Title

Dynamic gene networks reveal regulatory mechanisms orchestrating T cell function

Author

Arce, Maya

Publication Date

2025

Peer reviewed|Thesis/dissertation

Dynamic gene networks reveal regulatory mechanisms orchestrating T cell function

by
Maya Arce

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:
K. Mark Ansel
BC645044B450408...

K. Mark Ansel
Chair

DocuSigned by:
Alexander Marson
[illegible]

Alexander Marson

DocuSigned by:
Jonathan Pritchard
[illegible]

Jonathan Pritchard

DocuSigned by:
Qizhi Tang
9BFF9AD92ACD4BA...

Qizhi Tang

Committee Members

Dedication

To my parents, for their unconditional support. And to my sister, who is a constant source of inspiration and encouragement.

Acknowledgements

Thank you to my mentor, Alex, for your guidance throughout this experience, your advice has consistently redirected me at times when I have found myself going astray. I continue to marvel not only at your scientific vision, but your ability to communicate it so clearly in a way that resonates with the intended audience. I have turned to you countless times to distill my thoughts or evolve a narrative from a few pieces of information into a complete story. This time in your lab has been a tremendous opportunity to grow in these skills as well as in mindset. Your optimism has pushed me to pursue things that I would have never considered a possibility, thank you for encouraging me to continue to think bigger.

The members of my thesis committee have been particularly impactful to my experience at UCSF. Thank you, Dr. Qizhi Tang, for your thoughtful insight and inspiring vision, I continue to be a huge fan. Thank you, Dr. Mark Ansel, for the numerous forms of mentorship that you have graciously provided during my time here. I have burdened you with the bulk of my formal academic experiences, from NSF mentor to quals and thesis writing, and I am deeply appreciative of your guidance. Thank you, Dr. Jonathan Pritchard, for welcoming me as both a collaborator and mentee. I have tremendous admiration for both your scientific work and your leadership and feel fortunate to have been able to enter the orbit of your lab.

I am incredibly grateful to have been able to work with the many intelligent and helpful people that make up the Marson lab. First, Dr. Jacob Freimer, thank you for being a wonderful mentor during the first portion of my time in the Marson lab. You absolutely set me up for success and passed on a tremendous amount of knowledge during this period. Thank you for being so generous with both your ideas and time. I was also incredibly lucky to work alongside Dr. Jennifer Umhoefer, who helped me through every aspect of grad school. Jenny has acted as both my mentor and confidant and has become a better friend than I could have ever imagined finding during this time. Thank you for your positivity and your incredible generosity which saw me

through many rough patches. I have no interest in knowing what this experience would have been like without your support. Most recently, I have begun working in a small team which has brought so much excitement to the end of this experience. Thank you, Dr. Qi Liu, for collaborating enthusiastically with me at the busiest time in your postdoc. You are truly a force of nature. Thank you also to Sanjana Subramanya for joining us as a research associate and making everything easier and more pleasant, you are a wonderful addition to the team. I have been fortunate to mentor several trainees during my time in the Marson lab, who taught me many lessons and brought lots of good energy to the lab. Anika Wadhera, Jasmine Salazar, Charles Shen, and Melissa Bu, thank you for your patience and enthusiasm!

I am fortunate to have worked alongside and learned from many talented postdocs in the Marson lab. Dr. Youjin Lee, Dr. Murad Mamedov, Dr.Tori Yamamoto, Dr. Franziska Blaeschke, Dr. Ralf Schmidt, Dr. Zach Steinhart, Dr. Carl Ward, Dr. Ujjwal Rathore, Dr. Rosmely Hernandez, Dr. Mineto Ota, Dr. Hyuncheol Jung, Dr. Kristen Mengwasser, Dr. Edward Marsh, Dr. Ron Zhu, Dr. Emma Dann, Dr. Pascal Devant and Dr. Jun Yan; thank you for your wisdom and support. I am also incredibly lucky to overlapped with so many graduate students, who have been an essential support system in the lab. Thank you, Dr. Cody Mowry, Laine Goudy, Galen Xing, Eli Dugan, Jim Asaki, and Ben Lesch for your good advice and great company. Thank you to Ron Manlapaz, Jon Woo and Jackie Sawin, who keep the lab running and make it the productive environment that it is. Thank you also to the research associates and staff scientists; Christian Garrido, Viggy Kumar, Shane Vedova, Yan Yi Chen, Rama Dajani, Xinyi Feng, Carinna Tran, Hunter Thorton, Madeline Layeghi and Zhongmei Li, who are an essential source of knowledge and skills within the Marson lab.

The work in this thesis relied also on the contributions of wonderful scientists outside of the lab. My collaboration with Dr. Josh Weinstock began almost immediately after joining the lab and resulted in a rewarding partnership during the past three years, which would not have been possible without Josh's knowledge and openness to explore ideas. Thank you for patiently

answering my questions and trusting my contributions, even as a new and young member of the Marson lab. Thank you also to Dr. Alexis Battle, for your guidance at every stage of the project. Key collaborations emerged at the final stages of the MED12 story, which reinvigorated the project and helped get it across the finish line. Dr. Ansuman Satpathy provided insightful feedback in the nascent stages of this project (as a member of my quals committee) through the final version of the manuscript. I am deeply grateful for his contributions as well those of Dr. Siva Kasinathan in his lab, who generously provided time and resources to help us overcome technical challenges with incredible thoughtfulness. Likewise, I am thankful for Dr. Nevan Krogan's enthusiastic support of our efforts. Dr. Nadia Arang appeared from the Krogan lab, seemingly out of the blue, to execute experiments that I could only dream of. Thank you for demystifying proteomics, carrying the weight of the experimental burden, and maintaining an incredibly high standard. Your scientific contributions and friendship have gone a long way.

I am a proud Santa Clara University alumnus and will remain forever grateful for the faculty and staff there. Dr. Eric Tillman welcomed me into his organic chemistry lab my sophomore year and I owe him for every research experience I have had since then. He positioned me to start building my scientific resume before I even understood what is required to pursue a career in science. In my time before graduate school, I had the opportunity to work with many wonderful people who encouraged me to advance my career. Notably, Dr. Paul Van Hummelen and John Bell were instrumental in my ability to establish a path to graduate school. Thank you for your integrity, encouragement, and humor. Thank you also to the lovely ladies of Bristol Myers Squibb who helped me revision a future for myself in science.

My graduate school cohort is full of talented and kind people whom I am lucky to have shared this experience with. In particular, thank you to my quarantine circle; Grace Hernandez, Juliana Sucharov Costa, Laura Dwyer, and Maya Lopez-Ichikawa, for your support every step of the way. Our group chat and park dates saw me through the start of grad school in a pandemic and your encouragement continues to lift me up. Despite difficult times, I have been confident in

my decision to attend UCSF because of both the people and the place. So, thank you also to the city of San Francisco, for making me feel at home like no other place has before. You have provided both the peace to reflect and recharge and the chaos to become re-energized for challenges ahead.

Most importantly, thank you to my family, who this experience would not be possible without. My parents, Karen and Carlos Arce, have supported each of my aspirations in every way possible. Thank you for providing me the freedom to explore my interest in science and financially supporting me to pursue a career in a system that remains inaccessible to many. I feel so privileged to have been able to enjoy not only my education but poorly paid internships and summer research positions thanks to the security you provided. Mom, thank you for your sincere insistence that I should prioritize my happiness above anything else. Dad, thank you for encouraging me to challenge myself and resist self-imposed limitations. Your own careers have inspired me on the path I have taken. Ven, thank you for not outgrowing your older sister. You have always been my cheerleader and accomplice, thank you for your kindness, empathy and humor throughout my many ups and downs. Your arrival in San Francisco was a tremendous surprise and the return of family dinners has been a huge source of joy. My couch and kitchen are always open to you. Thank you also to my grandparents, Gloria and Carlos Arce, who have cheered me on all the way. Who knows where we would be without your prayers.

Finally, to my partner Tim, thank you for your endless care and patience. You have brought joy, excitement, and calm to this experience in so many unexpected ways and provided more support than I could ever ask for. Thank you for every ride home from lab and meal you prepared, thank you for relocating and the hundreds of other sacrifices you have made to accommodate grad school life. It feels impossible to reflect on this experience and not recount the evolution of our relationship, which have been entangled from the start. So, having marked our third Gladstone holiday party together, I can say that this experience has been undeniably better because of you and that more than anything else, I am excited about our life together.

Contributions

Author Contributions

The studies within this dissertation were performed under the supervision and guidance of Dr. Alexander Marson, MD, PhD. Critical feedback and guidance were provided by my thesis committee members Dr. K. Mark Ansel, PhD; Dr. Qizhi Tang, PhD; and Dr. Jonathan Pritchard, PhD. Dr. Ansuman Satpathy, MD, PhD also provided significant feedback at several stages of experimentation. Dr. Jacob Freimer, PhD guided the initiation of the projects described here.

Chapter 2 contains the contents of **Arce MM**, Umhoefer JM, Arang N, Kasinathan S, Freimer JW, Steinhart Z, Shen H, Ota M, Wadhera A, Pham MTN, Dajani R, Dorovskyi D, Chen YY, Liu Q, Shy BR, Carnevale J, Satpathy AT, Krogan NJ, Pritchard JK, Marson A. Central control of dynamic gene circuits governs T cell rest and activation. *Nature*. *Accepted*. This work was made possible by the contributions of many individuals, from multiple labs across UCSF, Stanford, and Gladstone Institutes. The contributions of each author are as follows; M.M.A, J.W.F, J.K.P and A.M. conceptualized the study. M.M.A, J.M.U, N.A, S.K, Z.S, A.W, M.T.N.P, H.S, R.D, D.D, Y.Y.C, Q.L, and D.L.S performed the experiments and generated essential reagents. M.M.A, S.K, N.A performed computational analysis and data visualization. M.O, Z.S, Y.Z provided statistical and computational direction. J.M.U, J.W.F, Z.S, K.O, A.T.S, J.C, B.R.S, N.J.K and J.K.P. helped design assays and interpret the results. M.M.A and A.M. wrote the manuscript with input from all the authors.

Chapter 3 contains an abridged version of Weinstock JS*, **Arce MM***, Freimer JW, Ota M, Marson A, Battle A, Pritchard JK. (2024) Gene regulatory network inference from CRISPR perturbations in primary CD4+ T cells elucidates the genomic basis of immune disease. *Cell Genomics*. The manuscript was co-authored by myself and Josh Weinstock, who created the models described in the work. **Chapter 3** has been structured to highlight the biological findings that resulted from

the application of the models, which are described in detail in the original manuscript. Contributions by all authors are as follows: formal analysis: J.S.W, M.A., M.O. Investigation: J.S.W, M.A., Supervision: J.K.P, A.B., A.M., Funding and acquisition: J.K.P, A.B., A.M., Experimental work: M.A., J.W.F. Writing – original draft preparation: J.S.W, M.A. Writing – review and editing: J.K.P, A.B., A.M., M.O., J.W.F.

Dynamic gene networks reveal regulatory mechanisms orchestrating T cell function

Maya Arce

Abstract

The immune system is complex, dynamic, and absolutely critical to maintain human health. The immune system comprises an enormous breadth of cell types, each able to respond to numerous extracellular environments. We have lacked adequate tools to dissect the mechanisms maintaining immune homeostasis through regulation of diverse cellular identities and stimulation responses. Human genetics has revealed key genes required to prevent immune dysfunction and many genomic loci associated with immune disease. However, natural constraint, statistical power limitations, and lack of causal data hinder genetic association studies, preventing the identification of all key regulators and limiting conclusions about the relationships between regulatory genes. Experimental manipulation of gene expression with CRISPR provides the ability to investigate the structure and function of immune regulatory systems without the inherent limitations of association studies.

In the T cell compartment, distinct lineages must respond to diverse signals to mount effective immune responses and maintain homeostasis, but the dynamic regulatory circuits that respond to extracellular cues in primary human cells remain poorly defined. To reveal the regulators of a core immune gene, *IL2RA*, expressed dynamically across the T cell compartment and required to prevent immune disease, we applied pooled CRISPR KO screens across cellular contexts. We defined critical context specific regulators of IL2RA expression as well as regulators that affect the overall rest and activation state of the cell. One regulator in particular, MED12, coordinated gene regulatory networks required to maintain both T cell rest and activation. CRISPR ablation of MED12 blunted the cell state transitions between rest and activation and protected from

activation-induced cell death, revealing a previously unappreciated gene regulatory mechanism governing T cell activity.

The effects of genetic variation on complex traits act mainly through changes in gene regulation. Although many genetic variants have been linked to target genes in cis, the trans-regulatory cascade mediating their effects remains largely uncharacterized. In a separate study, we investigated the function of and relationship between transcription factors associated with immune disease which are categorized as inborn errors of immunity (IEI) genes. We formed a large regulatory network consisting of regulators of IL2RA, IEI genes, and transcription factors without known immune disease associations. These connections revealed shared paths and novel regulatory nodes that enable control over specific immune traits.

As a whole, this thesis work uses gene perturbation in distinct human T cell populations under different conditions to discover critical mechanisms regulating cell type and state specific gene expression. Detailed gene regulatory maps resulting from these studies, coupled with functional immune assays and biochemical data, provide new insights into human genetic variants linked to immune dysfunction and have potential to predict new modification strategies that enhance immunotherapies for the betterment of human health.

Table of Contents

List of Figures

Chapter 2 Figures

Chapter 3 Figures

List of Tables

Chapter 1 Introduction

A dynamic immune system is required to maintain human health

The immune system as a whole is a dynamic entity, which must constantly adapt to protect against foreign pathogens and maintain self-tolerance. Failures within this system are therefore highly consequential and can result in the development of disease including chronic autoimmune disease, immunodeficiencies or cancer. Diverse cell types participate in the ebb and flow of immune responses and breakdowns can occur as the result of imbalanced proportions of distinct cell types or changes in cellular activity. Immunology research has resulted in the discovery and characterization of countless proteins that are required for the establishment of cellular identity or function within the immune system. This knowledge has resulted in improved diagnosis, treatment, and prevention of disease and continuously alters the trajectory of modern medicine[1–4]. However, there is a pressing need to improve understanding of the mechanisms that control cellular identity and function to advance the field of immunology and the treatment of immune related diseases.

The immune system has been parsed into hierarchical categories based on the features and function of each cell type. Temporal response to environmental triggers distinguishes the broadest categories of immune activity, referred to as innate and adaptive immunity. While the innate immune system poses a rapid response to pathogens, the adaptive immune system is slower to activate but provides a more specific response to the threat and can result in lasting memory towards the pathogen for years following the event[5]. A substantial portion of the adaptive immune compartment is comprised of T cells, which are informed about invading pathogens by the presentation of antigens in the form of short peptide sequences by major histocompatibility complex (MHC) molecules, which T cells use to find and eliminate the threat[6].

Stimulation of T cells via antigen presentation to the T cell receptor (TCR), costimulatory receptors present on the cell surface, and activating cytokines in the environment, leads to robust expansion of specific T cells[7]. Mature T cells include CD8+ T cells, known also as cytotoxic T cells, which are largely responsible for the elimination of pathogenic cells via direct killing mechanisms[8]. CD4+ T cells however, coordinate activation of other cells within the immune compartment and represent a diverse spectrum of cell subsets that secrete specific cytokines, enabling communication with a broad number of cell types[8,9].

CD4+ T cells include both pro and anti-inflammatory cells which functionally oppose one another. While the pro-inflammatory cell subsets, known as conventional T cells (Tconv) or effector T cells (Teffs) are much more abundant, their counterpart regulatory T cells (Tregs) are essential to establish dynamic immune homeostasis. Tregs make up less than 10% of the CD4+ T cell compartment but are required to prevent self-reactive T cells from expanding within the body, which can result in inflammation and autoimmune disease[10,11]. Tregs and Teffs each develop from a shared thymic precursor cell before cellular reprogramming leading to their distinct functions[12]. Due to this late divergence, both cell subsets have largely overlapped transcriptional profiles, making them responsive to many shared signaling molecules via expression of shared receptors. Mutual traits between these cell subsets enables like-minded localization and activation patterns that are required to mount a controlled immune response[13,14]. However, the underlying gene regulatory mechanisms that enable functional opposition with familial similarities across these subsets are not fully understood. This paradox highlights the complexity of balancing cellular identity with functional flexibility.

The cytokine IL-2 is particularly important for CD4+ T cell survival as it drives proliferation as well as cell intrinsic gene programming[15]. Expression of the IL-2 receptor itself is dynamically regulated across cell subsets. The receptor subunit IL2RA, (also known as CD25), improves cell

affinity for IL-2 and is carefully regulated to control sensitivity to the cytokine [16,17]. Tregs express high levels of IL2RA at a resting state and increase expression of the receptor slightly upon activation, whereas CD4+ Teffs express low levels of IL2RA at a resting state but upregulate the receptor to Treg equivalent levels for days following TCR activation[18]. For these reasons, IL2RA is known as a cell identity gene for Tregs and a functional marker of activation across the broader T cell compartment[19]. The expansive IL2RA enhancer extends over 50 Kb and contains a dense distribution of variants associated with autoimmune disease, suggesting even slight changes in regulation can disrupt immune function, while null mutations lead to widespread autoimmunity[20–22]. For decades, therapies have been developed in an attempt to harness IL-2 signaling, including targeted activation of IL2RA expressing cell subsets, to treat both cancer and autoimmune disease[15,16,23,24]. These strategies have been met with mixed success, including challenges in preventing widespread immune responses, suggesting that further research into the nuanced regulation of IL2RA may lead to improved and alternative routes of treatment. Owing to its context dependent regulation and great therapeutic significance, IL2RA represents an ideal gene to dissect mechanisms of cell type and stimulation specific gene regulation.

Disease associated variants provide insight into regulation of the immune system

In rare cases, a germline variant results in altered or diminished function of the encoded protein and leads to disease affecting the immune system. Record of these mutated genes, categorized as Inborn errors of immunity (IEI), and their associated diseases are currently maintained by the International Union of Immunological Societies (IUIS) and include 485 mutations as of 2022[25]. Information regarding these rare coding mutations provides valuable insight into the essential components of the immune system and the repercussions of altered expression in human systems. Investigations into the pathology of patients presenting with these monogenic diseases

has clarified the role of many proteins in the scope of the immune system as a whole. IPEX disorder, for example, resulting from mutations within *FOXP3*, affects only CD4+ regulatory T cells[25]. The link between *FOXP3* and the symptoms of IPEX, which generally presents as X-linked immune dysregulation including multi-organ autoimmune disease, was identified before Foxp3 expression had been observed in human Tregs[26,27]. Knowledge about IPEX instead led to the characterization of *FOXP3* as an essential gene in natural human Tregs through parallels to mouse models of autoimmune disease[28–30]. In this way, the identification of genetic associations with disease and investigations into the mechanism of pathology have expanded general knowledge of the immune system. In the case of IPEX, this knowledge is being directly translated into novel treatment strategies for patients as clinical trials are underway to insert functional *FOXP3* into CD4+ T cells from patients, bypassing the need for bone marrow transfusions from a donor which is the current standard of care[31,32].

Despite the significance of monogenic disease in immunology, variants in non-coding regions of the genome are much more commonly associated with immune disease. Genome wide association studies (GWAS) have led to the identification of thousands of these variants that are believed to contribute to the manifestation of numerous immune diseases and related blood traits[33–36]. Expression quantitative trait loci (eQTL) and protein quantitative trait loci (pQTL) studies have further revealed the functional effect of these variants on the expression of one or more genes or proteins. Overall, these analyses have been most successful in identifying variation in the expression of genes nearby the variant, in *cis*. Detecting the effects of the variants on the expression of genes outside the affected locus, in *trans*, has proven more difficult and highlights a limitation in understanding the broad effects of a non-coding variant on a complex trait[37,38]. Additionally, altered regulation as the result of a non-coding variant is often times cell type or context specific, especially within in the immune system where most cellular activity is dependent on external stimuli[39–42]. Finally, variants under substantial selective

constraint are less likely to be associated with an eQTL, resulting in reduced functional information for evolutionarily conserved sequences[37,43]. Comprehensively, reliance on QTL studies cannot fully reveal the functional implications of genetic variation that has been associated with immune disease.

CRISPR enables targeted perturbation and investigation of immune function

Advances in sequencing technology have propelled the discovery of genetic variants associated with immune disease and led to the characterization of a large number of genes and their encoded proteins as essential components of the immune system. However, this data is limited in the ability to extrapolate the role of each gene of interest in the relevant cell types and conditions, and to understand the relationship between important immune genes. Forward genetic approaches, which enable the identification of genes associated with a trait, allow us to probe the effect of individual genes in distinct cellular contexts without restraint. CRISPR, or clustered regularly interspersed short palindromic repeats, evolved as a defense system in bacteria and was eventually harnessed as a programable DNA nuclease capable of inducing a targeted double stranded break in DNA[44]. In 2012, Jennifer A. Doudna and Emmanuelle Charpentier published their engineered CRISPR-Cas9 system (CRISPR-associated 9), which includes a guide RNA that localizes the complex through complementary base paring to DNA[45]. This technology was then applied to mammalian systems to functionally ablate gene expression through a double stranded break in a gene which is then erroneously repaired via non-homologous end joining. In 2020, Douda and Charpentier were awarded the Nobel Prize in Chemistry for this work[44].

The development of CRISPR technology has enabled targeted manipulation of gene expression on a scale of a single gene to genome wide assays[46]. Notably, subsequent iterations of

CRISPR with altered domains and fusion proteins enable gene silencing or activation (CRISPRi, CRISPRa) without inducing double stranded breaks, targeted epigenetic modifications (e.g. CRISPRoff/CRISPRon), and editing of specific nucleotides (e.g. Cytosine base editors, adenine base editors, prime editors)[47,48]. The application of CRISPR gene editing within the immune system, in the form of arrayed and pooled high throughput assays, has resulted in the discovery of numerous genes essential to cellular identity and function, as well as novel therapeutic targets[20,49–58]. While the applications of CRISPR systems continue to evolve, enabling complex manipulation of the genetic landscape, the integration of the technology has already revolutionized experimentation and resulted in early therapeutic success[59,60].

Through the integration of knowledge gained from naturally occurring variants in the immune system and synthetic perturbation of genes *in vitro*, our understanding of the immune system and its regulatory components may expand without limitations. Pooled screening approaches, have become particularly invaluable for the identification of regulators of a specific cellular attribute, allowing for testing at a genome wide scale[61]. This approach, coupled with arrayed perturbation of genes of interest, or single cell sequencing after pooled perturbation, has led to the formation of a constellation of regulatory genes that control one another and important immune functions[52,62,63].

The application of CRISPR to dissect human T cell gene regulation

The induction of a genetic loss of function mutation through CRISPR in controlled cell types and conditions enables insight into the role of genes implicated in immune disease with nebulous function. Furthermore, transcript and chromatin level information from perturbed cells can be used to elucidate regulatory interactions both between genes of interest in immune disease and those without known associations to immune function alike. In this work, we focus exclusively

within the CD4+ T cell compartment, which contains diverse cells types relevant for immune disease with highly dynamic gene regulatory patterns. In **chapter 2**, we apply pooled CRISPR KO screens across conditions to define regulators of IL2RA in Tregs and resting and stimulated Teffs, revealing the regulatory components that participate in context specific gene expression of a core immune gene. Using arrayed CRISPR KO and genome wide sequencing approaches, we characterize a number of these regulators as factors that control T cell rest and activation and generate gene regulatory networks that clarify the mechanisms by which T cells transition between rest and activation. In **chapter 3**, we focus on IEI genes, perturbing transcription factors with a defined connection to immune disease and transcription factors with similar properties but no defined connection to immune disease. We constructed an interconnected network amongst these regulatory genes, distinguishing modules with comprehensible biological function and defining novel connections between regulators. Comprehensively, these investigations advanced our understanding of the regulation of essential T cell attributes with potential avenues to modify T cell behavior for therapeutic benefit.

