome profiling (**Figures 2-2B and 2-S2G**). Furthermore, our MPRA results correlate well with *in vivo* TE changes observed via ribosome profiling, a method that measures TE similar to polysome profiling, of tumor samples from the five UW PDX models. With this data, we correlate tissue-level TE changes for a 3'UTR-mutated gene with the behavior of the same mutation in our MPRA, finding that 5 out of 8 mutation effects observed by both MPRA and ribosome profiling agree, with a Spearman correlation coefficient of r=0.55 (**Figure 2-S2H**). These experiments show the power of single-nucleotide 3'UTR mutations to affect translation dynamics on a large scale and demonstrate the applicability of our high-throughput reporter technology to *in vivo* tissue effects.

Many functional 3'UTR mutations cause seemingly oncogenic changes in expression, either increasing TE of oncogenic mRNAs or decreasing TE of tumor suppressive genes (**Figure 2-2B**). For example, we find a particular 3'UTR mutation in *MSI2* (chr17:57681293 A→G) that increases its translation efficiency almost 16-fold. We would expect this change to be impactful in a tumor setting, as *MSI2* is an RNA-binding protein previously shown to promote progression in prostate and other cancer types[45–47]. Additionally, *GLIPR1* and *SSBP2* have tumor suppressive qualities in multiple cancers, including prostate, making their suppressive 3'UTR mutations (chr12:75503808 T→C, 4.8-fold decrease and chr5:81416226 G→A, 3.1-fold decrease, respectively) potentially clinically relevant[48–53]. Interestingly, we find enrichment of functional mutations in neuronal genes, suggesting dysregulation of neuroendocrine-related pathways that are often a means of cellular plasticity and therapy resistance in advanced prostate cancer[54] (**Figure 2-2D**). There is also significant enrichment across several known cancer-related gene sets, including p53, Wnt, receptor tyrosine kinase, and JAK-STAT signaling,

highlighting the many oncogenic pathways that seem to be functionally dysregulated by 3'UTR mutations.

To determine how these 3'UTR mutations affect TE on a molecular level, we analyzed whether they alter known regulatory sequences, finding that over 75% of functional mutations change RBP and/or miRNA binding motifs (**Figure 2-2E**). Specifically, most mutations are in RBP binding sites, particularly those that bind AU- and U-rich motifs, including hnRNPCL1, PUF60, RBM6, and RC3H1[55] (**Figure 2-2F**). Intriguingly, we also find 43 mutations that affect TE in the absence of a clear *cis*-regulatory element, suggesting the involvement of RNA structural changes or novel motifs. This MPRA uncovers the effects patient-derived 3'UTR mutations have on mRNA translation, establishing the broad importance of these mutations in regulating oncogenic expression through altering sequence elements.

Patient-based 3'UTR mutations significantly alter oncogenic mRNA stability

In addition to translation efficiency, the 3'UTR is an important mediator of mRNA stability. To determine how mCRPC patient mutations affect mRNA degradation, we designed and implemented a second, complementary MPRA using an RNA-seq time course of *in vitro* transcribed (IVT) mRNA (**Figure 2-3A**). We took advantage of the integrated T7 promoter to *in vitro* transcribe our plasmid library into a fully capped and poly-A tailed mRNA library. We transfected PC3 cells with this mRNA library, then collected library mRNA from the cells over a 24-hour time course in 6 biological replicates. An exogenous spike-in was used to normalize between time points and replicates, ensuring quantitative measurement of mRNA degradation. The RNA-seq time course was sequenced to an average depth of 443 reads per insert, with uniform depth and high correlation across samples and replicates (**Figures 2-S3A - 2-S3C**).

To define the decay of each mRNA from the 3'UTR library, we fit non-linear least squares regression curves with the formula $y_t=y_{plateau}+(y_0+y_{plateau})*e^{-\alpha (t)}$ to the data for each 3'UTR insert

and calculated mRNA half-lives using the relationship $half - life = \frac{\ln(2)}{\alpha}$ (**Figure 2-S3D**). Since most of the dynamic range between different 3'UTRs occurs at the 3- and 6-hour time-points, these were used for statistical analysis, in addition to filtering by raw read counts and changes in mRNA half-life. Using these criteria, we observe expected results in several internal control 3'UTRs. The insertion of miRNA seed sequences decreases mRNA stability relative to a blank (non-functional sequence) control and mutating the seed sequence abrogates this repression (**Figure 2-S3E**). Furthermore, the addition of RNA-binding protein motifs, including an AU-rich element and a Pumilio element, both cause expected decreases in mRNA stability as compared to the blank 3'UTR mRNA[19,56] (**Figure 2-S3F**). In addition to these internal MPRA controls, we externally validated several MPRA hits using separate dual luciferase reporter assays, finding that 3'UTR mutations in *SLC16A7*, *PRDM16*, and *LSAMP* significantly decrease protein expression as expected (**Figures 2-3B and 2-3C**). Additionally, changes in mRNA stability measured by our MPRA generally agree with tissue-level mRNA expression changes in UW PDX tumor samples (**Figure 2-S3G**, 6/8 mutation agreement). Interestingly, we see that correlation between MPRA measurements of mRNA stability and tissue-level RNA-seq are not as strong as correlations between translation MPRA results and ribosome profiling, potentially highlighting the difference between measuring steady-state RNA expression and RNA dynamics.

Overall, we find 150 patient-based 3'UTR mutations that significantly change mRNA stability (**Figure 2-3B**). Collectively, the functional mutations affecting mRNA stability are enriched in neuronal genes (**Figure 2-3D**). Together with our previous findings (**Figure 2-2D**), this strengthens the hypothesis that 3'UTR mutations can be an alternative way neuroendocrine features are dysregulated in advanced prostate cancer. While we do not observe enrichment of cancer-related gene sets, many functional 3'UTR mutations affect known oncogenes. Tumor mutations in *ASCL1*, a driver of advanced neuroendocrine prostate cancer[57], *MBD2*, an

epigenetic regulator that binds methylated DNA[58,59], and the RNA-binding protein *FXR1*[60,61] each increase the half-lives of their respective oncogenic mRNAs, establishing these as functional oncogenic 3'UTR mutations (**Figure 2-3B**, chr18:54154901 T→A, chr12:102960501 T→A, and chr3:180978239 A→G, respectively). Like those mutations that affect translation efficiency, most of these stability-affecting 3'UTR mutations alter known RBP binding sequences, particularly those with AU- and U-rich motifs, pointing to the dual effects many RBPs have on both mRNA translation and stability (**Figures 2-3E and 2-3F**). Our complementary MPRA technologies allow us to observe how 3'UTR mutations affect two distinct stages of post-transcriptional gene regulation while observing the common theme that functional mutations on either level target established cancer-related genes and ablate RBP binding elements.

3'UTR mutation functionality is determined by sequence conservation, regulatory motifs, and RNA structure

Using the power of our dual MPRAs, we have tested thousands of patient-derived 3'UTR mutations and identified hundreds that change gene expression, allowing us to delineate aspects of 3'UTRs connected to mutation function. We have noted that most functional 3'UTR mutations, either that change translation efficiency or mRNA stability, alter known *cis*-regulatory elements. Comparing this to the background level of passenger mutations that fall in regulatory elements without changing gene expression, we find that stability-modifying mutations alter motifs significantly more than passenger mutations (**Figure 2-4A**). As regulatory motifs are often under selective pressure to remain functional throughout evolution, it follows that these functional mutations may be targeting areas of evolutionary conservation. As expected, we observe that stability-related functional mutations are also found significantly more in areas of high sequence conservation (**Figure 2-4B**). These trends towards functional mutations preferentially being in conserved, *cis*-regulatory element sequences also holds true for 3'UTR

mutations altering translation efficiency, though they are less pronounced. A possible explanation may be that translation-mediating 3'UTR mutations are less dependent on specific sequence elements because they can act through structural changes instead. In fact, we see that 3'UTR mutations that change either mRNA translation or stability have significantly lower GC content and less structural stability (less negative ΔG of the predicted minimum free energy secondary structure) than passenger mutations, suggesting that areas of lower RNA structural stability are more likely to be functional when mutated (**Figures 2-4C & 2-4D**). From this analysis, we have uncovered important determinants of 3'UTR mutation function, whereby mutations in highly conserved regulatory motifs are more likely to affect mRNA stability, and those in areas of low structure are likely to change multiple levels of 3'UTR-mediated post-transcriptional gene regulation.

Patient-based 3'UTR mutations affect post-transcriptional gene expression and cancerous phenotypes in endogenous cellular contexts

Our systems-based analysis has expanded our understanding of 3'UTR mutation biology; however, it remains to be determined how these mutations impact cellular function. To address this, we used cutting-edge CRISPR-Cas9 base editing to explore the cellular consequences of 3'UTR mutations in cancer. CRISPR base editors allow us to introduce 3'UTR mutations precisely into their endogenous genomic location; however, there are specific constraints on what mutations can be created based off of PAM availability and the desired base change[62]. Therefore, we computationally sorted through the top hits of our MPRAs to determine which could be mutated using available CRISPR base editing systems, resulting in the choice of two mutations in the genes *ZWILCH* and *IGF1R* (chr15:66548998 A→G and chr15:98958058 G→A, respectively) (**Figure 2-5A**). Both of these tumor-based 3'UTR mutations significantly increased translation efficiency of their respective proliferation-associated mRNAs, making them potential oncogenic 3'UTR mutations. Therefore, we transfected cells with CRISPR base editing

systems specific to these mutations and selected single-cell clones, resulting in several wildtype and mutant cell lines for each mutation confirmed by Sanger sequencing (**Figures 2-S4A and 2-S4B**).

Western blotting for the endogenous ZWILCH and IGF1R proteins confirms that these single-nucleotide 3'UTR mutations cause increased protein expression (~45% increase in ZWILCH, ~35% increase in IGF1R, **Figures 2-5B and 2-5C**). These changes in overall protein levels are especially striking when considering that the 3'UTR mutation was only introduced to one genomic allele. We observe no change in mRNA expression in the 3'mutant cell lines, which validates our MPRA results that these mutations specifically change translation efficiency (**Figures 2-S4C and 2-S4D**). We next sought to determine whether the increased protein expression conferred by these 3'UTR mutations had cellular, potentially oncogenic, effects.

ZWILCH is an essential component of the mitotic spindle assembly checkpoint necessary for pausing metaphase transition to anaphase until chromosomes are properly attached to the kinetochore[63]. Overexpression of *ZWILCH* and other related mitotic checkpoint proteins has been previously associated with cancer progression[64–66], and there is a strong negative correlation between *ZWILCH* mRNA levels and patient survival in the SU2C dataset[6] (**Figure 2-S4E**). As such, we sought to determine how this 3'UTR mutation affects cell growth. While there is no difference between *ZWILCH* wildtype and 3'mutant cell lines in fully supplemented media (**Figure 2-S4F**), mutant lines grow significantly faster when under stress conditions. In particular, we challenged the cells with either nutrient deprivation or cisplatin, finding that the 3'UTR mutation confers a growth advantage to the cells in each condition (**Figure 2-5D**). The observed cisplatin resistance is especially striking, as this DNA-damaging agent is a direct challenge to the mitotic checkpoint and has clinical implications for the use of chemotherapy.

IGF1R is a growth factor receptor involved in the progression of multiple cancer types through its activation of several downstream oncogenic pathways[67–70]. The 3'UTR mutation we

identified in *IGF1R* increases its expression by ~35%, indicating its potential to increase cell proliferation. Under basal conditions, this *IGF1R* mutation, like the *ZWILCH* mutation, did not change cell growth rates (**Figure 2-S4G**). However, when the cells' access to IGF1R ligands was reduced by either decreasing FBS or replacing it with charcoal-stripped serum[71], we observe a growth advantage in the mutant cells (**Figure 2-5E**). This suggests that 3'UTR-mediated *IGF1R* overexpression enables the cells to activate proliferative signaling in response to minimal growth factor. These findings demonstrate that 3'UTR mutations can increase cancer cell oncogenicity by enhancing growth under stress, a quality particularly important within the harsh ecosystem of a tumor microenvironment.

Encouraged by the discovery of these oncogenic 3'UTR mutations, we turned to elucidating potential mechanisms by which these mutations drive improved translation. Examining the sequence around each mutation for predicted changes in *cis*-regulatory elements and RNA structure, we find that the *ZWILCH* 3'UTR mutation is within an HNRNPD (AUF1) binding element, altering a required terminal A position to a G (**Figure 2-5F**). AUF1 is known to decrease target gene expression by affecting either mRNA stability or translation efficiency in different contexts, consistent with our finding that interrupting its binding motif leads to an increase in mRNA translation[72]. In contrast, the mutation in *IGF1R* falls within the minority of functional mutations that do not alter a known regulatory motif. However, upon probing of the predicted RNA secondary structure using RNAfold[73,74], we find that this mutation causes a considerable change in the stability of a particular stem-loop structure. The G→A mutation introduces a proper A-U Watson-Crick base pair to the stem in place of a G-U pair, increasing the likelihood of this stem-loop structure being formed (**Figure 2-5G**). Though it is yet unknown how exactly this change in RNA structure may lead to increased translation, it provides an example of how mutations in low secondary structure areas without obvious sequence motifs can be functional. These CRISPR-based studies demonstrate that single-nucleotide patient

3’UTR mutations can have significant effects in endogenous contexts, both on a molecular and cellular level.

Functional oncogenic 3’UTR mutations correlate with poor patient outcomes

We have found individual 3’UTR mutations that cause oncogenic expression changes and increased cell growth, leading us to question how such mutations affect patient outcomes. Extensive patient outcome data are available for the UW TAN patients, allowing us to interrogate how 3’UTR mutations associate with patient prognosis (**Table 2-S6A**). To understand the nuances of this relationship, we first performed an extensive literature search to curate a list of presumed oncogenic mutations: those that increase expression of an established oncogenic mRNA or decrease expression of an established tumor suppressive gene (**Table 2-S6B-C**)**.**

Overall, we observe that patient tumors harboring any oncogenic 3’UTR mutations have significantly faster progression to androgen independence (AI) than tumors without these mutations (**Figure 2-6A**). Additionally, these patients trend towards shorter time to bone metastasis and overall survival (**Figures 2-6B and 2-6C**). These are important indicators of patient outcome, as progression from androgen sensitivity to castration-resistance, as well as the development of distant metastasis, are major causes of patient mortality. Interestingly, patients with translation-based oncogenic 3’UTR mutations have more moderate phenotypes, displaying a strong trend towards fast progression to androgen independence but no change in overall survival or metastasis (**Figures 2-6D-F**). However, stability-based oncogenic 3’UTR mutations have a larger effect on patient outcome, with tumors harboring these mutations causing patients to have significantly poorer prognosis on all three fronts (**Figures 2-6G-I**). Though it is a potential confounding factor, the difference between patient outcomes for the translation- versus stability-based oncogenic 3’UTR mutations is not simply due to differences in overall mutation burden. The patients harboring these oncogenic mutations do not have significantly different TMB than other patients in either case (**Figures 2-S5A and 2-S5B**). The

total patient population harboring oncogenic 3'UTR mutations has somewhat higher overall TMB (**Figure 2-S5C**), but since the differences in prognosis among these groups is not correlated with TMB trends, this does not seem to explain patient outcomes as a whole. Overall, these studies show that functional tumor-based 3'UTR mutations have significant implications for prostate cancer patients' outcomes.

DISCUSSION

Our study uses a combination of cutting-edge techniques to uncover the landscape of functional mutations in prostate cancer patient 3'UTRs. Whole-genome and UTR-specific sequencing identifies thousands of somatic 3'UTR mutations in patients, enriched in oncogenic signaling pathways and *cis*-regulatory motifs. These results suggest important functional consequences of 3'UTR mutations in cancer, leading us to design a novel dual-pronged MPRA to assess how these mutations affect mRNA translation and stability. To our knowledge, these MPRAs represent the first study to examine post-transcriptional effects of patient- and disease-based 3'UTR mutations in a high-throughput manner, expanding our current knowledge of oncogenic expression dynamics with further understanding of post-transcriptional gene dysregulation. We find hundreds of patient mutations, many affecting oncogenic proteins and pathways, that significantly change either mRNA stability or translation. These mutations represent novel insight into how genes like *IGF1R*, *MSI2*, *MBD2*, and *ASCL1* can be aberrantly regulated in cancer, highlighting the importance of expanding the field's view of cancer genomics to include alterations in UTRs. For example, the knowledge that a patient harbors a translation-increasing 3'UTR mutation in *IGF1R* may suggest anti-IGF1R agents as an effective route of therapy, even in the absence of CDS mutations or mRNA expression changes[75].

Interestingly, most functional 3'UTR mutations alter known regulatory motifs, and in fact the presence of a *cis*-regulatory element at the site of a mutation is found to predict altered function. Additionally, high sequence conservation and low RNA structural stability are associated with functional mutations, agreeing with previous studies that also observed correlations between 3'UTR function and low RNA structure[76,77]. These findings are the first steps toward a high-resolution understanding of 3'UTR mutations, where *in silico* analysis of the conservation, sequence, and structure of tumor 3'UTR mutations may accurately predict mutation function. This would likely require testing thousands more 3'UTR mutations across

different conditions and integrating these results with machine learning to fully understand the design principles of 3'UTRs.

Though our MPRA were limited to the mCRPC model of PC3 cells, it is encouraging to note that other UTR-based MPRAs have found high correlation in UTR function across cell types[17], suggesting that our results are likely applicable to other cellular contexts. Another limitation of our study is the necessity of using 201 base-pair fragments of 3'UTRs in our reporter library, as cloning full 3'UTRs would be impossible at an average length of almost 6,000 bases per 3'UTR and the maximum next-generation sequencing read length of only ~300 nucleotides. While this design does not incorporate long-range interactions between distant 3'UTR, 5'UTR, or CDS regulatory elements and the mutations of interest, it highlights the effects of altering specific regulatory motifs, which are usually 5-20 bases long. To truly capture such complex interaction networks, mutations must be studied in their endogenous genomic context.

Specificity and applicability of CRISPR single-nucleotide editing has increased significantly in the past few years with the ever-expanding toolkit of base editors and more recently prime editing, allowing for deeper understanding of single-nucleotide mutations in endogenous contexts[78]. We have leveraged this technology to demonstrate that individual patient 3'UTR mutations in *ZWILCH* and *IGF1R* found to significantly change mRNA translation via high-throughput reporter assays also have oncogenic effects in cells at both molecular and cellular levels. Importantly, we observed significant changes in both protein expression and growth even when these mutations were only introduced in a heterozygous manner. As mutations are very often clonal and heterozygous in tumors, this demonstrates that 3'UTR mutations can impactfully tune oncogenic expression even when present at low variant allele fractions and cause significant additive effects on cellular oncogenicity. The mutations in *ZWILCH* and *IGF1R* are predicted to affect gene expression through altering an RBP binding motif and RNA secondary structure, respectively. Future experiments could explore such molecular mechanisms of 3'UTR mutations robustly using CRISPR-edited endogenous cell

models, more so than in reporter systems. Expanding the use of CRISPR base- or prime-editing by designing large-scale 3'UTR endogenous mutation editing screens could be integral in further understanding tumor mutations. This has recently been accomplished for investigating variants of unknown significance in coding sequences and could be similarly applied to patient 3'UTR mutations as the next step in combining massively parallel assays with CRISPR technology[79–82].

Our work represents an unprecedented view of the extent to which disease-relevant 3'UTR mutations affect mRNA stability, translation efficiency, and cancer phenotypes, expanding the boundaries of functional cancer genomics and potentially uncovering novel therapeutic targets in previously unexplored regulatory regions.

ACKNOWLEDGEMENTS

We thank the patients who participated in this study and their families. We thank members of the A.C.H. laboratory for helpful advice and discussions. We thank the Seattle RNA Metabolism group for critical discussion of the work. This work was supported by NIH award R37 CA230617, R01 CA255526, R01 GM135362, Burroughs Welcome Fund, Career Award for Medicine Scientists (1012314.02), and grants from the Emerson Collective (691630), and the Robert J. Kleberg Jr. and Helen C. Kleberg Foundation to A.C.H. S.L.S. received funding through an NIH F31 NRSA grant (6F31CA260920-03), UW T32 training grant (GM007270), and Fred Hutchinson Cancer Center Co-operative Center for Excellence in Hematology pilot grant (U54 DK106829-P&F). We also thank Pete Nelson, Celestia Higano, Evan Yu, Elahe Mostaghel, Heather Cheng, Bruce Montgomery, Mike Schweizer, Jonathan Wright, Daniel Lin, Funda Vakar-Lopez, Xiaotun Zhang, Martine Roudier, Colm Morrissey and the rapid autopsy teams for their contributions to the University of Washington Medical Center Prostate Cancer Donor Rapid Autopsy Program. This work was supported by the Department of Defense Prostate Cancer Research Program (W81XWH-14-2-0183), the Pacific Northwest Prostate Cancer SPORE (P50CA97186), the PO1 NIH grant (PO1CA163227), the institute for Prostate Cancer Research, and the Richard M. LUCAS Foundation. The maintenance and characterization of the LuCaP PDX models were supported by the Pacific Northwest Prostate Cancer SPORE (P50CA97186), the Department of Defense Prostate Cancer Biorepository Network (W81XWH-14-2-0183), and NCI P01-CA163227. This research was also supported by the Genomics and Bioinformatics Shared Resource of the Fred Hutch/University of Washington Cancer Consortium (P30 CA015704) and the Scientific Computing Infrastructure at Fred Hutch funded by ORIP grant S10 OD028685.

AUTHOR CONTRIBUTIONS

Conceptualization, S.L.S. and A.C.H.; Methodology, S.L.S. and A.C.H.; Software, S.L.S., S.A., and L.R.C.; Investigation, S.L.S., C.L.W., and D.Y.; Resources, B.L.S., L.K., E.C., L.D.T, and P.J.P; Writing- Original draft, S.L.S; Writing- Review & Editing, S.A., C.L.W, E.C., and A.C.H.; Supervision, A.C.H.; Funding Acquisition, S.L.S and A.C.H.

COMPETING INTERESTS

No competing interests

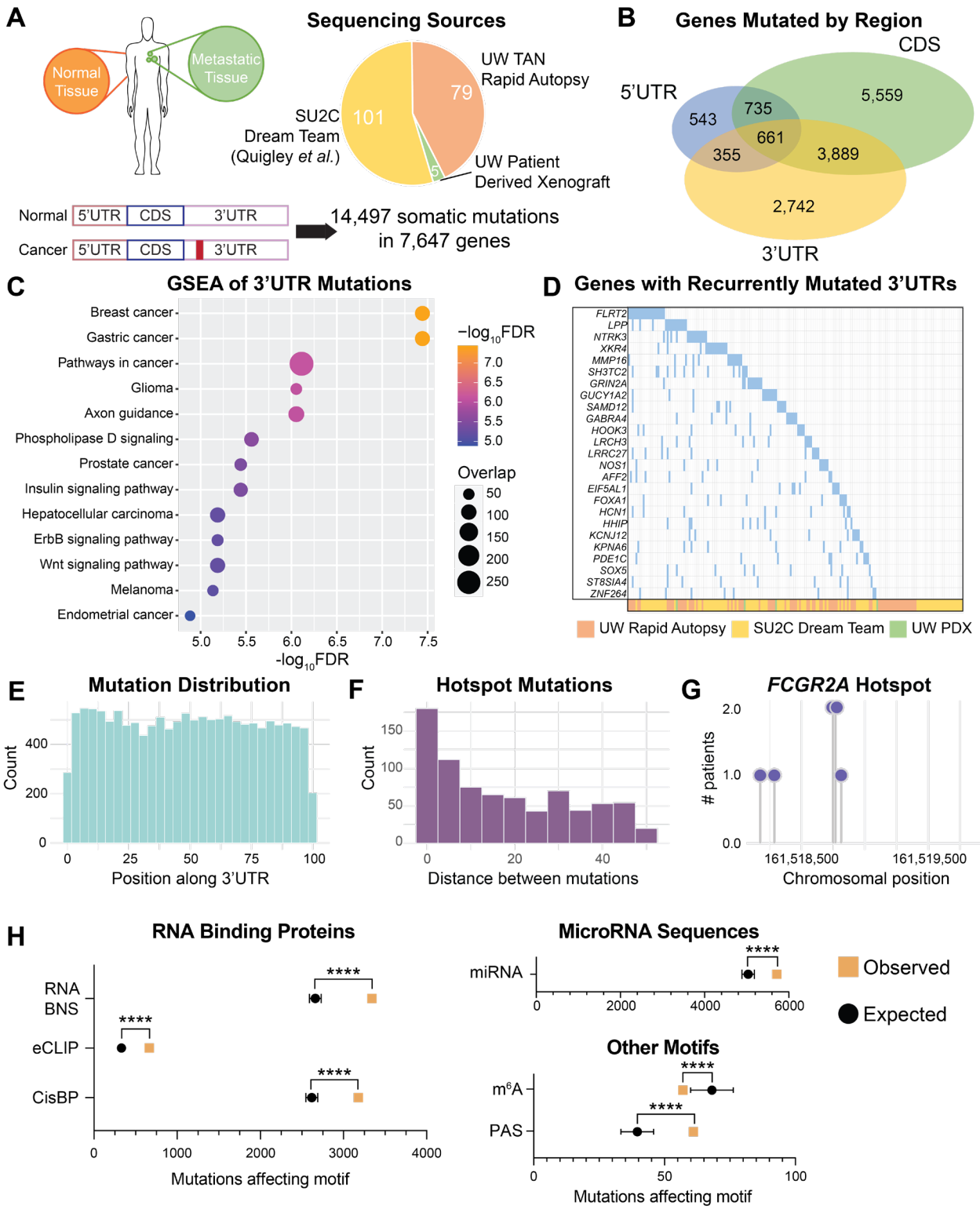

Figure 2-1: mCRPC patients harbor thousands of 3'UTR mutations in cancer-related genes and regulatory elements.

(A) Schematic of patient sequencing sources and mutation calling process.

(B) Venn diagram showing overlap between genes mutated in 5'UTR, CDS, and/or 3'UTR regions.

Figure legend continued on next page

Figure 2-1 continued:
(C) KEGG_2021_Human gene set enrichment analysis (GSEA) for the genes containing 3'UTR mutations in mCRPC patients, with color indicating -$\log_{10}$FDR and dot size indicating the number of genes overlapping between each gene set and the 3'UTR mutated genes.
(D) Waterfall plot of genes with top recurrently mutated 3'UTRs, with patients containing a 3'UTR in each gene indicated across the x-axis.
(E) Metaplot of mutation position along the 3'UTR for all 14,497 mutations, with 0 being the end of the coding sequence and 100 being the transcription stop site, normalized for each gene's 3'UTR length.
(F) Histogram of distances between all 3'UTR mutations found within 50 bases of another patient mutation.
(G) Lollipop plot showing example hotspot mutation area in *FCGR2A*, where multiple patients have clustered mutations along the 3'UTR axis.
(H) Results of *in silico* analysis where the observed number of mutations affecting each 3'UTR regulatory element is compared to the number expected if 3'UTR mutations were randomly distributed along the 3'UTRs, permuted ~10,000 times. One sample t-tests show significance of the observed value compared to the expected distribution (mean ± SD, ****p<0.0001). Abbreviations: RNA Bind-N-Seq (RNA BNS), Encyclopedia of DNA Elements (ENCODE) RNA eCLIP, Catalog of inferred sequence Binding Preferences of RBPs (CisBP), and polyadenylation signal (PAS).

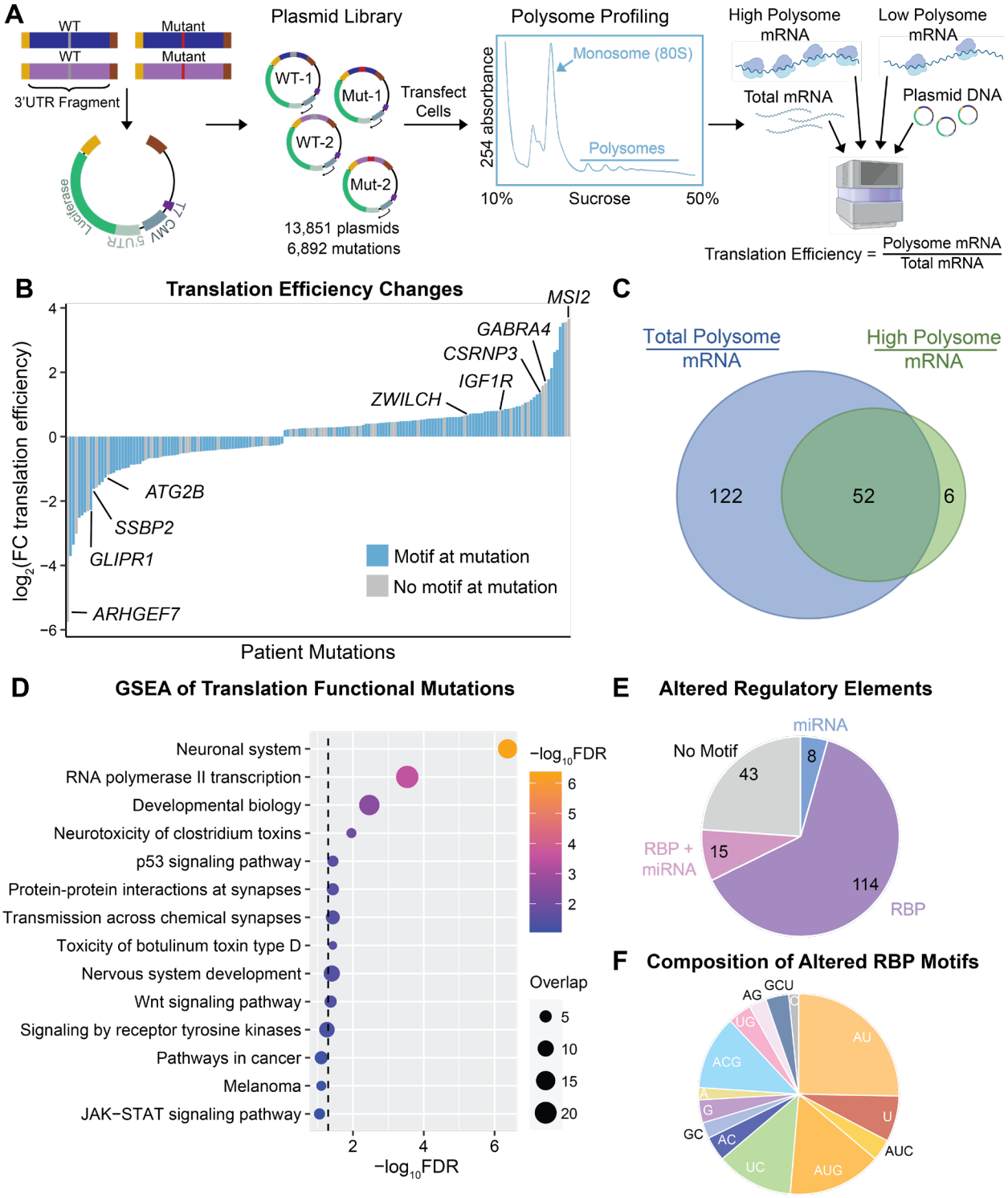

Figure 2-2: Patient-based 3'UTR mutations functionally impact gene-specific translation efficiency. **(A)** Schematic showing MPRA plasmid library construction by Gibson cloning 3'UTR sequence inserts into plasmid backbone, polysome profiling, and sequencing of distinct RNA and DNA samples. Created with Biorender.com.

Figure legend continued on next page

Figure 2-2 continued:
(B) Waterfall plot of all 3'UTR mutations causing significant (FDR<0.10) fold changes (FC) in translation efficiency as determined by total polysome:total mRNA ratio (n=6 biological replicates). Mutations in known regulatory elements indicated in blue.
(C) Overlap between 3'UTR mutations significantly changing TE according to total-polysome:mRNA ratio versus high-polysome:mRNA ratio.
(D) KEGG and Reactome (MSigDB) gene set enrichment analysis for the genes containing 3'UTR mutations that significantly affect TE. Dotted line shows significance cutoff at FDR<0.05.
(E) Proportion of functional 3'UTR mutations that alter miRNA, RBP, and/or no motif.
(F) Pie chart showing base compositions of RBP motifs altered by functional 3'UTR mutations.
Abbreviations: human cytomegalovirus promoter (CMV), T7 bacteriophage promoter (T7).

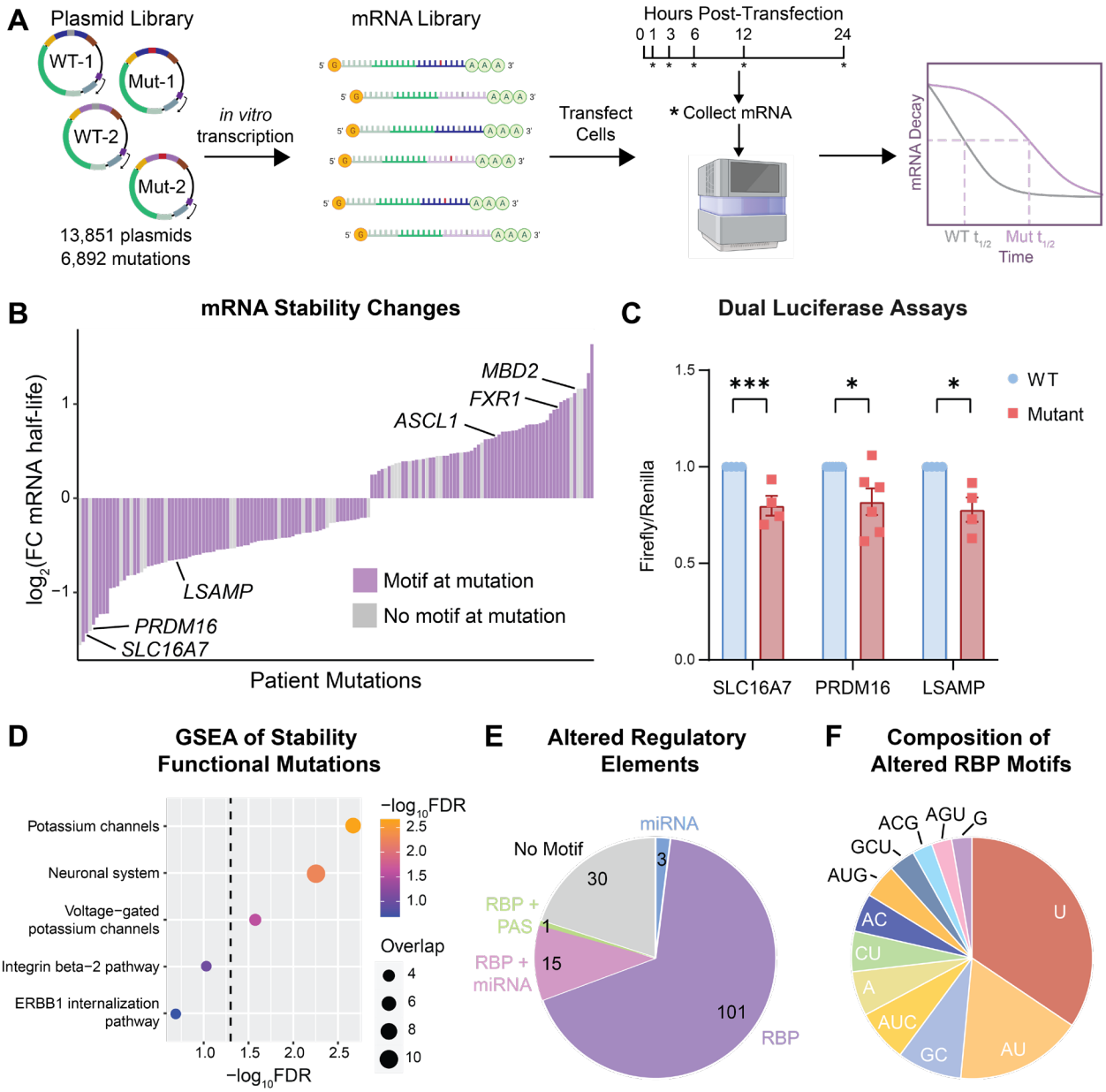

Figure 2-3: Patient-based 3'UTR mutations significantly alter oncogenic mRNA stability

(A) Schematic of mRNA stability MPRA. Plasmid library is *in vitro* transcribed to an mRNA library, which is transfected into cells, and an RNA-seq time course is conducted over 24 hours to measure changes in mRNA decay. Created with Biorender.com.

