High-throughput functional characterization of cancer specific alternative polyadenylation

Austin Max Gabel

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Robert K. Bradley, Chair

Daphne Avgousti

Celeste Berg

Lea Starita

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

High-throughput functional characterization of cancer specific alternative polyadenylation

Austin Max Gabel

Chair of the Supervisory Committee:
Robert K. Bradley
Department of Genome Sciences

RNA dysregulation is a ubiquitous feature of human cancers and misexpression of even a single RNA isoform is sufficient to drive cancer. Among types of RNA processing, alternative cleavage and polyadenylation (APA), a process whereby a messenger RNA is cleaved at a polyadenylation signal sequence and a poly(A) tail is added, stands out as poorly functionally studied yet broadly dysregulated in all cancer subtypes studied. Global patterns in APA are associated with patient outcomes, response to chemotherapy, and patterns of immune infiltration, all of which point to a functional role of APA in human cancers. An absence of molecular tools has led to a large number of cataloged APA events that have no functional understanding and a reliance on a sufficient number of adjacent healthy control tissues has led to an outsized focused on a small number of human cancer subtypes. In this dissertation, I address these problems and describe a computational framework to assess global patterns in poly(A) site selection that allows for intra and inter dataset comparisons without normalization to a healthy tissue control and a novel, CRISPR/Cas9 based method to functionally screen APA events in a live, immunocompetent host. I demonstrated that 3′ UTR lengthening in a single gene, *ATG7*, reduces melanoma tumor engraftment and boosts tumor immune evasion in both mouse and human melanoma. I then further characterize pan-cancer alterations in APA and identify that colorectal adenocarcinoma displays significantly distinct patterns in poly(A) site selection compared to all other datasets analyzed. I provide compelling evidence from human colorectal adenocarcinoma samples and genetically modified human colon organoid samples that adenomatous polyposis coli (APC) is an RNA binding protein that regulates poly(A) site selection. Together, this work demonstrates that APA is pervasive in human cancers and represent a novel source of potential targets for clinical therapy development.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

This dissertation and its encompassed works were only possible thanks to the support of many individuals. The support and mentorship of my advisor, Robert Bradley, has been critical to all the work included here as well as other works in my graduate education. I would also like to acknowledge the incredible mentorship I received from James D. Thomas who spent a large amount of time training me in both experimental and computational approaches to genomics, without which, this work would not have been possible. I would also like to thank all of the incredible lab members from the Bradley Lab including Guo-Liang Chew, Jake Polaski, Joey Pangallo, Mao Pineda, Emma J. DeNeef, James Thomas, Khrystyna North, Dylan Udy, Emma Hoppe, Andrea Belleville, Siegen McKellar, Taylor Nicholas, Toshi Banjo, Erik Kimble, and Edie Crosse, all of whom have been supportive as colleagues and as friends. I would like to thank my committee members, Celeste Berg, Daphne Avgousti and Lea Starita for their insightful feedback and encouragement throughout the past three years. I would also like to thank my classmates in the UW MSTP program, in particular Zak Yaffe, Sriram Pendyala and Maya Lewinsohn, who have all been incredible friends throughout the past six years together. My classmates in the Genome Sciences PhD program have also played an important role in my experience in graduate education as well as the multitude of post-docs, graduate students and administrative staff at the UW Genome Sciences, Fred Hutch and UW MSTP programs that have all made my experience in graduate school so fantastic.

I would like to thank my family, Debra, Gary, Brooks and Cole, as well as my partner, Tommy, for their endless support and patience over the past six years as I have worked towards this goal. In particular, I want to thank all of the encouragement, love and support from my parents I have received for as long as I can remember.

Chapter 1. INTRODUCTION

Despite advances in both the mechanistic underpinnings and therapeutic interventions, cancer remains a leading cause of human morbidity and mortality (Ferlay et al., 2019). While significant effort has been devoted to deep understandings of a small number of known tumor suppressor and protooncogenes, growing evidence suggests non-genetic adaptations in cancer cells contribute to or drive every cancer hallmark (Abdel-Wahab & Gebauer, 2018; Anczukow & Krainer, 2016; Dvinge et al., 2016). For example, aberrant RNA processing drives many cancers and expression of even a single RNA isoform can drive cancer formation. For example, in the context of SF3B1 mutant uveal melanoma, alternative splicing of the tumor suppressor *BRD9* can lead to increased poison exon inclusion, leading to transcript degradation of *BRD9* via nonsense-mediated decay (Inoue et al., 2019). This work further goes on to demonstrate that *BRD9* is located in on a chromosome that is frequently amplified in many cancer including uveal melanoma, explain why cancer's may select for a splicing based mechanism to nullify gene function rather than acquiring several loss-of-function mutations in several gene copies (Inoue et al., 2019). Non-genetic mechanisms, and specifically alternative RNA processing, represent an understudied contribution to cancer phenotypes and thus an untapped source of potential cancer therapeutics.

Among RNA processing, alternative polyadenylation (APA) stands out as being recurrently dysregulated in all human cancers studied, yet there have been no studies that comprehensively investigate how APA functionally contributes to disease progression (Elkon et al., 2013; Gruber & Zavolan, 2019). There is therefore a critical need to understand the molecular effects of APA and identify functionally relevant APA events in cancer.

More than 70% of all mRNAs in the cell contain multiple polyadenylation sites within their 3′ UTR and ample sequencing evidence suggests that most mRNAs are regulated via APA (Derti et al., 2012; Tian et al., 2005; H. Zhang et al., 2005). APA is altered throughout development where several studies have demonstrated that as cells divide and differentiate within the developing embryo, cells begin to use more distal poly(A) sites across most tissue with terminally differentiated, non-dividing cells such as neurons displaying the largest increase in use of more distal poly(A) sites (Ji et al., 2009). Single cell approaches have validated these findings at the global scale, where most genes display 3′ UTR lengthening as development progresses (Agarwal et al., 2021). However, several genes display temporal or cell type specific regulation that in some cases correlated with cell type specific patterns in gene expression (Agarwal et al., 2021). These observations from development support the idea that rapidly dividing or undifferentiated cells may globally prefer proximal poly(A) sites and more differentiated, slower dividing cells often use more distal poly(A) sites. Experiments in immune cell populations *ex vivo* have determined that stimulation, which induces rapid cell proliferation, of several cell types including monocytes, CD4$^+$ T cells and B cells causes global 3′ UTR shortening (Sandberg et al., 2008). In all these systems the relationship between 3′ UTR length and cell division rate or differentiation state are still correlative. A study focused on disentangling cell cycle kinetics and poly(A) site selection found that cell types that need to secrete huge amount of protein, in particular syncytiotrophoblast cells in the human placenta, display 3′ UTR shortening of several genes to stabilize mRNA levels in a manner that was independent of cell division rate (Cheng et al., 2020). APA alterations across development and biological contexts suggests it is a tool employed by cells to alter mRNA and protein levels, however APA can also alter other aspects of mRNAs and proteins.

Increased distal poly(A) site selection leads to 3′ UTR lengthening and for some mRNAs this leads to the inclusion of binding sites for miRNAs or RNA-binding proteins (RBPs) (Gruber & Zavolan, 2019). This can alter mRNA localization as RBPs may bind and transport mRNAs to specific cellular locations for local translation. In neurons, 3′ UTR lengthening of the gene *Calm1* allows for anterograde transport of the mRNA to axon terminals where the gene is translated locally (Bae et al., 2020). Specific ablation of the *Calm1* distal poly(A) site leads to dorsal root ganglion malformations in mice, suggesting the local translation of Calm1 is essential for establishing connections between neurons in the developing embryo (Bae et al., 2020). APA can also alter protein localizations. Use of the distal poly(A) site, but not the proximal poly(A) site, in the gene *CD47* leads to transient interactions with HuR, SET and activated RAC1 that leads to the protein localizing to the plasma membrane instead of the endoplasmic reticulum (Berkovits & Mayr, 2015). Thus, APA represents a mechanism via which cells can alter mRNA and protein localization without altering the final amino acid sequence.

Dysregulation of APA also leads to disease. Trinucleotide repeat expansion in *PABPN1*, which functions binds and blocks proximal poly(A) site use, leads to Oculopharyngeal Muscular

Dystrophy, a rare, progressive disease that slowly kills muscle cells in the eyes, upper body and pharynx (Raz et al., 2017). The exact pathophysiology is still unknown, but muscle cells which preferentially use distal poly(A) sites, are for some reason more susceptible to the toxicity of the repeat expansion in *PABPN1* even though the gene is ubiquitously expressed in all tissues (Raz et al., 2017). Poly(A) site selection is also globally dysregulated in Duchenne Muscular Dystrophy (Batra et al., 2014). Duchenne Muscular Dystrophy is most commonly due to a trinucleotide repeat expansion in the *DMPK* gene, which sequesters MBNL1, an RBP known to influence splicing and poly(A) site selection leading to muscle cell dysfunction and eventual muscle cell death (Batra et al., 2014). Pathological implication of disruption of APA in muscular dystrophies are still poorly understood due to a lack of large sequencing datasets; however, in other diseases, specifically cancer, the landscape of APA is more well defined.

Several independent observations support the hypothesis that APA is functionally relevant to cancer progression. Cells with higher rates of cell division, including cancer cells, tend to have significantly shorter 3′ UTRs, due to use of more proximal poly(A) cleavage sites (Sandberg et al., 2008). Shifts in 3′ UTR length affect protein production and protein localization (Berkovits & Mayr, 2015; Mayr & Bartel, 2009), both of which have the potential to fuel cancer phenotypes. Many non-mutated poly(A) complex proteins are overexpressed in cancer and drive universal changes in 3′UTR lengths. For example, *CSTF2* and *PABPC1* - both critical regulators of poly(A) site selection and cleavage - are commonly overexpressed in many human cancers (Xia et al., 2014). Analysis of RNA-seq data shows that expression of these key poly(A) complex proteins is strongly correlated with global 3′UTR lengths in many cancers (Goering et al., 2020; Xia et al., 2014). Studies of gene expression in cancer often attempt to correlate mRNA abundance with patient outcomes, thus providing insight to critical pathways or genes influencing tumor growth. Comparisons of different features of mRNA reveals that global measures of APA and 3′ UTR length correlate with patient outcomes better than mRNA abundance in a number of cancers (Goering et al., 2020; Xia et al., 2014; Xue et al., 2018). This observation underscores the importance of exploring the functional relevance of APA events.

While these findings are compelling and have been recapitulated by several independent groups, they are overwhelmingly correlative studies. The field still lacks comprehensive experimental evidence of whether recurrent APA events can drive cancer phenotypes like unchecked cell growth (Gruber & Zavolan, 2019). In Chapter 2, I present a high-throughput functional screening approach to identify APA events that drive melanoma formation in an immunocompetent host. I identify a specific APA event that alters both cancer cell growth *in vivo* and modifies tumor immune interactions in both mice and human melanoma. In Chapter 3, I look more broadly at pan-cancer trends of protein regulators of poly(A) site selection and identify that colorectal adenocarcinoma is a striking outlier, where expression of known poly(A) regulators is poorly correlated with 3′ UTR length. I provide strong evidence that the tumor suppressor protein Adenomatous Polyposis Coli (APC) regulates poly(A) site selection via direct RNA binding and demonstrate that the degree of 3′ UTR lengthening is correlated with patient survival and clinically actionable molecular signatures. Together, this work highlights that APA plays a functional role in shaping cancer phenotypes like unchecked growth and immune escape, and that APA events represent a novel source of potential therapeutic targets for human cancers.

Chapter 2. HIGH-THROUGHPUT FUNCTIONAL CHARACTERIZATION OF MELANOMA SPECIFIC ALTERNATIVE POLYADENYLATION

This research is in review at *Nature Communications*. RNA-seq and Poly(A)-seq mapping was completed by James D. Thomas and Robert Bradley. James D. Thomas and Robert Bradley assisted with paired-guide RNA library design, pgRNA data analysis, experimental design, and idea inception. *In vitro* and in vivo growth studies were completed with assistance from Andrea Belleville. Autophagy LC3-flow cytometry assays were completed with assistance from Siegen McKellar. The remaining work described here, including computational analyses of poly(A) site selection and survival, pgRNA CRISPR/Cas9 screen, in vitro and in vivo B16-F10 experiments, and immunotherapy cohort analyses represent my contribution to the work. All writing and figures were generated by me with revisions and review completed by Robert Bradley, Andrea Belleville, Siegen McKellar and James D. Thomas.

2.1 SUMMARY

Alternative polyadenylation (APA) is strikingly dysregulated in many cancers. Although APA dysregulation is frequently associated with poor prognosis, the biological importance of most APA events remains unclear simply because few have been functionally studied. Here, we performed a CRISPR/Cas9-based screen to assess individual APA events' contributions to tumor growth in vivo. Forcing use of specific polyadenylation sites altered mRNA and protein levels to modify mouse melanoma growth in an immunocompetent host. Our screen highlighted APA events of potential clinical relevance. For example, forced Atg7 3′ UTR lengthening in mouse melanoma reduced ATG7 protein levels and tumor immune infiltration; similarly, in human melanoma, a long ATG7 3′ UTR and low mRNA levels were significantly associated with reduced anti-tumor T cell activity and failure of immune checkpoint blockade. Our data demonstrate that cancer-associated APA plays a causative role in tumorigenesis and motivate future studies of the therapeutic potential of modulating APA.

2.2 INTRODUCTION

Alternative cleavage and polyadenylation (APA) – the process whereby a pre-mRNA can be cleaved and have a poly(A) tail added at multiple distinct locations, leading to expression of mRNAs with different 3′ untranslated regions (UTRs) – affects most human genes (Derti et al., 2012). Although the biological roles of most APA events are unknown, detailed studies have revealed that APA can affect mRNA levels, localization, and translation, among other molecular phenotypes (Berkovits & Mayr, 2015; Ji et al., 2009; Spies et al., 2013). Differential polyadenylation site usage is commonly observed in many biological processes, frequently in a biased manner indicative of coordinated regulatory changes (Lianoglou et al., 2013; Zhang et al., 2005). For example, rapidly dividing cells frequently utilize gene-proximal over gene-distal poly(A) sites, and thus express mRNAs with correspondingly shorter 3′ UTRs, compared to non-dividing, terminally differentiated cells. mRNAs are differentially polyadenylated throughout development, where terminally differentiated cells tend to utilize more distal poly(A) sites (Ji et al., 2009); immune cell subsets including monocytes, T cells, and B cells undergo global 3′ UTR shortening when stimulated by their respective chemokines to begin dividing (Sandberg et al., 2008); many cancers express mRNAs with markedly shorter 3′ UTRs than do peritumoral, healthy tissues (Goering et al., 2020; Mayr & Bartel, 2009; Xia et al., 2014).

It is well-established that APA is recurrently dysregulated in virtually all cancer types, yet it remains unknown whether most APA dysregulation is causally relevant to cancer progression or instead simply a downstream consequence of rapid cell division (Gruber & Zavolan, 2019; Hu et al., 2021; Xu & Zhang, 2018; Zhang et al., 2020, 2022). Consistent with a functional role for APA dysregulation in cancers, several studies have reported that APA dysregulation correlates with patient outcomes in a subset of tumor types and that overexpression of an mRNA utilizing a proximal poly(A) site, but not a distal poly(A) site, can promote enhanced cancer cell growth (Andres et al., 2019; Tan et al., 2021). Nonetheless, much work remains in order to test the potential functional roles of cancer-associated APA. The vast majority of cancer-associated APA events have never been studied. Most functional studies of APA have relied on transgenic expression, rather than manipulation of APA in an endogenous context, and no study has yet manipulated endogenous, cancer-associated APA events in the physiological context of tumorigenesis *in vivo*.

Here, we sought to test the hypothesis that cancer-associated APA dysregulation can play a causative, rather than simply correlative, role in tumorigenesis by functionally studying endogenous APA site selection in the physiologically relevant setting of tumorigenesis in vivo. To do so, we took advantage of recent reports that paired-guide RNAs (pgRNAs) can be utilized with CRISPR/Cas9 to ablate expression of individual isoforms and thereby force expression of an isoform of interest (Bae et al., 2020; Brumbaugh et al., 2018; Gonatopoulos-Pournatzis et al., 2020; Mitschka & Mayr, 2021; Pereira-Castro et al., 2022; Thomas et al., 2020; Q. Wang et al., 2018). This approach enabled us to develop a high-throughput platform for screening cancer-associated APA events during tumorigenesis in an immune-competent host, which in turn revealed a subset of APA events with marked functional consequences for tumor phenotypes.

2.3 RESULTS

2.3.1 *Global 3′ UTR lengthening predicts poor patient prognosis in human melanoma*

To assess whether cancer-associated changes in APA are relevant to clinical phenotypes, we tested whether global APA dysregulation was significantly associated with patient outcomes. Several prior studies have analyzed RNA sequencing data from The Cancer Genome Atlas (TCGA) and observed significant associations between global APA dysregulation and patient survival; however, those studies limited analyses to the 17 tumor types for which there were sufficient patient-matched, peritumoral normal tissue samples available (Xia et al., 2014; Goering et al., 2020). In order to extend such analyses to all tumor types, including those for which peritumoral samples are not available, we instead took a stratification-based approach that relied on data from tumor samples alone. For each cancer subtype, we stratified patients into terciles representing whether their tumor transcriptomes preferentially expressed short, medium, or long 3′ UTRs by computing a median 3′ UTR length for each tumor across 7,513 genes that are subject to alternative polyadenylation (**Figure 1A-H**; **Supp. Figure 1A-C**). This measure of global 3′ UTR length was significantly positively correlated with PABPN1 expression in 28 of 30 cancer subtypes and significantly negatively correlated with CSTF2 expression in 21 of 30 cancer subtypes, consistent with PABPN1 and CSTF2's known roles as a repressor or activator of proximal poly(A) site usage, respectively (**Supp. Figure 1D-I**) (Jenal et al., 2012; Li et al., 2015; Yao et al., 2012).

We then performed Kaplan-Meier survival analysis by comparing patients with short or long median 3′ UTRs and found significant survival differences that varied substantially by cancer type (**Figure 1I**; **Supp. Figure 2A-D; Supp. Table 1**). In ovarian carcinoma, renal cell carcinoma, breast carcinoma, colon adenocarcinoma, and lung adenocarcinoma, patients whose cancers expressed shorter 3′ UTRs exhibited significantly worse overall survival (**Figure 1I-J**; **Supp. Figure 2D**). Several cancer subtypes displayed the opposite trend, where patients with globally lengthened 3′ UTRs exhibited significantly worse overall survival, including head and neck squamous cell carcinoma, low grade glioma, and cutaneous melanoma (**Figure 1I-J**; **Supp.**

Figure 2D). These variable associations illustrate the potentially clinically relevant roles of APA and highlight the likely complexity of the relationship between 3′ UTR length trends and prognosis.

These results are necessarily based upon computational inference of 3′ UTR length using poly(A)-selected RNA-seq data, rather than assays like 3′-seq that provide nucleotide-level resolution of 3′ UTR length, due to the nature of most patient transcriptomic data. We therefore confirmed that our results were agnostic to the specific computational algorithm used to infer 3′ UTR usage by measuring 3′ UTR length using a distinct method, APAlyzer (Wang & Tian, 2020). We repeated our analyses for the cutaneous melanoma cohort, which demonstrated the most significant associations between 3′ UTR length and patient survival among TCGA cohorts, and found concordant and similarly strong associations between 3′ UTR length and survival (**Supp. Figure 2E-F; Supp. Table 2**). Together, these analyses highlight a subset of cancer types where global APA is strongly correlated with patient survival, suggesting possible functional relationships between APA and tumor growth.

2.3.2 *A mouse model of melanoma exhibits cancer-associated aberrant alternative polyadenylation*

Motivated by these strong associations between APA and patient prognosis, we sought to experimentally test the hypothesis that APA dysregulation affecting specific genes functionally contributes to tumorigenesis. As clinical melanoma exhibited the strongest correlations between 3′ UTR length and patient survival (**Figure 1I-J**), we turned to syngeneic mouse models of untransformed melanocytes and melanoma to functionally study APA in biologically relevant and experimentally tractable systems. B16-F10 cells are mouse melanoma cells that readily engraft and form aggressive tumors in immunocompetent C57BL/6 mice, while Melan-A cells are immortalized but non-tumorigenic melanocytes derived from the same C57BL/6 background (**Figure 2A**) (Bennett et al., 1987; Overwijk & Restifo, 2001). These two models therefore provided an opportunity to identify melanoma-associated APA by comparing immortalized melanocytes and melanoma cells in a consistent genetic background context.

We profiled global gene expression and poly(A) site selection in each of the two syngeneic cell lines with high-replicate RNA-seq and Poly(A)-seq (**Figure 2B**). We identified significant differences in APA site selection between the two cell lines (n = 6 replicates per cell line) with APAlyzer. This analysis revealed 654 significantly differentially polyadenylated transcripts for which differential APA site selection was supported by both Poly(A)-seq and RNA-seq (**Figure 2C-F; Supp. Table 3**). 204 genes displayed strong correlations between APA and differential gene expression, but the majority did not, consistent with the idea that many APA events have gene- or protein-specific effects rather than simply modulating transcript levels in an easily predictable manner (Spies et al., 2013) (**Figure 2G**).

We next compared APA in mouse and human melanoma. For each gene, we calculated the relative distal poly(A) site use in the B16-F10 mouse melanoma model system and for the human gene ortholog in the 424 TCGA melanoma samples and then computed the Pearson correlation for those gene-level values. This analysis revealed a modest but significant positive correlation between APA in mouse and human melanoma ($R = 0.24$, $p = 2.2x10^{-16}$; **Supp. Figure 3A-C**). Overall, these data demonstrate that widespread alterations in APA characterize the B16-F10 model of mouse melanoma, suggesting that it may be a useful system for dissecting the functional contributions of individual APA events to tumorigenesis.

2.3.3 *CRISPR-Cas9 paired-guide RNAs enable functional manipulation of alternative polyadenylation*

In order to identify possible functional roles for the APA events that we identified in melanoma, we sought to manipulate endogenous poly(A) site selection in order to test how each APA event influenced tumor growth. We utilized a well-established approach for endogenous poly(A) site manipulation, wherein CRISPR/Cas9 paired-guide RNAs (pgRNAs) are used to precisely delete an individual poly(A) site. This approach has been utilized by multiple groups to force usage of specific poly(A) sites in Trp53, Calm1, and other genes (Bae et al., 2020; Mitschka & Mayr, 2021; Pereira-Castro et al., 2022; Q. Wang et al., 2018).

We tested the efficacy of this approach in our melanoma model system using Sap30l, which has two distinct poly(A) sites separated by several kilobases. B16-F10 cells preferentially expressed SAP30L mRNA with the shorter 3′ UTR, while Melan-A cells preferentially utilized the longer 3′ UTR (**Figure 2C**). We generated constitutive Cas9-expressing B16-F10 cells that we then transduced with two distinct pgRNAs designed to delete the proximal polyadenylation signal sequence within the Sap30l 3′ UTR (**Figure 3A-B**). Cas9-expressing B16-F10 cells treated with either pgRNA designed to knock out the proximal polyadenylation signal sequence (pKO) of Sap30l exhibited appreciable levels of genomic DNA excision at the expected site (**Figure 3C**; **Supp Figure 3A-B**). We generated seven monoclonal lines with heterozygous excision of the proximal Sap30l polyadenylation signal, five monoclonal lines with homozygous excision, and six monoclonal negative control lines transduced with a non-targeting control (NTC) pgRNA (**Supp. Figure 3C**). Treatment with either of the two distinct Sap30l pKO pgRNAs significantly increased use of the distal polyadenylation signal sequence in both the polyclonal and monoclonal settings as measured by RT-PCR, where cells with homozygous deletion of the proximal poly(A) signal displayed the highest use of the distal poly(A) site (**Figure 3D-E**). Using qRT-PCR to measure SAP30L mRNA levels, we found that homozygous excision of the proximal poly(A) signal led to a median reduction in total SAP30L mRNA levels of 47.9% (**Figure 3F**; p = 0.017 computed across the monoclonal isolates). We also observed a strong negative correlation between distal polyadenylation site usage and total mRNA abundance for Sap30l (**Figure 3G**; p = 0.0084). We concluded that targeted deletion of proximal poly(A) sites using a CRISPR/Cas9 pgRNA approach can effectively force distal poly(A) site use in both polyclonal and monoclonal populations of B16-F10 cells.

2.3.4 *High-throughput functional screening identifies alternative polyadenylation that regulates murine melanoma growth in vitro and in vivo*

To test if APA is functionally relevant in melanoma, we designed a custom pgRNA library targeting 143 proximal polyadenylation signal sequences that we identified as cancer-associated from our genomic analyses (**Supp. Figure 5A; Supp. Table 4).** Approximately 25% of these APA events were dysregulated in both human and mouse melanoma (36 targets), while 75% were specific to mouse melanoma (107 targets) (**Supp. Figure 5B**). We created downloadable BED files to visualize pgRNA library design alongside the Poly(A)-seq data from B16-F10 and Melan-A cells that we used to select targets (**Supp. Figure 6**). The final library comprises 8-10 unique targeting pgRNAs per proximal poly(A) signal, 150 positive control pgRNAs designed to knock out 15 distinct genes known to be involved in tumor growth control, and 150 negative control pgRNAs targeting proximal polyadenylation signals in genes that are not expressed in B16-F10 cells. We included ample negative control pgRNAs in order to allow for accurate determination of empirical false discovery rates (FDRs) and robust statistical significance testing (**Supp. Figure 5B**)(Thomas et al., 2020). The resulting targeting, positive control, and negative control pgRNAs each have similar distributions of on- and off-target scores relative to pgRNA libraries that we and others have previously used for functional interrogation of tumorigenesis (**Supp. Figure 5C-D**) (Thomas et al., 2020). We then cloned this library of 1,718 unique pgRNAs via oligo array cloning as previously described (Gasperini et al., 2017) and confirmed excellent library diversity with next-generation sequencing of the final plasmid pool (**Supp. Figure 5E-F**).

We used this library to identify APA that influenced tumor growth by performing paired in vitro and in vivo screens in B16-F10 cells (**Figure 4A**). In brief, we infected Cas9-expressing B16-F10 cells with the lentiviral library at an MOI of 0.2. 24 hours later, we added 0.5 ug/mL puromycin, which we left for 2 days, and then cultured for one more day in puromycin-free media. We then split the cells into 8 replicates. For each replicate, we collected a day 0 fraction, plated 500,000 cells for culturing in vitro, or subcutaneously injected 500,000 cells into a mouse flank for tumor growth in vivo. After 20 days of monitored growth, cells and tumors were harvested, and total genomic DNA (gDNA) was extracted using established methods (Chen et al., 2015). gDNA from the plasmid pool, day 0 time points, and day 20 in vitro and in vivo time points was then PCR amplified and subjected to next-generation sequencing with ~2,500-fold coverage per pgRNA.

We performed enrichment analysis and FDR computation using our previously described statistical framework (Thomas et al., 2020). In brief, we normalized the fold-changes relative to day 0 for a given replicate such that the median of all pgRNAs targeting poly(A) sites in unexpressed genes was equal to 1 and pooled computed fold-changes across all replicates for a given condition to maximize statistical power (**Supp. Table 5**). We then calculated a p-value for significant enrichment/depletion associated with each poly(A) site target by comparing the distribution of fold-changes for all pgRNAs targeting that poly(A) site to the distribution of pgRNAs targeting poly(A) sites in unexpressed genes. We computed empirical false-discovery rates (FDRs) for each targeted poly(A) site using a subsampling procedure across negative control pgRNAs targeting unexpressed genes (n = 10,000 samples).

Both positive and negative control pgRNAs performed as expected in vitro and in vivo (**Figure 4B-D**; **Supp. Figure 7A-C**). pgRNAs designed to knock out genes that are essential for tumor growth exhibited marked depletion, while pgRNAs designed to delete proximal polyadenylation sites of unexpressed genes exhibited negligible enrichment or depletion, with 96.0% or 96.7% of pgRNAs falling within two standard deviations of the median fold-change for our in vitro and in vivo experiments, respectively (**Figure 4C-D**; **Supp. Figure 7A-C**). In contrast, numerous pgRNAs targeting proximal polyadenylation sites of expressed genes exhibited significant enrichment or depletion, with deletion of some proximal polyadenylation sites associated with alterations in B16-F10 cell growth in vitro and in vivo of similar magnitudes to positive controls (**Figure 4C-D**).

We next validated two particularly striking hits that emerged from our screen. pgRNAs targeting the proximal polyadenylation sites of Atg7 and Egln1 were strongly depleted and enriched, respectively, both in vitro and in vivo. We introduced individual pgRNAs targeting these polyadenylation sites into Cas9-expressing B16-F10 cells and systematically profiled polyadenylation site selection, protein levels, and effects on cell growth. All experiments were performed in comparison to Cas9-expressing B16-F10 cells transduced with a pgRNA targeting a poly(A) site in the unexpressed gene Crabp1, which was selected as a negative control because its normalized fold-change in the screen was close to the median of all control pgRNAs targeting poly(A) sites within unexpressed genes. These experiments confirmed that poly(A) site knock out (pKO) pgRNAs targeting Atg7 and Egln1 induced significant increases in use of the distal polyadenylation sites, as intended (**Figure 4E & 4H**). pKO pgRNA treatment significantly reduced ATG7 protein levels but was associated with only modest alterations in EGLN1 protein levels (**Figure 4F & 4I**). Introduction of pKO pgRNAs to force use of distal polyadenylation sites in Atg7 and Egln1 recapitulated the in vitro growth phenotypes observed in the screen format (**Figure 4G & 4J**). Overall, these data demonstrate that multiplexed pgRNA screening effectively nominates specific APA events that influence melanoma cell growth both in vitro and in vivo.

2.3.5 *A long Atg7 3′ UTR suppresses murine melanoma growth in vitro and in vivo*

We selected Atg7 for additional studies due to its significant depletion both in vitro and in vivo. Cas9-expressing B16-F10 cells were treated with either a pKO pgRNA targeting Atg7 or a control pgRNA targeting a poly(A) site in an unexpressed gene, and then engrafted into C57BL/6 mice. Consistent with results from the multiplexed screen, Atg7 pKO tumors exhibited reduced growth in vivo, leading to significantly prolonged host survival using Kaplan-Meier survival analysis ($p = 0.0053$) (**Figure 5A**).

We next sought to understand the molecular consequences of Atg7 APA. Autophagy related protein 7, ATG7, is considered a critical protein for autophagosome formation and function, and loss of ATG7 protein inhibits autophagy (Komatsu et al., 2005). We therefore investigated the consequences of Atg7 APA for autophagy. Complete Atg7 knock out and consequent loss of ATG7 protein caused a significant increase in p62 protein levels, as expected (**Supp. Figure 8A**). In contrast, treating cells with an Atg7 pKO pgRNA to force distal polyadenylation site usage did not cause detectable deficits in autophagy at baseline or following exposure to serum and amino acid starvation, measured by both p62 accumulation and a flow cytometry-based LC3-GFP-mCherry reporter (**Supp. Figure 8A-E**)(Leeman et al., 2018). These results are distinct from a previous study reporting that ATG7 3′ UTR lengthening was associated with autophagy inhibition in pro-B cells, a difference that may be attributable to cell type-specific effects or that study's reliance on KD of all ATG7 isoforms for functional studies rather than specific manipulation of poly(A) site usage (Park et al., 2016). Although forcing distal polyadenylation site usage reduced ATG7 protein levels by ~60% (**Figure 4F**), we detected none of the autophagy impairment that was readily apparent upon Atg7 KO, highlighting the utility of directly manipulating poly(A) site usage for functional inference.

We next performed immunohistochemistry (IHC) for the Ki67 proliferation marker on control and Atg7 pKO tumors. Although Atg7 pKO tumors grew more slowly (**Figure 5A**), they exhibited a significantly higher fraction of nuclei with moderate and strong Ki67 staining relative to control tumors (**Figure 5B-C**). The levels of Ki67 fluctuate significantly throughout the cell cycle, with RNA and protein levels rising through S phase and peaking in G2/M phase (Sobecki et al., 2017). We therefore performed cell cycle analysis with propidium iodide staining and found that Atg7 pKO increased the fraction of B16-F10 cells in S phase and G2/M phase. These altered cell cycle kinetics are consistent with the increased Ki67 staining intensity observed by tumor IHC and concordant with a previous study demonstrating that ATG7 KD by siRNA in human bladder cancer cells increased the fraction of cells in S phase and G2/M phase (**Figure 5D-E**)(Zhu et al., 2017).