References

1. Labanieh, L. & Mackall, C. L. CAR immune cells: design principles, resistance and the next generation. *Nature* **614**, 635–648 (2023).
2. Norberg, S. M. & Hinrichs, C. S. Engineered T cell therapy for viral and non-viral epithelial cancers. *Cancer Cell* vol. 41 58–69 Preprint at https://doi.org/10.1016/j.ccell.2022.10.016 (2023).
3. Mohammadi, P., Hesari, M., Chalabi, M., Salari, F. & Khademi, F. An overview of immune checkpoint therapy in autoimmune diseases. *International Immunopharmacology* vol. 107 Preprint at https://doi.org/10.1016/j.intimp.2022.108647 (2022).
4. Rappuoli, R., Alter, G. & Pulendran, B. Transforming vaccinology. *Cell* **187**, 5171–5194 (2024).
5. Yatim, K. M. & Lakkis, F. G. A brief journey through the immune system. *Clinical Journal of the American Society of Nephrology* **10**, 1274–1281 (2015).
6. Kotsias, F., Cebrian, I. & Alloatti, A. Antigen processing and presentation. in *International Review of Cell and Molecular Biology* vol. 348 69–121 (Elsevier Inc., 2019).
7. Azuma, M. & Yagita, H. *Co-Signal in T-Cell Activation*. http://www.springer.com/series/5584 (2019).
8. Sun, L., Su, Y., Jiao, A., Wang, X. & Zhang, B. T cells in health and disease. *Signal Transduction and Targeted Therapy* vol. 8 Preprint at https://doi.org/10.1038/s41392-023-01471-y (2023).

9. Hirahara, K. & Nakayama, T. CD4+ T-cell subsets in inflammatory diseases: Beyond the Th1/Th2 paradigm. *Int Immunol* **28**, 163–171 (2016).
10. Mason, G. M. *et al.* Phenotypic Complexity of the Human Regulatory T Cell Compartment Revealed by Mass Cytometry. *The Journal of Immunology* **195**, 2030–2037 (2015).
11. Santosh Nirmala, S. *et al.* Beyond FOXP3: a 20-year journey unravelling human regulatory T-cell heterogeneity. *Frontiers in Immunology* vol. 14 Preprint at https://doi.org/10.3389/fimmu.2023.1321228 (2023).
12. Lee, W. & Lee, G. R. Transcriptional regulation and development of regulatory T cells. *Exp Mol Med* **50**, e456-10 (2018).
13. Marangoni, F. *et al.* Expansion of tumor-associated Treg cells upon disruption of a CTLA-4-dependent feedback loop. *Cell* **184**, 3998-4015.e19 (2021).
14. Kumar, P., Bhattacharya, P. & Prabhakar, B. S. A comprehensive review on the role of co-signaling receptors and Treg homeostasis in autoimmunity and tumor immunity. *Journal of Autoimmunity* vol. 95 77–99 Preprint at https://doi.org/10.1016/j.jaut.2018.08.007 (2018).
15. Spolski, R., Li, P. & Leonard, W. J. Biology and regulation of IL-2: from molecular mechanisms to human therapy. *Nat Rev Immunol* **18**, 648–659 (2018).
16. Hernandez, R., Põder, J., LaPorte, K. M. & Malek, T. R. Engineering IL-2 for immunotherapy of autoimmunity and cancer. *Nat Rev Immunol* (2022) doi:10.1038/s41577-022-00680-w.
17. Chinen, T. *et al.* An essential role for the IL-2 receptor in T reg cell function. *Nat Immunol* **17**, 1322–1333 (2016).

18. Pekalski, M. L. *et al.* Postthymic Expansion in Human CD4 Naive T Cells Defined by Expression of Functional High-Affinity IL-2 Receptors. *The Journal of Immunology* **190**, 2554–2566 (2013).
19. Kmieciak, M. *et al.* Human T cells express CD25 and Foxp3 upon activation and exhibit effector/memory phenotypes without any regulatory/suppressor function. *J Transl Med* **7**, 1–7 (2009).
20. Simeonov, D. R. *et al.* Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549**, 111–115 (2017).
21. Goudy, K. *et al.* Human IL2RA null mutation mediates immunodeficiency with lymphoproliferation and autoimmunity. *Clinical Immunology* **146**, 248–261 (2013).
22. Li, Z. *et al.* T cell subset-selective IL2RA enhancers shape autoimmune diabetes risk. *bioRxiv* (2020).
23. Jiang, T., Zhou, C. & Ren, S. Role of IL-2 in cancer immunotherapy. *Oncoimmunology* **5**, 1–10 (2016).
24. Graßhoff, H. *et al.* Low-Dose IL-2 Therapy in Autoimmune and Rheumatic Diseases. *Front Immunol* **12**, (2021).
25. Tangye, S. G. *et al.* Human Inborn Errors of Immunity: 2022 Update on the Classification from the International Union of Immunological Societies Expert Committee. *J Clin Immunol* **42**, 1473–1507 (2022).
26. Sakaguchi, S. Regulatory T cells: history and perspective. *Methods Mol Biol* **707**, 3–17 (2011).

27. Bennett, C. L. *et al.* The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat Genet* **27**, 20–21 (2001).
28. Roncador, G. *et al.* Analysis of FOXP3 protein expression in human CD4+CD25+ regulatory T cells at the single-cell level. *Eur J Immunol* **35**, 1681–1691 (2005).
29. Graca, L. New tools to identify regulatory T cells. *Eur J Immunol* **35**, 1678–1680 (2005).
30. Mercer, F. & Unutmaz, D. *The Biology of FoxP3: A Key Player in Immune Suppression during Infections, Autoimmune Diseases and Cancer*. (2009).
31. Borna, S., Lee, E., Sato, Y. & Bacchetta, R. Towards gene therapy for IPEX syndrome. *European Journal of Immunology* vol. 52 705–716 Preprint at https://doi.org/10.1002/eji.202149210 (2022).
32. Passerini, L. *et al. CD4 + T Cells from IPEX Patients Convert into Functional and Stable Regulatory T Cells by FOXP3 Gene Transfer*. https://www.science.org.
33. Caliskan, M., Brown, C. D. & Maranville, J. C. A catalog of GWAS fine-mapping efforts in autoimmune disease. *Am J Hum Genet* **108**, 549–563 (2021).
34. Farh, K. K. H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
35. Goodin, D. S., Khankhanian, P., Gourraud, P. A. & Vince, N. *The Nature of Genetic and Environmental Susceptibility to Multiple Sclerosis*. *PLoS ONE* vol. 16 (2021).
36. Gerussi, A., Soskic, B., Asselta, R., Invernizzi, P. & Gershwin, M. E. GWAS and autoimmunity: What have we learned and what next. *J Autoimmun* **133**, 102922 (2022).

37. Umans, B. D., Battle, A. & Gilad, Y. Where Are the Disease-Associated eQTLs? *Trends in Genetics* vol. 37 109–124 Preprint at https://doi.org/10.1016/j.tig.2020.08.009 (2021).
38. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* **53**, (2021).
39. Chun, S. *et al.* Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat Genet* **49**, 600–605 (2017).
40. Flynn, E. D. *et al.* Transcription factor regulation of eQTL activity across individuals and tissues. (2021).
41. Jonkers, I. H. & Wijmenga, C. Context-specific effects of genetic variants associated with autoimmune disease. *Hum Mol Genet* **26**, R185–R192 (2017).
42. Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nat Genet* **49**, 139–145 (2017).
43. Mostafavi, H., Spence, J. P., Naqvi, S. & Pritchard, J. K. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nat Genet* **55**, 1866–1875 (2023).
44. Gostimskaya, I. CRISPR–Cas9: A History of Its Discovery and Ethical Considerations of Its Use in Genome Editing. *Biochemistry (Moscow)* vol. 87 777–788 Preprint at https://doi.org/10.1134/S0006297922080090 (2022).
45. Jinek, M. *et al. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity*. https://www.science.org.

46. Jinek, M. *et al. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity*. https://www.science.org.
47. Liu, G., Lin, Q., Jin, S. & Gao, C. The CRISPR-Cas toolbox and gene editing technologies. *Molecular Cell* vol. 82 333–347 Preprint at https://doi.org/10.1016/j.molcel.2021.12.002 (2022).
48. Simeonov, D. R. & Marson, A. CRISPR-Based Tools in Immunity. *Annu Rev Immunol* **12**, 44 (2019).
49. Freitas, K. A. *et al.* Enhanced T cell effector activity by targeting the Mediator kinase module. *Science (1979)* **378**, (2022).
50. Schmidt, R. *et al.* CRISPR activation and interference screens decode stimulation responses in primary human T cells. *Science (1979)* **375**, (2022).
51. Mowery, C. T. *et al.* Systematic decoding of cis gene regulation defines context-dependent control of the multi-gene costimulatory receptor locus in human T cells. *Nat Genet* **56**, 1156–1167 (2024).
52. Freimer, J. W. *et al.* Systematic discovery and perturbation of regulatory genes in human T cells reveals the architecture of immune networks. *Nat Genet* **54**, 1133–1144 (2022).
53. Schumann, K. *et al.* Functional CRISPR dissection of gene networks controlling human regulatory T cell identity. *Nat Immunol* **21**, 1456–1466 (2020).
54. Cortez, J. T. *et al.* CRISPR screen in regulatory T cells reveals modulators of Foxp3. *Nature* **582**, 416–420 (2020).
55. Lam, A. J. *et al.* Optimized CRISPR-mediated gene knockin reveals FOXP3-independent maintenance of human Treg identity. *Cell Rep* **36**, 109494 (2021).

56. Shifrut, E. *et al.* Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function. *Cell* **175**, 1958-1971.e15 (2018).
57. Loo, C. S. *et al.* A Genome-wide CRISPR Screen Reveals a Role for the Non-canonical Nucleosome-Remodeling BAF Complex in Foxp3 Expression and Regulatory T Cell Function. *Immunity* **53**, 143-157.e8 (2020).
58. Dubrot, J. *et al.* In vivo CRISPR screens reveal the landscape of immune evasion pathways across cancer. *Nat Immunol* **23**, 1495–1506 (2022).
59. Sheridan, C. The world's first CRISPR therapy is approved: who will receive it? *Nature Biotechnology* vol. 42 3–4 Preprint at https://doi.org/10.1038/d41587-023-00016-6 (2024).
60. Li, T. *et al.* CRISPR/Cas9 therapeutics: progress and prospects. *Signal Transduction and Targeted Therapy* vol. 8 Preprint at https://doi.org/10.1038/s41392-023-01309-7 (2023).
61. Joung, J. *et al.* Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* **12**, 828–863 (2017).
62. Lopes, R. *et al. Systematic Dissection of Transcriptional Regulatory Networks by Genome-Scale and Single-Cell CRISPR Screens*. *Sci. Adv* vol. 7 https://www.science.org (2021).
63. Rubin, A. J. *et al.* Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. *Cell* **176**, 361-376.e17 (2019).

Chapter 2 Context specific regulation of IL2RA

Abstract

The ability of cells to maintain distinct identities and respond to transient environmental signals requires tightly controlled regulation of gene networks[1–3]. These dynamic regulatory circuits that respond to extracellular cues in primary human cells remain poorly defined. The need for context-dependent regulation is prominent in T cells, where distinct lineages must respond to diverse signals to mount effective immune responses and maintain homeostasis[4–8]. Here, we performed CRISPR screens in multiple primary human CD4+ T cell contexts to identify regulators that control expression of IL2RA, a canonical marker of T cell activation transiently expressed by pro-inflammatory effector T cells and constitutively expressed by anti-inflammatory regulatory T cells where it is required for fitness[9–11]. Strikingly, ~90% of identified IL2RA regulators were cell type- or stimulation-state specific, and a subset even had opposite functional effects across conditions. Using single-cell transcriptomics after pooled perturbation of context-specific screen hits, we characterized specific factors as regulators of overall rest or activation and constructed state-specific regulatory networks. Remarkably, MED12 – a component of the Mediator complex – serves as a dynamic orchestrator of key regulators, controlling expression of distinct sets of regulators in different T cell contexts. Immunoprecipitation-mass spectrometry revealed that MED12 interacts with the histone methylating COMPASS complex. MED12 was required for histone methylation and expression of genes encoding key context-specific regulators, including rest maintenance factor KLF2 and the versatile regulator MYC. CRISPR ablation of MED12 blunted the cell state transitions between rest and activation and protected from activation-induced cell death. Overall, this work leverages CRISPR screens performed across conditions to define dynamic gene circuits required to establish resting and activated T cell states.

Introduction

Each cell type expresses a distinctive set of genes to maintain its identity and respond to external cues. Context-specific networks of trans-regulatory proteins are required to coordinate these gene expression programs but are not fully mapped in human cells[1,12,13]. The intricacies of conditional gene regulation are exemplified within the human immune system, where diverse cell types must specialize as well as sense and respond dynamically to stimuli to maintain homeostasis [5,7]. Cell type- and context-specific expression of receptors and other key molecules enable coordinated immune responses and have been targeted in immune modulating therapies[14–16]. However, the trans-regulatory mechanisms that allow for conditional expression of the genes encoding these proteins remain poorly understood. Deciphering these systems will advance our understanding of nuanced gene regulation required for human health and improve our ability to modulate the immune system with effective immunotherapies.

Within the CD4+ T cell compartment, regulatory T cells (Tregs) and effector T cells (Teffs) functionally oppose each other, serving immune-suppressive and immune-stimulating roles, respectively. However, their relatively late-stage differentiation results in a high degree of similarity at the gene expression level between the two cell types[4,5,8]. Both possess the ability to respond to a set of shared environmental signals, albeit with key differences. The cytokine IL-2 drives cellular fitness of Tregs and activated Teffs, and competition for this signal can shape immune responses in health and disease[10,17]. The IL-2 receptor high-affinity subunit IL2RA (also known as CD25) enhances receptor affinity for IL-2 and is carefully regulated to control sensitivity to the cytokine[11,15]. Tregs constitutively express high levels of IL2RA at rest and mildly increase expression of the receptor upon activation, whereas Teffs express low levels of IL2RA at rest but transiently upregulate the receptor for days following TCR stimulation[9].

Numerous therapeutic strategies have been employed to improve the cellular specificity and longevity of IL-2 signaling, some of which utilize the distinct expression patterns of IL2RA across subsets to promote efficacy and prevent adverse events[15,18]. IL2RA represents a clinically-relevant gene to study for mechanistic insights into cell type- and stimulation-specific gene regulation.

Results

Context-specific IL2RA regulator screens

We applied pooled CRISPR KO screens to identify upstream trans-regulators of IL2RA across cell type and stimulation conditions. We utilized a library of 6000 sgRNAs to target trans-factor genes expressed in T cells (~1350 transcription factors and chromatin modifiers) as well as select immune regulators and control genes[19,20].We isolated, edited and expanded primary human Tregs (CD4+CD25^{hi}CD127^{low}) and Teffs (CD4+CD25^{low}). We then screened for regulators of IL2RA in resting Teffs (IL2RA low) and resting Tregs (IL2RA high) 10 days after initial stimulation, as well as re-stimulated Teffs (IL2RA high, 72 hours post-stimulation) (**Figure 2.1a-b, Supplemental Figure 2.1a**). Screens were performed at high coverage (700-1000x cells/sgRNA per donor) and had similar positive control sgRNA effect sizes across conditions, as well as high donor-to-donor correlations and the resting Teff screen replicated published results[19] (**Supplemental Figure 2.1b-d**).

The screens collectively detected over 100 trans-regulators (FDR < 0.05, **Figure 2.1c, Table 2.1**) whose perturbation altered IL2RA surface expression in at least one context. Interestingly, only 16 regulators were hits in all three screens, 75% of which shared the same direction of effect across conditions (**Figure 2.1d**). These 12 “consistent regulators” of IL2RA included members of the JAK-STAT pathway. Among the consistent regulators, the effect sizes of

several trans-regulators varied greatly between conditions. GATA3, for example, was a particularly potent positive regulator of IL2RA in stimulated Teffs, with a median log_2 fold change in sgRNA enrichment in the IL2RA low/high bin of 2.47 compared to 0.95 and 0.49 in resting Teffs and resting Tregs, respectively (**Figure 2.1d**, **Supplemental Figure 2.1d**). The majority of identified IL2RA regulators were significant in only one or two conditions, demonstrating cell type- or stimulation-specific effects (**Figure 2.1c**) although most regulators were expressed (based on bulk RNAseq) across conditions (**Supplemental Figure 2.1e**). We compared the direction and magnitude of effect of the perturbations across the three screens to categorize context-dependent regulators of IL2RA. Notably, few strong negative regulators were identified in stimulated Teffs compared to resting Teffs. For example, KLF2, MYB, and ZNF217 were only identified as significant negative regulators in the resting state (**Figure 2.1e**). These data highlight broad differences in the network upstream of IL2RA between activation states, where negative regulatory forces are transiently reduced following stimulation to increase levels of IL2RA.

Although both Tregs and activated Teffs express high levels of IL2RA, there were fewer shared IL2RA regulators between the resting Treg and stimulated Teff screens than the resting Treg and resting Teff screens, indicating that Tregs and stimulated Teffs rely on different systems to achieve high expression (**Figure 2.1c-g**). Overall, the screen performed in Tregs yielded a particularly large number of significant hits, including both positive and negative regulators specific to the condition, such as FOXO1, USP22[21], and MYC (**Supplemental Figure 2.2a**). IL2RA is required for the fitness of Tregs[11] and the large network of positive and negative regulators likely acts as a buffering system to maintain relatively consistent expression.

Surprisingly, a few regulators exerted effects in opposing directions across conditions. Notably, MED12, CBFB, and PRDM1 were identified as strong positive regulators of IL2RA in stimulated

Teffs, but negative regulators in resting Teffs (**Figure 2.1e**). MED12 and, to a lesser extent, MED11 – components of Mediator of RNA polymerase II (Mediator) – were both identified as positive regulators of IL2RA in resting Tregs but negative regulators of IL2RA in resting Teffs. These strong "differential" context-dependent effects were particularly striking for components of a complex with general roles in transcription. Additionally, BATF and IRF4, which co-bind genomic sites in T cells[22], were identified as differential regulators with a negative effect on IL2RA levels in resting Tregs and a positive effect on IL2RA expression in both resting and stimulated Teffs (**Figure 2.1f-g**). Notably, BATF has been highlighted as a key regulator of Treg tissue homing and stability *in vivo*[23,24]. Our characterization of BATF and IRF4 as negative regulators of IL2RA in ex vivo human Tregs suggests a nuanced role with potential differences across species or contexts. Comprehensively, the screening approach led to the identification of cell type- and stimulation-specific regulators upstream of IL2RA, as well as the unexpected class of regulators that promote and repress expression of IL2RA in distinct contexts.

Dynamic Regulation of IL2RA

In order to validate and characterize the function of cell type-specific hits from our screens, we ablated 18 factors and quantified IL2RA protein expression in both Tregs and Teffs, prioritizing genes with discordant effects across cell types or Treg-specific effects. The arrayed CRISPR KO results confirmed condition-specific regulatory roles for many factors (**Supplemental Figure 2.2b-e**). Interestingly, components of Mediator (MED12, MED11, MED30) and SAGA (TAF5L, USP22, ATXN7L3), both ubiquitous transcriptional coactivator complexes, demonstrated cell type- and stimulation-specific effects. Consistent with screen results, MED12 had the most dynamic role (**Supplemental Figure 2.2e-f**) despite steady expression levels across conditions (**Supplemental Figure 2.1e**). MED12 ablation consistently increased IL2RA in resting Teffs but decreased IL2RA in stimulated Teffs and in Tregs (both resting and stimulated). We dissected

the kinetics of stimulation-response regulation through arrayed KO with an extended series of collection timepoints. Much like the screen, many perturbations that increased IL2RA in resting Teffs (e.g. ZNF217, MED12, PRDM1) had minimal effects on IL2RA expression or even decreased its expression 48-72 hours after stimulation (**Figure 2.2a-b**). In contrast, fewer IL2RA regulators with distinct stimulation-responsive effects were observed in Tregs than Teffs; no negative regulators of IL2RA in resting Tregs became positive regulators during activation or vice versa. Despite differences in activation responses, both Tregs and Teffs appeared reliant on KLF2 and CBFB to repress IL2RA at resting timepoints (**Figure 2.2a-b**).

While our pooled screens captured regulators of maximum and minimum levels of IL2RA expression at specific time-points, the arrayed KO time course experiments also revealed regulators that govern transitions between states. We identified several factors that enable the transition from activated IL2RA levels to rest levels (~96-144 hrs) with particularly large effects in Teffs, which undergo the greatest fluctuations in IL2RA expression. TAF5L, BPTF, and SOCS3 contributed to this reduction of surface IL2RA as cells returned to rest (**Figure 2.2b-c**). CTLA-4, another receptor transiently induced in stimulated Teffs and constitutively expressed in Tregs, exhibited similar patterns of temporal regulation by the perturbed genes, suggesting that the regulators control a broader network of dynamically expressed genes (**Supplemental Figure 2.2g**). In summary, many regulators contribute to activation and rest associated gene regulation in temporally defined stages, with some regulatory systems specific to each T cell subset.

MED12 facilitates rest and activation

Stimulation-induced expression of IL2RA is a canonical marker of T cell activation. We suspected that many of the regulators' effects were not limited to IL2RA and were reflections of

altered overall activation states. To characterize such global effects, we performed Perturb-CITE-seq (pooled CRISPR perturbations coupled with single cell RNA-seq and surface proteomics) in resting and stimulated (48 hours post-stimulation) Tregs and Teffs. We used CRISPR interference (CRISPRi), to knock down 28 regulators of IL2RA, prioritizing trans-factors with state-specific effects. We confirmed perturbation efficiency via transcript expression of the targeted regulator and observed significant changes to the transcriptome and key cell surface receptors (**Supplemental Figure 2.3a-c**). We next assessed resulting changes to the overall T cell activation states based on a global transcriptional signature[25]. Many context-specific regulators of IL2RA served as broad modulators of rest or activation, confirming our hypothesis. In resting Teffs, KLF2, MYB, and SOCS3 stood out as strong repressors of activation, whereas STAT5B, MYC, BATF and IRF4 appeared particularly important to promote activation in stimulated cells (**Figure 2.3a**). Strikingly, MED12 knock-down increased the activation scores of resting cells but lowered the activated score of stimulated cells in both cell types (**Figure 2.3a**, **Supplemental Figure 2.3d**). Collectively, these results reveal core regulators of global state-specific gene expression within our screen hits and notably distinguish MED12 as a dynamic factor governing both rest and activation programs.

Cell state regulators often operate in hierarchical networks[26]. We constructed state- and cell type-specific gene regulatory network maps to visualize how regulators affect one another (**Figure 2.3b-c**, **Supplemental Figure 2.3e-f**). Amongst rest maintenance factors, there were many positive feedforward loops converging on KLF2, ETS1 and MYC, which were downstream of the largest number of genes (**Figure 2.3b**, **Supplemental Figure 2.3e**). Notably, MED12 strongly promoted expression of these core resting state maintenance factors. The network structure of regulators controlling gene expression in stimulated Teffs was strikingly distinct from that in resting Teffs. In stimulated Teffs, we found few instances of strong positive connections, with the exception of MED12 promoting the expression of MYC (**Figure 2.3c**). Instead, MED12,

MYC, STAT5B, and BATF (all factors that promote activation following stimulation; white) were required to repress expression of several resting state maintenance factors which did not affect activation following stimulation, including SOCS3, NFKB2, and FOXO1 (colored in dark grey). Interestingly, this network structure echoes the general structure observed in our IL2RA screens and arrayed assays, where we found reduced negative regulatory effects following stimulation relative to the resting state. Even SOCS3, the strongest negative regulator of IL2RA identified in the stimulated Teff screen, was more specifically characterized as an early return-to-rest repressor of IL2RA (**Figure 2.2b**). Perturb-seq further clarifies that SOCS3 and other rest maintenance/promoting factors are repressed in stimulated cells by activation promoting factors, allowing the cells to transiently reach an activated state. These results lead to a model whereby the resting state is actively reinforced by a self-promoting network of regulators and the transition to peak activation state requires repression of factors that promote rest. Importantly, MED12 orchestrates expression of key regulators of both rest and activation within these networks.

MED12 controls key regulators of IL2RA

To further probe the mechanism of dynamic regulation by MED12, we knocked-out the gene and performed bulk RNAseq. In both Teffs and Tregs, MED12 KO caused resting cells to prematurely up or downregulate genes that are normally differentially expressed in response to stimulation (as assessed in AAVS1 KO control cells) (**Figure 2.4a**). Conversely, in stimulated cells, we observed dampening of stimulation-induced changes in gene expression in MED12 KO relative to AAVS1 KO cells. A binomial test also confirmed aberrant expression of stimulation-specific genes in all MED12 KO conditions (**Supplemental Figure 2.3f**). Together, along with our Perturb-seq activation scoring, these results demonstrate that without MED12,

CD4+ T cells are unable to reach a full rested state or achieve peak levels of activation and instead exist in an intermediate state.