(B) Waterfall plot showing fold changes in mRNA half-life of all 3'UTR mutations causing significant changes in mRNA stability as determined by unpaired t-test of 1hr-to-3hr or 1hr-to-6hr mRNA decay ($p<0.05$, n=6 biological replicates). Mutations in known regulatory elements indicated in purple.

(C) Results of individual dual luciferase assays performed on three functional 3'UTR mutations. Firefly:Renilla luciferase ratios shown (mean ± SEM) with mutant value normalized to wildtype for each biological replicate (n=4, 6, and 4 biological replicates, respectively, each biological replicate with n=5 technical replicates). Statistical analysis conducted using unpaired t-test and multiple comparisons correction of Benjamini, Krieger, and Yekutieli (*$p<0.05$, ***$p<0.001$).

Figure legend continued on next page

Figure 2-3 continued:
(D) Reactome 2019 gene set enrichment analysis for the genes containing 3'UTR mutations that significantly affect mRNA stability. Dotted line shows significance cutoff at FDR<0.05.
(E) Proportion of functional 3'UTR mutations that alter miRNA seed sequence, RBP motif, polyadenylation signal (PAS), and/or no motif.
(F) Pie chart showing base compositions of RBP motifs altered by functional 3'UTR mutations.

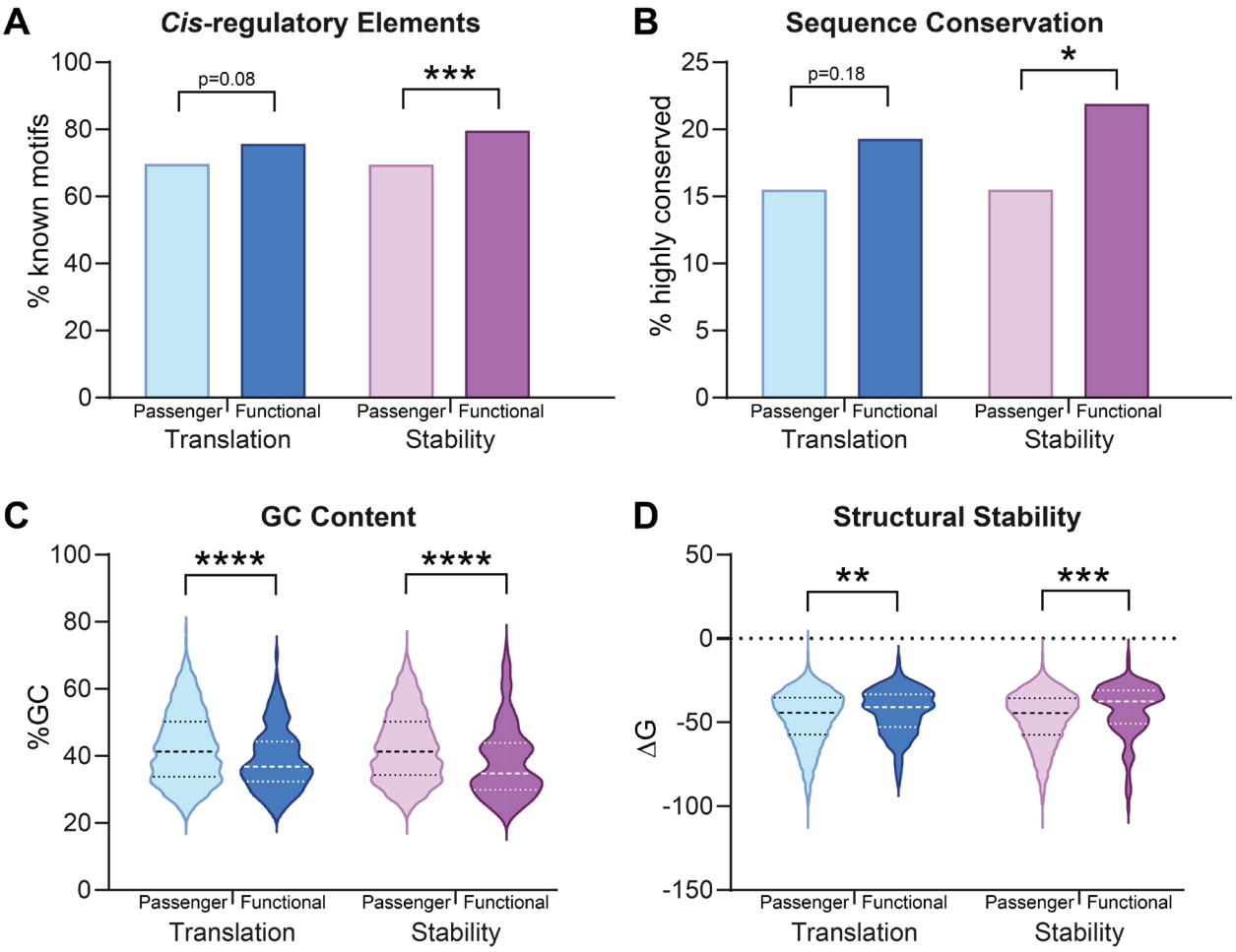

Figure 2-4: 3'UTR mutation functionality is determined by sequence conservation, regulatory motifs, and RNA structure

(A and B) Percentage of passenger (causing no significant expression change) or functional (causing a significant expression change) 3'UTR mutations that alter known regulatory motifs (A) or are highly evolutionarily conserved (PHASTCONS score>0.9) (B), separated by those that effect mRNA translation versus stability. Statistical analysis conducted using a chi-squared test (*p<0.05, ***p<0.001).

(C and D) GC content (C) and ΔG of RNAfold predicted secondary structure (more negative indicates more stable structure, D) for each passenger or functional 3'UTR mutation tested by translation or stability MPRAs. Statistical analysis conducted using two-tailed unpaired t-tests (**p<0.01, ***p<0.001, ****p<0.0001).

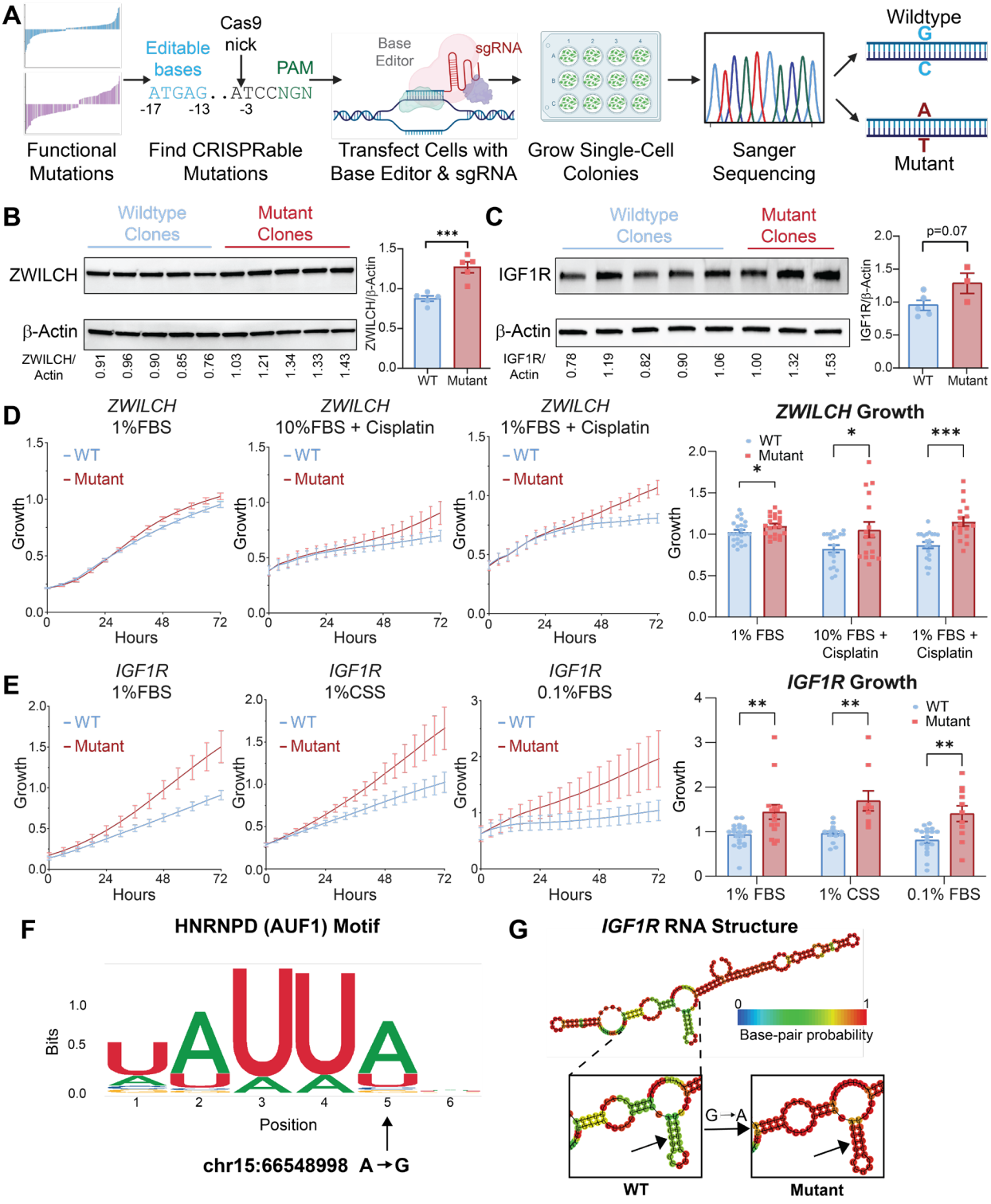

Figure 2-5: Patient-based 3'UTR mutations affect post-transcriptional gene expression and cancerous phenotypes in endogenous cellular contexts

(A) Top hits of MPRAs were analyzed to determine whether they could be CRIPSR base-edited. Two candidates were chosen and cells transfected with corresponding base editors and sgRNAs to grow out single-cell clonal wildtype (WT) and 3'mutant cell lines confirmed by Sanger sequencing. Created with Biorender.com. *Figure legend continued on next page*

Figure 2-5 continued:
(B and C) Western blot of endogenous ZWILCH and β-actin (B) or IGF1R and β-actin (C) for individual wildtype and mutant clonal cell lines in which the respective *ZWILCH* and *IGF1R* 3'UTR mutations have been edited. ImageJ-calculated ratios of ZWILCH or IGF1R to β-actin protein levels shown below blots and summarized at right in bar graphs (mean ± SEM). Statistical analysis conducted using two-tailed unpaired t-tests (***p<0.001). ZWILCH: n=5 WT and 5 Mutant lines; IGF1R: n=5 WT and 3 mutant lines.
(D and E) Growth kinetics of *ZWILCH* (D) and *IGF1R* (E) wildtype and 3'mutant cell lines under various stress conditions. Curves at left show mean ± SEM cell growth dynamics, with growth of all lines within an experiment normalized by setting a single WT line's end-point to 1, allowing comparisons between experiments. Bar graphs at right summarize end-point growth at 72hrs with statistical analysis conducted using unpaired t-test and multiple comparisons correction of Benjamini, Krieger, and Yekutieli (mean ± SEM, *p<0.05, **p<0.01, ***p<0.001). Replicates: ZWILCH, left-to-right: n=5, 4, 4 experimental replicates and n= 24/21, 20/17, 20/17 biological replicates (WT/Mutant lines); IGF1R, left-to-right: n=5, 4, 3 experimental replicates and n= 25/12, 20/12, and 15/9 biological replicates (WT/Mutant lines); all with n=5 technical replicates each.
(F) HNRNPD sequence logo, with base of *ZWILCH* A->G mutation indicated below.
(G) RNAfold predicted secondary structure of WT *IGF1R* sequence (200 bases around 3'UTR mutation). *IGF1R* G→A mutation and resultant change in secondary structure probability shown below.

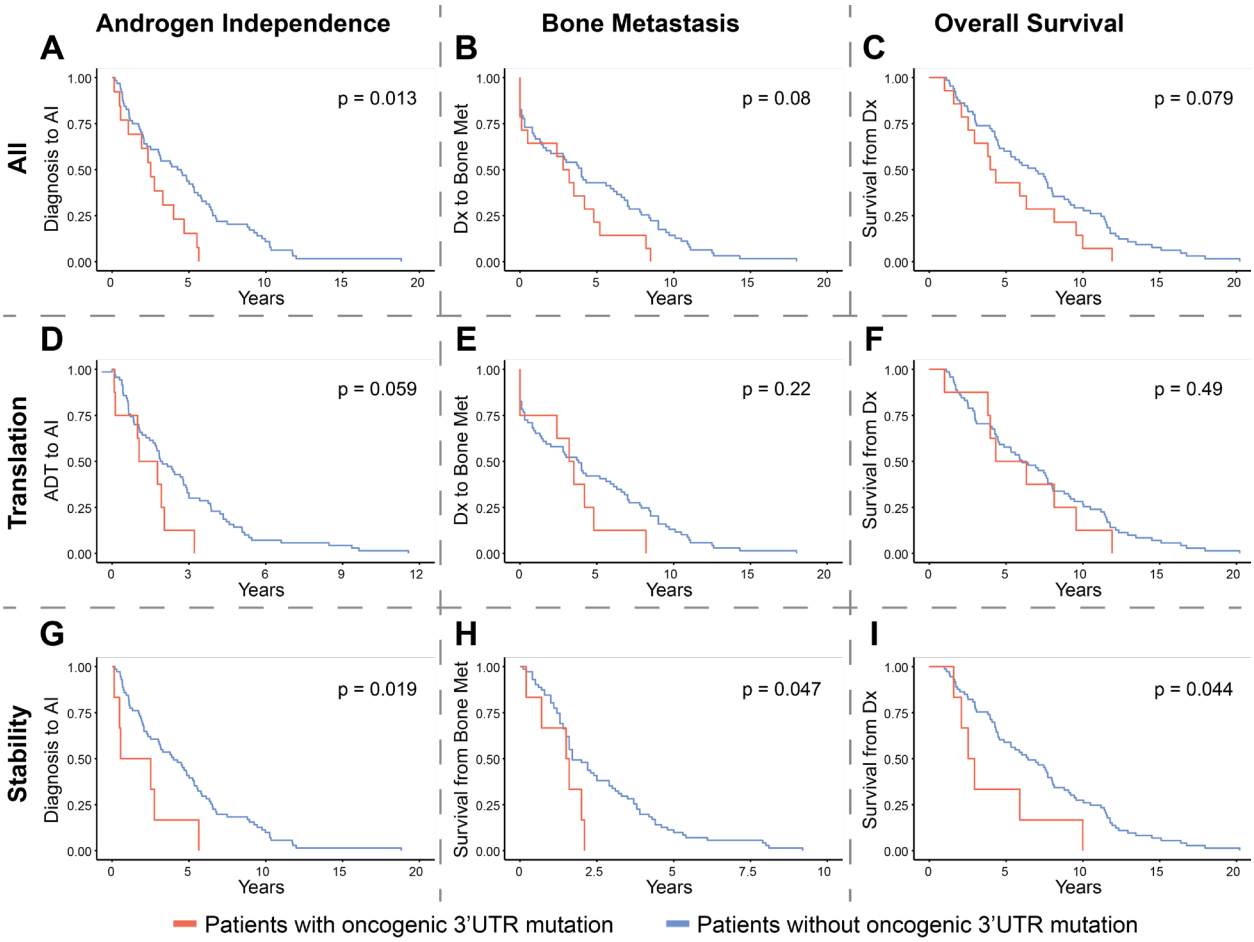

Figure 2-6: Functional oncogenic 3'UTR mutations correlate with poor patient outcomes
Kaplan-Meier plots comparing patients in which any **(A-C)**, translation-related **(D-F)**, or stability-related **(G-I)** functional oncogenic 3'UTR mutations were found versus patients without respective 3'UTR oncogenic mutations. Progression to androgen independence indicated by time to androgen independence (AI) from diagnosis (A, G) or start of androgen deprivation therapy (ADT) (D). Metastatic progression indicated by time from diagnosis (Dx) to bone metastasis (B, E) or survival from first bone metastasis (H). Overall survival measured from diagnosis (C, F, I). Statistical analysis conducted using Kaplan-Meier estimate. Every possible comparison of patients for each outcome statistic was run and select results shown.

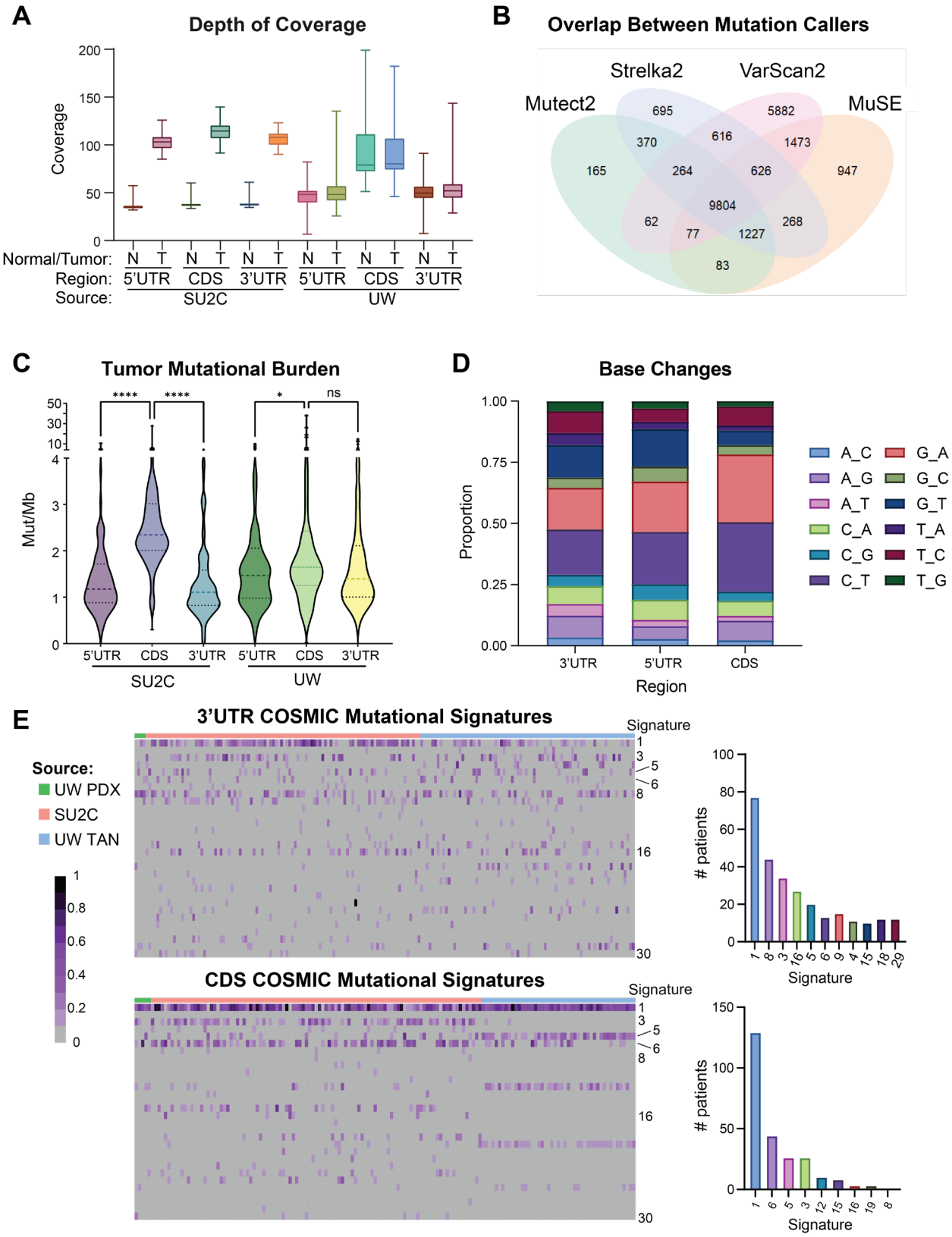

Figure 2-S1: Sequencing statistics and comparison of 3'UTR and CDS mutations

Figure legend continued on next page

Figure 2-S1 continued:
(A) Sequencing coverage of normal (N) and tumor (T) patient samples in each analyzed genomic region. UW group includes both TAN and PDX samples.
(B) Venn diagram showing overlap between 3'UTR mutations called by each mutation caller.
(C) Mutations per megabase in each patient by genomic region and patient source. Statistical analysis conducted using two-tailed unpaired t-tests (*p<0.05, ****p<0.0001, ns=p>0.05).
(D) Proportions of each base change called in mutations by genomic region.
(E) Heatmap showing strength of each of 30 COSMIC signatures in 3'UTR (top) and CDS (bottom) mutations called across patient dataset. Top signatures are quantified at right, showing the number of patients scoring over 0.2 for each signature.

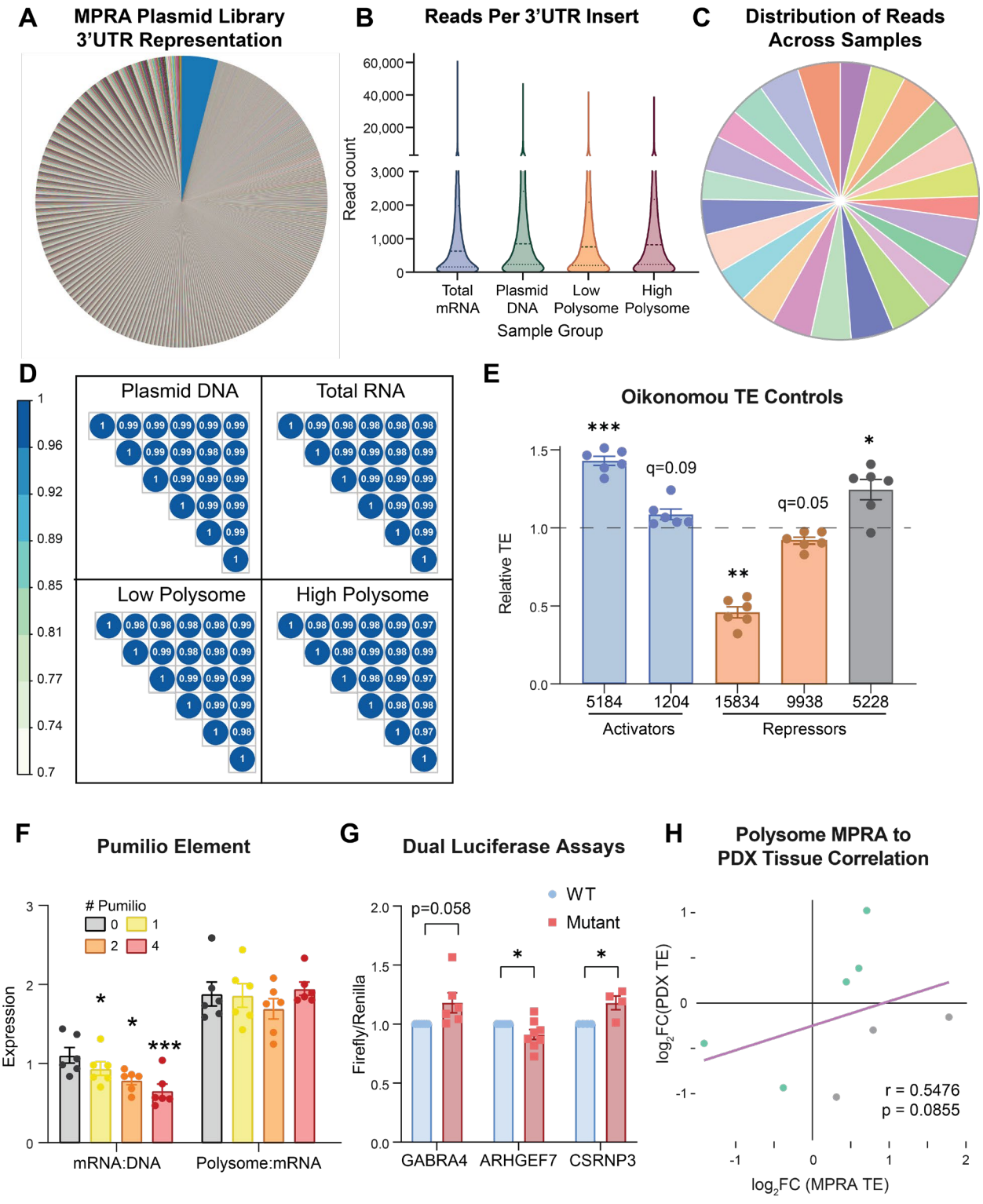

Figure 2-S2: Quality control and validation of MPRA plasmid library and polysome MPRA sequencing

(A) Distribution of 3'UTR inserts in plasmid library upon DNA sequencing. Blue slice represents blank control vector, amplified erroneously high in this sequencing run only.

Figure legend continued on next page

Figure 2-S2 continued:
(B) Sequencing coverage of each 3'UTR insert summed across 6 biological replicates upon sequencing of polysome MPRA results, separated by sample type.
(C) Distribution of total sequencing reads in each of 24 samples (6 biological replicates of 4 sample types) sequenced as a pool.
(D) Pearson's correlation between biological replicates for each sample type.
(E) "Activator" and "repressor" control sequences from Oikonomou *et al.* that were found to significantly change TE (FDR<0.10) in our MPRA results. The control 3'UTR TE in each biological replicate (n=6) was normalized to the average TE across all 13,851 3'UTR MPRA inserts for that replicate and this ratio was plotted (mean ± SEM). Colors indicate agreement with expected TE direction. Statistical analysis conducted using ratio paired t-test and multiple comparisons correction of Benjamini, Krieger, and Yekutieli (*q<0.05,**q<0.01, ***q<0.001).
(F) MPRA results for internal control 3'UTR inserts consisting of a blank vector sequence (0 Pumilio) or a Pumilio element sequence (1, 2, or 4 repeats). Results of these sequences shown for changes in RNA expression (mRNA:plasmid DNA ratio) and TE (total polysome:mRNA ratio) (mean ± SEM, n=6). Statistical analysis conducted using ratio paired t-test and multiple comparisons correction of Benjamini, Krieger, and Yekutieli (*q<0.05, ***q<0.001).
(G) Results of individual dual luciferase assays performed on three functional 3'UTR mutations. Firefly:Renilla luciferase ratios shown (mean ± SEM), with mutant value normalized to wildtype for each biological replicate (n=6, 8, 4, respectively, with 5 technical replicates each). Statistical analysis conducted using unpaired t-test and multiple comparisons correction of Benjamini, Krieger, and Yekutieli (*p<0.05).
(H) Correlation between TE values (total polysome:mRNA ratio) observed for 3'UTR mutations in MPRA (x-axis) and matching TE values obtained from ribosome profiling of UW PDX tissue samples (y-axis). Ribosome profiling expression data was matched by the gene and patient sample in which the 3'UTR mutation was originally found and reported as the change between expression in the mutated tumor sample and the average of the non-mutated samples. Statistical analysis conducted using simple linear regression and Spearman's correlation.

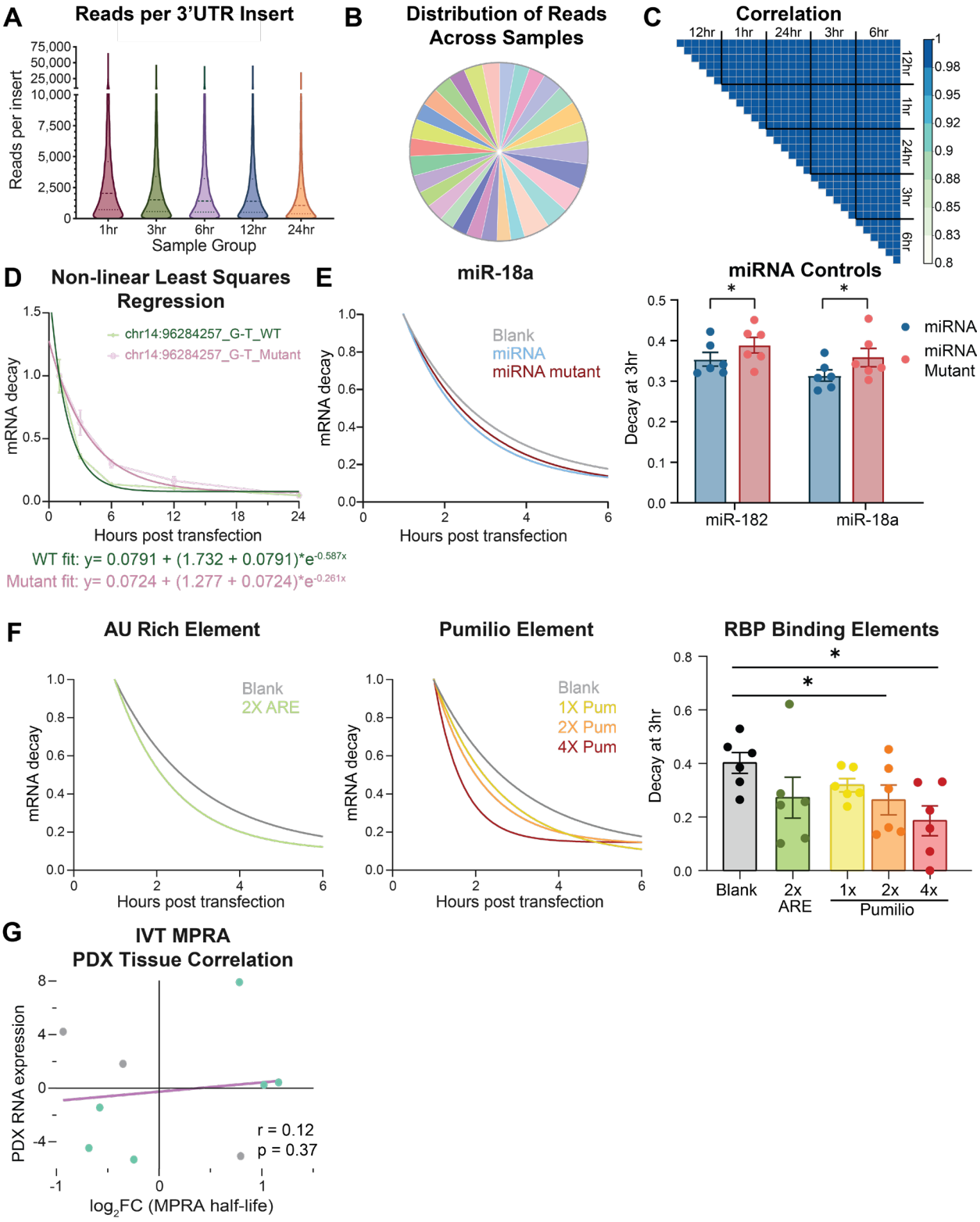

Figure 2-S3: Quality control and validation of IVT MPRA sequencing

(A) Sequencing coverage of each 3'UTR insert summed across 6 biological replicates upon sequencing of MPRA results, separated by time point.

Figure legend continued on next page

Figure 2-S3 continued:
(B) Distribution of total sequencing reads in each of 30 samples (6 biological replicates of 5 sample types) sequenced as a pool.
(C) Pearson's correlation between biological replicates for each sample type.
(D) Example fits of non-linear least-squares regression curves (smooth dark-colored lines) to observed data (straight light-colored lines, mean ± SEM) for a wildtype-mutant 3'UTR pair of inserts. Result fit parameters of each insert shown below.
(E) MPRA results for internal control 3'UTR inserts consisting of a blank vector sequence, miRNA seed sequences, or mutated miRNA seed sequence. NLS curves fit to MPRA data shown for miR-18a (left) and summary results of 1hr to 3hr decay shown for both miR-18a and miR-182 (right, mean ± SEM, n=6). Statistical analysis conducted using ratio paired t-test and multiple comparisons correction of Benjamini, Krieger, and Yekutieli (*q<0.05).
(F) MPRA results for internal control 3'UTR inserts consisting of a blank vector sequence, 2X repeated AU-rich element (ARE), or 1X, 2X, and 4X repeated Pumilio element (Pum). NLS curves fit to MPRA data shown for AU-rich element (left) and Pumilio element (center). Summary results of 1hr-to-3hr decay shown for both (right, mean ± SEM, n=6). Statistical analysis conducted using paired t-test and multiple comparisons correction of Benjamini, Krieger, and Yekutieli (*q<0.05).
(G) Correlation between mRNA half-life changes observed for 3'UTR mutations in MPRA (x-axis) and matching RNA expression values obtained from RNA sequencing of UW PDX tissue samples (y-axis). RNA expression was matched by the gene and sample in which the 3'UTR mutation was originally found and reported as the change between expression in the mutated tumor sample and the average of the non-mutated samples. Statistical analysis conducted using simple linear regression and Pearson's correlation.

