Title
Mapping and dissecting the post-transcriptional landscape in T cells

Author
Zhu, Wandi

Publication Date
2022

Peer reviewed|Thesis/dissertation

Mapping and dissecting the post-transcriptional landscape in T cells

by
Wandi Zhu

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Jason Cyster
5FFFC327038A40D...

Jason Cyster
Chair

DocuSigned by:
K. Mark Ansel
DocuSigned by:
David Erle
6E06DB0414654B9...

K. Mark Ansel

David Erle

Committee Members

Dedication

To my parents whose courage and resilience gave me the opportunity to pursue a career in science in the US. To my grandparents and relatives, whose support and care carried me through college and graduate school.

Acknowledgements

I would first like to thank my mentor Dr. Mark Ansel for guiding me through my project and graduate school. Whenever I felt doubt or pessimistic about my science, Mark would always be there at our one-on-one meetings or chats to highlight the positive things that I never saw at first. His immunological expertise, breadth of knowledge and scientific engagement have really helped me become a better scientist. I'm also very thankful to have had the opportunity to meet and hangout with the funniest dog, Roosevelt.

My project and thesis would not have been able to come together without the advice and support from my committee members Dr. Jason Cyster and Dr. David Erle. Whenever I got lost or wandered in the wrong direction, they were always there to help give clarity to my project and graduate school trajectory. I would also like to thank all the other mentors I've had through rotations and qualifying committees: Dr. Ari Molofsky, Dr. Alex Marson, Dr. Stephen Floor, Dr. Erin Gordon and Dr. Suneil Koliwad. I'd also like to thank the BMS program coordinator and administrative assistants (Demian Sainz, Ned Molyneaux and Manlio Maineri) for making the grad school experience a smooth one.

I am always thankful for the past and present members of the Ansel lab who've made the lab a fun and helpful space. Special thanks to Dr. Adam Litterman, who was my post-doc mentor in the beginning, and really helped guide the project off the ground. Thank you to Dr Marlys Fassett who has been a great neighbor in the lab bay and always willing to answer science and non-science questions. Thank you to the lab managers and research assistants Darryl Mar, Simon Zhou, Suparna Roy, Jaela Caston and Celeste Garza for helping the lab run smoothly, especially to Celeste who spent her weekends helping with EAE experiments. I am also grateful for the graduate students and other post-docs in the lab: John Gagnon, Eric Wigton, Ben Wheeler, Priscila Muñoz Sandoval and Kristina Johansson. Special thanks to Ben, Priscila and Kristina for helping with my experiments. Never would have been able to do the

assays without them. Also thank you to the Cyster and Lanier lab in HSE10. They graciously lent me reagents that saved a lot of my experiments.

I would not have been able to get into grad school without my undergrad mentor Katie Nicholas, who showed me how to be a great research scientist, and my post-bac mentor Dr. Jyoti Sen, who gave me the room and support to start working as an independent scientist.

Although grad school was long and at times really difficult, many good things came out of it, including all the friends I met in the BMS program. Their support, friendship and food adventures really helped me get through the rough times. The best thing that came out of grad school was meeting my partner Kamir Hiam Galvez. His continual support and presence brought a lot of joy into my life and really helped me make it through the mentally challenging parts of being a grad student. I'm eternally grateful for his support and affection for me.

Lastly, I'd like to thank my parents and grandparents for their love and support throughout my scientific career. I owe all my education, accomplishments and milestones to my parents, Jiajun Zhu and Dr Ying Sun, who have always supported my goals and never wavered in their belief in me. I'd like to especially acknowledge my mom, who is the most amazing and resilient scientist in my eyes. Her perseverance as a PI has been inspiring to see, and I hope to one day have even half the inner strength that she has.

Contributions

Text in Chapter 2 is a modified preprint published on bioRxiv that can be access here: https://www.biorxiv.org/content/10.1101/448654v1. The following authors contributed to the study: Adam J. Litterman, Harshaan S. Sekhon, Robin Kageyama, Maya M. Arce, Kimberly E. Taylor, Wenxue Zhao, Lindsey A. Criswell, Noah Zaitlen, David J. Erle, K. Mark Ansel. Kim and Lindsey provided the PICS SNP; Wenxue and David provided guidance on CRISPR dissection; Noah offered insight into statistical analysis; Maya helped with CRISPR dissection of 3'UTRs, Robin and Adam created GCLiPP technique and analysis pipeline; Harshaan performed computational analysis; Mark and Adam participated in project design, analysis and manuscript writing. Datasets in this paper are available on Gene Expression Omnibus accessions GSE94554 and GSE115886

Chapter 3 is a manuscript in preparation. The following authors contributed to the project: Branka Popovic, Benjamin Wheeler, Sander Engels, Zhongmei Li, Helen Wu, Marissa Chou, Alexander Marson, Monika Wolkers and K Mark Ansel. Branka and Sander helped with the proteomics study. Zhongmei and Helen helped generate mutant mouse models with oversight from Alex. Marissa helped with generating bone marrow chimeras. Benjamin helped with experiments and Mark helped with experimental design, data interpretation and manuscript writing.

Mapping and dissecting the post-transcriptional landscape in T cells

Wandi Zhu

Abstract

T cells undergo dynamic and rapid changes in gene expression and protein output in response to cognate antigen, processes aided by post-transcriptional regulation. RNA binding proteins (RPBs) are important *trans* factors that bind to the 3' untranslated region (3'UTR) of transcripts and modulate mRNA degradation and translation. Loss of function of these proteins, as well as mutations in their cis-regulatory binding sites, can alter T function and development and lead to immune-mediated diseases. While studies have characterized the network of RBPs present in T cells, as well as binding profiles of individual RBPs, the cis-regulatory landscape of the T cell transcriptome is yet to be defined. This body of work addresses this gap by developing a biochemical assay to map the RBP binding profile in T cells and identify post-transcriptional cis-regulatory regions that regulate protein expression and T cell function. Using our technique, GCLiPP, we identified biochemically shared binding sites across conserved regions of human and mouse 3'UTRs that govern transcript stability. Additional dissections of predicted RBP binding sites in several 3'UTRs of immune-related genes revealed cis-regulatory regions critical for modulating protein expression. We performed further analysis on CD69, a negative regulator of T cell egress, and identified a conserved destabilizing region within the 3'UTR. Absence of this region, as well as the whole 3'UTR, increased *CD69* transcript half-life and led to higher CD69 expression as well as more CD69+ T cells in mouse and human. The post-transcriptional circuitry within *Cd69* 3'UTR had a profound effect on T cell migration, as mice with homozygous deletion of this region showed impaired thymic egress. Taken together, our studies provide a resource for mapping post-transcriptional cis-elements in the T cell transcriptome and demonstrate its utility for discovering functional cis-regulatory regions that contain variants associated with human immune-mediated diseases.

Table of Contents

List of Figures

Chapter 2

Chapter 3

List of Tables

Chapter 3

CHAPTER 1: General Introduction

Overview of T cells

T cells are critical players in mounting an immune response against foreign pathogens and malignancies. Their development begins in the thymus where they first undergo β selection, a process that selects for cells that have re-arranged their T cell receptor (TCR) β chain locus, and form a pre-TCR complex that interacts with CD3. Cells with successfully formed TCR signaling complexes upregulate CD4 and CD8 to form a large population of double positive (DP) cells (Germain, 2002). In this stage, thymocytes undergo TCRα rearrangement, forming αβ TCRs, positive selection and down-regulation of CD4 or CD8 to transition into semi-mature single positive (SP) thymocytes (Koch & Radtke, 2011). Once past the negative selection checkpoint at the DP and semi-mature SP stages, SP thymocytes will upregulate protein markers including L-selectin (CD62L), CD69 and sphinegosine-1-phosphate receptor 1 (S1PR1) to become mature SP cells (Xu & Ge, 2014). Using S1PR1, these cells egress from the thymus through S1P-mediated migration and become part of the peripheral repertoire (Allende et al., 2004; Matloubian et al., 2004; Pappu et al., 2007). In the periphery, T cells continuously survey the body, circulating between blood and the lymphatic system, to be on alert for foreign pathogens.

αβ T cells can be broadly separated into two main subsets with differing functions. CD8+ T cells directly kill target cells by producing and secreting effector molecules while CD4+ T cells aid in inflammatory responses by signaling to other immune cells through cytokines and chemokines. As T cells must migrate and mount a quick response against insults at target sites, they must also be able to leave and dampen their response as prolonged proinflammatory state can lead to tissue damage (Mueller et al., 2013). During development and following activation by antigens, T cells undergo dynamic and rapid changes in gene expression and protein output

(Howden et al., 2019; Wolf et al., 2020). Transcriptional control is critical for regulating the abundance of transcripts generated in the cell by inducing or silencing mRNA production in response to specific signals (Kuo & Leiden, 1999). Post-transcriptional regulation is another key step for modulating gene expression and generating an appropriate output (Jurgens et al., 2021). Its regulatory mechanisms can not only aid in robust induction, but also dampen responses within T cells accordingly.

RNA binding proteins

RNA binding proteins (RBPs) and microRNAs (miRNAs) are important mediators of post-transcriptional processing of mRNAs. They bind to the coding region, introns and untranslated regions (UTR) of transcripts through linear and, in the case of RBPs, structural motifs to mediate transcript stability (Jurgens et al., 2021; Turner & DÍaz-Muñoz, 2018). Recent interest has focused on RBPs and their role in T cells. Thousands of proteins have been identified as RBPs (Queiroz et al., 2019; Trendel et al., 2019; van Nostrand et al., 2020), with over 1000 identified through RNA interactome capture methods in primary T cells (Hoefig et al., 2021). These proteins regulate many post-transcriptional processes including RNA editing, localization and translation. Extensive studies into individual RBPs have highlighted the critical role these *trans* factors play in regulating T cell function.

A well-studied class of RBPs are AU-rich element (ARE) binding RBPs, which includes zinc-finger protein 36 (ZFP36) family members, ELAV like RNA binding protein 1 (ELAVL1/HuR) and T cell intercellular antigen 1 (TIA-1). These proteins bind to the ARE motif AUUUA to initiate polyA tail shortening and 3’- and 5’-exonuclease degradation (Akira & Maeda, 2021). Other RBPs, such as Roquin and Regnase family members, exert similar function by binding to a specific stem loop structure (Leppek et al., 2013a; Schlundt et al., 2014; Tan et al., 2014). RBPs can act cooperatively, recruiting other proteins to the site, or competitively for the same binding site (Akira & Maeda, 2021). For example, in macrophages, ZFP36 and ELAVL1 compete for

binding to *Tumor Necrosis Factor α* (*TNFα)* and can either initiate transcript degradation through ZFP36, or stabilize the mRNA through ELAVL1 (Tiedje et al., 2012). Although the proteins can recognize the same motif and site, their capacity to bind to specific regions can differ. In activated primary human T cells, ZFP36 showed higher binding specificity to *Il-2* compared to ELAVL1 (Raghavan et al., 2001). Together, these examples demonstrate the complex and dynamic role RBPs play in post-transcriptional regulation. The following work in Chapter 2 and Chapter 3 will focus specifically on regulation of mRNA stability by RBPs.

RBPs are critical mediators of T cell responses and differentiation

Loss of function studies have shed light on the critical role of RBPs in modulating gene expression and T cell function, specifically restraining T cell responses. Mice with T cell specific deletion of Regnase-1 or Roquin-1/Roquin-2 develop systemic inflammation, autoantibodies and autoimmunity (Jeltsch et al., 2014; Minagawa et al., 2014; Uehata et al., 2013; Vogel et al., 2013). A similar phenotype was also observed in mice with an intact but non-functional Roquin-1 (Yu et al., 2007). These proteins were necessary for preventing inappropriate T cell activation and T follicular helper cell differentiation by degrading key transcripts such as *Ox40* and *Icos* (Essig et al., 2017; Vogel et al., 2013). ZFP36 is another RBP with a functional role in immune tolerance. Mice with condition deletion of this protein in T cells had accelerated T cell activation and quicker responses to viral challenge (Moore et al., 2018).

In contrast to the previously mentioned RBPs, ELAVL1 stabilizes certain transcripts such as *Gata3* (Stellato et al., 2011) in T cells and promotes early type 2 T helper cell differentiation in culture (Techasintana et al., 2017). ELAVL1 also stabilizes *Il17a*, promotes IL-17 production in CD4+ T helper 17 cells (Th17) and contributes to the pathology in the experimental autoimmune encephalomyelitis disease model (J. Chen et al., 2013).

RBPs are not only involved in mRNA degradation but also in translation. Memory T cells generated after an infection or challenge contain pre-formed mRNA necessary for rapid recall

responses, including *Ifng* and *Tnf*. To prevent continuous translation of these transcripts while in a quiescent state, ZFP36 like 2 (ZFP36L2) binds to the 3'UTR of *Ifng* and *Tnf* mRNAs and blocks translation of these rapidly induced proteins (Salerno et al., 2018). ELAVL1 exhibits a similar role as it was necessary for *Il2ra* translation (Techasintana et al., 2017). Together, these examples demonstrate the profound role post-transcriptional regulation and RBPs play in modulating T cell differentiation and responses.

Methods for identifying RBP binding sites

Various methods have been developed to identify RBP binding sites in the transcriptome (Lee & Ule, 2018). RNA immunoprecipitation (RIP) and cross-linking and immunoprecipitation methods in conjunction with sequencing (CLIP-Seq) are the most common techniques used to define individual RBP binding profiles (Hafner et al., 2021a). RIP approaches extract RNA-protein complexes in native conditions. In contrast, CLIP techniques utilize irradiation of cells by UV to irreversibly cross-link RNA and nearby proteins, which allows for more precise purification of ribonucleoprotein complexes (RNPs). After extraction, the RNA is separated from the protein and sequenced to map RBP binding sites. The datasets from these methods are deposited in large repositories such as ENCORE (Encyclopedia of RNA Elements) (van Nostrand et al., 2020), which have become valuable resources for the community. Although many RBP binding profiles have been generated through CLIP, very few have been performed in T cells. CLIP methods are also constrained by the availability of specific antibodies to perform protein IPs and by our knowledge of identified RBPs.

Recent methods have addressed this last concern, using broad approaches to characterize and define RBPs in cells. Some of these methods, such as orthogonal organic phase separation (OOPS) (Queiroz et al., 2019) and XRNAX (Trendel et al., 2019), leverage organic phase separation to extract RNA-bound RBPs and identify the proteins through mass spectrometry. OOPs and global RNA interactome capture (RIC) were used to identify RBPs in

primary mouse and human CD4+ T cells and characterize the networks of RBPs that interact and regulate mRNA targets (Hoefig et al., 2021). Although current work has expanded our understanding of the RBPome in T cells, a global RNA binding profile of the T cell transcriptome remains to be studied.

Cis-regulatory elements in 3'UTRs

The 3'UTR contains cis-regulatory elements that interact with RBPs and miRNAs to initiate post-transcriptional processes. Although a non-coding region, the 3'UTR contains many regions with potential functional activity. Fine-mapping of disease variants associated with immune-mediated disorders identified over hundreds of predicted causal single nucleotide polymorphisms (SNPs) in 3'UTRs (Farh et al., 2015; Taylor et al., 2021), suggesting a potential role for RNA cis-elements in disease pathology. Indeed, there have been examples of mutations in miRNA and RBP binding sites that have led to changes in gene expression and cell function (di Marco et al., 2001; Steri et al., 2018a). For instance, a genome-wide association study identified a SNP in *TNFSF13B,* associated with multiple sclerosis in Sardinians, that altered a miRNA binding site (Steri et al., 2017). This mutation yielded a shorter transcript, which escaped miRNA degradation, and led to increased production of soluble B-cell activating factor (Steri et al., 2017).

Dissection of 3'UTRs have yielded specific insight into individual cis-regulatory elements within the region. Development of genome-wide approaches, such as massively parallel reporter systems, have provided a way to systematically 1) identify novel cis-regulatory elements, 2) evaluate the effects of genetic variation on cis-regulatory regions (Zhao et al., 2014) and 3) analyze the relationships between sequence composition and post-transcriptional regulatory functions (Litterman et al., 2019). These methods, combined with CRISPR-Cas9 mediated functional dissections (Zhao et al., 2017), provide an unbiased approach to understanding the post-transcriptional regulatory networks in 3'UTRs. Despite recent work

systematically uncovering linear and structural motifs (Goodarzi et al., 2012), there remains a need for a map of global cis-regulatory RNA elements in T cells.

Overview of Thesis

The interaction of RBPs and cis-regulatory elements is critical for regulating T cell differentiation and responses, as well as for maintaining the fine balance between inflammatory and quiescent states. This body of work addresses the gaps in the field through the following:

1) Mapping the post-transcriptional cis-regulatory landscape in human and mouse T cells through a new biochemical technique (Chapter 2)
2) Evaluating how post-transcriptional machinery modulates protein expression and T cell function for a specific immunologically important gene (Chapter 3)

References

Akira, S., & Maeda, K. (2021). Control of RNA Stability in Immunity. *Annual Review of Immunology*, *39*(1), 481–509. https://doi.org/10.1146/annurev-immunol-101819-075147

Allende, M. L., Dreier, J. L., Mandala, S., & Proia, R. L. (2004). Expression of the Sphingosine 1-Phosphate Receptor, S1P1, on T-cells Controls Thymic Emigration . *Journal of Biological Chemistry*, *279*(15), 15396–15401. https://doi.org/10.1074/JBC.M314291200

Chen, J., Cascio, J., Magee, J. D., Techasintana, P., Gubin, M. M., Dahm, G. M., Calaluce, R., Yu, S., & Atasoy, U. (2013). Posttranscriptional Gene Regulation of IL-17 by the RNA-Binding Protein HuR Is Required for Initiation of Experimental Autoimmune Encephalomyelitis. *The Journal of Immunology*, *191*(11), 5441–5450. https://doi.org/10.4049/JIMMUNOL.1301188

di Marco, S., Hel, Z., Lachance, C., Furneaux, H., & Radzioch, D. (2001). Polymorphism in the 3′-untranslated region of TNFα mRNA impairs binding of the post-transcriptional regulatory protein HuR to TNFα mRNA. *Nucleic Acids Research*, *29*(4), 863–871. https://doi.org/10.1093/NAR/29.4.863

di Yu, Tan, A. H. M., Hu, X., Athanasopoulos, V., Simpson, N., Silva, D. G., Hutloff, A., Giles, K. M., Leedman, P. J., Lam, K. P., Goodnow, C. C., & Vinuesa, C. G. (2007). Roquin represses autoimmunity by limiting inducible T-cell co-stimulator messenger RNA. *Nature 2007 450:7167*, *450*(7167), 299–303. https://doi.org/10.1038/nature06253

Essig, K., Hu, D., Guimaraes, J. C., Alterauge, D., Edelmann, S., Raj, T., Kranich, J., Behrens, G., Heiseke, A., Floess, S., Klein, J., Maiser, A., Marschall, S., Hrabě de Angelis, M., Leonhardt, H., Calkhoven, C. F., Noessner, E., Brocker, T., Huehn, J., … Heissmeyer, V. (2017). Roquin Suppresses the PI3K-mTOR Signaling Pathway to Inhibit T Helper Cell Differentiation and Conversion of Treg to Tfr Cells. *Immunity*, *47*(6), 1067-1082.e12. https://doi.org/10.1016/J.IMMUNI.2017.11.008

Farh, K. K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., Shoresh, N., Whitton, H., Ryan, R. J. H., Shishkin, A. A., Hatan, M., Carrasco-Alfonso, M. J., Mayer, D., Luckey, C. J., Patsopoulos, N. A., de Jager, P. L., Kuchroo, V. K., Epstein, C. B., Daly, M. J., … Bernstein, B. E. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, *518*(7539), 337–343. https://doi.org/10.1038/nature13835

Germain, R. N. (2002). T-cell development and the CD4–CD8 lineage decision. *Nature Reviews Immunology 2002 2:5*, *2*(5), 309–322. https://doi.org/10.1038/nri798

Goodarzi, H., Najafabadi, H. S., Oikonomou, P., Greco, T. M., Fish, L., Salavati, R., Cristea, I. M., & Tavazoie, S. (2012). Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature 2012 485:7397*, *485*(7397), 264–268. https://doi.org/10.1038/nature11013

Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D., Ule, J., & Zavolan, M. (2021). CLIP and complementary methods. *Nature Reviews Methods Primers 2021 1:1*, *1*(1), 1–23. https://doi.org/10.1038/s43586-021-00018-1

Hoefig, K. P., Reim, A., Gallus, C., Wong, E. H., Behrens, G., Conrad, C., Xu, M., Kifinger, L., Ito-Kureha, T., Defourny, K. A. Y., Geerlof, A., Mautner, J., Hauck, S. M., Baumjohann, D., Feederle, R., Mann, M., Wierer, M., Glasmacher, E., & Heissmeyer, V. (2021). Defining the RBPome of primary T helper cells to elucidate higher-order Roquin-mediated mRNA regulation. *Nature Communications*, *12*(1), 5208. https://doi.org/10.1038/s41467-021-25345-5

Howden, A. J. M., Hukelmann, J. L., Brenes, A., Spinelli, L., Sinclair, L. v., Lamond, A. I., & Cantrell, D. A. (2019). Quantitative analysis of T cell proteomes and environmental sensors during T cell differentiation. *Nature Immunology 2019 20:11*, *20*(11), 1542–1554. https://doi.org/10.1038/s41590-019-0495-x

Jeltsch, K. M., Hu, D., Brenner, S., Zöller, J., Heinz, G. A., Nagel, D., Vogel, K. U., Rehage, N., Warth, S. C., Edelmann, S. L., Gloury, R., Martin, N., Lohs, C., Lech, M., Stehklein, J. E.,

Geerlof, A., Kremmer, E., Weber, A., Anders, H. J., … Heissmeyer, V. (2014). Cleavage of roquin and regnase-1 by the paracaspase MALT1 releases their cooperatively repressed targets to promote TH17 differentiation. *Nature Immunology*, *15*(11), 1079–1089. https://doi.org/10.1038/ni.3008

Jurgens, A. P., Popović, B., & Wolkers, M. C. (2021). T cells at work: How post-transcriptional mechanisms control T cell homeostasis and activation. *European Journal of Immunology*, *51*(9), 2178–2187. https://doi.org/10.1002/EJI.202049055

Koch, U., & Radtke, F. (2011). Mechanisms of T Cell Development and Transformation. *Http://Dx.Doi.Org/10.1146/Annurev-Cellbio-092910-154008*, *27*, 539–562. https://doi.org/10.1146/ANNUREV-CELLBIO-092910-154008

Kuo, C. T., & Leiden, J. M. (1999). TRANSCRIPTIONAL REGULATION OF T LYMPHOCYTE DEVELOPMENT AND FUNCTION. *Annual Review of Immunology*, *17*(1), 149–187. https://doi.org/10.1146/annurev.immunol.17.1.149

Lee, F. C. Y., & Ule, J. (2018). Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Molecular Cell*, *69*(3), 354–369. https://doi.org/10.1016/J.MOLCEL.2018.01.005/ATTACHMENT/D9C4D9B7-CB44-4400-9899-CB2149FA812B/MMC1.XLSX

Leppek, K., Schott, J., Reitter, S., Poetz, F., Hammond, M. C., & Stoecklin, G. (2013). Roquin Promotes Constitutive mRNA Decay via a Conserved Class of Stem-Loop Recognition Motifs. *Cell*, *153*(4), 869–881. https://doi.org/10.1016/J.CELL.2013.04.016

Litterman, A. J., Kageyama, R., le Tonqueze, O., Zhao, W., Gagnon, J. D., Goodarzi, H., Erle, D. J., & Ansel, K. M. (2019). A massively parallel 3′ UTR reporter assay reveals relationships between nucleotide content, sequence conservation, and mRNA destabilization. *Genome Research*, *29*(6), 896–906. https://doi.org/10.1101/GR.242552.118

Matloubian, M., Lo, C. G., Cinamon, G., Lesneski, M. J., Xu, Y., Brinkmann, V., Allende, M. L., Proia, R. L., & Cyster, J. G. (2004). Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. *Nature*, *427*(6972), 355–360. https://doi.org/10.1038/nature02284

Minagawa, K., Wakahashi, K., Kawano, H., Nishikawa, S., Fukui, C., Kawano, Y., Asada, N., Sato, M., Sada, A., Katayama, Y., & Matsui, T. (2014). Posttranscriptional Modulation of Cytokine Production in T Cells for the Regulation of Excessive Inflammation by TFL. *The Journal of Immunology*, *192*(4), 1512–1524. https://doi.org/10.4049/JIMMUNOL.1301619/-/DCSUPPLEMENTAL

Moore, M. J., Blachere, N. E., Fak, J. J., Park, C. Y., Sawicka, K., Parveen, S., Zucker-Scharff, I., Moltedo, B., Rudensky, A. Y., & Darnell, R. B. (2018). ZFP36 RNA-binding proteins restrain T cell activation and anti-viral immunity. *ELife*, *7*, 33057. https://doi.org/10.7554/eLife.33057

Mueller, S. N., Gebhardt, T., Carbone, F. R., & Heath, W. R. (2013). Memory T Cell Subsets, Migration Patterns, and Tissue Residence. *Annual Review of Immunology*, *31*(1), 137–161. https://doi.org/10.1146/annurev-immunol-032712-095954

Pappu, R., Schwab, S. R., Cornelissen, I., Pereira, J. P., Regard, J. B., Xu, Y., Camerer, E., Zheng, Y. W., Huang, Y., Cyster, J. G., & Coughlin, S. R. (2007). Promotion of lymphocyte egress into blood and lymph by distinct sources of sphingosine-1-phosphate. *Science*, *316*(5822), 295–298. https://doi.org/10.1126/SCIENCE.1139221/SUPPL_FILE/PAPPU.SOM.PDF

Queiroz, R. M. L., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.-M., Ramakrishna, M., Harvey, R. F., Dezi, V., Thomas, G. H., Willis, A. E., & Lilley, K. S. (2019). Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nature Biotechnology 2019 37:2*, *37*(2), 169–178. https://doi.org/10.1038/s41587-018-0001-2

Raghavan, A., Robison, R. L., McNabb, J., Miller, C. R., Williams, D. A., & Bohjanen, P. R. (2001). HuA and Tristetraprolin Are Induced following T Cell Activation and Display Distinct but Overlapping RNA Binding Specificities. *Journal of Biological Chemistry*, *276*(51), 47958–47965. https://doi.org/10.1074/JBC.M109511200

Salerno, F., Engels, S., van den Biggelaar, M., van Alphen, F. P. J., Guislain, A., Zhao, W., Hodge, D. L., Bell, S. E., Medema, J. P., von Lindern, M., Turner, M., Young, H. A., & Wolkers, M. C. (2018). Translational repression of pre-formed cytokine-encoding mRNA prevents chronic activation of memory T cells. *Nature Immunology*, *19*(8), 828–837. https://doi.org/10.1038/s41590-018-0155-6

Schlundt, A., Heinz, G. A., Janowski, R., Geerlof, A., Stehle, R., Heissmeyer, V., Niessing, D., & Sattler, M. (2014). Structural basis for RNA recognition in roquin-mediated post-transcriptional gene regulation. *Nature Structural & Molecular Biology 2014 21:8*, *21*(8), 671–678. https://doi.org/10.1038/nsmb.2855

Stellato, C., Gubin, M. M., Magee, J. D., Fang, X., Fan, J., Tartar, D. M., Chen, J., Dahm, G. M., Calaluce, R., Mori, F., Jackson, G. A., Casolaro, V., Franklin, C. L., & Atasoy, U. (2011). Coordinate Regulation of GATA-3 and Th2 Cytokine Gene Expression by the RNA-Binding Protein HuR. *The Journal of Immunology*, *187*(1), 441–449. https://doi.org/10.4049/JIMMUNOL.1001881

Steri, M., Idda, M. L., Whalen, M. B., & Orrù, V. (2018). Genetic variants in mRNA untranslated regions. *Wiley Interdisciplinary Reviews: RNA*, *9*(4), e1474. https://doi.org/10.1002/WRNA.1474

Steri, M., Orrù, V., Idda, M. L., Pitzalis, M., Pala, M., Zara, I., Sidore, C., Faà, V., Floris, M., Deiana, M., Asunis, I., Porcu, E., Mulas, A., Piras, M. G., Lobina, M., Lai, S., Marongiu, M., Serra, V., Marongiu, M., … Cucca, F. (2017). Overexpression of the Cytokine BAFF and Autoimmunity Risk. *New England Journal of Medicine*, *376*(17), 1615–1626.

https://doi.org/10.1056/NEJMOA1610528/SUPPL_FILE/NEJMOA1610528_DISCLOSURES.PDF

Tan, D., Zhou, M., Kiledjian, M., & Tong, L. (2014). The ROQ domain of Roquin recognizes mRNA constitutive-decay element and double-stranded RNA. *Nature Structural & Molecular Biology 2014 21:8*, *21*(8), 679–685. https://doi.org/10.1038/nsmb.2857

Taylor, K. E., Ansel, K. M., Marson, A., Criswell, L. A., & Farh, K. K.-H. (2021). PICS2: next-generation fine mapping via probabilistic identification of causal SNPs. *Bioinformatics*. https://doi.org/10.1093/BIOINFORMATICS/BTAB122

Techasintana, P., Ellis, J. S., Glascock, J., Gubin, M. M., Ridenhour, S. E., Magee, J. D., Hart, M. L., Yao, P., Zhou, H., Whitney, M. S., Franklin, C. L., Martindale, J. L., Gorospe, M., Davis, W. J., Fox, P. L., Li, X., & Atasoy, U. (2017). The RNA-Binding Protein HuR Posttranscriptionally Regulates IL-2 Homeostasis and CD4 + Th2 Differentiation . *ImmunoHorizons*, *1*(6), 109–123. https://doi.org/10.4049/IMMUNOHORIZONS.1700017/-/DCSUPPLEMENTAL

Tiedje, C., Ronkina, N., Tehrani, M., Dhamija, S., Laass, K., Holtmann, H., Kotlyarov, A., & Gaestel, M. (2012). The p38/MK2-Driven Exchange between Tristetraprolin and HuR Regulates AU–Rich Element–Dependent Translation. *PLOS Genetics*, *8*(9), e1002977. https://doi.org/10.1371/JOURNAL.PGEN.1002977

Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M. W., & Krijgsveld, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell*, *176*(1), 391-403.e19. https://doi.org/https://doi.org/10.1016/j.cell.2018.11.004

Turner, M., & DÍaz-Muñoz, M. D. (2018). RNA-binding proteins control gene expression and cell fate in the immune system review-article. In *Nature Immunology* (Vol. 19, Issue 2, pp. 120–129). Nature Publishing Group. https://doi.org/10.1038/s41590-017-0028-4

Uehata, T., Iwasaki, H., Vandenbon, A., Matsushita, K., Hernandez-Cuellar, E., Kuniyoshi, K., Satoh, T., Mino, T., Suzuki, Y., Standley, D. M., Tsujimura, T., Rakugi, H., Isaka, Y.,

Takeuchi, O., & Akira, S. (2013). Malt1-Induced Cleavage of Regnase-1 in CD4+ Helper T Cells Regulates Immune Activation. *Cell*, *153*(5), 1036–1049. https://doi.org/10.1016/J.CELL.2013.04.034

van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., Xiao, R., Blue, S. M., Chen, J.-Y., Cody, N. A. L., Dominguez, D., Olson, S., Sundararaman, B., Zhan, L., Bazile, C., Bouvrette, L. P. B., Bergalet, J., Duff, M. O., Garcia, K. E., Gelboin-Burkhart, C., … Yeo, G. W. (2020). A large-scale binding and functional map of human RNA-binding proteins. *Nature*, *583*(7818), 711–719. https://doi.org/10.1038/s41586-020-2077-3

Vogel, K. U., Edelmann, S. L., Jeltsch, K. M., Bertossi, A., Heger, K., Heinz, G. A., Zöller, J., Warth, S. C., Hoefig, K. P., Lohs, C., Neff, F., Kremmer, E., Schick, J., Repsilber, D., Geerlof, A., Blum, H., Wurst, W., Heikenwälder, M., Schmidt-Supprian, M., & Heissmeyer, V. (2013). Roquin Paralogs 1 and 2 Redundantly Repress the Icos and Ox40 Costimulator mRNAs and Control Follicular Helper T Cell Differentiation. *Immunity*, *38*(4), 655–668. https://doi.org/10.1016/J.IMMUNI.2012.12.004

Wolf, T., Jin, W., Zoppi, G., Vogel, I. A., Akhmedov, M., Bleck, C. K. E., Beltraminelli, T., Rieckmann, J. C., Ramirez, N. J., Benevento, M., Notarbartolo, S., Bumann, D., Meissner, F., Grimbacher, B., Mann, M., Lanzavecchia, A., Sallusto, F., Kwee, I., & Geiger, R. (2020). Dynamics in protein translation sustaining T cell preparedness. *Nature Immunology 2020 21:8*, *21*(8), 927–937. https://doi.org/10.1038/s41590-020-0714-5

Xu, X., & Ge, Q. (2014). Maturation and migration of murine CD4 single positive thymocytes and thymic emigrants. *Computational and Structural Biotechnology Journal*, *9*(15), e201403003-9. https://doi.org/10.5936/CSBJ.201403003

Zhao, W., Pollack, J. L., Blagev, D. P., Zaitlen, N., McManus, M. T., & Erle, D. J. (2014). Massively parallel functional annotation of 3′ untranslated regions. *Nature Biotechnology 2014 32:4*, *32*(4), 387–391. https://doi.org/10.1038/nbt.2851

Zhao, W., Siegel, D., Biton, A., le Tonqueze, O., Zaitlen, N., Ahituv, N., & Erle, D. J. (2017). CRISPR–Cas9-mediated functional dissection of 3′-UTRs. *Nucleic Acids Research*, *45*(18), 10800–10810. https://doi.org/10.1093/NAR/GKX675

CHAPTER 2: A global map of RNA binding protein occupancy guides functional dissection of post-transcriptional regulation of the T cell transcriptome

Abstract

RNA binding proteins (RBPs) mediate constitutive RNA metabolism and gene specific regulatory interactions. To identify RNA cis-regulatory elements, we optimized a biochemical technique, GCLiPP, that detects RBP occupancy transcriptome-wide in T cells. GCLiPP sequence tags corresponded with known RBP binding sites, specifically correlating to abundant cytosolic RBPs. Comparison of human Jurkat T cells and mouse primary T cells uncovered hundreds of biochemically shared peaks of GCLiPP signal across homologous regions of human and mouse 3' UTRs, including the 3' UTR of the proto-oncogene *PIM3*. To demonstrate the utility of our occupancy profiles, we performed functional dissection of 3'UTRs of immunological genes using CRISPR/Cas9 genome editing. Detailed mapping revealed a cis-regulatory element corresponding to a biochemically shared peak in *PIM3* 3'UTR that governs transcript stability in mouse and human, as well as potential stabilizing regions in *CD5, STAT6* and *IKZF1* 3'UTR. Our GCLiPP datasets and visualization tool provide a rich and easily accessible resource for investigation of post-transcriptional regulation in the immune system.