Overall, regulators of IL2RA identified in our pooled screens were enriched in the differentially expressed genes downstream of MED12 across all conditions and revealed routes of context-specific regulation by MED12 (**Figure 2.4b**). For example, MED12 KO caused increased expression of IRF4 in resting Teffs and Tregs, but decreased IRF4 levels in the stimulated cell conditions. (**Figure 2.4c**). MED12 ablation markedly decreased levels of positive IL2RA regulator GATA3 in stimulated Teffs, and decreased levels of negative IL2RA regulator (and rest maintenance factor) KLF2 in resting Teffs and Tregs (**Figure 2.4c-d**). Additional experiments revealed functional changes resulting from transcriptional reprogramming of MED12 KO cells, including reduced suppressive capacity *in vitro* by Tregs relative to AAVS1 KO cells, impaired IL-10 secretion by Tregs, and impaired Th2 associated cytokine secretion by Teffs (**Supplemental Figure 2.4a-e**). Collectively, these results reveal that MED12 directs a network composed of cell type- and stimulation-specific-regulators to achieve context dependent expression.

MED12 is part of the kinase domain of Mediator, which has been viewed as an inhibitory component because its presence prevents binding of the complex to RNA polymerase II[27–29]. We perturbed one subunit from each functional Mediator module and performed bulk RNAseq in resting and stimulated Tregs and Teffs. Much like the screen, MED12 and core Mediator KOs often shared the same direction of effect, reflected by a positive correlation between MED12 regulated genes and those regulated by core subunits MED11, MED14, and MED31 (**Figure 2.4e, Supplemental Figure 2.5a-b**). We also compared the effects of different Mediator component knockout on surface protein levels of IL2RA. Here we noted MED12 KO effects could be stronger and even partially discordant with other Mediator subunit KOs, depending on

the cell type and stimulation context (**Supplemental Figure 2.6a-b**). Collectively, these data reveal partially shared effects by MED12 and core Mediator, and depict MED12 as particularly important to promote expression of context-specific regulators of CD4+ T cell state.

We next assessed the effect of ablating key IL2RA regulators on the chromatin landscape at the *IL2RA* locus. H3K27ac is a mark of active enhancers and varies considerably between cell types and states[30]. We performed H3K27ac CUT&RUN following KO of select context-specific regulators; Mediator subunits (MED12, MED11, MED24), SAGA subunits (TAF5L, ATXN7L3, USP22), BATF and ZNF217. Perturbation of several regulators, especially MED12, resulted in significant changes in acetylation compared to AAVS1 KO samples (**Supplemental Figure 2.6c**). Downstream of the *IL2RA* transcription start site (TSS) showed significantly less H3K27ac in MED12 KO samples, specifically in Tregs. Most strikingly, MED12 KO Teffs had increased levels of acetylation in a region upstream of the TSS that is normally more acetylated in Tregs, which we previously characterized as a Treg-specific element called CaRE3[6] (**Supplemental Figure 2.6d**). Treg ChIP-seq data[31] contained prominent STAT5A peaks within CaRE3, suggesting that increased STAT5-signaling could contribute to more Treg-like gene expression in MED12 KO resting Teffs (**Supplemental Figure 2.6d**). Interestingly, we observed that genome-wide regions of differential acetylation in the TAF5L KO were highly correlated with MED12, including increased acetylation in Teffs at the *IL2RA* CaRE3 locus, suggesting a possible shared downstream regulator (**Supplemental Figure 2.6d-h**). Collectively, these changes demonstrate a loss of context-specific chromatin features required for cell identity and state dynamics as the result of regulator perturbations.

MED12 shapes chromatin at core genes

To probe the mechanism of context specific gene regulation by MED12 in human CD4+ T cells, we sought to define its interaction partners. MED12 lacks a DNA binding domain and enzymatic function but possesses several intrinsically disordered regions ideal for protein-protein interactions. We performed endogenous immunoprecipitation mass spectrometry (IP-MS) of MED12 in resting and stimulated Teffs and identified 203 significant interaction partners across conditions (Bayesian-FDR <= 0.05), including all members of the Mediator complex, except MED25, MED26 and MED12L (**Supplemental Figure 2.7a**). Surprisingly, the subset of proteins with over 100-fold enrichment in the MED12 pull down included numerous members of COMPASS, a histone methylating complex (**Figure 2.5a**). While COMPASS has several configurations, one particular assembly was represented including SETD1A, an H3K4me1-3 methyltransferase, and CXXC1, a DNA binding protein. Western blotting further confirmed co-precipitation of SETD1A and CXXC1 with MED12 (**Supplemental Figure 2.7c-d**).

MED12 interaction with COMPASS led us to suspect it may affect targeted H3K4 methylation. H3K4me1-2 are associated with poised and active enhancers while H3K4me3 is concentrated at active transcription start sites and has been shown to directly promote transcription[32–34]. CUT&RUN demonstrated widespread changes to H3K4me1-3 distribution in MED12 KO Teffs compared to control cells (**Figure 2.5b**, **Supplemental Figure 2.8a**). Changes in H3K4me3 were correlated strongly with altered gene expression in the MED12 KO samples (**Supplemental Figure 2.8b**). We defined genes bound by MED12 across states using ChIP-seq in AAVS1 KO T cells, using MED12 KO samples to establish the background, resulting in high-confidence MED12 peaks (Methods). We intersected these peaks with regions of differential methylation or gene expression in the MED12 KO samples and found that bound

regions were predominantly associated with decreased H3K4me3 and reduced expression (**Figure 2.5b**, **Supplemental Figure 2.8c**). Strikingly, a number of IL2RA regulators and T cell rest maintenance and activation promoting factors were among these genes, including *KLF2*, *MYC*, and *ETS1* at rest and *MYC* and *SATB1* after stimulation, suggesting that MED12 directly promotes their conditional expression. Again, using ChIP-seq we found CXXC1 was also present at many of these loci in resting cells, intersecting 52% of MED12 bound regions and 82% of MED12 bound genes (**Figure 2.5c, Supplemental Figure 2.8d-f**). Interestingly, CXXC1 peaks were less abundant in stimulated samples, suggesting that the protein may be displaced following activation, at which point MED12 localization also changes (**Supplemental Figure 2.8d-g**). H3K27ac was also affected at several regulators, including the *KLF2* locus in resting Teffs and Tregs (**Supplemental Figure 2.8h**). Comprehensively, these data suggest that COMPASS and MED12 colocalize at several key activation state regulatory genes where MED12 functions as a positive regulator of gene expression.

Loss of H3K4me3 has been associated with increased RNA polymerase II pausing, which results in an accumulation of polymerase at the TSS and coordinated loss in the gene body[33,34]. Using ChIP-seq in resting and stimulated Teffs, we found that pausing decreased globally following stimulation in AAVS1 KO cells. However, MED12 ablation increased pausing in stimulated Teffs relative to AAVS1 KO cells and decreased pausing in resting Teffs, resulting in diminished differences in pausing between states (**Supplemental Figure 2.9a-b**). This effect is consistent with hyper-activation in rest and blunted stimulation response as we described previously. Accumulation of NELFA at the TSS of genes in stimulated MED12 KO cells relative to AAVS1 KO samples and decreased polymerase PS5 and PS2 in the gene body relative to the TSS provided additional evidence of increased pausing after activation, including at the *IL2RA* locus which did not have significantly altered H3K4me3 (**Supplemental Figure 2.9c-d**). In resting MED12 KO cells, reduced polymerase at the TSS was a prominent global feature,

suggesting inhibited recruitment (**Supplemental Figure 2.9a**). Collectively, reduced pausing at rest and increased pausing after stimulation was a genome-wide trend for MED12 KO cells.

While global changes in polymerase activity reflect the altered activation state of the KO cells, core regulators of rest and activation bound by MED12 exhibited more consistent changes in transcription. The TSS of rest maintenance factor *KLF2* had significantly reduced H3K4me3 and increased pausing at rest in MED12 KO cells, deviating from the overall trend (**Figure 2.5c**). *MYC*, which participates in maintenance of both rest and promotion of activation response, had reduced H3K4me3 in both states and exhibited reduced polymerase recruitment at rest and increased pausing following stimulation. The resulting strong downregulation of these genes following MED12 ablation (**Figure 2.4c**) suggests a model in which MED12 drives expression of context-specific regulators. To assess the significance of these particular loci, we looked for evidence of their role in broader MED12 signaling. Using bulk RNAseq data in resting CD4+ T cells, we found that ablation of KLF2 alone was able to account for 30% of differentially expressed genes downstream of MED12 (**Figure 2.5d**). In lieu of bulk RNAseq for MYC KO T cells, we found a strong negative enrichment of MYC over-expression induced genes in both resting and stimulated MED12 KO samples (**Figure 2.5e**). Notably, MYC has been characterized as a pause release factor, and may contribute to global changes following activation[35]. While we did not exhaust the important factors downstream of MED12 likely contributing to its effects, we demonstrate that by promoting expression of several core regulatory factors, MED12 is able to establish central governance over broad regulatory networks.

We next asked if the Mediator kinase CDK8 and homolog CDK19 participate in MED12 driven regulation of activation. We used SEL120-34A, an inhibitor specific to CDK8 and CDK19. Treatment was sufficient to reproduce elevated levels of IL2RA in resting cells but when the

cells were stimulated, IL2RA remained significantly higher than vehicle (H_2O) control treated cells, whereas MED12 KO cells had lower IL2RA expression compared to AAVS1 KO cells (**Supplemental Figure 2.9e-f**). Next, we assessed changes in H3K4me1-3 following kinase inhibition. We first confirmed that the vehicle treated control cells responded to stimulation similarly to AAVS1 KO cells, finding high correlation between stimulation responsive sites (**Supplemental Figure 2.9g**). However, differentially methylated sites between SEL120-34A and vehicle treated cells were poorly correlated with MED12 KO effects, with the notable exception of shared reduced H3K4me3 at *KLF2*, *MYC*, and *ETS1* in the resting condition (**Supplemental Figure 2.9h**). Reduced expression of resting state maintenance factor *KLF2* is consistent with elevated levels of IL2RA observed in the kinase inhibitor treated cells. Collectively, these data suggest a complex role for the Mediator kinase, which contributes to the regulation of several rest maintenance factors, but does not explain the MED12-mediated regulation of stimulation responses.

MED12 KO limits activation induced death

Synthetic perturbation of key regulators is a promising strategy to improve adoptive T cell therapies. Recently, MED12 KO was nominated by a genome-wide CRISPR screen in CAR T cells to promote fitness. Ablation of MED12 resulted in improved CAR-T cells with sustained expansion and tumor control in preclinical models[36]. We speculated that an unreported but critical part of the therapeutic success of these experiments may be mediated by altered activation state transitions – avoiding complete rest and a state of peak activation. Using bulk RNAseq data from Freitas et al., we generated an activation score using genes upregulated in control CARs after stimulation. We then applied this score to the control and MED12 KO CARs and found a significant decrease in activation for the stimulated MED12 KO CARs compared to the controls (**Supplemental Figure 2.10a**). Within our Perturb-seq pool, MED12-targeted cells

experienced the largest increase in total Teff cell counts, especially stimulated cells (**Figure 2.6a**). Stimulated MED12-targeted Tregs also exhibited similar effects (**Supplemental Figure 2.10b**). We asked why MED12-targeted cells with reduced activation capacity would be more abundant than non-targeting controls. Based on cell state signatures in Perturb-seq, MED12 targeted cells showed a slight increase in proliferative cells in the resting condition but a substantial decrease in proliferative cells in the stimulated condition compared to non-targeting cells (**Supplemental Figure 2.10c**). Across the perturbed T cell pool, the percentage of proliferative cells and total cell abundance were not well-correlated, possibly due to decreased viability.

We reasoned that the reduced stimulation responses in MED12-targeted cells, may instead improve cell durability by limiting activation-induced cell death (AICD). Consistent with this hypothesis, genes associated with “apoptosis” were enriched among genes differentially expressed between MED12 KO and AAVS1 KO cells, driven by a mix of both up and downregulated genes (**Supplemental Figure 2.10d-e**). To determine if MED12 KO altered apoptosis in response to stimulation (AICD), we performed a dose-response of stimulation strength using anti-CD3/CD28/CD2 soluble tetramers and quantified apoptosis via caspase-3/7 activation. As expected, apoptosis increased with stimulation dose in AAVS1 KO control Teffs (**Figure 2.6b**, **Supplemental Figure 2.10f**). In marked contrast, MED12 KO cells underwent minimal apoptosis in response to even strong stimulation. Surprisingly, apoptosis initiating receptor FAS was elevated on MED12 KO cells throughout the assay, possibly the result of an affected feedback loop (**Supplemental Figure 2.10g**). The MED12 KO associated reductions in apoptosis translated to improved live cell abundance, providing an explanation for improved cell durability following MED12 ablation in the stimulated condition (**Figure 2.6c-d**).

Discussion

CRISPR screens in multiple primary cell conditions collectively defined a dynamic network of trans-regulators that enable cell type- and state-specific expression of IL2RA. We uncovered marked differences in regulation between Tregs with constitutively high levels of IL2RA and stimulated Teffs with transiently high levels. Teffs utilize waves of regulators to maintain rest (KLF2 and MED12), achieve peak expression (GATA3 and MED12), and return to a resting state (TAF5L and SOCS3); while Tregs appear to utilize a more static but expansive network of regulators to maintain IL2RA levels. One key insight that emerged is that the resting state depends on multiple rest maintenance factors that form a positive-feedback network, collectively promoting expression of a core rest factor KLF2. Activation promoting factors repress this rest maintenance network following stimulation to achieve a maximal activated cell state. MED12 serves as a dynamic regulator of state-specific gene expression via orchestration of downstream factors across these networks. Mediator accumulation is a feature of super-enhancers and has been associated with their activity in diverse cell types, suggesting that MED12 is recruited to the locus of key regulatory genes that are under precise enhancer regulation.[37,38]

Multiple studies have suggested that cycles between T cell rest and activation can influence the durability of adoptive T cell therapies[39,40]. Our findings suggest that loss of MED12 tunes T cell activation responses and limits activation-induced apoptosis to improve durability. This resistance to apoptosis appears to function independent of FAS (which was upregulated following MED12 ablation), another pathway that has been targeted to enhance cell therapy persistence[41]. Targeting MED12 and other context-dependent regulators may offer additional advantages by enhancing fitness of conventional T cells, but reducing IL2RA expression and suppressive function of Tregs (cells that could limit the efficacy of bulk CAR-T products[36]).

Increased expression of IL2RA by resting Teffs and decreased expression by Tregs may help Teffs compete more effectively for IL2.

Here we limited our study to trans-regulators within the CD4+ T cell compartment to comprehensively profile Teffs and Tregs across stimulation conditions, but additional screening conditions, perturbations and phenotypic readouts provide further insight into context-specific gene regulatory networks. The current work revealed the architecture of the gene regulatory networks controlling dynamic expression of IL2RA across cell types and activation states. More broadly, it provides fundamental insights into the regulation of T cell identity and mechanisms governing transitions between rest and activation.

Figures

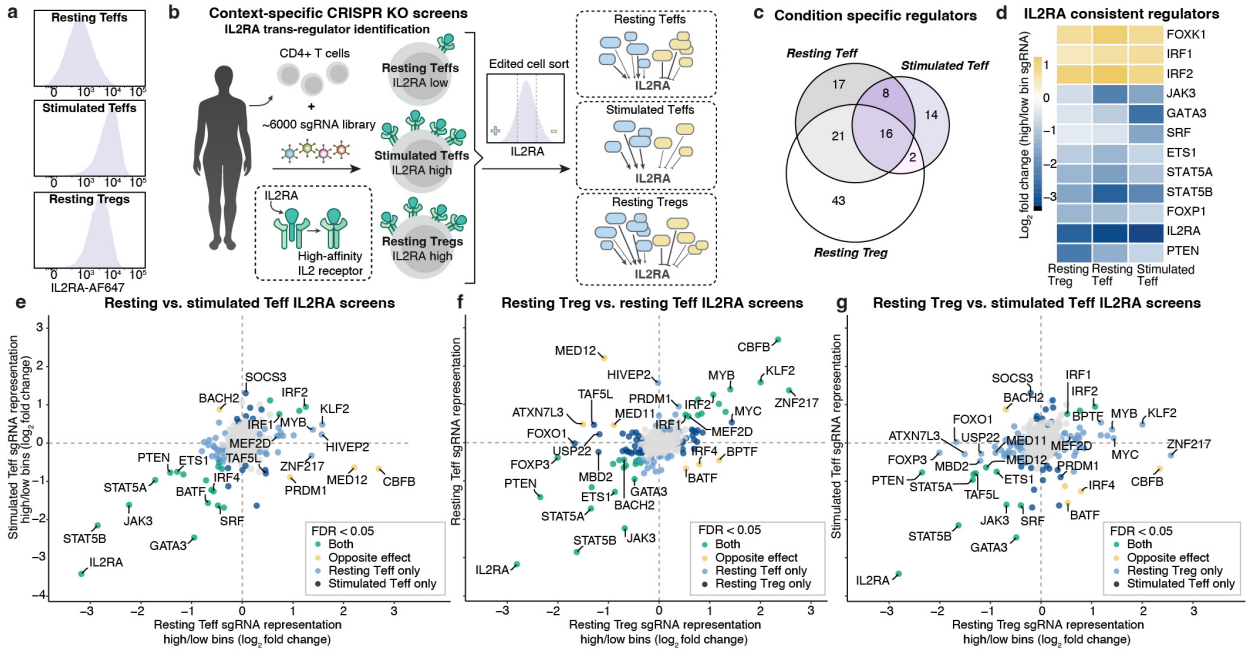


Figure 2.1 Identification of context-dependent regulators of IL2RA expression
a. IL2RA surface expression levels by flow cytometry. **b.** Schematic of context-specific trans-regulatory CRISPR screens **c.** Venn diagram of regulators identified across screen conditions. **d.** Consistent regulators of IL2RA identified as significant in the same direction across all 3 screens. (FDR < 0.05; Treg screen: n=2, resting Teff screen: n=3, stimulated Teff screen: n=3 donors). **e-g.** Comparisons of IL2RA screen results colored by significance and direction of effect in both screens (significant = FDR < 0.05; Treg screen: n=2, resting Teff screen: n=3, stimulated Teff screen: n=3 donors)

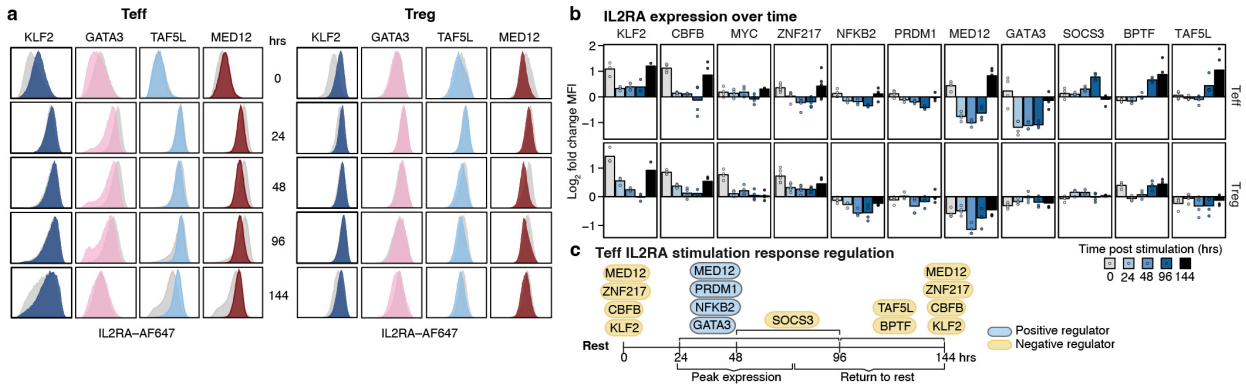


Figure 2.2 Temporal regulation of IL2RA following stimulation by distinct factors
a. Representative flow cytometry histograms of IL2RA expression after arrayed KO. AAVS1 safe harbor control KO results are shown in grey, and the y axis is normalized to the mode. Timepoints represent time after restimulation starting with 0 hours (no restimulation) **b.** Quantification of KO effect on IL2RA expression across stimulation timepoints. Log_2 fold change IL2RA median fluorescent intensity (MFI) calculated for KO compared to AAVS1 control KO samples from the same donor. Each point represents a donor and sgRNA combination. (n = 2 donors x 2 sgRNAs per KO except KLF2 KO where n = 3 and ZNF217 KO where n = 6). **c.** Schematic of select IL2RA regulators that enable temporal control of IL2RA in Teffs.

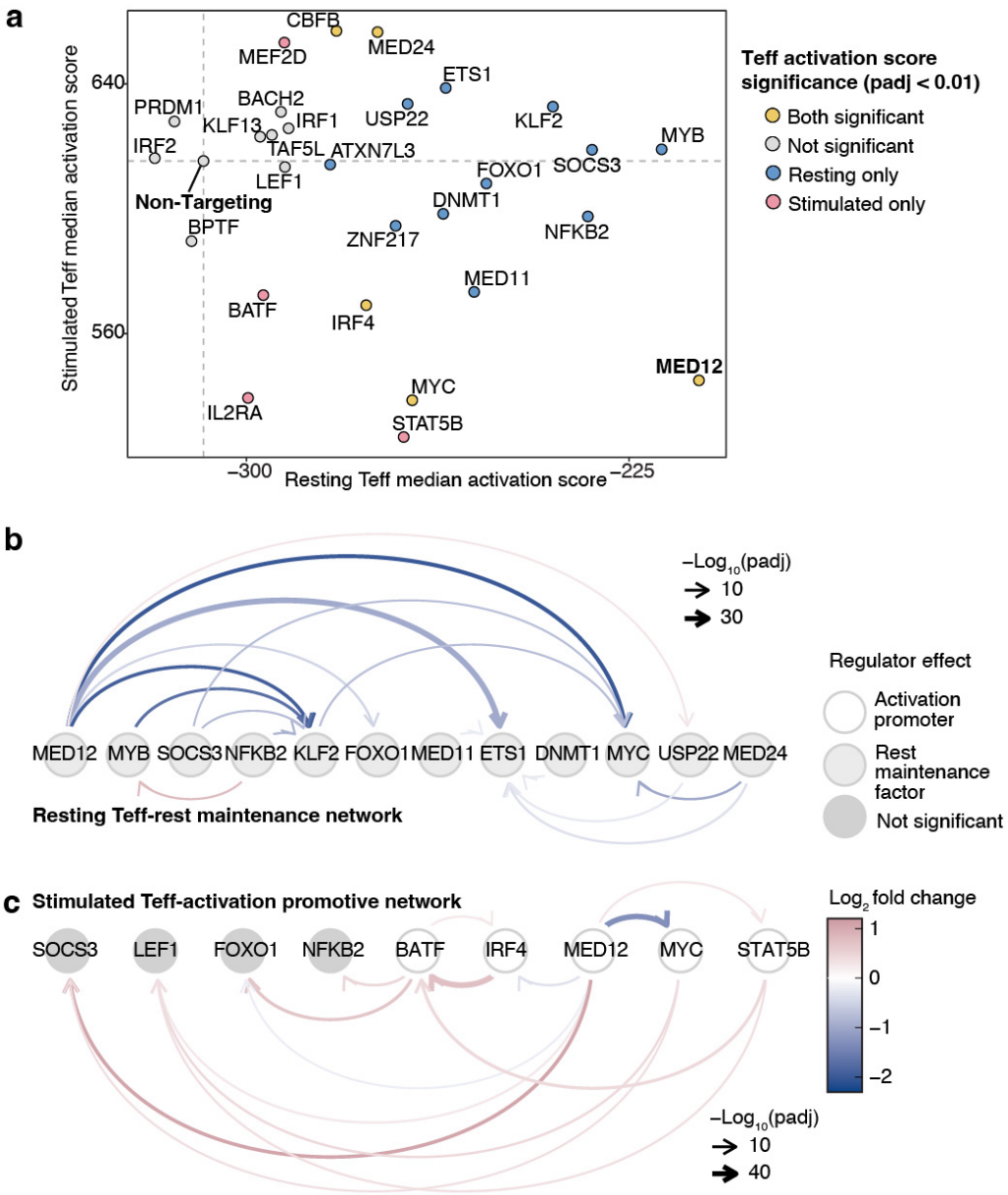


Figure 2.3 Perturb-seq reveals T cell rest and activation networks
a. Activation scores computed for each perturbed gene based on single cell gene signatures across resting and stimulated states. Each point represents the median activation score of cells targeted for CRISPRi knock-down of the indicated gene. Dashed grey lines indicate the activation scores for non-targeting control cells; colored points indicate perturbation with activation scores significantly different than control cells for each condition determined by a two-sided Wilcoxon rank sum test with continuity correction (padj < 0.01). **b-c.** Regulatory network of factors controlling rest and activation. Differentially expressed genes resulting from a perturbation (identified by pseudo-bulking knock-down vs. non-targeting cells) are represented as arrows from the perturbed gene (Wald test with Benjamini-Hochberg (BH) multiple test correction, padj < 0.05 threshold, n=2 donors per target gene). Light grey nodes indicate rest maintenance factors in resting Teffs, white nodes indicate activation-promoting factors in stimulated Teffs, and dark grey nodes indicate regulators without significant effects on activation scores in stimulated Teffs (categorization from a).

