UNIVERSITY OF CALIFORNIA,
IRVINE

Improvements on the Control of Gene Expression in Cancer Gene Therapies

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Molecular Biology and Biochemistry

by

Jingtian “Josh” Wang

Dissertation Committee:
Professor Robert Spitale, Chair
Professor Celia Goulding
Associate Professor Olga Razorenova
Professor Klemens Hertel
Associate Professor Scott Atwood

2022

DEDICATION

To

my parents, Guanrong and Ronghui
my colleagues, friends, and Jaclyn

for their endless support

But if you have no names in the hat,
how can you expect to be drawn from the hat? You can't.
Every day, you gotta put your name in it.

-Félix Lengyel

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to first express my appreciation to Dr. Renee Chosed. She personally taught me how to do my very first biology experiments. I would like to thank her for helping me start my academic career as a biologist, and her patience and kindness along the way.

I would like to thank Dr. Min-Ken Liao. She cared about me not only academically but also personally. She was not only my mentor but also someone to confide in.

I would like to thank Dr. Norma Frizzell and Dr. Geraldo Piroli. I was such a novice in science, but they treated me with patience. They played an instrumental role in shaping the biologist I am today.

I would like to thank Dr. Robert Spitale, for providing me with freedom and guidance during my PhD. I would like to thank him for injecting energy and enthusiasm into the lab. His passion and drive are an inspiration for me not only in science but also in life.

I would also like to thank my committee members: Drs. Celia Goulding, Scott Atwood, Klemens Hertel, and Olga Razorenova, for their words of advice and encouragement during our multiple meetings.

I would like to thank all current and past members of the Spitale Lab. I thank Dr. Whitney England, for teaching me so much about data analysis. She is the "full-stack" bioinformatician that I aspire to be. I thank Dr. Kim Nguyen, for her endless support, both technically and emotionally. She made me the biologist that I am today. I thank Dr. Jan Zimak, for changing the way I do experiments. Without his help, my splicing paper would not have been possible. I thank Dr. Sam Levine, Dr. Naoki Hida, Dr. Devon Pendlebury, Monika, Leslie, Abigail, Jay, Callie, Natalie, and Chely. All of you have helped me in some way. Thank you.

I thank my parents, Ronghui and Guanrong, and all my friends in China, South Carolina, and California. Sometimes just a text or call from them got me through the toughest days.

Last but not least, I thank my girlfriend Jaclyn for being someone I can always depend on. She has been incredibly supportive during my PhD and I am so glad she is in my life. To our cat Emma, thank you for always cheering me up.

The text of Chapter 2 is a reprint of the material as it appears in Nucleic Acids Research (England, W. E. *et al.* An atlas of posttranslational modifications on RNA binding proteins. *Nucleic Acids Res* **50**, 4329–4339 (2022).), used with the permission from Oxford University Press. The text of Chapter 5 is a reprint of the material as it appears in Oncogenesis (Zocchi, L. *et al.* Chromatin remodeling protein HELLS is critical for retinoblastoma tumor initiation and progression. *Oncogenesis 2020 9:2* **9**, 1–15 (2020).), used with the permission from Nature. The text of Chapter 6 is a reprint of the material as it appears in the journal PLOS ONE (Aboukilila, M. Y. *et al.* Identification of novel regulators of dendrite arborization using cell type-specific RNA metabolic labeling. *PLoS One* **15**, e0240386 (2020).), used with the permission from PLOS.

VITA

Jingtian 'Josh' Wang

Education:

2017-2022 Ph.D. Candidate in Molecular Biology and Biochemistry, University of California, Irvine

2012-2016 B.S. in Biology, Furman University

Experience:

2018-2022 University of California, Irvine
Advisor: Professor Robert C. Spitale
Developed a splicing-dependent cancer cell targeting gene therapy.
Developed a doxycycline-inducible promoter with undetectable transgene leakiness.

2016-2017 University of South Carolina School of Medicine
Advisor: Professor Norma Frizzell
Investigated the effect of DMF on Ndufs4 Knockout mice behavior as a model for Leigh Syndrome treatment.

2015 Furman University
Advisor: Professor Min-Ken Liao
Identified Rickettsia sp. bacteria carried by ticks in the Southeastern United States.

2014 Furman University
Advisor: Professor Renee Chosed
Investigated the roles of the accessory proteins that make up the COMPASS protein complex in yeast.

Publications:

6. (In preparation) Nguyen K., **Wang J.**, Zimak J., Spitale RC. (2022). Development of a doxycycline-inducible promoter with undetectable transgene leakiness.
5. (In preparation) **Wang J.**, Nguyen K., Zimak J., Spitale RC. (2022). Development of a splicing-dependent cancer cell targeting gene therapy. *ACS Synth. Biol.*
4. England W., **Wang J.**, Chen S., Baldi P., Flynn R., Spitale RC. (2022). An Atlas of Posttranslational Modifications on RNA Binding Proteins. *Nucleic Acids Res.*
3. Aboukilila MY., Sami JD., **Wang J.**, England W., Spitale RC., Cleary MD. (2020). Identification of novel regulators of dendrite arborization using cell type-specific RNA metabolic labeling. *PLOS One*
2. Zocchi L., Mehta A., Wu SC., Wu J., Gu Y., **Wang J.**, Suh S., Spitale RC., Benavente CA. (2020). Chromatin remodeling protein HELLS is critical for retinoblastoma tumor initiation and progression. *Oncogenesis*

1. Piroli GG., Manuel AM., Patel T., Walla MD., Shi L., Lanci SA., **Wang J.**, Galloway A., Ortinski PI., Smith DS., Frizzell N. (2019). Identification of Novel Protein Targets of Dimethyl Fumarate Modification in Neurons and Astrocytes Reveals Actions Independent of Nrf2 Stabilization. *Mol Cell Proteomics*

Teaching Experience:

Teaching Assistant, Department of Molecular Biology and Biochemistry, University of California, Irvine
Biochemistry Laboratory (Winter 2019)
Biochemistry (Winter 2019 and Winter 2020)

Technical Skills:

Experimental:	tissue culturing, molecular cloning, flow cytometry, protein extraction, western blotting, DNA/RNA extraction, gel electrophoresis, mouse handling
Analytical:	Python, R, Linux Shell, SQL, RNA-seq analysis

ABSTRACT OF THE DISSERTATION

Improvements on the Control of Gene Expression in Cancer Gene Therapies

by

Jingtian “Josh” Wang

Doctor of Philosophy in Molecular Biology and Biochemistry

University of California, Irvine, 2022

Professor Robert C. Spitale, Chair

A cancer gene therapy involves the expression of a foreign gene in cancer cells. However, the current methods to control the transgene expression are imperfect, with leaky expression usually found in normal cells. This leakiness can lead to harmful side effects in patients, especially if the transgene encodes a toxic protein. Herein, my studies aim to invent, as well as improve upon regulatory tools to control transgene expression in cancer gene therapies. Compared to the traditional promoter-based or cell surface marker-based approaches, my approach targets an intracellular protein defect caused by several mutations that are prevalent in cancers. Additionally, I improved the inducible gene expression system, further eliminating its background leakiness. Together, my studies offer novel cell-specific targeting and temporal control options, expanding the current gene therapy toolbox.

Chapter 1: Background

1.1 Introduction

Cancer and Cancer Stem Cells

According to the World Health Organization (WHO), cancer accounted for nearly one in six deaths worldwide. CDC projection shows that annual number of cancer cases in the US will increase by almost 50% from 1.5 million in 2015to 2.3 million in 2050, as a result of the aging US population[1].

Cancer is caused by the accumulation of genetic or epigenetic alterations of the genome. The resulting phenotypes of human cancers are extremely complex and dynamic and often comprise the following: sustained proliferative signaling, evasion of growth suppressors, ability to induce and access vasculature, reprogrammed cellular metabolism, and ability to avoid immune destruction[2]. In recent years, the existence of cancer stem cells (CSCs) has become increasingly accepted[2]. CSCs are a subpopulation of cancer cells that display enhanced self-renewal and tumorigenic capacity. Because of CSCs' slow proliferation and lack of typical cancer surface markers, they are largely resistant to and become enriched after traditional chemotherapies and targeted therapies[3]. Recent work suggests that after chemotherapies, CSCs may enter a dormant state that leads to cancer recurrence[4,5] As such, there is an urgent need for novel cancer therapies that specifically target cancer cells independent of the traditional hallmarks of cancer.

Cancer gene therapies

Thus far, 65% of all gene therapy trials have been aimed at treating cancer[6]. The very first commercialized cancer gene therapy, Gendicine, which dates back to 2003, is a recombinant adenovirus engineered to express wild type p53 to treat head and neck

squamous cell cancer[7]. Since entering the market, Gendicine has been administered to more than 30,000 patients, demonstrating superior efficacy and safety compared to standard therapy[8]; however, Gendicine has no targeting mechanism and exclusively treats p53-responsive cancers. The overwhelming majority of gene therapies are still in phase I and phase I/II of clinical trials[6]. The biggest challenge of cancer gene therapies is treating patients in a cancer cell-specific manner without inflicting damage on the normal tissues[9]. Currently, the specificity of cancer gene therapies is primarily achieved in two ways: cancer-specific promoters and targeted delivery[10].

Tissue/cancer-specific promoters (TCP) can be used for targeted transgene expression, however, TCPs such as hTERT are also active in germline cells and stem cells, inflicting adverse effects on the patients[11]. Furthermore, TCPs often require enhancers, introns and other elements to sustain expression at therapeutic levels[12]. Hybrid promoters have been used to achieve both the specificity and expression level, but they further increase the complexity and size of the cargo[13].

Targeting ligands have been coupled to or genetically incorporated to the vector capsid to improve the specificity of the gene therapy[14]. However, this specificity is reliant on the expression of surface receptors on the cancer cells. Current cancer-targeted therapeutic approaches rely on the identity of cell-type specific surface markers that enable selective drug delivery, but many of these receptors are also expressed on the surfaces of normal and germ cells[15]. Intracellular metabolic and gene expression profiles in cancer cells are instead drastically different from normal cells[16]. Moreover, cancer stem cell surface markers are also found on normal stem cells[17].

Cytotoxic proteins for cancer therapies

Proteins are highly modular and flexible macromolecules, allowing then to be synthetically engineered to achieve diverse functionalities or be combined with other protein species based on clinical needs[18]. Many proteins exhibiting cytotoxicity have been used in recent years as anticancer drugs. Toxins from scorpions, plants and bacteria have been rich sources of protein-based anticancer drugs[19–21]. For example, The Kid (for *killing determinant*) toxin is an RNase found in gram negative bacteria, capable of cleaving mRNAs at 5′-UUACU-3′ sites[22]. It has been found that Kid toxin induces cell death in eukaryotic cells[23]. Interestingly, Kid has an antidote protein, Kis (for *killing suppressor*), that neutralizes the RNase activity of Kid. In previous studies, the co-expression and subsequent silencing of Kis are required for Kid to exhibit its cytotoxicity in eukaryotic cells[23]. For the purpose of adeno-associated virus (AAV) packaging for gene therapy delivery, the current method to express the Kid toxin is cumbersome and leaves room for improvement. In addition, the Diphtheria Toxin segment A (DTA) has been extensively studied and used in clinical trials[24]. It catalyzes the ADP-ribosylation of the mammalian elongation factor EF2, leading to the inhibition of protein synthesis. DTA is extremely toxic: only one molecule of DTA is sufficient to kill a cell[25]. Therefore, it is imperative to design a stringent regulation system to tightly control the expression of DTA.

Developing a novel splicing-based cancer gene therapy

To further improve the safety and effectiveness of cancer gene therapies, we set out to search for a novel cancer-responsive gene regulatory element. Particularly, we were interested in RNA binding proteins (RBPs), as RBPs are constantly in contact with RNA molecules during their lifetimes, closely regulating the splicing, transport, expression, and

turnover of RNAs[26,27]. RBPs therefore are a great potential exploit to regulate the transgene expression in a gene therapy.

In our recent bioinformatic analysis on RBPs, we identified SF3B1 (*Splicing Factor 3B Subunit 1A*) as the most highly mutated residue in RBPs in breast cancer, with ~1% of the cancer cases harboring the mutation[28]. In addition to K700E in breast cancer, we also reported R625H or R625C in 17.5% of uveal melanoma.

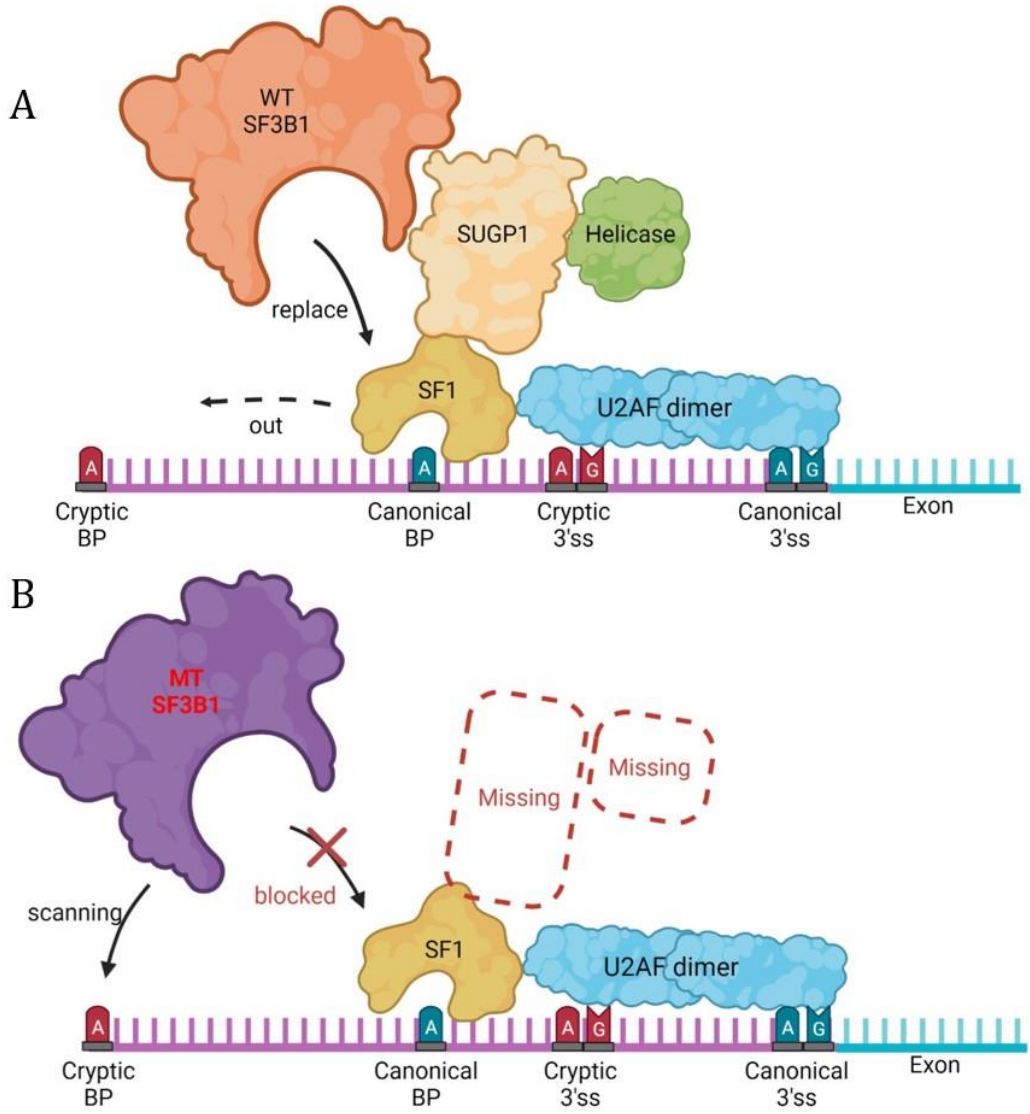


Figure 1-1. Mechanism of action of 3'ss selection by WT or K700E MT SF3B1 A. The WT SF3B1 can recruit SUGP1 to the 3'ss and uses the canonical BP and 3'ss. **B.** In cancers

where SF3B1 is mutated at K700 or other key residues, the interaction between SF3B1 and SUGP1 is disrupted. SF3B1 uses the cryptic BP and 3'ss.

We became particularly interested in this mutated protein because SF3B1 is a core component of the U2 small nuclear ribonucleoprotein (snRNP) complex of the spliceosome and is involved in 3'-splice site (3'ss) recognition during RNA splicing (Figure 1-1)[29,30]. During the 3'ss recognition, the U2AF heterodimer identifies and binds to the poly-pyrimidine tract and the AG dinucleotide of the 3'ss, followed by the binding of SF1 (*splicing factor 1*) to the branchpoint (BP). SUGP1 (*SURP and G-patch domain containing 1*) is then recruited to the 3' ss via interaction with SF3B1. SUGP1 then recruits an unknown helicase which uncouples SF1 from the BP. This allows SF3B1 to replace SF1 and occupy the BP. As a result, the canonical BP and 3'ss are used for splicing (Figure 1-1A). However, when SF3B1's key residues are mutated (e.g. K700E), the interaction between SF3B1 and SUGP1 are disrupted, and SF3B1 is no longer able to recruit SUGP1, and subsequently, the helicase, to the 3'ss. Consequently, the MT SF3B1 cannot displace SF1 from the BP, and is forced to keep scanning upstream for a suboptimal BP. As a result, the cryptic 3'ss is used and a longer mRNA transcript is produced (Figure 1-1B).

In Chapter 2 of my dissertation, I detail the exploratory data analysis on post translational modifications (PTM) on RNA binding proteins (RBP) that I was involved in. Specifically, I look at hotspot mutations of the PTM sites in cancers, and the K700E mutation in SF3B1 is one of the most common PTM mutations in RBPs in cancers. SF3B1 K700E is a pro-oncogenic mutation that leads to alternative splicing by altering branchpoint usage[31,32] .

In Chapter 3 of my dissertation, I exploit the splicing defect caused by SF3B1 K700E to develop a mutation-specific cancer gene therapy. It was found that several RNA transcripts mis-spliced by mutant SF3B1 could be used for mutation-specific protein expression. The ORAI2 transcript was used to express the Kid toxin in cells harboring the splicing mutation, killing them in a mutation-specific manner. Furthermore, the size of the ORAI2 splicing cassette was minimized, and it was found to be responsive to other SF3B1 mutations.

In Chapter 4 of my dissertation, I help develop an improved doxycycline-inducible gene expression system called OTetRA. This project was conceived when we were met with the difficulty cloning the Kid toxin. Because of the high toxicity of Kid, the original TetOn system is not acceptable due to its basal leakiness in its off state[33]. The OTetRA system greatly lowers the background signal while maintaining the induced signal.

In Chapter 5 of my dissertation, I help elucidate the roles of chromatin remodeling protein HELLS in retinoblastoma by conducting bioinformatic analysis. Retinoblastoma is an aggressive childhood cancer driven by epigenetic changes but the exact mechanism of these epigenetics changes is unknown. Through the mouse model, it was found that loss of HELLS had little effect on retinal development, but drastically reduced the incidence of retinoblastoma and increased overall survival.

In Chapter 6 of my dissertation, I help identify novel regulators of dendrite arborization in *Drosophila* by conducting bioinformatic analysis. It is difficult to isolate neuron transcriptomes due to the intricate microenvironments they are in. 5-ethynylcytosine (EC)-tagging was used to specifically tag neuron cells in *Drosophila* and top

genes enriched or depleted in neurons were identified via RNA-seq and experimentally validated to affect dendrite arborization.

Together, my research demonstrates that bioinformatic techniques can provide actionable information to accelerate and guide experimental biology. In particular, my data analysis helped discover a cancer-specific protein defect and this defect was then exploited to improve the specificity of cancer gene therapies. We hope this work will expand the current gene therapy toolkit and lay the groundwork for future gene therapies that target the biochemical defects in cancers.

1.2 References

1. Weir, H. K., Thompson, T. D., Stewart, S. L. & White, M. C. Cancer Incidence Projections in the United States Between 2015 and 2050. *Preventing Chronic Disease* **18**, 1–8 (2021).
2. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discovery* **12**, 31–46 (2022).
3. Pavet, V., Portal, M. M., Moulin, J. C., Herbrecht, R. & Gronemeyer, H. Towards novel paradigms for cancer therapy. *Oncogene* **30**, 1–20 (2011).
4. Kurtova, A. v *et al.* Blocking PGE 2-induced tumour repopulation abrogates bladder cancer chemoresistance. (2014) doi:10.1038/nature14034.
5. Lu, H. *et al.* Chemotherapy-Induced Ca 2+ Release Stimulates Breast Cancer Stem Cell Enrichment. *Cell Rep* **18**, 1946–1957 (2017).
6. Ginn, S. L., Amaya, A. K., Alexander, I. E., Edelstein, M. & Abedi, M. R. Gene therapy clinical trials worldwide to 2017: An update. *J Gene Med* **20**, (2018).
7. Raty, J., Pikkarainen, J., Wirth, T. & Yla-Herttuala, S. Gene therapy: the first approved gene-based medicines, molecular mechanisms and clinical indications. *Curr Mol Pharmacol* **1**, 13–23 (2008).
8. Zhang, W. W. *et al.* The First Approved Gene Therapy Product for Cancer Ad-p53 (Gendicine): 12 Years in the Clinic. *https://home.liebertpub.com/hum* **29**, 160–179 (2018).
9. Das, S. K. *et al.* Gene Therapies for Cancer: Strategies, Challenges and Successes. *J. Cell. Physiol* **230**, 259–271 (2015).
10. Chen, C. *et al.* Promoter-Operating Targeted Expression of Gene Therapy in Cancer: Current Stage and Prospect. *Molecular Therapy. Nucleic Acids* **11**, 508 (2018).
11. Dorer, D. E. & Nettelbeck, D. M. Targeting cancer by transcriptional control in cancer gene therapy and viral oncolysis. *Advanced Drug Delivery Reviews* **61**, 554–571 (2009).

12. ED, P., SA, N., AH, B. & SJ, W. Promoters and control elements: designing expression cassettes for gene therapy. *Curr Gene Ther* **4**, 89–113 (2004).

13. Kalinichenko, S. v., Shepelev, M. v., Vikhreva, P. N. & Korobko, I. v. A Novel Hybrid Promoter ARE-hTERT for Cancer Gene Therapy. *Acta Naturae* **9**, 66 (2017).

14. Waehler, R., Russell, S. J. & Curiel, D. T. Engineering targeted viral vectors for gene therapy. *Nature Reviews Genetics 2007 8:8* **8**, 573–587 (2007).

15. Payne, K. K., Bear, H. D. & Manjili, M. H. Adoptive cellular therapy of cancer: Exploring innate and adaptive cellular crosstalk to improve anti-tumor efficacy. *Future Oncology* Preprint at https://doi.org/10.2217/fon.14.97 (2014).

16. Pérez-Mancera, P. A., Young, A. R. J. & Narita, M. Inside and out: The activities of senescence in cancer. *Nature Reviews Cancer* Preprint at https://doi.org/10.1038/nrc3773 (2014).

17. Kim, W. T. & Ryu, C. J. Cancer stem cell surface markers on normal stem cells. *BMB Reports* **50**, 285 (2017).

18. Thi-Quynh Duong, M., Qin, Y., You, S.-H. & Min, J.-J. Bacteria-cancer interactions: bacteria-based cancer therapy. *Experimental & Molecular Medicine* doi:10.1038/s12276-019-0297-0.

19. Ortiz, E., Gurrola, G. B., Schwartz, E. F. & Possani, L. D. Scorpion venom components as potential candidates for drug development. *Toxicon* **93**, 125–135 (2015).

20. Bolognesi, A., Bortolotti, M., Maiello, S., Battelli, M. G. & Polito, L. Ribosome-Inactivating Proteins from Plants: A Historical Overview. *Molecules 2016, Vol. 21, Page 1627* **21**, 1627 (2016).

21. Michl, P. & Gress, T. Bacteria and Bacterial Toxins as Therapeutic Agents for Solid Tumors. *Current Cancer Drug Targets* **4**, 689–702 (2005).

22. de la Cueva-Méndez, G., Mills, A. D., Clay-Farrace, L., Diaz-Orejas, R. & Laskey, R. A. Regulatable killing of eukaryotic cells by the prokaryotic proteins Kid and Kis. *EMBO J* **22**, 246–251 (2003).

23. Turnbull, A., Bermejo-Rodríguez, C., Preston, M. A., Garrido-Barros, M. & de la Cueva-Me, G. Targeted Cancer Cell Killing by Highly Selective miRNA-Triggered Activation of a Prokaryotic Toxin−Antitoxin System. (2019) doi:10.1021/acssynbio.9b00172.

24. Zahaf, N.-I. & Schmidt, G. toxins Bacterial Toxins for Cancer Therapy. doi:10.3390/toxins9080236.

25. Yamaizumi, M., Mekada, E., Uchida, T. & Okada, Y. One molecule of diphtheria toxin fragment A introduced into a cell can kill the cell. *Cell* **15**, 245–250 (1978).

26. Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* **19**, 327–341 (2018).

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

28. England, W. E. *et al.* An atlas of posttranslational modifications on RNA binding proteins. *Nucleic Acids Research* **50**, 4329–4339 (2022).

29. Zhang, J. & Manley, J. L. Misregulation of pre-mRNA alternative splicing in cancer. *Cancer Discovery* **3**, 1228–1237 (2013).

30. Zhang, J. *et al.* Disease-Causing Mutations in SF3B1 Alter Splicing by Disrupting Interaction with SUGP1. *Molecular Cell* (2019) doi:10.1016/j.molcel.2019.07.017.

31. Liu, B. *et al.* Mutant SF3B1 promotes AKT- and NF-κB-driven mammary tumorigenesis. *J Clin Invest* **131**, (2021).

32. Alsafadi, S. *et al.* Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nature Communications 2016 7:1* **7**, 1–12 (2016).

33. Goverdhana, S. *et al.* Regulatable Gene Expression Systems for Gene Therapy Applications: Progress and Future Challenges. *Mol Ther* **12**, 189 (2005).

Chapter 2: An atlas of posttranslational modifications on RNA binding proteins

Publication note

This paper was originally published in the Nucleic Acids Research.

Whitney E. England, Jingtian Wang, Siwei Chen, Pierre Baldi, Ryan A. Flynn, and Robert C. Spitale. An atlas of posttranslational modifications on RNA binding proteins. Nucleic Acids Res. **50**, 4329-4339 (2022).

2.1 Contribution Statement

Members of the Spitale Lab helped collate the raw post translational modification data from numerous literature sources. Whitney England was the main analyst for this paper. Jingtian Wang queried the TCGA mutational data, conducted the GO-term analysis of the RBPs, and contributed to the analyses in Figure 2-2, 2-3, 2-4. Siwei Chen built the web interface to the database. Ryan Flynn conducted the mass spectrometry experiments. Pierre Baldi supervised the website and database building. Robert Spitale supervised the rest of the analysis and did the writing.

2.2 Abstract

RNA structure and function are intimately tied to RNA binding protein recognition and regulation. Posttranslational modifications are chemical modifications which can control protein biology. The role of PTMs in the regulation RNA-binding proteins (RBPs) is not well understood, in part due to a lacking analysis of PTM deposition on RBPs. Herein,

we present an analysis of posttranslational modifications (PTMs) on RNA binding proteins (RBPs; a PTM RBP Atlas). We curate published datasets and primary literature to understand the landscape of PTMs and use protein–protein interaction data to understand and potentially provide a framework for understanding which enzymes are controlling PTM deposition and removal on the RBP landscape. Intersection of our data with The Cancer Genome Atlas also provides researchers understanding of mutations that would alter PTM deposition. Additional characterization of the RNA–protein interface provided from in-cell UV crosslinking experiments provides a framework for hypotheses about which PTMs could be regulating RNA binding and thus RBP function. Finally, we provide an online database for our data that is easy to use for the community. It is our hope that our efforts will provide researchers with an invaluable tool to test the function of PTMs controlling RBP function and thus RNA biology.

2.3 Introduction

RNA molecules have emerged as critical players in nearly every biological process, from cell division to the regulation of chromatin state and gene expression[1,2]. Many diseases, ranging from cancer[3,4] to neurological disorders[5], can be traced back to faulty RNA structure, function, or posttranscriptional regulation. Signaling hubs inside cells can also be regulated by RNA molecules, further suggesting more dynamic processes can utilize RNA-centered signaling to control specific outputs to regulate cell fate and disease onset[6]. As such, understanding the molecular mechanisms utilized by cells to control RNA expression and fate are critical to creating a complete picture of how RNA molecules are controlled to impart their many biological outputs.

During their lifetimes, RNA molecules are constantly in contact with RBPs[7,8]. RNA-RBP interactions control and regulate alternative splicing, RNA export and RNA decay. Our expanded understanding of the catalog of such interactions suggests each of these processes can be dynamic with the catalog of RBP–RNA interactions changing to toggle RNA expression for desired cellular output[9–11]. Almost all characterization of RBP–RNA interactions has been focused on identifying the protein–RNA interface and have limited focus on how the RBPs themselves are regulated to control RNA fate[12]. There is much work to be done to fully understand how RBPs select their RNA targets in cells, how signaling networks can control RBP selection, how RBPs themselves are regulated, and how such regulation can lead to altered RNA expression.

Protein molecules are subject to a high level of regulation through chemical modifications termed posttranslational modifications (PTMs)[13]. PTMs are known to regulate protein structure, localization, decay and signaling activities[14,15]. PTMs vary widely in their chemical nature and can impart unique structural and charge differences associated with their deposition and removal. Further, proteins with evidence of multiple PTMs have been suggested to be more involved in disease biology than poorly modified proteins[16]. PTM marks are dynamic and can either be deposited through enzyme-regulated processes or changes in cellular chemistry (i.e. oxidation). While PTMs have been studied for quite some time through the lens of protein regulation, our understanding and characterization of RBP PTMs is quite limited with only a few examples well characterized to date.

There has been some important headway in characterizing PTM modifications and their role in regulating RNA binding proteins. The limited cases illustrate just how

impactful changes to the chemical composition of RBPs can be[17]. For example, in the interferon (IFN)-γ-activated inhibitor of translation (GAIT) system[18,19], IFN-γ induces formation of a multiprotein GAIT complex that binds structural GAIT elements in the 3' untranslated regions (UTR) of multiple inflammation-related mRNAs. This complex is recruited and regulated by IFN-γ-induced kinases that phosphorylate RNA binding proteins to control mRNA translation; PTM regulation of RBPs in this case are critical for regulating mRNA interactions and thus the IFN-γ translation response. In another example, there is growing evidence that PTMs can regulate phase separation[20]. In the context of tyrosine phosphorylation, it has been shown that tyrosine residues in RBPs can have weak interactions driving phase separation; however, once tyrosine residues are phosphorylated phase separation breaks down due to coulombic repulsion between the negatively charged residues. Overall, such examples highlight the importance of PTM regulation for RBPs, but likely represent only a small fraction of the total examples.

Expanding the suite of known and addressable of PTMs on RBPs requires rigorous curation and interrogation of existing proteome-scale mass spectrometry data coupled to other facets of RNA–protein interaction biology; together this could serve as a platform for novel hypotheses related to functional types and sites of PTMs on RBPs. Herein, we perform a thorough analysis of experimentally derived PTMs. Using recent large-scale datasets for the identification of RNA binding proteins, we investigate PTMs on the set of known RBPs (RBP-ome). We intersect this dataset with published protein–protein interaction databases to provide potential models for how PTMs are deposited and removed by cellular enzymes. We utilize mass spectrometry data on in-cell crosslinked RNA–protein interactions, that are close to identified RNA–protein interfaces, to provide

the community with PTMs that have the potential to regulate RNA–protein binding. We investigate the potential disease relevance of PTM sites by intersecting our data with The Cancer Genome Atlas. We catalog novel PTM sites that are mutated in these datasets and provide experimentally-derived evidence that a key and well known RBP is mutated in cancer. Finally, we provide the community with an online searchable database (http://PTM-RBP-ATLAS.igb.uci.edu) where this data can be easily gathered by the community for inspection to help with their research programs. Overall, we envision the data generated herein will serve as a powerful analysis for the growing labs interested in understanding how RBPs control the fate of RNA molecules.

2.4 Materials and Methods

Aggregation of PTM sites, RNA-binding proteins, and protein–protein interactions

Known sites of post-translational modifications in human proteins were downloaded from 13 public PTM databases covering a variety of PTM types: dbSNO, HPRD, MeMo, OGlycBase, PHOSIDA, PhosphoELM, Phosphositeplus, PTM_SD, RedoxDB, SwissPalm, Swiss-Prot, SysPTM and UbiProt[21–33]. Only experimentally determined PTMs were used; computational predictions were excluded. Additional PTM datasets not present in these databases were identified via literature search (Supplementary Table 2-S1). All protein identifiers were converted to UniProt identifiers, and proteins present in SwissProt[31] were retained for analysis.

A list of RNA-binding proteins was derived from three publications containing extensive lists of experimentally-determined RBPs[9,10,34,35].

Protein-protein interactions were collected from 8 publicly available databases: BIOGRID, CORUM, DIP, ELM, hu.MAP, IntAct, MINT and STRING[36–43].

PTM site conservation analysis

PhastCons conservation scores for a multiple alignment of 99 vertebrate genomes to the human genome (phastCons100way) were obtained from the UCSC Genome Browser [citation]. For each RNA binding protein, phastCons scores from 100 random positions were selected from its exons for background in comparison to scores at PTM positions.

eCLIP mRNA binding data

eCLIP narrowPeak files from 92 RNA-binding proteins in K562 cells was downloaded from the ENCODE website (https://www.encodeproject.org/). eCLIP peak regions overlapping genes were identified using[44] bedtools to intersect eCLIP peaks with GRCh38 Gencode v24 annotations. Hits where at 50% of the peak overlaps the gene on the correct strand in both eCLIP replicates were considered instances of the RBP binding that gene.

Cancer mutation data

Position, consequence, distribution, and frequency data were obtained for each mutation identified in The Cancer Genome Atlas and cross-referenced with identified PTM positions.

UV mass spec sites

RNA–protein crosslink data was retrieved from four publications[45–48] . As these studies employed varying techniques with differing resolutions, we adjusted all crosslink data to a 21-position range centered on the reported crosslink range. Ranges shorter than 21 positions were expanded, and those larger were truncated. PTM sites within these

crosslink ranges were identified and classified by their distance from the center of the range.

2.5 Results

Distribution and conservation of PTMs in the human RBP-ome

Our goal in this manuscript is to better understand the landscape of PTMs on RNA binding proteins, what enzymes are responsible for their deposition, what mutations in disease relevance could alter the PTM landscape, and if PTM sites are near RNA–protein interfaces. These data points, which are important to catalog to better understand the regulation of RBPs and thus RNA biology, are represented schematically in Figure 2-1.

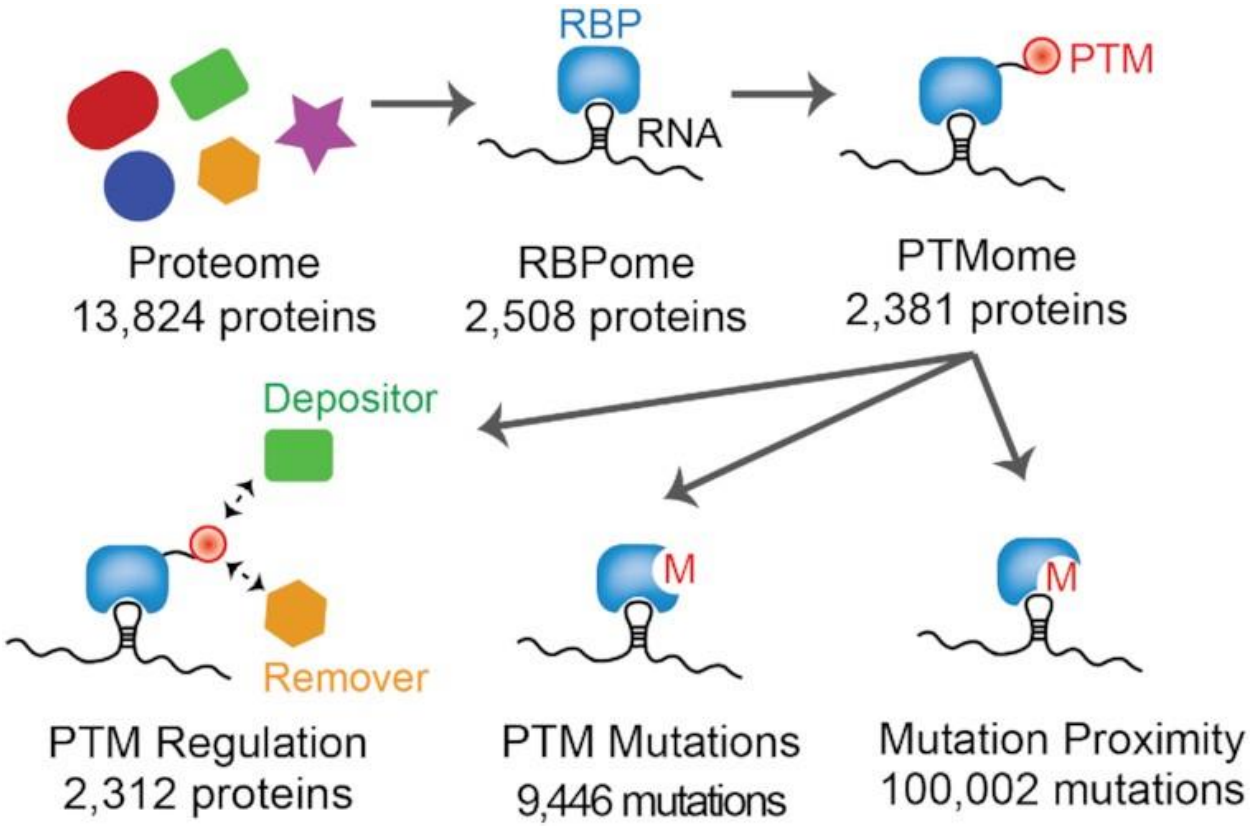


Figure 2-1. Schematic of PTM identification in the human proteome and RBPome. Clockwise from top left: count of PTM-containing proteins, total number of identified RNA binding proteins, count of RNA binding proteins with PTMs, number of cancer mutations within 10 residues of a PTM, number of cancer mutations at a PTM-bearing residue, number of modified proteins which interact with enzymes capable of depositing or removing the appropriate PTM.

We first worked to collate publicly available PTM sites from curated datasets and an extensive literature search. Experimentally determined PTM sites were aggregated from a set of 13 publicly available PTM databases[21–33] and an extensive literature search (Supplementary Table 2-S1). We found modified sites were widespread, with 13 824 of the 20 239 (68%) human proteins present in SwissProt[31] containing one or more PTM sites. The union of these datasets resulted in our input set (Figure 2-2A)

To classify proteins as RNA-binding (the RBP-ome), we combined three existing sets of annotated RBPs from[9,10,34,35] into a non-redundant set of 2508 RBPs. Of these RBPs the vast majority are highly modified, with 2381 RBPs (95%) modified at a minimum of one site (Supplementary Table 2-S2). This is a significantly higher proportion of modification than in the full set of human proteins (chi-squared test, $P = 9.87 \times 10–170$). Phosphorylation was the most common modification, though acetylation, ubiquitination, and methylation were also present in over half of RBPs (Figure 2-2A). A total of 19 individual PTMs were significantly more common in RBPs than in all proteins, including acetylation, caspase, citrullination, glutathionylation, glycosylation, hydroxylation, methylation, myristoylation, N6-malonyllysine, nitrosylation, oxidation, palmitoylation,

phosphorylation, sulfenylation, succinylation, sulfenic acid, sumoylation, ubiquitination, and Zn-Cys modifications (Chi-squared tests with Bonferroni correction, $P < 0.05$). These data highlight the potential complexity of PTM sites on RBPs.