Introduction

The life cycle of protein coding RNA transcripts involves their transcription from DNA, 5′ capping, splicing, 3′ polyadenylation, nuclear export, cellular localization, translation and degradation (Garneau et al., 2007; Martin & Ephrussi, 2009; Reed, 2003). RNA binding proteins (RBPs) coordinately regulate these processes through interaction with RNA cis-regulatory elements, often in the 5′ and 3′ untranslated regions (UTRs) whose sequences are not constrained by a functional coding sequence (Keene, 2007). Mammalian genomes encode hundreds of RBPs (Castello et al., 2012) and mutations in individual RBPs or binding sites can induce strong developmental, autoimmune and neurological defects in human patients and mouse models (Bassell & Kelic, 2004; Gebauer et al., 2021; Kafasla et al., 2014; Schwerk & Savan, 2015).

Recent systematic analyses that map protein-RNA interactions have expanded our understanding of post-transcriptional regulatory circuits. Methods utilizing organic phase separation such as protein-Xlinked RNA extraction (XRNAX) (Trendel et al., 2019), orthogonal organic phase separation (OOPS) (Queiroz et al., 2019) and others (van Ende et al., 2020) have expanded the repertoire of known RBPs through unbiased extraction of over 1000 distinct RNA-bound proteins in human cell lines and primary T cells (Hoefig et al., 2021). These studies, as well as those employing RNA interactome capture techniques (Perez-Perri et al., 2018a), have mostly focused on the *trans* factors involved in RNA regulation. Systematic analysis of the corresponding protein-occupied cis-regulatory regions in mRNAs remains limited to human cell lines (Baltz et al., 2012) and yeast (Freeberg et al., 2013). In addition, a biochemical study of 150 individual RBPs and their binding sites through enhanced CLIP (eCLIP) as part of the Encyclopedia of DNA elements (ENCODE) project has provided a valuable database for evaluating RNA cis-regulatory elements (a.k.a. ENCORE) (van Nostrand et al., 2020). New modes of transcriptome analysis inspired new methods for computational analysis of the

resulting data, which require dedicated tools to quantify, compare and catalogue transcriptomic features (Hafner et al., 2021b).

Post-transcriptional regulation plays an important role in fine-tuning T cell responses to external stimuli (Nicolet et al., 2021). As much as half of the extensive gene expression changes that occur during T cell activation occur post-transcriptionally (Raghavan et al., 2002), and several RBPs are known to be critical determinants of immune function and homeostasis (Kafasla et al., 2014). A large proportion of probable causal genetic variants associated with immune-mediated diseases map to non-coding regions with potential regulatory functions in immune cells (Farh et al., 2015; Steri et al., 2018b), but the mechanistic role of the large majority of these variants in immune cells is unknown. A map of RBP occupancy in T cells can be a powerful tool for interrogating post-transcriptional gene regulation in the immune system and, in combination with genetic analysis, dissecting the genetic basis of immune-mediated diseases.

Here, we create global RBP occupancy maps for a human T cell line, Jurkat, and primary mouse T cells. Comparing RBP occupancy for thousands of mRNAs across species identified biochemically shared regulatory sites, which are enriched for phylogenetically conserved sequences. With the map as guidance, we used a scalable system of CRISPR dissection to define regions of functional activity in 3’ UTRs of mouse and human transcripts of immunological importance through shared conservation. Targeted analysis revealed regulatory activity in regions of RBP occupancy that mapped to probable-causal genetic variants associated with human immune-mediated diseases. Our findings demonstrate the utility of RBP occupancy maps as valuable resources for functional analysis of post-transcriptional regulation in T cells.

Results

Transcriptome-wide analysis of RBP occupancy in T cells

To achieve transcriptome-wide RBP binding site profiling in T cells, we adapted biochemical methods for crosslinking purification of all mRNA-RBP complexes. Our Global CrossLinking Protein Purification method, abbreviated as GCLiPP, features: crosslinking of endogenous ribonucleoprotein complexes using high energy UV light (no photo-crosslinkable ribonucleotide analogues); oligo-dT pulldown prior to biotinylation to enrich for mRNA species; chemical biotinylation of primary amines using a water soluble reagent with a long, flexible linker; brief RNase digestion with RNase T1; and on-bead linker ligation with radiolabeled 3′ linker to facilitate downstream detection of ligated products (Fig 2.1A). We used the guanine specific ribonuclease T1 to favor larger average fragment sizes than would be produced with an RNA endonuclease with less stringent nucleotide specificity, such as RNase A. We first applied GCLiPP to interrogate RBP-occupied regions of RNA in human Jurkat T cells. Linker-ligated RBP-protected fragments were separated by PAGE and detected by radiography (Fig 2.1B, lanes 1-3). Single-stranded RNA oligonucleotides of 19 and 24 nt, the same length as the 5’ and 3’ linkers, were ligated to the radiolabeled 3’-linker and served as size markers (Fig 2.1B, lane 4). Material greater than 24 nt + 3’-linker in length were predicted to contain RBP-bound RNA fragments, and these were extracted and processed for small RNA library preparation and sequencing. Excluding the protein biotinylation or UV-crosslinking steps greatly diminished the yield of ligated RNA fragments (Fig 2.1B, lanes 5-8), indicating that the GCLiPP procedure preferentially captures RNA sequences interacting with RBPs in living cells.

We called local peaks of GCLiPP sequence read density and measured the distribution of GCLiPP reads within those peaks to assess the reproducibility of the technique. Local read density within individual transcripts was similar between experiments, as GCLiPP fragments yielded highly reproducible patterns in technical replicates (Fig 2.1C). The distribution of read

coverage from Jurkat GCLiPP libraries was strongly enriched within mature mRNAs and long non-coding RNAs (Fig 2.1D-E) compared to other transcriptome features.

RBPs bind to linear and structural motifs to regulate the stability and/or translation of the mRNAs that they bind (Corley et al., 2020). We observed GCLiPP read coverage corresponding to known RBP recognition motifs. Nuclear Receptor subfamily 4 group A member 1 (*NR4A1)*, which encodes the NUR77 protein that mediates T cell tolerance, is an example of RBP-mRNA interaction through linear sequence recognition. A local maximum of GCLiPP read density in the *NR4A1* 3'UTR corresponded with a region that contains multiple AU rich elements (AREs) that destabilize mRNA (C.-Y. A. Chen & Shyu, 1994) (Fig 2.1F). Similarly, the 3'UTR of *IER3*, an immediate early response gene that protects cells from Fas- or TNFα-induced apoptosis, contains a local maximum of GCLiPP read coverage at the previously characterized structurally-determined stem-loop binding motif regulated by the RBP Roquin (Leppek et al., 2013b) (Fig 2.1G). These examples provide snapshots of different motifs represented in GCLiPP protein occupancy maps. Further examination of individual 3'UTRs of interest can be accessed through our visualization tool, Thagomizer (http://thagomizer.ucsf.edu). Thagomizer utilizes a database of GCLiPP and Argonaute 2 (Ago2) HITS-CLIP experiments (Gagnon et al., 2019; Loeb et al., 2012) along with miRNA binding site predictions from the TargetScan database (Agarwal et al., 2015) to map RBP-mRNA and miRNA-mRNA interactions in mouse and human T cell 3' UTRs.

We also performed a systematic analysis to determine the dominant structural characteristics of protein-occupied RNA regions detected by GCLiPP. We used CLIPper (Lovci et al., 2013) to call peaks in our data, and calculated the base-pairing probability for every nucleotide pair in each 200bp sequence peak using RNAfold in the ViennaRNA package(Lorenz et al., 2011). Matrices for all peaks were averaged to generate an average base-pairing probability. This analysis revealed a decreased probability of base-pairing at the center of GCLiPP peaks compared to surrounding regions, indicating an enrichment for single-stranded RNA (ssRNA) at the center of GCLiPP peaks in Jurkat cell 3'UTRs (Fig 2.2A). A similar pattern

was observed in eCLIP peaks for Polypyrimidine Tract Binding Protein 1 (PTBP1), an RBP that binds to C/U-rich ssRNA through 4 RNA recognition motif (RRM) domains (Supplementary Fig 2.1A) (Oberstrass et al., 2005). Among known RBPs, those with ssRNA-binding RRM domains are the highest expressed in Jurkat cells (Fig 2.2B). RBPs captured in primary human CD4 T cells through RNA interactome capture (RNA-IC) (Hoefig et al., 2021) also predominantly contained the RRM motif compared to other domains (Fig 2.2C). Together, these data indicate that RBP-occupied regions detected by GCLiPP are predominantly composed of the most abundant structural motif, ssRNA, in T cells.

GCLiPP read density represents cytosolic RBP occupancy

Further analysis of GCLiPP and other interactome capture methods revealed shared characteristics. We compared CLIPper-called peaks in Jurkat GCLiPP data with compiled peaks from ENCORE eCLIP datasets (Sundararaman et al., 2016), and peaks detected in the phase separation-based RNA interactome methods, XRNAX (Trendel et al., 2019) and OOPs (Queiroz et al., 2019). We then assessed the phylogenetic conservation of RBP-RNA interaction sites detected by each of these techniques. PhyloP scores for each 200nt CLIPper called peak were averaged for all binding sites and then normalized around a mean of 0. Both GCLiPP and ENCORE peaks displayed greater sequence conservation at peak centers, although GCLiPP showed a slightly broader local maximum of conservation, perhaps indicating lower resolution of RBP binding sites (Fig 2.2D). We observed a similar, but even broader pattern of phylogenetic conservation in global RBP binding data from XRNAX and OOPS. We conclude that all of these methods, including GCLiPP, capture conserved RBP binding sites in the transcriptome.

To determine whether GCLiPP detects previously identified predicted RBP binding sites, we systematically compared GCLiPP occupancy maps and eCLIP analyses of RBP binding profiles from ENCORE (Sundararaman et al., 2016). We examined pairwise correlations of normalized read density across individual 3’ UTRs between GCLiPP and individual RBP eCLIP

samples (Fig 2.2E, Supplementary Fig 2.1B). In parallel, we compared GCLiPP to the input control for each eCLIP experiment. eCLIP for many RBPs, such as TIA1 and IGF2BP1, matched GCLiPP read density much more closely than the eCLIP control input across the transcriptome (Fig 2.2F, Supplementary Fig S2.1C), indicating a relatively high contribution of these RBPs to the overall GCLiPP signal. For other proteins, such as PUM2, this comparison showed poor correlation, indicating a low contribution to total RBP occupancy transcriptome-wide. Yet we found evidence that GCLiPP captured focal RBP binding to specific sites (UGUA motifs in the case of PUM2) that were overrepresented in GCLiPP reads, although they did not dominate the overall signal (Supplementary Fig 2.1B, bottom panel). This was revealed when we called GCLiPP peaks with CLIPper (Lovci et al., 2013) and compared these peaks with CLIPper called peaks in eCLIP datasets. The observed fraction of PUM2 eCLIP peaks that overlap GCLiPP peaks (0.56) was much greater than the fraction overlapping eCLIP peaks randomly shuffled across the 3' UTRs from which they were derived (Supplementary Fig 2.1D, bottom panel). Similar results were obtained for TIA-1 (Fig 2.2G) and IGF2BP1 (Supplementary Fig 2.1D, top panel). These enrichments above background binding for IGF2BP1, TIA1 and PUM2 were among the highest 8 of the 87 RBPs whose eCLIP signals were examined (Supplementary Fig 2.2).

We performed genome wide correlation analysis for 87 RBPs obtained from eCLIP data, and compared the correlation between eCLIP and GCLiPP with RBP abundance previously determined via mass spectrometry (Baltz et al., 2012). There was an overall significant correlation between RBP abundance and correspondence between RBP eCLIP and GCLiPP profiles ($r=0.28$, $p=0.02$). However, stratifying RBPs by their predominant cellular localization (Binder et al., 2014) showed that this correlation was driven almost entirely by cytosolic RBPs with no correlation for non-cytoplasmic RBPs (Fig 2.2H, Supplementary Fig 2.2E). The fraction of eCLIP peaks that overlapped GCLiPP peaks above a shuffled background was also significantly greater for cytosolic versus non-cytosolic RBPs ($p=0.003$, Supplementary Fig 2.2

inset). These findings were expected, as the GCLiPP experimental protocol preferentially samples the cytosol by eliminating most nuclear material. In summary, we conclude that GCLiPP read density represents conserved sequences and predominantly reflects cytosolic RBP occupancy.

RBP Occupancy of RNA cis-regulatory elements in primary T cells

Previous global RBP profiling has been conducted with cell lines. To examine transcriptome-wide RBP occupancy in primary T cells, we performed GCLiPP on primary mouse CD8 and CD4 type 2 helper T cells (Th2) (Fig 2.3A). Local read density at peaks showed reproducible patterns between multiple pooled experiments for the two T cell subsets (Supplementary Fig 2.3A). Similar to Jurkat cells (Fig 2.1D-E), distribution of reads in primary mouse T cells was enriched in mature transcripts and long non-coding RNAs (Supplementary Fig 2.3B-C). The most striking difference was the greater proportion of reads derived from transposable elements in mouse GCLiPP libraries. This increase is likely due to the greater amount of annotated transposable elements in the mouse genome since the relative coverage of these elements was similar between species. We examined the GCLiPP profiles at previously characterized cis-regulatory elements of various functional and structural categories in primary mouse T cells. As in Jurkat cells, we observed GCLiPP read density at Roquin/Regnase binding site in the 3′ UTR of *Ier3* (Fig 2.3B).

Known cis-regulatory elements involved in transcript localization were also represented by local regions of GCLiPP read density. The Beta-actin “zipcode” element is responsible for localization of *Actb* mRNA to the cellular leading edge in chicken embryo fibroblasts (Kislauskis et al., 1994) and contains conserved linear sequence elements separated by a variable linker. These conserved sequence elements are thought to form the RNA/protein contacts in a complex involving the actin mRNA and the RNA binding protein Igf2bp1 (previously known as Zbp1) where the non-conserved sequence winds around the RBP (Chao et al., 2010). This

sequence corresponds to the center of the second highest peak of GCLiPP read density in the *Actb* transcript (Fig 2.3C).

The canonical polyadenylation signal AAUAAA is a known linear sequence motif that binds to a number of RBPs in the polyadenylation complex, including CPSF and PABP (Millevoi & Vagner, 2010), as part of constitutive mRNA metabolism. We examined T cell lineage-defining transcripts with well-resolved GCLiPP profiles (due to their high expression levels), including *Cd3g* (Fig 2.3D), *Cd3e, Cd4*, and *Cd8b1* (Supplementary Fig 2.4). The canonical polyadenylation signal sequences in these transcripts were contained within called GCLiPP peaks, often as the peak with the highest GCLiPP read density in the entire transcript. Interestingly, the GCLiPP profile of *Cd8b1* contained direct biochemical evidence for alternative polyadenylation signal usage (Supplementary Fig 2.4C), a phenomenon that has previously been described to be important in activated T cells (Sandberg et al., 2008). GCLiPP peaks appeared in multiple canonical polyadenylation signal sequences in *Cd8b1*, coincident with clear evidence for both short and long 3′ UTR isoform usage indicated by lower RNAseq read counts after the initial canonical polyadenylation signal. A similar pattern was apparent in *Hifa* (Supplementary Fig 2.4D) and a number of other highly expressed transcripts.

The insertion of the selenium containing amino acid selenocysteine into selenoproteins represents a unique case of RBP regulation of protein translation. Selenoproteins are redox enzymes that use selenocysteine at key reactive residues (Johansson et al., 2005; Vanda Papp et al., 2007). Selenocysteine is encoded by the stop codon UGA. This recoding occurs only in mRNAs that contain 3′ UTR cis-regulatory elements (termed SECIS elements) that bind to RBPs that recruit the elongation factor Eefsec and selenocysteine-tRNA (Berry et al., 1993; Tujebajeva et al., 2000). SECIS elements were prominent peaks of GCLiPP read coverage in selenoprotein mRNAs. For example, the predicted SECIS element (Mariotti et al., 2013) in the 3′ UTR of *Gpx4* was entirely covered by GCLiPP reads (Fig 2.3E). Indeed, a canonical polyadenylation signal and the full hairpin structure containing the SECIS element account for

essentially all of the GCLiPP reads in the *Gpx4* 3′ UTR (Fig 2.3F). Comparing transcriptome-wide *in vivo* folding data from icSHAPE (Spitale et al., 2015) and GCLiPP data supports the identification of an RBP bound, structured SECIS element (Fig 2.3G-H). Furthermore, this analysis suggests that the folded, RBP bound structure is even larger than that predicted by SECISearch 3, with regions of GCLiPP read density and apposed high and low icSHAPE signals spanning almost the entire 3′ UTR. Thus, GCLiPP recapitulated previously described structured and single-stranded RNA cis-regulatory elements that mediate constitutive RNA metabolism, transcript localization, regulation of gene expression, and translation.

Cross-species comparison of GCLiPP reveals patterns of biochemically shared post-transcriptional regulation

Next, we sought to compare RBP occupancy in mouse and human T cells. To do so, we performed Clustal Omega sequence alignments of thousands of human 3' UTRs and their corresponding sequences in the mouse genome, and then designed an algorithm to identify correlated peaks of normalized GCLiPP read density along the aligned nucleotides (Fig 2.4A). Using this approach, we identified 1047 high-stringency biochemically shared GCLiPP peaks derived from 901 3' UTRs. As a class, biochemically shared peaks exhibited significantly higher sequence conservation than the full 3' UTRs in which they reside (Fig 2.4B). The highly conserved, biochemically shared peak in *USP25* exemplifies this general pattern (Fig 2.4C, right panel). However, many biochemically shared peaks did not exhibit corresponding increases in local sequence conservation. For example, the *ARRB2* mRNA that encodes β-arrestin, another regulator of T cell migration in response to chemoattractant gradients (Fong et al., 2002), exhibited a common peak of RBP occupancy in Jurkat cells and primary mouse T cells that is roughly equally conserved as the rest of the 3' UTR (Fig 2.4C, left panel).

To examine which RBPs contributed to biochemically shared peaks more than other GCLiPP peaks, we used HOMER motif calling software (Heinz et al., 2010) to identify enriched

motifs. Strikingly, of the six linear sequence motifs present in >10% of biochemically shared peaks with $p \leq 10^{-10}$, five resemble well-known regulatory sequences (Fig 2.4D). The two most common appeared to represent canonical CELF (Timchenko et al., 1996) and PUM (Hafner et al., 2010) binding motifs. Three other identified motifs corresponded to runs of homo-polymers: an A-rich motif that resembled the canonical polyadenylation signal (Proudfoot, 2011); a poly-U containing motif similar to a sequence that has long been known to stabilize mRNAs (Zubiaga et al., 1995) and a poly-C containing motif similar to the C-rich RNAs bound by poly-C binding proteins (Makeyev & Liebhaber, 2002). We used Metascape (Tripathi et al., 2015) to identify categories of biologically related genes enriched among mRNAs that contained biochemically shared GCLiPP peaks (Fig 2.4E). Interestingly, 3 of the 5 most enriched categories were related to RNA regulation ("regulation of mRNA metabolism," "large Drosha complex," "RNA splicing"), with the broad category "post-transcriptional regulation of gene expression" also in the top 10. Thus, biochemically shared GCLiPP binding sites are generally more well conserved than their local sequence context, enriched for well-studied RBP binding motifs, and occur preferentially in genes that encode proteins involved in post-transcriptional gene regulation. Together, these observations suggest the presence of conserved autoregulatory gene expression networks.

GCLiPP-guided CRISPR dissection of biochemically shared post-transcriptional cis-elements

We hypothesized that functionally conserved destabilizing cis-regulatory elements could be identified by examining biochemically shared GCLiPP peaks in 3' UTRs of labile transcripts. To prioritize candidates, we computed Pearson correlation coefficients for the normalized GCLiPP profiles of 3' UTRs of genes expressed in both Jurkat cells and primary mouse T cells (Fig 2.5A, black histogram) and examined transcript instability by RNAseq analysis of primary mouse T cells treated with actinomycin D (Fig 2.5A, red histogram). The proto-oncogene *PIM3* emerged as an outstanding candidate with both strong interspecies GCLiPP correlation and

very high transcript instability. Alignment of the GCLiPP profiles of human and mouse *PIM3* revealed a dominant shared peak of GCLiPP read density (Fig 2.5B). This peak corresponded to a highly conserved region of the transcript that contains a G-quadruplex, followed by a putative AU-rich element (ARE) and a CELF binding motif (Fig 2.5C). Another conserved region with G-quadruplex followed by a putative ARE appeared upstream of the biochemically share GCLiPP peak. We numbered these conserved regions ARE1 and ARE2 according to their order in the 3'UTR, and hypothesized that ARE2 would exert greater cis-regulatory activity than ARE1, given its RBP occupancy in both species and the relative lack of occupancy in ARE1. To test this hypothesis, we performed CRISPR dissections of both the human and mouse *PIM3* 3' UTRs (Fig 2.5). These analyses produced largely concordant patterns of post-transcriptional cis-regulatory activity in the human and mouse 3'UTR with the greatest significant destabilizing effect corresponding to the shared region of GCLiPP read intensity covering the CR2 element (Fig 2.5 D-K). Consistent with this portrait of the entire 3' UTR, when we filtered specifically for mutations that completely deleted either CR1 or CR2, we observed significantly greater expression of transcripts derived from cells with CR2 deleted versus CR1 (Fig 2.5 L-M). Thus, *PIM3* is a very unstable transcript with highly concordant RBP occupancy in human and mouse. Functional dissection of the post-transcriptional regulatory landscape of this gene revealed that this biochemical concordance between mouse and human cells is mirrored at a functional level, with the most highly occupied region indicated by GCLiPP read density corresponding to the most destabilizing region of the 3' UTR.

GCLiPP-guided CRISPR dissection of biochemically shared post-transcriptional cis-elements

We hypothesized that RBP occupancy maps may help guide functional annotation of sequence variants in cis-regulatory elements. To test this, we intersected our Jurkat GCLiPP peaks with probable casual single nucleotide polymorphisms (SNPs) associated with human

immune-mediated diseases. A previously developed algorithm, Probabilistic Identification of Casual SNPs (PIC2) (Farh et al., 2015) identified candidate causal SNPs through fine-mapping that were linked to an immune-mediated disease. PICS2 (Taylor et al., 2021) has expanded that list to include variants identified with more recently collected GWAS data. Within these variants, we identified 63 SNPs that occurred within a GCLiPP peak in a 3'UTR in Jurkat cells. These variants were associated with a variety of immune-mediated disorders and appeared in a variety of genes that are expressed in T cells (Fig 2.6A).

To test whether disease-associated probable causal variants overlapping GCLiPP peaks mark functional RNA cis-regulatory elements, we deleted 4 individual RBP binding sites in the 3'UTRs of 3 distinct immunologically important genes using a dual guide RNA (gRNA) CRISPR-Cas9 editing approach. *CD5* encodes an inhibitory receptor expressed on T cells (Voisinne et al., 2018). It's 3'UTR contains a probable causal SNP associated with rheumatoid arthritis that lies within a GCLiPP peak detected in Jurkat cells (Fig 2.6B). Deletion of this region with paired gRNAs (at 50-60% editing efficiency; data not shown) decreased CD5 expression (Fig 2.6C), suggesting the presence of a cis-regulatory element in the 3'UTR.

SNP rs1059513 in the 3'UTR of *STAT6* had a PICS2 probability score of 0.985 for association with allergy, making it by far the most likely causal variant in the locus for this trait. STAT6 is an important signaling protein and transcription factor that is pivotal for mounting a type 2 inflammatory response. It is activated by Janus kinase (JAK)-mediated phosphorylation downstream of IL-4 and IL-13 signaling (Goenka & Kaplan, 2011). To determine whether the identified RBP binding site affected STAT6 expression and function, we used CRISPR-Cas9 to generate a small deletion (Fig 2.6D) and treated the edited cells with IL-4 to measure phospho-STAT6 (pSTAT6). *STAT6* 3'UTR edited cells showed similar phosphorylation kinetics as control (Supplementary Fig 2.5A), but overall decreased pSTAT6 expression compared to controls (Fig 2.6E, Supplementary 2.5B).

Finally, we intersected PICS2 and GCLiPP data to identify functional elements in the 3'UTR of *Ikaros family zinc finger 1 (IKZF1)*, a pleiotropic transcription factor involved in lymphocyte differentiation (Heizmann et al., 2018). We generated two separate deletions (Fig 2.7F), both of which contained a probable causal SNP associated with systemic lupus erythematosus. Jurkat cells lacking either region expressed less IKZF1 protein compared to control cells (Fig 2.6G-H). In summary, a GCLiPP-guided analysis of probable causal SNPs in 3'UTRs efficiently identified functional RNA cis-regulatory elements that regulate protein expression. These findings demonstrate the utility of a transcriptome-wide profile of RBP occupancy in the T cell transcriptome.

Discussion

Interconnected networks of bound RBPs and RNAs form a complex layer of post-transcriptional regulation that affects all biological processes. Understanding these networks remains one of the key challenges in deciphering how the genome encodes diverse cell identities and behaviors (Farh et al., 2015; Simeonov et al., 2017). Methods like DNase I hypersensitivity and ATAC-seq that query regulatory element accessibility and occupancy without prior knowledge of their protein binding partners have proven themselves as powerful techniques for the systematic mapping of cis-regulatory sequences in DNA (Buenrostro et al., 2013; Thurman et al., 2012). Their development has allowed for comparisons in the regulatory structure of diverse cell types (Corces et al., 2016) and for functional analysis of genetic variants (Simeonov et al., 2017). Large-scale analyses of individual RBPs have begun the intensive process of documenting RBP binding sites in the transcriptome of a few model cell systems, providing a useful repository of RNA regulatory data (van Nostrand et al., 2020). Pioneering work utilizing RNA-interactome capture (RIC) (Perez-Perri et al., 2021) and organic phase separation (Smith et al., 2020) have furthered our understanding of RNA-RBP interactome.

Using GCLiPP, an RNA interaction capture technique, we generated and validated a RBP binding map of the transcriptome in T cells and used it as a guide to identify cis-regulatory elements in 3'UTRs. As ATAC-seq has been used to define global regulatory elements involved in transcription, we used GCLiPP to discover RNA regulatory elements that mediate post-transcriptional gene regulation.

Dissection of the human *PIM3* and mouse *Pim3* 3'UTRs demonstrated the utility of GCLiPP for decoding biochemically shared and functionally conserved post-transcriptional regulation. The PIM family of serine/threonine kinases exert profound regulatory effects on MYC activity, cap-dependent translation independent of MTOR, and BAD mediated antagonism of apoptosis (Narlik-Grassow et al., 2014). Post-transcriptional regulation of PIM kinases is important, as proviral integrations in the *Pim1* 3' UTR are highly oncogenic (Nawijn et al., 2010). *Pim3* mRNA was abundant but highly labile in T cells, with a turnover rate in the top 2% of expressed mRNAs. PIM family members contain multiple ARE like repeats of AUUU(A), but the specific sequences responsible for rapid mRNA decay have not been described and cannot be predicted from the primary sequence alone. The *PIM3* 3'UTR contains two phylogenetically conserved regions with very similar predicted ARE sequences. Of these regions, we predicted that greater regulatory activity would be exerted by the region with GCLiPP evidence for RBP occupancy in both human and mouse cells. CRISPR dissection bore out this prediction in both species. The inactive conserved region may be structurally inaccessible to RBP occupancy, or it may be occupied and exert regulatory activity only in other cell types or signaling conditions.

Targeted dissection of GCLiPP peaks within 3'UTRs of immunologically relevant genes also led to discovery of cis-regulatory regions that modulate protein expression. Decreased expression of both CD5 and IKZF1 after deletion of the targeted regions suggests the presence of a post-transcriptional stabilizing or translational element. Lower levels of pSTAT6 similarly indicate stabilizing activity in *STAT6* 3'UTR. The mechanism by which these elements affect protein expression, and their role in regulating T cell biology is not yet well-defined. Changes in

CD5 and IKZF1 through the dissected regions could potentially contribute to changes in T cell activation and immune cell differentiation, respectively (Alexander et al., 1995; Heizmann et al., 2018; Voisinne et al., 2018). Phosphorylation of STAT6 in T cells and its important role in immune-associated allergy and asthma has been well-defined (Goenka & Kaplan, 2011). It will be interesting to investigate how the RBP binding region, and potentially the asthma-associated SNP, regulates protein activity and its effect on T cell biology in the context of type 2 immune responses. Together, these targeted dissections further highlight the utility of unbiased biochemical determination of RBP occupancy for annotating the regulatory transcriptome in conjunction with genetic data. Future studies will evaluate the effects of these regulatory regions on T cell biology and their mechanism.

Systematic comparison with eCLIP data for 87 individual RBPs (Sundararaman et al., 2016) indicated that GCLiPP roughly represented a weighted average of all potential eCLIP experiments for cytosolic RBPs. GCLiPP peaks overlapped eCLIP peaks at a frequency much greater than would be expected by chance, even though different cell types were used for the GCLiPP and eCLIP experiments. These findings are consistent with the prior observation that binding sites for individual proteins detected by eCLIP generally differ little between cell types with different tissue origin (van Nostrand et al., 2020). Nevertheless, the precise profiles of RBP occupancy and regulation of individual transcripts may be subject to cell type and context-dependent differences in RBP expression, binding activity, and site accessibility. Overall GCLiPP read density correlated with eCLIP read density in a manner that corresponded with the relative abundance of a given RBP in purified cellular mRNPs (Baltz et al., 2012). Still, the eCLIP peaks for some low abundance RBPs were significantly enriched in GCLiPP profiles. The strongest correlations were observed for abundant cytosolic RBPs, and the correspondence between eCLIP and GCLiPP was only apparent for cytosolic, but not non-cytosolic RBPs. This result was expected since the GCLiPP protocol selectively enriches for cytosolic polyadenylated RNA. GCLiPP could be modified to intentionally enrich for nuclear RBPs to examine the

regulatory landscape of mRNA biogenesis. Although GCLiPP cannot identify the proteins bound at specific sites in the transcriptome, we overcame this limitation by utilizing available RBP CLIP datasets to predict and map RBP-RNA interactions.

We leveraged the matched datasets from similar cell types to perform a cross species comparison of the post-transcriptional regulatory landscape. As might be expected, the sequences of 3' UTR regions that appeared as peaks of RBP occupancy in both species were in general more conserved than the full length 3' UTRs in which they occurred. These biochemically shared peaks were enriched in well-known RBP-binding cis-regulatory sequences including PUM motifs, CELF motifs and canonical polyadenylation signals. We also found clear biochemically shared peaks with relatively poor sequence conservation. These regions retain RBP occupancy despite an evident lack of strong selective pressure on their primary sequence, perhaps due to highly degenerate and/or structural determinants of RBP occupancy. RNAs with conserved structure and RBP binding but poorly conserved primary sequence have been reported before, and they are enriched in gene regulatory regions (Seemann et al., 2017; Weinreb et al., 2016). Finally, we noted that transcripts with biochemically shared peaks tended to encode proteins that were themselves involved in post-transcriptional gene regulation. This pattern is consistent with previous suggestions that auto-regulatory or multi-component feedback loops may be a conserved mode of post-transcriptional gene regulation (Kanitz & Gerber, 2010).

