From Genetics to Anti-Tumor Immunity: Exploring Poison Exons, 3' UTRs, and Vault Organelles

Andrea Belleville

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Robert Bradley, Chair

Alice Berger

Evan Newell

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

From Genetics to Anti-Tumor Immunity: Exploring Poison Exons, 3' UTRs, and Vault Organelles

Andrea Belleville

Chair of the Supervisory Committee:
Robert Bradley
Public Health Sciences and Basic Sciences Divisions
Fred Hutchinson Cancer Center

This work integrates three distinct yet interconnected projects, providing novel insights into genomic regulation, cancer biology, and immunotherapy. The first project reveals the critical role of a 'poison exon' in the gene *Smndc1*, demonstrating that poison exon deletion in mouse and *Arabidopsis thaliana* models disrupts mRNA processing and impacts organism phenotypes. This highlights the importance of conserved poison exons in molecular and organism fitness. The second project focuses on alternative polyadenylation in melanoma, demonstrating that manipulation of the alternative polyadenylation site of *Atg7* alters mRNA stability and ATG7 protein levels and ultimately impacts cell proliferation. This suggests alternative polyadenylation as a potential target in cancer therapy. The final project investigates Major Vault Protein (MVP) in the immune response to cancer, establishing its correlation with improved survival and enhanced response to immune checkpoint blockade in melanoma and renal cell carcinoma. These findings underscore MVP as a promising biomarker and therapeutic target in cancer, illustrating the interconnected nature of molecular mechanisms across diverse biological contexts. All three

projects collectively underscore the intricate network of genomic mechanisms governing cellular processes, from mRNA processing to immune responses, highlighting the potential for targeted genetic and molecular interventions in advancing cancer therapy and improving patient outcomes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation reflects not just my efforts but the invaluable contributions of many. My advisor, Rob Bradley, has been a patient and guiding force in shaping both the research presented here as well as my broader scientific perspective. I credit the independence that Rob allowed me in the lab to a great deal of personal and professional growth I experienced while in graduate training. I appreciate the many ways he has prepared me to continue learning throughout my career. I am also indebted to James Thomas for his generous mentorship in both experimental and computational methods, which has been fundamental to my work.

I extend my heartfelt thanks to all members of the Bradley Lab, including Austin Gabel, Toshihiro Banjo, Edie Crosse, Dylan Udy, Jose Pineda, Emma Hoppe, Emma De Neef, Siegen McKellar, Erik Kimble, Taylor Nicholas, Jacob Polaski, Joey Pangallo, and Khrystyna North, all of whom have been supportive scientists and wonderful colleagues. Their support played a significant role in my academic journey. My gratitude also goes to my committee members, Evan Newell, Alice Berger, Edith Wang, and Sita Kugel, for their insightful feedback and encouragement over the years. Special acknowledgment goes to my friends in the UW Medical Scientist Training Program, for their camaraderie through many exciting and difficult years. My time at Fred Hutchinson Cancer Center (FHCC) has been significantly enriched by many postdocs, graduate students, technicians, scientific support, and administrative staff. The vibrant scientific community at FHCC has been instrumental in shaping my education and career path, and I am profoundly inspired by their work.

Finally, my deepest appreciation is reserved for my family, especially my parents and siblings, and my husband, whose endless support and belief have carried me through my training. My parents first instilled in me the understanding that learning is an immense blessing and privilege, and that I have a responsibility to always seek improvement. Their encouragement, love, and support have been a constant in my life, for which I cannot express sufficient gratitude.

Chapter 1. INTRODUCTION

Alternative mRNA splicing is critical for many biological processes (Wang et al., 2008, Baralle et al., 2017), both expanding proteome diversity and regulating mRNA abundance (Graveley 2001, Soergel et al., 2013). The importance of alternative splicing is underscored by its conservation across eukaryotes (Nilsen & Graveley, 2010), as well as the many ways splicing and mRNA processing can be dysregulated in disease (Dvinge et al., 2016, Carvill & Mefford, 2020, Gruber & Zavolan, 2019). I will begin this introduction by highlighting the significance of alternatively spliced exons called 'poison exons,' before discussing the importance of processing within the pre-mRNA 3' untranslated regions (UTR), especially in the context of cancer.

Poison exons are alternatively spliced exons that disrupt their host transcript's reading frame, thereby introducing a premature termination codon, and ultimately leading to degradation via Nonsense-Mediated mRNA Decay (NMD) (Kurosaki et al., 2019). Poison exons are an intriguing class of exons as they do not contribute to the protein-coding potential of mRNA transcripts, yet are highly conserved (Lareau et al., 2007, Ni et al., 2007). In fact, many poison exons overlap ultraconserved elements (UCE) within the genome, which are 200 nucleotide regions conserved between humans, mice, and rats (Bejerano et al., 2004). There are 481 UCE sequences in the genome, over 75% of which are non-coding elements including poison exons, ncRNA, repressors, and enhancers (Snetkova et al., 2021). UCEs are perfectly sequence conserved, they demonstrate reduced rates of sequence variants, and they are depleted within copy number variants in healthy tissue—all supporting the idea that purifying selection acts upon these regions (Snetkova et al., 2021). Here we encounter an interesting dilemma—why are non-coding regions of the genome under such strong purifying selection and what does this tell us

about their biological role or essentiality? In Chapter 2, I address this question of poison exon essentiality for life by focusing on a highly conserved poison exon in the gene Survival Motor Neuron Domain Containing Protein 1 (*Smndc1*). We also previously addressed the biological role of poison exons in cancer using a CRISPR/Cas9 screen to specifically deplete poison exons in lung adenocarcinoma cells (Thomas et al., 2019). We found many poison exons are essential for tumor growth, while others have a tumor-suppressive role. This highlights a dynamic process by which cancer cells may regulate the alternative splicing of highly conserved poison exons to modulate transcript abundance.

In cancer, the modulation of mRNA transcript abundance is not only managed through poison exon splicing, but through many additional alternative mechanisms, including alternative polyadenylation (APA) of pre-mRNA 3' UTR (Lianoglou et al., 2013). APA impacts around 70% of genes, with many poly(A) sites being common across mammalian species (Derti et al., 2012). The selection of poly(A) sites is often influenced by upstream intronic splicing (Rigo and Martinson, 2008; Movassat et al., 2016), and involves a network of RNA-binding proteins and polyadenylation machinery (Tian and Manley, 2016; Mitschka and Mayr, 2021). APA's influence is context-dependent; for instance, during mouse myogenesis, there is a tendency for mouse genes to produce mRNAs with longer 3' UTRs (Ji et al., 2009). Cells that divide rapidly typically favor proximal poly(A) sites and therefore have shorter 3′UTRs (Sandberg et al., 2008). In the context of RNA processing, APA is notably dysregulated in all human cancers studied, yet the comprehensive investigation of APA's functional role in disease progression remains unexplored (Gruber & Zavolan, 2019). In Chapter 3, I will discuss a specific APA event in the critical autophagy gene *Atg7*. We apply CRISPR/Cas9-directed editing to force usage of a distal

poly(A) site of *Atg7* and demonstrate the resulting reduction in tumor cell proliferation and improved host survival.

Understanding cancer as a largely genomic disease, which includes the dysregulation of processes like alternative polyadenylation (APA) and poison exon splicing, has laid the foundation for advances in immunotherapy. The effectiveness of immune checkpoint blockade (ICB), which activates the patient's immune system to target cancer cells, can be predicted based on the genomic characteristics of an individual's cancer. For example, tumor mutational burden (TMB) is a predictive biomarker for ICB response, especially when also considering MHC and T-cell receptor repertoire to present and recognize subsequent neoantigens (Jardim et al., 2020). Other intuitive biomarkers for ICB therapy include PD-L1 expression on tumor cells (Patel & Kurzrock, 2015). While there are a growing number of FDA-approved inhibitors for multiple malignancies, and some patients maintain remarkable responses to ICB therapy, the overall response rates to ICB monotherapy leave critical room for improvement. Beyond targeting immune checkpoints on tumor-infiltrating T-cells to enhance anti-tumor immunity, other critical cell types govern suppressive or permissive tumor microenvironments (TME) and warrant additional study. Bridging the gap between the limitations of current therapies and the exploration of novel targets, in Chapter 4 I explore Major Vault Protein (MVP), which has been understudied in the context of the tumor microenvironment. MVP is the major structural protein in the vault particle, which has been implicated in cancer chemotherapeutic response, but is not well understood in the context of immunotherapy. *MVP* gene expression is positively correlated with improved patient survival in both melanoma and renal cell carcinoma. We demonstrate preliminary evidence that *Mvp* expression may be important for tumor-intrinsic signaling and tumor-suppressive function in mice.

In this dissertation, I use genomic and transcriptomic data in a number of ways. The first is to analyze highly conserved regions of the genome and investigate paradoxically conserved non-coding exons. The second is to assess how alterations in RNA transcript processing impact tumorigenesis. The third is to assess gene expression changes that are associated with cancer outcomes and identify cell populations that may be important for improved immunotherapy response.

REFERENCES

Baralle, F. E., & Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nature Reviews Molecular Cell Biology*, *18*(7), 437–451. https://doi.org/10.1038/nrm.2017.27

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved Elements in the Human Genome. *Science*, *304*(5675), 1321–1325. https://doi.org/10.1126/science.1098119

Bradley, R. K., & Anczuków, O. (2023). RNA splicing dysregulation and the hallmarks of cancer. *Nature Reviews Cancer*, *23*(3), 135–155. https://doi.org/10.1038/s41568-022-00541-7

Carvill, G. L., & Mefford, H. C. (2020). Poison exons in neurodevelopment and disease. *Current Opinion in Genetics & Development*, *65*, 98–102. https://doi.org/10.1016/j.gde.2020.05.030

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

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

Graveley, B. R. (2001). Alternative splicing: increasing diversity in the proteomic world. *Trends in Genetics*, *17*(2), 100–107. https://doi.org/10.1016/S0168-9525(00)02176-4

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

Jardim, D. L., Goodman, A., de Melo Gagliato, D., & Kurzrock, R. (2021). The Challenges of Tumor Mutational Burden as an Immunotherapy Biomarker. *Cancer Cell*, *39*(2), 154–173. https://doi.org/10.1016/j.ccell.2020.10.001

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*, *106*(17), 7028–7033. https://doi.org/10.1073/pnas.0900028106

Kurosaki, T., Popp, M. W., & Maquat, L. E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nature Reviews Molecular Cell Biology*, *20*(7), 406–420. https://doi.org/10.1038/s41580-019-0126-2

Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., & Brenner, S. E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, *446*(7138), 926–929. https://doi.org/10.1038/nature05676

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 & Development*, *27*(21), 2380–2396. https://doi.org/10.1101/gad.229328.113

Mitschka, S., & Mayr, C. (2021). Endogenous p53 expression in human and mouse is not regulated by its 3′UTR. *eLife*, *10*. https://doi.org/10.7554/eLife.65700

Movassat, M., Crabb, T. L., Busch, A., Yao, C., Reynolds, D. J., Shi, Y., & Hertel, K. J. (2016). Coupling between alternative polyadenylation and alternative splicing is limited to terminal introns. *RNA Biology*, *13*(7), 646–655. https://doi.org/10.1080/15476286.2016.1191727

Ni, J. Z., Grate, L., Donohue, J. P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T. A., Blume, J. E., & Ares, M. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes & Development*, *21*(6), 708–718. https://doi.org/10.1101/gad.1525507

Nilsen, T. W., & Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature*, *463*(7280), 457–463. https://doi.org/10.1038/nature08909

Patel, S. P., & Kurzrock, R. (2015). PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy. *Molecular Cancer Therapeutics*, *14*(4), 847–856. https://doi.org/10.1158/1535-7163.MCT-14-0983

Rigo, F., & Martinson, H. G. (2008). Functional Coupling of Last-Intron Splicing and 3′-End Processing to Transcription In Vitro: the Poly(A) Signal Couples to Splicing before Committing to Cleavage. *Molecular and Cellular Biology*, *28*(2), 849–862. https://doi.org/10.1128/MCB.01410-07

Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., & Burge, C. B. (2008). Proliferating Cells Express mRNAs with Shortened 3' Untranslated Regions and Fewer MicroRNA Target Sites. *Science*, *320*(5883), 1643–1647. https://doi.org/10.1126/science.1155390

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., Weinstock, G. M., Wilson, R. K., Gibbs, R. A., Kent, W. J., Miller, W., & Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, *15*(8), 1034–1050. https://doi.org/10.1101/gr.3715005

Snetkova, V., Pennacchio, L. A., Visel, A., & Dickel, D. E. (2022). Perfect and imperfect views of ultraconserved sequences. *Nature Reviews Genetics*, *23*(3), 182–194. https://doi.org/10.1038/s41576-021-00424-x

Soergel, D. A. W., Lareau, L. F., & Brenner, S. E. (n.d.). Regulation of Gene Expression by Coupling of Alternative Splicing and NMD. In *Madame Curie Bioscience Database*.

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

Tian, B., & Manley, J. L. (2017). Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology*, *18*(1), 18–30. https://doi.org/10.1038/nrm.2016.116

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, *456*(7221), 470–476. https://doi.org/10.1038/nature07509

Chapter 2. AN AUTOREGULATORY POISON EXON IS CONSERVED ACROSS KINGDOMS AND INFLUENCES ORGANISM GROWTH

A version of this chapter is currently under review at *PloS Genetics* as:
An autoregulatory poison exon is conserved across kingdoms and influences organism growth. Andrea E. Belleville, James D. Thomas, Jackson Tonnies, Austin M. Gabel, Priti Singh, Christine Queitsch, Robert K. Bradley

James D. Thomas assisted with mouse experiments, RNA-sequencing, and read mapping. Priti Singh performed CRISPR/Cas9 mouse model generation, with further model development assisted by Austin Gabel and James D. Thomas. Jackson Tonnies assisted and conducted plant experiments, with input from Christine Queitsch. James D. Thomas and Robert Bradley guided experimental design. Unless noted above, I designed and performed experiments, generated figures, and am the primary author of the work detailed in this Chapter. I received edits from Robert Bradley, James D. Thomas, and Austin Gabel, and I am grateful for the feedback of many collaborators and anonymous reviewers.

2.1 SUMMARY

Many of the most highly conserved elements in the human genome are 'poison exons,' alternatively spliced exons that contain premature termination codons and permit post-transcriptional regulation of mRNA abundance through induction of nonsense-mediated mRNA decay (NMD). Poison exons are widely assumed to be highly conserved due to their importance for organismal fitness, but this functional importance has never been tested in the context of a whole organism. Here, we report that a poison exon in *Smndc1* is conserved across mammals and plants and plays a molecular autoregulatory function in both kingdoms. We generated mouse and *A. thaliana* models lacking this poison exon to find its loss leads to deregulation of SMNDC1

protein levels, pervasive alterations in mRNA processing, and organismal size restriction. Together, these models demonstrate the importance of poison exons for both molecular and organismal phenotypes that likely explain their extraordinary conservation.

2.2 INTRODUCTION

Alternative splicing (AS) is a post-transcriptional mechanism that regulates the function and expression of multi-exonic genes. While AS expands an organism's proteomic diversity, it is estimated that up to one third of human genes produce non-coding mRNAs are immediately subject to degradation (Lewis et al., 2003). Specifically, inclusion of a class of cassette exons termed 'poison exons' introduces one or more premature termination codons (PTC) into the mature mRNA, subjecting the mRNA transcript to degradation via nonsense-mediated mRNA decay (NMD). Poison exon AS coupled to NMD serves as a means of post-transcriptional gene regulation, and it has been implicated as a critical mechanism in neurodevelopment, environmental adaptation, and disease (Carvill & Mefford, 2020; de Oliveira Freitas Machado et al., 2023; Neumann et al., 2020; Thomas et al., 2020; Zhang et al., 2016).

Alternative splicing of poison exons is a conserved regulatory process across metazoan, yeast, and plants (Kurihara et al., 2009; Lareau & Brenner, 2015; Rayson et al., 2012). Many poison exons exhibit significant nucleotide sequence conservation, despite lacking protein coding potential, and conserved poison exons are enriched in RNA-binding proteins, such as serine-arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (Lareau et al., 2007; Ni et al., 2007). Within these splicing factor families, it is proposed that poison exons may provide a context-specific auto- and cross-regulatory mechanism to regulate splicing factor abundance (Leclair et al., 2020). Therefore, poison exon conservation within splicing factors is

believed to be due to their important role in maintaining splicing factor homeostasis and consequently global splicing.

Outside of their role in regulating splicing factor abundance, little is known about the strong purifying selection driving poison exon sequence conservation. Our recent work demonstrated that endogenous deletion of poison exons influenced growth rates in cancer cell lines, an indication of their potential impact on cellular fitness (Thomas et al., 2020). Beyond this, no study has yet tested the essentiality of poison exons in an *in vivo* setting. This underlines the pressing need for further investigation into their functional relevance in organismal physiology.

In this study, we undertook a detailed study of a poison exon that is present in both mammals and plants despite an evolutionary distance exceeding 1 billion years (Parfrey et al., 2011). Specifically, we studied the poison exon within the murine gene Survival motor neuron domain containing 1 (*Smndc1*), which encodes a well-conserved spliceosome component SMNDC1. *Smndc1* contains a Tudor domain by which it binds methylated arginines on other splicing proteins, playing an essential role in spliceosome assembly (Tripsianes et al., 2011, Meister et al., 2001; Rappsilber et al., 2001, Côté et al., 2005). Like many other genes encoding splicing factors, human *SMNDC1* contains an NMD-targeted poison exon which is thought to regulate *SMNDC1* transcript abundance (Saltzman et al., 2008). Recently, an ortholog to *SMNDC1* was identified in plants called splicing factor 30 (*SPF30*), which undergoes extensive alternative splicing in many plant species (Zhang et al., 2019). Given the conservation of *SMNDC1* and *SPF30* across kingdoms, we hypothesized that this splicing factor gene might prove a compelling exemplar to investigate the potential essentiality of poison exons in physiological contexts.

2.3 RESULTS

2.3.1 Identification of SMNDC1 orthologs and unannotated poison exons

We first sought to understand the extent of similarity between the structure and sequence of the human *SMNDC1* and mouse *Smndc1* genes. In both human and mouse orthologs, the poison exon is located between the second and third coding exons and introduces a frameshift when spliced into the mRNA transcript, introducing a PTC in the downstream coding exon (**Fig. 1A**). At the nucleotide level, the poison exons retain greater sequence similarity between humans and mice (94.7% similar), than the upstream and downstream coding exons (86.7% and 90.9%, respectively; **Supplementary Fig. 1A**). We sought to identify other *SMNDC1* orthologs using protein sequence conservation and reciprocal best hit analysis on a range of animal, fungal, and plant model species (**Fig. 1B**, **Supplementary Table 1**). Consistent with previously published reports (Mier & Pérez-Pulido, 2012), we did not find evidence of *SMNDC1* orthologs in fungal species with reduced genomic intron counts, such as *S. cerevisiae*, suggesting the loss of those orthologs, but identified *SMNDC1* orthologs in all other queried species (**Fig. 1B**).

We next tested whether the *SMNDC1* poison exon was conserved outside of the human and mouse genomes. We first searched for unproductive splice isoforms of each orthologous gene using publicly available expressed sequence tags (ESTs). We identified a poison exon-containing isoform in *G. gallus* from EST data but were unable to conclusively determine whether similar unproductive splice isoforms were present in *SMNDC1* orthologs in other genomes due to the relative paucity of EST data from those species (**Fig. 1B**). For *A. thaliana*, a recent study reported alternative splicing within *SPF30* (Zhang et al., 2019), potentially resulting in inclusion of an in-frame PTC. To determine if the reported *SPF30* alternative splicing

reflected use of a previously unannotated poison exon, we searched publicly available RNA-seq data from cycloheximide (CHX)-treated *A. thaliana* (Drechsel et al., 2013). Indeed, we identified the accumulation of poison exon-containing mRNA transcripts when NMD was inhibited with CHX treatment (**Supplementary Fig. 1B**). Overall, we were able to conclusively identify *SMNDC1* orthologs in most queried species and poison exon-containing isoforms in human, mouse, *G. gallus*, and *A. thaliana*.

To further experimentally confirm that the putative poison exons in both mouse *Smndc1* and *A. thaliana SPF30* are NMD-sensitizing, we treated a murine melanocyte cell line and *A. thaliana* seedlings with CHX. Subsequent RT-PCR demonstrated high levels of inclusion of the poison exon in both mouse cell lines and plant seedlings (**Fig. 1C-D**). These data indicate that an alternatively spliced exon that triggers NMD lies in the same intronic position of the human, mouse, and *A. thaliana* genes (**Fig. 1A**).

2.3.2 A conserved regulatory role of the Smndc1 poison exon

Poison exons are believed to play a role in controlling RNA splicing factor abundance by targeting mRNA transcripts for degradation. This was previously reported for human *SMNDC1*, where overexpression of SMNDC1 protein led to increased inclusion of the poison exon *in vitro* (Saltzman et al., 2008). We therefore hoped to test whether this role is conserved in mice and plant orthologs. We transiently overexpressed murine SMNDC1 protein in murine cells, and then treated with an NMD-inhibitor to allow for accumulation of poison exon-included mRNA transcripts. Using RT-PCR, we determined the corresponding degree of poison exon inclusion at different levels of SMNDC1 protein overexpression. Increased expression of SMNDC1 led to more than a tenfold increase in the poison exon-containing transcript in mouse cells (**Fig. 1E**,

Supplementary Fig. 2A).

Building on these findings, we aimed to determine whether the *Smndc1* poison exon maintains a conserved regulatory role in the presence of orthologous SPF30 protein. We ectopically expressed *A. thaliana* SPF30 in murine cell culture and performed RT-PCR to amplify endogenous *Smndc1* mRNA transcripts (**Fig. 1F**). Similar to the overexpression of murine SMNDC1, the overexpression of SPF30 led to increased poison exon inclusion in mouse cells. This result is consistent with a conserved autoregulatory cycle which increases inclusion of the *Smndc1* poison exon in the context of high protein abundance. Across highly divergent species, this may be a conserved mechanism of regulating total SMNDC1 protein abundance.

2.3.3 Generation of a mouse model lacking the Smndc1 poison exon

Given the conservation of SMNDC1 orthologs and the poison exon, we next aimed to test the essentiality of the *Smndc1* poison exon as a regulatory element *in vivo* by generating a genetically engineered mouse model. The murine *Smndc1* locus is composed of seven NCBI RefSeq annotated exons— five constitutive exons and two alternative 5′ exons encoding untranslated regions of the mRNA (**Fig. 2A**). We selected flanking paired guide RNAs (pgRNAs) with optimal off-target scores to generate an allele with deletion of the *Smndc1* poison exon and an allele with isolated deletion of the poison exon 3′ splice site (**Supplementary Table 2** and **Fig. 2A**). We previously demonstrated that pgRNA targeting of 3′ splice sites minimally impacts genomic DNA sequence while efficiently abrogating exon inclusion *in vitro* (Thomas et al., 2020). Lastly, to control for unanticipated effects of local editing, we generated an allele with a deletion in the upstream intronic region (**Fig 2A**).

To perform CRISPR/Cas9 editing, we injected pgRNAs and Cas9 protein into the

pronucleus of one-cell mouse embryos, which were then transferred to surrogate mothers. Viable pups were derived from all three targeting strategies, with successful generation of 27, 6, and 20 founder pups from the whole exon deletion line, 3′ splice site deletion line, and intronic control line, respectively (**Fig. 2B**). Subsequent genotyping was carried out using PCR, Sanger sequencing, and next-generation amplicon sequencing, indicating many mutated founder mice (**Fig. 2B-C**). From each targeting strategy, individual founder mice were selected based on sequencing results, and the lines were subsequently backcrossed for multiple generations (**Supplementary Fig. 3A**).