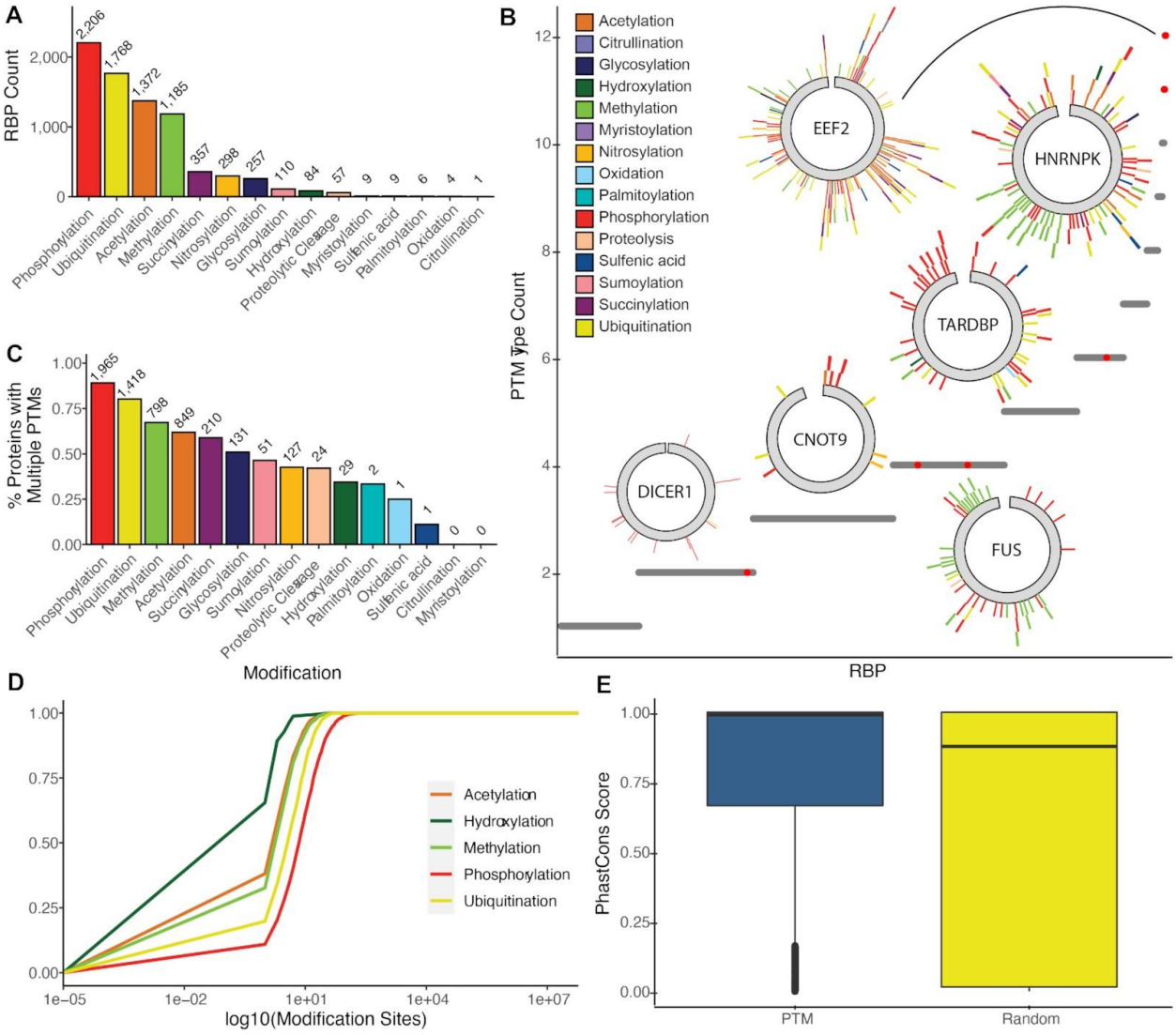


Figure 2-2. Sites of post-translational modification in RNA binding proteins. A. Count of RBPs with each PTM type. **B**. Counts of different PTM types per RBP. Location and PTM type are indicated on circular inset plots for six example RBPs with low and high PTM type counts. **C**. Proportion of modified RBPs which contain multiple PTMs of the same type. **D**. Cumulative distribution of modified sites per protein for five common PTM types. **E**.

Boxplot of PhastCons 100-way conservation scores for nucleotides at PTM-modified residues in RBPs versus randomly-selected positions from the same protein set. Higher scores represent more conserved positions based on genome-wide alignments of human and 99 other vertebrates.

We next worked to analyze the complexity of PTMs observed on the same RBP. Most RBPs had multiple types of PTMs (Figure 2-2B), with 2074 of the 2381 modified proteins having multiple modification types (87%). RBPs have significantly more PTMs per amino acid than all proteins (0.1992 versus 0.0181 modifications per amino acid, Student's *t*, $P = 4.93 \times 10–60$). The most common set of modifications was phosphorylation, ubiquitination, acetylation, and methylation (Supplementary Figure 2-S1), followed by phosphorylation alone or in combination with one or more of the aforementioned modifications. In addition, most RBPs (2202) have multiple instances of a PTM type (Figure 2-2C and D). The vast majority of phosphorylated and ubiquitinated proteins are multiply modified (88% and 81%, respectively), while multiple methylation and acetylation sites are less frequent (58% and 48%, respectively). This data may be the result of a bias in generated data as phosphorylation profiling is the most common PTM analyzed in the literature; however, the complexity of PTM classes and their dynamic nature is likely to contribute to RBP function in development and disease, as has been observed in other protein classes[13,14,49].

While most proteins have a small number of repeated modifications, there is a long tail of highly modified proteins (Figure 2-2D). Up to 65 acetylation, 75 methylation and 147 ubiquitination sites were observed in proteins in this dataset. Phosphorylation in particular

exhibits many modifications per protein, with approximately 1% containing at least 100 phosphorylated sites. The most phosphorylated protein, serine-rich SRRM2, contained 587 sites among its 2752 residues, consistent with its known biology to be regulated by dynamic serine phosphorylation[50–52].

The presence of PTM sites is often regarded to suggest conserved function. It has been observed that PTM sites can be important for regulation of protein–protein interactions and regulatory hot spots for signaling in other classes of functional proteins[53]. We worked to better understand if such high conservation would be observed in RNA binding proteins as well. Examining the conservation of modified sites, we found that within an RBP, PTM sites were more conserved across vertebrates than randomly-selected sites in 74.5% of proteins (Figure 2-2E). Across proteins of all types, 74.8% of modified sites were more conserved, indicating that this high level of conservation is not specific to RNA-binding proteins. For 38.6% of RBPs and 36.3% of all proteins, the mean conservation score is >0.9, indicating a high probability that they lie in elements highly conserved across species, and the higher conservation of PTM sites in RBPs is statistically significant (Kolmogorov–Smirnov test, $P < 2e{-}16$). This high evolutionary conservation of many PTM sites indicates they may have important functions that remain intact through evolutionary time.

Overall, our analysis of PTM sites from curated datasets and published manuscripts suggests that RBPs have highly complex experimentally observed PTM profiles, that RBPs have many different types of PTMs, and that the sites of PTM deposition are highly conserved. These resources provide a large dataset focused on RNA binding proteins for analysis by the community.

Identification of probable PTM-depositing enzymes

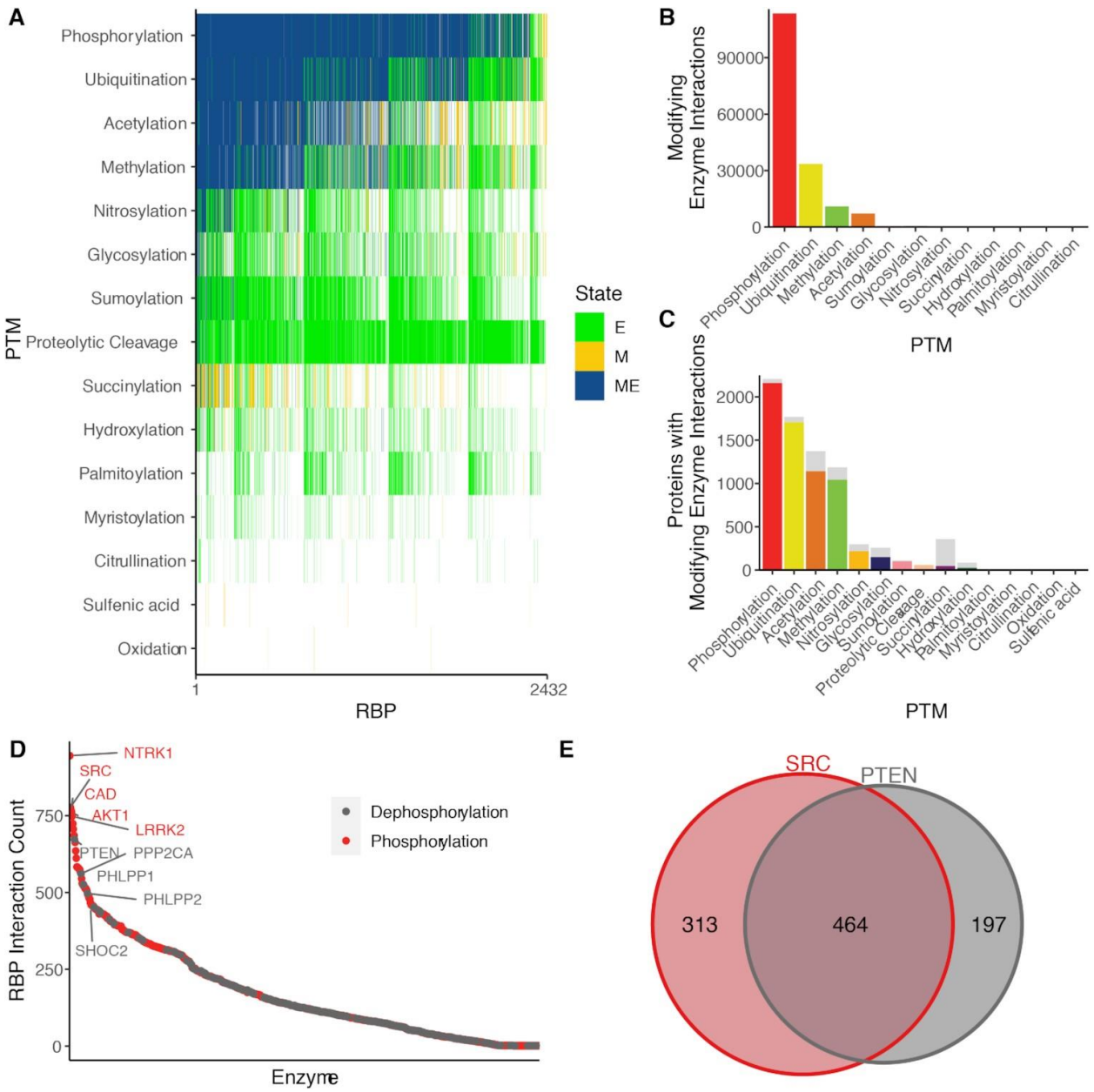


Figure 2-3. Interaction of modified RNA binding proteins with PTM-associated enzymes. A. Heatmap of interaction types for each RBP and PTM type. 'ME' indicates that the RBP contains the indicated PTM and also interacts with an enzyme known to regulate that modification. 'M' indicates a protein that contains the PTM, but is not known to interact with an enzyme which regulates it. 'E' indicated proteins that interact with

enzymes which regulate the PTM, but do not contain that PTM. **B**. Total count of potentially modifying interactions per PTM type. Each enzyme interacting with a modified protein is counted as a separate interaction. **C**. Count of RBPs with modifying interactions for each PTM type. Gray bars represent the total number of PTMs with the given modification. **D**. Number of RBPs which interact with each phosphorylating or dephosphorylating enzyme. **E**. Number of RBPs with interact with the kinase SRC and/or dephosphorylase PTEN.

PTM modifications on proteins are deposited, read and removed by distinct classes of enzymes that ultimately control PTM stoichiometry[53,54]. To identify the specific enzymes with may be responsible for depositing each PTM, we cross-referenced our list of RBPs with 8 protein–protein interaction (PPI) databases [36–43]. We then identified enzymes of the appropriate class which interact with RBPs carrying the corresponding PTM (from our analysis above). RBPs with known PTMs that require protein enzymes for deposition or removal, and which are known to interact with enzymes that carry out these functions, are commonly identified in these data ('ME'; Figure 2-3A; Supplementary Figure 2-S1). In addition, specific RBPs were cataloged to interact with PTM-depositing enzymatic proteins ('E'; Figure 2-3A), but we did not observe these PTMs in our catalog. This may be due in part to degradation of proteins bearing certain modifications, including proteolytic cleavage and sumoylation, making these less likely to be profiled by large-scale PTM characterization efforts[55]. Lastly, we identified a class of PTMs on RBPs were not associated with enzymes in protein–protein interaction databases ('M'; Figure 2-3A); however, for some, this can be attributed to their nature of deposition. For example, both 'sulfenic acid'

and ‘oxidation’ refer to oxidized forms of cysteine that are due to changes in the oxidation state of cells and do not require enzyme deposition[54,56].

Our approach identified tens of thousands of potentially modifying interactions, or cases in which a modified RBP interacts with an enzyme capable of making the same modification. Most modified proteins had at least one potentially modifying interaction; however, the percentage of proteins with such an interaction varied by PTM type (Figure 2-3B). For the four most common modifications (phosphorylation, ubiquitination, acetylation, and methylation), over 80% of modified RBPs have evidence for a protein–protein interaction that could control PTM stoichiometry, including >113 000 phosphorylation events. Others, like hydroxylation (29%) and succinylation (13%), had far fewer explanatory interactions. This may be due to a bias in the study of particular modifications, as kinase, methylation and acetylation signaling are well-studied for their diverse roles in regulating protein function. For many of the modifications, analysis of protein–protein interactions revealed potential enzymes that are controlling deposition. As shown in (Figure 2-3C), nearly all RBPs that have observed phosphorylations have associated enzymes that interact.

The large landscape of identified RBP-PTMs in which a potential enzyme pair was identified through analysis of protein–protein interactions implored us to look deeper at the identified enzyme-RBP pairs. We focused on phosphorylation as it is well studied and altered kinase activities are known to be associated with tumorigenesis[57]. We ranked the interacting deposit and removal enzymes (kinase and dephosphorylase, respectively). As is shown in Figure 2-3D, many of the kinase and dephosphorylase enzymes interacted with a large portion of the total RBP-ome (up to 945 RBPs for kinases and 686 for

dephosphorylases). Two notable proteins emerged from this analysis: PTEN and SRC kinase. Phosphatase and TENsin homolog deleted on chromosome 10 (PTEN) is a classical tumor suppressor gene, which possesses lipid and protein phosphatase activities. It has been implicated in numerous studies to control gene expression in tumorigenesis. SRC kinase is a non-receptor protein tyrosine kinase that transduces signals involved in the control of a variety of cellular processes and has been demonstrated to be important for the development, growth, progression, and metastasis of a number of different cancer types[58,59]. An intersection of the PTEN-SRC target sets revealed strong overlap between the RBPs that they interact with and have been demonstrated to be phosphorylated, with 464 RBPs the binding partners of both important proteins (Figure 2-3E). One such RBP stuck out as a target of both enzymes: DDX5. DDX5 is notable in this case due to its interaction with oncogenic transcriptional regulators, such as nuclear factor-κβ (NF-κβ), estrogen receptor α (ERα), β-catenin, and P53, and androgen receptor. These interactions, consistent with many other protein–protein interactors could be regulated by PTMs and their deposition/removal enzymes, of which many have been observed but are not fully studied. Furthermore, DDX5 has been described as a potential general cancer target due to its many roles described above[60,61]. Overall, our analysis of PTMs on RBPs, and the proteins that are likely responsible for their regulation is an important hypothesis generating analysis to provide a framework to the community.

Mutations in cancer at RBP PTM sites

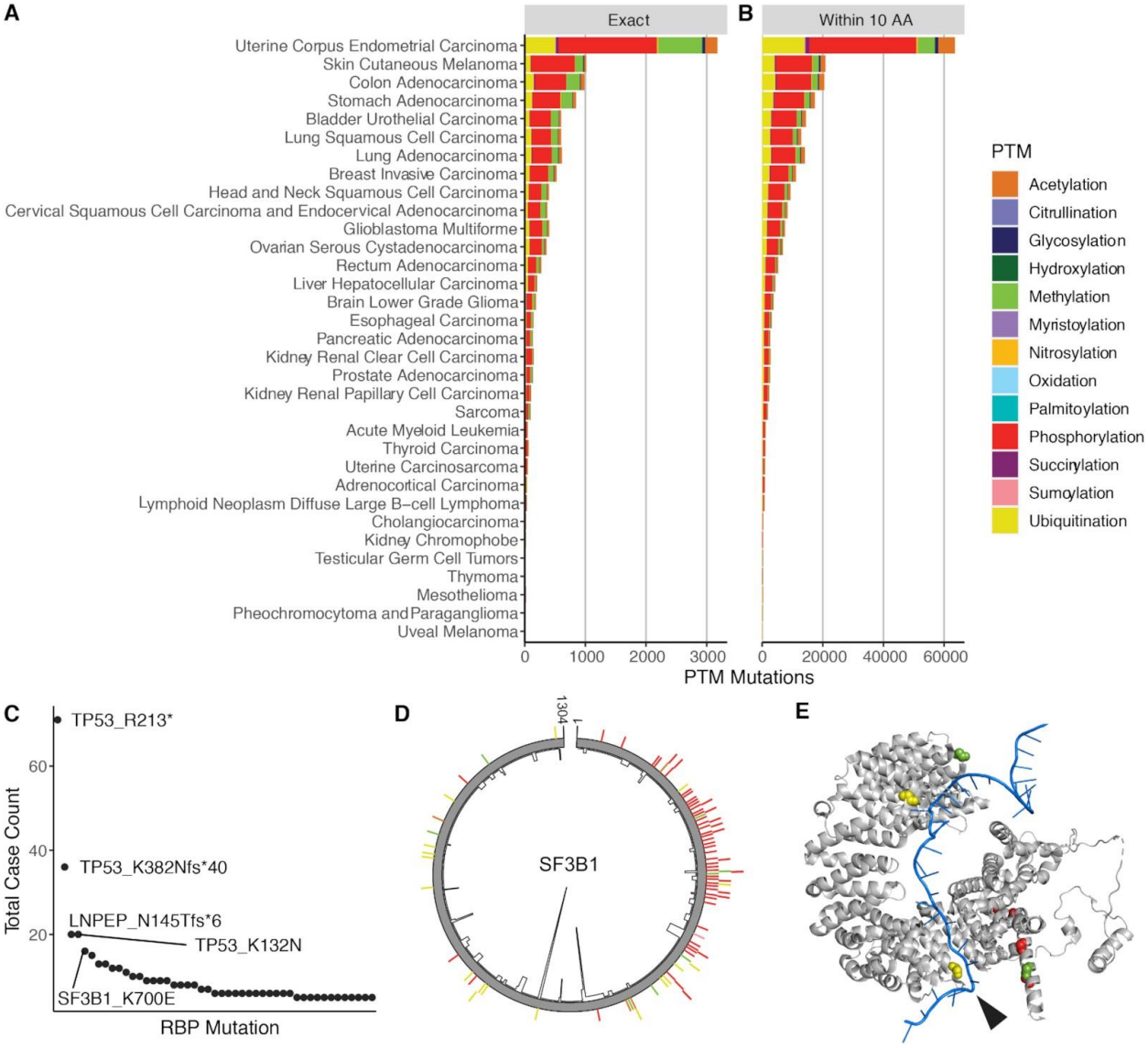


Figure 2-4. Cancer-related mutations at PTM sites. A, B. Count of mutations at **A** or within 10 residues **B** of PTM sites. **C**. Count of cancer cases with each PTM site mutation. **D**. PTMs and mutation frequencies in the splicing protein SF3B1. The inner black line shows per-position mutation frequency; the outer colored bars show PTM types and locations. **E**. Structure of SF3B1 (gray) with RNA (blue). Modified positions are colored according to their PTM type. The K700 position is indicated by the black arrow.

The known importance of regulation brought about by PTM sites on proteins, and identified examples in which mutation at a site of PTM altering the function of a protein prompted us to analyze mutational profiles in PTM sites in our RBP-focused analysis. To investigate the potential role of PTMs in human disease, we intersected our PTM data with mutation data from The Cancer Genome Atlas (https://www.cancer.gov/tcga). We identified 9446 cancer mutations that coincide with the position of categorized PTM modified sites in 1727 RBPs (Supplementary Table 2-S3) (Figure 2-4A). We also characterized mutations surrounding PTM sites, as the local structure of proteins is known to be critical for the recognition of deposition enzymes to control PTM stoichiometry[62]. From this analysis we identified 100 002 cancer mutations that fall within 10 residues of PTM modified sites in 2285 RBPs, with a similar distribution across cancer types (Figure 2-4B). The most mutation-rich cancer type was endometrioid carcinomas, which are known to be very unstable and as such much richer in mutations overall[63].

Analysis of the sites of mutations along RNA binding proteins provided greater understanding of which proteins are more enriched in mutations and how diverse the mutation sites are. As illustrated in Figure 2-4C, the most frequently observed PTM-linked mutation was at methylated position 213 in TP53, the well-known and frequently-mutated tumor suppressor. This mutation was observed in multiple patients with several cancer types, including colon, skin, breast, stomach, and uterine cancers. Consistent with p53's known role as the 'guardian of the genome', its high mutation rate is associated with many cancer types[63].

One of the most common PTM-linked mutations outside of TP53 was a mutation from lysine to glutamic acid at position 700 in SF3B1 (Figure 2-4C). This mutation is observed in multiple patients with breast carcinoma (0.9% frequency) and thymoma (1.6% frequency), as well as single patients with prostate adenocarcinoma, cutaneous melanoma, sarcoma, and acute myeloid leukemia. This lysine is shown to be ubiquitinated in our analysis of PTM sites. Plotting all the PTM sites and mutational frequencies across SF3B1 demonstrate that the K700E site is by far the most frequently mutated (Figure 2-4D). SF3B1 is involved in splicing and is the most frequently mutated splicing factor in cancer, and mutation at this position is linked to loss of splicing function[64]. As such, it seems that the altered K700E ubiquitin site is likely altering ubiquitin deposition and splicing. It has been demonstrated previously that ubiquitin can control protein–protein interactions involving splicing component assembly and that alterations in ubiquitin stoichiometry can lead to defects in RNA splicing[65,66]. The lysine at position 700 is in close proximity to the pre-mRNA in the complex (Figure 2-4E); disruption of ubiquitination by mutation to glutamic acid may influence interactions with the pre-mRNA and recognition of the branchpoint sequence, leading to increased splicing at cryptic sites[67]. The analysis provided herein provides a roadmap for testing such hypotheses as it relates to PTM sites on RBPs that are mutated in disease.

Relationship between PTM sites and the RNA–protein interface

RNA binding proteins function through their interaction with RNA molecules to regulate RNA structure and function. The presence of thousands of PTM sites prompted us to make headway in understanding the relationship between RNA binding and PTM occupancy in protein sequence space. There are now many published reports cataloging

the cross-linked sites between RNAs and proteins through UV-mediated crosslinking. These reports provide amino-acid resolution of the RNA–protein interface (Figure 2-5A)[45,47,68]. We sought to intersect these datasets with our PTM database to better understand the potential relationship between RNA binding and PTM deposition.

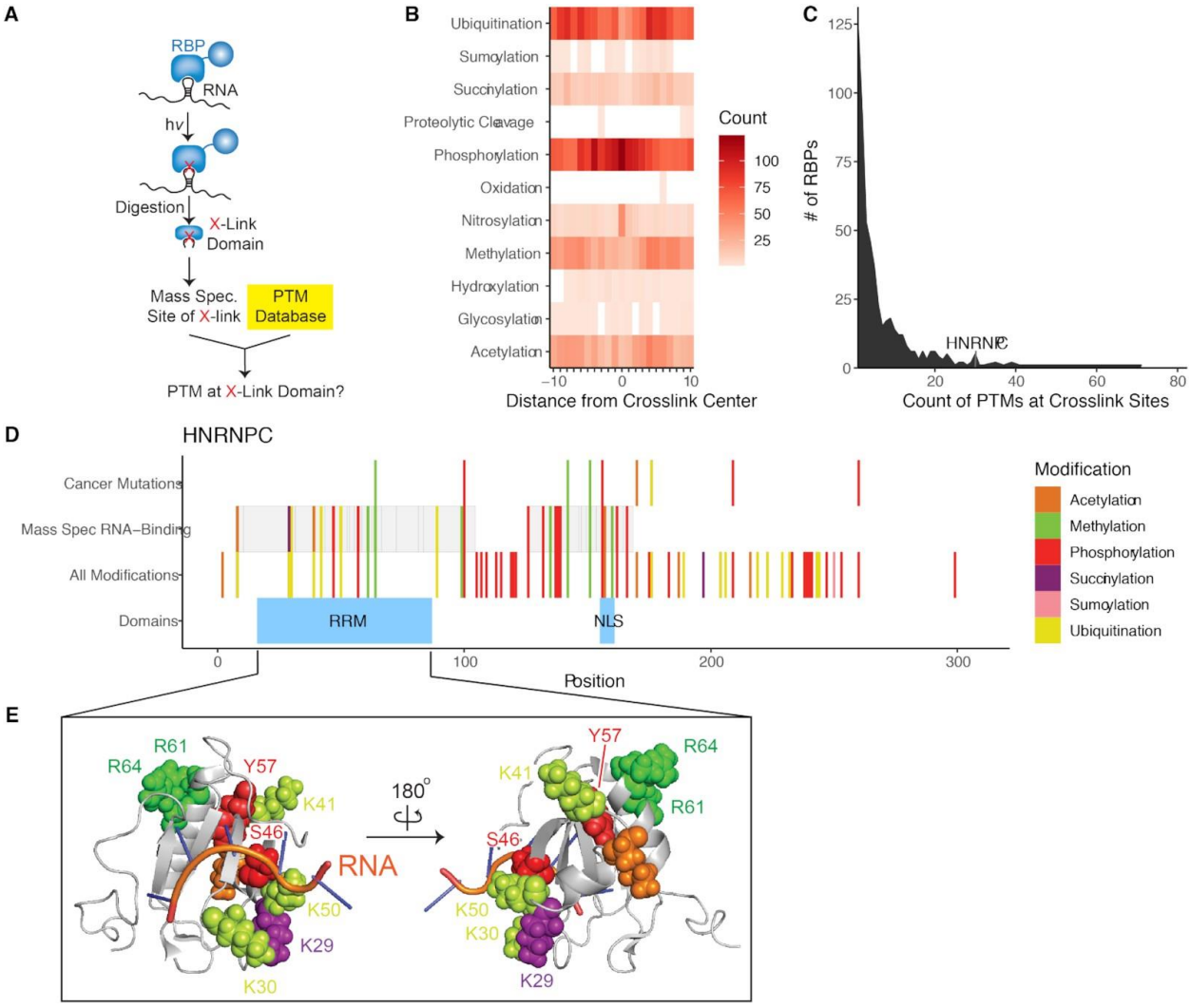


Figure 2-5. A. Schematic of identification of PTMs at crosslinked domains. **B**. Count of PTMs within 10 residues of crosslink locations. **C**. Distribution of per-RBP PTM count at crosslink sites. **D**. PTMs at crosslink and cancer mutation sites in HNRNPC. **E**. Structure of hnRNP-C. Modified positions are colored according to PTM type.

We first calculated the frequency of intersection between PTM sites and RBP–RNA crosslink peptides. In this scenario, peptides crosslinking to RNA in cells would also have PTM sites within their peptide range. In most cases PTM sites did not overlap with crosslink sites, but in rare instances proteins had a substantial amount of PTM sites that corresponded to crosslink peptide sequences (Figure 2-5C)[45–48]. Using this data, we worked to understand the primary sequence resolution of the interactions and PTM sites, counting the number of PTMs that clustered near the crosslink peptide sites. CLIP crosslinking protocols rely on reactivity of nucleic acids and residues to form UV crosslinks; however, not all are equally reactive[69]. Specific residues including lys, cys, tyr, phe and trp are more reactive for UV crosslinking, and are also subject to posttranslational modification. In cases where the reactive residue is modified, the modification could potentially interfere with the crosslinking reaction. To avoid this potential source of bias, we additionally examined PTMs up to 10 positions away from a crosslink site. As shown in Figure 5b, some PTMs had observed depositions near the center of the crosslink site, suggesting that they may be near the RNA interface of RNA binding proteins. Across all PTMs, there is an inverse relationship between distance from the nearest crosslink site and PTM frequency (Spearman's rho = –0.84, $P < 2e{-}16$). For individual PTMs, each had a negative association between frequency and distance from crosslink site, which was significant for 16/33 PTMs (Supplementary Table 2-S4). The strongest association was observed for phosphorylation (rho = –0.83), followed by acetylation, ubiquitination, and methylation (rho = –0.65, –0.64 and –0.62 respectively). This observation is very exciting as the accounts of direct competition between PTM deposition and RNA binding are scarce, despite reports that PTM

modifications can alter protein structure when deposited at sites other than the direct RNA–protein interface[70]. Further, some types of RNA-binding proteins have been found to be intrinsically disordered, particularly in non-RNA-binding regions, and these disordered regions contain more PTM sites[71] and as such these interactions could serve as important relationships between PTMs and protein/RNA regulation to be further explored. Inspection of the proteins with higher overlap revealed central RBPs that control numerous RNA processing functions. In particular, hnRNP proteins have a significant amount of cross link sites that are proximal to PTM sites (Supplementary Table 2-S5). The mean distance from a PTM to the nearest crosslink site is significantly shorter in hnRNPs (36 residues) versus all other RBPs (268.4 residues; *t*-test *P*-value < 2.2e–16). We worked to analyze PTM sites at RNA–protein crosslinks from available crystal structures in the PDB. Among those, hnRNP-C stood out due to the high number of PTMs and those that intersect with cross link peptides from RNA–protein crosslinking experiments (Figure 2-5C). In total hnRNP-C has 63 PTM sites and 28 of them fall within the annotated RNA recognition domain (Figure 2-5D). hnRNP-C is a centralized RNA binding protein as part of a larger class of ubiquitously expressed heterogeneous nuclear ribonucleoproteins (hnRNPs). hnRNP-C has been demonstrated to be associated with associated with pre-mRNAs in the nucleus and appear to influence pre-mRNA processing and polyA tailing[72,73]. Inspection of the crystal structure overlaid with the PTM sites we observed diverse PTMs surrounding the RNA interface of the RNA recognition motif (RRM). One notable site of PTM mutation in cancers at the RRM motif is arginine 64 (R64; Figure 2-5D). Arginines are known to be sites of methylation on hnRNPs; this methylation controls their nuclear export[74] and can alter their ability to bind to nucleic acids by altering arginine

charge. In addition, hnRNPs contain about 65% of the total methylated arginine found in the cell nucleus[75]. The interface of methylated sites on hnRNPs may be a central hub of advanced regulation. Overall, our analysis provides a better understanding of the PTM landscape and how this intersects with the RNA–RBP interface. Further inspection of individual PTM sites and their mutations in cancer should prove valuable for analysis of specific PTMs that may contribute to RBP function, RNA regulation, and potentially oncogenesis.

An online database for access to the RBP PTM atlas

To provide the community with a working database for the ease of access to the data analyses herein, we have hosted an online database for the PTM atlas. We created a MySQL database and imported the data corresponding to 2436 RNA-binding proteins (RBPs) together with the corresponding information about 47 types of post-translational modification (PTMs). In total, there are 280 708 entries in the database, and each entry contains information about the RBP's Uniprot ID, gene name, position, PMID, Ssource, Amino Acid (AA), modification type, cell line, sample type, body site, and disease state. The database is publicly searchable by RBP Uniprot ID or by gene name and corresponding PTM type through the web server located at: http://PTM-RBP-ATLAS.igb.uci.edu/.

2.6 Discussion

RNA biology is tightly controlled through its interactions with RNA binding proteins. Chemical modifications, termed posttranslational modifications, can control protein function. Herein we provide the community with the first global analysis of PTM marks on RNA binding proteins. Our analysis, using experimentally-determined RNA binding

proteins and PTM databases reveals that RBPs are modified by a wide variety of PTMs, in all totaling 69 658 distinct modifications. We also demonstrate that the sites of PTM deposition are conserved, with many RNA binding proteins have a high level of diversity of their cataloged PTM modification types.

We have attempted to create the broadest possible sets of data for PTMs and PPIs; however, sources of bias still exist. Our data are derived from existing literature, and PTMs are not studied evenly—phosphorylation is the focus of far more studies than other PTMs, and just four modifications (phosphorylation, acetylation, ubiquitination and methylation) make up the bulk of the available data. Less-studied PTMs may not have enough data to draw significant conclusions observed with more common modifications. For PPIs, different methodologies may miss certain interactions, including transient interactions with substrates. We included eight PPI databases, which include both high-throughput and more focused experiments, using both physical and genetic techniques, to collect the most complete set of interactions possible; however, it is possible that difficult to detect interactions are underrepresented in our interaction data.

Intersection of protein–protein interaction data onto RBP modifications reveals complexity of PTM-depositing enzymes and those that could be responsible for PTM removal. Focusing on phosphorylation, an analysis of overlapping protein–protein interactions suggests master regulators that could be controlling PTM deposition and removal. This analysis highlights the potential of such enzymes, which have known roles in regulating cancer phenotypes, at center stage, thus likely controlling signaling through RBP function. Two enzymes highlighted, PTEN and SRC kinase, are known to be critically important in controlling oncogenic phenotypes, but there has yet to be a connection in their

role as master regulators of RBP function and as such the potential for re-wiring the transcriptome. Our database provides a framework for generating hypotheses about potential communication between PTM regulation and RBP function ripe for testing.

We also analyzed the relationship between PTM deposition sites and mutations identified in The Cancer Genome Atlas. Our analysis revealed that many of the catalogued RBPs contain mutations, with some familiar nucleic acid binding proteins such as p53 dominating the landscape, but further analysis revealed lesser-known yet likely important mutations. For example, the SF3B1 gene encodes part of the U2 snRNP portion of the spliceosome B complex. Consistent with the literature, we observe that SF3B1 is enriched in mutations; one in particular, K700E, is now established to play an important role in regulating alternative splicing through an unknown mechanism[67]. Our PTM atlas may provide a clue by virtue of the complementary analysis of PTM mass spectrometry and structural datasets demonstrating that K700 is a site of ubiquitination which may interact with the pre-mRNA undergoing splicing. A complete mechanism, further analyzed through the lens of controlling ubiquitin deposition, is sure to provide additional understanding of this modified enriched mutation site.

PTM sites control protein function, but it is unclear if proximity to the RNA–protein interface would allow them to influence RNA–protein interactions. We provide the first comparison of PTM data in the context of in-cell captured RNA–protein crosslinks. This is critical to understand if there is an overlap between the sites of PTM deposition and RNA binding. Through this analysis, the complexity of RNA recognition motifs, experimentally validated, and the relationship of PTM sites is revealed. PTM complexity and number of the hnRNP proteins, a critically important class of RBPs, comes to light. Highlighting this

relationship suggests an enrichment for arginine methylation among these proteins, suggesting that neutralization of the arginine positive charge could be critically important for controlling the RBP–RNA interface. Such analysis in a fuller context of these modifications and their relationship to controlled RBP function will be invaluable to fully understand their roles.

Finally, we provide the community with an online database that is fully searchable. We hope this database, and the analyses herein provide a catalyst for experiments to fully understand the role of PTMs in regulating RBP and RNA function.

2.7 Data Availability

The web interface to the database is available at: http://PTM-RBP-ATLAS.igb.uci.edu/. This website is free, open to all users and no login or password is required.

2.8 Supplementary Data

Supplementary Data (https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkac243#supplementary-data) are available at NAR Online.

2.9 References

1. J.S., M. The state of long non-coding RNA biology. *Noncoding RNA* **4**, 17.
2. K.V., M. & J.S., M. The rise of regulatory RNA. *Nat. Rev. Genet.* **15**, 423–437.
3. V., D. & M., E. Disruption of long noncoding RNAs targets cancer hallmark pathways in lung tumorigenesis. *Cancer Res.* **79**, 3028–3030.
4. A.M., S. & H.Y., C. Long noncoding RNAs in cancer pathways. *Cancer Cell* **29**, 452–463.
5. A.M., A., H., A. A.-S., R.K., S. & A., A. The role of RNA metabolism in neurological diseases. *Balkan J. Med. Genet.* **18**, 5–14.
6. Wang, K. C. & Chang, H. Y. Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell* **43**, 904–914 (2011).
7. M.W., H., A., C., T., S. & T., P. A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* **19**, 327–341.
8. E., D. Handshakes and fights: the regulatory interplay of RNA-binding proteins. *Front Mol Biosci* **4**, 67.
9. E.L., V. N. *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719.
10. E.L., V. N. *et al.* Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol.* **21**, 90.
11. M., C., M.C., B. & G.W., Y. How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol. Cell* **78**, 9–29.
12. F.C.Y., L. & J., U. Advances in CLIP technologies for studies of protein-RNA interactions. *Mol. Cell* **69**, 354–369.

13. D.G., K., N.V., K. & T.S., G. Chemical and functional aspects of posttranslational modification of proteins. *Acta Naturae* **1**, 29–51.

14. Y.C., W., S.E., P. & J.F., L. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.* **24**, 143–160.

15. V., P. *et al.* The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci.* **23**, 1077–1093.

16. Huang, Q. *et al.* Human proteins with target sites of multiple post-translational modification types are more prone to be involved in disease. *Journal of Proteome Research* **13**, 2735–2748 (2014).

17. Hartley, A. V. & Lu, T. Modulating the modulators: regulation of protein arginine methyltransferases by post-translational modifications. *Drug Discovery Today* **25**, 1735–1743 (2020).

18. Mukhopadhyay, R., Jia, J., Arif, A., Ray, P. S. & Fox, P. L. The GAIT system: a gatekeeper of inflammatory gene expression. *Trends in Biochemical Sciences* **34**, 324–331 (2009).

19. Arif, A. *et al.* The GAIT translational control system. *Wiley Interdiscip Rev RNA* **9**, (2018).

20. Owen, D. H. *et al.* KRAS G12V mutation in acquired resistance to combined BRAF and MEK inhibition in papillary thyroid cancer. *JNCCN Journal of the National Comprehensive Cancer Network* **17**, 409–413 (2019).

21. Chen, Y. J. *et al.* DbSNO 2.0: A resource for exploring structural environment, functional and disease association and regulatory network of protein S-nitrosylation. *Nucleic Acids Research* **43**, D503–D511 (2015).

22. Keshava Prasad, T. S. *et al.* Human Protein Reference Database - 2009 update. *Nucleic Acids Research* **37**, (2009).

23. Chen, H., Xue, Y., Huang, N., Yao, X. & Sun, Z. MeMo: A web tool for prediction of protein methylation modifications. *Nucleic Acids Research* **34**, (2006).

24. Hansen, J. E., Lund, O., Nielsen, J. O., Hansen, J. S. & Brunak, S. O-GLYCBASE: A revised database of O-glycosylated proteins. *Nucleic Acids Research* **24**, 248–252 (1996).

25. Gnad, F. *et al.* PHOSIDA (phosphorylation site database): Management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biology* **8**, (2007).