Atg7 undergoes APA in both mouse and human cells, raising the possibility that ATG7 3′ UTR length could influence human melanoma just as it does mouse melanoma. To understand if our results from B16-F10 cells are relevant to clinical melanoma, we performed Kaplan-Meier survival analysis after stratifying 424 cutaneous melanoma patients studied by The Cancer Genome Atlas by ATG7 3′ UTR length. We observed that patients whose tumors expressed a longer ATG7 3′ UTR exhibited significantly better progression-free survival, consistent with our finding that forcing Atg7 long 3′ UTR usage slowed mouse melanoma growth (**Figure 5F**). ATG7 3′ UTR length is modestly but significantly negatively correlated with ATG7 expression in clinical melanoma, mirroring our observations in mouse melanoma (**Figure 5G**). Although ATG7 is not subject to recurrent loss-of-function mutations in clinical melanoma, with no nonsense or frame-disrupting mutations observed in the 424 TCGA patients, ATG7 APA is highly variable across melanomas (**Figure 5G-H**). Given the association between ATG7 APA and survival that we observed in mouse and clinical human cancer, it is interesting to speculate that ATG7 APA could provide a post-transcriptional mechanism for cancers to regulate ATG7 protein levels in the absence of ATG7 mutations.

2.3.6 *A long Atg7 3′ UTR is associated with reduced immune infiltration in mouse and human melanoma*

To gain further insights into potential means by which APA influences tumorigenesis, we performed RNA sequencing of control and Atg7 pKO B16-F10 cells and tumors (n = 3 replicates each in vitro and n=5 replicates each in vivo). Quantifying and comparing gene expression between Atg7 pKO cells and tumors revealed that Atg7 is the most significantly downregulated gene in both settings, consistent with the on-target APA modulation that we observed by RT-

PCR (**Supp. Figure 9A-C; Supp. Table 6**). We performed a Gene Ontology enrichment analysis of differentially expressed genes to identify biological pathways whose regulation was significantly different in Atg7 pKO relative to control cells (**Supp. Figure 9E**). Although this analysis revealed no significantly enriched or depleted terms in Atg7 pKO cells cultured in vitro, Atg7 pKO tumors exhibited markedly reduced expression of genes associated with inflammation as well as both the innate and adaptive immune responses (**Figure 6A-B; Supp. Figure 9E**). This result was notable in the context of a recent study, which reported that Atg7 gene KO sensitized B16-F10 cells to $CD8^+$ T cell-mediated cytotoxicity (Lawson et al., 2020). We therefore reasoned that modulation of ATG7 protein levels by ATG7 APA might similarly modify tumor-immune interactions, potentially by influencing antigen processing or presentation.

To test this hypothesis, we measured $H\text{-}2K^b$ expression following interferon-gamma stimulation, which revealed that Atg7 KO cells expressed higher levels of cell-surface $H\text{-}2K^b$ than did control cells. In contrast, B16-F10 Atg7 pKO cells displayed reduced $H\text{-}2K^b$ expression relative to control and Atg7 KO cells, demonstrating that Atg7 APA influences antigen presentation and highlighting that alterations caused by Atg7 APA are functionally distinct from those resulting from Atg7 KO (**Supp. Figure 10A-E**). We next tested whether the gene expression signature of reduced inflammation and immune responses observed in Atg7 pKO tumors arose from differences in immune cell infiltration. We quantified $CD8^+$ T cell infiltrates by performing CD8A IHC of control and Atg7 pKO B16-F10 tumors, which revealed significantly reduced numbers of tumor-infiltrating $CD8^+$ T cells in pKO relative to control tumors (**Figure 6C**).

We then extended these analyses to clinical melanoma. We returned to The Cancer Genome Atlas cutaneous melanoma dataset and completed differential gene expression analysis of patients stratified into terciles based on tumor ATG7 3′ UTR length. We detected reduced expression of immune- and inflammation-related genes in tumor samples that expressed longer ATG7 3′ UTRs, consistent with the signals from mouse melanoma (**Supp. Figure 9D & 9F; Supp. Table 7**). Clinical samples expressing long ATG7 3′ UTRs preferentially exhibited modestly reduced CD8A expression and a strikingly decreased gene expression signature of cytolytic T cell activity (**Supp. Figure 11A-C**), defined as the geometric mean of GZMB and

PRF1 expression (Rooney et al., 2015). Overall, these data demonstrate that usage of the ATG7 distal polyadenylation site is associated with concordant signatures of reduced anti-tumor immune activity in both mouse and human tumors.

2.3.7 *Low ATG7 expression is associated with reduced response to immune checkpoint blockade in clinical melanoma*

As ATG7 distal polyadenylation site usage is strongly associated with reduced CD8$^+$ T cell infiltration in melanoma, we hypothesized that ATG7 APA might be relevant to clinical responses to T cell-dependent therapies such as immune checkpoint blockade. We therefore analyzed two independent clinical cohorts of melanoma patients treated with immune checkpoint blockade (ICB; anti-PD-1 monotherapy or anti-CTLA-4 + anti-PD-1 combination therapy) (Gide et al., 2019; Liu et al., 2019). As these cohorts utilized exome capture in order to extract RNA from FFPE samples for RNA-seq library preparation, the resulting RNA-seq library coverage was restricted to coding sequences and did not include UTRs. We therefore stratified patients based on ATG7 expression levels, which correlates with ATG7 3′ UTR length in mouse and human melanoma, with ATG7 3′ UTR lengthening associated with lower ATG7 expression (**Figure 5G**; **Supp. Figure 9A-C**). Samples with low ATG7 expression exhibited significantly reduced expression of a core set of CD8$^+$ T cell relevant genes (GZMA, GZMB, IFNG, STAT1, CD274, CD8A, PRF1) and significantly lower cytolytic activity relative to samples with high ATG7 expression, as expected from our previous mouse and human analyses based on direct measurement of ATG7 3′ UTR length (**Figure 6D-E and 6H**; **Supp. Figure 12A-F; Supp. Table 8**)(Ayers et al., 2017; van Allen et al., 2015). In both ICB cohorts, patients with low ATG7 expression were significantly less likely to exhibit partial or complete responses to therapy as defined by objective Response Evaluation In Solid Tumors (RECIST) criteria (Eisenhauer et al., 2009) (**Figure 6F & 6I**). We further analyzed one cohort for which both pre-treatment (PRE) and early-during therapy samples (EDT) were available to find that the quantitative increase in ATG7 expression following therapy initiation was significantly associated with better clinical outcomes (**Supp. figure 12G-I**). Low ATG7 expression was significantly associated with worse progression-free survival by Kaplan-Meier analysis in both cohorts (**Figure 6G & 6J**). Together, these data suggest that ATG7 APA-dependent alterations in anti-tumor immune activity are relevant to clinical response to immune checkpoint blockade, a particularly important therapy for metastatic melanoma.

2.4 DISCUSSION

Dysregulated alternative polyadenylation is a pervasive feature of most cancers, and thousands of cancer-associated APA events have been identified. However, very few of these APA events have been functionally studied or linked to cancer phenotypes. Here, we describe a high-throughput, in vivo screening platform to interrogate the functional roles of individual APA events in cancer and identify APA isoforms which enhance or reduce tumor growth in an immunocompetent mouse model. Our study provides strong support for the hypothesis that pervasive APA dysregulation can play a causative, rather than simply a correlative, role in influencing cancer phenotypes such as rapid cell growth and immune evasion.

We primarily focused on APA events whose modulation was associated with notable effects on tumor growth. However, it is interesting to note that most APA events queried in our screen were not associated with detectable alterations in tumor cell growth either in vitro or in vivo. Although we cannot rule out the possibility that such negative results arise from ineffective pgRNA targeting and APA modulation, this seems unlikely given the relatively high efficiency of APA modulation that we observed even in polyclonal settings for all APA events that we studied individually. These data therefore suggest that many APA events have modest or no effects on tumor growth kinetics in our system and therefore either alter other facets of tumorigenesis or are bystander events of low functional relevance. This dichotomy, where some APA events play critical functional roles and others do not, may help to resolve the ongoing controversy regarding the likely functional relevance of most APA (Gruber & Zavolan, 2019, 2020; Xu & Zhang, 2018). In either case, our data highlight the utility of high-throughput screens for identifying functionally relevant events for detailed study.

We utilized the CRISPR/Cas9 pgRNA system for endogenous APA modulation because it has been extensively tested and validated (Bae et al., 2020; Mitschka & Mayr, 2021; Pereira-Castro et al., 2022; Q. Wang et al., 2018). Future studies of APA may take advantage of the diverse other genome editing technologies that are now amenable to high-throughput screens. For example, dead Cas9 (dCas9) targeted downstream of the poly(A) signal has been used to enhance proximal poly(A) site selection (Shin et al., 2022), providing a useful counterpart to the suppression of poly(A) site use via pgRNAs that we relied upon. A fusion of the RNA-targeting dead Cas13b with adenosine deaminase 2 (ADAR2) can specifically modify adenosine residues in RNA (Cox et al., 2017) and as such could potentially be used to disrupt the A-rich canonical polyadenylation signal sequences AATAAA and ATTAAA, thereby allowing for functional manipulation of poly(A) sites at the level of RNA, rather than DNA.

Although we focused on cancer in the current study, we speculate that high-throughput functional interrogation of APA will prove equally fruitful in other biological contexts. For example, embryonic development and tissue differentiation (Agarwal et al., 2021; Cheng et al., 2020) display coordinated, temporal regulation of poly(A) site selection, while poly(A) site selection is globally dysregulated in many genetic diseases, including Oculopharyngeal Muscular Dystrophy and Duchenne Muscular Dystrophy (Batra et al., 2014; Raz et al., 2017). Targeted sequencing has cataloged thousands of context-specific APA alterations in these and other systems, the vast majority of which have never been functionally studied. High-throughput manipulation of endogenous poly(A) site selection offers an efficient way to bridge the gap between APA identification and functional understanding.

2.5 FIGURES

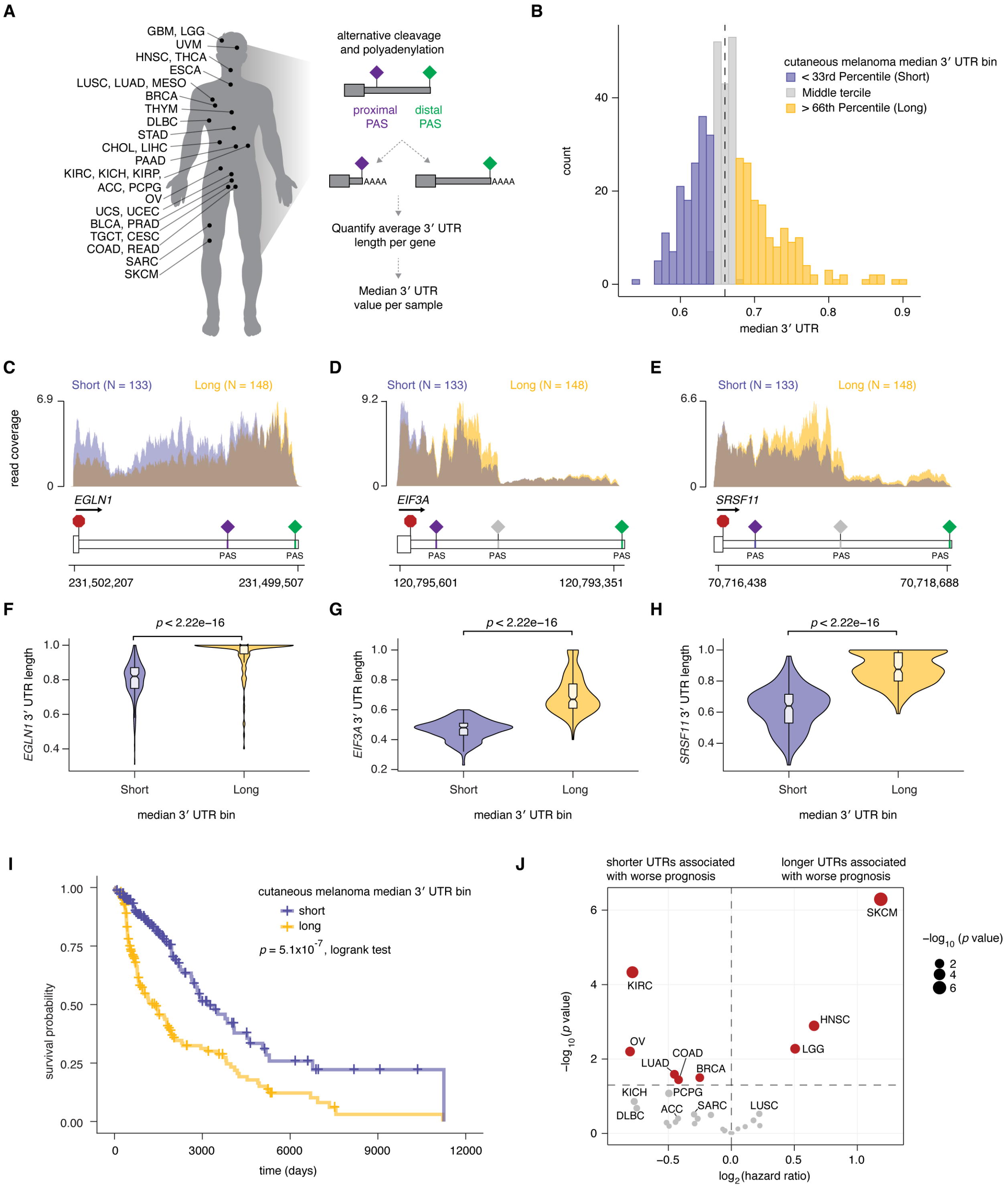

A
GBM, LGG
UVM
HNSC, THCA
ESCA
LUSC, LUAD, MESO
BRCA
THYM
DLBC
STAD
CHOL, LIHC
PAAD
KIRC, KICH, KIRP,
ACC, PCPG
OV
UCS, UCEC
BLCA, PRAD
TGCT, CESC
COAD, READ
SARC
SKCM
alternative cleavage and polyadenylation
proximal PAS
distal PAS
AAAA
AAAA
Quantify average 3′ UTR length per gene
Median 3′ UTR value per sample
B
cutaneous melanoma median 3′ UTR bin
< 33rd Percentile (Short)
Middle tercile
> 66th Percentile (Long)
count
0
20
40
0.6
0.7
0.8
0.9
median 3′ UTR
C
Short (N = 133)
Long (N = 148)
read coverage
6.9
0
EGLN1
PAS
PAS
231,502,207
231,499,507
D
Short (N = 133)
Long (N = 148)
9.2
0
EIF3A
PAS
PAS
PAS
120,795,601
120,793,351
E
Short (N = 133)
Long (N = 148)
6.6
0
SRSF11
PAS
PAS
PAS
70,716,438
70,718,688
F
p < 2.22e−16
EGLN1 3′ UTR length
1.0
0.8
0.6
0.4
Short
Long
median 3′ UTR bin
G
p < 2.22e−16
EIF3A 3′ UTR length
1.0
0.8
0.6
0.4
0.2
Short
Long
median 3′ UTR bin
H
p < 2.22e−16
SRSF11 3′ UTR length
1.0
0.8
0.6
0.4
Short
Long
median 3′ UTR bin
I
cutaneous melanoma median 3′ UTR bin
short
long
p = 5.1x10^-7, logrank test
survival probability
1.00
0.75
0.50
0.25
0.00
0
3000
6000
9000
12000
time (days)
J
shorter UTRs associated with worse prognosis
longer UTRs associated with worse prognosis
−log10 (p value)
6
4
2
0
SKCM
KIRC
HNSC
LGG
OV
COAD
LUAD
BRCA
KICH
PCPG
DLBC
ACC
SARC
LUSC
−0.5
0.0
0.5
1.0
log2 (hazard ratio)
−log10 (p value)
2
4
6

Figure 2.1. Global 3′ UTR length correlates with clinical outcomes across tumor types.

(**A**) TCGA cancer subtypes analyzed and schematic of quantification of gene-level 3′ UTR lengths for all genes per sample. We followed standard practice (Xia et al., 2014) to compute the median 3′ UTR length per sample, normalized such that a value closer to 0 indicates, on average, globally shorter 3′ UTRs and a value closer to 1 indicates globally longer 3′ UTRs. Values are based on the percent of distal poly(A) site usage index computed by the DaPars algorithm (Xia et al., 2014).

(**B**) Distribution of median 3′ UTR per sample for the 424 tumor samples in The Cancer Genome Atlas cutaneous melanoma cohort. Samples were binned into terciles corresponding to short (blue), medium (gray), and long (yellow) global 3′ UTR samples. The median of all samples is marked with a dashed black line.

(**C-E**) Individual BAM coverage plots (read coverage reflective of TPM values) for three exemplar genes that exhibit significant differences in 3′ UTR length between samples with globally shorter (blue) and longer (yellow) UTRs as indicated in panel B for *EGLN1*, *EIF3A* and *SRSF11* in cutaneous melanoma (SKCM). Poly(A) signal sequences (PAS) sites indicated as diamonds.

(**F-H**) Violin plots of gene level 3′ UTR length for *EGLN1*, *EIF3A* and *SRSF11* from panels C-E comparing short and long median 3′ UTR stratified cutaneous melanoma samples. A value closer to 0 indicates higher use of the proximal poly(A) site and a value closer to 1 indicates higher use of the distal poly(A) site (*p* values from two-sided Wilcoxon rank-sum test).

(**I**) Kaplan-Meier analysis comparing overall survival of TCGA cutaneous melanoma samples binned as short or long median 3′ UTR samples (*p* values from a two-sided logrank test).

(**J**) Volcano plot of the $\log_2$(hazard ratio) calculated from univariate cox regression models comparing overall survival of short versus long UTR stratified samples plotted against the -$\log_{10}$(*p* value). Cancer subtypes with a p value < 0.05 are indicated in red.

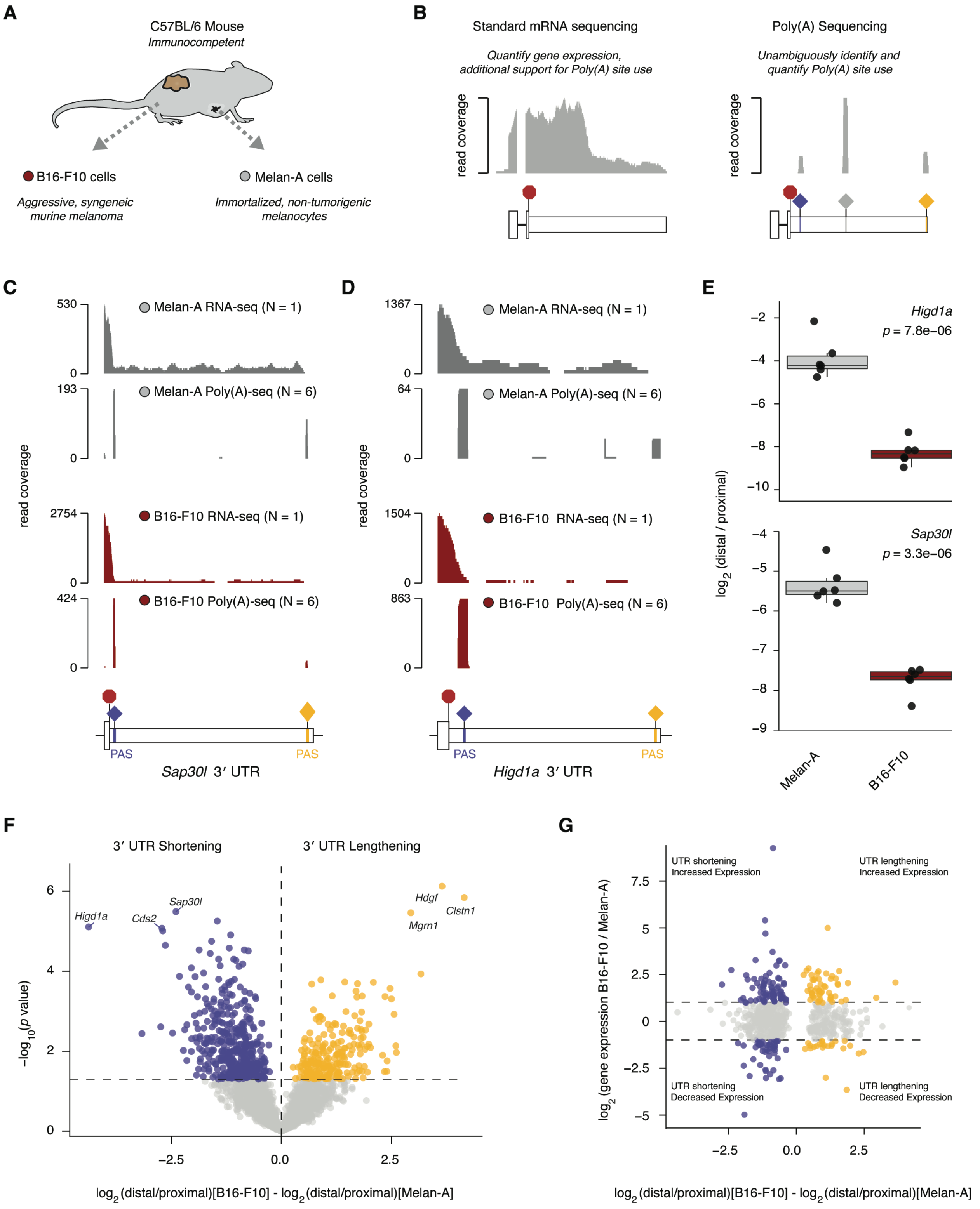

A
C57BL/6 Mouse
Immunocompetent
B16-F10 cells
Aggressive, syngeneic murine melanoma
Melan-A cells
Immortalized, non-tumorigenic melanocytes
B
Standard mRNA sequencing
Quantify gene expression, additional support for Poly(A) site use
read coverage
Poly(A) Sequencing
Unambiguously identify and quantify Poly(A) site use
read coverage
C
530
0
Melan-A RNA-seq (N = 1)
193
0
Melan-A Poly(A)-seq (N = 6)
2754
0
B16-F10 RNA-seq (N = 1)
424
0
B16-F10 Poly(A)-seq (N = 6)
read coverage
PAS
PAS
Sap30l 3′ UTR
D
1367
0
Melan-A RNA-seq (N = 1)
64
0
Melan-A Poly(A)-seq (N = 6)
1504
0
B16-F10 RNA-seq (N = 1)
863
0
B16-F10 Poly(A)-seq (N = 6)
read coverage
PAS
PAS
Higd1a 3′ UTR
E
Higd1a
p = 7.8e−06
−2
−4
−6
−8
−10
Sap30l
p = 3.3e−06
−4
−5
−6
−7
−8
−9
log2 (distal / proximal)
Melan-A
B16-F10
F
3′ UTR Shortening
3′ UTR Lengthening
Higd1a
Cds2
Sap30l
Hdgf
Clstn1
Mgrn1
6
4
2
0
−log10(p value)
−2.5
0.0
2.5
log2 (distal/proximal)[B16-F10] - log2 (distal/proximal)[Melan-A]
G
UTR shortening Increased Expression
UTR lengthening Increased Expression
UTR shortening Decreased Expression
UTR lengthening Decreased Expression
7.5
5
2.5
0
−2.5
−5
log2 (gene expression B16-F10 / Melan-A)
−2.5
0.0
2.5
log2 (distal/proximal)[B16-F10] - log2 (distal/proximal)[Melan-A]

Figure 2.2. Identification of differentially polyadenylated RNAs in a murine model of melanoma.

(**A**) Description of two syngeneic model cell lines, B16-F10 cells (red) and Melan-A cells (gray) both derived from a C57BL/6 background melanocyte origin.

(**B**) Differences in information provided by RNA-seq and Poly(A)-seq methods, and examples of BAM coverage plots reflective of the reads generated by each sequencing approach across the terminal exon including the 3′ UTR of an example gene.

(**C**) BAM coverage plots of RNA-seq and Poly(A)-seq completed for Melan-A cells (gray) B16-F10 cells (red) of the *Sap30l* terminal exon and 3′ UTR with the stop codon and the annotated poly(A) signal sequences (PAS).

(**D**) BAM coverage plot of RNA-seq and Poly(A)-seq completed for Melan-A cells (gray) and B16-F10 cells (red) of the *Higd1a* terminal exon and 3′ UTR with the stop codon and the annotated poly(A) signal sequences (PAS).

(**E**) Box plot demonstrating relative 3′ UTR length calculated as the $\log_2$(distal reads / proximal reads) per sample for the gene *Higd1a* and *Sap30l*. *P* values from two-sided Wilcoxon rank-sum test.

(**F**) Volcano plot of all differentially polyadenylated transcripts between B16-F10 cells and Melan-A cells quantified using the APAlyzer pipeline. Data reflects six Poly(A)-seq runs per cell line, significantly altered events were determined using a two-sided Student’s *t*-test. Significantly shortened 3′ UTRs are indicated in blue and significantly lengthened 3′ UTRs are indicated in yellow.

(**G**) Scatter plot of all genes identified as differentially polyadenylated from Poly(A)-seq data, comparing gene-level 3′ UTR length differences and gene expression differences between B16-F10 and Melan-A cells. Blue indicates genes that are significantly shortened in B16-F10 cells and display a significant difference in expression levels, yellow indicates genes that are significantly lengthened and display a significant difference in expression levels, and gray indicates the gene shows no significant change in expression between the two cell lines.

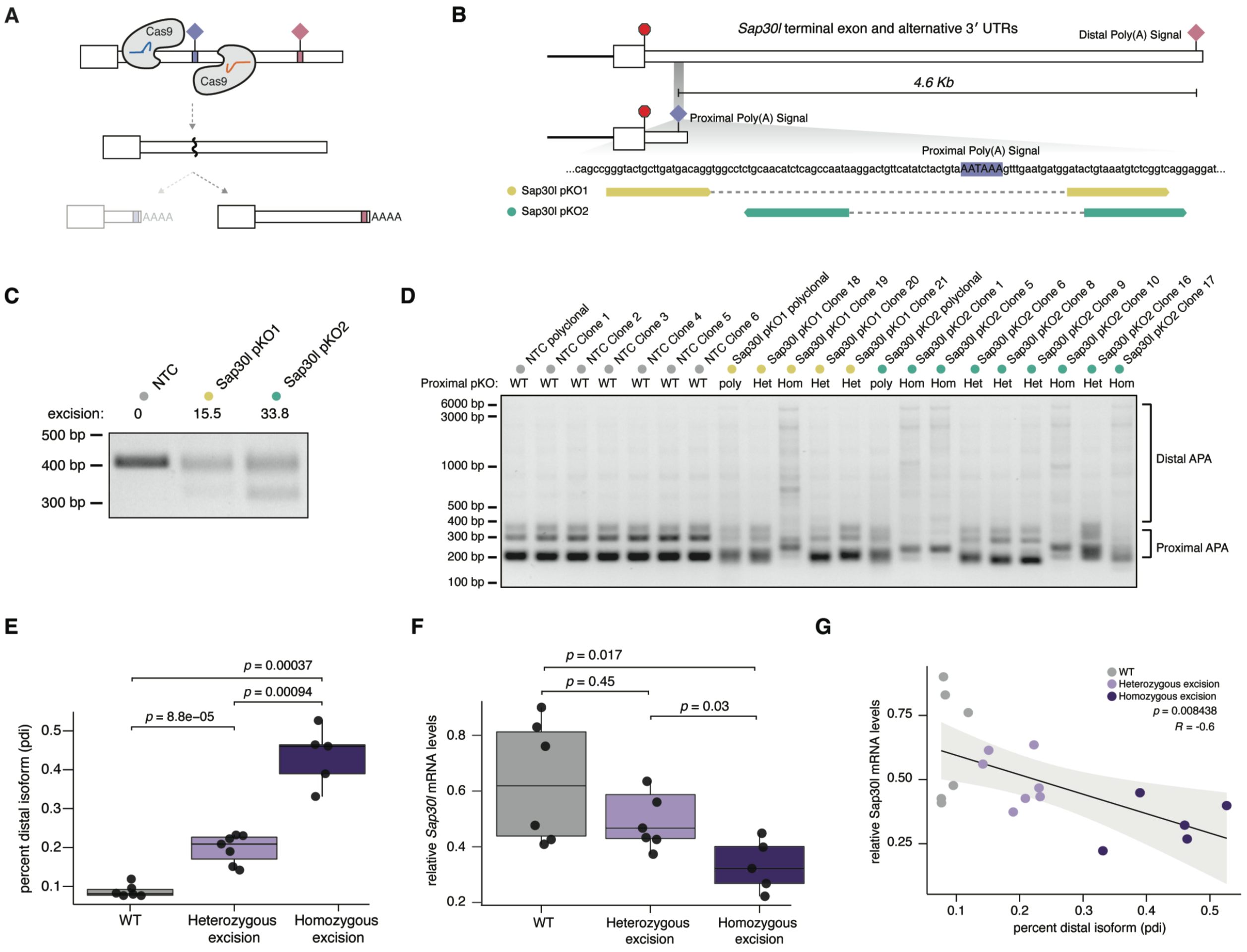


Figure 2.3. A CRISPR-Cas9 paired-guide RNA strategy can force targeted distal poly(A) site usage.

(**A**) Schematic of CRISPR-Cas9 paired-guide RNA (pgRNA) approach to excise proximal poly(A) sites and force use of distal poly(A) sites.
(**B**) Diagram of *Sap30l* terminal exon with two distinct 3′ UTRs which vary depending on use of a proximal or distal poly(A) signal. Schematic of two distinct sets of proximal poly(A) KO (pKO) pgRNAs designed to delete the proximal poly(A) signal in *Sap30l*, annotated as *Sap30l* pKO1 and pKO2.
(**C**) Genotyping PCR of polyclonal B16-F10 Cas9-expressing cells treated with a non-targeting control (NTC) or one of two distinct *Sap30l* pKO pgRNAs, with the calculated percent of signal showing DNA excision.

(**D**) Nested RT-PCR of the *Sap30l* 3′ UTR in polyclonal and monoclonal B16-F10 Cas9-expressing cell lines treated with a non-targeting (NTC) pgRNA, *Sap30l* pKO1 or *Sap30l* pKO2. Proximal poly(A) signal presence is noted as proximal pKO, where WT = two wild type poly(A) signals, Het = single excision of poly(A) signal and Hom = double excision of poly(A) signal.

(**E**) Box plots of the percent distal isoform (pdi) usage per cell line quantified from the nested RT-PCR of the *Sap30l* 3′ UTR for each monoclonal cell line grouped by genotype (WT, heterozygous or homozygous deletion of the proximal poly(A) signal). *P* value from a two-sided Wilcoxon rank-sum test.

(**F**) Box plots of *Sap30l* mRNA abundance per cell line measured by q-RT-PCR for each monoclonal cell line grouped by genotype (WT, heterozygous or homozygous deletion of the proximal poly(A) signal). *P* value from a two-sided Wilcoxon rank-sum test.

(**G**) Scatter plot and Pearson correlation of *Sap30l* percent distal isoform usage versus *Sap30l* mRNA abundance.