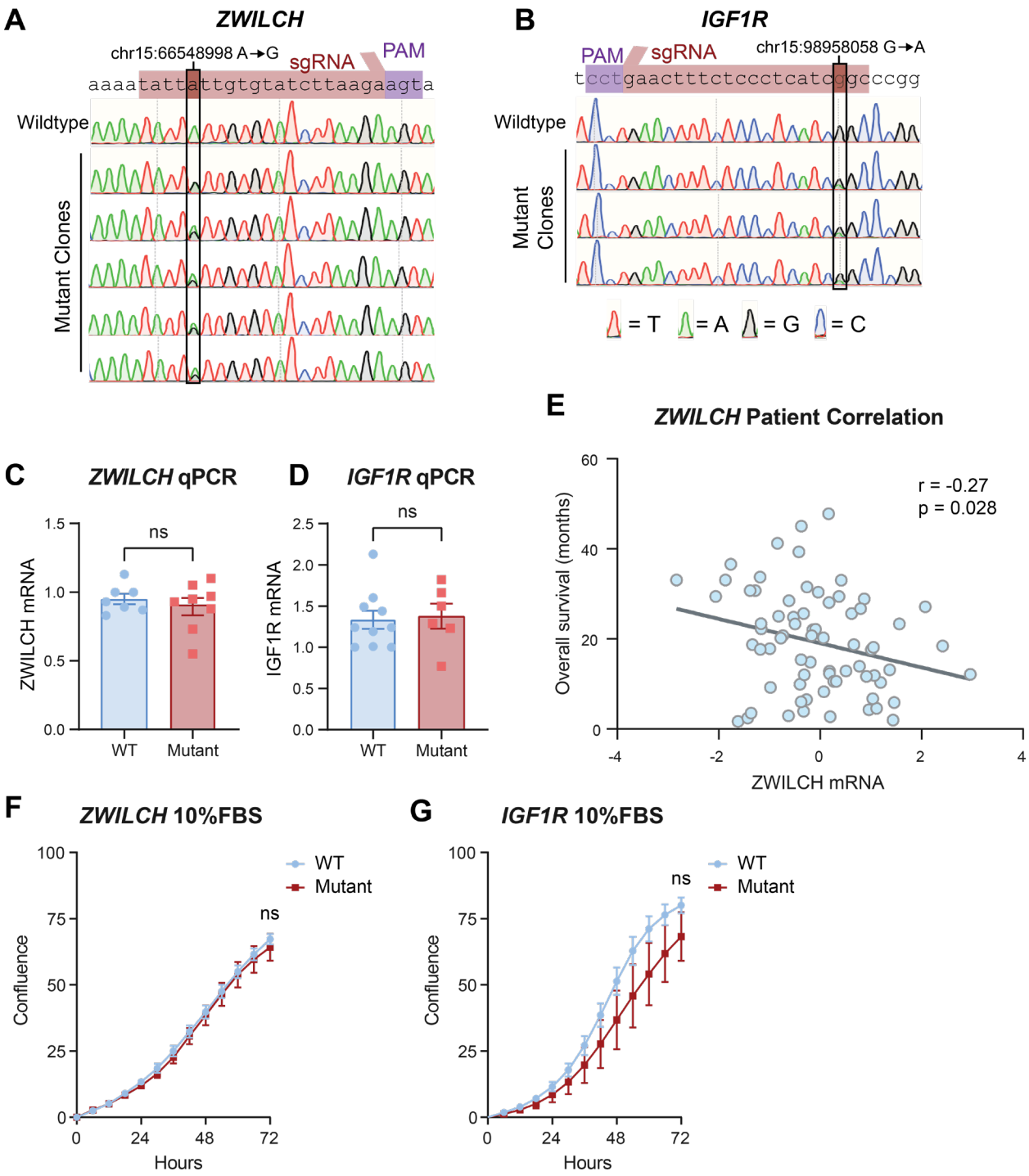

Figure 2-S4: Baseline characterization of CRISPR-edited clonal cell lines and cBioportal mining of *ZWILCH* patient outcome association

(A and B) Sanger sequencing results of *ZWILCH* (A) and *IGF1R* (B) CRISPR-edited clonal cell lines. **(C and D)** qPCR results measuring changes in endogenous *ZWILCH* (C) and *IGF1R* (D) mRNA levels relative to housekeeping β-actin in respective wildtype and mutant CRISPR cell lines (mean ± SEM). Statistical analysis conducted using two-tailed unpaired t-tests (ns=p>0.05). Two experimental replicates for each gene shown, with n=7/8 and 10/6 (WT/Mutant lines), for *ZWILCH* and *IGF1R*, respectively.

Figure legend continued on next page

Figure 2-S4 continued:
(E) Patient data obtained from cBioportal (mPCa SU2C/PCF Dream Team 2019 study) with Pearson's correlation between overall patient survival and *ZWILCH* mRNA expression (*ZWILCH*: mRNA expression z-scores relative to all samples, log FPKM capture).
(F and G) Growth kinetics of *ZWILCH* (F) and *IGF1R* (G) wildtype and 3'mutant cell lines under normal conditions (media with 10%FBS), shown as mean ± SEM confluence over time. Result of unpaired t-test of confluence difference at 72 hours is shown (ns=p>0.05). ZWILCH: n=4 experimental replicates and n= 20/19 biological replicates (WT/Mutant lines); IGF1R: n=2 experimental replicates and n= 10/6 biological replicates (WT/Mutant lines); all with n=5 technical replicates each.

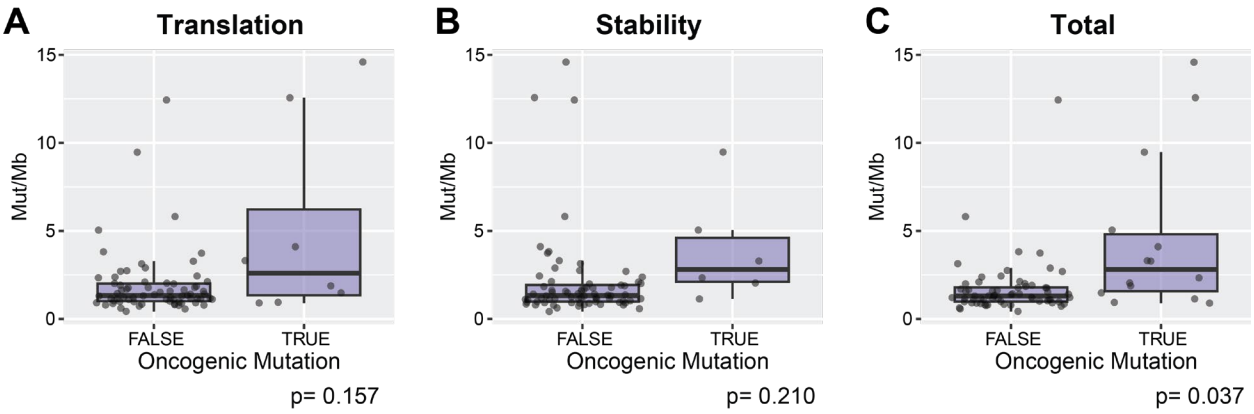

Figure 2-S5: Tumor mutational burden across patient sets

Mutations per megabase (Mut/Mb) in each patient in which translation-related **(A)**, stability-related **(B)**, or any **(C)** functional oncogenic 3'UTR mutations were found versus patients without respective 3'UTR oncogenic mutations. Statistical analysis conducted using two-tailed unpaired t-tests.

MATERIALS & METHODS

Resource Availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Andrew C. Hsieh (ahsieh@fredhutch.org)

Materials availability

Plasmids and mutant cell lines generated in this study are available upon request to the lead contact.

Data and code availability

- MPRA sequencing data for plasmid library sequencing, polysome profiling MPRA, and IVT MPRA is deposited under superseries GSE200304 with reviewer token gbifaywgtxyfjed.
- All original code has been deposited at Zenodo and is publicly available. DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Experimental Models and Subject Details

Cell culture

Cell lines used in this study are PC3 and HEK293T cells. Both were cultured primarily in RPMI 1640 media + L-glutamine (Gibco) supplemented with 10% fetal bovine serum (Cytiva), 1% penicillin/streptomycin (Gibco), and 1% L-glutamine (Gibco), except where noted for treatment. Cell lines were cultured in an incubator at 37°C with 5% CO_2.

Method Details

MPRA plasmid library construction

A pool of 13,851 250bp 3'UTR sequences (Table S2) was manufactured by TWIST Bioscience and resuspended to 10ng/μL in TE buffer. A 5ng portion of the pool was amplified by PCR using Phusion High-Fidelity DNA Polymerase (NEB) following manufacturer protocol for 25 cycles (Table S7, primers #1-2). The 250bp product was purified by agarose gel size selection and extracted using Monarch DNA Gel Extraction Kit (NEB). An additional purification was done using Monarch PCR & DNA Cleanup Kit (NEB). Plasmid backbone pLuc2CP-noARE was linearized by PCR using Q5 High-Fidelity DNA Polymerase (NEB) following manufacturer protocol for 35 cycles, using 5ng input DNA with a 6 minute extension time (Table S7, primers #3-4). The 6,895bp product was purified using Monarch PCR & DNA Cleanup Kit (NEB). To remove template plasmid, the linearized vector was digested with DpnI (NEB) in rCutSmart Buffer (NEB) at 37°C for 15 minutes.

The 3'UTR insert sequence pool and linearized pLuc2CP-noARE were assembled into plasmids using Gibson Assembly Master Mix (NEB) with a ratio of 1:15 backbone to fragment DNA. The reaction was incubated at 50°C for 60 minutes. Cloning efficiency was enhanced by incubating the completed Gibson reaction at room temp for 60min on a 0.025μm filter (Millipore) floating in a 10cm petri dish of UltraPure DNase/RNase-Free Distilled Water (ThermoFisher). The reaction was recovered and immediately cloned into Stellar Electrocompetent Cells (Takara) following manufacturer protocol. Transformed cells were plated on 4 LB-ampicillin 24.5cm assay plates (Fisher) to allow for individual colony growth. An aliquot of transformed cells was plated on 10.0cm LB ampicillin plates in a dilution series for quality control and representation calculation.

All plates were incubated overnight at 37°C. The bacteria were removed from the assay plates and stored at -20°C until plasmid isolation. 4 colonies from each assay plate were sequenced to confirm 3'UTR sequence insertion. The Gibson and cloning process was repeated until insert representation of 259x was achieved. Plasmid DNA was isolated using PureLink HiPure Plasmid Filter Maxiprep Kit (Invitrogen).

Polysome profiling

PC3 cells were plated at a density of 3 million cells per 15cm plate, with three 15cm plates pooled to make each biological replicate. Cells were transfected with 16µg MPRA plasmid library using 48µL Fugene HD (Promega) per plate 24 hours after plating. Cell media was changed to fresh media 16 hours after transfection and collected 24 hours after media change. At time of collection, cells were trypsinized and pelleted at 300xg for 5 minutes, then treated with 100 µg/mL (final concentration) cycloheximide (Sigma) for 10 min on ice. After treatment, cells were pelleted and flash frozen until all biological replicates were collected.

Cell pellets were lysed on ice in 220µL of polysome lysis buffer (10 mM Tris-HCl pH 7.4 (Ambion), 132 mM NaCl (Ambion), 1.4 mM MgCl2 (Ambion), 19 mM DTT (Sigma), 142 µg/mL cycloheximide (Sigma), 0.1% Triton X-100 (Fisher), 0.2% NP-40 (Pierce), 607 U/mL SUPERase-In RNase Inhibitor (Invitrogen) with periodic vortex mixing. Lysates were clarified by centrifugation at 9300xg for 5 min and supernatants were transferred to fresh tubes. Three 220µL lysates were combined for each replicate for a total volume of 660µL per sample. This total lysate was split into three parts: 60µL for plasmid DNA extraction, 150µL for total mRNA isolation, and 450µL for polysome profiling.

For each sample, the 450µL lysate fraction was layered onto a 10% to 50% (w/v) linear sucrose gradient (Fisher) containing 2 mM DTT (Sigma) and 100 µg/mL heparin (Sigma). The gradients were centrifuged at 37,000 rpm for 2.5hr at 4°C in a Beckman SW41Ti rotor in Seton 7030 ultracentrifuge tubes. After centrifugation, samples were fractionated using a Biocomp Gradient Station by upward displacement into collection tubes, through a Bio-Rad EM-1 UV monitor (Bio-Rad) for continuous measurement of the absorbance at 254 nm. 820µL of TRIzol Reagent (Invitrogen) were added to each RNA fraction and stored at -80°C.

In vitro transcription

MPRA IVT template preparation: DNA template for MPRA *in vitro* transcription was prepared by first digesting 65µg MPRA plasmid library DNA with FseI restriction enzyme (NEB) in rCutSmart buffer (NEB) for ~90 minutes. FseI singly cuts the MPRA plasmid backbone directly downstream of the luciferase coding sequence and inserted 3'UTRs. Digested DNA was run on a 1% agarose gel and size extracted using NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel). Phenol:chloroform and isopropanol DNA isolation was used to purify extracted DNA, resulting in ~20µg purified, gel-extracted MPRA template.

MPRA in vitro *transcription:* MPRA template was *in vitro* transcribed, capped, tailed, and purified using the mMESSAGE mMACHINE T7 ULTRA Transcription Kit (Invitrogen). Eleven reactions in total were carried out, each using 1µg of MPRA template in the manufacturer recommended reaction with the addition of 1µL SUPERase-In RNase Inhibitor (Invitrogen). Reactions were incubated at 37°C for 4 hours, then 1µL Turbo DNase (Invitrogen) added and incubated at 37°C for an additional 15 minutes for removal of DNA template. Poly-A tailing protocol was followed to manufacturer's instructions with a 45-minute incubation at 37°C. RNA recovery was performed using lithium chloride precipitation according to kit protocol. All eleven reaction products (~350µg) were pooled to make the MPRA mRNA library.

Exogenous spike-in control template preparation: DNA template for exogenous control spike-in RNA was prepared from pDualLuc (a generous gift from Arvind Subramaniam) using a two-step PCR approach to add MPRA-compatible ends and T7 promoter to a portion of the nanoluciferase coding sequence. PCRs were carried out using 50µL reactions in Q5 High Fidelity 2X Master Mix (NEB) using pDualLuc-specific primers (Table S7, primers #5-8) and purified using DNA Clean & Concentrator (Zymo Research).

Exogenous spike-in control in vitro *transcription:* Nanoluciferase exogenous spike-in control RNA was *in vitro* transcribed using MAXIscript T7 Transcription Kit (Invitrogen). Protocol was followed to manufacturer's specifications using 1µg starting material, 1 hour incubation at 37°C, and Turbo DNase (Invitrogen) digestion. RNA was purified using lithium chloride precipitation as described in mMESSAGE mMACHINE T7 ULTRA Transcription Kit (Invitrogen).

In vitro transcription RNA-seq time-course

PC3 cells were plated in 10cm plates at 1.5 million cells per plate. Cells were transfected the day after plating with 8µg *in vitro* transcribed MPRA mRNA library using 40µL Lipofectamine MessengerMAX (Invitrogen) transfection reagent per plate. Cells were washed and media changed one hour after mRNA transfection to remove residual RNA in media before collection. One 10cm plate of transfected cells was collected for each of 6 biological replicates at 1, 3, 6, 12, and 24 hours after transfection, amounting to 30 total plates/samples. Collection was performed on ice by gently scraping cells into cold PBS and pelleting at 350xg for 5 minutes at 4°C. Pellets were each resuspended in 1mL pre-spiked TRIzol reagent (Invitrogen). TRIzol was spiked with 0.7ng nanoluciferase spike-in control RNA in 35mL total TRIzol (20pg spike-in RNA per sample) for normalization across samples.

MPRA Sequencing

Polysome profiling cDNA and plasmid DNA preparation: Total (input), monosome-associated (fraction 5), low polysome-associated (fractions 8 and 9), and high polysome-associated (fractions 10, 11, and 12) mRNA samples were isolated from TRIzol (Invitrogen) using the Direct-zol RNA Miniprep Plus Kit (Zymo Research) with DNaseI treatment according to manufacturer's directions. RNA was eluted in 50µL nuclease-free water and, for polysome samples, the multiple fractions were then pooled and concentrated using RNeasy MinElute Cleanup Kit (Qiagen). cDNA was made from 1µg RNA per sample using SuperScript III First-Strand Synthesis System (Invitrogen) and MPRA-specific primer (Table S7, primer #9). Plasmid DNA was isolated from 60µL cell lysate fractions using a modified QIAprep Spin Miniprep Kit (Qiagen) protocol, starting with addition of P2 to lysate and continuing manufacturer's directions from this step.

IVT RNA-seq time course cDNA preparation: RNA was isolated from samples collected in TRIzol (Invitrogen) using a chloroform-isopropanol protocol. 200µL chloroform was added to 1mL TRIzol samples, transferred to a heavy phase-lock gel tube (QuantaBio), and spun in a microcentrifuge at 12,000xg for 15 minutes at 4°C. The aqueous layer was transferred to a new tube containing 1µL of glycogen (Fisher Scientific) and equal volume isopropanol added. After 10 minutes of incubation at room temperature, sample was spun at 20,000xg for 20 minutes at 4°C, then RNA pellet was washed with fresh 75% ethanol before drying and resuspending in 40µL nuclease-free water. Isolated mRNA was DNase treated using Turbo DNase (Invitrogen) in reactions consisting of 11µL (~15µg) RNA and 1.5µL Turbo DNase according to manufacturer's protocol. DNase-treated RNA was purified using lithium chloride precipitation as described in mMESSAGE mMACHINE T7 ULTRA Transcription Kit (Invitrogen). 2µg RNA per sample was reverse transcribed into cDNA using SuperScript III First-Strand Synthesis System (Invitrogen) and MPRA-specific primer (Table S7, primer #2).

Two-step PCR amplicon approach to prepare library pools for sequencing: Library pools for MPRA plasmid library, polysome-based MPRA, and IVT-based MPRAs were assembled separately using the same protocol. Each sample (1 for plasmid library, 24 for polysome MPRA, and 30 for IVT MPRA) was individually amplified using Q5 High Fidelity 2X Master Mix (NEB) in a 50μL reaction with 250ng plasmid DNA (MPRA plasmid library), 9.5μL of cDNA or 250ng of plasmid DNA (Polysome MPRA), or 10μL cDNA (IVT MPRA) and MPRA-specific primers (Table S7, primers #10-11). Thermocycling consisted of: 98°C for 30 seconds; 10 cycles of 98°C for 10 seconds, 71°C for 30 seconds, 72°C for 30 seconds; and a final extension at 72°C for 5 minutes. 5μL PCR product was run on an agarose gel to confirm size and remaining 45μL was purified using AMPure XP beads (Beckman Coulter) at a 1.8x ratio. Second round PCR was performed using 20μL round 1 purified PCR product in a 50μL total volume Q5 High Fidelity 2X Master Mix (NEB) reaction and IDT® for Illumina® DNA/RNA UD Indexes (Illumina) primers (Set C used for polysome MPRA, Set A used for IVT MPRA). Thermocycling consisted of: 98°C for 30 seconds; 10 cycles (8 cycles for MPRA plasmid library) of 98°C for 10 seconds, 67°C for 30 seconds, 72°C for 30 seconds; and a final extension at 72°C for 5 minutes. 5μL PCR product was run on an agarose gel to confirm size and remaining 45μL was purified using AMPure XP beads (Beckman Coulter). MPRA plasmid library sample was bead purified at a 1.8x ratio. Polysome MPRA samples were bead purified twice, once at a 1.8x ratio then once at a 0.7x ratio. IVT MPRA samples were bead purified once at a 0.7x ratio.

The concentrations of final purified PCR products were measured using a Qubit Fluorometer (Invitrogen) and size/purity was verified using an Agilent 4200 Tapestation. For the polysome and IVT MPRAs, all samples were pooled for sequencing using 1.869ng per sample (polysome MPRA) or 1ng per sample (IVT MPRA). All sequencing was performed by the Fred Hutchinson Cancer Center genomics core. MPRA plasmid library was sequenced on a MiSeq using the

Nano PE150 run configuration. Polysome MPRA was sequenced on a HiSeq 2500 using the Rapid PE150 run configuration. IVT MPRA was sequenced on a NextSeq 2000 using the P3 PE150 run configuration.

Dual luciferase assays

Individual luciferase plasmids with target 3'UTRs were cloned out of the total MPRA plasmid library using inverse PCR followed by plasmid ligation. Briefly, PCRs were performed for each plasmid of interest using the MPRA plasmid DNA library as a template, Q5 High-Fidelity 2X Master Mix (NEB), and opposing 3'UTR mutation-specific primers that would amplify the entirety of the plasmid beginning with the unique mutation region (Table S7, primers #12-29). After amplification, a mixture of Dpn1 (NEB), T4 PNK (NEB), and T4 DNA Ligase (NEB) was added for linearization of the PCR product and removal of the plasmid library template. These were transformed into NEB 5-alpha Competent *E. coli* and resultant plasmids confirmed by Sanger sequencing.

PC3 cells were plated at 8,000 cells per well in 96-well plates and transfected 24 hours later using 0.4μL of Fugene HD (Promega), 100ng pLuc2CP-noARE plasmid containing WT or mutant 3'UTR sequence, and 1ng of control Renilla plasmid (pRL-UBC) per well. Each WT or mutant construct was transfected into five technical replicate wells per experiment and repeated in four or more biological replicates. Promega's Dual Luciferase Assay Reporter System was used to lyse cells in PLB and measure luciferase activity 24 hours after transfection according to manufacturer's instructions. Luminescence was read using a Cytation 5 plate reader (BioTek).

CRISPR base editing

A custom R script was used to filter MPRA results by whether they could be CRISPR edited using available base editing systems. This included filtering for A→G or C→T mutations and

searching for nearby PAM sites situated at the proper distance from the mutation of interest. 3'UTR mutations in *ZWILCH* (chr15:66548998 A→G) and *IGF1R* (chr15:98958058 G→A) were chosen for editing.

To edit the *ZWILCH* 3'UTR mutation into PC3 cells, we used a system consisting of the NG-ABE8e adenine base editor with NG PAM specificity (a gift from David Liu, Addgene plasmid #138491) and a mutation-specific sgRNA plasmid with sgRNA: 5'-TATTATTGTGTATCTTAAGA-3'. To edit the *IGF1R* mutation into HEK293T cells, we used a system consisting of the evoAPOBEC1-BE4max cytosine base editor with NGG PAM specificity (a gift from David Liu, Addgene plasmid #122611) and a mutation-specific sgRNA plasmid with sgRNA: 5'-GCCGATGAGGGAGAAAGTTC-3'. sgRNA plasmids were cloned from the pFYF1548 EMX1 sgRNA plasmid backbone (a gift from Keith Joung, Addgene plasmid #47508) using the Q5 Site-Directed Mutagenesis Kit (NEB) and sgRNA-specific primers (Table S7, primers #30-33).

To introduce editing systems to cells, 1.25 million PC3 cells (ZWILCH) or 1.75 million HEK293T cells (IGF1R) were plated in a 10cm dish and transfected the next day with 6µg base editor plasmid, 2µg sgRNA plasmid, 200ng pMaxGFP, and 24µL Fugene HD (Promega). For PC3 *ZWILCH* editing, transfected cells were treated with 10nM romidepsin (Selleckchem) to increase editing efficiency[83] 8 hours after transfection, then media was changed to remove reagents 24 hours after romidepsin treatment. For 293T *IGF1R* editing, no romidepsin was used, but media was changed 24 hours after transfection to remove transfection reagents. Transfected cells were collected by trypsinization 72 hours after transfection and resuspended in 500µL PBS for flow cytometry. Cells were sorted based off live/dead and GFP+/- criteria on a Sony SH800 sorter. Live, GFP+ cells were sorted first as single cells directly into 96-well plates for clonal growth, then a second portion of cells was collected in bulk for polyclonal Sanger sequencing.

As single-cell colonies grew in 96-well plates, they were transferred to 24-well plates and then 6-well plates before a pellet was collected for monoclonal Sanger sequencing.

Sanger sequencing was completed by isolating genomic DNA from cell pellets using the DNeasy Blood & Tissue Kit (Qiagen), PCR amplifying the region around the 3'UTR mutations, and Sanger sequencing through Genewiz. PCR amplification was conducted in a 25µL reaction of Phusion High Fidelity PCR Master Mix (Thermo Scientific) using locus-specific primers surrounding each mutation (Table S7, primers #34-37).

Western blotting

Cell pellets were lysed using RIPA buffer (Fisher Scientific) supplemented with protease and phosphatase inhibitors (Sigma). 30µg (IGF1R) or 20µg (ZWILCH) protein per sample was run on a polyacrylamide gel and transferred to a PVDF membrane. Proteins were detected using respective primary and HRP-conjugated secondary antibodies, including rabbit anti-IGF1R (Cell Signaling 3027, 1:500), rabbit anti-ZWILCH (Abcam 202898, 1:1,000), mouse anti-β-actin (Sigma A5316, 1:1,000), goat anti-rabbit (Invitrogen 31460, 1:5,000), and goat anti-mouse (Invitrogen 31430, 1:5,000).

RT-qPCR

Total RNA was extracted from cell pellets collected from CRISPR WT and mutant cell lines at two separate times using TRIzol reagent (Invitrogen). Briefly, TRIzol and chloroform were added to lyse cell pellets, and isopropanol used to precipitate RNA from the resultant aqueous layer. iScript Reverse Transcription Supermix (Bio-Rad) was used to reverse transcribe cDNA from 1µg RNA per sample. qPCR was performed using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) for ZWILCH, IGF1R, and β-actin (Table S7, primers #38-43).

Cell growth assays

4,000 cells per well were plated in 96-well plates using the media (10%FBS, 1%FBS, 0.1%FBS, or 1%CSS [Gemini Bioproducts]) specific to the experimental condition. For cisplatin-treated conditions, media was changed 24 hours after plating to include 15μM cisplatin (Sigma Aldrich). Each cell line was plated in five technical replicate wells per condition and three or more biological replicate experiments were performed. Cells were moved 24 hours after plating (or immediately following cisplatin treatment) to an IncuCyte S3 or IncuCyte Zoom (Sartorius) for continuous imaging and confluence analysis. Cells were imaged over the course of 72 hours and IncuCyte software used to measure change in confluence over this time. For comparison between experimental replicates, all values within an experiment were normalized around a single wildtype cell line.

Quantification and Statistical Analysis

Obtaining publicly available datasets

BAM files for 101 SU2C project tumor/matched normal mCRPC patients were obtained from Quigley *et al.* (dbGaP accession code phs001648.v2.p1)[3]. BEDtools[84] "bamtofastq" was used to extract raw sequencing data from BAM files. Fastq files for UW TAN and PDX UTR sequencing data, UW TAN exome sequencing data, and UW PDX ribosome profiling data were downloaded from Lim *et al.* (UW UTR sequencing data: dbGaP accession code phs001825.v1.p1; UW exome sequencing data: GEO repositories GSE147250 and GSE171729; and PDX ribosome profiling data: GEO repository GSE130465)[17].

Alignment to hg38 for WGS and WES samples

All downloaded fastq files were aligned to hg38 using Bowtie2 (v2.4.2)[85]. The reference used for alignment was downloaded from GDC portal (https://gdc.cancer.gov/about-data/gdc-data-

processing/gdc-reference-files/GRCh38.d1.vd1.fa.tar.gz). The reads with low quality were filtered and duplicates were marked using Picard (v2.25.0, http://broadinstitute.github.io/picard) and GATK[86] (v4.2.5). UTR and CDS coverage were calculated using GATK DepthOfCoverage.

Subtraction of reads that aligned to mouse genome

For genome and ribosome profiling sequencing of UW PDX tissues, mouse genome subtraction was used to separate mouse contamination from human tissue sequencing. Short reads were aligned to both human reference genome hg38 and mouse reference genome mm10 separately using STAR2. XenofiltR[87] was used to retain those alignment records which were found with much higher fidelity in the human genome compared to mouse genome.

Somatic mutation analysis

MuTect (v2)[26], Strelka (v2.9.2)[27], MuSE(v2.0)[28], and VarScan2[29] were used to identify somatic single-nucleotide variants within the UTRs and CDS for each tumor and matched normal pair. For Mutect2, separate panels of normal were constructed for the SU2C and UW datasets using the matched normal samples. Three different bed files were used in separate runs for obtaining 3′UTR, 5'UTR, and CDS mutations: ncbi_hg38_knownGene_utr3_bed.bed, ncbi_hg38_knownGene_utr5_bed.bed, and ncbi_hg38_knownGene_cds_bed.bed, respectively. Annovar[88] was used for annotating the variants. For SU2C samples obtained from Quigley *et al.*[3], the following cutoffs were applied to derive a final list of mutations: Ref_reads_in_normal>=8, Total_reads_in_Tumor>=28, Alt_reads_in_Tumor>=10, Tumor_VAF>=0.1, Normal_VAF<=0.05. For UW TAN and PDX samples obtained from Lim *et al.*[17], the following cutoffs were applied: Ref_reads_in_normal>=8, Total_reads_in_Tumor>=14, Alt_reads_in_Tumor>=5, Tumor_VAF>=0.1, Normal_VAF<=0.05. Variant allele frequency (VAF) refers to the fraction of sequencing reads overlapping a genomic coordinate that supports the non-reference (mutant/alternate) allele. Mutations obtained from 2 or more callers were retained

for all further mutation analysis. Mutations in hypermutant samples 10-068 (UW TAN), DTB-083 (SU2C), and LuCaP147 (UW PDX) were removed from mutational analysis.

Tumor mutational burden

Tumor mutational burden (mutations per megabase) for each patient was calculated from all variants (passing filters and called by two or more tools) separately in the 3'UTR, 5'UTR and CDS regions as defined by the UCSC knownGene Table.

Mutational signature analysis

Mutational signatures for 3'UTR mutations and CDS mutations were evaluated separately using DeconstructSigs[89]. The deconvoluted mutation signature frequency was derived using 30 predefined COSMIC SBS V2 signatures.

3'UTR mutational motif analysis

Analysis of 3'UTR mutations within *cis*-element regulatory regions was performed by examining if the observed mutations in our patient cohort disrupt RNA-binding protein motifs, miRNA seed sequences, the polyadenylation signal AAUAAA, or the m^6A RRACH motif. Position weight matrices describing RNA binding protein motifs were obtained from three sources: the CisBP-RNA database[90], RNA Bind-N-Seq project[55], and mCross analysis of ENCODE eCLIP data[91]. Seed sequences of human miRNAs were obtained from miRBase[92] and the expression of miRNAs in the PC3 cell lines obtained from ENCODE project ENCSR387TQN. Only the top 100 expressed miRNAs in PC3 cells were used for motif analysis to filter for motif changes that could affect binding of only well-expressed miRNAs. A custom set of Python scripts was written to determine whether the observed counts of 3'UTR mutations were statistically enriched within these regulatory elements when compared to a random model preserving sequence-specific characteristics such as trinucleotide context. In this analysis, mutations that impacted pre-

existing regulatory elements or introduced new elements were both considered. First, each observed mutation was analyzed as to whether it added or removed the sequence of each motif. To generate the background distribution, permutations of all 3'UTR mutation locations found within our dataset were performed ~10,000 times. The original mutational frequency of all specific transversions, transitions, and trinucleotide context (a total of 288 possible mutations are possible under this scheme—64 possible codons plus 32 additional with no nucleotide in exclusively the first or third position, each with three possible mutations to the middle base) were taken into account. The number of mutations in these permutations that affected each type of motif in each database or specific element was counted. The total number of observed mutations impacting each regulatory element type was compared to the background distribution of the permutation data using a one-sample t-test.

Design of MPRA plasmid library

The MPRA plasmid library consists of 66 control sequences and 6,892 pairs of WT and Mutant sequences based on patient mutations and their surrounding endogenous 3'UTR sequence. All 13,851 3'UTR insert sequences are 201 nucleotides long. There are three types of control sequences based on miRNA seed sequences[92], the PTRE-Seq 3'UTR MPRA[19], and a 3'UTR MPRA by Oikonomou *et al*[20]. For the patient mutation sequences, transcript ids, genomic coordinates, and transcription stop sites for the 3′UTR of each mutated gene were obtained from UCSC's Refseq Table using R/Bioconductor package GenomicFeatures[93]. 3′-UTR sequences were retrieved using R/Bioconductor package "BSgenome.Hsapiens.UCSC.hg38" (DOI:10.18129/B9.bioc.BSgenome.Hsapiens.UCSC.hg38). These were used to extract the 100 bases upstream and 100 bases downstream of each mutation of interest.

MPRA sequencing and statistical analysis

Alignment and read counts: Fastq files were aligned to a custom reference genome consisting of our MPRA library 3'UTR sequences using Bowtie2[85] (v2.4.2) allowing for 0 mismatches. WT and mutant barcodes were segregated to make samples for each sample group. Log_2CPM values were calculated using edgeR[94]. The normalized count data ($log_2(cpm+1)$) was used for Pearson's correlation analysis in R and visualized using ggplot2[95].

Polysome MPRA statistics: For polysome MPRA statistical analysis, xtail[96] was used to identify differentially regulated 3'UTRs. Translation efficiency was calculated by total polysome (high polysome + low polysome) to total RNA and high polysome to total RNA ratios for each 3'UTR. RNA expression changes, used for internal control validation, were calculated by total RNA to plasmid DNA ratios. A ratio of ratios was then calculated to compare Mutant TE to WT TE for each 3'UTR mutation and an FDR<0.10 in this comparison was considered significant.

IVT MPRA statistics: For IVT MPRA statistical analysis, all counts within each sample were first normalized using the spike-in control counts of that sample. Then, decay ratios were calculated from 1hr-to-3hr and from 1hr-to-6hr. These decay values were compared between paired WT and Mutant 3'UTR sequences across the 6 replicates using two-sample t-tests. Additionally, non-linear least squares regression curves were fitted to each 3'UTR insert to calculate mRNA half-lives. This was done using the nlstools package[97] in R using the exponential decay function y ~ yf + (y0 - yf) * exp(-alpha * t) and the starting parameters y0 = 0.5, yf = 0, alpha = 0.1. Half-lives were calculated based on resultant alpha values using the relationship half-life = ln(2)/alpha. Mutations that were considered to have significantly changed mRNA stability passed all the following filters: (A) sum of 1hr read counts ≥ 24 for both WT and Mutant 3'UTR inserts, (B) $p<0.05$ and $|log_2FC|>0.3$ in either 3 hour or 6 hour decay comparison, and (C) $|log_2FC$ half-life$|>0.2$.

Ribosome profiling data analysis

STAR (v2.7.3a)[98] was used to align the downloaded fastq files from Lim *et al.*[17] (GEO repository GSE130465) to hg38 and subtraction of mouse sequences was performed using XenofilteR[87]. Aligned reads were counted for gene associations against the UCSC genes database with HTSeq (0.11.0)[99]. Five UW PDX and five normal prostate tissue samples were each sequenced twice. In each analysis, the two replicates for each UW PDX were considered as the test group and five normal prostate tissue samples as the control group. Xtail[96] (v1.1.15) was used to find translationally regulated genes individually for each LuCaP. DESeq2[100] was used to determine changes in RNA expression. 3'UTR mutation-mediated changes in either translation efficiency or RNA expression were calculated by comparing the gene expression values for the tissue sample in which the mutation was found to the average values across the other four non-mutant tissue samples.

GSEA analysis

Enrichr[101] was used to compute overlaps and enrichment of gene sets within the KEGG_2021_Human and Reactome_2019 databases, and gene sets with FDR<0.05 were considered significant.