26. Diella, F. *et al.* Phospho.ELM: A database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics* **5**, (2004).

27. Hornbeck, P. v. *et al.* PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Research* **40**, (2012).

28. Craveur, P., Rebehmed, J. & de Brevern, A. G. PTM-SD: A database of structurally resolved and annotated posttranslational modifications in proteins. *Database* **2014**, (2014).

29. Sun, M. A. *et al.* RedoxDB-a curated database for experimentally verified protein oxidative modification. *Bioinformatics* **28**, 2551–2552 (2012).

30. Blanc, M. *et al.* SwissPalm: Protein Palmitoylation database. *F1000Res* **4**, (2015).

31. Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M. & Bairoch, A. UniProtKB/Swiss-Prot. *Methods Mol Biol* **406**, 89–112 (2007).

32. Li, H. *et al.* SysPTM: A systematic resource for proteomic research on post-translational modifications. *Molecular and Cellular Proteomics* **8**, 1839–1849 (2009).

33. Chernorudskiy, A. L. *et al.* UbiProt: A database of ubiquitylated proteins. *BMC Bioinformatics* **8**, (2007).

34. Castello, A. *et al.* Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular Cell* **63**, 696–710 (2016).

35. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nature Reviews Genetics* **15**, 829–845 (2014).

36. Oughtred, R. *et al.* The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci* **30**, 187–200 (2021).

37. Ruepp, A. *et al.* CORUM: The comprehensive resource of mammalian protein complexes-2009. *Nucleic Acids Research* **38**, (2009).

38. Salwinski, L. *et al.* The Database of Interacting Proteins: 2004 update. *Nucleic Acids Research* **32**, (2004).

39. Kumar, M. *et al.* ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Res* **48**, D296–D306 (2020).

40. Drew, K. *et al.* Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes. *Molecular Systems Biology* **13**, 932 (2017).

41. Hermjakob, H. *et al.* IntAct: An open source molecular interaction database. *Nucleic Acids Research* **32**, (2004).

42. Licata, L. *et al.* MINT, the molecular interaction database: 2012 Update. *Nucleic Acids Research* **40**, (2012).

43. Szklarczyk, D. *et al.* The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research* **45**, D362–D368 (2017).

44. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. doi:10.1002/0471250953.bi1112s47.

45. Bae, J. W., Kwon, S. C., Na, Y., Kim, V. N. & Kim, J. S. Chemical RNA digestion enables robust RNA-binding site mapping at single amino acid resolution. *Nature Structural and Molecular Biology* **27**, 678–682 (2020).

46. Castello, A. *et al.* Comprehensive Identification of RNA-Binding Proteins by RNA Interactome Capture. *Methods Mol Biol* **1358**, 131–139 (2016).

47. Kramer, K. *et al.* Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nature Methods* **11**, 1064–1070 (2014).

48. Trendel, J. *et al.* The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell* **176**, 391-403.e19 (2019).

49. Bagwan, N., el Ali, H. H. & Lundby, A. Proteome-wide profiling and mapping of post translational modifications in human hearts. *Scientific Reports* **11**, (2021).

50. Wojcechowskyj, J. A. *et al.* Quantitative phosphoproteomics reveals extensive cellular reprogramming during HIV-1 entry. *Cell Host and Microbe* **13**, 613–623 (2013).

51. Tanaka, H. *et al.* The intellectual disability gene PQBP1 rescues Alzheimer's disease pathology. *Molecular Psychiatry* **23**, 2090–2110 (2018).

52. Shinde, M. Y. *et al.* Phosphoproteomics reveals that glycogen synthase kinase-3 phosphorylates multiple splicing factors and is associated with alternative splicing. *Journal of Biological Chemistry* **292**, 18240–18255 (2017).

53. Beltrao, P. *et al.* Systematic functional prioritization of protein posttranslational modifications. *Cell* **150**, 413–425 (2012).

54. Millar, A. H. *et al.* The Scope, Functions, and Dynamics of Posttranslational Protein Modifications. *Annu Rev Plant Biol* **70**, 119–151 (2019).

55. Gill, G. SUMO and ubiquitin in the nucleus: Different functions, similar mechanisms? *Genes and Development* **18**, 2046–2059 (2004).

56. Alcock, L. J., Perkins, M. v. & Chalker, J. M. Chemical methods for mapping cysteine oxidation. *Chemical Society Reviews* **47**, 231–268 (2018).

57. Cicenas, J., Zalyte, E., Bairoch, A. & Gaudet, P. Kinases and cancer. *Cancers (Basel)* **10**, (2018).

58. Irby, R. B. & Yeatman, T. J. Role of Src expression and activation in human cancer. *Oncogene* **19**, 5636–5642 (2000).

59. Dehm, S. M. & Bonham, K. SRC gene expression in human cancer: The role of transcriptional activation. *Biochemistry and Cell Biology* **82**, 263–274 (2004).

60. Dai, T. Y. *et al.* P68 RNA helicase as a molecular target for cancer therapy. *Journal of Experimental and Clinical Cancer Research* **33**, (2014).

61. Hashemi, V. *et al.* The role of DEAD-box RNA helicase p68 (DDX5) in the development and treatment of breast cancer. *J Cell Physiol* **234**, 5478–5487 (2019).

62. Barber, K. W. & Rinehart, J. The ABCs of PTMs. *Nature Chemical Biology* **14**, 188–192 (2018).

63. Bansal, N., Yendluri, V. & Wenham, R. M. The molecular biology of endometrial cancers and the implications for pathogenesis, classification, and targeted therapies. *Cancer Control* **16**, 8–13 (2009).

64. Zhou, Z. *et al.* The biological function and clinical significance of SF3B1 mutations in cancer. *Biomarker Research* **8**, (2020).

65. Song, E. J. *et al.* The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. *Genes and Development* **24**, 1434–1447 (2010).

66. Bellare, P. *et al.* A role for ubiquitin in the spliceosome assembly pathway. *Nature Structural and Molecular Biology* **15**, 444–451 (2008).

67. Obeng, E. A. *et al.* Physiologic Expression of Sf3b1K700E Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. *Cancer Cell* **30**, 404–417 (2016).

68. Castello, A. *et al.* Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149**, 1393–1406 (2012).

69. Ule, J., Jensen, K., Mele, A. & Darnell, R. B. CLIP: A method for identifying protein-RNA interaction sites in living cells. *Methods* **37**, 376–386 (2005).

70. Thapar, R. Structural basis for regulation of rna-binding proteins by phosphorylation. *ACS Chemical Biology* **10**, 652–666 (2015).

71. Zhao, B. *et al.* Intrinsic Disorder in Human RNA-Binding Proteins. *Journal of Molecular Biology* **433**, (2021).

72. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Human Genetics* **135**, 851–867 (2016).

73. Gruber, A. J. *et al.* A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Research* **26**, 1145–1159 (2016).

74. Shen, E. C. *et al.* Arginine methylation facilitates the nuclear export of hnRNP proteins. *Genes and Development* **12**, 679–691 (1998).

75. Liu, Q. & Dreyfuss, G. In vivo and in vitro arginine methylation of RNA-binding proteins. *Molecular and Cellular Biology* **15**, 2800–2808 (1995).

Chapter 3: Selective cancer cell killing by exploiting a defective splicing machinery

3.1 Contribution Statement

Robert Spitale conceived the project idea. Minigene plasmids were acquired from Jian Zhang and James Manley. Jingtian Wang designed and cloned all the plasmids and conducted all the experiments. Kim Nguyen and Jan Zimak guided some of the experiments. Jan Zimak provided the Kid toxin cloning strategy. K562 CRISPR cell lines were acquired from Katrina Lappin and Kenneth Mills. Jingtian Wang wrote the manuscript with comments provided by Kim Nguyen and Robert Spitale.

3.2 Abstract

Currently, the two main strategies of cancer gene therapies to target cancer cells are vector engineering and cancer-specific promoters. These strategies are reliant on cancer cells displaying the hallmarks of cancer. Cancers frequently carry mutations in their splicing machinery which leads to the production of aberrant mRNA transcript. Here, we engineered a mis-spliced mRNA transcript to respond to a hotspot mutation in splicing factor protein SF3B1. We demonstrated that this splicing-responsive RNA switch can be shrunken down to ~200bp in size and can enable specific protein expression in cells harboring K700E or several other mutations in SF3B1. When the Kid toxin is fused downstream of the RNA switch, it can specifically kill cancer cells carrying the SF3B1 K700E mutation. This RNA switch offers a novel strategy to target the faulty biochemistry of proteins in cancer cells.

3.3 Introduction

One of the great remaining challenges in cancer therapy is the design of drugs that will selectively kill cancer cells and leave healthy cells unharmed. Traditional cytotoxic chemotherapy drugs target rapidly dividing cells, killing both cancer cells and certain normal cells. These drugs lead to side effects and are not effective against quiescent cancer stem cells. Current targeted cancer therapies largely rely on detecting the extracellular surface markers on cancer cells. However, many of the cancer surface markers can also be found on germ cells[1]. On the other hand, the intracellular environment of cancer cells can be drastically different from that of normal cells[2]. Therefore, if a cancer therapy can detect the unique intracellular environment and even the cancer-causing mutations of cancer cells, it will selectively act upon cancer cells, while leaving normal cells unharmed, thus minimizing adverse side effects.

Almost 90% of the human genes undergo some splicing event[3]. Consequently, mutations in the splicing machinery can lead to many diseases, including cancer[4,5]. In our recent bioinformatic analysis of RNA binding proteins[6], we identified a hotspot mutation in invasive breast cancer: K700E in splicing factor 3B 1 (SF3B1). SF3B1 is an essential component of U2 snRNP, and it directly interacts with both the 5′ and 3′ sides of the branch point sequence. In addition to the components within the U2 snRNP, SF3B1 also interacts with other proteins in the spliceosome[7]. Several studies have reported SF3B1 K700E recognizes alternative 3' splice sites (3'ss) upstream of canonical 3'ss[8,9]. Importantly, Liu and colleagues recently demonstrated that SF3B1 K700E mutant suppresses PPP2R5A and MAP3K7 via aberrant splicing, leading to accelerated cell migration, thus directly linking the

SF3B1 K700E mutation to tumorigenesis[10]. The Kid toxin is a ribonuclease in *Escherichia coli* that preferentially cleaves the RNA at the UUACU of single-stranded regions, while the Kis antitoxin neutralizes the cell growth inhibition ability of Kid[11]. In this paper we aim to exploit the defective splicing mechanism in mutant SF3B1, and express an in-frame, active Kid toxin in cancer cells, and an out-of-frame, inactive form of the protein in normal cells. The development of such a targeted therapy will serve as a new paradigm for cancer cell death that can be extended toward other splicing-defective cancers.

3.4 Results

Identification of mis-spliced transcripts suitable for mutant selective protein expression

Zhang et al. first identified introns most mis-spliced at the 3' ss by mutant SF3B1 via RNA seq and cloned select mis-spliced targets into minigenes driven by the cytomegalovirus (CMV) promoter[12]. Seeing this as an opportunity for a gene therapy, we requested the minigene plasmids from them, and first co-transfected the minigene plasmids with SF3B1 wild type (WT) or K700E mutant (MT) expression plasmids into HEK cells. After three days post transfection, longer isoforms of the minigene transcripts in MT samples were observed via RT-PCR (Figure 3-1a). We were able to recapitulate the mutant cryptic splicing of the minigenes.

Next, to identify the minigene candidate suitable for protein expression, we cloned three of the minigenes (ORAI2, MAP3K7, ZNF91) upstream of GFP (Figure 3-1b). To enable the minigene for protein expression, bases were inserted or removed between the cryptic and the canonical 3'ss to ensure frameshifts. Hidden stop codons in the exons of the

minigenes were also removed by point mutagenesis. The mutant-dependent GFP expression was validated via flow cytometry and western blotting (Figure 3-1c-d). It was apparent that the ORAI2 splicing cassette had the highest mutant-to-wild type protein expression ratio.

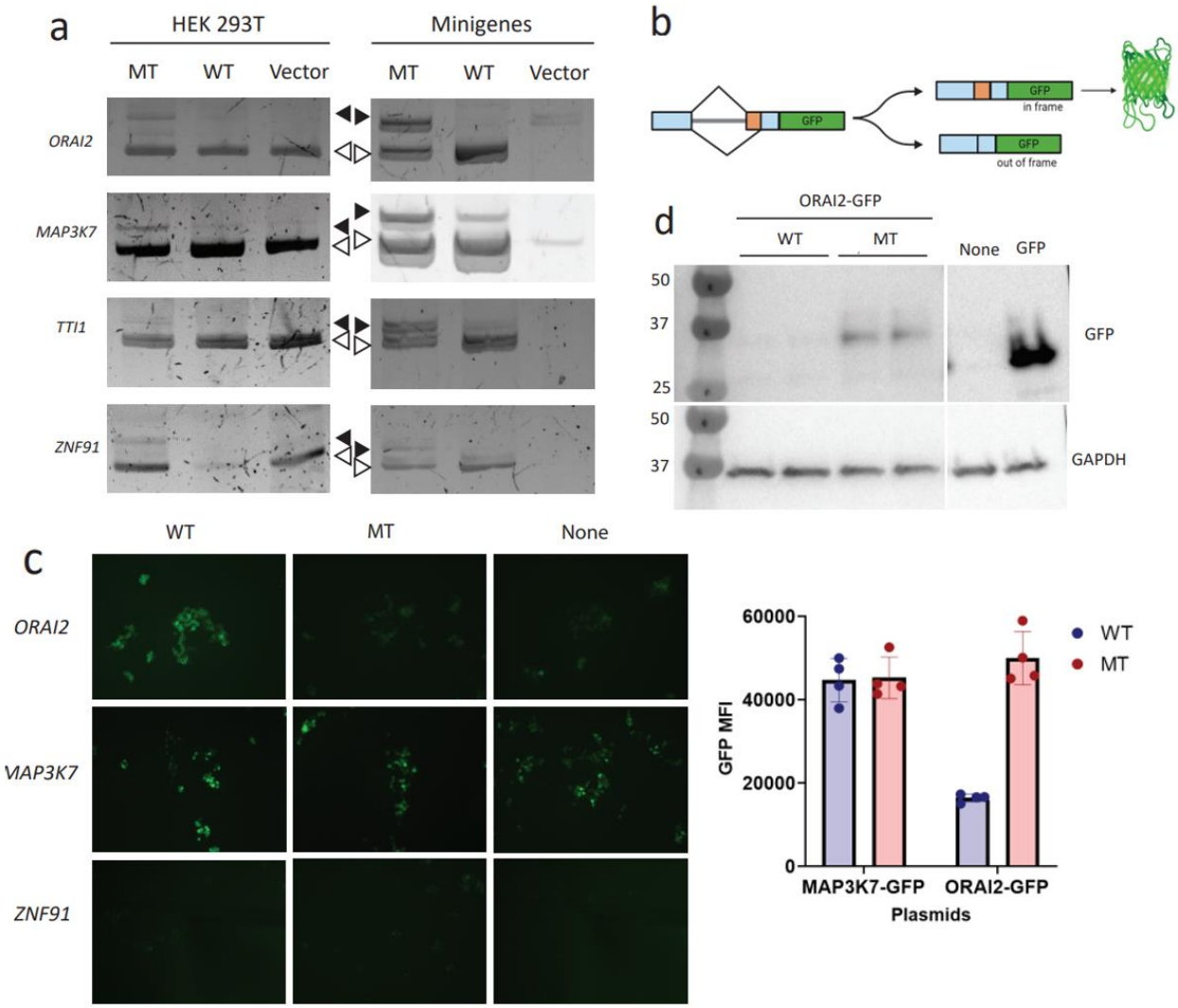


Figure 3-1. Identification of mis-spliced transcripts suitable for mutant selective protein expression. A. HEK293T cells were transfected with expression plasmid for wild type (WT) or K700E (MT) SF3B1, followed by total RNA extraction and RT-PCR of the 3' splice sites. Total RNA of HEK293T cells co-transfected with minigenes were RT-PCR'ed for

the 3' splice sites with primers specific to the minigene vectors. Solid arrow heads denote cryptic splice sites. Open arrow heads denote canonical splice sites. **B.** Schematic of minigene-GFP. **C.** HEK293T cells were transfected with GFP expression plasmid with ORAI2, MAP3K7 or ZNF91 splicing cassette fused to its N-terminus and co-transfected with WT or MT expression plasmid (microscopy result-left), followed by flow cytometry on cells transfected with the MAP3K7 and ORAI2 minigene-GFP constructs (right plot). **D.** Confirmation of ORAI2-GFP's MT-specific gene expression (anti-GFP) via western blotting. Followed by GAPDH loading control.

Optimizing the ORAI2 splicing cassette

Because the ~1.1kb length of the original ORAI2 splicing cassette is over 20% of the packaging limit of the AAV vector, and we plan on incorporating other regulatory elements (promoter, IRES, PolyA tail, etc.) into the gene therapy, we next investigated whether the ORAI2 splicing cassette can be shortened while still maintaining the SF3B1 mutation selectivity (Figure 3-2a)[13]. ORAI2 intron was progressively trimmed down to 32bp downstream of the 5' ss and 65bp upstream of the 3'ss, preserving any identifiable cis-regulatory elements of splicing[14]. The 5' exon was trimmed down to 29bp upstream of the 5'ss and the 3'exon was trimmed down to 23bp downstream of the 3'ss. A T2A self-cleaving peptide was inserted between ORAI2 and GFP to avoid any potential interference to the cargo by ORAI2. These shortened constructs were tested by co-transfection with SF3B1 WT/MT plasmids followed by flow cytometry. It was shown that the ORAI2 splicing cassette can be minimized from 1076bp to 172bp while still maintaining relatively high

mutant-selectivity (Figure 3-2b). Two ORAI2 splicing cassettes in tandem did not offer any significant improvement on the specificity (Figure 3-S1).

The SF3B1 expression plasmid was mutagenized to two other common SF3B1 mutants in cancers: K666T and R625G[15]. It was found that the ORAI2 splicing cassette is applicable to those mutations as well, in addition to K700E (Figure 3-2c).

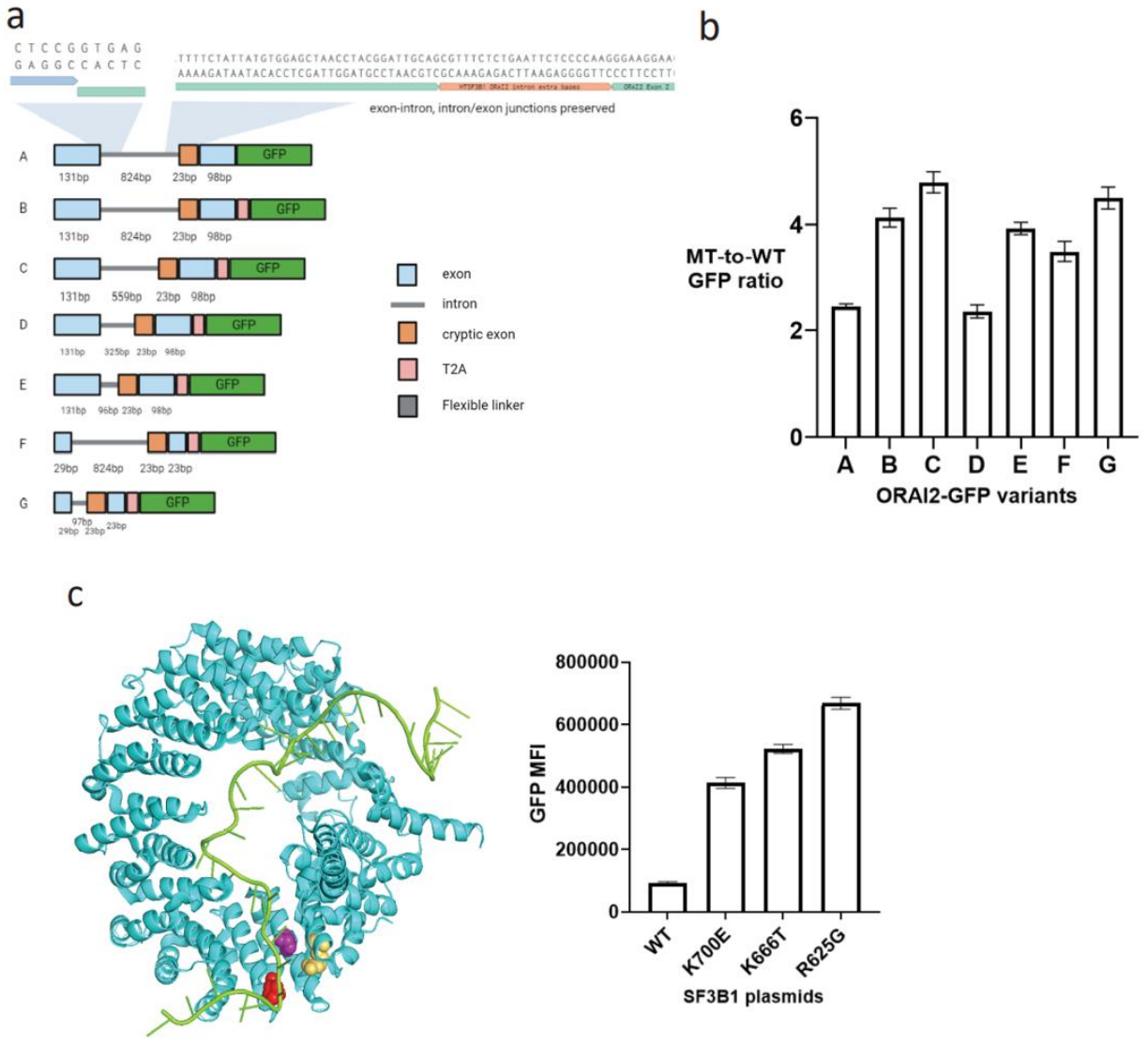


Figure 3-2. Optimization of the ORAI2 splicing cassettes. A. Schematics of all versions of ORAI2-GFP constructs. Untouched sequences at the splice junctions were shown. **B.** Flow

cytometry of HEK293T cells co-transfected with WT or MT plasmids and variants of ORAI2-GFP constructs. C. Structure of SF3B1 with mutated residues highlighted. Red = K700, Purple = K666, Yellow = R625. Flow cytometry results of HEK293T cells co-transfected with WT, K700E, K666T or R625G plasmids and ORAI2-GFP construct. Flow cytometry of HEK293T cells co-transfected with WT or MT plasmids and variants of 2xORAI2 tandem-GFP constructs.

The use of Kid/Kis toxin/antitoxin pair for mutant-specific cell killing

The Kid (*killing determinant*) bacterial toxin was chosen as the cargo due to its small size. It was noticed that a Kid toxin plasmid could not be cloned alone despite using a mammalian expression promoter. We hypothesize this was because the prokaryotic gene expression system allows for the leaky expression of Kid, killing any bacteria carrying the Kid plasmid. Eventually, we were able to clone the Kid plasmid for constitutive mammalian expression with the Kis (*killing suppressor*) antitoxin under a prokaryotic promoter (Figure 3-3a). To test the toxicity of the Kid toxin in mammalian cells, the Kid plasmid (with mCherry under IRES) was transfected into HEK 293T cells. Three days post transfection, microscopy showed the complete ablation of mCherry by Kid. mCherry expression is restored when Kid is co-expressed with Kis (Figure 3-3b).

We next investigated whether the ORAI2 exons affect the toxicity of Kid. Intron-less ORAI2 was cloned upstream of the Kid toxin. MTT cell viability assay showed that ORAI2 exons negatively affect the toxicity of Kid, however the toxicity is restored with the insertion of a T2A self-cleaving peptide between intron-less ORAI2 and Kid (Figure 3-3c).

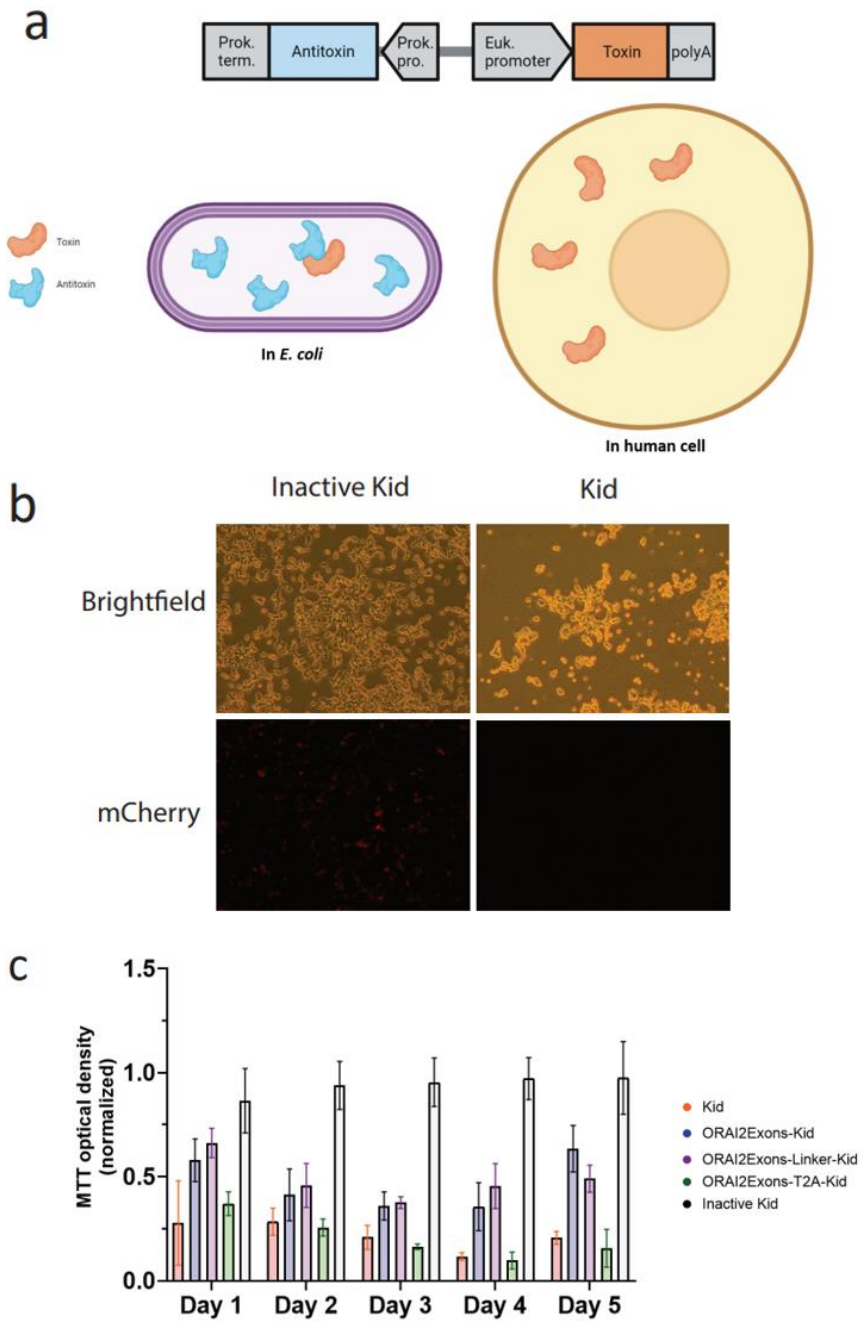


Figure 3-3. Assessment of Kid toxin potency in cell killing. A. Cloning schematics of the Kid toxin for constitutive expression in mammalian cells. In E. coli, the Kis antitoxin is highly expressed under the prokaryotic promoter, suppressing any leaked Kid toxin. In mammalian cells, the mammalian expression machinery would only recognize the eukaryotic promoter, leading to high toxicity in cells. **B.** Microscopy of HEK293T cells three

days post transfection of active or inactive Kid constructs (both had mCherry under IRES). **C.** MTT growth assay of HEK293T cells transfected with 5 variations of Kid toxin. n=4.

To recapitulate the mutation more closely in an isogenic context, we obtained a K562 CRISPR cell line harboring the SF3B1 K700E mutation from Liberante el al. [16] Cryptic splicing in K562 MT was confirmed via flow cytometry (Figure 3-4a). The ORAI2 splicing cassette was then cloned upstream of Kid and transfected into wild type K562 cells (WT) or K562 CRISPR cell line with the SF3B1 K700E mutation (MT). MTT cell viability assay showed that ORAI2-T2A-Kid negatively affected MT cells growth more than WT cells (Figure 3-4b). However, the toxicity was high in both cell types.

To lower the unwanted toxicity of our design in WT, the Kis antitoxin was introduced downstream of Kid, under the control of an Internal Ribosomal Entry Site (IRES) (Figure 3-4c). Since MAP3K7 is the "leakier" splicing cassette and expresses more cargo protein in WT compared to ORAI2, it was inserted between IRES and Kis. In an alternative design, two T2A peptides were introduced between IRES and Kis to lower the expression of Kis[17]. These two designs will constitutively express low trace amount of Kis, neutralizing the Kid leaked in WT cells, but still sparing enough Kid toxin in MT cells to exert cell-killing effect. To test these scenarios, MTT assay was deployed to assess cell viability. MTT cell viability assay showed that the new designs had less toxicity in WT cells while still maintaining similar toxicity in MT cells compared to the original ORAI2-T2A-Kid (Figure 3-4d). When ORAI2 splicing cassette was applied to the pro-apoptotic factor PUMA, it also showed MT-specific cell killing (Figure 3-S2).

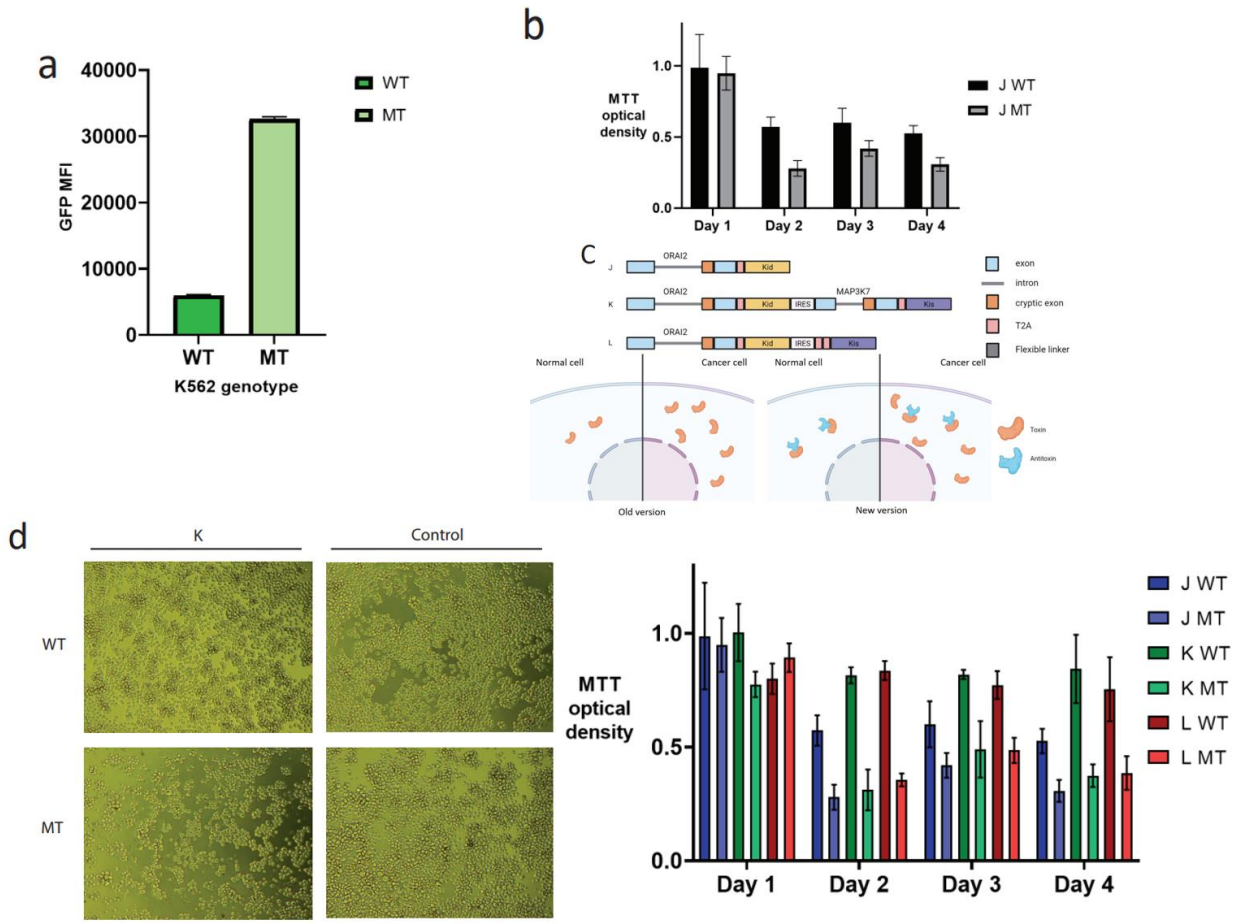


Figure 3-4. Assessment of the ORAI2 containing Kid toxin and Kis antitoxin cassettes in killing K562 cancer cells harboring WT or mutant SF3B1. A. Flow cytometry of WT or MT K562 cells nucleofected with the ORAI2-T2A-GFP plasmid. **B.** MTT growth assay of WT or MT K562 cells nucleofected with ORAI2-T2A-Kid (J) plasmid. n=5. **C.** Schematics of 3 versions of ORAI2-Kid constructs and their proposed mechanisms. **D.** Microscopy of WT or MT K562 cells nucleofected with construct K 3 days post nucleofection. MTT growth assay of WT or MT K562 cells nucleofected with three versions of ORAI2-Kid constructs. MTT reads were normalized to controls. n=5.

Exploring other mis-spliced transcripts as potential gene therapies

Rank	Gene Name	Related Cancer	Event	Event type	Number of patients inflicted
4	EPHX1	Lung[20]	alt_5prime_9779	A5	341
39	PTMA	Colorectal, bladder, lung, and liver[21–23]	alt_5prime_126313	A5	174
41	TPT1	Breast, liver and prostate[24–26]	alt_3prime_67141	A3	174
42	PTMA	Colorectal, bladder, lung, and liver	mutex_exons_3025134	MEX	171
43	TPT1	Breast, liver and prostate	alt_5prime_49328	A5	169
44	LTBR	Acute myeloid leukemia[20]	alt_5prime_37658	A5	165

Table 3-1 Table of select top CASEs and the cancers they are related to.

In our previous study on RNA binding proteins (RBP), we showed that defect-causing SF3B1 mutations account for ~20% of all uveal melanoma cases and only ~1% of all breast cancers cases[6]. Since alteration of splicing is prevalent across all cancer types, we set out to discover more candidates that could be used in place of ORAI2 to target more cancer cases[18]. To that end, we adopted the alternative splicing data from a recent bioinformatic analysis on the splicing landscape across TCGA cancer patients[18]. Because alternative splicing is usually tissue-specific, we first removed the cancer patients without normal controls with matching tissue of origin [19]. Two filters were used to determine the top candidate alternative splicing events (CASEs): 1. The average Percent spliced-in (PSI) of

the CASE in normal samples must be the same as that of ORAI2's or better (minimizing the "leakiness" of the CASE in WT). 2. The disparity between the cancer and normal PSIs of the CASE must be the same as the SF3B1 K700E to normal PSI ratio of ORAI2 or better, and this difference must be statistically significant (filtering for cancer-specific protein expression). After the filtering steps, the CASEs were ranked based on the number of cancer patients they cover.

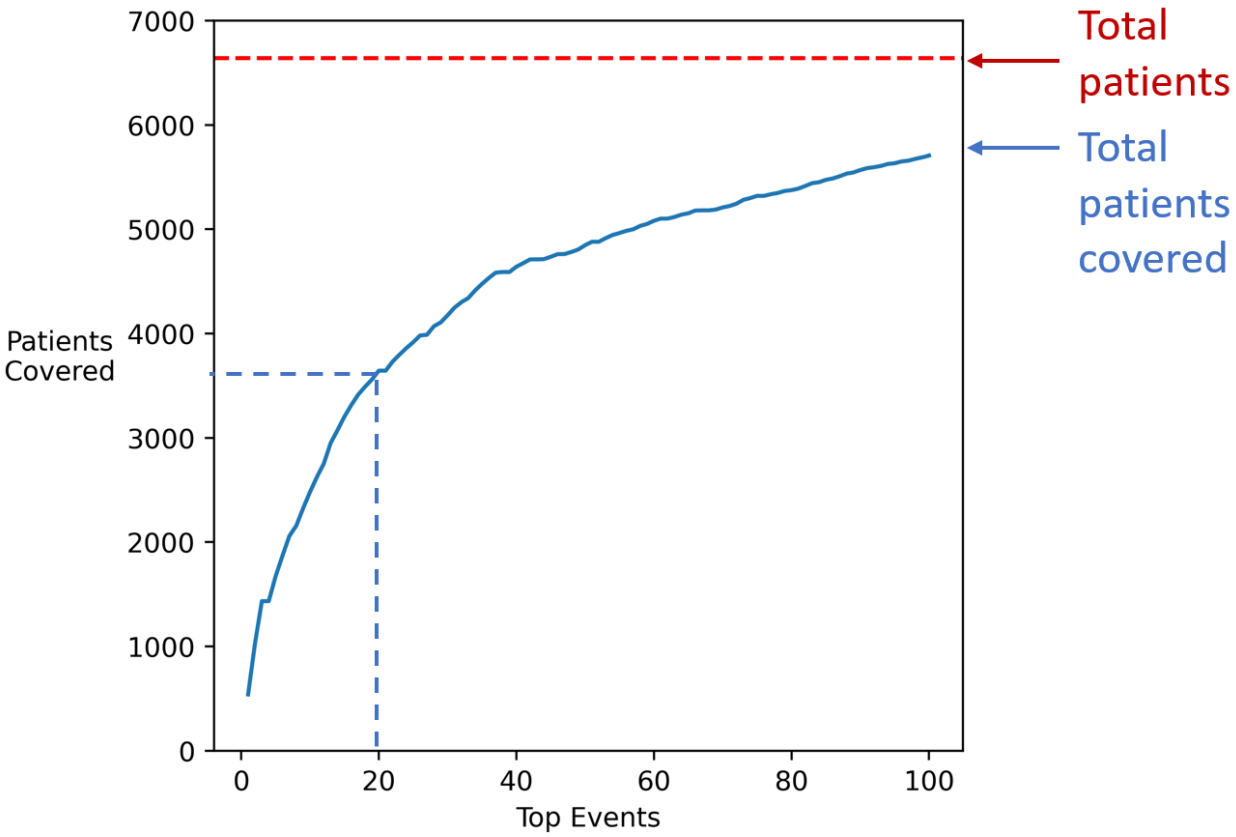


Figure 3-5. Number of patients covered by the CASES vs number of top CASEs considered. Red dotted line indicates total number of TCGA patients eligible for this analysis.

We found that among the top CASEs, several events are at the loci of genes implicated in cancers (Table 3.1). We hypothesize that the mis-splicings of those genes serve as potential drivers of tumor progression. We also found that top 20 ranked CASEs are sufficient to cover over half of all considered TCGA cancer patients (3644/6641 = 54.9%) (Figure 3-3.5). These top CASEs have the potential to increase the generalizability of our system and be applied to a wide range of cancer patients.

3.5 Discussion

Currently, the targeting methods of cancer gene therapies are largely limited to cancer-specific promoters or engineered cancer-specific vectors. If a gene therapy can target the inherent faulty biochemistry of a housekeeping protein of the cancer cell, the therapy can potentially overcome the limitations in targeting dormant or stem-like cancer cells. In this paper we show the potential to kill the cancer cell by exploiting a defect in its splicing machinery. A portion of the ORAI2 gene was chosen and modified to be the splicing RNA switch of the system. The Kid RNase was used as the toxin of choice and its cloning and toxicity were optimized. The leakiness of the design in WT cells was lessened by the addition of the antitoxin. Moreover, we have shown that ORAI2 has the potential to be engineered to facilitate the gene therapy design. It can be used to target mutations other than SF3B1 K700E and can regulate the expression of a variety of protein cargos such as bacterial toxins and pro-apoptotic factors. Lastly, we have shown the potential of expanding the use case of our gene therapy by exploiting other alternative splicing events. However, the mechanism of action of the mis-splicing of those expanded events are unknown.