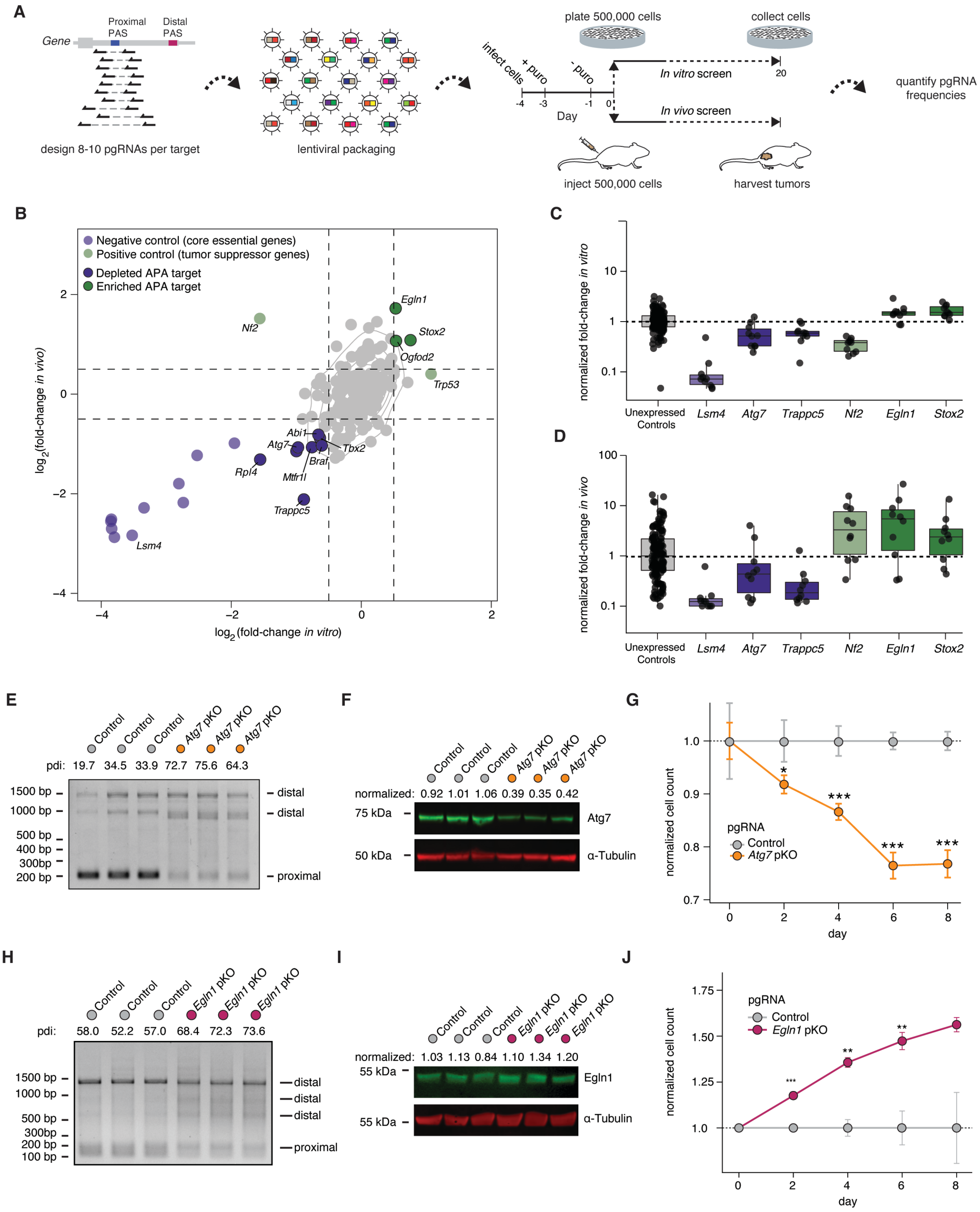

A
Gene
Proximal PAS
Distal PAS
design 8-10 pgRNAs per target
lentiviral packaging
infect cells
+ puro
- puro
-4
-3
-1
0
Day
plate 500,000 cells
collect cells
In vitro screen
20
In vivo screen
inject 500,000 cells
harvest tumors
quantify pgRNA frequencies
B
Negative control (core essential genes)
Positive control (tumor suppressor genes)
Depleted APA target
Enriched APA target
Egln1
Stox2
Ogfod2
Trp53
Nf2
Abi1
Atg7
Tbx2
Braf
Rpl4
Mtfr1l
Trappc5
Lsm4
log2(fold-change in vivo)
log2(fold-change in vitro)
2
0
−2
−4
C
normalized fold-change in vitro
10
1
0.1
Unexpressed Controls
Lsm4
Atg7
Trappc5
Nf2
Egln1
Stox2
D
normalized fold-change in vivo
100
10
1
0.1
Unexpressed Controls
Lsm4
Atg7
Trappc5
Nf2
Egln1
Stox2
E
Control
Control
Control
Atg7 pKO
Atg7 pKO
Atg7 pKO
pdi: 19.7 34.5 33.9 72.7 75.6 64.3
1500 bp
1000 bp
500 bp
400 bp
300bp
200 bp
distal
distal
proximal
F
Control
Control
Control
Atg7 pKO
Atg7 pKO
Atg7 pKO
normalized: 0.92 1.01 1.06 0.39 0.35 0.42
75 kDa
50 kDa
Atg7
α-Tubulin
G
normalized cell count
1.0
0.9
0.8
0.7
pgRNA
Control
Atg7 pKO
*
***
***
***
0
2
4
6
8
day
H
Control
Control
Control
Egln1 pKO
Egln1 pKO
Egln1 pKO
pdi: 58.0 52.2 57.0 68.4 72.3 73.6
1500 bp
1000 bp
500 bp
300bp
200 bp
100 bp
distal
distal
distal
proximal
I
Control
Control
Control
Egln1 pKO
Egln1 pKO
Egln1 pKO
normalized: 1.03 1.13 0.84 1.10 1.34 1.20
55 kDa
55 kDa
Egln1
α-Tubulin
J
normalized cell count
1.75
1.50
1.25
1.0
pgRNA
Control
Egln1 pKO
***
**
**
0
2
4
6
8
day

Figure 2.4. A high-throughput functional CRISPR-Cas9 screen reveals APA events that influence melanoma growth.

(**A**) Schematic of CRISPR-Cas9 pgRNA library design (8-10 pgRNAs per target), cloning and paired *in vitro* and *in vivo* screen format and timing.
(**B**) Mean $\log_2$ (fold-change) of all pgRNAs targeting a specific gene or proximal poly(A) site normalized to 150 pgRNAs targeting proximal poly(A) sites of unexpressed genes compared between paired *in vitro* and *in vivo* screens in B16-F10 Cas9-expressing cells. Significantly enriched (green) or depleted (purple) proximal poly(A) targets are outlined in black. Positive (light green) and negative (light purple) control genes are plotted without black outlines.
(**C**) Performance of 150 pgRNAs targeting proximal poly(A) sites of unexpressed genes or pgRNA performance per indicated growth control or poly(A) site *in vitro*.
(**D**) Performance of 150 pgRNAs targeting proximal poly(A) sites of unexpressed genes or pgRNA performance per indicated growth control or poly(A) site *in vivo*.
(**E**) Nested RT-PCR of the *Atg7* 3′ UTR of B16-F10 Cas9-expressing cells treated with a pgRNA targeting a proximal poly(A) site in an unexpressed gene (control) or *Atg7* proximal poly(A) site knock out (*Atg7* pKO). Percent distal isoform (pdi) is quantified per sample.
(**F**) Western blot of lysates from B16-F10 Cas9-expressing cells treated with a control pgRNA or *Atg7* pKO pgRNA. Atg7 protein level normalized to alpha-tubulin control.
(**G**) *In vitro* cell growth of B16-F10 Cas9-expressing cells treated with a control pgRNA or *Atg7* pKO pgRNA measured by CellTitre Glo, measurement is the average of three replicates +/- standard error of the mean. Significance denoted as *$p < 0.05$, **$p < 0.01$ or ***$p < 0.001$ using a two-sided Student's *t*-test (exact *p* values 0.023, 0.00011, $9\text{x}10^{-8}$, and 0.00096).
(**H**) Nested RT-PCR of the *Egln1* 3′ UTR of B16-F10 Cas9-expressing cells treated with a pgRNA targeting a proximal poly(A) site in an unexpressed gene (control) or *Egln1* proximal poly(A) site knock out (*Egln1* pKO). Percent distal isoform (pdi) is quantified per sample.
(**I**) Western blot of lysates from B16-F10 Cas9-expressing cells treated with a control pgRNA or *Egln1* pKO pgRNA. Egln1 protein level normalized to alpha-tubulin control.
(**J**) *In vitro* cell growth of B16-F10 Cas9-expressing cells treated with a control pgRNA or *Egln1* pKO pgRNA measured by CellTiter-Glo. Measurement is the average of three replicates +/- standard error of the mean. Significance denoted as *$p < 0.05$, **$p < 0.01$ or ***$p < 0.001$ using a two-sided Student's *t*-test (exact *p* values 0.0005, 0.0011, 0.0079, and 0.059).

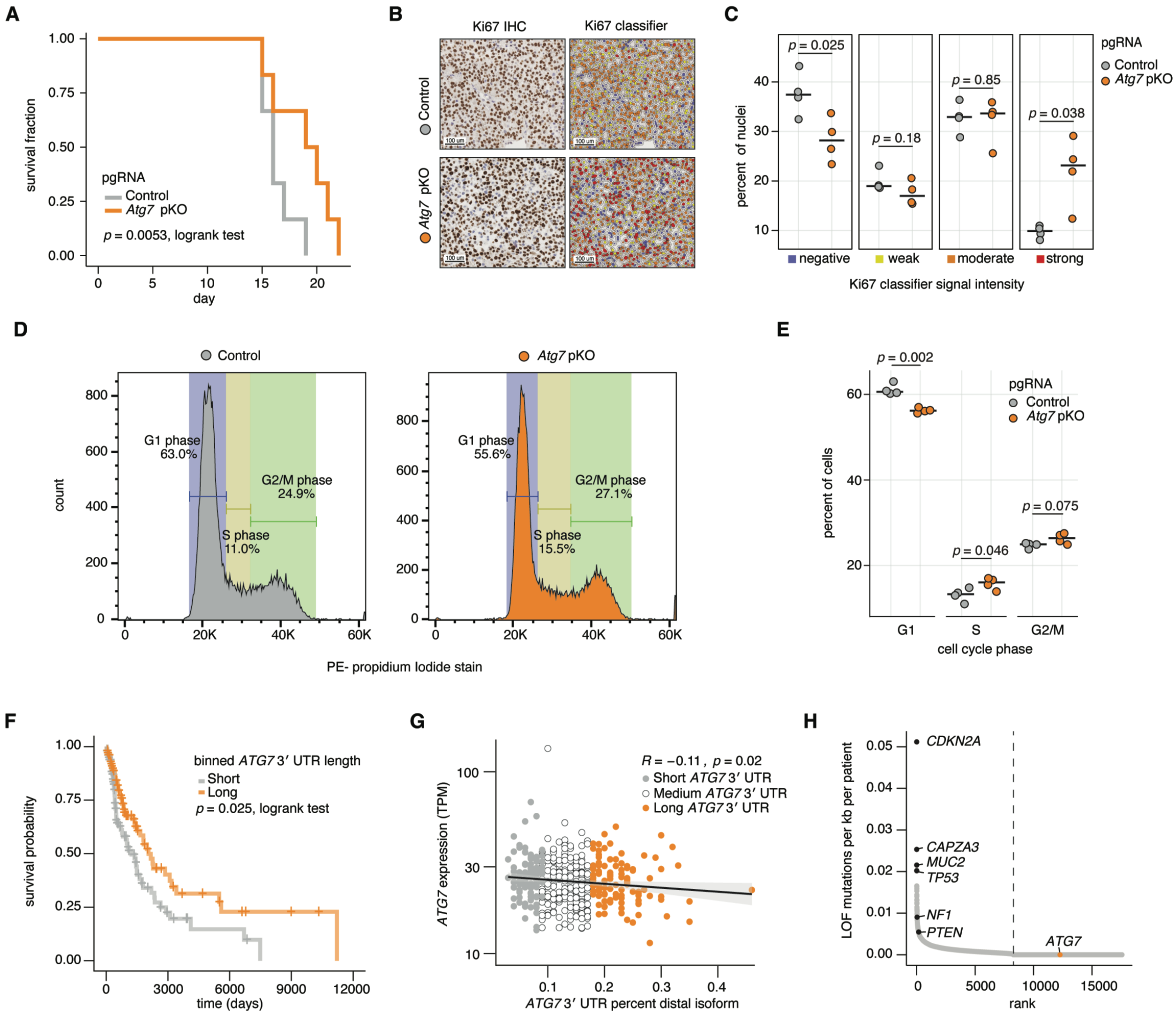


Figure 2.5. Atg7 alternative poly(A) site selection alters melanoma cell growth in *vitro* and in *vivo*.

(**A**) Survival data from a cohort of C57BL/6 mice injected with B16-F10 Cas9-expressing cells treated with a control pgRNA targeting the poly(A) site of an unexpressed gene (control) or an *Atg7* pKO pgRNA (n = 6 mice, 12 tumors, per condition).

(**B**) Representative immunohistochemistry images of tumor B16-F10 Cas9-expressing control or *Atg7* pKO pgRNA tumor sections stained for Ki67. Those same representative images now with nuclei classified using HALO software to count each nucleus as either negative (blue), weak (yellow), moderate (orange) or strong (red) staining for Ki67 signal.

(**C**) Quantification of percent of nuclei quantified as negative, weak, moderate or strong staining for Ki67 signal (n = 4 images from 4 distinct tumors per genotype). For each image, the entire slide is processed, only excluding areas if they are easily discernible as non-tumor tissue. Significance calculated with a two-sided Student's *t*-test.

(**D**) Representative flow cytometry histograms of B16-F10 Cas9-expressing cells treated with a control or *Atg7* pKO pgRNA cells stained with propidium iodide and then analyzed using Dean-Jett-Fox classification for cell cycle stage from FlowJo v10.

(**E**) Quantification of cell cycle stage from propidium iodide staining and flow cytometry (n = 4 samples per genotype). Significance calculated with a two-sided Student's *t*-test.

(**F**) Kaplan-Meier survival analysis of TCGA cutaneous melanoma cohort (SKCM) patients binned into terciles based on *ATG7* 3′ UTR length, and then comparing the short (gray) and long (orange) *ATG7* 3′ UTR bins in terms of progression-free survival. *P* value from a two-sided logrank test.

(**G**) Scatter plot comparing *ATG7* 3′ UTR length represented as the percent distal isoform use index, where a value of 0 indicates all transcripts use the proximal poly(A) site and a value of 1 indicates all transcripts use the distal poly(A) site, and *ATG7* gene expression (TPM) for each of 424 TCGA cutaneous melanoma samples analyzed. *R* and *p* values are from a Pearson correlation.

(**H**) Rank order plot of loss-of-function (LOF) mutations normalized per kilobase per patient detected in genomic DNA sequencing from the TCGA cutaneous melanoma cohort of 424 patients analyzed. Vertical dashed line indicates the point at which all remaining ranked genes have no detected LOF mutations per kilobase per patient in the cutaneous melanoma cohort.

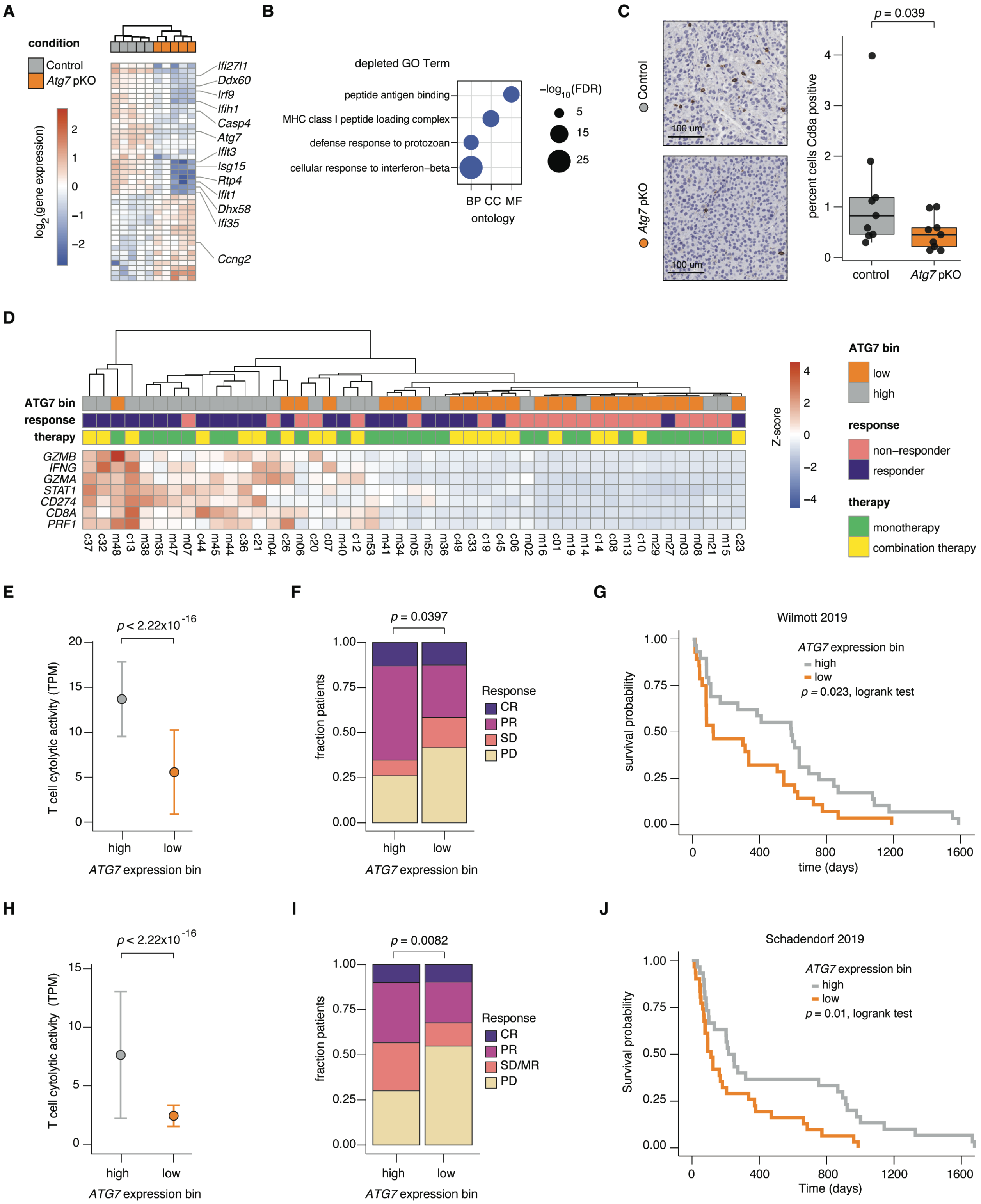


Figure 2.6. ATG7 expression influences the tumor-immune microenvironment and correlates with response to immune checkpoint blockade therapy in clinical melanoma.

(**A**) Heatmap of differentially expressed genes identified from RNA-seq between B16-F10 Cas9-expressing control and *Atg7* pKO pgRNA tumors (n = 5 samples per genotype). Each box is the $\log_2$(gene-expression) difference from the median across all 10 samples.

(**B**) Gene Ontology (GO) analysis of differentially expressed genes between B16-F10 Cas9-expressing control and *Atg7* pKO pgRNA tumors.

(**C**) Representative immunohistochemistry images of tumor B16-F10 Cas9-expressing control pgRNA or *Atg7* pKO pgRNA tumor sections stained for CD8A. CD8A$^+$ cells were classified using HALO software to count each cell as either negative or positive for signal. Box plots represent percent of CD8A$^+$ cells from 9 immunohistochemistry images (each from a distinct tumor) per genotype, where the entire slide was quantified. *P* value from a one-sided Wilcoxon rank-sum test.

(**D**) Heatmap for Wilmott 2019 cohort showing Z-scores for each of seven immune checkpoint blockade relevant genes across 47 samples. Each sample is labeled as receiving either monotherapy or combination therapy (noted in sample identification number with an m for monotherapy or a c for combination therapy), whether clinical response was deemed a responder (CR, PR) or non-responder (SD, PD), and the *ATG7* expression bin (high or low representing the highest 33% or lowest 33% *ATG7* expression, respectively).

(**E**) Median T cell cytolytic activity (geometric mean of *PRF1* and *GZMB* expression) for samples stratified into high or low *ATG7* expression bins +/- 95% confidence intervals obtained from bootstrapping for the Wilmott 2019 cohort. *P* value from a two-sided Wilcoxon rank-sum test.

(**F**) Stacked bar plot of clinical responses of patients with high or low *ATG7* expression as defined by Response Evaluation Criteria in Solid Tumors (RECIST) criteria from the Wilmott 2019 cohort. *P* values were calculated using a two-sided multinomial proportion test.

(**G**) Kaplan-Meier survival analysis of the Wilmott 2019 cohort comparing progression-free survival of patients with high and low *ATG7* expression*,* *p* value from a two-sided logrank test.

(**H**) Median T cell cytolytic activity (geometric mean of *PRF1* and *GZMB* expression) for samples stratified into high or low *ATG7* expression bins +/- 95% confidence intervals obtained from bootstrapping for the Schadendorf 2019 cohort. *P* value from a two-sided Wilcoxon rank-sum test.

(**I**) Stacked bar plot of clinical responses of patients with high or low *ATG7* expression as defined by Response Evaluation Criteria in Solid Tumors (RECIST) criteria from the Schadendorf 2019 cohort. *P* values were calculated using a two-sided multinomial proportion test.

(**J**) Kaplan-Meier survival analysis of the Schadendorf 2019 cohort comparing progression-free survival of patients with high and low *ATG7* expression, *p* value from a two-sided logrank test.

2.6 SUPPLEMENTAL FIGURES

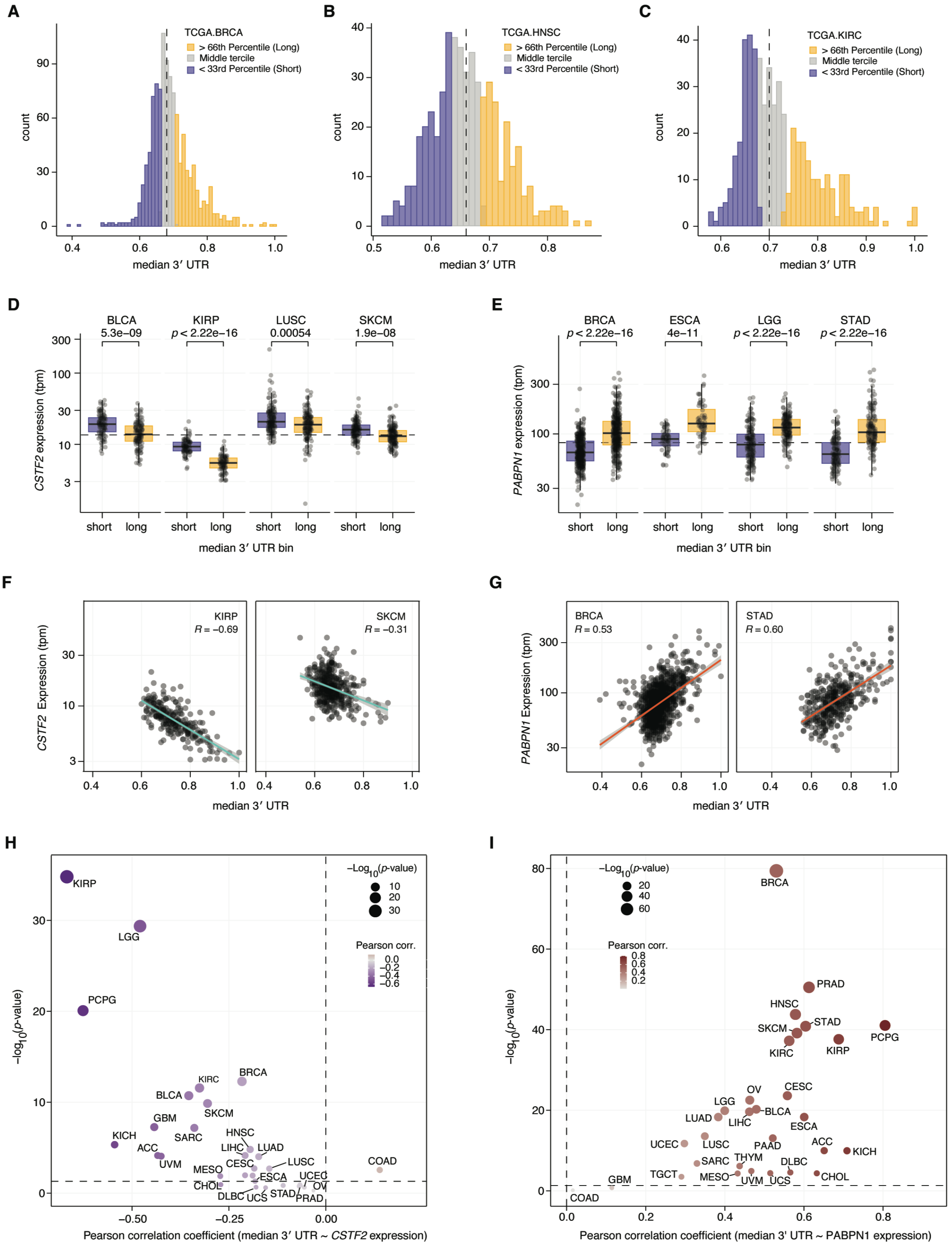


Figure 2.7. Supplemental Figure 1. Median 3′ UTR correlates with known global regulators of 3′ UTR length.

(**A-C**) Distributions and binning of median 3′ UTR length for invasive breast carcinoma (A), head and neck squamous cell carcinoma (B), and kidney renal cell carcinoma (KIRC), calculated using the DaPars algorithm to quantify gene-level 3′ UTR lengths per sample (Xia et al., 2014).
(**D**) Box plot comparing *CSTF2* expression (TPM) per median 3′ UTR bin for four TCGA cancer subtypes, bladder urothelial carcinoma (BLCA), kidney renal papillary cell carcinoma (KIRP), lung squamous cells carcinoma (LUSC), and cutaneous melanoma (SKCM). *P* values calculated with a two-sided Wilcoxon rank-sum test.
(**E**) Box plot comparing *PABPN1* expression (TPM) per median 3′ UTR bin for four TCGA cancer subtypes, invasive breast carcinoma (BRCA), esophageal carcinoma (ESCA), brain lower grade glioma (LGG), and stomach adenocarcinoma (STAD). *P* values calculated with a two-sided Wilcoxon rank-sum test.
(**F**) Scatter plots of median 3′ UTR length versus *CSTF2* expression (TPM) per sample for two TCGA cancer subtypes, kidney renal papillary cell carcinoma (KIRP) and cutaneous melanoma (SKCM). *R* values from Pearson correlation.
(**G**) Scatter plots of median 3′ UTR length versus *PABPN1* expression (TPM) per sample for two TCGA cancer subtypes, invasive breast carcinoma (BRCA) and stomach adenocarcinoma (STAD). *R* values from Pearson correlation.
(**H**) Volcano plot of the Pearson correlation per TCGA cancer subtype plotted against the -$\log_{10}$(*p* value) of that Pearson correlation coefficient for median 3′ UTR length versus *CSTF2* expression (TPM).
(**I**) Volcano plot of the Pearson correlation per TCGA cancer subtype plotted against the -$\log_{10}$(*p* value) of that Pearson correlation coefficient for median 3′ UTR length versus *PABPN1* expression (TPM).

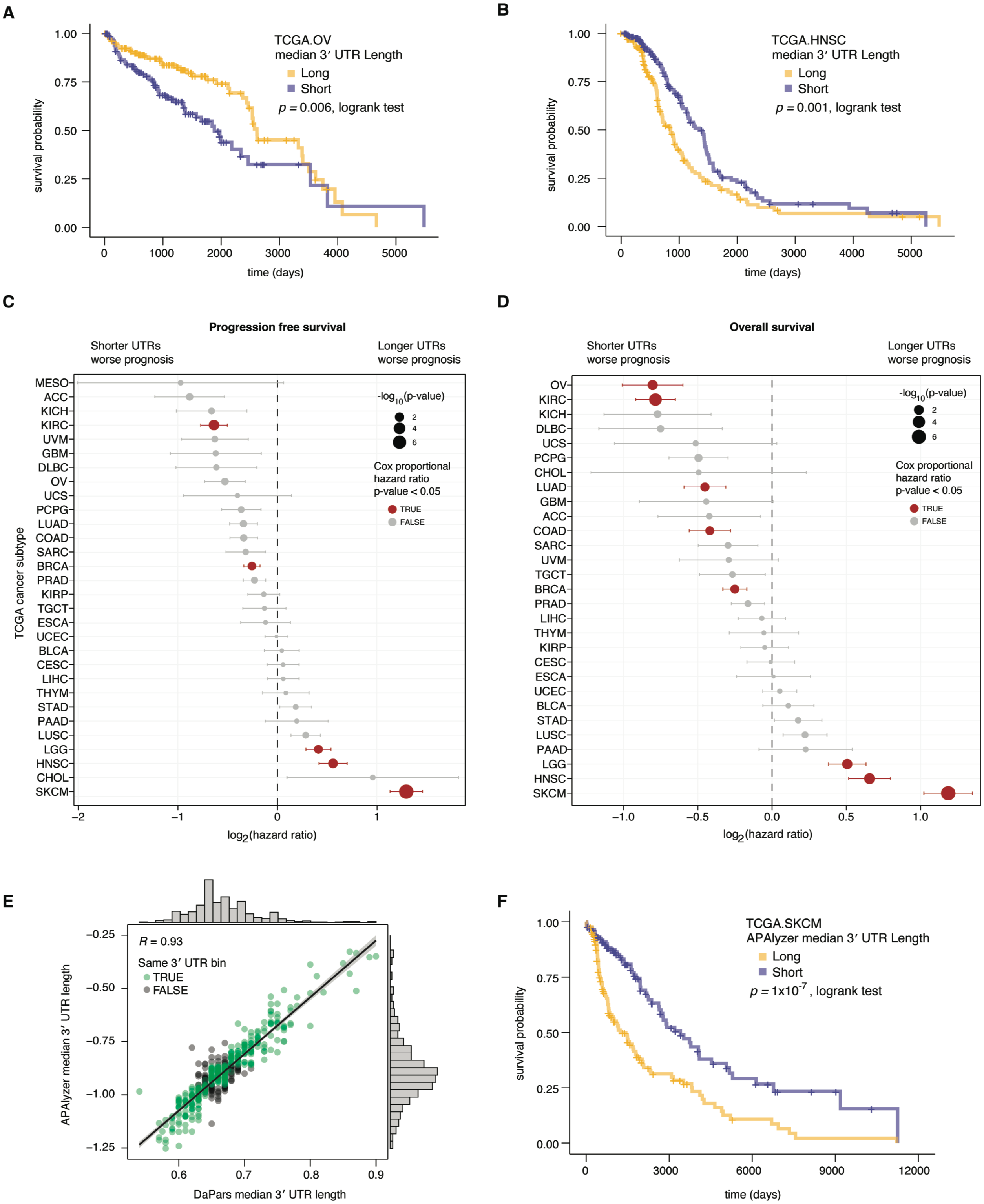


Figure 2.8. Supplemental Figure 2. Global median 3′ UTR length correlates with patient outcomes in several cancer subtypes.

(**A**) Kaplan-Meier analysis comparing overall survival of TCGA ovarian serous cystadenocarcinoma (OV) samples binned as short or long median 3′ UTR samples. *P* values from a two-sided logrank test.

(**B**) Kaplan-Meier analysis comparing overall survival of TCGA head and neck squamous cell carcinoma (HNSC) samples binned as short or long median 3′ UTR samples. *P* values from a two-sided logrank test.

(**C-D**) Forest plot of the $\log_2$(hazard ratio) per TCGA cancer subtype comparing progression-free survival (C) or overall survival (D) for long versus short median median 3′ UTR bin. Point sizes are scaled to $-\log_{10}$(*p* value) of the hazard ratio per subtype.

(**E**) Scatter plot comparing median 3′ UTR length calculated using DaPars (Xia et al., 2014) or APAlyzer (Wang & Tian, 2020) computational pipelines, *R* from Pearson correlation. Points are colored by if they are stratified into the same bin (short, medium or long) by both computational algorithms, where green indicates they are classified similarly by both algorithms and black indicates they are not.

(**F**) Kaplan-Meier analysis comparing overall survival of cutaneous melanoma samples with shorter versus longer median 3′ UTRs calculated using APAlyzer, *P* values from a two-sided logrank test.

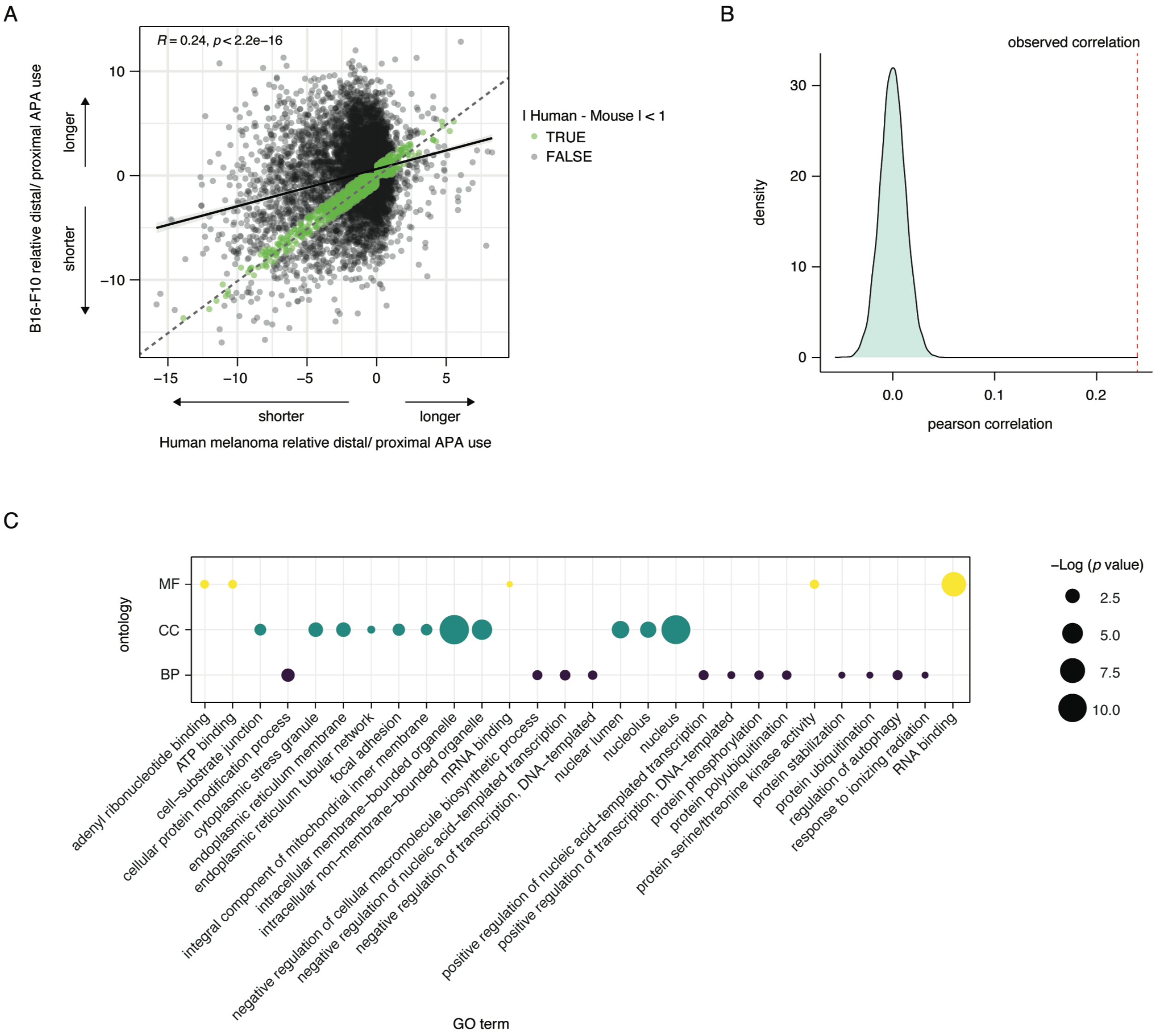


Figure 2.9. Supplemental Figure 3. APA in human and mouse melanoma are correlated.