To test whether pgRNA-directed CRISPR/Cas9 editing generated alleles impacted the *Smndc1* poison exon inclusion, we generated primary fibroblast lines from wild-type, heterozygous, and homozygous littermates of each of the three deletion lines. We treated fibroblasts with CHX to inhibit NMD, as the inclusion of the poison exon is artificially low in steady state cells. Using RT-PCR, we observed complete loss of the *Smndc1* poison exon-containing transcripts in mice lacking the entire poison exon. Similarly, a 19 nucleotide deletion disrupting the poison exon 3′ splice site was sufficient to completely exclude the poison exon in fibroblast mRNA. In contrast, poison exon inclusion was not impacted by upstream intronic deletion (**Fig. 2D**). This result suggests a limited deletion of necessary splicing sequences surrounding the *Smndc1* poison exon is sufficient to prevent inclusion of the poison exon in our mouse model. As both the poison exon and 3′ splice site deletion alleles effectively prevented exon inclusion and are therefore functionally equivalent, we aggregated the results of both the poison exon deletion model as well as the 3′ splice site deletion model in subsequent analyses.

2.3.4 Generation of an A. thaliana model lacking the SPF30 poison exon

We next sought to model the corresponding loss of the *SPF30* poison exon in *A. thaliana*. The *SPF30* locus is composed of eight exons, with four annotated splice variants. To generate CRISPR/Cas9-edited *A. thaliana* lines, we used floral dip transformations to transfer nuclear-localized Cas9 protein under egg cell-specific promotion and gRNA sequences under constitutive promotion. pgRNAs were selected for limited off-target scores and were designed to flank the poison exon 3′ splice site, as well as the upstream intron as a control for effects of local editing (**Supplementary Table 2** and **Fig. 2E**). We confirmed deletion alleles present in T1 plants through PCR and Sanger sequencing at the appropriate DNA locus, and selected a poison exon deletion allele that removed the 3′ splice site as well as 48 nt of the *SPF30* poison exon (**Supplementary Fig. 3B**). To ensure the deletion at the *SPF30* poison exon indeed prevented exon inclusion and did not introduce aberrant splice isoforms, we collected and pooled T2 seedlings, treated with CHX to inhibit NMD, and performed RT-PCR. RT-PCR was performed using a mixed pool of both heterozygous and homozygous T2 seedlings, the inclusion of the poison exon was 6-fold lower in the *SPF30* poison exon deleted pooled sample relative to the pooled intronic control sample (**Fig. 2F**). Seeds from additional heterozygous T1 plants were planted, resulting plants were genotyped, and homozygous plants were selected for further phenotyping.

2.3.5 The Smndc1 poison exon regulates transcript and protein abundance in vivo

Having established murine *Smndc1* poison exon knockout model, we next sought to determine how loss of the poison exon impacted global *Smndc1* mRNA and protein abundance at organism scale. We began by gathering tissues, including primary fibroblasts, for RNA-seq from

poison exon null mice, heterozygous mice, and wild-type littermates. Consistent with RT-PCR data, poison exon null mice demonstrated little to no junction reads aligning to the *Smndc1* poison exon (**Fig. 3A**). Meanwhile, heterozygous mice demonstrated intermediate poison exon-aligning reads, consistent with reduced allelic dose. Loss of the poison exon further correlated with increases (between 10.50–36.15%) in total *Smndc1* mRNA transcript abundance across all tissues and cell lines assayed (**Fig. 3A**, **Supplementary Fig. 4A,** and **Supplementary Table 3**).

To test whether an increased abundance of *Smndc1* mRNA transcript corresponds with greater protein abundance, we performed Western blotting for SMNDC1. We found SMNDC1 mean abundance was modestly but significantly higher (14.50% increase; $p = 0.012$) in poison exon null mice relative to wild-type mice, consistent with de-repression of *Smndc1* mRNA transcript abundance (**Fig. 3B-C** and **Supplementary Fig. 4B**). Similarly, SMNDC1 immunohistochemistry staining demonstrated increased nuclear protein abundance in poison exon null tissues across multiple tissues assayed (**Fig. 3D**). Using a HALO software image classifier, we quantified cells from each tissue with no, low, moderate, or high SMNDC1 nuclear staining signal. In tested brain, lung, and quadriceps tissue, we detected significant increases ($p < 2.2 \times 10^{-16}$) in the proportions of cells staining for SMNDC1 in all three tissues (**Fig. 3E**). Together, these results demonstrate that loss of the *Smndc1* poison exon impacts global mRNA and protein abundance in steady-state cells and tissues. Interestingly, the impact of poison exon loss was tissue dependent, with especially high dysregulation of SMNDC1 staining in the brain of poison exon null mice.

Given the *Smndc1* poison exon loss increased SMNDC1 protein abundance and because SMNDC1 is known to interact with many mRNA splicing and processing proteins (Casteels et al., 2022), we next sought to determine if global splicing was impacted by shifting SMNDC1

abundance via poison exon loss. We focused on RNA-seq data from mouse cerebellar transcriptome, given the high levels of protein staining in brain samples and the dramatic increase in protein signal with the loss of the poison exon. Loss of the *Smndc1* poison exon contributed to increased splicing of hundreds of intronic events compared to wild-type transcriptomes, such as increased intronic splicing of *Npb* (**Fig. 3F-G** and **Supplementary Table 4**). Increased intronic splicing is consistent with increased abundance of SMNDC1 protein, given its importance in spliceosome assembly, and also consistent with our previous findings in the *SMNDC1* poison exon null neoplastic cell lines (Thomas et al., 2020). In addition to intron splicing, loss of the *Smndc1* poison exon impacted polyadenylation and 3′UTR processing of many transcripts within the cerebellum (**Supplementary Table 4**). Loss of the *Smndc1* poison exon was associated with 11% of queried transcripts with tandem polyadenylation sites shifting usage toward the distal polyadenylation site, while a similar percentage of transcripts shifted usage toward a more proximal polyadenylation site, such as the 3′UTR of *Galk1* (**Fig. 3F, 3H**). In addition to cerebellum, hundreds of differential splicing events were identified between wild-type and poison exon null samples sourced from both liver and primary fibroblast cell lines (**Supplementary Fig. 4C-D** and **Supplementary Table 5-6**).

2.3.6 Knockout of the Smndc1 poison exon and the SFP30 poison exon results in reduced growth in mouse and A. thaliana models

We next investigated the developmental and morphological phenotypes associated with alteration of *Smndc1* in our mouse models. *Smndc1* is considered an essential gene *in vitro* (Meister et al., 2001), so to serve as a positive control we generated a *Smndc1* gene knockout mouse model through pgRNA-directed CRISPR/Cas9 editing (**Supplementary Fig. 5A**). We

targeted pgRNA disruption to coding exon 4, resulting in a founder mouse with a 5-bp frameshifting deletion allele which was subsequently backcrossed (**Supplementary Fig. 5A-B**). Resulting backcrossed, heterozygous mice expressed less SMNDC1, indicating that the deletion allele may have impeded the translation process of *Smndc1* mRNA transcripts into functional proteins (**Supplementary Fig. 5C**). Furthermore, no homozygous *Smndc1* knockout mice were born ($n = 26$), which is a statistically significant deviation from normal Mendelian inheritance ($p = 0.011$; **Supplementary Fig. 5D**). These data suggest *Smndc1* gene knockout is not developmentally viable.

We next aimed to determine if there was embryonic lethality associated with alteration of non-coding elements of *Smndc1*, specifically in the poison exon deletion model. To test this, we established heterozygous crosses and genotyped all offspring derived from these crosses at weaning (**Fig. 4A**). Genotyped offspring from both the poison exon deletion and control lines adhered to expected Mendelian ratios, suggesting that loss of the *Smndc1* poison exon does not impair viability under heterozygous breeding conditions (**Fig. 4B**). We also found no significant difference in litter sizes or sex ratios of pups born to either the Smndc1 poison exon or control deletion lines (**Supplementary Fig. 6A-B**). This demonstrates that the reproductive success of poison exon heterozygous mice is comparable to that of the control counterparts.

To assess gross morphology of the *Smndc1* poison exon null mice generated from heterozygous crosses, we weighed mice at weaning and over time. At weaning (21 days old), the *Smndc1* poison exon null mice were equivalently sized to wild-type and heterozygous littermates and do not demonstrate gross morphological differences (**Fig. 4C**). Poison exon null mice exhibit a trend towards statistically significant reduced birth rate and reduced size at weaning compared to littermates, but these trends do not reach significance (**Fig. 4B-C**). Aging poison exon null

mice up to 20 months demonstrated no change in overall weight nor mortality rate when compared to wild-type and heterozygous littermates (**Supplementary Fig. 6C-E**). Additionally, we performed complete gross necropsy on male poison exon null, heterozygous, and wild-type littermates (2 replicates per genotype). The following tissues were evaluated by a board-certified staff pathologist: liver, spleen, kidney, heart, gonad, and brain. There was no genotype-dependent change in weight or relative organ size (data not shown). Loss of the poison exon was associated with proliferative and necrotizing polyarteritis in one poison exon null mouse, though this could not be differentiated from idiopathic systemic inflammation (**Supplementary Fig. 7A-E**).

A key component of organism fitness is the ability to reproduce and give rise to fit offspring. Therefore, we next evaluated the reproductive success of the *Smndc1* poison exon null mice by establishing poison exon null breeder pairs (**Fig. 4D**). These breeder pairs were able to reproduce at normal frequency and with typical litter sizes (data not shown). We weighed offspring of poison exon null crosses every week from one week after birth to eight weeks of age. Compared to control offspring, the *Smndc1* poison exon null mice were significantly smaller during postnatal development. We observed 7.58% ($p = 0.037$) and 9.21% ($p = 0.006$) weight reduction over postnatal weeks three and four, respectively, corresponding to the critical developmental time point of weaning (**Fig. 4E**). The postnatal weight reduction was resolved by five weeks of age. This phenotype was notable given the striking weight reduction, its high temporal specificity, and presence solely for homozygous, not heterozygous, crosses.

We next tested whether the *A. thaliana* model exhibited gross morphological phenotypes or other signs of reduced fitness corresponding to the loss of the *SPF30* poison exon. We chose to examine commonly measured life history traits of *A. thaliana*, including longest leaf length at the

onset of flowering, internode length on the primary inflorescence, time to flowering in days, and total seed yield (**Fig. 4F**). Total seed yield and days to flowering were not impacted by loss of the poison exon (**Fig. 4G-H**). In contrast, *A. thaliana* homozygous for the *SPF30* poison exon demonstrated a significant reduction in leaf length and internode length relative to intronic deletion plants and Col-0 plants (**Fig. 4I-J**). We conclude that loss of the *SPF30* poison exon in *A. thaliana* results in growth restriction without impact to reproductive phenotypes.

2.4 DISCUSSION

Regulatory mechanisms like unproductive splicing play a critical role in controlling the abundance of RNA-binding proteins. Yet, most highly conserved splice isoforms in RNA-binding proteins have not been functionally examined. To specifically modulate isoform splicing *in vivo*, we applied CRISPR/Cas9 pgRNA targeting to generate specific and minimally disruptive gDNA edits. Using this approach, this study is the first to directly knockout poison exons in multiple model organisms. This technique offers potential for exploring the functional significance of other poison exons in RNA-binding proteins.

Many splicing factors contain poison exons that are either highly conserved or overlap with ultraconserved DNA elements (UCEs). Of the 481 UCEs present in the human genome, the majority are found within non-coding genomic regions, such as poison exons or enhancers (Bejerano et al., 2004, Snetkova et al., 2022). Numerous deletion studies have targeted ultraconserved enhancers *in vivo*, but only a small number have reported developmental or tissue-specific phenotypes associated with homozygous or hemizygous mice (Ahituv et al., 2007, Dickel et al., 2018, Nolte et al., 2014). Among these deletion models, there are no obvious changes to overall viability or reproduction. It is possible that outside of a controlled, sterile laboratory setting, or across multiple generations, the phenotypic consequence of poison exon loss or mutation would be greater, as the current study does not recapitulate native environmental selective pressures. We suggest that negative selection against detrimental phenotypes may play a role in *Smndc1* and *SPF30* poison exon conservation.

Our study provides clear evidence that highly conserved poison exons are important for organismal fitness. Although this provides clues to the extraordinary conservation that characterizes many poison exons, many questions remain. First, what are the mechanistic origins

of size restriction in *Smndc1* poison exon-null organisms? One may postulate that *Smndc1*-associated missplicing of metabolically relevant genes, such as *Npb* and *Galk1*, could direct further mechanistic studies of factors responsible for the siz-restricted phenotype characterizing our models. Secondly, among highly conserved poison exons, what fraction are important for organism fitness? Does the autoregulatory role of poison exons explain their sequence conservation across genomes? Much work remains to fully understand the roles and evolutionary dynamics of poison exons.

2.5 FIGURES

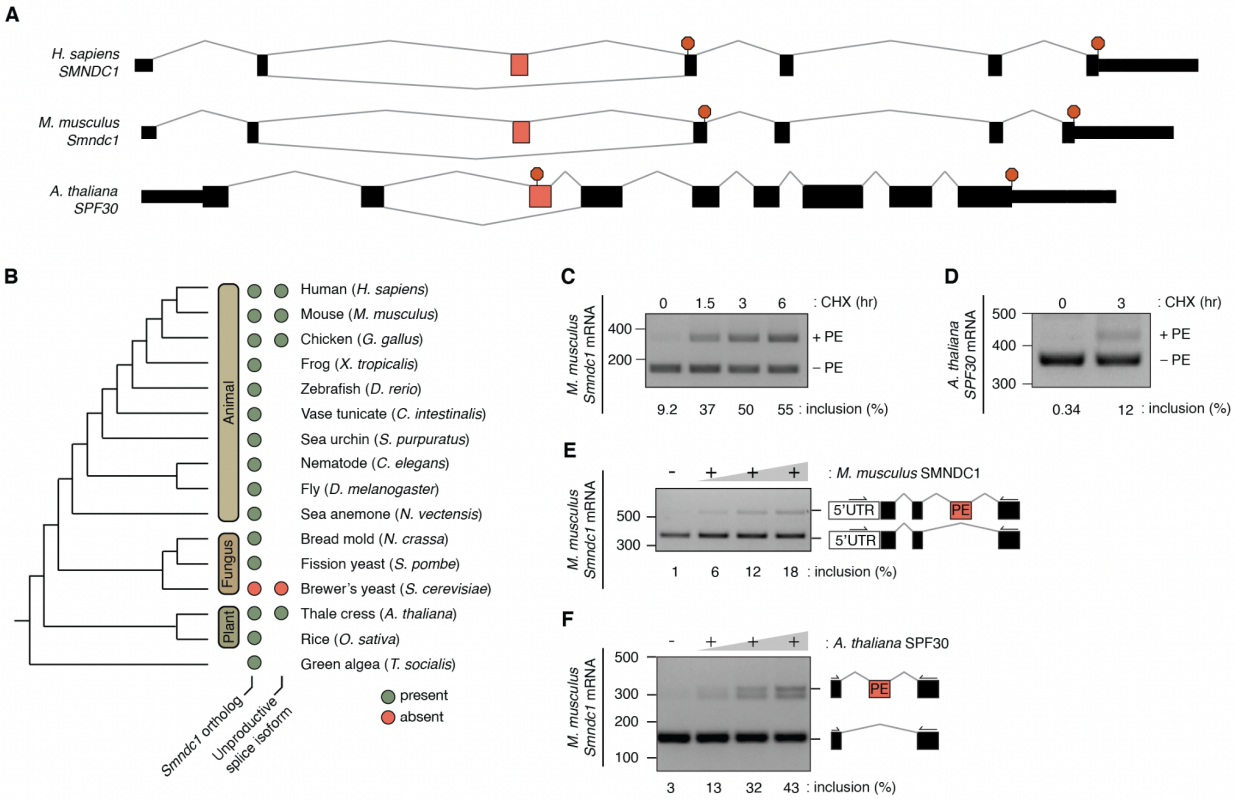

Figure 2.1. A poison exon in *Smndc1* is functionally conserved between mammals and plants.

A, Schematic demonstrating *SMNDC1* and *Smndc1* exon structure in humans and mice, respectively, and *SPF30* exon structure in *A. thaliana*. Black boxes indicate coding exons, and the alternatively spliced poison exon is highlighted in red in each gene structure. Termination codons indicated by red octagons.
B, Unscaled phylogenetic tree of *Smndc1* and *SPF30* indicating species with the presence or absence of identified ortholog (left circles) and unproductive spliced RNA isoform (right circles).
C, RT-PCR analysis of *Smndc1* poison exon (PE) inclusion in mouse Melan-A cells following nonsense-mediated mRNA decay (NMD) inhibition with cycloheximide (CHX) for increasing amounts of time (hours). Inclusion of the poison exon-containing transcript is calculated as a percentage of total transcript abundance.
D, as in **C**) but treatment of *A. thaliana* Col-0 seedlings and analysis of the *SPF30* poison exon inclusion following CHX treatment.
E, RT-PCR analysis of *Smndc1* PE inclusion in mouse cells with increasing overexpression of mouse SMNDC1. PCR was performed with primers within the 5′UTR and third coding exon to capture endogenous transcript.

F, RT-PCR analysis of *Smndc1* PE inclusion in mouse cells with increasing overexpression of *A. thaliana* SPF30. PCR was performed with primers within the second and third coding exons of *Smndc1*.

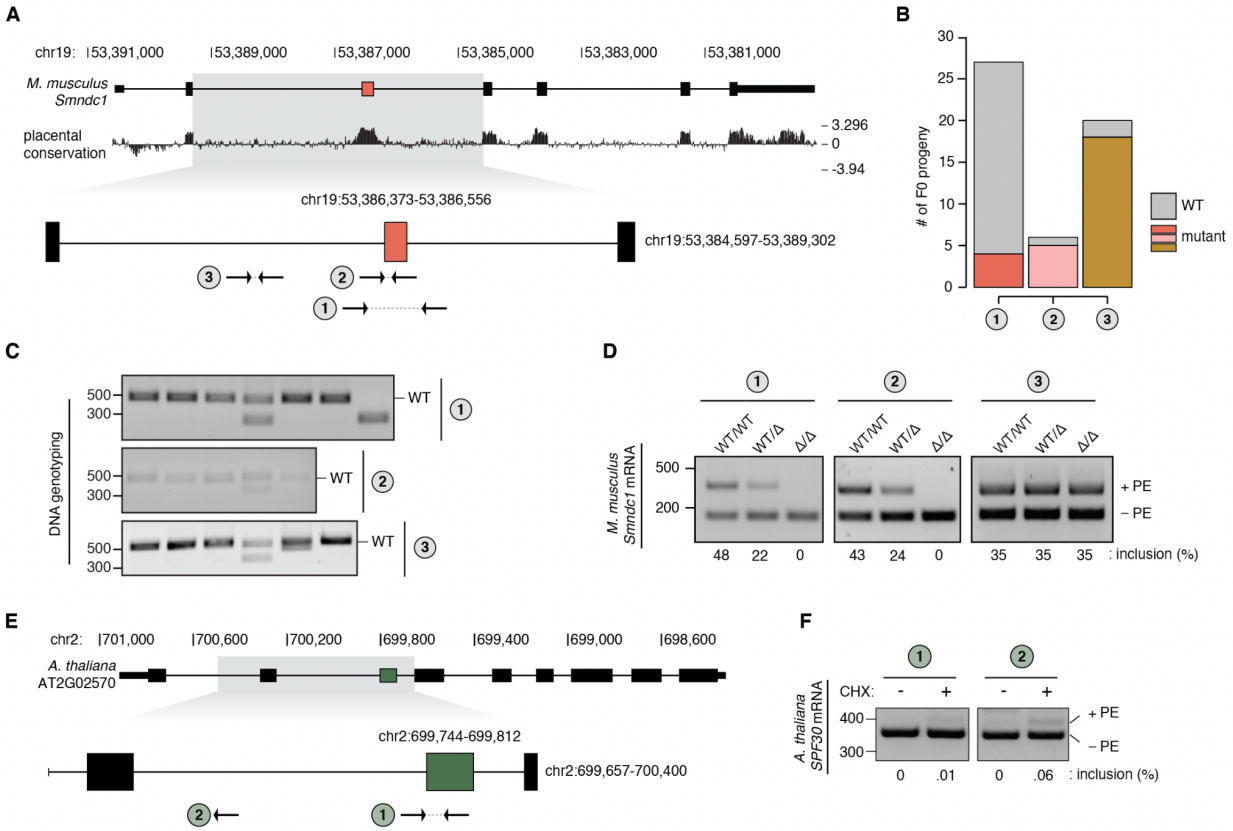

Figure 2.2. Generation of parallel poison exon null models in mouse and *A. thaliana*

A, Paired guide RNAs (pgRNA, arrows) designed to disrupt the *Smndc1* poison exon (red box). pgRNA-directed editing produced three mouse lines with deletion of the poison exon (#1), the poison exon 3′ splice site (#2), or upstream intronic region (#3).
B, Quantification of efficient genomic DNA (gDNA) targeting of founder mice (F0 progeny) from the three pgRNA targeting strategies designated in (**A**). PCR amplicon sequencing was used to determine gDNA targeting.
C, Genotyping gels demonstrating DNA edits in a selection o
f founder mice from the three targeting strategies designated in (**A**). Wild-type (WT) band labeled.
D, RT-PCR analysis of poison exon (PE) inclusion from primary mouse fibroblast lines generated from backcrossed littermates. Littermates were either wild-type (WT/WT), heterozygous (WT/Δ), or homozygous (Δ/Δ) for the respective edited alleles for each of the targeting strategies designated in (**A**). Fibroblasts were treated with cycloheximide (CHX) to inhibit NMD. Inclusion values represent inclusion of the poison exon transcript as a percentage of total transcript quantity.
E, pgRNAs designed to disrupt the *A. thaliana SPF30* poison exon (green box). pgRNA-directed editing produced two *A. thaliana* lines with deletion of the poison exon (#1), or upstream intronic region (#2).
F, RT-PCR analysis of PE inclusion from heterogeneous pooled populations of *A. thaliana* seedlings from both lines designated in **C**. Seedlings were treated with CHX to inhibit NMD. Inclusion values represent inclusion of the poison exon transcript as a percentage of total transcript quantity.

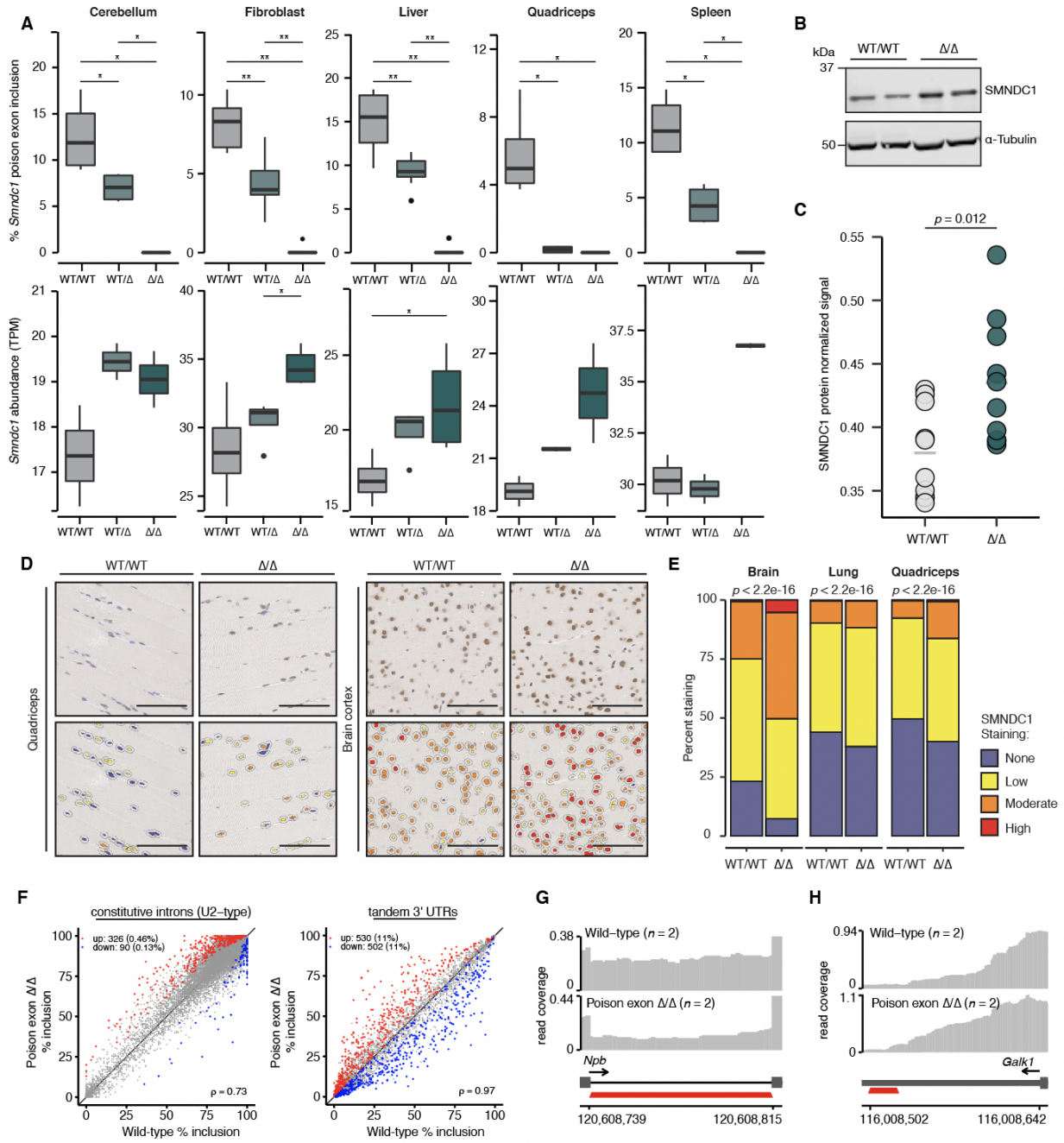

Figure 2.3. Loss of the *Smndc1* poison exon alters global transcript and protein abundance, while steady state RNA splicing is modestly affected.