The GCLiPP datasets reported here provide a rich resource for the annotation and experimental dissection of cis-regulatory function in mRNAs. GCLiPP detected RBP occupancy at many known cis-regulatory regions, including canonical polyadenylation signals and elements that control mRNA localization, translation and stability. These data are provided to the scientific community for browsing and mining in a readily accessible form online. Combining GCLiPP with unbiased biochemical assays, genetic analyses and other future datasets probing RNA

regulatory circuits will yield a roadmap for the dissection of post-transcriptional regulatory networks and hypothesis generation of multi-omics studies.

Materials and Methods

Cells

Jurkat cells were grown in RPMI supplemented with fetal bovine serum (Omega). Primary $CD4^+$ and $CD8^+$ mouse T cells were isolated from C57BL/6J mouse peripheral lymph nodes and spleen using positive and negative selection Dynabeads, respectively, according to the manufacturer's instructions (Invitrogen). All mice were housed and bred in specific pathogen-free conditions in the Animal Barrier Facility at the University of California, San Francisco. Animal experiments were approved by the Institutional Animal Care and Use Committee of the University of California, San Francisco. Cells were stimulated with immobilized biotinylated anti-CD3 (clone 2C11, 0.25 μg/mL, BioXcell) and anti-CD28 (clone 37.51, 1 μg/mL, BioXcell) bound to Corning 10 cm cell culture dishes coated with Neutravidin (Thermo) at 10 μg/mL in PBS for 3 h at 37 °C. Cells were left on stimulation for 3 days before being transferred to non-coated dishes in T cell medium (Steiner et al., 2011) supplemented with recombinant human IL-2 (20 U/mL). Th2 cell cultures were also supplemented with murine IL-4 (100 U/mL) and anti-mouse IFN-γ (10 μg/mL). CD8 T cell cultures were also supplemented with 10 ng/mL recombinant murine IL-12 (10 ng/mL). For re-stimulation, cells were treated with 20 nM phorbol 12-myristate 13-acetate (PMA) and 1 μM ionomycin (Sigma) for 4 hours before harvest.

Measurement of mRNA Decay

Cells were stimulated with PMA and Ionomycin for 4 hours and then additionally treated with Actinomycin-D (Sigma-Aldrich) at 5 μg/mL for an additional 0, 1, 2 or 4 hours. After treatment, cells were lysed with Trizol LS (Life Technologies) and processed with Direct-zol ™

96 well RNA (Zymogen). RNA was quantified with an ND-1000 spectrophotometer (NanoDrop) and reverse transcribed with SuperScript III First Strand Synthesis Kit (Invitrogen). cDNA was quantified using TB Green Premix Ex Taq (Takara Bio).

GCLiPP and RNAseq

~100×10^6 mouse T cells cultured from 3 mice or ~100×10^6 Jurkat T cells were washed and resuspended in ice cold PBS and UV irradiated with a 254 nanometer UV crosslinker (Stratagene) in three doses of 4000 mJ, 2000 mJ and 2000 mJ, swirling on ice between doses. Cells were pelleted and frozen at -80 °C. Thawed pellets were rapidly resuspended in 400 µL PXL buffer without SDS (1X PBS with 0.5% deoxycholate, 0.5% NP-40, Protease inhibitor cocktail) supplemented with 2000 U RNasin (Promega) and 10 U DNase (Invitrogen). Pellets were incubated at 37 °C with shaking for 10 min, before pelleting of nuclei and cell debris (17000 g for 5 min). Supernatants were biotinylated by mixing at room temperature for 30 min with 500 µL of 10 mM EZ-Link NHS- SS-Biotin (Thermo) and 100 µL of 1 M sodium bicarbonate. Supernatants were mixed with 1 mg of washed oligo-dT beads (New England Biolabs) at room temperature for 30 min and washed 3 times with magnetic separation. Oligo-dT selected RNA was eluted from beads by heating in poly-A elution buffer (NEB) at 65 °C with vigorous shaking for 10 min. An aliquot of eluted RNA was treated with proteinase K and saved for RNAseq analysis using Illumina TruSeq Stranded Total RNA Library Prep Kit according to the manufacturer's instructions. Cells treated with Actinomycin-D as described above were also collected for RNAseq to generate transcriptome wide measurements of transcript stability.

The remaining crosslinked, biotinylated mRNA-RBP complexes were captured on 250 µL of washed M-280 Streptavidin Dynabeads (Invitrogen) for 30 min at 4 °C with continuous rotation to mix. Beads were washed 3 times with PBS and resuspended in 40 µL of PBS

containing 1000 U of RNase T1 (Thermo) for 1 min at room temperature. RNase activity was stopped by addition of concentrated (10% w/v) SDS to a final concentration of 1% SDS. Beads were washed successively in 1X PXL buffer, 5X PXL buffer and twice in PBS. 24 pmol of 3′ radiolabeled RNA linker was ligated to RBP bound RNA fragments by resuspending beads in 20 µL ligation buffer containing 10 U T4 RNA Ligase 1 (New England Biolabs) with 20% PEG 8000 at 37 degrees for 3 h. Beads were washed 3X with PBS and free 5′ RNA ends were phosphorylated with polynucleotide kinase (New England Biolabs). Beads were washed 3X with PBS and resuspended in ligation buffer containing 10 U T4 RNA Ligase 1, 50 pmol of 5′ RNA linker and 20% PEG 8000 and incubated at 15 °C overnight with intermittent mixing. Beads were again washed 3 times in PBS and linker ligated RBP binding fragments were eluted by treatment with proteinase K in 20 µL PBS with high speed shaking at 55 °C. Beads and supernatant were mixed 1:1 with bromophenol blue formamide RNA gel loading dye (Thermo) and loaded onto a 15% TBE-Urea denaturing polyacrylamide gel (BioRad). Ligated products with insert were visualized by autoradiography and compared to a control ligation (19 and 24 nt markers). Gel slices were crushed and soaked in gel diffusion buffer (0.5 M ammonium acetate; 10 mM magnesium acetate; 1 mM EDTA, pH 8.0; 0.1% SDS) at 37 °C for 30 min with high speed shaking, ethanol precipitated and resuspended in 20 µL of RNase free water. Ligated RNAs were reverse transcribed with Superscript III reverse transcriptase (Invitrogen) and amplified with Q5 polymerase (New England Biolabs). PCR was monitored using a real time PCR thermal cycler and amplification was discontinued when it ceased to amplify linearly. PCR products were run on a 10% TBE polyacrylamide gel, size selected for an amplicon with the predicted 20-50 bp insert size to exclude linker dimers and purified from the gel (Qiagen). Cleaned up library DNA was quantified on an Agilent 2100 Bioanalyzer using the High Sensitivity DNA Kit before being sequenced. All GCLiPP and RNAseq sequencing runs were carried out on an Illumina HiSeq 2500 sequencer.

GCLiPP and RNAseq bioinformatics analysis pipeline

FastQ files were de-multiplexed and trimmed of adapters. Each experiment was performed on three technical replicates per condition (resting and stimulated) per experiment. Cloning replicates and experiments were pooled in subsequent analyses. Jurkat and mouse T cell trimmed sequence reads were aligned to the hg38 human or mm10 mouse genome assembly using bowtie2, respectively. After alignment, PCR amplification artifacts were removed by de-duplication using the 2-nt random sequence at the 5′ end of the 3′ linker using a custom script that counted only a single read containing a unique linker sequence and start and end position of alignment per sequenced sample. Peaks of GCLiPP read density were called by convolving a normal distribution against a sliding window of the observed read distribution with a custom script (utr_peak_finder.pl). A 70 nucleotide window was analyzed centered on every nucleotide within the 3' UTR. For each window, the observed distribution of read density was compared to a normal distribution of the same magnitude as the nucleotide in the center of the window. The Pearson's correlation coefficient was computed for each nucleotide and peaks were defined as local maxima of goodness of fit between observed GCLiPP read density and the normal distribution, requiring a read depth above 20% of the maximum read depth in the 3' UTR global minimum of 10 reads. RNAseq reads were aligned using STAR Aligner (https://github.com/alexdobin/STAR) (Dobin et al., 2013) to align against the mm10 genome, and gene expression data were calculated as fragments per kilobase per million reads. Source code for data visualization software Thagomizer can be found at https://github.com/sskhon-2014/Graphy.

Comparison of GCLiPP to individual eCLIP datasets

eCLIP data (Sundararaman et al., 2016) from K562 cell line were downloaded via the ENCODE data portal (http://www.encodeproject.org/). The first replicate set of bigwig files were downloaded for each RBP deposited online at the time of analysis (December 2017) as well as

CLIPper called peaks for the same. To facilitate comparisons with GCLiPP we called GCLiPP peaks in the Jurkat data using CLIPper (Lovci et al., 2013) after re-aligning Jurkat GCLiPP reads to hg19. Correlation analysis was performed with a custom perl script that calculated the Spearman correlation for read depth at each nucleotide in the 3' UTR of all genes that were expressed in each dataset (as determined by CLIP read depth). ~5000-15000 expressed genes were included in the correlation analysis for each RBP. For comparison to mRNP abundancy, log10 RBP mass spectrometry spectra counts of HEK293 cells were utilized from (Baltz et al., 2012). To stratify RBPs by subcellular localization, data were taken from the COMPARTMENTS database, with RBPs with a localization score of 5 in the cytosol counted as cytosolic and lower counted as non-cytosolic (Binder et al., 2014). All custom scripts are available as STAR Methods Key Resource.

RBP Domain Analysis

For the following analysis, we called Jurkat GCLiPP peaks aligned to hg38 using CLIPper2.0 (Lovci et al., 2013). Each peak was resized to 200bp and oriented at the original peak center. The 200bp RNA sequence of each peak was analyzed using pf_fold method from ViennaRNA (RNAlib version 2.4.13) (Lorenz et al., 2011) to calculate base pairing probability for each pair of nucleotides and presented as an average for all the identified RBP binding sites. Matrix from Figure 2.2A is zoomed into central 150bp region. PTBP1 eCLIP dataset (hg38) from K562 cells was downloaded from ENCORE (GSM2424223) and processed in similar manner as described above.

We used available resting and activated Jurkat expression data (Ling Felce et al., 2021) (GSE145453) to calculate read counts mapped to RBP domains using annotations from RBPDBv1.3 (Cook et al., 2011) as a reference. Proteomics data of RBPs expressed in human Th0 cells was obtained and identified as described (Hoefig et al., 2021). RBPs that contained

more than one annotated domain based on RBPDBv1.3 were considered as an individual count in each appropriate category.

Conservation of RBP binding sites across various datasets and methods

To evaluate sequence conservation across various datasets, we performed CLIPper2.0 on our Jurkat data from GCLiPP, as well as sequencing data obtained through XRNAX (Trendel et al., 2019) and OOPS (Queiroz et al., 2019). The average PhyloP conservation score, obtained from UCSC genome browser as a bigwig of PhyloP scores of conservation 100 vertebrates, was calculated across all the sites within each method. This average was then standardized to contain a mean of 0 and a standard deviation of 1. Sequencing data for XRNAX (PRJEB26441; run accession ERR2537875) and OOPS (PRJEB26736; run accession SAMEA4663545, SAMEA4663546, SAMEA4663547, SAMEA4663548) was retrieved from EMBL-EBI ENA server and mapped to hg38 before CLIPper2.0 analysis. Specifically, our analysis used XRNAX data without ribosomal depletion and OOPS data performed using 150mJ/cm^2 crosslinking condition.

CRISPR editing

Guide RNA sequences were selected using the Benchling online CRISPR design tool (https://benchling.com/crispr) with guides selected to target genomic regions of GCLiPP read density. Synthetic crRNAs and tracrRNA (Dharmacon) were resuspended in water at 160 μM at 1:1 ratio and allowed to hybridize at 37 c for 30 m. For CRISPR dissection experiments, all crRNAs were mixed at an equimolar ratio before annealing to tracrRNA. This annealed gRNA complex (80 μM) was then mixed 1:1 by volume with 40 μM *S. pyogenes* Cas9-NLS (University of California Berkeley QB3 Macrolab) to a final concentration of 20 μM Cas9 ribonucleotide complex (RNP). This complexed gRNA:Cas9 RNP was mixed with a carrier solution of salmon sperm DNA (Invitrogen) and diluted to a final concentration between 5-20 μM. The diluted

gRNA:Cas9 RNPs were nucleofected into primary mouse T cells (24 hours after stimulation) with the P3 Primary Cell 96-well Nucleofector ™ Kit and into Jurkat cells with the SE Cell Line 96-well Nucleofector ™ Kit using a 4-D Nucleofector following the manufacturer's recommendations (Lonza). Cells were pipetted into pre-warmed media and then returned to CD3/CD28 stimulation for another two days and expanded an additional 3 days for primary mouse T cells or an additional 3-10 days in resting conditions for Jurkat cells.

3' UTR dissection

3' UTR dissection was performed as described (Zhao et al., 2017). Gene edited cells were harvested into Trizol reagent (Invitrogen) and total RNA was phase separated and purified from the aqueous phase using the Direct-zol™ RNA Miniprep Kit with on-column DNase treatment (Zymogen). Genomic DNA was extracted from the remaining organic phase by vigorous mixing with back extraction buffer (4 M guanidine thiocyanate, 50 mM sodium citrate, 1 M Tris base). cDNA was prepared with oligo-dT using the SuperScript III reverse transcription kit (Invitrogen). cDNA and genomic DNA were used as a template for PCR using MyTaq 2X Red Mix (Bioline). To equilibrate the number of target molecules and number of PCR cycles between samples, we performed semi-quantitative PCR followed by agarose gel electrophoresis to determine a PCR cycle number where genomic DNA first showed visible bands. This cycle number was then used with a titration of cDNA concentrations. A concentration that amplified equivalently was selected for analysis by deep sequencing. To quantify relative RNA/DNA ratios, cDNA and genomic DNA amplicons were purified using a QIAquick PCR purification up kit (Qiagen) and quantified on an Agilent 2100 Bioanalyzer using the High Sensitivity DNA Kit (Agilent).

Amplicons were tagmented with the Nextera XT kit (Illumina) and sequenced on an Illumina 2500 HiSeq. Reads were aligned to a custom genome consisting of the targeted PCR amplicon using STAR aligner and mutations were scored using an awk script

(https://github.com/alexdobin/STAR/blob/master/extras/scripts/sjFromSAMcollapseUandM.awk). RNA/DNA read ratios were calculated for all mutations over 20 nucleotides long and less than 250 nucleotides long, and relative expression was quantified as the median normalized RNA/DNA ratio for this subset of mutations. Mutations had to have at least 10 reads in both the RNA and gDNA amplicons and mutations with an RNA/DNA ratio of greater than 10 were excluded as outliers. Effect sizes for each nucleotide of the amplicon in each experiment were computed by comparing this median normalized RNA/DNA ratio for all mutations spanning a given nucleotide to all other mutations. Combined p-values were calculated using a Welch's two sample t-test comparing all mutations spanning a given nucleotide with all other mutations.

Shared peak calling, motif analysis and icSHAPE and Phylogenetic analyses

3′ UTR alignments of mouse and human were performed by downloading hg38 RefSeq 3' UTRs from UCSC genome browser, (http://genome.ucsc.edu), identifying syntenic regions of the mouse genome in mm10 with the KentUtils liftOver program (https://github.com/ucscGenomeBrowser/kent) and aligning UTRs with Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers et al., 2011). Alignments were programmatically performed for all human 3' UTRs with a custom perl script (get_alignment_from_fasta.pl). Biochemically shared peaks were called by the following algorithm (implemented in conserved_peak_finder.pl). This algorithm normalizes GCLiPP read density (i.e. the fraction of the maximal read depth within that 3' UTR) at each position, and calculates the correlation between mouse and human normalized signal. To favor regions with a clear local peak of GCLiPP read density, the algorithm further calculates the correlation between the observed data and a normal distribution centered at the point being examined in both the mouse and human data tracks. These three Spearman correlations were added together to calculate a numerical score, and shared peaks were defined as local maxima of these scores. To identify high stringency peaks, peaks were only accepted if they 1) had a

correlation of >0.75 between mouse and human, 2) had a peak that had a read density of >0.5 of the maximum read density within that 3' UTR in one data track (mouse or human) and >0.2 in the other and 3) had >10 reads at that location in both mouse and human datasets. Biological enrichment of genes with shared peaks was calculated using the Metascape (Tripathi et al., 2015) online interface (http://metascape.org) using the default settings, with the exception that a background set of genes was included in the analysis, specifically all genes that contain a called GCLiPP peak in both human and mouse datasets that do not contain a biochemically shared peak.

For motif calling, HOMER (Heinz et al., 2010) was used in RNA mode with the "noweight" option to turn off GC correction to search for motifs of width 5, 6 or 7 nucleotides, with otherwise default parameters. The positive sequence set was the mouse and human sequences of the biochemically shared GCLiPP peaks, the negative sequence set was all other GCLiPP called peaks from Jurkat and mouse T cells that were not shared across species. For icSHAPE we used a published bigwig file of locally normalized icSHAPE signal intensity generated in mouse ES cell (Spitale et al., 2015). Conservation of loci in the mouse and human genomes were obtained from the UCSC genome browser as a bigwig of PhyloP scores of conservation across 60 placental mammals (mouse) and 100 vertebrates (human) (http://hgdownload.cse.ucsc.edu/goldenpath/mm10/phyloP60way/, http://hgdownload.cse.ucsc.edu/goldenpath/hg38/phyloP100way/).

Mapping SNPs within GCLiPP peaks

We intersected our list of 3'UTR RBP peaks, determined using our peak calling algorithm, with a curated list of predicted disease causal SNPs (Taylor et al., 2021) to identify SNPs within predicted RBP binding regions. We limited our analysis to SNPs located in the 3'UTR of genes that contained at least 1 GCLiPP peak. Specific regions in the 3'UTR of *CD5*, *IKZF1* and *STAT6* were deleted in resting Jurkats using CRISPR-Cas9 RNPs as previously

mentioned. Protein expression of the edited genes was measured by flow cytometry 3-5 days after nucleofection. Briefly, cells were stained with Live/Dead efluor780 (Invitrogen) and anti-human CD5 (UCHT2) or IKZF1 (16B5C71). For pSTAT6 expression, Jurkat cells were treated with recombinant IL-4 (R&D, 12.5ng/mL) for 0, 5, 10, 15 or 30min, immediately fixed with 1.5% PFA for 10min and permeabilized with ice-cold methanol for 15min before staining with pSTAT6 (18/P-stat6) for 1 hour at room temperature. Cells were analyzed on LSRII and FACSAria cytometers. Graphpad Prism was used for data visualization and for Mann-Whitney two-tailed *t*-test.

Oligonucleotide and primer sequences

GCLiPP 3′ RNA linker: 5′-NNGUGUCUUUACACAGCUACGGCGUCG-3′

GCLiPP 5′ RNA linker: 5′-CGACCAGCAUCGACUCAGAAG-3′

GCLiPP Reverse transcription primer: 5′-CAAGCAGAAGACGGCATACGAGATNNNNNNCGCTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCCGACGCCGTAGCTGTGTAAA-3′ (NNNNNN is barcode for demultiplexing)

GCLiPP 3′ PCR primer: 5′-CAAGCAGAAGACGGCATACGAGAT-3′

GCLiPP 5′ PCR primer: 5′-AATGATACGGCGACCACCGAGATCTACACTGGTACTCCGACCAGCATCGACTCAGAAG-3′

Read1seq sequencing primer for GCLiPP: 5′-ACACTGGTACTCCGACCAGCATCGACTCAGAAG-3′Index sequencer primer for GCLiPP: 5′-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3′

PIM3 (human) gRNA1: TGTGCAGGCATCGCAGATGG

PIM3 (human) gRNA2: GACTTTGTACAGTCTGCTTG

PIM3 (human) gRNA3: GTGGCTAACTTAAGGGGAGT

PIM3 (human) gRNA4: AAACAATAAATAGCCCCGGT

PIM3 (human) gRNA5: TTGAGAAAACCAAGTCCCGC

PIM3 (human) gRNA6: CAGGAGGAGACGGCCCACGC

PIM3 (human) gRNA7: TTTATGGTGTGACCCCCTGG

PIM3 (human) gRNA8: CCAAGCCCCAGGGGACAGTG

Pim3 (mouse) gRNA1: GTTCAATTCTGGGAGAGCGC

Pim3 (mouse) gRNA2 CTGGTTCAAGTATCCACCCA

Pim3 (mouse) gRNA3: CCATAAATAAGAGACCGTGG

Pim3 (mouse) gRNA4: GCTTCCTCCCGCAAACACGG

Pim3 (mouse) gRNA5: CTGGTGTGACTAAGCATCAG

Pim3 (mouse) gRNA6: TGGAGAAGGTGGTTGCTTGG

CD5 gRNA1: GGAGCCTCGGGTCTGATCAA

CD5 gRNA2: GCTCTTCCAGACTTATTATG

IKZF1 R1 gRNA1: AAGGCTGACTTGTGTTCATG

IKZF1 R1 gRNA2: GCAACAAACTGACTCTAAGA

IKZF1 R2 gRNA1: TTATCATTGCATATCAGCAA

IKZF1 R2 gRNA2: ACATAATGCTTTTGGTGCGA

STAT6 gRNA1: GGGGTTAGCATATGTCAGAG

STAT6 gRNA2: CCAAATTCCTGTTAGCCAGG

Primers

PIM3 F (human): TCCAGCAGCGAGAGCTTGTGAGGAG

PIM3 R(human): TGATCTCCAGACATCTCACTTTTGAACTG

PIM3 R2(human):

TGAGATAGGTGCCTCACTGATTAAGCATTGGTGATCTCCAGACATCTCACTTTTGAACTG

Pim3 F (mouse): GCGTTCCAGAGAACTGTGACCTTCG

Pim3 R (mouse): TATGATCTTCAGACATTTCACACTTTTG

Figures

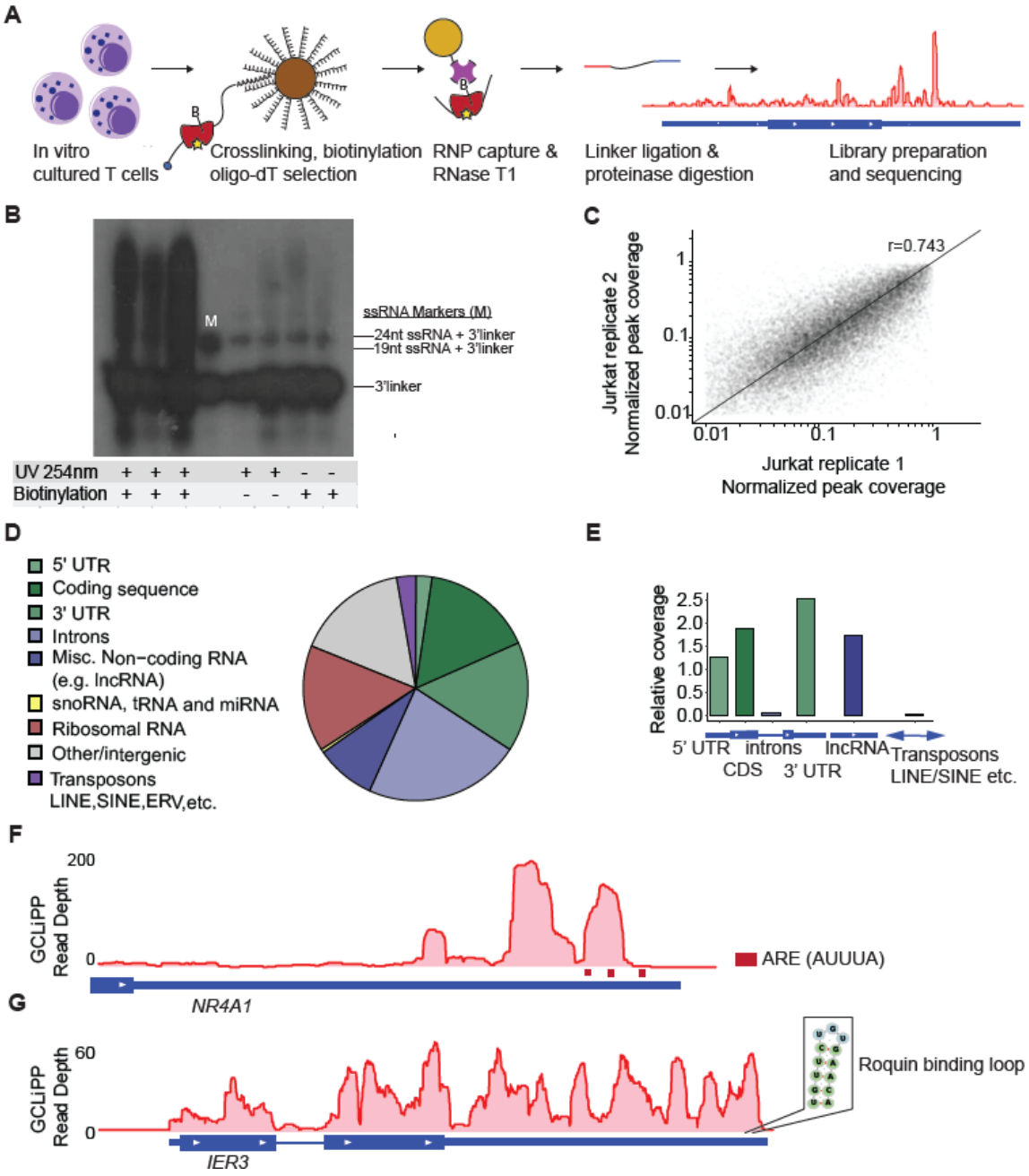


Figure 2.1. GCLiPP sequencing reveals RNA transcript protein occupancy.

(**A**) GCLiPP method of global RBP profiling. T cell RNAs are crosslinked to RBPs and lysates are biotinylated on primary amines. mRNAs are enriched with oligo-dT beads, and RBP protected sites are digested, captured, sequenced and aligned to the genome. **(B)** Film image of RBP-bound RNAs captured from Jurkats that underwent either UV crosslinking (UV 254nm), protein biotinylation or both. Lane marked “M” contains 19 and 24 nucleotide (nt) ssRNA ligated to radiolabeled 3’linker. RNA greater than 24nt+3’linker size were extracted and processed for

sequencing. **(C)** Normalized GCLiPP read depth (fraction of reads in called peak relative to all GCLiPP reads in annotated 3′ UTR) in two replicates of Jurkat cells. ρ represents Pearson correlation. **(D)** Proportion of mapped GCLiPP reads derived from genomic features. (**E**) Relative coverage of genomic features in GCLiPP sequencing reads relative to total length of genomic features of indicated class. **(F)** GCLiPP track of *NR4A1* focusing on 3'UTR. Red bars indicate presence of ARE motif (AUUUA). **(G)** GCLiPP track of *IER3* gene along with predicted roquin binding loop in the 3'UTR.

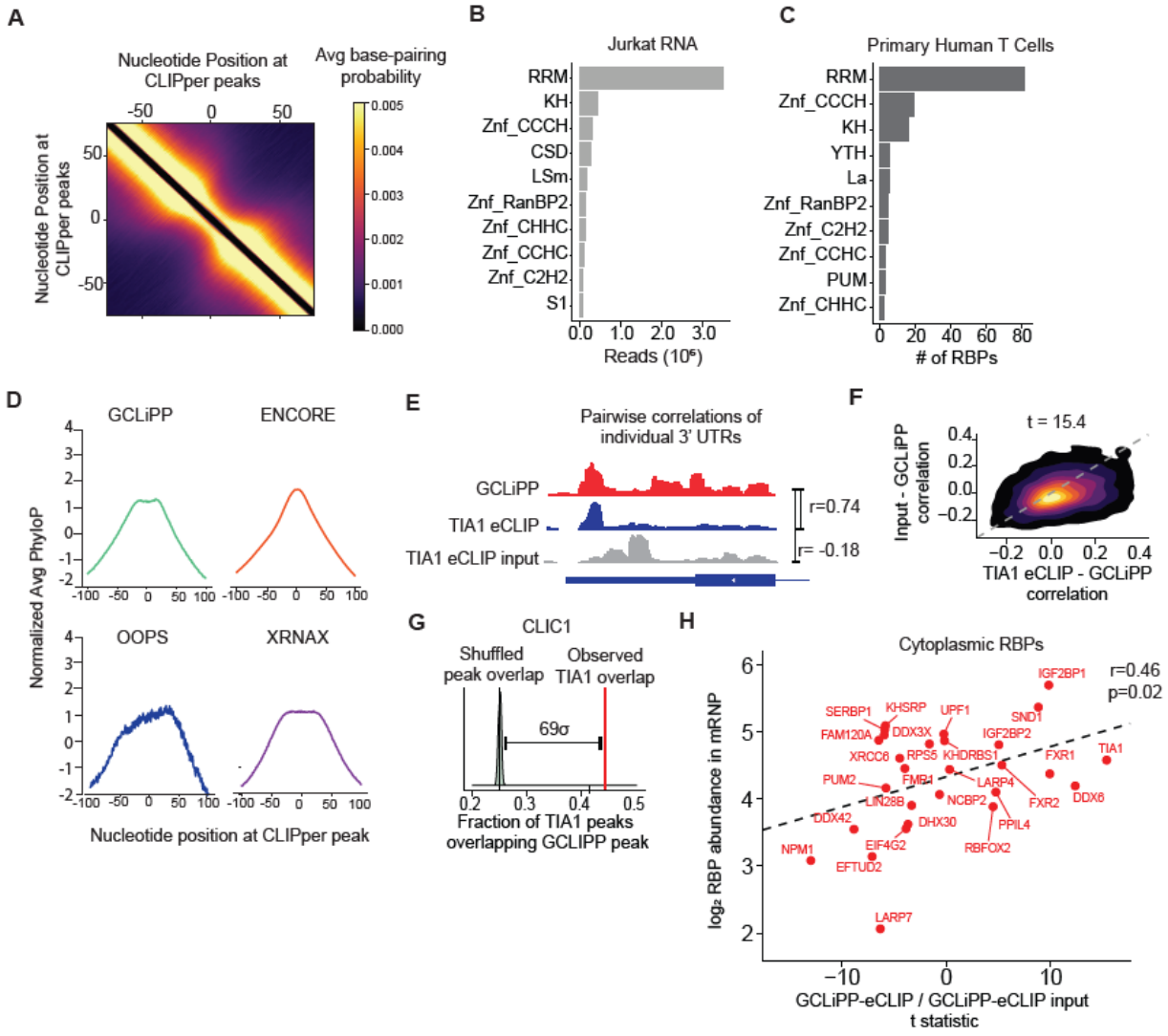


Figure 2.2. GCLiPP detects cytosolic RBP binding sites with characteristic sequence conservation and structural properties.

(A) Base-pairing probability was calculated for each pair of nucleotides within 200bp peak called by CLIPper2.0 in Jurkat cells. The average base-pairing matrices for all peaks in the 3'UTR is shown here as a heatmap. **(B)** Jurkat RNA-seq reads mapped to known RBPs were categorized into different RBDs. Top 10 occurring domains were determined by total reads that can be ascribed to specific domain motif. **(C)** Number of RBPs identified through RNA-IC in activated primary human T cells that contain certain domain. Only top 10 occurring motifs are shown.

RBPDB databased was used as a reference for categorizing RBPs in **(B)** and **(C). (D)** Sequence conservation of called peaks from various CLIP-Seq methods. Datasets used include RBP-bound RNAs detected by GCLiPP, phase-separation methods XRNAX and OOPS, as well as amalgamation of 87 RBP eCLIP data from ENCORE. Peaks in all data were identified using CLIPper2.0. **(E)** Genomic snapshots of individual 3' UTR showing exemplary correlation between TIA1 eCLIP dataset and GCLiPP. GCLiPP is shown in red, while the indicated RBP eCLIP data is shown in blue, and matched control input samples are shown in gray, shown for the 3' UTRs of the indicated gene. r indicates Pearson correlation between pairs of normalized read density at a given nucleotide for the indicated comparisons**. (F)** 2D density plots showing matched correlations between GCLiPP and TIA1 eCLIP (X-axis) and GCLiPP and the matched control input sample (Y-axis) for individual 3' UTR for all expressed genes in eCLIP and GCLiPP datasets. The t-statistic shown is for a paired t-test of the correlations. **(G)** Overlap of CLIPper called peaks in 3' UTRs in GCLiPP and TIA1 eCLIP. Red lines indicate observed overlap of GCLiPP peaks and eCLIP peaks. Grey distribution represents bootstrapped expected overlap, computed by shuffling called eCLIP peaks within the same 3' UTR and computing overlap of shuffled set with GCLiPP called peaks. This analysis was repeated 500 times. The indicated distance represents the number of standard deviations above the mean shuffled overlap of the observed overlap. **(H)** Correlation of eCLIP-GCLiPP paired t-tests from **(D)** and RBP abundance in mRNPs. RBPs in red are cytosolic localized (5 cytosolic according to COMPARTMENTS).