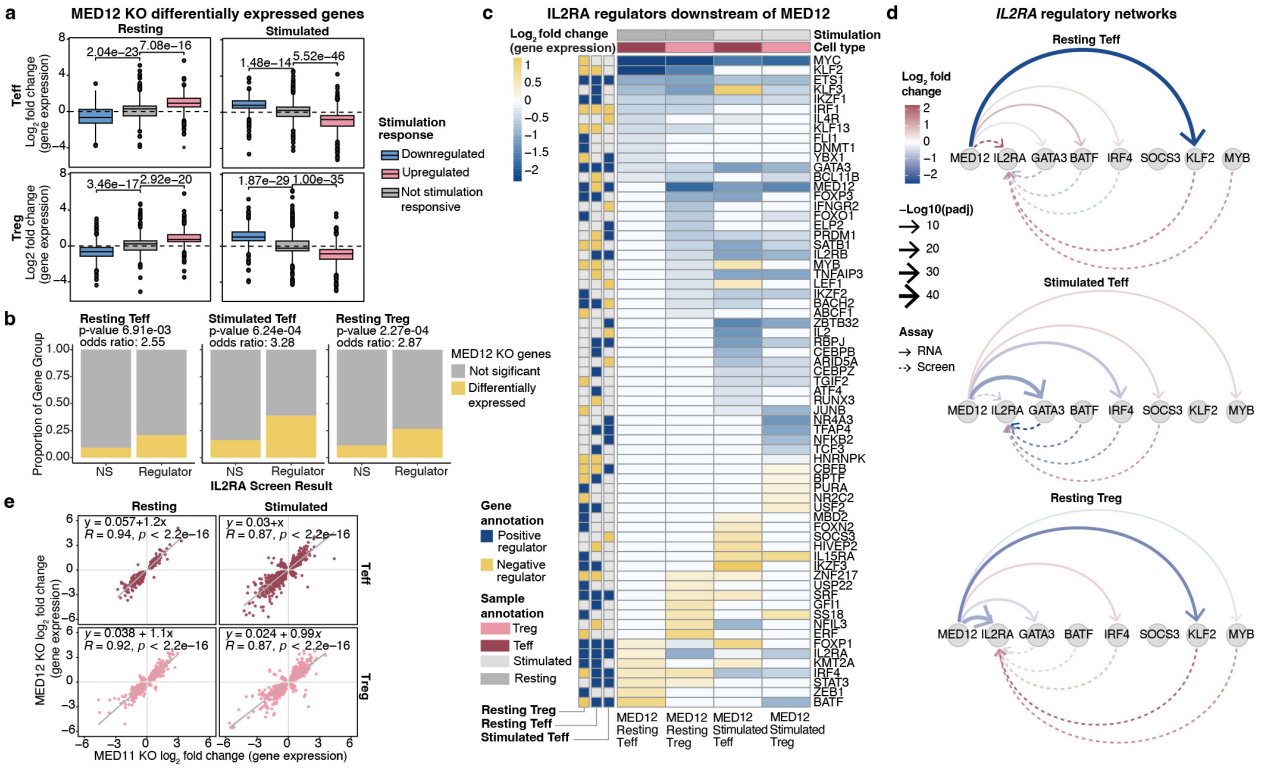


Figure 2.4 MED12 coordinates expression of IL2RA regulators across CD4+ T cell conditions

a. Genes differentially expressed in MED12 KO samples compared to control AAVS1 KO samples (Wald test with BH multiple test correction, padj < 0.05, n=3 donors per KO) are grouped according their stimulation-responsive behavior in AAVS1 KO control cells. The Bonferroni adjusted p value resulting from a two-tailed T test is displayed comparing each stimulation response group to the non-stimulation responsive group (Methods). **b.** Proportions of IL2RA regulators vs. non-regulators (NS) whose expression levels are affected by MED12 KO. One-sided Fisher's exact test for regulators of IL2RA downstream of MED12 (Methods). **c.** Heatmap of IL2RA regulators differentially expressed between MED12 KO cells and control cells (as described in a). Gene annotation boxes represent the IL2RA screens result (FDR < 0.05; navy = positive regulator of IL2RA, gold = negative regulator of IL2RA). **d.** Directed network plots depicting select trans-regulators downstream of MED12. Solid lines depict effects of MED12 based on significant gene expression changes as described in a and dashed lines represent effects on IL2RA based on screen results as described in 1e-g. **e.** Comparison of transcriptional effects of MED12 ablation versus ablation of core Mediator subunit MED11. Each point represents the effect on genes significantly differentially expressed in both KOs, as described in a. Linear regression equation and Pearson coefficient provided for each condition.

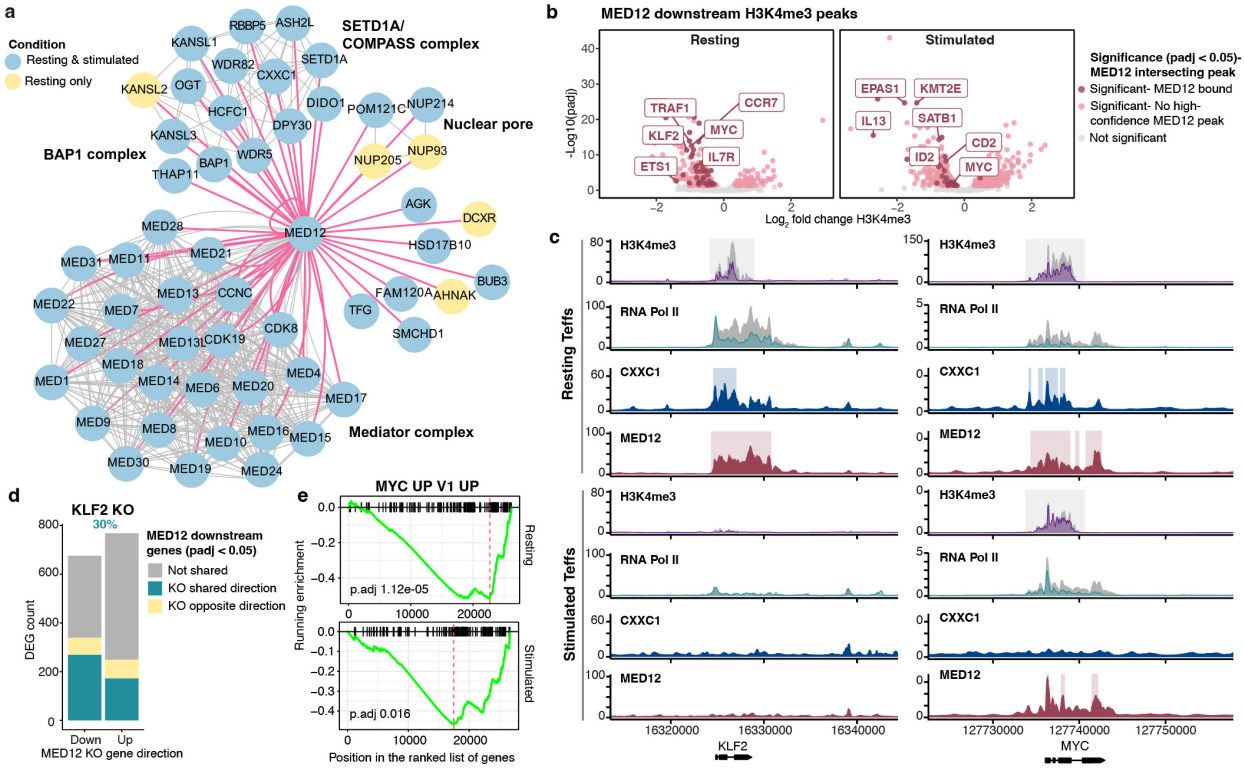


Figure 2.5 MED12 shapes chromatin landscapes to promote cell type and stimulation-specific regulation

a. Proteins enriched in CD4+ Teff MED12 immunoprecipitation mass spectrometry with > 100-fold enrichment relative to IgG control (BFDR <= 0.05, n = 3 donors). Pink lines indicate enrichment in immunoprecipitation and grey lines are reported physical interactions in the STRING database. **b.** Gene loci with H3K4me3 altered by MED12 KO relative to AAVS1 KO control Teffs determined by CUT&RUN (n = 3 donors per condition). Significant regions intersecting MED12 high-confidence ChIP-seq peaks (Methods, n=2 donors per condition), are colored in red with select genes labeled. **c.** *KLF2* and *MYC* loci depicting differential H3K4me3 and RNA Pol II CTD occupancy between the MED12 KO (purple and turquoise) and AAVS1 KO (grey) conditions from a representative donor. Light grey boxes indicate region of differential H3K4me3 between the MED12 KO and AAVS1 KO (padj < 0.05, n=3 donors). Colored boxes indicate CXXC1 peaks and MED12 high-confidence peaks in AAVS1 KO Teffs (n=2 donors). **d.** Differentially expressed genes (DEG) downstream of MED12 and KLF2 in resting CD4+ Teffs as described in 4c. KLF2 regulated genes from Freimer et al. (padj < 0.05)[19]. **e.** GSEA with BH multiple test correction depicting significantly reduced enrichment of MYC signature genes (MYC_UP.V1_UP from msigdb) in MED12 KO cells vs. AAVS1 KO control cells.

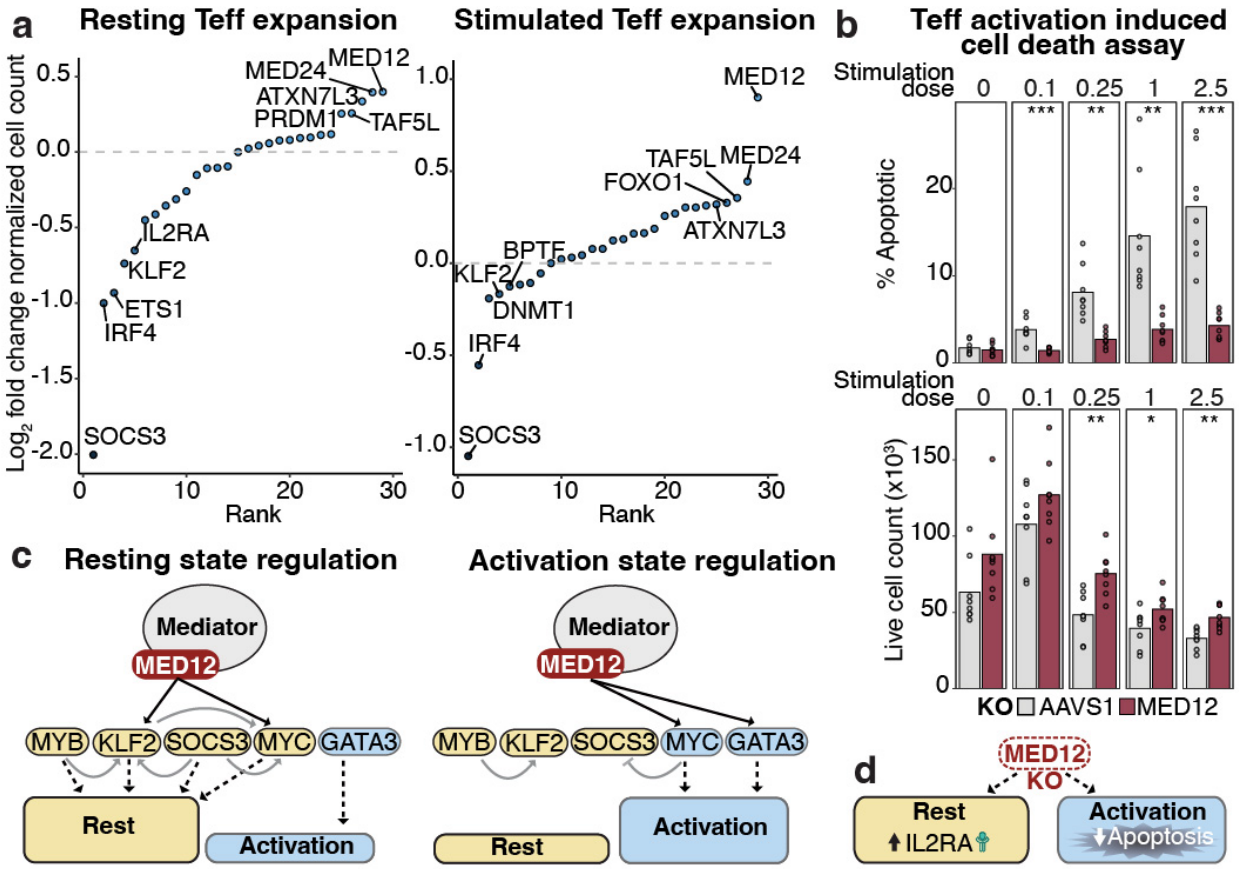
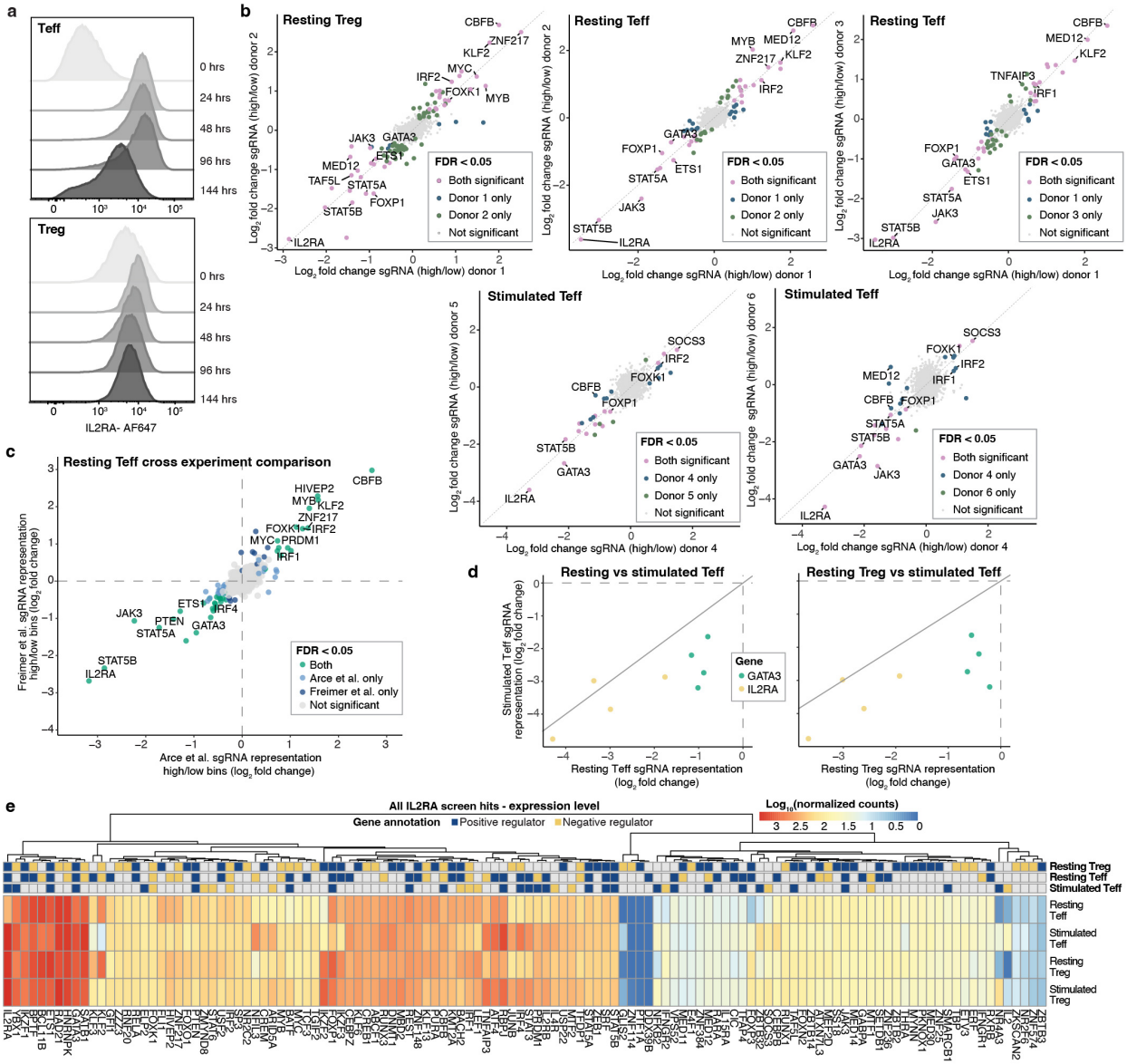


Figure 2.6 MED12 ablation limits activation-induced T cell apoptosis

a. Total cell abundance for each gene knockdown within the indicated Perturb-seq pool of single cells normalized using the sgRNA distribution in the plasmid library and represented as the $\log_2$ fold change compared to non-targeting cells (dashed line). **b.** Percentage of apoptotic cells and live Teff cell count following various dosages of anti-CD3/CD28/CD2 stimulation reported relative to manufacturer recommended dose. Two-tailed T test comparing groups (n=4 donors x 2 sgRNAs per target gene; Apoptosis- Dose 0: p=0.56, 0.1: p=0.00087, 0.25: p=0.0011, 1: p=0.003, 2.5: p=0.00032; Live counts- Dose 0: p=0.68, 0.1: p=0.14, 0.25: p=0.0026, 1: p=0.036, 2.5: p= 0.0017). **c.** Model of core regulatory networks controlling T cell rest and activation, both coordinated by MED12. Solid lines indicate regulatory effects on other factors and dashed lines represent effects on overall states. Solid black lines indicate potential direct regulation by MED12 as supported by ChIP-seq data. **d.** Phenotypic effects of MED12 ablation in CD4+ Teff cells.

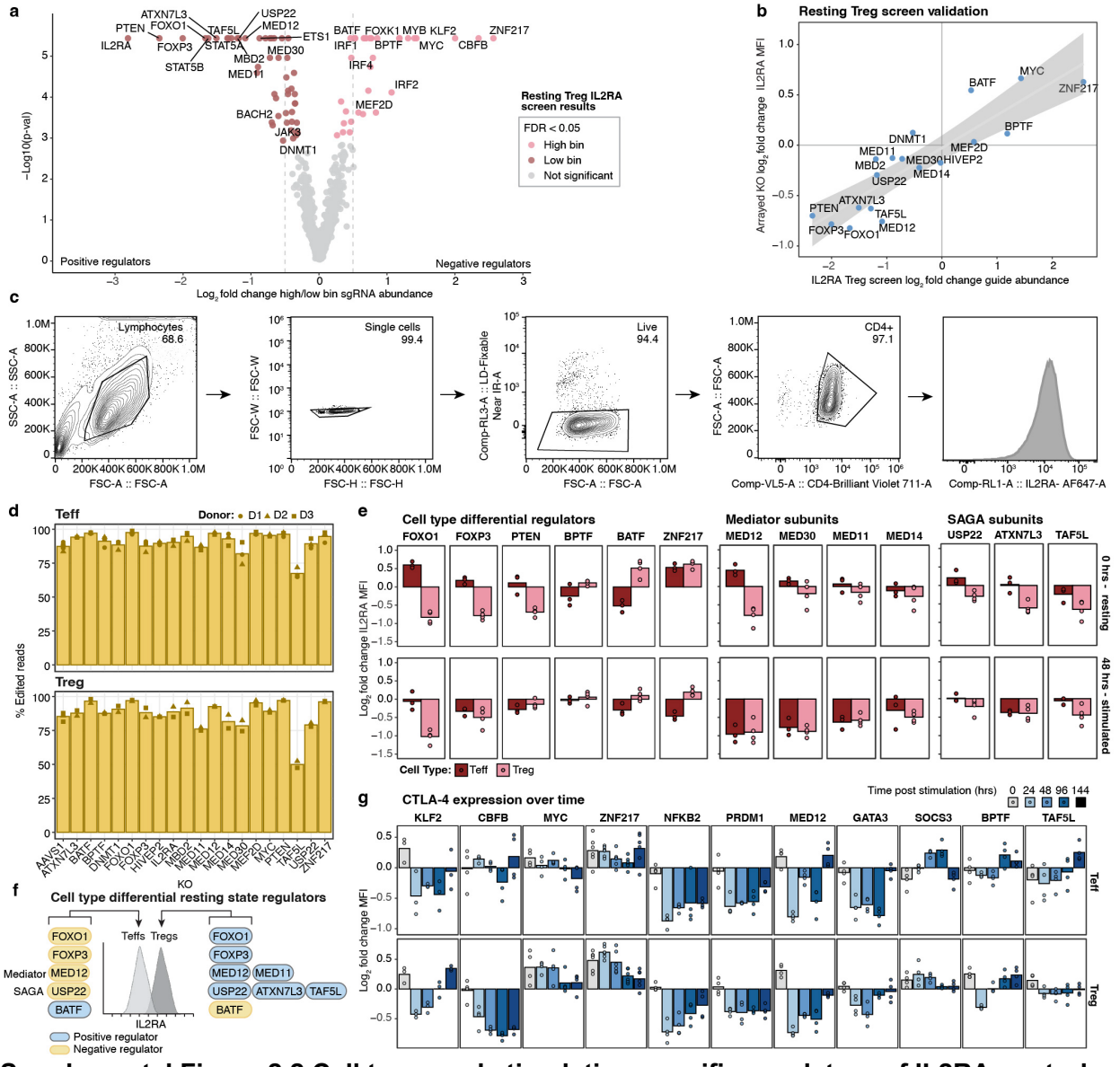
Supplemental Figures



Supplemental Figure 2.1 Pooled KO screens across cell states and lineages reveal context-specific regulators

a. Kinetics of IL2RA expression in Teffs and Tregs following restimulation. Representative histograms of IL2RA expression assessed via flow cytometry adjusted to the mode of each sample. **b.** Donor-to-donor correlations for all screening conditions (FDR < 0.05; Treg screen: n=2, Teff screen: n=3, Stimulated Teff screen: n=3 donors). **c.** Comparison of resting IL2RA screen results between Freimer et al. and new screen data. Non-significant genes are shown in grey and significant hits (FDR < 0.05) colored by direction of effect in both screens. **d.** Comparison of IL2RA screen results for IL2RA KO and GATA3 KO sgRNAs. Each dot represents an individual sgRNA average effect for the respective gene KO (Treg screen: n=2, Teff screen: n=3, Stimulated Teff screen: n=3 donors). **e.** Heatmap showing expression of genes encoding screen hits. Transcript levels were assessed in AAVS1 KO control cells by bulk RNAseq. (Figure caption continued on the next page.)