A, (Top) Quantification of RNA-sequencing (RNA-seq) reads supporting *Smndc1* poison exon inclusion. (Bottom) Total *Smndc1* mRNA transcript abundance (TPM). RNA was collected from mice which were wild-type (WT/WT), heterozygous (WT/Δ), or homozygous (Δ/ Δ) for the *Smndc1* poison exon. P-values computed using two-sided Wilcoxon rank-sum test. $N = 6$ biological replicates per genotype.

B, Representative Western blot of SMNDC1 signal from wild-type (WT/WT) and poison exon null (Δ/Δ) mouse liver lysate. $N = 2$ biological replicates per genotype.
C, Quantification of SMNDC1 signal from Western blots generated with wild-type (WT/WT) and poison exon null (Δ/Δ) mouse liver lysate. Signal normalized to alpha-tubulin. P-value computed using two-sided Wilcoxon rank-sum test. $N = 10$ animals per genotype.
D, (Top) Representative immunohistochemistry images of quadriceps and brain cortex collected from wild-type and *Smndc1* poison exon null littermates, stained with diaminobenzidine (DAB) for SMNDC1. (Bottom) The same representative images with SMNDC1 nuclear staining classified using HALO software as negative (blue), weak (yellow), moderate (orange), or strong (red). Scale bar: 100 μm.
E, Stacked barchart of SMNDC1 staining intensities from tissues using quantification method in (**D**). P-values computed using multinomial proportion test. $N = 3$ mice (poison exon Δ/Δ) and $n = 2$ mice (WT/WT).
F, Scatterplots of constitutive intron splicing and tandem 3′ UTR reads from cerebellum RNA-seq. Comparison between wild-type and *Smndc1* poison exon null mice. Red and blue points represent significantly increased and decreased spliced isoforms in the poison exon null samples, respectively. $N = 2$ biological replicates per genotype.
G, Representative RNA-seq coverage plot of intronic splicing of *Npb* from cerebellum samples. $n = 2$ biological replicates per genotype.
H, as in (**G**) demonstrating a tandem 3′ UTR event in *Galk1*.

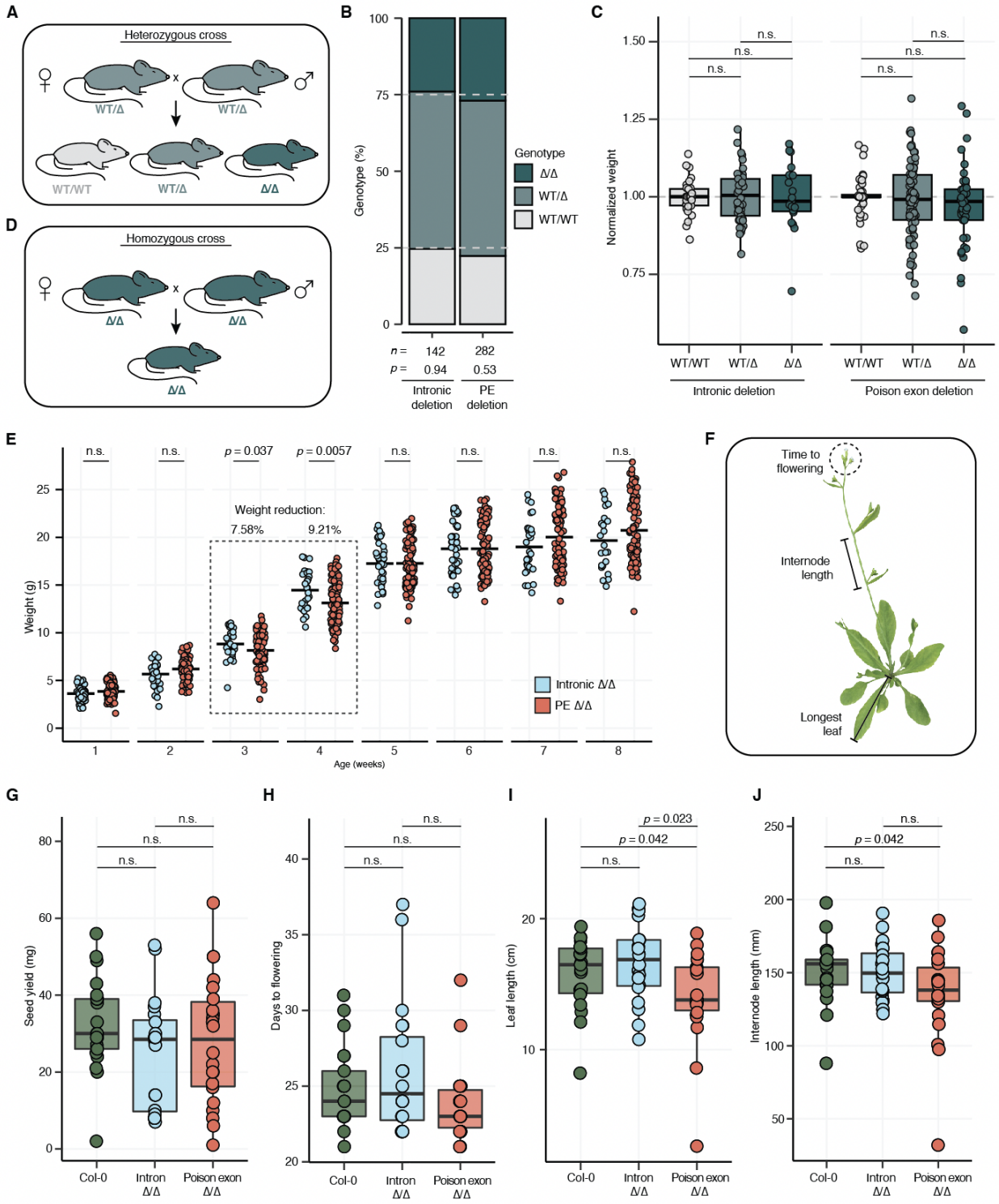

Figure 2.4. Reduced size characterizes the *Smndc1* and *SPF30* poison exon null genetically engineered organisms.

A, Schematic of heterozygous mouse crosses that generated the data in panels (**B**) and (**C**). Possible pup genotypes are wild-type (WT/WT), heterozygous (WT/Δ), or homozygous (Δ/Δ) for *Smndc1* poison exon.
B, Genotype percentages of pups born to heterozygous crosses of intronic deletion and *Smndc1* poison exon deletion lines. P-values generated through Chi-squared testing.
C, Quantification of pup weight at weaning from heterozygous crosses of either the intronic control line (n = 93 mice) or *Smndc1* poison exon deletion lines (n = 172 mice) . Weight normalized to the median weight of all wild-type pups in each respective litter. Statistical significance was assessed by the Wilcoxon-rank sum test.
D, Schematic of homozygous mouse crosses generating the data in panel **C**.
E, Quantification of homozygous pup weight at postnatal weeks 1-8. Weight measured for all pups born to homozygous breeder pairs from both the *Smndc1* intronic deletion and *Smndc1* poison exon deletion lines.
F, Schematic showing phenotypes measured in *A. thaliana* genetically engineered plants.
G-J, Phenotypic assessment of *wild-type* Col-0, intronic deletion, and the *SPF30* poison exon deletion lines. N = 20 biological replicates per line. Statistical significance was assessed by the Wilcoxon-rank sum test. In (**G**), quantification of total seed mass yielded per plant in milligrams. In (**H**), measurement of days to flowering (onset of flowering, 1 centimeter (cm) inflorescence). In (**I**), quantification of maximal plant leaf length in cm at onset of flowering. In (**J**), quantification of internode length in millimeters.

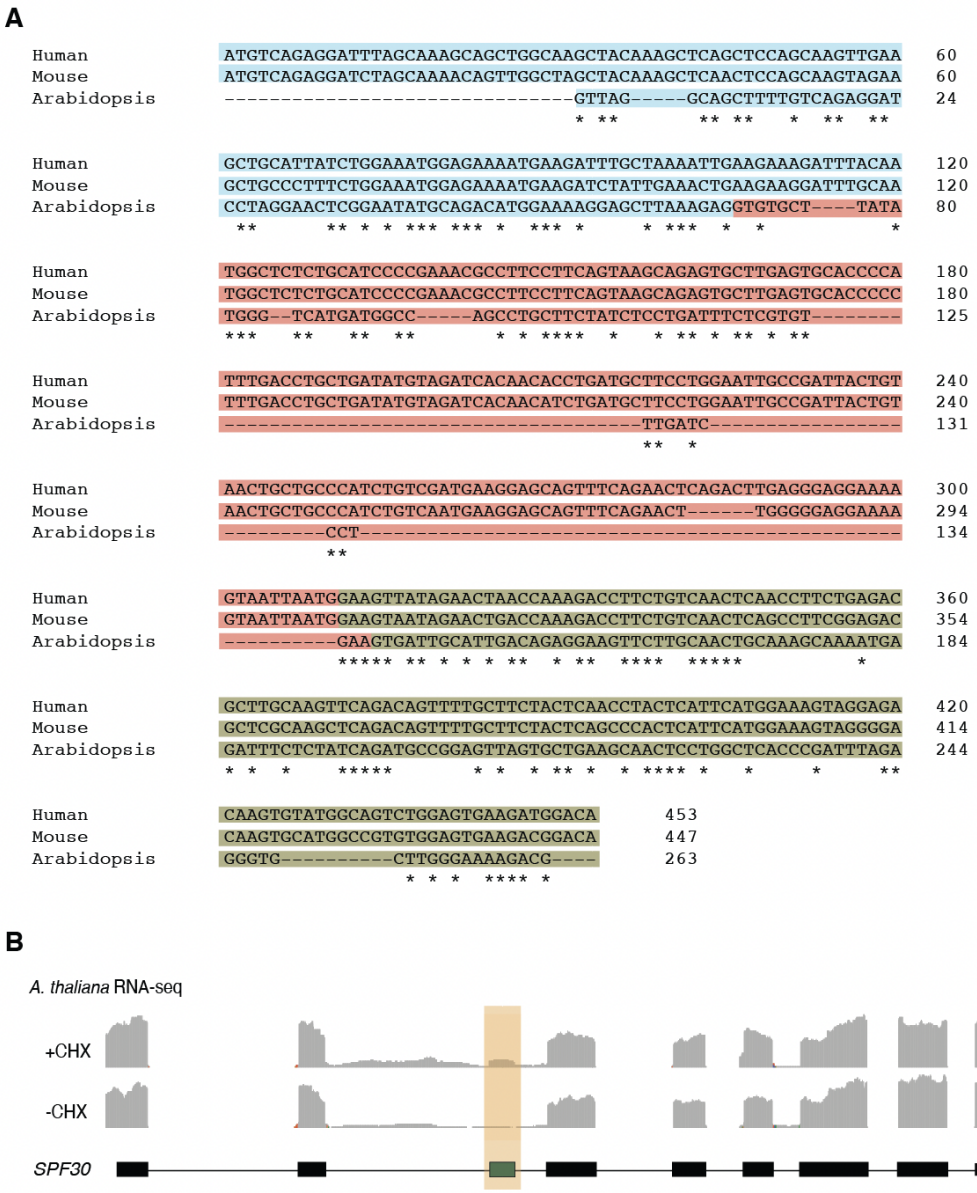

Figure 2.5. Supplementary Figure 1. Conservation of nucleotide sequence and NMD-targeting in the human *SMNDC1*, mouse *Smndc1*, and *A. thaliana SPF30* poison exons.

A, Alignment of human, mouse, and *A. thaliana* DNA sequences from coding and poison exons of *SMNDC1*, *Smndc1* and *SPF30*, respectively. Highlighted are upstream coding exon (blue), poison exon (red), and downstream coding exon (green).
B, BAM coverage plot of *SPF30* gene structure and mapped read coverage in samples treated with or without cycloheximide (CHX). Poison exon supporting reads highlighted in yellow. The poison exon (green box) was previously unannotated.

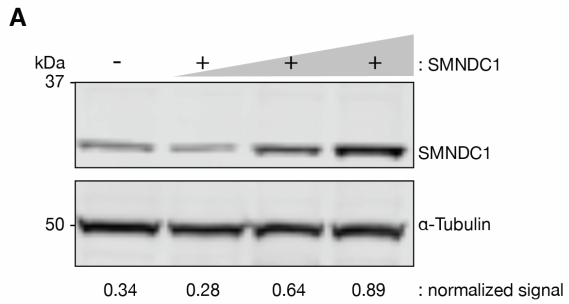

Figure 2.6. Supplementary Figure 2. Transient overexpression of SMNDC1 in mouse melanocytes.

A, Western blot indicated SMNDC1 signal from mouse melanocytes using a transient overexpression system with increasing amounts of SMNDC1 expression. SMNDC1 signal is normalized to alpha-tubulin signal.

.

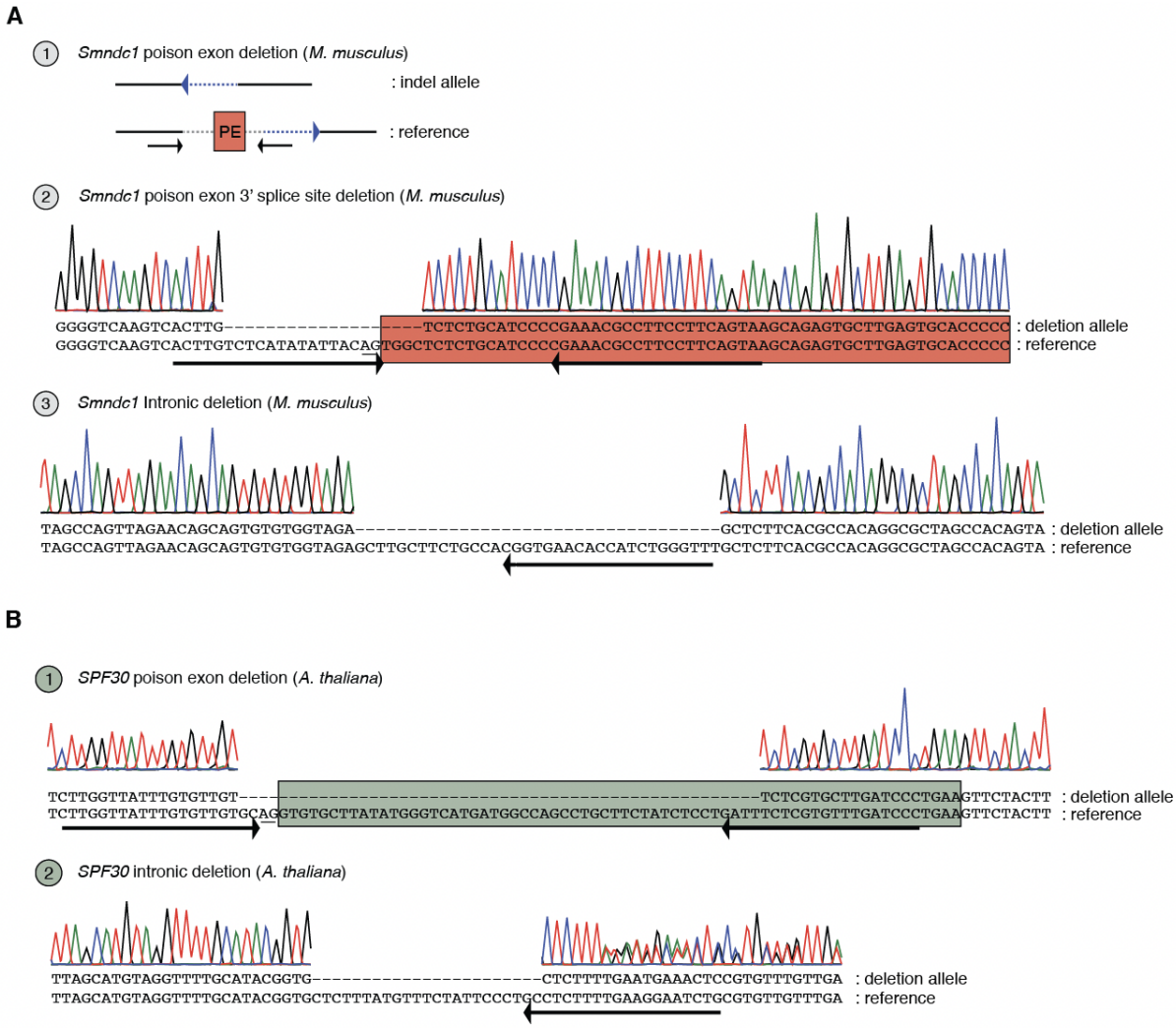

Figure 2.7. Supplementary Figure 3. Validation of *Smndc1* and *SPF30* deletion alleles.

A, Schematic and Sanger sequencing of mouse lines with 254 base pair (bp) deletion of the *Smndc1* poison exon (#1), 19 bp deletion of the *Smndc1* poison exon 3′ splice site (#2), and 35 bp *Smndc1* intronic deletion (#3). Red box indicates the *Smndc1* poison exon (PE) and arrows indicate guide RNA sequences. Blue dotted arrow in deletion allele #1 indicates a 77 bp sequence originally downstream of the PE and inverted in the genetically engineered allele. **B**, Sanger sequence of *A. thaliana* lines with deletion of the *SPF30* poison exon (#1), and *SPF30* intronic deletion (#2). Green box indicates poison exon and arrows indicate guide RNA sequences.

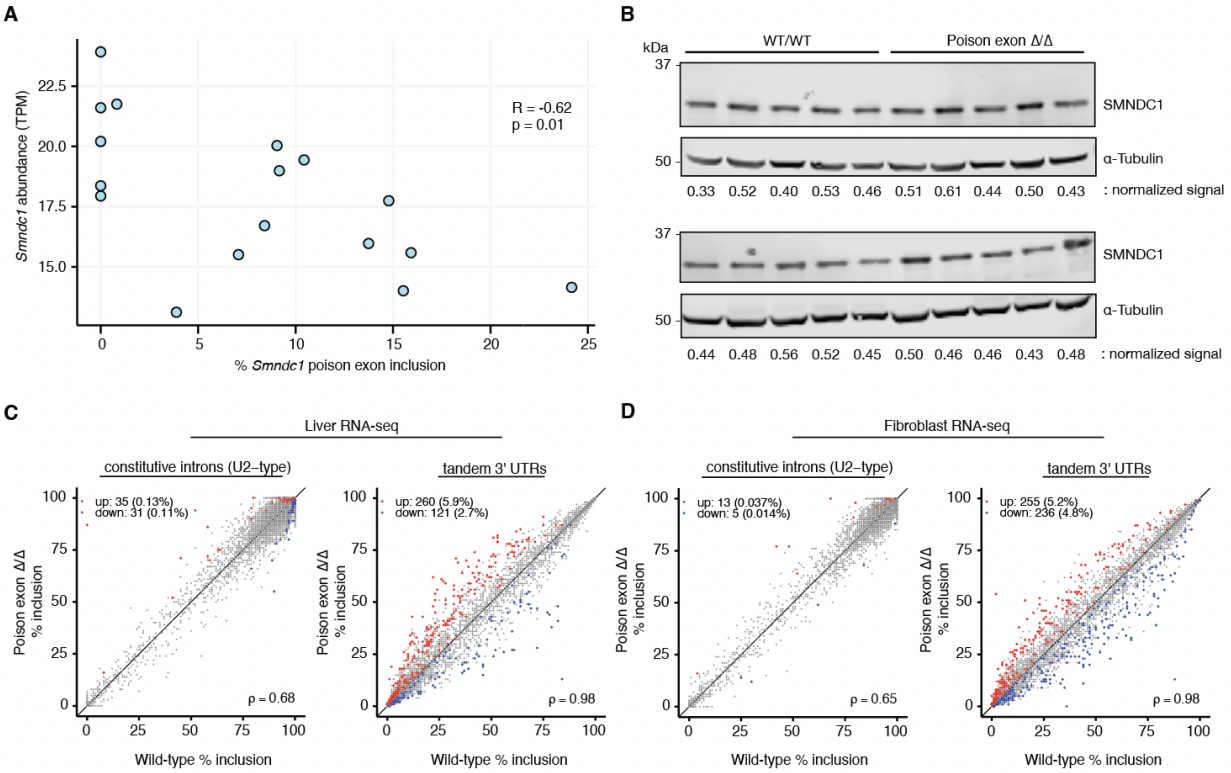

Figure 2.8. Supplementary Figure 4. *Smndc1* poison exon deletion increases transcript and protein abundance and impacts global splicing.

A, Correlation between percent of *Smndc1* poison exon read coverage and total *Smndc1* transcript abundance across liver samples. R, Pearson correlation coefficient. $N = 17$ animals. **B,** Western blot demonstrating SMNDC1 and alpha-tubulin signal and normalized ratio in liver lysate from wild-type (WT/WT) and homozygous (Δ/Δ) mice. $n = 10$ animals per genotype. **C-D**, Scatterplots of constitutive intron splicing (**C**) and tandem 3′ UTR reads (**D**) from RNA-sequencing of liver and fibroblast. Comparison between wild-type and *Smndc1* poison exon homozygous mice. Red and blue dots represent significantly increased and decreased spliced isoforms in the poison exon null samples, respectively. $N = 5$ biological replicates per genotype.