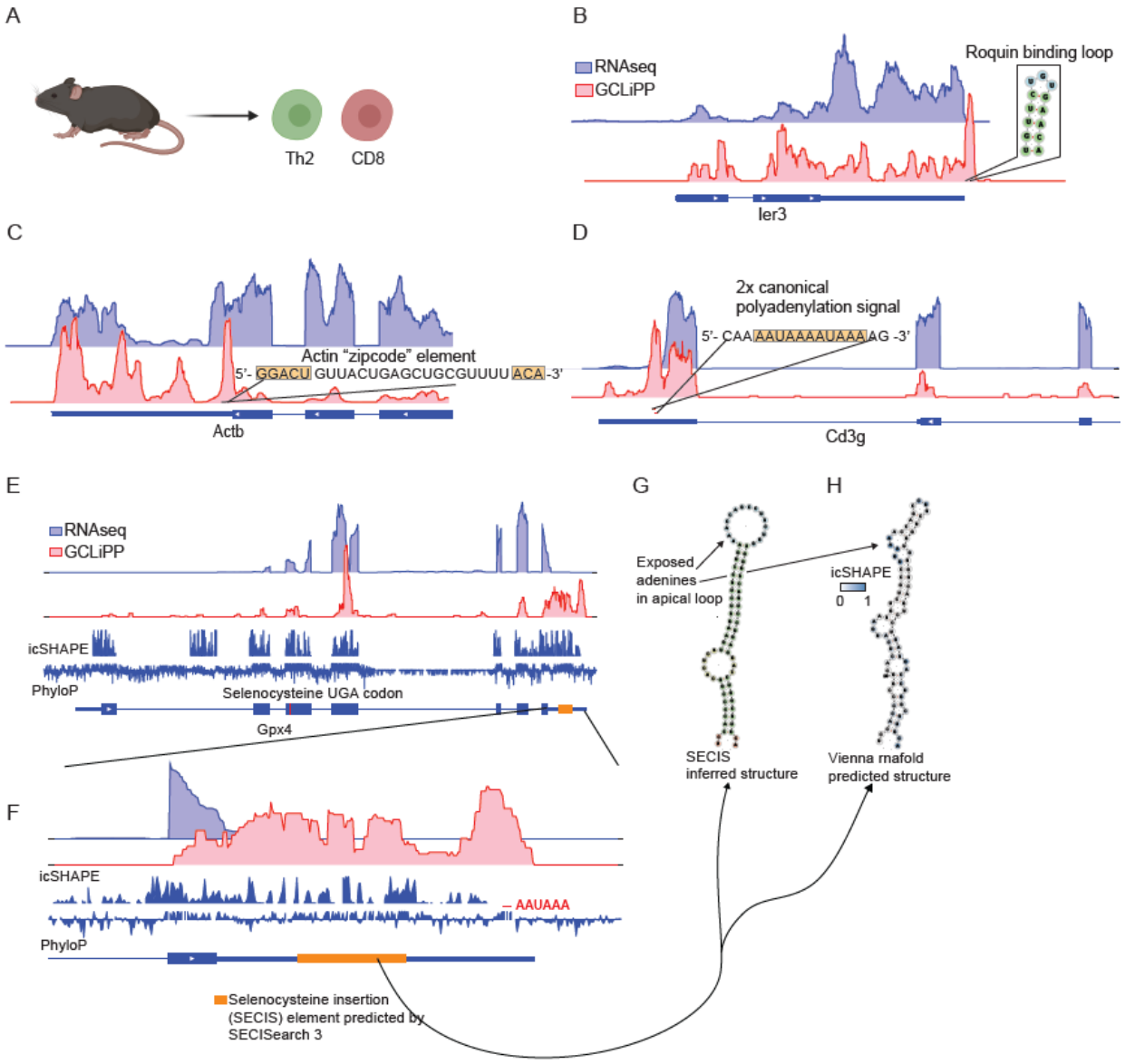


Figure 2.3. GCLiPP recapitulates previously described mRNA-RBP interactions in primary T cells.

(A) GCLiPP was performed on primary mouse Th2 and CD8 T cells as shown in the schematic. RNAseq and GCLiPP tracks for **(B)** *Ier3* (**C**) *Actb* **(D)** *Cd3g* and (**E-G**) *Gpx4*. RNAseq track is from resting Th2 cells. GCLiPP is sum of five experiments, three in Th2 and two in CD8 T cells. Locations of known RBP binding determinants are shown as insets.

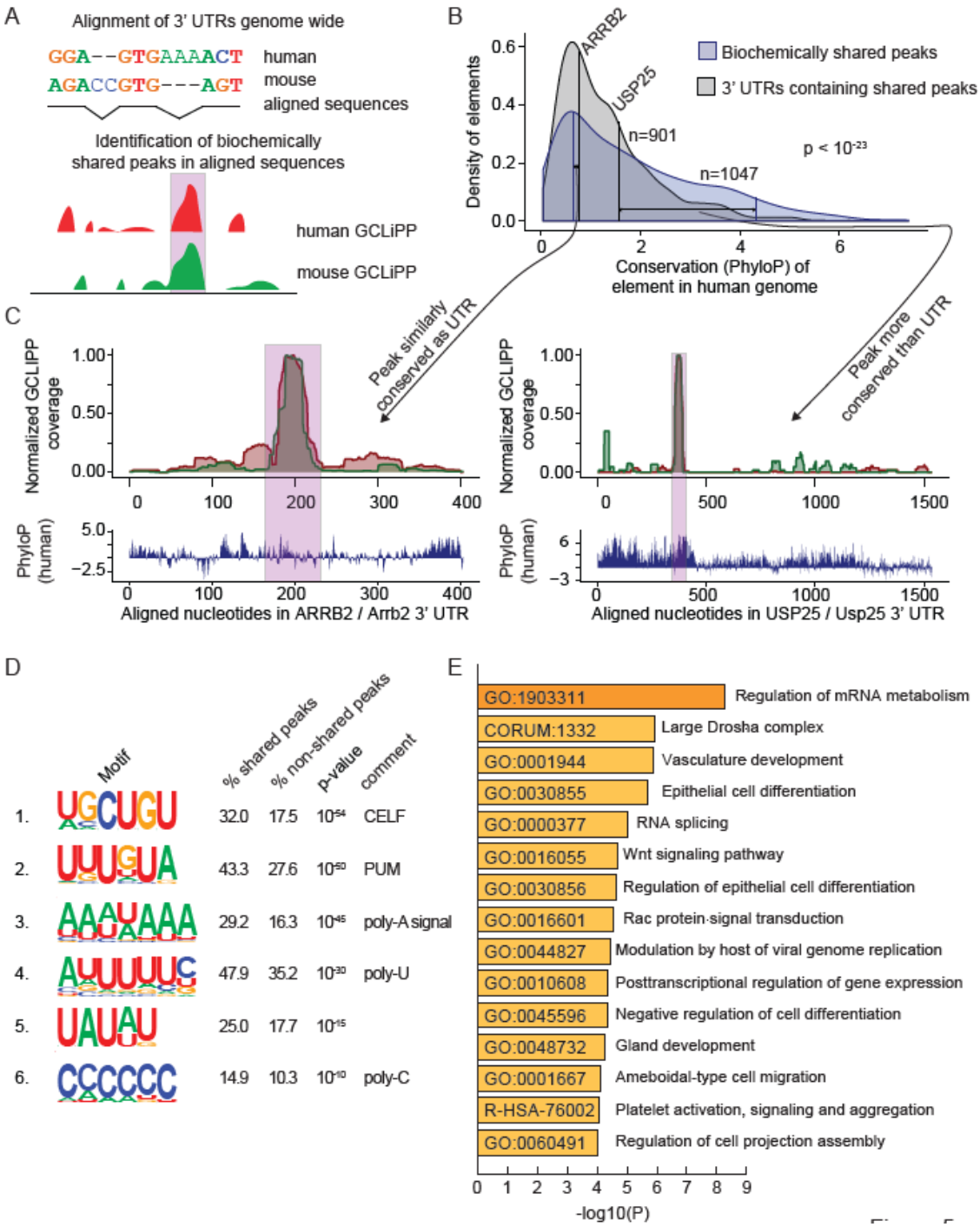


Figure 2.4. Comparison between mouse and human GCLiPP reveals principles of shared post-transcriptional regulation.

(A) Schematic illustration of 3' UTR alignment and biochemically shared GCLiPP peak calling. **(B)** Distribution of conservation across 100 vertebrates (PhyloP score) of regions in the human genome. Blue indicates biochemically shared peaks and gray indicates the 3' UTRs of the transcripts that those peaks are contained within. For both peaks within *ARRB2* and *USP25*, their matched conservation of peak and UTR are indicated by connected vertical lines. **(C)** Human and mouse normalized GCLiPP density and conservation (PhyloP) across aligned

nucleotides of the indicated 3’ UTRs. Biochemically shared peaks of GCLiPP read density are indicated in pink. **(D)** HOMER called motifs enriched in biochemically shared peaks. Percentages indicate the frequency of occurrence of the indicated motif in biochemically shared peaks versus non-shared background peaks. P-value indicates HOMER calculated p-value of enrichment. **(E)** Metascape called biological enrichment categories of genes containing biochemically shared peaks. The background set was all genes that contained peaks in both mouse and human GCLiPP datasets that did not contain a shared peak.

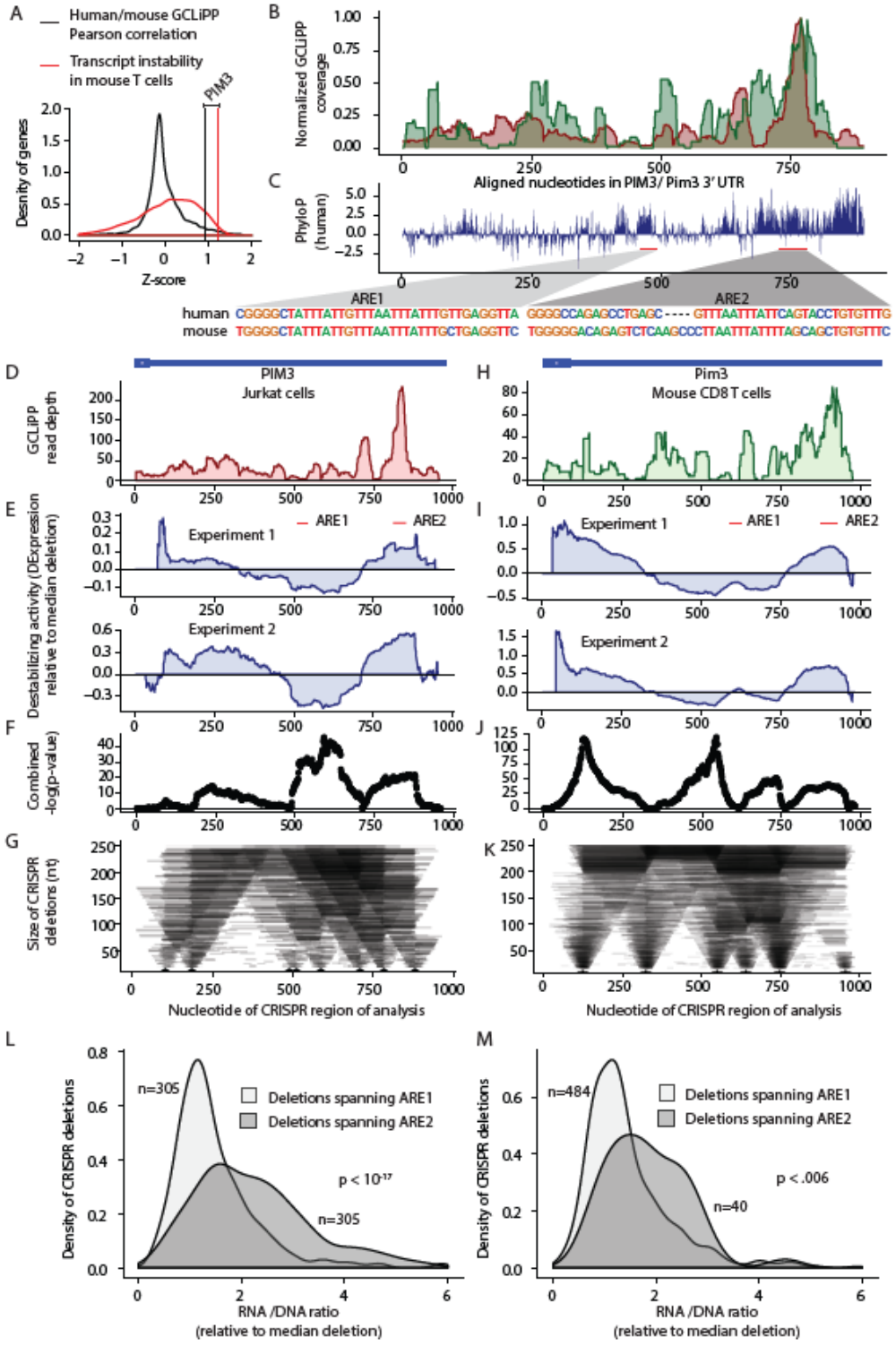


Figure 2.5. Biochemically and functionally shared post-transcriptional regulation of *PIM3* in human and mouse cells.

(A) Z-scores of Pearson correlation between mouse and human GCLiPP (black distribution) and transcript instability as measured by comparing transcript read abundance in untreated versus actinomycin-D treated mouse T cells (red distribution) for 7541 genes with matched data.

Vertical lines indicate observations for PIM3. **(B)** Normalized human and mouse GCLiPP read density and **(C)** PhyloP across aligned nucleotides of *PIM3* 3' UTR (as depicted in Figure 5). Insets show sequences of putative regulatory elements. **(D-G)** Dissection of human *PIM3* 3'UTR in Jurkat T cells. **(D)** GCLiPP peaks aligned to schematic illustration of 3'UTR. (**E**) Change in expression along the 3'UTR relative to median expression of all possible deletions. Per-nucleotide effect score was calculated by comparing median normalized RNA/gDNA ratio for all shown deletions spanning a given nucleotide with median of all shown deletions. Experiment 1 and 2 are biological duplicates which were transfected with 80μM or 120μM of gRNAs respectively. Red bars indicate putative ARE-containing cis regulatory elements. **(F)** Unadjusted -log10 p-values from Welch's two sample t-test comparing all deletions spanning a nucleotide with all other deletions across both experiments. **(G)** Size of deletions generated using CRISPR-Cas9. Arrow heads represent gRNA placement. **(H-K)** Dissection of mouse *Pim3* 3'UTR. Data are represented identically to human data, except that mouse primary CD8 T cells were used, and both mouse experiments 1 and 2 used a gRNA concentration of 80μM. **(L)** Effect of deletions spanning putative ARE containing cis-regulatory elements. The RNA/DNA ratio for mutants deleting ARE1 and ARE2 are shown in human Jurkat T cells. **(M)** Same as in **(L)** but using data from mouse primary T cells.

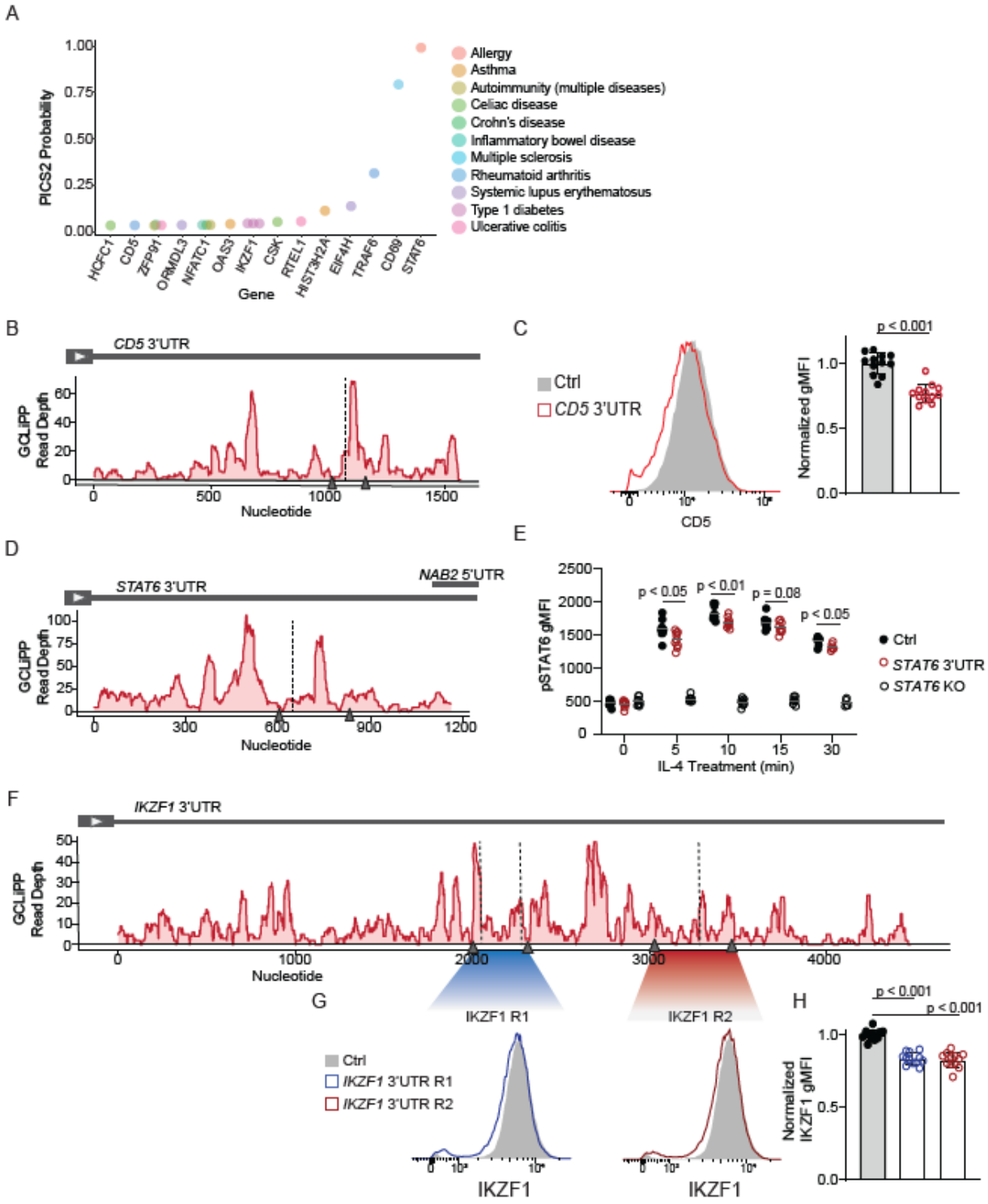
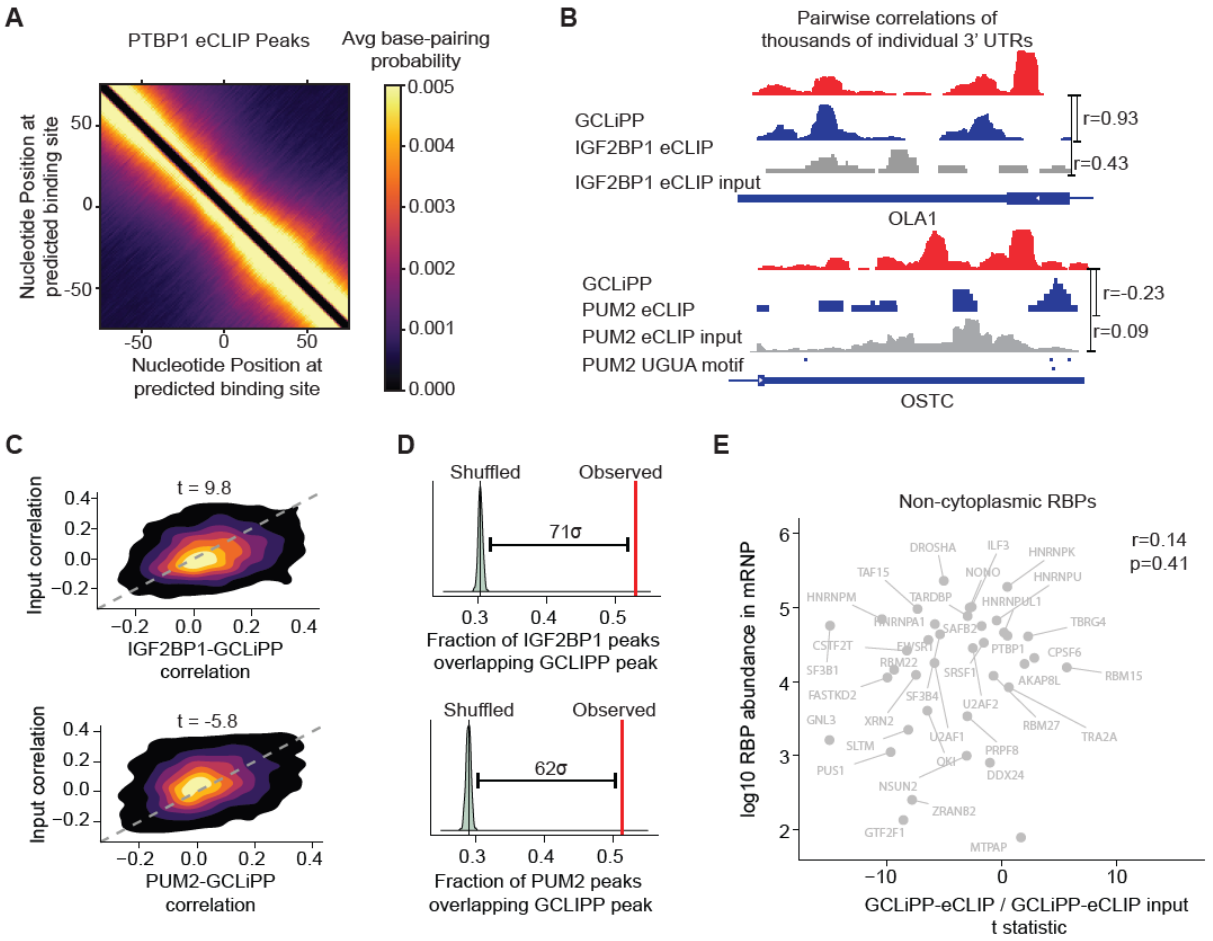


Figure 2.6. GCLiPP and PICS2 identified probable causal SNPs guide dissection of cis-regulatory elements in 3'UTR.

(A) Top 15 PICS2 SNPs within GCLiPP peaks with gene location (x-axis) and ranked by PICS2 probability score (y-axis). Various diseases associated with the SNPs are marked by color. **(B)** GCLiPP track of *CD5* 3'UTR in Jurkats. Arrow heads represent gRNA placement for deletion. Vertical dotted line indicates variant location **(C)** Representative flow plot of CD5 expression in

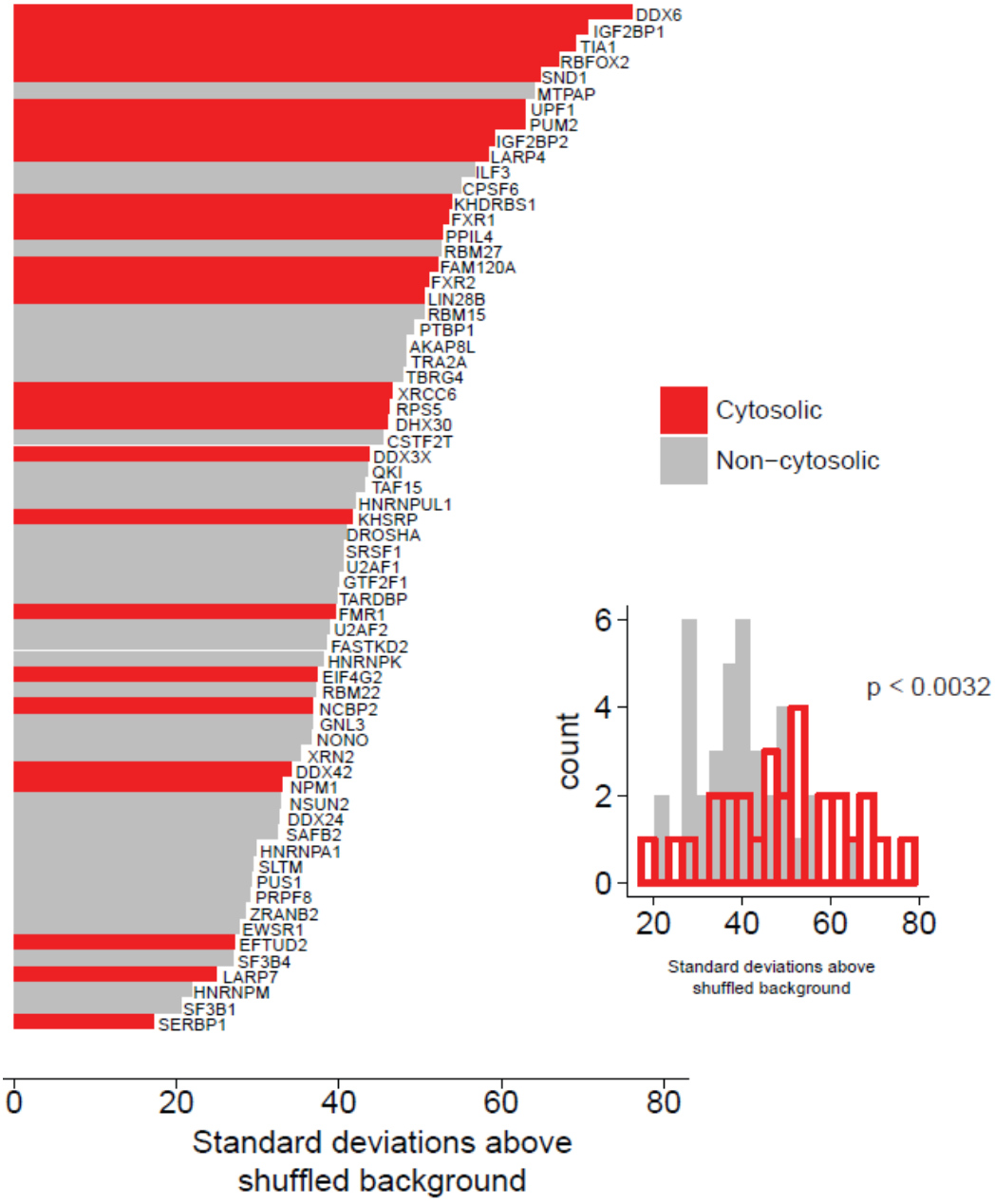
control (Ctrl) (shaded grey) and *CD5* 3'UTR edited Jurkats (red line) (left). Ctrl Jurkats were given control gRNAs and *CD5* 3'UTR Jurkats given paired gRNAs to generate deletion. Normalized gMFI is depicted in a bar graph (right). **(D)** GCLiPP track of *STAT6* 3'UTR in Jurkats, similar annotations as (B). **(E)** pSTAT6 gMFI for Ctrl, *STAT6* KO or *STAT6* 3'UTR-edited Jurkats after treatment with IL-4 for 0, 5, 10, 15 or 30 minutes. **(F)** CRISPR dissection of two regions in *IKZF1* 3'UTR, labeled as region 1 (R1) and region 2 (R2). Same annotations and GCLiPP track as seen in (B) and (D). **(G)** Representative flow plots of IKZF1 expression in Jurkats with *IKZF1* 3'UTR R1 (blue) or R2 deletion (red), as well as controls (grey). Normalized IKZF1 gMFI expression quantified in **(H)**.



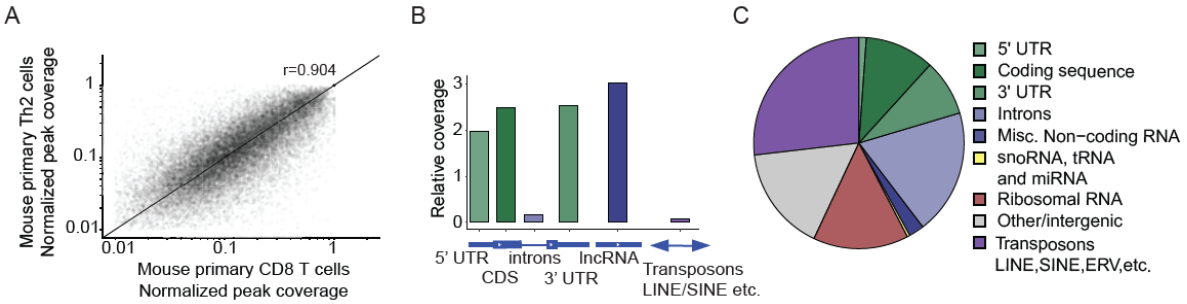
Supplementary Figure 2.1. Comparison GCLiPP with eCLIP datasets.

(A) Average base-pairing probability for all PTBP1 binding sites captured through eCLIP in K562 cells. Peaks were called using CLIPper and matrix is shown as a heatmap similar to Fig 2.2A. **(B)** Snapshots of individual 3'UTRs showing correlation between GCLiPP and IGF2BP1 (top) and PUM2 (bottom) eCLIP datasets. GCLiPP shown in red, eCLIP in blue and matched control input samples in gray. r indicates Pearson correlation between pairs of normalized read density at a given nucleotide for the indicated comparison. **(C)** 2D density plots showing matched correlations between GCLiPP and eCLIP for indicated RBP (X-axis), and GCLiPP and control input sample (Y-axis) for individual 3'UTRs for all expressed genes in GCLiPP and eCLIP datasets. Paired t-test was used for the correlations. **(D)** Overlap of CLIPper called peaks in 3'UTRs of GCLiPP and eCLIP for indicated RBP. Red lines indicated observed overlap of

GCLiPP peaks with eCLIP peaks. Green distribution represents bootstrapped expected overlap. Computed shuffling was conducted as described in Fig 2.2G. **(E)** Correlation of eCLIP-GCLiPP paired t-tests from (B) and RBP abundance in mRNPs not localized in the cytosol (<5 cytosolic according to COMPARTMENTS).

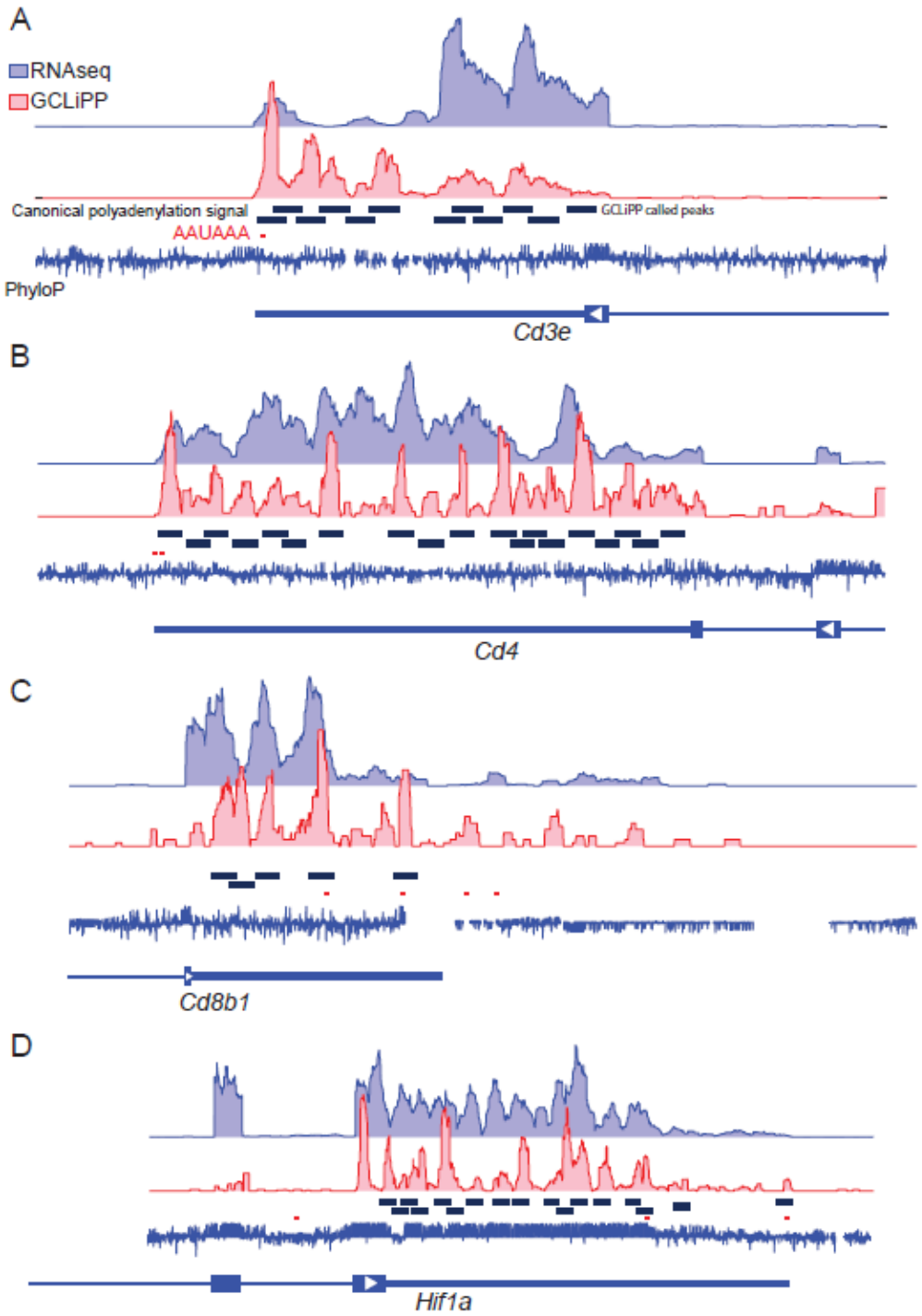


Supplementary Figure 2.2. Overlap of GCLiPP peaks and cytosolic RBP eCLIP peaks. Overlap of CLIPper called peaks in 3' UTRs in GCLiPP and eCLIP as in Fig 2.2G. The number of standard deviations between the observed fraction of eCLIP peaks that overlap with GCLiPP peaks above a background of 500 shuffled versions of the eCLIP peaks is depicted as the size of the bar. Peaks are shuffled across 3' UTRs containing eCLIP and GCLiPP peaks. Inset depicts a comparison of the scores of RBPs defined as cytosolic versus non-cytosolic as in Figure 2.2.