Survival Analysis

Literature search was conducted on the genes containing 3'UTR mutations that significantly changed translation with a $\log_2$FC>0.75 and or changed stability with a $\log_2$FC>0.6. Mutations in established oncogenic genes that increased mRNA translation or stability and mutations in established tumor suppressive genes that decreased mRNA translation or stability were noted as "oncogenic 3'UTR mutations" (Table S6B-C). UW TAN patients in which these mutations were originally called constituted the test set of patients and comparisons made between this

group and the patients not bearing oncogenic 3'UTR mutations. The survminer R package was used to plot Kaplan-Meier curves and perform default statistical analysis (Table S6D).

Sequence conservation analysis

Sequence conservation scores at each mutation were retrieved using the GenomicScores[102] package in R. The PHASTCONS 100-way vertebrate alignment score was extracted for the location of each mutation of interest from the "phastCons100way.UCSC.hg38" database[103].

RNA structural stability analysis

RNA secondary structure for each 201bp 3'UTR fragment of the MPRA plasmid library was predicted using the ViennaRNA package[73,74]. The minimum free energy algorithm of Zuker & Stiegler 1981, which yields a single optimal structure, was used and the ΔG of these structures was reported.

Filtering CRISPR-able mutations

All mutations that significantly altered either mRNA translation or stability were computationally sorted to determine whether they were valid targets for CRISPR base editing. This was done via custom R script that included filtering out non-transition base changes (retaining C→T, G→A, A→G, and T→C), obtaining the sequence around each mutation, and subsequently determining whether there was a PAM sequence nearby in the proper orientation. This required searching the positive or negative DNA strand depending on the base change, as the mutation must be C→T or A→G on the strand that contains the PAM. Multiple possible PAM sequences were allowed, including NGG, NG, NGAN (VQR), NGCG (VRER), NNGRRT (Sa), and NNNRRT (SaKKH), but NGG and NG sites were prioritized. The distance between the PAM and desired mutation was required to be 13-18 bases, and no other bystander mutations were allowed within this editing window (eg: if the desired mutation was C→T, no other cytosines in the window).

Plots and statistical analysis

All analysis was done either in R (v4.2.0) with plots made using ggplot2[95] or in GraphPad Prism.

Data and Code availability

All data needed to evaluate the conclusions in the paper are described in the paper and/or the methods and are available at the following repositories or databases. The Lim *et al*.[17] publicly available data used in this study are available in the European Genome-Phenome Archive under accession code phs001825.v1.p1. The Quigley *et al.*[3] publicly available data used in this study are available in dbGaP under accession code phs001648.v2.p1. The MPRA sequencing data is deposited under superseries GSE200304. All code used for analysis in this paper can be found under https://zenodo.org/badge/latestdoi/566554304 (DOI:10.5281/zenodo.7331811).

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-ZWILCH	Abcam	202898
Rabbit anti-IGF1R	Cell Signaling	3027
Goat anti-mouse	Invitrogen	31430
Goat anti-rabbit	Invitrogen	31460
Mouse anti-b-Actin	Sigma	A5316
Bacterial and virus strains		
5-alpha Competent E. coli (High Efficiency)	NEB	C2987I
Stellar® Electrocompetent Cells	Takara	636765
Biological samples		
N/A		
Chemicals, peptides, and recombinant proteins		
AMPure XP for PCR Purification	Beckman Coulter	A63880
HyClone Fetal Bovine Serum	Cytiva	SH30396.03
Charcoal:Dextran Stripped Fetal Bovine Serum	Gemini Bio-Products	100-119
Lipofectamine MessengerMAX	Invitrogen	LMRNA003
SUPERase-In RNase Inhibitor	Invitrogen	AM2696
TRIzol Reagent	Invitrogen	15596-018
TURBO DNase	Invitrogen	AM2238
DpnI	NEB	R0176
FseI	NEB	R0588S
T4 DNA Ligase (2,000,000 U/mL)	NEB	M0202T
T4 PNK	NEB	M0201S
Fugene HD	Promega	E2311
Phase Lock Gel Tube- Heavy	QuantaBio	10847-802
Romidepsin (FK228, Depsipeptide)- 1mg	Selleckchem	S3020
Cisplatin	Sigma Aldrich	P4394-25MG
Cycloheximide	Sigma Aldrich	C7698
DTT	Sigma Aldrich	43815
Critical commercial assays		
iScript Reverse Transcription Supermix	Bio-Rad	1708841
SsoAdvanced Universal SYBR Green Supermix	Bio-Rad	172-5272
MAXIscript T7 Transcription Kit	Invitrogen	AM1312
mMESSAGE mMACHINE T7 ULTRA Transcription Kit	Invitrogen	AM1345
PureLink™ HiPure Plasmid Filter Maxiprep Kit	Invitrogen	K210016
SuperScript III First-Strand Synthesis System	Invitrogen	18080-051
NucleoSpin Gel and PCR Clean-up Kit	Macherey-Nagel	740609.5
Gibson Assembly® Master Mix	NEB	E2611
Monarch® DNA Gel Extraction Kit	NEB	T1020
Monarch® PCR & DNA Cleanup Kit	NEB	T1030
Phusion® High-Fidelity DNA Polymerase	NEB	M0530

Q5 High Fidelity 2X Master Mix	NEB	M0492S
Q5 Site-Directed Mutagenesis Kit	NEB	E0554
Q5® High-Fidelity DNA Polymerase	NEB	M0491
Dual Luciferase Reporter Assay System	Promega	E1980
DNeasy Blood & Tissue Kit	Qiagen	69504
QIAprep Spin Miniprep Kit	Qiagen	27104
RNeasy MinElute Cleanup Kit	Qiagen	74204
Phusion High-Fidelity PCR Master Mix with HF Buffer	Thermo Scientific	F531S
Direct-zol RNA Miniprep Plus Kit	Zymo Research	ZR2070
DNA Clean and Concentrator	Zymo Research	11-302C
Deposited data		
MPRA sequencing data (plasmid, polysome, and IVT)	Gene Expression Omnibus	GSE200304
All custom code	Zenodo/Github	DOI:10.5281/zenodo.7331811
Experimental models: Cell lines		
HEK293T cells	ATCC	CRL-3216
PC3 cells	ATCC	CRL-1435
Experimental models: Organisms/strains		
N/A		
Oligonucleotides		
IDT for Illumina Nextera UDI, Set A	Illumina	20027213
IDT for Illumina Nextera UDI, Set C	Illumina	20042666
Individual primers for PCR, etc in Table S7	IDT	N/A
Recombinant DNA		
evoAPOBEC1-BE4max cytosine base editor	Thuronyi *et al*.[104]	Addgene 122611
NG-ABE8e adenine base editor	Richter *et al*.[105]	Addgene 138491
pFYF1548 EMX1 sgRNA plasmid backbone	Fu *et al*.[106]	Addgene 47508
pLuc2CP-ARE	Younis *et al*.[107]	Addgene 62857
pDualLuc	Arvind Subramaniam	N/A
Nanoluciferase exogenous spike-in control	This study	N/A
pIGF1R_sgRNA	This study	N/A
pZWILCH_sgRNA	This study	N/A
pLuc2CP-noARE	This study	N/A
pRL-UBC	This study	N/A
Software and algorithms		
ImageJ	Schneider *et al*.[108]	https://imagej.nih.gov/ij/
GraphPad Prism	GraphPad	www.graphpad.com
IncuCyte Base Analysis Software	Sartorius	N/A
BEDtools	Quinlan & Hall[84]	https://bedtools.readthedocs.io/en/latest/t
Bowtie2 (v2.4.2)	Langmead & Salzberg[85]	https://bowtie-bio.sourceforge.net/bowtie2/index.shtmlt

Picard (v2.25.0)	N/A	http://broadinstitute.github.io/picard
GATK (v4.2.5)	Van der Auwera & O'Connor[86]	https://gatk.broadinstitute.org/hc/en-us
XenofiltR	Kluin *et al*.[87]	https://github.com/NKI-GCF/XenofilteR
Mutect2 (v2)	Benjamin *et al*.[26]	https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2
Strelka (v2.9.2)	Kim *et al*.[27]	https://github.com/Illumina/strelka
MuSE (v2.0)	Fan *et al*.[28]	https://bioinformatics.mdanderson.org/public-software/muse/
VarScan2	Koboldt *et al*.[29]	https://varscan.sourceforge.net/
ANNOVAR	Wang *et al*.[88]	https://annovar.openbioinformatics.org/en/latest/
DeconstructSigs	Rosenthal *et al*.[89]	https://github.com/raerose01/deconstructSigs
GenomicFeatures	Lawrence *et al*.[93]	https://bioconductor.org/packages/GenomicFeatures
BSgenome.Hsapiens.UCSC.hg38	N/A	DOI:10.18129/B9.bioc.BSgenome.Hsapiens.UCSC.hg38
edgeR	Robinson *et al*.[94]	https://bioconductor.org/packages/edgeR
ggplot2	Wickham[95]	https://ggplot2.tidyverse.org/
xtail (v1.1.15)	Xiao *et al*.[96]	https://github.com/xryanglab/xtail
nlstools	Baty *et al*.[97]	https://github.com/aursiber/nlstools
STAR (v2.7.3a)	Dobin *et al*.[98]	https://github.com/alexdobin/STAR
HTSeq (0.11.0)	Anders *et al*.[99]	https://htseq.readthedocs.io/en/master/
DESeq2	Love *et al*.[100]	https://github.com/mikelove/DESeq2
Enrichr	Chen *et al*.[101]	https://maayanlab.cloud/Enrichr/
survminer	N/A	rpkgs.datanovia.com/survminer/
GenomicScores	Puigdevall and Castelo[102]	https://github.com/rcastelo/GenomicScores
ViennaRNA	Lorenz *et al*.[73]	https://www.tbi.univie.ac.at/RNA/
Other		
N/A		

CHAPTER 2 REFERENCES

1. Siegel, R.L., Miller, K.D., Fuchs, H.E., and Jemal, A. (2022). Cancer statistics, 2022. CA. Cancer J. Clin. *72*, 7–33. 10.3322/CAAC.21708.
2. Fraser, M., Sabelnykova, V.Y., Yamaguchi, T.N., Heisler, L.E., Livingstone, J., Huang, V., Shiah, Y.-J., Yousif, F., Lin, X., Masella, A.P., et al. (2017). Genomic hallmarks of localized, non-indolent prostate cancer. Nature *541*, 359–364. 10.1038/nature20788.
3. Quigley, D.A., Dang, H.X., Zhao, S.G., Maher, C.A., Small, E.J., and Feng, F.Y. (2018). Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. Cell *174*, 758–769. 10.1016/j.cell.2018.06.039.
4. Abeshouse, A., Ahn, J., Akbani, R., Ally, A., Amin, S., Andry, C.D., Annala, M., Aprikian, A., Armenia, J., Arora, A., et al. (2015). The Molecular Taxonomy of Primary Prostate Cancer. Cell *163*, 1011–1025. 10.1016/j.cell.2015.10.025.
5. Robinson, D., Van Allen, E.M., Wu, Y.M., Schultz, N., Lonigro, R.J., Mosquera, J.M., Montgomery, B., Taplin, M.E., Pritchard, C.C., Attard, G., et al. (2015). Integrative clinical genomics of advanced prostate cancer. Cell *161*, 1215–1228. 10.1016/j.cell.2015.05.001.
6. Abida, W., Cyrta, J., Heller, G., Prandi, D., Armenia, J., Coleman, I., Cieslik, M., Benelli, M., Robinson, D., Van Allen, E.M., et al. (2019). Genomic correlates of clinical outcome in advanced prostate cancer. Proc. Natl. Acad. Sci. U. S. A. *166*, 11428–11436. 10.1073/pnas.1902651116.
7. Taylor, B.S., Lash, A.E., Reva, B., Scher, H.I., Wilson, M., Dolgalev, I., Socci, N.D., Schultz, N., Kaushik, P., Antipin, Y., et al. (2010). Integrative Genomic Profiling of Human Prostate Cancer. Cancer Cell *18*, 11–22. 10.1016/j.ccr.2010.05.026.
8. Buccitelli, C., and Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene expression control. Nat. Rev. Genet. 2020 2110 *21*, 630–644. 10.1038/s41576-020-0258-4.
9. Schuster, S.L., and Hsieh, A.C. (2019). The Untranslated Regions of mRNAs in Cancer. Trends in Cancer *5*, 245–262. 10.1016/j.trecan.2019.02.011.
10. Zhang, Y., Wang, X.W., Jelovac, D., Nakanishi, T., Yu, M.H., Akinmade, D., Goloubeva, O., Ross, D.D., Brodie, A., and Hamburger, A.W. (2005). The ErbB3-binding protein Ebp1 suppresses androgen receptor-mediated gene transcription and tumorigenesis of prostate cancer cells. Proc. Natl. Acad. Sci. U. S. A. *102*, 9890–9895. 10.1073/PNAS.0503829102.
11. Zhang, Y., Linn, D., Liu, Z., Melamed, J., Tavora, F., Young, C.Y., Burger, A.M., and Hamburger, A.W. (2008). EBP1, an ErbB3-binding protein, is decreased in prostate cancer and implicated in hormone resistance. Mol. Cancer Ther. *7*, 3176–3186. 10.1158/1535-7163.MCT-08-0526.
12. Zhou, H., Mazan-Mamczarz, K., Martindale, J.L., Barker, A., Liu, Z., Gorospe, M., Leedman, P.J., Gartenhaus, R.B., Hamburger, A.W., and Zhang, Y. (2010). Post-transcriptional regulation of androgen receptor mRNA by an ErbB3 binding protein 1 in prostate cancer. Nucleic Acids Res. *38*, 3619–3631. 10.1093/NAR/GKQ084.
13. Ribas, J., Ni, X., Haffner, M., Wentzel, E.A., Salmasi, A.H., Chowdhury, W.H., Kudrolli, T.A., Yegnasubramanian, S., Luo, J., Rodriguez, R., et al. (2009). miR-21: An androgen receptor-regulated microRNA that promotes hormone-dependent and hormone-independent prostate cancer growth. Cancer Res. *69*, 7165–7169. 10.1158/0008-5472.CAN-09-1448/655379/P/MIR-21-AN-ANDROGEN-RECEPTOR-REGULATED-MICRORNA.
14. Kanagasabai, T., Li, G., Shen, T.H., Gladoun, N., Castillo-Martin, M., Celada, S.I., Xie, Y., Brown, L.K., Mark, Z.A., Ochieng, J., et al. (2022). MicroRNA-21 deficiency suppresses prostate cancer progression through downregulation of the IRS1-SREBP-1 signaling pathway. Cancer Lett. *525*, 46–54. 10.1016/J.CANLET.2021.09.041.
15. Ziqiang, Y., Joongho, S., Wilson, A., Goel, S., Ling, Y.H., Ahmed, N., Dopeso, H., Jhawer, M., Nasser, S., Montagna, C., et al. (2009). An A13 repeat within the 3′-untranslated region of epidermal growth factor receptor (EGFR) is frequently mutated in microsatellite instability colon

cancers and is associated with increased EGFR expression. Cancer Res. *69*, 7811–7818. 10.1158/0008-5472.CAN-09-0986/655287/P/AN-A13-REPEAT-WITHIN-THE-3-UNTRANSLATED-REGION-OF.

16. Fan, Z., Yang, J., Zhang, D., Zhang, X., Ma, X., Kang, L., Liu, Y., Yan, X., Ji, Q., Wang, J., et al. (2018). The risk variant rs884225 within EGFR impairs miR-103a-3p's anti-tumourigenic function in non-small cell lung cancer. Oncogene *38*, 2291–2304. 10.1038/s41388-018-0576-6.
17. Lim, Y., Arora, S., Schuster, S.L., Corey, L., Fitzgibbon, M., Wladyka, C.L., Wu, X., Coleman, I.M., Delrow, J.J., Corey, E., et al. (2021). Multiplexed functional genomic analysis of 5' untranslated region mutations across the spectrum of prostate cancer. Nat. Commun. *12*. 10.1038/S41467-021-24445-6.
18. Wissink, E.M., Fogarty, E.A., and Grimson, A. (2016). High-throughput discovery of post-transcriptional cis-regulatory elements. BMC Genomics *17*, 177. 10.1186/s12864-016-2479-7.
19. Cottrell, K.A., Chaudhari, H.G., Cohen, B.A., and Djuranovic, S. (2018). PTRE-seq reveals mechanism and interactions of RNA binding proteins and miRNAs. Nat. Commun. *9*, 301. 10.1038/s41467-017-02745-0.
20. Oikonomou, P., Goodarzi, H., and Tavazoie, S. (2014). Systematic Identification of Regulatory Elements in Conserved 3′ UTRs of Human Transcripts. Cell Rep. *7*, 281–292. 10.1016/J.CELREP.2014.03.001.
21. Zhao, W., Pollack, J.L., Blagev, D.P., Zaitlen, N., McManus, M.T., and Erle, D.J. (2014). Massively parallel functional annotation of 3′ untranslated regions. Nat. Biotechnol. *32*, 387–391. 10.1038/nbt.2851.
22. Litterman, A.J., Kageyama, R., Tonqueze, O. Le, Zhao, W., Gagnon, J.D., Goodarzi, H., Erle, D.J., and Ansel, K.M. (2019). A massively parallel 3′ UTR reporter assay reveals relationships between nucleotide content, sequence conservation, and mRNA destabilization. Genome Res. *29*, 896–906. 10.1101/GR.242552.118.
23. Rabani, M., Pieper, L., Chew, G.-L., and Schier, A.F. (2017). A Massively Parallel Reporter Assay of 3′ UTR Sequences Identifies In Vivo Rules for mRNA Degradation. Mol. Cell *68*, 1083-1094.e5. 10.1016/J.MOLCEL.2017.11.014.
24. Roudier, M.P., True, L.D., Higano, C.S., Vesselle, H., Ellis, W., Lange, P., and Vessella, R.L. (2003). Phenotypic heterogeneity of end-stage prostate carcinoma metastatic to bone. Hum. Pathol. *34*, 646–653. 10.1016/S0046-8177(03)00190-4.
25. Labrecque, M.P., Coleman, I.M., Brown, L.G., True, L.D., Kollath, L., Lakely, B., Nguyen, H.M., Yang, Y.C., Da Costa, R.M.G., Kaipainen, A., et al. (2019). Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. J. Clin. Invest. *129*, 4492–4505. 10.1172/JCI128212.
26. Benjamin, D., Sato, T., Cibulskis, K., Getz, G., Stewart, C., and Lichtenstein, L. (2019). Calling Somatic SNVs and Indels with Mutect2. bioRxiv, 861054. 10.1101/861054.
27. Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Källberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., et al. (2018). Strelka2: fast and accurate calling of germline and somatic variants. Nat. Methods 2018 158 *15*, 591–594. 10.1038/s41592-018-0051-x.
28. Fan, Y., Xi, L., Hughes, D.S.T., Zhang, J., Zhang, J., Futreal, P.A., Wheeler, D.A., and Wang, W. (2016). MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. Genome Biol. *17*, 178. 10.1186/S13059-016-1029-6/FIGURES/3.
29. Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568–576. 10.1101/GR.129684.111.
30. Ryan, M.J., and Bose, R. (2019). Genomic Alteration Burden in Advanced Prostate Cancer and Therapeutic Implications. Front. Oncol. *9*, 1287. 10.3389/FONC.2019.01287.

31. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. *47*, D941–D947. 10.1093/NAR/GKY1015.
32. Guo, X., Song, C., Fang, L., Li, M., Yue, L., and Sun, Q. (2020). FLRT2 functions as Tumor Suppressor gene inactivated by promoter methylation in Colorectal Cancer. J. Cancer *11*, 7329–7338. 10.7150/JCA.47558.
33. Song, C.Q., Li, Y., Mou, H., Moore, J., Park, A., Pomyen, Y., Hough, S., Kennedy, Z., Fischer, A., Yin, H., et al. (2017). Genome-Wide CRISPR Screen Identifies Regulators of Mitogen-Activated Protein Kinase as Suppressors of Liver Tumors in Mice. Gastroenterology *152*, 1161-1173.e1. 10.1053/J.GASTRO.2016.12.002.
34. Bae, H., Kim, B., Lee, H., Lee, S., Kang, H.S., and Kim, S.J. (2017). Epigenetically regulated Fibronectin leucine rich transmembrane protein 2 (FLRT2) shows tumor suppressor activity in breast cancer cells. Sci. Rep. *7*. 10.1038/S41598-017-00424-0.
35. Leung, C.S., Yeung, T.L., Yip, K.P., Wong, K.K., Ho, S.Y., Mangala, L.S., Sood, A.K., Lopez-Berestein, G., Sheng, J., Wong, S.T.C., et al. (2018). Cancer-associated fibroblasts regulate endothelial adhesion protein LPP to promote ovarian cancer chemoresistance. J. Clin. Invest. *128*, 589–606. 10.1172/JCI95200.
36. Ngan, E., Stoletov, K., Smith, H.W., Common, J., Muller, W.J., Lewis, J.D., and Siegel, P.M. (2017). LPP is a Src substrate required for invadopodia formation and efficient breast cancer lung metastasis. Nat. Commun. *8*. 10.1038/NCOMMS15059.
37. Kuriyama, S., Yoshida, M., Yano, S., Aiba, N., Kohno, T., Minamiya, Y., Goto, A., and Tanaka, M. (2016). LPP inhibits collective cell migration during lung cancer dissemination. Oncogene *35*, 952–964. 10.1038/ONC.2015.155.
38. Wang, H., Wu, J., Ling, R., Li, F., Yang, Q., He, J., Lei, X., Wu, C., Zhang, G., Zheng, B., et al. (2022). Fibroblast-derived LPP as a biomarker for treatment response and therapeutic target in gastric cancer. Mol. Ther. oncolytics *24*, 547–560. 10.1016/J.OMTO.2022.01.008.
39. Grimm, D., Bauer, J., Wise, P., Krüger, M., Simonsen, U., Wehland, M., Infanger, M., and Corydon, T.J. (2020). The role of SOX family members in solid tumours and metastasis. Semin. Cancer Biol. *67*, 122–153. 10.1016/J.SEMCANCER.2019.03.004.
40. Hu, J., Tian, J., Zhu, S., Sun, L., Yu, J., Tian, H., Dong, Q., Luo, Q., Jiang, N., Niu, Y., et al. (2018). Sox5 contributes to prostate cancer metastasis and is a master regulator of TGF-β-induced epithelial mesenchymal transition through controlling Twist1 expression. Br. J. Cancer *118*, 88–97. 10.1038/BJC.2017.372.
41. Del Giudice, M., Foster, J.G., Peirone, S., Rissone, A., Caizzi, L., Gaudino, F., Parlato, C., Anselmi, F., Arkell, R., Guarrera, S., et al. (2022). FOXA1 regulates alternative splicing in prostate cancer. Cell Rep. *40*, 111404. 10.1016/J.CELREP.2022.111404.
42. Cai, H., Agersnap, S.N., Sjøgren, A., Simonsen, M.K., Blaavand, M.S., Jensen, U. V., and Thomsen, M.K. (2022). In Vivo Application of CRISPR/Cas9 Revealed Implication of Foxa1 and Foxp1 in Prostate Cancer Proliferation and Epithelial Plasticity. Cancers (Basel). *14*. 10.3390/CANCERS14184381.
43. Sun, Y., Zhang, Y., Hamilton, K., Manley, J.L., Shi, Y., Walz, T., and Tong, L. (2018). Molecular basis for the recognition of the human AAUAAA polyadenylation signal. Proc. Natl. Acad. Sci. U. S. A. *115*, E1419–E1428. 10.1073/pnas.1718723115.
44. Zhang, C., Chen, Y., Sun, B., Wang, L., Yang, Y., Ma, D., Lv, J., Heng, J., Ding, Y., Xue, Y., et al. (2017). M6A modulates haematopoietic stem and progenitor cell specification. Nature *549*, 273–276. 10.1038/nature23883.
45. Zhao, J., Zhang, Y., Liu, X. sheng, Zhu, F. ming, Xie, F., Jiang, C. yi, Zhang, Z. ye, Gao, Y. li, Wang, Y. chuan, Li, B., et al. (2020). RNA-binding protein Musashi2 stabilizing androgen receptor drives prostate cancer progression. Cancer Sci. *111*, 369–382. 10.1111/CAS.14280.

46. Minuesa, G., Albanese, S.K., Xie, W., Kazansky, Y., Worroll, D., Chow, A., Schurer, A., Park, S.M., Rotsides, C.Z., Taggart, J., et al. (2019). Small-molecule targeting of MUSASHI RNA-binding activity in acute myeloid leukemia. Nat. Commun. *10*. 10.1038/S41467-019-10523-3.
47. Palacios, F., Yan, X.J., Ferrer, G., Chen, S.S., Vergani, S., Yang, X., Gardner, J., Barrientos, J.C., Rock, P., Burack, R., et al. (2021). Musashi 2 influences chronic lymphocytic leukemia cell survival and growth making it a potential therapeutic target. Leukemia *35*, 1037–1052. 10.1038/S41375-020-01115-Y.
48. Karantanos, T., Tanimoto, R., Edamura, K., Hirayama, T., Yang, G., Golstov, A.A., Wang, J., Kurosaka, S., Park, S., and Thompson, T.C. (2014). Systemic GLIPR1-ΔTM protein as a novel therapeutic approach for prostate cancer. Int. J. cancer *134*, 2003–2013. 10.1002/IJC.28529.
49. Sheng, X., Bowen, N., and Wang, Z. (2016). GLI pathogenesis-related 1 functions as a tumor-suppressor in lung cancer. Mol. Cancer *15*. 10.1186/S12943-016-0508-4.
50. Li, K., Ren, C., Yang, G., Fattah, E.A., Goltsov, A.A., Kim, S.M., Lee, J.S., Park, S., Demayo, F.J., Ittmann, M.M., et al. (2011). GLIPR1 suppresses prostate cancer development through targeted oncoprotein destruction. Cancer Res. *71*, 7694–7704. 10.1158/0008-5472.CAN-11-1714.
51. Liu, J.W., Nagpal, J.K., Sun, W., Lee, J., Kim, M.S., Ostrow, K.L., Zhou, S., Jeronimo, C., Henrique, R., Van Criekinge, W., et al. (2008). ssDNA-binding protein 2 is frequently hypermethylated and suppresses cell growth in human prostate cancer. Clin. Cancer Res. *14*, 3754–3760. 10.1158/1078-0432.CCR-07-4763.
52. Liang, H., Samanta, S., and Nagarajan, L. (2005). SSBP2, a candidate tumor suppressor gene, induces growth arrest and differentiation of myeloid leukemia cells. Oncogene *24*, 2625–2634. 10.1038/SJ.ONC.1208167.
53. Wang, Y., Klumpp, S., Amin, H.M., Liang, H., Li, J., Estrov, Z., Zweidler-Mckay, P., Brandt, S.J., Agulnick, A., and Nagarajan, L. (2010). SSBP2 is an in vivo tumor suppressor and regulator of LDB1 stability. Oncogene *29*, 3044–3053. 10.1038/ONC.2010.78.
54. Davies, A.H., Beltran, H., and Zoubeidi, A. (2018). Cellular plasticity and the neuroendocrine phenotype in prostate cancer. Nat. Rev. Urol. *15*, 271–286. 10.1038/NRUROL.2018.22.
55. Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P.A., and Burge, C.B. (2014). RNA Bind-n-Seq: Quantitative Assessment of the Sequence and Structural Binding Specificity of RNA Binding Proteins. Mol. Cell *54*, 887–900. 10.1016/j.molcel.2014.04.016.
56. Barreau, C., Paillard, L., and Osborne, H.B. (2006). AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res. *33*, 7138–7150. 10.1093/NAR/GKI1012.
57. Nouruzi, S., Ganguli, D., Tabrizian, N., Kobelev, M., Sivak, O., Namekawa, T., Thaper, D., Baca, S.C., Freedman, M.L., Aguda, A., et al. (2022). ASCL1 activates neuronal stem cell-like lineage programming through remodeling of the chromatin landscape in prostate cancer. Nat. Commun. *13*. 10.1038/S41467-022-29963-5.
58. Kim, M.Y., Na, I., Kim, J.S., Son, S.H., Choi, S., Lee, S.E., Won, H.S., Jang, K., Alterovitz, G., Chen, Y., et al. (2019). Rational discovery of antimetastatic agents targeting the intrinsically disordered region of MBD2. Sci. Adv. *5*. 10.1126/SCIADV.AAV9810/SUPPL_FILE/AAV9810_SM.PDF.
59. Zhou, K., Zhou, M., Cheng, L., Chen, X., Wang, X., Chu, Y., Yu, Q., Zhang, S., Wang, N., Zhao, L., et al. (2021). Loss of MBD2 attenuates MLL-AF9-driven leukemogenesis by suppressing the leukemic cell cycle via CDKN1C. Oncogenesis *10*. 10.1038/S41389-021-00366-3.
60. Deng, M., Wang, N., Li, Z., Chen, R., Duan, J., Peng, Y., Wu, Z., Zhang, Z., Jiang, L., Zheng, X., et al. (2022). FXR1 can bind with the CFIm25/CFIm68 complex and promote the progression of urothelial carcinoma of the bladder by stabilizing TRAF1 mRNA. Cell Death Dis. *13*. 10.1038/S41419-022-04614-1.
61. George, J., Li, Y., Kadamberi, I.P., Parashar, D., Tsaih, S.W., Gupta, P., Geethadevi, A., Chen, C., Ghosh, C., Sun, Y., et al. (2021). RNA-binding protein FXR1 drives cMYC translation by recruiting

eIF4F complex to the translation start site. Cell Rep. *37*. 10.1016/J.CELREP.2021.109934.
62. Rees, H.A., and Liu, D.R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. Nat. Rev. Genet. *19*, 770–788. 10.1038/s41576-018-0059-1.
63. Karess, R. (2005). Rod-Zw10-Zwilch: A key player in the spindle checkpoint. Trends Cell Biol. *15*, 386–392. 10.1016/j.tcb.2005.05.003.
64. Zhang, L., Luo, Y., Cheng, T., Chen, J., Yang, H., Wen, X., Jiang, Z., Li, H., and Pan, C. (2021). Development and Validation of a Prognostic N6-Methyladenosine-Related Immune Gene Signature for Lung Adenocarcinoma. Pharmgenomics. Pers. Med. *14*, 1549–1563. 10.2147/PGPM.S332683.
65. Tong, H., Liu, X., Peng, C., Shen, B., and Zhu, Z. (2022). Silencing of KNTC1 inhibits hepatocellular carcinoma cells progression via suppressing PI3K/Akt pathway. Cell. Signal., 110498. 10.1016/J.CELLSIG.2022.110498.
66. Mou, K., Zhang, J., Mu, X., Wang, L., Liu, W., and Ge, R. (2021). Zwint facilitates melanoma progression by promoting c-Myc expression. Exp. Ther. Med. *22*. 10.3892/ETM.2021.10250.
67. Klinakis, A., Szabolcs, M., Chen, G., Xuan, S., Hibshoosh, H., and Efstratiadis, A. (2009). Igf1r as a therapeutic target in a mouse model of basal-like breast cancer. Proc. Natl. Acad. Sci. U. S. A. *106*, 2359–2364. 10.1073/PNAS.0810221106/SUPPL_FILE/0810221106SI.PDF.
68. Yaktapour, N., Übelhart, R., Schüler, J., Aumann, K., Dierks, C., Burger, M., Pfeifer, D., Jumaa, H., Veelken, H., Brummer, T., et al. (2013). Insulin-like growth factor-1 receptor (IGF1R) as a novel target in chronic lymphocytic leukemia. Blood *122*, 1621–1633. 10.1182/BLOOD-2013-02-484386.
69. Wu, J., and Yu, E. (2014). Insulin-like growth factor receptor-1 (IGF-IR) as a target for prostate cancer therapy. Cancer Metastasis Rev. *33*, 607–617. 10.1007/S10555-013-9482-0.
70. Bitelman, C., Sarfstein, R., Sarig, M., Attias-Geva, Z., Fishman, A., Werner, H., and Bruchim, I. (2013). IGF1R-directed targeted therapy enhances the cytotoxic effect of chemotherapy in endometrial cancer. Cancer Lett. *335*, 153–159. 10.1016/J.CANLET.2013.02.009.
71. Tu, C., Fiandalo, M. V., Pop, E., Stocking, J.J., Azabdaftari, G., Li, J., Wei, H., Ma, D., Qu, J., Mohler, J.L., et al. (2018). Proteomic Analysis of Charcoal-Stripped Fetal Bovine Serum Reveals Changes in the Insulin-like Growth Factor Signaling Pathway. J. Proteome Res. *17*, 2963–2977. 10.1021/ACS.JPROTEOME.8B00135.
72. Moore, A.E., Chenette, D.M., Larkin, L.C., and Schneider, R.J. (2014). Physiological networks and disease functions of RNA-binding protein AUF1. Wiley Interdiscip. Rev. RNA *5*, 549–564. 10.1002/WRNA.1230.
73. Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*. 10.1186/1748-7188-6-26.
74. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. Nucleic Acids Res. *36*. 10.1093/NAR/GKN188.
75. Lin, S.-L.;, Lin, C.-Y.;, Lee, W.;, Teng, C.-F.;, Shyu, W.-C.;, Jeng, L.-B., Lin, S.-L., Lin, C.-Y., Lee, W., Teng, C.-F., et al. (2022). Mini Review: Molecular Interpretation of the IGF/IGF-1R Axis in Cancer Treatment and Stem Cells-Based Therapy in Regenerative Medicine. Int. J. Mol. Sci. 2022, Vol. 23, Page 11781 *23*, 11781. 10.3390/IJMS231911781.
76. Griesemer, D., Xue, J.R., Reilly, S.K., Ulirsch, J.C., Kukreja, K., Davis, J.R., Kanai, M., Yang, D.K., Butts, J.C., Guney, M.H., et al. (2021). Genome-wide functional screen of 3′UTR variants uncovers causal variants for human disease and evolution. Cell, 1–14. 10.1016/j.cell.2021.08.025.
77. Litterman, A.J., Kageyama, R., Le Tonqueze, O., Zhao, W., Gagnon, J.D., Goodarzi, H., Erle, D.J., and Ansel, K.M. (2019). A massively parallel 3′ UTR reporter assay reveals relationships between nucleotide content, sequence conservation, and mRNA destabilization. Genome Res. *29*, 896–906. 10.1101/gr.242552.118.
78. Anzalone, A. V., Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. Nat. Biotechnol. *38*, 824–844. 10.1038/S41587-

020-0561-9.
79. Kweon, J., Jang, A.H., Shin, H.R., See, J.E., Lee, W., Lee, J.W., Chang, S., Kim, K., and Kim, Y. (2019). A CRISPR-based base-editing screen for the functional assessment of BRCA1 variants. Oncogene 2019 391 *39*, 30–35. 10.1038/s41388-019-0968-2.
80. Cuella-Martin, R., Hayward, S.B., Fan, X., Chen, X., Huang, J.W., Taglialatela, A., Leuzzi, G., Zhao, J., Rabadan, R., Lu, C., et al. (2021). Functional interrogation of DNA damage response variants with base editing screens. Cell *184*, 1081-1097.e19. 10.1016/J.CELL.2021.01.041.
81. Xu, P., Liu, Z., Liu, Y., Ma, H., Xu, Y., Bao, Y., Zhu, S., Cao, Z., Wu, Z., Zhou, Z., et al. (2021). Genome-wide interrogation of gene functions through base editor screens empowered by barcoded sgRNAs. Nat. Biotechnol. 2021 3911 *39*, 1403–1413. 10.1038/s41587-021-00944-1.
82. Hanna, R.E., Hegde, M., Fagre, C.R., DeWeirdt, P.C., Sangree, A.K., Szegletes, Z., Griffith, A., Feeley, M.N., Sanson, K.R., Baidi, Y., et al. (2021). Massively parallel assessment of human variants with base editor screens. Cell *184*, 1064-1080.e20. 10.1016/J.CELL.2021.01.012.
83. Shin, H.R., See, J.E., Kweon, J., Kim, H.S., Sung, G.J., Park, S., Jang, A.H., Jang, G., Choi, K.C., Kim, I., et al. (2021). Small-molecule inhibitors of histone deacetylase improve CRISPR-based adenine base editing. Nucleic Acids Res. *49*, 2390–2399. 10.1093/NAR/GKAB052.
84. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. 10.1093/BIOINFORMATICS/BTQ033.
85. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 2012 94 *9*, 357–359. 10.1038/nmeth.1923.
86. Van der Auwera, G.A., and O'Connor, B.D. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition) (O'Reilly Media, Inc.).
87. Kluin, R.J.C., Kemper, K., Kuilman, T., de Ruiter, J.R., Iyer, V., Forment, J. V., Cornelissen-Steijger, P., de Rink, I., ter Brugge, P., Song, J.Y., et al. (2018). XenofilteR: Computational deconvolution of mouse and human reads in tumor xenograft sequence data. BMC Bioinformatics *19*, 1–15. 10.1186/S12859-018-2353-5/FIGURES/6.
88. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. 10.1093/NAR/GKQ603.
89. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S., and Swanton, C. (2016). deconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. *17*, 1–11. 10.1186/S13059-016-0893-4/FIGURES/5.
90. Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature *499*, 172–177. 10.1038/nature12311.
91. Feng, H., Bao, S., Rahman, M.A., Weyn-Vanhentenryck, S.M., Khan, A., Wong, J., Shah, A., Flynn, E.D., Krainer, A.R., and Zhang, C. (2019). Modeling RNA-Binding Protein Specificity In Vivo by Precisely Registering Protein-RNA Crosslink Sites. Mol. Cell *74*, 1189-1204.e6. 10.1016/j.molcel.2019.02.002.
92. Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Res. *47*, D155–D162. 10.1093/NAR/GKY1141.
93. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for Computing and Annotating Genomic Ranges. PLOS Comput. Biol. *9*, e1003118. 10.1371/JOURNAL.PCBI.1003118.
94. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. 10.1093/BIOINFORMATICS/BTP616.
95. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York) 10.1007/978-3-319-24277-4.