(**A**) Gene level 3′ UTR lengths correlated between human clinical melanoma samples and B16-F10 mouse melanoma cells both calculated using APAlyzer, each value reflects the log2 (distal reads / proximal reads) for that given gene ortholog in mouse and human. *R* and *P* value calculated from Pearson correlation. Dotted line indicates x = y and points colored in green are genes where the absolute difference between the mouse and human values are less than 1, indicating the 3′ UTR lengths are similar in mouse and human.

(**B**) Distribution of Pearson coefficients obtained from 10,000 random pairings of gene level 3′ UTR lengths correlated between Human clinical melanoma samples and B16-F10 mouse melanoma cells (data from panel A), dashed red line indicates the observed *R* from Pearson correlation of the actual data.

(**C**) GO analysis of genes where the 3′ UTR lengths are highly correlated between human and mouse melanoma (indicated in green in panel A).

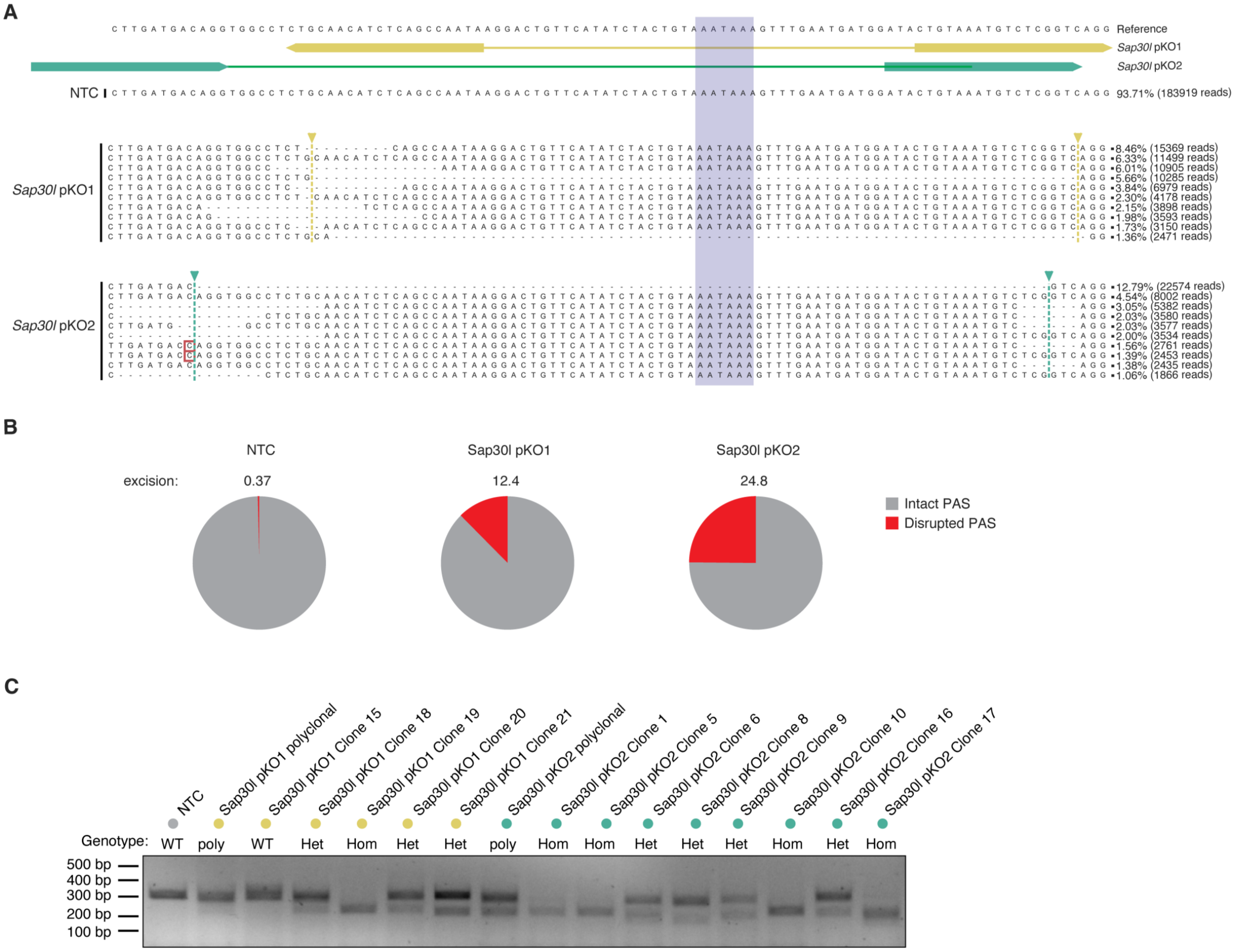


Figure 2.10. Supplemental Figure 4. CRISPR-Cas9 paired-guide RNAs can be used to delete polyadenylation signals.

(**A**) Schematic of *Sap30l* proximal poly(A) sequence targeted pgRNAs, *Sap30l* pKO1 and pKO2. Next-generation sequencing results showing the major detected alleles from three polyclonal B16-F10 Cas9-expressing cells treated with either a non-targeting control (NTC), *Sap30l* pKO1 or *Sap30l* pKO2 pgRNA as indicated and submitted for next-generation sequencing of the target locus.

(**B**) Pie chart demonstrating the fraction of next-generation sequencing reads with an intact or disrupted polyadenylation signal sequence for each cell line treated with the indicated pgRNA.

(**C**) Genotyping PCR of genomic DNA obtained from the indicated polyclonal or monoclonal or monoclonal cell line confirming single or dual KO of the proximal poly(A) signal sequence (lower band indicates excision) and genotypes indicated as wild-type (WT), heterozygous excision (Het) or homozygous excision (Hom).

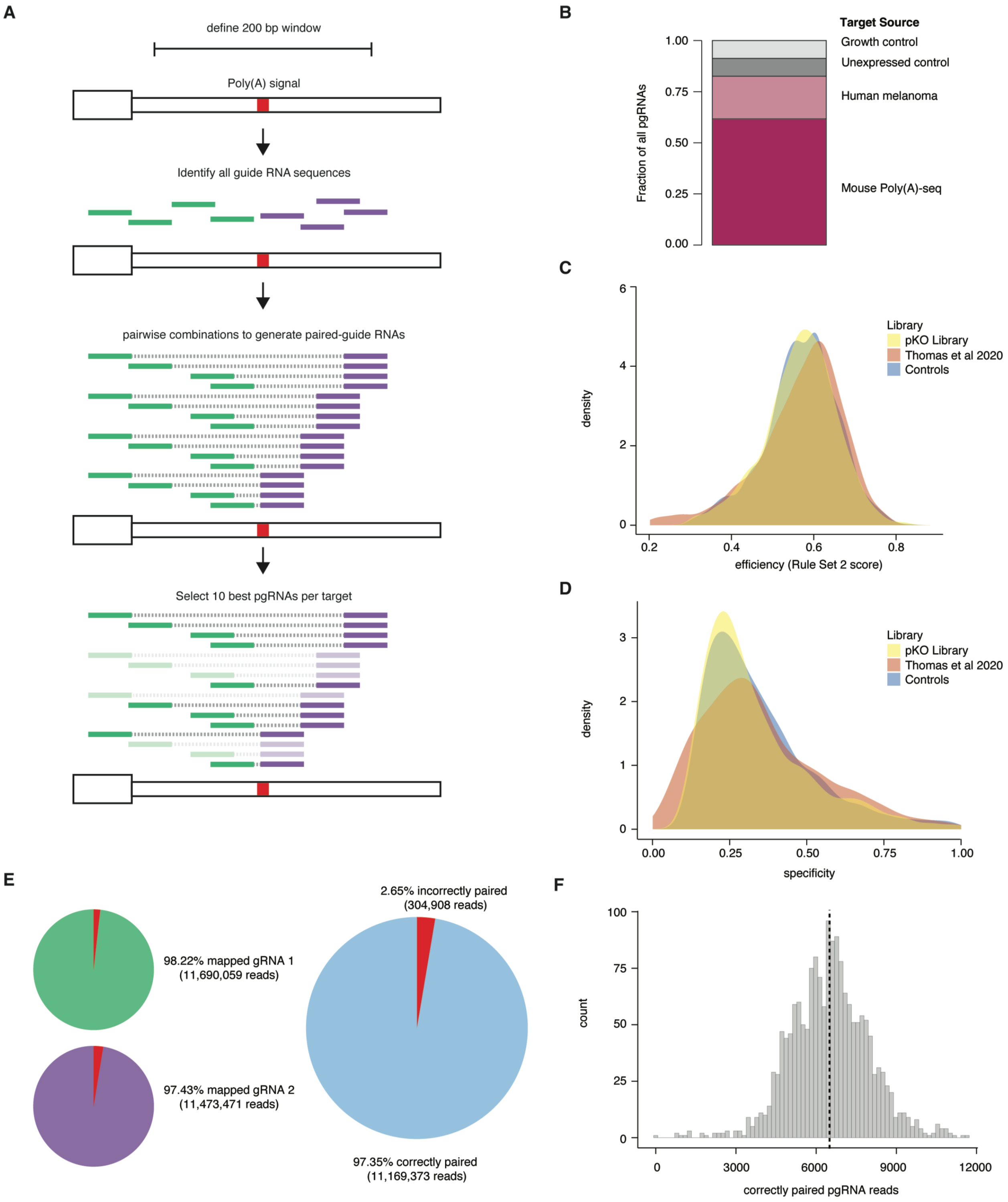
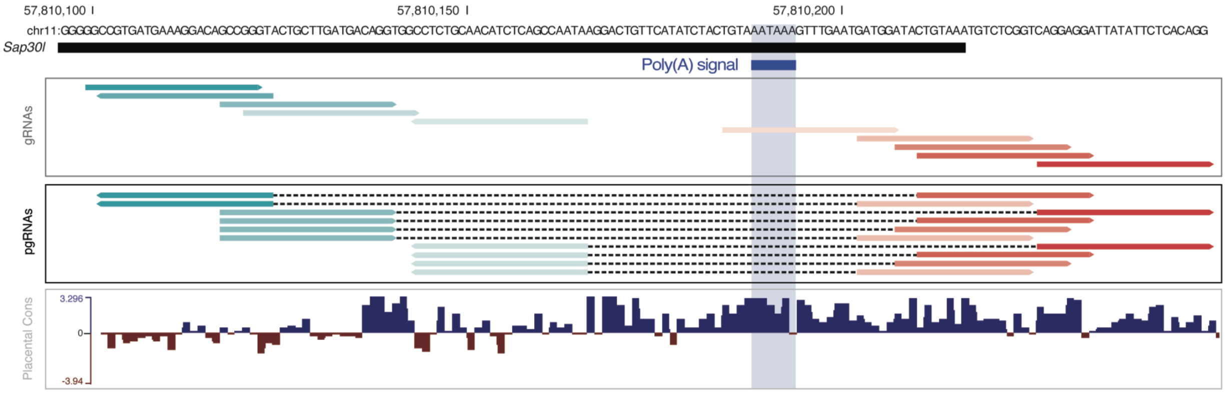

A
define 200 bp window
Poly(A) signal
Identify all guide RNA sequences
pairwise combinations to generate paired-guide RNAs
Select 10 best pgRNAs per target
B
Fraction of all pgRNAs
1.00
0.75
0.50
0.25
0.00
Target Source
Growth control
Unexpressed control
Human melanoma
Mouse Poly(A)-seq
C
density
6
4
2
0
0.2
0.4
0.6
0.8
efficiency (Rule Set 2 score)
Library
pKO Library
Thomas et al 2020
Controls
D
density
3
2
1
0
0.00
0.25
0.50
0.75
1.00
specificity
Library
pKO Library
Thomas et al 2020
Controls
E
98.22% mapped gRNA 1
(11,690,059 reads)
97.43% mapped gRNA 2
(11,473,471 reads)
2.65% incorrectly paired
(304,908 reads)
97.35% correctly paired
(11,169,373 reads)
F
count
100
75
50
25
0
0
3000
6000
9000
12000
correctly paired pgRNA reads

Figure 2.11. Supplemental Figure 5. Rational design of a CRISPR-Cas9 paired-guide RNA library to target polyadenylation signals.

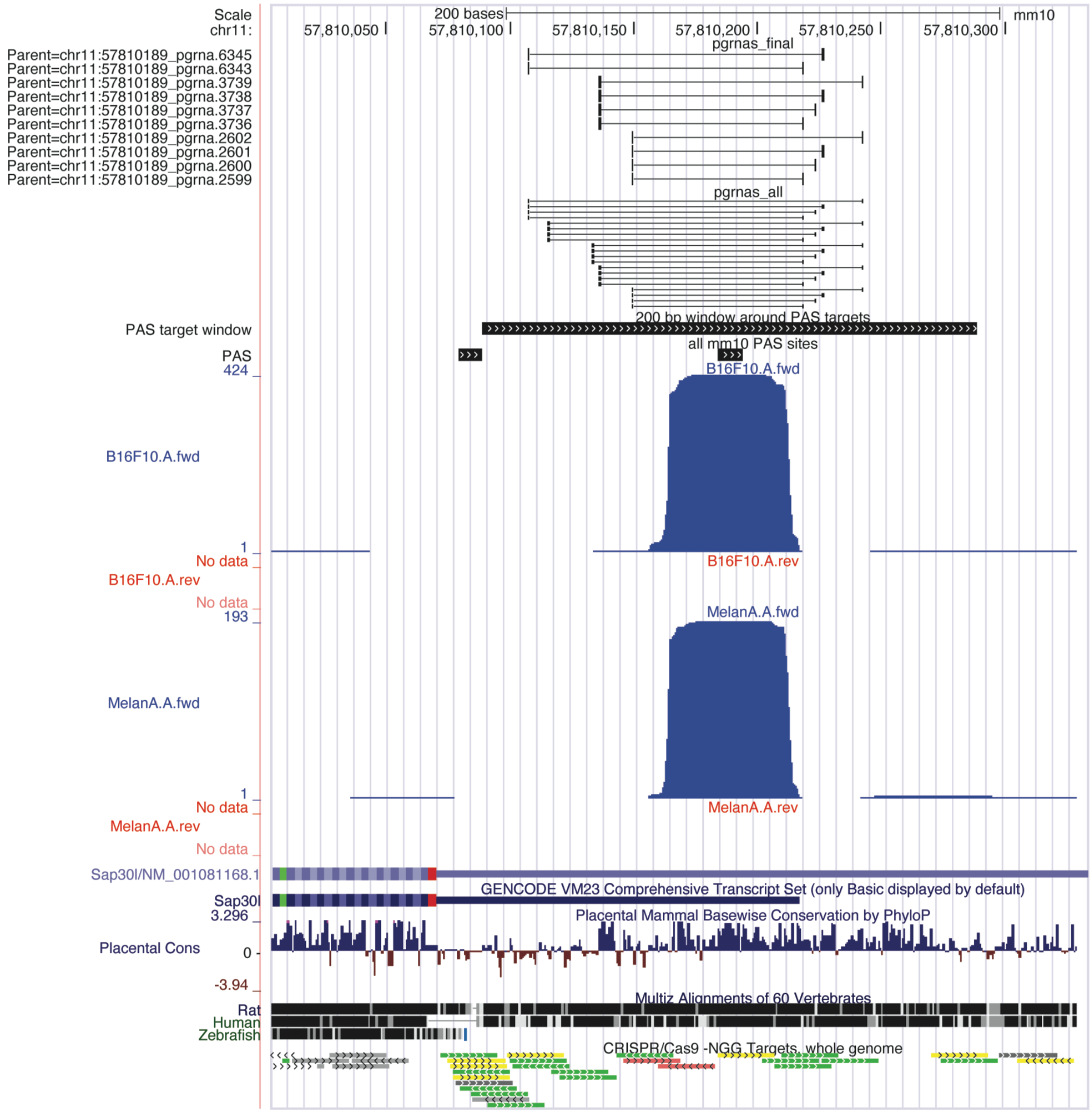
(**A**) Diagram of computational steps taken to generate a CRISPR-Cas9 paired-guide RNA library to delete proximal polyadenylation signal sequences of interest.
(**B**) Library divided into the fraction of total pgRNAs which were selected based on how the specific target was selected to be included, referred to as the target source (n = 1718 pgRNAs).
(**C**) Distribution of on-target efficiency, using Rule Set 2 scores (Doench et al., 2016) per pgRNA for the pgRNAs targeting proximal poly(A) sites (pKO Library) compared to a previously published pgRNA library (Thomas et al. 2020) or the 150 pgRNAs targeting poly(A) sites in unexpressed genes (Controls), which were randomly sampled from all possible control pgRNAs to match the distribution of on and off-target scores of the pKO library.
(**D**) Distribution of target specificity per pgRNA for the pgRNAs targeting proximal poly(A) sites (pKO Library) compared to a previously published pgRNA library (Thomas et al. 2020) or the pgRNAs targeting poly(A) sites in unexpressed controls.
(**E**) Next-generation sequencing of the final plasmid pool of the cloned proximal poly(A) site KO library. The number of reads and the fraction of those reads correctly mapping to gRNA1, gRNA2 and then the correct pairing of those gRNAs together are shown.
(**F**) Histogram of the number of reads with the correct gRNA mapping and pairing per each of the 1718 pgRNAs included in the final library. Dotted line indicates median number of reads across all pgRNAs.

A



57,810,100
57,810,150
57,810,200
chr11:GGGGGCCGTGATGAAAGGACAGCCGGGTACTGCTTGATGACAGGTGGCCTCTGCAACATCTCAGCCAATAAGGACTGTTCATATCTACTGTAAATAAAGTTTGAATGATGGATACTGTAAATGTCTCGGTCAGGAGGATTATATTCTCACAGG
Sap30l
Poly(A) signal
gRNAs
pgRNAs
Placental Cons
3.296
0
-3.94

B



Scale
chr11:
200 bases
mm10
57,810,050
57,810,100
57,810,150
57,810,200
57,810,250
57,810,300
pgrnas_final
Parent=chr11:57810189_pgrna.6345
Parent=chr11:57810189_pgrna.6343
Parent=chr11:57810189_pgrna.3739
Parent=chr11:57810189_pgrna.3738
Parent=chr11:57810189_pgrna.3737
Parent=chr11:57810189_pgrna.3736
Parent=chr11:57810189_pgrna.2602
Parent=chr11:57810189_pgrna.2601
Parent=chr11:57810189_pgrna.2600
Parent=chr11:57810189_pgrna.2599
pgrnas_all
PAS target window
200 bp window around PAS targets
PAS
all mm10 PAS sites
424
B16F10.A.fwd
1
No data
B16F10.A.rev
193
MelanA.A.fwd
MelanA.A.rev
Sap30l/NM_001081168.1
GENCODE VM23 Comprehensive Transcript Set (only Basic displayed by default)
Sap30l
3.296
Placental Mammal Basewise Conservation by PhyloP
Placental Cons
0
-3.94
Multiz Alignments of 60 Vertebrates
Rat
Human
Zebrafish
CRISPR/Cas9 -NGG Targets, whole genome

Figure 2.12. Supplemental Figure 6. Visualization of proximal poly(A) knock out library.

(**A**) *Sap30l* 3′ UTR DNA sequence with the annotated proximal poly(A) signal sequence, individual gRNAs identified within a 200 base-pair window centered on the poly(A) signal sequence, the 10 selected final pgRNAs designed to delete the proximal poly(A) site and the genetic conservation per base-pair across placental mammals.
(**B**) *Sap30l* 3′ UTR DNA sequence in the UCSC Genome Browser window. The session contains several tracks including annotated poly(A) signal sites (PAS), all pgRNAs generated within the 200 base-pair window per poly(A) site and the final selected 10 pgRNAs. There are also 4 tracks of Poly(A- seq mapped in a stranded fashion from Melan-A and B16-F10 cells (a + and - track for each cell line).

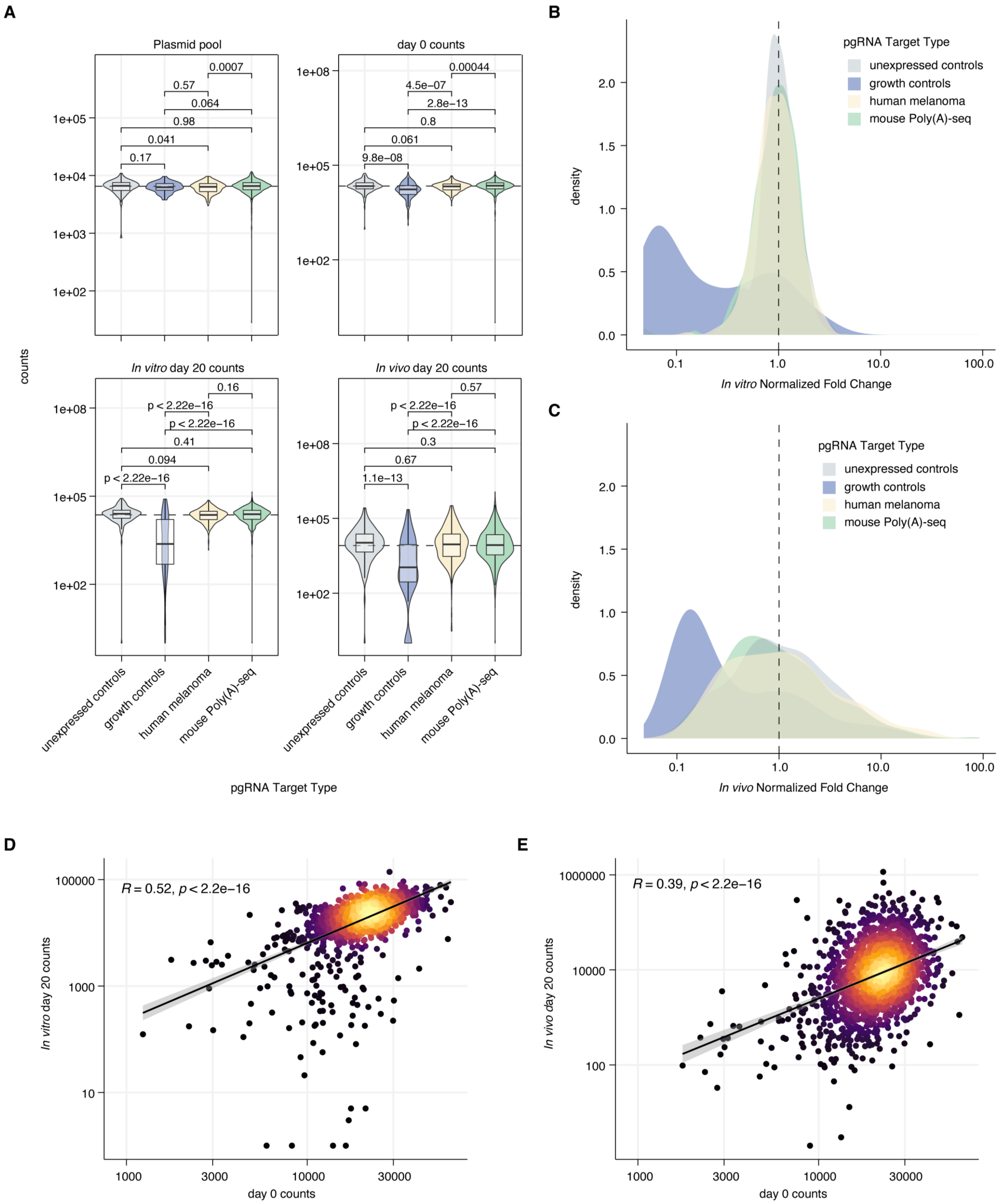


Figure 2.13. Supplemental Figure 7. Performance of CRISPR-Cas9 pgRNA screening platform.

(**A**) Violin plots of all raw counts per each pgRNA at each experimental time point (plasmid pool, day 0, *in vitro* day 20 or *in vivo* day 20) separated by the target type. *P* values from two-sided Wilcoxon rank-sum tests.

(**B**) Density plot of the *in vitro* fold-change of each pgRNA separated by target type, all normalized to the median of 150 pgRNAs targeting poly(A) sites in unexpressed genes. Dotted line at the normalized median of the control pgRNAs which was set to 1.

(**C**) Density plot of the *in vivo* fold-change of each pgRNA separated by target type, all normalized to the median of 150 pgRNAs targeting poly(A) sites in unexpressed genes. Dotted line at the normalized median of the control pgRNAs which was set to 1.

(**D**) Scatter plot of day 0 pgRNA reads versus *in vitro* day 20 pgRNA reads (summed counts across 8 replicates), *R* and *P* value from Pearson correlation.

(**E**) Scatter plot of day 0 pgRNA reads versus *in vivo* day 20 pgRNA reads (summed counts across 8 replicates), *R* and *P* value from Pearson correlation.

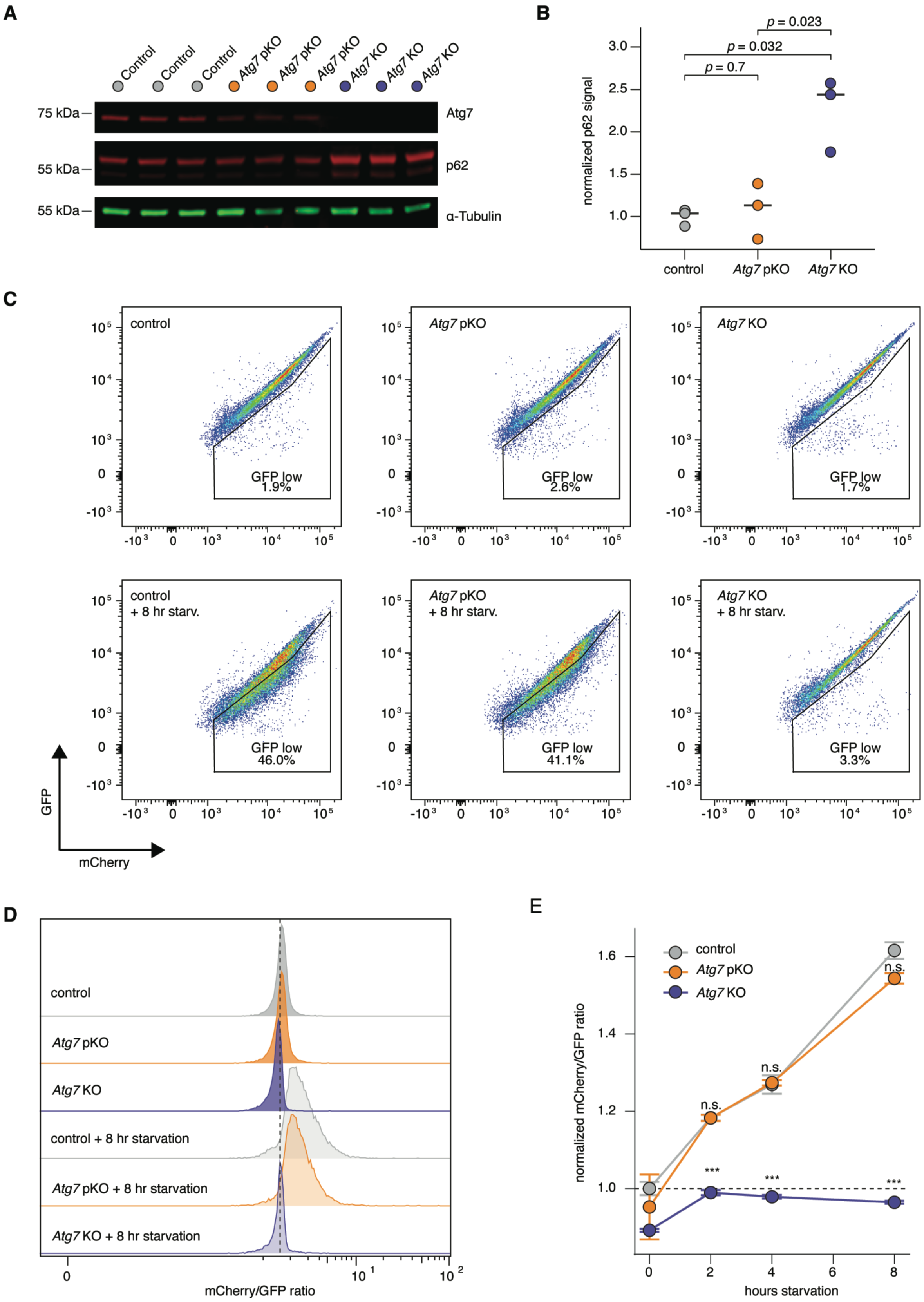

A
Control
Control
Control
Atg7 pKO
Atg7 pKO
Atg7 pKO
Atg7 KO
Atg7 KO
Atg7 KO
75 kDa
Atg7
55 kDa
p62
55 kDa
α-Tubulin
B
normalized p62 signal
3.0
2.5
2.0
1.5
1.0
p = 0.023
p = 0.032
p = 0.7
control
Atg7 pKO
Atg7 KO
C
control
Atg7 pKO
Atg7 KO
GFP low 1.9%
GFP low 2.6%
GFP low 1.7%
control + 8 hr starv.
Atg7 pKO + 8 hr starv.
Atg7 KO + 8 hr starv.
GFP low 46.0%
GFP low 41.1%
GFP low 3.3%
10⁵
10⁴
10³
0
-10³
GFP
mCherry
D
control
Atg7 pKO
Atg7 KO
control + 8 hr starvation
Atg7 pKO + 8 hr starvation
Atg7 KO + 8 hr starvation
0
10¹
10²
mCherry/GFP ratio
E
control
Atg7 pKO
Atg7 KO
normalized mCherry/GFP ratio
1.6
1.4
1.2
1.0
n.s.
n.s.
n.s.
***
***
***
0
2
4
6
8
hours starvation

Figure 2.14. Supplemental Figure 8. Melanoma cells with lengthened *Atg7* 3′ UTRs display no obvious autophagy deficits.

(**A**) Immunoblot of protein collected from B16-F10 Cas9-expressing cells treated with either a control, *Atg7* pKO or *Atg7* gene KO pgRNA completed in triplicate. Western blot is stained for Atg7 and p62/Sqstm1, as well as alpha-tubulin as a loading control.
(**B**) Quantification of p62/Sqstm1 protein levels per genotype, normalized to loading control and then normalized to the median of B16-F10 Cas9-expressing cells treated with a control pgRNA. *P* values from a two-sided Student's *t*-test.
(**C**) B16-F10 Cas9-expressing cells treated with control, *Atg7* pKO or *Atg7* gene KO pgRNAs were stably integrated with a LC3-mCherry-GFP autophagy reporter. mCherry and GFP signals were measured by flow cytometry following different durations of serum and amino acid starvation. Representative flow plots of live, single cells showing mCherry and GFP signal per genotype with 0 and 8 hours of amino acid and serum starvation.
(**D**) Representative histograms of mCherry/GFP signal ratio per genotype with 0 and 8 hours of starvation, cells are gated for live, single cells.
(**E**) Line plots of normalized mCherry/GFP signal in B16-F10 Cas9-expressing cells stably expressing the LC3-mCherry-GFP autophagy reporter treated with the indicated pgRNAs and increasing time under starvation conditions. Each point is the mean of 3 replicates +/- SEM, *P* value from two-sided Student's *t*-test.