Recently, North et al. reported a similar design to selectively express the herpes simplex virus–thymidine kinase (HSV–TK) in SF3B1 mutant cancer cells[27]. Compared to our design, their design involved inserting a "mutant-splice-out" intron that was spliced out in SF3B1-mutant cancer cells, but spliced in in WT cells, thus disrupting the translation of the toxic protein. Their design requires a deliberate choice of the insertion location of the synthetic intron, as the ORF of the cargo is involved in the alternative splicing. In our design, the synthetic intron is inserted upstream of the cargo ORF. The splicing of the synthetic intron is not affected by the ORF of the toxin. This enables easy swapping the cargo and better generalizability of the design.

3.6 Methods

Expression Plasmid Constructs

The SF3B1 WT and K700E MT plasmids were obtained from addgene (#82576 and #82577). The SF3B1 K666T and R625G plasmids were made by point mutagenesis based on the WT plasmid. The ORAI2, MAP3K7, ZNF91 minigene plasmids were gifts from Dr. James Manley's lab [12]. The GFP sequence and plasmid backbone for making the minigene-GFP plasmids were obtained from addgene (#13031). The Kid toxin and the Kis antitoxin sequences were from addgene (#58299 and #58306). The PUMA DNA fragment was obtained from addgene (#16588). All plasmids were constructed using the Gibson Assembly Master Mix (New England Biolabs).

Cell handling and transfection

HEK293T cells were maintained in DMEM media (Thermo Fisher Scientific) supplemented with 10% FBS and 1% Pen/Strep. For all overexpression experiments,

300,000 cells were seeded to each well of a six-well plate. On the next day, 2ug of total plasmid DNA were transfected to each well using jetPRIME Transfection Reagent (Polyplus) according to the manufacturer's instructions.

K562 WT and MT cells were maintained in RPMI media (Cytiva) supplemented with 10% FBS and 1% Pen/Strep. For all overexpression experiments, media was refreshed two days prior to the transfection. On the day of transfection, 1 million cells were carefully resuspended in 100uL of room-temperature supplemented Nucleofector Solution and transferred to the provided cuvette (Lonza). Cells were then nucleofected using the T-016 program on the Nucleofector 2b (Lonza) and immediately transferred to pre-warmed media in a 6-well plate.

Minigene assay

Minigene overexpression experiments and the subsequent reverse transcription were conducted as described by Zhang et al. [12]

Western Blot

Whole cell lysates were dissolved in homemade RIPA buffer and protein concentrations were calibrated using the BCA Assay (Thermo Scientific). Protein was then ran on SDS-PAGE gel and transferred to nitrocellulose membranes (BIO-RAD), followed by immunoblotting with primary and secondary antibodies. Antibodies used were: anti-GFP antibody (Life Technologies, A11122), anti-rabbit secondary antibody (Cell Signaling Technology, 7074S), and Direct-Blot HRP anti-GAPDH antibody (Biolegend, 607904). Immunofluoresence was detected using the ChemiDoc Imaging System (BIO-RAD).

Flow Cytometry

On the third day post transfection, cells were lifted with trypsin (0.5% EDTA) (Fisher Scientific) and resuspended in room-temperature flow buffer (PBS with 5% FBS). Fluorescence was measured in the NovoCyte Flow Cytometer (Agilent).

MTT Cell Viability Assay

The day after transfection, K562 cells of each condition were centrifuged at 500xg for 5 minutes and resuspended in 500uL of fresh complete RPMI media with no phenol red (Life Technologies). K562 cells of each condition were transferred to 96 well plates at 20,000 cells (100uL) per well, 5 wells per plate. To minimize the edge effect, 100uL of PBS was added to the wells on the edges of each plate. For MTT assays, 1ug of MTT from the Vybrant MTT Cell Proliferation Assay Kit was dissolved in 1ml of PBS (Fisher Scientific). The 96-well plate was centrifuged at 500xg for minutes. Fifty microliter of media was carefully removed from each well. MTT solution was diluted with 5 parts of RPMI media with no phenol red and added to each well at 60uL per well. The plate was then incubated at 37C for 30 minutes. The 96-well plate was centrifuged at 500xg for minutes and 85uL of MTT/media mixture was carefully removed from each well. Formazan crystals in each well was dissolved with 150uL of DMSO by pipetting up and down slowly. Optical density was read in the CLARIOstar Plus Microplate Reader at 560nm (BMG LABTECH). Reads were blank-corrected and normalized to WT or MT negative controls (negative controls were transfected with filler plasmid).

Splicing Data Analysis

Three prime alternative splicing PSI data was downloaded from https://gdc.cancer.gov/about-data/publications/PanCanAtlas-Splicing-2018 [18]. Cancer samples were discerned from normal samples based on the TCGA barcodes. Cancer

samples harboring the SF3B1 K700E mutation were identified using the TCGA mutational data queried in our previous study[6].

The requirements for an alternative splicing event to be considered were as follows:

- The event must have at least 20 non-null PSI values in normal samples.
- The mean PSI values of the event must be >= 0.979 (mean of ORAI2 in normal samples) or <= 0.021 (1 - mean of ORAI2 in normal samples).
- The difference between the cancer PSI and normal PSI must be statistically significant (adjusted p< 0.001, Wilcoxon signed-rank test, adjusted to multiple testing correction)

The analysis was done in Bash and Python. Visualization was done with the matplotlib Python module.

3.7 Supplemental Data

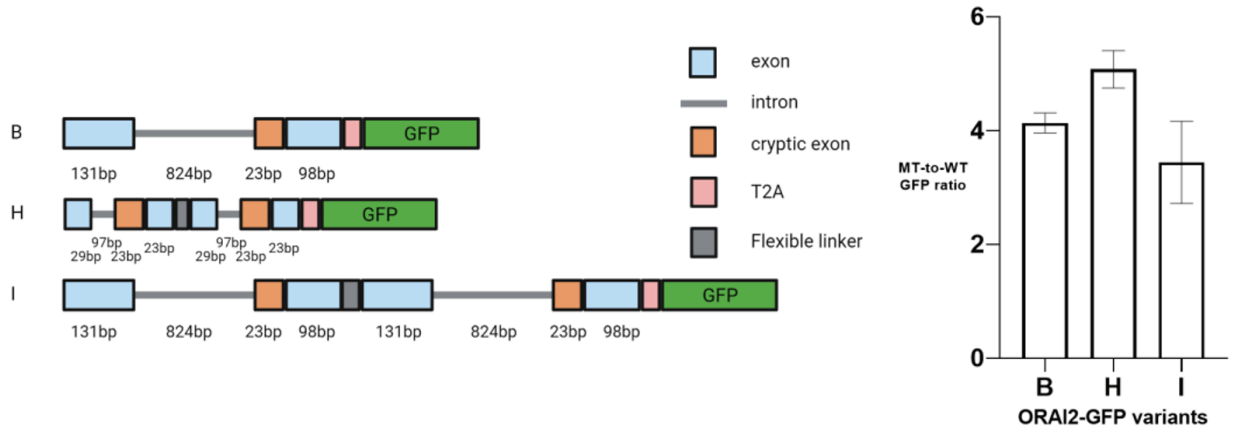


Figure 3-S1. Schematics and flow cytometry of HEK cells co-transfected with three variants of ORAI2-GFP plasmids and WT or MT SF3B1 plasmid.

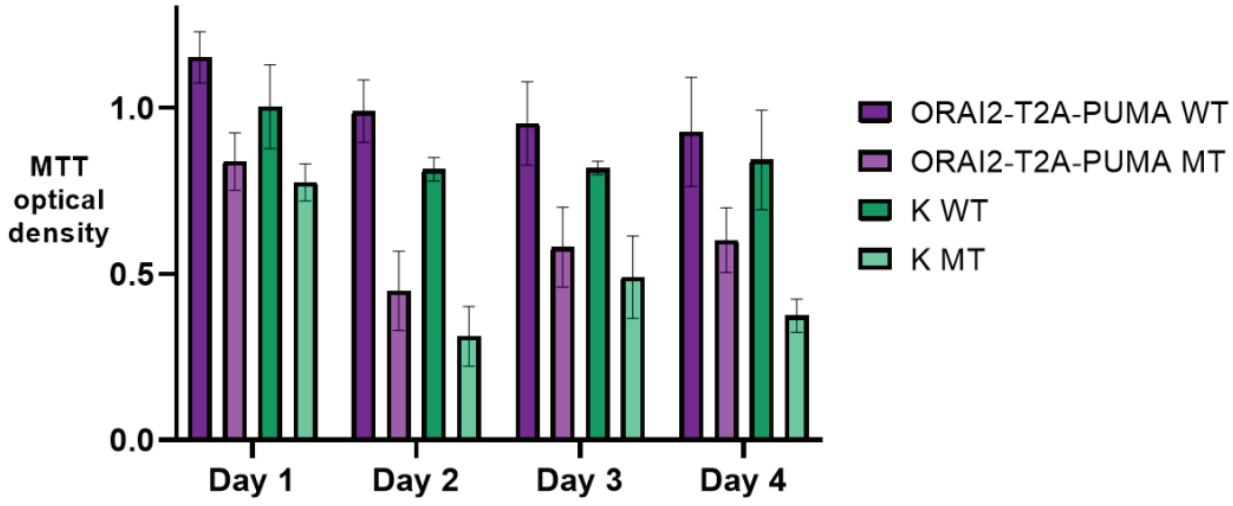


Figure 3-S2. Comparison of single ORAI2 Kid vs. PUMA cell killing potency. MTT growth assay of WT or MT K562 cells nucleofected with ORAI2-T2A-PUMA or construct K. (n=5).

3.8 References

1. Payne, K. K., Bear, H. D. & Manjili, M. H. Adoptive cellular therapy of cancer: Exploring innate and adaptive cellular crosstalk to improve anti-tumor efficacy. *Future Oncology* Preprint at https://doi.org/10.2217/fon.14.97 (2014).
2. Pérez-Mancera, P. A., Young, A. R. J. & Narita, M. Inside and out: The activities of senescence in cancer. *Nature Reviews Cancer* Preprint at https://doi.org/10.1038/nrc3773 (2014).
3. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* (2001) doi:10.1038/35057062.
4. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* (2011) doi:10.1038/nature10496.
5. van Roosmalen, W. *et al.* Tumor cell migration screen identifies SRPK1 as breast cancer metastasis determinant. *Journal of Clinical Investigation* (2015) doi:10.1172/JCI74440.
6. England, W. E. *et al.* An atlas of posttranslational modifications on RNA binding proteins. *Nucleic Acids Research* **50**, 4329–4339 (2022).
7. Wan, Y. & Wu, C. J. SF3B1 mutations in chronic lymphocytic leukemia. *Blood* Preprint at https://doi.org/10.1182/blood-2013-02-427641 (2013).
8. Darman, R. B. *et al.* Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. *Cell Reports* (2015) doi:10.1016/j.celrep.2015.09.053.
9. Kesarwani, A. K. *et al.* Cancer-associated SF3B1 mutants recognize otherwise inaccessible cryptic 3' splice sites within RNA secondary structures. *Oncogene* (2017) doi:10.1038/onc.2016.279.

10. Liu, B. *et al.* Mutant SF3B1 promotes AKT and NF-kB driven mammary tumorigenesis. *J Clin Invest* (2020) doi:10.1172/JCI138315.

11. de la Cueva-Méndez, G., Mills, A. D., Clay-Farrace, L., Diaz-Orejas, R. & Laskey, R. A. Regulatable killing of eukaryotic cells by the prokaryotic proteins Kid and Kis. *EMBO J* **22**, 246–251 (2003).

12. Zhang, J. *et al.* Disease-Causing Mutations in SF3B1 Alter Splicing by Disrupting Interaction with SUGP1. *Molecular Cell* (2019) doi:10.1016/j.molcel.2019.07.017.

13. Wu, Z., Yang, H. & Colosi, P. Effect of Genome Size on AAV Vector Packaging. *Molecular Therapy* **18**, 80 (2010).

14. Brent, M. R. & Guigó, R. Recent advances in gene structure prediction. *Current Opinion in Structural Biology* **14**, 264–272 (2004).

15. Alsafadi, S. *et al.* Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat Commun* **7**, (2016).

16. Liberante, F. G. *et al.* Altered splicing and cytoplasmic levels of tRNA synthetases in SF3B1-mutant myelodysplastic syndromes as a therapeutic vulnerability. *Scientific Reports* **9**, (2019).

17. Liu, Z. *et al.* Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific Reports 2017 7:1* **7**, 1–9 (2017).

18. Kahles, A. *et al.* Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell* **34**, 211-224.e6 (2018).

19. Grosso, A. R. *et al.* Tissue-specific splicing factor gene expression signatures. *Nucleic Acids Research* **36**, 4823 (2008).

20. Fernandes, M. T. *et al.* Lymphotoxin-β receptor in microenvironmental cells promotes the development of T-cell acute lymphoblastic leukaemia with cortical/mature immunophenotype. *Br J Haematol* **171**, 736–751 (2015).

21. Ha, S. Y., Song, D. H., Hwang, S. H., Cho, S. Y. & Park, C. K. Expression of prothymosin α predicts early recurrence and poor prognosis of hepatocellular carcinoma. *Hepatobiliary and Pancreatic Diseases International* **14**, 171–177 (2015).

22. Zhang, M. *et al.* Increased expression of Prothymosin-α, independently or combined with TP53, correlates with poor prognosis in colorectal cancer. *International Journal of Clinical and Experimental Pathology* **7**, 4867 (2014).

23. Tsai, Y. S. *et al.* Aberrant Prothymosin-alpha Expression in Human Bladder Cancer. *Urology* **73**, 188–192 (2009).

24. Jung, J. *et al.* Translationally controlled tumor protein induces human breast epithelial cell transformation through the activation of Src. *Oncogene* **30**, 2264–2274 (2011).

25. Arcuri, F. *et al.* Translationally controlled tumor protein (TCTP) in the human prostate and prostate cancer cells: Expression, distribution, and calcium binding activity. *Prostate* **60**, 130–140 (2004).

26. Chan, T. H. M. *et al.* Translationally controlled tumor protein induces mitotic defects and chromosome missegregation in hepatocellular carcinoma development. *Hepatology* **55**, 491–505 (2012).

27. North, K. *et al.* Synthetic introns enable splicing factor mutation-dependent targeting of cancer cells. *Nat Biotechnol* (2022) doi:10.1038/s41587-022-01224-2.

Chapter 4: Novel design of Tet-ON inducible O-Tet-A system provides robust and tight temporal control in gene expression

4.1 Contribution Statement

Kim Nguyen conceived the project idea. The O-Tet-RA system design, cloning and testing was done by Kim Nguyen and Myhanh Nguyen. Experiments involving Diphtheria toxin and flow cytometry (Figure 4-2b, Figure 4-3d, g, Figure 4-4, Figure 4-5b, c) were done by Jingtian Wang. Several mini promoter ideas were offered by Jan Zimak. Manuscript writing was done by Kim Nguyen and commented by Jingtian Wang. The project was supervised by Robert Spitale.

4.2 Abstract

The ability to control when and where a target gene expressed often employs an inducible transcription system. Tet-inducible (Tet-ON and Tet-OFF) vectors are widely adopted to such studies due to its robustness. However, inherited level of leakiness in classic design of Tet-inducible system has limited its usage in certain studies such as regulation of T-cell activation or suicide gene expression. In this study, we re-designed Tet-ON inducible system and subjected the novel system to different stringent tests including expression of Diptheria toxin to assess both its robustness and leakiness. We demonstrated that the new Tet-ON design called Long-SV40-TATA O-TetR-A system comprised of novel promoter design and dual tetracycline regulators effectively respond to low Doxycycline concentration with high expression output and low basal expression resulting significantly high signal-to-noise ratio of an induced gene.

4.3 Introduction

Current gene therapy approach employed for treating non-monogenetic defect neurological diseases such as Alzheimer and Parkinson often involved with delivery nerve growth factor (NGF) to the brain via injection AAV vector or introduction of genetically-modified cells to Alzheimer inflicted patients[1], or administered neurturin, a ligand in the family of glial cell line derived neurotrophic factor (GDNF) using AAV vector to Parkinson patients[2]. However, uncontrolled delivery of neurotrophic factors led to adverse effects[3–5].

T-cell immunotherapy for cancer utilized conventional chimeric antigen receptor (CAR)-expressing T cells has CAR expressed in constitutive mode; consequently, suffer frequent on-target and off-tumor toxicity[6–9]. Therefore, to improve safety for this type of therapy, it is necessary to tightly regulate CAR expression, restrict antigen recognition and T-cell activation.

To address such unmet need for improvement in therapy safety, several inducible systems have been developed and modified. One of such is Tet-inducible system including Tet-ON and Tet-OFF versions. Essentially, the Tet-inducible system has a Tet-responsive promoter (pTet) comprising of 7xTet operon (TetO) sequences, followed by the minimal CMV immediate early gene without enhancer sequence (miniCMV) and a fusion protein (tTA) of Tet repressor (TetR) to transcriptional activator VP16 protein from herpex simplex virus (Tet-OFF)[10] or a fusion protein (rtTA, reverse-tTA) of VP16 with TetR variants which functions in reverse mode (Tet-ON)[11]. In the absence of tetracycline or its derivative such as doxycycline (Dox), tTA from Tet-OFF system will dimerize and bind to TetO and mediate transcription, whereas rtTA from Tet-ON system remains unbound to TetO and results no

expression. Upon introduction of Dox, tTA binds to Dox, changes conformation, releases from TetO and stops mediating transcription, whereas rtTA binds to Dox, changes conformation and is able to bind to TetO and mediates transcription. However, significant reduction of Dox sensitivity of rtTA in comparison to tTA was reported[12] and triggered immense effort to improve the system.

Efforts in optimization of Tet-ON system include systematic modification of pTet promoter[13,14], extensive mutagenesis of tTA to isolate an improved function rtTA such as rtTA3 which harbors six mutations and a tandem of 12 amino acid sequence of VP16 activation domain instead the classic 127 aa, resulting 25-fold more sensitive to Dox[15,16].

Additional efforts focusing on using additional regulators such as a chicken b-globin insulator element to restrict viral enhancer effect on the inducible promoter in a 1-unit tet-inducible transgene expression AAV construct[17] was also reported. Development of additional Tet-inducible repressor such as fusion protein of TetR-KRAB domain of Kox1[18] or TetR-HDAC4[19] to be used in combination with rtTA gave greater control in Tet-ON inducible system. However, fusion protein of TetR-eukaryotic repressors potentially posed unintended complications due to interaction of eukaryotic repressors with other host elements that might confound overall experimental results or side effects in therapy.

The inherent leakiness of the Tet-On system[20] poses a challenge for researchers in need for a truly stringent gene regulation system. Thus, we aimed to develop an improved Tet-ON system that increases Dox sensitivity and robust activation of transgene expression and yet decreases basal noise by focusing on pTet promoter design and dual Tet-responsive regulators in a single unit.

Here in this study, we demonstrate a novel design of Tet-ON inducible system called Long-SV40-TATA O-TetR-A vector that improves both expression robustness and level of leakiness as well as Dox sensitivity.

4.4 Results and Discussion

To overcome the high basal expression of a gene of interest from a widely used Tet-ON inducible system, we incorporate both tetracycline repressor (TetR) and tetracycline-controlled transactivator (rtTA), a variant of TetR fused with VP16, a single Tet-inducible expression vector.

Both of transcription factors, TetR and rtTA, are able to bind to tetracycline responsive element (TetO) and tetracycline or doxycycline but work in an opposite manner. In the absence of doxycycline, rtTA (or its third generation rtTA3) does not bind to TetO, therefore, does not activate transcription (Figure 4-1a, left); however, TetR binds to TetO and represses transcription (Figure 4-1a, middle). In the presence of doxycycline, rtTA binds to doxycycline and is able to bind TetO and activate transcription (Figure 4-1a, left) whereas TetR is unable to bind to TetO and thus no longer represses transcription (Figure 4-1a, middle). This unique regulatory property of rtTA and TetR on the same Tet responsive elements enables a tight control on expression of a gene of interest if expressing in the same cell or in the same expression cassette. This led to a novel design for a Tet-ON inducible system including pTet promoter (TetO, miniCMV promoter) to drive target gene expression, restriction sites for cloning in a gene of interest, two regulators-TetR and rtTA3 driven by full version of CMV promoter and separated by the internal ribosome entry sequence (IRES). This new expression vector is named the O-TetR-A system (Figure 4-1a, right).

To assess basal expression from a classic Tet-ON v.s. the O-TetR-A system, we constructed two expression vectors: (1) O-rtTA3 containing a single regulator-rtTA3, and (2) O-TetR-A containing both rtTA3 and TetR transcription factor regulators (Figure 4-1b). Both vectors share the same DNA backbone sequence, promoter (miniCMV) downstream of TetO and 6xHis-mCherry downstream the promoter except the O-TetR-A contains the IRES sequence and a second regulator, TetR, downstream of rtTA3. Due to a weak expression nature of the original TetO-TetR system, the O-TetR was excluded from this study.

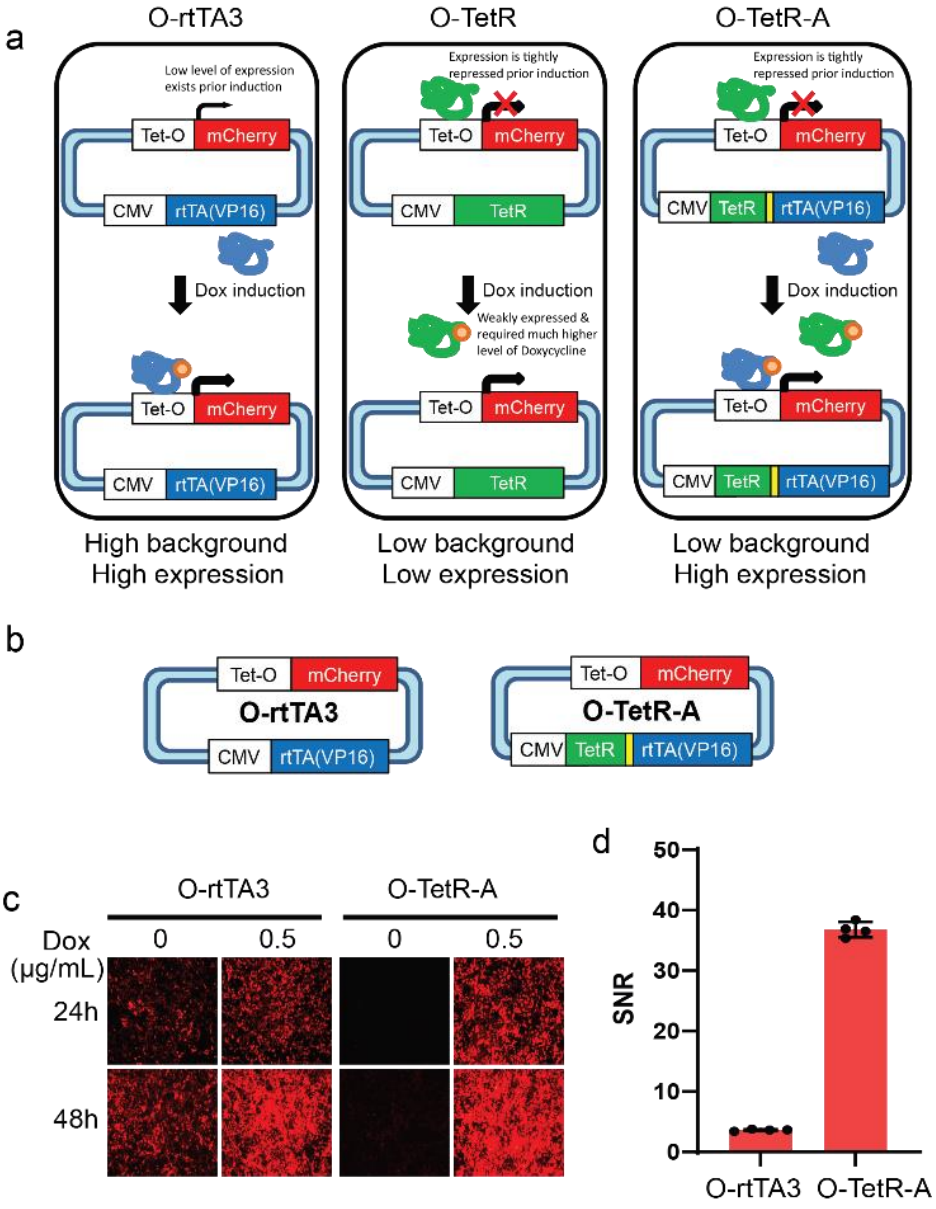


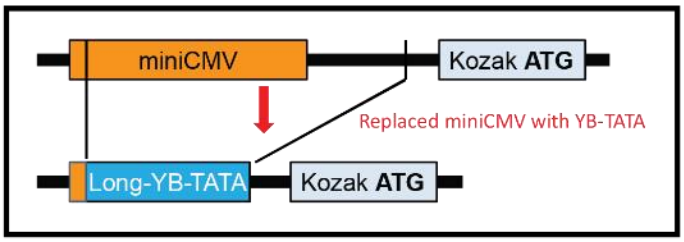
Figure 4-1. Novel design of the O-TetR-A (Tet-ON) expression system. A. Schematic of different Tet-ON inducible systems. **B**. Schematic of mCherry expression O-rtTA3 and O-TetR-A vectors. Yellow region represents IRES sequence. **C**. Microscopy analysis of mCherry expression in un-induced and induced (+/- 0.5 µg/mL Dox) HEK293 cells transfected with either vectors shown in (b). Un-induced cells were exposed for 2 seconds to capture all lowly expressed mCherry signal. **D.** Flow analysis to assess signal to noise (SNR) level of the vectors shown in (B). TetO: Tet Operon, TetR: repressor, rtTA(VP16): reverse tetracycline transactivator virion protein 16, mCherry: red fluorescent protein, single regulator containing O-rtTA3: TetR & transactivator vers 3 and O-TetR: TetO & repressor, dual regulator containing O-TetR-A: TetO-TetR-rtTA3. Flow replicates: n = 4. Error bars are standard deviation of the mean.

HEK293 cells were transfected with either O-rtTA3 or O-TetR-A mCherry expression vectors and induced with 0.5µg/mL Doxycycline (Dox) for 24h and 48h. mCherry expression signal was captured using Iive imaging at two different exposure times: 395 miliseconds (short) and 2 seconds (long). Both the O-rtTA3 and O-TetR-A transfected cells showed robust red fluorescent signal at 24h and 48h post induction under short exposure. For the corresponding un-induced cells, using the short exposure setting we did not see any difference between the two systems; however, when the imaging was taken at long exposure time, a strikingly intense red fluorescent signal was observed in cells transfected with the O-rtTA3 but not with O-TetR-A system (Figure 4-1c). To quantify basal expression of the two systems, a similar experiment was performed but the cells were subjected to flow cytometry and the ratio of raw mCherry values of induced to un-induced (SNR) was taken for

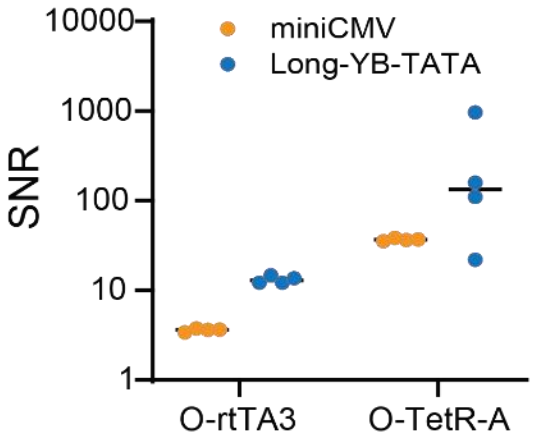
assessment of the basal expression of the two systems. As shown in Figure 4-1d, by just including a second regulator, TetR, as in the O-TetR-A vector, the SNR is improved by 10 folds compared to the classic Tet-ON system.

Synthetic promoter further improves the O-TetR-A basal expression

a



b



c

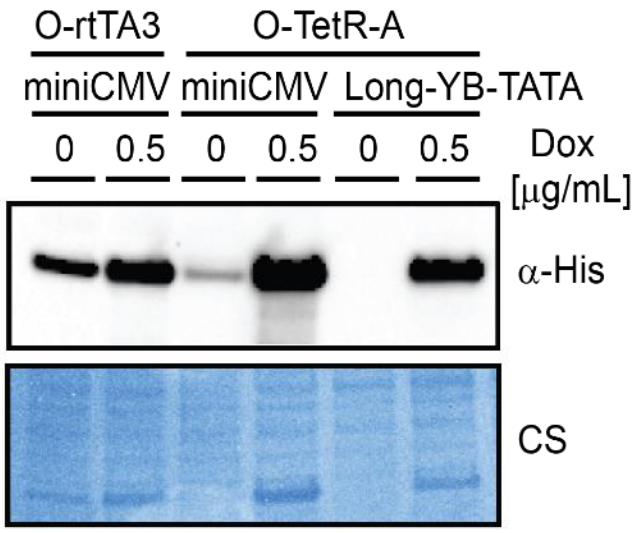


Figure 4-2. Assessment 6xHis-mCherry expression driven by a synthetic promoter from the dual regulators-containing O-TetR-A (Tet-ON) system. A. Schematic of swapping promoters (miniCMV to Long-YB-TATA) in the mCherry expressing O-TetR-A system. **B**. Flow analysis: to assess mCherry signal driven by miniCMV and Long-YB-TATA promoters in different Tet-ON inducible vector designs (O-rtTA3 vs O-TetR-A). **C**. Immunoblot analysis to assess 6xHis-mCherry level from un-induced and induced cells transfected with different vectors. 2.5µg total protein/lane. a-His: 6xHis Antibody, CS: Coomassie stain serving as loading control. Flow replicates: n = 4. Error bars are standard deviation of the mean.

MiniCMV is derived from a robust constitutive CMV promoter and widely used in the Tet-ON system but it also inherently comes with high basal expression level when utilized in an inducible vector context. Such promoter design would not be a viable option for temporally controlling expression of a toxic gene. Thus, to further cut down basal expression of the O-TetR-A, we next focused on promoter design which normally sits downstream of TetO.

Studies by Ede, C et al.[21] and Hansen, J et al.[22] showed that the synthetic YB-TATA promoter, originally designed by Benenson's laboratory, has very low level of basal expression comparing to the miniCMV in various inducible vectors. Thus, we replaced the entire miniCMV sequence with the YB-TATA promoter sequence (Figure 4-2a) resulting Long-YB-TATA vectors in both versions of the tet-inducible mCherry expression O-rtTA3 and the O-TetR-A. These Long-YB-TATA vectors were used to transfect HEK293 cells. After 48h

post induction with 0.5μg/mL Dox, cells were harvested for flow cytometry and immunoblot analyses.

By flow cytometry, both single (O-rtTA3) and dual (O-TetR-A) regulators containing expression vectors showed a 3-fold increase in SNR indicating a reduction in basal expression of Long-YB-TATA promoter (Figure 4-2b). To further evaluate background signal, we took advantage of the signal amplification nature of the sandwich immunoblot approach to stringently assess any basal signal from these constructs. Total protein from transfected cells was isolated and subjected to immunoblot analysis using 2.5μg/sample. The blot was probed with anti-His Ab and displayed robust 6xHis-mCherry signal in all induced samples as well as the un-induced sample from the classic single regulator containing O-rtTA3 system. However, the un-induced miniCMV containing O-TetR-A vector showed a much lower signal of 6xHis-mCherry, reconfirming the advantage of the dual regulator containing O-TetR-A system over the single regulator containing O-rtTA3 system in noise reduction. Excitingly, 6xHis-mCherry signal was not detectable in the un-induced sample derived from cells transfected with the synthetic Long-YB-TATA containing O-TetR-A vector (Figure 4-2c). Together, these results and differences of the two promoters including (1) TATA box, (2) the primary sequences and (3) the distance between the TATA box and ATG translation start codon (Table S1) indicate the contribution of these differences in reduction of basal expression.

Novel promoter designs enable tightly and robustly controlled gene expression

To further improve the O-TetR-A expression system, we aimed to develop a minimal promoter that gives lowest basal expression but robustly expresses upon induction by tackling the differences of promoters discussed above. To that end, first we focused on our

best working Long-YB-TATA promoter in the O-TetR-A expression vector by further trimming the distance between its TATA box and the start codon to a bare minimal which resulted the minYB-TATA promoter (Table S1). We reasoned that by decreasing the distance between TATA box and the start codon, we would also reduce features, indicative of the highest promoter activity normally located within the region of 75 nt upstream and 50 nt downstream of the main transcription start site[23]; therefore, we would further decrease basal noise. Also, for clarification, from this point forward all vectors described in this study contain dual regulators (TetR and rtTA3), which constitutes the O-TetR-A Tet-ON inducible system.

As we shortened the promoter, we primarily addressed two of the three aforementioned promoter differences, specifically, # 3 and also # 2 regarding the primary sequence variation between Long-YB-TATA and minYB-TATA. In our new vector design, the minYB-TATA promoter has 33 nt instead of 61 nt as in the original, counting from the first nucleotide of its TATA box to the last base right upstream of the start codon (Table S1). Accompanying with shortening the promoter, the T-richness signal, normally located upstream to a start codon and associated with high promoter activity in constitutive & regulated genes[23], also decreases. Therefore, the constitutive expressing nature of the promoter (or the leakiness) is predicted to reduce with mCherry signal highest in miniCMV (13 Ts) then YB-TATA (7 Ts) and lastly minYB-TATA (2 Ts).

By microscopy and immunoblot analyses, HEK 293T cells transfected with minYB-TATA showed no change or below detectable signal in basal expression but also decreased in the overall induced mCherry signal (Figure 4-2c and Figure 4-S1), indicating the new design reached the distance and T-richness limit for optimal promoter activity. Overall, the

observation agrees with our prediction, and also this vector can be used as a starting promoter design for screening TATA box sequences (aforementioned promoter difference #1) contributing to highest promoter activity.

Instead of relying on typical reported TATA box consensus sequences of 5-7 nt, we used 8 nt consensus sequences supported by the evidence of the TATA binding protein (TBP) crystal structure in complex with 8 nt TATA box[24,25] and by the report of Basehoar et al.[26]. We selected three TATA box candidates derived from viral promoters: CMV, SV40 and Thymidine kinase-TK from herpes simplex virus that have been widely adapted in transgene expression as shown in Figure 4-3a. We replaced the 8 nt synthetic YB-TATA in the minYB-TATA promoter with these 8 nt TATA box variants resulting three new mCherry expression vectors: (a) CMV-TATA, (b) SV40-TATA and (c) TK-TATA (Table S2). Along with minYB-TATA vector, they were subjected to microscopy, immunoblot analysis and flow cytometry.

Both CMV and YB-TATA boxes share seven identical nucleotides except the 8th position. Given everything else in the minimal promoters stayed the same and yet with just one residue changed: a "G" in the CMV and a "T" in the synthetic YB-TATA, a strike contrast in mCherry expression was observed. In this promoter context, a "G" at the 8th position of the CMV TATA box sequence (TATATAA**G**) is detrimental to promoter function, suggesting the interaction between TBP and TATA box was at suboptimal level. The result from this experiment also showcased that the 8 nt TATA box with a "T" at the last position from YB-TATA (TATATAA**T**) is the candidate of choice (Figure 4-3b-c). Surprisingly, both TK-TATA (ATATTAAG) and SV-TATA box (TATTTATG) sequences which are deviated from the canonical TATA box consensus sequences also showed high level of mCherry expression. Strikingly, SV40-TATA displayed much higher signal than minYB-TATA. All, as predicted,

showed little or no detectable signal in un-induced (0 Dox), even when imaging was performed with long exposure (2 seconds) (Figure 4-3c).

To isolate the best TATA box that demonstrates both robust mCherry expression and minimal leakiness, we analyzed mCherry expression from un-induced and induced cells transfected with the above vectors by flow cytometry. A clear separation of an excellent SV40-TATA box from the rest was illustrated: the signal-to-noise ratio (SNR) of SV40-TATA promoter is 1000 fold and twice as high as the SNR of minYB-TATA and TK-TATA (Figure 4-3d, Table S1). Collectively, we have obtained a better design of the Tet-ON inducible O-TetR-A system.

As previously demonstrated, the 61nt version of Long-YB-TATA promoter, facilitating higher mCherry expression without elevating observable leakiness than minYB-TATA, (Figure 4-S1) led us to revisit this promoter design. In this experiment, we replaced YB-TATA box with the best non-canonical SV40-TATA box sequence (Figure 4-3e) resulting Long-SV40-TATA. HEK293 was transfected with either Long-SV40-TATA or Long-YB-TATA vector and subjected to imaging and flow cytometry analyses.

Imaging results showed that mCherry signal driven by Long-YB-TATA promoter was higher than the signal from Long-SV40-TATA when induced with Dox (0.5μg/mL). In order to detect basal expression driven by the two promoters, images were taken under long (2 seconds) exposure and revealed low but detectable red fluorescence signal from the un-induced Long-YB-TATA transfected cells. Under the same experimental conditions, un-induced Long-SV40-TATA transfected cells showed little or below detectable level of mCherry (Figure 4-3f). HEK293 cells transfected with Long-SV40-TATA, SV40-TATA (short), Long-YB-TATA or minYB-TATA were analyzed via flow cytometry 48h post induction.

Consistent with imaging data, flow results showed higher SNRs for both SV40-TATA versions than those of the two YB-TATA vectors. These results indicate that the noise level from un-induced cells transfected with either versions of YB-TATA is much higher than that of SV40-TATA. Collectively, the Long-SV40-TATA promoter design is the best candidate – with robust expression upon induction and very low basal expression when un-induced (Figure 4-3g).