96. Xiao, Z., Zou, Q., Liu, Y., and Yang, X. (2016). Genome-wide assessment of differential translations with ribosome profiling data. Nat. Commun. 2016 71 *7*, 1–11. 10.1038/ncomms11194.
97. Baty, F., Ritz, C., Charles, S., Brutsche, M., Flandrois, J.P., and Delignette-Muller, M.L. (2015). A Toolbox for Nonlinear Regression in R: The Package nlstools. J. Stat. Softw. *66*, 1–21. 10.18637/JSS.V066.I05.
98. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15. 10.1093/BIOINFORMATICS/BTS635.
99. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166. 10.1093/BIOINFORMATICS/BTU638.
100. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 1–21. 10.1186/S13059-014-0550-8/FIGURES/9.
101. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 1–14. 10.1186/1471-2105-14-128/FIGURES/3.
102. Puigdevall, P., and Castelo, R. (2018). GenomicScores: seamless access to genomewide position-specific scores from R and Bioconductor. Bioinformatics *34*, 3208–3210. 10.1093/BIOINFORMATICS/BTY311.
103. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.D.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. *15*, 1034–1050. 10.1101/GR.3715005.
104. Thuronyi, B.W., Koblan, L.W., Levy, J.M., Yeh, W.H., Zheng, C., Newby, G.A., Wilson, C., Bhaumik, M., Shubina-Oleinik, O., Holt, J.R., et al. (2019). Continuous evolution of base editors with expanded target compatibility and improved activity. Nat. Biotechnol. *37*, 1070–1079. 10.1038/s41587-019-0193-0.
105. Richter, M.F., Zhao, K.T., Eton, E., Lapinaite, A., Newby, G.A., Thuronyi, B.W., Wilson, C., Koblan, L.W., Zeng, J., Bauer, D.E., et al. (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nat. Biotechnol., 1–9. 10.1038/s41587-020-0453-z.
106. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. *31*, 822–826. 10.1038/NBT.2623.
107. Younis, I., Berg, M., Kaida, D., Dittmar, K., Wang, C., and Dreyfuss, G. (2010). Rapid-response splicing reporter screens identify differential regulators of constitutive and alternative splicing. Mol. Cell. Biol. *30*, 1718–1728. 10.1128/MCB.01301-09.
108. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 2012 97 *9*, 671–675. 10.1038/nmeth.2089.

CHAPTER 3

Building functional genomics tools able to measure mRNA stability from patient-based mutations

This chapter presents unpublished work.

ABSTRACT

In this chapter, I discuss how I optimized and built the massively parallel reporter assay (MPRA) I used to measure mRNA stability in the previous chapter. There are a variety of methods to determine differences in mRNA decay dynamics in cells; however, established protocols do not exist for the performance of an mRNA-stability MPRA in human cells. To ensure that we used the optimal methodology for this MPRA, I adapted two widely-used protocols, based on either *in vitro* transcription (IVT) of an mRNA library or 4-thiouracil (4sU) labeling of nascent RNA and compared them head-to-head. Through extensive optimization of each method to perform reliably in an MPRA format, I conclude that use of an *in vitro* transcribed mRNA library is more reproducible and robust than 4sU-based methods. Therefore, the IVT-based MPRA was used for further analysis as detailed in Chapter 2.

INTRODUCTION

The gene expression cascade that ultimately leads to functional protein consists of many parts, including transcription, messenger RNA (mRNA) stability, translation, and protein stability. While transcription is the most widely studied aspect of gene expression, mRNA stability also plays a significant role in gene regulation. This role is emphasized by the tight regulation of mRNA stability, which has been illustrated by several important observations. In both yeast and humans cells, it has been found that decay rates are coordinated across functional classes of mRNAs. In yeast, mRNAs encoding the different components of protein complexes are known to have similar half-lives[1]. In humans, transcription factor mRNAs, whose expression can vary to affect cell state, were found overall to have relatively short half-lives, while housekeeping genes had much longer half-lives[2]. This coordination of mRNA half-lives therefore seems to be an important, intrinsic part of gene regulation that allows certain genes to be regulated as a group in response to stimuli. A striking example of this is in the inflammatory response, where groups of early, mid, and late-expressed genes display significant differences in mRNA turnover, with early response mRNAs decaying more quickly, and those expressed later being more stable[3]. This observation suggests that high mRNA expression together with rapid turnover may be an important mechanism for genes whose expression must be “poised” to respond quickly to stimuli.

mRNA stability is regulated through a myriad of processes. Some of the earliest papers on the subject noted that particular sequence motifs like AU-rich elements (AREs) in the 3’ untranslated regions (3’UTRs) of mRNAs were associated with shorter half-lives[4]. In fact, 3’UTRs are major hubs of mRNA decay regulatory elements such as miRNA binding sequences and RNA-binding protein (RBP) motifs[5–7]. More recently, the influence of RNA modification (epitranscriptomics), alternative polyadenylation, and translation elongation on mRNA stability has been uncovered[8]. Multiple epitranscriptomic marks, including N^6-methyladenosine (m^6A), 8-

oxo-7,8-dihydroguanosine (8-oxoG), and 5-methylcytidine (m^5C) can be deposited onto mRNAs, where their presence can be read by specific proteins that increase or decrease the mRNA's stability[9]. Alternative polyadenylation, by changing the length of the 3'UTR and thereby the regulatory elements contained within the mRNA, can also have a profound effect on mRNA stability. In fact, the coordinated 3'UTR shortening of oncogenic mRNAs in cancer has been found to increase the stability and protein expression of oncogenes as an important mechanism of disease[10]. Interestingly, in addition to being regulated by mRNA sequence and modifications, mRNA stability is also tightly coupled to translation elongation[9,11,12]. Ribosome transit rates and pausing can be sensed by the CCR4/NOT complex, leading to deadenylation, de-capping, and degradation of mRNAs experiencing slow elongation from nonoptimal codon usage or pause sites[13,14]. In total, mRNA stability is a critical aspect of gene expression and cellular response to stimuli, which is encoded in the gene sequence through UTR-based motifs and codon usage, as well as RNA modifications such as m^6A methylation and alternative polyadenylation.

Though much is known about mRNA stability, it remains relatively less characterized than steady-state mRNA expression, in part due to the difficulty of studying such a dynamic process. Many techniques exist to study mRNA degradation, including transcriptional manipulation via actinomycin D or Tet-responsive promoters, mRNA labeling using 4-thiouridine (4sU), 5-bromouridine (5-BrU), or E-ethynyluridine (5-EU)[15], and transfection of *in vitro* transcribed mRNAs. Each of these methods have pros and cons, which can be magnified when used on a whole-transcriptome scale. The simplest and most-commonly used method of measuring mRNA stability is shutting off transcription via treatment with actinomycin D, which ensures that subsequent measurements of mRNA levels reflect changes in mRNA decay and not production. While this method is simple, popular, and can be easily used to study endogenous mRNA dynamics genome-wide, there are concerns over the indirect effects of complete transcriptional shutoff on mRNA dynamics[16]. In fact, actinomycin D treatment has

been shown to alter mRNA stability of particular transcripts, changing the degradation of some mRNAs more than others, creating artifacts in the data from these experiments[17–19]. One way to overcome this is to instead measure mRNA stability of a gene of interest synthetically cloned under a tetracycline-regulated promoter[20,21]. These assays can either be conducted using the Tet-ON promoter, where a short pulse of active transcription creates a distinct population of mRNAs to observe over time, or the Tet-OFF system, where transcriptional shutoff is used similarly to actinomycin D, but with specificity to the gene of interest. Either way, measurement of mRNA decay is possible without perturbing the entire cellular environment. Unfortunately, these systems can be leaky, with inefficient transcriptional shutoff leading to artifacts in calculations of mRNA half-lives[22]. They also require the use of plasmids to express exogenous genes of interest under non-endogenous promoters, rather than measuring dynamics of endogenous mRNAs.

Labeling mRNA with various modified nucleotides (4sU, 5-BrU, or 5-EU) during normal transcription can allow for the capture of newly synthesized endogenous mRNAs and subsequent determination of mRNA half-life[23]. Such methods can either use a single pulse of modified nucleotide, wherein the ratio of labeled to unlabeled mRNA after a certain time period can be used to calculate the mRNA half-life, or a pulse-chase method. In the pulse-chase method, a population of mRNAs is labeled over a time period, then the system is chased with unlabeled nucleotide such that the labeled mRNA population can be measured over time. The strength of these labeling methods lies in being able to measure global changes in endogenous mRNA dynamics, without causing toxicity and downstream indirect effects on gene expression. However, these protocols are fairly technically complex, as the treatment with labeled nucleotide, chase treatment, and enrichment of the labeled nascent mRNA all must be rigorously optimized[24,25].

The most straightforward method of measuring mRNA decay is to transcribe, cap, and polyadenylate a population of mRNA *in vitro*, then introduce this mRNA into cells and measure its degradation over time[26,27]. As the introduced mRNA is synthetic, it can be easily distinguished from endogenous mRNAs without labeling, and transcriptional dynamics are not a confounding factor. However, using synthetic mRNA raises different issues, as these experiments inherently will not capture any contribution to mRNA dynamics by co-transcriptional processes, such as epitranscriptomics, alternative polyadenylation, or splicing. Therefore, this method is best used to study aspects of mRNA stability that are expected to behave independently from these co-transcriptional processes.

While several methods to assay mRNA stability exist, it can be difficult to determine which protocol is best suited for a given scenario or question. In this chapter, I describe and compare the use of 4-thiouracil (4sU) labeling versus an *in vitro* transcribed mRNA library for high-throughput measurement of mRNA stability changes in a massively parallel reporter assay (MPRA). Massively parallel reporter assays are useful tools to simultaneously measure how thousands of different sequences affect gene expression using plasmid reporter libraries. My goal was to adapt both a 4sU- and an IVT-based method to the MPRA platform to determine how the 3'UTR, and particularly single-nucleotide mutations within this region, affects mRNA stability. To answer this question, we built a reporter plasmid library where different 3'UTR inserts were cloned downstream of the firefly luciferase coding region. The 3'UTR inserts include control sequences with known effects on mRNA stability as well as pairs of wildtype and mutant 3'UTRs, where a single-nucleotide mutation was introduced into the mutant sequence such that the effect of this point mutation could be assessed. In order to determine whether a 4sU- or IVT-based method was better suited for studying mRNA stability changes in our MPRA format, each was optimized through pilot experiments and then compared head-to-head in final sequencing of the MPRAs. Both methods presented different strengths and challenges,

including the technical complexity, replicability, and biological interpretability of the data. In this particular application of mRNA stability measurement in an MPRA format, *in vitro* transcription was concluded to be the best method producing robust results that reliably agreed with expected biological outcomes.

RESULTS

Optimization of an MPRA using an in vitro *transcribed mRNA library*

I began the process of optimizing an MPRA using an *in vitro* transcribed mRNA library by adapting similar established protocols, particularly that of Rabani and colleagues, where they injected an *in vitro* transcribed mRNA pool into zebrafish embryos to study mRNA dynamics during early embryogenesis[27–29]. As suggested by this study, I used the commercially available and highly cited mMESSAGE mMACHINE T7 Ultra Transcription Kit to *in vitro* transcribe my mRNA library.

This kit requires a highly pure DNA template for *in vitro* transcription (IVT), and therefore my first optimization step was to ensure that restriction enzyme digest of my plasmid DNA library resulted in a proper linear template. After testing, it was determined that gel purification with the Macherey Nagel NucleoSpin Gel and PCR Clean-up Kit resulted in impurities in the product, such that the gel-purified template must be further purified with phenol:chloroform isolation to be usable. This further purification increased IVT mRNA yields from 300ng to the expected 30µg mRNA from 1µg of DNA template. Additional minor adjustments to the kit protocol were the addition of 1µL SUPERase-In RNase Inhibitor to the IVT reaction and incubating the IVT reaction for 4 hours versus 1 hour. The suggested lithium chloride precipitation protocol was followed for purification of the end mRNA product.

Once high yields of *in vitro* transcribed mRNA were achieved, I performed quality control on the IVT mRNA library product to ensure that the kit was producing full-length and polyadenylated mRNA. RT-PCR with PCR primers spanning the entire length of the expected mRNA (~2kb) was performed, and results confirmed that the reverse transcribed IVT cDNA is an equivalent length to the template plasmid DNA (**Figure 3-1A**). This experiment also confirms that there is no template DNA carry-over in the IVT product, as no product is seen from the no RT control PCR. Further, IVT reactions from both an individual plasmid DNA template and the plasmid library DNA template were performed with and without addition of PolyA Polymerase (E-PAP) and analyzed via Tapestation to confirm a size shift in the +E-PAP samples, indicating the proper tailing necessary for cellular mRNA dynamics (**Figure 3-1B**).

For introduction of the mRNA library to cells, 8µg IVT mRNA was transfected into 1.5 million PC3 cells in each 10cm plate using 40µL of Lipofectamine MessengerMAX. This mRNA-specific reagent was determined to be significantly better at mRNA transfection than Lipofectamine 3000, though both are stated by the manufacturer as being suitable for mRNA transfection. When luciferase mRNA was transfected at the recommended doses with each reagent, MessengerMAX resulted in ~10x higher luciferase protein expression than Lipofectamine 3000 (**Figure 3-1C**). Using MessengerMAX to introduce *in vitro* transcribed mRNA into cells, we can clearly observe its decay over time, and can differentiate between 3'UTR inserts that appear to decay at different rates (**Figure 3-1D**). The amount of mRNA and transfection reagent was optimized for the 10cm plate of PC3 cells with tested doses between 2-14µg mRNA and 25-50µL MessengerMAX, finding that 8µg RNA with 40µL MessengerMAX resulted in 4-5x higher mRNA transfection efficiency as the lowest doses and sufficient product for all downstream processing.

Once both the production of *in* vitro transcribed mRNA and its transfection into cells were optimized, the last issue to contend with was maintaining low technical variability across

samples while scaling up to perform the full MPRA protocol. In order to quantitatively normalize across all 30 RNA-seq samples collected in the final MPRA (5 time points in 6 replicates), a synthetic RNA was spiked into the TRIzol used for cell collection and RNA isolation. This RNA was built from the nanoluciferase coding sequence, flanked by the same homologous ends as each 3'UTR in the MPRA plasmid library, such that the internal sequence was unique and would be isolated along with the MPRA mRNA library using a MPRA-specific RT primer. The proper dosage of this RNA into the TRIzol was optimized such that it would get sufficient sequencing coverage in each sample, without overwhelming the MPRA reads of interest. This was particularly an issue with the samples collected at 24 hours, where the transfected IVT mRNA had decayed considerably and so could easily be overwhelmed by an exogenous spike-in. Rabani and colleagues used ratios of spike-in mRNA to injected mRNA of interest between 1:1,000 and 1:10,000,000 in their study, and therefore I tested a similar range of ratios[27]. RT-qPCR was performed on RNA samples collected 1, 6, and 24 hours after transfection of 8μg of mRNA with 0.2-10ng of nanoluciferase RNA spiked into each sample (**Figure 3-1E**). While the samples collected 1 or 6 hours after transfection are not overwhelmed by 2ng of spike-in, a 0.2ng spike-in is necessary to achieve no significant increase in the MPRA+Spike-in value versus the MPRA value. This is further shown when these samples were processed through sequencing library preparation and analyzed by gel electrophoresis (**Figure 3-1F**). Again, 10ng of spike-in easily overwhelms the RNA sample collected at 24 hours after transfection, while there is no significant addition to the sample at 0.2 or 0.4ng of spike-in RNA. Since when sequencing the library, a 1:1 ratio of spike-in to MPRA 3'UTR inserts means the spike-in would still be taking 50% of the total reads away from the constructs of interest, we further decreased the spike-in amount for the final protocol to 0.02ng per sample (a 1:400,000 ratio).

Having optimized the *in vitro* transcription, mRNA transfection, RNA isolation, and library prep for this IVT MPRA protocol, we proceeded with sequencing of the six biological replicates

collected, each consisting of mRNA isolated 1, 3, 6, 12, and 24 hours after IVT transfection, for a total of 30 samples sequenced. Preliminary quality control analysis of the nanoluciferase spike-in showed expected results. Approximately 1-15% of the total reads for each sample were the nanoluciferase spike-in, with this amount increasing in the samples collected later, due to the MPRA mRNA of interest degrading while the spike-in remains constant (**Figure 3-1G**). We also observe the utility of this spike-in control when degradation of individual 3'UTR inserts are analyzed, where the expected mRNA decay over time is much clearer after spike-in normalization (**Figure 3-1H**). This normalization also succeeds in decreasing variability between replicates, demonstrating that much of the original variability stems from minor technical differences in sample processing. Overall, optimization of this IVT-based MPRA method of measuring mRNA stability required few deviations from established protocols and manufacturer's recommendations to adapt this method to our system and achieve results with high reproducibility.

Optimization of an MPRA using 4-thiouracil labeling of nascent mRNA

I assessed many variations of 4-thiouracil (4sU)-based protocols that label endogenous nascent mRNA for their potential to adapt to the MPRA format and my question of how single-nucleotide 3'UTR mutations affect mRNA stability. In pulse-chase 4sU experiments, 4sU is first pulsed onto cells to label a particular population of nascently transcribed mRNA, then this is chased by a many-fold higher treatment of unlabeled uridine[30,31]. In this way, only mRNA transcribed during the initial 4sU treatment is labeled, and its degradation can be followed out over a time course. In contrast, some 4sU experiments perform only a single 4sU treatment without a chase, instead using the ratio of newly transcribed +4sU mRNA to total or pre-existing (unlabeled) mRNA at various time points after treatment to calculate mRNA half-lives[32]. Usually, either of these protocols are followed by biotinylation of the thiol group of 4sU mRNA and subsequent streptavidin-based enrichment of the labeled mRNA to isolate it from the cell's total

mRNA. Newer pulldown-independent methods have also been described to avoid the inefficiencies inherent in such biotin-streptavidin enrichment; however, these rely on the conversion of incorporated 4sU into cytidine and the readout of unexpected point mutations in the final sequencing[33,34]. These techniques, therefore, are not suitable for our assessment of how single point mutations affect mRNA stability, as there would be no way to determine whether a point mutation in the sequencing reads indicated an intended test mutation or the labeling of nascent RNA from 4sU conversion. Therefore, we chose to perform pulse-chase 4sU labeling followed by biotinylation and streptavidin pulldown using advanced methane thiosulfonate (MTS)-biotin chemistry, building on protocols optimized by the Simon and Wilusz labs[25,32,35]. Using a pulse-chase protocol versus a pulse-only protocol requires only the final quantification of 4sU labeled nascent RNA over time, instead of quantification of the nascent:total RNA ratio, leaving less room for technical error in the pulldown process.

Some studies have raised concerns that high or prolonged treatment of cells with 4sU may be toxic to cells and therefore confound results by inhibiting rRNA synthesis and causing nucleolar stress[36], while others maintain there is no ill effects of moderate treatment[33]. To determine whether 4sU treatment is acutely toxic to cells in our system, we treated PC3 cells with 50-500μM of 4sU for up to 5 hours and measured cell viability visually and via Trypan blue, where <5% cell death was observed over this time course (**Figure 3-2A**). To ensure that the subcellular nucleolar stress others have found upon 4sU treatment was not occurring in the absence of whole cell toxicity, nucleolar staining was performed on cells treated with 500μM 4sU for 5 hours, allowing visualization of normal nucleolar morphology in 4sU treated PC3 cells (**Figure 3-2B**). Additionally, rRNA production was examined through Tapestation analysis, where normal rRNA levels and 18S:28S ratio (RIN = 10.0) from cells treated with 5 hours of 500μM 4sU demonstrates normal rRNA processing (**Figure 3-2C**). We conclude that up to 5 hours of 4sU treatment at 500μM is non-toxic to PC3 cells.

Our experimental setup requires the cells to be challenged not only with 4sU treatment but also plasmid DNA transfection. In order to ensure that cells are allowed to recover from transfection before 4sU treatment, a 16 hour transfection period was followed by a media change to remove the transfection reagent and a 24 hour recovery period before 4sU treatment. In this way, 24 hour treatment with 500μM 4sU after plasmid transfection was found to cause no appreciable toxicity to cells (**Figure 3-2D**). As 4sU is only incorporated into a relatively small percentage of nascent mRNA at a time, this 24 hour treatment period was chosen to allow build-up of sufficient levels of labeled mRNA in the cells before the uridine chase was applied.

The quality of RNA isolated from cells after transfection and 4sU treatment is integral to the robustness of the subsequent biotinylation and pulldown steps. We determined that RNA isolation using TRIzol, followed by DNase treatment using TURBO DNase, and final purification using the RNeasy MinElute Cleanup Kit was necessary to yield highly pure RNA free from contamination of DNA, salts, or protein. During such optimization, TURBO DNase was found to perform ~35% better than Thermo Fisher's DNase I in decreasing the contamination of RNA prep with residual DNA.

Biotinylation of the purified RNA was performed according to the optimized protocol of Duffy and colleagues using MTSEA-biotin-XX, which results in greater efficiency, higher yields, and less bias than previously used HPDP-biotin[25,35]. In order to normalize between samples for variation in biotinylation and streptavidin pulldown efficiency, an *in vitro* transcribed RNA labeled with 4sU was spiked equally into each RNA sample before biotinylation. We found that the Miltenyi μMACS streptavidin kit used by Russo *et al.* improves the pulldown efficiency over the Invitrogen Dynabeads technology used by Duffy *et al*. (**Figure 3-2E**). The μMACS kit yielded an enrichment ratio of ~40-100x between nascent mRNA measured from +4sU treated versus -4sU untreated cells, whereas Dynabeads only achieved +4sU/-4sU ratios ranging from 1.5-4x.

This protocol's ability to observe mRNA degradation and differentiate between mRNAs with different decay rates was determined using a set of endogenous control mRNAs with various known relative stabilities: Actin (long half-life), DHX9 (moderate half-life), and miR17HG (short half-life). After the optimization detailed above, we observed these expected trends in endogenous mRNA half-lives (**Figure 3-2F**). With this quality control check, combined with the high +4sU enrichment observed using μMACS beads, we concluded that this protocol was robust enough to go forward with the massively parallel reporter assay.

Before sequencing the six biological replicates of the 4sU MPRA that were collected and processed, we first checked the +/- 4sU enrichment and endogenous control mRNA decay in these samples. The enrichment of +4sU samples over -4sU controls ranged from 0.3x to 16.5x (median 3.0x, mean 5.6x), versus the 40-100x enrichment we saw in pilot assays (**Figure 3-2G**). One possibility for this wide range and seeming decrease in overall efficiency is our use of the 4sU-labeled spike-in for pulldown normalization. Though this spike-in was added equivalently to all samples, there are some in which substantially lower amounts (high Cq values) were observed via qPCR (**Figure 3-2H**). This led to the spike-in control displaying highly variable 1hr +4sU/1hr -4sU enrichment scores. While this ratio is expected to equal 1, indicative of the same amount of spike-in being pulled down in each control sample, we see that replicates 1 and 6 are far from this expected value (**Figure 3-2I**). Looking closely, either the -4sU or +4sU control samples of replicates 1 and 6, which had the most variable enrichment scores, had had abnormally high Cq values. This calls into question the validity of normalizing to the +4sU spike-in control, though technically it plays an important role in controlling for processing variation. It remains unclear whether the spike-in is quantifying true technical variability in the biotin-streptavidin pulldown, or whether using it for normalization introduces further noise and artifacts into the system.

When decay trends for endogenous miR17HG and DHX9 and the transfected MPRA luciferase mRNA were analyzed, we see further evidence of replicate-to-replicate variability (**Figure 3-2J**). The overall decay of each mRNA over time is not clear in replicates 1, 2, and 3, part of which may be due to the abnormal spike-in control values in the 10hr samples for replicates 1 and 3, which artificially inflate the normalized nascent mRNA values at the 10hr timepoint. Additionally, in several biological replicates, particularly 4 and 5, nascent mRNA is not observed to decay between the 1 hour and 3 hour time points after the uridine chase was applied. This indicates that the switch between the 4sU pulse and uridine chase within cells is not precise and that newly transcribed mRNA is likely still incorporating 4sU in the first 1-3 hours of uridine treatment, introducing more noise into the system. Despite this, encouragingly, we do observe faster decay for miR17HG versus DHX9 and the MPRA luciferase in all replicates (**Figure 3-2J**). However, since decay of the MPRA luciferase mRNA must be quantifiable over the time-course for further analysis, we sequenced only biological replicates 4, 5, and 6 of the 4sU MPRA to compare these with results seen in the *in vitro* transcribed MPRA.

Comparison of IVT and 4sU massively parallel reporter assays

In total, six biological replicates of the IVT-based MPRA and three replicates of the 4sU-based MPRA were sequenced. Control 3'UTR inserts built into the MPRA plasmid library were used to evaluate each MPRA's ability to accurately capture expected differences in mRNA degradation. Such controls included 3'UTRs containing AU-rich elements (AREs) and Pumilio binding elements (Pum), both of which are known to decrease mRNA stability[37,38]. These trends are clearly observed in the IVT-based MPRA, where addition of either regulatory element leads to faster mRNA decay and a dose-dependent effect is observed for the Pumilio element (**Figure 3-3A-C**). In comparison, the same controls in the 4sU MPRA behave more variably. Using the 4sU method, we observe decreased mRNA stability from the 2X ARE control sequence, as expected (**Figure 3-3D**); however, the results of the Pumilio regulatory element are less clear

(**Figure 3-3E**). Though both the 1X and 2X Pumilio-containing 3'UTRs decay faster than the blank vector, the dose dependence trends the opposite direction as expected, with the 3'UTR containing four Pumilio elements showing increased stability from that containing only one repeat (**Figure 3-3F**). From analysis of these controls, we conclude that the IVT-based MPRA captures more true differences in mRNA decay than the 4sU-based MPRA.

As the IVT-based MPRA demonstrated less variability and more expected trends when analyzing internal controls, we went forward with analysis of mRNA stability changes between the wildtype and mutant pairs of 3'UTRs in this IVT MPRA library (these results detailed in Chapter 2). Nine of the top hits of the IVT MPRA (log_2FC half-life >0.75) were sequenced at sufficient coverage in the 4sU MPRA for comparison across these two methods. For each of these 3'UTR mutations, the IVT method displays clear decay over time and significance is reached due to the relatively small variability between replicates (**Figure 3-3G**). Analyzing the same mutations using the 4sU MPRA results, some, but not all, of the same trends are observed (**Figure 3-3H**). Mutations 2 and 4 reproduce the same increase in mutant stability as found in the IVT MPRA, perhaps even with higher effect sizes; however, the high variability between replicates makes statistical analysis difficult. Mutations 1, 3, and 8 display slight trends toward the same decay patterns as in the IVT MPRA at the 6 hour time point, but overall could not be concluded to agree between methods. Further, mutations 5, 6, and 7 seem to show opposite trends between methods, and mutation 9 shows no change using the 4sU method. Overall, the IVT-based MPRA displays higher reproducibility between replicates and expected trends in mRNA decay are more clearly observed. Importantly, we were also able to replicate results from the IVT MRPA using orthogonal methods such as dual luciferase assays and by using actinomycin D, as detailed in the previous chapter.

DISCUSSION

After optimization of two methods to study changes in mRNA decay, we conclude that, though each method has pros and cons, an IVT-based MPRA was better suited for our purposes than a 4sU-based method. The 4sU protocol is more technically difficult, with numerous steps that introduce higher variability between replicates. Since our goal was to quantify relatively small effect size changes caused by single-nucleotide 3'UTR mutations, the higher reproducibility of the IVT-based assay was important for attaining statistical power and significance. It is not wholly surprising that these MPRAs did not have high concordance with each other, as various methods for measuring mRNA stability genome-wide have resulted in widely variable conclusions, with the median mRNA half-life of the transcriptome being reported anywhere from 3.4 to 10 hours[23].

Overall, the IVT MPRA ended up being more well-controlled than the 4sU MPRA. The spike-in control RNA used in the IVT MPRA was found to be critical in normalizing between each sample, illustrated by the decrease in variability and clearer decay trends in normalized versus unnormalized data. The 4sU MPRA also used a similar spike-in to normalize for variation in the enrichment of nascent mRNA; however, it did not behave fully as intended. Though in some cases it seemed to be important for normalizing between small amounts of sample-to-sample variability (eg: replicates 2-5), there were a few outlier samples in which the spike-in normalization may have introduced more noise into the system. From the observed variability in nascent mRNA enrichment, we conclude that the protocol used for 4sU tagging and pulldown was not fully optimized to a high degree of efficiency, though it was based largely on published protocols and pilot experiments were promising. This inefficiency of biotin-streptavidin pulldown has been noted by other groups as well and likely accounts for the general move in the field towards nucleotide recoding instead of such biochemical enrichment[24]. These newer protocols, such as TimeLapse-Seq and SLAMseq, improve upon the 4sU method, minimizing noise in the system, as well as allowing for significantly less starting material[31,33].

We recognize that there are limitations to using *in vitro* transcribed mRNA as the basis of our MPRA, particularly in that synthetic mRNA is not subjected to co-transcriptional processes that would happen in the cell, such as RNA editing, mRNA modifications, splicing, and alternative polyadenylation. The use of transfection to introduce the mRNA into cells may also bias or change the localization of the mRNAs in the cytoplasm. Further, there could be mRNA left in the cell media able to contaminate mRNA isolated from cells, creating artifactual mRNA dynamics from mRNA that never entered cells[39]. However, these issues are generally controlled by the fact that our final analysis focuses specifically on the pairwise difference between wildtype and mutant constructs. Since all the luciferase mRNA analyzed in our assays encounters the same lack of co-transcriptional processing, both wildtype and mutant constructs will be affected the same, maintaining any difference in decay between them. Therefore, though some effects specifically related to co-transcriptional processing may be missed, the effects observed in the assay are likely to hold true in an endogenous cellular setting.

It is worth noting that, in a practical sense, an IVT-based protocol is simpler, quicker, and less expensive than 4sU-based protocol (**Table 3-1**). While the IVT method has only four major parts (*in vitro* transcription of the library, transfection, RNA isolation, and sequencing preparation), the 4sU method has eight (transfection, 4sU pulse treatment, uridine chase, RNA isolation, biotinylation, streptavidin pulldown, nascent mRNA purification, and sequencing preparation), most of which required more optimization and contain more individual steps than the IVT method. From cell plating to final mRNA isolation, the 4sU method takes a minimum of 6 days, whereas the IVT method takes approximately three. The IVT method does require more preparation before transfection; however, to create the mRNA library. Each IVT reaction, starting with undigested plasmid library, takes 1-2 days, and many such reactions may be needed to accumulate enough mRNA library for large-scale transfection. While this adds to the total

amount of time needed from beginning to end, the IVT method remains faster and simpler overall.

In our hands, fewer modifications had to be made to established protocols to arrive at a rigorous and reproducible MPRA protocol based on *in vitro* transcription than 4sU nascent RNA labeling. This is not to say that 4sU labeling cannot be used in a reliable manner, but that it seems to be more complex, and for our purposes, the additional complexity of further 4sU optimization was not justifiable when *in vitro* transcription produced robust data with a simpler workflow. In other cases, where the effects of co-transcriptional processes or mRNA localization may be a significant part of the experimental question, or when mRNA stability of endogenous mRNA, not that of a reporter library, is being measured, a 4sU-based MPRA may be necessary and appropriate. In this case, we would recommend experimenting with nucleotide recoding as an alternative to biochemical enrichment, and extensive optimization and QC before proceeding with large-scale sequencing. Based on the optimization and comparison of techniques detailed in this chapter, the IVT-based MPRA for measurement of mRNA stability changes caused by 3'UTR mutations was used in Schuster *et al*. as detailed in the previous chapter, where further analysis of the results elucidated interesting facets of 3'UTR biology and these mutations' relevance to cancer.

ACKNOWLEDGEMENTS

We would like to thank Jeremy A. Schofield for helpful discussion of 4sU-based methods during the optimization of this assay.