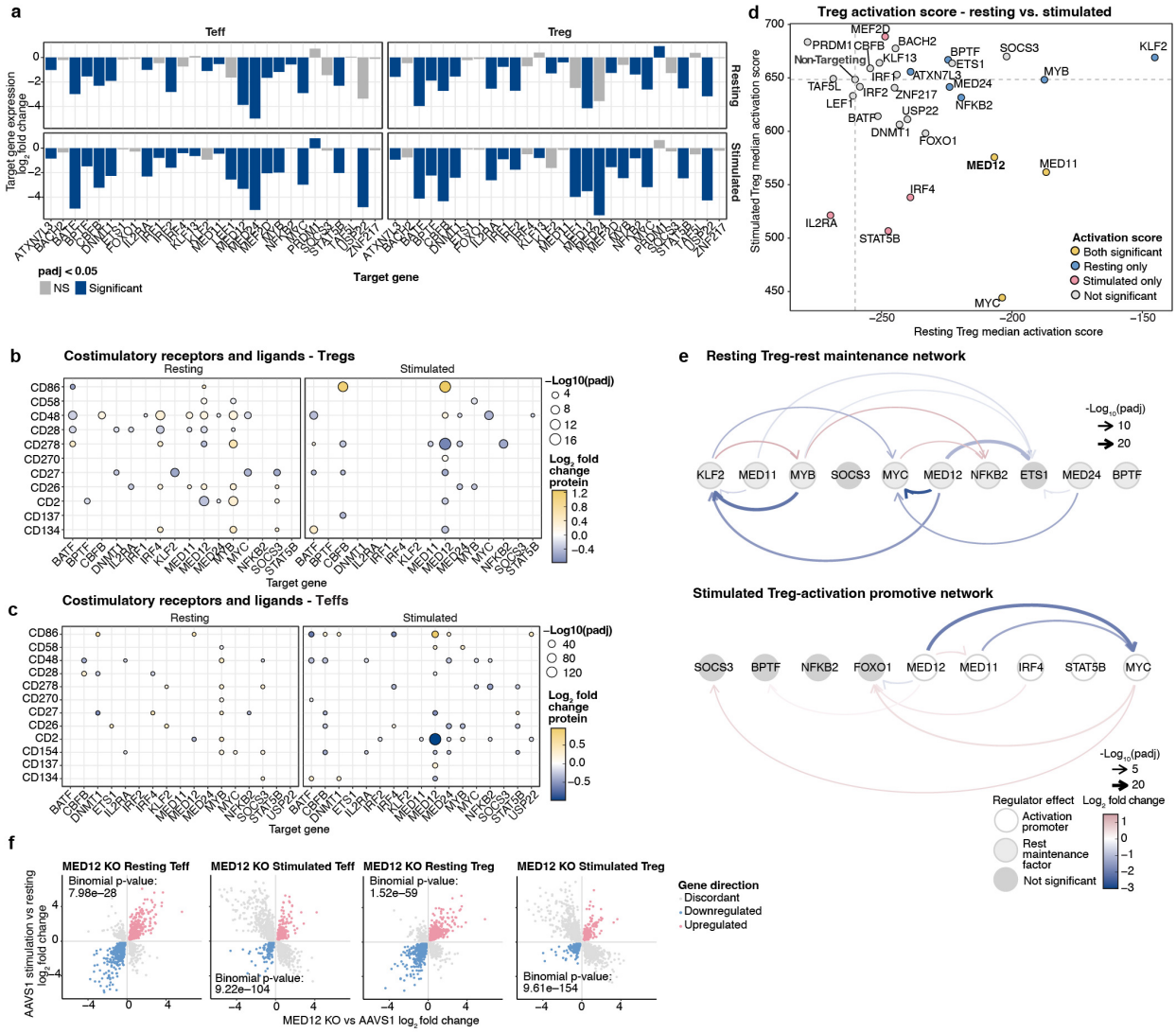
(Figure caption continued on the previous page.) The color bar represents the log10(mean normalized counts) for the gene expression level of each IL2RA regulator across each cell type and stimulation condition (n=3 donors). The annotation bars on the top illustrate the direction of effect for each regulator in the three IL2RA screens (colored boxes = FDR < 0.05, navy = positive regulator of IL2RA and gold = negative regulator of IL2RA).



Supplemental Figure 2.2 Cell type- and stimulation-specific regulators of IL2RA control dynamic gene expression

a. Volcano plot of screen results for regulators of IL2RA in resting primary human Tregs. Significant hits (FDR < 0.05, n=2 donors) colored by direction of effect. **b.** Validation of select Treg screen hits using arrayed KO and flow cytometry. IL2RA expression displayed as the mean $\log_2$ fold change median fluorescent intensity (MFI) of the perturbed samples compared to control AAVS1 KO control samples vs. the $\log_2$ fold change IL2RA Treg screen effect (flow cytometry n=4 donors; Treg screen n=2 donors). **c.** Flow cytometry gating strategy for IL2RA expression in arrayed assays displayed as contour plots with outliers. (Figure caption continued on the next page.)

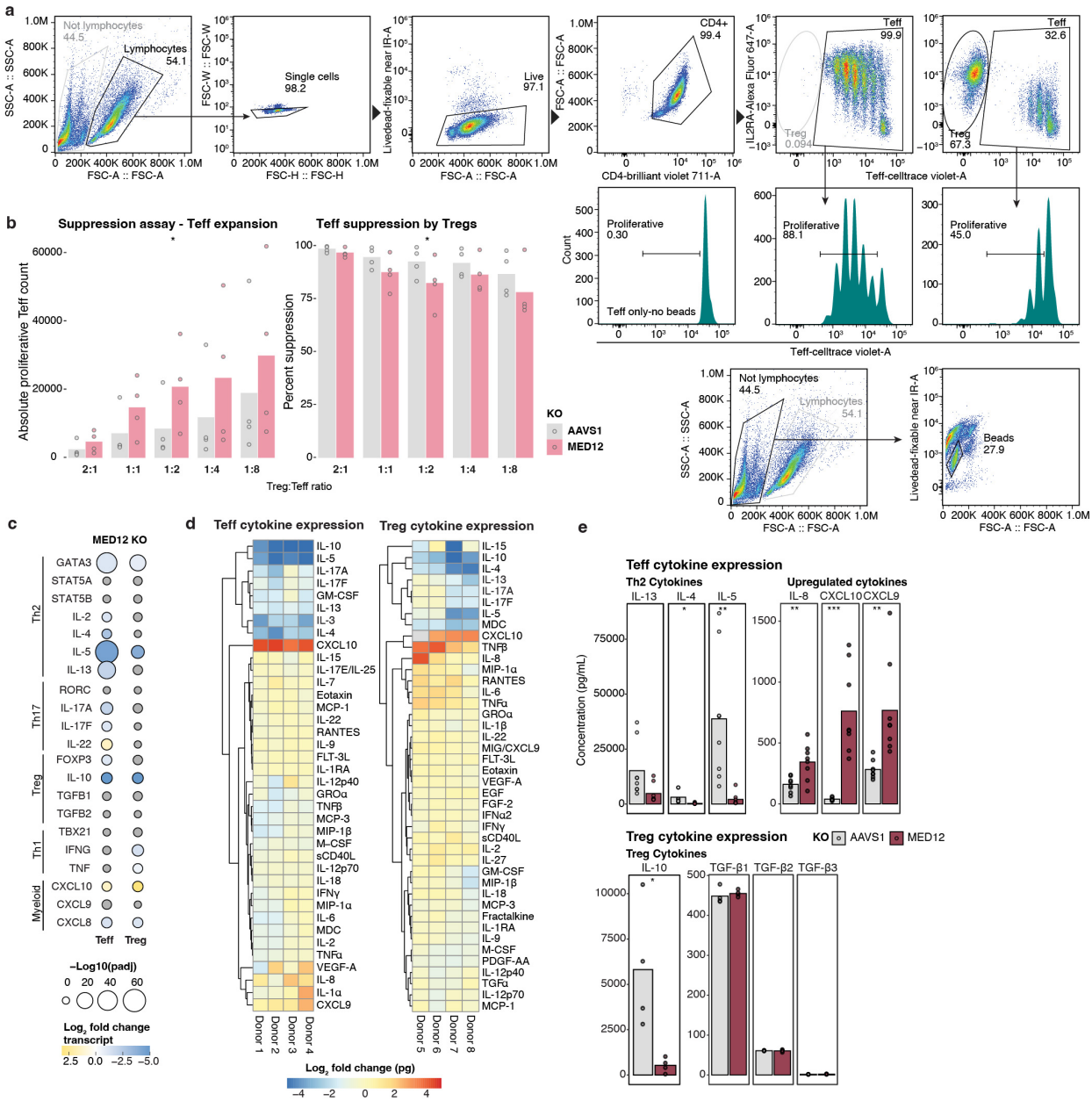
(Figure caption continued on the previous page.) **d.** Amplicon-seq genotyping of arrayed validation KOs to confirm editing. The mean percent modified (edited) reads as quantified by NGS is shown on the y axis for each of the targeted genes. (Teff genotyping n=3, Treg genotyping n=2 donors) **e.** Select regulators of IL2RA demonstrate cell type-specific effects. Regulators from arrayed KO in b selected for visualization based on apparent cell type differential effect and membership to Mediator or SAGA. IL2RA surface expression displayed as the $\log_2$ fold change median fluorescent intensity (MFI) of the perturbed samples compared to control AAVS1 KO sample from the same donor. (Teffs: n=3 donors, Tregs: n=4 donors). **f.** Schematic of IL2RA screen hits with cell type-differential regulatory roles **g.** Intracellular CTLA-4 expression is affected by perturbation of stimulation-responsive regulators of IL2RA. Data displayed as described in e., but for CTLA-4 expression. (n = 2 donors x 2 sgRNAs per KO except KLF2 KO where n = 3 and ZNF217 KO where n = 6) ZNF217 KO and AAVS1 KO data of CTLA-4 for time 0 Tregs was published in Mowery et al[42]. Gating strategy is as displayed in c.



Supplemental Figure 2.3 MED12 controls expression of stimulation-responsive genes in Tregs and Teffs

a. Differential gene expression results for CRISPRi targeted genes using pseudobulked Perturb-seq counts relative to non-targeting control cells (Wald test with BH multiple test correction, padj < 0.05, n=2 donors). **b-c.** Differentially expressed cell surface proteins from pseudobulked Perturb-CITE-seq samples (Wald test with Benjamini-Hochberg (BH) multiple test correction, padj < 0.05, n=2 donors). **d.** Activation scores for each perturbed gene computed based on single cell gene signatures. Each point represents the median activation score of cells targeted for CRISPRi knock-down of the indicated gene in Tregs. Dashed grey lines indicate the activation score for non-targeting control cells within each respective condition; colored points indicate perturbation with activation scores significantly different than control cells for each condition as determined by a two-sided Wilcoxon rank sum test with continuity correction (padj < 0.01). **e.** Regulatory network of factors controlling rest and activation in Tregs. Differentially expressed genes resulting from a perturbation (identified by pseudo-bulking knock-down vs. non-targeting cells) are represented as arrows from the perturbed gene (padj < 0.05, n=2 donors per target gene). (Figure caption continued on the next page.)

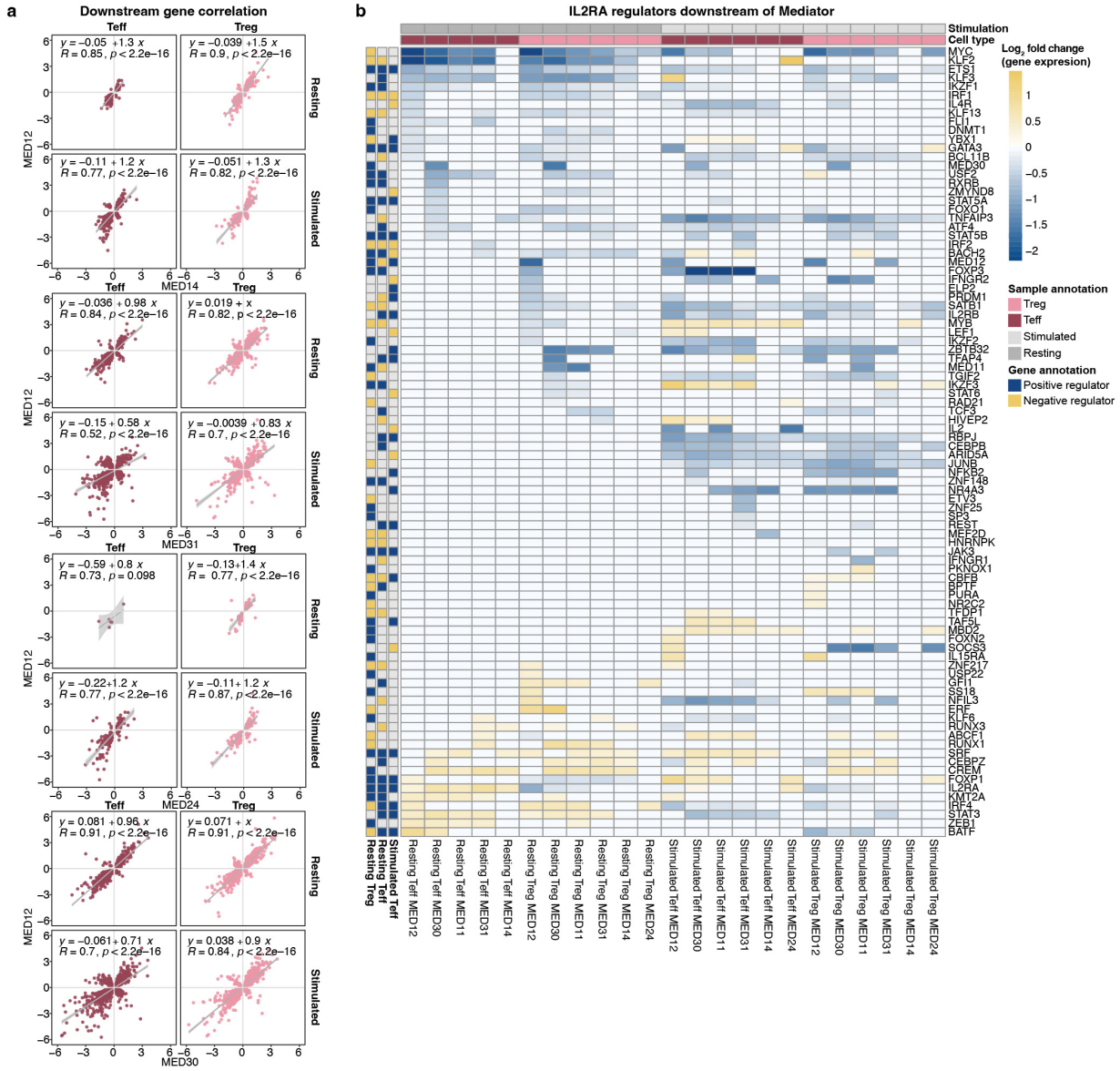
(Figure caption continued on the previous page.) Light grey nodes indicate rest maintenance factors in resting Tregs, dark grey nodes indicate regulators without significantly different activation scores in the respective condition, and white nodes indicate activation promoting regulators in stimulated Tregs **f.** Log_2 fold change of differentially expressed genes in MED12 KO vs AAVS1 KO bulk RNAseq samples (padj < 0.05 as described in 4a) compared to the log_2 fold change of differentially expressed genes between stimulated and resting control AAVS1 KO cells (Wald test with Benjamini-Hochberg (BH) multiple test correction, padj <0.05, n=3 donors). Two-sided binomial test results are displayed comparing the proportion of genes downstream of MED12 that are concordant in direction with stimulation-responsive genes to genes discordant in direction.



Supplemental Figure 2.4 MED12 is required for distinct functional features in specific CD4+ T cell subsets

a. Flow gating strategy for suppression assays. **b.** *In vitro* Treg suppression assays showing absolute proliferative Teff count and percent suppression by Tregs (Methods). Unedited (wild type) CD4+ Teffs were used in the assay with MED12 KO Tregs or AAVS1 KO control Tregs. Paired two-tailed T test comparing MED12 KO and AAVS1 KO control samples (n=4 donors per KO; Teff count: Ratio 2:1 p=0.11, 1:1 p=0.067, 1:2 p=0.023, 1:4 p=0.12, 1:8 p=0.15; Suppression: Ratio 2:1 p=0.067, 1:1 p=0.061, 1:2 p=0.045, 1:4 p=0.24, 1:8 p=0.14). **c.** Gene expression of selected genes associated with indicated cell identities. Color indicates the $\log_2$ fold change of differentially expressed genes in bulk RNAseq (Wald test with BH multiple test correction, padj <0.05, n=3 donors) comparing stimulated MED12 KO to AAVS1 KO control samples. Data are shown in Teffs (left) and Tregs (right). **d.** Cytokine expression measured by Luminex. (Figure caption continued on the next page.)

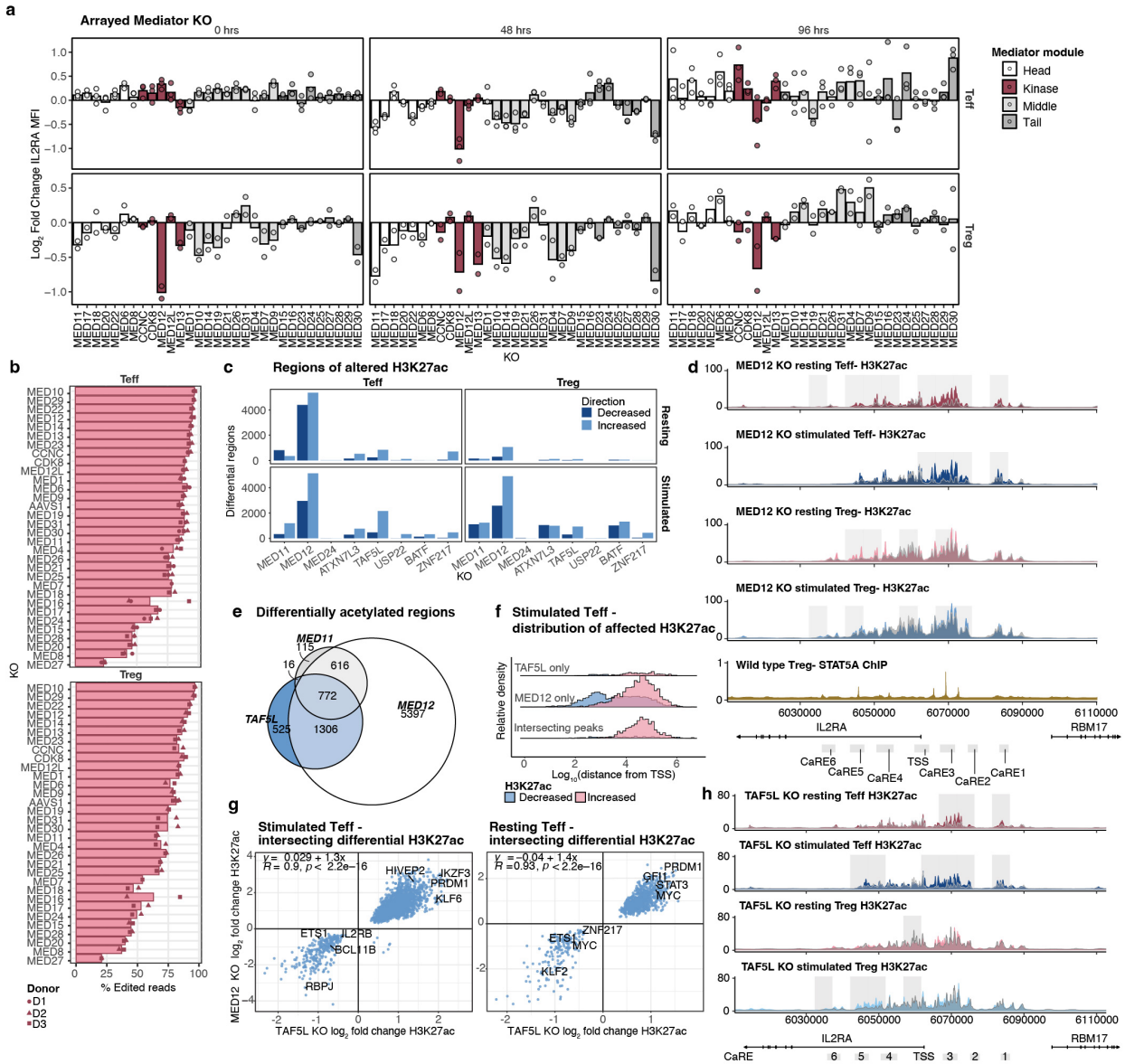
(Figure caption continued on the previous page.) Heatmaps represent the log_2 fold change cytokine concentration in the MED12 KO sample supernatant relative to AAVS1 KO control supernatant. Teff heatmap values display the average concentration of 2 sgRNAs per gene KO (n=4 donors per cell type x 2 sgRNA for Teff or 1 sgRNA for Tregs). **e.** Cytokine concentrations as represented in d for select cytokines. Two-tailed T test comparing MED12 KO and AAVS1 KO samples (paired for Tregs only) (n=4 donors x 2 sgRNAs per KO (Teff only); Teff cytokines: IL-13: p=0.054, IL-4: p=0.025, IL-5: p=0.0099, IL-8: p=0.0095, CXCL10 (IP-10): p=0.00072, CXCL9: p=0.0096; Treg cytokines: IL-10: p=0.043, TGF-β1: p=0.54, TGF-β2: p=0.84, TGF-β3: p=0.19).



Supplemental Figure 2.5 Partially-shared transcriptional effects of MED12 and other Mediator subunits

a. Comparison of the transcriptional effects of MED12 KO compared to KO of Mediator subunits MED14 (middle/backbone), MED31 (middle), MED24 (tail), MED30 (tail) as described in Figure 4e. **b.** Regulators of IL2RA in the network downstream of several Mediator complex subunits. (Figure caption continued on the next page.)

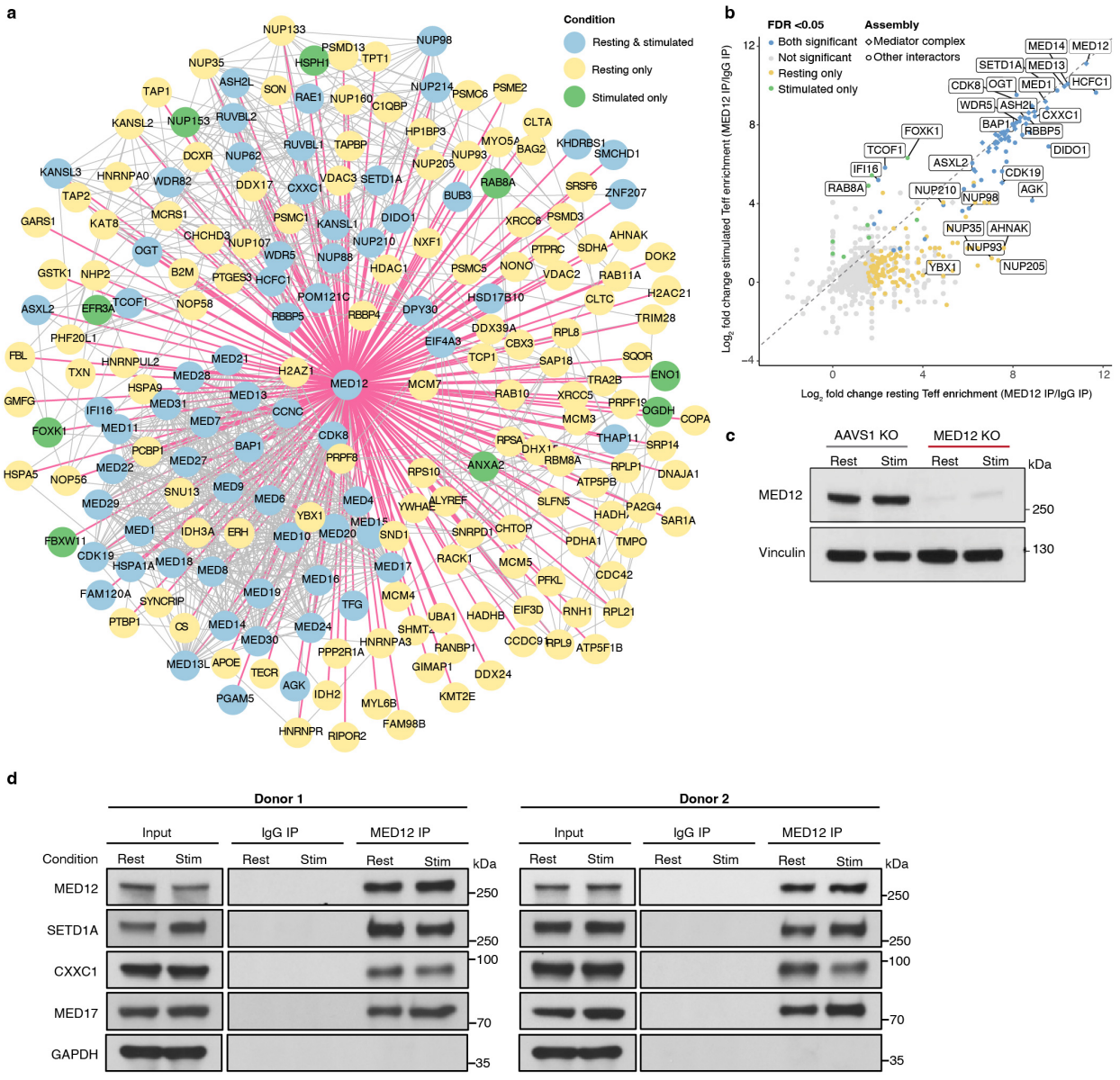
(Figure caption continued on the previous page.) Differentially expressed genes as assessed by bulk RNAseq are displayed as the $\log_2$ fold change gene expression in the subunit KO/AAVS1 KO samples for the respective condition (Wald test with Benjamini-Hochberg (BH) multiple test correction, padj < 0.05, n=3 donors per KO). The horizontal annotation bars on the top of the figure represent the stimulation condition (dark grey for resting and light for restimulated) and the vertical annotation bars represent the effect of the regulator in the IL2RA KO screens (colored boxes = statistically significant FDR < 0.05, navy = positive regulator of IL2RA and gold = negative regulator of IL2RA).