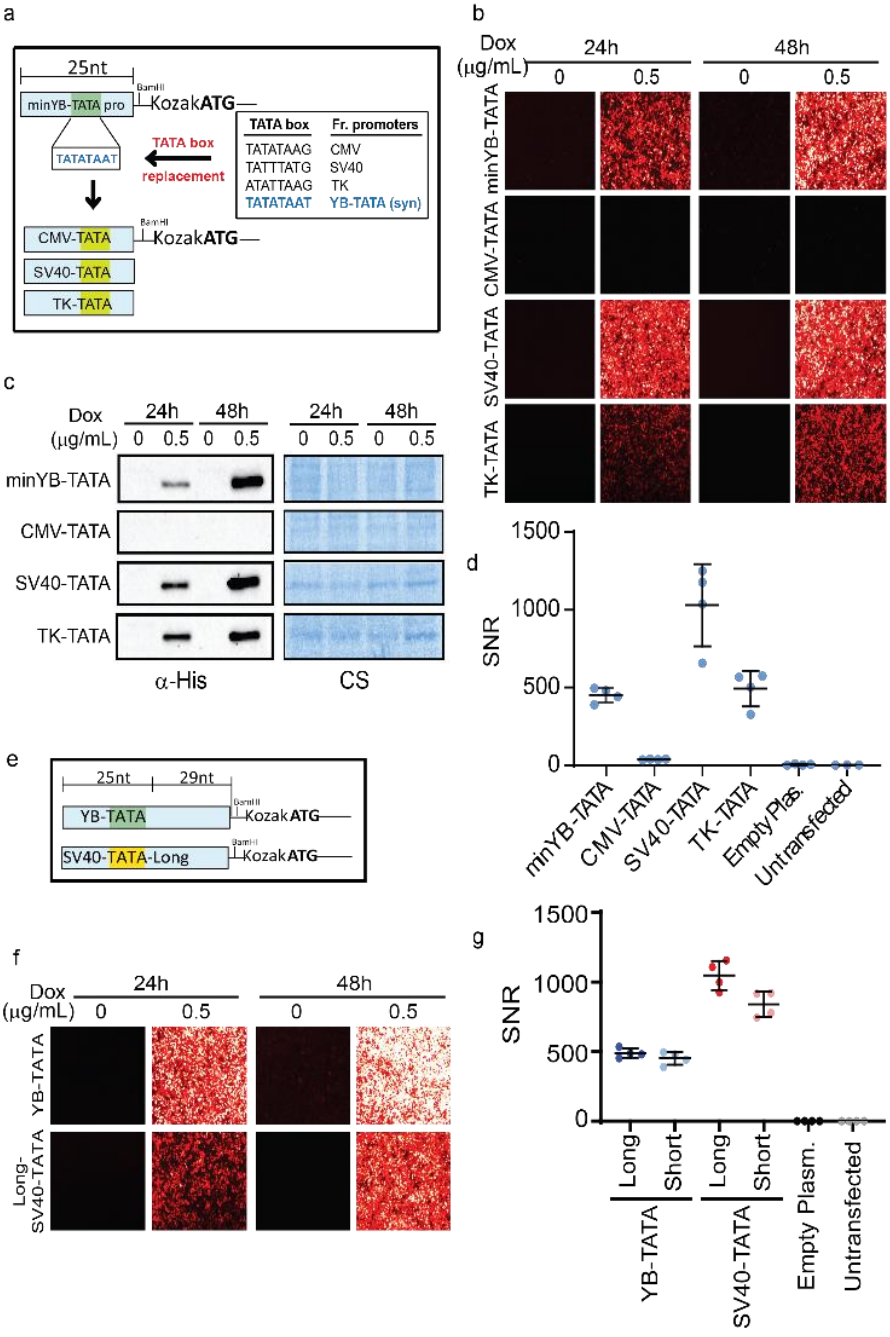


Figure 4-3. Optimization of the O-TetR-A system with novel designed promoters. A & **E.** Schematic of strategy used to replace TATA box variants. **B.** & **F.** Microscopy analysis of mCherry expression under the same promoter design except bearing different TATA boxes (b) or under different promoter primary sequences but sharing the same SV40-TATA box (f). **C**. Immunoblot analysis of 6xHismCherry from total protein of cells transfected with different expression vectors shown in (a) and (b). **D** & **G**. Flow analysis of uninduced and induced HEK293 cells transfected with different promoter designs for assessment of signal-to-noise ratio. 0: un-induced and 0.5µg/mL Dox: induced condition. α-His: His Ab, CS: Coomassie stain. Flow replicates: n = 4. Error bars are standard deviation of the mean.

Titration of Doxycycline concentration

Tetracycline or its derivative doxycycline is an effective inducer for Tet inducible system. However, it was shown to interfere with mitochondrial functions in different eukaryotic study models[27,28]. Consequently, disrupted mitochondria could lead to confounding experimental results. To minimize side effects of doxycycline, we aimed to identify the lowest concentration of Dox that is sufficient for inducing robust expression of a target gene by performing doxycycline titration experiment. HEK cells were transfected with mCherry expression Long-SV40-TATA, Long-YB-TATA, or an empty O-TetR-A plasmid, and induced for 48h with different doxycycline concentrations ranging from 8 to 1500 ng/mL prior being subjected to flow analysis.

As shown in Figure 4-4a, Dox concentration of at least 0.063 µg/mL is needed to induce mCherry expression from Long-SV40-TATA vector above the background, but beyond 0.5 µg/mL no observable improvement of mCherry expression was detected; (Figure

4-4a, b). This range of working concentration of Dox is about 20 folds less than the recommended usage concentration (0.1-10μg/mL) for controlling transcription by Tet-ON/Tet-OFF system in cell culture[29].

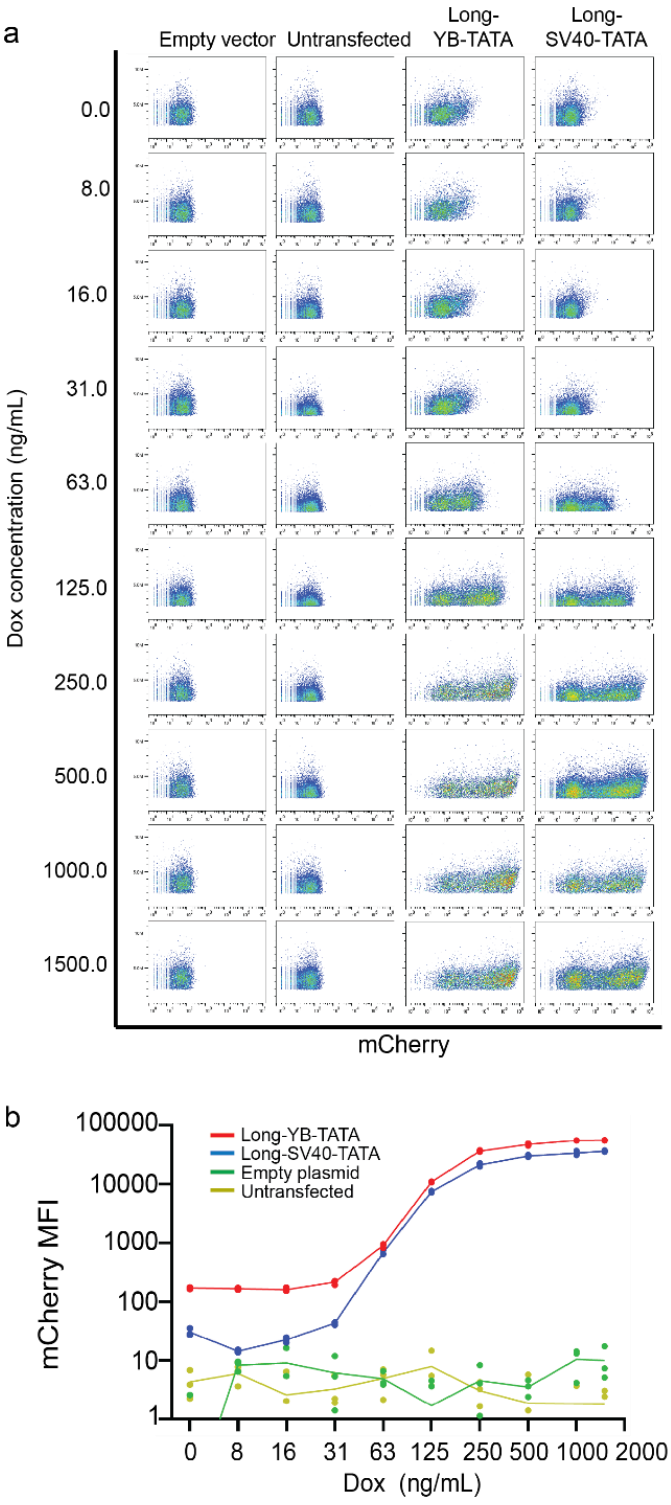


Figure 4-4. Titration of Doxycycline concentration for mCherry induction by the O-TetR-A vectors bearing different promoter designs. **A.** Flow analysis: profile of cell distribution(x axis = Y586::PE-H, y axis = FSC-H) based on mCherry expression signal at each Dox concentration (0.008 to 1.5μg/mL). **B.** Quantification of induced mCherry signal from different O-TetR-A vectors and controls plasmid at different Dox concentrations. MFI: mean fluorescent intensity, Flow replicates: n = 3. Error bars are standard deviation of the mean.

Similar to Long-SV40-TATA, Long-YB-TATA vector requires the same working concentration of Dox, ranging from at least 0.063 - 0.5 μg/mL, to induce mCherry expression above noise level. In comparison to Long-SV40-TATA under un-induced condition, mCherry expression from Long-YB-TATA vector was five times greater, indicating a higher basal expression of Long-YB-TATA vector (Figure 4-4a, b). As previously shown in Figure 4-2c, Long-YB-TATA already outperforms both the miniCMV-containing O-TetR-A and the classic Tet-ON O-rtTA3 vectors, yet its performance is still suboptimal when comparing to Long-SV40-TATA vector. The result shown here is consistent with previous data from Long-YB-TATA vector (Figure 4-3g). In short, we have established a range of Dox working concentrations for the O-TetR-A system that is lower than the currently recommended concentration for Tet-inducible systems. Consequently, potential confounding effects caused by Doxycycline toxicity on mitochondrial functions due to usage of high Dox concentration can be reduced.

Long-SV40-TATA O-TetR-A system withstands stringent challenge in expression Diptheria toxin

To further challenge the Long-SV40-TATA O-TetR-A system, we replaced mCherry coding sequence with an exotoxin gene encoded by a lysogenic corynephage β which infected all pathogenic bacterial strains, *Corynebacterium diptheriae*[30–33]. The gene encodes for Diptheria toxin fragment A (DTA) which inhibits protein synthesis in host cells when expressed by inactivating an essential factor in protein synthesis, eukaryotic elongation factor 2 (eEF2) via ADP-ribosylation a dipthamide residue of eEF2[34–37]. In this experiment, we also cloned DTA into miniCMV promoter containing O-TetR-A vector and Long-SV40-TATA promoter containing O-rtTA3 for comparison (Figure 4-5a).

Live cell imaging analysis of HEK293 cells transfected with one of the three DTA-containing plasmids demonstrated how well the novel design of the Long-SV40-TATA O-TetR-A vector performed. Without induction (Figure 4-5B-left panel), both constructs of the Long-SV40-TATA O-TetR-A and the empty vector control (top & bottom) had high density of cells indicating DTA toxin had not expressed. However, a great drop in cell density was observed in cells transfected with either the Long-SV40-TATA promoter in single regulator-containing O-rtTA3 plasmid or the miniCMV promoter in dual regulator-containing O-TetR-A vector, suggesting DTA toxin had been expressed and caused cell death even when un-induced. The right panel of Figure 4-5b showed cell density dropped in all except the negative control empty vector, demonstrating deathly effect of DTA toxin when expressed. To quantify toxic effect of DTA toxin, flow analysis was employed. HEK293 cells were transfected with both Lonza pMAX-GFP and one of the DTA-containing plasmids or DTA-free (empty control) vector. After 72h post induction, cells were subjected to flow cytometry analysis: high or low GFP signal serves as a visual tracker of functional or dysfunctional protein synthesis, respectively. Without induction, DTA toxin should not express, thus GFP

signal will be high (Figure 4-5C-Left, dark green bars), and upon induction, DTA toxin expresses and disrupts protein synthesis, thus causing reduction of GFP signal. Since empty vector control has no DTA, no reduction of GFP was observed (Figure 4-5C-Left light green bars); however, comparing to empty vector, Long-SV40-TATA O-TetR-A exhibited reduction of GFP signal suggesting low level noise remained. When comparing to the other two DTA-containing vectors in un-induced cells, Long-SV40-TATA O-TetR-A decreased basal expression by 5 folds than the O-rtTA3 version simply due to the presence of additional TetR regulator and 8 folds than the miniCMV O-TetR-A vector due to different promoter.

Upon induction, the reduction of GFP signal from Long-SV40-TATA (O-TetR-A) was not as low comparing to the other constructs (Figure 4-5C-Left, light green bars). Perhaps, due to robust induction of DTA expression by the Long-SV40-TATA O-TetR-A vector, the toxin effectively inhibits the host protein synthesis and consequently inhibits its own production.

Because DTA is an extremely toxic protein, with one molecule being sufficient to shut down protein production and kill the cell[38], the tight control of DTA expression is especially important. We have shown that Long-SV40-TATA enables the most stringent expression control of DTA. Collectively, in this study our data demonstrated the Tet-ON inducible design of Long-SV40-TATA O-TetR-A system comprising of both novel design of promoter and a combination of dual regulators (TetR and rtTA3) provides robust induction of target gene expression accompanied with minimal noise by tight control of leakiness. In addition, potential confounding effects due to toxicity of Doxycycline to mitochondrial functions is also lowered due to reduction of Doxycycline concentration required for induction. The Long-SV40-TATA O-TetR-A system will facilitate studies that require tight temporal control such

as T-cell activation regulation and/or are sensitive to minute expression such as suicide gene expression.

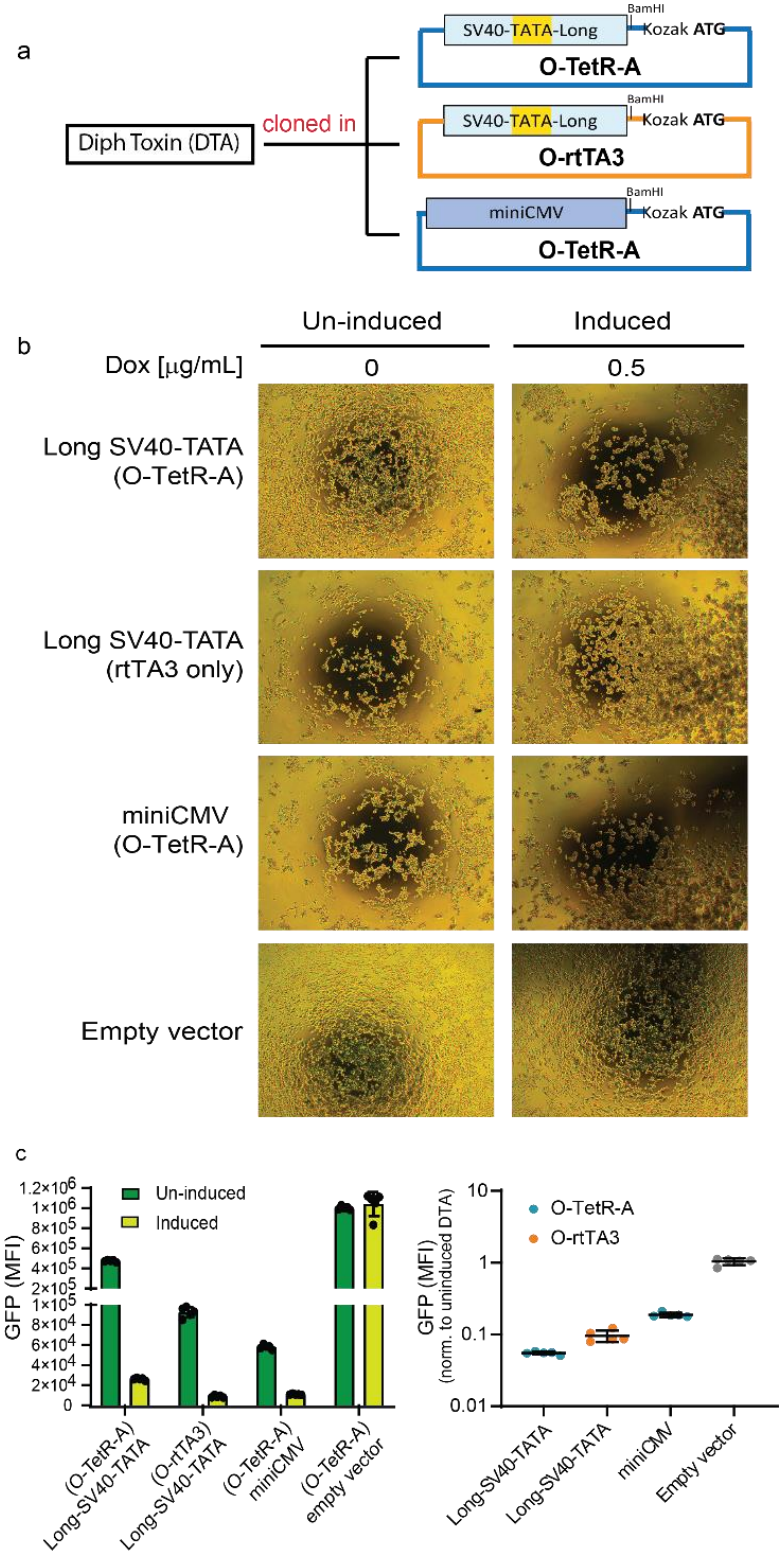


Figure 4-5. Expression of Diphteria toxin (DTA) from different designs of novel Tet-ON inducible vectors. **A**. Schematic of DTA-containing vectors. **B**. Microscopy analysis: assessment of cell density upon un-induced and induced expression of different DTA-containing vectors. **C.** Flow analysis: **(Left)** HEK293 cells transfected with both GFP expression plasmid and an inducible DTA-containing vector to visually assess protein synthesis functionality via raw green fluorescent intensity upon induced and un-induced DTA expression, **(Right)** GFP signal from DTA-induced cells was normalized to GFP signal from DTA-un-induced cells to assess expression robustness and leakiness of different designs of inducible DTA-containing vectors. Flow replicates, n = 5, Error bars are standard deviation of the mean.

4.5 Methods

Plasmid cloning

All plasmids used in this study are listed in Table S2.

a. Construction of O-rtTA3 (single regulator) vector: Sequence of CMV promoter and rtTA3 from pLenti CMV rtTA3a Blast (w756-1) (a gift from Dr. Eric Campeau, Addgene#26429) and bGH-PA terminator from pCDNA3.3-GFP (Addgene # 13031) were amplified individually then fused together via PCR. Fusion amplicons were cloned into pTREtight2 (a gift from Dr. Markus Ralser, Addegene# 19407) via PciI site to make pKN411 = version 1 of O-rtTA3 vector. *Escherichia coli* cyclic AMP receptor protein (CAP) binding sequence located upstream of TetO in pKN411 was removed, resulting pKN426 = O-rtTA3.

b. Construction of O-TetR-A (dual regulators) vector: Sequence of TetR from pJZ300 and sequence of IRES from pWIG-nFLAG-PGA-IRES-EGFP (a gift from Dr Michael S. Cohen at

Oregon Health & Science University) were amplified individually then fused together via PCR. Fusion amplicons were cloned into pKN426 via BsrGI to make pKN430 = O-TetR-A vector.

c. Construction of YB-TATA promoter in O-rtTA3 and O-TetR-A vectors via two steps:

Step 1: miniCMV sequence located downstream TetO of pTREtight2 was removed by digesting with RsaI resulting 3 fragments (71nt, 923 nt & 1938nt). The two large fragments were gel purified and self-ligated to become pKN437. YB-TATA promoter (long version, dsOligo oKN586F & 587R) was cloned into pKN437 via XmaI to become pKN438.

Step 2: TetO-YB-TATA promoter fragment was cleaved from pKN438 then gel-purified and subsequently cloned into pKN426 and pKN430 via PspXI and BamHI to make pKN442 (YB-TATA O-rtTA3) and pKN443 (YB-TATA O-TetR-A).

d. Construction of different TATA boxes into O-TetR-A backbone using the same approach described in (c).

e. Construction of Diptheria toxin containing vectors: “Long SV40 TATA OTetRA DTA" (plJW0036), "Long SV40 TATA rtTA3 only DTA" (plJW0037), "miniCMV OTetRA DTA" (plJW0038) were cloned by swapping mCherry for Diphtheria Toxin fragment A (from addgene#58536) in pKN522, pKN427, and pKN431, via Gibson Assembly.

Cell culture, transfection and Dox induction

Cell cultures: HEK293 cells from ATCC were cultured in DMEM (Corning, Cat#: 10-017-CM) supplemented with 10% FBS, 1% (1 mg/mL) penicillin and streptomycin and grown at 37°C, 5% CO_2.

E. coli NEBStable cells and grown in LB medium at 37°C with ampicillin at 100 μg/mL.

Transfection: After 48h seeding of $2.5x10^5$ cells/well, HEK293 cells in 6-well plates were transfected with 1ug desirable plasmid using JetPrime Transfection reagent (Polyplus Transfection, France) according to manual instruction except adding 5x higher than the manual recommended volume of JetPrime Reagent.

Induction: Right after adding transfection reagent to cells, Dox was supplemented directly to cell medium in each well at final concentration of 0.5µg/mL.

Microscopy

24h or 48h post induction, live cell Imaging was performed using Leica microscope, 10x Objective, exposure time: 2 seconds (long).

Protein isolation and Immunoblot analysis

24h or 48h post induction, and imaging, cells were scraped off plates and spun down at 1000xg for 5min at room temperature. Cell pellets subjected to protein isolation were re-suspended with 100 -150 µL (10x volume of dried cell pellet) RIPA lysis buffer [50mM Tris-HCl at pH7.5, 150mM NaCl, 0.1% Triton, 0.5% Na-deoxycholate & 0.1%SDS, 5mM EDTA), supplemented with 4 µL of proteinase inhibitor cocktail Set VII (EMB, Cat#: 539138-1ML), 4 µL Turbo DNase (Life Tech., Cat# AM2239), 4 µL RNaseA (Thermo Fisher, Cat# FEREN0531) per milliliter of RIPA lysis buffer], was added. Cells were re-suspended and incubated on ice for 0.5h then subjected to 5 cycles of freeze-thaw using dry-ice and 22°C water bath, then followed with cell shearing (15 times) to further break cells using 20G x1 ½ (0.9mm x 40mm) needle (BD PrecisionGlide Needle, Ref# 305179) and 1mL syringe. Total protein lysate was cleared by centrifugation at 15K rpm for 30min at 4°C. The cleared lystate was subjected to BCA assay (BioSciences, Cat# 786-570) to determine protein concentration according to manufacturer instruction at 37°C for 30 min and quantified using NanoDrop BCA program.

SDS-PAGE analysis was performed: 2.5 µg of total protein per well was resolved in 4-20% gradient 10-well MiniProtean gel (Bio-Rad, Cat# 4561094). Proteins were transferred onto nitrocellulose membrane using Trans-Blot Turbo Transfer unit (Bio-Rad, Cat#1704150). Membranes were blocked with 5% non-fat dry milk in 1xPBST (1xPBS: 137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4 + 0.1%Tween-20) for 1h at room temperature and probed overnight at 4°C with fresh blocking buffer + mouse anti-His (GenScript, Cat#: A00186-100) at 1:2000dil. Blot was washed 3x10min/each with 1xPBST at room temperature (RT) and incubated in 1xPBST+ 5% non-fat dry milk + anti-mouse conjugated Horseradish peroxidase (1:20000 dilution) (Cell Signaling, MA, USA; Cat# 7076S) for 1h at RT. Finally, blot was washed 3x10min/each with 1xPBST at RT and imaged using ECL Chemiluminescent Substrate (Fisher Scientific, Cat# PI32106) and imaged on a ChemiDoc MP imaging system (Bio-Rad).

Flow cytometry

24h or 48h post induction, cells were lifted with 0.5% Trypsin-EDTA (Fisher Scientific, Cat# 15400054) and re-suspended in room-temperature flow buffer (1xPBS with 5% FBS). Fluorescence was measured in the NovoCyte Flow Cytometer (Agilent).

4.6 Supplemental Data

Table 4-S1: Promoter sequences.

Promoter	Sequence (5' -> 3')
miniCMV	GTAGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGT GAACCGTCAGATCGCCTGGAGAATTCGAGCTCGGTACCCGGGGATC CGCCACCATG
Long-YB-TATA	GTAGGCGTGTACCCGGTCTAGAGGGTATATAATGGGGGCCACCGGT CTGGGTAAAGCCGAATTCCTAGCGCCGGGGATCCGCCACCATG

minYB-TATA	GTAGGCGTGTACCCGGTCTAGAGGGTATATAATGGGGGGCCACCGGT GGATCCGCCACCATG
CMV-TATA	GTAGGCGTGTACCCGGTCTAGAGGGTATATAAGGGGGGCCACCGGT GGATCCGCCACCATG
Long-SV40-TATA	GTAGGCGTGTACCCGGTCTAGAGGGTATTTATGGGGGGCCACCGGT CTGGGTAAAGCCGAATTCCTAGCGCCGGGGATCCGCCACCATG
SV40-TATA	GTAGGCGTGTACCCGGTCTAGAGGGTATTTATGGGGGGCCACCGGT GGATCCGCCACCATG
TK-TATA	GTAGGCGTGTACCCGGTCTAGAGGGATATTAAGGGGGGCCACCGGT GGATCCGCCACCATG

Underlined sequence in black color denotes flanking spacer derived from YB-TATA developed by Benenson's laboratory**.** TATA box in red. Kozak sequence in blue**.** Translational start codon in green.

Table 4-S2: Plasmids

Plasmid ID	Description	Backbone	Alternative name & Reference
pTREtight2	miniCMV promoter upstream of TetO	n/a	Addgene # 19407
pLenti CMV rtTA3a Blast	CMV driven reverse tetracycline transactivator VP16	n/a	Addgene # 26429
pJZ300	Tet repressor	n/a	This study
pKN411	Empty vector. miniCMV promoter downstream of TetO and CMV promoter driven rtTA3 with bGH-PA terminator	O-rtTA3	This study
pKN426	Empty vector. pKN11 without CAP binding site upstream TetO, miniCMV promoter downstream of TetO and CMV promoter driven rtTA3 with bGH-PA terminator	O-rtTA3	O-rtTA3. This study
pKN427	mCherry in pKN426 driven by miniCMV promoter downstream of TetO	O-rtTA3	miniCMV O-rtTA3. This study
pKN430	Empty vector. No CAP binding site upstream TetO, miniCMV promoter downstream of TetO and CMV promoter driven TetR-IRES-rtTA3 with bGH-PA terminator	O-rtTA3	O-TetR-A. This study
pKN431	mCherry in pKN430 (O-TetR-A) driven by miniCMV promoter downstream of TetO	O-TetR-A	miniCMV or miniCMV O-TetR-A. This study
pKN437	Intermediate vector. MiniCMV promoter was remove from pTREtight2	n/a	
pKN438	Intermediate vector. Long-YB-TATA was cloned downstream TetO of pKN437	n/a	
pKN439	mCherry driven by Long-YB-TATA promoter (O-TetR-A)	O-TetR-A	Long-YB-TATA or Long-YB-TATA O-TetR-A. This study
pKN442	Empty vector, Long-YB-TATA promoter downstream TetO (O-rtTA3)	O-rtTA3	This study
pKN443	Empty vector, Long-TATA promoter downstream TetO (O-TetR-A)	O-TetR-A	This study
pKN451	Empty vector, minYB-TATA promoter downstream TetO (O-TetR-A)	O-TetR-A	This study
pKN452	mCherry driven by minYB-TATA promoter downstream TetO (O-TetR-A)	O-TetR-A	minYB-TATA or minYB-TATA O-TetR-A. This study
pKN461	Empty vector, SV40-TATA promoter downstream TetO (O-TetR-A)	O-TetR-A	This study
pKN462	Empty vector, CMV-TATA promoter downstream TetO (O-TetR-A)	O-TetR-A	This study
pKN463	Empty vector, Long-YB-TATA promoter downstream TetO (O-TetR-A)	O-TetR-A	This study

Plasmid ID	Description	Backbone	Alternative name & Reference
pKN471	mCherry driven by SV40-TATA promoter downstream TetO (O-TetR-A)	O-TetR-A	SV40-TATA or SV40-TATA O-TetR-A. This study
pKN472	mCherry driven by CMV-TATA promoter downstream TetO (O-TetR-A)	O-TetR-A	CMV-TATA or CMV-TATA O-TetR-A. This study
pKN473	mCherry driven by TK-TATA promoter downstream TetO (O-TetR-A)	O-TetR-A	TK-TATA or TK-TATA O-TetR-A. This study
pKN495	Empty vector, Long-SV40-TATA promoter downstream TetO (O-TetR-A)	O-TetR-A	This study
pKN522	mCherry driven by Long-SV40-TATA promoter downstream TetO (O-TetR-A)	O-TetR-A	Long-SV40-TATA or Long-SV40-TATA O-TetR-A. This study
pJW0036	Diphteria toxin driven by Long-SV40-TATA promoter downstream TetO (O-TetR-A)	O-TetR-A	DTA-containing Long-SV40-TATA O-TetR-A. This study
pJW0037	Diphteria toxin driven by Long-SV40-TATA promoter downstream TetO (O-rtTA3)	O-rtTA3	DTA-containing Long-SV40-TATA O-rtTA3. This study
pJW0038	Diphteria toxin driven by miniCMV promoter downstream TetO (O-TetR-A)	O-TetR-A	DTA-containing miniCMV O-TetR-A. This study

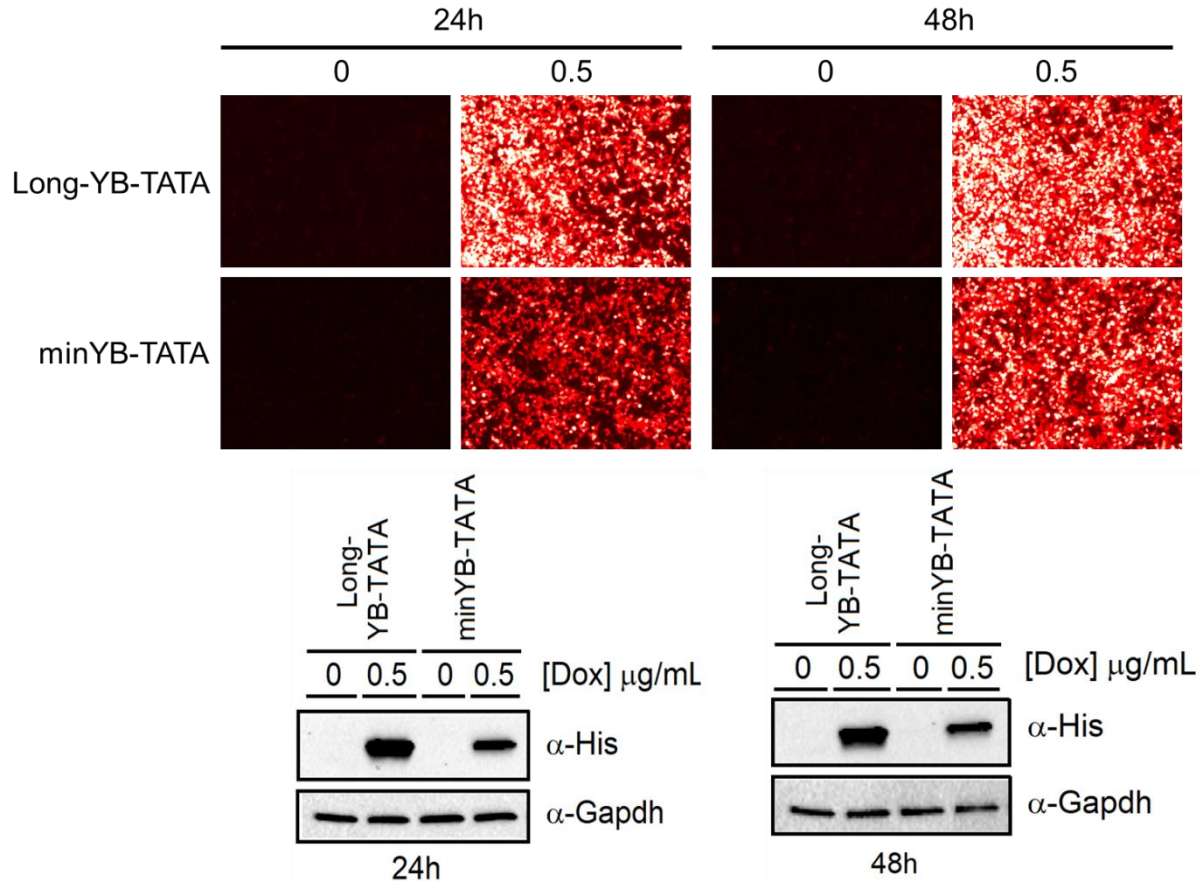


Figure 4-S1. Assessment promoter activity of Long-YB-TATA vs. minYB-TATA O-TetR-A vectors via mCherry expression. **A.** Microscopy analysis: Un-induced and induced HEK293 cells transfected with either 6xHis-mCherry containing Long-YB-TATA or minYB-TATA O-TetR-A vectors were imaged at 24h and 48h post induction using Leica microscopy, objective 10x, exposure for 2 seconds. **B.** Total protein extracted from cells described in (**A**) were subjected immunoblot analysis: 2.5μg total protein/lane from un-induced and induced cells. Probed with α-His and subsequent α-Gapdh (loading control) antibodies.

4.7 References

1. Tuszynski, M. H. *et al.* Nerve Growth Factor Gene Therapy: Activation of Neuronal Responses in Alzheimer Disease. *JAMA Neurol* **72**, 1139–1147 (2015).
2. Marks, W. J. *et al.* Gene delivery of AAV2-neurturin for Parkinson's disease: a double-blind, randomised, controlled trial. *Lancet Neurol* **9**, 1164–1172 (2010).
3. Cayouette, M., Behn, D., Sendtner, M., Lachapelle, P. & Gravel, C. Intraocular gene transfer of ciliary neurotrophic factor prevents death and increases responsiveness of rod photoreceptors in the retinal degeneration slow mouse. *J Neurosci* **18**, 9282–9293 (1998).
4. Nutt, J. G. *et al.* Randomized, double-blind trial of glial cell line-derived neurotrophic factor (GDNF) in PD. *Neurology* **60**, 69–73 (2003).
5. Manfredsson, F. P. *et al.* Nigrostriatal rAAV-mediated GDNF Overexpression Induces Robust Weight Loss in a Rat Model of Age-related Obesity. *Molecular Therapy: the Journal of the American Society of Gene Therapy* **17**, 980 (2009).
6. Brentjens, R., Yeh, R., Bernal, Y., Riviere, I. & Sadelain, M. Treatment of Chronic Lymphocytic Leukemia With Genetically Targeted Autologous T Cells: Case Report of an Unforeseen Adverse Event in a Phase I Clinical Trial. *Molecular Therapy* **18**, 666 (2010).
7. Kochenderfer, J. N. *et al.* B-cell depletion and remissions of malignancy along with cytokine-associated toxicity in a clinical trial of anti-CD19 chimeric-antigen-receptor-transduced T cells. *Blood* **119**, 2709–2720 (2012).
8. Morgan, R. A. *et al.* Case report of a serious adverse event following the administration of T cells transduced with a chimeric antigen receptor recognizing ERBB2. *Mol Ther* **18**, 843–851 (2010).

9. Linette, G. P. *et al.* Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. *Blood* **122**, 863–871 (2013).

10. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* **89**, 5547–5551 (1992).

11. Gossen, M. *et al.* Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766–1769 (1995).

12. Baron, U. & Bujard, H. Tet repressor-based system for regulated gene expression in eukaryotic cells: principles and advances. *Methods Enzymol* **327**, 401–421 (2000).

13. Loew, R., Heinz, N., Hampf, M., Bujard, H. & Gossen, M. Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol* **10**, (2010).

14. Agha-Mohammadi, S. *et al.* Second-generation tetracycline-regulatable promoter: repositioned tet operator elements optimize transactivator synergy while shorter minimal promoter offers tight basal leakiness. *J Gene Med* **6**, 817–828 (2004).

15. Markusic, D., Oude-Elferink, R., Das, A. T., Berkhout, B. & Seppen, J. Comparison of single regulated lentiviral vectors with rtTA expression driven by an autoregulatory loop or a constitutive promoter. *Nucleic Acids Research* **33**, e63 (2005).

16. Das, A. T. *et al.* Viral evolution as a tool to improve the tetracycline-regulated gene expression system. *J Biol Chem* **279**, 18776–18782 (2004).

17. Bell, A. C., West, A. G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387–396 (1999).

18. Deuschle, U., Meyer, W. K. & Thiesen, H. J. Tetracycline-reversible silencing of eukaryotic promoters. *Mol Cell Biol* **15**, 1907–1914 (1995).

19. Bockamp, E. *et al.* Generation and characterization of tTS-H4: a novel transcriptional repressor that is compatible with the reverse tetracycline-controlled TET-ON system. *J Gene Med* **9**, 308–318 (2007).

20. Costello, A. *et al.* Leaky Expression of the TET-On System Hinders Control of Endogenous miRNA Abundance. *Biotechnol J* **14**, (2019).

21. Ede, C., Chen, X., Lin, M. Y. & Chen, Y. Y. Quantitative Analyses of Core Promoters Enable Precise Engineering of Regulated Gene Expression in Mammalian Cells. *ACS Synth Biol* **5**, 395–404 (2016).

22. Hansen, J. *et al.* Transplantation of prokaryotic two-component signaling pathways into mammalian cells. *Proc Natl Acad Sci U S A* **111**, 15705–15710 (2014).

23. Lubliner, S., Keren, L. & Segal, E. Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. *Nucleic Acids Res* **41**, 5569–5581 (2013).

24. Kim, J. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527 (1993).

25. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520 (1993).

26. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699–709 (2004).

27. Moullan, N. *et al.* Tetracyclines Disturb Mitochondrial Function across Eukaryotic Models: A Call for Caution in Biomedical Research. *Cell Rep* **10**, 1681–1691 (2015).

28. Chatzispyrou, I. A., Held, N. M., Mouchiroud, L., Auwerx, J. & Houtkooper, R. H. Tetracycline antibiotics impair mitochondrial function and its experimental use confounds research. *Cancer Res* **75**, 4446–4449 (2015).

29. Sigma Aldrich. Doxycycline Hyclate Datasheet. *https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/402/297/d5207dat.pdf*.

30. Murphy, J. R. Corynebacterium Diphtheriae. in *Medical Microbiology, 4th edition* vol. 69 565–568 (University of Texas Medical Branch at Galveston, 1996).

31. FREEMAN, V. J. Studies on the virulence of bacteriophage-infected strains of Corynebacterium diphtheriae. *J Bacteriol* **61**, 675–688 (1951).

32. FREEMAN, V. J. & MORSE, I. U. Further observations on the change to virulence of bacteriophage-infected a virulent strains of Corynebacterium diphtheria. *J Bacteriol* **63**, 407–414 (1952).

33. Todar, K. Diphtheria. in *Todar's Online Textbook of Bacteriology* (2009).

34. Bell, C. E. & Eisenberg, D. Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide. *Biochemistry* **35**, 1137–1149 (1996).

35. Bennett, M. J. & Eisenberg, D. Refined structure of monomeric diphtheria toxin at 2.3 A resolution. *Protein Sci* **3**, 1464–1475 (1994).

36. Bennett, M. J., Choe, S. & Eisenberg, D. Refined structure of dimeric diphtheria toxin at 2.0 A resolution. *Protein Science : A Publication of the Protein Society* **3**, 1444 (1994).

37. Narrowe, A. B. *et al.* Complex Evolutionary History of Translation Elongation Factor 2 and Diphthamide Biosynthesis in Archaea and Parabasalids. *Genome Biol Evol* **10**, 2380–2393 (2018).

38. Yamaizumi, M., Mekada, E., Uchida, T. & Okada, Y. One molecule of diphtheria toxin fragment A introduced into a cell can kill the cell. *Cell* **15**, 245–250 (1978).

Chapter 5: Chromatin remodeling protein HELLS is critical for retinoblastoma tumor initiation and progression

Publication note

This paper was originally published in Oncogenesis.

Loredana Zocchi, Aditi Mehta, Stephanie C. Wu, Jie Wu, Yijun Gu, Jingtian Wang, Susie Suh, Robert C. Spitale, and Claudia A. Benavente. Chromatin remodeling protein HELLS is critical for retinoblastoma tumor initiation and progression. Oncogenesis **9**, 1-15 (2020).