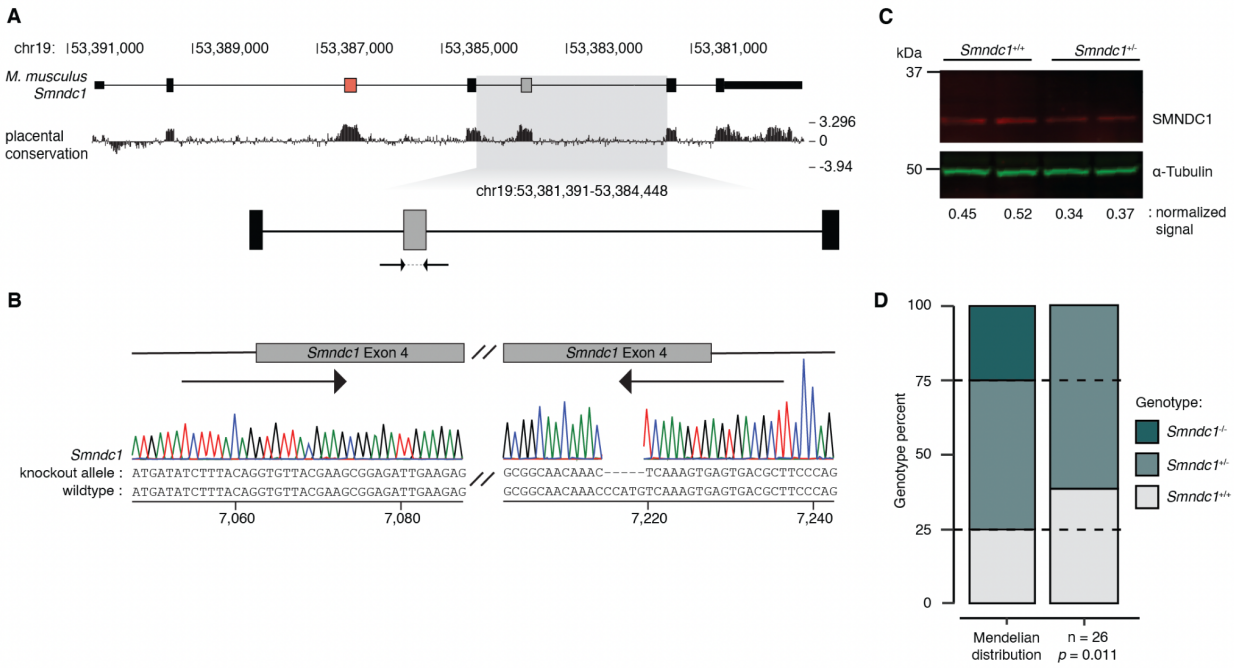

Figure 2.9. Supplementary Figure 5. Generation of *Smndc1* knockout mouse model suggests essentiality of SMNDC1.

A, Paired guide RNAs (pgRNAs, arrows) were designed to disrupt the exon 4 (gray box) of *Smndc1*.
B, Sanger sequencing trace of 5 bp deletion allele at exon 4 (gray box), including pgRNA-targeted sequences (arrows).
C, Representative Western blot of SMNDC1 signal from wild-type ($Smndc1^{+/+}$) and heterozygous ($Smndc1^{+/-}$) mouse liver lysate, normalized to alpha-tubulin signal. $N = 2$ animals per genotype.
D, Mendelian analysis of genotyped tissue from three week old offspring. Dotted lines represent normal genotypic percentages. P-values generated through Chi-squared testing.

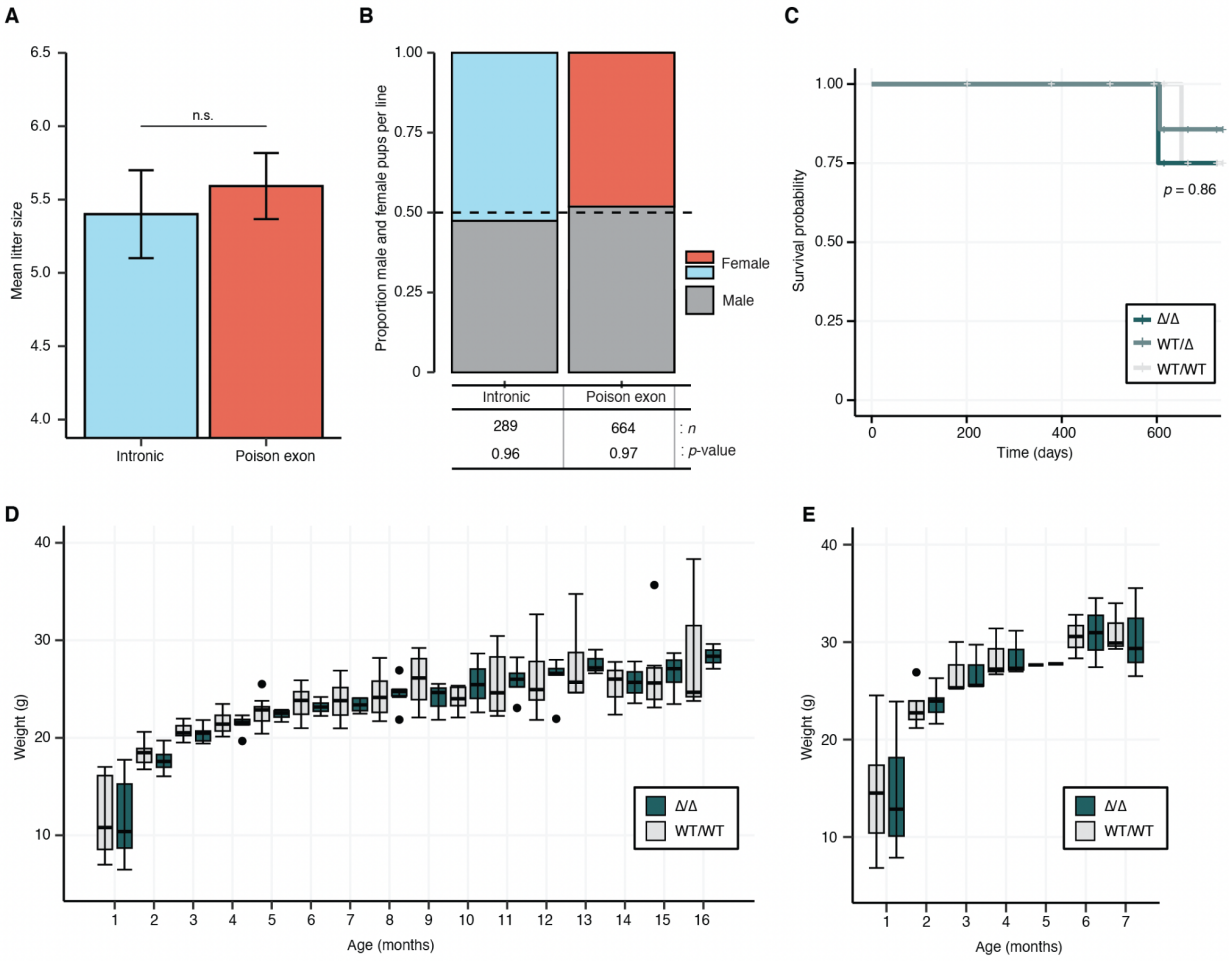

Figure 2.10. Supplementary Figure 6. *Smndc1* poison exon deletion does not impact litter composition, survival, or aging.

A, Quantification of mean litter sizes born to intercrossed and backcrossed mice from the intronic deletion (blue) and combined poison exon and 3′ splice site deletion (red) lines. Error bars represent +/- standard error. *P*-value computed using two-sided Wilcoxon rank-sum; $n = 54$ litters (intronic deletion line) and 116 litters (combined poison exon and 3′ splice site deletion lines).
B, Proportional representation of male and female pup births across deletion lines as in (**A**). *P*-value computed using chi-squared test.
C, Kaplan-Meier survival curve of mice either wild-type (WT/WT), heterozygous (WT/Δ), or homozygous (Δ/Δ) for *Smndc1* poison exon ($n = 30$ total, both male and female mice).
D-E, Weight quantification during aging for female (**D**, $n = 27$) and male mice (**E,** $n = 9$) of either WT/WT or Δ/Δ genotype. Boxplot upper whisker extends to the largest value no further than 1.5 * IQR.

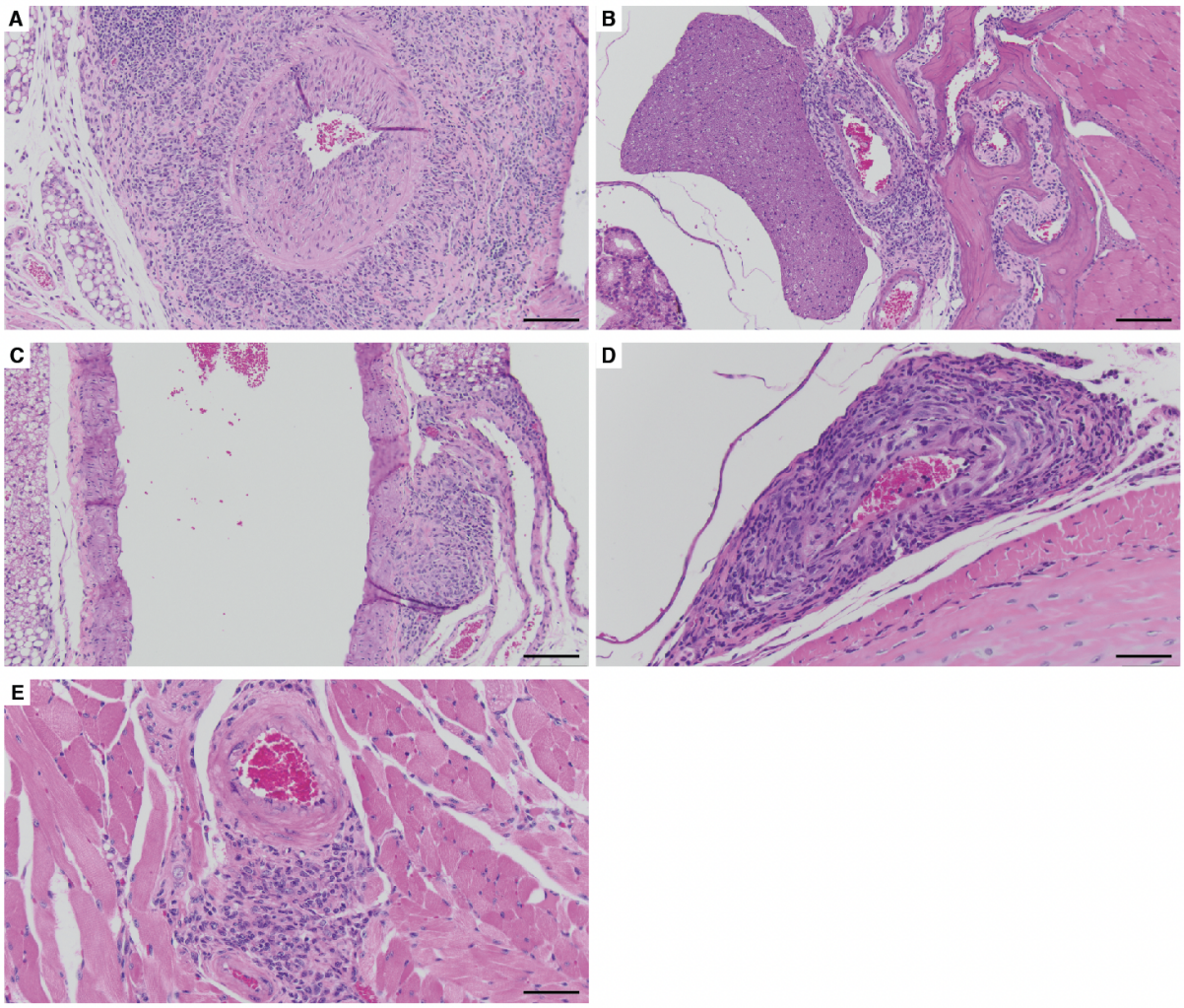

Figure 2.11. Supplementary Figure 7. Proliferative and necrotizing polyarteritis in *Smndc1* poison exon null mouse.

A-E, Hematoxylin and eosin staining of multiple tissues from a 6-month-old, male *Smndc1* poison exon null mouse. Imaging depicts arteritis lesions containing neutrophilic inflammation and necrosis, with more chronic intimal and medial proliferation, fibrosis, and lymphocytic inflammation. Lower-magnification of prostatic artery, artery of skull (presumed temporal artery), and thoracic aorta (**A-C**, respectively); objective lens 20X, scale bar: 100 µm. Higher-magnification of lumbar arteriole and tongue arteriole (**D-E**, respectively); objective lens 40X, scale bar: 50 µm.

2.6 MATERIALS AND METHODS

Identification of orthologs and sequence alignment

To identify protein orthologs of SMNDC1, we employed Basic Local Alignment Search Tool (BLAST) and reciprocal best hit (RBH) analysis. The nucleotide sequence alignment was generated with the Clustal Omega online tool (Madeira et al., 2022). We performed a BLASTP (protein-protein) search using human SMNDC1 and *A. thaliana* SPF30 sequences as queries against the protein database of each target species. The BLASTP search was conducted using the NCBI BLAST+ suite (version 2.14.0+) with default parameters and using the BLOSUM62 substitution matrix. The top hits from the initial BLASTP search were subjected to a reciprocal BLASTP search against the protein database of the query species. Exon DNA sequences were aligned using EMBOSS Needle pairwise sequence alignment and default settings.

Mouse model generation and genotyping

All deletion alleles were generated by CRISPR/Cas9-mediated genome editing with paired guide RNAs (pgRNAs), as previously described (Singh et al., 2015; Thomas et al., 2020). High-quality pgRNA seed sequences (**Supplementary Table 2**) were selected using the Benchling guide design tool, confirmed by GuideScan, and analyzed for specificity using CasOffinder. pgRNAs were produced through cloning-free *in vitro* transcription as previously described (Varshney et al., 2015) using the MEGAshortscript T7 Transcription Kit (Life Technologies). Products were purified using the MEGAclear Transcription Clean-up Kit (Life Technologies). Cas9 mRNA (50 ng/µL) and pgRNA (25 ng/µL) were microinjected into single-cell zygotes derived from B6(Cg)-Tyrc-2J/J inbred mice (JAX strain #000664). Microinjected embryos were transplanted into pseudopregnant CD-1 recipients. Founder animals were genotyped using ear punch tissue with

primers that flanked the expected deletion loci by 100 to 200 bases (**Supplementary Table 7**). The resulting amplicons were further characterized by Sanger sequencing and Amplicon-EZ Next-Generation Sequencing (Azenta Life Sciences). All founder animals were backcrossed to C57BL/6 breeders (JAX strain #000664) for at least four generations.

To genotype founder mice and subsequent generations, ear tissue was collected when the mice were two weeks old and digested in Lysis M Buffer (Macherey-Nagel) and Proteinase K (Qiagen) for five minutes at room temperature. This was followed by standard PCR using appropriate primers and gel electrophoresis (**Supplementary Table 7**). PCR products were subsequently Sanger sequenced when alleles were indistinguishable through gel electrophoresis (**Supplementary Table 7**). After the establishment of backcrossed lines, genotyping was performed using quantitative genomic PCR (Transnetyx).

A. thaliana model generation

Seeds from CRISPR/Cas9-generated genetically engineered T1 plants were provided by Creative Biogene (Shirley, NY, USA), after careful selection of gRNA with optimal off-target scores (**Supplementary Table 2**). Due to local sequence constraints, only one gRNA was used to target the upstream intronic region. Genetically engineered plants contain targeting gRNA sequences driven by the Arabidopsis U6-26 promoter, an EC1.1-promoted Cas9 protein with a nuclear localization signal, and a bar gene selectable marker (basta resistance gene). To screen plants, we collected genomic DNA from leaf samples using the cetyltrimethyl ammonium bromide (CTAB) method, performed standard PCR with appropriate primers (**Supplementary Table 7**), and followed this with gel electrophoresis and Sanger sequencing.

Isolation of mouse fibroblasts

Isolation of primary fibroblasts was performed from mouse ear and tail tissue as previously outlined (Khan & Gasser, 2016), with the following modifications. We euthanized eight-week-old wild-type, heterozygous, and homozygous littermates, collected ear and tail tissues, and immediately carried out digestion in a solution of 15 U/mL collagenase and 0.13% dispase. Following digestion and filtration, the cells were cultured in Ham's F10 medium with 10% FBS, 1% penicillin, and 1% streptomycin. Once the fibroblasts reached confluency, they were dissociated and split at a ratio of 1:5.

RT-PCR

To detect poison exon-containing RNA isoforms in cell lines and *A. thaliana* seedlings, we used cycloheximide (CHX) to inhibit NMD prior to RNA extraction. Cell lines were treated with medium containing 100 μg ml−1 CHX for three hours before collection in TRIzol reagent (Invitrogen). *A. thaliana* seedlings (24 days) were removed from the soil, rinsed in buffer (0.046 g/L Murashige and Skoog Plant Salt Mixture, 0.3 g/L sucrose, pH 5.8), and vacuum-infiltrated with 10 μg ml−1 CHX in buffer for 10 minutes, as previously described (Kurihara et al., 2009). After vacuum infiltration, seedlings were incubated in the CHX buffer for three additional hours at room temperature prior to flash-freezing and immediately homogenization in TRIzol reagent (Invitrogen).

Total RNA was extracted using the Direct-zol RNA MiniPrep (Zymo Research). cDNA was synthesized using SuperScript IV Reverse Transcriptase (Thermo Scientific) according to the manufacturer's protocol. RT–PCR was performed with Phire Green Hot Start II DNA

polymerase (Thermo Scientific) and gene-specific primers (**Supplementary Table 7**). Amplicons were analyzed and quantified using agarose gel electrophoresis. Band intensity was then quantified using FIJI/ImageJ.

Mouse aging and necropsy

Mice were kept under standard housing conditions. To evaluate the impact of aging, we tracked their body weight monthly. The mice were gently restrained and weighed on a digital scale with a precision of 0.01 grams. This weight data was utilized to monitor age-related changes and assess general health throughout the study period.

For necropsy, 6-month-old male mice were euthanized with CO_2 to retrieve organs (6 replicates total). Organs were washed with deionized water before fixation in 4% paraformaldehyde. The tissues were processed routinely, and sections were stained with hematoxylin and eosin. The specimens were interpreted by a board-certified staff pathologist, in a blinded fashion.

Immunohistochemistry and HALO analysis

Tissue embedding and staining was performed by the Experimental Histopathology Core at Fred Hutchinson Cancer Center. Mouse SMNDC1 was detected using a mouse polyclonal antibody (Thermo Fisher PA5-31148) at 1:8000 dilution following optimization by the Experimental Histopathology core facility. Staining was performed using a BOND RX autostainer (Leica Biosystems). Imaging was performed using an Aperio ImageScope (Leica Biosystems), and image analysis was completed using HALO Image Analysis software.

Western blotting

Total protein lysates were prepared in RIPA buffer (Cell Signaling) with 1mM Pefabloc (Sigma-Aldrich) and quantified using the Pierce 660 nm Protein Assay Reagent (Thermo Fisher). Protein lysates were electrophoretically separated and transferred to nitrocellulose membrane using NuPAGE systems (Thermo Fisher). Membranes were blocked with Odyssey Blocking Buffer (LI-COR Biosciences) for one hour at room temperature. Primary antibody incubation was performed overnight at 4°C in Intercept Antibody Diluent (LI-COR Biosciences), with SMNDC1 (Thermo Fisher PA5-31148, 1:1000) and alpha-tubulin (Sigma-Aldrich T6199, 1:1000) primary antibodies. IRDye (LI-COR Biosciences) secondary antibodies were used for detection and imaging using the Odyssey CLx Imager (LI-COR Biosciences). Protein blot analysis and quantification were performed using FIJI/ImageJ.

Protein overexpression in B16-F10 cells

Mouse *Smndc1* coding sequence was expressed in the PCMV6-Kan/Neo backbone (OriGene). To generate a construct containing *A. thaliana SPF30* coding sequence, mouse *Smndc1* cDNA sequence was dropped out with EcoRI and MluI digestion and *A. thaliana* cDNA fragment was inserted using Gibson assembly. The constructs were transiently transfected in serum-free media with Lipofectamine 3000, using 0, 0.25, 1, or 2.5 μg per 400,000 B16-F10 cells in a 6 well plate. RNA was collected for RT-PCR 24 hr post-transfection.

RNA-seq library preparation and data analysis

Fibroblast RNA was extracted using Direct-zol RNA MiniPrep kit (Zymo Research), according to the manufacturer's instructions. For tissue RNA extractions, tissue was homogenized in TRIzol reagent with zirconium beads prior to RNA extraction, using the Zymo Direct-zol RNA Miniprep kit. A minimum of 500 ng of high-quality RNA was used as input for library

preparation. Poly(A)-selected unstranded libraries were prepared using a TruSeq protocol and sequenced using an Illumina sequencer to obtain 100 bp paired-end reads.

RNA-seq reads were processed as previously described (Dvinge et al., 2014). In brief, FASTQ files were trimmed to remove sequencing adapters and then aligned to the hg19/GRCh37 reference assembly. This assembly was created by merging the UCSC knownGene gene annotation (Meyer et al., 2012), Ensembl gene annotation (Flicek et al., 2012), and MISO isoform annotation (Katz et al., 2010). Read alignment and expression estimates were generated with RSEM (RNA-seq by expectation maximization; Li & Dewey, 2011), Bowtie (Langmead et al., 2009), and TopHat (Trapnell et al., 2009), and isoform ratios were quantified with MISO v.2.0 (Katz et al., 2010). Aligned reads from mapping steps were merged to generate BAM coverage plots. Significantly differentially spliced events met the following criteria: at least 20 identifying reads in each sample, a minimum Bayes factor of 5, and a minimum of 10% change (absolute scale) in isoform ratio or minimum fold-change of 2 (log2 scale) in absolute isoform ratio.

As the mouse *Smndc1* poison exon was not annotated in the described RNA-seq mapping method, RNA-seq reads were manually mapped to the *Smndc1* poison exon inclusion and exclusion isoforms using Bowtie. No mismatched nucleotides were allowed, junction-spanning reads were filtered to require a minimum overhang of 6 nucleotides, and resulting counts were normalized by the length of the mapping index.

All analyses and visualization were performed in the R programming environment, with tools from Bioconductor and dplyr, ggplot2, tidyverse packages.

Plant growth and phenotyping

For all experiments, control and edited plant lines were grown under long day conditions (16 hour light/8 hour dark), with fluorescent bulbs at 70–80 μmol m−2 s−1. The room was maintained at 22°C and 50% relative humidity. The wild-type Col-0 seeds were collected, stored, and planted at the same time as the seeds of the transgenic lines. All genotypes were grown in a random block design and plant trays were regularly rotated to account for location-dependent growth effects. Longest leaves were measured as the longest leaf at time of bolting (onset of flowering, defined by the presence of a 1 cm primary inflorescence). Internode length was measured from the primary inflorescence. Seed yield was measured as the total mass of all seeds collected from a fully dried plant.

Animal use

Animal husbandry was conducted in accordance with the Guidelines for the Care and Use of Laboratory Animals and was approved by the Institutional Animal Care and Use Committees at Fred Hutchinson Cancer Center.

2.7 TABLES

Table 2.1. Supplementary Table 1: Reciprocal best hit analysis across species.