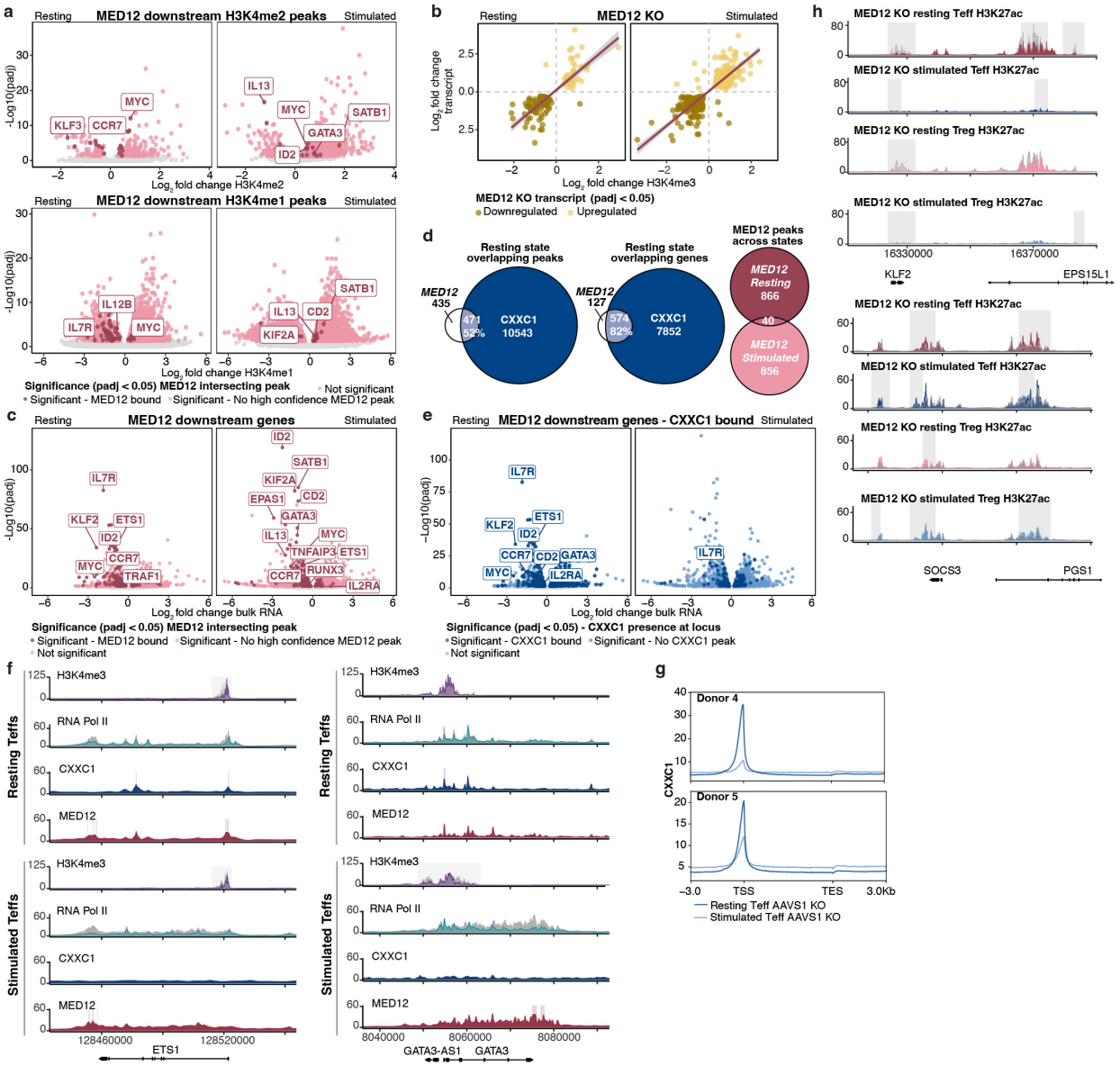
Supplemental Figure 2.6 Mediator and SAGA complexes shape context-dependent regulation of IL2RA

a. Mediator complex subunit KO effects on expression of IL2RA as quantified by flow cytometry across contexts. Each color represents a Mediator complex structural module. (n=3 donors for Teffs and 2 donors for Tregs x 1 sgRNA per condition) (Figure caption continued on the next page.)

(Figure caption continued on the previous page.) **b.** Amplicon-seq genotyping of arrayed validation KOs to confirm editing. The mean percent modified (edited) reads as quantified by NGS is depicted for each of the targeted genes. (Teff genotyping n=3 donors, Tregs genotyping n=2 donors) **c.** Select regulators of IL2RA affect H3K27ac distribution in CD4+ T cells. The number of differentially H3K27 acetylated regions in each KO cell population compared to control AAVS1 KO cells is displayed on the y axis (Two-tailed Wald test with BH multiple test correction, padj < 0.05, n=2 donors per KO). The direction in which H3K27ac was altered is depicted as the color of the bar. **d.** Trackplot of the *IL2RA* locus depicting regions of differential acetylation between the MED12 KO (solid color) and AAVS1 control KO (grey transparent) conditions from a representative donor. Light grey boxes distinguish regions of significantly differential acetylation between the MED12 KO and the AAVS1 KO (Two-tailed Wald test with BH multiple test correction, padj < 0.05, n=2 donors per KO). *IL2RA* CaRE enhancer elements[6] are annotated in grey boxes below gene tracks. STAT5A ChIP data sourced from public data (Methods). **e.** Venn diagram depicting differentially acetylated regions (relative to AAVS1 KO control cells) for MED12, TAF5L, and MED11 KO samples. Differentially acetylated regions determined as described in c. **f.** Histogram depicting the distribution of differentially acetylated regions based on distance to the transcription start site of the nearest gene. The peak height is the proportional to the number of differentially acetylated regions across the samples. **g.** Correlation of TAF5L and MED12 differentially acetylated regions. Regions as described in c. depicted as the $\log_2$ fold change acetylation for the respective perturbations. **h.** Trackplot of H3K27ac as described in d for TAF5L KO instead of MED12 KO.



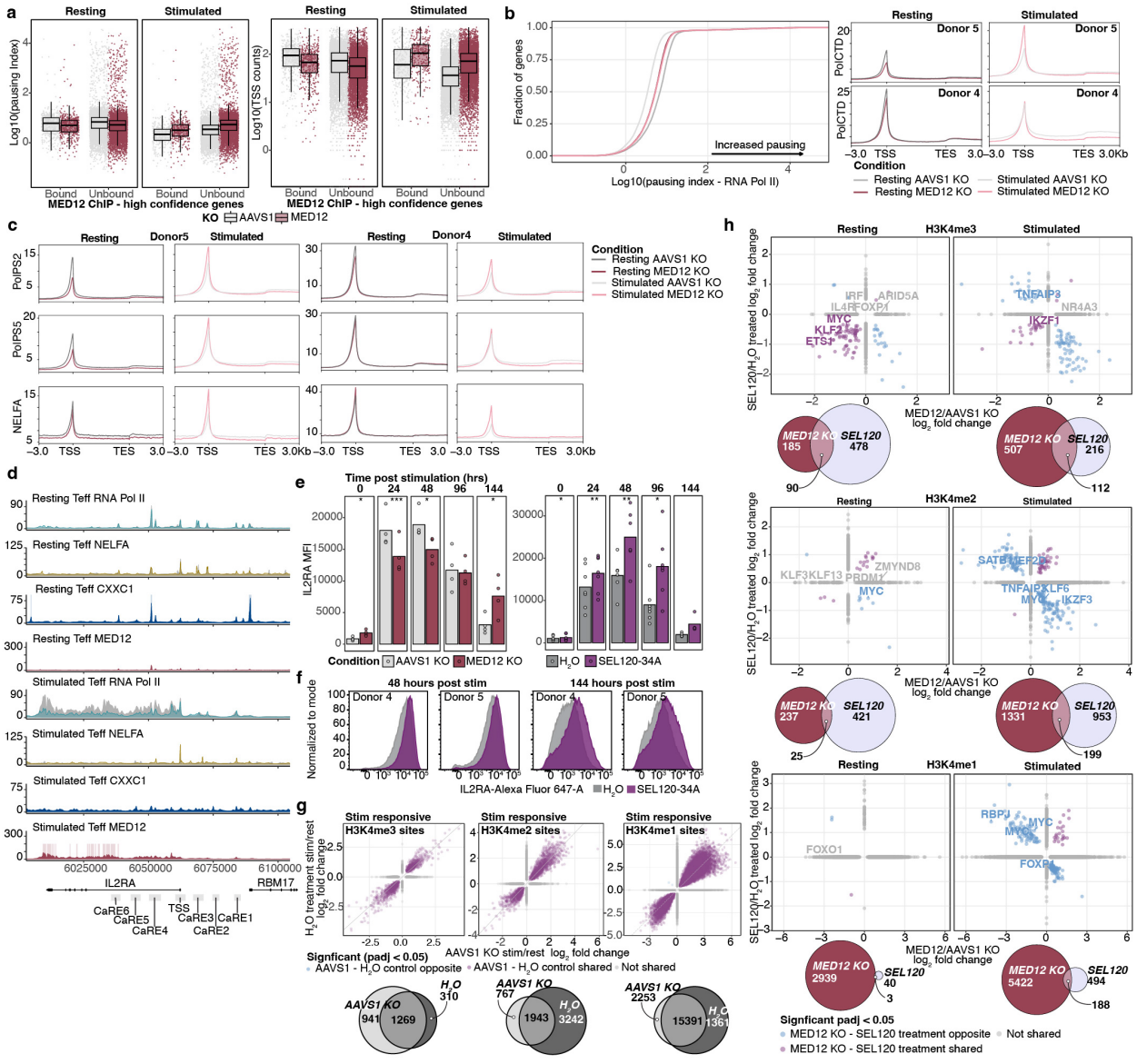
Supplemental Figure 2.7 SETD1A/COMPASS complex members are enriched within an extensive MED12 protein interaction network within CD4+ Teffs.
a. MED12 interaction partners in CD4+ Teffs. Proteins enriched in MED12 immunoprecipitation mass spectrometry (IP-MS) relative to IgG control (BFDR <= 0.05, n = 3 donors). Pink lines indicate enrichment in immunoprecipitation and grey lines are derived from reported physical interactions in STRING database[64]. **b.** Proteins from a. plotted by $\log_2$ fold change enrichment demonstrate high representation of SETD1A/COMPASS members. **c.** Western blot confirmation of MED12 protein loss with MED12 KO sgRNA s2770 targeting MED12 using antibody clone D9K5J used in IP experiments. Two donors were processed in one experiment. For gel source data, see Supplemental Figure 1. **d.** MED12 IP validation of select interactors by western blot. Western blot of IgG and MED12 IP samples demonstrating interaction between MED12 and SETD1A, CXXC1, and MED17 in resting and stimulated Teffs in two human donors. For gel source data, see Supplemental Figure 1.



Supplemental Figure 2.8 MED12 ablation disrupts chromatin at regulators of rest and activation

a. Sites of H3K4me2/1 altered by MED12 KO relative to AAVS1 KO Teffs (n = 3 donors per condition) determined by CUT&RUN. Significant regions intersecting high-confidence MED12 peaks, assessed by ChIP-seq (Methods, n=2 donors per condition), are colored in red with select genes labeled. **b.** Correlation between genes with differential H3K4me3 and transcript expression comparing MED12 KO to AAVS1 KO Teffs. **c.** Genes differentially expressed in MED12 KO relative to AAVS1 KO (n = 3 donors per condition) colored as in a. **d.** Intersection of MED12 high-confidence peaks and CXXC1 peaks (left plot) and MED12 high-confidence bound genes and CXXC1 bound genes (middle plot) in resting Teffs. Intersection of MED12 high-confidence peaks between resting and stimulated states (right plot, n=2 donors per condition) **e.** Genes differentially expressed in MED12 KO relative to AAVS1 KO Teffs. CXXC1 bound genes determined by ChIP-seq (Methods, n=2 donors per condition) are colored in dark blue with select genes labeled. (Figure caption continued on the next page.)

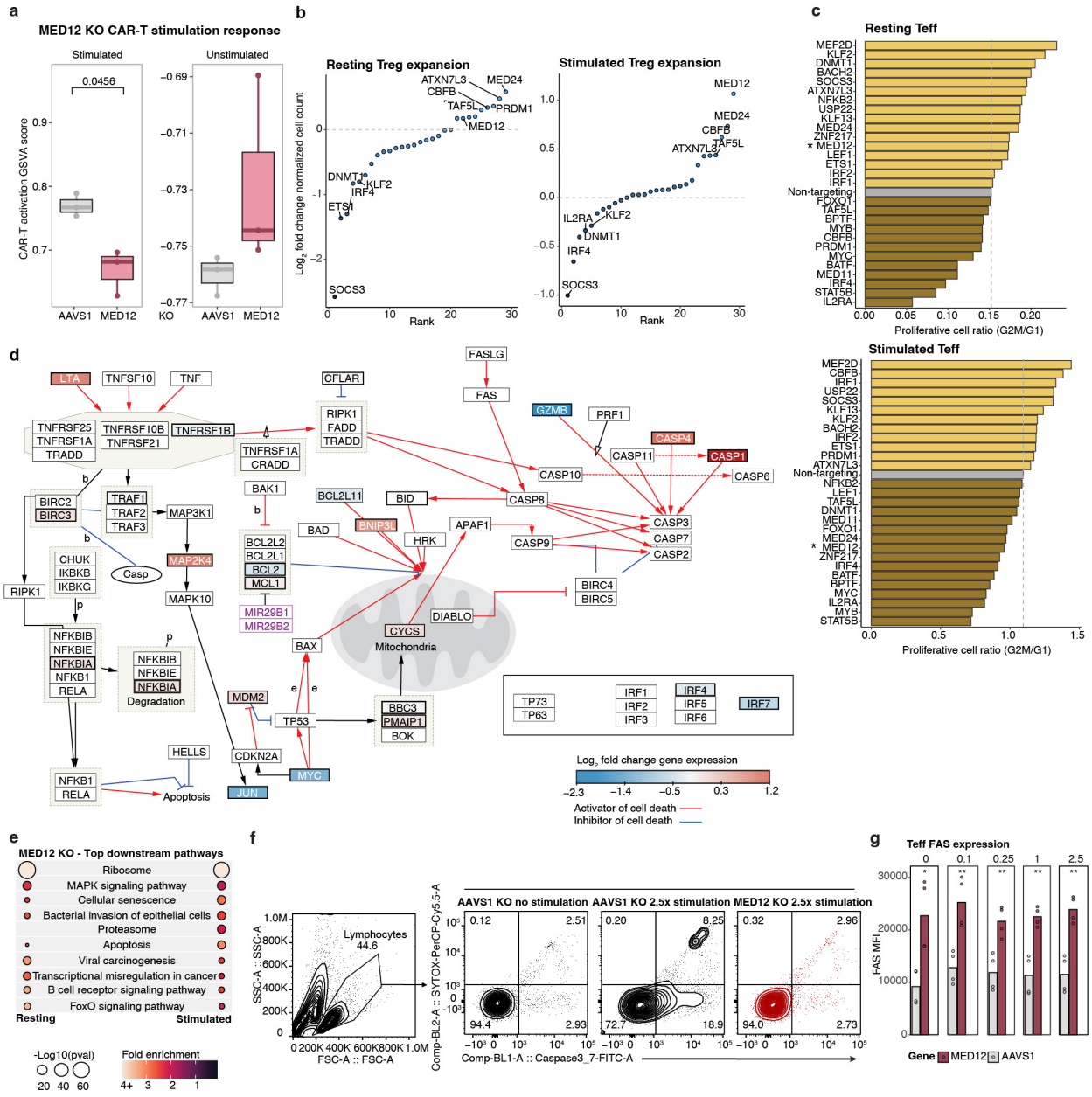
(Figure caption continued on the previous page.) **f.** Trackplots of *ETS1* and *GATA3* loci depicting differential H3K4me3 and RNA Pol II CTD occupancy between the MED12 KO (purple and turquoise) and AAVS1 KO control (grey) conditions from a representative donor. Light grey boxes define significantly differential H3K4me3 peaks comparing MED12 KO and AAVS1 KO control (padj < 0.05, n=3 donors). Colored boxes indicate CXXC1 peaks and MED12 high-confidence peaks (padj < 0.05, n=2 donors, Methods). **g.** CXXC1 binding distribution at expressed genes in CD4+ Teffs determined via ChIP-seq **h.** Trackplots of *KLF2* and *SOC3* loci depicting regions of differential H3K27ac between the MED12 KO (solid color tracks) and AAVS1 control KO (grey transparent tracks) conditions from a representative donor. Light grey boxes distinguish regions of significantly differential acetylation between the MED12 KO and the AAVS1 KO control (padj < 0.05, n=2 donors).



Supplemental Figure 2.9 MED12 ablation results in widespread changes in polymerase pausing

a. Mean Pausing Index (left) and polymerase abundance at the TSS's depicted for all genes expressed in Teffs (n=2 donors per condition). (Figure caption continued on the next page.)

(Figure caption continued on the previous page.) Boxplots are separated by presence or absence of MED12 high-confidence peak(s) at the loci. Boxplot center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. **b.** Empirical cumulative distribution function plot displaying the mean pausing index for all genes expressed in Teffs (n=2 donors per condition, left). RNA Pol CTD distribution across genes (right). **c.** Teff ChIP-seq binding distributions across genes expressed in Teffs (n=2 donors per condition). **d.** Trackplot of the *IL2RA* locus depicting differences in occupancy of RNA Polymerase II CTD and NELFA in MED12 KO (turquoise and gold) and AAVS1 control KO (grey tracks) conditions from a representative donor. Colored boxes indicate CXXC1 (blue) peaks and MED12 high-confidence peaks (red, Methods). **e.** IL2RA expression of CRISPR perturbed or kinase inhibitor (SEL120-34A) treated samples. Statistics performed using a paired two-tailed T test (MED12 KO: n=4 donors, SEL120-34A: n=7 donors all timepoints except T144 where n=4 donors and T48 where n=6 donors for SEL120-34A assay; MED12 KO assay: 0 hrs: p=0.015, 24 hrs: p=0.00013, 48 hrs: p=0.025, 96 hrs: p=0.75, 144 hrs: p=0.02; SEL120-34A assay: 0 hrs: p=0.045, 24 hrs: p=0.002, 48 hrs: p=0.0015, 96 hrs: p=0.024, 144 hrs: p=0.052). **f.** Representative flow plots from e. **g.** Stimulation responsive histone methylation shared between AAVS1 KO control cells and vehicle-treated control cells. Top plot compares significantly differentially methylated sites in AAVS1 KO resting vs stimulated Teffs (x axis) and vehicle (H_2O)-treated resting vs stimulated Teffs (y axis). Euler plots below depict the overlap of all differentially methylated stimulation responsive sites displayed in the scatter plot (padj < 0.05 significance threshold, n=3 donors per condition) **h.** Top plot compares significantly differentially methylated sites in MED12 vs AAVS1 KO Teffs (x axis) and SEL120-34A vs vehicle (H_2O)-treated Teffs (y axis). Euler plots below depict the overlap of all differentially histone methylated sites displayed in scatter plot (padj < 0.05, n=3 donors per condition).



Supplemental Figure 2.10 Activation-induced cell death pathways are dysregulated in MED12 KO T cells

a. MED12 KO CAR-T gene expression activation scores based on gene expression signature analysis of RNAseq data from Freitas et al. (GEO series GSE174279). The GSVA activation score for each group of samples is compared using a two-tailed T test and the Bonferroni adjusted p value displayed for significantly different groups. Boxplot center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range (n=3 donors per KO). **b.** Rank plot of cell abundance within the Perturb-seq single cell pools for each CRISPRi gene target for resting and stimulated Tregs. Cell abundance normalized by the sgRNA distribution in the plasmid library and represented as the $\log_2$ fold change compared to non-targeting control cells. The dashed line indicates the abundance of non-targeting control cells. **c.** Proliferative cell ratios assessed by gene signature of cycling cells within each CRISPRi regulator knockdown condition in the Perturb-seq pool. (Figure caption continued on the next page.)

(Figure caption continued on the previous page.) The estimated ratio of G2M/G1 cells within each condition is represented on the x axis. * Used to denote MED12. **d.** Differentially expressed genes (padj < 0.05) from stimulated Teff MED12 KO vs AAVS1 KO control cell bulk RNAseq data are illustrated on a schematic of the Apoptosis-Homo sapiens pathway from WikiPathways. **e.** Top pathways affected by MED12 KO in bulk RNAseq data. Apoptosis pathway genes were dysregulated by MED12 KO in stimulated Teffs. **f.** Flow cytometry gating strategy for activation-induced cell death (AICD) apoptosis assays. **g.** FAS cell surface expression in MED12 KO and AAVS1 KO control cells (FAS MFI as quantified by flow cytometry is compared across perturbation conditions). Stars indicate significantly different MFI in MED12 KO cells compared to control AAVS1 KO cells using a two-tailed T test (n=2 donors x 2 sgRNAs per gene; Dose 0: p=0.019, 0.1: p=0.0071, 0.25: p =0.0056, 1: p=0.0047, 2.5: p=0.0018).