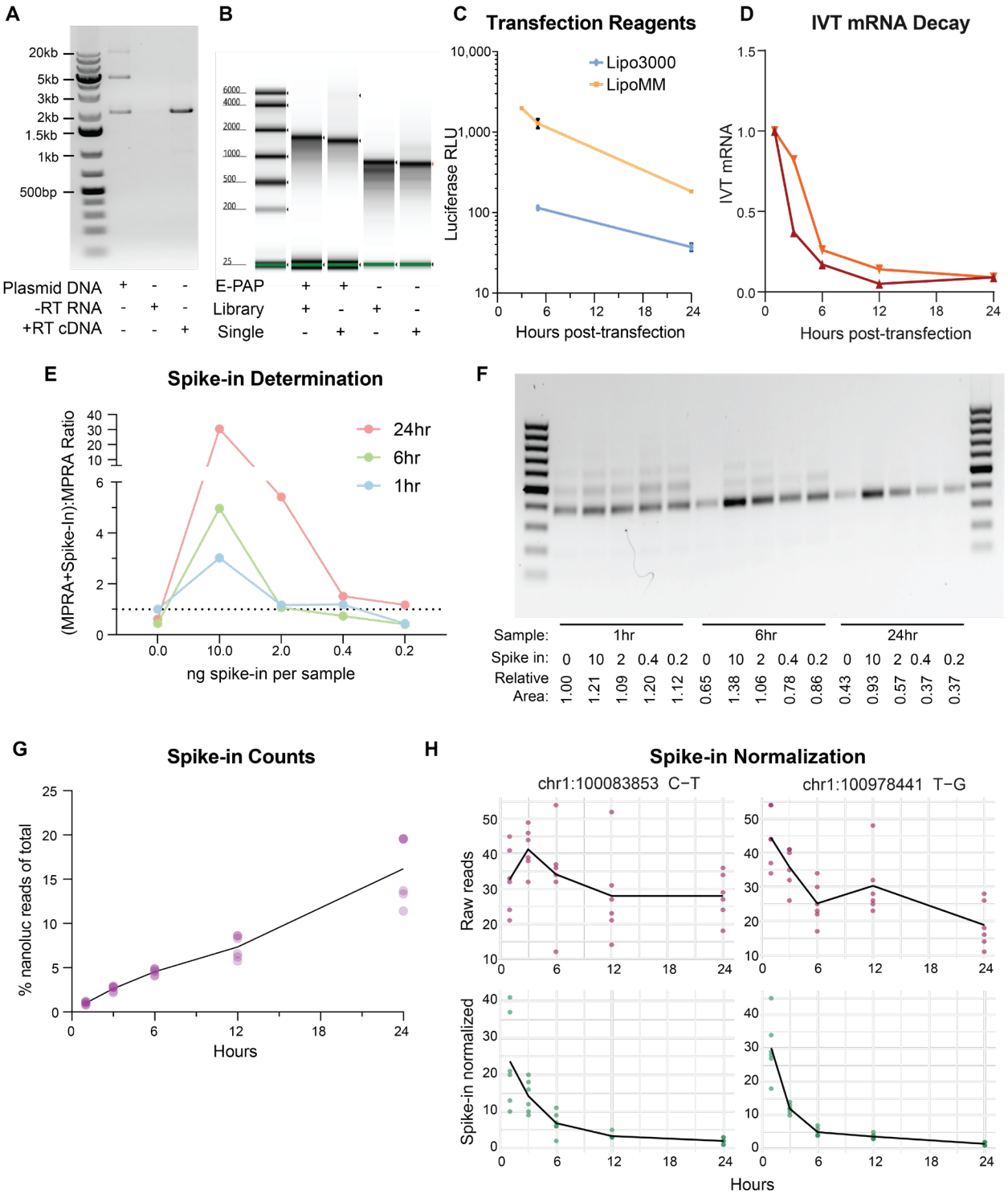

Figure 3-1: Optimization of an IVT-based MPRA

(A) Agarose gel image of RT-PCR performed on template DNA, no RT control IVT RNA, and +RT IVT cDNA using primers that amplify the full-length cDNA.

Figure legend continued on next page

Figure 3-1 continued:

(B) Tapestation RNA assay conducted on IVT products from either plasmid library template or a single plasmid template with or without E. coli Poly-A Polymerase (E-PAP) present in tailing reactions.

(C) Firefly luciferase relative light unit (RLU) values from cells lysed 3, 6, or 24 hours after IVT mRNA transfection with either Lipofectamine 3000 or Lipofectamine MessengerMAX (100ng mRNA + 0.3uL either reagent).

(D) qPCR amplifying luciferase mRNA 1, 3, 6, 12, and 24 hours post-transfection of two different IVT products

(E) RT-qPCR results of spike-in dilution testing, where 0, 10, 2, 0.4, and 0.2ng of exogenous spike-in RNA was added to 1hr, 6hr, and 24hr post-transfection collected samples.

(F) Agarose gel image of library preparation products, where samples spiked with differing amounts of exogenous spike-in were processed by luciferase-specific reverse transcription and two rounds of library prep PCR amplification. The expected product is ~380bp.

(G) The percentage of total sequencing reads from each IVT MPRA sample coming from the nanoluciferase spike-in.

(H) mRNA decay of two individual constructs before and after raw reads were normalized to the amount of spike-in control measured in each sample.

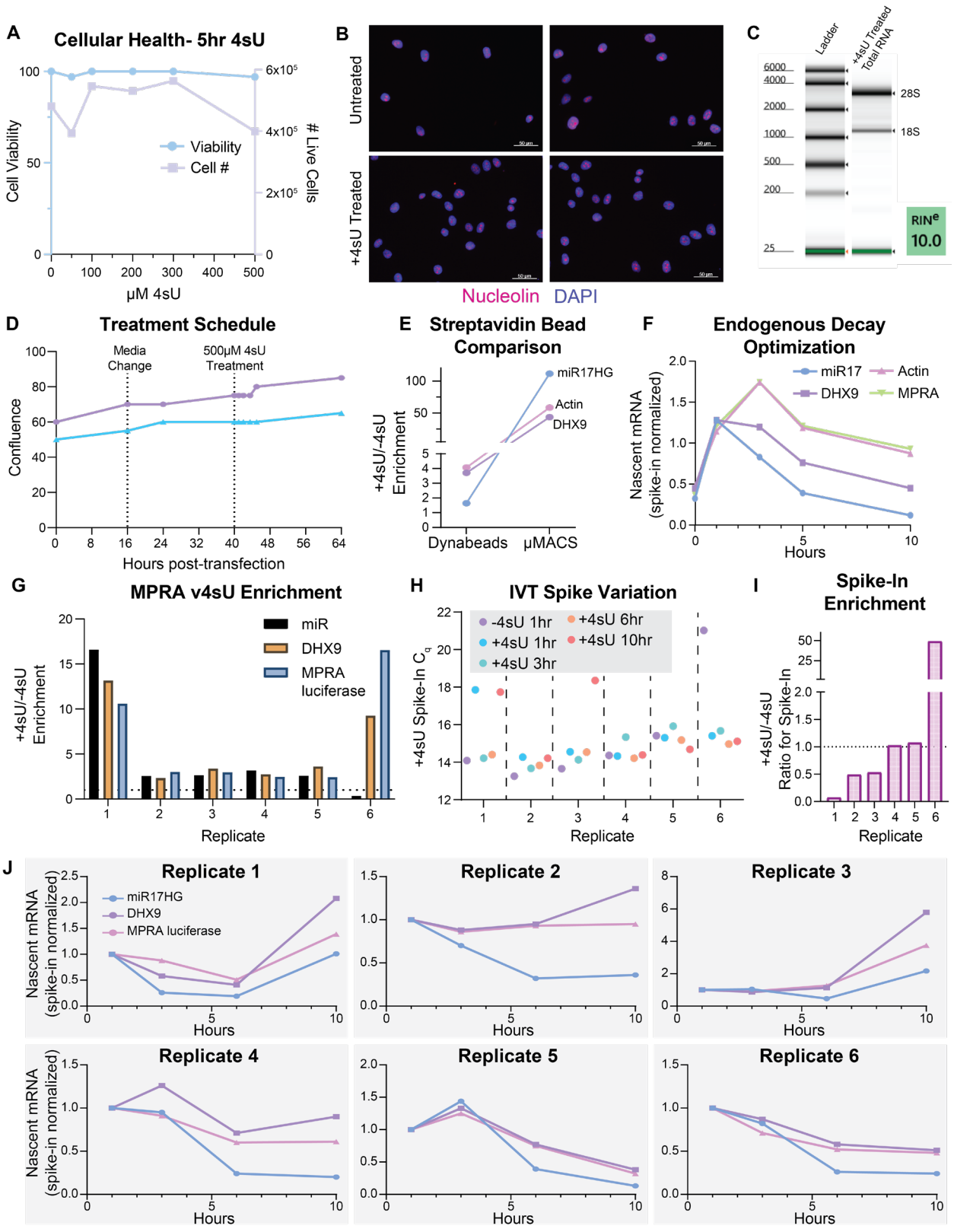

Figure 3-2: Optimization of a 4-thiouracil-based MPRA

Figure legend continued on next page

Figure 3-2 continued:

(A) Health of PC3 cells after 5 hours of 4sU treatment. Percent viability measured by Trypan blue displayed on the left y-axis and number of live cells collected after treatment on the right y-axis.

(B) Immunoflorescent images of PC3 cells taken after 5 hours of 500uM 4sU treatment. Nucleolar morphology visualized using anti-Nucleolin antibody.

(C) Tapestation RNA assay conducted on RNA isolated from cells after 5 hours of 500uM 4sU treatment.

(D) Cellular confluence measured from two populations of cells over an experimental timeline of DNA transfection, media change at 16 hours, recovery period, and 24 hour 500uM 4sU treatment.

(E) Enrichment ratio of miR17HG, Actin, and DHX9 levels recovered after RNA isolation, biotinylation, and streptavidin bead pulldown to elute only 4sU-tagged mRNA from cells cultured with or without 4sU.

(F) Nascent mRNA levels measured by qPCR after 24 hour 100uM 4sU treatment, 0, 1, 3, 5, and 10 hours of 10mM uridine chase, and biotin-streptavidin enrichment of 4sU-tagged nascent mRNA. Results are representative of three optimization experiments.

(G) Enrichment ratios quantified by qPCR of miR17HG, DHX9 and transfected MPRA luciferase construct between 4sU treated and untreated cell samples across the six biological replicate MPRA samples. Dotted line indicates a ratio of one, below which all signal would be considered noise from unlabeled mRNA contamination.

(H) qPCR C_q values measuring amounts of the exogenous 4sU-labeled spike-in control across the 30 MPRA samples.

(I) Enrichment ratios quantified by qPCR of the exogenous spike-in control between 4sU treated and untreated cell samples across the six biological replicate MPRA samples. Dotted line indicates the expected ratio of one.

(J) Decay of miR17HG, DHX9, and transfected MPRA luciferase mRNA measured by qPCR across the six biological replicate MPRA samples.

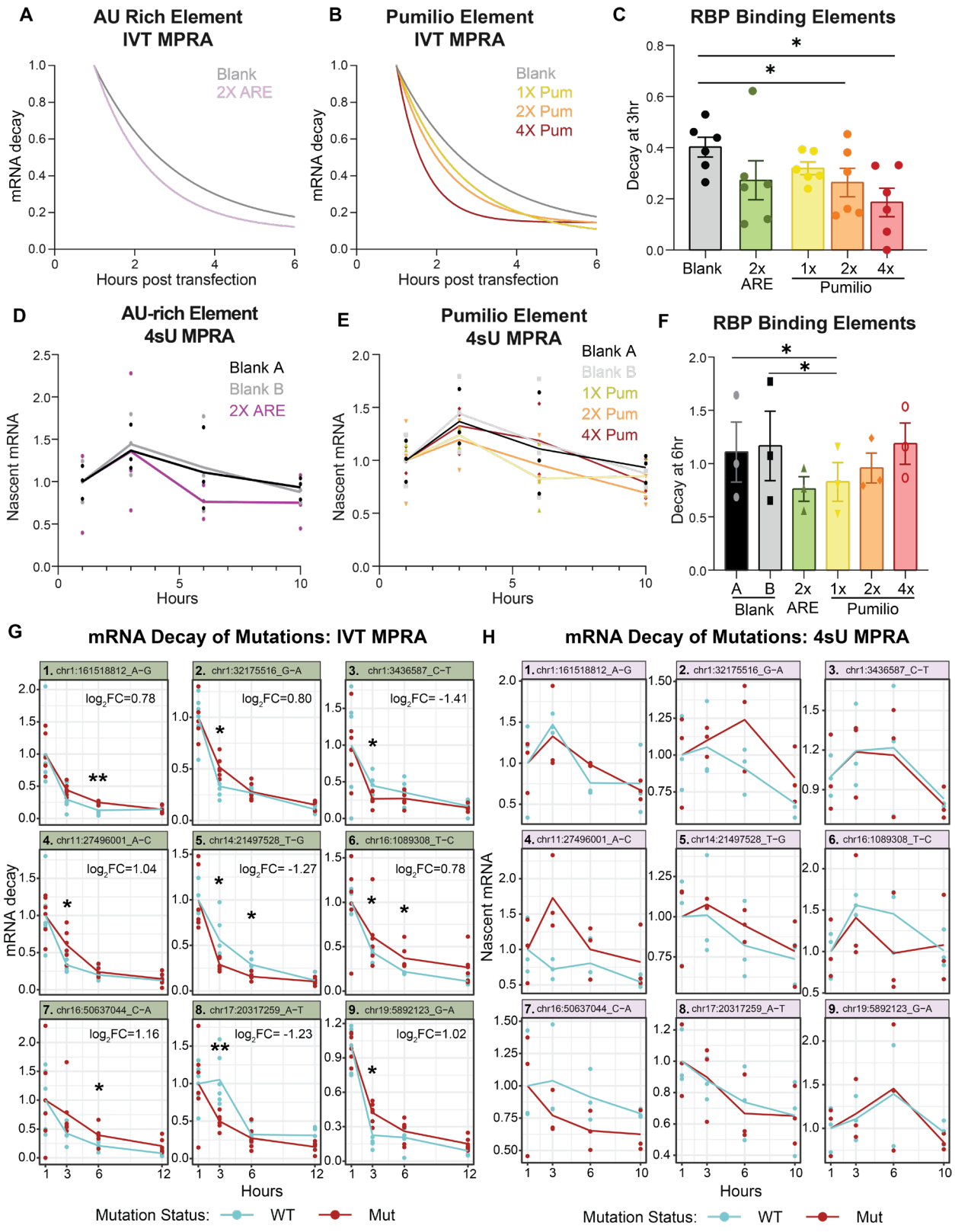

Figure 3-3: Comparison of two MPRAs for measuring mRNA stability changes

Figure 3-3 continued:

(A) Nonlinear least-squares regression curves fit to IVT MPRA data for internal control 3'UTR inserts consisting of a blank vector sequence versus 2X repeated AU-rich element (ARE).

(B) Nonlinear least-squares regression curves fit to IVT MPRA data for internal control 3'UTR inserts consisting of a blank vector sequence versus 1X, 2X, and 4X repeated Pumilio element (Pum).

(C) Summary results of 1hr-to-3hr decay corresponding to 3A & 3B (mean ± SEM, n=6). Statistical analysis conducted using paired t-test and multiple comparisons correction of Benjamini, Krieger, and Yekutieli (*q<0.05).

(D) Nascent mRNA decay of 2X ARE-containing mRNA versus two variations of blank 3'UTR controls in the 4sU MPRA. Points show data from all 3 replicates sequenced with lines connecting corresponding mean values.

(E) Nascent mRNA decay of 1X, 2X, and 4X Pumilio element-containing mRNAs versus two variations of blank 3'UTR controls in the 4sU MPRA. Points show data from all 3 replicates sequenced with lines connecting corresponding mean values.

(F) Summary results of 1hr-to-6hr decay corresponding to 3D & 3E (mean ± SEM, n=3). Statistical analysis conducted using paired ratio t-test.

(G) mRNA decay values (each sample normalized first to the spike-in control, then each construct normalized such that the mean at 1hr equals 1) for nine mutations within the IVT MPRA. Log_2FC mRNA half-life values shown for each in upper right. Stars indicate significance (*:p<0.05, **:p<0.01) using t-test to compare WT and mutant values at either 3hr or 6hr time points.

(H) Nascent mRNA decay values (each sample normalized first to the spike-in control, then each construct normalized such that the mean at 1hr equals 1) for corresponding nine mutations within the 4sU MPRA.

Table 3-1: Overview of IVT- versus 4sU-based methods of measuring mRNA stability

Feature	IVT MPRA Method	4sU MPRA Method
Reproducibility	High	Low
Simplicity of Workflow	Simple	Complex
Time Needed	IVT mRNA: 1-2 days/reaction Transfection to RNA: 2 days	Transfection to RNA: 4 days Enrichment: 1-2 days
Ease of Optimization	Some optimization needed	Extensive optimization needed
Ability to analyze endogenous mRNA	No, only transfected mRNA decay can be observed	Yes, observe decay of both endogenous mRNAs and transfected DNA/mRNA
Effects of co-transcriptional processing	Not observed	Can be observed
Reagent cost	~$3,200	~$5,500

MATERIALS & METHODS

The methods used for the performance of the final IVT-based MPRA are detailed in Chapter 2 (pages 74-91). Here I describe additional methods used in the optimization of the IVT MPRA, as well as the methods pertaining to the 4sU-based MPRA. A finalized step-by-step protocol used for each of the MPRAs is included in addition.

IVT Optimization

RT-PCR to test IVT mRNA length: Messenger RNA *in vitro* transcribed from the MPRA base plasmid using the mMESSAGE mMACHINE T7 Ultra Kit as described in the previous chapter was reverse transcribed into cDNA using the SuperScript III First-Strand Synthesis System and a MPRA-specific RT primer. PCR amplification of cDNA, no RT control, and MPRA base plasmid DNA was performed using primers flanking the entire mRNA sequence and Phusion High-Fidelity PCR Master Mix. The forward primer sits just upstream of the luciferase coding sequence start, downstream of the T7 promoter, and the reverse primer sits just downstream of the inserted 3'UTR sequence, upstream of the polyadenylation signal.

Testing lipofectamine transfection reagents: PC3 cells were plated in a 96-well plate at 8,000 cells per well. The next day, cells were transfected with either 100ng of IVT mRNA +0.3µL Lipofectamine 3000 or 100ng IVT mRNA + 0.3µL Lipofectamine MessengerMAX. Cells were collected 3, 5, and 24 hours after mRNA transfection. Firefly luciferase luminescence was measured using the LAR II solution from Promega's Dual Luciferase Reporter Assay System.

Testing spike-in concentrations: Test 1 (RT-qPCR): RNA was isolated from cells transfected with IVT mRNA using 1mL of TRIzol per sample as described in the previous chapter. Transfected cells were collected 1, 6, and 24 hours after mRNA transfection. Each 1mL of TRIzol for RNA isolation was spiked with 0, 10, 2, 0.4, or 0.2ng of IVT nanoluciferase control RNA. RNA was reverse transcribed into cDNA using the SuperScript III First-Strand Synthesis System and a

MPRA-specific RT primer, then qPCR performed using SsoAdvanced Universal SYBR Green Supermix and primers amplifying (A) specifically the nanoluciferase spike-in, (B) specifically the MPRA firefly luciferase IVT mRNA, or (C) a region common to both the spike-in and MPRA IVT mRNA. The ratio of C/B was used to determine whether the spike-in was significantly increasing the total amount of spike-in+MPRA RNA over the base amount of MPRA RNA. *Test 2 (Library prep PCR):* The same cDNA made in Test 1 was processed through the full sequencing preparation protocol using a two-step PCR amplification approach (described in the previous chapter). Briefly, the 3'UTR insert of the MPRA IVT mRNA was amplified using MPRA-specific primers adding homologous adapter ends in a 50µL Q5 MasterMix PCR reaction, then this PCR product was purified using the Zymo Research Clean & Concentrator kit. A second PCR amplification step was performed using primers specific to the homologous ends added in the previous step that added the necessary Illumina sequencing adapter sequence. This 50µL Q5 MasterMix PCR reaction was also purified using the Zymo Research kit. These purified products were run on a 1% agarose gel and band intensities quantified in ImageJ.

4sU Optimization

Making 4sU solutions: 4sU solution was made fresh for each experiment. A minimal amount of powdered 4sU (~1mg, Fisher Scientific AAJ60679MC) was transferred to an eppendorf tube and weighed. Nuclease-free water was added to a working concentration of 50-75mM (4sU molecular weight: 260.264 g/mol). 4sU solutions are light-sensitive and always kept in the dark. All experiments using 4sU were performed in the dark, from cell treatment through biotinylation and enrichment.

General course of transfection and 4sU treatment: PC3 cells were plated (175,000 cells per well for 6-well plates or 1.5 million cells per well for 10cm plates) on Day 0. On Day 1, cells were

transfected with plasmid using Fugene HD according to manufacturer's directions (1.5µg DNA + 4.5µL of Fugene HD for 6-well plates or 8µg DNA + 24µL Fugene HD for 10cm plates). On Day 2, approximately 16 hours after transfection, media was changed on the cells to remove transfection reagents. On Day 3, cells were treated with 4sU.

Assessing 4sU toxicity: PC3 cells were plated and transfected with a control luciferase-expressing plasmid in 6-well plates, then treated with 0, 50, 100, 200, 300, or 500µM 4sU. Cells were observed over 5 hours before being collected. Collected cells were mixed with Trypan Blue and counted using a Bio-Rad TC20 to determine cell number and viability.

Nucleolar staining after 4sU treatment: Coverslips were first prepared by washing in 70% ethanol, air drying, and coating in 0.01% Poly-L-Lysine, followed by rinsing in MilliQH_2O and UV sterilization. PC3 cells were plated on coverslips placed in 6-well plates using ~20,000 cells in a 200µL volume. After cells settled onto the coverslip for 4 hours, additional media was added to wells. The next day, cells were treated with 0 or 500µM 4sU for 5 hours before fixing with 1mL of 2% paraformaldehyde for 10 minutes at room temperature. Staining was performed by first permeabilizing cells with 2mL of 0.1% Triton X-100 for 10 minutes on ice, followed by 30 minutes of blocking with 5% BSA, incubation with primary antibody overnight (1:1,000 α-Nucleolin ab136649), 1 hour of secondary antibody incubation (1:500 goat α-mouse Alexa Flour 555, Thermo Fisher #A-21422), and mounting with Vectashield mounting media with DAPI. All images were taken using the same exposure times across samples.

Tapestation analysis: PC3 cells were plated and transfected with luciferase plasmid DNA in 10cm plates, then treated with 500µM 4sU. Cells were collected after 5 hours of 4sU treatment. Total RNA was isolated using the Qiagen RNeasy Miniprep Plus Kit, then DNase treatment was performed on 50µg RNA using 50µL DNase I (Thermo EN0525) incubated for 30 minutes at 37°C. Phenol:chloroform purification followed by ethanol precipitation was performed after

DNase treatment to remove DNase. This final product was then sent to the Fred Hutchinson Genomics Core to analyze via RNA Assay on the Agilent 4200 Tapestation.

DNase optimization: Thermo Fisher DNase I (EN0525) was tested using a ratio of 1µg RNA:1µL DNase I and a 30 minute incubation at 37°C. TURBO DNase (Invitrogen AM1907) was tested at a ratio of 10µg RNA:1µL TURBO DNase and a 30 minute incubation at 37°C.

Using an in vitro *transcribed 4sU-labeled RNA spike-in:* A short *in vitro* transcription template was created from the nanoluciferase coding sequence using two rounds of PCR to amplify the sequence, add homologous ends identical to those flanking the 3'UTR sequences of the MPRA plasmid library, and add a T7 promoter. The MAXIscript T7 Transcription Kit was then used to transcribe a 4sU-labeled control RNA. The manufacturer's protocol was followed with the alteration of using 0.5µL of unlabeled UTP plus 0.5µL s^4UTP (TriLink N-1025-1) in place of 1µL UTP. The reaction was incubated for 1 hour at 37°C, then optional DNase treatment was performed followed by phenol:chloroform:IAA purification. This protocol was adapted from the "RNA ladder synthesis" protocol from Duffy *et al.* 2016[35]. 50ng of this +4sU labeled control RNA was added to each sample during biotinylation to control for technical variation in enrichment.

Biotinylation: DNase-treated RNA from 4sU-treated samples was biotinylated as described in Duffy *et al*. 2016[35]. 8-10µg of RNA, 1µL 1M HEPES (pH 7.4), 1µL 0.5M EDTA, 50ng +4sU IVT control RNA, and water to a 40µL total volume was combined in an Eppendorf tube. MTSESA Biotin-XX (Biotium 90066) stocks (1mg/mL) were diluted 1:20 in dimethylformamide (4µL biotin + 76µL DMF), then 10µL of diluted biotin was added to each reaction for a total volume of 50µL. Reactions were covered with foil and incubated at room temperature for 30 minutes with rotation. Unreacted biotin was then removed using a modified RNeasy MinElute Cleanup protocol. First, 50µL water and 100µL chloroform was added to the biotin reaction in a Heavy Phase Lock Gel Tube (QuantaBio 2302830), shaken vigorously, and incubated at room temperature for 2 minutes before centrifuging at 12,000xg for 5 minutes at 4°C. The aqueous

phase (~100µL) was then transferred to a new tube for RNeasy MinElute Cleanup following the manufacturer's protocol. All centrifugation steps were performed at 4°C and final biotinylated RNA was eluted in 50µL.

4sU enrichment using Invitrogen Dynabeads: Invitrogen Dynabeads were tested for 4sU enrichment exactly following the protocol described in Duffy *et al*. 2016[35]. Briefly, streptavidin beads were blocked starting with 2 washes in water, followed by 2 washes in high salt wash buffer, then blocking for 1 hour at room temperature using a mixture of high salt wash buffer and glycogen. For fractionation, 5µL of wash buffer was added to 50µL product after biotinylation and biotin cleanup. 5µL of this 55µL total was saved as the "total RNA"/input sample. The rest was added to 100µL of blocked beads and incubated in the dark for 15 minutes at room temperature with rotation. A magnetic rack was then used to separate beads, washing thrice with 100µL high salt wash buffer. 25µL elution buffer (100mM DTT, 20mM HEPES, 1mM EDTA, 100mM NaCl, 0.05% Tween) was added to beads to elute nascent 4sU-tagged and biotinylated RNA twice, for a total of 50µL nascent RNA sample.

4sU enrichment using Miltenyi µMACS: Miltenyi µMACS system was tested for 4sU enrichment as described in Russo *et al.* 2017[32]. Of the 50µL biotinylated sample, 5µL was reserved as the "total RNA" sample. The rest of the sample was mixed with 100µL streptavidin bead solution and incubated in the dark for 15 minutes at room temperature. The µMACS column was equilibriated twice with 100µL nucleic acid buffer immediately before the RNA/bead solution was added to the column. The flow-through was collected into a tube labeled "pre-existing RNA". The column was then washed with column wash buffer (100mM Tris-HCl, 10mM EDTA, 1M NaCl, 0.1% Tween-20), once with 100µL (added to "pre-existing RNA"), then again with 500µL (discarded). Nascent RNA was eluted by applying 100µL of 100mM DTT and letting this flow through the beads by gravity for 5 minutes. This elution was repeated for a total elution volume of 200µL.

RNA cleanup after 4sU nascent RNA enrichment: A modified RNeasy MinElute Cleanup protocol was used to purify the total, pre-existing, and total RNA samples after 4sU enrichment. 350µL Buffer RLT and 250µL 100% ethanol was added to each 100µL RNA sample before applying to MinElute columns. Samples were centrifuged 15 seconds (all centrifugation done at 12,000xg, 4°C), then 500µL Buffer RPE supplemented with 35µL 1% β-mercaptoethanol (10mM βME) was added to the column. Sample was centrifuged for 15 seconds, followed by application of 500µL fresh 80% ethanol and further centrifugation for 2 minutes. Finally, samples were spun at max speed for 5 minutes and then purified RNA eluted using 14µL of nuclease-free water.

Quantitative PCR for endogenous mRNA decay: RNA samples were reverse transcribed using iScript Reverse Transcription Supermix according to manufacturer's protocol using equal volumes of RNA sample. qPCR was performed using SsoAdvanced Universal SYBR Green Supermix and primers amplifying miR17HG, DHX9, Actin, and/or the luciferase plasmid used for the MPRA library.

Sequencing and computational analysis of the 4sU-based MPRA: Nascent RNA samples from replicates 4, 5, and 6 of the 4sU-based MPRA were prepped for sequencing and alignment of reads to the MPRA plasmid library "genome" were performed as described for the IVT-based MPRA in the previous chapter. For normalization, all cDNA reads were first normalized using the spike-in control ratios across samples. Then normalized cDNA counts were normalized to plasmid DNA counts to account for variation in transfection efficiency. Finally, counts across each 3'UTR insert were normalized by setting the average of the 1hr counts was to 1, making comparison across inserts feasible.

Primer sequences:

Primer Name	Sequence
DHX9 qPCR For	CCGATTCCTCCATGCGAGTT
DHX9 qPCR Rev	TCTGGCCTTCTACCGAGACA
MIR17HG qPCR For	GACTCCTGACAAAATGCAGCC
MIR17HG qPCR Rev	TCTGAAGTCTCAAGTGGGCAT
Actin qPCR For	GGCATGGGTCAGAAGGATT
Actin qPCR Rev	CACACGCAGCTCATTGTAGA
+4sU Spike-in Template PCR 1 For	GCCAGCGCCAGGATCAACGTCTAACATGGTCTTCACACTCGA
+4sU Spike-in Template PCR 1 Rev	CGCCCCGACTCTAGCTAGACTGCAAGACCTTCATACGGGAT
+4sU Spike-in Template PCR 2 For	TAATACGACTCACTATAGGGAGAGCCAGCGCCAGGATCAACGTCTAAC
+4sU Spike-in Template PCR 2 Rev	GGTTTGTCCAAACTCATCAATGTCGCCCCGACTCTAGCTAGACTGCA
MPRA Specific RT Primer	CGCCCCGACTCTAGC
Spike-in Test MPRA Firefly+Spike Total For	GCCAGCGCCAGGATC
Spike-in Test MPRA Firefly+Spike Total Rev	CGCCCCGACTCTAGC
MPRA Firefly Luciferase For (for Spike-in and 4sU endogenous qPCRs)	GCACATATCGAGGTGGACATTA
MPRA Firefly Luciferase Rev (for Spike-in and 4sU endogenous qPCRs)	CCACGATCCGATGGTTTGTAT
Spike in Test Nanoluc qPCR For	GGAGGTGTGTCCAGTTTGTT
Spike in Test Nanoluc qPCR Rev	ATGTCGATCTTCAGCCCATTT

CHAPTER 3 REFERENCES

1. Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., and Brown, P.O. (2002). Precision and functional specificity in mRNA decay. Proc. Natl. Acad. Sci. U. S. A. *99*, 5860–5865. 10.1073/PNAS.092538799/ASSET/934EF8FA-C0CF-418B-8265-DC6A5EE4A0BA/ASSETS/GRAPHIC/PQ0925387004.JPEG.
2. Yang, E., van Nimwegen, E., Zavolan, M., Rajewsky, N., Schroeder, M., Magnasco, M., and Darnell, J.E. (2003). Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. Genome Res. *13*, 1863–1872. 10.1101/GR.1272403.
3. Hao, S., and Baltimore, D. (2009). The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. Nat. Immunol. 2009 103 *10*, 281–288. 10.1038/ni.1699.
4. Chen, C.Y.A., and Shyu, A. Bin (1995). AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem. Sci. *20*, 465–470. 10.1016/S0968-0004(00)89102-1.
5. Mayr, C. (2019). What Are 3' UTRs Doing? Cold Spring Harb. Perspect. Biol. *11*. 10.1101/CSHPERSPECT.A034728.
6. Fabian, M.R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. Annu. Rev. Biochem. *79*, 351–379. 10.1146/ANNUREV-BIOCHEM-060308-103103.
7. Li, W., Deng, X., and Chen, J. (2022). RNA-binding proteins in regulating mRNA stability and translation: roles and mechanisms in cancer. Semin. Cancer Biol. *86*, 664–677. 10.1016/J.SEMCANCER.2022.03.025.
8. Chen, C.Y.A., and Shyu, A. Bin (2017). Emerging Themes in Regulation of Global mRNA Turnover in cis. Trends Biochem. Sci. *42*, 16–27. 10.1016/J.TIBS.2016.08.014.
9. Boo, S.H., and Kim, Y.K. (2020). The emerging role of RNA modifications in the regulation of mRNA stability. Exp. Mol. Med. *52*, 400–408. 10.1038/S12276-020-0407-Z.
10. Mayr, C., and Bartel, D.P. (2009). Widespread Shortening of 3′UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. Cell *138*, 673–684. 10.1016/j.cell.2009.06.016.
11. Bae, H., and Coller, J. (2022). Codon optimality-mediated mRNA degradation: Linking translational elongation to mRNA stability. Mol. Cell *82*, 1467–1476. 10.1016/J.MOLCEL.2022.03.032.
12. Hanson, G., and Coller, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. Nat. Rev. Mol. Cell Biol. *19*, 20–30. 10.1038/NRM.2017.91.
13. Burke, P.C., Park, H., and Subramaniam, A.R. (2022). A nascent peptide code for translational control of mRNA stability in human cells. Nat. Commun. *13*. 10.1038/S41467-022-34664-0.
14. Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al. (2015). Codon Optimality Is a Major Determinant of mRNA Stability. Cell *160*, 1111–1124. 10.1016/J.CELL.2015.02.029.
15. Chen, C.Y.A., Ezzeddine, N., and Shyu, A. Bin (2008). Chapter 17 Messenger RNA Half-Life Measurements in Mammalian Cells. Methods Enzymol. *448*, 335–357. 10.1016/S0076-6879(08)02617-7.
16. Bensaude, O. (2011). Inhibiting eukaryotic transcription. Which compound to choose? How to evaluate its activity? http://dx.doi.org/10.4161/trns.2.3.16172 *2*, 103–108. 10.4161/TRNS.2.3.16172.
17. Shyu, A.B., Greenberg, M.E., and Belasco, J.G. (1989). The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways. Genes Dev. *3*, 60–72. 10.1101/GAD.3.1.60.
18. Seiser, C., Posch, M., Thompson, N., and Kuhn, L.C. (1995). Effect of Transcription Inhibitors on the Iron-dependent Degradation of Transferrin Receptor mRNA. J. Biol. Chem. *270*, 29400–29406. 10.1074/JBC.270.49.29400.