5.1 Contribution Statement

Claudia Benavente conceived the project. Loredana Zocchi, Aditi Mehta, Stephanie C. Wu, Yijun Gu and Claudia Benavente performed the experiments. Jie Wu performed the P21 ATAC-Seq and RNA-seq data analysis. Jingtian Wang performed the tumor RNA-seq analysis (Figure 5-4D, Figure 5-5A, B, C). Susie Suh performed the ERG experiments and ERG data analysis. Robert Spitale supervised the tumor RNA-Seq data analysis. Loredana Zocchi, Aditi Mehta and Claudia Benavente wrote the paper.

5.2 Abstract

Retinoblastoma is an aggressive childhood cancer of the developing retina that initiates by biallelic *RB1* gene inactivation. Tumor progression in retinoblastoma is driven by epigenetics, as retinoblastoma genomes are stable, but the mechanism(s) that drive these epigenetic changes remain unknown. Lymphoid-specific helicase (HELLS) protein is an epigenetic modifier directly regulated by the RB/E2F pathway. In this study, we used

novel genetically engineered mouse models to investigate the role of HELLS during retinal development and tumorigenesis. Our results indicate that *Hells*-null retinal progenitor cells divide, undergo cell-fate specification, and give rise to fully laminated retinae with minor bipolar cells defects, but normal retinal function. Despite the apparent nonessential role of HELLS in retinal development, failure to transcriptionally repress *Hells* during retinal terminal differentiation due to retinoblastoma (RB) family loss significantly contributes to retinal tumorigenesis. Loss of HELLS drastically reduced ectopic division of differentiating cells in *Rb1/p107*-null retinae, significantly decreased the incidence of retinoblastoma, delayed tumor progression, and increased overall survival. Despite its role in heterochromatin formation, we found no evidence that *Hells* loss directly affected chromatin accessibility in the retina but functioned as transcriptional co-activator of E2F3, decreasing expression of cell cycle genes. We propose that HELLS is a critical downstream mediator of E2F-dependent ectopic proliferation in RB-null retinae. Together with the nontoxic effect of HELLS loss in the developing retina, our results suggest that HELLS and its downstream pathways could serve as potential therapeutic targets for retinoblastoma.

5.3 Introduction

The RB pathway can directly regulate genes that control cell cycle exit but can also regulate the expression of genes that control cell-fate specification and differentiation through mechanisms, including chromatin remodeling and epigenetic control[1]. In the developing retina, the RB pathway is critical in development and tumorigenesis. Genetic ablation of RB is characterized by alterations in ganglion cell, bipolar, and rod photoreceptor terminal differentiation that results in cell death of these retinal cells and

disruption of horizontal cell synaptogenesis[2,3]. Further, almost all retinoblastomas have *RB1* gene inactivation[4,5]. While genetic alterations in the *RB1* gene are required for tumor initiation, retinoblastoma tumors have stable genomes, with tumor progression occurring through epigenetic dysregulation of several cancer pathways[6].

HELLS (helicase, lymphoid specific; also known as LSH, ICF4, PASG, and SMARCA6), a gene transcriptionally controlled by the RB/E2F pathway[7–9] , encodes a chromatin remodeling protein thought to be responsible for the epigenetic changes seen in retinoblastoma and required for tumor survival[10]. HELLS is a SWI/SNF-related matrix-associated regulator of chromatin that contributes to global genome methylation[11]. HELLS remodels chromatin to render DNA accessible to DNA methyltransferase enzymes Dnmt3a or Dnmt3b, supporting de novo DNA methylation and stable gene silencing during cellular differentiation[12,13]. Binding of HELLS to DNMT1, HDAC1, and HDAC2 has also been reported[14]. HELLS plays an important role in normal development as HELLS knockout mice display perinatal lethality[15]. Furthermore, HELLS mutant mice exhibit signs of growth retardation, premature aging, and impaired neural stem/progenitor cells self-renewal and maintenance during development [16,17]. Thus, consideration of HELLS as a therapeutic target in the treatment of cancers requires careful consideration of the potential toxicities that could arise in tissues exposed to treatment.

The vertebrate neural retina is organized with remarkable precision into laminar structure formed by multiple types of neurons and glial cells. In mouse retina, the seven major retinal cell types (rod and cone photoreceptors; horizontal, amacrine, and bipolar interneurons; Müller glia; and ganglion cells) are differentiated from a common population of multipotent retinal progenitor cells (RPCs) that appear between embryonic day 11 (E11)

and postnatal day 10 (P10) in a conserved temporal order (Figure 5-1A)[18,19]. The process of self-renewal and differentiation required for RPCs to differentiate into the distinct neural lineages is regulated by various molecular mechanisms, including transcriptional and post-transcriptional regulation. Epigenetic regulation plays a fundamental role in the maintenance of cell identity as well as the stepwise control toward cellular differentiation. Chromatin states of RPCs change gradually during this process and alterations of epigenetic markers during retinal development have been linked to several ocular diseases, including retinoblastoma[6,10,20–23]. Thus, understanding the molecular mechanism and identifying key chromatin factors that regulate retinal stem cell maintenance and neurogenesis is critical for understanding normal retinal development and discovering molecular pathways that could be targeted for cancer therapy and intraocular delivery.

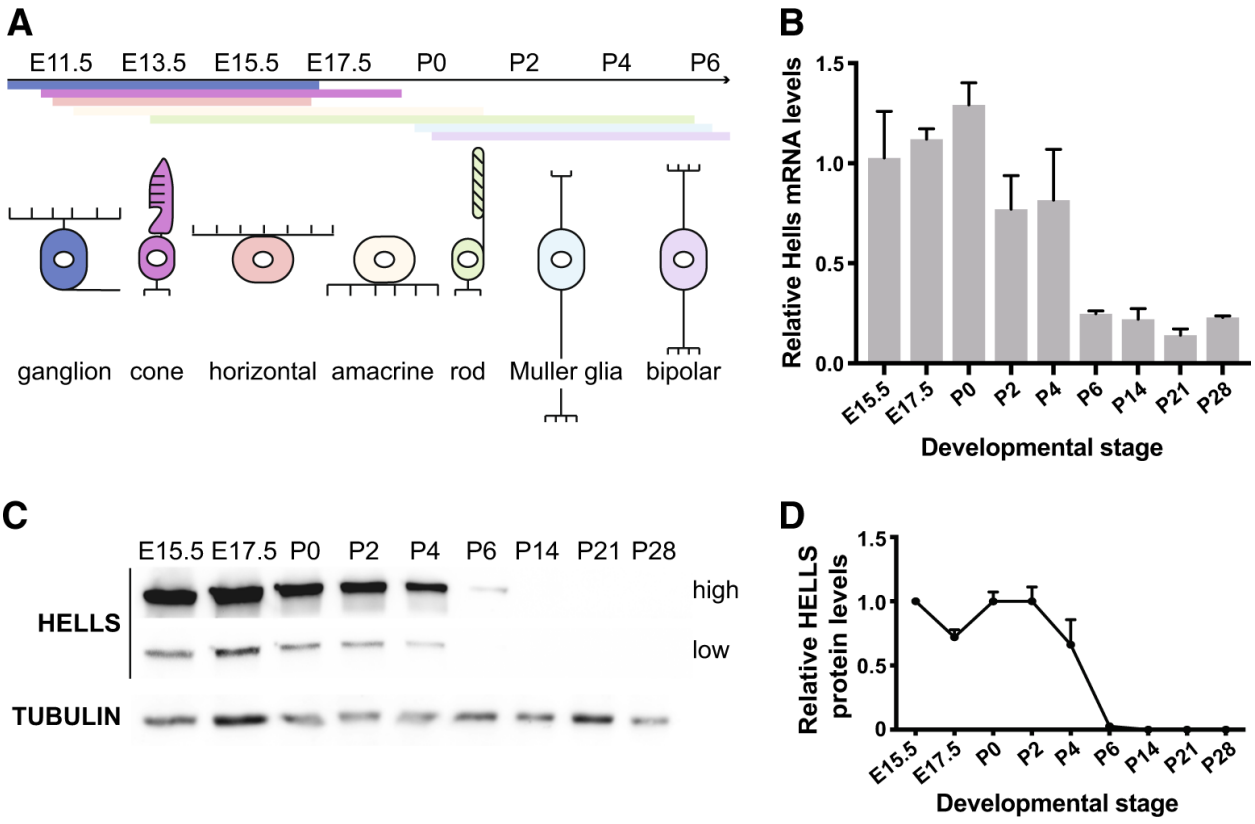


Figure 5-1. HELLS is repressed during the late stages of retinal development. A. Illustration of the developmental stages and specification timing of the seven major retinal cell types: retinal ganglion cells, amacrine, horizontal, bipolar, Müller, cone, and rod photoreceptors. **B.** Time course of *Hells* mRNA expression in the developing retina in wild-type mice. Levels of *Hells* mRNA were measured by RT-qPCR and normalized to the levels at E15.5 taken as 1 (n = 3). Mean ± SD. **C.** Representative western blot analysis of HELLS protein levels with high and low exposure. Tubulin was used as a loading control. **D.** Quantification of the relative HELLS protein levels on western blots were E15.5 was taken as 1 (n = 3). Mean ± SD.

To investigate the role of HELLS in normal retinal development and tumorigenesis, we generated a series of genetically engineered mouse models. Using a *Hells* conditional knockout mouse model driven by *Chx10*-cre recombinase, our results demonstrate that *Hells*-null RPCs divide, undergo cell-fate specification, and give rise to normally laminated retinae with all seven major retinal cell types present in normal ratios. While minor synaptogenesis defects were observed in a subset of bipolar cells, retinal function is unaltered in the absence of HELLS. Despite the apparent dispensable role of HELLS in retinal development, here we present evidence that loss of HELLS significantly decreases the incidence of retinoblastoma, delays tumor progression, and increases overall survival.

5.4 Results

Hells is expressed during retinal development, and repressed following terminal differentiation

We first investigated the normal expression pattern of HELLS in developing retinae. Real-time reverse transcriptase PCR (RT-qPCR) analysis of retinal tissue at different developmental stages showed that *Hells* mRNA is robustly expressed starting at early stages of retinal development (E15.5), reaching its maximal expression around postnatal day 0 (P0) and then gradually declining until it reaches its minimal expression after P6, and is maintained at low levels thereafter (Figure 5-1B). The mRNA levels correlate with HELLS protein expression, with high expression of HELLS protein during the proliferative stages of retinal development (E15.5-P2), followed by a steep decrease that falls under the limits of detection after retinal cell-fate specification has been completed at P6 (Figure 5-1C, D). The expression pattern of HELLS in the developing retina suggests that HELLS is expressed from an early stage of retinogenesis and likely required during cell-fate specification and differentiation.

Hells is nonessential for normal retinal development

To elucidate the function of HELLS during retinal development, we generated genetically engineered mice to ablate the *Hells* gene (Figure 5-2A) in early retinal progenitors using the *Chx10-Cre* allele (*Chx10-Cre; Hells*$^{lox/lox}$; referred hereafter as *Hells* cKO). The *Hells* cKO mouse model used in this study targets exon 12, which contains the conserved helicase domain IV. Since *Chx10-Cre* is expressed in a mosaic pattern in progenitor cells, we generated mice that also carry the *Z/EG* transgene to serve as a reporter for Cre recombination. Retinal cells that arise from the early progenitor cells that stochastically expressed Cre recombinase during early development will express enhanced green fluorescent protein (EGFP). Western blot analysis using P0 retinae, the postnatal stage with highest HELLS expression in wild-type retinae, confirmed a significant

twofold reduction in HELLS protein levels in retinae expressing *Chx10-Cre* compared with *Cre*-negative littermate controls (Figure 5-2B). A short form of HELLS protein was not observed using a rabbit antibody against the N-terminal sequence of HELLS, suggesting that truncated HELLS protein is not stable or below limit of detection. As mentioned, *Chx10-Cre* is expressed stochastically, therefore complete abrogation of protein expression was not expected when using whole retinae.

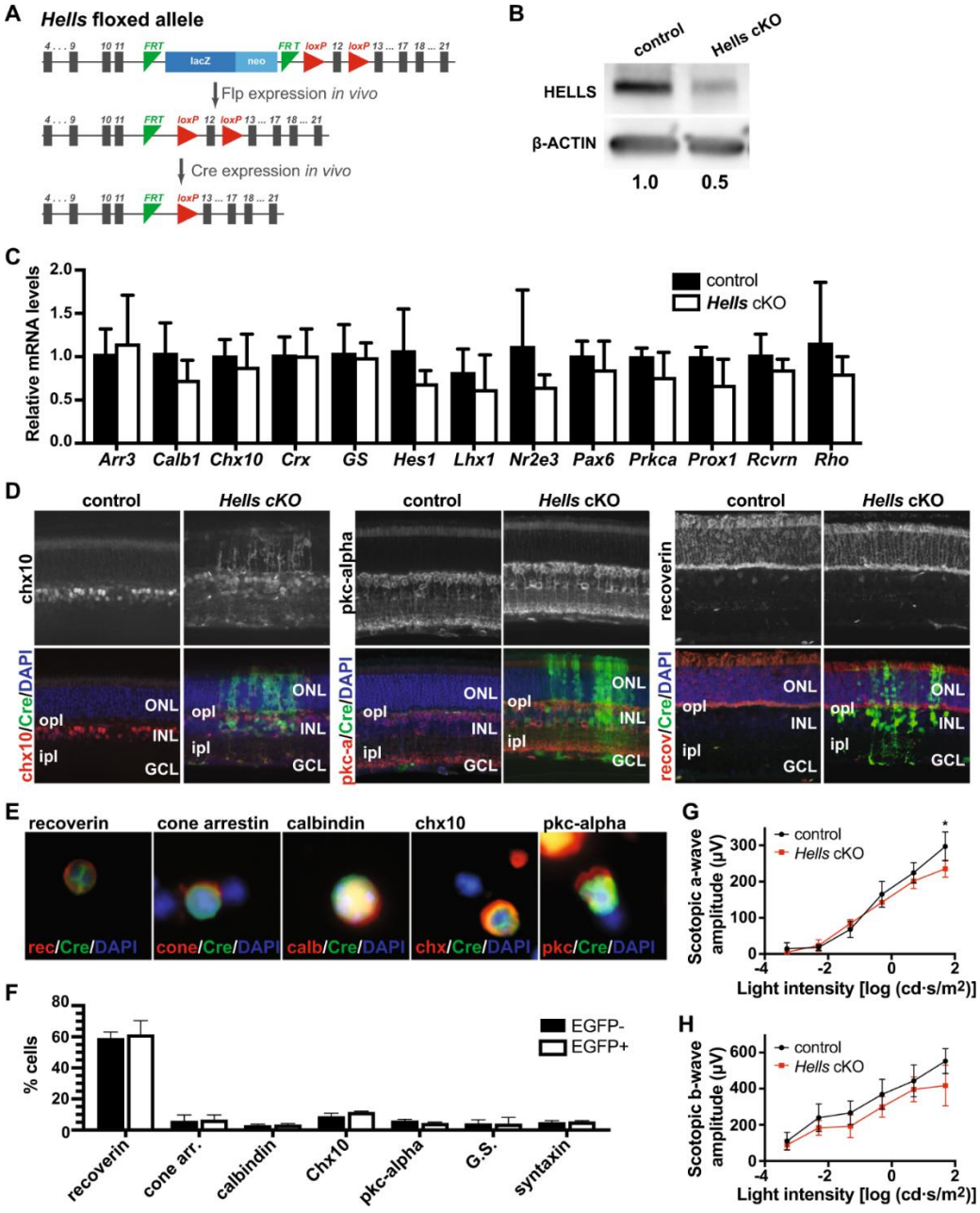


Figure 5-2. *Hells* is nonessential for normal retinal development. A. Schematic diagram of flippase (Flp) removal of the neo-cassette and conditional excision of the floxed HELLS allele. **B.** Western blot analysis of HELLS expression in *Hells* cKO (*Chx10-Cre* $Hells^{lox/lox}$) and littermate control ($Hells^{lox/lox}$) shows effective reduction of HELLS protein upon Cre recombination of floxed HELLS alleles. Actin was used as a loading control. Band intensities were quantified by densitometry and normalized to littermate control. **C.** RT-qPCR analysis of retinal cell marker genes. The genes of the seven major retina cell types are proportionately expressed in EGFP+ cells from *Hells* cKO mice compared with cells from Cre-negative littermate controls. All data are mean ± SD normalized to control littermates (n = 5). **D.** Representative images of P21 retina cross-sections from *Hells* cKO *Z/EG* and littermate control mice immunostained with chx10 (bipolar), pkc-alpha (bipolar), and recoverin (photoreceptors) antibodies (red). Retinae were double immunostained with anti-EGFP to capture areas of *Chx10-Cre*-mediated EGFP expression. Nuclei were counterstained with DAPI (blue). ONL outer nuclear layer, INL inner nuclear layer, GCL ganglion cell layer, ipl inner plexiform layer, opl outer plexiform layer. **E.** Representative images of P21 retina dissociated cells from *Hells* cKO *Z/EG* mice immunostained with recoverin (photoreceptors), cone arrestin (cone photoreceptors), calbindin (horizontal and a subset of amacrine cells), chx10 (bipolar), and pkc-alpha (bipolar) antibodies (red). Cells were double immunostained with anti-EGFP to capture areas of *Chx10-Cre*-mediated EGFP expression. Nuclei were counterstained with DAPI (blue). **F.** Quantification of the proportion of immunoreactive cells for each cellular marker antibody shown in E and Supplementary Figure 5-3 was determined for EGFP+ and EGFP− cells from *Hells* cKO *Z/EG* mice from independent litters (n = 3). Each bar represents the

mean ± SD of 500 cells scored from each retina. **G**, **H.** ERGs were recorded from 5-week-old *Hells* cKO (red line) and littermate control (black line). a-wave amplitude (**G**) and b-wave amplitude (**H**) were recorded at various light intensities. All measurements are mean ± SD (n = 4).

RT-qPCR analysis of the expression of retinal cell markers specific for the seven major retinal cell types using flow cytometry sorted EGFP + retinal cells from P21 retinae showed no significant differences in the expression of *Arr3*, *Calb1*, *Chx10*, *Crx*, *GS*, *Hes1*, *Lhx1*, *Nr2e3*, *Pax6*, *Prkca*, *Prox1*, *Rcvrn*, and *Rho* in *Hells* cKO retinae compared with *Cre*-negative littermate controls, suggesting normal retinal differentiation in the absence of HELLS (Figure 5-2C).

Further evaluation of potential alterations in the development of the retina upon loss of HELLS was achieved by examining retinae cross-sections using immunohistochemistry (Figure 5-2D; Supplementary Figure 5-1) and immunocytochemistry of dissociated retinae (Figure 5-2E; Supplementary Figures 5-2 and 5-3). We performed immunostaining using seven different specific antibodies against each major retinal cell type. *Hells* cKO displayed no major alterations in retinal cell birthing and cell-fate specification as evidenced by the uniform layer and thickness of the outer nuclear layer (ONL), inner nuclear layer (INL), and ganglion cell layer (GCL) (Figure 5-2D; Supplementary Figure 5-1). All cells are located in their corresponding retinal layer. The only abnormality observed was found in a subset of bipolar cells (Chx10-immunopositive cells). We observed bipolar cellular processes extending through the outer plexiform layer (OPL) into the ONL. In comparison, the bipolar cell bodies in *Cre*-negative

littermate controls and in the *Cre*-negative areas of the same retina are limited to the INL (Figure 5-2D). No differences in spatial localization were observed for any of the other retinal cell types analyzed, including pkc-alpha-immunoreactive rod-bipolar cells[24] and recoverin-immunoreactive photoreceptor cells (Figure 5-2D; Supplementary Figure 5-1). Immunocytochemical analysis of dissociated retinae was able to detect double-immunostained cells for every retinal cell-type-specific antibody and the EGFP reporter of Cre recombination (Figure 5-2E; Supplementary Figures 2 and 3). Quantification of EGFP + and EGFP− cell types co-stained with retinal cell markers indicated that all cells are present in the correct ratios (Figure 5-2F). This confirms that all retinal cell types that are born, can terminally differentiate and survive in the absence of HELLS.

In order to determine if the aberrant bipolar processes observed in *Hells* cKO mice affect normal retinal function, we performed electroretinography (ERG) on dark-adapted *Hells* cKO mice with retinae that had at least 50% EGFP + penetrance compared with *Cre*-negative littermate controls (Figure 5-2G, H). ERG measures the electrical activity of the various cell types in the retina in response to light stimulation. The a-wave hyperpolarization measures photoreceptor (cone and rod) function and the b-wave depolarization measures the inner retinal function, predominantly Müller and bipolar cells. Although *Hells* cKO mice tend to show a lower a-wave and b-wave amplitude, ERG recordings show neither significant changes in the amplitude of the a-wave, except at the highest light intensity (Figure 5-2G), nor the amplitude of the b-wave (Figure 5-2H), consistent with the observation that photoreceptors and inner neuron cells are comparable in numbers between *Hells* cKO and littermate controls.

***Hells* is transcriptionally repressed by the Rb family**

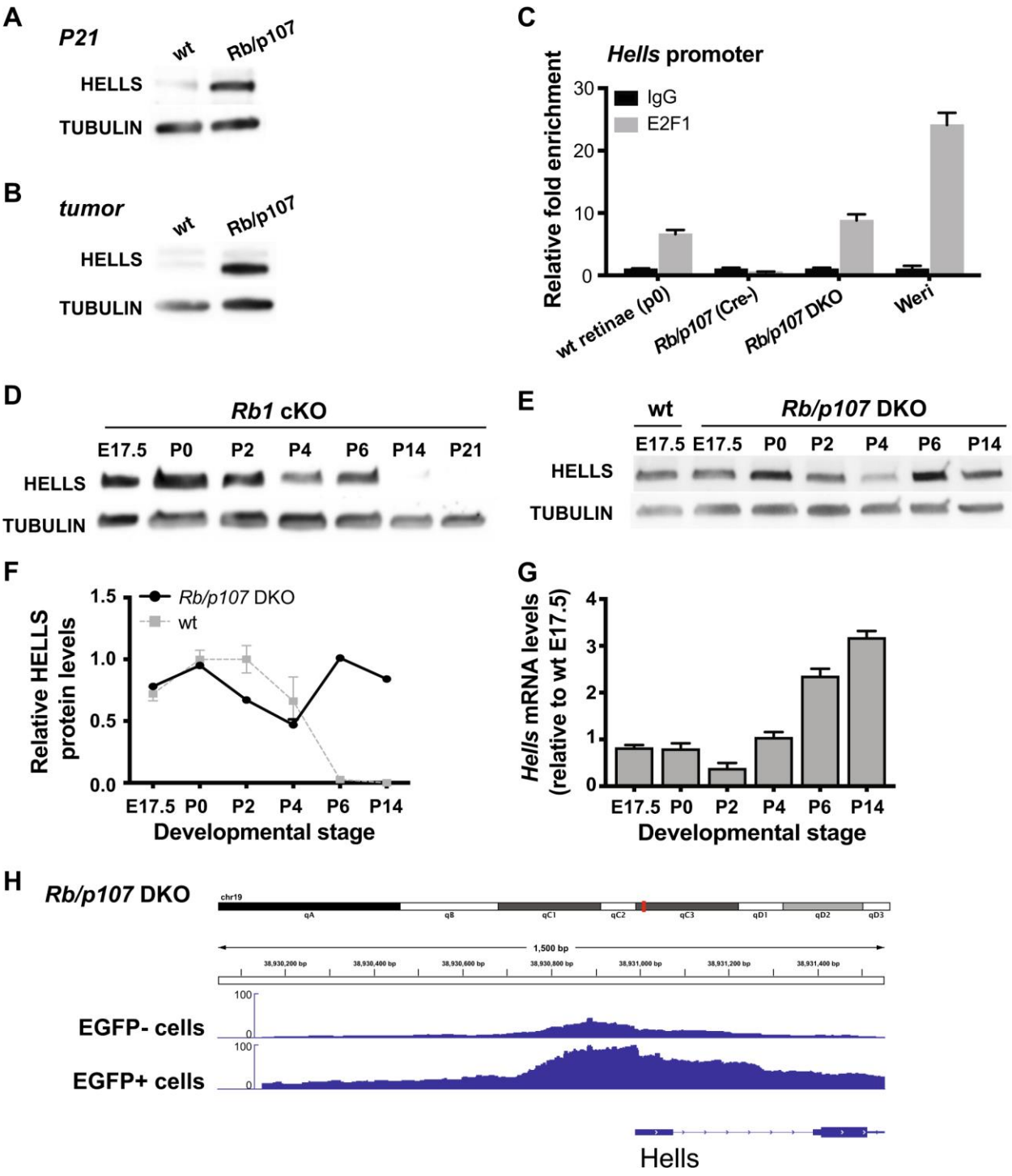


Figure 5-3. Hells is transcriptionally repressed by RB and p107. A, B. HELLS protein level in mouse (**A**) P21 retinae and (**B**) retinoblastoma tumor samples were compared with

littermates' controls (wt). Tubulin was used as a loading control. **C.** Chromatin immunoprecipitation (ChIP) assay in P0 and P21 *Rb1/p107* DKO mouse retinae and Weri retinoblastoma human cell line reveals enrichment of E2F1 within the *HELLS* promoter. IgG pull-down was used as negative control. **D, E.** Western blot analysis of HELLS protein level at different retinal development stages in *Rb1* cKO mice and *Rb1/p107* DKO. Tubulin was used as a loading control. **F.** Quantification of HELLS protein expression in *Rb1/p107* DKO (black line) compared to wt (gray) retinae. **G.** RT-qPCR analysis of *Hells* mRNA level in *Rb1/p107* DKO retinae relative to E17.5 wild-type (wt) retinae. **H.** Representative sequencing tracks for the *Hells* locus show increased peaks at the promoter in P21 *Rb1/*p107 DKO mouse retinal cells (EGFP+) compared with control retinal cells (EGFP−). The ATAC-Seq data have been normalized to take sequencing depth into account. All measurements are mean ± SD (*n* = 3).

Previous studies have shown that *HELLS* is regulated through the RB/E2F pathway and that loss of *RB1* results in upregulation and overexpression of HELLS, a process that may contribute to epigenetic changes that drive tumor progression[7,8,10]. We verified that HELLS overexpression occurs soon after tumor initiation in *Chx10-Cre* $Rb1^{lox/lox}$ $p107^{-/-}$ (*Rb1/p107* DKO; Figure 5-3A) retinae and this overexpression persists in retinoblastoma tumors (Figure 5-3B). To confirm that the RB/E2F pathway directly controls *Hells* transcription in the retina, we performed chromatin immunoprecipitation (ChIP) analysis using P0 wild-type and P21 *Rb1/p107* DKO and Cre-negative littermate control mouse retinae, as well as human retinoblastoma tumor cells (Weri). Pull-down using an E2F1 antibody showed enrichment of E2F1 at a consensus-binding site within

the *Hells* promoter in samples that highly express HELLS protein: P0 wild-type and P21 *Rb1/p107* DKO retinae and human tumor cells, but not in P21 Cre-negative *Rb1/p107* (Figure 5-3C).

To determine whether HELLS is transcriptionally repressed specifically by RB or also by other members of the RB family, we analyzed HELLS expression in *Chx10-Cre* $Rb1^{lox/lox}$ (*Rb1* cKO) and *Rb1/p107* DKO retinae. Interestingly, we found that in *Rb1* cKO retinae, a mouse model that does not develop retinoblastoma due to compensation by p107, HELLS mRNA and protein expression follows a similar dynamic as wild-type retinae, with maximal expression during embryonic stages (E17.5-P0), repression after postnatal day 2 (P2) and significant reduction in protein levels after postnatal day 6 (P6; Figure 5-3D). On the other hand, analysis of *Rb1/p107* DKO retinae at different stages of retinal development, alongside wild-type littermate controls indicate an absence of HELLS repression in *Rb1/p107* DKO mice after P4, observed both at protein (Figure 5-3E, F) and transcriptional (Figure 5-3G) levels. Analysis of chromatin accessibility at the *Hells* locus using ATAC-seq of P21 retinae confirmed a significant increase in reads at the *Hells* promoter and along the first two exons in EGFP-positive ($Rb^{-/-}/p107^{-/-}$) compared with EGFP-negative ($p107^{-/-}$) flow-sorted *Rb/p107* DKO retinal cells (Figure 5-3H). Since HELLS expression levels in *Rb1/p107* DKO retinae are comparable with those observed in wild-type retinae between E17.5 and P2 and chromatin remains open in EGFP + *Rb/p107* DKO P21 retinal cells, these results suggest that RB-family members are required for *Hells* repression during terminal differentiation of one or several retinal cell fates to block access of E2F1 at the *Hells* promoter.

Loss of *Hells* reduces tumorigenesis in a genetic mouse model of retinoblastoma

A previous study in retinoblastoma identified HELLS as a key target gene that promotes tumor proliferation[10]. To determine if HELLS overexpression is critical for retinoblastoma development, we generated *Chx10-Cre* $Rb^{lox/lox}$ $p107^{-/-}$ $Hells^{lox/lox}$ triple knockout (*Rb1/p107/Hells* TKO) mice and compared them to *Rb1/p107* DKO littermate controls. *Rb1/p107/Hells* TKO and *Rb1/p107* DKO mice were followed for 1 year beginning at birth for signs of tumor development. We found that abrogation of *Hells* led to a significant increase in survival and decreased morbidity in *Rb1/p107/Hells* TKO mice compared with their *Rb1/p107* DKO littermate controls (Figure 5-4A). We observed that only 16% of *Rb1/p107/Hells* TKO mice (n = 25) developed tumors compared with 72% of *Rb1/p107* DKO mice (n = 18; $p < 0.0001$) in the span of 1 year (Figure 5-4A). We also found that once tumors were detected, the time to moribund status for *Rb1/p107/Hells* TKO mice was on average 13.6 weeks compared to 6.8 weeks in *Rb1/p107* DKO mice (p = 0.0361; Figure 5-4B). This translates to a significant increase of ~16 weeks in the mean survival of tumor-bearing mice from 33.8 weeks for *Rb1/*p107 DKO to 49.6 weeks for *Rb1/p107/Hells* TKO (p = 0.0052). We confirmed that tumors arising from *Rb1/p107/Hells* TKO mice had effectively recombined the conditional *Hells* alleles using western blot analysis (Figure 5-4C).

Tumors from *Rb1/p107* DKO and *Rb1/p107/Hells* TKO mice were analyzed for gene expression using RNA-seq (n = 3 each). Surprisingly, despite the role of HELLS in heterochromatin formation, the gene expression profile from *Hells*-null tumors were virtually indistinguishable from controls, as portrayed by the PCA plot (Figure 5-4D). Comparative transcriptome analysis of tumors from *Rb1/p107* DKO

and *Rb1/p107/Hells* TKO mice rendered only 17 significant differential expressed genes (DEGs; Supplementary Figure 5-4). None of the DEGs provided insightful information regarding the mechanism by which *Hells* loss reduced the incidence of retinoblastoma or improved survival.

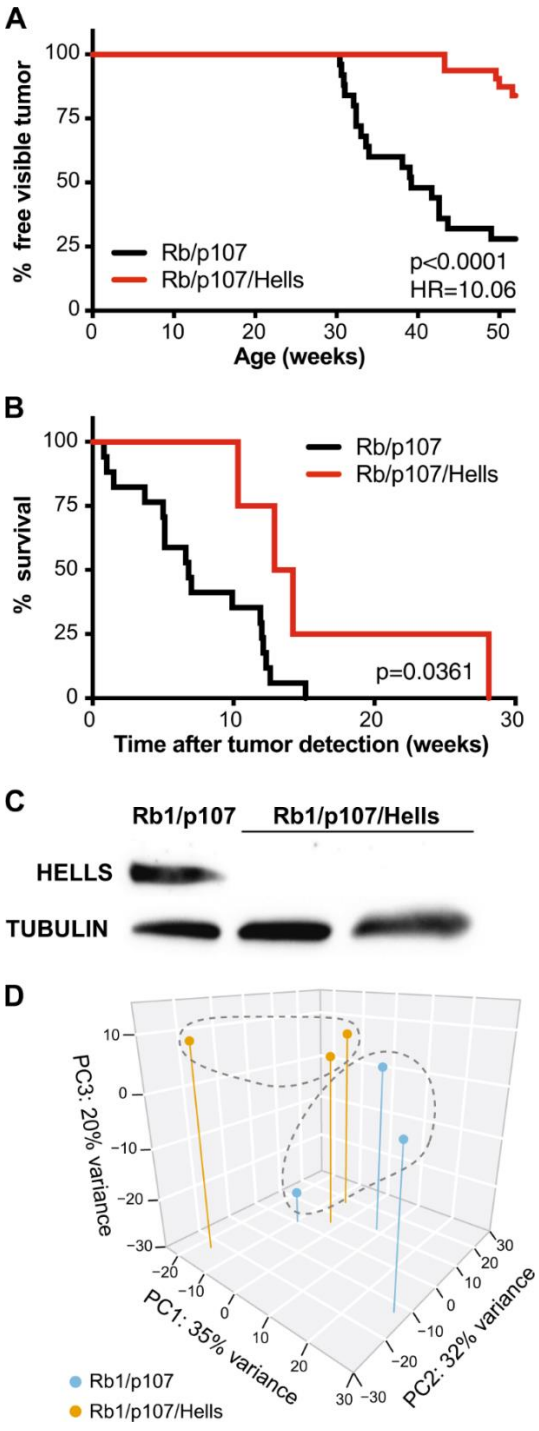


Figure 5-4. Loss of *Hells* decreases morbidity and increased survival in retinoblastoma. A, B. Kaplan–Meier curves showing (**A**) the percentage of mice free of visible retinoblastoma tumors in *Rb1/p107* DKO (*n* = 18) and *Rb1/p107/Hells* (*n* = 25) and (**B**) the time to moribund status after tumor detection of retinoblastoma-bearing animals mice in *Rb1/p107* DKO (*n* = 13) and *Rb1/p107/Hells* (*n* = 4). Mantel–Cox test was used for curve comparisons. **C.** Western blot detection of HELLS confirms loss of HELLS protein in *Rb1/p107/Hells* TKO tumors and HELLS expression in *Rb1/p107* DKO tumors. Tubulin used as a loading control. **D.** Principal components analysis (PCA) (3D loading plot) of RNA-seq data from three *Rb/p107* DKO and three *Rb1/p107/Hells* TKO tumors show the absence of distinct transcriptome clusters.

Hells loss decreases cellular proliferation in Rb1/p107-deficient retinae

Since *Rb1/p107/Hells* TKO mice develop tumors at a reduced frequency and with slower progression rates than *Rb1/p107* DKO mice, we aimed to identify pathways regulated by HELLS that aid in the delayed tumor progression. Since transcriptome analysis from tumors provided limited insight, we also performed RNA-seq and ATAC-seq analyses of P21 retinae from *Rb1/p107* DKO and *Rb1/p107/Hells* mice (*n* = 3 each) to determine if early transcriptomic differences could deepen our understanding of the mechanism through which *Hells* loss contributes to delayed tumorigenesis.

Unlike what we found in tumors, at P21 the transcriptome profile of *Rb1/p107/Hells* TKO retinae was markedly distinctive from *Rb1/p107* DKO retinae (Figure 5-5A). None of the 17 genes identified from the tumor DEG analysis (Supplementary Figure 5-4) were differentially expressed in the same direction in the P21

retinae DEG analysis (data not shown). We identified a total of 1455 DEGs, with 434 upregulated genes and 1021 downregulated genes in *Rb1/p107/Hells* TKO compared with *Rb1/p107* DKO P21 retinae. Gene ontology (GO) analysis for biological processes of the 434 upregulated genes show an enrichment of genes involved in visual system-related signaling pathways and functions, as seen by the classification of the top ten most significant GO terms (Figure 5-5B). Among the 1021 downregulated genes, we observed enrichment of pathways involved in DNA replication and cell division (Figure 5-5C).

Since HELLS has a known role in heterochromatin formation, we performed genome-wide analysis of chromatin accessibility to determine which genes were differentially expressed due to chromatin changes following loss of *Hells*. For this, we used ATAC-seq to analyze EGFP + flow-sorted *Rb1/p107/Hells* TKO and *Rb1/p107* DKO P21 retina cells along with EGFP- control cells. In line with our previous findings[25], *Rb1/p107* DKO retinae displayed increased areas of chromatin accessibility compared with control cells (Figure 5-5D). Interestingly, *Rb1/p107/Hells* TKO compared with *Rb1/p107* DKO retinae showed only 22 regions of the genome with significant changes in chromatin accessibility, 19 of which were chromatin repression and only 3 resulting in chromatin relaxation (Figure 5-5D and data not shown). Further, analysis of global levels of DNA methylation assessed by dot blot using a 5-mC-specific antibody showed no changes in genomic 5-mC levels in *Rb1/p107/Hells* TKO compared with *Rb1/p107* DKO and *Cre*-negative control retinae (Supplementary Figure 5). These data suggest that the transcriptional changes observed upon *Hells* loss that result in decreased tumorigenesis are likely mediated through a mechanism independent of chromatin remodeling.

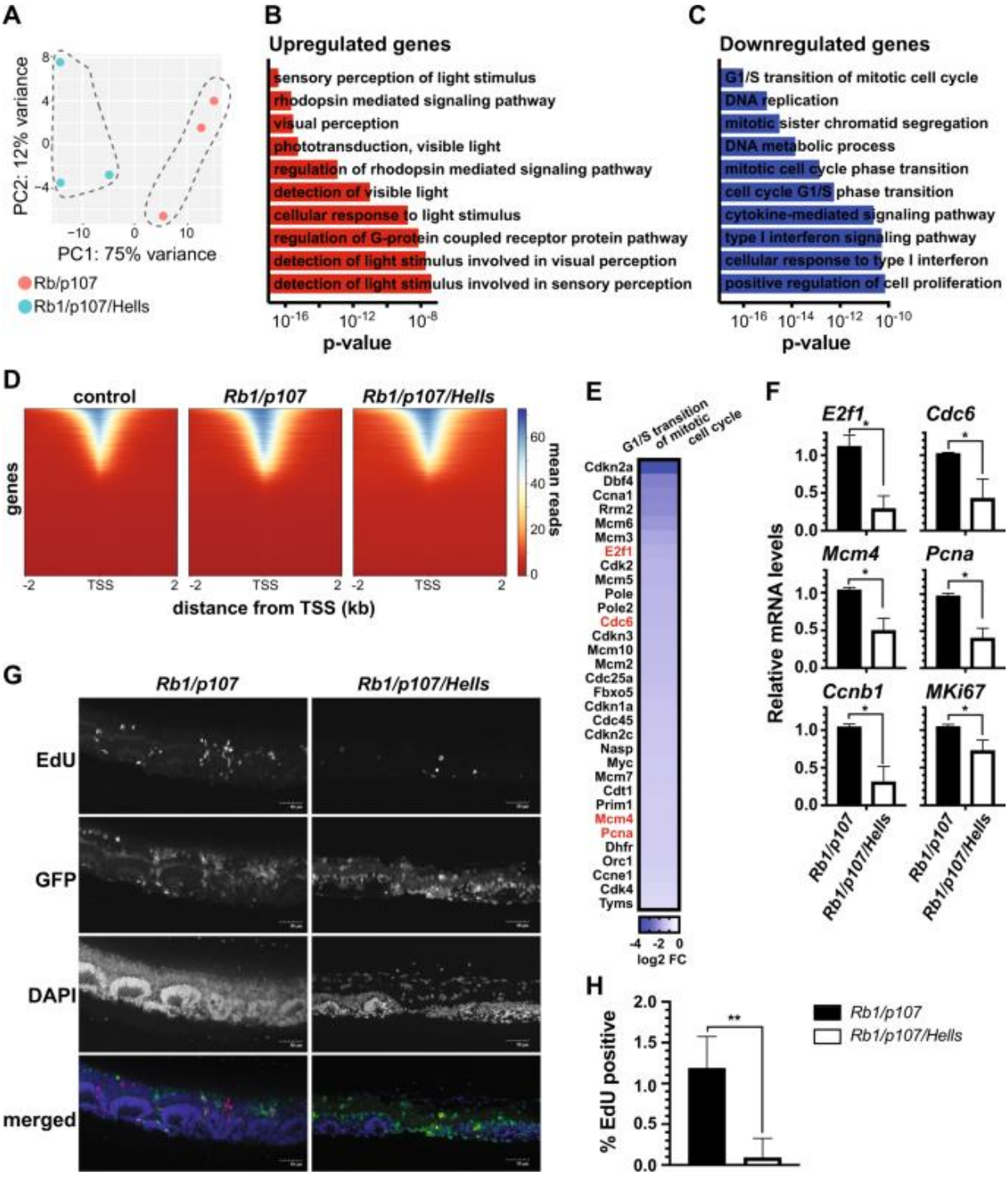


Figure 5-5. HELLS is necessary for ectopic proliferation in *Rb1/p107*-null retinae.