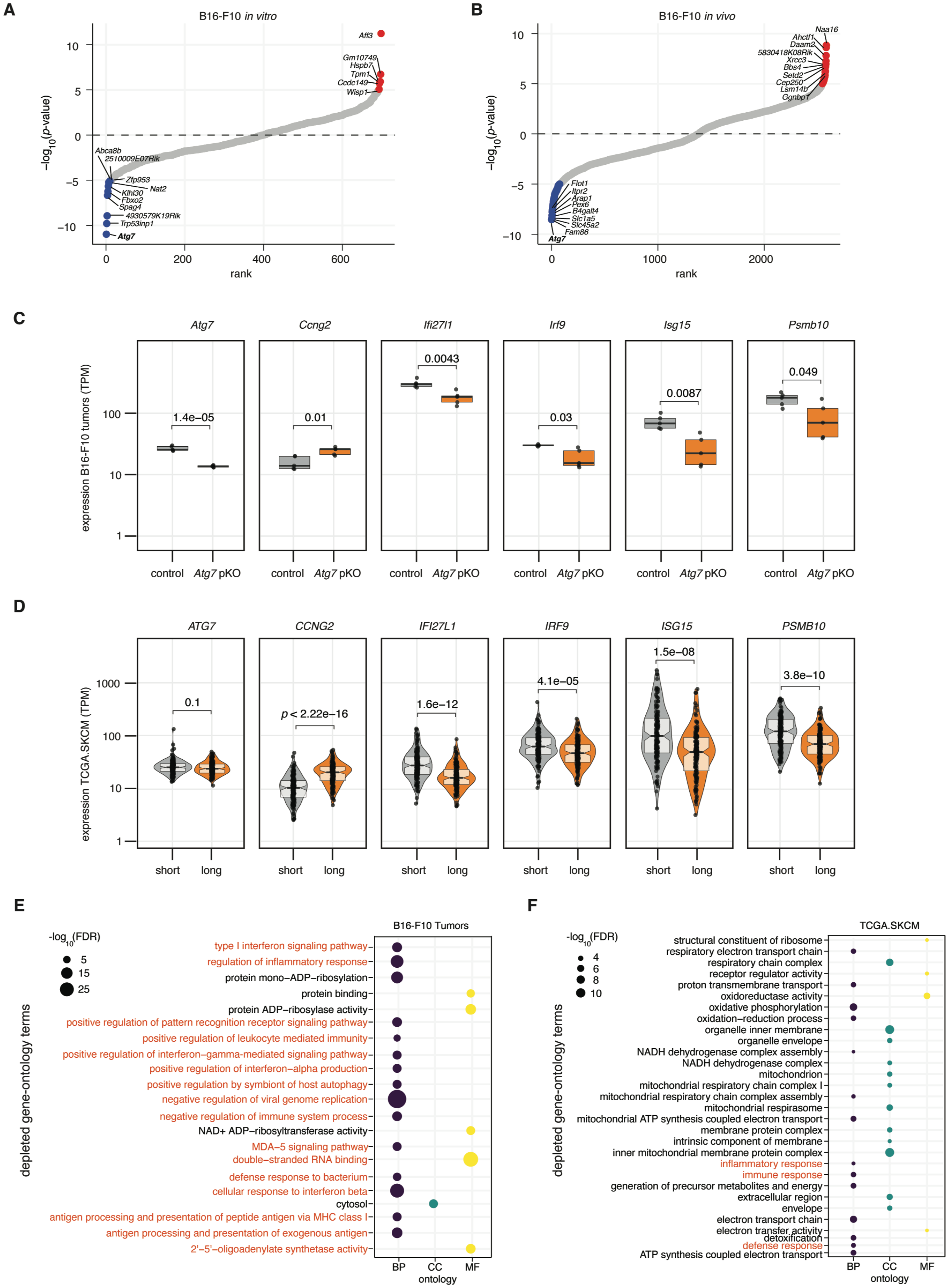

A
B16-F10 in vitro
-log10(p-value)
rank
10
5
0
−5
−10
0
200
400
600
Aff3
Gm10749
Hspb7
Tpm1
Ccdc149
Wisp1
Abca8b
2510009E07Rik
Zfp953
Nat2
Klhl30
Fbxo2
Spag4
4930579K19Rik
Trp53inp1
Atg7
B
B16-F10 in vivo
-log10(p-value)
rank
10
5
0
−5
−10
0
1000
2000
Naa16
Ahctf1
Daam2
5830418K08Rik
Xrcc3
Bbs4
Setd2
Cep250
Lsm14b
Ggnbp1
Flot1
Itpr2
Arap1
Pex6
B4galt4
Slc1a5
Slc45a2
Fam86
Atg7
C
expression B16-F10 tumors (TPM)
100
10
1
Atg7
1.4e−05
Ccng2
0.01
Ifi27l1
0.0043
Irf9
0.03
Isg15
0.0087
Psmb10
0.049
control
Atg7 pKO
D
expression TCGA.SKCM (TPM)
1000
100
10
1
ATG7
0.1
CCNG2
p < 2.22e−16
IFI27L1
1.6e−12
IRF9
4.1e−05
ISG15
1.5e−08
PSMB10
3.8e−10
short
long
E
B16-F10 Tumors
-log10(FDR)
5
15
25
depleted gene-ontology terms
type I interferon signaling pathway
regulation of inflammatory response
protein mono−ADP−ribosylation
protein binding
protein ADP−ribosylase activity
positive regulation of pattern recognition receptor signaling pathway
positive regulation of leukocyte mediated immunity
positive regulation of interferon−gamma−mediated signaling pathway
positive regulation of interferon−alpha production
positive regulation by symbiont of host autophagy
negative regulation of viral genome replication
negative regulation of immune system process
NAD+ ADP−ribosyltransferase activity
MDA−5 signaling pathway
double−stranded RNA binding
defense response to bacterium
cellular response to interferon beta
cytosol
antigen processing and presentation of peptide antigen via MHC class I
antigen processing and presentation of exogenous antigen
2'−5'−oligoadenylate synthetase activity
BP
CC
MF
ontology
F
TCGA.SKCM
-log10(FDR)
4
6
8
10
depleted gene-ontology terms
structural constituent of ribosome
respiratory electron transport chain
respiratory chain complex
receptor regulator activity
proton transmembrane transport
oxidoreductase activity
oxidative phosphorylation
oxidation−reduction process
organelle inner membrane
organelle envelope
NADH dehydrogenase complex assembly
NADH dehydrogenase complex
mitochondrion
mitochondrial respiratory chain complex I
mitochondrial respiratory chain complex assembly
mitochondrial respirasome
mitochondrial ATP synthesis coupled electron transport
membrane protein complex
intrinsic component of membrane
inner mitochondrial membrane protein complex
inflammatory response
immune response
generation of precursor metabolites and energy
extracellular region
envelope
electron transport chain
electron transfer activity
detoxification
defense response
ATP synthesis coupled electron transport
BP
CC
MF
ontology

Figure 2.15. Supplemental Figure 9. A common set of genes are differentially expressed in mouse and human melanomas when the Atg7 3′ UTR is lengthened.

(**A**) Log rank plot of differentially expressed genes detected by RNA-seq of B16-F10 Cas9-expressing cells treated with control or Atg7 pKO pgRNAs (n = 3 replicates per genotype).

(**B**) Log rank plot of differentially expressed genes detected by RNA-seq of B16-F10 Cas9-expressing tumors treated with control or Atg7 pKO pgRNAs (n = 5 replicates per genotype).

(**C**) Box plots showing expression (TPM) of individual genes that are dysregulated in B16-F10 Cas9 tumors treated with Atg7 pKO pgRNAs, P values calculated using two-sided Wilcoxon rank-sum test.
(**D**) Violin plots showing expression (TPM) of individual genes that are dysregulated in the TCGA cutaneous melanoma (SKCM) samples with short (bottom tercile) or long (top tercile) ATG7 3′ UTRs.
(**E**) GO analysis of differentially expressed genes in B16-F10 tumors treated with Atg7 pKO pgRNAs, indicated terms are depleted in Atg7 pKO tumor samples relative to control tumors. Go terms are separated as terms related to a biological process (BP), cellular compartment (CC), or molecular function (MF) and point sizes are scaled to the -$\log_{10}$(p value). GO terms in red are related to the immune system.
(**F**) GO analysis of differentially expressed genes detected in the TCGA cutaneous melanoma cohort comparing samples short versus long ATG7 3′ UTRs. Indicated terms are depleted in tumor samples with long ATG7 3′ UTRs. Go terms are separated as terms related to a biological process (BP), cellular compartment (CC), or molecular function (MF) and point sizes are scaled to the -$\log_{10}$(p value). GO terms in red are related to the immune system.

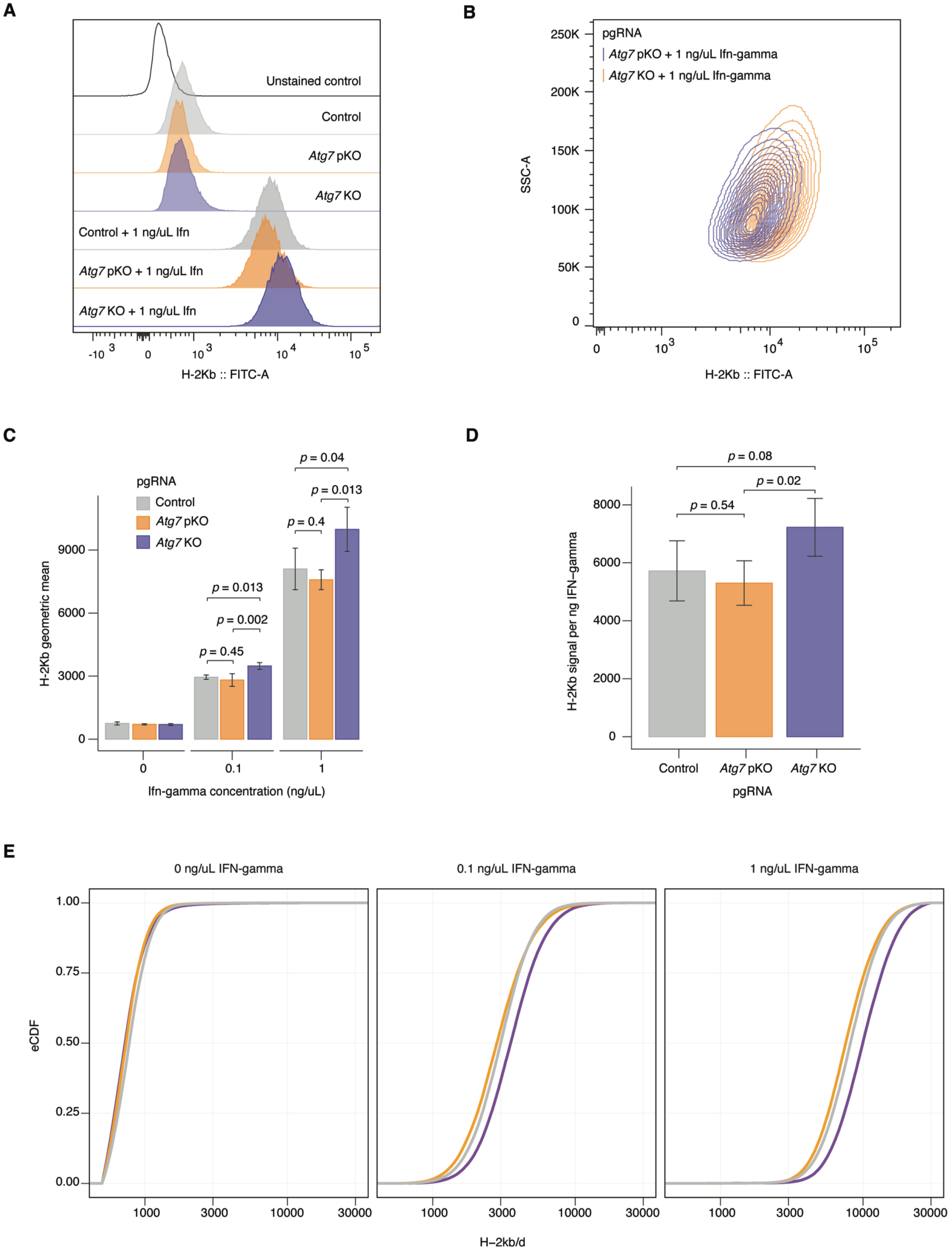

A
Unstained control
Control
Atg7 pKO
Atg7 KO
Control + 1 ng/uL Ifn
Atg7 pKO + 1 ng/uL Ifn
Atg7 KO + 1 ng/uL Ifn
-10^3
0
10^3
10^4
10^5
H-2Kb :: FITC-A
B
pgRNA
Atg7 pKO + 1 ng/uL Ifn-gamma
Atg7 KO + 1 ng/uL Ifn-gamma
250K
200K
150K
100K
50K
0
SSC-A
0
10^3
10^4
10^5
H-2Kb :: FITC-A
C
pgRNA
Control
Atg7 pKO
Atg7 KO
p = 0.04
p = 0.013
p = 0.4
p = 0.013
p = 0.002
p = 0.45
9000
6000
3000
0
H-2Kb geometric mean
0
0.1
1
Ifn-gamma concentration (ng/uL)
D
p = 0.08
p = 0.02
p = 0.54
8000
6000
4000
2000
0
H-2Kb signal per ng IFN-gamma
Control
Atg7 pKO
Atg7 KO
pgRNA
E
0 ng/uL IFN-gamma
0.1 ng/uL IFN-gamma
1 ng/uL IFN-gamma
1.00
0.75
0.50
0.25
0.00
eCDF
1000
3000
10000
30000
H–2kb/d

Figure 2.16. Supplemental Figure 10. Atg7 3′ UTR lengthening alters the cellular response to IFN-gamma.

(**A**) Representative histograms of B16-F10 Cas9-expressing cells treated with control, *Atg7* pKO or *Atg7* gene KO pgRNAs treated with no or 1 ng/uL of Interferon-gamma (Ifn-gamma) for 24 hours and subsequently stained for H-2k$^{b/d}$, compared to an unstained control. Cells are gated for live, single cells.

(**B**) Overlaid contour plots of B16-F10 Cas9-expressing cells treated with *Atg7* pKO or *Atg7* gene KO pgRNAs treated with 1 ng/uL Ifn-gamma after 24 hours and subsequently stained for H-2k$^{b/d}$. Cells are gated for live, single cells.

(**C**) Bar plots of H-2k$^{b/d}$ per indicated cell line treated with increasing concentrations of Interferon-gamma for 24 (n = 4 replicates per genotype per condition). *P* values from a two-sided Student's *t*-test.

(**D**) Bar plots of the slope representing the increase in H-2k$^{b/d}$ signal per ng of Interferon-gamma, calculated from the data in panel C (n = 4 replicates per genotype condition). *P* values from a two-sided Student's *t*-test.

(**E**) Cumulative distribution function (CDF) plots of the summed raw flow data at each indicated interferon-gamma concentration (ng/uL).

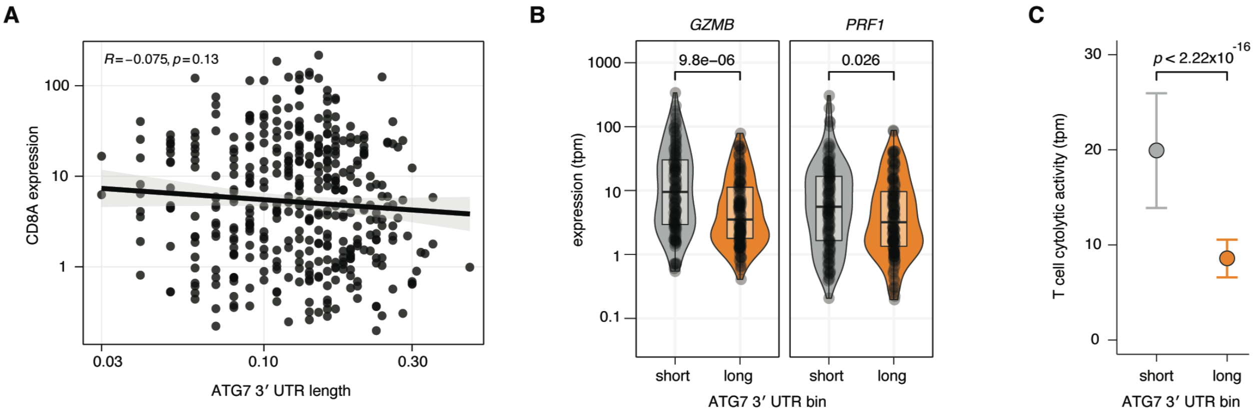


Figure 2.17. Supplemental Figure 11. ATG7 3′ UTR lengthening correlates with reduced T cell cytolytic activity in clinical melanoma.

(**A**) Correlation of *ATG7* 3′ UTR length and *CD8A* expression (TPM) per sample from the TCGA cutaneous melanoma (SKCM) cohort. *R* and *P* value from Pearson correlation.
(**B**) Expression of *GZMB* and *PRF1* (TPM) in TCGA cutaneous melanoma samples stratified into terciles by *ATG7* 3′ UTR length, comparing the bottom 33% (short) and the top 33% (long) samples. *P* values from two-sided Wilcoxon rank-sum test.
(**C**) Median T cell cytolytic activity (geometric mean of *PRF1* and *GZMB* expression) for samples stratified into short or long bins based on *ATG7* 3′ UTR length. Medians are shows +/- a 95% confidence interval obtained from bootstrapping. *P* value from a two-sided Wilcoxon rank-sum test.

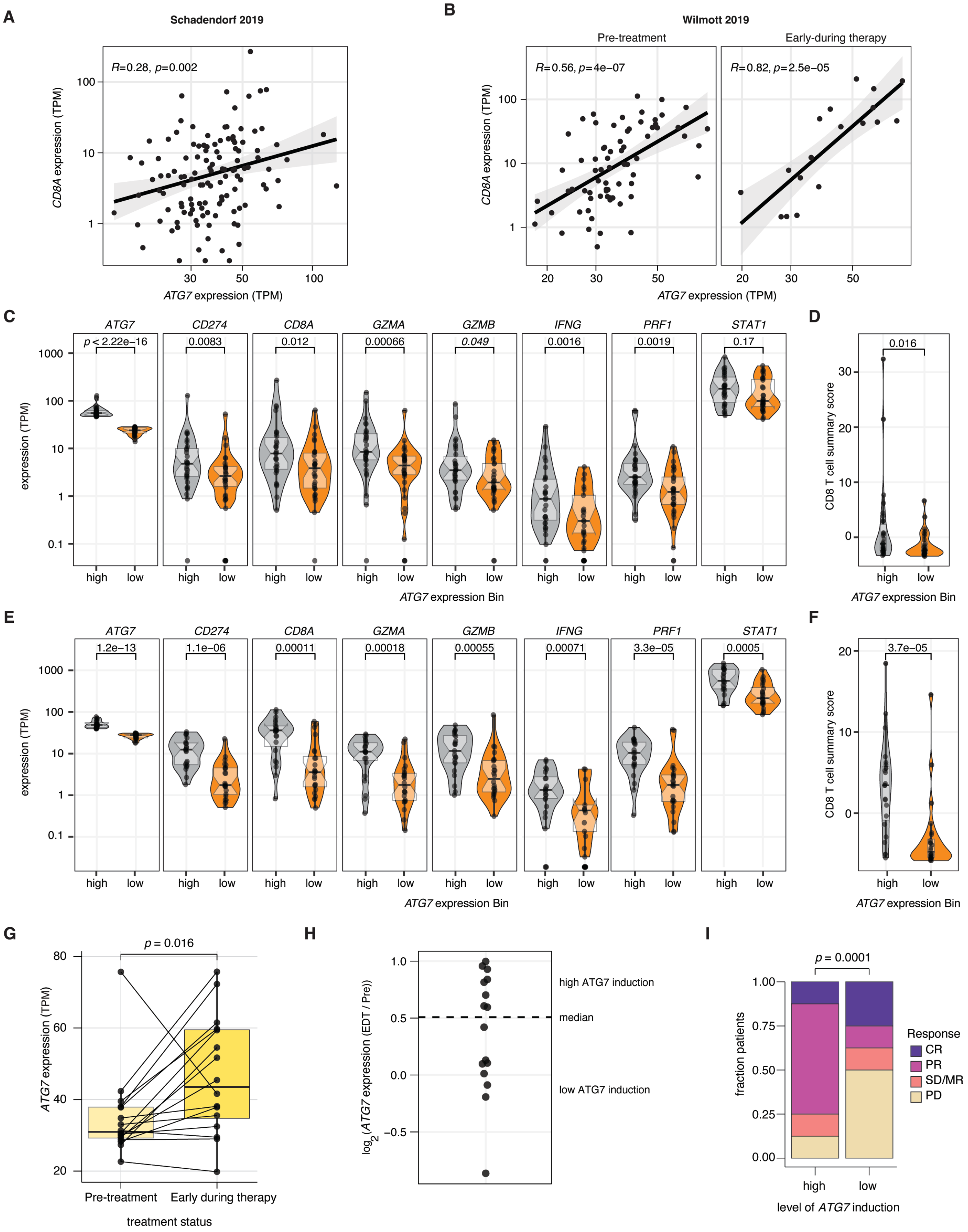

A
Schadendorf 2019
R=0.28, p=0.002
CD8A expression (TPM)
ATG7 expression (TPM)
100
10
1
30
50
100
B
Wilmott 2019
Pre-treatment
Early-during therapy
R=0.56, p=4e−07
R=0.82, p=2.5e−05
CD8A expression (TPM)
ATG7 expression (TPM)
100
10
1
20
30
50
C
ATG7
CD274
CD8A
GZMA
GZMB
IFNG
PRF1
STAT1
p < 2.22e−16
0.0083
0.012
0.00066
0.049
0.0016
0.0019
0.17
expression (TPM)
1000
100
10
1
0.1
high
low
ATG7 expression Bin
D
0.016
CD8 T cell summary score
30
20
10
0
high
low
ATG7 expression Bin
E
ATG7
CD274
CD8A
GZMA
GZMB
IFNG
PRF1
STAT1
1.2e−13
1.1e−06
0.00011
0.00018
0.00055
0.00071
3.3e−05
0.0005
expression (TPM)
1000
100
10
1
0.1
high
low
ATG7 expression Bin
F
3.7e−05
CD8 T cell summary score
20
10
0
high
low
ATG7 expression Bin
G
p = 0.016
ATG7 expression (TPM)
80
60
40
20
Pre-treatment
Early during therapy
treatment status
H
log2(ATG7 expression (EDT / Pre))
1.0
0.5
0.0
−0.5
high ATG7 induction
median
low ATG7 induction
I
p = 0.0001
fraction patients
1.00
0.75
0.50
0.25
0.00
high
low
level of ATG7 induction
Response
CR
PR
SD/MR
PD

Figure 2.18. Supplemental Figure 12. Low ATG7 expression correlates with reduced expression of genes predictive of stronger response to immune checkpoint blockade.

(**A**) Scatter plot of *ATG7* and *CD8A* expression (TPM) per sample from the Schadendorf 2019 cohort. *R* and *p* value from Pearson correlation.
(**B**) Scatter plot of *ATG7* and *CD8A* expression (TPM) from the Wilmott 2019 cohort. *R* and *p* value from Pearson correlation. Samples are stratified by sample type as either pre-treatment (PRE) or early during therapy (EDT) samples.
(**C**) Violin plots showing expression (TPM) of *ATG7* and several genes of interest related to $CD8^+$ T cell infiltration and cytotoxic activity in the Schadendorf 2019 cohort. Samples are stratified into low *ATG7* expression (bottom 25% of samples) and high *ATG7* expression (top 25% of samples). *P* values from two-sided Wilcoxon rank-sum test.
(**D**) CD8 T cell summary score, defined as the sum of the *Z*-scores for each gene listed in panel C excluding *ATG7*, per sample grouped by *ATG7* expression bin. *P* values from two-sided Wilcoxon rank-sum test.
(**E**) Violin plots showing expression (TPM) of *ATG7* and several genes of interest related to $CD8^+$ T cell infiltration and cytotoxic activity in the Wilmott 2019 cohort, excludes early during therapy samples. Samples are stratified into low *ATG7* expression (bottom 33% of samples) and high *ATG7* expression (top 33% of samples). *P* values from two-sided Wilcoxon rank-sum test.
(**F**) CD8 T cell summary score, defined as the sum of the *Z*-scores for each gene listed in panel E excluding *ATG7*, per sample grouped by *ATG7* expression bin. *P* values from two-sided Wilcoxon rank-sum test.
(**G**) Box plots showing *ATG7* expression (TPM) in the Wilmott 2019 cohort with both a pretreatment (PRE) versus early-during therapy (EDT) RNA-seq sample. *P* value from a two-sided paired Wilcoxon rank-sum test.
(**H**) Log_2(fold-change) of *ATG7* expression per matched EDT and PRE sample. The samples were then divided into either a low *ATG7* induction group (bottom 50%) or a High *ATG7* induction (top 50%). Dotted line indicates the split in the data.
(**I**) Stacked bar plots of clinical responses of patients with high or low *ATG7* expression as defined by RESIST criteria for samples stratified by level of *ATG7* induction as defined in panel H. *P* values were calculated using a two-sided multinomial proportion test.

2.7 MATERIAL AND METHODS

2.7.1 *Cell lines and culture*

B16-F10 cells were obtained from ATCC (CRL-6475) and cultured per the manufacturer's instructions. The Melan-A cell line was obtained from the Wellcome Trust Functional Genomics Cell Bank and cultured per their instructions (Bennett et al., 1987). B16-F10 cells were infected at 0.3 MOI with Cas9 lentivirus (Addgene 52962-LV) and selected with 2 ug/mL Blasticidin for 7 days. Polyclonal cells were then single cell sorted, and individual clones were assessed for on-target editing efficiency, all subsequent CRISPR assays were completed with this B16-F10 Cas9 clone with the highest editing rate.

2.7.2 *Poly(A)-Seq library preparation*

RNA was isolated from cell pellets with the Direct-zol RNA MiniPrep kit (Zymo Research). Protocol was adapted from Derti et al. 2012. In brief, 5 ug of RNA was heated at 65 degrees for 2 minutes and then placed on ice. Poly(A) selected mRNA was isolated using DynaBeads (ThermoFischer 61006) per the manufacturer's instructions. 1st strand cDNA synthesis using SuperScript IV Reverse Transcriptase (ThermoFischer 18090010) and the primer RKB4087 (a modified oligo d(TVN) primer) to generate cDNA reads which begin synthesis just before the poly(A) tail. This mixture was then purified using a 2x SPRI bead clean up and then followed by 2nd strand synthesis and then a 1x SPRI bead clean up. This cDNA library was then subject to PCR (NEBNExt High Fidelity 2x Master Mix) and primers RKB4089 (universal primer) and RKB4090-RKB4101 (indexed primers). The final libraries generate a broad range of fragment sizes, as such we gel extracted libraries from a 2% agarose gel corresponding to 300-500 bp size. Following extraction and purification, library size was analyzed with a 4200 TapeStation System before sequencing.

2.7.3 *RNA-seq library preparation*

RNA was isolated from cell pellets with the Direct-zol RNA MiniPrep kit (Zymo Research). Poly(A)-selected, unstranded Illumina libraries were prepared following the TruSeq protocol per the manufacturer's instructions. Library size and distribution was analyzed with a 4200

TapeStation System before sequencing on an Illumina HiSeq as 2x50 bp to obtain ~40 million reads per sample.

2.7.4 *RNA-seq and Poly(A)-seq data analysis*

RNA-seq was analyzed as previously described (Dvinge et al., 2014). RNA-seq reads were mapped to an annotated transcriptome created using Ensembl 71 (Flicek et al., 2013), UCSC knownGene (Meyer et al., 2013) and Misov2.0 (Katz et al., 2010) annotations using RSEM version 1.2.4(Li & Dewey, 2011) (modified to call Bowtie (Langmead et al., 2009) with option '-v 2'). Unaligned reads were then mapped to the corresponding genome (hg19/GRCh37 assembly, mm10/GRCmc38 assembly) and a database containing all possible pairings of 5' and 3′ splice sites per gene in our merged transcriptome annotation using TopHat version 20.8b (Trapnell et al., 2009). Mapped reads were then merged and input into MISO v2.0. For TCGA studies, we analyzed 9,045 available samples across 29 cancer types.

For the immunotherapy cohorts analyses, we analyzed 72 pre-treatment samples as well as 18 early-during therapy samples (16 of which had matched samples at each of the time-points) which we refer to as the Wilmott 2019 cohort (Gide et al., 2019), and 144 samples all taken pre-treatment which we refer to as the Schadendorf 2019 cohort (Liu et al., 2019).

For Poly(A)-seq, data was mapped similarly but in a stranded fashion. BAM files were input into the APAlyzer package in R (R. Wang & Tian, 2020) and the gene level $\log_2$(distal reads / proximal reads) was computed per sample for each respective strand.

2.7.5 *Survival analyses*

Cancer type abbreviations are the same as TCGA standards (https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations). Survival analyses were completed with the Kaplan-Meier estimator and statistical test were performed with a logrank test (R package survival). Stratification per cancer subtype was completed for Figure 1b, Supp. Figure 1A-C and 2A-F by computing the median 3′ UTR length per sample and dividing each cancer subtype into terciles (short, medium and long) and comparing the short (≤ 33%) vs long (≥66%) bins. Gene level 3′ UTR measurements were downloaded from a previously published database (Feng et al., 2018) or computed using the APAlyzer package in R (R. Wang & Tian, 2020).

For Figure 5F, patients from the TCGA SKCM cohort were stratified based on the ATG7 3 UTR length per sample, again into terciles and downstream analyses were performed identically as described above. For the two ICB cohorts, Wilmott 2019 (Gide et al., 2019) and Schadendorf 2019 (Liu et al., 2019), patients were stratified into terciles or quartiles, respectively, due to the larger size of the Schadendorf 2019 cohort. Once stratified, the low and high expression bins were subject to downstream analysis identically as described above.

2.7.6 *pgRNA library construction and cloning*

Target genes selected from the Poly(A) sequencing and TCGA SKCM analysis were largely selected based on manual inspection of BAM coverage plot quality. For each gene, the most utilized APA sites within the 3′ UTR were identified from PolyA-DB, filtered based off number of tissues where a particular APA site is utilized and the fraction of samples it has been detected in (R. Wang et al., 2018). Of those Poly(A) sites, the most proximal site was selected for targeting. The 100 base-pairs up and downstream of the given poly(A) site were selected as the genomic window wherein all possible gRNAs were generated using Guidescan Version 1.0 (Perez et al., 2017). The downloaded gRNAs were then identified as up or downstream of the poly(A) signal (PAS), and then the guides were combined in a pairwise fashion to generate all possible pgRNAs which disrupt the poly(A) signal sequence. For each target we excluded gRNAs with more than two two-nucleotide off-target sites, more than thirty three-nucleotide off-target sites or an on-target editing efficiency of less than thirty according to Rule Set 2 scoring (Doench et al., 2016). All input target sites were then filtered to only include targets with at least eight pgRNAs. For sites with more than ten possible pgRNAs, pgRNAs were ranked based first on targeting efficiency and then specificity to select the ten pgRNAs to be utilized in the library. This same process was used to generate all possible pgRNAs targeting highly utilized PAS sites of genes that are not expressed in B16-F10 or Melan-A cells. In this control library, 150 pgRNAs were selected randomly as controls to be included in the library, of note this random selection process was run several times until the distribution of control pgRNAs matched the true library in terms of efficiency and specificity. An additional 15 genes were included as growth controls based on previous literature, either core essential genes identified from DepMap or genes which promote more rapid B16-F10 cell growth in vivo identified by mining data generated from previous genome wide gene knock out CRISPR screens in B16-F10 cells (Manguso et al., 2017).

2.7.7 *Animal use*

All animal work and procedures were completed in accordance with the Guidelines for the Care and Use of Laboratory Animals and approved by the Fred Hutchinson Cancer Center Institutional Animal Care and Use Committee. C57BL/6 mice were obtained from the Jackson Laboratory.

2.7.8 *gDNA PCR and on-target editing verification*

gDNA was extracted using the DNeasy Blood and Tissue kit (Qiagen) per the manufacturer's protocol. A window around the Sap30l proximal poly(A) site was selected to allow for simple detection of deletion by band separation on gel electrophoresis. PCR amplicons were then purified and submitted for AmpliconEZ sequencing (Azenta/Genewiz). Reads were trimmed, mapped and the fraction of reads with an intact or disrupted poly(A) signal sequence were identified using the CRISPresso2 software (Clement et al., 2019).