Tables

Table 2.1 Significant screen hits across all screens
Genes where FDR < 0.05 according to MAGeCK gene level results. Reported as negative regulator of IL2RA or Positive regulator of IL2RA.

id	Stimulated Teff	Resting Teff	Resting Treg
STAT5B	Positive	Positive	Positive
IL2RA	Positive	Positive	Positive
PTEN	Positive	Positive	Positive
STAT5A	Positive	Positive	Positive
FOXP1	Positive	Positive	Positive
ETS1	Positive	Positive	Positive
JAK3	Positive	Positive	Positive
GATA3	Positive	Positive	Positive
KMT2A	NS	Positive	Positive
RELA	NS	Positive	Positive
BATF	Positive	Positive	Negative
CIC	NS	Positive	Positive
SETDB1	NS	Positive	Positive
ZNF148	NS	Positive	Positive
VPS52	Positive	Positive	NS
ZBTB14	NS	Positive	NS
IKZF1	NS	Positive	Positive
GFI1	NS	Positive	NS
SRF	Positive	Positive	Positive
DDX39B	NS	Positive	NS
ZNF236	NS	Positive	NS
BPTF	NS	Positive	Negative
STAT3	Positive	Positive	NS

id	Stimulated Teff	Resting Teff	Resting Treg
RBPJ	Positive	Positive	NS
IL2RB	Positive	Positive	NS
BACH2	Negative	Positive	Positive
TFAP4	Positive	Positive	NS
GABPA	NS	Positive	NS
CEBPZ	NS	Positive	NS
TCF3	NS	Positive	NS
IKZF3	NS	Positive	Positive
IRF4	Positive	Positive	Negative
REST	Positive	Positive	NS
KLF3	NS	Positive	NS
RXRB	NS	Positive	Positive
FOXP3	NS	Positive	Positive
RARA	NS	Positive	NS
CEBPB	NS	Positive	NS
USF2	NS	Positive	Positive
ATF4	NS	Positive	NS
ZNF25	NS	NS	Positive
ELP2	Positive	NS	NS
FLI1	NS	NS	Positive
IL15RA	NS	NS	Positive
PTF1A	NS	NS	Positive
ZBTB3	NS	NS	Positive
PURA	NS	NS	Positive
ZEB1	NS	NS	Positive
MYNN	NS	NS	Positive
MBD2	NS	NS	Positive
THRA	NS	NS	Positive
CREM	NS	NS	Positive
CREB1	NS	NS	Negative
PKNOX1	NS	NS	Positive
YBX1	Positive	NS	Negative
RNF20	NS	NS	Negative
FOXO1	NS	NS	Positive
ZBTB2	NS	NS	Positive
ZMYND8	Negative	NS	NS
IL2	Negative	NS	NS

id	Stimulated Teff	Resting Teff	Resting Treg
KLF6	NS	NS	Positive
NR4A3	Positive	NS	NS
ZNF574	NS	NS	Negative
ZZZ3	NS	NS	Negative
SP3	NS	NS	Positive
STAT6	Negative	NS	NS
ABCF1	NS	NS	Negative
GLIS2	NS	NS	Negative
ZNF114	NS	NS	Negative
FOXN2	NS	NS	Positive
IKZF2	NS	NS	Positive
ERF	NS	NS	Negative
SMARCB1	Positive	NS	NS
IL4R	Negative	NS	NS
ZBTB32	Positive	NS	NS
E4F1	NS	NS	Negative
USP22	NS	NS	Positive
EGFP	NS	NS	Negative
TAF5L	Positive	NS	Positive
TBP	NS	NS	Negative
ARID5A	Negative	NS	NS
TGIF2	NS	NS	Negative
ZKSCAN2	NS	NS	Negative
SOCS3	Negative	NS	NS
ZNF384	NS	NS	Positive
JUNB	NS	NS	Negative
SS18	NS	NS	Positive
ETV3	NS	NS	Negative
MTF2	Negative	NS	NS
NFIL3	NS	Negative	NS
IFNGR2	Negative	NS	NS
MED30	NS	NS	Positive
NFKB2	Positive	NS	NS
NR2F6	NS	NS	Negative
ATXN7L3	NS	Negative	Positive
MED14	NS	NS	Positive
RAD21	NS	NS	Negative

id	Stimulated Teff	Resting Teff	Resting Treg
RUNX1	NS	NS	Negative
NR2C2	NS	NS	Negative
MYC	NS	NS	Negative
LEF1	Negative	NS	NS
DNMT1	NS	NS	Positive
IFNGR1	NS	Negative	NS
RUNX3	NS	Negative	NS
MED11	NS	Negative	Positive
IRF1	Negative	Negative	Negative
MTF1	Negative	Negative	NS
PRDM1	Positive	Negative	NS
HNRNPK	NS	Negative	Negative
BCL11B	NS	Negative	NS
SATB1	NS	Negative	Negative
MEF2D	NS	Negative	Negative
TNFAIP3	NS	Negative	NS
IRF2	Negative	Negative	Negative
TFDP1	NS	Negative	Negative
KLF2	NS	Negative	Negative
KLF13	NS	Negative	Negative
ZNF217	NS	Negative	Negative
FOXK1	Negative	Negative	Negative
MYB	NS	Negative	Negative
HIVEP2	NS	Negative	NS
MED12	Positive	Negative	Positive
CBFB	Positive	Negative	Negative

Table 2.2 sgRNA sequences used in arrayed KO experiments
Guides from Brunello library used in arrayed experiments

Gene	sgRNA	sequence
TAF5L	s_3244	TAAGGTGAGGACTTTGCACA
ZNF217	s_2471	GGACACATAATGGCAAATCG
BPTF	s_496	GCCTTCGATCAAGTGCACTG
MED12	s_2770	CAGTGAGTAGTGCCAAACCA
SMARCB1	s_1812	TGAGAACGCATCTCAGCCCG
PRDM1	s_91	GGGGAGCGAGTGATGTACGT
BACH2	s_3748	GTTCAGCATAACGAAAAGTG
AAVS1	AAVS1_1	GGGGCCACTAGGGACAGGAT
IL2RA	s_994	GGATACAGGGCTCTACACAG
MYC	s_1296	GCTGCACCGAGTCGTAGTCG
CBFB	s_135	GCCGACTTACGATTTCCGAG
AAVS1	AAVS1_3	TAAGCAAACCTTAGAGGTTC
SOCS3	s_2642	GCGGATCAGAAAGGTGCCGG
NFKB2	s_1388	GGGACCAGCCAAGATCGAGG
NR4A3	s_2529	CCTGCGTGTACCAAATGCAG
TAF5L	s_3243	GCGGACCAGTGTACAGCACG
ZNF217	s_2469	CAAAATCTCACCCTGAAACG
BPTF	s_494	ACTGCTATCCTGAATAGACA
MED12	s_2772	TTCACATTATGACCAACACC
AAVS1	AAVS1_2	GGACGCACCATTCTCACAAA
IL2RA	s_995	TGGCTTTGAATGTGGCGTGT
MYC	s_1295	CTTCGGGGAGACAACGACGG
AAVS1	AAVS1_4	CCTCTAAGGTTTGCTTACGA
MEF2D	s_1220	GGTGAGCGAATGAGTAGACT
MEF2D	s_1217	CAAGTACCGACGCGCCAGCG
KLF2	s_2844	TCGGGGTAATAGAACGCAGG
GATA3	s_602	CAGGGAGTGTGTGAACTGTG
GATA3	s_601	AGGTACCCTCCGACCCACCA
FOXO1	s_557	ACAGGTTGCCCCACGCGTTG
MBD2	s_2640	CGAAAATCTGGGCTAAGTGC

Gene	sgRNA	sequence
MED30	s_4235	GGACATCGTGTACCGCACCA
FOXP3	s_3330	CCCACCCACAGGGATCAACG
USP22	s_3110	CCTCGAACTGCACCATAGGT
ATXN7L3	s_3594	CATCGCTCAGGAGATATACG
DNMT1	s_313	GAGGCAAAAAGAAATCCCCA
MED14	s_2673	ATCACACATAGCGACGAAGT
MED11	s_5200	CACCGCTTCAGTGCAACACG
PTEN	s_1640	CCAATTCAGGACCCACACGA
BATF	s_2879	ATGTGAGAAGAGTTCAGAGG
HIVEP2	s_694	GACAAGATGTCAGACCTAGG
MED20	s_7052	AACAGACATGATGCGGTCTA
MED19	s_7024	CACCTGGCAGTTCCCTCATG
MED24	s_7009	CCTCGAGCAGGATCACGGCG
MED9	s_7017	TTGATGATGTTGTGAACCAA
MED12L	s_7023	TGCCAATGGTTCGAGCAACG
MED26	s_7006	CCTCGGAACTCACGGCATGA
MED28	s_7019	TCCACCAAAGTACTGCTGGA
MED22	s_7002	CGAGATCATCAAGACCGCCA
MED23	s_7004	AAGGCCATGCGTAGTCACGA
MED31	s_7014	TCTAACATGTGTAAACACTG
MED10	s_7021	GGAGAAGTTTGACCACCTAG
MED27	s_7007	CCATCACTCACTGAGGTGGT
MED21	s_7003	ATTGGAGTATTGCAGCAATG
MED16	s_7012	GGTGTGGTAGTCGCACACGC
MED1	s_7001	CCAACCAACACCTTTCCGGG
MED18	s_7016	TGTCACACAAACCACGAAGG
MED25	s_7020	GGGAGCGTGACATTGTACGT
MED17	s_7005	TGAAGAGGTTAACTGTGCCG
MED29	s_7018	CAGTCGTAACGCACTTCCGG
MED8	s_7022	ATGACCTGGTTACGGAACAG
MED15	s_7015	GGAATTGGCATGCCTCCTCG
MED6	s_7011	TCATCATTCGGAAGCAACAG

Gene	sgRNA	sequence
CDK8	s_7050	CGAGGACCTGTTTGAATACG
MED7	s_7008	GACCTTGAGAGTCATGATGG
MED13	s_7010	ACAAGGGTGAGGACTAAGTG
CCNC	s_7000	CTTGAAATATACCGTAGCAG
MED4	s_7013	AAATAATTAAGTATGCACAT
KLF2	s_2841	AAACCAGGGCCACCGAAAGG

Table 2.3 sgRNA sequences used in CRISPRi Perturb-CITE-seq
Guides from Dolcetto library used in Perturb-seq

Gene	sequence
BACH2	GCGCTGTGCGACCGCAGCCC
BACH2	CAGCAGCGGCCGTGCACGCC
BATF	CCTGCGTCCTCCTCACTCTG
BATF	TCCCTCTGCACCCCAGAGTG
BPTF	AAGGCTCAATCCGAATTGCT
BPTF	GATGGCGGCTGAAGGCGATC
CBFB	GCGGCAGGCAACGGCTGAGG
CBFB	GGCAACGGCTGAGGCGGCGG
DNMT1	GGGCAGCGAGATGGCCGGGA
DNMT1	TGCTGAAGCCTCCGAGATGC
ETS1	CGAGGGCCGGGCAGGAGGAG
ETS1	GGCGCGGCGCGGGCGAGGCT
FOXO1	CAGGGCCGCGGACGGAAGGA
FOXO1	GCGCAGCGGGCGCGCCGCTG
IL2RA	GGCATCGCGCCGGAGGATGT
IL2RA	TCTTCCCATCCCACATCCTC
IRF1	GCCCGAGCCCCGCCGAACCG
IRF1	CGGCCGGCGTGGACTGGGCA
IRF2	TTTCGATCTGGACTGTTCTC
IRF2	GACTGTTCTCAGGCAAGCCG
IRF4	GTCCAGGGCGAGGTAAGGGC
IRF4	TCGGAGCTGAGGGCAGCGGT

Gene	sequence
KLF13	CCGGTTCTAAGGATGCCGAG
KLF13	GGGACCACCTCGCCAAGTCG
LEF1	ATCTGCCTTCTCTATCCCAA
LEF1	TCGTACTCACCTCTGCCATT
MED11	GAACAAGCGTCGCGTTTCTG
MED11	GTAGGTAGCCATTATCACTC
MED12	ACGGCGGCCGAGAGACAACA
MED12	GGCGGCCGAGAGACAACAAG
MED24	CCACCTAGAACTGGATTGTG
MED24	CCGCACAATCCAGTTCTAGG
MYB	GGAGCGCCGCTGCGCAGCCG
MYB	GTCCCTCCCCGGCTGCGCAG
MYC	AGGCAGAGGGAGCGAGCGGG
MYC	GTAATTCCAGCGAGAGGCAG
NFKB2	TGGAGAGCGAGATCCGGAGT
NFKB2	CGGATCTCACCCGCCACACC
PRDM1	AGAGGCAAGAGCAGCGACCG
PRDM1	GGCCCTCCAGTGTTGCGGAG
STAT5B	GGCCGAGGGAGGGAGCGAGC
STAT5B	CCAGCGCAGGCAACTCCGCG
USP22	GCGCCGAGAACAAAGCGCGG
USP22	CGGGCGCCGAGAACAAAGCG
ZNF217	TGAGCTGACACCACTCGGGC
ZNF217	TGACACCACTCGGGCCGGCC
SOCS3	CGGGGCCGAAGCGGCAGCAG
SOCS3	TGGGCCGGGCGGGCGGCTGG
ATXN7L3	ACTGCTCGCGCCTGCTAGAA
ATXN7L3	GAGCGCGTGCATCTGCCCCG
TAF5L	GCGCGGGGGCCCAGCTGAGC
TAF5L	CGGCCGCCCAGAGCGGCGGC
MEF2D	GCCGAGCGCCTGAGCCGCCT
MEF2D	CAAAATATCAACAACAGCCG

Gene	sequence
KLF2	CCGGGGAGAAAGGACGCGGA
KLF2	GTGAGTGGCTGCCCGAGGCC
NO-TARGET	CGCGGAAATTTTACCGACGA
NO-TARGET	GGGATGCGTCTTGCTAAACC
NO-TARGET	GAGGACCTTAAGGTGACATG
NO-TARGET	GCTGTTCCGAAGTTGAGAAT
NO-TARGET	ACGCCTCCTCAAATTAGCTC
NO-TARGET	GAAACGAGAAGTTTGTACTA
NO-TARGET	TAATGCTGCACACGCCGAAT
NO-TARGET	GGCTGGTTGACGACTCCTGA
NO-TARGET	TAACCGATACTCCCCACATT

Table 2.4 Antibodies used in experiments
Antibodies used in flow cytometry, mass spectrometry, and western blots

Target	Species/Isotype/Clone	Vendor	Cat. No.	Amout per IP (µg)	Assay
MED12	Rabbit polyclonal IgG	Bethyl/Thermo	A300-774A	8	ChIPseq
CXXC1	Rabbit monoclonal IgG (D1R5R)	Cell Signaling	40672S	2	ChIPseq
NELF-A	Mouse monoclonal IgG2bk (G-11)	Santa Cruz	sc-365004	2	ChIPseq
RNA PolII CTD	Mouse monoclonal IgG1 (4H8)	Cell Signaling	2629	2.5	ChIPseq
RNA PolII phospho-Ser2	Rabbit polyclonal IgG	Abcam	ab5095	2.5	ChIPseq
RNA PolII phosphor-Ser5	Rabbit polyclonal IgG	Abcam	ab5131	2.5	ChIPseq
H3K27ac	Rabbit Monoclonal (2114-3E4)	EpiCypher	13-0045	0.5	CUT&RUN
H3K4me1	Rabbit Monoclonal (2088-1F4)	EpiCypher	13-0057	0.5	CUT&RUN
H3K4me2	Rabbit Monoclonal	EpiCypher	13-0027	0.5	CUT&RUN
H3K4me3	Rabbit Monoclonal (2909-3D7)	EpiCypher	13-0041	0.5	CUT&RUN
IgG	Rabbit	EpiCypher	13-0042	0.5	CUT&RUN
MED12	Rabbit monoclonal IgG (D9K5J)	Cell Signaling	14360	10	IP

Target	Species/Isotype/Clone	Vendor	Cat. No.	Amout per IP (µg)	Assay
IgG	Normal Rabbit IgG	Cell Signaling	2729S	10	IP
MED12	Rabbit monoclonal IgG (D9K5J)	Cell Signaling	14360	NA	Western blot
IgG-HRP conjugate	Mouse Anti-rabbit IgG (Conformation Specific) (L27A9)	Cell Signaling	5127S	NA	Western blot
CXXC1	Rabbit monoclonal IgG (D1R5R)	Cell Signaling	40672S	NA	Western blot
MED17	Rabbit mAb (E3V6Y)	Cell Signaling	64733S	NA	Western blot
SET1A	Rabbit mAb (D3V9S)	Cell Signaling	61702	NA	Western blot
GAPDH	Rabbit mAb D16H11	Cell Signaling	5174	NA	Western blot
Alexa Fluor® 647 anti-human IL2RA	Mouse IgG1, κ (BC96)	Biolegend	302618	NA	Flow cytometry
Ghost Dye™ Red 780	NA	Tonobo	13-0865-T500	NA	Flow cytometry
BV711 anti-human CD4	Mouse IgG1, κ (SK3)	Biolegend	344648	NA	Flow cytometry
PE anti-mouse/human Helios	Armenian Hamster IgG, 22F6	Biolegend	137216	NA	Flow cytometry
KIRAVIA Blue 520™ anti-human CD152 (CTLA-4)	Mouse IgG1, κ (L3D10)	Biolegend	349938	NA	Flow cytometry
Pacific Blue™ anti-human FOXP3	Mouse IgG1, κ (206D)	Biolegend	320116	NA	Flow cytometry
PE anti-Human CD127	Mouse IgG1, κ (HIL-7R-M21)	Beckon Dickinson	557938	NA	Flow cytometry
Pacific Blue™ anti-human CD4	Mouse IgG1, κ (SK3)	Biolegend	344620	NA	Flow cytometry
PE anti-human CD95 (Fas)	Mouse IgG1, κ (DX2)	Biolegend	305608	NA	Flow cytometry

Methods

Primary human T cell isolation and expansion

CD4+ regulatory and effector T cells were isolated from fresh Peripheral Blood Leukopaks (STEMCELL Technologies, #70500) from healthy human donors with institutional review board–approved informed written consent (STEMCELL Technologies). The contents of the Leukopaks were washed twice with a 1X volume of EasySep buffer (DPBS, 2% fetal Bovine Serum (FBS), 1mM pH 8.0 EDTA) using centrifugation. The washed cells were resuspended at $200x10^6$ cells/mL in EasySep buffer and isolated with the EasySep™ Human CD4+CD127lowCD25+ Regulatory T Cell Isolation Kit (STEMCELL Technologies, #18063), according to the manufacturer's protocol. Following isolation with the kit, Tregs were stained Alexa Fluor® 647 anti-human IL2RA (CD25) Antibody (Biolegend, #302618, diluted 1:25), PE anti-Human CD127 (Beckon Dickinson, #557938, diluted 1:50), and Pacific Blue™ anti-human CD4 Antibody (Biolegend, #344620, diluted 1:50) and isolated with FACS performed on a BD FACS ARIA Fusion 1 (#656700) to ensure a pure population without contaminating effector cells. After sorting pure CD4+CD127lowCD25+ Regulatory T Cells, the cells were seeded at $1x10^6$ cells/mL in XVIVO-15 (Lonza, #02-053Q) supplemented with supplemented with 5% FCS, 55 µM 2-mercaptoethanol, 4 mM N-acetyl L-cysteine, and 200 U/mL IL-2 (Amerisource Bergen, #10101641). Teffs were seeded at $1x10^6$ cells/mL in RPMI-1640 supplemented with 10% FCS, 2 mM L-Glutamine (Fisher Scientific #25030081), 10 mM HEPES (Sigma, #H0887-100ML), 1X MEM Non-essential Amino Acids (Fisher, #11140050), 1 mM Sodium Pyruvate (Fisher Scientific #11360070), 100 U/mL Penicillin-Streptomycin (Sigma, #P4333-100ML), and 50 U/mL IL-2 (Amerisource Bergen, #10101641). Both cell subsets were then stimulated with ImmunoCult™ Human CD3/CD28/CD2 T Cell Activator (STEMCELL Technologies, #10990) at 25 µL/mL for Tregs and 6.25 µL/mL for Teff. Cells were cultured at 37C with 5% CO_2. Following activation

and electroporation, cells were split 1:2 every 48 hours to maintain an approximate density of $1x10^6$ cells/mL and supplemented with respective doses of IL-2.

Pooled CRISPR knock-out screen trans-regulator editing

Pooled screens were performed following the protocol described in Freimer et al[19]. In brief, 24 hours after stimulating and plating the T cells, the trans-regulator lentiviral library[19] was added to each culture. The cells were counted prior to transduction, and virus was added at a multiplicity of infection (MOI) of 0.8, using gentle mixing to disperse the viral media without disrupting cell bundling. The cells were then incubated at 37°C for an additional 24 hours, pelleted by centrifugation, and viral media was replaced with fresh media supplemented with IL-2.

24 hours after washing, the cells were pelleted by centrifugation at 150 g for 10 minutes, resuspended at 1.5 $x10^6$ cells per 17.8 µL supplemented P3 Primary Cell Nucleofector Buffer (Lonza, component of #V4SP-3960) and combined with 7.2 µL RNP/1.5 $x10^6$ cells in a sterile 10 mL reservoir. After mixing the cells and RNPs, 25 µL of the mixture was distributed to the wells of a 96-well Nucleocuvette Plate (Lonza, component of #V4SP-3960). Cells were nucleofected using code EO-115 for Tregs and EH-115 for Teffs on the Lonza 4D-Nucleofector System with the 96-well Shuttle. Immediately after nucleofection, 90 µL pre-warmed cell-appropriate media was added to each well, and the cells were incubated at 37°C for 15 minutes. Following incubation, cells were seeded at 1 $x10^6$ cells/mL in media supplemented with IL-2.

IL2RA screen sorting and library preparation

Transduced and electroporated cells were expanded for a minimum of 6 days following editing prior to sorting. Cell sorting was performed 10 days following isolation for the resting screens. For the stimulated Teff screen, cells were restimulated with ImmunoCult™ Human

CD3/CD28/CD2 T Cell Activator (STEMCELL Technologies, #10990) 9 days following initial isolation and sorting was performed 72 hours after restimulation, at the time of peak IL2RA expression. Prior to sorting, cells were counted, washed once with EasySep buffer, and stained with Alexa Fluor® 647 anti-human IL2RA (CD25) Antibody (Biolegend, #302618, diluted 1:25). Cells were then washed and resuspended in EasySep buffer. During sorting, cells were gated on the GFP+ population (lentiviral sgRNA library marker) and the top and bottom 20% of IL2RA expressing cells were sorted into 15 mL conical tubes coated with FCS. Isolated cells were pelleted, counted and lysed. gDNA extraction was performed using phenol-chloroform extractions and sgRNA libraries were amplified and prepared for sequencing using custom primers. Libraries were sequenced on an Illumina HiSeq 4000 at the UCSF CAT.

Screen analysis

All pooled screens were analyzed with MAGeCK[43] (v0.5.9.5). MAGeCK count was performed on all donors using --norm-method none followed by MAGeCK test --sort-criteria pos to identify genes that resulted in a statistically significant change in IL2RA expression. Results are calculated as the IL2RA low bin/IL2RA high bin. Screen visualization is represented as the IL2RA high bin/IL2RA low bin by flipping the sign for the fold change. All genes with an FDR-adjusted $P < 0.05$ were considered significant.

Arrayed CRISPR KO of select regulators

Guide-loaded Cas9 RNPs were assembled with custom crRNAs (Dharmacon) which were resuspended in IDT duplex buffer (IDT, #11-01-03-01) at 160 µM. Sequences are provided in **Table 2.2**. Dharmacon Edit-R CRISPR-Cas9 Synthetic tracrRNA (Dharmacon, #U-002005-20) also resuspended in Nuclease Free Duplex Buffer at 160 µM was combined at a 1:1 molar ratio in a 96 well plate and incubated at 37°C for 30 minutes. Single-stranded donor oligonucleotides

(ssODN; sequence:

TTAGCTCTGTTTACGTCCCAGCGGGCATGAGAGTAACAAGAGGGTGTGGTAATATTACGGT

ACCGAGCACTATCGATACAATATGTGTCATACGGACACG, 100 μM stock) was added to the complex at a 1:1 molar ratio and incubated at 37°C for 5 minutes. Finally, Cas9 protein (MacroLab, Berkeley, 40 μM stock) was added at a 1:2 molar ratio and incubated at 37°C for 15 minutes. The resulting RNPs were frozen at -80°C until the day of electroporation and were thawed to room temperature prior to use. 48 hours following T cell activation, the cells were pelleted at 100xg for 10 minutes and resuspended in room temperature P3 Primary Cell Nucleofector Buffer (Lonza, catalog no. V4XP-3032) at $1.5x10^6$ cells per 17.8 μL. $1.5x10^6$ cells were transferred to each RNP containing well and mixed gently. 25 μL of the combined RNP cell solution was transferred to a 96-well electroporation cuvette plate (Lonza, #VVPA-1002) and nucleofected with pulse code DS-137. Immediately following electroporation, the cells were gently resuspended in 90 μL warmed media and incubated at 37°C for 15 minutes. After recovery, the cells were cultured in 96 well round-bottom plates at $1x10^6$ cells/mL for the duration of the experiment. To prevent edge effects, the sgRNAs were randomly distributed across each plate and the first and last columns and rows of each plate was filled with PBS to prevent evaporation. Unless otherwise specified, CRISPR-Cas9 edited cells were restimulated on day 8 following isolation for stimulation response arrayed assays with ImmunoCult™ Human CD3/CD28/CD2 T Cell Activator (STEMCELL Technologies, #10990).