A. Principal components analysis (PCA) (2D loading plot) of RNA-seq data from three *Rb1/p107* DKO and three *Rb1/p107/Hells* TKO P21 retinae show distinct

transcriptome clusters. **B, C.** *p* value ranking bar graph representing the ten most significant gene ontology (GO) biological processes for (**B**) upregulated genes and (**C**) downregulated genes in *Rb1/p107/Hells* TKO compared with *Rb1/p107* DKO P21 retinae. **D.** Genome-wide heatmap plot of chromatin accessibility peaks (ATAC-seq) from flow-sorted EGFP-negative Cre-negative *Rb1/p107/Hells* and EGFP-positive *Rb1/p107* DKO and *Rb1/p107/Hells* TKO P21 retina grouped by mean reads and by distance from the transcription starting site (TSS). Each row represents a gene ordered in descending accessibility mean reads. **E.** Heatmap of all differentially expressed genes GO clustered as G1/S transition of mitotic cell cycle in P21 retinae from *Rb/p107* DKO compared with *Rb1/p107/Hells* TKO mice. Highlighted in red are known genes transcriptionally co-activated by HELLS. **F.** Real-time RT-qPCR validation of some of the RNA-seq downregulated genes involved in cell cycle progression. $n = 3$; $^{*}p < 0.05$ by unpaired *t* test. **G.** Representative images of P21 retina cross-sections from *Rb1/p107* DKO and *Rb1/p107/Hells* TKO *Z/EG* mice labeled with Click-iT EdU (proliferating cells; red). Nuclei were counterstained with DAPI (blue). EGFP fluorescence by the Z/EG reporter transgene captured areas of *Chx10-Cre*-mediated EGFP expression (green). Images taken on ×20 power field. Scale bar = 50 μm. **H.** Quantification of the proportion of EdU-positive cells from dissociated retinae from *Rb1/p107* DKO and *Rb1/p107/Hells* TKO mice from independent litters ($n = 3$). Each bar represents the mean ± SD of 2000 cells scored from each retina. $^{**}p < 0.0021$ by unpaired *t* test.

Previous in vitro studies have described HELLS as a transcriptional co-activator of cell proliferation, binding directly to the promoter regions of cell cycling genes, including

CDC6, MKi67, PCNA, CCNB1, MCM4, E2F1, and CCNA2[26,27]. Our RNA-seq analysis detected significant downregulation in all these cell cycle genes. Four of them, *E2f1*, *Cdc6*, *Mcm4*, and *Pcna*, are part of the G1/S transition of mitotic cell cycle cluster, our most significantly enriched GO term for downregulated genes (Figure 5-5E). Using RT-qPCR, we were able to independently validate downregulation of these genes in *Rb1/p107/Hells* TKO compared with *Rb1/p107* DKO P21 retinae (Figure 5-5F).

To evaluate if the transcriptional downregulation of cell cycle genes in *Rb1/p107/Hells* TKO retinae results in limited proliferation, we examined P21 retinae cross-sections using immunohistochemistry (Figure 5-5G; Supplementary Figure 6) and immunocytochemistry of dissociated retinae (Figure 5-5H) from mice injected with EdU 2 h prior to collection. We observed a clear decrease in the number of EdU-positive cells found along the retinae sections in *Rb1/p107/Hells* TKO P21 retinae compared with *Rb1/p107* DKO (Figure 5-5G; Supplementary Figure 6). We quantified the extent of the decrease in proliferating cells using dissociated retinae and found that the percentage of EdU-positive cells was significantly lower in *Rb1/p107/Hells* TKO (0.076 ± 0.131%) compared with *Rb1/p107* DKO (1.22 ± 0.35%; $p = 0.0064$) retinae (Figure 5-5H). Together, these data suggest that HELLS overexpression (or persistent expression) may contribute to the maintenance of a de-differentiated and proliferative state in *Rb1/p107*-null retinae that benavente progression may result in retinoblastoma formation.

5.5 Discussion

In recent years, the chromatin remodeling protein HELLS has been increasingly considered as a therapeutic target in several cancers, including gliomas and

carcinomas[26,28,29]. Our own studies, including this one, indicate that HELLS might be an attractive target for retinoblastoma therapeutics[10]. However, since HELLS is critical for the survival of mice and is required for normal tissue development, including the brain[17], understanding the function of HELLS during retinal development is relevant if ocular delivery of HELLS inhibitors were to be considered as a therapeutic option for this retinal malignancy. Fortunately, unlike a previous report showing that HELLS mutant mice exhibit impaired neural progenitor cells self-renewal and maintenance during development[17], we show in this study that *Hells* depletion does not affect RPC division, cell-fate specification, nor retinal function. Despite the seemingly dispensable role of *Hells* in retinal development, we found that persistent expression of *Hells* in *Rb1/p107*-null retinal cells is critical for ectopic proliferation and retinoblastoma tumor progression.

Hells is expressed during retinal cell-fate specification stages, but is nonessential for normal retinal development

Epigenetic regulation plays a fundamental role in the maintenance of cell identity as well as the stepwise control towards cellular differentiation. Here, we studied the role of chromatin remodeler HELLS during retinal development. We found that *Hells* mRNA is expressed in early stages of retinal development and declines around postnatal day 6. We confirmed that this reduction in gene expression translated into decreased protein expression. These patterns of expression suggest that HELLS is involved in the development and cell differentiation of the retina, but is not necessary for maintenance of the retina once the major cell types have developed.

We also assessed whether HELLS is critical for normal retinal development using a conditional knockout mouse model driven by *Chx10-Cre* recombinase. We found that *Hells*-

null mature (P21) retina display normal retinal cell specification and differentiation compared with their littermate controls. This suggests that HELLS is either unnecessary for retinal differentiation and development or its loss of function is compensated by other chromatin remodeling proteins.

The only difference we noted in *Hells* cKO retinae was a modest abnormal lamination of bipolar cells when compared with the surrounding Cre-negative tissue and littermate controls. Bipolar cells are the last cell type to terminally differentiate in the mouse retina, around postnatal day 6 when *Hells* expression is seen to decrease in the wild-type retina. It is possible that transcriptional repression of *Hells* helps orchestrate bipolar cell migration such that when HELLS is not present, lamination of bipolar cells is slightly altered. Despite this mild developmental defect, ERG measurements of the electrical activity of the retina in response to light stimulation indicated that *Hells*-null retina function is unaltered.

***Hells* is downregulated by Rb and p107 during late retinal differentiation**

Other groups and we have identified HELLS as a transcriptional downstream target gene of E2F1 that is overexpressed in cancer and contributes to tumor progression[9,10,30]. The evidence of direct regulation of *HELLS* by the RB family was limited to in vitro studies in gliomas and osteosarcoma[7,9]. In this study, we confirmed that the RB/E2F signaling pathway directly regulates transcriptional activation of HELLS in the developing retinae. ChIP analysis revealed enrichment of E2F1 at a known binding area on the promoter region of *HELLS*[8] in retinal tissue during early development or in the absence of *Rb1/p107*. Interestingly, loss of *Rb1*, unlike humans, is insufficient to drive tumorigenesis in the developing murine retina, nor does it exert *Hells* derepression during terminal retinal

differentiation. Our results show that in the murine retinae, loss of both *Rb1* and *p107* are necessary for the persistent expression of HELLS. These data suggest that the compensation by *p107* observed in the *Rb1* cKO mouse model is sufficient to effectively repress *Hells* expression in the absence of *Rb1*, an event that likely contributes to the prevention of tumor formation in *Rb1*-null mice[31]. Since HELLS expression levels during the early stages of retinal development (E17.5 through P2) in *Rb1/p107* DKO retinae are comparable with those observed in wild-type retinae, it is likely that the RB family is required for *Hells* repression during terminal differentiation of one or several retinal cell fates and failure to do so results in the aiding of tumor progression.

HELLS overexpression is critical for proliferation and retinoblastoma formation in Rb1/p107-null retina

In this study, we also sought to determine whether HELLS plays a critical role during retinoblastoma development. We found that HELLS abrogation during retinoblastoma development leads to a significant decrease in morbidity and increase in survival. The mechanism(s) by which HELLS contributes to tumorigenesis, however, remain unknown. One hypothesis is that HELLS overexpression leads to increased de novo DNA methylation of tumor suppressor genes, thereby decreasing the transcriptional level of tumor suppressor genes and increasing the incidence of tumor. Given the role of HELLS in facultative heterochromatin formation, we found it intriguing that we did not observe significant changes in gene transcription in *Rb1/p107/Hells* TKO tumors compared with *Rb1/p107* DKO tumors. Given the absence of a retinal phenotype upon loss of *Hells*, it is plausible that compensation by other chromatin remodelers occurs during retinal development, something that should be explored in future studies. It is also interesting that

more than twice as many gene expression changes observed in *Rb1/p107/Hells* TKO compared with *Rb1/p107* DKO P21 retinae were gene downregulations and that these changes were not associated with chromatin structure variations and that global DNA methylation was unaltered upon *Hells loss*.

Among the most significantly downregulated transcripts in *Rb1/p107/Hells*, TKO retinae are genes involved in cell cycle regulation. As reflected by the normal functioning *Hells* cKO retina, loss of *Hells* showed no signs of alterations in the proliferation in retinal progenitor cells; however, we showed that loss of *Hells* in *Rb1/p107* differentiating retinae drastically reduced cellular proliferation. These in vivo results are in line with our previous observation in human retinoblastoma cell lines, showing reduced proliferative capacity upon *HELLS* knockdown[10]. There is a remarkable similarity between the observations in this study and studies on the E2F family. Activating E2Fs, (aE2Fs: E2F1/2/3) control the transcription of genes required for DNA replication and proliferation of quiescent cells, but are not required for normal progenitor division in several tissues, including the developing retina[32–35]. However, aE2Fs are essential for abnormal division of differentiating RB-null cells [32,34–36] and removing E2F1 or E2F3 completely block retinoblastoma formation[10,37]. Taken together, we propose that HELLS is a critical downstream mediator of E2F-dependent ectopic division in the *Rb1/p107*-null retina (Figure 5-6). In the absence of RB and p107, E2F1 transcriptional derepression drives the expression of *Hells*. HELLS then functions as transcriptional co-activator of E2F3, stimulating expression of growth promoting genes during G1/S phase transitions[27]. The resulting ectopic proliferation of differentiating retinal cells promotes retinoblastoma formation.

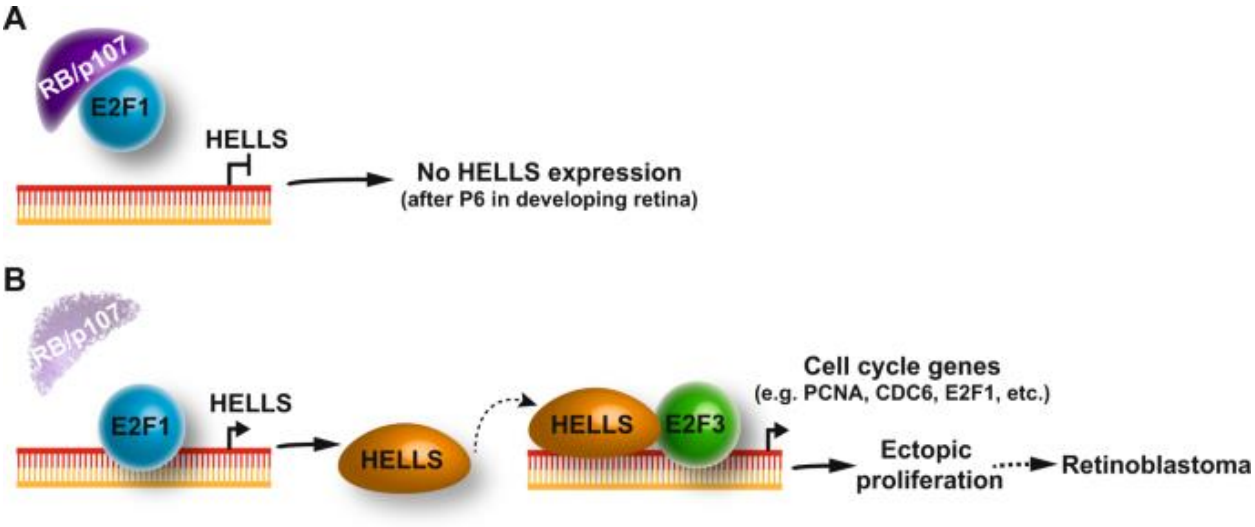


Figure 5-6. Model for HELLS function as driver of tumorigenesis in the developing retina. A. In the normal developing retina, RB or p107 represses E2F1 from driving HELLS expression by postnatal day 6 (P6). **B.** In the absence of RB and p107, E2F1 drives *Hells* transcription beyond of P6. HELLS is known to interact physically with E2F3 and function as transcriptional co-activator of E2F3. The resulting HELLS overexpression in the absence of RB and p107 mediate expression of several genes that stimulate G1/S transition and cell proliferation.

The decrease in tumor burden and morbidity observed in *Rb1/p107/Hells* TKO mice presents HELLS as an attractive potential therapeutic target for the treatment of retinoblastoma. This is particularly true given our observations indicating that HELLS expression is not essential for normal retinal development and that HELLS is not expressed in terminally differentiated retinal cells. This suggests that ocular delivery of HELLS inhibitors might reduce tumor burden with low retinal toxicities.

5.6 Materials and Methods

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus[38] and are accessible through GEO Series accession number GSE144429 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144429).

Mouse models

The *Rb*$^{lox/lox}$ mice were obtained from the Mouse Models of Human Cancer Consortium at the National Cancer Institute; the *p107*$^{-/-}$ mice were obtained from Dr. Tyler Jacks (Massachusetts Institute of Technology); *Chx10-Cre* mice were obtained from Dr. Connie Cepko (Harvard Medical School). Z/EG mice were obtained from Dr. Itsuki Ajioka (Tokyo Medical and Dental University). *Hells*$^{lox/lox}$ mice were obtained from the European Mouse Mutant Archive, backcrossed to Flp mice for removal of the neo-cassette (tm1c conversion), and then backcrossed to C57BL/6N mice for final Flp removal (Figure 5-2A). All mice are maintained in a C57BL/6N background, and both sexes were used in the analysis. Mice were monitored weekly, unblinded, for signs of retinoblastoma and anterior chamber invasion for 1 year from the time of birth. Moribund status was defined as the point when tumor cells invaded the anterior chamber and intraocular pressure increased to the point of imminent ocular rupture. The University of California Irvine Institutional Animal Care and Use Committee approved all animal procedures. Survival curves were generated using GraphPad Prism. Mantel–Cox test was used for statistical analysis of the Kaplan–Meier curves. Using trends from previous studies[10], we have found that 20 mice per group were sufficient to accurately measure differences in survival with an 80% power at 0.05 significance level in *Rb1/p107/Hells* TKO compared with *Rb1/p107* DKO mice.

Real-time reverse transcriptase PCR (RT-qPCR)

RNA was isolated from retinal tissue by homogenizing samples using Trizol Reagent and then isolated using chloroform. We used 1 µg of RNA to make cDNA according to SuperScript™III First-strand synthesis system (Invitrogen) manufacturer's protocol at a reaction volume of 20 µl. Quantitative PCR amplification was performed using 1 µl of reverse-transcribed product in Power SYBR Green PCR Master Mix (4367659, Life Technologies). Reaction were carried out using 7500 Real-Time PCR system (Applied Biosciences). Data were normalized to endogenous 18S and GAPDH controls, and analyzed using the $\Delta\Delta C_t$ method. Primers used are listed in Supplementary Table 5-1.

Western blotting

Retinae were homogenized by pellet pestle in RIPA buffer (50 mM Tris-HCl, pH = 8, 150 mM NaCl, 1% NP-40, 0.5% Sodium deoxycholate, 0.1% SDS, 1 mM EDTA) with added protease inhibitor (Mini cOmplete™, Roche). Samples were placed to lyse on ice for 30 min and then centrifuged at 14,000 rpm for 30 min at 4 °C. Assessed protein concentration using the BCA protein assay (Pierce™ BCA Protein Assay Kit). In all, 30 µg of the total protein was used, and blot was run on SDS-PAGE gel (Mini-PROTEAN, Bio-rad). The gel was then transferred onto a PVDF membrane (Immobilon-P Membrane, EMD Millipore) using a semi-dry transfer apparatus (Bio-rad). The membrane was incubated in 3% nonfat dry milk dissolved in tris-buffered saline (TBS) with 0.25% Tween (TBS-T) at room temperature for an hour in order to prevent nonspecific binding. Primary antibodies were diluted in 0.5% nonfat dry milk in TBS-T as follows: 1:1000 anti-Hells (sc-28202, Santa Cruz Biotechnology), 1:2000 anti-tubulin (2144S, Cell Signaling Technology). Membranes were incubated in primary antibody overnight at 4 °C. The membranes were washed three times with TBS-T for 5 min per wash then incubated with corresponding secondary

antibody for 30 min at room temperature. The secondary antibodies were prepared by diluting in 0.5% nonfat dry milk in TBS-T as follows: 1:1000 peroxidase labeled anti-mouse IgG (PI-2000, Vector Laboratories), 1:1000 peroxidase labeled anti-rabbit IgG (PI-1000, Vector Laboratories). After incubation for 30 min with secondary antibodies, the membranes were again rinsed three times with TBS-T for 5 min per wash. Bands were visualized using chemiluminescence (SuperSignal™ West Pico Chemiluminescent Substrate by Thermo Scientific). Band intensity assessed using Image J software.

Immunohistochemistry

Retinae were isolated in phosphate-buffered saline (PBS) and fixed overnight in 4% (w/v) paraformaldehyde. Whole retinae were embedded in 4% (w/v) agarose in PBS. The retinae were then cut into 50-µm slices using a vibratome. Retinal sections were blocked in 5% (v/v) normal donkey, goat, or rabbit serum, 0.5% Triton X-100 in PBS for 4 h at room temperature and then placed in primary antibody in the same blocking solution at 4° overnight. Mouse anti-calbindin antibody (C-9848, Sigma) was used at 1:100, rabbit anti-recoverin antibody (AB5585, Millipore) was used at 1:5000 dilution, mouse anti-pH3 antibody (H6409, Sigma) was used at 1:200 dilution, mouse anti-syntaxin antibody (S0664, Sigma) was used at 1:500 dilution, mouse anti-glutamine synthetase (610518, BD Biosciences) was used at 1:100 dilution, rabbit anti-cone-arrestin antibody (AB15282, Millipore) was used at 1:5000 dilution, and sheep anti-chx10 antibody (X1180P, Exalpha Biological) was used at 1:200 dilution. The retinal sections were washed three times with PBS and incubated in corresponding secondary antibody diluted 1:500 in respective blocking buffer (goat, rabbit, or donkey) for 1 h at room temperature in the dark. Again, retinae were washed three times with PBS and then incubated in 300 µl of Vectastain ABC

kit (Vector Laboratories) for 30 min at room temperature. Retina washed three times in PBS, and then placed in 300 μL tyramide Cy3 1:125 in amplification buffer (Perkin Elmer) for 10 min. Retinal sections were washed with PBS three times, and DAPI was added (1:1000 dilution in PBS) for 10 min to stain the nuclei. Retina washed two times with PBS, and mounted on slides. Imaging done using a Zeiss Confocal microscope.

Immunocytochemistry of dissociated retina

Retina were isolated and dissociated using trypsin 100X stock and incubation at 37 °C for 10 min. Trypsin inhibitor and 1000X DNase I were added to sample and incubated at 37 °C for 5 min. Complete culture medium was then added, and cells were transferred to chamber slides. The chamber slides were prepared with 1Xpoly-L-lysine incubated for 5 min. Cells were allowed to incubate on the slides for 30 min at 37 °C. Then, media was aspirated, and cells were fixed using 4% (w/v) paraformaldehyde overnight at 4 °C. Slides washed with PBS twice. Slides were then incubated with primary antibody overnight at 4 °C. Mouse anti-calbindin antibody (C-9848, Sigma) was used at 1:100, rabbit anti-recoverin antibody (AB5585, Millipore) was used at 1:5000 dilution, mouse anti-pH3 antibody (H6409, Sigma) was used at 1:200 dilution, mouse anti-syntaxin antibody (S0664, Sigma) was used at 1:500 dilution, mouse anti-glutamine synthetase (610518, BD Biosciences) was used at 1:100 dilution, rabbit anti-cone-arrestin antibody (AB15282, Millipore) was used at 1:5000 dilution, and sheep anti-chx10 antibody (X1180P, Exalpha Biological) was used at 1:200 dilution. Slides washed three times in PBS then incubated with secondary antibody in 1:500 dilution with the corresponding blocking buffer for 30 min at room temperature in the dark. Slides were washed three times with PBS, and incubated in 300 μl of Vectastain ABC kit (Vector Laboratories) for 30 min at room

temperature. Slides were washed three times in PBS and incubated with tyramide Cy3 1:150 in amplification buffer (Perkin Elmer) for 10 min. After washing with PBS three times, DAPI was added (1:1000 dilution in PBS) for 5 min to stain the nuclei. Slides were washed twice with PBS and mounted using gelvatol containing DABCO. Imaging was completed using Life Technologies EVOS microscope. Cell scoring was performed by a blinded investigator.

Electroretinography (ERG)

For scotopic ERG measurement, mice were dark adapted for 24 h prior to ERG recording. Each group consisted of four 5-week-old mice. Under dim red light, mice were anesthetized by intraperitoneal injection of a cocktail consisting of 20 mg/ml ketamine and 5 mg/ml xylazine in phosphate-buffered saline at a dose of 0.1 ml per 20 g body weight. Pupils were dilated with 1% tropicamide (Henry Schein, Melville, NY), and applied with 2.5% hypromellose (Akorn, Lake Forest, IL) to keep corneas hydrated. Contact electrodes were placed on corneas while the reference electrode needle was positioned subdermally between the ears. The a-wave and b-wave responses were recorded followed by a white light stimulus of different flash intensities (−3.3 to 1.7 log cd·s/m^2). For each intensity, 3–20 recordings were made with the resting intervals for recovery from photobleaching and were averaged for the final amplitude. All ERGs were recorded with the Celeris ophthalmic electrophysiology system (Diagnosys LLC, Lowell, MA) and analyzed with Espion V6 software (Diagnosys LLC). Statistical analysis was performed with paired *t* test. Data are represented as means ± SD. The investigator performing ERG measurements was blinded to the mouse genotype.

Chromatin immunoprecipitation (ChIP assay)

ChIP assays were performed on Weri retinoblastoma human cell line, postnatal day 0 retinae collected from wild-type mice, or postnatal day 21 retinae collected from *Rb1/p107 DKO* mice as previously described[25]. The antibodies used for chromatin pull-down were anti-E2F1 (3742; Cell Signaling) and rabbit IgG (sc-2027, Santa Cruz Biotechnologies). ChIP DNA was analyzed by qPCR with SYBR Green (Bio-Rad) in ABI-7500 (Applied Biosystems). Primers used are listed in Supplementary Table 5-1. The primers flank areas of homology in the mouse and human promoter region locater between position −94 and 114 bp and −115 and −134 bp, respectively. In human, this sequence falls within a functionally validated E2F1-binding region[8]. ChIP experiments were run in triplicate for each of the three independent retinal samples per condition (n = 3).

RNA sequencing

The total RNA was isolated using the RNA Micro Kit (Qiagen). Subsequently, 500 ng of the total RNA was used to create the RNA-seq library following the manufacturer's protocol from purification, mRNA fragmentation through the adenylation of end-repaired cDNA fragments and cleanup (TruSeq Stranded mRNA, Illumina). The collected sample was cleaned with AMPure XP beads (Beckman Coulter) and eluted in 20 µl of 10 mM Tris buffer, pH 8, 0.1% Tween 20. A paired-end 100-bp sequencing run was performed on HiSeq 4000 yielding 348 M PE reads with a final library concentration of 2 nM as determined by qPCR (KAPA Biosystem).

Sequencing reads were aligned to the mouse reference genome (GRCm38, ENSEMBL v.92) using STAR (v2.5.2a)[39]. Each read pair was allowed a maximum number of mismatches of 10. Each read was allowed a maximum number of multiple alignments of 3. Gene counts for each sample were produced using HTSeq (v0.6.1p1)[40]. Read count

normalization and differential expression analysis were performed using DESeq2 (v1.22.2) in R (v3.5.2)[41]. Genes with low reads (sum across samples <10) were removed. Differentially expressed genes (DEG) were calculated by comparing *Rb1/p107/Hells* TKO to *Rb1/p107* DKO, genes with base mean ≥10, Log2FoldChange ≥ 1 or ≤ −1, and adjusted *p* value (Benjamini–Hockberg) ≤ 0.05 were called differentially expressed genes. 3D PCA plot was generated using the R package plot3D (v1.1.1).

Chromatin profiling

Approximately 50,000 cells were harvested for ATAC-seq for each replicate. Briefly, retinae were dissociated, flow sorted into EGFP− or EGFP + , assessed for cell viability, counted, and washed with PBS. ATAC-seq was performed as previously described[42]. ATAC-seq libraries were sequenced with the Illumina HiSeq 4000 using 100 bp paired-end single indexed run. Raw reads were first QCed (*FASTQC*) and quality and adapter trimmed using *Trimmomatic*. Trimmed reads were then aligned to mm10 build of the mouse genome using Bowtie2 (v2.2.5) with alignment parameters: bowtie2 -X 2000– local–dovetail. Potential PCR duplicate reads were removed using *MarkDuplicates* from the *Picard* tools. Peaks were called in each sample using *MACS2* and further filtered using the ENCODE consensus blacklist regions (http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/mm10-mouse/). Differential peaks were identified using R package *diffbind* across the consolidated peak sets and metrics such as adjusted *p* values were reported. Heatmap plot was generated using deepTools3.

5-mC DNA dot blot

Genomic DNA was extracted from P21 retinae (Wizard SV genomic DNA purification Kit). Purified DNA was quantified, sonicated, denatured, and transferred to the Nylon membrane (RPN303B, GE Healthcare) according to the Cell Signaling DNA Dot Blot Protocol. 5methylcytosine (5-mC) Ab (#28692, Cell Signaling) was used to detect global DNA methylation of the samples. Methylene blue staining was used to verify equal DNA loading across samples.

5.7 References

1. Ferreira, R., Naguibneva, I., Pritchard, L. L., Ait-Si-Ali, S. & Harel-Bellan, A. The Rb/chromatin connection and epigenetic control: opinion. *Oncogene* **20**, 3128–3133 (2001).
2. Donovan, S. L. & Dyer, M. A. Developmental defects in Rb-deficient retinae. *Vis. Res* **44**, 3323–3333 (2004).
3. Chen, D. *et al.* Cell-specific effects of RB or RB/p107 loss on retinal development implicate an intrinsically death-resistant cell-of-origin in retinoblastoma. *Cancer Cell* **5**, 539–551 (2004).
4. Friend, S. H. *et al.* A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**, 643–646 (1986).
5. Rushlow, D. E. *et al.* Characterisation of retinoblastomas without RB1 mutations: genomic, gene expression, and clinical studies. *LANCET Oncol.* **14**, 327–334 (2013).
6. Zhang, J. *et al.* A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature* **481**, 329–334 (2012).
7. SC Wu, C. B. Chromatin remodeling protein HELLS is upregulated by inactivation of the RB-E2F pathway and is nonessential for osteosarcoma tumorigenesis. *Oncotarget* **9**, 32580–32592 (2018).
8. Niu, J. *et al.* Transcriptional activation of the senescence regulator Lsh by E2F1. *Mech. Ageing Dev.* **132**, 180–186 (2011).
9. Xiao, D. *et al.* Chromatin remodeling factor LSH is upregulated by the LRP6-GSK3beta-E2F1 axis linking reversely with survival in gliomas. *Theranostics* **7**, 132–143 (2017).

10. Benavente, C. A. *et al.* Chromatin remodelers HELLS and UHRF1 mediate the epigenetic deregulation of genes that drive retinoblastoma tumor progression. *Oncotarget* **5**, 9594–9608 (2014).
11. Dennis, K., Fan, T., Geiman, T., Yan, Q. & Muegge, K. Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev* **15**, 2940–2944 (2001).
12. Zhu, H. *et al.* Lsh is involved in de novo methylation of DNA. *EMBO J.* **25**, 335–345 (2006).
13. Burrage, J. *et al.* The SNF2 family ATPase LSH promotes phosphorylation of H2AX and efficient repair of DNA double-strand breaks in mammalian cells. *J. Cell Sci* **125**, 5524–5534 (2012).
14. Myant, K. & Stancheva, I. LSH cooperates with DNA methyltransferases to repress transcription. *Mol. Cell Biol.* **28**, 215–226 (2008).
15. Geiman, T. M. *et al.* Lsh, a SNF2 family member, is required for normal murine development. *Biochim. Biophys. Acta* **1526**, 211–220 (2001).
16. Sun, L. Q. *et al.* Growth retardation and premature aging phenotypes in mice with disruption of the SNF2-like gene, PASG. *Genes Dev.* **18**, 1035–1046 (2004).
17. Han, Y. *et al.* Lsh/HELLS regulates self-renewal/proliferation of neural stem/progenitor cells. *Sci. Rep.* **7**, (2017).
18. Turner, D. L. & Cepko, C. L. A common progenitor for neurons and glia persists in rat retina late in development. *Nature* **328**, 131–136 (1987).
19. Turner, D. L., Snyder, E. Y. & Cepko, C. L. Lineage-independent determination of cell type in the embryonic mouse retina. *Neuron* **4**, 833–845 (1990).
20. Aldiri, I. *et al.* The dynamic epigenetic landscape of the retina during development, reprogramming, and tumorigenesis. *Neuron* **94**, 550-568. e510 (2017).

21. Cvekl, A. & Mitton, K. P. Epigenetic regulatory mechanisms in vertebrate eye development and disease. *Heredity (Edinb)* **105**, 135–151 (2010).

22. KL Pennington, M. D. Epigenetic mechanisms of the aging human retina. *J. Exp. Neurosci.* **9**, 51–79 (2015).

23. S He, X. L. N. C. D. H. Review: epigenetic mechanisms in ocular disease. *Mol. Vis.* **19**, 665–674 (2013).

24. Haverkamp, S., Ghosh, K. K., Hirano, A. A. & Wässle, H. Immunocytochemical description of five bipolar cell types of the mouse retina. *J. Comp. Neurol.* **455**, 463–476 (2003).

25. Benavente, C. A. *et al.* Cross-species genomic and epigenomic landscape of retinoblastoma. *Oncotarget* **4**, 844–859 (2013).

26. Zhang, G. *et al.* Chromatin remodeler HELLS maintains glioma stem cells through E2F3 and MYC. *JCI Insight* **4**, (2019).

27. von Eyss, B. *et al.* The SNF2-like helicase HELLS mediates E2F3-dependent transcription and cellular transformation. *EMBO J.* **31**, 972–985 (2012).

28. Law, C. T. *et al.* HELLS regulates chromatin remodeling and epigenetic silencing of multiple tumor suppressor genes in human hepatocellular carcinoma. *Hepatology* **69**, 2013–2030 (2019).

29. Liu, S. & Tao, Y. G. Chromatin remodeling factor LSH affects fumarate hydratase as a cancer driver. *Chin. J. Cancer* **35**, 72 (2016).

30. He, X. *et al.* Chromatin remodeling factor LSH drives cancer progression by suppressing the activity of fumarate hydratase. *Cancer Res* **76**, 5743–5755 (2016).

31. Donovan, S. L., Schweers, B., Martins, R., Johnson, D. & Dyer, M. A. Compensation by tumor suppressor genes during retinal development in mice and humans. *BMC Biol.* **4**, (2006).

32. Chen, D. *et al.* Rb-mediated neuronal differentiation through cell-cycle-independent regulation of E2f3a. *PLoS Biol.* **5**, 1504–1519 (2007).

33. Chen, D. *et al.* Division and apoptosis of E2f-deficient retinal progenitors. *Nature* **462**, 925–929 (2009).

34. Chong, J. L. *et al.* E2f1-3 switch from activators in progenitor cells to repressors in differentiating cells. *Nature* **462**, 930–934 (2009).

35. Wenzel, P. L. *et al.* Cell proliferation in the absence of E2F1-3. *Dev. Biol.* **351**, 35–45 (2011).

36. Ziebold, U., Reza, T., Caron, A. & Lees, J. A. E2F3 contributes both to the inappropriate proliferation and to the apoptosis arising in Rb mutant embryos. *Genes Dev.* **15**, 386–391 (2001).

37. Sangwan, M. *et al.* Established and new mouse models reveal E2f1 and Cdk2 dependency of retinoblastoma, and expose effective strategies to block tumor initiation. *Oncogene* **31**, 5019–5028 (2012).

38. Edgar, R., Domrachev, M. & Lash, A. E. Gene expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).

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

40. Anders, S., Pyl, P. T. & Huber, W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

41. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).

42. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).

Chapter 6: Identification of novel regulators of dendrite arborization using cell type-specific RNA metabolic labeling

Publication note

This paper was originally published in PLOS One.

Mohamed Y. Aboukilila, Josephine D. Sami, Jingtian Wang, Whitney England, Robert C. Spitale, Michael D. Cleary. Identification of novel regulators of dendrite arborization using cell type-specific RNA metabolic labeling **15**, 12 (2020).

6.1 Contribution Statement

Mohamed Y. Aboukilila conducted the conceptualization, formal analysis, investigation, methodology, and writing – review & editing; Josephine D. Sami conducted the investigation, methodology, and writing – review & editing; Jingtian Wang conducted the conceptualization, data curation, investigation, and methodology; Whitney England conducted the data curation, investigation, and methodology; Robert C. Spitale conducted the conceptualization, formal analysis, funding acquisition, supervision, visualization, and writing – review & editing; Michael D. Cleary conducted the conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, supervision, validation, visualization, writing – original draft, and writing – review & editing

6.2 Abstract

Obtaining neuron transcriptomes is challenging; their complex morphology and interconnected microenvironments make it difficult to isolate neurons without potentially altering gene expression. Multidendritic sensory neurons (md neurons) of *Drosophila* larvae are commonly used to study peripheral nervous system biology, particularly dendrite arborization. We sought to test if EC-tagging, a biosynthetic RNA tagging and purification method that avoids the caveats of physical isolation, would enable discovery of novel regulators of md neuron dendrite arborization. Our aims were twofold: discover novel md neuron transcripts and test the sensitivity of 5-ethynylcytosine (EC)-tagging. RNAs were biosynthetically tagged by expressing CD:UPRT (a nucleobase-converting fusion enzyme) in md neurons and feeding 5-ethynylcytosine to larvae. Only CD:UPRT-expressing cells are competent to convert EC into 5-ethynyluridine-monophosphate which is subsequently incorporated into nascent RNA transcripts. Tagged RNAs were purified and used for RNA-sequencing. Reference RNA was prepared in a similar manner using 5-ethynyluridine (EUd) to tag RNA in all cells and negative control RNA-seq was performed on "mock tagged" samples to identify non-specifically purified transcripts. Differential expression analysis identified md neuron enriched and depleted transcripts. Three candidate genes encoding RNA-binding proteins (RBPs) were tested for a role in md neuron dendrite arborization. Loss-of-function for the m6A-binding factor Ythdc1 did not cause any dendrite arborization defects while RNAi of the other two candidates, the poly(A) polymerase Hiiragi and the translation regulator Hephaestus, caused significant defects in dendrite arborization. This work provides an expanded view of transcription in md neurons and a technical framework for combining EC-tagging with RNA-seq to profile transcription in cells that may not be amenable to physical isolation.

6.3 Introduction

Neuron development requires regulation of gene expression at the transcriptional and post-transcriptional levels. *Drosophila* peripheral nervous system (PNS) neurons provide a useful model for investigating these mechanisms. Sensory neurons of the larval PNS are classified according to dendrite morphology: external sensory and chordotonal neurons have a single dendrite, bipolar dendrite neurons have two unbranched dendrite projections, and multidendritic (md) neurons have more complex dendritic arborization. Md neurons innervate the larval body wall and function as touch receptors, proprioceptors, thermoreceptors or nociceptors. Md neurons have proven useful for investigating the molecular mechanisms that control dendrite arborization[1]. Foundational work used a mutagenesis screen to identify genes that regulate dendrite arborization[2]. Others have taken a reverse genetics approach and tested candidates through RNA-interference (RNAi) based on the functional properties of candidates[3,4]. The ability to induce neuron-specific RNAi in *Drosophila* makes reverse genetics an attractive approach and the efficacy of candidate choice can be improved by selecting genes from neuron-specific transcriptome data.

Transcriptome profiling of dendritic arborization (da) neurons, a subclass of md neurons, has been performed using fluorescence-activated cell sorting (FACS)[5] or magnetic bead-based purification[6]. These studies used differential expression data to select candidates and demonstrated roles for those genes in da neuron dendrite arborization. One caveat of physical isolation is that neurons are removed from their natural environment and undergo processing prior to RNA extraction, possibly inducing transcriptional and

post-transcriptional responses that do not reflect *in vivo* gene expression. An alternative approach is to use biosynthetic RNA tagging methods that do not require physical isolation and enrich for nascent and recently-transcribed mRNAs[7]. These methods use metabolic labeling, under *in vivo* conditions, to generate tagged RNAs in the cells of interest. Tagged RNAs are subsequently purified from total RNA of animals or tissues. We recently described a cell type-specific biosynthetic RNA tagging method called EC-tagging[8]. EC-tagging works via targeted expression of a nucleobase-converting fusion enzyme composed of cytosine DEminase and uracil phosphoribosyltransferase (CD:UPRT). Metazoans lack cytosine DEminase activity and have varying endogenous uracil phosphoribosyltransferase activity, depending on organism or cell type[9,10]. Only cells expressing CD:UPRT are competent to convert the bioorthogonal base 5-ethynylcytosine (EC) into 5-ethynyluracil (via CD activity) and subsequently into 5-ethynyluridine monophosphate (via transgenic UPRT and endogenous pathways) which is ultimately incorporated into nascent RNAs. The 5-ethynyl group allows click-chemistry-based biotinylation of tagged RNAs and subsequent purification on streptavidin beads. We previously demonstrated the utility of EC-tagging in the *Drosophila* central nervous system (CNS)[8]. This initial work used a microarray platform to analyze transcriptomes of relatively large populations of neurons (the entire larval CNS and the mushroom body neurons). To further test the specificity and sensitivity of this technique and to discover novel regulators of dendrite arborization, here we combine EC-tagging with RNA-sequencing to generate md neuron transcriptome profiles.