2.7.9 *pgRNA screen*

$9x10^7$ Cas9 expressing B16-F10cells were infected at an MOI of 0.2. Twenty-four hours post infection, cells were then selected with 1 ug/mL Puromycin for 48 hours. Remaining cells were then pooled and divided into eight replicates. 500,000 cells were isolated from each replicated and frozen as a time point 0. Of the remaining cells, 500,000 cells were then plated into 10 cm plates and another 500,000 cells were injected subcutaneously into C57BL/6 mice. Cells grown in culture were passed every 3-4 days, and tumors were monitored and then harvested when reaching 1.5 cm in any single dimension or after 20 days, whichever condition was met first. On the day a given tumor was harvested, digested and frozen, the corresponding cells growing in vitro were also frozen. gDNA was then extracted from all samples using previously published methods (Chen et al., 2015), and quantified via Nanodrop. 1.5 ug of each sample was then utilized as input for PCR to amplify out integrated pgRNA constructs as previously described

(Thomas et al., 2020). Final, purified libraries were then submitted for Next-Generation Sequencing using a custom approach performed as previously described.

2.7.10 *pgRNA sequencing and data analysis*

Analyses were performed identically as previously described (Thomas et al., 2020). In brief, each read includes independent reads for gRNA1 and gRNA2 as well as an index read. Each gRNA was separately mapped to a database of pgRNAs using BowTie, and correct pairings were identified by ensuring gRNA1 and gRNA2 come from the same pgRNA, incorrect pairings were discarded. We generated a per pgRNA pseudocount which was added to the raw counts and served to regularize the fold-change computations so that they are proportional to the relative representation of each individual pgRNA in the library.

Fold-changes were then computed by comparing each time point to the counts at the day 0 time point. These fold-changes were then normalized so that the median of all pgRNAs targeting poly(A) sites in unexpressed genes (150 pgRNAs in total) was equal to 1. For each poly(A) site target or gene KO control target a p value was computed by comparing the distribution of normalized log-fold changes per pgRNA (8-10 per target) to the distribution of all 150 pgRNAs targeting poly(A) sites in unexpressed genes using a two-sided Wilcoxon rank-sum test. FDRs were computed by generating a distribution of fake target p values by randomly subsampling 10 pgRNAs from our control pgRNAs 10,000 times to estimate the distribution of expected p values. We then estimated FDR per real poly(A) site or gene KO target via the cumulative distribution function of the distribution of the fake target p values.

Statistical analyses were performed in R with Bioconductor (Huber et al., 2015), and tables and plots were generated using dplyr (Wickham et al., 2018) and ggplot2 (Wickham, 2009) packages.

2.7.11 *In vitro validation studies*

Cas9 expressing B16-F10 cells were grown in standard conditions and then transduced with lentivirus containing indicated pgRNAs and then selected in 1ug/mL Puromycin for 72 hours.

For cell growth assays, 100,000 cells were plated into each well of a 24 well plate and 20% of the well was passaged every 2 days into a new 24 well plate. 100 uL of the remaining cell suspension was used as input for a CellTiter-Glo (ProMega Catalog Number G9242) assay performed per the manufacturer's instructions.

2.7.12 *In vivo validation studies*

Cas9 expressing B16-F10 cells were grown in standard conditions and then transduced with lentivirus containing indicated pgRNAs and then selected in 1ug/mL Puromycin for 72 hours. $5x10^5$ cells were then injected subcutaneously into each flank of adult male C57BL/6 mice, and tumors were measured every 3 days using calipers. Animals were euthanized when a tumor reached 1.5 cm in any dimension, at which time tumor material was isolated either for an archived flash frozen sample, fixed in 10% formalin at room temperature for histology studies or placed in TRIzol reagent (Thermo Fischer 15596018) for subsequent RNA isolation.

2.7.13 *Immunohistochemistry*

Tissues from tumors were processed, embedded and stained through the Fred Hutch Experimental Histopathology core. Mouse Ki-67 (CST Clone 12202 1:2000 dilution) and Cd8a (CST Clone D4W27 1:200 dilution) staining were performed using rabbit monoclonal antibodies. Staining was performed with a BOND RX autostainer (Leica Biosystems) and images were then acquired with an Aperio ImageScope at 40x magnification (Leica Biosystems). Image analysis was completed using HALO Image Analysis software.

2.7.14 *RT-PCR analyses*

RNA was isolated using the Direct-zol RNA MiniPrep (Zymo Research). SuperScript IV reverse Transcriptase was used to synthesize cDNA per the manufacturer's instructions, but using a specific oligo d(TVN) primer to amplify the DNA directly upstream of the poly(A) tail (Thermo Fisher Scientific). Nested RT-PCR was performed with a universal poly(A) tail primer and two gene specific primers to amplify the 3′ UTR. The final product visualized using agarose gel electrophoresis and band intensity quantification was performed with FIJI/ImageJ and reported as the percent distal isoform (pdi) defined as the percent of total signal arising from bands corresponding to distal poly(A) site use.

2.7.15 *Western blotting*

Total protein lysates were isolated in 1x RIPA buffer and quantified with the Pierce 660 nm Protein Assay Reagent. Total protein lysates were electrophoretically separated and then transferred onto a nitrocellulose membrane using the NuPAGE system (Thermo Fisher Scientific). Each membrane was blocked for 1 hour at room temperature and then probed with primary antibody diluted in a blocking buffer overnight at 4 degrees Celsius. Atg7 (AbCam Ab133528, 1:1000), Egln1 (Cell-Signaling Technology D31E11, 1:1000), p62/Sqstm1 (AbCam Ab109012, 1:1000), and Alpha-tubulin (Sigma-Aldrich Clone DM1, 1:2000) primary antibodies were used. Anti-mouse or anti-rabbit IRDye (LI-COR Biosciences) secondary antibodies and the Odyssey CLx Imager (LI-COR Biosciences) were utilized for detection and imaging.

2.7.16 *Fluorescent LC3 reporter assay*

The FUW mCherry-GFP-LC3 reporter plasmid was obtained from Addgene (Plasmid #110060) which was utilized to generate lentivirus in 293T cells. B16-F10 Cas9-expressing cells were then infected and selected using 1 ug/mL puromycin for 48 hours. Selected cells were then plated and treated with standard DMEM with 10% FBS or Hanks Balanced Salt Solution for the indicated time. Following the time course all cells were washed 3x with PBS and then resuspended in 100 uL of PBS and stained with LIVE/DEAD Fixable Violet stain per the manufacturer's instructions.

2.7.17 *Flow cytometry*

Cells were stained with LIVE/DEAD Fixable Violet stain and passed through a 40 micron filter to generate a single cell suspension and then run on a BD FACSCelesta. Single cells were gated for live cells and then the individual GFP and mCherry signal were measured in the FITC and CF594 channels, respectively. Data were analyzed in FlowJo v10.

2.8 DATA AVAILABILITY

Our proximal-Poly(A) KO (pKO) library has been deposited with addgene (Pooled Library #81543). RNA-seq data we generated for this study have been added to the Gene Expression Omnibus. RNA-seq data generated by TCGA were downloaded from the Genomic Data Commons (GDC) and Cancer Genomics Hub (CGHub). RNA-seq data generated by the two immunotherapy treated melanoma cohorts (Gide et al., 2019; Liu et al., 2019) were downloaded from the European Nucleotide Archive or dbGaP, respectively (accession numbers PRJEB23709 and phs000452.v3.p1). Gene-level 3′ UTR measurements for all TCGA samples were downloaded from http://tc3a.org. Source data for Figures 1-2, 4-6 and Supplementary Figures 1-3, 5, 7, 9, 11-12 are included in Supplementary Tables 1-8.

2.9 ACKNOWLEDGEMENTS

We thank A. Rajan for feedback and advice on autophagy related studies. We thank D. Bennett and the Wellcome Trust Functional Genomics Cell Bank for providing the Melan-A cells. A.M.G is an ARCS Foundation scholar. R.K.B. was supported in part by the NIH/NCI (R01 CA251138), NIH/NHLBI (R01 HL128239, R01 HL151651) and the Blood Cancer Discoveries Grant program through the Leukemia & Lymphoma Society, Mark Foundation for Cancer Research, and Paul G. Allen Frontiers Group (8023-20). R.K.B is a Scholar of The Leukemia & Lymphoma Society (1344-18) and holds the McIlwain Family Endowed Chair in Data Science. Computational studies were supported in part by FHCC’s Scientific Computing Infrastructure (ORIP S10 OD028685). Experimental studies were supported in part by the Experimental Histopathology, Flow Cytometry, and Genomics Shared Resources of the Fred Hutch/University of Washington Cancer Consortium (NIH/NCI P30 CA015704). The results in this publication are based in part on data from The Cancer Genome Atlas Research Network (http://cancergenome.nih.gov).

Chapter 3. APC LOSS-OF-FUNCTION MUTATIONS PROMOTE 3′ UTR LENGTHENING IN COLON ADENOCARCINOMA

This research is unpublished. RNA-seq mapping and processing was completed by James D. Thomas and Robert Bradley and RNA-seq was obtained from previously published data (Ringel et al., 2020). The remaining analyses described here represent my contribution to the work.

3.1 INTRODUCTION

Alternative cleavage and polyadenylation (APA) is a process whereby a precursor mRNA can be differentially cleaved and polyadenylated to become a mature messenger RNA (Gruber & Zavolan, 2019). An estimated 70% of human genes can be differentially polyadenylated and poly(A) site selection can vary by tissue type, throughout development, and is often dysregulated in disease states including cancer (Ji et al., 2009; Lianoglou et al., 2013; Xiang et al., 2018).

Proteins regulating poly(A) site selection are misexpressed in cancer relative to healthy tissues, leading to widespread dysregulation of poly(A) site selection (Xia et al., 2014). Changes in poly(A) site use and thus 3′ untranslated region (UTR) length can alter mRNA abundance, localization, and translation kinetics (Elkon et al., 2013; Tian & Manley, 2016). Downstream functional consequences of APA of individual genes remain poorly characterized but a wealth of computational approaches now exist to rapidly assess patterns of poly(A) site selection (Goering et al., 2020; Xia et al., 2014).

Here, we sought to identify unappreciated protein regulators of poly(A) site selection in human cancers that could serve as a novel source of therapeutic targets. To achieve this goal, we assessed thousands of correlations between gene expression and global trends in APA across 34 large scale RNA-seq datasets. These datasets include the 29 TCGA solid tumor studies, as well as RNA-seq from the 1000 genomes project, a TCGA metastatic cancer cohort, a large cohort of acute myeloid leukemia (AML) samples and a large cohort of primary prostate cancer samples. This transcriptomic analysis revealed that loss-of-function mutations in the gene Adenomatous Polyposis Coli (APC), which are frequently observed in colorectal adenocarcinoma (Kinzler & Vogelstein, 1996), significantly disrupt poly(A) site selection globally. The severity of APA dysregulation in colorectal adenocarcinoma samples with APC loss-of-function mutations correlated with significant alterations in patient survival and expression of clinically actionable immune checkpoint genes.

3.2 RESULTS

3.2.1 *Canonical poly(A) regulators correlate with global polyadenylation site selection in all cancer subtypes except colorectal adenocarcinoma*

To identify potential regulators of APA in human cancer, we assessed the correlation between gene expression and a summary statistic of 3′ UTR across 34 large RNA-seq datasets. Previous studies have assessed the correlation between gene expression and measures of poly(A) site selection for individual 3′ UTRs, but they have been restricted to datasets that contain a sufficient number of matched tumor and adjacent normal control tissues to perform robust statistical analyses (Goering et al., 2020; Xia et al., 2014). To extend these analyses to any dataset regardless of the presence of health tissue controls, we took a novel approach where we quantified relative poly(A) site selection per gene and then took the median of all quantified UTRs to obtain the global median 3′ UTR per sample, where a larger value indicates a sample on average has longer 3′ UTRs. We then computed gene expression of all coding genes per sample and obtained the Pearson correlation coefficient per gene per dataset comparing gene expression and median 3′ UTRs (**Figure 1A**). We assessed correlations of several genes where expression is known to alter 3′ UTR length globally (*CPSF6*, *CSTF2*, *CSTF2T*, *PABPN1*, and *SYMPK*) and all observed correlations were highly significant and in the expected direction (**Figure 1B-1D; Supp. Figure 1A-1B**) (Gruber & Zavolan, 2019; Xia et al., 2014). However, colorectal adenocarcinoma was a major exception where median 3′ UTR length and gene expression of known poly(A) site selection regulators was poorly correlated (**Supp. Figure 1B**). We next compared all obtained Pearson correlation coefficients for each of the 34 datasets analyzed by completing pairwise correlations of each dataset (**Figure 1E**). The median pairwise correlation across datasets was 0.699, reflecting the fact that most datasets demonstrate similar patterns of gene expression and median 3′ UTR length, and by extension that global patterns of regulation of poly(A) site selection are likely similar across datasets. However, colorectal adenocarcinoma was an extreme outlier with a median pairwise correlation of -0.1496. Together, these data indicate that correlations of global regulators are surprisingly homogenous across datasets except in colorectal adenocarcinoma, suggesting some common feature of colorectal adenocarcinomas uniquely dysregulates APA.

3.2.2 *APC loss-of-function mutations enhance distal poly(A) site selection in colorectal adenocarcinoma*

Motivated by the observation that patterns of poly(A) site selection are poorly correlated with known APA regulators in colorectal adenocarcinoma, I sought to further investigate genetic differences that could explain these trends. I first assessed common driver mutations in colorectal adenocarcinoma that have experimental evidence of RNA binding function. Several studies have independently demonstrated that Adenomatous Polyposis Coli (APC), the most frequently mutated gene in colorectal adenocarcinoma, is an RNA binding protein that preferentially binds RNA in the 3′ UTR (Preitner et al., 2014).

We first identified colorectal adenocarcinoma samples that harbor at least one APC loss-of-function mutation (nonsense or frameshift mutations) in The Cancer Genome Atlas cohort. While greater than 80% of patients harbor some kind of mutation in APC, we identified 326 (65.3%) that harbored at least one APC loss-of-function mutation and 176 (34.6%) without an APC loss-of-function mutation. We then compared the median 3′ UTR length of samples with or without an APC loss-of-function mutation and found samples with an APC loss-of-function mutation on average have significantly longer 3′ UTRs (**Figure 2A**). Since poly(A) site selection and 3′ UTR length are inferred based on computational algorithms applied to RNA-seq data, we confirmed these trends were independent of the computational pipeline used by implementing a distinct tool, APAlyzer, and again confirmed the trends we found (**Supp. Figure 2A-2B**) (R. Wang & Tian, 2020; Xia et al., 2014). We then compared individual UTR length alterations and identified 1198 lengthened and 55 shortened individual 3′ UTRs in the APC loss-of-function samples (**Figure 2B-2C**; **Supp. Figure 2C-2D**).

To assess molecular differences between lengthened and unchanged 3′ UTRs we completed *de novo* motif analysis comparing the sequences of the lengthened 3′ UTRs to the sequence of unchanged 3′ UTRs. We detected several significantly enriched C and G rich motifs in 3′ UTRs that are significantly lengthened (**Figure 2D**). These identified motifs are highly concordant with crosslink immunoprecipitation sequencing (CLIP-seq) data which experimentally determined the APC RNA binding motif as C and G rich motifs within 3′ UTRs (Preitner et al., 2014). Lengthened

UTRs also contained a significantly higher fraction of guanine nucleotides than unchanged 3′ UTRs (**Figure 2E**). Overall, these data demonstrate APC loss-of-function in colorectal adenocarcinoma is associated with increased use of distal poly(A) sites in 3′ UTRs.

3.2.3 *Targeted APC knockout in human colon organoids alter poly(A) site selection*

As RNA-seq obtained from cancer samples are often extremely heterogeneous in terms of cell populations, treatment history, and sample quality (Dvinge et al., 2014), we wanted to assess alterations in APA in a more controlled experimental setting. Thus, we analyzed data from human colon organoids cultured for 24 hours with targeted knockout of APC using CRISPR/Cas9 compared to a time zero control (Ringel et al., 2020). We computed differentially polyadenylated 3′ UTRs and identified 196 shortened and 207 lengthened 3′ UTRs 24 hours after APC knockout (**Figure 3A**). Unlike in the colorectal adenocarcinoma samples, we observed similar levels of shortening and lengthening events which could be due to differences in time scale between the two datasets.

To assess if results from our analysis of colorectal adenocarcinoma were similar to 3′ UTR changes in an organoid model, we assayed guanine content of significantly altered shortening or lengthening events. We found that only the lengthening events showed significantly higher guanine content when compared to 3′ UTRs that were unchanged (**Figure 3B**). We previously identified several G and C rich motifs that were significantly enriched in 3′ UTRs that were lengthened in APC loss-of-function colorectal adenocarcinoma samples (**Figure 2D**). We screened all analyzed 3′ UTRs in our organoid data for presence of at least one of the top five enriched motifs (CSGGCCMC, GCCCCS, GGGGGAS, CGGSCC, CCCWGSCC) and found that lengthened 3′ UTRs, but not shortened 3′ UTRs, contained significantly more of these motifs (73.8%) compared to unchanged 3′ UTRs (61.4%) (**Figure 3C**). We identified 78 lengthened and 19 shortened transcripts shared between the two datasets (**Supp. Figure 3A-3B**) and bam coverage plots of the RNA-seq data validated imputed lengthening events (**Figure 3D**).

These data demonstrate that APC loss-of-function perturbs poly(A) site selection and leads to 3′ UTR lengthening of a subset of 3′ UTRs enriched in G and C rich motifs.

3.2.4 *Degree of 3′ UTR lengthening is associated with patient survival in APC loss-of-function colorectal adenocarcinoma*

We next sought to understand the clinic implications of APA in the context of APC loss-of-function colorectal adenocarcinoma. We first validated a set of high confidence 3′ UTR lengthening events that were detected using more stringent statistical cutoffs and only included events detected by distinct computational algorithms, DaPars and APAlyzer (R. Wang & Tian, 2020; Xia et al., 2014). This led us to a subset of 181 high confidence 3′ UTR lengthening events in colorectal adenocarcinoma harboring APC loss-of-function mutations. We then stratified patients with or without APC loss-of-function mutations into quartiles based on the degree of 3′ UTR lengthening observed. In patients with APC loss-of-function mutations, but patients without loss-of-function mutations, shorter 3′ UTRs were associated with significantly worse overall survival (**Figure 4A-4B**). To further characterize differences between short and long UTR samples we completed differential gene expression analysis and found that short UTR samples expressed significantly higher levels of immune checkpoint genes including CTLA4, CD70, PDL1, and TIGIT (**Figure 4C-4D**; **Supp. Figure 4A-D**). All these molecules serve to dampen immune response, specifically CD8+ T cell mediated cytotoxicity and are frequently employed by cancer cells to prevent immune mediated clearance of tumors (Conway et al., 2018; Flieswasser et al., 2022; He & Xu, 2020). Taken together, this data shows that 3′ UTR length correlates with patient outcomes and clinically actionable molecular signatures in APC loss-of-function colorectal adenocarcinoma.

3.3 DISCUSSION

Alternative cleavage and polyadenylation is globally dysregulated in human cancers and known regulators of APA are frequently over or under expressed in different cancer subtypes (Kargapolova et al., 2017; Lee et al., 2021; Xia et al., 2014). However, most studies of poly(A) site selection have been limited to datasets with an appropriate number of healthy tissue control samples (Xia et al., 2014; Goering et al., 2020). Here, we describe a novel approach which circumvents this issue, and we demonstrate that expression of known protein regulators of poly(A) site selection are strongly correlated with a global measure of 3′ UTR length in most datasets. We show that colorectal adenocarcinoma is a striking outlier where expression of genes that are responsible for poly(A) site selection are poorly or not at all correlated with 3′ UTR length. A unique feature of colorectal adenocarcinoma is the high prevalence of loss-of-function mutations in polyposis coli (APC) (Kinzler & Vogelstein, 1996). APC is well established as an RNA binding protein that binds RNAs predominantly in the 3′ UTR and transports them along microtubules for location specific translation along growing axons (Baumann et al., 2020; Preitner et al., 2014; T. Wang et al., 2017). The role of APC as an RNA binding protein is well studied in neurogenesis, but this function remains entirely unexplored in the context of colorectal adenocarcinoma.

We focused on how APC defects alter poly(A) site selection globally and demonstrate that APC loss-of-function mutations lead to global 3′ UTR lengthening in colorectal adenocarcinoma. We validate these results in a human colon organoid model where APC knockout using CRIPSR/Cas9 globally perturbs poly(A) site selection (Ringel et al., 2020). In both colorectal adenocarcinoma with APC loss-of-function mutations and APC knockout colon organoids we find lengthened 3′ UTRs are significantly enriched in GC rich motifs as well as guanine nucleotide content. These motifs are very similar to G and C rich APC RNA binding motifs that have been experimentally determined in mice (Baumann et al., 2020; Preitner et al., 2014). We see that global 3′ UTR lengthening correlates with patient outcomes in APC loss-of-function colorectal cancer, highlighting clinical relevance of the RNA binding function of APC.

We analyzed previous existing RNA-seq datasets from colorectal adenocarcinoma cancer samples as well as human colon organoids given the systems have large samples and represent a clean genetic knockout, respectively. However, identifying poly(A) site selection from RNA-seq data requires computational inference. Future work is needed to generate unambiguous evidence of shifts in poly(A) site use over appropriate time scales. Given that development of colorectal adenocarcinoma is a process that typically takes years, conditional gene knockout of APC studied over the course of 24 hours may not reflect the most important 3′ UTR alterations that drive colorectal adenocarcinoma phenotypes. In addition, further work is needed to understand the specific APA events that play a functional role in cancer progression.

Although we focus on colorectal adenocarcinoma in this study, APC mutations are commonly found in other cancer subtypes and cancer syndromes that predispose to colon, endometrial and brain tumors among others (Tanley et al., 1995; Y. Zhang et al., 2022). APC mutations have also been linked to complex neurological diseases including schizophrenia (Cui et al., 2005; Mohn et al., 2014). The combination of computational cataloging of differentially polyadenylated isoforms and high-throughput functional assessment of APA in relevant biological systems represents an important source of novel therapeutic targets for a number of human diseases.

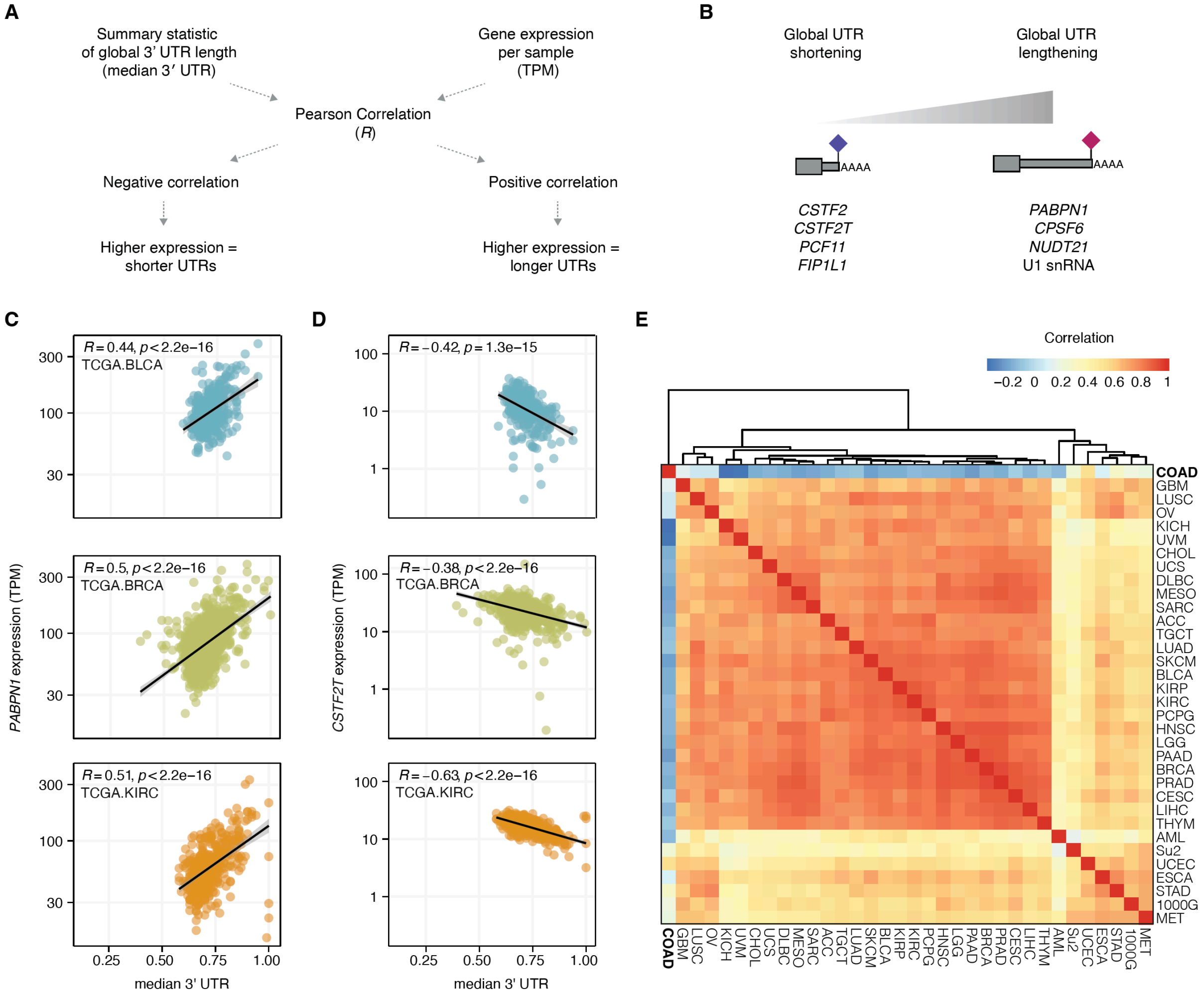


Figure 3.1. **Figure 1.** Global regulators of poly(A) site selection are poorly correlated with 3′ UTR length in colorectal adenocarcinoma.

(**A**) Workflow summary of first independently quantifying gene expression (transcripts per million, TPM) for all coding genes per RNA-seq sample across 34 datasets. For each sample, the 3′ UTR length is calculated per gene using either DaPars (Xia et al., 2014) for all TCGA data or APAlyzer (R. Wang & Tian, 2020) for non-TCGA datasets (D. Robinson et al., 2015; D. R. Robinson et al., 2017; Tyner et al., 2018). For each sample, a summary statistic of 3′ UTR length was calculated by taking the median of all imputed 3′ UTR lengths, referred to as median 3′

UTR. Then for each gene a Pearson correlation was completed comparing median 3′ UTR length and gene expression per gene, per dataset.

(**B**) Proteins involved in poly(A) site selection and how they are purported to act globally when expressed highly.

(**C**) Scatter plot and Pearson correlation of 3′ UTR length and *PABPN1* expression (TPM) for three datasets (bladder adenocarcinoma = BLCA, Invasive breast carcinoma = BRCA, kidney renal cell carcinoma = KIRC). *R* and *p* value reflective of Pearson correlation.

(**D**) Scatter plot and Pearson correlation of 3′ UTR length and *CSTF2T* expression (TPM) for three datasets (bladder adenocarcinoma = BLCA, Invasive breast carcinoma = BRCA, kidney renal cell carcinoma = KIRC). *R* and *p* value reflective of Pearson correlation.

(**E**) Correlation matrix of all 34 datasets comparing calculated Pearson correlation coefficients for all genes versus median 3′ UTR. Pairwise dataset correlations calculated using Pearson correlation.

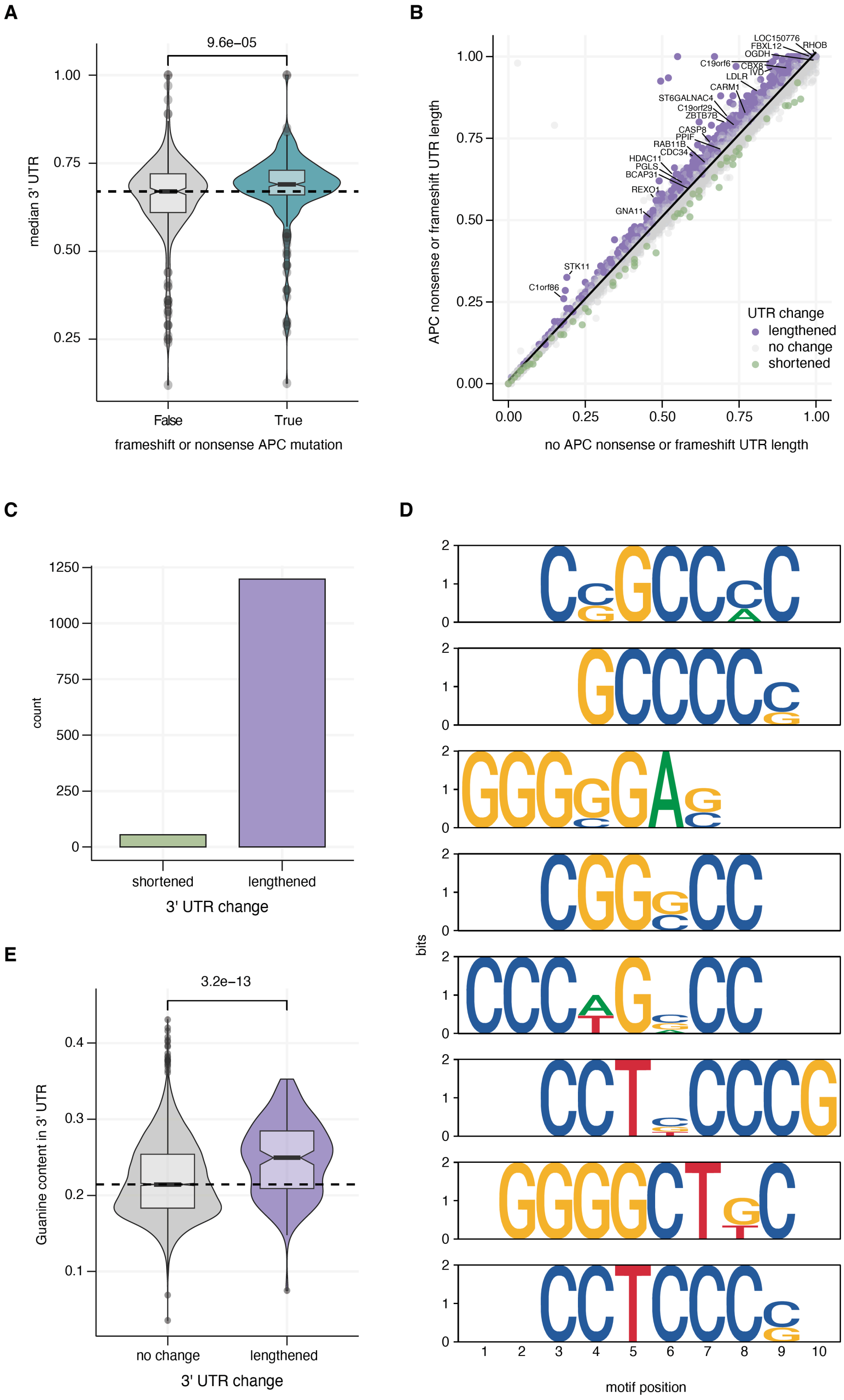

A
9.6e−05
median 3' UTR
1.00
0.75
0.50
0.25
False
True
frameshift or nonsense APC mutation
B
APC nonsense or frameshift UTR length
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
no APC nonsense or frameshift UTR length
LOC150776
RHOB
FBXL12
OGDH
C19orf6
CBX8
IVD
LDLR
CARM1
ST6GALNAC4
C19orf29
ZBTB7B
CASP8
PPIF
RAB11B
CDC34
HDAC11
PGLS
BCAP31
REXO1
GNA11
STK11
C1orf86
UTR change
lengthened
no change
shortened
C
count
1250
1000
750
500
250
0
shortened
lengthened
3' UTR change
D
bits
2
1
0
1
2
3
4
5
6
7
8
9
10
motif position
E
3.2e−13
Guanine content in 3' UTR
0.4
0.3
0.2
0.1
no change
lengthened
3' UTR change

Figure 3.2. **Figure 2. APC loss-of-function mutations are associated with enhanced distal poly(A) site use in colorectal adenocarcinoma.**

(**A**) Violin plots of median 3′ UTR per sample comparing samples with a detected APC loss-of-function mutation (nonsense, frameshift insertion or frameshift deletion) in the TCGA colorectal adenocarcinoma dataset. *P* value from two-sided Wilcoxon rank-sum test.

(**B**) Scatter plot of the median imputed 3′ UTR length for samples without APC loss-of-function mutations versus samples with a loss-of-function APC mutation. All 3′ UTR measurements per group were compared using a two-sided Wilcoxon rank-sum test. Comparisons which a Benjamin-Hochberg corrected *p* value < 0.05 were called as significantly altered. Points labeled in green indicate significant shortening in APC loss-of-function samples, points labeled in purple indicate significant lengthening in APC loss-of-functions samples, and points labeled in gray indicate no significant difference.

(**C**) Bar plot of significantly altered lengthening (purple, n = 1198) or shortening events (green, n = 55) in APC loss-of-function samples compared to APC no loss-of-function samples from the TCGA colorectal adenocarcinoma cohort.