Species	BLASTP query	E-Value	Percent identity	Protein identifier	Reciprocal BLASTP to query
H. sapiens	NA	NA	NA	SMNDC1	NA
M. musculus	*H. sapiens*	6.00E-177	99.58%	Q8BGT7	TRUE
R. norvegicus	*H. sapiens*	2.00E-176	99.16%	NP_001020571.1	TRUE
C. familiaris	*H. sapiens*	3.00E-178	99.58%	E2QYH2	TRUE
G. gallus	*H. sapiens*	3.00E-150	95.80%	E1BQK7	TRUE
X. tropicalis	*H. sapiens*	6.00E-139	88.66%	Q6DEY1	TRUE
D. rerio	*H. sapiens*	1.00E-129	78.99%	Q7ZV80	TRUE
T. rubripes	*H. sapiens*	4.00E-125	74.37%	XP_003964002	TRUE
C. intestinalis	*H. sapiens*	6.00E-69	48.95%	XP_002129900	TRUE
S. purpuratus	*H. sapiens*	3.00E-49	51.28%	XP_030830516	TRUE
C. elegans	*H. sapiens*	4.00E-21	37.50%	NP_001022932	TRUE
D. melanogaster	*H. sapiens*	2.00E-39	41.77%	NP_001138001	TRUE
N. vectensis	*H. sapiens*	2.00E-52	43.55%	XP_032236631	TRUE
N. crassa	*A. thaliana*	2.00E-05	40.98%	NCU04794	TRUE
S. pombe	*A. thaliana*	0.015	40.32%	NP_588166	TRUE
A. thaliana	NA	NA	NA	Q84K41_ARATH	NA
O. sativa	*A. thaliana*	2.00E-131	63.25%	XP_015650899	TRUE
T. socialis	*A. thaliana*	6.00E-33	34.19%	PNH12932.1	TRUE

Table 2.2. Supplementary Table 2: Sequences of pgRNAs used in model generation.

gene	target	pgrna_id	species	grna1	grna1 PAM	grna2	grna2 PAM
Smndc1	poison exon	pgSmndc1.1	mouse	GAATAAACTGGTTCTAAAGG	NGG	GAACTACTTTACGTATGAGG	NGG
Smndc1	poison exon	pgSmndc1.2	mouse	ACTTGTCTCATATATTACAG	NGG	TACTGAAGGAAGGCGTTTCG	NGG
Smndc1	intron 2	pgSmndc1.3	mouse	CAAAAACGCAAATCAGGTGG	NGG	AACCCAGATGGTGTTCACCG	NGG
Smndc1	exon 4	pgSmndc1.4	mouse	TCTTTACAGGTGTTACGAAG	NGG	AGCGTCACTCACTTTGACAT	NGG
SFP30	poison exon	pgSPF30.1	arabidopsis	CTTGGTTATTTGTGTTGTGC	NGG	GGGATCAAACACGAGAAATC	NGG
SFP30	exon 2	pgSPF30.2	arabidopsis	GTCAGAGGATCCTAGGAACT	NGG	TCTGCATATTCCGAGTTCCT	NGG
SFP30	intron 2	pgSPF30.3	arabidopsis	CAGATTCCTTCAAAAGAGGC	NGG	-	-

Table 2.3. Supplementary Table 3: Smndc1 expression across mouse models and tissue samples

mouse_line	mouse_genotype	mouse_id	sample_origin	Smndc1 expression (TPM)
3' splice site	homozygous	210115_17	fibroblast	33.458561
3' splice site	heterozygous	210115_18	fibroblast	28.119879
3' splice site	wild-type	210115_19	fibroblast	24.421763
poison exon	wild-type	210115_30	fibroblast	27.856688
poison exon	homozygous	210115_31	fibroblast	34.440598
poison exon	heterozygous	210115_33	fibroblast	32.370823
3' splice site	homozygous	210115_17	liver	21.757213
3' splice site	heterozygous	210115_18	liver	18.990131
3' splice site	wild-type	210115_19	liver	17.745878
poison exon	wild-type	210115_30	liver	15.972025
poison exon	homozygous	210115_31	liver	23.924761
poison exon	heterozygous	210115_33	liver	19.439494
poison exon	homozygous	210910_1	fibroblast	37.082112
poison exon	homozygous	210910_2	fibroblast	36.353529
poison exon	wild-type	210910_5	fibroblast	26.293468
poison exon	wild-type	210921_30	fibroblast	36.844394
poison exon	wild-type	210921_32	fibroblast	36.884562
poison exon	homozygous	210910_1	liver	21.605446
poison exon	homozygous	210910_2	liver	20.208177
poison exon	wild-type	210910_5	liver	13.116939
poison exon	wild-type	210921_30	liver	14.149304
poison exon	wild-type	210921_32	liver	15.507439
poison exon	wild-type	210503_1	cerebellum	17.809182
poison exon	heterozygous	210503_2	cerebellum	19.247741
poison exon	homozygous	210503_3	cerebellum	18.990067
poison exon	homozygous	210503_6	cerebellum	17.894433
poison exon	wild-type	210503_7	cerebellum	15.569345
poison exon	heterozygous	210503_8	cerebellum	18.326943
poison exon	wild-type	210503_1	liver	14.007783
poison exon	heterozygous	210503_2	liver	16.711969
poison exon	homozygous	210503_3	liver	17.944239
poison exon	homozygous	210503_6	liver	18.366804
poison exon	wild-type	210503_7	liver	15.588578

poison exon	heterozygous	210503_8	liver	20.036803
poison exon	wild-type	210503_1	quad	20.433079
poison exon	heterozygous	210503_2	quad	22.174108
poison exon	homozygous	210503_3	quad	28.227602
poison exon	homozygous	210503_6	quad	22.694817
poison exon	wild-type	210503_7	quad	19.085531
poison exon	heterozygous	210503_8	quad	21.951409
poison exon	wild-type	210503_1	fibroblast	31.656086
poison exon	heterozygous	210503_2	fibroblast	31.769339
poison exon	homozygous	210503_3	fibroblast	35.671046
poison exon	homozygous	210503_6	fibroblast	36.232673
poison exon	wild-type	210503_7	fibroblast	29.020742
poison exon	heterozygous	210503_8	fibroblast	31.522736
poison exon	wild-type	210503_1	spleen	33.354154
poison exon	heterozygous	210503_2	spleen	30.776749
poison exon	homozygous	210503_3	spleen	36.243246
poison exon	homozygous	210503_6	spleen	37.443884
poison exon	wild-type	210503_7	spleen	30.788902
poison exon	heterozygous	210503_8	spleen	31.116246

Table 2.4. Supplementary Table 7. Primers used for Sanger sequencing, gDNA PCR, and RT-PCR.

primer_F_id	primer_F_seq	primer_R_id	primer_R_seq	use
RKB3931	AGCTCAACTCCAGCAAGTAGAAGC	RKB3933	GAAGCAAAACTGTCTGAGCTTGCG	RT-PCR; mouse *Smndc1* poison exon
RKB4441	GCGACAAAGAACCATCGGAG	RKB4442	CGTCTTTTCCCAAGCACCCT	RT-PCR; *A.thaliana SPF30* poison exon
RKB4453	GTAGGTCTCGCGAGATCTCG	RKB4454	CCGAAGGCTGAGTTGACAGA	RT-PCR; mouse *Smndc1* poison exon, 5' UTR priming
Sm1Poex1CRIS-F	TTTGACTTAGGAAGGAAAGTAGGT	Sm1Poex1CRIS-R	CCAGAGAATTTTCTCCTGATG	gDNA PCR; mouse *Smndc1* poison exon deletion
Sm1spCRIS-F	TCTTTTAAAGTCCTAACTGCTTGt	Sm1spCRIS-R	TACGTATGAGGGGGTTCTT	gDNA PCR; mouse *Smndc1* poison exon 3' splice site deletion
Sm1ContCRIS-F	GTAGACAGTTTTTACCTAAAAGGCTTA	Sm1ContCRIS-R	TGCACTGTCCAATATGTAATCAA	gDNA PCR; mouse *Smndc1* intronic deletion
RKB4475	GGCTTTTGATACCAGGTGCC	RKB4476	ACAGTCTCCACTTTACCTAGGC	gDNA PCR; mouse *Smndc1* exon 4 deletion
RKB4472	TGTGGATTCGTTTTATTAGTCTCCC	RKB4473	CAACTGCTTTTGGCTGACCC	gDNA PCR; *A.thaliana SPF30* poison exon deletion
RKB4474	AGGTGAGTATGTTTTTGAGTGGC	RKB4432	GCAGTTGCAAGAACTTCCTCT	gDNA PCR; *A.thaliana SPF30* intronic deletion

2.8 Data Availability

RNA-seq data generated as part of this study were deposited at the Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) under accession number PRJNA1027874. RNA-seq data generated by CHX treatment of *A. thaliana* (Drechsel et al., 2013) were accessed through GEO accession number GSE41432.

2.9 Acknowledgements

We thank C. Queitsch and J. Cuperus for feedback and advice on *A. thaliana* related studies. We thank Creative Biogene for providing *A. thaliana* models. We thank A. Koehne for assistance with mouse necropsy and histopathology. 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, Preclinical Modeling, and Genomics Shared Resources of the Fred Hutch/University of Washington Cancer Consortium (NIH/NCI P30 CA015704).

2.10 REFERENCES

Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L. A., & Rubin, E. M. (2007). Deletion of Ultraconserved Elements Yields Viable Mice. *PLoS Biology*, *5*(9), e234. https://doi.org/10.1371/journal.pbio.0050234

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved Elements in the Human Genome. *Science*, *304*(5675), 1321–1325. https://doi.org/10.1126/science.1098119

Carvill, G. L., & Mefford, H. C. (2020). Poison exons in neurodevelopment and disease. *Current Opinion in Genetics & Development*, *65*, 98–102. https://doi.org/10.1016/j.gde.2020.05.030

Casteels, T., Bajew, S., Reiniš, J., Enders, L., Schuster, M., Fontaine, F., Müller, A. C., Wagner, B. K., Bock, C., & Kubicek, S. (2022). SMNDC1 links chromatin remodeling and splicing to regulate pancreatic hormone expression. Cell Reports, 40(9), 111288. https://doi.org/10.1016/j.celrep.2022.111288

Côté, J., & Richard, S. (2005). Tudor Domains Bind Symmetrical Dimethylated Arginines. Journal of Biological Chemistry, 280(31), 28476–28483. https://doi.org/10.1074/jbc.M414328200

de Oliveira Freitas Machado, C., Schafranek, M., Brüggemann, M., Hernández Cañás, M. C., Keller, M., Di Liddo, A., Brezski, A., Blümel, N., Arnold, B., Bremm, A., Wittig, I., Jaé, N., McNicoll, F., Dimmeler, S., Zarnack, K., & Müller-McNicoll, M. (2023). Poison cassette exon splicing of *SRSF6* regulates nuclear speckle dispersal and the response to hypoxia. *Nucleic Acids Research*, *51*(2), 870–890. https://doi.org/10.1093/nar/gkac1225

Dickel, D. E. et al. Ultraconserved enhancers are required for normal development. Cell 172, 491–499.e15 (2018).

Drechsel, G., Kahles, A., Kesarwani, A. K., Stauffer, E., Behr, J., Drewe, P., Rätsch, G., & Wachter, A. (2013). Nonsense-Mediated Decay of Alternative Precursor mRNA Splicing Variants Is a Major Determinant of the *Arabidopsis* Steady State Transcriptome. *The Plant Cell*, *25*(10), 3726–3742. https://doi.org/10.1105/tpc.113.115485

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*, *111*(47), 16802–16807. https://doi.org/10.1073/pnas.1413374111

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. (2012). Ensembl 2013. *Nucleic Acids Research*, *41*(D1), D48–D55. https://doi.org/10.1093/nar/gks1236

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

Khan, M., & Gasser, S. (2016). Generating Primary Fibroblast Cultures from Mouse Ear and Tail Tissues. *Journal of Visualized Experiments*, *107*. https://doi.org/10.3791/53565

Kurihara, Y., Matsui, A., Hanada, K., Kawashima, M., Ishida, J., Morosawa, T., Tanaka, M., Kaminuma, E., Mochizuki, Y., Matsushima, A., Toyoda, T., Shinozaki, K., & Seki, M. (2009). Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, *106*(7), 2453–2458. https://doi.org/10.1073/pnas.0808902106

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), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lareau, L. F., & Brenner, S. E. (2015). Regulation of Splicing Factors by Alternative Splicing and NMD Is Conserved between Kingdoms Yet Evolutionarily Flexible. *Molecular Biology and Evolution*, *32*(4), 1072–1079. https://doi.org/10.1093/molbev/msv002

Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., & Brenner, S. E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, *446*(7138), 926–929. https://doi.org/10.1038/nature05676

Leclair, N. K., Brugiolo, M., Urbanski, L., Lawson, S. C., Thakar, K., Yurieva, M., George, J., Hinson, J. T., Cheng, A., Graveley, B. R., & Anczuków, O. (2020). Poison Exon Splicing Regulates a Coordinated Network of SR Protein Expression during Differentiation and Tumorigenesis. *Molecular Cell*, *80*(4), 648-665.e9. https://doi.org/10.1016/j.molcel.2020.10.019

Lewis, B. P., Green, R. E., & Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences*, *100*(1), 189–192. https://doi.org/10.1073/pnas.0136770100

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 323. https://doi.org/10.1186/1471-2105-12-323

Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*, *50*(W1), W276–W279. https://doi.org/10.1093/nar/gkac240

Meister, G., Hannus, S., Plöttner, O., Baars, T., Hartmann, E., Fakan, S., Laggerbauer, B., & Fischer, U. (2001). SMNrp is an essential pre-mRNA splicing factor required for the formation of the mature spliceosome. *The EMBO Journal*, *20*(9), 2304–2314. https://doi.org/10.1093/emboj/20.9.2304

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. (2012). The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Research*, *41*(D1), D64–D69. https://doi.org/10.1093/nar/gks1048

Mier, P., & Pérez-Pulido, A. J. (2012). Fungal Smn and Spf30 homologues are mainly present in filamentous fungi and genomes with many introns: Implications for spinal muscular atrophy. *Gene*, *491*(2), 135–141. https://doi.org/10.1016/j.gene.2011.10.006

Ni, J. Z., Grate, L., Donohue, J. P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T. A., Blume, J. E., & Ares, M. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes & Development*, *21*(6), 708–718. https://doi.org/10.1101/gad.1525507

Nolte, M. J., Wang, Y., Deng, J. M., Swinton, P. G., Wei, C., Guindani, M., Schwartz, R. J., & Behringer, R. R. (2014). Functional analysis of limb transcriptional enhancers in the mouse. *Evolution & Development*, *16*(4), 207–223. https://doi.org/10.1111/ede.12084

Neumann, A., Meinke, S., Goldammer, G., Strauch, M., Schubert, D., Timmermann, B., Heyd, F., & Preußner, M. (2020). Alternative splicing coupled mRNA decay shapes the temperature-dependent transcriptome. *EMBO Reports*, *21*(12). https://doi.org/10.15252/embr.202051369

Parfrey, L.W., Lahr, D.J., Knoll, A.H., & Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences, 108*(33), 13624–136249. https://doi.org/10.1073/pnas.1110633108

Rappsilber, J., Ajuh, P., Lamond, A. I., & Mann, M. (2001). SPF30 Is an Essential Human Splicing Factor Required for Assembly of the U4/U5/U6 Tri-small Nuclear Ribonucleoprotein into the Spliceosome. *Journal of Biological Chemistry*, *276*(33), 31142–31150. https://doi.org/10.1074/jbc.M103620200

Rayson, S., Arciga-Reyes, L., Wootton, L., De Torres Zabala, M., Truman, W., Graham, N., Grant, M., & Davies, B. (2012). A Role for Nonsense-Mediated mRNA Decay in Plants: Pathogen Responses Are Induced in *Arabidopsis thaliana* NMD Mutants. *PLoS ONE*, *7*(2), e31917. https://doi.org/10.1371/journal.pone.0031917

Saltzman, A. L., Kim, Y. K., Pan, Q., Fagnani, M. M., Maquat, L. E., & Blencowe, B. J. (2008). Regulation of Multiple Core Spliceosomal Proteins by Alternative Splicing-Coupled Nonsense-Mediated mRNA Decay. *Molecular and Cellular Biology*, *28*(13), 4320–4330. https://doi.org/10.1128/MCB.00361-08

Singh, P., Schimenti, J. C., & Bolcun-Filas, E. (2015). A Mouse Geneticist's Practical Guide to CRISPR Applications. *Genetics*, *199*(1), 1–15. https://doi.org/10.1534/genetics.114.169771

Snetkova, V., Pennacchio, L. A., Visel, A., & Dickel, D. E. (2022). Perfect and imperfect views of ultraconserved sequences. *Nature Reviews Genetics*, *23*(3), 182–194. https://doi.org/10.1038/s41576-021-00424-x

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

Tripsianes, K., Madl, T., Machyna, M., Fessas, D., Englbrecht, C., Fischer, U., Neugebauer, K. M., & Sattler, M. (2011). Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins. *Nature Structural & Molecular Biology*, *18*(12), 1414–1420. https://doi.org/10.1038/nsmb.2185

Varshney, G. K., Pei, W., LaFave, M. C., Idol, J., Xu, L., Gallardo, V., Carrington, B., Bishop, K., Jones, M., Li, M., Harper, U., Huang, S. C., Prakash, A., Chen, W., Sood, R., Ledin, J., & Burgess, S. M. (2015). High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Research*, *25*(7), 1030–1042. https://doi.org/10.1101/gr.186379.114

Wojciechowicz, T., Billert, M., Jasaszwili, M., Strowski, M. Z., Nowak, K. W., & Skrzypski, M. (2021). The Role of Neuropeptide B and Its Receptors in Controlling Appetite, Metabolism,

and Energy Homeostasis. *International Journal of Molecular Sciences*, *22*(12), 6632. https://doi.org/10.3390/ijms22126632

Zhang, D., Yang, J. F., Gao, B., Liu, T. Y., Hao, G. F., Yang, G. F., Fu, L. J., Chen, M. X., & Zhang, J. (2019). Identification, evolution and alternative splicing profile analysis of the splicing factor 30 (SPF30) in plant species. *Planta*. https://doi.org/10.1007/s00425-019-03146-x

Zhang, X., Chen, M. H., Wu, X., Kodani, A., Fan, J., Doan, R., Ozawa, M., Ma, J., Yoshida, N., Reiter, J. F., Black, D. L., Kharchenko, P. V., Sharp, P. A., & Walsh, C. A. (2016). Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. *Cell*, *166*(5), 1147-1162.e15. https://doi.org/10.1016/j.cell.2016.07.025

Chapter 3. ATG7 ALTERNATIVE POLY(A) SITE SELECTION ALTERS MELANOMA CELL GROWTH *IN VITRO* AND *IN VIVO*

This chapter includes work referenced in the following publication as well as unpublished results:

Multiplexed screening reveals how cancer-specific alternative polyadenylation shapes tumor growth in vivo. Gabel, A. M**., Belleville, A. E.**, Thomas, J. D., McKellar, S. A., Nicholas, T. R., Banjo, T., Crosse, E. I., & Bradley, R. K. (2024). *Nature Communications*, *15*(1), 959. https://doi.org/10.1038/s41467-024-44931-x

Please note that Austin Gabel is the primary author of the referenced publication above. Text within this Chapter reflects work published by Austin Gabel, including modifications from me. This Chapter includes work that I performed with Austin Gabel to support the project and my contributions to manuscript revisions. Many experiments described in this chapter were designed by Austin Gabel, Robert Bradley, and me, and performed by Austin Gabel and me. I have noted instances of work performed by Austin Gabel, for which I cannot take credit. I am grateful for the opportunity to collaborate on this project.

3.1 SUMMARY

Alternative polyadenylation (APA) is a complex biological process that is dysregulated in many cancers. It has previously not been clear whether differential poly(A) site selection is causative or merely correlative to tumorigenesis. Recent studies focused on the endogenous role of poly(A) site selection in cancer systems indicate that many APA events do not impact tumorigenesis, while certain key APA events can control mouse melanoma in immunocompetent hosts. We investigated one such APA event in the critical autophagy gene *Atg7*. Forced use of the distal *Atg7* poly(A) site lengthens the 3′UTR, destabilizes mRNA, and reduces total ATG7 protein levels. Consequently, mouse melanoma cells with forced use of the *Atg7* distal poly(A) site

demonstrate reduced proliferation, improving host survival. This result can be rescued with increased ATG7 protein, suggesting a direct proliferative phenotype tied to ATG7 levels in an autophagy-independent manner.

3.2 INTRODUCTION

RNA processing is a complex and dynamic means to modulate gene expression. Alternative cleavage and polyadenylation (APA) is one step in RNA processing whereby a pre-mRNA transcript is cleaved and a poly-adenosine monophosphate, or poly(A), tail is added. The addition of the poly(A) tail can occur at more than one distinct location, ultimately impacting the length of the 3' untranslated region (UTR). The alternative use of distinct poly(A) sites impacts about 70% of human genes, and many sites are shared between mammalian species (Derti et al., 2012). Poly(A) site selection is frequently impacted by upstream intronic splicing (Rigo and Martinson, 2008, Movassat et al., 2016), and is coordinated by many RNA-binding proteins and polyadenylation machinery (Tian and Manley, 2016; Mitschka and Mayr, 2021). The functional significance of most APA events is unknown, but mechanistic studies of specific APA events demonstrate that APA can regulate mRNA abundance (Lianoglou et al., 2013), translation (Spies et al., 2013), and cellular localization (Berkovits and Mayr, 2015). APA is context-dependent—for example as mouse myogenesis progresses, mouse genes tend to express mRNAs with longer 3' UTRs (Ji et al., 2009). Rapidly dividing cells preference use of the proximal poly(A) and have shorter 3' UTRs (Sandberg et al., 2008).

In cancer cells, APA is frequently dysregulated. Similar to our understanding in healthy cells, the functional relevance of many cancer-associated APA events is not fully understood. It has

previously been unclear if APA dysregulation is a causative factor in cancer progression or if APA dysregulation is a correlative byproduct of other cancer phenotypes, such as rapid cell division. Recent work by our lab has begun to address the causative impact of APA in tumorigenesis using a CRISPR/Cas9 paired guide RNA (pgRNA) murine system. By comparing the APA landscape in a syngeneic context—mouse melanoma (B16-F10 cells) and melanocytes (MelanA cells)—we identified cancer-specific polyadenylated events in mouse melanoma (Gabel et al., 2024). From this comparison, we targeted the gene-proximal poly(A) sites of 143 genes, 25% of which are also dysregulated in human TCGA melanoma samples. The targeted deletion of proximal poly(A) sites is a well-established method to force usage of the distal poly(A) site (Bae et al., 2020, Pereira-Castro et al., 2022). We then systematically quantified how APA events contribute to tumor growth *in vivo* and found that many APA events have modest or no effects on tumor growth kinetics. However, specific APA events play critical functions in either promoting or inhibiting tumorigenesis (Gabel et al., 2024). For example, pgRNAs targeting the proximal poly(A) site of *Atg7* were strongly depleted *in vitro* and *in vivo*, suggesting distal polyadenylation of *Atg7* decreases tumorigenesis.

Autophagy related protein 7, ATG7, is a critical protein for autophagosome formation, a key cellular maintenance and survival mechanism. Autophagosomes traffic cytosolic components to lysosomes for degradation, and ATG7 plays a key role in conjugating and loading phagosome components. Correspondingly, the loss of ATG7 protein inhibits autophagy (Komatsu et al., 2005). Autophagy is frequently reported to be dysregulated in cancers, with either tumor-promoting or tumor-suppressive effects depending on the tumor setting, stage, or context (Levy and Thorburn, 2015; Galluzzi et al., 2015,). However, in a pancreatic ductal adenocarcinoma model (PDAC), *Atg7* hemizygous loss reduced tumor progression, metastasis,

and cellular invasion in an autophagy-independent manner (Long et al., 2022). This model corresponds with our initial screen data targeting *Atg7* proximal poly(A) site, and we aimed to validate this phenotype and understand the mechanistic implications of *Atg7* APA in melanoma.

3.3 RESULTS

3.3.1 A long Atg7 3'UTR suppresses murine melanoma growth in vitro and in vivo

We first sought to assess the mRNA stability of the *Atg7* short and long 3′UTR isoforms by an Actinomycin D time course. The percent of *Atg7* mRNA transcripts utilizing the distal poly(A) site decreased to a greater extent than those utilizing the proximal poly(A) site in the setting of Actinomycin D (**Figure 3.1 A-B**). This trend suggests the long *Atg7* 3' UTR transcript is less stable than the shorter *Atg7* 3' UTR transcript. This may in part explain the reduced protein levels observed in B16-F10 *Atg7* proximal poly(A) site knock-out (pKO) cells (**Figure 3.2 B**).

In the setting of a previous CRISPR/Cas9 screen, alteration to *Atg7* 3' UTR was demonstrated to lead to reduced growth of murine melanoma *in vivo* (Gabel et al., 2024). To validate that alteration of 3' UTR polyadenylation impacts *in vivo* growth, we treated B16-F10 cells with either pgRNAs targeting the proximal poly(A) site of *Atg7* (pKO) or a control pgRNA targeting a poly(A) site of an unexpressed gene. Targeting the proximal poly(A) site is shown to promote usage of the distal poly(A) site and therefore lengthen transcript 3' UTR length (Gabel et al., 2024). The generated cell lines were engrafted subcutaneously into C57BL/6 mice and survival was measured. B16-F10 *Atg7* pKO tumors exhibited reduced growth and consequent significantly prolonged survival when compared to control pKO tumors ($p = 0.0053$, **Figure 3.1C**).