19. Harrold, S., Genovese, C., Kobrin, B., Morrison, S.L., and Milcarek, C. (1991). A comparison of apparent mRNA half-life using kinetic labeling techniques vs decay following administration of transcriptional inhibitors. Anal. Biochem. *198*, 19–29. 10.1016/0003-2697(91)90500-S.
20. Xu, N., Loflin, P., Chen, C.Y.A., and Shyu, A. Bin (1998). A broader role for AU-rich element-mediated mRNA turnover revealed by a new transcriptional pulse strategy. Nucleic Acids Res. *26*, 558–565. 10.1093/NAR/26.2.558.
21. Baird, T.D., and Hogg, J.R. (2018). Using tet-off cells and RNAi knockdown to assay mRNA decay. Methods Mol. Biol. *1720*, 161–173. 10.1007/978-1-4939-7540-2_12/FIGURES/1.
22. Costello, A., Lao, N.T., Gallagher, C., Capella Roca, B., Julius, L.A.N., Suda, S., Ducrée, J., King, D., Wagner, R., Barron, N., et al. (2019). Leaky Expression of the TET-On System Hinders Control of Endogenous miRNA Abundance. Biotechnol. J. *14*. 10.1002/BIOT.201800219.
23. Tani, H., and Akimitsu, N. (2012). Genome-wide technology for determining RNA stability in mammalian cells. https://doi.org/10.4161/rna.22036 *9*, 1233–1238. 10.4161/RNA.22036.
24. Duffy, E.E., Schofield, J.A., and Simon, M.D. (2019). Gaining insight into transcriptome-wide RNA population dynamics through the chemistry of 4-thiouridine. Wiley Interdiscip. Rev. RNA *10*, e1513. 10.1002/wrna.1513.
25. Duffy, E.E., Rutenberg-Schoenberg, M., Stark, C.D., Kitchen, R.R., Gerstein, M.B., and Simon, M.D. (2015). Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. Mol. Cell *59*, 858–866. 10.1016/j.molcel.2015.07.023.
26. Jia, L., Mao, Y., Ji, Q., Dersh, D., Yewdell, J.W., and Qian, S.B. (2020). Decoding mRNA translatability and stability from the 5' UTR. Nat. Struct. Mol. Biol. *27*, 814–821. 10.1038/S41594-020-0465-X.
27. Rabani, M., Pieper, L., Chew, G.-L., and Schier, A.F. (2017). A Massively Parallel Reporter Assay of 3′ UTR Sequences Identifies In Vivo Rules for mRNA Degradation. Mol. Cell *68*, 1083-1094.e5. 10.1016/J.MOLCEL.2017.11.014.
28. Sample, P.J., Wang, B., Reid, D.W., Presnyak, V., McFadyen, I., Morris, D.R., and Seelig, G. (2018). Human 5′ UTR design and variant effect prediction from a massively parallel translation assay. bioRxiv, 310375. 10.1101/310375.
29. Litterman, A.J., Kageyama, R., Le Tonqueze, O., Zhao, W., Gagnon, J.D., Goodarzi, H., Erle, D.J., and Ansel, K.M. (2019). A massively parallel 3′ UTR reporter assay reveals relationships between nucleotide content, sequence conservation, and mRNA destabilization. Genome Res. *29*, 896–906. 10.1101/gr.242552.118.
30. Munchel, S.E., Shultzaberger, R.K., Takizawa, N., and Weis, K. (2011). Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. Mol. Biol. Cell *22*, 2787–2795. 10.1091/MBC.E11-01-0028/ASSET/IMAGES/LARGE/2787FIG5.JPEG.
31. Herzog, V.A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., Wlotzka, W., Von Haeseler, A., Zuber, J., and Ameres, S.L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. Nat. Methods 2017 1412 *14*, 1198–1204. 10.1038/nmeth.4435.
32. Russo, J., Heck, A.M., Wilusz, J., and Wilusz, C.J. (2017). Metabolic labeling and recovery of nascent RNA to accurately quantify mRNA stability. Methods *120*, 39–48. 10.1016/j.ymeth.2017.02.003.
33. Schofield, J.A., Duffy, E.E., Kiefer, L., Sullivan, M.C., and Simon, M.D. (2018). TimeLapse-seq: Adding a temporal dimension to RNA sequencing through nucleoside recoding. Nat. Methods *15*, 221–225. 10.1038/nmeth.4582.
34. Herzog, V.A., Fasching, N., and Ameres, S.L. (2020). Determining mRNA Stability by Metabolic RNA Labeling and Chemical Nucleoside Conversion. Methods Mol. Biol. *2062*, 169–189. 10.1007/978-1-4939-9822-7_9.
35. Duffy, E.E., and Simon, M.D. (2016). Enriching s 4 U-RNA using methane thiosulfonate (MTS) chemistry. Curr Protoc Chem Biol. 10.1002/cpch.12.

36. Burger, K., Mühl, B., Kellner, M., Rohrmoser, M., Gruber-Eber, A., Windhager, L., Friedel, C.C., Dölken, L., and Eick, D. (2013). 4-Thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. RNA Biol. *10*, 1623–1630. 10.4161/rna.26214.
37. Cottrell, K.A., Chaudhari, H.G., Cohen, B.A., and Djuranovic, S. (2018). PTRE-seq reveals mechanism and interactions of RNA binding proteins and miRNAs. Nat. Commun. *9*, 301. 10.1038/s41467-017-02745-0.
38. Barreau, C., Paillard, L., and Osborne, H.B. (2006). AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res. *33*, 7138–7150. 10.1093/NAR/GKI1012.
39. Barreau, C., Dutertre, S., Paillard, L., and Osborne, H.B. (2006). Liposome-mediated RNA transfection should be used with caution. RNA *12*, 1790. 10.1261/RNA.191706.

DETAILED PROTOCOLS

PROTOCOL FOR IVT-BASED MASSIVELY PARALLEL REPORTER ASSAY

IVT template preparation

Restriction enzyme digest with FseI

- FseI: NEB R0588S

1. Set up reaction:
    - 5 µg DNA
    - 2.5 µL 10X rCutSmart Buffer
    - 2.5 µL (5 units) FseI
    - H_20 to 25µL
2. Incubate at 37°C for 90 minutes followed by heat inactivation at 65°C for 20min
3. Run product on a gel & check for cutting
    - Uncut plasmid: supercoiled 6,895bp -> run ~4500bp
    - Cut plasmid: linear 6,895bp -> run at ~7,000bp
4. Gel extraction with Macherey-Nagel NucleoSpin Gel and PCR Clean-up Kit
    - As manufacturer's directions
5. Phenol:chloroform extraction with isopropanol precipitation:
    a. Add H20 to digest reaction to bring to 200µL
    b. Add 200µL (equal volume) phenol:chloroform:IAA (pH 8.0) & vortex/mix well
    c. Move sample to pre-spun (12,000xg for 30 seconds) phase lock light tube
    d. Centrifuge 10 minutes at top speed (room temperature)
    e. Carefully transfer top (aqueous) layer to RNase-free Eppendorf tube
    f. Add 20µL 3M NaAc & 1µL glycogen
    g. Add 200µL isopropanol (RT) & pipet up/down to mix well

h. Centrifuge 20 minutes at top speed at 4°C

i. Carefully remove supernatant from pellet

j. Wash with 1mL cold, RNase-free 75% ethanol

k. Centrifuge 5 minutes at top speed at 4°C

l. Carefully remove all supernatant

m. Repeat wash

n. Resuspend in 10µL H_20

o. Heat 15 min at 50°C to fully dissolve

***In vitro* transcription of MPRA mRNA**

RNA transcription & capping

- mMESSAGE mMACHINE T7 ULTRA Kit

1. Thaw reagents:
    - Place RNA polymerase enzyme mix on ice
    - Vortex 10X T7 Reaction Buffer & T7 2X NTP/ARCA until completely in solution
    - Once thawed, keep ribonucleotides on ice, but *buffer at room temperature*
2. Assemble reaction *at room temperature* in the following order:
    - H20 to 20µL
    - 10µL T7 2X NTP/ARCA
    - 2µL 10X T7 Reaction Buffer
    - 1µg linearized plasmid template (up to 6µL)
    - 2µL T7 Enzyme mix
    - 1µL SUPERase-In RNase Inhibitor
3. Pipette up and down gently & spin down briefly
4. Incubate at 37°C for 4hrs
5. Add 1µL TURBO DNase & mix well

6. Incubate 15 min at 37°C

PolyA tailing

1. Add tailing reagents to 20µL T7 reaction in the following order:
    - 36µL H_20
    - 20µL 5X E-PAP Buffer
    - 10µL 25mM $MnCl_2$
    - 10µL ATP solution
2. Remove 2.5µL of the reaction mixture as a “no E-PAP” control
3. Add 4µL E-PAP & mix gently
4. Incubate at 37°C for 45min, then place on ice

mRNA Purification

- Lithium chloride precipitation using reagents provided with mMESSAGE mMACHINE kit

1. Add 50µL LiCl precipitation solution
2. Mix thoroughly
3. Chill for >= 30min at -20°C
4. Centrifuge at 4°C for 15min at maximum speed
5. Carefully remove supernatant
6. Wash once with 1mL 75% ethanol
7. Centrifuge at 4°C for 15min at maximum speed
8. Carefully remove 70% ethanol
9. Resuspend in 20µL H20

Transfection of IVT mRNA

Day 1: Plating cells

- Plate 1.5 million PC3 cells per 10cm dish

Day 2: Transfect mRNA & collect cells

Transfection

- Cells should be ~75% confluent and healthy
- Transfection per 10cm plate: 8µg IVT mRNA + 40µL Lipofectamine MessengerMAX

1. Dilute 40µL MessengerMAX in 500µL Opti-MEM media and incubate at room temperature for 10 minutes
2. Dilute 8µg IVT mRNA in 500µL Opti-MEM
3. Add diluted MessengerMAX to diluted mRNA
4. Incubate at room temperature 5 minutes
5. Add 1mL of complex to each plate of cells

Wash out mRNA transfection

- One hour after mRNA transfection, wash cells to remove any remaining mRNA from the media

1. Aspirate media containing transfection reagents from cells
2. Wash cells thoroughly with PBS, then aspirate PBS
3. Add fresh media to cells

Collection

- Cells are collected 1, 3, 6, 12, and 24 hours after mRNA transfection
- Spike TRIzol used for collecting entire set of samples at once:
    - Add 0.02ng spike-in control RNA per 1mL TRIzol

1. Place cell culture plate on ice and aspirate media
2. Rinse with cold PBS, then aspirate PBS
3. Add 500µL cold PBS to plate and gently scrape cells to collect in an Eppendorf tube
4. Pellet cells by centrifugation at 350xg for 5 minutes at 4°C
5. Aspirate PBS from pellet
6. Resuspend in 1mL spiked TRIzol

7. Freeze samples at -80°C until all are ready for RNA isolation

Isolation of IVT mRNA

1. Let samples come to room temperature for 5 minutes
2. Add 200µL chloroform and shake 15 seconds
3. Transfer to pre-spun heavy phase-lock tube
    a. Pre-spin: 12,000xg for 30 seconds
4. Allow sample to stand for 2 min at RT
5. Centrifuge sample at 12,000xg for 15 min in pre-chilled (4°C) microcentrifuge.
    a. Prep fresh Eppendorf tubes with labels & 1µL glycogen each
6. Transfer 500 µl of aqueous layer to fresh tube
7. Add 500 µl (1 eq) of isopropanol. Invert tube ~10 times or until thoroughly mixed.
8. Incubate samples 10 min at RT
9. Pellet RNA at ~20,000xg for 20 min in pre-chilled (4°C) microcentrifuge.
10. Remove supernatant from pellet, wash pellet once with 1mL freshly prepared 75% EtOH & shake for 10 seconds
11. Centrifuge at 12,000xg for 10 min at 4°C.
12. Remove EtOH and dry pellet.
13. Resuspend pellet in 40 µL of nuclease-free water and freeze in -80°C.

DNase Treatment

1. Set up DNase reaction:
    - Equal volume RNA (~15µg)
    - 1.5 µL TURBO DNase
    - 7.5µL 10X DNase buffer
    - Water to 75µL
2. Invert to mix and spin down briefly

3. Incubate at 37°C for 30 minutes
4. Add 7.5µL DNase inactivation agent (from TURBO DNA-free kit) and mix well
5. Incubate at room temperature for 5 minutes, mixing 2-3 times during incubation
6. Centrifuge at 10,000xg for 90 seconds
7. Transfer supernatant to a new tube

Final RNA Purification

- Using lithium chloride purification reagent provided in mMESSAGE mMACHINE Kit

1. Add 50µL LiCl precipitation solution and 25µL water to DNase-treated RNA
2. Mix thoroughly
3. Chill for >= 30min at -20°C
4. Centrifuge at 4°C for 15min at maximum speed
5. Carefully remove supernatant
6. Wash once with 1mL 75% ethanol
7. Centrifuge at 4°C for 15min at maximum speed
8. Carefully remove 70% ethanol
9. Resuspend in 15µL H_20

PROTOCOL FOR 4SU-BASED MASSIVELY PARALLEL REPORTER ASSAY

Day 1: Plate Cells

Plating cells

- 10cm dish: 1.25 million PC3 cells per plate (yields ~20µg RNA)
- 15cm dish: 2.5 million PC3 cells per plate (yields ~40µg RNA)

Day 2: Transfect Cells

Transfection

- Cells should be ~60% confluent & healthy
- Usually done in late afternoon
- Transfection per plate:
    - 10cm dish: 8µg DNA + 24µL Fugene HD
    - 15cm dish: 16µg DNA + 48µL Fugene HD

1. Dilute 8µg/16µg DNA in 420µL/840µL Opti-MEM per plate & vortex
2. Add 24µL/48µL Fugene HD to Opti-MEM mixture *Fugene at room temperature before use*
3. Mix gently by flicking and/or inversion
4. Incubate mixture 5-10 minutes at room temperature
    a. During incubation, change media on cells to RPMI containing no antibiotics (8mL/16mL)
5. Add 400µL/800µL transfection mixture per plate

Day 3: Change media on transfected cells

Changing media

- ~16 hours post-transfection, change the media on the transfected cells to fresh, full (10% FBS) media
- If transfected at 5pm, change media at 9am next day

Day 4: 4sU treatment

s^4U treatment

- **4sU is light sensitive, all steps from here on should be done in dark as much as possible**
- Cells should be ~75% confluent & healthy
- Treat ~24 hours after changing transfection media
- Always make one -4sU control plate with +4sU samples

1. Make 50mM 4sU in water *fresh from powder*
    a. Molecular weight = 260.67 g/mol
2. Make 100uM 4sU media in 20% FBS RPMI
    a. Ex: add 120µL of 50mM 4sU to 20% FBS RPMI to make 60mL of 100uM 4sU media
3. Change media on transfected cells to 100uM 4sU media (20% FBS RPMI)

Day 5: Uridine chase, cell collection, & RNA extraction

Uridine chase

- Perform uridine chase 24 hours after 4sU treatment

1. Make 10mM uridine media from fresh powder:
    a. Molecular weight = 244.20 g/mol
    b. Use 20% FBS RPMI as media base
2. Aspirate 4sU media from cells
3. Wash twice with PBS
4. Add new 10mM uridine media to cells

Cell harvesting

1. Pre-chill 4°C centrifuge
2. Place cell culture plates on ice and aspirate media

3. Gently rinse plate once with ice cold PBS and aspirate again
4. Add 500µl/1000µl ice cold PBS to cells. Gently scrape cells from plate using a cell scraper, and transfer cell suspension to a 1.7 ml epi tube.
5. Pellet cells in a pre-chilled (4°C) centrifuge at 350xg for 5 min. Carefully aspirate PBS from cell pellet.
6. Resuspend pellet in 1 mL Trizol by gently pipetting up and down ~10 times.
    a. Trizol is toxic, use with care and in well ventilated area.
7. Keep Trizol samples on ice in dark for RNA isolation.
    a. Better to isolate immediately, but if needed, can freeze overnight in -80°C

RNA isolation

1. Let samples come to room temperature for 5 minutes in dark.
2. Add 200µl chloroform to each Trizol sample. Shake samples vigorously for 15 seconds.
3. Transfer to pre-spun heavy phase-lock tube
    a. Pre-spin: 12,000xg for 30 seconds
4. Allow sample to stand for 2 min at RT in dark.
5. Centrifuge sample at 12,000xg for 10 min in pre-chilled (4°C) microcentrifuge.
    a. Prep fresh Eppendorf tubes with labels & 1µL glycogen each
6. Transfer 500 µl of aqueous layer to fresh epi.
7. Add 500 µl (1 eq) of isopropanol and 10 µL freshly prepared 100mM DTT (final concentration: 1mM) to sample. Invert tube ~10 times or until thoroughly mixed.
8. Incubate samples 10 min at RT in dark.
9. Pellet RNA at ~20,000xg for 20 min in pre-chilled (4°C) microcentrifuge.
10. Remove supernatant from pellet, wash pellet once with 1mL freshly prepared 75% EtOH & shake for 10 seconds
11. Centrifuge at 12,000xg for 10 min at 4°C.
12. Remove EtOH and dry pellet.

a. Use gentle vacuum to remove most of ethanol from pellet
b. Spin 1 minute
c. Use gentle vacuum to remove rest of ethanol from pellet
d. Dry ~10 minutes open under Kimwipe & foil

13. Resuspend pellet in 30 µL of nuclease-free water and freeze in -80°C.

Day 6: DNase treatment, biotinylation, and enrichment

Genomic DNA depletion

1. Set up DNase reaction:
    - For 10-40µg of RNA:
        - Per 10µg of RNA: combine 10µg RNA, 1µL Turbo DNase, 5µL 10X Turbo DNase buffer, and nuclease-free water to 50µL total volume in a PCR tube
        - Ex: 20µg RNA, 2µL Turbo DNase, 10µL 10X buffer, water to 100µL volume
    - For over 40µg of RNA:
        1. Per 10µg of RNA: combine 10µg RNA and 2µL Turbo DNase
        2. Add 20µL of 10X Turbo DNase buffer
        3. Bring volume of RNA + DNase + buffer to total volume of 200µL (not scaled up for >40µg RNA)
        - Ex: 80µg RNA, 16µL Turbo DNase, 20µL 10X buffer, water to 200µL volume
2. Invert to combine and spin down quickly
3. Incubate in thermocycler for 30 minutes at 37°C
4. Add 0.2 volumes of DNase Inactivation Reagent (10µL inactivation per 50µL DNase reaction) & mix well

5. Incubate in dark at room temperature for 5 minutes, flicking/mixing 2-3 times during incubation
6. Centrifuge at 10,000xg for 1.5 minutes
7. Carefully transfer supernatant containing RNA (volume equal to step 1) to a fresh tube without disturbing the pellet of inactivation reagent.

RNeasy MinElute Cleanup

For DNase reactions <=100µL/For DNase reactions >100µL <=200µL

**Minelute columns can hold up to 45µg of RNA. For DNase reactions over this amount, split over multiple columns.

1. Bring total DNase reaction volume to 100µL/200µL
2. Add 350µL/700µL buffer RLT and mix well
3. Add 250µL/500µL 100% Ethanol to RNA sample & mix well by pipetting
4. Apply 700µL of sample to RNeasy minelute columns
5. Centrifuge 15s at 12,000xg at 4°C & discard flow-through
    a. Apply extra 700µL of sample to RNeasy minelute columns and spin again if starting with 200µL volume
6. Per column, add 500µL RPE buffer supplemented with 35 µL 1% β-me
7. Centrifuge 15s at 12,000xg at 4°C & discard flow-through
8. Add 500 µL fresh 80% ethanol
9. Centrifuge 2 min at 12,000xg at 4°C & discard flow-through
10. Move to new 2mL collection tubes & centrifuge at max speed for 5 min at 4°C
11. Transfer to 1.7mL tube and add 15µL water
12. Centrifuge 1 min at 12,000xg at 4°C

Biotinylation

- Making MTSEA-biotin-XX stock: dilute ~1mg solid biotin in dry DMF to make 1 mg/mL stock

- Stable up to 3 months at -20°C

1. Dilute 1 mg/mL biotin stock 1:20 in DMF (ex: dilute 4µL of stock into 76µL DMF) and mix by vortexing
2. Mix in a 1.7mL epi tube per 10µg of RNA:
    a. 10µg RNA
    b. 1µL 1M HEPES, pH 7.4
    c. 1µL 0.5M EDTA
    d. 50ng +4sU IVT positive control
    e. H20 to 40µL total volume
    f. 10µL of diluted biotin (to bring to 50µL total volume, 20% DMF)
3. Cover with foil and incubate at room temperature in the dark for 30 minutes with rotation.

Remove unreacted biotin

1. If necessary, bring volume of biotin reaction to 100µL by adding 50µL water to sample
2. Move sample to pre-spun heavy phase lock tube
    a. Spin: 30 seconds at 12,000xg
3. Add 1 volume (100µL) chloroform to each tube
4. Shake vigorously for 15sec & let sit 2 min
5. Centrifuge for 5 min at 12,000xg at 4°C
6. Transfer aqueous phase (~100µL) to new epi tube
7. Add 350µL buffer RLT & 250µL 100% ethanol
8. Mix well by pipetting and apply to an RNeasy minelute column
9. Centrifuge columns for 15sec at 12,000xg at 4°C & discard flow-through
10. Add 500µL buffer RPE
11. Centrifuge for 15s at 12,000xg at 4°C & discard flow-through
12. Add 500µL fresh 80% ethanol
13. Centrifuge for 2 min at 12,000xg at 4°C & discard flow-through

14. Move columns to new 2mL collection tubes
15. Centrifuge 1 min at max speed at 4°C
16. Move columns to new 1.7mL tubes & add 50µL water
17. Centrifuge for 1 min at 12,000xg at 4°C

µMacs Streptavidin Pulldown

1. Reserve a 5 µL aliquot of biotinylated RNA sample -> “total RNA”
2. Mix 45µL of biotinylation RNA (rest of sample) with 100 µL of streptavidin magnetic bead solution
3. Incubate bead-RNA solution at room temperature in the dark with gentle agitation for 15 minutes
4. With “waste” tube as collection: After incubation, place column on the magnetic stand and equilibrate column with 100µL of nucleic acid buffer. Repeat with another 100µL nucleic acid buffer
    a. Equilibrating too soon can reduce RNA recovery
5. With “pre-existing RNA” tube as collection: apply the RNA/bead solution to the column immediately after equilibration
6. With “pre-existing RNA” tube as collection: wash column once with 100µL column wash buffer
    a. 100 mM Tris-HCl (pH 7.4), 10 mM EDTA, 1M NaCl, 0.1% Tween-20
7. With “waste” tube as collection: wash column a second time with 500 µL column wash buffer
8. Elute RNA:
    1. Remove column from magnetic stand and set on “nascent RNA” tube for collection
    2. Apply 100µL 100mM DTT (freshly diluted from 1M stock) & let the DTT flow through the beads by gravity into the “nascent RNA” tube
    3. After 5 minutes, add another 100 µL 100mM DTT

µMacs Streptavidin Cleanup

Starting volumes: 5µL Total RNA, 250µL Pre-existing RNA, 200µL Nascent RNA

Option A: Modified RNeasy Minelute Cleanup

1. Adjust sample volumes:
    a. add 95µL to Total RNA to bring to 100µL volume
    - For 100µL starting volume (Total RNA)
    - For 200µL/250µL starting volume (Pre-existing/Nascent RNA)
2. Add 350µL/700µL buffer RLT + 250µL/500µL 100% Ethanol to RNA sample & mix well by pipetting
3. Apply samples to RNeasy minelute columns (700µL at a time)
4. Centrifuge 15s at 12,000xg at 4°C & discard flow-through
    a. Repeat with remaining sample volume if needed
5. Per column, add 500 µL RPE buffer supplemented with 35 µL 1% b-me (final concentration 10mM b-me)
6. Centrifuge 15s at 12,000xg at 4°C & discard flow-through
7. Add 500 µL fresh 80% ethanol
8. Centrifuge 2 min at 12,000xg at 4°C & discard flow-through
9. Move to new 2mL collection tubes & centrifuge at max speed for 5 min at 4°C
10. Transfer to 1.7mL tube and add 15µL water
11. Centrifuge 1 min at 12,000xg at 4°C

Option B: NaCl precipitation (Russo protocol)

1. Adjust sample volumes:
    - Add 45µL water and 150µL 100mM DTT to “Total RNA” fraction and incubate 5 minutes
2. Adjust NaCl to ~0.5M:

- +6µL to “pre-existing RNA”
- +25µL to “total RNA”
- +20µL to “nascent RNA”

3. Add 1µL glycogen to all fractions
4. Add 2.5 volumes of ethanol & mix well by vortexing
    - 625µL to “pre-existing RNA”
    - 500µL to “total RNA” & “nascent RNA”
5. Incubate in -80°C for at least 20 minutes to precipitate RNA
6. Centrifuge at 16,000xg for 10 minutes at 4°C
7. Carefully remove supernatant from RNA pellet and wash with 1mL fresh 70% ethanol
8. Centrifuge at 16,000xg for 10 minutes at 4°C
9. Carefully remove ethanol from RNA pellet and air dry in dark under kimwipe
10. Resuspend in 15µL water

CHAPTER 4

Elucidating consequences of a 5'UTR mutation that inserts a novel upstream ATG by CRISPR-Cas9 base editing

This chapter presents published work.

Introduction section partially adapted from: **Schuster SL**, Hsieh AC. The Untranslated Regions of mRNAs in Cancer. Trends Cancer. 2019 Apr;5(4):245-262. doi: 10.1016/j.trecan.2019.02.011. Epub 2019 Mar 22. PMID: 30961831; PMCID: PMC6465068.

Results section adapted from: Lim Y, Arora S, **Schuster SL**, Corey L, Fitzgibbon M, Wladyka CL, Wu X, Coleman IM, Delrow JJ, Corey E, True LD, Nelson PS, Ha G, Hsieh AC. Multiplexed functional genomic analysis of 5' untranslated region mutations across the spectrum of prostate cancer. Nat Commun. 2021 Jul 9;12(1):4217. doi: 10.1038/s41467-021-24445-6. PMID: 34244513; PMCID: PMC8270899.

Figure 4-1 displays data from experiments performed by Yiting Lim. Figure 4-2 displays data from experiments performed by Samantha L. Schuster. All text written by Samantha L. Schuster.

INTRODUCTION

The 5' untranslated region (5'UTR) lies upstream of the coding sequence (CDS) on messenger RNA (mRNA) and controls gene expression through a myriad of regulatory elements. Whereas the 3'UTR largely mediates mRNA stability and translation, the 5'UTR mainly affects the initiation of mRNA translation, as well as pre-mRNA processing. Regulatory elements within the 5'UTR include many sequence-specific or structural motifs, upstream open reading frames (uORFs), and internal ribosome entry sites. Importantly, each of these elements has been found to be co-opted in cancer to drive pathogenesis[1]. Therefore, we questioned how somatic mutations within the 5'UTR could affect mRNA translation in prostate cancer.

Through the use of a polysome profiling-based massively parallel reporter assay called PLUMAGE (Pooled full-length UTR Multiplex Assay on Gene Expression), Yiting Lim, then a post-doc in the Hsieh lab, measured how 545 somatic mutations found in prostate cancer tumors affected both steady-state mRNA levels and translation[2]. One of the most striking functional mutations observed in this study was within *CKS2* (CDC28 protein kinase regulatory subunit 2) (**Figure 4-1**). This mutation (chr9:91,926,143 in the *hg19* human genome) caused a significant increase in mRNA translation by mutating an ACG triplet to an ATG start codon, thereby creating a novel uORF within the 5'UTR.

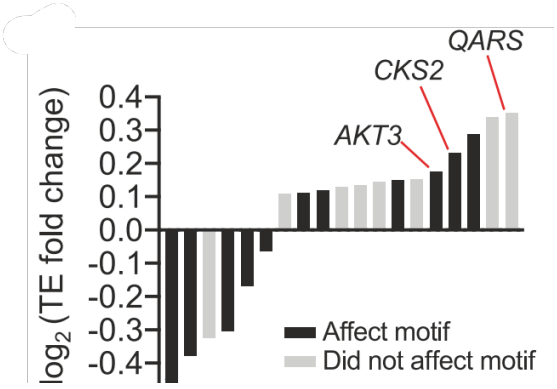

Figure 4-1: 5'UTR mutations functionally alter mRNA translation. 5′UTR mutations that significantly affect mRNA translation efficiency (Mann–Whitney U test, FDR < 0.1) shown by magnitude fold change compared to unmutated 5′UTRs. Black bars indicate mutations that impact known RNA-binding protein-binding motifs.

Upstream open reading frames are particularly unique 5'UTR elements that can regulate translation initiation of specific transcripts. A uORF consists of a translatable open reading frame with a start codon upstream of the primary CDS start codon. Under normal conditions, uORFs decrease protein expression through

induction of either translation repression or nonsense-mediate decay (NMD)[3]. A phenomenon called “leaky scanning” determines whether uORFs are recognized or disregarded based on the strength of the sequence context around the uORF AUG. If a uORF is recognized by the translation pre-initiation complex (PIC), it can repress translation initiation by preventing access to the downstream main ORF or by blocking the scanning of other PICs. If the uORF stop codon is recognized as a premature stop codon, it can trigger NMD, degrading the transcript and likewise decreasing protein expression[3].

Previous germline or somatic mutations that create, delete, or alter uORFs have been shown to contribute to cancer phenotypes. For example, a single-point mutation that creates a novel uORF in the tumor suppressor *CDKN2A* decreases CDKN2A protein levels in hereditary melanoma[4]. Likewise, examination of the *CDKN1B* 5’UTR in a patient with inherited multiple endocrine neoplasia syndrome uncovered a 4-nucleotide deletion, which induces a frameshift and lengthening of the uORF reading frame. This decreases re-initiation of translation at the *CDKN1B* primary ORF and lowers protein production of this tumor suppressor[5]. With these previous examples of functional uORF-altering mutations in mind, we set out to functionally characterize the 5’UTR mutation in *CKS2* using CRISPR-Cas9 base editing.

RESULTS

We observed that a C → T mutation in the 5′UTR of oncogenic CKS2 (chr9:91,926,143) creates a new upstream AUG (uAUG) within the 5′UTR in-frame with the main reading frame (**Figure 4-2A**). Interestingly, this uAUG increased overall translation through the CKS2 5′UTR in reporter assays (**Figure 4-1**). To determine whether this mutation would increase translation of endogenous CKS2, we engineered the C → T mutation into 293T cells using CRISPR cytosine base editing. Nucleotide-specific base editing utilizes a complex of a cytosine deaminase (APOBEC1), a Cas9-nickase, and uracil-DNA glycosylase inhibitor (UGI) (**Figure 4-2B**). The

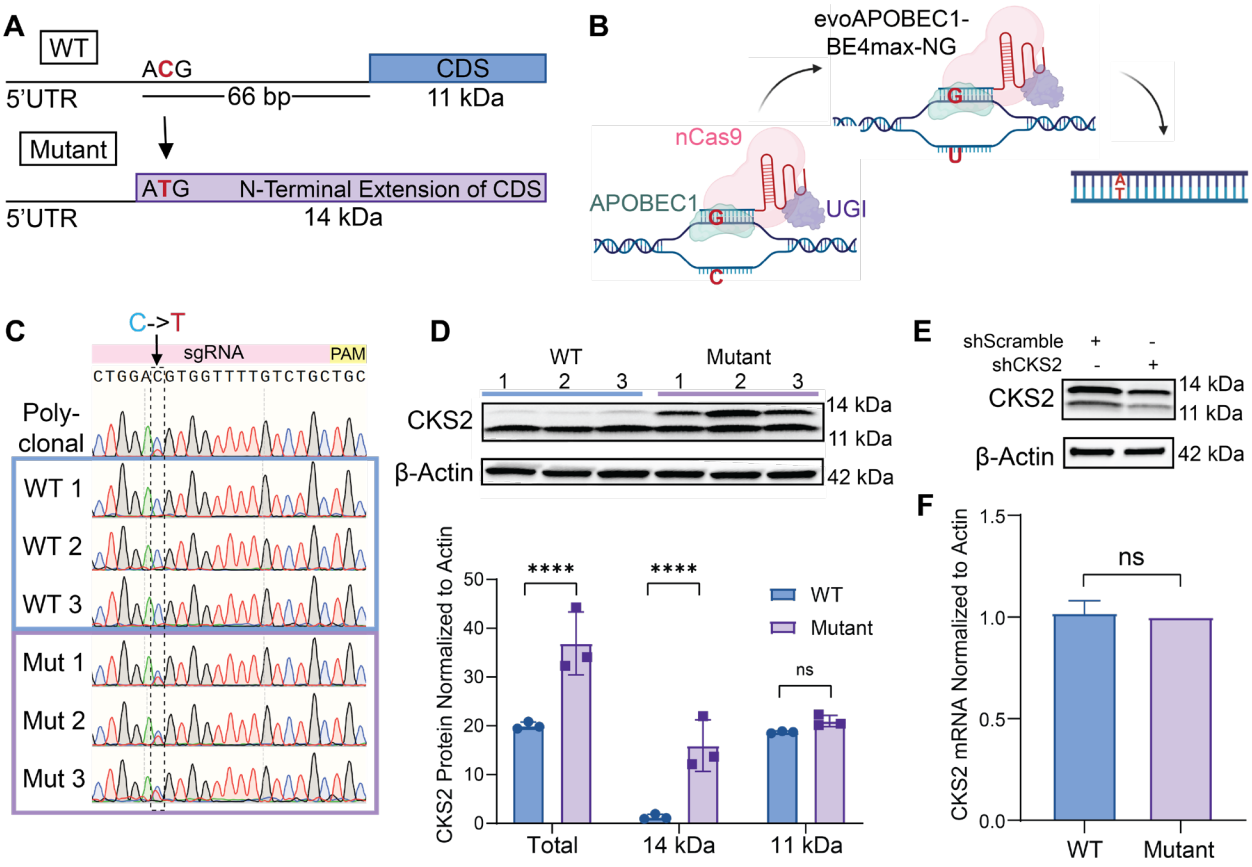

Figure 4-2: CRISPR-Cas9 base editing of a point mutation in the CKS2 5′UTR increased translation efficiency in its endogenous context. **(A)** Schematic showing wild-type (WT, top) and mutant (bottom) versions of *CKS2* transcript, including 5′UTR, normal coding sequence (CDS), and mutant N terminally extended CDS. The C to T mutation (chr9:91926143) within the 5′UTR of *CKS2* generates a start codon that extends the coding sequence of *CKS2*.The mutated nucleotide is represented in red. **(B)** Method of CRISPR-Cas9 base editing using evoAPOBEC1-BE4max-NG, which is composed of APOBEC1, a Cas9-nickase domain, and uracil-DNA glycosylase inhibitor (UGI). This base editor deaminates target cytosines to uracil, which changes the original G–C base pair into an A–T base pair after DNA repair. **(C)** Sanger sequencing traces from polyclonal population of CRISPR-transfected 293T cells and six individual single-cell clones selected from this pool for further study. The target C (blue) → T (red) mutation in the 5′UTR of *CKS2* is shown within the dashed box. **(D)** Western blot of the three WT and three CKS2 mutant clonal cell lines created by CRISPR base editing with antibodies against CKS2 and β-actin. The graph shows these results quantified using ImageJ, where each CKS2 band intensity was measured and normalized to the intensity of the corresponding β-actin loading control. Statistics show two-sided Student's t test with multiple comparisons correction using the three WT (blue) vs three CKS2 mutant (purple) biological replicates (total p value = 0.00001 and 14 kDa p value= 0.00023). Data are presented as mean values ± StDev. **(E)** Immunoblot of CKS2 5'UTR knock-in mutant cell line after shRNA knockdown of CKS2 demonstrates the specificity of the antibody. Data is representative of 2 independent experiments with similar results. **(F)** *CKS2* qPCR shows no change in mRNA levels between three WT and three mutant clonal cell lines created from CRISPR base editing. *CKS2* mRNA levels in each sample were normalized to β-actin as a loading control (n=6 biological replicates). Two-sided Student's t-test, data are presented as mean values ± SEM ns: not statistically significant.