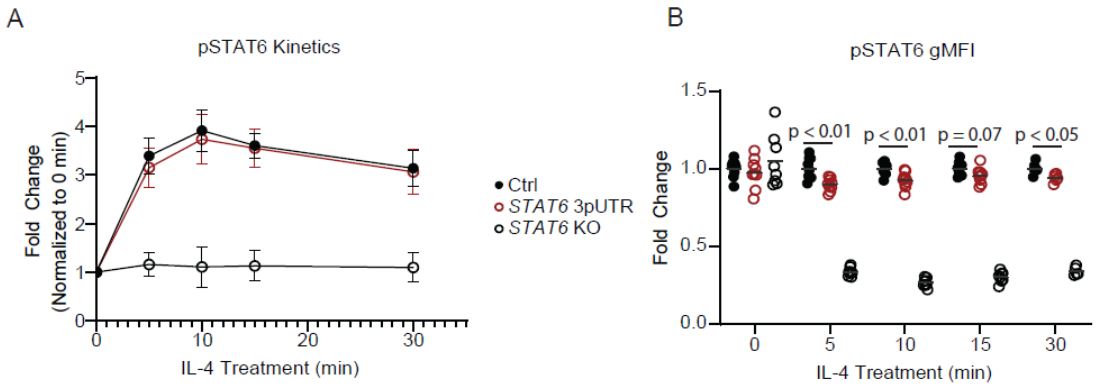
Supplementary Figure 2.3. GCLiPP read coverage in primary mouse T cells.

(A) Normalized GCLiPP read depth (fraction of reads in called peak relative to all GCLiPP reads in annotated 3'UTR) in mouse primary Th2 and CD8 T cells. r represents Pearson correlation. **(B)** Proportion of mapped GCLiPP reads derived from genomic features and **(C)** coverage of features relative to total length of genomics features of indicated class.



Supplementary Figure 2.4. GCLiPP detects RBP binding of canonical polyadenylation signal.

(A-D) RNAseq and GCLiPP read densities, conservation, called GCLiPP peaks and location of canonical polyadenylation signals (red lines) for indicated genes.



Supplementary Figure 2.5. STAT6 expression in 3'UTR edited Jurkat cells.

(A) Change in pSTAT6 gMFI for 0, 5, 10, 15 and 30min treatment with IL-4 relative to 0min for control, *STAT6* 3'UTR-edited and STAT6 knockout (KO) Jurkat cells. **(B)** Change in pSTAT6 gMFI normalized to control within individual timepoints for all groups.

References

Agarwal, V., Bell, G. W., Nam, J. W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *ELife*, *4*(AUGUST2015). https://doi.org/10.7554/ELIFE.05005

Alexander, T., Kanner, S. B., Joachim, H., Ledbetter, J. A., Werner, M., Nigel, K., & Klaus, R. (1995). A Role for CD5 in TCR-Mediated Signal Transduction and Thymocyte Selection. *Science*, *269*(5223), 535–537. https://doi.org/10.1126/science.7542801

Baltz, A. G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C., & Landthaler, M. (2012). The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Molecular Cell*, *46*(5), 674–690. https://doi.org/10.1016/J.MOLCEL.2012.05.021

Bassell, G. J., & Kelic, S. (2004). Binding proteins for mRNA localization and local translation, and their dysfunction in genetic neurological disease. *Current Opinion in Neurobiology*, *14*(5), 574–581. https://doi.org/10.1016/J.CONB.2004.08.010

Berry, M. J., Banu, L., Harney, J. W., & Larsen, P. R. (1993). Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons. *The EMBO Journal*, *12*(8), 3315–3322. https://doi.org/10.1002/J.1460-2075.1993.TB06001.X

Binder, J. X., Pletscher-Frankild, S., Tsafou, K., Stolte, C., O'Donoghue, S. I., Schneider, R., & Jensen, L. J. (2014). COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database*, *2014*. https://doi.org/10.1093/DATABASE/BAU012

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*(12), 1213–1218. https://doi.org/10.1038/nmeth.2688

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsveld, J., & Hentze, M. W. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell*, *149*(6), 1393–1406. https://doi.org/10.1016/J.CELL.2012.04.031/ATTACHMENT/3EF9B7DB-485C-4DA2-A14E-2C67F71CC55A/MMC7.XLS

Chao, J. A., Patskovsky, Y., Patel, V., Levy, M., Almo, S. C., & Singer, R. H. (2010). ZBP1 recognition of β-actin zipcode induces RNA looping. *Genes & Development*, *24*(2), 148–158. https://doi.org/10.1101/GAD.1862910

Chen, C.-Y. A., & Shyu, A.-B. (1994). Selective degradation of early-response-gene mRNAs: functional analyses of sequence features of the AU-rich elements. *Molecular and Cellular Biology*, *14*(12), 8471–8482. https://doi.org/10.1128/MCB.14.12.8471-8482.1994

Cook, K. B., Kazan, H., Zuberi, K., Morris, Q., & Hughes, T. R. (2011). RBPDB: a database of RNA-binding specificities. *Nucleic Acids Research*, *39*(suppl_1), D301–D308. https://doi.org/10.1093/nar/gkq1069

Corces, M. R., Buenrostro, J. D., Wu, B., Greenside, P. G., Chan, S. M., Koenig, J. L., Snyder, M. P., Pritchard, J. K., Kundaje, A., Greenleaf, W. J., Majeti, R., & Chang, H. Y. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nature Genetics 2016 48:10*, *48*(10), 1193–1203. https://doi.org/10.1038/ng.3646

Corley, M., Burns, M. C., & Yeo, G. W. (2020). How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Molecular Cell*, *78*(1), 9–29. https://doi.org/10.1016/J.MOLCEL.2020.03.011

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Farh, K. K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., Shoresh, N., Whitton, H., Ryan, R. J. H., Shishkin, A. A., Hatan, M., Carrasco-Alfonso, M. J., Mayer, D., Luckey, C. J., Patsopoulos, N. A., de Jager, P. L., Kuchroo, V. K., Epstein, C. B., Daly, M. J., … Bernstein, B. E. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, *518*(7539), 337–343. https://doi.org/10.1038/nature13835

Fong, A. M., Premont, R. T., Richardson, R. M., Yu, Y. R. A., Lefkowitz, R. J., & Patel, D. D. (2002). Defective lymphocyte chemotaxis in β-arrestin2- and GRK6-deficient mice. *Proceedings of the National Academy of Sciences*, *99*(11), 7478–7483. https://doi.org/10.1073/PNAS.112198299

Freeberg, M. A., Han, T., Moresco, J. J., Kong, A., Yang, Y. C., Lu, Z. J., Yates, J. R., & Kim, J. K. (2013). Pervasive and dynamic protein binding sites of the mRNA transcriptome in Saccharomyces cerevisiae. *Genome Biology*, *14*(2), 1–20. https://doi.org/10.1186/GB-2013-14-2-R13/FIGURES/7

Gagnon, J. D., Kageyama, R., Shehata, H. M., Fassett, M. S., Mar, D. J., Wigton, E. J., Johansson, K., Litterman, A. J., Odorizzi, P., Simeonov, D., Laidlaw, B. J., Panduro, M., Patel, S., Jeker, L. T., Feeney, M. E., McManus, M. T., Marson, A., Matloubian, M., Sanjabi, S., & Ansel, K. M. (2019). miR-15/16 Restrain Memory T Cell Differentiation, Cell Cycle, and Survival. *Cell Reports*, *28*(8), 2169-2181.e4. https://doi.org/10.1016/J.CELREP.2019.07.064

Garneau, N. L., Wilusz, J., & Wilusz, C. J. (2007). The highways and byways of mRNA decay. *Nature Reviews Molecular Cell Biology 2007 8:2*, *8*(2), 113–126. https://doi.org/10.1038/nrm2104

Gebauer, F., Schwarzl, T., Valcárcel, J., & Hentze, M. W. (2021). RNA-binding proteins in human genetic disease. *Nature Reviews Genetics*, *22*(3), 185–198. https://doi.org/10.1038/s41576-020-00302-y

Goenka, S., & Kaplan, M. H. (2011). Transcriptional regulation by STAT6. *Immunologic Research*, *50*(1), 87. https://doi.org/10.1007/s12026-011-8205-2

Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D., Ule, J., & Zavolan, M. (2021). CLIP and complementary methods. *Nature Reviews Methods Primers*, *1*(1). https://doi.org/10.1038/S43586-021-00018-1

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A. C., Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M., & Tuschl, T. (2010). Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell*, *141*(1), 129–141. https://doi.org/10.1016/J.CELL.2010.03.009/ATTACHMENT/2F68D9DB-BCEB-4222-9358-02C18AD54FFC/MMC7.XLS

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, *38*(4), 576–589. https://doi.org/10.1016/J.MOLCEL.2010.05.004/ATTACHMENT/AC76CFA3-57BD-4789-B1BD-E1F2EE8F5776/MMC1.PDF

Heizmann, B., Kastner, P., & Chan, S. (2018). The Ikaros family in lymphocyte development. *Current Opinion in Immunology*, *51*, 14–23. https://doi.org/10.1016/J.COI.2017.11.005

Hoefig, K. P., Reim, A., Gallus, C., Wong, E. H., Behrens, G., Conrad, C., Xu, M., Kifinger, L., Ito-Kureha, T., Defourny, K. A. Y., Geerlof, A., Mautner, J., Hauck, S. M., Baumjohann, D., Feederle, R., Mann, M., Wierer, M., Glasmacher, E., & Heissmeyer, V. (2021). Defining the RBPome of primary T helper cells to elucidate higher-order Roquin-mediated mRNA regulation. *Nature Communications*, *12*(1), 5208. https://doi.org/10.1038/s41467-021-25345-5

Johansson, L., Gafvelin, G., & Arnér, E. S. J. (2005). Selenocysteine in proteins—properties and biotechnological use. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1726*(1), 1–13. https://doi.org/10.1016/J.BBAGEN.2005.05.010

Kafasla, P., Skliris, A., & Kontoyiannis, D. L. (2014). Post-transcriptional coordination of immunological responses by RNA-binding proteins. *Nature Immunology 2014 15:6*, *15*(6), 492–502. https://doi.org/10.1038/ni.2884

Kanitz, A., & Gerber, A. P. (2010). Circuitry of mRNA regulation. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *2*(2), 245–251. https://doi.org/10.1002/WSBM.55

Keene, J. D. (2007). RNA regulons: coordination of post-transcriptional events. *Nature Reviews Genetics 2007 8:7*, *8*(7), 533–543. https://doi.org/10.1038/nrg2111

Kislauskis, E. H., Zhu, X., & Singer, R. H. (1994). Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. *Journal of Cell Biology*, *127*(2), 441–451. https://doi.org/10.1083/JCB.127.2.441

Leppek, K., Schott, J., Reitter, S., Poetz, F., Hammond, M. C., & Stoecklin, G. (2013). Roquin Promotes Constitutive mRNA Decay via a Conserved Class of Stem-Loop Recognition Motifs. *Cell*, *153*(4), 869–881. https://doi.org/10.1016/j.cell.2013.04.016

Ling Felce, S., Farnie, G., Dustin, M. L., Felce, J. H., Dobrovinskaya, O., McComb, S., & Canada, C. (2021). RNA-Seq analysis of early transcriptional responses to activation in the leukaemic Jurkat E6.1 T cell line. *Wellcome Open Research 2021 5:42*, *5*, 42. https://doi.org/10.12688/wellcomeopenres.15748.2

Loeb, G. B., Khan, A. A., Canner, D., Hiatt, J. B., Shendure, J., Darnell, R. B., Leslie, C. S., & Rudensky, A. Y. (2012). Transcriptome-wide miR-155 Binding Map Reveals Widespread Noncanonical MicroRNA Targeting. *Molecular Cell*, *48*(5), 760–770. https://doi.org/10.1016/J.MOLCEL.2012.10.002/ATTACHMENT/71E28334-BB00-454E-BC52-6B57157CE7AB/MMC4.PDF

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, *6*(1), 1–14. https://doi.org/10.1186/1748-7188-6-26/TABLES/2

Lovci, M. T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T. Y., Stark, T. J., Gehman, L. T., Hoon, S., Massirer, K. B., Pratt, G. A., Black, D. L., Gray, J. W., Conboy, J. G., & Yeo, G. W. (2013). Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nature Structural & Molecular Biology*, *20*(12), 1434–1442. https://doi.org/10.1038/nsmb.2699

Makeyev, A. v., & Liebhaber, S. A. (2002). The poly(C)-binding proteins: A multiplicity of functions and a search for mechanisms. *RNA*, *8*(3), 265–278. https://doi.org/10.1017/S1355838202024627

Mariotti, M., Lobanov, A. v, Guigo, R., & Gladyshev, V. N. (2013). SECISearch3 and Seblastian: new tools for prediction of SECIS elements and selenoproteins. *Nucleic Acids Research*, *41*(15), e149–e149. https://doi.org/10.1093/nar/gkt550

Martin, K. C., & Ephrussi, A. (2009). mRNA Localization: Gene Expression in the Spatial Dimension. *Cell*, *136*(4), 719–730. https://doi.org/10.1016/J.CELL.2009.01.044

Millevoi, S., & Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3′ end processing regulation. *Nucleic Acids Research*, *38*(9), 2757–2774. https://doi.org/10.1093/NAR/GKP1176

Narlik-Grassow, M., Blanco-Aparicio, C., & Carnero, A. (2014). The PIM Family of Serine/Threonine Kinases in Cancer. *Medicinal Research Reviews*, *34*(1), 136–159. https://doi.org/10.1002/MED.21284

Nawijn, M. C., Alendar, A., & Berns, A. (2010). For better or for worse: the role of Pim oncogenes in tumorigenesis. *Nature Reviews Cancer 2011 11:1*, *11*(1), 23–34. https://doi.org/10.1038/nrc2986

Nicolet, B. P., Zandhuis, N. D., Lattanzio, V. M., & Wolkers, M. C. (2021). Sequence determinants as key regulators in gene expression of T cells. *Immunological Reviews*, *n/a*(n/a). https://doi.org/https://doi.org/10.1111/imr.13021

Oberstrass, F. C., Auweter, S. D., Michèle, E., Yann, H., Anke, H., Philipp, W., Luc, R., Batoul, A.-A., Stefan, P., Black, D. L., & Allain, F. H. T. (2005). Structure of PTB Bound to RNA: Specific Binding and Implications for Splicing Regulation. *Science*, *309*(5743), 2054–2057. https://doi.org/10.1126/science.1114066

Perez-Perri, J. I., Noerenberg, M., Kamel, W., Lenz, C. E., Mohammed, S., Hentze, M. W., & Castello, A. (2021). Global analysis of RNA-binding protein dynamics by comparative and enhanced RNA interactome capture. *Nature Protocols*, *16*(1), 27–60. https://doi.org/10.1038/s41596-020-00404-1

Perez-Perri, J. I., Rogell, B., Schwarzl, T., Stein, F., Zhou, Y., Rettel, M., Brosig, A., & Hentze, M. W. (2018). Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nature Communications*, *9*(1), 4408. https://doi.org/10.1038/s41467-018-06557-8

Proudfoot, N. J. (2011). Ending the message: poly(A) signals then and now. *Genes & Development*, *25*(17), 1770–1782. https://doi.org/10.1101/GAD.17268411

Queiroz, R. M. L., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.-M., Ramakrishna, M., Harvey, R. F., Dezi, V., Thomas, G. H., Willis, A. E., & Lilley, K. S. (2019). Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nature Biotechnology 2019 37:2*, *37*(2), 169–178. https://doi.org/10.1038/s41587-018-0001-2

Raghavan, A., Ogilvie, R. L., Reilly, C., Abelson, M. L., Raghavan, S., Vasdewani, J., Krathwohl, M., & Bohjanen, P. R. (2002). Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. *Nucleic Acids Research*, *30*(24), 5529–5538. https://doi.org/10.1093/NAR/GKF682

Reed, R. (2003). Coupling transcription, splicing and mRNA export. *Current Opinion in Cell Biology*, *15*(3), 326–331. https://doi.org/10.1016/S0955-0674(03)00048-6

Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., & Burge, C. B. (2008). Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science*, *320*(5883), 1643–1647. https://doi.org/10.1126/SCIENCE.1155390/SUPPL_FILE/SANDBERG.SOM.PDF

Schwerk, J., & Savan, R. (2015). Translating the Untranslated Region. *The Journal of Immunology*, *195*(7), 2963–2971. https://doi.org/10.4049/JIMMUNOL.1500756

Seemann, S. E., Mirza, A. H., Hansen, C., Bang-Berthelsen, C. H., Garde, C., Christensen-Dalsgaard, M., Torarinsson, E., Yao, Z., Workman, C. T., Pociot, F., Nielsen, H., Tommerup, N., Ruzzo, W. L., & Gorodkin, J. (2017). The identification and functional annotation of RNA structures conserved in vertebrates. *Genome Research*, *27*(8), 1371–1383. https://doi.org/10.1101/GR.208652.116

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1), 539. https://doi.org/10.1038/MSB.2011.75

Simeonov, D. R., Gowen, B. G., Boontanrart, M., Roth, T. L., Gagnon, J. D., Mumbach, M. R., Satpathy, A. T., Lee, Y., Bray, N. L., Chan, A. Y., Lituiev, D. S., Nguyen, M. L., Gate, R. E., Subramaniam, M., Li, Z., Woo, J. M., Mitros, T., Ray, G. J., Curie, G. L., … Marson, A. (2017). Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature*, *549*(7670), 111–115. https://doi.org/10.1038/nature23875

Smith, T., Villanueva, E., Queiroz, R. M. L., Dawson, C. S., Elzek, M., Urdaneta, E. C., Willis, A. E., Beckmann, B. M., Krijgsveld, J., & Lilley, K. S. (2020). Organic phase separation opens up new opportunities to interrogate the RNA-binding proteome. *Current Opinion in Chemical Biology*, *54*, 70–75. https://doi.org/https://doi.org/10.1016/j.cbpa.2020.01.009

Spitale, R. C., Flynn, R. A., Zhang, Q. C., Crisalli, P., Lee, B., Jung, J. W., Kuchelmeister, H. Y., Batista, P. J., Torre, E. A., Kool, E. T., & Chang, H. Y. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. *Nature 2015 519:7544*, *519*(7544), 486–490. https://doi.org/10.1038/nature14263

Steiner, D. F., Thomas, M. F., Hu, J. K., Yang, Z., Babiarz, J. E., Allen, C. D. C., Matloubian, M., Blelloch, R., & Ansel, K. M. (2011). MicroRNA-29 Regulates T-Box Transcription Factors and Interferon-γ Production in Helper T Cells. *Immunity*, *35*(2), 169–181. https://doi.org/10.1016/J.IMMUNI.2011.07.009/ATTACHMENT/284A1DFD-A340-4162-9F4A-FA38A1CEC8FA/MMC4.ZIP

Steri, M., Idda, M. L., Whalen, M. B., & Orrù, V. (2018). Genetic variants in mRNA untranslated regions. *WIREs RNA*, *9*(4), e1474. https://doi.org/https://doi.org/10.1002/wrna.1474

Sundararaman, B., Zhan, L., Blue, S. M., Stanton, R., Elkins, K., Olson, S., Wei, X., van Nostrand, E. L., Pratt, G. A., Huelga, S. C., Smalec, B. M., Wang, X., Hong, E. L., Davidson, J. M., Lécuyer, E., Graveley, B. R., & Yeo, G. W. (2016). Resources for the Comprehensive Discovery of Functional RNA Elements. *Molecular Cell*, *61*(6), 903–913. https://doi.org/10.1016/J.MOLCEL.2016.02.012/ATTACHMENT/2E756C08-DBB9-46F3-8303-252DEF80229D/MMC7.XLSX

Taylor, K. E., Ansel, K. M., Marson, A., Criswell, L. A., & Farh, K. K.-H. (2021). PICS2: next-generation fine mapping via probabilistic identification of causal SNPs. *Bioinformatics*. https://doi.org/10.1093/BIOINFORMATICS/BTAB122

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K., Diegel, M., Dunn, D., Ebersol, A. K., … Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature 2012 489:7414*, *489*(7414), 75–82. https://doi.org/10.1038/nature11232

Timchenko, L. T., Miller, J. W., Timchenko, N. A., Devore, D. R., Datar, K. v., Lin, L., Roberts, R., Thomas Caskey, C., & Swanson, M. S. (1996). Identification of a (CUG) n Triplet Repeat RNA-Binding Protein and Its Expression in Myotonic Dystrophy. *Nucleic Acids Research*, *24*(22), 4407–4414. https://doi.org/10.1093/NAR/24.22.4407

Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M. W., & Krijgsveld, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell*, *176*(1), 391-403.e19. https://doi.org/https://doi.org/10.1016/j.cell.2018.11.004

Tripathi, S., Pohl, M. O., Zhou, Y., Rodriguez-Frandsen, A., Wang, G., Stein, D. A., Moulton, H. M., Dejesus, P., Che, J., Mulder, L. C. F., Yángüez, E., Andenmatten, D., Pache, L., Manicassamy, B., Albrecht, R. A., Gonzalez, M. G., Nguyen, Q., Brass, A., Elledge, S., … Chanda, S. K. (2015). Meta- and Orthogonal Integration of Influenza "oMICs" Data Defines a Role for UBR4 in Virus Budding. *Cell Host and Microbe*, *18*(6), 723–735. https://doi.org/10.1016/J.CHOM.2015.11.002/ATTACHMENT/EE2DCEA6-58D0-4855-BD8D-6FCC27A2D482/MMC9.XLSX

Tujebajeva, R. M., Copeland, P. R., Xu, X. M., Carlson, B. A., Harney, J. W., Driscoll, D. M., Hatfield, D. L., & Berry, M. J. (2000). Decoding apparatus for eukaryotic selenocysteine insertion. *EMBO Reports*, *1*(2), 158–163. https://doi.org/10.1093/EMBO-REPORTS/KVD033

van Ende, R., Balzarini, S., & Geuten, K. (2020). Single and Combined Methods to Specifically or Bulk-Purify RNA–Protein Complexes. *Biomolecules*, *10*(8). https://doi.org/10.3390/biom10081160

van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., Xiao, R., Blue, S. M., Chen, J.-Y., Cody, N. A. L., Dominguez, D., Olson, S., Sundararaman, B., Zhan, L., Bazile, C., Bouvrette, L. P. B., Bergalet, J., Duff, M. O., Garcia, K. E., Gelboin-Burkhart, C., … Yeo, G. W. (2020). A large-scale binding and functional map of human RNA-binding proteins. *Nature*, *583*(7818), 711–719. https://doi.org/10.1038/s41586-020-2077-3

Vanda Papp, L., Holmgren, A., Kum Khanna, K., Finley, J. W., Sies, H., Stolz, J. F., & Williamson, G. (2007). From Selenium to Selenoproteins: Synthesis, Identity, and Their Role in Human Health. *Https://Home.Liebertpub.Com/Ars*, *9*(7), 775–806. https://doi.org/10.1089/ARS.2007.1528

Voisinne, G., Gonzalez de Peredo, A., & Roncagalli, R. (2018). CD5, an Undercover Regulator of TCR Signaling. *Frontiers in Immunology*, *9*, 2900. https://www.frontiersin.org/article/10.3389/fimmu.2018.02900

Weinreb, C., Riesselman, A. J., Ingraham, J. B., Gross, T., Sander, C., & Marks, D. S. (2016). 3D RNA and Functional Interactions from Evolutionary Couplings. *Cell*, *165*(4), 963–975. https://doi.org/10.1016/J.CELL.2016.03.030

Zhao, W., Siegel, D., Biton, A., le Tonqueze, O., Zaitlen, N., Ahituv, N., & Erle, D. J. (2017). CRISPR–Cas9-mediated functional dissection of 3′-UTRs. *Nucleic Acids Research*, *45*(18), 10800–10810. https://doi.org/10.1093/NAR/GKX675

Zubiaga, A. M., Belasco, J. G., & Greenberg, M. E. (1995). The nonamer UUAUUUAUU is the key AU-rich sequence motif that mediates mRNA degradation. *Molecular and Cellular Biology*, *15*(4), 2219–2230. https://doi.org/10.1128/MCB.15.4.2219

CHAPTER 3: Post-transcriptional control of mRNA stability modulates CD69 expression and thymic egress

Abstract

T cells undergo dynamic and rapid changes upon activation to migrate and respond to foreign pathogens in tissues. CD69, an inducible cell surface protein, regulates T cell migration by inhibiting sphingosine-1-phosphate (S1P)-mediated egress from lymphoid and peripheral organs. While the mechanisms regulating CD69 transcriptionally and post-translationally have been characterized, how it's regulated post-transcriptionally remains to be studied. Here, we evaluate how post-transcriptional regulatory networks in the 3'UTR of *CD69* modulate CD69 and T cell function. Absence of *CD69* 3'UTR in human and mouse T cells resulted in higher *CD69* transcript and protein expression. This un-regulated expression of CD69 inhibited T cell egress from the thymus in mice. Further dissection of the 3'UTR revealed a conserved region containing multiple predicted RNA binding protein sites that was necessary for regulating CD69 expression. Homozygous deletion of the destabilizing region in mice resulted in T cells with higher CD69 levels and higher proportion of CD69+ T cells after activation. Together, our data demonstrates the crucial role post-transcriptional circuitry plays in modulating CD69 expression and T cell migration.

Introduction

T cells are critical for mounting an adaptive immune response against foreign pathogens. Upon antigen presentation, these cells must respond rapidly to exert their effector function as well as migrate to specific sites of infection (Mueller et al., 2013). Just as T cells must quickly alter their gene expression profile and protein composition when activated, they must also alter their programming to enter a resting state as well as egress from tissues when infections are resolved, as prolonged proinflammatory T cell responses can lead to tissue damage (Jurgens et al., 2021; Mueller et al., 2013). Targeting the mechanisms that regulate T cell migration can therefore provide therapeutic benefits for T cell-mediated inflammatory disorders as well as be harnessed to improve vaccine efficacy.

CD69 is a type II C-lectin membrane protein that is highly regulated. It's rapidly upregulated on the cell surface in response to T cell receptor stimulation (Testi et al., 1989; Yamashita et al., 1993) and interferon I exposure (Shiow et al., 2006), and down-regulated within ~48-72 hours after activation. Commonly used as an early T cell activation marker, CD69 is also a negative regulator of T cell egress. On the cell surface, it couples with sphingosine-1-phosphate receptor 1 (S1PR1) and leads to internationalization and degradation of the receptor (Bankovich et al., 2010), which inhibits S1P-mediated egress and migration (Shiow et al., 2006). This interaction also leads to downregulation of CD69 on the surface and is a post-translational mechanism that governs CD69 expression. Modulating levels of CD69 on the cell surface has profound effects on T cell migration. T cells lacking CD69 can not accumulate and retain in tissues after challenge (Miki-Hosokawa et al., 2009; Takamura et al., 2016), which can lead to decreased formation of tissue resident memory cells (MacKay et al., 2013; Mackay et al., 2015). Overexpression of CD69 impairs T cell egress from the thymus, leading to low numbers of peripheral T cells (Feng et al., 2002; Kimura et al., 2002). While studies have elucidated the transcriptional circuitry that regulates CD69 (Laguna et al., 2015; López-Cabrera et al., 1995), how the mRNA is regulated post-activation and in a quiescent T cell state remains unclear.

T cells employ several regulatory mechanisms, including post-transcriptional regulation, to rewire their gene expression in response to environmental signals (Jurgens et al., 2021). RNA binding proteins (RBPs) are one of the *trans* factors that regulate post-transcriptional processes including mRNA localization, editing, splicing, translation, editing and decay (Turner & DÍaz-Muñoz, 2018). They recognize and bind to secondary structural motifs like the constitutive decay element (Leppek et al., 2013a; Schlundt et al., 2014; Tan et al., 2014) or short linear sequences such as AU-rich elements (AREs) in the 3'untranslated region (3'UTRs) of mRNA to mediate post-transcriptional processes (Chen & Shyu, 1995). The 3'UTRs of cytokines and other immunological mediators contain several RBP binding motifs that play a necessary role in tightly regulating the expression of these proteins in appropriate cellular contexts (Chen et al., 2013; Salerno et al., 2018; Techasintana et al., 2017). Mutations in RBP binding sites, as well as loss of function of RBPs, have been associated with changes in T cell fate and function that induce strong immune-related diseases (Gebauer et al., 2021; Steri et al., 2017). Previous work using reporter constructs identified AREs in the 3'UTR of *CD69* as destabilizing regulatory elements (Santis et al., 1995). How the 3'UTR regulates CD69 in its native contexts, as well as its effect on T cell function, is unknown.

In this study, we show that the 3'UTR of *CD69* is necessary for regulating CD69 expression and T cell function, as absence of the region resulted in impaired T cell egress from tissues. Further dissection of the 3'UTR found a conserved region in mouse and human T cells that contains a decay element critical for regulating degradation of *CD69* transcript and for regulating CD69 expression on activated T cells. These results provide insight into the post-transcriptional circuitry that regulates CD69 and its necessary role in modulating T cell migration.

Result

Post-transcriptional circuitry in the 3'UTR regulates CD69 protein expression in T cells

To understand how post-transcriptional regulation affects CD69 expression, paired CRISPR-Cas9 ribonucleoproteins (RNPs) were used to delete the 3'UTR of *CD69* in primary human CD4 T cells (Fig 3.1A). The polyadenylation signal at the 3' end of the region was kept intact. Bulk edited cells that lacked *CD69* 3'UTR (Δ3'UTR) showed increased CD69 in resting condition as well as higher proportion of cells that were CD69+ (Fig 3.1B). Activating the Δ3'UTR cells with PMA and Ionomycin for 4 hours similarly led to higher expression of the protein (Fig 3.1C), indicating that the 3'UTR is necessary for modulating CD69.

CD69 3'UTR is well-conserved between mouse and human, with ~70% sequence conservation. To determine whether the effect observed in human T cells also occurs in mice, the 3'UTR of *Cd69* was deleted in primary mouse CD4 cells *in vitro* as was done in human T cells (Fig 3.1D). In mouse T cells, there was higher expression of CD69 and a greater proportion of CD69+ cells in resting condition (Fig 3.1E) as well as higher expression after activation (Fig 3.1F).

As the previous studies were conducted in bulk edited CD4 cells, a mouse model with the same 3'UTR deletion as conducted in Fig 3.1D-F was generated to evaluate the region's effect on CD69 expression. CD4 and CD8 T cells from secondary lymphoid organs (SLO), spleen and inguinal lymph nodes (iLNs), of Δ3'UTR mouse had higher CD69 expression (Fig 3.1G-H), similar to *in vitro* bulk edited cells. In the thymus, there was higher CD69 expression on CD4+CD8+ double positive (DP), CD8 single positive (SP) and CD4 SP thymocytes in Δ3'UTR mice compared to wild-type (WT) (Fig 3.1I-J). These data suggest that the *Cd69* 3'UTR contains conserved cis-regulatory elements critical for regulating protein expression.

Absence of *Cd69* 3'UTR leads to decreased peripheral T cells and accumulation of SP cells in the thymus

The Δ3'UTR mice were used to evaluate how absence of post-transcriptional control of CD69 affects T cell maintenance *in vivo*. At baseline, Δ3'UTR had significantly lower frequency of CD8 T cells in iLNs with a trend towards lower cell numbers for both CD4 and CD8 T cells (Fig 3.2A-B) compared to WT. In the spleen, Δ3'UTR had significantly lower proportion of CD4 and CD8 T cells, also with a trend towards lower cell numbers (Fig 3.2C-D). Previous work on CD69 transgenic mice demonstrated a similar decrease in peripheral T cells which resulted from an accumulation of SP thymocytes (Feng et al., 2002; Kimura et al., 2002). To determine whether Δ3'UTR mice express the same phenotype, we examined the thymus and found that the thymocyte populations in Δ3'UTR mice skewed towards higher proportion of CD4 SP and CD8 SP cells and lower frequency of DP cells (Fig 3.2E-F). Δ3'UTR had comparable cell numbers of DP thymocytes and a trend towards higher SP thymocyte numbers (Fig 3.2E-F).

CD69 negatively regulates T cell egress through its interaction with S1PR1 (Allende et al., 2004; Matloubian et al., 2004), their expression mutually exclusive. Because the Δ3'UTR mice showed accumulation of SP thymocytes, similar to S1PR1-deficient mice (Matloubian et al., 2004), we hypothesized that thymic egress in Δ3'UTR mice may be impaired due to decreased S1PR1 expression. Δ3'UTR mice had higher proportion and numbers of CD69+ SP cells and lower proportions of S1PR1+ cells compared to WT (Fig 3.2G-H). These data indicate that without the *Cd69* 3'UTR, thymic T cells overexpress CD69, which decreases S1PR1 on the surface and impairs T cell egress into the periphery.