To determine how the use of the *Atg7* distal poly(A) site may lead to reduced cellular proliferation and consequently improved survival, we performed immunohistochemistry (IHC) for the proliferation marker Ki67. When analyzing *Atg7* pKO tumors, we noted a significantly higher proportion of nuclei with moderate and high levels of Ki67 when compared to control tumors (**Figure 3.1D-E**). The higher proportion of Ki67 may initially seem at odds with slower tumor growth in *Atg7* pKO tumors. However, since Ki67 levels fluctuate through the cell cycle, we performed cell cycle analysis with propidium iodide staining. This analysis demonstrated that *Atg7* pKO cells spend an increased fraction of time in S phase and G2/M phase, relative to control B16-F10 cells (**Figure 3.1F**). This is consistent with tumor IHC and previous work with *ATG7* KD in human bladder cancer cells (Zhu et al., 2017).

The occurrence of APA in *Atg7* across mouse and human cells points to the potential that the *ATG7* 3′ UTR length could be a factor in human melanoma growth, mirroring its effect in mouse melanoma. After binning the 424 patients in TCGA melanoma cohort into terciles based on *ATG7* 3' UTR length, we observed a correlation between long *ATG7* 3' UTR length and improved progression-free survival (**Figure 3.1G**). This correlation suggests a possible mechanism by which melanoma may regulate ATG7 protein levels to influence growth.

3.3.2 Restoration of ATG7 protein rescues the Atg7 pKO growth phenotype

To bolster these conclusions regarding the functional relevance of Atg7 APA isoforms, we performed ectopic ATG7 rescue experiments. We generated two *Atg7* cDNA constructs with identical *Atg7* coding sequences followed downstream by either 1) a SV40 poly(A) signal or 2) the endogenous *Atg7* proximal poly(A) signal (**Figure 3.2A**). Constructs were stably transduced into *Atg7* pKO cell line which harbors the specific deletion of the proximal poly(A) site. We

validated ATG7 protein expression from both constructs was 3.47 and 2.74 fold higher than *Atg7* pKO protein expression (in the *Atg7* SV40 and *Atg7* short 3′UTR constructs, respectively; **Figure 3.2B**). Both constructs modestly increased ATG7 levels relative to endogenous ATG7.

Next, we tested cell growth *in vitro* using a Cell Titer-Glo proliferation assay. We found that expression with either the *Atg7* SV40 or *Atg7* short 3' UTR constructs rescued the cell proliferation phenotype of *Atg7* pKO cells (**Figure 3.2C**). While not yet completely validated from the endogenous locus, these data are consistent with the idea that ATG7 protein levels drive the proliferation phenotypes present in *Atg7* pKO cells. These experiments allowed us to demonstrate the restoration of ATG7 protein and consequent rescue of reduced growth phenotype in *Atg7* pKO cells lacking the proximal poly(A) site.

3.3.3 Reduced metastatic potential of Atg7 pKO cells

Enlightened by human melanoma cohort data in which *Atg7* long 3' UTR usage is associated with increased overall survival (**Figure 3.1G**), we performed many experiments using a murine melanoma model (B16-F10 cells) and focused on proliferation phenotypes. Our mouse tumor experiments focused on tumor growth at the site of engraftment and did not consider tumor metastasis. However, patient survival data is frequently tied to tumor dissemination, and not primary tumor growth *per se* (Steeg, 2006). To test migratory potential of our cell model, cells were grown to confluence, a thin scratch was introduced, and cell regrowth into the scratched area was assayed. This experiment demonstrated reduced migratory potential in B16-F10 *Atg7* pKO cells relative to B16-F10 *Crabp1* pKO (control) cells (**Figure 3A-B**). These data reflect a reproducible and statistically significant reduction in *Atg7* pKO migration at all time points assayed.

3.4 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. Through our study of *Atg7* pKO melanoma cells, we reveal a causative link between *Atg7* APA and reduced proliferative and migratory capacity, which could influence metastatic capacity. Mechanistically, forced use of the distal *Atg7* poly(A) site destabilizes the mRNA transcript and reduces ATG7 protein levels. The proliferative phenotype may be rescued by increasing ATG7 protein levels, and interestingly this phenotype seems to be autophagy-independent in line with other recent reports in PDAC (data not shown; Long et al., 2022). It is possible that reducing ATG7 protein by roughly 60% impacts growth phenotypes but is not sufficient to impact ATG7's role in autophagy.

It should be emphasized that our study was centered on understanding the role of APA modulation in the proliferative capacity of tumor cells. Cancer cell proliferation is one of several vital factors that affect the prognosis of cancer patients. Our study, however, did not include a full evaluation of metastatic potential, which is commonly associated with overall patient survival (Steeg 2006). Consequently, a natural progression of this study would be to investigate how APA specific to cancer influences the spread of tumor cells. Indeed, recent work in metastatic breast cancer established that the downregulation of heterogenous nuclear ribonucleoprotein C encoding the RNA-binding protein HNRNPC in highly metastatic cells was associated with increased distal poly(A) site usage in many downstream genes (Navickas et al., 2023). Correspondingly, increasing HNRNPC levels reduced metastatic capacity, likely through rescuing normal APA signature of downstream genes.

While our study of *Atg7* pKO utilized a polyclonal pool of cancer cells and utilized bulk sequencing techniques, future studies of APA in complex cancer microenvironments may take advantage of increasingly available scRNA-seq datasets and techniques. The development of 3'-tagged scRNA-seq protocols and computational tools will greatly increase the number of publicly available datasets that can be used to quantify gene and 3′UTR isoform expression from specific cell types or developmental stages (Li et al., 2021; Agarwal et al., 2021, Kim et al., 2019, Shulman and Elkon, 2019). Tumor or immune-specific APA signatures may be prognostic markers or potentially therapeutic targets, as our understanding of APA-related cancer phenotypes develops.

3.5 FIGURES

Please note that Figure 3.1 below is the work of Austin Gabel, included in the publication Gabel et al., *Nature Communications* (2024). I have reduced the Figures to those elements that I collaborated on or those that provide context to this Chapter. I generated and analyzed the data represented in Figures 3.2 and 3.3 with input from Austin Gabel.

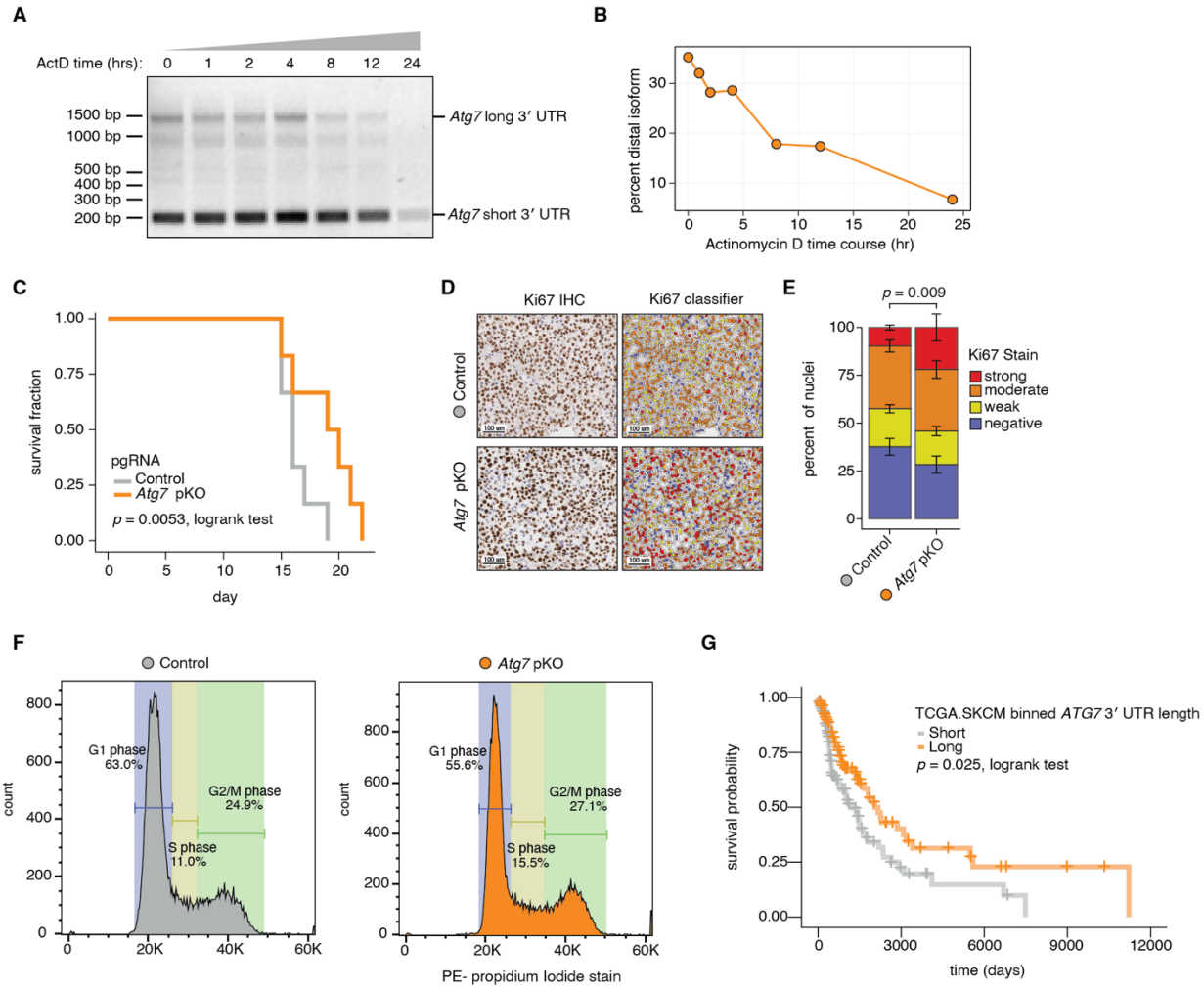

Figure 3.1. Atg7 alternative poly(A) site selection alters melanoma cell growth *in vitro* and *in vivo*.

A) Nested RT-PCR of the *Atg7* 3′ UTR in Cas9-expressing B16-F10 cell lines treated with 1 ug/mL Actinomycin D to inhibit transcription. Expected *Atg7* 3′ UTR size when utilizing the proximal poly(A) site (short) or the distal poly(A) site (long) are indicated. Representative gel from n = 2 biological replicates.
B) Quantification of the percent distal isoform from (A). Gel intensities quantified using FIJI.

C) Survival data from a cohort of C57BL/6 mice injected with Cas9-expressing B16-F10 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).
D) Left, representative immunohistochemistry images of control or *Atg7* pKO pgRNA tumor sections stained for Ki67. Right, images with nuclei classified using HALO as negative (blue), weak (yellow), moderate (orange) or strong (red) staining.
E) Stacked bar plot quantifying data from (D). 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. *P* value calculated with a two-sided binomial proportion test.
F) Representative flow cytometry histograms of Cas9-expressing B16-F10 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.
G) Kaplan-Meier analysis of progression-free survival in TCGA cutaneous melanoma cohort for patients binned into terciles based on *ATG7* 3′ UTR length. *P* value from a two-sided log-rank test.

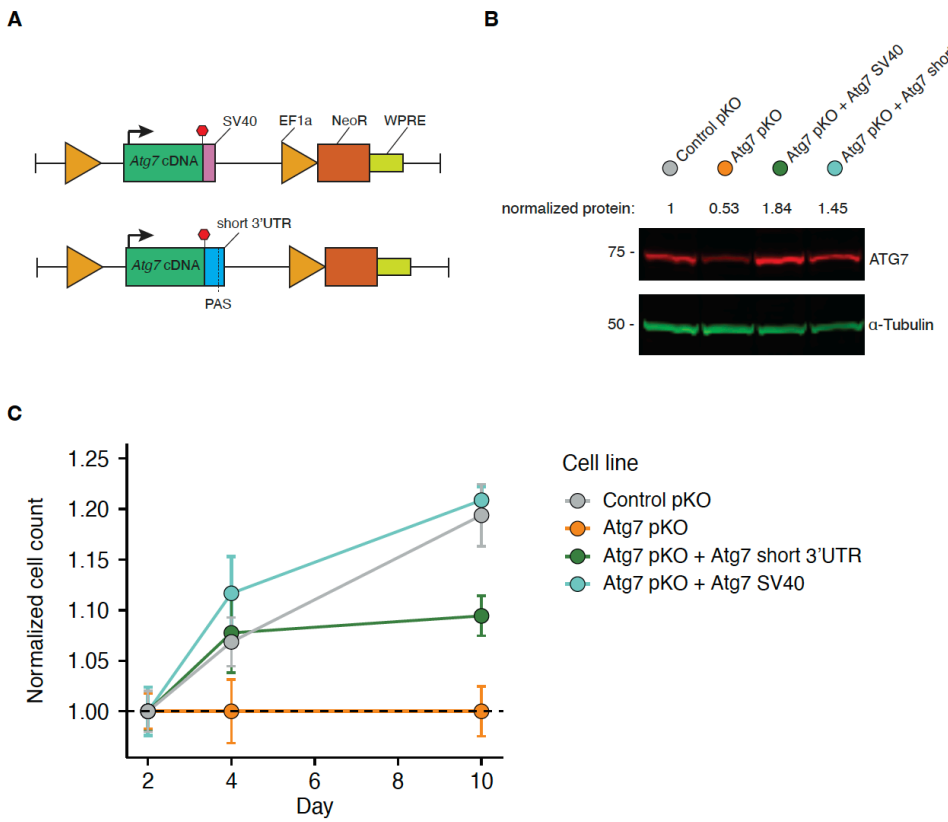

Figure 3.2. Restoration of ATG7 protein rescues *Atg7* pKO growth phenotype

A) Schematic of two *Atg7* cDNA constructs harboring identical coding sequences, but distinct 3′ UTRs, with either an SV40 poly(A) signal or the endogenous 3′ UTR and the proximal poly(A) site.
B) Immunoblot of protein collected from Cas9-expressing B16-F10 cells treated with either a control pgRNA or *Atg7* pKO pgRNA with stable expression of the indicated *Atg7* cDNA construct. Protein ratio normalized to a-Tubulin concentration and then to the control pKO protein ratio.
C) *In vitro* cell growth of Cas9-expressing B16-F10 cells treated with a control pgRNA or *Atg7* pKO pgRNA with the indicated cDNA constructs as measured by CellTiter-Glo. Measurement is the average of three replicates +/- standard error of the mean.

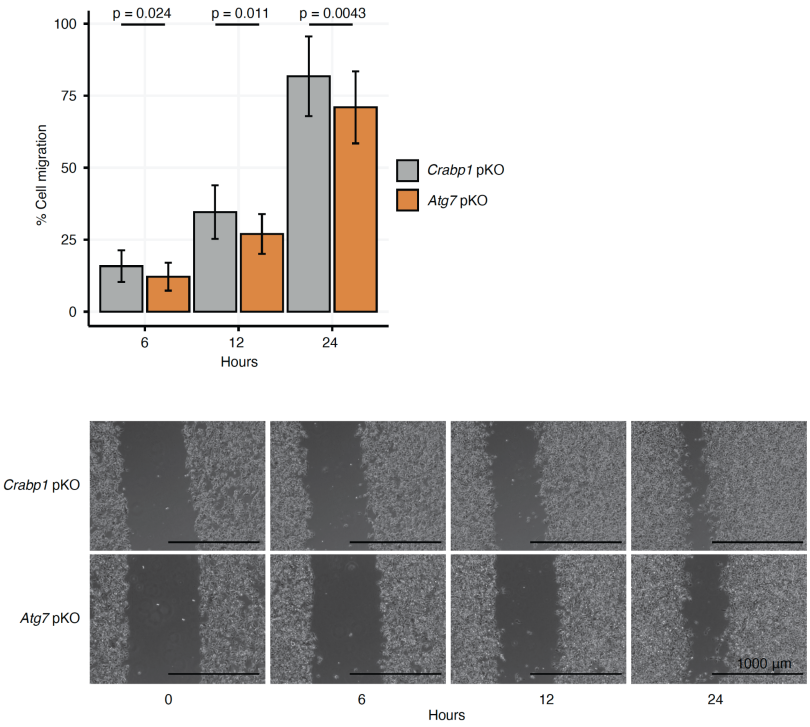

Figure 3.3. Reduced metastatic potential of *Atg7* pKO cells

A) Percent cell migration of B16-F10 *Crabp1* pKO or *Atg7* pKO cells at multiple time timepoints. Scratch width values are normalized to scratch width at 0 hr and percent migration is calculated as a percent traveled from the original scratch border. Two-sided Wilcoxon rank sum, error bars represent +/- SD. n = 20 image sets (*Crabp1* pKO), n = 18 image sets (*Atg7* pKO). **B**) Representative imaging of B16-F10 *Crabp1* pKO (top) or *Atg7* pKO (bottom) cells at 0, 6, 12, and 24 hours post-scratch. Bars represent 1000 um.

3.6 MATERIALS AND METHODS

Please note that Materials and Methods below are the work of Austin Gabel, included in publication Gabel et al., *Nature Communications* (2024). I have reduced Materials and Methods to those elements that I collaborated on or those that provide context to this Chapter. I specifically wrote and performed sections 3.6.8 and 3.6.9 and collaborated in performing additional sections.

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. Six-week-old C57BL/6 male mice were obtained from the Jackson Laboratory for all animal experiments listed here. Animals were housed for 1 week prior to initiation of any experiments in line with IACUC guidance.

Cell culture and in vitro validation studies

B16-F10 cells were obtained from ATCC (CRL-6475) and cultured per the manufacturer's instructions. 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.

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. $5x10^6$ cells were then injected subcutaneously into each flank of adult male C57BL/6 mice, and monitored using calipers. Animals were euthanized when a tumor reached 1.5 cm in any dimension in accordance with IACUC guidelines, and no tumors ever exceeded this threshold in any dimension. 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 (ThermoFisher 15596018) for subsequent RNA isolation.

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 log-rank 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 (Wang and Tian, 2020). For this Chapter, 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.

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) staining was performed using a rabbit monoclonal antibody. 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.

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 (ThermoFisher 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. For the Actinomycin D treatment time course, cells were plated and treated with 1 ug/mL Actinomycin D (Sigma-Aldrich A1410-2MG) for the indicated time frame, whereafter cells were harvested and RNA was isolated as described above. Specific primer sequences are listed in Supplementary Data File 6.

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 BOLT system (ThermoFisher 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) 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.

Scratch migration assay

B16-F10 cells were plated on 12 well plates and grown to 90% confluency. Scratches were generated in a straight line using sterile pipette tips, and dislodged cells were removed through two successive PBS rinses. Cells were grown in serum free media and imaged using BioTek Cytation at 0, 6, 12, and 24 hours. Gap width analysis was performed using the Wound_healing_size_tool plugin in ImageJ (Suarez-Arnedo et al., 2020). Multiple sites were measured in each well.

Overexpression assay

Transfer plasmids were generated on the pLentiGuide_CMV_EF1a_Neo backbone. Lentivirus was generated by transfecting HEK293T with 2 ug plasmid DNA (transfer plasmid, psPAX2, VSV-G) and the Lipofectamine 3000 system (ThermoFisher Scientific). Lentivirus titer was calculated and equivalent lentivirus was added to 100,000 B16-F10 cells. After 24 hr, cells were selected with G418 (3 mg/mL). Western blot and proliferation assays were performed after two weeks of selection.

3.7 REFERENCES

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), 5101. https://doi.org/10.1038/s41467-021-25388-8

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

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

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

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

Gabel, A. M., Belleville, A. E., Thomas, J. D., McKellar, S. A., Nicholas, T. R., Banjo, T., Crosse, E. I., & Bradley, R. K. (2024). Multiplexed screening reveals how cancer-specific alternative polyadenylation shapes tumor growth in vivo. *Nature Communications*, *15*(1), 959. https://doi.org/10.1038/s41467-024-44931-x

Galluzzi, L., Pietrocola, F., Bravo-San Pedro, J. M., Amaravadi, R. K., Baehrecke, E. H., Cecconi, F., Codogno, P., Debnath, J., Gewirtz, D. A., Karantza, V., Kimmelman, A., Kumar, S., Levine, B., Maiuri, M. C., Martin, S. J., Penninger, J., Piacentini, M., Rubinsztein, D. C., Simon, H., … Kroemer, G. (2015). Autophagy in malignant transformation and cancer progression. *The EMBO Journal*, *34*(7), 856–880. https://doi.org/10.15252/embj.201490784

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

Kim, N., Chung, W., Eum, H. H., Lee, H.-O., & Park, W.-Y. (2019). Alternative polyadenylation of single cells delineates cell types and serves as a prognostic marker in early stage breast cancer. *PLOS ONE*, *14*(5), e0217196. https://doi.org/10.1371/journal.pone.0217196

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

Li, W. V., Zheng, D., Wang, R., & Tian, B. (2021). MAAPER: model-based analysis of alternative polyadenylation using 3′ end-linked reads. *Genome Biology*, *22*(1), 222. https://doi.org/10.1186/s13059-021-02429-5

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

Long, J. S., Kania, E., McEwan, D. G., Barthet, V. J. A., Brucoli, M., Ladds, M. J. G. W., Nössing, C., & Ryan, K. M. (2022). ATG7 is a haploinsufficient repressor of tumor

progression and promoter of metastasis. *Proceedings of the National Academy of Sciences*, *119*(28). https://doi.org/10.1073/pnas.2113465119

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

Movassat, M., Crabb, T. L., Busch, A., Yao, C., Reynolds, D. J., Shi, Y., & Hertel, K. J. (2016). Coupling between alternative polyadenylation and alternative splicing is limited to terminal introns. *RNA Biology*, *13*(7), 646–655. https://doi.org/10.1080/15476286.2016.1191727

Mulcahy Levy, J. M., & Thorburn, A. (2020). Autophagy in cancer: moving from understanding mechanism to improving therapy responses in patients. *Cell Death & Differentiation*, *27*(3), 843–857. https://doi.org/10.1038/s41418-019-0474-7

Navickas, A., Asgharian, H., Winkler, J., Fish, L., Garcia, K., Markett, D., Dodel, M., Culbertson, B., Miglani, S., Joshi, T., Yin, K., Nguyen, P., Zhang, S., Stevers, N., Hwang, H.-W., Mardakheh, F., Goga, A., & Goodarzi, H. (2023). An mRNA processing pathway suppresses metastasis by governing translational control from the nucleus. *Nature Cell Biology*, *25*(6), 892–903. https://doi.org/10.1038/s41556-023-01141-9

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

Rigo, F., & Martinson, H. G. (2008). Functional Coupling of Last-Intron Splicing and 3′-End Processing to Transcription In Vitro: the Poly(A) Signal Couples to Splicing before Committing to Cleavage. *Molecular and Cellular Biology*, *28*(2), 849–862. https://doi.org/10.1128/MCB.01410-07

Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science. 2008 Jun 20;320(5883):1643-7. doi: 10.1126/science.

Shulman, E. D., & Elkon, R. (2019). Cell-type-specific analysis of alternative polyadenylation using single-cell transcriptomics data. *Nucleic Acids Research*, *47*(19), 10027–10039. https://doi.org/10.1093/nar/gkz781

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

Steeg, P. S. (2006). Tumor metastasis: mechanistic insights and clinical challenges. *Nature Medicine*, *12*(8), 895–904. https://doi.org/10.1038/nm1469

Suarez-Arnedo A, Torres Figueroa F, Clavijo C, Arbeláez P, Cruz JC, Muñoz-Camargo C. An image J plugin for the high throughput image analysis of in vitro scratch wound healing assays. PLoS One. 2020 Jul 28;15(7):e0232565. doi: 10.1371/journal.pone.0232565.

Tian, B., & Manley, J. L. (2017). Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology*, *18*(1), 18–30. https://doi.org/10.1038/nrm.2016.116

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

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

Chapter 4. VAULT'S ROLE IN ANTI-TUMOR IMMUNITY

The following work was performed in collaboration with Siegen McKellar (Rob Bradley lab, Fred Hutchinson Cancer Center). The work represented here is ongoing and will serve as the first chapters of the story that will be continued by Siegen McKellar. I developed the mouse model and performed the computational analysis demonstrated below; Siegen and I collaborated on mouse studies and molecular work.