(**D**) Sequence logo plot of de novo motif enrichment of 3′ UTRs exhibiting significant lengthening in APC loss-of-function colorectal adenocarcinoma samples compared to all other UTRs analyzed in panel B that exhibit no significant change in 3′ UTR length. Top 8 significantly enriched motifs displayed.

(**E**) Violin plots of the fraction of nucleotides within the longest annotated 3′ UTR that are guanine in UTRs that are either unchanged (gray) or lengthened (purple) in APC loss-of-function colorectal adenocarcinoma samples. *P* value from two-sided Wilcoxon rank-sum test.

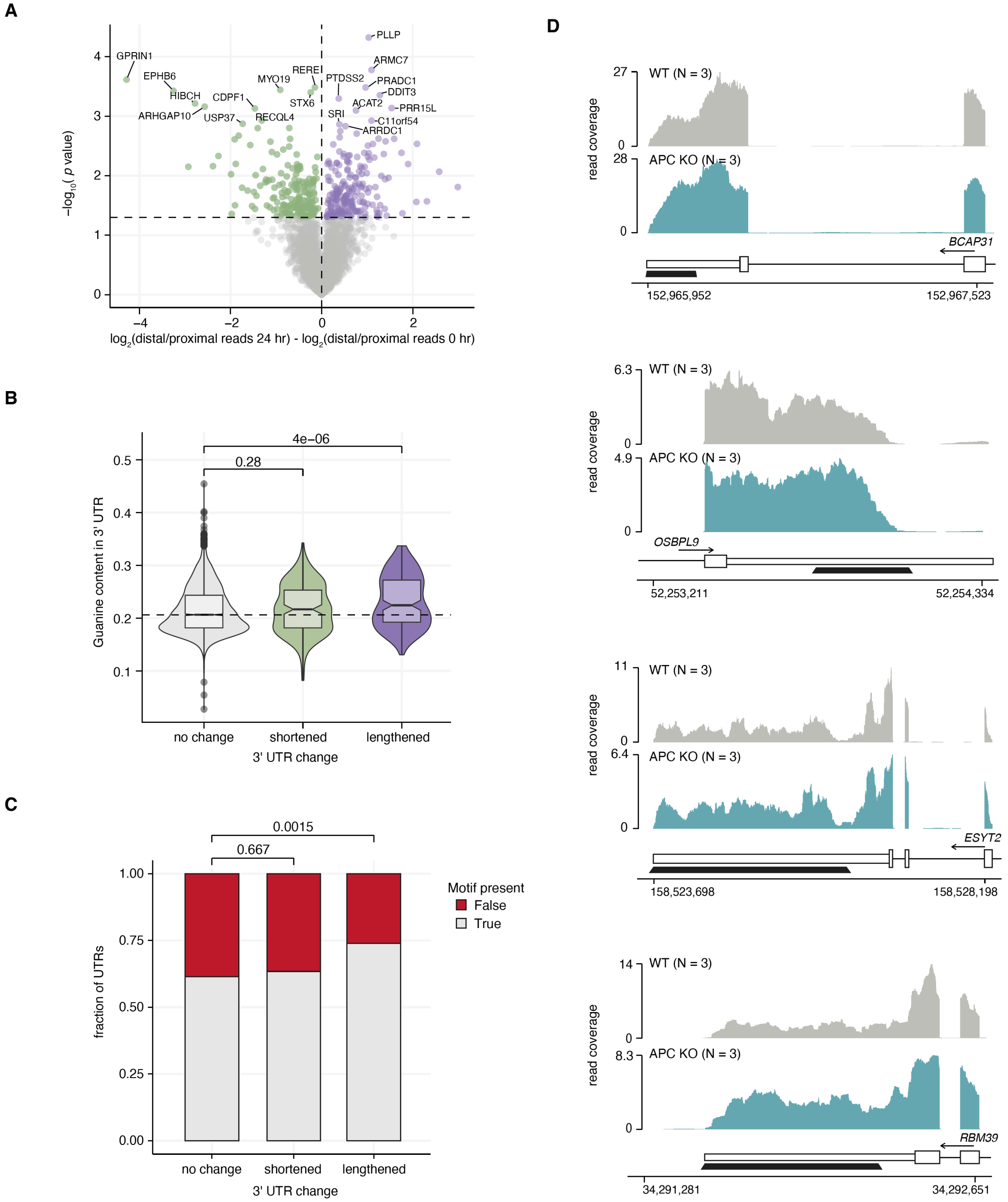


Figure 3.3. **Figure 3. APC knockout in colon organoids recapitulate global dysregulation of poly(A) site selection.**

(**A**) Scatter plot of the difference in gene level 3′ UTR lengths in human colon organoids with APC gene knockout using CRISPR/Cas9 comparing 0 versus 24 hour timepoints. 3′ UTRs that show significant shortening are indicated in green lengthening are indicated in purple.

(**B**) Violin plots of the fraction of nucleotides within the longest annotated 3′ UTR that are guanine in UTRs that are either unchanged (gray), shortened (green) or lengthened (purple) in APC knockout colon organoids. *P* values from two-sided Wilcoxon rank-sum test.

(**C**) Fraction of 3′ UTRs identified in colon organoids as lengthened, shortened or no change (Panel A) that contain at least one of the five most significantly enriched motifs (CSGGCCMC, GCCCCS, GGGGGAS, CGGSCC, CCCWGSCC) identified from colorectal adenocarcinoma de novo motif enrichment analyses (**Figure 2D**). *P* values were calculated using a Chi-squared test compared to the 3′ UTRs in the no change group.

(**D**) Bam coverage plots of RNA-seq data for individual 3′ UTRs (*BCAP31*, *ESYT2*, *OSBPL9*, *RBM39*) that display lengthening for WT human colon organoids (gray, n = 3) or APC knockout human colon organoids (yellow, n = 3).

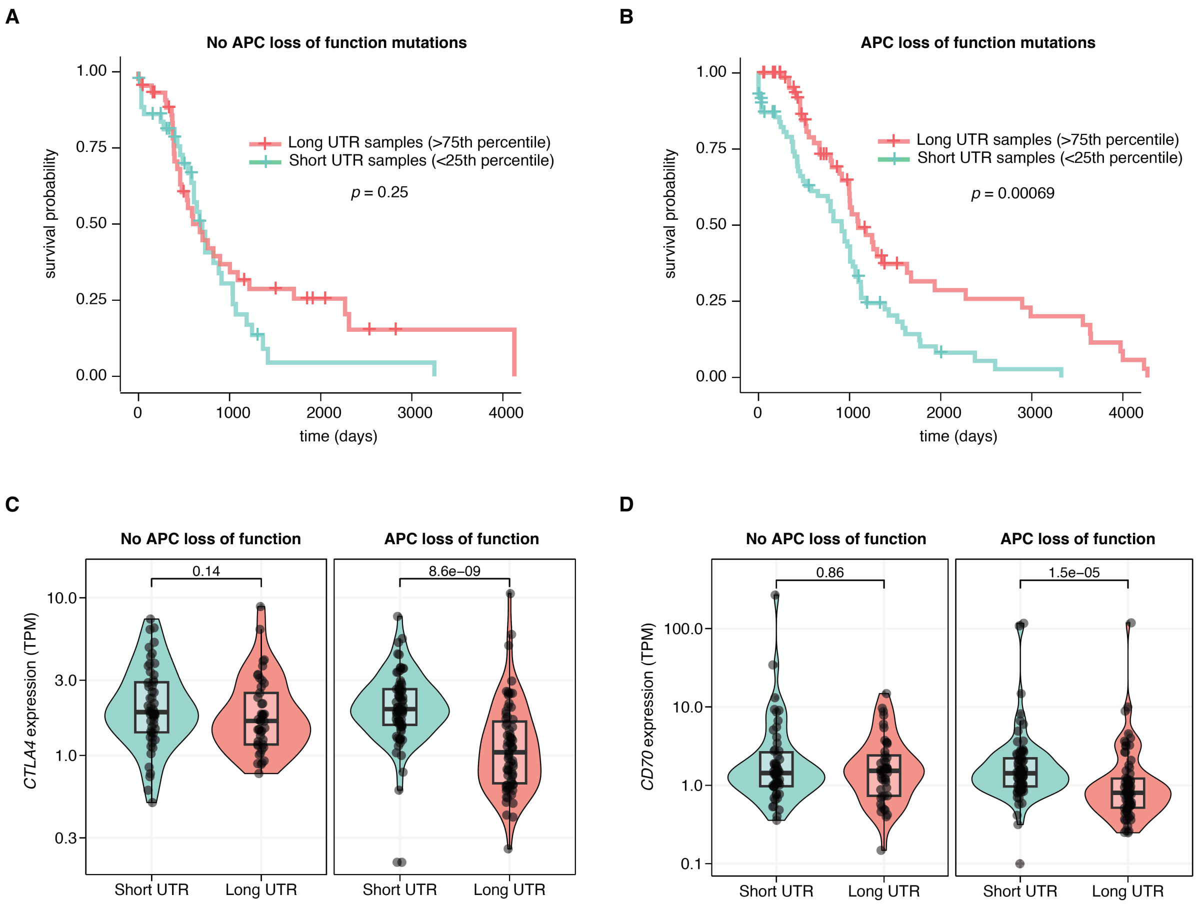


Figure 3.4. Figure 4. Degree of 3′ UTR lengthening correlates with patient outcomes in APC loss-of-function colorectal adenocarcinoma.

(**A**) Colorectal adenocarcinoma samples without APC loss-of-function mutations were stratified into quartiles based on median 3′ UTR length. Overall survival of patients with the shortest (<25th percentile) and the longest (>75th percentile) 3′ UTRs was compared using Kaplan-Meier survival analysis and cox proportional hazard ratios were calculated. *P* value was calculated using a logrank test.

(**B**) Colorectal adenocarcinoma samples with APC loss-of-function mutations were stratified into quartiles based on median 3′ UTR length. Overall survival of patients with the shortest (<25th percentile) and the longest (>75th percentile) 3′ UTRs was compared using Kaplan-Meier survival analysis and cox proportional hazard ratios were calculated. *P* value was calculated using a logrank test.

(**C**) Violin plots of gene expression (TPM) of the immune checkpoint marker *CTLA4* in short or long 3′ UTR stratified samples with or without APC loss-of-function mutations in colorectal adenocarcinoma. *P* value from two-sided Wilcoxon rank-sum test.

(**D**) Violin plots of gene expression (TPM) of the immune checkpoint marker *CD70* in short or long 3′ UTR stratified samples with or without APC loss-of-function mutations in colorectal adenocarcinoma. *P* value from two-sided Wilcoxon rank-sum test.

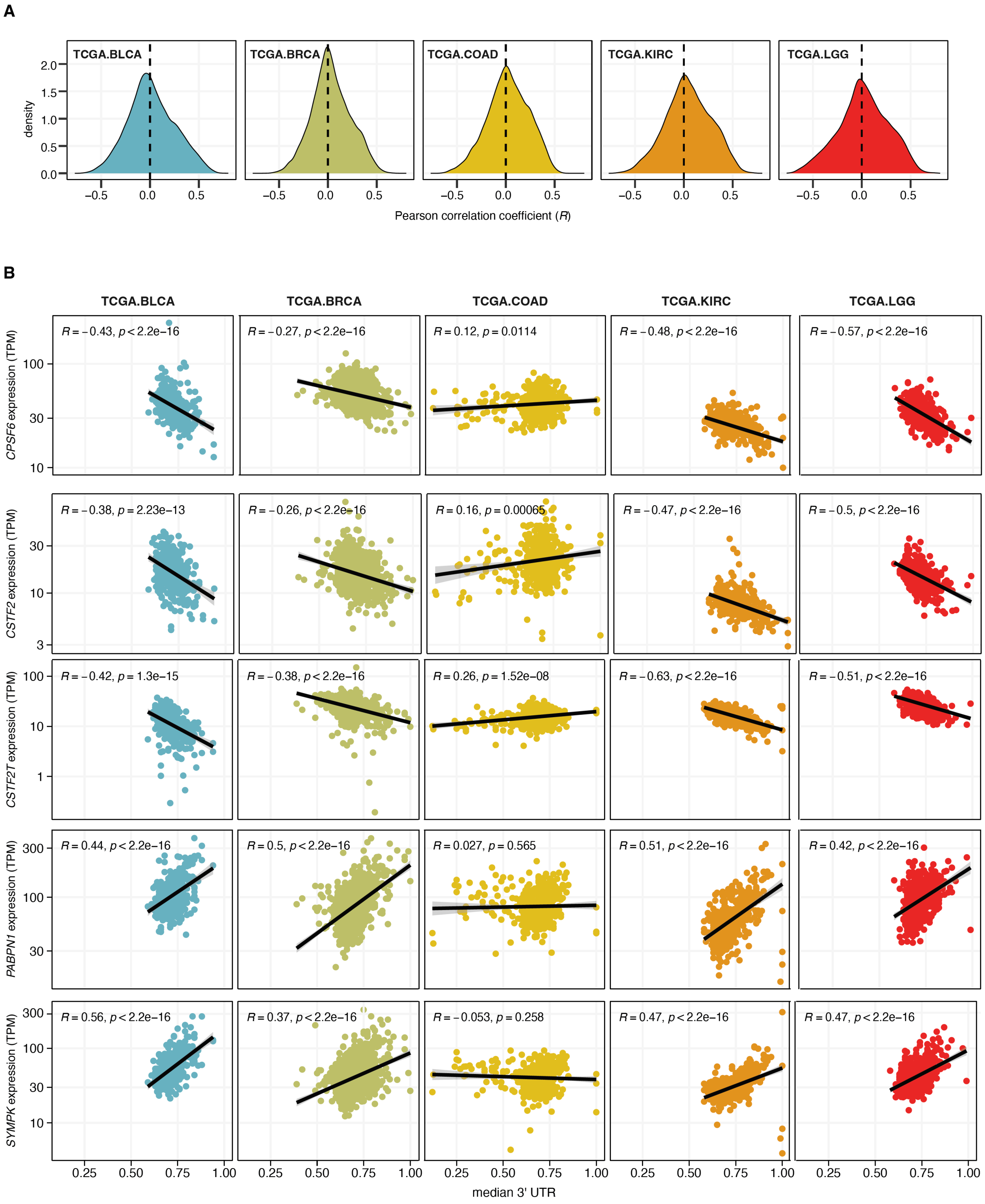


Figure 3.5. Supplemental Figure 1. Canonical poly(A) site selection regulators correlate with global 3′ UTR length in all cancer subtypes except colorectal adenocarcinoma.

(**A**) Density plots of all obtained Pearson correlation coefficients per dataset indicated (BLCA = bladder adenocarcinoma, BRCA = invasive breast carcinoma, COAD = colorectal adenocarcinoma, KIRC = kidney renal cell carcinoma, LGG = low grade glioma). Correlations were calculated by comparing median 3′ UTR per sample versus gene expression (TPM) per gene, per dataset.

(**B**) Scatterplots of median 3′ UTR length per sample versus gene expression of the indicated gene (TPM). Plots shown for known poly(A) site regulators *CPSF6*, *CSTF2*, *CSTF2T*, *PABPN1*, and *SYMPK*. *R* and *p* values are calculated from Pearson correlation.

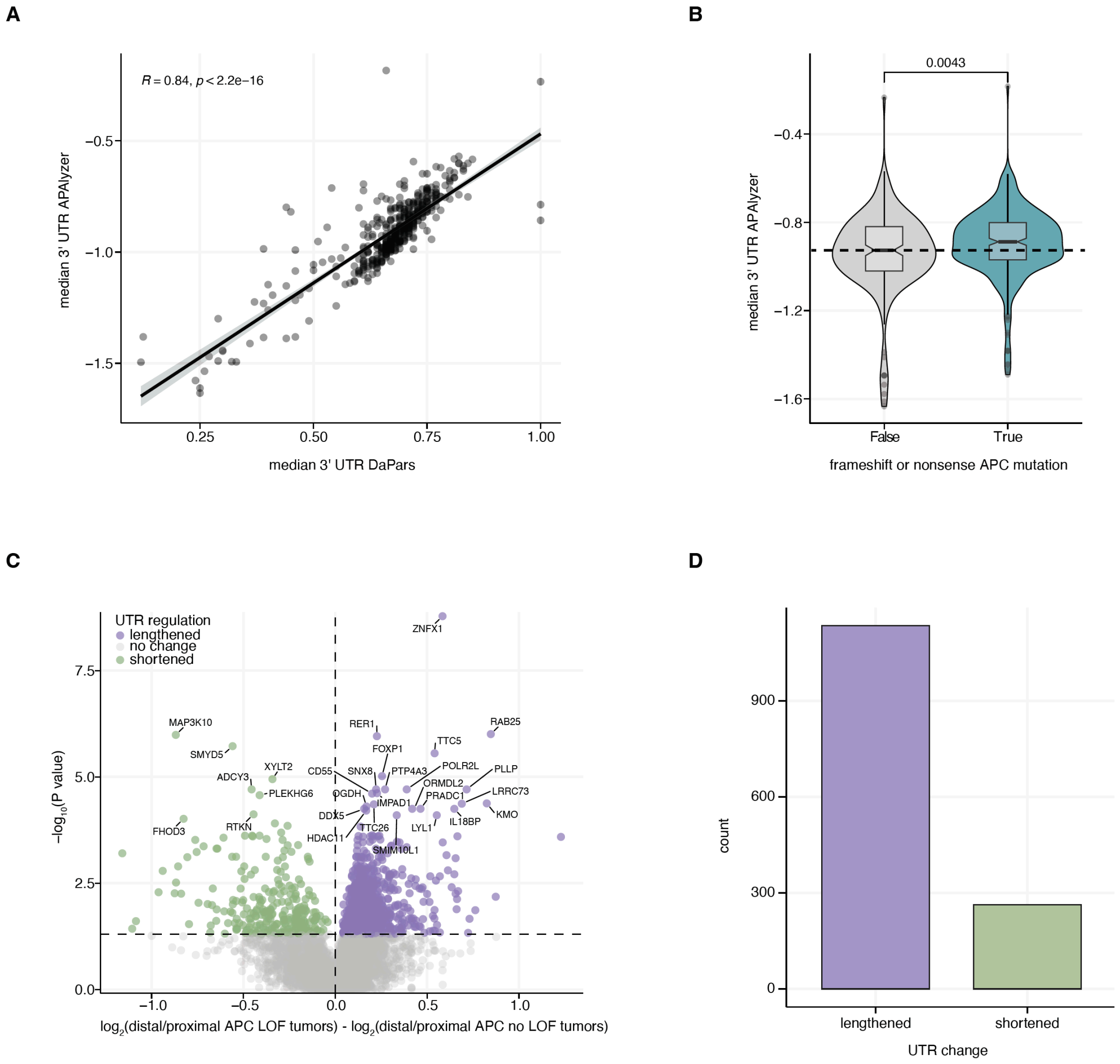


Figure 3.6. Supplemental Figure 2. APC loss-of-function mutations are associated with global 3′ UTR lengthening.

(**A**) Scatter plot of median 3′ UTR value per colorectal adenocarcinoma sample calculated using the DaPars or APAlyzer algorithm. *R* and *p* values are calculated from Pearson correlation.

(**B**) Violin plots of median 3′ UTR per sample (calculated using APAylzer pipeline) with a detected APC loss-of-function mutation (nonsense, out of frame insertion or deletion mutation) in the TCGA colorectal adenocarcinoma dataset. Dashed line indicates the median of all data. *P* value from two-sided Wilcoxon rank-sum test.

(**C**) Volcano plot comparing 3′ UTR length in APC loss-of-function samples versus samples without a detected APC loss-of-function mutation in the TCGA colorectal adenocarcinoma dataset. Individual UTRs are indicated as shortened (green), lengthened (purple) or no change (gray) based on a p value < 0.05 using a two-sided Wilcoxon rank-sum test.

(**D**) Bar plot of significantly altered lengthening (purple, n = 1135) or shortening events (green, n = 263) in APC loss-of-function samples.

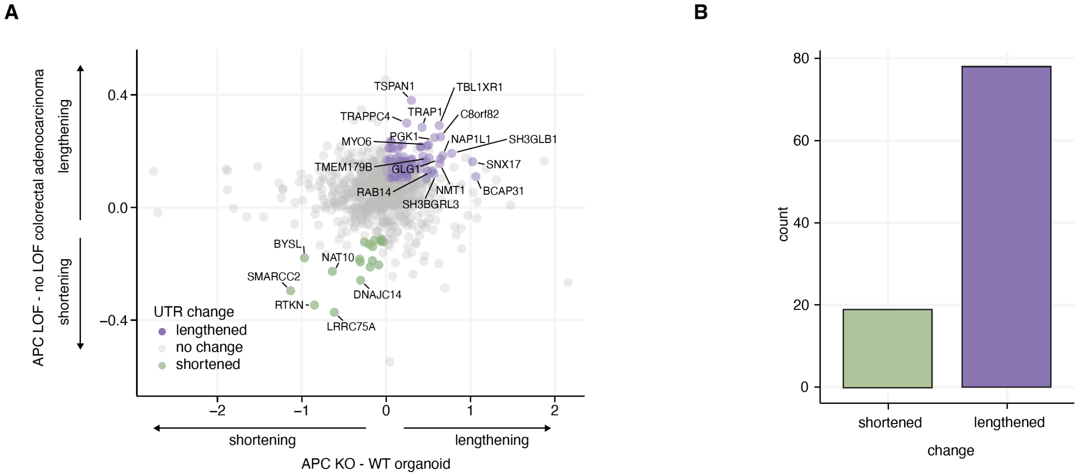


Figure 3.7. Supplemental Figure 3. Comparative poly(A) site selection in colorectal adenocarcinoma and colon organoids.

(**A**) Scatter plot of relative changes in poly(A) site selection in APC KO versus WT organoids compared to poly(A) site selection in APC loss-of-function colorectal adenocarcinoma samples versus no loss-of-function samples. APA events that are significantly shortened (green) or lengthened (purple) have an adjusted *p* value < 0.05 in both datasets, with a change in UTR length in the same direction.

(**B**) Barplot of the total number of significant events called as shortened (green) or lengthened (purple) in both datasets.

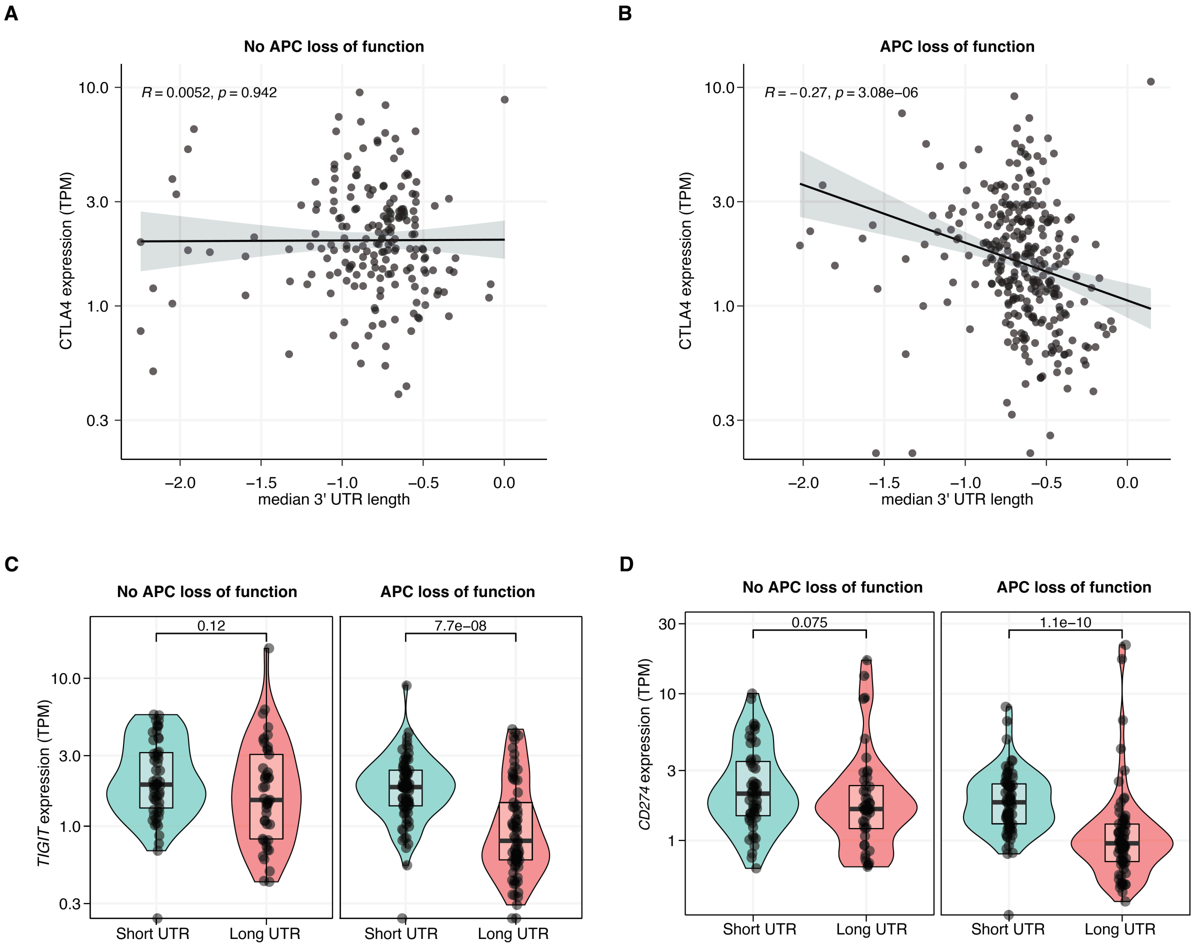


Figure 3.8. Supplemental Figure 4. Immune checkpoint gene expression correlates with 3′ UTR length In APC loss-of-function colorectal adenocarcinoma.

(**A**) Scatter plot of median 3′ UTR length versus CTLA4 expression (TPM) per sample in colorectal adenocarcinoma samples without APC loss-of-function mutations. R and *p* value from Pearson correlation.

(**B**) Scatter plot of median 3′ UTR length versus CTLA4 expression (TPM) per sample in colorectal adenocarcinoma samples with at least one APC loss-of-function mutation. R and *p* value from Pearson correlation.

(**C**) Violin plots of gene expression (TPM) of the immune checkpoint marker *TIGIT* in short or long 3′ UTR stratified samples with or without APC loss-of-function mutations in colorectal adenocarcinoma. *P* value from two-sided Wilcoxon rank-sum test.

(**D**) Violin plots of gene expression (TPM) of the immune checkpoint marker *CD274 (PDL1)* in short or long 3′ UTR stratified samples with or without APC loss-of-function mutations in colorectal adenocarcinoma. *P* value from two-sided Wilcoxon rank-sum test.

3.4 MATERIAL AND METHODS

3.4.1 *RNA-seq data analysis*

RNA-seq was analyzed as previously described (Dvinge et al., 2014). RNA-seq reads were mapped to an annotated transcriptome created using Ensembl 71 (Flicek et al., 2013), UCSC knownGene (Meyer et al., 2013) and Misov2.0 (Katz et al., 2010) annotations using RSEM version 1.2.4(Li & Dewey, 2011) (modified to call Bowtie (Langmead et al., 2009) with option '-v 2'). Unaligned reads were then mapped to the corresponding genome (hg19/GRCh37 assembly) and a database containing all possible pairings of 5' and 3′ splice sites per gene in our merged transcriptome annotation using TopHat version 20.8b (Trapnell et al., 2009). Mapped reads were then merged and input into MISO v2.0. For TCGA studies, we analyzed available samples across 30 cancer types. Cancer type abbreviations are the same as TCGA standards (https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations). We also completed analysis for four additional datasets including RNA-seq from the Stand up to Cancer Prostate Cancer (Robinson et al., 2015), 1000 Genomes project (Lappalainen et al., 2013), Beat AML (Tyner et al., 2018) and a large cohort of metastatic cancer samples (Robinson et al., 2017). For figure 3, previously published RNA-seq data was obtained from wildtype or APC knockout colon organoids (N = 3 per genotype) was mapped in the same fashion as described above (Ringel et al., 2020).

3.4.2 *Poly(A) site selection and 3′ UTR length measurements*

Gene level 3′ UTR measurements were downloaded from a previously published database (Feng et al., 2018) for all TCGA cohorts which utilizes the DaPars algorithm to quantify relative 3′ UTR length. For other datasets, we computed gene level 3′ UTR measurements using the APAlyzer package in R (R. Wang & Tian, 2020). For each sample, we took the median of all gene level 3′ UTR measurements to obtain the median 3′ UTR length which was used as the sample summary statistic of 3′ UTR length (a higher value indicates, on average, the sample uses globally longer UTRs).

For differential poly(A) site analysis in the TCGA colorectal adenocarcinoma (Figure 2B) we used a two-sided Wilcoxon-Rank sum test of the gene level 3′ UTR measurements comparing samples with or without an identified loss-of-function mutation in APC. We confirmed these results by running the data through a distinct computational algorithm, APAlyzer, and obtained highly concordant results (Supp. figure 2) (R. Wang & Tian, 2020).

For differential poly(A) site analysis in the colon organoid data we utilized the APAlyzer pipeline to obtain gene level 3′ UTR measurements and then used a two-sided Student's *t*-test to identify significantly altered APA events between wild-type and APC knockout organoids (R. Wang & Tian, 2020).

3.4.3 *Survival analyses*

Survival analyses were completed with the Kaplan-Meier estimator and statistical test were performed with a logrank test (R package survival). Stratification for APC loss-of-function samples or no loss-of-function samples was completed for Figure 5A and 5B by computing the median 3′ UTR length per sample and dividing each cancer subtype into quartiles (short, medium, long and longest) and comparing the short (≤ 25%) vs longest (≥75%) bins. Gene level 3′ UTR measurements were computed using the APAlyzer package in R (R. Wang & Tian, 2020).

3.4.4 *Motif discovery*

Significantly lengthened 3′ UTRs identified in TCGA colorectal adenocarcinoma samples with APC loss-of-function mutations (**Figure 2B**) using the APAlyzer pipeline (Wang & Tian, 2020) were selected as the target group for motif discovery, N = 166 of high confidence UTRs that have multiple annotated poly(A) sites. A group of background 3′ UTRs were selected based on showing no significant change in 3′ UTR length in that TCGA colorectal adenocarcinoma dataset, N =5805. The sequence of longest annotated 3′ UTR was obtained from RefSeq and then fasta files for the target and background 3′ UTRs were input into a de novo motif enrichment tool, MEME Suite version 5.5.0, using the default E-value cutoff of 0.5 (Bailey et al., 2009).

Chapter 4. DISCUSSION AND FUTURE DIRECTIONS

These works together provide a computational and experimental framework for functional interrogation of alternative polyadenylation in biologically relevant contexts. We perform the first screen of poly(A) sites in any context and demonstrate that APA can drive cancer associated phenotypes including tumor growth. Specifically, 3′ UTR lengthening of *ATG7* is associated with reduced tumor burden in mice, but also reduced CD8$^+$ T cell infiltration. Our observations extend to human clinical melanoma where *ATG7* 3′ UTR length correlated with overall survival in a cohort of cutaneous melanoma and was significantly associated with gene expression signatures of cytotoxic CD8$^+$ T cells. *ATG7* expression is associated with clinical response to immune checkpoint blockade in two independent melanoma cohorts and, considering that ATG7 is not subject to recurrent loss-of-function mutations, fine tuning of *ATG7* expression via APA could be mechanism employed by cancers to evade immune mediated tumor clearance. Further work is needed to characterize how *ATG7* 3′ UTR lengthening specifically modulates immune response in melanoma and if ATG7 APA or targeted inhibition of protein function could be a useful therapeutic strategy. Given that *ATG7* gene knock out sensitizes several cancer cell lines to CD8+ T cell killing (Lawson et al., 2020), *ATG7* seems like a promising target for therapy development. Beyond *ATG7*, we identified several other UTR lengthening events that impact cancer cell growth that still require mechanistic follow up. Interestingly, most events screened seem to have little impact on cancer cell growth *in vitro* or *in vivo*. This suggests many APA events may have negligible functional effects in the context of melanoma cell growth; however, functional screening of APA events in distinct biological contexts is needed.

More work is needed to improve the number of poly(A) sites targeted. Given that many 3′ UTRs are AT rich sequences, they often make poor targets for canonical CRISPR/Cas9 guide RNA design, but utilization of improved Cas variants with less guide RNA design requirements would boost the number of targetable poly(A) sites in the genome (Nidhi et al., 2021). Though our technology is applicable to distal poly(A) sites, we need to thoroughly test and confirm the efficacy of distal poly(A) site excision. Our library also neglects other more nuanced features of RNA processing including 3′ UTR splicing, intronic polyadenylation, and alternative last exons, all of which are more complicated RNA processing events that, in some cases, can alter coding sequences or target an RNA for degradation by nonsense mediated decay (Anczukow & Krainer, 2016; Chan et al., 2022; Dvinge et al., 2016). Incorporating more complex RNA processing events into high-throughput screening approaches will hopefully allow us to identify individual RNA isoforms that drive therapeutically actionable cancer phenotypes.