Higher CD69 expression and impaired T cell egress in Δ3'UTR mice is cell intrinsic

To determine whether the 3'UTR affects CD69 expression and T cell egress in a cell intrinsic manner, 1:1 ratio of WT:WT or Δ3'UTR:WT bone marrow were reconstituted in WT recipients to generate bone marrow chimeras (Fig 3.3A). As expected, thymocytes generated

from Δ3'UTR donor marrow expressed higher CD69 on the surface compared to cells from WT marrow in the same recipient (Supplementary Fig 3.1A-B). In the spleen, recipients reconstituted with Δ3'UTR:WT marrow had lower ratio of CD4 and CD8 cells compared to those with WT:WT donors (Fig 3.3B-C). Interestingly, the ratio of Δ3'UTR:WT was only lower in CD8 cells in iLNs (Fig 3.3B-C). The decreased presence of Δ3'UTR CD8 cells in the periphery may be a consequence of SP thymocyte accumulation, as the chimeras also showed greater ratio of Δ3'UTR:WT CD8 SP cells compared to mice with WT:WT marrow (Fig 3.3D).

To determine whether this accumulation is due to impaired S1PR1-mediated egress, S1PR1 and CD69 expression was measured for SP thymocytes. There was a decrease in S1PR1+ CD4 SP and CD8 SP from Δ3'UTR:WT recipients (Fig 3.3E). S1PR1 FMO control was used to determine S1PR1+ expression (Supplementary Fig 3.1C). Ratio of S1PR1+ CD4 and CD8 cells in iLNs (Supplementary Fig 3.1D-E) and spleen (Supplementary Fig 3.1F) was also lower in Δ3'UTR:WT compared to WT:WT. Further analysis revealed an increase in mature-like SP thymocytes in Δ3'UTR:WT mice, which had higher ratios of CD62Lhi and Qa-2+ CD4 SP (Fig 3.3F) and CD8 SP (Fig 3.3G) cells compared to WT:WT. Immature SP thymocytes were unaffected, as CD62Llo and Qa-2- populations were comparable (Fig 3.3F-G). This suggests that the increase in SP cells is likely due to accumulation of mature thymocytes as seen in other models (Feng et al., 2002; Kimura et al., 2002; Matloubian et al., 2004) rather than alterations in SP thymocyte maturation. Together these data demonstrate the cell intrinsic role the 3'UTR plays in regulating CD69 expression and thymic egress.

Guided dissection of decay elements in *CD69* 3'UTR

A previous biochemical technique had identified potential RBP binding sites in *CD69* 3'UTR (Fig 3.4A), suggesting the presence of cis-regulatory elements in the region (Litterman et al., 2018). To identify the cis-regulatory region modulating the post-transcriptional effect on CD69 expression, various regions along the 3'UTR were deleted using Cas9 RNPs in primary

human CD4 T cells (Fig 3.4A). Deletion of Region A (ΔA) and B (ΔB) significantly increased CD69 expression and proportion of CD69+ resting T cells compared to cells with intact 3'UTR (Ctrl) (Fig 3.4B), with Region B having the greater effect compared to Region A. Cells stimulated with PMA and Ionomycin for 4 hours had a trend towards increased CD69 expression in cells with Region B (Fig 3.4C).

As deletions were induced in bulk edited populations (~40-50% editing efficiency), single-cell Jurkat clones expressing homozygous deletions of the different regions in the 3'UTR were generated. There was a greater abundance of *CD69* mRNA in clones that lacked the 3'UTR and all the individual regions, with Region B expressing the highest relative amount compared to the other regions (Fig 3.4D). To determine whether these regions contained destabilizing elements that were affecting transcript and protein expression, Jurkat clones were activated and then treated with actinomycin D to inhibit transcription before measuring the mRNA half-life. Removing the 3'UTR and Region B led to longer *CD69* mRNA half-life (Fig 3.4E) compared to cells with intact 3'UTR. This led to significantly higher protein expression for Δ3'UTR and ΔB clones (Fig 3.4F-G). While the cells with an intact *CD69* 3'UTR downregulated CD69 after 24 and 48 hours of rest, Δ3'UTR cells maintained high CD69 expression, a pattern also observed in ΔB (Fig 3.4F-G). Previous work had identified AU rich elements in Region D of the 3'UTR that exhibited destabilizing activity (Santis et al., 1995). Measuring the destabilizing activity of the individual regions using a dual luciferase reporter showed a decrease in reporter expression for the construct with Region D, but also for Regions A and B (Supplementary Fig 3.2). While Region D contained destabilizing AREs, Region B was necessary for controlling *CD69* mRNA and protein expression *in vitro*.

Post-transcriptional activity in Region B maps to RBP binding sites

Region B is a ~250bp sequence that contains various RBP binding sites and motifs. To fine-map the exact regions that expressed destabilizing activity, constructs with 20bp deletions

that spanned across Region B were generated and cloned into a dual luciferase vector (Fig 3.5A). Luciferase activity was measured in Jurkat cells. Constructs Δ2, Δ3, Δ5 and Δ6 had significantly higher reporter expression compared to the construct with the full length Region B (Fig 3.5B), indicating the presence of destabilizing elements within these specific regions. Δ12 also displayed a trend towards higher reporter activity but did not reach statistical significance (Fig 3.5B). Constructs with scrambled (scr) sequences at sections 2-3, 5-6 and 12 showed the same increase in reporter activity (Fig 3.5C-E). These sections with destabilizing activity mapped to RBP binding peaks (Fig 3.5A), which suggests the regulation may be mediated by the interaction between RBPs and cis-regulatory elements.

To determine the RBP that may be binding to the identified regions, proteins from cytoplasmic Jurkat lysates that bound to Region B were isolated using the streptavidin-binding RNA aptamer 4xS1m (Leppek & Stoecklin, 2014), which expresses the sequence from Region B. To control for false positive candidates, aptamers that expressed the individual scrambled sequences (scr 2-3, scr5-6 and scr12) were included. The analysis identified several proteins that bound to WT but not scr2-3 (Fig 3.5F), scr 5-6 (Fig 3.5G) or scr12 (Fig 3.5H) sequences. Notable proteins include ARE binding proteins ZFP36L2, which binds to *Cd69* transcript in mouse T cells (Salerno et al., 2018), and ELAVL1 (Fig 3.5F). These proteins may bind to the AU rich region (CUUUAUAUUAUU) in section 2-3 in the 3'UTR. IGF2BP2, an RBP that recognizes N6-methyladenosine modifications (Huang et al., 2018) and splicing protein KHDRBS3 (Vernet & Artzt, 1997) were both differentially expressed between WT and scr5_6 (Fig 3.5G). Their function in T cells has not been studied. RBMS1, which has been shown to regulate stability of certain transcripts in T cells (Hoefig et al., 2021) is another interesting candidate as well as PUM2 (Fig 3.5H) which can bind to the motif UGUA (White et al., 2001) in section 12. Future studies will validate the RBPs that bind and regulate CD69 through Region B.

Post-transcriptional destabilizing elements in *CD69* 3'UTR are conserved between mouse and human

As the sequence and overall function of the *CD69* 3'UTR is conserved between mouse and human, there may also be destabilizing elements in mouse *Cd69* 3'UTR that are necessary for regulation of CD69. Various regions in *Cd69* 3'UTR that contained RBP binding peaks were deleted in primary mouse CD4 cells using Cas9 RNPs (Fig 3.6A). In resting CD4 cells, deleting Region β (Δβ) significantly increased CD69 expression and proportion of cells that were CD69+ (Fig 3.6B). This increase was also observed in activated cells (Fig 3.6C). Δβ cells expressed a trend towards higher amounts of *Cd69* transcript compared to non-edited cells (Ctrl) (Fig 3.6D). As the data was from bulk edited cells, the presence of WT alleles may obscure the results from edited alleles in the qPCR assay. Further analysis revealed that deleting the 3'UTR extended *Cd69* mRNA half-life as did deleting Regions α and β (Fig 3.6E). Together, this suggests that Region β is most likely necessary for regulating CD69 protein expression and transcript levels in mouse T cells.

Overlapping the RBP binding profile and sequences from mouse and human *CD69* 3'UTR revealed shared RBP peaks (Fig 3.6F). Interestingly, Region B, which contains destabilizing activity in human T cells, and Region β share RBP binding sites (Fig 3.6F, gray bar). Although the coverage at the site is different between mouse and human, the region contains two conserved sequences as indicated by the PhyloP score, suggesting a potential shared RBP and cis-regulatory element interaction.

Post-transcriptional element in Region β regulates CD69 expression in T cells *in vitro* and *in vivo*

As previous *in vitro* studies on Region β in *Cd69* 3'UTR was performed in bulk edited cells, a mouse with germline deletion of Region β (Δβ) was generated to further analyze the effect of this region on CD69 expression *in vitro* and *in vivo*. To understand how Region β

affects *Cd69* transcript and protein expression in T cells during activation and rest, CD4+CD69- cells were isolated from WT, Δβ and Δ3'UTR mice and stimulated *ex vivo* for 3 days and then rested for 4 days afterwards (Fig 3.7A). During activation, Δβ and Δ3'UTR cells expressed significantly higher amounts of *Cd69* transcript compared to WT cells at different timepoints during stimulation, peaking at 4 hours (Fig 3.7B). Δβ and Δ3'UTR maintained higher levels of transcript even after cells were taken off stimulants and rested for 2-4 days (Fig 3.7C). On the protein level, Δβ cells displayed a trend towards higher CD69 expression when activated, although not as high as Δ3'UTR cells (Fig 3.7D). After cells were rested, Δ3'UTR consistently expressed higher CD69 at various timepoints and maintained residual CD69 expression even after WT cells had returned to baseline (Fig 3.7E). Δβ cells in the first 24 hours after activation expressed higher levels of CD69 compared to WT but expressed similar levels of the protein 2-4 days of rest (Fig 3.7E).

This higher upregulation was also observed in an *in vivo* challenge. CD69- CD4 naïve cells from WT (WT-OTII) and Δβ mice (Δβ-OTII) mice that expressed OTII, ovalbumin(OVA)-specific TCR transgene, were transferred into WT recipients at 1:1 ratio. Mice were then sensitized and challenged with OVA and house dust mite (HDM) through the airways to activate the donor cells *in vivo* (Fig 3.7F). After acute challenge, a significantly greater proportion of lung memory Δβ-OTII cells expressed CD69 compared to WT independent of congenic background (Fig 3.7G-H). Within CD69+ cells, Δβ cells also expressed higher levels of CD69 compared to co-transferred WT cells (Fig 3.7I). Together, this suggests that Region β is necessary for regulating CD69 and the absence of the destabilizing element within the region leads to uncontrolled protein and transcript expression.

Δβ mice do not have impaired T cell egress at baseline

To understand whether Region β affects T cell maintenance and egress *in vivo*, 1:1 ratio of WT and WT or WT and Δβ donor bone marrow was reconstituted in WT recipients. As

expected, Δβ thymocytes expressed higher CD69 at various T cell development stages compared to WT (Supplementary Fig 3.3 A-B). Comparison of WT:WT and Δβ:WT thymocytes revealed similar ratios of DP and SP cells (Supplementary Fig 3.3C). In the periphery, Δβ:WT and WT:WT chimeras had similar ratios of CD4 and CD8 cells in the spleen, but higher ratios of both subsets in iLNs from Δβ:WT (Supplementary Fig 3.3D). In both SLOs, there were more CD69+ Δβ T cells compared to WT, but comparable number of cells expressing S1PR1 (Supplementary Fig 3.3D). Further analysis showed higher CD69+ and significantly lower ratios of S1PR1+ cells in Δβ:WT recipients (Supplementary Fig 3.3E-F). In contrast, S1PR1 expression was similar between WT:WT and Δβ:WT T cells from iLNs (Supplementary Fig 3.3G) and spleen (Supplementary Fig 3.3H) despite greater proportion of CD69+ cells in Δβ:WT. Unlike Δ3'UTR, Δβ mice do not exhibit abnormal thymic egress and are phenotypically similar to WT at baseline.

Discussion

Previous studies have shown how CD69 expression is controlled on the transcriptional level through transcription factor binding at enhancers and the promoter (Laguna et al., 2015; Vazquez et al., 2009), as well as post-translationally (Shiow et al., 2006). In this work, we demonstrate how post-transcriptional regulation is also critical for regulating CD69 expression in T cells. Human and mouse T cells that lacked *CD69* 3'UTR expressed higher amounts of *CD69* transcript and protein. Further dissection of the 3'UTR revealed a specific region in mouse and human that contained a decay element necessary for regulating CD69 protein output.

CD69 3'UTR contains many AU-rich motifs, particularly clustered towards the 3'end of the region. When expressed in a reporter construct in T cells, we, along with others (Santis et al., 1995), found this AU-rich region to be destabilizing in T cells. Although deleting this AU rich region (Region D) in Jurkat cells resulted in higher CD69 protein, eliminating Region B had the

greatest effect on transcript decay, expression, and protein output *in vitro*. The exact motif and *trans* factor driving the post-transcriptional effect is yet undefined. As Region B also contains an ARE, this element, rather than the ones in Region D, may be the primary regulators of *CD69* mRNA degradation in the cells. Preferential use of specific AREs within the same 3'UTR have been observed in other genes such as *PIM3* (Litterman et al., 2018). It's also possible that the post-transcriptional decay of *CD69* is driven by another RBP. Our proteomics analysis identified a list of RBPs that may bind to Region B, including ZFP36L2 which binds to *Cd69* transcript in mouse and human T cells (Salerno et al., 2018). On-going studies will determine whether ZFP36L2, as well as the other RBPs, bind to Region B in the 3'UTR and regulate *CD69* transcript and protein expression.

The destabilizing activity discovered in Region B of human *CD69* 3'UTR was conserved in mouse 3'UTR. Although the local peak maxima of GCLiPP read coverage differed between mouse and human at these sites, the aligned sequences at the peaks were highly conserved and contained AREs, which suggests a shared post-transcriptional mechanism between the two species. At baseline, Δβ mice were phenotypically similar to WT in terms of T cell distributions in the thymus and SLOs despite higher CD69 expression. The difference in phenotype between Δβ and Δ3'UTR may be driven by different levels of CD69 expressed on the cell surface. In CD69 transgenic mice, transgene copy number was inversely correlated with the number and proportion of peripheral T cells (Feng et al, 2002). Mice with high copy numbers also had increased SP thymocyte accumulation similar to Δ3'UTR mice. Overexpression of CD69 down-regulates S1PR1 from the cell surface (Bankovich et al., 2010). It's possible that CD69 expression on Δβ thymocytes may not reach the threshold in which it alters the ratio of CD69:S1PR1 on the surface and inhibits S1P-mediated egress. This balance between CD69 and S1PR1 may be especially sensitive in CD8 T cells, which we observed to be significantly decreased in the periphery of Δ3'UTR mice. In contrast, CD4 T cells have higher S1PR1 expression and may tolerate higher CD69 on the cell surface without affecting egress.

We demonstrate an effect of Δβ on CD69 expression *in vivo* and show that sensitization and challenge with antigen can induce more CD69+ memory cells on Δβ OTII cells in the lung, which also express higher CD69 levels compared to CD69+ WT memory cells. Many questions remain about the function of Δβ T cells in peripheral tissues that is not addressed in this study. How the absence of Region β affects T cell maintenance in peripheral tissues is unclear. CD69+ memory resident T cells in non lymphoid tissues need to upregulate S1PR1 and subsequently down-regulate CD69 to egress and re-circulate (Fonseca et al., 2020). The absence of Region β in T cells may therefore impair the cell's ability to down-regulate CD69 and delay their migration out of the tissue. On-going studies will address this area.

Recent genetic analysis has suggested that post-transcriptional elements in the 3'UTR may be involved in immune-mediated diseases. Two single nucleotide polymorphisms (SNPs) associated with multiple sclerosis (MS) (Beecham et al., 2013) and type 1 diabetes (Onengut-Gumuscu et al., 2015) were identified to be likely causal SNPs through the Probabilistic Identification of Causal SNPs algorithm (PICS) (Farh et al., 2015; Taylor et al., 2021). The variant associated with MS, rs11052877, is especially interesting as it has a high PICS probability (0.79) and lies within an RBP binding site in Region B in human *CD69* 3'UTR. Treatment for MS commonly includes S1PR1 agonist FTY720 and indicates a critical role for S1P-mediated migration (Chun & Hartung, 2010). Therefore having regulatory mechanisms, even on the post-transcriptional level, to modulate CD69 expression and T cell egress may be critical for preventing a proinflammatory states in tissues like the central nervous system. Together, the genetic and biochemical analysis suggest an important functional role of this region in regulating CD69 and T cell function in autoimmune disorders. Future studies will examine how the 3'UTR and Region β affects T cell function *in vivo* by inducing experimental autoimmune encephalomyelitis in mice, an animal model of MS.

In conclusion, we demonstrate the critical role post-transcriptional regulation has in modulating *CD69* transcript and protein expression in mouse and human T cells. To our

knowledge, our work is the first to identify a specific region within the 3'UTR that contains a destabilizing element with significant effects on transcript half-life and protein output *in vitro* and *in vivo*. On-going work will decipher the RBPs involved in regulating *CD69* as well as the impacts of these post-transcriptional elements on T cell function in inflammatory settings.

Materials and Methods

Mice

Mice with *Cd69* 3'UTR deletion (Δ3'UTR) and region β deletion (Δβ) were generated by the UCSF Mouse Genetic Core by microinjection of Cas9 ribonucleoprotein into C57BL/6 zygotes. Briefly, paired crRNAs (160μM) (Δβ used mCD69 3'UTR 2 and 3; Δ3'UTR used mCD69 3'UTR 1 and 5) were incubated with tracrRNA (160μM) for 30min at 37°C in injection buffer (10mMTris, 150mM KCl) and then incubated with Cas9 (40uM) for 15min at 37°C. The mixture was microinjected into the cytoplasm of C57BL/6 single-cell zygotes isolated from super-ovulated females. Δ3'UTR and Δβ mouse line was established by backcrossing founder for at least one generation before breeding to homozygosity. Guide sequences can in Supplementary Table 4. Δβ mice were crossed to TCR OT-II transgenic mice (Jackson Laboratory) to generate Δβ-OTII mice that were homozygous for deletion and expressed OT-II transgene. WT-OTII and Δβ-OTII on CD45.2 background mice were bred to BoyJ CD45.1+ mice to generate strains that were CD45.1/CD45.2. Male and female age and sex matched mice were used between 6-12 weeks of age. All mice were housed and bred in specific pathogen-free conditions in the Animal Barrier Facility at the University of California, San Francisco. Animal experiments were approved by the Institutional Animal Care and Use Committee of the University of California, San Francisco.

Mixed Bone Marrow Chimeras

To generate mixed bone marrow chimeras, 8-9 week old CD45.1+ BoyJ mice (The Jackson Laboratory) recipients were lethally irradiated (2X360 rad) and reconstituted with 1:1 ratio of WT (CD45.1/CD45.2) and either WT (CD45.2), Δβ (CD45.2) or Δ3'UTR (CD45.2) bone marrow cells. Animals were analyzed 8 weeks later.

Allergic Inflammation Model

Naïve CD4 cells were isolated from spleen and lymph nodes of age-matched female WT-OTII and Δβ-OTII mice using EasySep Mouse Naïve CD4+ T Cell Isolation Kit (Stem Cell Technologies). 2μg of anti-mouse CD69 biotin was added along with the CD4 Isolation cocktail from the kit to select against CD69+ cells. Naïve CD4+CD69- cells from WT-OTII and Δβ-OTII animals were mixed at 1:1 ratio and injected into anesthetized female CD45.1+ BoyJ through *i.v*. The next day, recipients were anesthetized and sensitized with 10 μg of house dust mite extract (HDM) (Greer Laboratories) and 50μg of ovalbumin (OVA) in 20uL PBS administered to the lung by oropharyngeal (o.p.) aspiration. Seven days later mice were then challenged with 10 μg of HDM and 10 μg OVA by o.p. daily for 5 consecutive days. Mice were euthanized for tissue processing 2 days after the last challenge.

Tissue Processing

Spleen, lymph nodes and thymus were dissociated into single-cell suspension and passed through 70μM filter. Lungs were digested with 1.2mg/mL of Collagenase D (Roche) and 10U/mL of deoxyribonuclease I (Roche) for 30min at 37°C and dissociated with gentleMACS dissociator (Miltenyi Biotec). Enzymatic reaction was inhibited with cold PBS with 2% fetal bovine serum (FBS) (Omega). Cells were filtered through 70μM filter, pelleted through centrifugation and incubated with ACK Lysis Buffer (Gibco) to remove red blood cells before final suspension in PBS with 2% FBS.

Flow Cytometry

FC receptors of cells were blocked with anti-mouse CD16/32 (BioXCell). Cells were surface stained with anti-mouse monoclonal antibodies from BioLegend and BD Biosciences including: anti-CD4 (RM4-5), anti-CD8 (53-6.7), anti-CD44 (IM7), anti-CD45.1 (A20), anti-CD45.2 (104), anti-CD69 (H1.2F3), anti-CD62L (MEL-14) and anti-Qa2 (695H1-9-9). Dead cells were excluded with Fixable Viability Dye eFluor780 (eBiosciences). Samples were collected with BD Fortessa and analyzed with FlowJo software (TreeStar).

For intravascular staining, mice were anesthetized and injected with 2µg of anti-mouse CD45 (30-F11) in 100uL PBS by *iv* for ~3min and euthanized by CO^2 asphyxiation.

For intracellular Foxp3 expression, cells were fixed and permeabilized using Foxp3 Transcription Factor Staining Buffer Set (eBiosciences) and stained with anti-mouse Foxp3 (FJK-16s).

For S1PR1 staining, cells were incubated with rat anti-mouse S1PR1 Antibody (MAB7089, R&D Systems), followed by APC donkey anti-rat IgG Fc-fragment (Jackson ImmunoResearch) as previously described (Arnon et al., 2011). Samples given secondary APC IgG fragment but not S1PR1 antibody (S1PR1 FMO) was used as background stain for S1PR1.

***In Vitro* Cell Cultures**

Jurkat cells were grown in RPMI supplemented with 10% FBS and 2mM L-glutamine (Invitrogen). Primary $CD4^+$ mouse T cells were isolated from mouse peripheral lymph nodes and spleen using EasySep Mouse CD4 Negative Isolation Kit (StemCell Technologies) according to the manufacturer's instructions. Cells were stimulated with immobilized biotinylated anti-CD3

(clone 2C11, 0.25 μg/mL, BioXcell) and anti-CD28 (clone 37.51, 1 μg/mL, BioXcell) bound to Corning cell culture dishes coated with Neutravidin (Thermo) at 10 μg/mL in PBS for 3 h at 37 °C. Cells were left on stimulation and treated with 10μg/mL of anti-interferon gamma (αIFNγ) for 3 days before being transferred to non-coated dishes in T cell medium supplemented with recombinant human IL-2 (20 U/mL) and 10μg/mL of αIFNγ . For re-stimulation, cells were treated with 20nM phorbol 12-myristate 12-acetate (PMA) and 1 μM ionomycin (Sigma) for 4 hours before harvest.

Measurement of mRNA Decay

Cells were stimulated with PMA and Ionomycin for 4 hours and then additionally treated with Actinomycin-D (Sigma-Aldrich) at 5 μg/mL for an additional 0, 1, 2 or 4 hours. After treatment, cells were lysed with Trizol LS (Life Technologies) and processed with Direct-zol ™ 96 well RNA (Zymogen). RNA was quantified with an ND-1000 spectrophotometer (NanoDrop) and reverse transcribed with SuperScript III First Strand Synthesis Kit (Invitrogen). Primers used are listed in Supplementary Table 1. qPCR was performed in technical duplicates on Realplex² (Eppendorf) using TB Green Advantage qPCR Premix (Takara Bio). Ct values were normalized to housekeeping gene described in figure legends.

Luciferase Assays

Jurkats were transfected with luciferase reporter constructs using E6.1 Jurkat program on Amaxa 4D-Nucleofector 96-well Unit (Lonza Biosciences). Luciferase activity was measured 20-24 h after transfection with the Dual Luciferase Reporter Assay System (Promega) and a FLUOstar Optima plate-reader (BMG Labtech). Fragments of human *CD69* 3'UTR were cloned into the psiCHECK-2 luciferase reporter construct using Xho1 and Not1 sites (Promega). Primers for cloning are listed in Supplementary Table 2. Site-directed mutagenesis (SDM) was

performed using Q5 Site-Directed Mutagenesis Kit (New England BioLabs) following manufacturer's protocol using primers in Supplementary Table 3.

CRISPR Editing of T cells

Guide RNA sequences were selected using the Benchling online CRISPR design tool (https://benchling.com/crispr) with guides selected to target genomic regions of GCLiPP read density. Synthetic crRNAs and tracrRNA (Dharmacon) were resuspended in water at 160 µM at 1:1 ratio and allowed to hybridize at 37 c for 30 m. This annealed gRNA complex (80 µM) was then mixed 1:1 by volume with 40 µM *S. pyogenes* Cas9-NLS (University of California Berkeley QB3 Macrolab) to a final concentration of 20 µM Cas9 ribonucleotide complex (RNP). The diluted gRNA:Cas9 RNPs along with a random 200bp single-stranded DNA fragment were nucleofected into primary human (48 hours after stimulation) or mouse CD4 T cells (24 hours after stimulation) with the P3 Primary Cell 96-well Nucleofector ™ Kit and into Jurkat cells with the SE Cell Line 96-well Nucleofector ™ Kit using the 4-D Nucleofector following the manufacturer's recommendations (Lonza Biosciences). Cells were pipetted into pre-warmed media and then returned to anti-CD3 and anti-CD28 stimulation for another 2 days for primary mouse T cells, 1 day for primary human T cells or an additional 3-10 days in resting conditions for Jurkat cells. crRNA sequences are listed in Supplementary Table 4.

Generation of Single-Cell Clones

Jurkats were nucleofected with Cas9 RNPs to delete various regions within *CD69* 3'UTR as described above. The cells were rested in culture for a few days and then single-cell sorted into 96-well plate using FACS Aria (BD Biosciences) and propagated for 2-3 weeks. Genomic DNA from clones was extracted with QuickExtract Buffer (Lucigen) and mixed with *CD69* 3'UTR primers (Table S1) as well as MyTaq Red Mix (Meridian Biosciences) and run on Mastercycler

(Eppendorf). The resulting PCR product was run on 2% agarose gel to identify clones that were homozygous for the specific deletions. These clones were propagated for analysis**.**

Cloning *CD69* 3'UTR Fragments into 4xS1m Aptamer

Human *CD69* 3'UTR variants were cloned into pSP73-4xS1m vector (Leppek) through BamHI and EcoRI sites as previously described (Salerno et al., 2018). To generate 4xS1m RNA aptamers alone or containing *CD69* 3'UTR fragments, plasmids were linearized with nDel, *in vitro* transcribed using Ampliscribe T7-flash transcription kit (Epicentre), and RNA was purified using the mini Quick Spin Oligo columns (Roche). The RNA quality and quantity was determined by RNAnano Chip assay (Agilent). Primers used to clone fragments into vector are listed in Supplementary Table 5.

4xS1m RNA Aptamer-Protein Pull Down

Resting Jurkat T cells were washed with PBS and the pellet was snap frozen on dry ice. Cells were homogenized using 5-mm steel beads and a tissue lyser (Qiagen TissueLyser II) 6x at 25 Hz for 15 s. The homogenate was then solubilized and precleared as previously described (Salerno et al., 2018). Cell lysates were incubated with RNA-aptamer-coupled beads for 3.5 h at 4 °C under rotation in the presence of 60 U RNasin (Ambion). For each pull-down, about 30 µg of *in vitro* transcribed RNA, coupled to Streptavidin Sepharose High Performance beads (GE Healthcare), and 5-10 mg cell lysate protein was used. RNA-bound proteins were eluted by adding 1 µg RNaseA (Thermo Scientific) in 100 µ 100 mM Tris-HCl, pH 7.5 (Gibco-Invitrogen). Proteins were reduced, alkylated and digested into peptides using trypsin (Promega). Peptides were desalted and concentrated using Empore-C18 StageTips and eluted with 0.5 % (v/v) acetic acid, 80 % (v/v) acetonitrile. Sample volume was reduced by SpeedVac and supplemented with 2 % acetonitrile and 0.1 % TFA.

Mass Spectrometry Data Acquisition

Tryptic peptides were separated by nanoscale C18 reverse chromatography coupled online to an Orbitrap Fusion Tribrid mass spectrometer via a NanoElectroSpray Ion Source (both Thermo Scientific). Peptides were loaded on a 20 cm 75-360 µm inner-outer diameter fused silica emitter (New Objective) packed in-house with ReproSil-Pur C18-AQ 1.9 µm resin (Dr Maisch GmbH). The column was installed on a Dionex Ultimate3000 RSLC NanoSystem (Thermo Scientific) using a MicroTee union formatted for 360 µ outer diameter columns (IDEX) and a liquid junction. The spray voltage was set to 2.15 kV. Buffer A was composed of 0.5 % acetic acid and buffer B of 0.5 % acetic acid, 80 % acetonitrile. Peptides were loaded for 17 min at 300 nl/min at 5 % buffer B, equilibrated for 5 min at 5 % buffer B (17-22 min) and eluted by increasing buffer B from 5 to 15 % (22-87 min) and 15 to 38 % (87-147 min), followed by a 10 min wash to 90 % and a 5 min regeneration to 5 %. Survey scans of peptide precursors from 400 to 1,500 m/z were performed at 120 K resolution (at 200 m/z) with a 4×10^5 ion count target. Tandem mass spectrometry was performed by isolation with the quadrupole with isolation window 1.6, HCD fragmentation with normalized collision energy of 30 and rapid scan mass spectrometry analysis in the ion trap. The MS2 ion count target was set to 10^4 and the maximum injection time was 35 ms. Only those precursors with charge states 2-7 were sampled for MS2. The dynamic exclusion duration was set to 30 s with a 10-p.p.m. tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on. The instrument was run in top speed mode with 3 s cycles. All data were acquired with Xcalibur software.

Mass spectrometry data analysis

The raw mass spectrometry files were processed with the MaxQuant computational platform, 1.6.2.10 (Cox & Mann, 2008). Proteins and peptides were identified using the Andromeda search engine by querying the human Uniprot database (downloaded February 2017 and

February 2019, 89,796 entries) for the RNA pull-down. Standard settings with the additional options match between runs, and unique peptides for quantification were selected. The generated ‘proteingroups.txt’ data were imported in R and processed with the Differential Enrichment analysis of Proteomics data (DEP) R package (Zhang et al., 2018). Identified peptides were filtered for potential contaminants, only identified by site and reverse hits. The raw intensity values were transformed in log2 scale and averaged, and log2 fold change (LFC) was calculated. We only included proteins that were identified in at least 2 out of 3 replicates in this analysis. To select for RBPs, we compiled the 1153 RBPs identified by RNA-interactome capture on HeLa and Jurkat cells (Castello et al., 2016; Perez-Perri et al., 2018b) with 1542 computationally predicted RBPs based on the presence of a defined list of RNA-binding domains (RBDs) (Gerstberger et al., 2014). Filtered data are shown as log2 median-centered intensities.

Quantification and Statistical Analysis

Excel (Microsoft), Prism (GraphPad), Flowjo (TreeStar) and R Studio (Rstudio Team) were used for data analysis. Individual statistical tests performed are included in the data legends. All data was assumed to be normally distributed unless stated otherwise.

Data Availability

The GCLiPP data reported in this paper can be accessed at GSE115886 on Gene Expression Omnibus.