4.1 SUMMARY

We have preliminary evidence that Major Vault Protein (MVP), the major constituent of the vault particle, is positively correlated with patient survival in both renal cell carcinoma and melanoma contexts. We demonstrate that *MVP* expression in both malignant cells and immune populations is associated with improved response to immune checkpoint blockade (ICB). We suggest vault's potential role in modulating interferon gamma (IFNγ) response in murine malignant cells. Lastly, we demonstrate that host-expressed *MVP* in mice is significantly associated with reduced tumor burden and improved survival outcomes. There is still much to learn about vault's tumor-intrinsic and immune cell-intrinsic roles in anti-tumor immunity.

4.2 INTRODUCTION

Over the past two centuries, scientists have been engaged in the characterization of cell organelles, beginning with the advent of the first light microscopes and later with electron microscopy. It wasn't until 1986 that researchers first documented vaults—which are large cellular ribonucleoparticles— despite their size (13 MDa) and widespread expression in animal cells (Kedersha and Rome, 1986). Vaults were aptly named after their barrel-shaped symmetry,

which resembles the vaulted ceilings of cathedrals (Rome et al., 1991). The primary structure of the vault particle is composed of thirty-nine copies of Major Vault Protein (MVP), which self-oligomerize to form half the domed vault structure (Stephen et al., 2001). The MVP-dependent formation of the vault particle is highly dynamic, allowing for the incorporation of two vault-associated proteins, poly(ADP-ribose)-polymerase (vPARP) and telomerase-associated protein 1 (TEP1) (Poderycki et al., 2006). Additional small untranslated vault RNAs (vtRNAs) are associated with vault RNPs to differing degrees across species (Stadler et al., 2009). Because the primary structural component of vault is MVP, and MVP particles readily form vault structures, most studies approximate MVP expression and localization to that of the vault particle. There is some evidence that MVP may have functional significance outside of the vault particle, for example on the cell-surface (Lee et al., 2017), but more work is needed to fully characterize vault-independent MVP function.

Given the intriguing hollow nanocapsule structure of the vault particle, as well as its conservation across many eukaryotic species (Kedersha et al., 1990), studies have focused on vault's fundamental cellular function. Many publications followed the discovery of the vault particle, largely focused on the role of vault, and consequently MVP, in chemotherapeutic drug resistance. Multiple *in vitro* studies in human chemoresistant cell models described a correlation between drug resistance and MVP overexpression (reviewed in Mossink et al., 2003 and Berger et al., 2009). This is an intriguing correlation, especially given vault's localization at the nuclear pore complex (Dickenson et al., 2007) which may point to vault's role in cyto-nuclear trafficking. Despite these correlations, mechanistic evidence directly linking vault particles to drug resistance remains scant. Furthermore, *in vivo* studies utilizing *Mvp* knockout mice do not exhibit increased sensitivity to cytostatic drugs (Mossink et al., 2002), indicating that while MVP

may be indicative of a drug resistance phenotype, its functional role in this process has not yet been elucidated.

Compelling insights into vault's cellular function emerge from its involvement in immune cell development and innate immune response to pathogens. Notably, macrophages, dendritic, and epithelial cells are among the cell types that exhibit the highest levels of *MVP* expression. Depleting MVP impedes proper dendritic cell differentiation and antigen presentation capabilities (Schroeijers et al., 2002). MVP is upregulated in the setting of viral infection, where MVP is necessary for optimal interferon production (Liu et al., 2012). In mouse models, MVP loss leads to decreased antiviral cytokine production and results in suboptimal viral response (Peng et al., 2016). In epithelial cells exposed to *Pseudomonas aeruginosa*, MVP is rapidly recruited to lipid rafts, which are dynamic protein assemblies within the cell membrane, where MVP plays a role in cell internalization and clearance of bacteria (Kowalski et al., 2007). Many lines of evidence, therefore, point to vault's important role in immune cell development and pathogen response.

Lastly, due in part to vault's association with lipid rafts in the cell membrane as well as localization at nuclear membrane, vault has been postulated to be an important regulator of multiple intracellular signaling pathways. Among the many pathways MVP has been indicated to play a role, MVP was first suggested to regulate the tumor suppressor PTEN by binding in a Ca^{2+} dependent manner (Yu et al., 2002). MVP was demonstrated to interact with multiple components of the epidermal growth factor (EGFR)-induced MAPK pathway, such as Erk and Shp2 (Kolli et al., 2004), and may act as a scaffold for protein components within this pathway. *MVP* transcription is also responsive to many cytokines, including type II interferon. *MVP*'s promoter contains a STAT1-binding site and transcription is stimulated by exposure to interferon

gamma (IFNγ; Steiner et al., 2006). Interestingly, in *MVP*-negative lung cancer models, MVP overexpression leads to reduced STAT1 phosphorylation and thereby reduces IFNγ response. These studies suggest that while *MVP* is stimulated by IFNγ, it may also modulate interferon response in various cellular contexts and with different external stimuli (Steiner et al., 2006). MVP's involvement in these diverse pathways implicates MVP and potentially other proteins within the vault structure as important scaffolds for many protein interactions.

Tumor-intrinsic MVP has been studied in cancer outside of multidrug resistance, but its associations with tumorigenesis seem to be largely context-dependent. For example, MVP is overexpressed in papillary thyroid cancer where it was shown to activate the PI3K/AKT/mTOR and MAPK/ERK pathways (Dong et al., 2022). Similarly, MVP upregulation in colon cancer mouse models was demonstrated to alter miRNA loading into exosomes and thereby promote tumorigenesis (Teng et al., 2017). In contrast, MVP is upregulated in a lung adenocarcinoma model where it is believed to inhibit the STAT3 signaling pathway and inhibit tumorigenesis (Bai et al., 2019). It is clear that MVP plays an important but varied role in solid tumors that is likely context-specific. As the paradigm of solid tumor cancer therapy has shifted towards blockade of T-cell inhibitory pathways, there is still much unknown about the role MVP may play in the critical response to immune checkpoint blockade (ICB). I will present our findings demonstrating that *MVP* expression is associated with enhanced patient survival across various cancer subtypes. Notably, this survival correlation is strengthened in the context of ICB treatment. Considering MVP's pivotal role in the immune response to pathogens, we hypothesize that MVP functions as more than a marker of improved treatment response, and may be involved in enhancing tumor cell recognition and ICB response in malignant cells.

4.3 RESULTS

4.3.1 MVP expression status is associated with improved survival and ICB response in renal cell carcinoma and metastatic melanoma.

In order to assess *MVP* expression across multiple cancers, we used bulk RNA-seq data from The Cancer Genome Atlas (TCGA). We stratified patients from each cohort into lowest and highest terciles of *MVP* expression and generated Kaplan-Meier estimates of overall survival. Notably among primary cancer types, *MVP* expression was significantly correlated (p = 4.8e-3) with improved survival in the papillary renal cell carcinoma cohort (**Figure 4.1A**). Given MVP's known role in modulating response to IFNγ, we also searched available ICB-treated bulk RNA-seq datasets to see if *MVP* expression correlated with improved survival in this context. Indeed, in two metastatic melanoma cohorts treated with either anti-PD1 therapy (Liu et al., 2019) or anti-CTLA-4 therapy (Van Allen et al., 2015), *MVP* expression was significantly correlated (p = 1.6e-3 and p = 9.04e-4, respectively) with improved patient survival (**Figure 4.1B**).

We next investigated if *MVP* expression was derived from malignant or immune cell subsets within the tumor, and specifically how *MVP* expression within these cell subsets is correlated with an ICB-responsive phenotype. Given the association between *MVP* expression and patient survival in renal cell carcinoma, we obtained published single-cell RNA sequencing (scRNA-seq) from patients diagnosed with either papillary renal cell carcinoma or clear cell renal cell carcinoma (Bi et al., 2021). We separated cell expression data based on previously described cell identities (Bi et al., 2021). Interestingly, malignant cells, tumor associated macrophages (TAMs), and dendritic cells (DCs) all expressed the highest levels of MVP in ICB-naïve patients (**Figure 4.1C**), aligning with previous findings that high MVP expression is critical for monocyte differentiation and DC function (Schroeijers et al., 2002). Furthermore,

MVP expression is significantly upregulated in all three cell subsets upon ICB exposure (**Figure 4.1C**). ICB response amongst the patient samples was categorized by RECIST criterion, and *MVP* expression was significantly higher in patients with partial response (PR) relative to those with stable disease (SD) or progressive disease (SD) (**Figure 4.1D**). These data suggest that *MVP* expression is driven by both malignant cell and monocyte-derived cells within the tumor microenvironment and is linked with an improved response to ICB therapy.

4.3.2 Murine Mvp modulates immune signaling

To first explore the intrinsic role of MVP in malignant cells, we used a common murine melanoma cell model (B16-F10). Using pooled gRNAs and transient Cas9 expression, we generated a polyclonal B16-F10 line with 50% reduction in MVP protein ($MVP^{-/-}$, **Figure 4.2A**). To determine how reduced MVP protein abundance would impact transcriptomic response to IFNγ, we stimulated $MVP^{-/-}$ cells and cells treated with non-targeting control gRNA with IFNγ for 24 hours and performed RNA-seq. At baseline, $Mvp^{-/-}$ cells demonstrated modest global transcriptomic dysregulation compared to control cells (**Figure 4.2B**). Transcriptomic differences were accentuated by IFNγ stimulation, leading to significant upregulation of 1.4% of mRNA transcripts and significant downregulation of 0.91% of mRNA transcripts (**Figure 4.2C**). Among the gene set pathways most differentially upregulated in $MVP^{-/-}$ B16-F10 cells were TNFa signaling via NFkB, p53 pathway, and both type I and type II IFN responses (**Figure 4.2D**). We next plotted differentially expressed IFNγ-stimulated genes (ISGs) between $MVP^{-/-}$ and control cells (**Figure 4.2E**). This confirmed *Mvp* as a differentially expressed ISG, with reduced expression in $MVP^{-/-}$ cells upon IFNγ stimulation when compared to IFNγ stimulated control cells. In addition, there is widespread upregulation of ISGs in B16-F10 $Mvp^{-/-}$ cells, with

a select number of ISGs significantly downregulated (Figure 4.2E). These data suggest a potential role of MVP in modulating the immune response to IFNγ.

4.3.3 Genetic depletion of Mvp results in tumor proliferation and reduced survival

Given MVP's role in immune recognition and because MVP expression in TAM and DC significantly correlates with improved patient survival, we next questioned a potential causative role MVP might play in modulated anti-tumor immune response. We generated an *Mvp* knockout (*Mvp*$^{-/-}$) mouse to address MVP's role in murine immune response to cancer. Using pgRNAs targeting *Mvp* coding exon 4, we generated a knockout allele with limited genomic alteration (**Supplementary Figure 1A**). DNA editing at both pgRNAs introduced two single nucleotide insertions, with the insertion in exon 4 predicted to introduce a premature termination codon and target the transcript for degradation via NMD (**Supplementary Fig. 1B**). We validated total loss of MVP protein in *Mvp*$^{-/-}$ mice and backcrossed the model at least four generations to C57BL/6 to reduce the impact of any off-target genomic editing (**Supplementary Figure 1C**).

We next sought to determine whether melanoma tumors engrafted in *Mvp*$^{-/-}$ mice had a tumorigenic advantage, using B16-F10 cells engineered to express the model antigen ovalbumin (OVA) to increase B16-F10 immunogenicity (**Supplementary Figure 2A**).
Following tumor engraftment, we initiated dual ICB treatment with both anti-PD-1 and anti-CTLA-4 antibodies (**Figure 4.3A**) and noted a significant reduction in tumor burden in both control and MVP$^{-/-}$ mice treated with ICB (p = 6.9e-4 and p = 8.9e-5, respectively; **Figure 4.3B**). Interestingly, the tumors engrafted in MVP$^{-/-}$ mice and treated with control antibodies tended to be larger than those engrafted in control mice (p = 0.079, **Figure 4.3B**). Most strikingly, there was a significant reduction in survival for *Mvp*$^{-/-}$ mice harboring B16-F10-OVA tumors when

compared to control mice ($p = 0.0027$, **Figure 4.3C**). These data suggest that while MVP may not alter ICB responsiveness in this melanoma model, host-expressed MVP is protective against melanoma growth and is correlated with improved survival.

4.4 DISCUSSION

We have demonstrated preliminary evidence in both human and mouse contexts that MVP is likely important for anti-tumor immunity. There is much more work to be done to fully characterize MVP's tumor cell-intrinsic and immune cell-intrinsic roles.

To begin, we have shown correlative evidence from human renal cell carcinoma and melanoma cohorts, but we have only begun work in murine melanoma model. It will be essential to also use a relevant renal cell carcinoma model, especially given differences in population and density of immune infiltration in the TME of both melanoma (Erdag et al., 2012) and renal cell carcinoma (Becht et al., 2016). The genetic background, cell type, and tumor context undoubtedly influence MVP's context-specific role and warrant appropriate modeling.

In the context of melanoma treated with anti-CTLA-4 therapy, the IFNγ response pathway within tumor cells is critical to ICB response, and knockdown of genes within this response pathway reduces survival (Gao et al., 2016). We have demonstrated MVP plays a role in modulating the IFNγ response, largely by negatively regulating the expression of ISGs in B16-F10 (**Figure 4.2E**). Mechanistic work to characterize how MVP protein interacts with IFNγ pathway components will enlighten this regulatory role. Additionally, using $MVP^{-/-}$ cells to study clinical response markers like major histocompatibility complex (MHC) presentation or antigen presentation could help to characterize how MVP enhances ICB therapy effectiveness.

When considering the role of MVP in immune cell response to ICB therapy, it will be critical to use scRNAseq approaches to define cell types and cell states of interest. While we have shown preliminary evidence that TAMs upregulate MVP in response to an interferon-rich microenvironment in ICB-exposed tumors, there is notable TAM diversity in renal cell carcinoma (Chevrier et al., 2017) that we have not yet addressed with publicly available scRNAseq data. It will be important to determine if specific TAM identities or additional markers are correlated with MVP expression. Furthermore, we are actively interested in defining the TME of melanoma cells engrafted in MVP-deficient mice through scRNAseq. It is attractive to speculate that MVP loss in TAM and DC prevents proper antigen presentation or anti-tumor immune response. Lastly, it is possible and intriguing to consider that MVP may play a role in ICB response in both the tumor and host compartments—as has been demonstrated of PD-L1 expression in murine colon adenocarcinoma models (Lau et al., 2017).

4.5 FIGURES

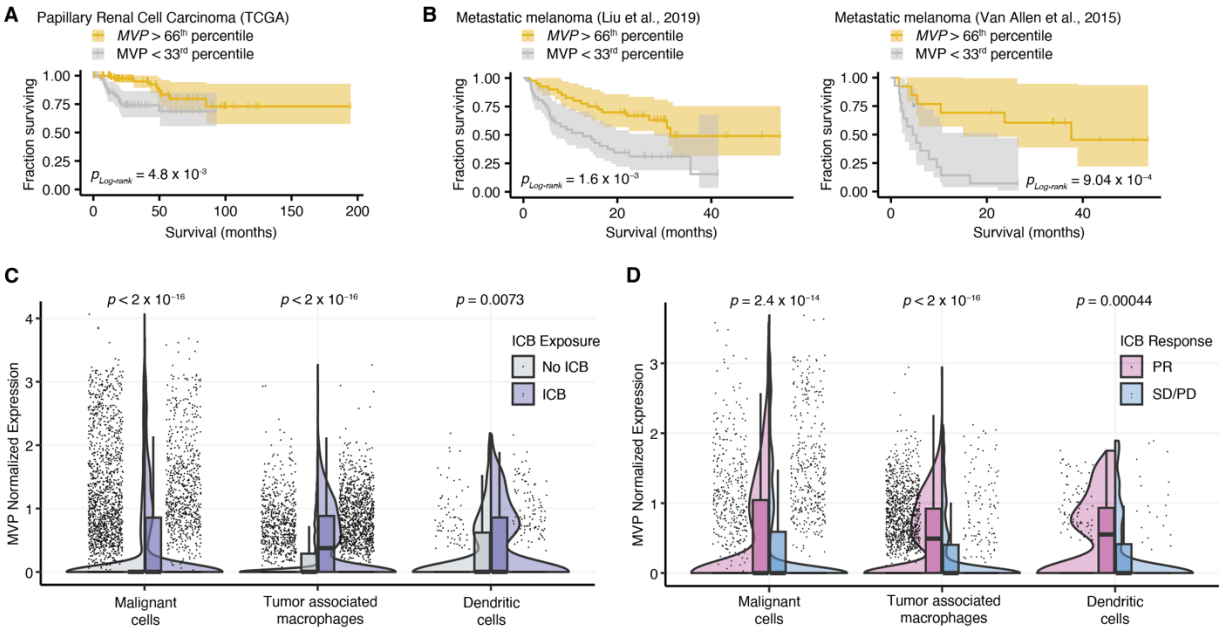

Figure 4.1 MVP expression status is associated with improved survival and ICB response in renal cell carcinoma and metastatic melanoma.

A, Kaplan-Meier (KM) estimate of overall survival (solid lines) for patients with papillary renal cell carcinoma stratified by *MVP* expression status (95% confidence interval, log-rank test).
B, KM estimate of overall survival for patients with ICB-treated metastatic melanoma from two separate datasets, stratified by *MVP* expression status (95% confidence interval, log-rank test).
C, Violin and boxplots comparing *MVP* expression distributions between all malignant cells, tumor associated macrophages, and dendritic cells from ICB-naïve and ICB-exposed patients. Significance of differential expression was determined by two-sided Wilcoxon rank-sum test. Boxplot represents median expression (centerline), upper and lower quartiles, and whiskers extend to the largest value no further than 1.5 * IQR.
D, as in **C**, but comparing cells from ICB PR versus ICB SD/PD patients. P-values computed using two-sided Wilcoxon rank-sum test.

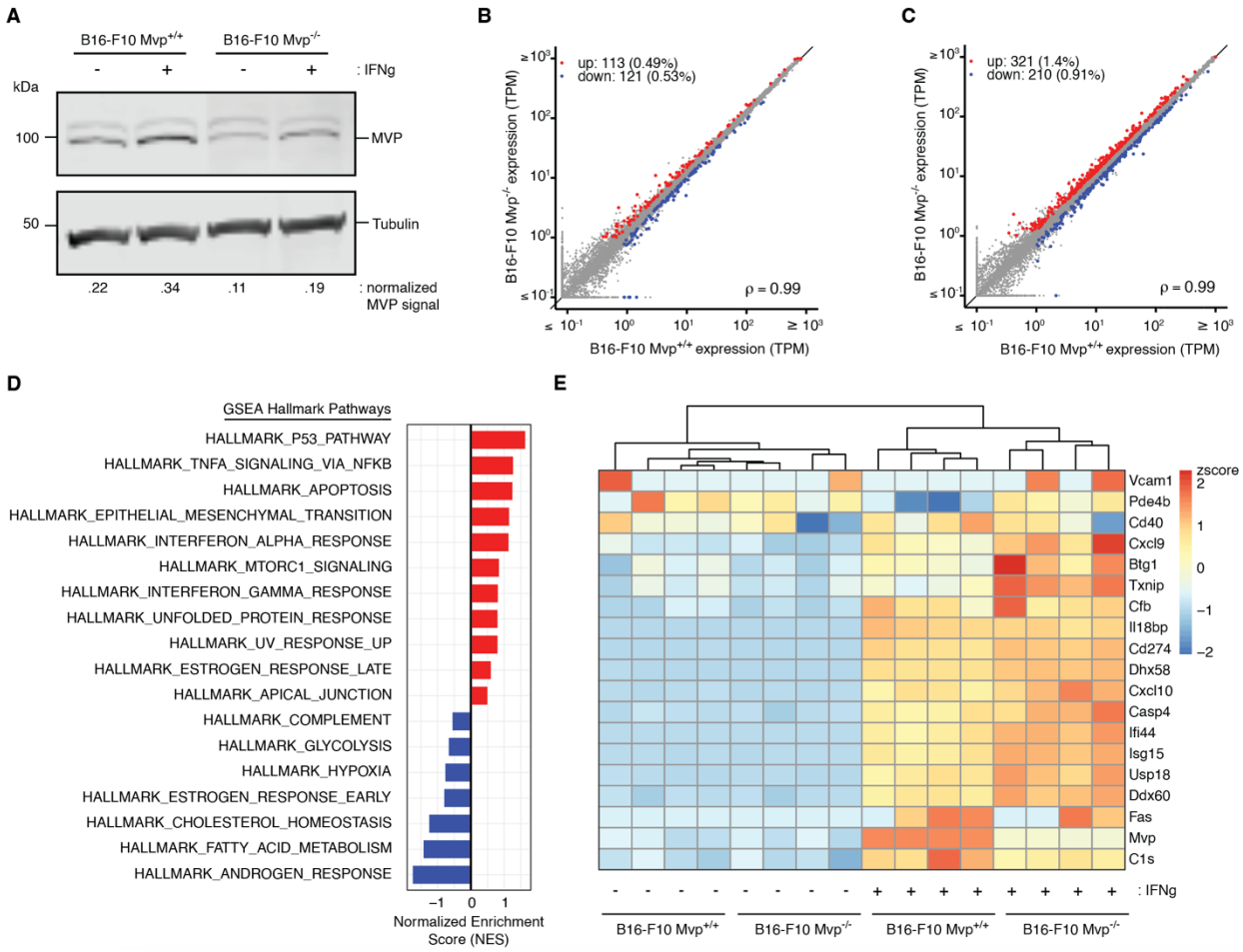

Figure 4.2. MVP modulates gene expression and response to Type II Interferon.

A, Representative Western blot of MVP signal in B16-F10 control cells (B16-F10 MVP$^{+/+}$) and B16-F10 MVP knockout (MVP$^{-/-}$) cells treated with Interferon gamma (IFNγ). MVP signal is normalized to alpha-tubulin signal.

B-C, Scatterplots of differentially expressed transcripts from B16-F10 control and B16-F10 MVP$^{-/-}$ cells. Red and blue points represent transcripts with significantly increased and decreased expression in the MVP$^{-/-}$ cells, respectively. Cells are untreated (**A**) or stimulated with IFNγ (**B**).

D, *Data forthcoming.*

E, IFNγ response gene expression in B16-F10 control or B16-F10 MVP$^{-/-}$ cells.

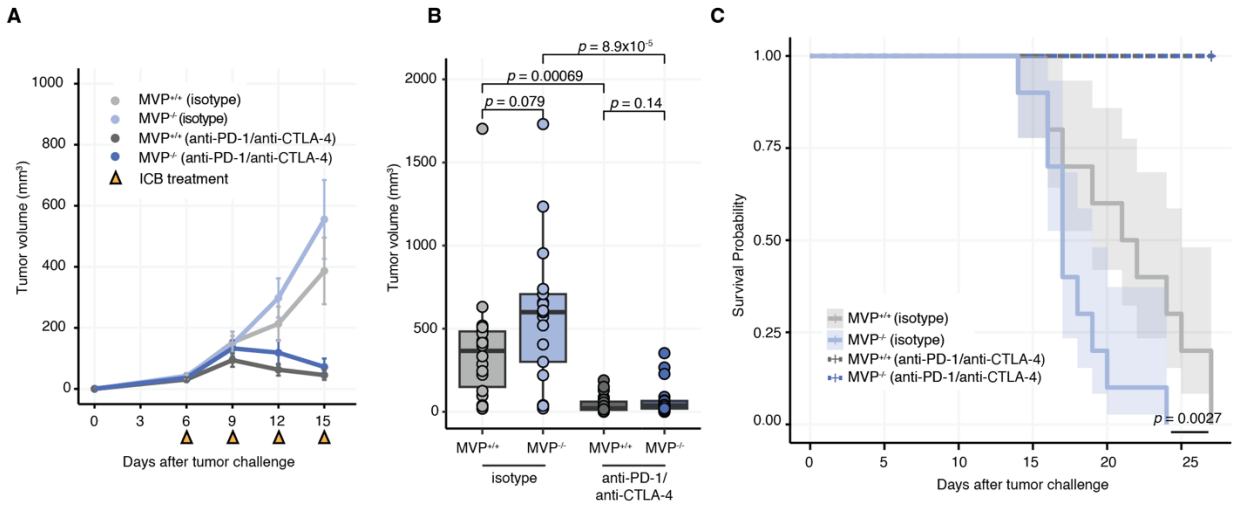

Figure 4.3. Host MVP is tumor-suppressive and correlates with improved survival.