6.4 Results

EC-tagging enriches for larval md neuron-specific transcripts

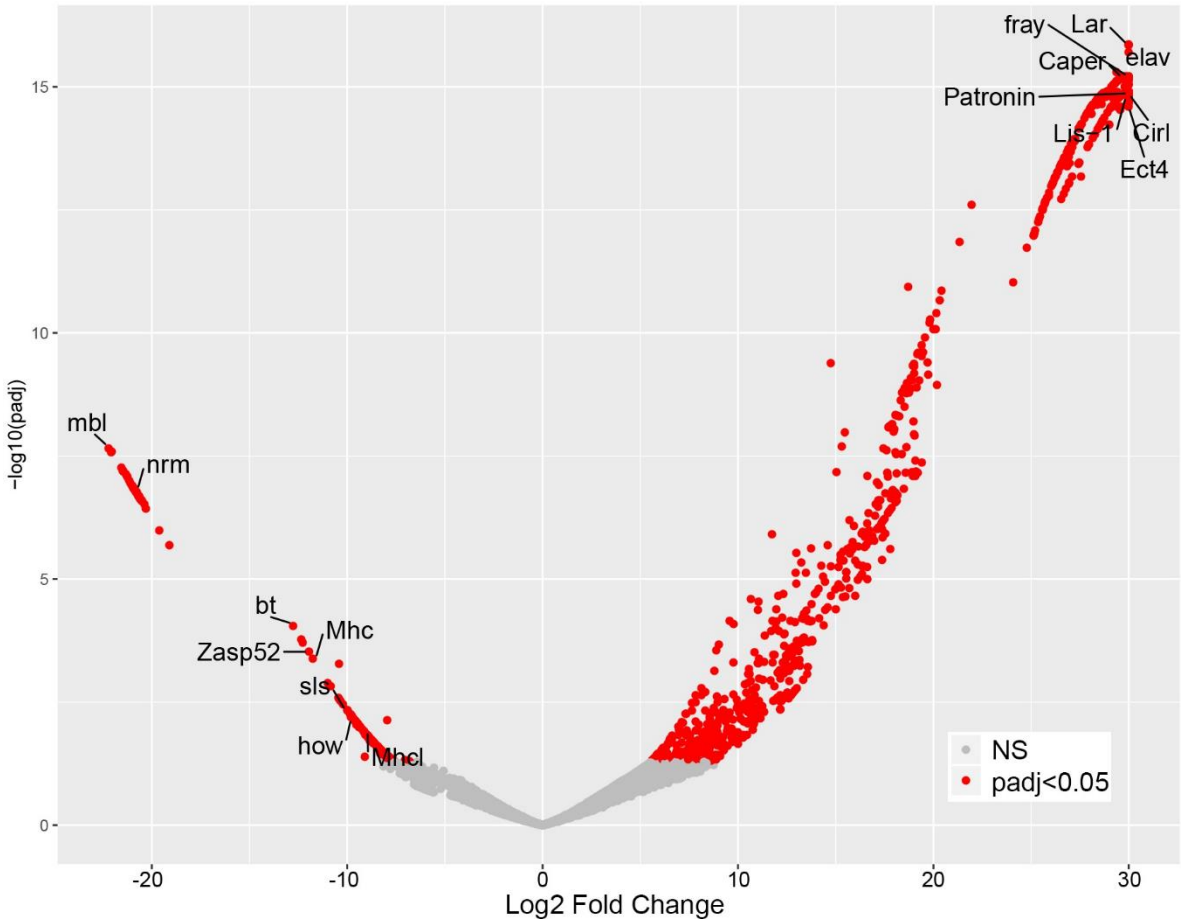


Figure 6-1. Identification of md neuron enriched and depleted transcripts.

Volcano plot of differential expression results showing the $\log_2$ fold change for EC-tagged TPM / EUd-tagged TPM (x-axis) and the adjusted p-value for each transcript (y-axis). NS = not significant. Select genes with known md neuron expression (enriched side of plot) and known muscle expression (depleted side of plot) are labeled.

To identify genes transcribed in all md neurons, we used *Gal4*$^{109(2)80}$ [2]to drive *UAS-CD*:*UPRT* and fed 5EC for 12 hours to L3 larvae staged between 72–84 hours after hatching. Though formed in embryos, md neuron dendrites continue growth, elongation and/or arborization throughout the L3 stage[5], thus we predicted relevant genes of interest would

be transcribed during this timeframe. Based on a previously described neural EC-tagging time-course[8], we estimate it takes six hours before ingested 5EC is metabolized to the point that widespread RNA tagging occurs in target neurons. The twelve-hour 5EC feeding is therefore expected to generate a population of tagged RNAs synthesized over approximately six hours. At the end of EC feeding, larval carcass (containing primarily muscle, epidermis and the peripheral neurons of interest) was dissected to remove CNS neurons that express *Gal4*$^{109(2)80}$. A reference sample was prepared by feeding 5-ethynyluridine (5EUd) to stage-matched *UAS-CD*:*UPRT* larvae, in the absence of any Gal4. 5EUd is incorporated into RNA independent of CD:UPRT and thus provides a reference containing mRNAs transcribed in all cells over the same labeling period. As a negative control, we prepared "mock-tagged" samples in which larvae were not fed 5EC or 5EUd but were subjected to the same carcass dissection and RNA processing. The mock sample serves as a control for the stringency of the purification and allows identification of transcripts that may be purified independent of EC-tagging. This type of mock reference has proven useful in other biosynthetic RNA labeling experiments[11,12].

For all three sample types (5EC-tagged, 5EUd-tagged and mock-tagged), equal amounts of total RNA were biotinylated and applied to streptavidin beads. RNA captured on the beads was directly used to prepare sequencing libraries. The number of mapped reads per sample agreed with the expected yield of tagged RNA: 5EUd-tagging (expected high yield, RNA tagging in all cells) gave 48–56.6 million reads, 5EC-tagging (expected low yield, RNA tagging only in rare md neurons) gave 2.7–7.7 million reads, and mock-tagging (background) gave 0.11–0.13 million reads. 5EC-tagged biological replicates and 5EUd-tagged biological replicates had a high degree of RNA-seq correlation, while the correlation

for mock-tagged replicates was much lower (**S6-1 Fig**). To identify transcripts enriched in md neurons, we performed differential expression (DE) analysis to compare 5EC-tagged RNA (EC-RNA) and 5EUd-tagged RNA (ref-RNA). This differential expression analysis was performed using two versions of the ref-RNA: 1) the complete RNA-seq dataset and 2) a randomly generated subset of 2.7 million reads (see methods). The random subset matches the read depth of the lowest yield EC-RNA library, thus controlling for possible sample size effects.

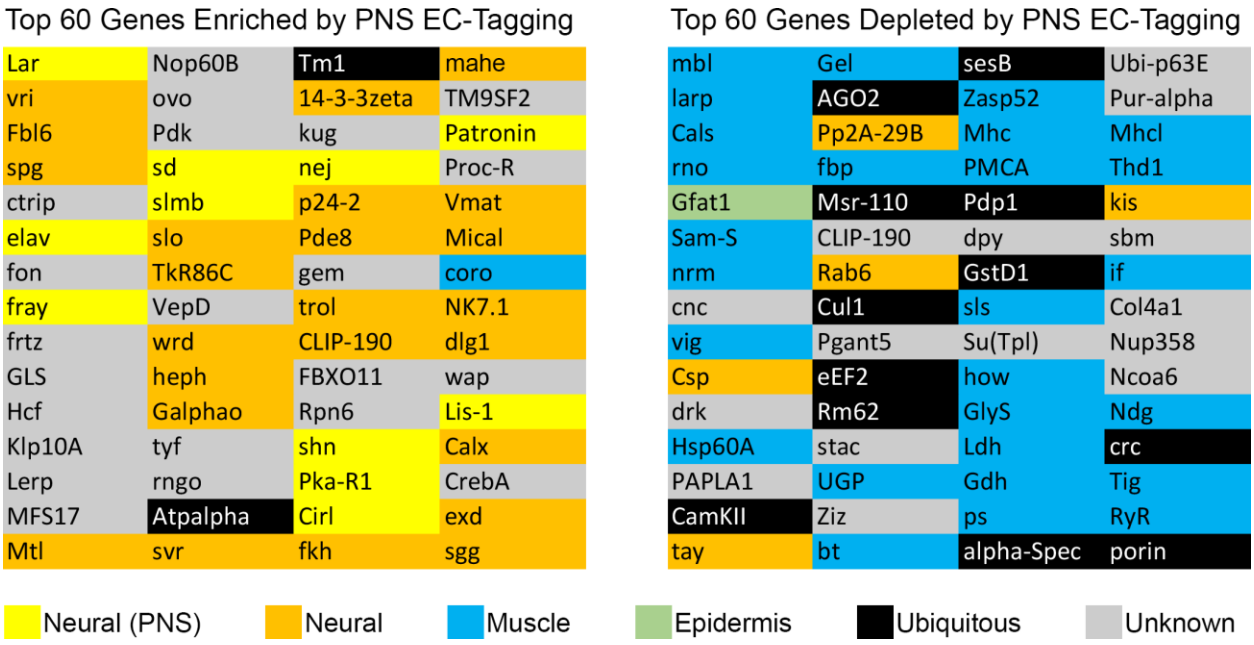


Figure 6-2. Top md neuron enriched and depleted transcripts.

Only named genes are included, genes known only by "CG number" were excluded. The primary tissue expression pattern is color coded. Tissue expression was determined by searching Flybase annotations and published literature. "Neural (PNS)" indicates known expression in larval md neurons. "Neural" indicates expression in the central nervous system at any stage. "Muscle" indicates expression in any type of muscle at any stage. "Epidermis" indicates expression in epidermis at any stage. "Ubiquitous" indicates either

widespread expression or evidence of expression in both neurons and muscle. “Unknown” indicates the literature do not support assignment to any of the other categories.

DE analysis identified 937 enriched transcripts and 236 depleted transcripts (minimum two-fold difference and adjusted p-value < 0.05). To address potential non-specific purification among the enriched transcripts, we compared transcript levels between EC-RNA and mock-RNA. This comparison identified 85 transcripts with no significant enrichment compared to mock-RNA, thus reducing the list of md neuron transcripts to 852. Similar background correction for depleted transcripts is not possible since transcripts that are rare or absent in md neurons are expected to be low in EC-RNA and mock-RNA. In the background-corrected dataset, known md neuron transcripts are among the most significantly enriched and known muscle-specific transcripts are among the most significantly depleted **(Figure 6-1).** The top 60 enriched or depleted genes are listed in **Figure 6-2** and the complete DE results are provided in **S6-1 Table.** EC-RNA compared to the subset ref-RNA yielded 571 enriched transcripts and 130 depleted transcripts (minimum two-fold difference and adjusted p-value < 0.05) and the background correction procedure removed 39 transcripts from the enriched list (**S6-1 Table)**. Enriched and depleted transcripts were similar regardless of the type of ref-RNA used (commonly enriched transcripts are denoted in **S6-1 Table**). We also performed gene ontology (GO) analysis on the complete ref-RNA DE results (**Figure 6-3**) and the subset ref-RNA DE results (**S6-2 Fig**). Both approaches yielded multiple neuron-specific GO categories including “peripheral nervous system development” and “dendrite morphogenesis”. The “synaptic growth at neuromuscular junction” GO category reflects the fact that enriched

genes in this category function in the synapses of motor neurons and md sensory neurons. Non-neural GO categories, “dorsal closure” and “border follicle cell migration”, reflect the fact that many neurite growth or morphogenesis genes (e.g. *stathmin*, *shot*, *kay*, *shn*, *aop*) are also involved in these processes. In addition to the EC-tagged data, we performed GO analysis on DE results comparing mock-RNA to the complete or subset ref-RNA. Importantly, the mock-tagged DE data did not result in any significant GO category enrichment.

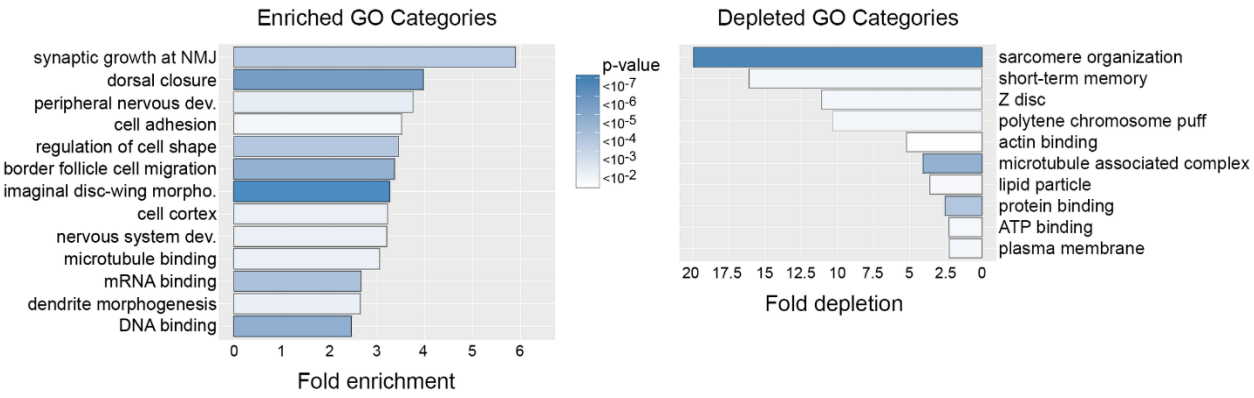


Figure 6-3. Gene ontology of enriched and depleted transcripts.

GO categories over-represented among EC-RNA enriched and depleted genes. Observed / expected value = frequency of category genes in EC-RNA / frequency in the *Drosophila* genome. Heatmap = Bonferroni-corrected P-values. NMJ = neuromuscular junction.

Given the agreement between EC-RNA compared to the full or subset ref-RNA, we focused subsequent analyses on the background-corrected EC-RNA versus full ref-RNA. In addition to the genes listed in **Figure 6-2**, this dataset contains many previously described md neuron genes (according to Flybase annotations and associated references), including

transcription regulators (*ab*, *ttk*, *stan*, *cnc*, *jim*, *kay*, *gro*), RNA-binding factors *(bel*, *Caper*, *Fmrp*, *sqd*, *stau*, *rump*), ion channels (*Piezo*, *SK*), signal receptors and transducers (*EcR*, *Egfr*, *Rac1*, *spin*, *puc*), and cytoskeletal factors (*spas*, *shot*). We also identified multiple transcripts encoding general regulators of neurotransmission, including *Frq1*, *brp*, *Csp*, *and Rim*. While differences in experimental design and target cell populations limit the validity of broad comparisons between our data and prior transcriptome studies, we selected two gene lists from prior studies to compare with our enriched gene set. Iyer et al. identified 40 transcription factor genes enriched in class I and/or class IV da neurons compared to whole larvae[6]. Our enriched gene set contains 6 of these 40 genes, a moderate but significant over-representation (Fisher's exact test comparing representation in the EC-RNA dataset to representation in the *Drosophila* genome, p-value = 0.03). Hattori et al. identified 24 genes expressed in class I and/or class IV da neurons for which RNAi caused dendritic arborization phenotypes[5]. Our md neuron enriched gene set contains 10 of these 24 genes, a significant over-representation (Fisher's exact test, p-value = 5×10^{-6}).

Candidate testing identifies novel regulators of md neuron dendrite arborization

We next sought to test the function of novel candidate genes in md neuron dendrite arborization. We focused on genes encoding mRNA-binding proteins (a significantly enriched GO category, **Figure 6-3**)) based on our interest in post-transcriptional control of mRNA processing[13]. Out of 20 enriched RNA-binding proteins (RBPs), 10 (*rump*, *fus*, *bru1*, *sqd*, *stau*, *shep*, *elav*, *Fmr1*, *BicD*, and *Caper*) have previously described functions in md neuron dendrite arborization (based on Flybase annotations or a RNAi screen of RBPs[4]). The RNAi screen of Olesnicki et al. [4]tested 7 additional RBPs that we

identified as md neuron-enriched but did not detect any dendrite arborization defects, suggesting these RBPs have functions irrelevant to dendrite morphogenesis. We focused on the remaining 3 RBPs with no available md neuron information: *ythdc1* (also known as *YT521-B*), *hiiragi* (*hrg*), and *hephaestus* (*heph*). Ythdc1 is a nuclear-localized m6A binding protein that regulates alternative splicing[14]. Hiiragi is a poly(A) polymerase that acts on nascent mRNAs and regulates poly(A) tail length of cytoplasmic mRNAs in oocytes[15]. Hephaestus is a polypyrimidine tract binding protein that represses translation of *oskar* in oocytes[16]. We selected a known md neuron RBP enriched in our dataset, *Caper*, as a positive control. Caper regulates alternative splicing and is required for proper dendrite arborization of the ddaC class IV da neuron[17].

We crossed a single *UAS-RNAi* line for each gene of interest to $Gal4^{477}$, *UAS-mCD8*::GFP^{4}. The membrane-tethered GFP allowed us to measure the number of dendrite branch termini and dendritic field size in ddaC neurons of late L3 larvae. $Gal4^{477}$ expression begins in newly differentiated class IV da neurons of embryos and continues throughout larval development[18]. $Gal4^{477}$, *UAS-mCD8*::*GFP* crossed to wildtype served as a negative control. *Caper* RNAi caused a significant increase in dendrite branch termini but no increase in dendritic field size since the terminal branches tended to be short and tightly clustered (**Figure 6-4**). These *Caper* RNAi results are very similar to the previously described *Caper* loss-of-function phenotype[17]. For the analysis of *ythdc1*, we used a previously described loss of function mutant[19] combined with $Gal4^{477}$, *UAS-mCD8*::*GFP* as well as a *ythdc1* RNAi line. Neither *Ythdc1* loss of function (**Figure 6-4**) nor ythdc1 RNAi (data not shown) affected the number of dendrite branch termini or field coverage. In contrast, *Heph* RNAi caused a significant decrease in the number of branch termini and

field size (**Figure 6-4**). *Hrg* RNAi also caused a significant decrease in branch termini and field size but the termini / field size value was not significantly different from controls (**Figure 6-4**). While limited to the individual RNAi lines tested, these results suggest that *heph* knockdown primarily affects branching, resulting in fewer termini and reduced dendritic field size, while *hrg* knockdown primarily affects dendrite growth without directly affecting branching.

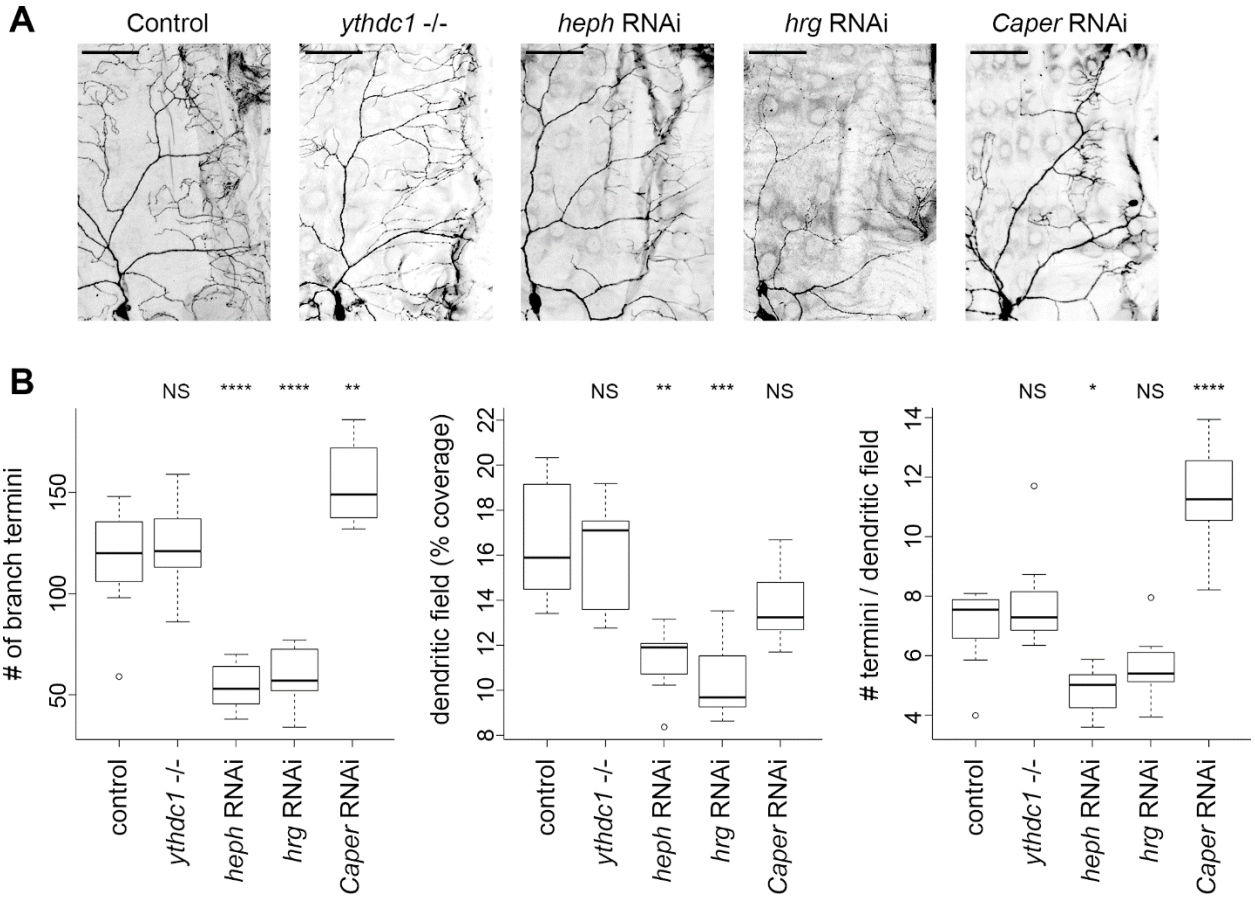


Figure 6-4. Identification of novel regulators of md neuron dendrite arborization.

A. Dendrite arborization in control and mutant class IV ddaC neurons. A representative image is shown for each genotype. The same region was analyzed in all cases, with the cell body positioned at lower left. Larva anterior is to the left and dorsal is at the top. Scale bar

= 50 μm. **B.** Quantification of dendrite arborization. Within the window applied to all neurons (as shown in panel A), the number of branch termini were counted and the total dendrite length was traced and quantified (pixels) then divided by the total pixels of the window to calculate the dendritic field (% coverage). The number of termini per neuron was divided by the dendritic field % coverage to calculate # termini / dendritic field. Sample sizes: control n = 9, *ythdc1-/-* n = 9, *heph* RNAi n = 8, *hrg* RNAi n = 8, *Caper* RNAi n = 8. Data were analyzed using ANOVA with Tukey HSD post-test. P-value code: **** ≤ 0.0001, *** ≤ 0.001, ** ≤ 0.01, * ≤ 0.05.

6.5 Discussion

We selected larval multidendritic sensory neurons for transcription profiling for two reasons: 1) this cell population presents a good target for testing the specificity and sensitivity of EC-tagging and 2) md neurons have previously been used to identify post-transcriptional regulators of dendrite arborization and we reasoned that EC-tagging would enable identification of additional genes in this category. Our prior EC-tagging work[8] used a microarray platform and focused on relatively large populations of cells (all neurons of the central nervous system, mushroom body neurons); in contrast, this work demonstrates that EC-tagging can be combined with RNA-seq and is sensitive enough to identify transcripts in a rare neuronal population. We estimate there are 484 md neurons per larva (22 md neurons per hemisegment[20]) compared to tens of thousands of cells in the larval carcass. Through the use of EUd-tagged and mock-tagged references, we obtained significant enrichment of expected md sensory neuron transcripts and a list of potential novel regulators of md neuron development and function.

Neuronal metabolic RNA labeling experiments have revealed that profiling newly transcribed mRNAs can reveal gene expression dynamics that are missed by more traditional steady-state measurements[21,22]. Analysis of mRNAs synthesized over a relatively short period likely aided our detection of transcripts that have a high rate of decay, such as transcription factors, signaling factors, and RNA binding proteins (classes of genes we have shown to encode low stability mRNAs[13]). For example, in the embryonic nervous system, mRNAs encoding the three RBPs we selected for analysis (*ythdc1*, *heph*, and *hrg*) all have below average half-lives[13]. As previously described[7], metabolic RNA tagging approaches are most effective when comparing purified target cell RNA to a reference generated by metabolic tagging in all cell types of the starting material (tissue or whole animal). For this reason, any genes transcribed in more abundant cells of the carcass are unlikely to be identified as md neuron-enriched in this study. For example, Iyer et al. identified the transcription factors *Poxm* and *Hand* as transcribed in da neurons but these genes are also transcribed in larval muscle and are abundant in our EUd ref-RNA. EC-tagging may be best suited for the discovery of differentially transcribed genes while physical isolation methods such as FACS are better suited for defining complete transcriptomes. Depending on experiment goals, the potential for a less comprehensive transcriptome profile generated by EC-tagging should be weighed against the risks of perturbing gene expression through the use of physical isolation.

RNA-binding proteins have previously been shown to play important roles in md neuron dendrite arborization[4]. Given the significance of this class of post-transcriptional regulators in dendrite arborization, we prioritized the analysis of RBPs from among all the candidate regulators of md neuron development identified in this study. Olesnicky et al.

used a comprehensive RNAi-based screen to identify a large number of RBPs that affect class IV da neuron dendrite morphogenesis. We identified three additional md neuron-enriched RBPs of interest: *ythdc1*, *hrg* and *heph*. Knockdown of *hrg* and *heph* resulted in dendrite arborization defects while *ythdc1* loss of function and knockdown had no effect. The lack of a dendrite arborization phenotype in the *ythdc1* mutants may indicate this is a false positive (not expressed in md neurons). It is difficult to test this possibility without a Ythdc1 antibody, but the widespread expression of Ythdc1 in CNS neurons[14] suggests a general neural function. An alternative explanation is that Ythdc1 controls splicing of transcripts that are irrelevant to dendrite arborization. *Drosophila* motor neurons that lack m6A, the RNA modification recognized by Ythdc1, do not have growth or patterning defects but do have a moderate increase in the number of synaptic boutons and active zones per bouton[14]. Loss of Ythdc1 in md neurons may similarly affect synapses or synaptic activity, phenotypes that would not be detected in our analysis. The *hrg* RNAi phenotype may be explained by Hrg's interaction with Orb, an ortholog of human cytoplasmic polyadenylation element binding protein 1. Hypomorphic alleles of *orb* cause dendrite arborization defects in class IV da neurons[23], with decreased branching and decreased field size similar to what we observed in *hrg* RNAi neurons. Hrg and Orb may regulate cytoplasmic polyadenylation in md neurons, likely targeting mRNAs encoding regulators of dendrite growth. Heph may affect dendrite arborization via its repression of *oskar* (*osk*) translation. Oskar is necessary for proper localization of *nanos* mRNA in class IV da neurons and *osk* loss of function decreases dendrite branching[24], similar to the phenotype we observe in *heph* RNAi neurons. *Osk* mRNA is transported along dendrites[24] and Heph likely represses *osk* translation during transport, as it does in oocytes[16]. In this

model, *heph* knockdown may cause a phenotype similar to *osk* loss of function due to altered Osk distribution in dendrites. Confirming these predicted interactions and further defining mechanisms by which Hrg and Heph control dendrite arborization will be important areas of future investigation.

6.6 Materials and methods

Drosophila genetics

The following lines were obtained from the Bloomington *Drosophila* Stock Center: Oregon-R-P2 (wildtype) (stock # 2376), *Gal4*^{477}, *UAS-mCD8*::*GFP* (stock # 8746), *UAS-ythdc1{RNAi}* (stock #34627), UAS-hrg{RNAi} (stock # 33378), UAS-heph{RNAi} (stock # 55655). For EC-tagging, *Gal4*$^{109(2)80}$ (stock # 8769) was combined with *UAS-CD*:*UPRT* on the 3rd chromosome (stock # 77120) to make the stable line *Gal4*$^{109(2)80}$*; UAS-CD*:*UPRT*. The ythdc1 loss of function mutant, *YT521-NP2]/ TM6C -1*, was provided by Dr. Eric Lai.

EC-tagging, RNA purification and sequencing library preparation

5-ethynylcytosine was synthesized as previously described[8]. Biological replicates were prepared by carrying out 5EC, 5EUd, or mock feeding, carcass dissections and RNA processing independently. Larvae were reared at 25°C and fed 1 mM 5EC or 5EUd from 72–84 hours after hatching. Total carcass RNA was extracted using Trizol. For each treatment, duplicate 20 µg RNA samples (obtained from 25–30 carcasses) were biotinylated using Click-iT Nascent RNA Capture reagents (ThermoFisher) and purified on Dynabeads MyOne Streptavidin T1 magnetic beads (ThermoFisher) as previously described[8]. After the final wash, beads were combined with NuGen Ovation Universal RNA-Seq reagents, following the manufacturer's protocol beginning at first strand cDNA synthesis. Primer annealing and

cDNA synthesis was performed in a heated lid thermomixer to ensure beads did not settle. After cDNA synthesis, beads were washed three times with 500 µl 1X PBS, discarding the supernatant each time. Beads were resuspended in 50 µl RNaseA/T1/H elution mix (1X RNAse H buffer, 12.5 mM D-biotin, RNase A/T1 cocktail (0.1 U/µl), RNase H (0.1 U/µl)) and incubated at 37°C for 30 minutes at 1,000 rpm in a thermomixer. The reaction was stopped by adding 1 µl of DMSO and heating at 95°C for 4 minutes. Beads were collected on a magnet and the supernatant was mixed with DNA binding buffer and applied to a Zymogen DNA Clean and Concentrator-5 column to purify first-strand cDNA. Following purification, the duplicate samples were combined into a single tube and the volume was reduced to 10 µl using a SpeedVac concentrator. The samples were then used to make second-strand cDNA according to the NuGen Ovation Universal RNA-Seq protocol, including adapter ligation and ribosomal RNA depletion using a *Drosophila*-specific AnyDeplete rRNA primer mixture. Libraries were amplified and purified according to the NuGen protocol and quality was assessed using an Agilent Bioanalyzer DNA high-sensitivity chip.

RNA-sequencing and bioinformatics

Sequencing was performed on a HiSeq 2500. Sequence data were trimmed using *Trimmomatic* prior to mapping to the *Drosophila melanogaster* cDNA transcriptome (BDGP6) using *kallisto*. Differential expression analysis was performed using *DESeq2*[25]. The EUd-RNA reference subset was obtained using the *shuf* command in UNIX to randomly select 2.7 million reads from the EUd-RNA BAM files. Gene ontology analysis was performed using DAVID functional annotation clustering[26], with the pre-loaded *Drosophila melanogaster* gene set as background, high classification stringency and default settings for all other parameters. Unnamed genes (those identified only by CG number) were excluded

from GO analysis. Only categories with an enrichment of ≥ 2.0 and Bonferonni-corrected p-values of < 0.01 were considered significant.

Imaging and quantification of dendrite morphology

Dendrite morphology was analyzed in wandering larval stages. Larval fillet preparations were fixed using formaldehyde and stained with rat anti-mCD8 (ThermoFisher) at 1:100 followed by Alexa Fluor 488 anti-rat secondary antibody (ThermoFisher) at 1:200. Imaging was performed using a Zeiss LSM 880 confocal microscope. The number of branch termini and total dendrite length were quantified in Z-series projections. Branch terminal counting and dendrite tracing was performed using Adobe Photoshop and dendrite length (pixels) were quantified using Zeiss Zen Blue software. Dendritic field (% coverage) values were calculated by dividing the number of pixels in the dendrite trace by the total number of pixels in the area analyzed (the fixed-size analysis window applied to all neurons). All statistical analyses were performed in R. Normal distribution of the data for each genotype was confirmed using the Wilk-Shapiro test. Statistical significance was determined using ANOVA with Tukey HSD post-test.

6.7 Supporting information

Supporting information can be downloaded here:

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0240386

6.8 References

1. Jan, Y. & Jan, L. Branching out: mechanisms of dendritic arborization. *Nat Rev Neurosci* **11**, (2010).
2. Gao, F. B., Brenman, J. E., Jan, L. Y. & Jan, Y. N. Genes regulating dendritic outgrowth, branching, and routing in Drosophila. *Genes and Development* **13**, 2549–2561 (1999).
3. Parrish, J., Kim, M., Jan, L. & Jan, Y. Genome-wide analyses identify transcription factors required for proper morphogenesis of Drosophila sensory neuron dendrites. *Genes Dev* **20**, (2006).
4. Olesnicky, E. *et al.* Extensive use of RNA-binding proteins in Drosophila sensory neuron dendrite morphogenesis. *G3 (Bethesda)* **4**, (2014).
5. Hattori, Y. *et al.* Sensory-neuron subtype-specific transcriptional programs controlling dendrite morphogenesis: genome-wide analysis of Abrupt and Knot/Collier. *Dev Cell* **27**, (2013).
6. Iyer, E. *et al.* Functional genomic analyses of two morphologically distinct classes of Drosophila sensory neurons: post-mitotic roles of transcription factors in dendritic patterning. *PLoS One* **8**, (2013).
7. Cleary, M. Uncovering cell type-specific complexities of gene expression and RNA metabolism by TU-tagging and EC-tagging. *Wiley Interdiscip Rev Dev Biol* **7**, (2018).
8. Hida, N. *et al.* EC-tagging allows cell type-specific RNA analysis. *Nucleic Acids Res* **45**, (2017).
9. Ghosh, A. C., Shimell, M., Leof, E. R., Haley, M. J. & O'Connor, M. B. UPRT, a suicide-gene therapy candidate in higher eukaryotes, is required for Drosophila larval growth and normal adult lifespan. *Sci. Rep* **5**, (2015).

10. Nainar, S. *et al.* An optimized chemical-genetic method for cell- specific metabolic labeling of RNA. *Nat Methods* **17**, (2020).

11. Tomorsky, J., DeBlander, L., Kentros, C., Doe, C. & Niell, C. TU-Tagging: A Method for Identifying Layer-Enriched Neuronal Genes in Developing Mouse Visual Cortex. *eNeuro* **4**, (2017).

12. Erickson, T. & Nicolson, T. Identification of sensory hair-cell transcripts by thiouracil-tagging in zebrafish. *BMC Genomics* **16**, (2015).

13. Burow, D. A. *et al.* Dynamic regulation of mRNA decay during neural development. *Neural Development* **10**, (2015).

14. Lence, T. *et al.* m6A modulates neuronal functions and sex determination in Drosophila. *Nature* **540**, (2016).

15. Juge, F., Zaessinger, S., Temme, C., Wahle, E. & Simonelig, M. Control of poly(A) polymerase level is essential to cytoplasmic polyadenylation and early development in Drosophila. *EMBO J* **21**, (2002).

16. Besse, F., de Quinto, S. L., Marchand, V., Trucco, A. & Ephrussi, A. Drosophila PTB promotes formation of high-order RNP particles and represses oskar translation. *Genes and Development* **23**, 195–207 (2009).

17. Olesnicky, E. C., Bono, J. M., Bell, L., Schachtner, L. T. & Lybecker, M. C. The RNA-binding protein caper is required for sensory neuron development in Drosophila melanogaster. *Developmental Dynamics* **246**, 610–624 (2017).

18. Grueber, W., Jan, L. & Jan, Y. Different levels of the homeodomain protein cut regulate distinct dendrite branching patterns of Drosophila multidendritic neurons. *Cell* **112**, (2003).

19. Kan, L. *et al.* The m6A pathway facilitates sex determination in Drosophila. *Nat Commun* **8**, (2017).

20. Singhania, A. & Grueber, W. B. Development of the embryonic and larval peripheral nervous system of Drosophila. *Wiley Interdisciplinary Reviews: Developmental Biology* **3**, 193–210 (2014).

21. Chatzi, C., Zhang, Y., Shen, R., Westbrook, G. & Goodman, R. Transcriptional Profiling of Newly Generated Dentate Granule Cells Using TU Tagging Reveals Pattern Shifts in Gene Expression during Circuit Integration. *eNeuro* **3**, (2016).

22. Zajaczkowski, E. *et al.* Bioorthogonal Metabolic Labeling of Nascent RNA in Neurons Improves the Sensitivity of Transcriptome-Wide Profiling. *ACS Chem Neurosci* **9**, (2018).

23. Olesnicky, E. & Killian, D. The cytoplasmic polyadenylation element binding protein (CPEB), Orb, is important for dendrite development and neuron fate specification in Drosophila melanogaster. *Gene* **738**, (2020).

24. Xu, X., Brechbiel, J. & Gavis, E. Dynein-dependent transport of nanos RNA in Drosophila sensory neurons requires Rumpelstiltskin and the germ plasm organizer Oskar. *J Neurosci* **33**, (2013).

25. Love, M., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, (2014).

26. Huang, W. da, Sherman, B. & Lempicki, R. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, (2009).

Chapter 7: Conclusions

7.1 Conclusions and Future Directions

Since the 1930s, we have had a myriad of ways of treating cancer, however, specifically targeting cancers while sparing normal cells remains a challenge[1]. Complicating the treatment is the existence of cancer stem cells and quiescent cancer cells that do not display the typical characteristics of cancer[2]. Currently, the typical targeting strategies of viral-based gene therapies largely rely on engineering the viral capsid and the cancer-specific promoters, but they come with their own limitations, such as reliance on traditional cancer hallmarks and surface receptors[3,4].

My proposed therapy involves an RNA switch that detects several mutations in SF3B1 that are frequently found in cancers. The basal level of Kid toxin expressed in WT cells is neutralized by constitutively expressed Kis antitoxin. Moreover, the ORAI2 splicing cassette can be shrunken down to ~200bp, facilitating the gene therapy design. Instead of targeting the typical markers of cancer (unchecked proliferation, sustained angiogenesis, evasion from apoptosis, etc.), my proposed therapy targets the defective biochemistry of a pro-oncogenic housekeeping protein[5]. Therefore, my therapy can selectively attack cancer cells as long as they harbor the SF3B1 mutation, regardless of whether the cancer cells are stem-like, differentiated, actively proliferating, or quiescent. The strategy that we developed to clone the Kid toxin expression plasmid achieved constitutive expression in mammalian cells without needing an inducible component, simplifying the gene therapy design. This strategy provides a platform for the constitutive mammalian expression of any bacterial toxin that has an antitoxin pair.

Lastly, the OTetRA inducible gene expression system improves upon the original TetOn system by eliminating the basal leakiness expression in the "off" state. This tight gene expression control is especially important for gene therapies that involve highly toxic payloads such as Kid and Diphtheria toxin.

As cancers typically have more alternative splicing events compared to normal tissues, it is likely that similar splicing-based cancer-recognizing RNA switch modules exist and can be used to treat a wider range of cancers[6,7]. The gene therapy I described in this work offers a new framework for improving the specificity of cancer gene therapies.

7.2 References

1. Arruebo, M. *et al.* Assessment of the Evolution of Cancer Treatment Therapies. *Cancers (Basel)* **3**, 3279 (2011).
2. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
3. Waehler, R., Russell, S. J. & Curiel, D. T. Engineering targeted viral vectors for gene therapy. *Nat Rev Genet* **8**, 573–587 (2007).
4. Payne, K. K., Bear, H. D. & Manjili, M. H. Adoptive cellular therapy of cancer: exploring innate and adaptive cellular crosstalk to improve anti-tumor efficacy. *Future Oncol* **10**, 1779–1794 (2014).
5. Pavet, V., Portal, M. M., Moulin, J. C., Herbrecht, R. & Gronemeyer, H. Towards novel paradigms for cancer therapy. *Oncogene* **30**, 1–20 (2011).
6. Kahles, A. *et al.* Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell* **34**, 211-224.e6 (2018).
7. North, K. *et al.* Synthetic introns enable splicing factor mutation-dependent targeting of cancer cells. *Nat Biotechnol* (2022) doi:10.1038/s41587-022-01224-2.