Figures

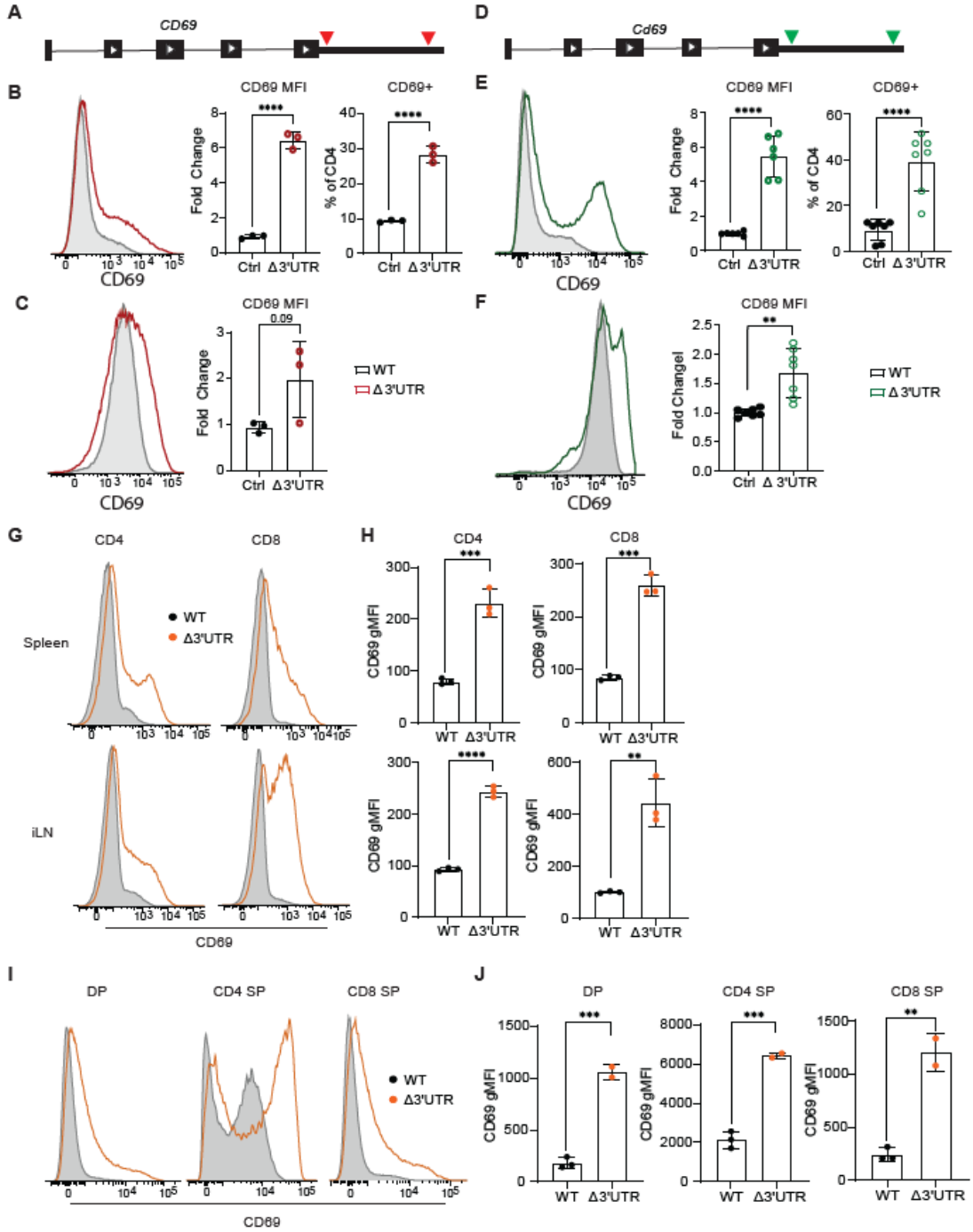


Figure 3.1. *CD69* 3'UTR modulates CD69 protein expression.

(A) Schematic of human *CD69* gene. Red arrows represent cut sites to remove the 3'UTR using CRISPR-Cas9 editing. **(B)** Representative flow cytometry plot of CD69 expression on resting primary human CD4 T cells (left). Red line represents cells with 3'UTR deletion and grey histogram shows background cells given control (Ctrl) gRNAs. Quantification of mean fluorescence intensity (MFI) and percentage of CD69+ CD4 cells shown on right. MFI was

normalized to average Ctrl values for each experiment and plotted as fold change (right). **(C)** CD69 expression on human primary CD4 cells activated with PMA and Ionomcyin for 4 hours (left). CD69 MFI was normalized to Ctrl and quantified on the right. **(D)** Schematic of mouse *Cd69* 3'UTR. Green arrows represent cut sites for CRISPR-Cas9 editing. **(E)** Representative flow cytometry plot of CD69 expression on resting mouse CD4 T cells (left). Grey histogram represents cells given Ctrl gRNA and green line represents edited cells without the 3'UTR. MFI was normalized to average Ctrl. **(F)** CD69 MFI on mouse CD4 cells activated with PMA and Ionomcying for 4 hours (left). MFI normalized to Ctrl (right). **(G)** CD69 expression on CD4 and CD8 T cells from spleen and inguinal lymph nodes (iLN), quantified in **(H)**. Gray histogram represents wild-type (WT) mice and orange line represents mice with homozygous deletion of *Cd69* 3'UTR (Δ3'UTR). **(I)** CD69 expression on double positive (DP), CD4 single positive (SP) and CD8 SP thymocytes. MFI is quantified in **(J).** Each dot represents a single mouse or human subject. Data shows mean ± SEM. Significance was calculated using student's t-test. **** $p<0.0001$, *** $p<0.001$, ** $p<0.01$, * $p<0.05$

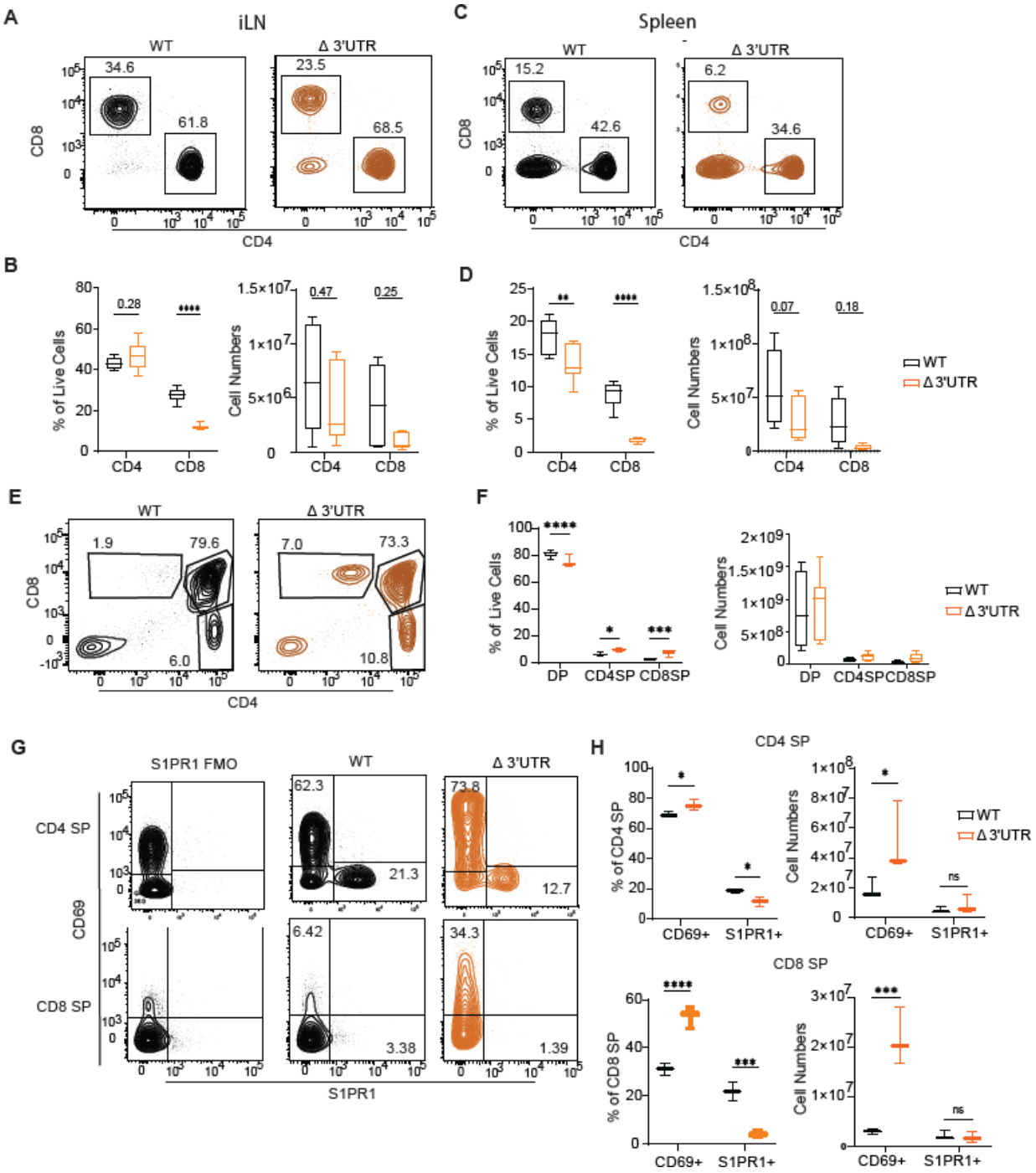


Figure 3.2. Δ3'UTR mice have impaired thymic egress.

(A) Representative flow cytometry plot of CD4 and CD8 T cells in iLNs. Proportion of CD4 and CD8 T cells, as well as cell numbers, are shown in **(B)**. **(C)** Representative flow cytometry plot of CD4 and CD8 T cells in spleen with numbers and percentage of live cells shown in **(D)**. **(E)** Representative flow cytometry plot with gates on DP, CD4 SP and CD8 SP thymocytes. The percentage of live cells and cell numbers are quantified in **(F)**. **(G)** Representative flow cytometry plots of CD4 SP and CD8 SP showing CD69 and S1PR1 expression. Gates for

S1PR1 were set using S1PR1 fluorescence minus one (FMO) in which S1PR1 antibody was not added. **(H)** Percentage of CD69+ and S1PR1+ cells and cell numbers for CD4 SP (top) and CD8 SP (bottom). Black represents WT and orange represents Δ3'UTR mice. Results are from 2-3 independent experiments with N=6-7 total. Šídák multiple comparisons tests was used to calculate significance. **** $p<0.0001$, *** $p<0.001$, **$p<0.01$, *$p<0.05$

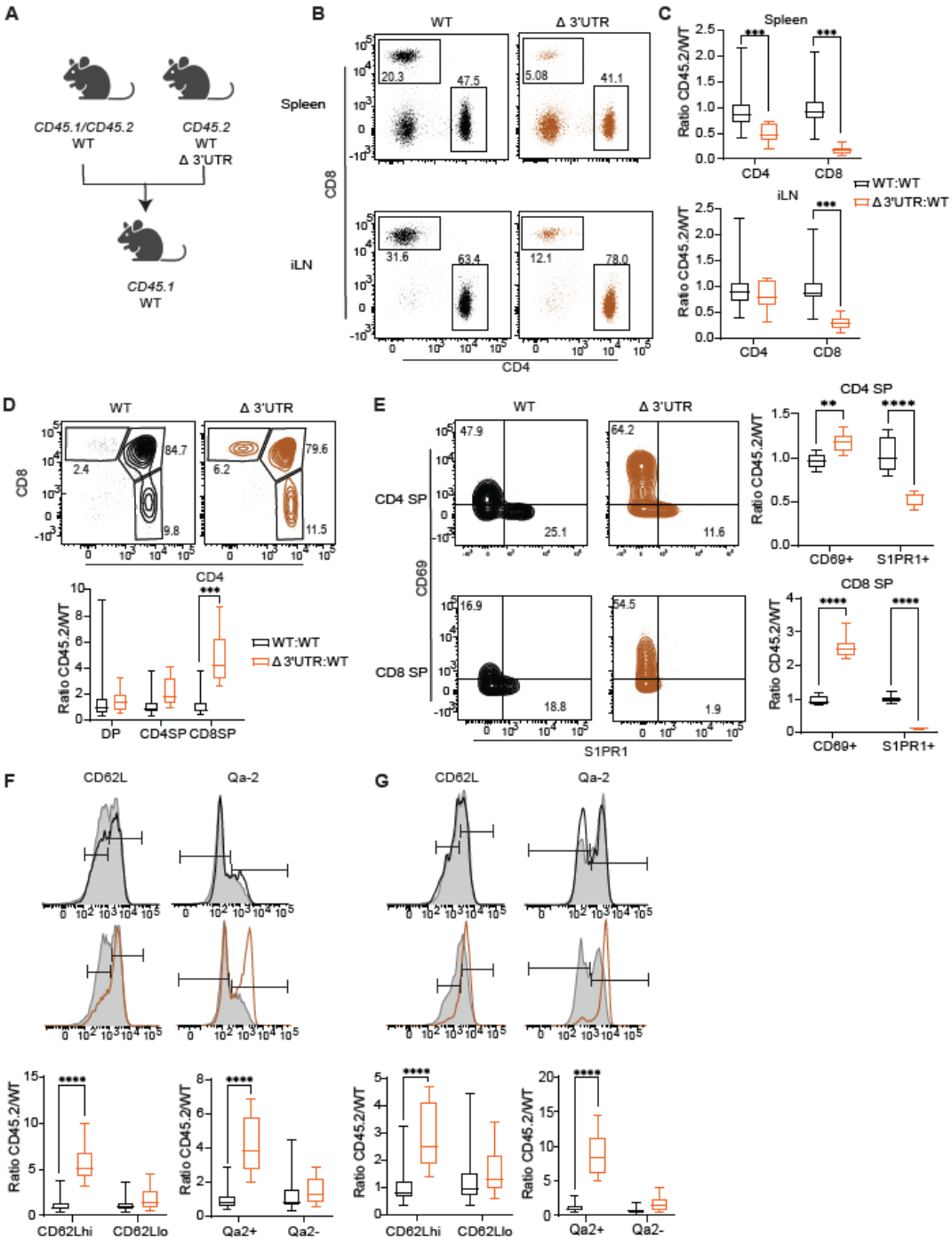


Figure 3.3. 3'UTR regulates CD69 protein expression and T cell egress in cell-intrinsic manner.

(A) Diagram of mixed bone marrow chimeras. WT recipients were re-constituted with 1:1 ratio of WT:WT or Δ3'UTR:WT bone marrow. **(B)** Representative flow cytometry plots showing CD4 and CD8 T cells in spleen (top) and iLNs (bottom) from WT (CD45.2) or Δ3'UTR (CD45.2). Ratio of WT/WT or Δ3'UTR/WT plotted in **(C)**. **(D)** Representative flow cytometry plot showing WT or Δ3'UTR thymocyte populations. Ratio of WT:WT or Δ3'UTR:WT plotted in **(D)**. **(E)** CD69 and

S1PR1 expression on CD4 SP (top) and CD8 SP (bottom). Ratios of WT or Δ3'UTR over co-transferred WT quantified on right. **(F-G)** CD62L (left) and Qa-2 (right) expression on CD4 SP thymocytes **(F)** and CD8 SP **(G)**. Gates demarcate CD62Llo and CD62Lhi, as well as Qa-2- and Qa-2+. Ratio of WT:WT and Δ3'UTR:WT plotted on bottom for each population. Black represents WT:WT and orange represents Δ3'UTR:WT. Results are from 2-3 independent experiments with N=16-26 total. Šídák multiple comparisons tests was used to calculate significance. **** p<0.0001, *** p<0.001, **p<0.01, *p<0.05

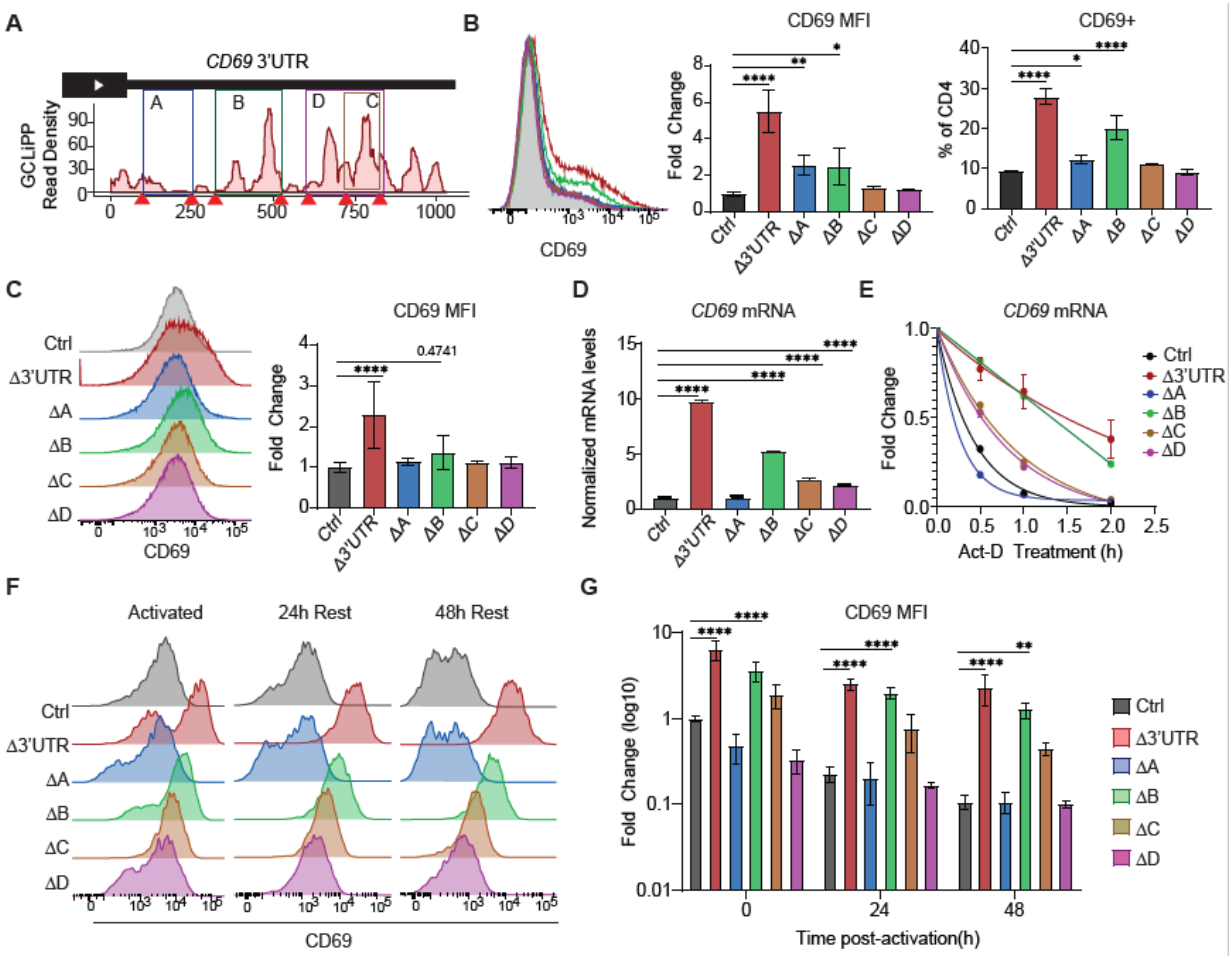


Figure 3.4. Dissection of 3'UTR identifies destabilizing elements that regulate *CD69* mRNA and protein expression.

(A) Schematic of *CD69* 3'UTR with red arrows showing gRNA placement and boxes indicating regions that were deleted. **(B)** Representative flow cytometry plot of CD69 expression on resting primary human CD4 T cells (left). CD69 MFI and percentage of CD69+ cells are quantified on right. Ctrl samples were given control gRNAs and express intact 3'UTR. MFI normalized to average Ctrl per experiment. **(C)** CD69 expression in Jurkat clones that contained various deletions indicated in (A) shown in a representative flow cytometry plot (left). MFI is quantified on the right. **(D)** *CD69* mRNA levels were normalized to housekeeping gene *GAPDH*. **(E)** *CD69* mRNA levels, normalized to housekeeping gene and then to 0 hour timepoint, at various times during treatment with actinomycin-D (act-D). **(F)** Representative plots showing CD69 expression

in the different clones after 4h activation with PMA and Ionomycin and resting for 24 or 48h post-stimulation. MFI was normalized to Ctrl for all timepoints and quantified in **(G)**. N=3 for experiments with primary human T cells and N=2-4 for assays with Jurkat clones. Statistics were calculated using Dunnett's multiple comparisons test, using Ctrl as control. **** p<0.0001, *** p<0.001, **p<0.01, *p<0.05

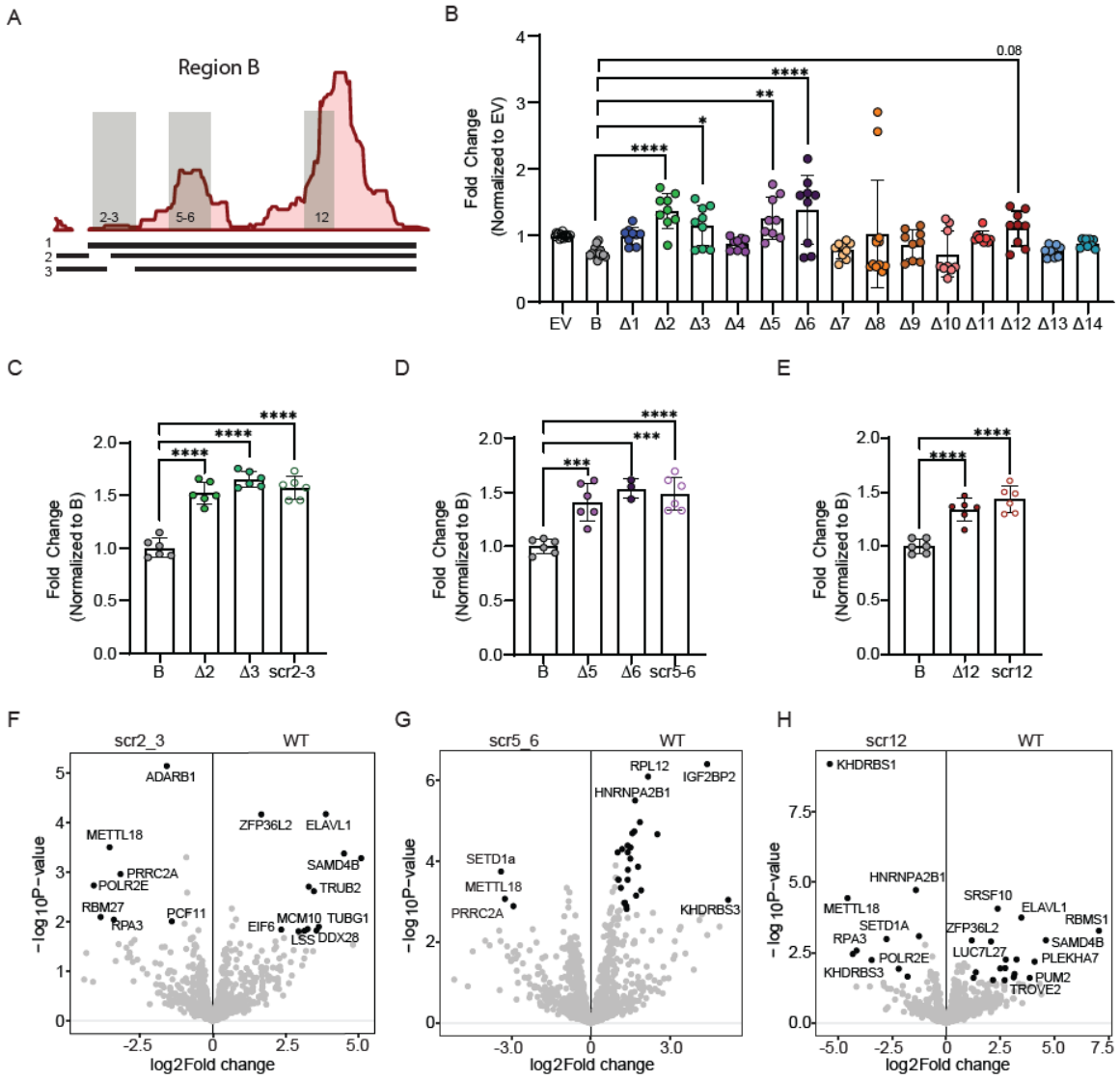


Figure 3.5. Fine-mapping reveals destabilizing activity in RBP binding sites.

(A) GCLiPP read coverage of Region B in *CD69* 3'UTR depicting RBP binding peaks. Lines below are examples of constructs with 20bp deletion spanning the region. Gray bars indicate sites with destabilizing activity. **(B)** Dual luciferase reporter expression for empty vector (EV), construct with intact Region B (B) and constructs with deletions cross Region B(Δ). Expression is normalized to EV. Data is compiled from 2-3 individual experiments. Reporter expression from constructs with deletions and scrambled sequences in section 2-3 **(C)**, 5-6 **(D)** and 12 **(E)** were normalized to construct B. **(F-H)** Volcano plots represent RBPs quantified by mass

spectrometry from pull down with Jurkat lysates using 4xS1m RNA aptamer expressing WT, scr2_3 **(F)** scr 5-6 **(G)** or scr 12 **(H)** as shown in **(C-E)**. Relative abundance of proteins between scrambled and WT constructs was calculated using raw protein intensity values expressed as log2 fold change. Statistics were calculated using Dunnett's multiple comparisons test, using Ctrl as control. **** p<0.0001, *** p<0.001, **p<0.01, *p<0.05

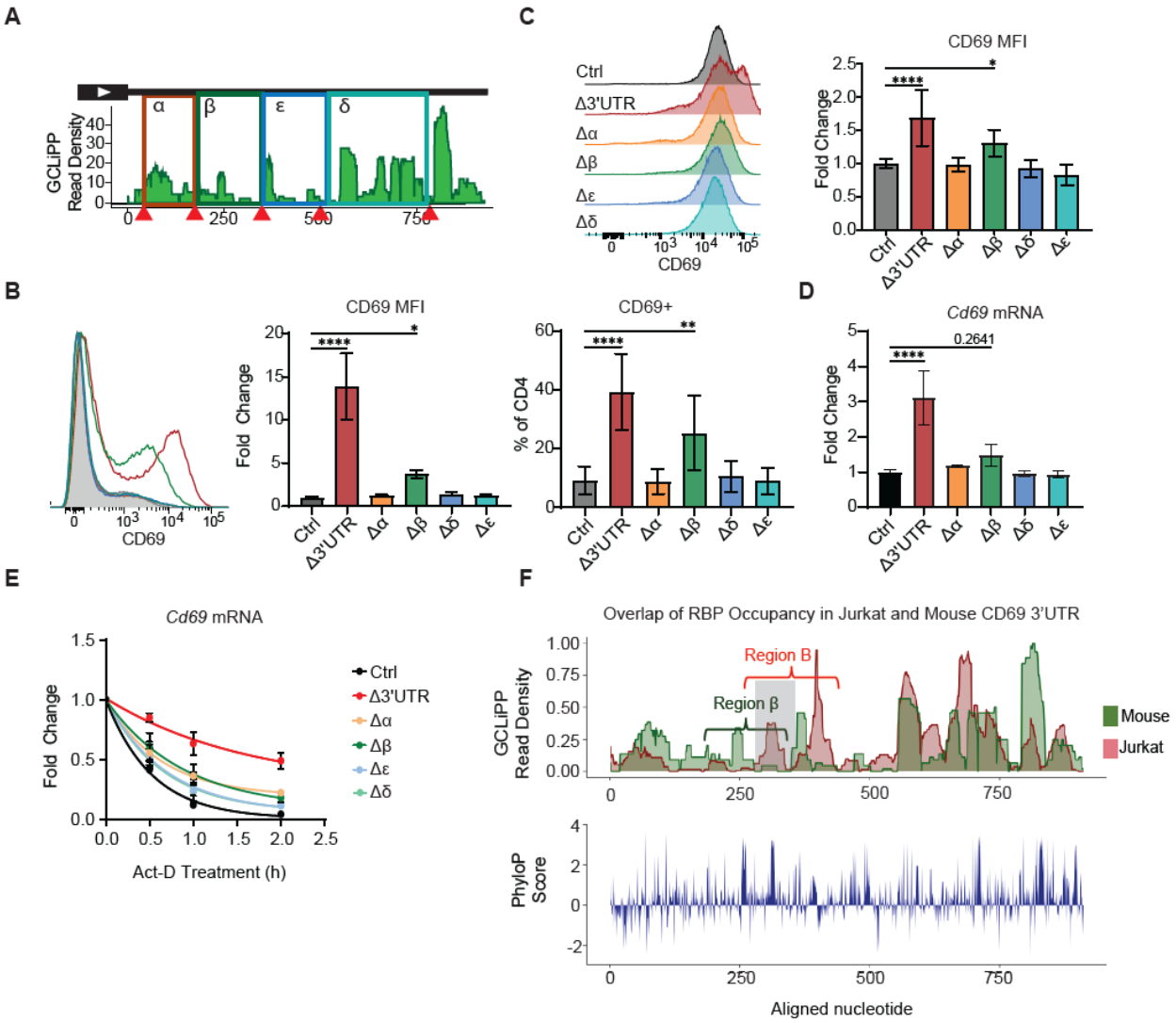


Figure 3.6. Conserved destabilizing activity in *Cd69* 3'UTR between mouse and human.

(A) Schematic of mouse *Cd69* 3'UTR with predicted RBP binding peaks represented by GCLiPP read density. Red arrows indicate gRNA placement and boxes show deleted region. **(B)** CD69 expression in resting primary mouse CD4 T cells shown as a representative flow cytometry plot (left). CD69 MFI and percentage of CD69+ CD4 cells is quantified on the right. MFI was normalized to average Ctrl for each independent experiment. **(C)** CD69 expression in primary mouse T cells activated for 4 hours with PMA and Ionomycin shown in representative flow cytometry plot (left) and normalized MFI values (right). **(D)** *Cd69* mRNA abundance in activated mouse CD4 T cells. After activation, mRNA abundance was measured during treatment with act-D in **(E). (F)** GCLiPP read coverage of human and mouse *CD69* 3'UTR were aligned by sequence. RBP occupancy is plotted on top, and sequence conservation is indicated

below by the PhyloP score. Green represents mouse and red represents human RBP binding occupancy. Brackets show region with mapped destabilizing activity. Statistics were calculated using Dunnett's multiple comparisons test, using Ctrl as control. **** p<0.0001, *** p<0.001, **p<0.01, *p<0.05

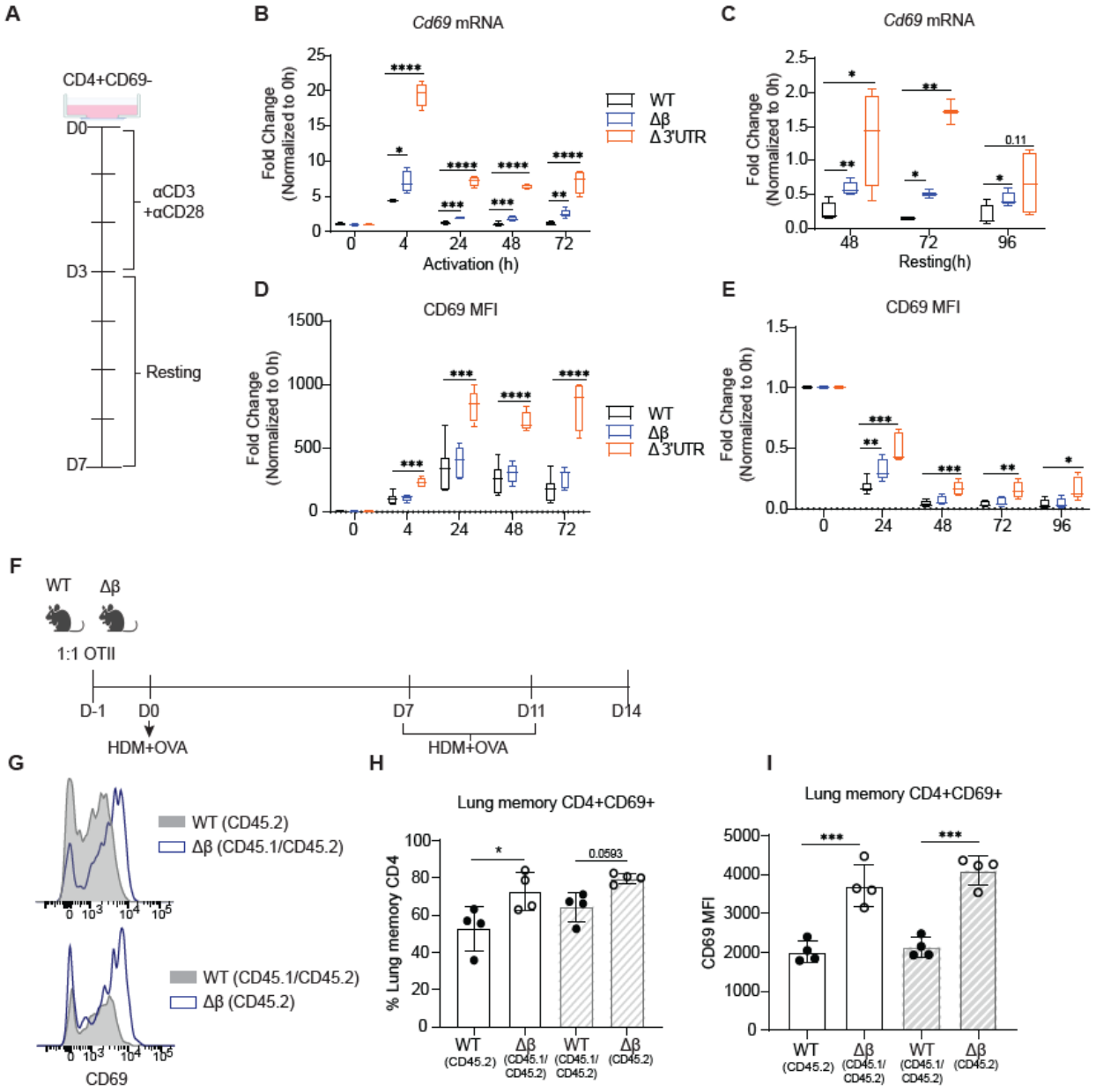
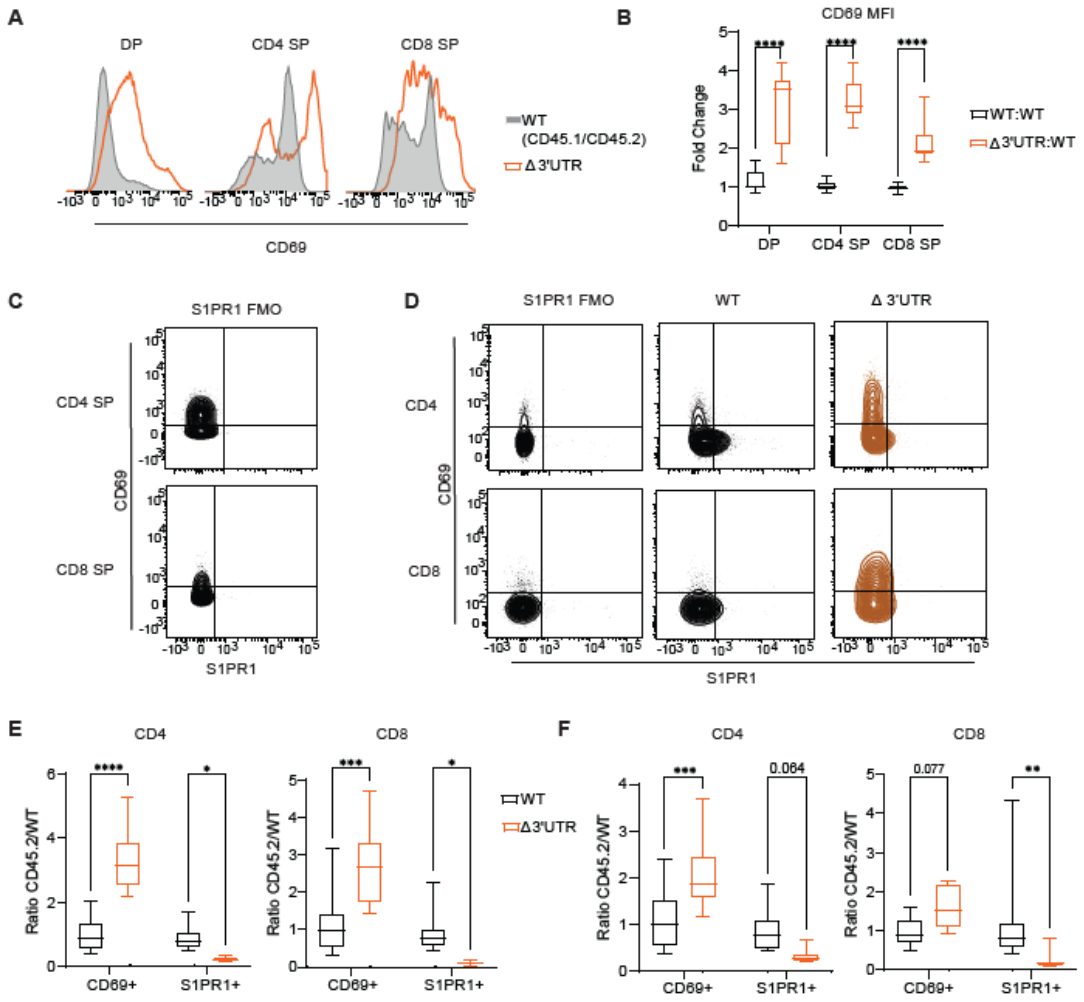


Figure 3.7. Region β contains destabilizing elements that regulate CD69 expression in T cells.