Cas9 domain directs the complex to a target locus, where APOBEC1 deaminates a cytosine to uracil. This results in a G–U base pair that is protected from excision by UGI, which is replaced

by cellular mismatch repair to an A–T base pair in place of the original C–G[6]. Using a single guide RNA (sgRNA) targeting the CKS2 5′UTR and the evolved evoAPOBEC1-BE4max-NG base editor, we successfully generated three CKS2 5′-UTR mutant cell lines, each with one allele possessing the C → T mutation (**Figure 4-2C**). Next, we measured CKS2 protein levels by western blot analysis and observed a 2-fold increase in overall CKS2 protein in mutant cells compared to WT controls (**Figure 4-2D**). This increase was due to the novel expression of the 14 kDa extended CKS2 coding sequence without any noticeable sacrifice of the normal 11 kDa CKS2 coding sequence. Importantly, the specificity of the CKS2 antibody for the 11 and 14 kDa N-terminally extended CKS2 was confirmed by short hairpin RNA (shRNA) knockdown (**Figure 4-2E**). Since CKS2 mRNA expression did not differ between the WT and mutant cells (**Figure 4-2F**), we concluded that the CKS2 chr9: 91926143 C → T mutation was sufficient to increase mRNA-specific translation. This observation corroborates our PLUMAGE findings and demonstrates that 5′UTR mutations can coordinately impact mRNA translation by altering RNA-based *cis*-regulatory elements in their endogenous context.

DISCUSSION

CRISPR-Cas9 base editing is an ideal method to assess the function of single-nucleotide mutations in their endogenous genomic contexts. Here, we used this technology to validate a 5'UTR point mutation found in an MPRA to increase translation. The use of CRISPR base editing allows a much more thorough investigation of how this 5'UTR mutation is acting. By altering the *CKS2* genomic loci, we have been able to demonstrate that this single-nucleotide mutation creates a functional upstream AUG that causes an N-terminal extension of the CKS2 protein.

This is particularly interesting because novel upstream start codons in the 5'UTR would generally be predicted to decrease translation of the downstream ORF, whereas this mutation actually increases overall translation. This is likely because, rather than creating a new out-of-frame uORF that competes with the downstream ORF, this mutation simply extends the existing main ORF. Still, further experiments would have to be performed to understand exactly how the change from two wildtype alleles to one wildtype and one uAUG allele does not decrease the production of wildtype protein, but only increases the translation of N-terminally extended CKS2. It is possible that the translation of the uORF creates a positive feedback loop increasing the translation of the wildtype transcript as well, or that there are compensatory mechanisms that ensure that even half the amount of wildtype mRNA produces normal amounts of protein. The poor sequence context around the new uAUG (T at -3 and +4) may mean that few ribosomes actually begin translating at the upstream start codon, allowing for normal levels of the 11kDa protein. If the 14kDa protein is translated less, but more stable, this may account for the observed protein expression. Closely measuring the translation initiation rates from wildtype versus mutant *CKS2* mRNA might help answer these questions.

It is currently unknown how the extended 14kDa CKS2 protein functions in the cell as compared to wildtype CKS2. Normal CKS2 protein interacts with cyclin-dependent kinases (CDKs) to control cell cycle progression[7,8]. Overexpression of CKS2 in mammary or breast cancer cells inhibits the DNA-damage cell cycle checkpoint, allowing increased cellular proliferation despite DNA damage[9]. This seems to be a common mechanism of stress resistance in cancer, as high expression of CKS2 has been found in several cancers, where it is necessary for proliferation and migration of cells[10–12]. From these known CKS2 functions, it would be expected that a 5'UTR mutation that increases CKS2 overall protein expression might be pro-tumorigenic by increasing either proliferation or migration. However, the addition of an extra 22 amino acids to the N-terminus could alter the function of the protein. It would be

interesting to know whether this mutant CKS2 interacts with CDKs, localizes normally within the cell, or has any altered function regarding cell cycle control. This N-terminal extension could even potentially create a neoantigen in the cancer cell, altering immunologic responses within the tumor. Further studies would have to be performed to answer these mechanistic questions. Nevertheless, our experiments have demonstrated that a single-nucleotide mutation found in a prostate cancer tumor has the ability to increase mRNA translation, where the heterozygous mutation not only increases wildtype translation but also produces a novel N-terminally extended CKS2 protein with potentially neo-oncogenic functions.

MATERIALS & METHODS

CRISPR base editing

Plasmid to express CKS2-targeting sgRNA was cloned using the Q5 Site-Directed Mutagenesis Kit (NEB) according to the manufacturer's instructions. The pFYF1320 sgRNA expression plasmid was used as a template for Q5 mutagenesis PCR to replace the existing sgRNA sequence with the CKS2-targeting sgRNA sequence (CTGGACGTGGTTTTGTCTGC). 293T cells were plated in six-well plates at 375,000 cells/well, incubated at 37°C overnight, and transfected with 1,125 ng evoAPOBEC1-BE4max-NG (Addgene: 125616), 375 ng CKS2 sgRNA expression plasmid, and 30 ng pMaxGFP using Fugene HD (Promega) according to the manufacturer's protocol. At 72 hours post transfection, cells were washed with PBS, harvested with 0.05% Trypsin-EDTA (Gibco), and centrifuged at 400 × g for 5 min. This cell pellet was resuspended in PBS and sorted using flow cytometry for live, singlet, GFP+ cells on a Sony SH800 sorter. GFP+ cells were plated using limiting dilution in 10 cm plates to grow out single-cell clones. After clones had grown sufficiently (~3 weeks), DNA was extracted using Zymo's MicroPrep Quick-DNA Kit, the CKS2 locus PCR amplified using the Phusion High Fidelity

Master Mix (Thermo Fisher) in a 25 µL reaction and primers according to the manufacturer's protocol. PCR products were then Sanger sequenced to determine if the intended CKS2 mutation (chr9: 91926143 C → T) had been introduced. Six individual clonal cell lines were chosen for further testing: three mutant clones each mutated at one of two *CKS2* alleles, and three WT clones that were not mutated.

shRNA knockdown

An shRNA construct targeting CKS2 (hairpin sequence: TG CTGTTGACAGTGAGCGAACAGCAACAGAGCTCAGTTAATAGTGAAGCC ACAGATGTATTAACTGAGCTCTGTTGCTGTGTGCCTACTGCCTCGGA) in the pGIPZ backbone was obtained as a gift from the Paddison Lab (Fred Hutchinson Cancer Center). The shCKS2 construct was transfected into the CKS2 mutant 2 clonal cell line created by CRISPR base editing due to its high endogenous expression of CKS2. Transfection was performed by plating 375,000 cells per well in six-well plates, incubating overnight at 37°C, and next day adding 1.5 µg of plasmid DNA with 4.5 µL Fugene HD (Promega) according to the manufacturer's instructions. At 24 hours post transfection of shCKS2, cells were harvested and lysed for Western blotting.

Western blotting

A total of 1×10^6 cells were collected from each CKS2 WT and mutant 293T cell line and lysed in RIPA lysis buffer (Thermo Scientific) supplemented with 10% Complete Mini Protease Inhibitor (Sigma) and 10% PhosSTOP Phosphatase Inhibitor (Roche). After incubating on ice for 30 min, lysates were centrifuged at 13,000 × g for 10 min at 4°C. The supernatant was collected and protein concentration was measured using a Bradford assay (Bio-Rad). Twenty-five to 50 µg of extract per cell line was separated by sodium dodecyl sulfate-PAGE and transferred onto PVDF membranes for immunoblot analysis. Primary antibodies used were CKS2 (Abcam,

catalog #: ab155078; dilution: 1:1000) and β-actin (Sigma-Aldrich; catalog #: A5316; dilution: 1:1000).

Quantitative PCR

RNA was extracted from ~500,000 cells per 293T CKS2 WT or mutant cell line using the RNeasy Plus Kit (Qiagen) following the manufacturer's protocol. cDNA was synthesized using 500 ng RNA and iScript RT Supermix (Bio-Rad) or iScript NRT Supermix for negative controls. qPCR was performed using SsoAd- vanced Universal SYBR Green Supermix (Bio-Rad) on 1 µL of each cDNA, no reverse transcription, and no-template control sample in triplicate using primers specific to *CKS2* and *β-actin* as a housekeeping control.

Primers used:

Primer Name	Primer Sequence
CKS2 qPCR Forward primer	CACTACGAGTACCGGCATGTT
CKS2 qPCR Reverse primer	ACCAAGTCTCCTCCACTCCT
β-actin qPCR Forward primer	AAATCTGGCACCACACCTTC
β-actin qPCR Reverse primer	GGGGTGTTGAAGGTCTCAAA
CKS2 Q5 mutagenesis Forward primer	TTTTGTCTGCGTTTTAGAGCTAGAAATAGCAAG
CKS2 Q5 mutagenesis Reverse primer	CCACGTCCAGGGTGTTTCGTCCTTTCCAC
CKS2 PCR Forward primer	ACTTCCGCAGAAGGTGATTG
CKS2 PCR Reverse primer	TACTCGTAGTGTTCGTCGAAGT

CHAPTER 4 REFERENCES

1. Schuster, S.L., and Hsieh, A.C. (2019). The Untranslated Regions of mRNAs in Cancer. Trends in Cancer *5*, 245–262. 10.1016/j.trecan.2019.02.011.
2. Lim, Y., Arora, S., Schuster, S.L., Corey, L., Fitzgibbon, M., Wladyka, C.L., Wu, X., Coleman, I.M., Delrow, J.J., Corey, E., et al. (2021). Multiplexed functional genomic analysis of 5' untranslated region mutations across the spectrum of prostate cancer. Nat. Commun. *12*. 10.1038/S41467-021-24445-6.
3. Barbosa, C., Peixeiro, I., and Romão, L. (2013). Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. PLoS Genet. *9*, e1003529. 10.1371/journal.pgen.1003529.
4. Liu, L., Dilworth, D., Gao, L., Monzon, J., Summers, A., Lassam, N., and Hogg, D. (1999). Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. Nat. Genet. *21*, 128–132. 10.1038/5082.
5. Occhi, G., Regazzo, D., Trivellin, G., Boaretto, F., Ciato, D., Bobisse, S., Ferasin, S., Cetani, F., Pardi, E., Korbonits, M., et al. (2013). A Novel Mutation in the Upstream Open Reading Frame of the CDKN1B Gene Causes a MEN4 Phenotype. PLoS Genet. *9*, e1003350. 10.1371/journal.pgen.1003350.
6. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420–424. 10.1038/nature17946.
7. Frontini, M., Kukalev, A., Leo, E., Ng, Y.M., Cervantes, M., Cheng, C.W., Holic, R., Dormann, D., Tse, E., Pommier, Y., et al. (2012). The CDK subunit CKS2 counteracts CKS1 to control cyclin A/CDK2 activity in maintaining replicative fidelity and neurodevelopment. Dev. Cell *23*, 356–370. 10.1016/J.DEVCEL.2012.06.018.
8. Spruck, C.H., De Miguel, M.P., Smith, A.P.L., Ryan, A., Stein, P., Schultz, R.M., Lincoln, A.J., Donovan, P.J., and Reed, S.I. (2003). Requirement of Cks2 for the first metaphase/anaphase transition of mammalian meiosis. Science *300*, 647–650. 10.1126/SCIENCE.1084149.
9. Liberal, V., Martinsson-Ahlzén, H.S., Liberal, J., Spruck, C.H., Widschwendter, M., McGowan, C.H., and Reed, S.I. (2012). Cyclin-dependent kinase subunit (Cks) 1 or Cks2 overexpression overrides the DNA damage response barrier triggered by activated oncoproteins. Proc. Natl. Acad. Sci. U. S. A. *109*, 2754–2759. 10.1073/PNAS.1102434108.
10. Esser, L.K., Branchi, V., Shakeri, F., Simon, A.G., Stephan, C., Kristiansen, G., Buness, A., Schorle, H., and Toma, M.I. (2022). Overexpression of Parkin in clear cell renal cell carcinoma decreases tumor aggressiveness by regulating CKS2 levels. Int. J. Oncol. *60*. 10.3892/IJO.2022.5310.
11. Yu, K., Ji, Y., Liu, M., Shen, F., Xiong, X., Gu, L., Lu, T., Ye, Y., Feng, S., and He, J. (2022). High Expression of CKS2 Predicts Adverse Outcomes: A Potential Therapeutic Target for Glioma. Front. Immunol. *13*. 10.3389/FIMMU.2022.881453.
12. Chen, M., Zhao, Z., Wu, L., Huang, J., Yu, P., Qian, J., Ni, T., and Zhao, C. (2022). E2F1/CKS2/PTEN signaling axis regulates malignant phenotypes in pediatric retinoblastoma. Cell Death Dis. *13*. 10.1038/S41419-022-05222-9.

CHAPTER 5: CONCLUSIONS & FUTURE DIRECTIONS

My dissertation research has aimed to answer the question “How do 3’UTR mutations affect gene expression and cancer?” Underneath this question lies the important and potentially controversial question of “Do 3’UTR mutations matter at all?” Of course, these are both multifaceted and difficult questions. My research concludes that 3’UTR mutations can affect multiple levels of gene expression, as well as cancerous phenotypes like cellular growth under stress, thereby demonstrating their significance. However, the definition of “significance” can vary - are we talking about discovering a molecular phenomenon for the first time or changing patient lives? In this concluding chapter, I will address the contributions my research makes to the field of functional non-coding cancer genomics, the question of how clinically applicable such conclusions are, how tool building may bring us closer to bridging the gaps between molecular and clinical significance, and future directions stemming from these areas.

5.1 Putting 3’UTR mutations on the cancer genomics map

Historically, the untranslated regions of mRNA have been generally discounted in the field of cancer genomics. This stems in large part from our inability to easily assign function to non-coding genomic alterations. In Chapter 2 of this dissertation, I have shown that 3’UTR mutations are not only widespread, but functional on both the molecular and cellular levels, with considerable relevance to cancer. While previous massively parallel reporter assays (MPRAs) have established the importance of various 3’UTR regulatory elements (Chapter 1, Table 1-1), mine are the first to directly test how mutations found in a specific patient population affect post-transcriptional gene regulation. Importantly, I not only discovered hundreds of patient-based mutations that significantly affect either mRNA translation or stability, but my analysis also showed that the presence of these mutations in patients is associated with worse prognosis.

As expected, most of the functional mutations discovered by our MPRAs were found in sites of predicted RBP or miRNA binding. However, many other mutations that altered regulatory motifs were not functional in changing gene expression. Therefore, more details are necessary to fully discern which motif-altering mutations are functional. Perhaps combining predicted motifs with sequence conservation data may better separate functional from passenger mutations. Functional RBP and miRNA binding analysis such as eCLIP data may also help narrow down predicted sites to those of functional importance. Interestingly, we found many more functional mutations in RBP motifs than miRNA sites; however, it is unclear whether this is simply due to technical differences in how we classified RBP and miRNA sites or whether it is representative of true biology. In some respects, it would be expected that miRNA binding sites are more sensitive to point mutation than RBP motifs, as miRNAs bind based off of base-pairing alone, while RBP motifs can often be degenerate. This could potentially be answered by building an MPRA based on mutated RBP and miRNA motifs to see which are more sensitive. Importantly, these answers may likely change depending on the genomic context and the particular RBP/miRNA in question; therefore, it would be important to use a wide variety of RBP and miRNA sites as found in their endogenous contexts. Such experiments would increase our overall understanding of 3'UTR point mutations and potentially help predict the function of *de novo* mutations.

It is important to note the difficulties of using MPRAs to study gene expression. As discussed in Chapter 3, it took extensive optimization to find a method to study mRNA stability reliably. Even with such optimization, technical variation between replicates remained an issue that decreased the overall statistical power of these tools and the conclusions we were able to make from them. From the tool-building that I have performed throughout my research, including adapting polysome profiling for my MPRA, testing multiple mRNA stability methods, and using three generations of CRISPR editing techniques, I have learned that using the right,

well-optimized and controlled tools for a particular purpose is essential for obtaining robust data and making conclusions. In this process, reliable controls are vital for benchmarking new techniques. It is also important to understand the causes of variability in data, as knowing the source can assist with mitigating it, either by building better normalization techniques into experiments or otherwise altering the method.

Different experimental techniques have various strengths and limitations, which underscores the importance of using orthogonal methods for validating our MPRA hits. The most convincing method used to validate top mutations from the MPRAs was CRISPR base editing of the ZWILCH and IGF1R mutations. These experiments prove that 3'UTR mutations in their endogenous contexts can affect protein expression and cellular growth. Importantly, these two 3'UTR mutations were only introduced into one genomic allele, and yet still caused significant molecular and cellular effects. This argues for the relevance of subclonal non-coding mutations in tumors, as even mutations with a low variant allele frequency (making up a relatively small proportion of the total tumor, as most of the 3'UTR mutations in our dataset did) may be able to cause significant effects. CRISPR base editing takes an important step beyond MPRAs, identifying not only molecular changes but cellular consequences of point mutations. It would be exciting to take these even further to discover whether such *in vitro* cellular growth advantages are recapitulated *in vivo* if the CRISPR-edited cells were put into mice.

Using a combination of novel massively parallel reporter assays, CRISPR base editing, and computational analyses, we have determined that 3'UTR mutations in cancer patients can affect mRNA stability, translation efficiency, and cancer growth, solidifying that these mutations should not continue to be ignored in cancer genomics.

5.2 The challenge of patient applicability

When I discuss my research with friends and family, one of their most common questions is “so what does your research mean for cancer patients?” It’s a difficult question to answer, because as of yet research into 3’UTR mutations or even much of non-coding genomics isn’t directly relevant to patients in a therapeutic sense. Some of the major issues in current cancer treatment, particularly in advanced prostate cancer, are tumor heterogeneity and the development of drug resistance. Both of these issues arise from our incomplete understanding of and therefore inability to predict the myriad of ways a cell can be transformed, increase its proliferation, and circumvent normal growth inhibition pathways. In my research, I have discovered that 3’UTR mutations can significantly affect oncogenic pathways and cellular growth, providing novel mechanisms of cancer progression. This supports the idea that to fully understand mechanisms of drug resistance and oncogenesis, we must also take into consideration the non-coding genome. Perhaps, if we understand the function of non-coding alterations to the genome, we can find more actionable novel cancer drivers in tumors. For example, a tumor containing the IGF1R 3’UTR mutation I studied may be a good candidate for IGF1R-targeting therapies, and the ZWILCH mutation-harboring patient may benefit from cisplatin treatment.

One of the major issues with this idea, however, is that non-coding mutations do not seem to be very recurrent across patients. The ZWILCH and IGF1R 3’UTR mutations I studied were each found only in one patient, making it difficult to say just how practical knowledge of their functionality is. The PCAWG study used 2,658 whole genome sequencing samples to call non-coding variants across pan-cancer genomes and still found very few hotspot mutations with recurrency across patients[1]. As could be expected, we observe the same in our cohort of 185 mCRPC patients, where only 116 of 14,497 3’UTR mutations called were found in more than one patient, and only 11 mutations were found in 3 or 4 patients (**Figure 5-1A**). Unexpectedly, we did not find a correlation between recurrence and function in our dataset, where those

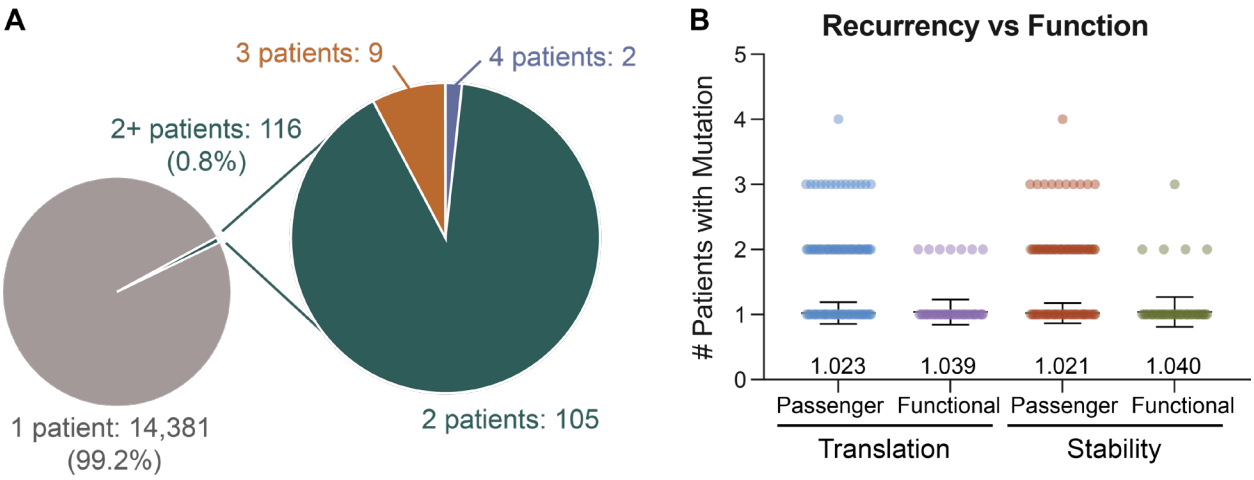

Figure 5-1: Recurrency and function in 3'UTR mutations. **(A)** Proportions of 3'UTR mutations called across 185 patient dataset found in more than one patient. **(B)** Number of patients harboring each 3'UTR mutation, separated by whether the mutation was functional at either the translation or stability levels. The average # of patients for each set is shown above the x-axis.

116 mutations found in multiple patients were no more likely to be functional than the thousands found in only one patient (**Figure 5-1B**). While this does not discount the fact that many of these mutations found only in one patient functionally affect mRNA translation, stability, or even cellular growth, it does make applying this knowledge difficult, as we would not necessarily expect new patients to carry any of these tested mutations. Current precision genomics (such as the use of PARP inhibitors) is based off using small, specialized sequencing panels to find common mutations in new tumors, but the non-coding space seems to be littered with low-frequency passenger mutations with only occasional functional variants. While it would be great to suggest every cancer patient undergo whole genome sequencing to determine all of the somatic mutations their tumor harbors, not only would this be prohibitively expensive, but we don't yet have the knowledge to discern passenger from functional mutations without individually testing them. Therefore, there are still significant gaps in our knowledge preventing non-coding genomics from being broadly clinically relevant.

5.3 Towards a unified model of UTR-based mRNA dynamics

If our goal is being able to sequence a tumor and determine that a patient is a good candidate for a PARP inhibitor from finding a 3'UTR mutation in BRCA2, we must unify the field's knowledge about 3'UTR regulatory elements into a predictive model. Many MPRAs studying the 3'UTR have been performed (Chapter 1, Table 1-1), each focusing on various aspects of gene regulation (mRNA stability, translation, protein expression, alternative polyadenylation, etc.) and testing 3'UTR sequences from different natural or synthetic sources. If this knowledge could be combined into a deep learning algorithm, we may have the beginnings of a predictive model of 3'UTR dynamics that could confidently assign function to novel 3'UTR mutations. This model would combine many of the ideas analyzed and proposed here, such as the importance of sequence conservation, predicted regulatory motifs, known sites of RBP binding from eCLIP data, RNA structure predication, etc. Many published MPRAs have already built predictive models from their experimental results, which suggests the possibility of expanding such individual analyses to incorporate multiple datasets[2–6]. Though it is likely difficult to unite such disparate datasets collected using different methods and readouts, all of this published information would be useful in piecing together a robust model of mRNA dynamics. Common themes of the importance of RNA structure and conservation are already becoming apparent. Such a model could be used not only for better understanding the role of UTR variants in disease, but also in precise design of synthetic gene expression systems.

5.4 Future Directions: CRISPR base editing screening for the non-coding genome

MPRAs are useful tools to discover how regulatory elements affect gene expression; however, a major drawback is that they cannot be used to delineate cellular consequences of the tested mutations. CRISPR base-editing screens present a logical step from MPRAs, moving from high-throughput discovery of 3'UTR mutation molecular effects to cellular function. It is exciting to consider what new questions can be answered by this relatively new technology.

CRISPR base editing screens allow the discovery of how single-nucleotide mutations genome-wide affect cellular phenotypes. The published studies so far have measured the enrichment of genomically-integrated sgRNAs over time to determine how introduced mutations affect cellular growth[7,8]. Though these screens are proving to be powerful tools to define the cellular effects of point mutations genome-wide, there are some limitations to consider. Unlike CRISPR-knockout screens, base editing screens lack the ability to use multiple sgRNAs to mutate the same target, as each mutation of interest is usually only achievable from one sgRNA. This reduces the confidence and power of any given analysis by essentially removing technical replicates. Another issue is that each sgRNA in base editing can potentially introduce multiple mutations. In CRISPR-KO, it is reasonable to assume that a given sgRNA knocks out the gene of interest, particularly if multiple sgRNAs for the same gene produce the same cellular consequence. In CRISPR base editing, however, the use of only one sgRNA per target, ambiguity of the introduced mutation, and overall inefficiency of editing makes it much more difficult to pinpoint exactly what genomic change is causing a growth advantage for the cell. Therefore, simply measuring the change in sgRNAs in the cell population over time is not a perfect readout for what genomic alterations are affecting cellular growth.

Further iterations of this technology will likely start addressing these issues. Every new edition of base editing technology improves its efficiency and selectivity, mitigating the question of whether each sgRNA is actually inserting the specific mutation of interest[9]. However, to fully resolve the incongruency between sgRNA insertion and actual gene editing, the analysis of what mutations confer growth advantages must be measured not by enrichment of integrated sgRNAs, but by sequencing the mutations themselves. Of course, with sgRNAs targeting mutations throughout the endogenous genome, it is difficult to capture all the introduced mutations and analyze how their proportions are changing on a cell-by-cell basis. Single-cell whole genome sequencing may solve this issue. Though such technologies are still in

development, some studies have successfully identified single nucleotide variants genome wide[10,11]. Therefore, this method could be used to determine how many cells contain each mutation of interest at different time points along a growth period to definitively determine which mutations increase or decrease cellular growth. On a smaller scale, CRISPR base editing screens with more limited targets (eg: deep mutational scanning of a few genes, or a mini-screen of mutations of interest from an MPRA) may be able to simply use custom targeted amplicon sequencing to pull out the genomic sites of interest and sequence them at high depth to identify mutation proportions at different time points. Either way, base editing screening will prove to be a valuable tool in determining the functional consequences of non-coding mutations, particularly if such improvements are made to discern exactly what mutation is causing each functional outcome.

It is intriguing to also consider combining CRISPR base editing screens measuring how point mutations affect cellular growth with measurements of gene expression changes. While MPRAs are limited by their inability to measure changes in cellular function, current CRISPR base editing screens have the opposite problem- the molecular mechanisms of how a sgRNA is causing a cellular change are unclear. Perhaps if both DNA-seq and gene expression sequencing (RNA-seq, polysome-seq, etc.) were used together after CRISPR base editing, we would be able to understand the entire cascade of gene expression: from endogenous genomic mutation to molecular change in gene expression to cellular change in growth.

5.5 Future Directions: Further mechanistic exploration of 3'UTR mutations

My dissertation research utilized two MPRAs to determine how 3'UTR mutations affect mRNA translation and stability, as these are the two major aspects of gene expression mediated by the 3'UTR. However, these are not the only facets of mRNA dynamics that the 3'UTR can affect. It is likely that 3'UTR mutations may also be able to change splicing, alternative polyadenylation, RNA modifications, mRNA localization, or RNA structure. In some ways,

measuring how 3'UTR mutations affect APA, epitranscriptomics, and structure are redundant to measuring changes in mRNA stability and translation. If a mutation affects RNA structure, it likely affects gene expression through altering stability or translation, and therefore the functional mutations identified in either assay would be the same. Even so, further assays designed to measure each of these could provide valuable insight as to the molecular mechanisms of the gene expression changes I observed in my MPRAs. Such assays may also help explain the 20-25% of functional mutations we discovered that were not associated with any RBP, miRNA, or polyA signal that I analyzed. It is likely that with further analysis, we could determine that many of these are acting through structural changes, dORFs, m^6A alterations, or other known or novel regulatory elements.

The 3'UTR's ability to mediate mRNA localization is especially intriguing, as it's possible that alterations to the 3'UTR can change localization and thereby cause cellular changes without affecting gene expression[12]. Therefore, understanding such changes could unveil an entirely new set of functional 3'UTR mutations that affect cellular processes despite not altering gene expression. It is likely that these mutations, if they exist, would either work by changing APA, 3'UTR splicing, or RBP motifs, as mRNA localization would likely have to be mediated by a change in RBP binding.

One limitation of both my MPRAs and many previously published MPRAs is that they have been conducted in cell culture in one or two cell types. Therefore, it is unclear the extent to which 3'UTR mutations or regulatory elements function differently across cell types. For example, differential expression of 3'UTR-interacting factors like RBPs and miRNAs would likely change the effects of certain mutations that alter these motifs. Further, cells in culture can differ greatly from cells in actual tissue, and very few *in vivo* experiments have been done analyzing 3'UTR dynamics. Therefore, future experiments using multiple cell types or animal models

would greatly expand our knowledge of cell type-specific 3'UTR function, which would be of particular importance when trying to use this knowledge to help patients.

Another question that would be particularly interesting is determining which types of molecular changes can have the largest effects on gene expression. For example, do mutations that change APA, therefore affecting the presence of many regulatory elements at once, have a much greater average affect than those that change a single miRNA site? While very few of the functional mutations I identified in my MPRAs were predicted to affect polyA signals (**Chapter 2, Figure 2-2E and 2-3E**), this may be due to our technical definition of these sites, which are much more complex than could be captured in our simple *in silico* analysis. These answers would contribute greatly to our ability to predict what 3'UTR mutations would have the greatest effect size and use this information to decide the clinical relevance of mutations, as discussed above.

5.6 Concluding Remarks

My research demonstrates the significance of 3'UTR mutations on molecular gene expression, cellular growth, and even patient outcomes, setting the stage for the discovery of further non-coding drivers of cancer progression. Building on this foundation of 3'UTR research using novel MPRAs designed to uncover additional molecular mechanisms of mutation function, CRISPR base editing able to assign cellular function to single-nucleotide mutations, and combining this information into a deep learning algorithm may bridge the gap between our current understanding of non-coding genomics and application of this knowledge to cancer patients.

CHAPTER 5 REFERENCES

1. Rheinbay, E., Nielsen, M.M., Abascal, F., Wala, J.A., Shapira, O., Tiao, G., Hornshøj, H., Hess, J.M., Juul, R.I., Lin, Z., et al. (2020). Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nat. 2020 5787793 *578*, 102–111. 10.1038/s41586-020-1965-x.
2. Cottrell, K.A., Chaudhari, H.G., Cohen, B.A., and Djuranovic, S. (2018). PTRE-seq reveals mechanism and interactions of RNA binding proteins and miRNAs. Nat. Commun. *9*, 301. 10.1038/s41467-017-02745-0.
3. Griesemer, D., Xue, J.R., Reilly, S.K., Ulirsch, J.C., Kukreja, K., Davis, J.R., Kanai, M., Yang, D.K., Butts, J.C., Guney, M.H., et al. (2021). Genome-wide functional screen of 3′UTR variants uncovers causal variants for human disease and evolution. Cell, 1–14. 10.1016/j.cell.2021.08.025.
4. Rabani, M., Pieper, L., Chew, G.-L., and Schier, A.F. (2017). A Massively Parallel Reporter Assay of 3′ UTR Sequences Identifies In Vivo Rules for mRNA Degradation. Mol. Cell *68*, 1083-1094.e5. 10.1016/J.MOLCEL.2017.11.014.
5. Siegel, D.A., Le Tonqueze, O., Biton, A., Zaitlen, N., and Erle, D.J. (2022). Massively parallel analysis of human 3′ UTRs reveals that AU-rich element length and registration predict mRNA destabilization. G3 Genes|Genomes|Genetics *12*. 10.1093/G3JOURNAL/JKAB404.
6. Bogard, N., Linder, J., Rosenberg, A.B., and Seelig, G. (2019). A Deep Neural Network for Predicting and Engineering Alternative Polyadenylation. Cell *178*, 91-106.e23. 10.1016/J.CELL.2019.04.046.
7. Hanna, R.E., Hegde, M., Fagre, C.R., DeWeirdt, P.C., Sangree, A.K., Szegletes, Z., Griffith, A., Feeley, M.N., Sanson, K.R., Baidi, Y., et al. (2021). Massively parallel assessment of human variants with base editor screens. Cell *184*, 1064-1080.e20. 10.1016/J.CELL.2021.01.012.
8. Cuella-Martin, R., Hayward, S.B., Fan, X., Chen, X., Huang, J.W., Taglialatela, A., Leuzzi, G., Zhao, J., Rabadan, R., Lu, C., et al. (2021). Functional interrogation of DNA damage response variants with base editing screens. Cell *184*, 1081-1097.e19. 10.1016/J.CELL.2021.01.041.
9. Wang, J.Y., and Doudna, J.A. (2023). CRISPR technology: A decade of genome editing is only the beginning. Science (80-.). *379*. 10.1126/SCIENCE.ADD8643/ASSET/F32937F7-FE8D-4FDA-B0EC-7DB7B79A1B94/ASSETS/IMAGES/LARGE/SCIENCE.ADD8643-F4.JPG.
10. Xing, D., Tan, L., Chang, C.H., Li, H., and Xie, X.S. (2021). Accurate SNV detection in single cells by transposon-based whole-genome amplification of complementary strands. Proc. Natl. Acad. Sci. U. S. A. *118*. 10.1073/PNAS.2013106118/-/DCSUPPLEMENTAL.
11. Zhang, L., Dong, X., Lee, M., Maslov, A.Y., Wang, T., and Vijg, J. (2019). Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. Proc. Natl. Acad. Sci. U. S. A. *116*, 9014–9019. 10.1073/PNAS.1902510116/-/DCSUPPLEMENTAL.
12. Berkovits, B.D., and Mayr, C. (2015). Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. Nature *522*, 363–367. 10.1038/nature14321.