A-C, B16-F10-Ova tumors engrafted in Mvp knockout ($Mvp^{-/-}$) or Mvp wild-type ($Mvp^{+/+}$) mice treated with dual therapy (anti-PD-1 and anti-CTLA-4) or isotype control. n = 10 animals, 20 tumors, per group. **A**, Tumor growth and immune-checkpoint blockade (ICB) treatment, data are mean +/- SEM. **B**, Tumor volume at day 15 post-engraftment, one-sided Student's T-test. **C**, Survival data from the same cohort, Log-rank test.

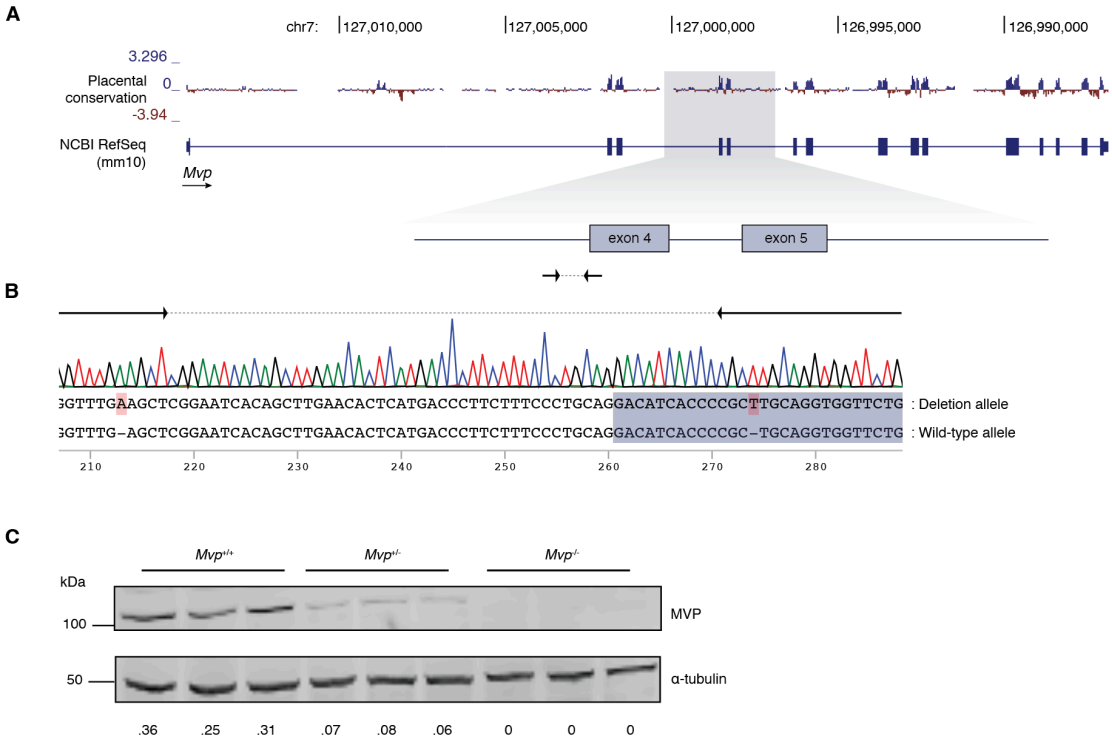

Figure 4.4. Supplementary Figure 1. Generation of Mvp$^{-/-}$ mouse.

A, Paired guide RNAs (pgRNAs, arrows) were designed to disrupt exon 4 (purple box) of *Mvp*.
B, Sanger sequencing trace of 1 bp insertion in intron 3 and 1 bp insertion (red highlights) in exon 4 (purple box), including pgRNA-targeted sequences (arrows).
C, Representative Western blot of MVP signal from wild-type (*Mvp*$^{+/+}$), heterozygous (*Mvp*$^{+/-}$) and homozygous (*Mvp*$^{-/-}$) mouse liver lysate. *n* = 1 animal per genotype, run in technical triplicate.

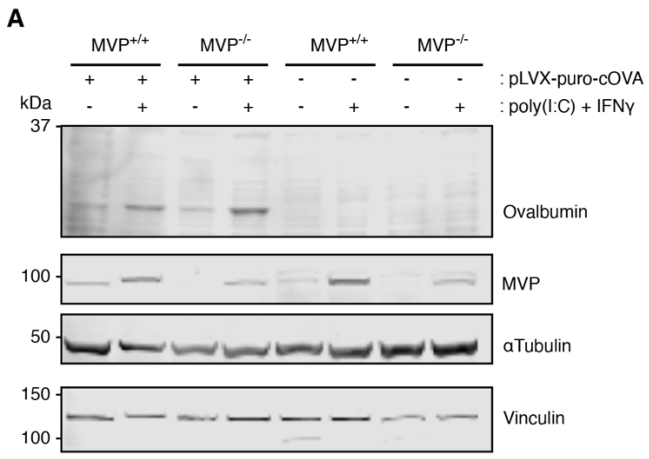

Figure 4.5. Supplementary Figure 2. Generation of Ova-expressing CRISPR-edited B16-F10 cells.

A, Representative Western blot of CRISPR-edited control or *Mvp* knockout (*Mvp*$^{-/-}$) B16-F10 cells stably transduced with pLVX-puro-cOVA to express ovalbumin. Cells were stimulated by IFNγ and poly(I:C) to induce *Mvp* expression.

4.6 MATERIALS AND METHODS

Generation of CRISPR-edited and OVA expressing tumor cell lines

gRNAs were generated by duplexing tracrRNA with either pre-designed crRNAs targeting two *Mvp* coding exons or pre-designed crRNAs targeting two non-coding regions (Supplementary Table 4.2; Integrated DNA Technologies, Alt-R system). The duplexed gRNA was combined with Cas9 endonuclease (ThermoFisher, A36499) prior to transfection of $5x10^5$ B16-F10 cells (Lonza, Nucleofector Kit V). B16-F10 editing was later validated using T7 digestion (New England Biolabs, E3321) and Western blotting.

To generate Ova-expressing lentivirus, $1x10^6$ HEK293T cells were transfected with 1.2 ug of pLVX-puro-cOVA (Addgene plasmid #135073) and 1:2 ratio of psPAX2 (Addgene plasmid #12260) and pMD2.G (Addgene plasmid #12259) to generate lentiviral particles. Lentiviral media was collected 48 hours post-transfection and filtered using a 0.45 uM PES filter. $1x10^5$ CRISPR-edited B16-F10 cells were transduced with 250 uL of lentivirus and were selected with puromycin (1.5 ug/mL) for 7 days.

B16-F10 stimulation was performed with 50 ug/mL poly(I:C) (InvivoGen) and or 10 ng/mL IFNγ (Cell Signaling Technology, 39127).

Western blotting

Total protein lysates were prepared in RIPA buffer (Cell Signaling) with 1mM Pefabloc (Sigma-Aldrich) and quantified using the Pierce 660 nm Protein Assay Reagent (ThermoFisher). Protein lysates were electrophoretically separated and transferred to nitrocellulose membrane using NuPAGE systems (ThermoFisher). Membranes were blocked with Odyssey Blocking Buffer

(LI-COR Biosciences) for one hour at room temperature. Primary antibody incubation was performed overnight at 4°C in Intercept Antibody Diluent (LI-COR Biosciences), with MVP (ThermoFisher 16478-1-AP, 1:1000) and alpha-tubulin (Sigma-Aldrich T6199, 1:1000) primary antibodies. IRDye (LI-COR Biosciences) secondary antibodies were used for detection and imaging using the Odyssey CLx Imager (LI-COR Biosciences). Protein blot analysis and quantification was performed using FIJI/ImageJ.

RNA preparation and RNA-sequencing

B16-F10 RNA was extracted using Direct-zol RNA MiniPrep kit (Zymo Research), according to the manufacturer's instructions. For tissue RNA extractions, tissue was homogenized in TRIzol reagent with zirconium beads prior to RNA extraction, using the Zymo Direct-zol RNA Miniprep kit. A minimum of 500 ng of high-quality RNA was used as input for library preparation. Poly(A)-selected unstranded libraries were prepared using a TruSeq protocol and sequenced using an Illumina sequencer to obtain 100 bp paired-end reads.

RNA-seq reads were processed as previously described (Dvinge et al., 2014). In brief, FASTQ files were trimmed to remove sequencing adapters and then aligned to the hg19/GRCh37 reference assembly. This assembly was created by merging the UCSC knownGene gene annotation (Meyer et al., 2012), Ensembl gene annotation (Flicek et al., 2012), and MISO isoform annotation (Katz et al., 2010). Read alignment and expression estimates were generated with RSEM (RNA-seq by expectation maximization; Li & Dewey, 2011), Bowtie (Langmead et al., 2009), and TopHat (Trapnell et al., 2009), and isoform ratios were quantified with MISO v.2.0 (Katz et al., 2010).

All analyses and visualization were performed in the R programming environment, with tools from Bioconductor and dplyr, clusterProfiler, ggplot2, tidyverse packages.

Mouse model generation and genotyping

All deletion alleles were generated by CRISPR/Cas9-mediated genome editing with paired guide RNAs (pgRNAs), as previously described (Singh et al., 2015; Thomas et al., 2020). High-quality pgRNA seed sequences (**Supplementary Table 4.2**) were selected using the Benchling guide design tool, confirmed by GuideScan, and analyzed for specificity using CasOffinder. pgRNAs were produced through cloning-free *in vitro* transcription as previously described (Varshney et al., 2015) using the MEGAshortscript T7 Transcription Kit (Life Technologies). Products were purified using the MEGAclear Transcription Clean-up Kit (Life Technologies). Cas9 mRNA (50 ng/μL) and pgRNA (25 ng/μL) were microinjected into single-cell zygotes derived from B6(Cg)-Tyrc-2J/J inbred mice (JAX strain #000664). Microinjected embryos were transplanted into pseudopregnant CD-1 recipients. Founder animals were genotyped using ear punch tissue with primers that flanked the expected deletion loci by 100 to 200 bases (**Supplementary Table 4.1**). The resulting amplicons were further characterized by Sanger sequencing and Amplicon-EZ Next-Generation Sequencing (Azenta Life Sciences). All founder animals were backcrossed to C57BL/6 breeders (JAX strain #000664) for at least four generations.

To genotype founder mice and subsequent generations, ear tissue was collected when the mice were two weeks old and digested in Lysis M Buffer (Macherey-Nagel) and Proteinase K (Qiagen) for five minutes at room temperature. This was followed by standard PCR using appropriate primers and gel electrophoresis (**Supplementary Table 4.1**). PCR products were subsequently Sanger sequenced when alleles were indistinguishable through gel electrophoresis

(**Supplementary Table 4.1**). After the establishment of the backcrossed line, genotyping was performed using quantitative genomic PCR (Transnetyx).

In vivo engraftment studies

CRISPR-edited B16-F10 cells were grown in standard conditions prior to engraftment. $0.5x10^6$ cells were injected subcutaneously into each flank of adult female C57BL/6 mice or age-matched $MVP^{-/-}$ mice (8-11 weeks), and measured using calipers. Mice were treated with intraperitoneal injection of 250 μg anti-PD-1 (BioXCell, clone RMP1-14) and 100 μg anti-CTLA-4 (BioXCell, clone 9H10) or 250 μg rat IgG (BioXCell, BE0089) and 100 μg hamster IgG (BioXCell, BE0087) on days 6, 9, 12, and 15 post-engraftment. Researchers were blinded to group while measuring tumors. Animals were euthanized when a tumor reached 1.5 cm in any dimension in accordance with IACUC guidelines, and no tumors exceeded this threshold in any dimension. Tumor volume was estimated with the formula $width^2$ x length. Tumor material was either fixed in 10% formalin for 48 hours for histology studies, homogenized in TRIzol reagent (ThermoFisher 15596018) for subsequent RNA isolation, of fixed in paraformaldehyde for future scRNA-seq experimentation (10X Genomics, PN-1000414).

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. Seven to eleven-week-old C57BL/6 female mice were obtained from the Jackson Laboratory or bred at Fred Hutchinson Cancer Center for all animal experiments listed here.

4.7 TABLES

Table 4.1. Primers used for Sanger sequencing and gDNA PCR.

primer_F_id	primer_F_seq	primer_R_id	primer_R_seq	use
MVP_5F	GCATCAGCTGTCTCCTCCTC	MVP_5R	AGTTCTTCAGACCTGGTGGC	mouse *Mvp* PCR genotyping, Sanger sequencing

Table 4.2. Sequences of guide RNAs used in model generation and gene knockout

gene	target	species	grna1	grna1 PAM	grna2	grna2 PAM
Mvp	exon 4	mouse	CCAAGTATCAGGTTTGAGCT	NGG	CGCTGCAGGTGGTTCTGCCC	NGG
Mvp	exon 2	mouse	ACCAAAGACCTACATCCGGC	NGG	NA	NA
Mvp	exon 3	mouse	GTTCGCATGGTGACGGTCCC	NGG	NA	NA

4.8 REFERENCES

Bai, H., Wang, C., Qi, Y., Xu, J., Li, N., Chen, L., Jiang, B., Zhu, X., Zhang, H., Li, X., Yang, Q., Ma, J., Xu, Y., Ben, J., & Chen, Q. (2019). Major vault protein suppresses lung cancer cell proliferation by inhibiting STAT3 signaling pathway. *BMC Cancer*, *19*(1), 454. https://doi.org/10.1186/s12885-019-5665-6

Becht, E., Giraldo, N. A., Dieu-Nosjean, M.-C., Sautès-Fridman, C., & Fridman, W. H. (2016). Cancer immune contexture and immunotherapy. *Current Opinion in Immunology*, *39*, 7–13. https://doi.org/10.1016/j.coi.2015.11.009

Berger, W., Steiner, E., Grusch, M., Elbling, L., & Micksche, M. (2009). Vaults and the major vault protein: Novel roles in signal pathway regulation and immunity. In *Cellular and Molecular Life Sciences* (Vol. 66, Issue 1, pp. 43–61). https://doi.org/10.1007/s00018-008-8364-z

Chevrier, S., Levine, J. H., Zanotelli, V. R. T., Silina, K., Schulz, D., Bacac, M., Ries, C. H., Ailles, L., Jewett, M. A. S., Moch, H., van den Broek, M., Beisel, C., Stadler, M. B., Gedye, C., Reis, B., Pe'er, D., & Bodenmiller, B. (2017). An Immune Atlas of Clear Cell Renal Cell Carcinoma. *Cell*, *169*(4), 736-749.e18. https://doi.org/10.1016/j.cell.2017.04.016

Dickenson, N. E., Moore, D., Suprenant, K. A., & Dunn, R. C. (2007). Vault Ribonucleoprotein Particles and the Central Mass of the Nuclear Pore Complex. *Photochemistry and Photobiology*, *83*(3), 686–691. https://doi.org/10.1111/j.1751-1097.2007.00050.x

Dong, X., Akuetteh, P. D. P., Song, J., Ni, C., Jin, C., Li, H., Jiang, W., Si, Y., Zhang, X., Zhang, Q., & Huang, G. (2022). Major Vault Protein (MVP) Associated With BRAFV600E Mutation Is an Immune Microenvironment-Related Biomarker Promoting the Progression of Papillary Thyroid Cancer via MAPK/ERK and PI3K/AKT Pathways. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/fcell.2021.688370

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*, *111*(47), 16802–16807. https://doi.org/10.1073/pnas.1413374111

Erdag, G., Schaefer, J. T., Smolkin, M. E., Deacon, D. H., Shea, S. M., Dengel, L. T., Patterson, J. W., & Slingluff, C. L. (2012). Immunotype and Immunohistologic Characteristics of Tumor-Infiltrating Immune Cells Are Associated with Clinical Outcome in Metastatic Melanoma. *Cancer Research*, *72*(5), 1070–1080. https://doi.org/10.1158/0008-5472.CAN-11-3218

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. (2012). Ensembl 2013. *Nucleic Acids Research*, *41*(D1), D48–D55. https://doi.org/10.1093/nar/gks1236

Gao, J., Shi, L. Z., Zhao, H., Chen, J., Xiong, L., He, Q., Chen, T., Roszik, J., Bernatchez, C., Woodman, S. E., Chen, P.-L., Hwu, P., Allison, J. P., Futreal, A., Wargo, J. A., & Sharma, P. (2016). Loss of IFN-γ Pathway Genes in Tumor Cells as a Mechanism of Resistance to Anti-CTLA-4 Therapy. *Cell*, *167*(2), 397-404.e9. https://doi.org/10.1016/j.cell.2016.08.069

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

Kedersha, N. L., Miquel, M. C., Bittner, D., & Rome, L. H. (1990). Vaults. II. Ribonucleoprotein structures are highly conserved among higher and lower eukaryotes. *The Journal of Cell Biology*, *110*(4), 895–901. https://doi.org/10.1083/jcb.110.4.895

Kedersha, N. L., & Rome, L. H. (1986). Isolation and Characterization of a Novel Ribonucleoprotein Particle: Large Structures Contain a Single Species of Small RNA. In *The Journal of Cell Biology* (Vol. 103).

Kolli, S., Zito, C. I., Mossink, M. H., Wiemer, E. A. C., & Bennett, A. M. (2004). The Major Vault Protein Is a Novel Substrate for the Tyrosine Phosphatase SHP-2 and Scaffold Protein in Epidermal Growth Factor Signaling. *Journal of Biological Chemistry*, *279*(28), 29374–29385. https://doi.org/10.1074/jbc.M313955200

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), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lau, J., Cheung, J., Navarro, A., Lianoglou, S., Haley, B., Totpal, K., Sanders, L., Koeppen, H., Caplazi, P., McBride, J., Chiu, H., Hong, R., Grogan, J., Javinal, V., Yauch, R., Irving, B., Belvin, M., Mellman, I., Kim, J. M., & Schmidt, M. (2017). Tumour and host cell PD-L1 is required to mediate suppression of anti-tumour immunity in mice. *Nature Communications*, *8*(1), 14572. https://doi.org/10.1038/ncomms14572

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 323. https://doi.org/10.1186/1471-2105-12-323

Liu, D., Schilling, B., Liu, D., Sucker, A., Livingstone, E., Jerby-Arnon, 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

Liu, S., Hao, Q., Peng, N., Yue, X., Wang, Y., Chen, Y., Wu, J., & Zhu, Y. (2012). Major vault protein: A virus-induced host factor against viral replication through the induction of type-I interferon. *Hepatology*, *56*(1), 57–66. https://doi.org/10.1002/hep.25642

Lee, H. M., Joh, J. W., Seo, S.-R., Kim, W.-T., Kim, M. K., Choi, H. S., Kim, S. Y., Jang, Y.-J., Sinn, D. H., Choi, G. S., Kim, J. M., Kwon, C. H. D., Chang, H. J., Kim, D. S., & Ryu, C. J. (2017). Cell-surface major vault protein promotes cancer progression through harboring mesenchymal and intermediate circulating tumor cells in hepatocellular carcinomas. *Scientific Reports*, *7*(1), 13201. https://doi.org/10.1038/s41598-017-13501-1

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. (2012). The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Research*, *41*(D1), D64–D69. https://doi.org/10.1093/nar/gks1048

Mossink, M. H., van Zon, A., Fränzel-Luiten, E., Schoester, M., Kickhoefer, V. A., Scheffer, G. L., Scheper, R. J., Sonneveld, P., & Wiemer, E. A. C. (2002). Disruption of the murine major vault protein (MVP/LRP) gene does not induce hypersensitivity to cytostatics. *Cancer Research*, *62*(24), 7298–7304.

Mossink, M. H., van Zon, A., Scheper, R. J., Sonneveld, P., & Wiemer, E. A. (2003). Vaults: a ribonucleoprotein particle involved in drug resistance? *Oncogene*, *22*(47), 7458–7467. https://doi.org/10.1038/sj.onc.1206947

Peng, N., Liu, S., Xia, Z., Ren, S., Feng, J., Jing, M., Gao, X., Wiemer, E. A. C., & Zhu, Y. (2016). Inducible Major Vault Protein Plays a Pivotal Role in Double-Stranded RNA– or Virus-Induced Proinflammatory Response. *The Journal of Immunology*, *196*(6), 2753–2766. https://doi.org/10.4049/jimmunol.1501481

Rome, L., Kedersha, N., & Chugani, D. (1991). Unlocking vaults: organelles in search of a function. *Trends in Cell Biology*, *1*(2–3), 47–50. https://doi.org/10.1016/0962-8924(91)90088-Q

Schroeijers, A. B., Reurs, A. W., Scheffer, G. L., Stam, A. G. M., de Jong, M. C., Rustemeyer, T., Wiemer, E. A. C., de Gruijl, T. D., & Scheper, R. J. (2002). Up-Regulation of Drug Resistance-Related Vaults During Dendritic Cell Development. *The Journal of Immunology*, *168*(4), 1572–1578. https://doi.org/10.4049/jimmunol.168.4.1572

Singh, P., Schimenti, J. C., & Bolcun-Filas, E. (2015). A Mouse Geneticist's Practical Guide to CRISPR Applications. *Genetics*, *199*(1), 1–15. https://doi.org/10.1534/genetics.114.169771

Stadler, P. F., Chen, J. J.-L., Hackermuller, J., Hoffmann, S., Horn, F., Khaitovich, P., Kretzschmar, A. K., Mosig, A., Prohaska, S. J., Qi, X., Schutt, K., & Ullmann, K. (2009). Evolution of Vault RNAs. *Molecular Biology and Evolution*, *26*(9), 1975–1991. https://doi.org/10.1093/molbev/msp112

Steiner, E., Holzmann, K., Pirker, C., Elbling, L., Micksche, M., Sutterlüty, H., & Berger, W. (2006). The major vault protein is responsive to and interferes with interferon-γ-mediated STAT1 signals. *Journal of Cell Science*, *119*(3), 459–469. https://doi.org/10.1242/jcs.02773

Stephen, A. G., Raval-Fernandes, S., Huynh, T., Torres, M., Kickhoefer, V. A., & Rome, L. H. (2001). Assembly of Vault-like Particles in Insect Cells Expressing Only the Major Vault Protein. *Journal of Biological Chemistry*, *276*(26), 23217–23220. https://doi.org/10.1074/jbc.C100226200

Teng, Y., Ren, Y., Hu, X., Mu, J., Samykutty, A., Zhuang, X., Deng, Z., Kumar, A., Zhang, L., Merchant, M. L., Yan, J., Miller, D. M., & Zhang, H.-G. (2017). MVP-mediated exosomal sorting of miR-193a promotes colon cancer progression. *Nature Communications*, *8*(1), 14448. https://doi.org/10.1038/ncomms14448

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

Van Allen, E. M., Miao, D., Schilling, B., Shukla, S. A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Geukes Foppen, M. H., 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

Yu, Z., Fotouhi-Ardakani, N., Wu, L., Maoui, M., Wang, S., Banville, D., & Shen, S.-H. (2002). PTEN Associates with the Vault Particles in HeLa Cells. *Journal of Biological Chemistry*, *277*(43), 40247–40252. https://doi.org/10.1074/jbc.M207608200

VITA

Andrea E. Belleville was born in Fort Collins, Colorado, and grew up in Whitefish, Montana before returning to Fort Collins to graduate from Poudre High School. She matriculated to Seattle Pacific University (SPU) where she graduated with a B.S. in Physiology while enrolled in the University Scholars honors program. During her time at SPU, she engaged with research conducted at SPU and also received a fateful internship with Dr. Catherine Peichel at Fred Hutchinson Cancer Center (FHCC). Her work with threespine stickleback fish and excellent mentorship from Dr. Peichel changed the trajectory of her career path. After graduating, she worked as a technician with Dr. Stephen Tapscott at FHCC before enrolling in the University of Washington's Medical Scientist Training Program. When not in the lab or working, Andrea enjoys hiking, observing people, watching all types of sports (with a beer), and spending time with her husband and family.