(A) Diagram showing timeline of CD4 T cell cultures. Primary mouse CD69- CD4 cells were activated with anti-CD3 (αCD3) and anti-CD28 (αCD28) for 3 days and rested for 4 days afterwards (N=4-6). *Cd69* mRNA transcript expression was measure during activation **(B)** and during rest **(C)**. Values were normalized to housekeeping gene *B2m* and then to unstimulated cells (0h activation timepoint). Similar normalization was performed for CD69 MFI in activated

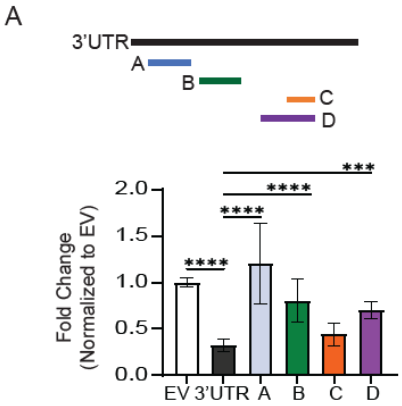
(D) and resting cells **(E)**. **(F)** Allergen sensitization and challenge model. CD69- naïve CD4 T cells from WT-OTII and Δβ-OTII mice were adoptively transferred at 1:1 ratio into WT recipients. Mice were sensitized with HDM and OVA at day 0, challenged 5 times from day 7 to day 11 and euthanized on day 14 for analysis. **(G)** CD69 expression of lung memory CD4 T cells (CD45*iv*-CD4+CD44+CD62L-) from WT-OTII or Δβ-OTII donors. Grey histogram represents WT and blue lines shows Δβ cells on different congenic backgrounds. Percentage of CD69+ lung memory cells shown in **(H)**. **(I)** CD69 MFI in CD69+ lung memory T cells. Open bars represent one mixture with CD45.2 WT-OTII and CD45.1/CD45.2 Δβ-OTII, while striped bars represent mixture with CD45.1/CD45.2 WT-OTII and CD45.2 Δβ-OTII. Šídák multiple comparisons tests or student's t-test was used to calculate significance. **** p<0.0001, *** p<0.001, **p<0.01, *p<0.05



Supplementary Figure 3.1. CD69 and S1PR1 expression on thymocytes in bone marrow chimeras.

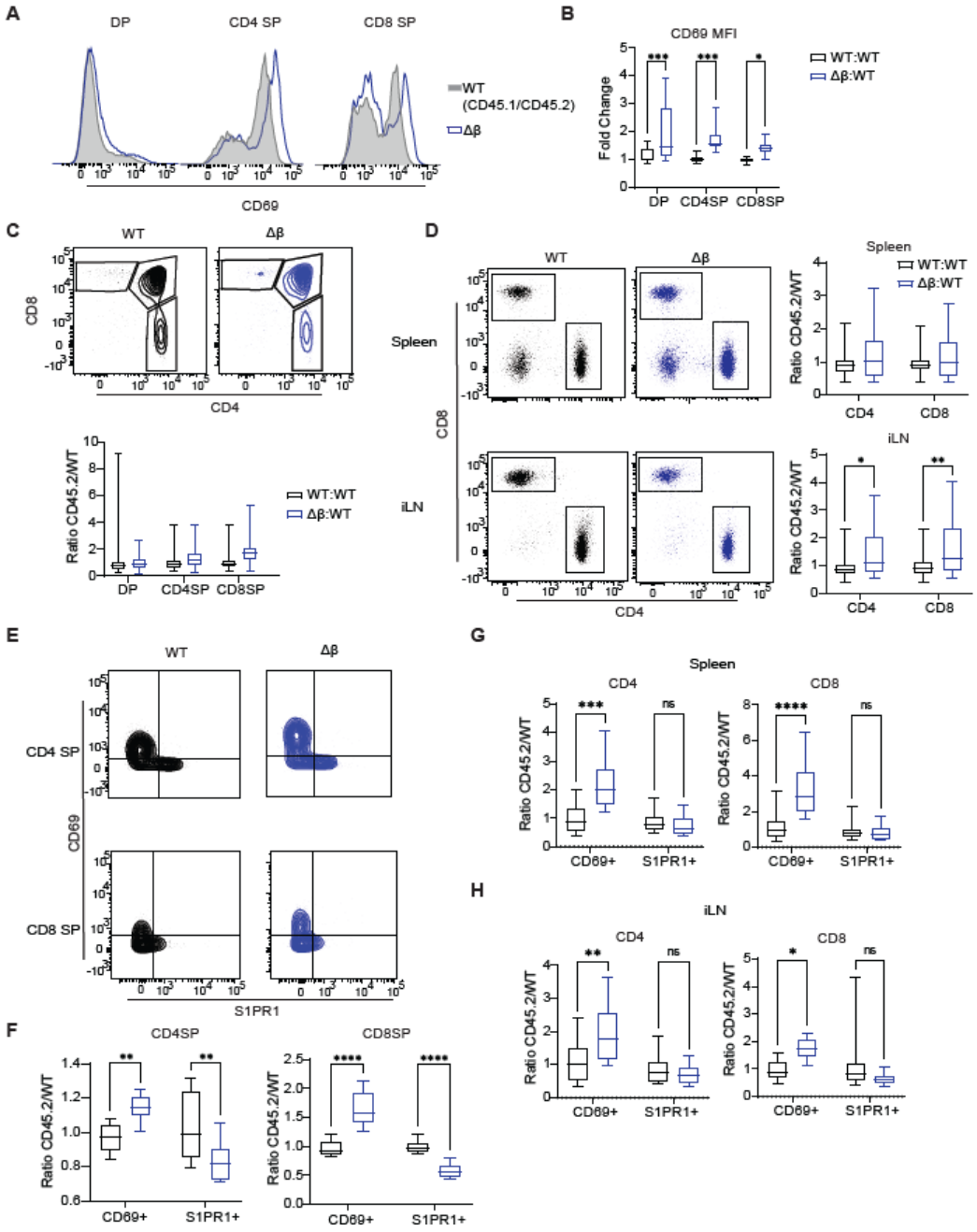
(A) CD69 expression on thymocyte populations from WT (CD45.1/.CD45.2) and Δ3'UTR (CD45.2) in WT: Δ3'UTR chimeras. **(B)** Plot showing CD69 MFI for WT:WT or Δ3'UTR:WT chimeras in DP and SP thymocytes calculated as fold change. **(C)** S1PR1 FMO samples were used to set S1PR1 gate. **(D)** Representative plots of CD69+ and S1PR1+ CD4 and CD8 T cells in iLNs. Ratio of WT:WT and Δ3'UTR:WT cells that are CD69+ or S1PR1+ quantified in **(E)**.

Same ratios for spleen in **(F).** Black represents WT:WT and orange represents Δ3'UTR:WT. Šídák multiple comparisons tests was used to calculate significance. **** p<0.0001, *** p<0.001, **p<0.01, *p<0.05



Supplementary Figure 3.2. Dual luciferase reporter expression of fragments in *CD69* 3'UTR.

Fragments of *CD69* 3'UTR were cloned into dual luciferase reporter. Expression was normalized to empty vector (EV). Dunnett's multiple comparison test was performed. Statistic was calculated using Dunnett's multiple comparisons test, using 3'UTR as control. **** p<0.0001, *** p<0.001, **p<0.01, *p<0.05



Supplementary Figure 3.3. Δβ mice are similar to WT mice at baseline.

(A) CD69 expression on WT (CD45.1/.CD45.2) and Δβ (CD45.2) thymocytes from Δβ:WT chimeras. **(B)** Fold change of CD69 MFI in WT:WT or Δβ:WT chimeras. **(C)** Thymocyte populations in WT:WT (black) or Δβ:WT (blue) chimeras. Representative flow cytometry plots on top. Ratios quantified on bottom. **(D)** Representative flow cytometry plots of CD4 and CD8 T cells in spleen (top) and iLNs (bottom) from chimeras with W:T or WT:Δβ marrow. Ratio of WT:WT and Δβ:WT within each T cell population plotted on right. CD69 and S1PR1 expression in WT (CD45.2) or Δβ (CD45.2) on CD4SP and CD8 SP thymocytes (**E-F)** as well as CD4 and CD8 T cells in spleen **(G)** and iLNs **(H)**. Ratios were calculated for CD69+ and S1PR1+ cells. Color

scheme for WT:WT and Δβ:WT are consistent throughout figure. Šídák multiple comparisons tests was used to calculate significance. **** $p<0.0001$, *** $p<0.001$, **$p<0.01$, *$p<0.05$

Tables

Supplementary Table 3.1. Human and mouse primers.

	Primers	Sequence
Mouse	Cd69 3'UTR F	GTGGACTGTGAGGCAAACTTCCACTG
	Cd69 3'UTR R	GGCAATAAATAGTAACTCTAGAGCTAGTCAG
	Cd69 F	GCTCTCATTGCCTTAAATGTGGGCAAG
	Cd69 R	GTAGCAGCATCTTCAGAACAAGAGCG
	B2m F	TGGTCTTTCTGGTGCTTGTC
	B2m R	GGGTGGAACTGTGTTACGTAG
Human	CD69 3'UTR F	TGGAATGTGAGAAGAATTTATACTGG
	CD69 3'UTR R	GAACTTTCTTGGAAACCATTGTG
	CD69 F	TTTGCATCCGGAGAGTGGAC
	CD69 R	TTGGCCCACTGATAAGGCAA
	GAPDH F	GACCACTTTGTCAAGCTCATTTC
	GAPDH R	CTCTCTTCCTCTTGTGCTCTTG

Supplementary Table 3.2. Primers for cloning into dual luciferase vector.

CD69 3'UTR Fragments	Sequence (include XhoI and Not1 sites)
Cd69 3'UTR F1	TAAGCAGCTCGAGTAAGGAAACATGTTCACTTATTGAC
Cd69 3'UTR F2	TGCTTAGGCGGCCGCCTAGTATTATTCTAGGTCTGAAG
Region A F	TAAGCAGCTCGAGTAAGGAAACATGTTCACTTATTG
Region A R	TGCTTAGGCGGCCGCGTCATTCTTCTCATTCTTGGGC
Region B F	TAAGCAGCTCGAGGAATGACTATGCAACCTTTGGATGCAC
Region B R	TGCTTAGGCGGCCGCCCCTGAGTTTAAGGGATTCAATTAG
Region C F	TAAGCAGCTCGAGGCTGTAGACAGGTCCTTTTCGATGG
Region C R	TGCTTAGGCGGCCGCCTAGTATTATTCTAGGTCTGAAG
Region D F	TAAGCAGCTCGAGGAGAAATTTGCCAATTTACTTTG
Region D R	TGCTTAGGCGGCCGCCTAGTATTATTCTAGGTCTGAAG

Supplementary Table 3.3. Primers for site directed mutagenesis.

SDM Primers	Sequence
scr2_3 F	tagtatgtattcTCCAGAAATAATGAAATAACTAG
scr2_3 R	taactaatcgaatAAGGTTGCATAGTCATTC
scr5_6 F	gatatcattgTTGCTGAATGACTACCAAC
scr5_6 R	gattgcaaccAGTTATTTCATTATTTCTGGATTC
scr12 F	ttcctaatgctagttatGACTACCAACAGTGAGAG
scr12 R	ctcgaattatgttcaaccTTCATTATTTCTGGATTCAAAATAATATAAAG

Supplementary Table 3.4. Human and mouse crRNA.

	crRNA	Sequence
Mouse	mCd69 3'UTR 1	TGACTGTGCCATAGCACCAC
	mCd69 3'UTR 2	TCCAGGAAGATCCCTTGACC
	mCd69 3'UTR 3	GTGCAAATGCGTAAAGGCAC
	mCd69 3'UTR 4	GATGGAGTTACTGCAAAGCT
	mCd69 3'UTR 5	TTATTCCAAGGTCAAACCAC
Human	CD69 3'UTR 1	CTCAAGGAAATCTGTGTCAG
	CD69 3'UTR 2	TCATTCTTGGGCATGGTTAT
	CD69 3'UTR 3	AAGAATGACTATGCAACCTT
	CD69 3'UTR 4	TAATTGAATCCCTTAAACTC
	CD69 3'UTR 5	TGTAGACAGGTCCTTTTCGA
	CD69 3'UTR 6	TATTTTTCCCCAAAAAGAAT
	CD69 3'UTR 7	TGATGTGGCAAATCTCTATT
Control	Ctrl1	GGTTCTTGACTACCGTAATT
	Ctrl2	TCGGATGTAAATTATGCCGT
	Ctrl2	CAGCTCATCGGTGTCCTACT

Supplementary Table 3.5. Primers for RNA aptamer cloning.

Fragment	Original Sequence	Scrambled Sequence
Region B 2_3	TGGATGCACTTTATATTATTttgaa	ATTCGATTAGTTATAGTATGTATTC
Region B 5_6	taaataGTAAGTCCACGCCT	GGTTGCAATCGATATCATTG
Region B 12	gaatgtaaacaaaggAATTA	AagAgataagAtacaTagaT

References

Allende, M. L., Dreier, J. L., Mandala, S., & Proia, R. L. (2004). Expression of the Sphingosine 1-Phosphate Receptor, S1P1, on T-cells Controls Thymic Emigration *. *Journal of Biological Chemistry*, *279*(15), 15396–15401. https://doi.org/10.1074/JBC.M314291200

Arnon, T. I., Ying, X., Charles, L., Trung, P., Jinping, A., Shaun, C., Dorn, G. W., & Cyster, J. G. (2011). GRK2-Dependent S1PR1 Desensitization Is Required for Lymphocytes to Overcome Their Attraction to Blood. *Science*, *333*(6051), 1898–1903. https://doi.org/10.1126/science.1208248

Bankovich, A. J., Shiow, L. R., & Cyster, J. G. (2010). CD69 suppresses sphingosine 1-phosophate receptor-1 (S1P1) function through interaction with membrane helix 4. *The Journal of Biological Chemistry*, *285*(29), 22328–22337. https://doi.org/10.1074/jbc.M110.123299

Beecham, A. H., Patsopoulos, N. A., Xifara, D. K., Davis, M. F., Kemppinen, A., Cotsapas, C., Shah, T. S., Spencer, C., Booth, D., Goris, A., Oturai, A., Saarela, J., Fontaine, B., Hemmer, B., Martin, C., Zipp, F., D'Alfonso, S., Martinelli-Boneschi, F., Taylor, B., … (IIBDGC), I. I. B. D. G. C. (2013). Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nature Genetics*, *45*(11), 1353–1360. https://doi.org/10.1038/ng.2770

Castello, A., Fischer, B., Frese, C. K., Horos, R., Alleaume, A. M., Foehr, S., Curk, T., Krijgsveld, J., & Hentze, M. W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular Cell*, *63*(4), 696–710. https://doi.org/10.1016/J.MOLCEL.2016.06.029/ATTACHMENT/B75C1C3A-2C6A-438E-8F2C-91C43BF8366C/MMC3.XLSX

Chen, C. Y. A., & Shyu, A. bin. (1995). AU-rich elements: characterization and importance in mRNA degradation. *Trends in Biochemical Sciences*, *20*(11), 465–470. https://doi.org/10.1016/S0968-0004(00)89102-1

Chen, J., Cascio, J., Magee, J. D., Techasintana, P., Gubin, M. M., Dahm, G. M., Calaluce, R., Yu, S., & Atasoy, U. (2013). Posttranscriptional Gene Regulation of IL-17 by the RNA-Binding Protein HuR Is Required for Initiation of Experimental Autoimmune Encephalomyelitis. *The Journal of Immunology*, *191*(11), 5441–5450. https://doi.org/10.4049/JIMMUNOL.1301188

Chun, J., & Hartung, H.-P. (2010). Mechanism of Action of Oral Fingolimod (FTY720) in Multiple Sclerosis. *Clinical Neuropharmacology*, *33*(2). https://journals.lww.com/clinicalneuropharm/Fulltext/2010/03000/Mechanism_of_Action_of_Oral_Fingolimod__FTY720__in.7.aspx

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, *26*(12), 1367–1372. https://doi.org/10.1038/nbt.1511

Farh, K. K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., Shoresh, N., Whitton, H., Ryan, R. J. H., Shishkin, A. A., Hatan, M., Carrasco-Alfonso, M. J., Mayer, D., Luckey, C. J., Patsopoulos, N. A., de Jager, P. L., Kuchroo, V. K., Epstein, C. B., Daly, M. J., … Bernstein, B. E. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, *518*(7539), 337–343. https://doi.org/10.1038/nature13835

Feng, C., Woodside, K. J., Vance, B. A., El-Khoury, D., Canelles, M., Lee, J., Gress, R., Fowlkes, B. J., Shores, E. W., & Love, P. E. (2002). A potential role for CD69 in thymocyte emigration. *International Immunology*, *14*(6), 535–544. https://doi.org/10.1093/intimm/dxf020

Fonseca, R., Beura, L. K., Quarnstrom, C. F., Ghoneim, H. E., Fan, Y., Zebley, C. C., Scott, M. C., Fares-Frederickson, N. J., Wijeyesinghe, S., Thompson, E. A., Borges da Silva, H., Vezys, V., Youngblood, B., & Masopust, D. (2020). Developmental plasticity allows outside-in immune responses by resident memory T cells. *Nature Immunology 2020 21:4*, *21*(4), 412–421. https://doi.org/10.1038/s41590-020-0607-7

Gebauer, F., Schwarzl, T., Valcárcel, J., & Hentze, M. W. (2021). RNA-binding proteins in human genetic disease. *Nature Reviews Genetics*, *22*(3), 185–198. https://doi.org/10.1038/s41576-020-00302-y

Gerstberger, S., Hafner, M., Ascano, M., & Tuschl, T. (2014). Evolutionary Conservation and Expression of Human RNA-Binding Proteins and Their Role in Human Genetic Disease. In G. W. Yeo (Ed.), *Systems Biology of RNA Binding Proteins* (pp. 1–55). Springer New York. https://doi.org/10.1007/978-1-4939-1221-6_1

Hoefig, K. P., Reim, A., Gallus, C., Wong, E. H., Behrens, G., Conrad, C., Xu, M., Kifinger, L., Ito-Kureha, T., Defourny, K. A. Y., Geerlof, A., Mautner, J., Hauck, S. M., Baumjohann, D., Feederle, R., Mann, M., Wierer, M., Glasmacher, E., & Heissmeyer, V. (2021). Defining the RBPome of primary T helper cells to elucidate higher-order Roquin-mediated mRNA regulation. *Nature Communications*, *12*(1), 5208. https://doi.org/10.1038/s41467-021-25345-5

Huang, H., Weng, H., Sun, W., Qin, X., Shi, H., Wu, H., Zhao, B. S., Mesquita, A., Liu, C., Yuan, C. L., Hu, Y.-C., Hüttelmaier, S., Skibbe, J. R., Su, R., Deng, X., Dong, L., Sun, M., Li, C., Nachtergaele, S., … Chen, J. (2018). Recognition of RNA N6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nature Cell Biology*, *20*(3), 285–295. https://doi.org/10.1038/s41556-018-0045-z

Jurgens, A. P., Popović, B., & Wolkers, M. C. (2021). T cells at work: How post-transcriptional mechanisms control T cell homeostasis and activation. *European Journal of Immunology*, *51*(9), 2178–2187. https://doi.org/10.1002/EJI.202049055

Kimura, F. Z., Didierlaurent, A., Koseki, H., Yamashita, S., Schubert, L. A., Gillard, G., Nakayama, T., & Kasprowicz, D. J. (2002). Blockade or Overexpression CD69 Thymocytes In Vivo Is Dysregulated by The Generation of Mature, Single-Positive. *The Journal of Immunology*, *168*, 87–94. https://doi.org/10.4049/jimmunol.168.1.87

Laguna, T., Notario, L., Pippa, R., Fontela, M. G., Vázquez, B. N., Maicas, M., Aguilera-Montilla, N., Corbí, Á. L., Odero, M. D., & Lauzurica, P. (2015). New insights on the transcriptional regulation of CD69 gene through a potent enhancer located in the conserved non-coding sequence 2. *Molecular Immunology*, *66*(2), 171–179. https://doi.org/10.1016/j.molimm.2015.02.031

Leppek, K., Schott, J., Reitter, S., Poetz, F., Hammond, M. C., & Stoecklin, G. (2013). Roquin Promotes Constitutive mRNA Decay via a Conserved Class of Stem-Loop Recognition Motifs. *Cell*, *153*(4), 869–881. https://doi.org/10.1016/J.CELL.2013.04.016

Leppek, K., & Stoecklin, G. (2014). An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Research*, *42*(1), e13. https://doi.org/10.1093/nar/gkt956

Litterman, A. J., Zhu, W. S., Kageyama, R., Zhao, W., Zaitlen, N., Erle, D. J., & Ansel, K. M. (2018). A global map of RNA binding protein occupancy guides functional dissection of post-transcriptional regulation of the T cell transcriptome. *BioRxiv*, 448654. https://doi.org/10.1101/448654

López-Cabrera, M., Muñoz, E., Blázquez, M. V., Ursa, M. A., Santis, A. G., & Sánchez-Madrid, F. (1995). Transcriptional Regulation of the Gene Encoding the Human C-type Lectin Leukocyte Receptor AIM/CD69 and Functional Characterization of Its Tumor Necrosis Factor-α-responsive Elements (∗). *Journal of Biological Chemistry*, *270*(37), 21545–21551. https://doi.org/https://doi.org/10.1074/jbc.270.37.21545

Mackay, L. K., Braun, A., Macleod, B. L., Collins, N., Tebartz, C., Bedoui, S., Carbone, F. R., & Gebhardt, T. (2015). Cutting Edge: CD69 Interference with Sphingosine-1-Phosphate Receptor Function Regulates Peripheral T Cell Retention. *The Journal of Immunology*, *194*(5), 2059–2063. https://doi.org/10.4049/jimmunol.1402256

MacKay, L. K., Rahimpour, A., Ma, J. Z., Collins, N., Stock, A. T., Hafon, M. L., Vega-Ramos, J., Lauzurica, P., Mueller, S. N., Stefanovic, T., Tscharke, D. C., Heath, W. R., Inouye, M.,

Carbone, F. R., & Gebhardt, T. (2013). The developmental pathway for CD103+ CD8+ tissue-resident memory T cells of skin. *Nature Immunology*, *14*(12), 1294–1301. https://doi.org/10.1038/ni.2744

Matloubian, M., Lo, C. G., Cinamon, G., Lesneski, M. J., Xu, Y., Brinkmann, V., Allende, M. L., Proia, R. L., & Cyster, J. G. (2004). Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. *Nature*, *427*(6972), 355–360. https://doi.org/10.1038/nature02284

Miki-Hosokawa, T., Hasegawa, A., Iwamura, C., Shinoda, K., Tofukuji, S., Watanabe, Y., Hosokawa, H., Motohashi, S., Hashimoto, K., Shirai, M., Yamashita, M., & Nakayama, T. (2009). CD69 controls the pathogenesis of allergic airway inflammation. *Journal of Immunology (Baltimore, Md. : 1950)*, *183*(12), 8203–8215. https://doi.org/10.4049/jimmunol.0900646

Mueller, S. N., Gebhardt, T., Carbone, F. R., & Heath, W. R. (2013). Memory T Cell Subsets, Migration Patterns, and Tissue Residence. *Annual Review of Immunology*, *31*(1), 137–161. https://doi.org/10.1146/annurev-immunol-032712-095954

Onengut-Gumuscu, S., Chen, W.-M., Burren, O., Cooper, N. J., Quinlan, A. R., Mychaleckyj, J. C., Farber, E., Bonnie, J. K., Szpak, M., Schofield, E., Achuthan, P., Guo, H., Fortune, M. D., Stevens, H., Walker, N. M., Ward, L. D., Kundaje, A., Kellis, M., Daly, M. J., … Consortium, T. 1 D. G. (2015). Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nature Genetics*, *47*(4), 381–386. https://doi.org/10.1038/ng.3245

Perez-Perri, J. I., Rogell, B., Schwarzl, T., Stein, F., Zhou, Y., Rettel, M., Brosig, A., & Hentze, M. W. (2018). Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nature Communications*, *9*(1), 4408. https://doi.org/10.1038/s41467-018-06557-8

Salerno, F., Engels, S., van den Biggelaar, M., van Alphen, F. P. J., Guislain, A., Zhao, W., Hodge, D. L., Bell, S. E., Medema, J. P., von Lindern, M., Turner, M., Young, H. A., & Wolkers, M. C. (2018). Translational repression of pre-formed cytokine-encoding mRNA prevents chronic activation of memory T cells. *Nature Immunology*, *19*(8), 828–837. https://doi.org/10.1038/s41590-018-0155-6

Santis, A. G., López-Cabrera, M., Sánchez-Madrid, F., & Proudfoot, N. (1995). Expression of the early lymphocyte activation antigen CD69, a C-type lectin, is regulated by mRNA degradation associated with AU-rich sequence motifs. *European Journal of Immunology*, *25*(8), 2142–2146. https://doi.org/10.1002/eji.1830250804

Schlundt, A., Heinz, G. A., Janowski, R., Geerlof, A., Stehle, R., Heissmeyer, V., Niessing, D., & Sattler, M. (2014). Structural basis for RNA recognition in roquin-mediated post-transcriptional gene regulation. *Nature Structural & Molecular Biology 2014 21:8*, *21*(8), 671–678. https://doi.org/10.1038/nsmb.2855

Shiow, L. R., Rosen, D. B., Brdičková, N., Xu, Y., An, J., Lanier, L. L., Cyster, J. G., & Matloubian, M. (2006). CD69 acts downstream of interferon-α/β to inhibit S1P1 and lymphocyte egress from lymphoid organs. *Nature*, *440*(7083), 540–544. https://doi.org/10.1038/nature04606

Steri, M., Orrù, V., Idda, M. L., Pitzalis, M., Pala, M., Zara, I., Sidore, C., Faà, V., Floris, M., Deiana, M., Asunis, I., Porcu, E., Mulas, A., Piras, M. G., Lobina, M., Lai, S., Marongiu, M., Serra, V., Marongiu, M., … Cucca, F. (2017). Overexpression of the Cytokine BAFF and Autoimmunity Risk. *New England Journal of Medicine*, *376*(17), 1615–1626. https://doi.org/10.1056/NEJMOA1610528/SUPPL_FILE/NEJMOA1610528_DISCLOSURES.PDF

Takamura, S., Yagi, H., Hakata, Y., Motozono, C., McMaster, S. R., Masumoto, T., Fujisawa, M., Chikaishi, T., Komeda, J., Itoh, J., Umemura, M., Kyusai, A., Tomura, M., Nakayama, T., Woodland, D. L., Kohlmeier, J. E., & Miyazawa, M. (2016). Specific niches for lung-

resident memory CD8+ T cells at the site of tissue regeneration enable CD69-independent maintenance. *The Journal of Experimental Medicine*, *213*(13), 3057–3073. https://doi.org/10.1084/jem.20160938

Tan, D., Zhou, M., Kiledjian, M., & Tong, L. (2014). The ROQ domain of Roquin recognizes mRNA constitutive-decay element and double-stranded RNA. *Nature Structural & Molecular Biology 2014 21:8*, *21*(8), 679–685. https://doi.org/10.1038/nsmb.2857

Taylor, K. E., Ansel, K. M., Marson, A., Criswell, L. A., & Farh, K. K.-H. (2021). PICS2: next-generation fine mapping via probabilistic identification of causal SNPs. *Bioinformatics*. https://doi.org/10.1093/BIOINFORMATICS/BTAB122

Techasintana, P., Ellis, J. S., Glascock, J., Gubin, M. M., Ridenhour, S. E., Magee, J. D., Hart, M. L., Yao, P., Zhou, H., Whitney, M. S., Franklin, C. L., Martindale, J. L., Gorospe, M., Davis, W. J., Fox, P. L., Li, X., & Atasoy, U. (2017). The RNA-Binding Protein HuR Posttranscriptionally Regulates IL-2 Homeostasis and CD4 + Th2 Differentiation . *ImmunoHorizons*, *1*(6), 109–123. https://doi.org/10.4049/IMMUNOHORIZONS.1700017/-/DCSUPPLEMENTAL

Testi, R., Phillips, J. H., & Lanier, L. L. (1989). Leu 23 induction as an early marker of functional CD3/T cell antigen receptor triggering. Requirement for receptor cross-linking, prolonged elevation of intracellular [Ca++] and stimulation of protein kinase C. *The Journal of Immunology*, *142*(6), 1854. http://www.jimmunol.org/content/142/6/1854.abstract

Turner, M., & DÍaz-Muñoz, M. D. (2018). RNA-binding proteins control gene expression and cell fate in the immune system review-article. In *Nature Immunology* (Vol. 19, Issue 2, pp. 120–129). Nature Publishing Group. https://doi.org/10.1038/s41590-017-0028-4

Vazquez, B. N., Laguna, T., Carabana, J., Krangel, M. S., & Lauzurica, P. (2009). CD69 gene is differentially regulated in T and B cells by evolutionarily conserved promoter-distal elements. *Journal of Immunology (Baltimore, Md. : 1950)*, *183*(10), 6513–6521. https://doi.org/10.4049/jimmunol.0900839

Vernet, C., & Artzt, K. (1997). STAR, a gene family involved in signal transduction and activation of RNA. *Trends in Genetics*, *13*(12), 479–484. https://doi.org/https://doi.org/10.1016/S0168-9525(97)01269-9

White, E. K., Moore-Jarrett, T., & Ruley, H. E. (2001). PUM2, a novel murine puf protein, and its consensus RNA-binding site. *RNA*, *7*(12), 1855. /pmc/articles/PMC1370223/?report=abstract

Yamashita, I., Nagata, T., Tada, T., & Nakayama, T. (1993). CD69 cell surface expression identifies developing thymocytes which audition for T cell antigen receptor-mediated positive selection. *International Immunology*, *5*(9), 1139–1150. https://doi.org/10.1093/INTIMM/5.9.1139

Zhang, X., Smits, A. H., van Tilburg, G. B. A., Ovaa, H., Huber, W., & Vermeulen, M. (2018). Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nature Protocols*, *13*(3), 530–550. https://doi.org/10.1038/nprot.2017.147

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Wandi Zhu
CBA3491981004F5...
Author Signature

7/24/2022
Date