Our more global approach to correlations of 3′ UTR length and gene expression revealed that colorectal adenocarcinoma is a striking outlier, where unlike other datasets analyzed, expression of canonical regulators of poly(A) site selection are poorly correlated with global UTR length. We establish that loss-of-function mutations in adenomatous polyposis coli (APC), which are present in more than half of colorectal adenocarcinoma cancer samples (Aghabozorgi et al., 2019; Y. Zhang et al., 2022), are associated with significantly longer 3′ UTRs. APC loss-of-function mutations are widely considered an initiating event for the formation of colon polyps that eventually lead to development of colorectal adenocarcinoma (L. Zhang & Shay, 2017), however 3′ UTR lengthening is classically associated with slowed cell cycle progression and cell growth rates (Sandberg et al., 2008). There are reports of biological contexts where 3′ UTR lengthening is dissociated from cell cycle dynamics (Cheng et al., 2020), but further work is needed to understand if cell cycle kinetics and poly(A) site selection are linked in APC mutant colorectal adenocarcinoma. We see that APC knock out globally affects 3′ UTR lengths enriched in G and C rich motifs, which is consistent with previous reports of APC RNA binding motifs identified from mouse neurogenesis studies (Baumann et al., 2020; Preitner et al., 2014; L. Zhang & Shay, 2017). Although we establish that the degree of 3′ UTR lengthening is associated with significantly altered clinical outcomes and expression of immune checkpoint genes, further work is required to elucidate 3′ UTR lengthening events that are functionally relevant in colorectal adenocarcinoma. The association of 3′ length and immune checkpoint expression is particularly relevant as colorectal adenocarcinoma responds poorly to immune checkpoint blockade (Almquist et al., 2020). We observe widespread 3′ UTR lengthening, more than 900 APA events, many of which could be potential therapeutic targets to synergize with immune checkpoint blockade.

Finally, this work focuses on establishing a mechanistic link between APA and cancer phenotypes, but we also catalog a large number of cancer specific alternative polyadenylation events in clinical melanoma and colorectal adenocarcinomas. There are now databases of annotated poly(A) sites along with distinct types of computational and experimental evidence (R. Wang et al., 2018), enabling generation of highly multiplexed CRIPSR/Cas9 based approaches to survey poly(A) site selection. Generation of tools comparable to publicly available CRISPR/Cas9 libraries targeting all coding genes for poly(A) sites will be critical to the proliferation of functional studies in a variety of experimental systems. Identification of specific APA events can also allow for targeted therapeutic interventions using antisense oligonucleotides to block poly(A) site selection (Elkon et al., 2013; Vickers et al., 2001). Pharmacological modulation of poly(A) site selection is also possible (Araki et al., 2018), although those drugs impact APA transcriptome wide, they could also potentially modulate critical APA events with therapeutic value. The work shown here indicates that poly(A) site selection is a mechanism employed by cancers and motivates future studies to develop novel anti-cancer treatments.

BIBLIOGRAPHY

Abdel-Wahab, O., & Gebauer, F. (2018). Editorial overview: Cancer genomics: RNA metabolism and translation in cancer pathogenesis and therapy. *Current Opinion in Genetics and Development*, *48*, iv–vi. https://doi.org/10.1016/j.gde.2018.01.007

Agarwal, V., Lopez-Darwin, S., Kelley, D. R., & Shendure, J. (2021). The landscape of alternative polyadenylation in single cells of the developing mouse embryo. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-25388-8

Aghabozorgi, A. S., Bahreyni, A., Soleimani, A., Bahrami, A., Khazaei, M., Ferns, G. A., Avan, A., & Hassanian, S. M. (2019). Role of adenomatous polyposis coli (APC) gene mutations in the pathogenesis of colorectal cancer; current status and perspectives. In *Biochimie* (Vol. 157, pp. 64–71). Elsevier B.V. https://doi.org/10.1016/j.biochi.2018.11.003

Almquist, D. R., Ahn, D. H., & Bekaii-Saab, T. S. (2020). The Role of Immune Checkpoint Inhibitors in Colorectal Adenocarcinoma. In *BioDrugs* (Vol. 34, Issue 3, pp. 349–362). Adis. https://doi.org/10.1007/s40259-020-00420-3

Anczukow, O., & Krainer, A. R. (2016). Splicing-factor alterations in cancers. *Rna*, *22*(9), 1285–1301. https://doi.org/10.1261/rna.057919.116

Andres, S. F., Williams, K. N., Plesset, J. B., Headd, J. J., Mizuno, R., Chatterji, P., Lento, A. A., Klein-Szanto, A. J., Mick, R., Hamilton, K. E., & Rustgi, A. K. (2019). IMP1 3′ UTR shortening enhances metastatic burden in colorectal cancer. *Carcinogenesis*, *40*(4), 569–579. https://doi.org/10.1093/carcin/bgy153

Araki, S., Nakayama, Y., Sano, O., Nakao, S., Shimizu-Ogasawara, M., Toyoshiba, H., Nakanishi, A., & Aparicio, S. (2018). Decoding Transcriptome Dynamics of Genome-Encoded Polyadenylation and Autoregulation with Small-Molecule Modulators of Alternative Polyadenylation. *Cell Chemical Biology*, *25*(12), 1470-1484.e5. https://doi.org/10.1016/j.chembiol.2018.09.006

Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D. R., Albright, A., Cheng, J. D., Kang, S. P., Shankaran, V., Piha-Paul, S. A., Yearley, J., Seiwert, T. Y., Ribas, A., & McClanahan, T. K. (2017). IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *Journal of Clinical Investigation*, *127*(8), 2930–2940. https://doi.org/10.1172/JCI91190

Bae, B., Gruner, H. N., Lynch, M., Feng, T., So, K., Oliver, D., Mastick, G. S., Yan, W., Pieraut, S., & Miura, P. (2020). *Elimination of Calm1 long 3'-UTR mRNA isoform by CRISPR-Cas9 gene editing impairs dorsal root ganglion development and hippocampal neuron activation in mice*. https://doi.org/10.1261/rna

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, *37*(SUPPL. 2). https://doi.org/10.1093/nar/gkp335

Batra, R., Charizanis, K., Manchanda, M., Mohan, A., Li, M., Finn, D. J., Goodwin, M., Zhang, C., Sobczak, K., Thornton, C. A., & Swanson, M. S. (2014). Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Molecular Cell*, *56*(2), 311–322. https://doi.org/10.1016/j.molcel.2014.08.027

Baumann, S., Komissarov, A., Gili, M., Ruprecht, V., Wieser, S., & Maurer, S. P. (2020). C E L L B I O L O G Y A reconstituted mammalian APC-kinesin complex selectively transports defined packages of axonal mRNAs. In *Sci. Adv* (Vol. 6). https://www.science.org

Bennett, D. C., Cooper, P. J., & Hart, I. R. (1987). A line of non-tumorigenic mouse melanocytes, syngeneic with the B16 melanoma and requiring a tumour promoter for growth. *International Journal of Cancer*, *39*(3), 414–418. https://doi.org/10.1002/ijc.2910390324

Berkovits, B. D., & Mayr, C. (2015). Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. *Nature*, *522*(7556), 363–367. https://doi.org/10.1038/nature14321

Brumbaugh, J., di Stefano, B., Wang, X., Borkent, M., Forouzmand, E., Clowers, K. J., Ji, F., Schwarz, B. A., Kalocsay, M., Elledge, S. J., Chen, Y., Sadreyev, R. I., Gygi, S. P., Hu, G., Shi, Y., & Hochedlinger, K. (2018). Nudt21 Controls Cell Fate by Connecting Alternative Polyadenylation to Chromatin Signaling. *Cell*, *172*(1–2), 106-120.e21. https://doi.org/10.1016/j.cell.2017.11.023

Chan, J. J., Zhang, B., Chew, X. H., Salhi, A., Kwok, Z. H., Lim, C. Y., Desi, N., Subramaniam, N., Siemens, A., Kinanti, T., Ong, S., Sanchez-Mejias, A., Ly, P. T., An, O., Sundar, R., Fan, X., Wang, S., Siew, B. E., Lee, K. C., … Tay, Y. (2022). Pan-cancer pervasive upregulation of 3′ UTR splicing drives tumourigenesis. *Nature Cell Biology*, *24*(6), 928–939. https://doi.org/10.1038/s41556-022-00913-z

Chen, S., Sanjana, N. E., Zheng, K., Shalem, O., Lee, K., Shi, X., Scott, D. A., Song, J., Pan, J. Q., Weissleder, R., Lee, H., Zhang, F., & Sharp, P. A. (2015). Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell*, *160*(6), 1246–1260. https://doi.org/10.1016/j.cell.2015.02.038

Cheng, L. C., Zheng, D., Baljinnyam, E., Sun, F., Ogami, K., Yeung, P. L., Hoque, M., Lu, C. W., Manley, J. L., & Tian, B. (2020). Widespread transcript shortening through alternative polyadenylation in secretory cell differentiation. *Nature Communications*, *11*(1), 1–14. https://doi.org/10.1038/s41467-020-16959-2

Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., Cole, M. A., Liu, D. R., Joung, J. K., Bauer, D. E., & Pinello, L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. In *Nature Biotechnology* (Vol. 37, Issue 3, pp. 224–226). Nature Publishing Group. https://doi.org/10.1038/s41587-019-0032-3

Conway, J. R., Kofman, E., Mo, S. S., Elmarakeby, H., & van Allen, E. (2018). Genomics of response to immune checkpoint therapies for cancer: implications for precision medicine. In *Genome Medicine* (Vol. 10, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13073-018-0605-7

Cox, D. B. T., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., & Zhang, F. (2017). *RNA editing with CRISPR-Cas13*. https://www.science.org

Cui, D. H., Jiang, K. D., Jiang, S. D., Xu, Y. F., & Yao, H. (2005). The tumor suppressor adenomatous polyposis coli gene is associated with susceptibility to schizophrenia. *Molecular Psychiatry*, *10*(7), 669–677. https://doi.org/10.1038/sj.mp.4001653

Derti, A., Garrett-Engele, P., MacIsaac, K. D., Stevens, R. C., Sriram, S., Chen, R., Rohl, C. A., Johnson, J. M., & Babak, T. (2012). A quantitative atlas of polyadenylation in five mammals. *Genome Research*, *22*(6), 1173–1183. https://doi.org/10.1101/gr.132563.111

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, *34*(2), 184–191. https://doi.org/10.1038/nbt.3437

Dvinge, H., Guenthoer, J., Porter, P. L., & Bradley, R. K. (2019). RNA components of the spliceosome regulate tissueand cancer-specific alternative splicing. *Genome Research*, *29*(10), 1591–1604. https://doi.org/10.1101/gr.246678.118

Dvinge, H., Kim, E., Abdel-Wahab, O., & Bradley, R. K. (2016). RNA splicing factors as oncoproteins and tumour suppressors. *Nature Reviews Cancer*, *16*(7), 413–430. https://doi.org/10.1038/nrc.2016.51

Dvinge, H., Ries, R. E., Ilagan, J. O., Stirewalt, D. L., Meshinchi, S., & Bradley, R. K. (2014). Sample processing obscures cancer-specific alterations in leukemic transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(47), 16802–16807. https://doi.org/10.1073/pnas.1413374111

Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., Rubinstein, L., Shankar, L., Dodd, L., Kaplan, R., Lacombe, D., & Verweij, J. (2009). New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer*, *45*(2), 228–247. https://doi.org/10.1016/j.ejca.2008.10.026

Elkon, R., Ugalde, A. P., & Agami, R. (2013). Alternative cleavage and polyadenylation: Extent, regulation and function. *Nature Reviews Genetics*, *14*(7), 496–506. https://doi.org/10.1038/nrg3482

Feng, X., Li, L., Wagner, E. J., & Li, W. (2018). TC3A: The Cancer 3′ UTR Atlas. *Nucleic Acids Research*, *46*(D1), D1027–D1030. https://doi.org/10.1093/nar/gkx892

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International Journal of Cancer*, *144*(8), 1941–1953. https://doi.org/10.1002/ijc.31937

Flicek, P., Ahmed, I., Amode, M. R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gil, L., García-Girón, C., Gordon, L., Hourlier, T., Hunt, S., Juettemann, T., Kähäri, A. K., Keenan, S., … Searle, S. M. J. (2013). Ensembl 2013. *Nucleic Acids Research*, *41*(D1). https://doi.org/10.1093/nar/gks1236

Flieswasser, T., van den Eynde, A., van Audenaerde, J., de Waele, J., Lardon, F., Riether, C., de Haard, H., Smits, E., Pauwels, P., & Jacobs, J. (2022). The CD70-CD27 axis in oncology: the new kids on the block. In *Journal of Experimental and Clinical Cancer Research* (Vol. 41, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13046-021-02215-y

Gasperini, M., Findlay, G. M., McKenna, A., Milbank, J. H., Lee, C., Zhang, M. D., Cusanovich, D. A., & Shendure, J. (2017). CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. *American Journal of Human Genetics*, *101*(2), 192–205. https://doi.org/10.1016/j.ajhg.2017.06.010

Gide, T. N., Quek, C., Menzies, A. M., Tasker, A. T., Shang, P., Holst, J., Madore, J., Lim, S. Y., Velickovic, R., Wongchenko, M., Yan, Y., Lo, S., Carlino, M. S., Guminski, A., Saw, R. P. M., Pang, A., McGuire, H. M., Palendira, U., Thompson, J. F., … Wilmott, J. S. (2019). Distinct Immune Cell Populations Define Response to Anti-PD-1 Monotherapy and Anti-PD-1/Anti-CTLA-4 Combined Therapy. *Cancer Cell*, *35*(2), 238-255.e6. https://doi.org/10.1016/j.ccell.2019.01.003

Goering, R., Engel, K. L., Gillen, A. E., Fong, N., David, L., & Taliaferro, J. M. (2020). *LABRAT reveals association of alternative polyadenylation with transcript localization , RNA binding protein expression , transcription speed , and cancer survival*.

Gonatopoulos-Pournatzis, T., Aregger, M., Brown, K. R., Farhangmehr, S., Braunschweig, U., Ward, H. N., Ha, K. C. H., Weiss, A., Billmann, M., Durbic, T., Myers, C. L., Blencowe, B. J., & Moffat, J. (2020). Genetic interaction mapping and exon-resolution functional genomics with a hybrid Cas9–Cas12a platform. *Nature Biotechnology*, *38*(5), 638–648. https://doi.org/10.1038/s41587-020-0437-z

Gruber, A. J., & Zavolan, M. (2019). Alternative cleavage and polyadenylation in health and disease. *Nature Reviews Genetics*, *20*(10), 599–614. https://doi.org/10.1038/s41576-019-0145-z

He, X., & Xu, C. (2020). Immune checkpoint signaling and cancer immunotherapy. In *Cell Research* (Vol. 30, Issue 8, pp. 660–669). Springer Nature. https://doi.org/10.1038/s41422-020-0343-4

Hu, C., Liu, C., Li, J., Yu, T., Dong, J., Chen, B., Du, Y., Tang, X., & Xi, Y. (2021). Construction of Two Alternative Polyadenylation Signatures to Predict the Prognosis of Sarcoma Patients. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/fcell.2021.595331

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MaCdonald, J., Obenchain, V., Oleš, A. K., … Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, *12*(2), 115–121. https://doi.org/10.1038/nmeth.3252

Inoue, D., Chew, G. L., Liu, B., Michel, B. C., Pangallo, J., D’Avino, A. R., Hitchman, T., North, K., Lee, S. C. W., Bitner, L., Block, A., Moore, A. R., Yoshimi, A., Escobar-Hoyos, L., Cho, H., Penson, A., Lu, S. X., Taylor, J., Chen, Y., … Bradley, R. K. (2019). Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature*, *574*(7778), 432–436. https://doi.org/10.1038/s41586-019-1646-9

Ji, Z., Lee, J. Y., Pan, Z., Jiang, B., & Tian, B. (2009). Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(17), 7028–7033. https://doi.org/10.1073/pnas.0900028106

Kargapolova, Y., Levin, M., Lackner, K., & Danckwardt, S. (2017). SCLIP-An integrated platform to study RNA-protein interactomes in biomedical research: Identification of CSTF2tau in alternative processing of small nuclear RNAs. *Nucleic Acids Research*, *45*(10), 6074–6086. https://doi.org/10.1093/nar/gkx152

Katz, Y., Wang, E. T., Airoldi, E. M., & Burge, C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods*, *7*(12), 1009–1015. https://doi.org/10.1038/nmeth.1528

Kinzler, K. W., & Vogelstein, B. (1996). Lessons from Hereditary Review Colorectal Cancer. *Cell* (Vol. 87).

Komatsu, M., Waguri, S., Ueno, T., Iwata, J., Murata, S., Tanida, I., Ezaki, J., Mizushima, N., Ohsumi, Y., Uchiyama, Y., Kominami, E., Tanaka, K., & Chiba, T. (2005). Impairment of starvation-induced and constitutive autophagy in Atg7-deficient mice. *Journal of Cell Biology*, *169*(3), 425–434. https://doi.org/10.1083/jcb.200412022

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3). https://doi.org/10.1186/gb-2009-10-3-r25

Lawson, K. A., Sousa, C. M., Zhang, X., Kim, E., Akthar, R., Caumanns, J. J., Yao, Y., Mikolajewicz, N., Ross, C., Brown, K. R., Zid, A. A., Fan, Z. P., Hui, S., Krall, J. A., Simons, D. M., Slater, C. J., de Jesus, V., Tang, L., Singh, R., … Moffat, J. (2020). Functional genomic landscape of cancer-intrinsic evasion of killing by T cells. *Nature*, *586*(7827), 120–126. https://doi.org/10.1038/s41586-020-2746-2

Lee, K., Zheng, Q., Lu, Q., Xu, F., Qin, G., Zhai, Q., Hong, R., Chen, M., Deng, W., & Wang, S. (2021). CPSF4 promotes triple negative breast cancer metastasis by upregulating MDM4. In

Signal Transduction and Targeted Therapy (Vol. 6, Issue 1). Springer Nature. https://doi.org/10.1038/s41392-021-00565-9

Leeman, D. S., Hebestreit, K., Ruetz, T., Webb, A. E., McKay, A., Pollina, E. A., Dulken, B. W., Zhao, X., Yeo, R. W., Ho, T. T., Mahmoudi, S., Devarajan, K., Passegué, E., Rando, T. A., Frydman, J., & Brunet, A. (2018). Lysosome activation clears aggregates and enhances quiescent neural stem cell activation during aging. *Science*, *359*(6381), 1277–1283. https://doi.org/10.1126/science.aag3048

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*. https://doi.org/10.1186/1471-2105-12-323

Li, W., You, B., Hoque, M., Zheng, D., Luo, W., Ji, Z., Park, J. Y., Gunderson, S. I., Kalsotra, A., Manley, J. L., & Tian, B. (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 Genetics*, *11*(4). https://doi.org/10.1371/journal.pgen.1005166

Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S., & Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes and Development*, *27*(21), 2380–2396. https://doi.org/10.1101/gad.229328.113

Liu, D., Schilling, B., Liu, D., Sucker, A., Livingstone, E., Jerby-Amon, L., Zimmer, L., Gutzmer, R., Satzger, I., Loquai, C., Grabbe, S., Vokes, N., Margolis, C. A., Conway, J., He, M. X., Elmarakeby, H., Dietlein, F., Miao, D., Tracy, A., … Schadendorf, D. (2019). Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nature Medicine*, *25*(12), 1916–1927. https://doi.org/10.1038/s41591-019-0654-5

Manguso, R. T., Pope, H. W., Zimmer, M. D., Brown, F. D., Yates, K. B., Miller, B. C., Collins, N. B., Bi, K., la Fleur, M. W., Juneja, V. R., Weiss, S. A., Lo, J., Fisher, D. E., Miao, D., van Allen, E., Root, D. E., Sharpe, A. H., Doench, J. G., & Haining, W. N. (2017). In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. *Nature*, *547*(7664), 413–418. https://doi.org/10.1038/nature23270

Mayr, C., & Bartel, D. P. (2009). Widespread Shortening of 3′UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell*, *138*(4), 673–684. https://doi.org/10.1016/j.cell.2009.06.016

Meyer, L. R., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Kuhn, R. M., Wong, M., Sloan, C. A., Rosenbloom, K. R., Roe, G., Rhead, B., Raney, B. J., Pohl, A., Malladi, V. S., Li, C. H., Lee, B. T., Learned, K., Kirkup, V., Hsu, F., Heitner, S., … Kent, W. J. (2013). The UCSC Genome Browser database: Extensions and updates 2013. *Nucleic Acids Research*, *41*(D1). https://doi.org/10.1093/nar/gks1048

Mitschka, S., & Mayr, C. (2021). Endogenous p53 expression in human and mouse is not regulated by its 30utr. *ELife*, *10*. https://doi.org/10.7554/eLife.65700

Mohn, J. L., Alexander, J., Pirone, A., Palka, C. D., Lee, S. Y., Mebane, L., Haydon, P. G., & Jacob, M. H. (2014). Adenomatous polyposis coli protein deletion leads to cognitive and autism-like disabilities. *Molecular Psychiatry*, *19*(10), 1133–1142. https://doi.org/10.1038/mp.2014.61

Nidhi, S., Anand, U., Oleksak, P., Tripathi, P., Lal, J. A., Thomas, G., Kuca, K., & Tripathi, V. (2021). Novel crispr–cas systems: An updated review of the current achievements, applications, and future research perspectives. In *International Journal of Molecular Sciences* (Vol. 22, Issue 7). MDPI AG. https://doi.org/10.3390/ijms22073327

Overwijk, W. W., & Restifo, N. P. (2001). B16 as a melanoma model for human. In *Curr Protoc Immunol*. https://doi.org/10.1002/0471142735.im2001s39.B16

Park, S. M., Ou, J., Chamberlain, L., Simone, T. M., Yang, H., Virbasius, C. M., Ali, A. M., Zhu, L. J., Mukherjee, S., Raza, A., & Green, M. R. (2016). U2AF35(S34F) Promotes Transformation by Directing Aberrant ATG7 Pre-mRNA 3′ End Formation. *Molecular Cell*, *62*(4), 479–490. https://doi.org/10.1016/j.molcel.2016.04.011

Pereira-Castro, I., Garcia, B. C., Curinha, A., Neves-Costa, A., Conde-Sousa, E., Moita, L. F., & Moreira, A. (2022). MCL1 alternative polyadenylation is essential for cell survival and mitochondria morphology. *Cellular and Molecular Life Sciences*, *79*(3). https://doi.org/10.1007/s00018-022-04172-x

Perez, A. R., Pritykin, Y., Vidigal, J. A., Chhangawala, S., Zamparo, L., Leslie, C. S., & Ventura, A. (2017). GuideScan software for improved single and paired CRISPR guide RNA design. *Nature Biotechnology*, *35*(4), 347–349. https://doi.org/10.1038/nbt.3804

Preitner, N., Quan, J., Nowakowski, D. W., Hancock, M. L., Shi, J., Tcherkezian, J., Young-Pearse, T. L., & Flanagan, J. G. (2014). APC is an RNA-binding protein, and its interactome provides a link to neural development and microtubule assembly. *Cell*, *158*(2), 368–382. https://doi.org/10.1016/j.cell.2014.05.042

Ransom, B., Goldman, S. A., Meldolesi, J., Zhou, L., Murai, K. K., Harris, K. M., Mccarthy, K. D., Li, N., Doyle, R. T., Haydon, P. G., Zielke, H. R., Ni, Y., Sunjara, V., Hua, X., Parpura, V., Jiang, L., Goldman, S. A., Nedergaard, M., Xu, J., … Roman, R. J. (2008). Proliferating Cells Express mRNAs with. *Science*, *320*(June), 1643–1647.

Raz, V., Dickson, G., & 't Hoen, P. A. C. (2017). *Dysfunctional transcripts are formed by alternative polyadenylation in OPMD* (Vol. 8, Issue 43). www.impactjournals.com/oncotarget

Ringel, T., Frey, N., Ringnalda, F., Janjuha, S., Cherkaoui, S., Butz, S., Srivatsa, S., Pirkl, M., Russo, G., Villiger, L., Rogler, G., Clevers, H., Beerenwinkel, N., Zamboni, N., Baubec, T., & Schwank, G. (2020). Genome-Scale CRISPR Screening in Human Intestinal Organoids Identifies Drivers of TGF-β Resistance. *Cell Stem Cell*, *26*(3), 431-440.e8. https://doi.org/10.1016/j.stem.2020.02.007

Robinson, D. R., Wu, Y. M., Lonigro, R. J., Vats, P., Cobain, E., Everett, J., Cao, X., Rabban, E., Kumar-Sinha, C., Raymond, V., Schuetze, S., Alva, A., Siddiqui, J., Chugh, R., Worden, F., Zalupski, M. M., Innis, J., Mody, R. J., Tomlins, S. A., … Chinnaiyan, A. M. (2017). Integrative clinical genomics of metastatic cancer. *Nature*, *548*(7667), 297–303. https://doi.org/10.1038/nature23306

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., Beltran, H., Abida, W., Bradley, R. K., Vinson, J., Cao, X., Vats, P., Kunju, L. P., Hussain, M., Feng, F. Y., … Chinnaiyan, A. M. (2015). Integrative clinical genomics of advanced prostate cancer. *Cell*, *161*(5), 1215–1228. https://doi.org/10.1016/j.cell.2015.05.001

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., & Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*, *160*(1–2), 48–61. https://doi.org/10.1016/j.cell.2014.12.033

Shin, J., Ding, Q., Wang, L., Cui, Y., Baljinnyam, E., Guvenek, A., & Tian, B. (2022). CRISPRpas: Programmable regulation of alternative polyadenylation by dCas9. *Nucleic Acids Research*, *50*(5), E25. https://doi.org/10.1093/nar/gkab519

Sobecki, M., Mrouj, K., Colinge, J., Gerbe, F., Jay, P., Krasinska, L., Dulic, V., & Fisher, D. (2017). Cell-cycle regulation accounts for variability in Ki-67 expression levels. *Cancer Research*, *77*(10), 2722–2734. https://doi.org/10.1158/0008-5472.CAN-16-0707

Spies, N., Burge, C. B., & Bartel, D. P. (2013). 3′ UTR-Isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Research*, *23*(12), 2078–2090. https://doi.org/10.1101/gr.156919.113

Tan, S., Zhang, M., Shi, X., Ding, K., Zhao, Q., Guo, Q., Wang, H., Wu, Z., Kang, Y., Zhu, T., Sun, J., & Zhao, X. (2021). CPSF6 links alternative polyadenylation to metabolism adaption in hepatocellular carcinoma progression. *Journal of Experimental and Clinical Cancer Research*, *40*(1). https://doi.org/10.1186/s13046-021-01884-z

Tanley, S., Amilton, R. H., Amon, R., Arsons, E. P., Apadopoulos, I. P., En, I. J., Teven, S., Owell, M. P., Rush, N. J. K., Heresa, T., Erk, B. A., Ane, Z., Ohen, C., Etu, T., Urger, E. C. B., Ood, A. A. W., Owzia, F., Aqi, T., Usan, S., … Inzler, W. K. (1995). The Molecular Basis of Turcot’s Syndrome (Vol. 332, Issue 13).

Thomas, J. D., Polaski, J. T., Feng, Q., de Neef, E. J., Hoppe, E. R., McSharry, M. v., Pangallo, J., Gabel, A. M., Belleville, A. E., Watson, J., Nkinsi, N. T., Berger, A. H., & Bradley, R. K. (2020). RNA isoform screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison exons. *Nature Genetics*, *52*(1), 84–94. https://doi.org/10.1038/s41588-019-0555-z

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105–1111. https://doi.org/10.1093/bioinformatics/btp120

Tyner, J. W., Tognon, C. E., Bottomly, D., Wilmot, B., Kurtz, S. E., Savage, S. L., Long, N., Schultz, A. R., Traer, E., Abel, M., Agarwal, A., Blucher, A., Borate, U., Bryant, J., Burke, R., Carlos, A., Carpenter, R., Carroll, J., Chang, B. H., … Druker, B. J. (2018). Functional genomic landscape of acute myeloid leukaemia. *Nature*, *562*(7728), 526–531. https://doi.org/10.1038/s41586-018-0623-z

van Allen, E. M., Miao, D., Schilling, B., Shukla, S. A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Foppen, M. H. G., Goldinger, S. M., Utikal, J., Hassel, J. C., Weide, B., Kaehler, K. C., Loquai, C., Mohr, P., Gutzmer, R., Dummer, R., Gabriel, S., … Garraway, L. A. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*, *350*(6257), 207–211. https://doi.org/10.1126/science.aad0095

Vickers, T. A., Wyatt, J. R., Burckin, T., Bennett, C. F., & Freier, S. M. (2001). Fully modified 2′ MOE oligonucleotides redirect polyadenylation. In *Nucleic Acids Research* (Vol. 29, Issue 6).

Wang, Q., He, G., Hou, M., Chen, L., Chen, S., Xu, A., & Fu, Y. (2018). Cell Cycle Regulation by Alternative Polyadenylation of CCND1. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-25141-0

Wang, R., & Tian, B. (2020). APAlyzer: A bioinformatics package for analysis of alternative polyadenylation isoforms. *Bioinformatics*, *36*(12), 3907–3909. https://doi.org/10.1093/bioinformatics/btaa266

Wang, T., Hamilla, S., Cam, M., Aranda-Espinoza, H., & Mili, S. (2017). Extracellular matrix stiffness and cell contractility control RNA localization to promote cell migration. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00884-y

Xia, Z., Donehower, L. A., Cooper, T. A., Neilson, J. R., Wheeler, D. A., Wagner, E. J., & Li, W. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′ 2-UTR landscape across seven tumour types. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms6274

Xiang, Y., Ye, Y., Lou, Y., Yang, Y., Cai, C., Zhang, Z., Mills, T., Chen, N. Y., Kim, Y., Ozguc, F. M., Diao, L., Karmouty-Quintana, H., Xia, Y., Kellems, R. E., Chen, Z., Blackburn, M. R., Yoo, S. H., Shyu, A. bin, Mills, G. B., & Han, L. (2018). Comprehensive characterization of alternative polyadenylation in human cancer. *Journal of the National Cancer Institute*, *110*(4), 379–389. https://doi.org/10.1093/jnci/djx223

Xu, C., & Zhang, J. (2018). Alternative Polyadenylation of Mammalian Transcripts Is Generally Deleterious, Not Adaptive. *Cell Systems*, *6*(6), 734-742.e4. https://doi.org/10.1016/j.cels.2018.05.007

Xue, Z., Warren, R. L., Gibb, E. A., MacMillan, D., Wong, J., Chiu, R., Hammond, S. A., Yang, C., Nip, K. M., Ennis, C. A., Hahn, A., Reynolds, S., & Birol, I. (2018). Recurrent tumor-specific regulation of alternative polyadenylation of cancer-related genes. *BMC Genomics*, *19*(1), 1–12. https://doi.org/10.1186/s12864-018-4903-7

Yao, C., Biesinger, J., Wan, J., Weng, L., Xing, Y., Xie, X., & Shi, Y. (2012). Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(46), 18773–18778. https://doi.org/10.1073/pnas.1211101109

Zhang, H., Lee, J. Y., & Tian, B. (2005). Biased alternative polyadenylation in human tissues. *Genome Biology*, *6*(12), 1–13. https://doi.org/10.1186/gb-2005-6-12-r100

Zhang, L., & Shay, J. W. (2017). Multiple Roles of APC and its Therapeutic Implications in Colorectal Cancer. In *Journal of the National Cancer Institute* (Vol. 109, Issue 8). Oxford University Press. https://doi.org/10.1093/jnci/djw332

Zhang, Y., Liu, X., Li, A., & Tang, X. (2022). A pan-cancer analysis on the carcinogenic effect of human adenomatous polyposis coli. *PLoS ONE*, *17* (3 March). https://doi.org/10.1371/journal.pone.0265655

Zhang, Y., Wang, Y., Li, C., & Jiang, T. (2020). Systemic Analysis of the Prognosis-Associated Alternative Polyadenylation Events in Breast Cancer. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.590770

Zhang, Y., Xu, Y., & Wang, Y. (2022). Alternative polyadenylation associated with prognosis and therapy in colorectal cancer. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-11089-9

Zhu, J., Li, Y., Tian, Z., Hua, X., Gu, J., Li, J., Liu, C., Jin, H., Wang, Y., Jiang, G., Huang, H., & Huang, C. (2017). ATG7 Overexpression Is Crucial for Tumorigenic Growth of Bladder Cancer In Vitro and In Vivo by Targeting the ETS2/miRNA196b/FOXO1/p27 Axis. *Molecular Therapy - Nucleic Acids*, *7*, 299–313. https://doi.org/10.1016/j.omtn.2017.04.012