Genotyping of arrayed KOs

On the final day of the respective assay, genomic DNA was isolated using DNA QuickExtract (Lucigen, Cat #QE09050) according to the manufacturer's protocol. Primers were designed to flank each sgRNA target site. Amplicons of the region were generated by adding 1.25 μL each of forward and reverse primer at 10 μM to 5 μL of sample in QuickExtract, 12.5 μL of NEBNext Ultra II Q5 master mix (NEB, Cat #M0544L), and H2O to a total 25 μL reaction volume.

Touchdown PCR was used with the following cycling conditions: 98°C for 3 minutes, 15 cycles of 94°C for 20 seconds followed by 65°C-57.5°C for 20 seconds (0.5°C incremental decreases per cycle), and 72°C for 1 minute, and a subsequent 20 cycles at 94°C for 20 seconds, 58°C for 20 seconds and 72°C for 1 minute, and a final 10 minute extension at 72°C. Amplicons were diluted 1:200 and Illumina sequencing adapters were then added in a second PCR reaction. Indexing reactions included 1 µL of the diluted PCR1 sample, 2.5 µL of each the forward and reverse Illumina TruSeq indexing primers at 10 µM each, 12.5 µL of NEB Q5 master mix, and H2O to a total 25 µL reaction volume. The following PCR cycling conditions were used: 98°C for 30 seconds, followed by 98°C for 10 seconds, 60°C for 30 seconds, and 72°C for 30 seconds for 12 cycles, and a final extension period at 72°C for 2 minutes. Samples were pooled at an equivolume ratio and SPRI purified prior to sequencing on an Illumina MiSeq with PE 150 reads. Analysis with performed with CRISPResso2[44] (v2.2.7) CRISPRessoBatch --skip_failed --n_processes 4 --exclude_bp_from_left 5 --exclude_bp_from_right 5 --plot_window_size 10.

Flow Cytometry analysis of arrayed KOs

The Biolegend FoxP3 Fix/Perm kit (Biolegend, #421403) was used for staining according to the manufacturer protocol. Cells were washed in EasySep buffer prior to extracellular staining. Cells were stained with Alexa Fluor® 647 anti-human IL2RA (CD25) Antibody diluted 1:25 (Biolegend, #302618), Ghost Dye™ Red 780 diluted 1:1000 (Tonbo, #13-0865-T500) and BV711 anti-human CD4 diluted 1:50 (Biolegend, #344648) for 20 minutes at 4C and then washed once with EasySep buffer. After fixing and permeabilizing according to the kit, intracellular staining was performed with PE anti-mouse/human Helios Antibody (Biolegend #137216), KIRAVIA Blue 520™ anti-human CD152 (CTLA-4) Antibody (Biolegend #349938), and Pacific Blue™ anti-human FOXP3 Antibody (Biolegend, #320116) each diluted 1:50 in permeabilization buffer for 30 minutes at room temperature. Cells were subsequently washed in permeabilization buffer and resuspended in EasySep buffer before running on the ThermoFisher Attune NxT flow

cytometer (#A29004). Analysis of flow data was performed in FlowJo (v10.8.1). Gating was performed to select for lymphocytes, singlets, live cells (Ghost Dye negative), and CD4+ cells in the specified order. This population was then used to calculate the median fluorescence intensity (MFI) for IL2RA or CTLA-4. Visualization was performed in R using ggplot2 (v3.4.1).

Cloning and lentivirus preparation

CRISPRi sgRNAs for Perturb-seq were selected from the Dolcetto library[45] and cloned into the LGR2.1 plasmid backbone (Addgene #108098). A lenti EF1a-Zim-3-dCas9-P2A-BSD was generated using Gibson assembly as described in Pacalin et al[46]. Lentivirus was prepared according to the protocol in Schmidt et al[25].

Perturb-Seq

24 hours after stimulation of isolated human Tregs and Teffs from 2 donors, the cells were transduced with Zim3-dCas9 lentivirus at 3% v/v. The following day, Perturb-seq sgRNA library lentivirus was added at 0.75% v/v (MOI 0.3, **Table 2.2**). 48 hours after transduction with Zim3-dCas9, 10 mg/ml blasticidin (Gibco, #A1113903) was added to each sample to select for dCas9+ cells. Blasticidin was replenished every 48 hours until the cells were processed for sequencing. 8 days after initial isolation and stimulation of cells, half of the Treg and Teff culture was restimulated with ImmunoCult™ Human CD3/CD28/CD2 T Cell Activator (STEMCELL Technologies, #10990). On the 10th day after initial isolation, the samples were collected for 10X single cell sequencing. First, cells from each donor within the same stimulation and cell type condition were pooled at equal concentrations. Sorting was performed to isolate live GFP+ cells from each condition. Sorted cells were processed according to the Chromium Next GEM Single Cell 5' HT Reagent Kits v2 (Dual Index) with Feature Barcode technology for CRISPR Screening and Cell Surface Protein guide User Guide, CG000513. In brief, sorted cells were

pelleted and washed once with Cell Staining Buffer (Biolegend, #420201). Next, the samples were blocked with Human TruStain FcX™ Fc Blocking reagent (Biolegend, #422302). Meanwhile, TotalSeq™-C Human Universal Cocktail V1.0 (Biolegend, #399905) was prepared using Cell Staining Buffer (Biolegend, #420201) and TotalSeq™-C0251 anti-human Hashtag Antibodies 1-4 (Biolegend, #394661) were added to aliquots of the cocktail. After blocking, cells were stained with TotalSeq-C cocktail including one Hashtag per cell and stimulation condition. After staining, the cells were washed three times in Cell Staining Buffer. The samples were then resuspended in PBS with 1% BSA (Gibco) for final counting. The resulting samples were pooled across conditions and approximately 65,000 cells per well were loaded into 8 wells of a Chromium Next GEM Chip N Single Cell Kit (10X Genomics, #1000375) for GEM generation. The samples were prepared for sequencing using the Chromium Next GEM Single Cell 5' HT Kit v2 (#1000374), 5' Feature Barcode Kit, (#1000256), and 5' CRISPR Kit (#1000451) according to the manufacturer's protocol. GEM generation and library preparation was performed by the Gladstone Genomics Core. The resulting libraries were sequenced using a NovaSeqX Series 10B flowcell (Illumina, #20085595) at the UCSF CAT.

Perturb-seq analysis

Fastqs for each 10X well were concatenated across lanes and flow cells. Alignment of perturb-seq data and count aggregation for the gene expression, CRISPR sgRNA, and Antibody Derived Tag (ADT) libraries was performed with cellranger[47] count (v7.1.0) using the default settings and –expect-cells=45000 –chemistry=SC5P-R2. Gene expression fastqs were aligned to "refdata-gex-GRCh38- 2020-A" human transcriptome reference acquired from 10x Genomics. SgRNA sequences were aligned to a custom reference file using the pattern TAGCTCTTAAAC(BC) while ADTs were aligned to the TotalSeq-C-Human-Universal-Cocktail-399905-Antibody-reference-UMI-counting.csv provided by Biolegend, also including the hashtag oligo (HTO) sequences which were used to distinguish each cell type and stimulation condition.

Counts for each respective library were aggregated across wells with cellranger aggr using the default settings. Cells were assigned to a donor using genetic demultiplexing with Souporcell[48] (https://github.com/wheaton5/souporcell). For each well, souporcell_pipeline.py was run using the bam file and cellranger count output barcodes.tsv as input in addition to the reference fasta. Donor calls shared across wells were identified using shared_samples.py using the vcf file outputs from Souporcell.

Perturb-seq analysis was performed in R (v4.3.1) using Seurat[49] (v4.3.0.1) based on code published in Steinhart 2022[50]. Count matrices were imported into R using the Seurat Read10X function. After creating a Seurat object with CreateSeuratObject, quality filtering was performed to retain cells with more than 1000 RNA features identified and less than 7.5% mitochondrial RNA. Cells without a singular donor assignment were also excluded from the object as well as cells with more than one HTO assignment as determined after running HTODemux. Low abundance transcripts were filtered using the threshold of 10 cells per feature and TCR genes were removed from the primary RNA assay as they were found to be a major source of variance in the dataset. No sgRNA targets were removed as the number of cells in each condition exceeded the threshold set of 150 cells. After filtering, gene-expression counts were normalized and transformed using the Seurat SCTransform function with regression of both S-phase score and G2/M-phase score, as described on the Satija website (https://satijalab.org/seurat/articles/cell_cycle_vignette.html). ADT counts were normalized using the CLR normalization method of NormalizeData. After generating PCAs of both normalized and transformed RNA and ADT data, Harmony[51] (v0.1.1) was used to correct for donor associated variability in the dataset. The resulting normalized and transformed counts were used for downstream analysis unless otherwise specified. UMAPs were generated using the transformed and corrected RNA and ADT counts with Seurat function FindMultiModalNeighbors followed by RunUMAP using weighted.nn. Prior to cell type specific analysis, Treg cells were manually filtered to include only

cells belonging to clusters with FOXP3 and IKZF2 expression to maximize cell purity (clusters 1,7,8,15,6,4,19,20,17,23).

Activation scoring was performed according to Schmidt et at.[25,50] In brief, Seurat FindMarkers was used to identify differentially expressed genes between stimulated and resting non-targeting control cells within the Teffs and Tregs individually. Genes that had a $\log_2$ fold change >0.25 and were detected in 10% of restimulated or resting cells were used to generate gene weights for the score calculated as sum(GE × GW/GM), where GE is a gene's normalized/transformed expression count, GW is the gene's weight, and GM is the gene's mean expression in non-target control cells of the respective cell type. Wilcox tests were performed to determine significance compared to non-targeting control cells with Bonferroni correction for multiple hypothesis testing. To observe the effect of each sgRNA within independent cell and stimulation conditions, the cells were subset by HTO. RNA and ADT normalization, transformation, and donor variability correction was repeated for each subset as described above for the combined dataset. UMAPs were generated using the transformed and corrected RNA and ADT counts with Seurat function FindMultiModalNeighbors followed by RunUMAP using weighted.nn. Cell cycle quantification for each subset was performed using cycle assignments generated using the Satija cell cycle vignette referenced above.

Pseudobulking of resting and stimulated Treg and Teff samples was performed using Seurat AggregateExpression grouped by HTO, target gene, and donor pulling from the counts slot (sgRNAs targeting the same gene were collapsed within the same donor). Differential expression analysis was performed with the resulting pseudobulked raw counts for both RNA and ADTs. DESeq2[52] (v1.32.0) was used to identify differentially expressed genes and proteins between each sgRNA and non-targeting control sample within each cell type and stimulation condition, using donor information as a covariate. Network plots of differentially expressed gene

connections were visualized in R using influential[53] (v2.2.7) and ggraph[54] (v2.1.0), including only genes with an adjusted p-value < 0.05. Other visualization of differentially expressed genes and surface proteins was performed using ggplot2 (v3.4.1).

Bulk RNAseq

At their respective timepoints, resting and 48-hour restimulated cells were pelleted and resuspended at $1x10^6$ cells per 300 µL of RNA lysis buffer (Zymo, #R1060-1-100). Cells were pipette mixed and vortexed to lyse and frozen at -80 until RNA isolation was performed. RNA was isolated using the Zymo-Quick RNA micro prep kit (#R1051) according to the manufacturer's protocol with the following modifications: After thawing the samples, each sample was vortexed vigorously to ensure total lysis prior to loading into the extraction columns. The optional kit provided DNAse step was skipped, and instead RNA was eluted from the isolation column after the recommended washes and digested with Turbo-DNAse (Fisher Scientific, AM2238) at 37 C for 20 minutes. Following digestion, RNA was purified using the RNA Clean & Concentrator-5 kit (Zymo, #R1016) according to the manufacturer's protocol. The purified RNA was submitted to the UC Davis DNA Technologies and Expression Analysis Core to generate 3′ Tag-seq libraries with unique molecular indices (UMIs). Barcoded sequencing libraries were prepared using the QuantSeq FWD kit (Lexogen) for multiplexed sequencing on a NextSeq 500 (Illumina).

Bulk RNAseq analysis

RNAseq data was processed using the pipeline described in Freimer et al[19]. In brief, fastq adapter trimming was performed with cutadapt (v2.10). Low-quality bases were trimmed with seqtk (v0.5.0). Reads were then aligned with STAR[55] (v 2.7.10a) and mapped to GRCh38. UMI counting and deduplication was performed with umi_tools[56] (v1.0.1) and gene counts were

generated from the deduplicated reads using featureCounts (subread v2.0.1) using Gencode v41 basic transcriptome annotation. Quality control metrics were generated for each sample with Fastqc[57] (v0.11.9), rseqc[58] (v3.0.1), and Multiqc[59] (v1.9). Differentially expressed genes between Mediator KOs and AAVS1 KO samples as well as stimulated and resting AAVS1 KO samples were identified from the deduplicated count matrix using DESeq2[52] (v1.32.0) in R (v4.1.0). Comparisons were made within each cell type and stimulation condition across 3 donors, using donor ID as a covariate in the model. Normalized counts were generated using a DESeqDataSet containing all samples, followed by estimateSizeFactors and counts(normalized=TRUE). AAVS1 KO normalized sample counts were then subset and averaged across donors for visualization.

Differentially expressed genes for MED12 vs AAVS1 KO samples were defined by a cut off of padj < 0.05. Comparison of the effects of MED12 KO differentially expressed genes across stimulation-response categories was performed by grouping MED12 vs AAVS1 KO differentially expressed genes according to their stimulation-responsive behavior in control cells (stimulation response = padj < 0.05 and abs(log_2 fold change) > 1). The Bonferroni adjusted p value resulting from a two-tailed T test is displayed (Figure 4a) comparing each stimulation response group to the non-stimulation responsive group. Boxplot center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range (genes per group (Downregulated, Not Stimulation Responsive, Upregulated) = Resting Teff:272, 954, 218; Stimulated Teff:242, 1432, 467; Resting Treg:269, 1491, 241; Stimulated Treg:245, 1945, 426). One-sided Fisher's exact test for regulators of IL2RA within the differentially-expressed genes downstream of MED12 was determined using screen results from the matched cell type- and stimulation-conditions (Figure 4b). Genes were subset to those targeted in the screen library and detected in CD4+ T cell bulk RNAseq (genes per group:Regulators, Non-Regulators = Resting Teff:62, 807; Stimulated Teff:41, 824; Resting Treg:82, 787). Pathway analysis was performed using PathfindR[60] (v1.6.4)

including KEGG, Reactome, and GO-BP gene sets. Visualization was performed after removing KEGG disease pathways. Apoptosis pathway visualization was performed using Cytoscape[61] (v3.8.2). GSEA analysis was performed with clusterProfiler[62] (v4.10.1) using msigdbr (v7.5.1) all human gene sets.

SEL120-34A treatment

SEL120-34A (Selleckchem, #S8840) was reconstituted in ultrapure H2O according to the manufacturer's recommendations. Cells were treated every 48 hours with a 1 μM dose and treatment was started 48 hours following cell isolation to align with the time at which cells are edited in CRISPR based experiments. Restimulation of cells for flow cytometry and CUT&RUN was performed 10 days after initial isolation.

Endogenous immunoprecipitation of MED12

IP base buffer (0.05M Tris-HCl pH 7.5, 0.15M NaCl, 0.001M EDTA, AP MS Water) was prepared the day of the experiment. 20 x10^6 resting and 48-hour restimulated cells per sample and immunoprecipitation were washed twice with PBS. Samples were then lysed in 500 μl Lysis buffer per 10 x10^6 cells (Base buffer, 1X PhosphoStop (Roche, #04906837001), 1X Complete mini-EDTA protease inhibitor cocktail tablets (Sigma Aldrich, #11836170001), 0.50% NP-40 Surfact-Amps™ Detergent Solution (Thermo Scientific #85124) and incubated on nutator for 30 minutes at 4ºC. To digest chromatin, tip sonication was performed in round with incubation on ice between each step- 7 sec 12%, 7 sec 12%, 7 sec 12%, 7 sec 15% with 4 rounds of sonication total. Cell lysate was clarified by centrifugation at 3500xg for 10 minutes at 4ºC. A BCA was performed for each sample and protein concentrations were normalized across conditions. 10% of whole cell lysate was reserved for input and samples were split into MED12 (Cell Signaling Technologies, #14360) IP and Rabbit IgG isotype control (Cell Signaling

Technologies, #3900) IP conditions. In each case, 10 µg antibody was added to a 1.5 mL protein lo bind tube containing clarified protein and samples were incubated overnight at 4°C, with rotation on a nutator. In the morning, Pierce Protein A+G magnetic beads (Thermo Fisher, #88802) were washed four times using 1 mL of lysis buffer per 1 mL of bead slurry, allowing the beads to bind to a magnet between each wash before removing the buffer. After the final wash, beads were resuspended in Lysis buffer at the original bead slurry volume and 50 µl was added to each sample. The lysate-antibody-bead mixture was then incubated at 4°C for 2 hours with rotation on a nutator. After incubation, beads were bound to a magnetic tube rack and washed one time with IP buffer + NP40 (IP buffer + 0.05% NP40) followed by three washes with a 900 µL IP buffer. The resulting purified proteins were processed for mass spectrometry or western blot.

Mass Spectrometry

After immunoprecipitation, bound proteins were lysed in 8 M urea + 25 mM ammonium bicarbonate followed by reduction (5 mM DTT for 1 h at 37 °C), alkylation (10 mM iodoacetamide for 45 min at room temperature in the dark) and digestion overnight with 1 µg of trypsin (Promega). Peptide samples were applied to activated columns and the columns were washed three times with 200 µl of 0.1% TFA. Peptides were eluted with 140 µl of 50% ACN and 0.1% trifluoroacetic acid (TFA) and dried down by speedvac.

Samples were resuspended in 0.1% FA and separated by reversed-phase (RP) chromatography using an EASY-nLC instrument (Thermo Fisher Scientific) with a 15-cm PepSep column (150 µm inner diameter) (Bruker). Samples were acquired by data-dependent acquisition (DDA). Mobile phase A consisted of 0.1% FA in water and mobile phase B consisted of 80% ACN and 0.1% FA. Peptides were separated at a flow rate of 500 nl min−1 over the following 60 min gradient: 4–35% B in 44 min, 35–45% B in 5 min and 10 min at 88% B.

Peptides were analyzed by an Orbitrap Lumos MS instrument (Thermo Fisher Scientific). Data were collected in positive ion mode with MS1 resolution of 240,000, 350–1,350 m/z scan range, maximum injection time - 50ms, RF lens – 30%. For DDA, MS2 fragmentation was performed on charge states 2–5 with a 20 sec dynamic exclusion after a single selection and 10 ppm ± mass tolerance. All raw MS data were searched using MaxQuant (v2.4.7) against the human proteome (UniProt canonical protein sequences downloaded in September 2022) using default settings and with a match-between-runs enabled[63].

Mass Spectrometry analysis

Protein spectral counts as determined by MaxQuant search results were used for PPI confidence scoring by SAINTexpress[64] (v3.6.1). Rabbit IgG pulldown samples were used as control. The total list of candidate PPIs was filtered to those that met the criteria of SAINTexpress BFDR <= 0.05. To quantify changes in interactions between resting and stimulated T cell states, we used a label-free quantification approach in which statistical analysis was performed using MSstats (v4.8.7)[65] from the artMS (v1.18.0) R package. Visualization was performed in Cytoscape with additional connections included from the STRING database[66].

Western blots

After affinity purification of proteins, beads were resuspended in 100 µl 2X sample buffer (BioRad 4x Laemmli Sample Buffer #1610747) with 1:10 B-mercaptoethanol (Sigma #63689-25ML-F) diluted 1:1 with 500 µL Lysis buffer). Samples were boiled for 5 minutes at 95°C and stored at -20C until further processing. Western blots were performed as previously published[67]. Briefly, cell lysates were subjected to SDS–PAGE on 4-15% acrylamide gels and electroblotted to polyvinylidene difluoride membranes. Blocking and primary (diluted 1:1000) and secondary antibody incubations of immunoblots were performed in Tris-buffered saline + 0.1% Tween-20

supplemented with 5% (w/v) bovine serum albumin (Antibodies provided in **Table 2.4**). Horseradish peroxidase–conjugated goat anti-rabbit and immunoglobulin G (Southern Biotech) was used at a dilution of 1:30,000, and immunoreactive bands were detected using Pierce™ ECL Western Blotting Substrate (32106)
according to the manufacturer's instructions.

CUT&RUN

CUT&RUN was performed on resting and 48-hour restimulated cells according to the manufacturer's protocol with the EpiCypher CUTANA™ ChIC/CUT&RUN Kit and provided reagents. Samples for H3K27ac CUT&RUN were lightly crosslinked prior to isolation using 0.1% formaldehyde (Sigma, #252549) for 1 minute and quenched with 125 mM Glycine (Sigma, #50046). In brief, $5x10^5$ T cells per reaction were washed with PBS before nuclear isolation using the EpiCypher recommended lysis buffer consisting of 20 mM HEPES pH 7.9 (Sigma-Aldrich), 10 mM KCl (Sigma-Aldrich), 0.1% Triton X-100 (Sigma-Aldrich), 20% Glycerol (Sigma-Aldrich), 1 mM MnCl2 (Sigma-Aldrich), 1X cOmplete Mini-Tablet (Roche, # 11873580001), and 0.5 mM Spermidine (Sigma-Aldrich). The cells were resuspended in 100 µL per reaction cold nuclear extraction buffer and incubated on ice for 10 minutes. Following lysis, nuclei were pelleted and resuspended in 100 µL per reaction of nuclear extraction buffer. The isolated nuclei were then frozen at -80°C in extraction buffer until DNA isolation. After thawing the samples at 37°C, the nuclei were bound to activated conA beads. After adsorption of nuclei to beads, permeabilization was performed with 0.01% digitonin containing buffer. Antibodies for H3K27ac (EpiCypher, #13-0045), H3K4me1 (EpiCypher, #13-0057), H3K4me2 (EpiCypher, #13-0027), H3K4me3 (EpiCypher, #13-0041) and IgG (EpiCypher, #13-0042) were added at 500 ng per reaction. Following overnight antibody binding, pAG-MNase addition, and chromatin cleavage, 0.5 ng of the provided E. coli DNA was added to each sample following chromatin cleavage by MNase. Prior to DNA isolation, crosslinked samples were digested overnight with Proteinase K

(Invitrogen, #AM2546) as recommended. The provided spin columns and buffers were used for DNA isolation and purification. The resulting DNA was prepared for sequencing using the CUTANA™ CUT&RUN Library Prep Kit (Cat 14-1002) according to the manufacturer's protocol.

CUT&RUN analysis

Pooled libraries were sequenced on a NextSeq 500 (H3K27ac) and NextSeq 2000 with 2x75 or 2x50 paired end reads, respectively. Bcl2fastq (v2.19) with the settings --minimum-trimmed-read-length 8 was used to generate fastqs. CUT&RUN data analysis was performed according Zheng et al. with the recommended settings unless otherwise specified below[68]. In brief, the fastqs were trimmed with cutadapt (v1.18). Bowtie2[69] (v2.2.5) was used to align the trimmed fastqs to GRCh38 using settings --local --very-sensitive --no-mixed --no-discordant --phred33 --dovetail -I 10 -X 700 -p 8 -q and E. coli (EMBL accession U00096.2) with settings --local --very-sensitive --no-overlap --no-dovetail --no-mixed --no-discordant --phred33 -I 10 -X 700 -p 8 -q. Bam files were generated with samtools[70,71] (version 1.9) view -bS -F 0x04 and bam to bed conversion performed with bedtools (v2.30.0) bamtobed -bedpe. Bedfiles were filtered to include only paired reads of less than 1000 bp with the command awk '$1==$4 && $6-$2 < 1000 {print $0}' samplename.bed before generating bedgraph files using bedtools (version 2.30.0) genomecov -bg. Peak calling was performed using the bedgraph files as input with SEACR[72] (v1.3). Each target bedgraph file was compared to the respective donor and KO condition IgG file to identify peaks above the background using the norm and stringent options for H3K27ac samples. Spike-in scaling was performed prior to methylation peak calling with SEACR using the IgG file as background and non and stringent options.

Prior to generating a peak by sample matrix for each target, ChIP-seq blacklist regions were removed from the data. The sample matrix was reduced across all peaks within the dataset and

H3K27ac peaks were segmented into regions of 5000 bps maximum length. Regions of differential acetylation or methylation between the regulator KOs and AAVS1 KO samples were identified for the peaks called across any of the samples from bam files using DESeq2[52] (v1.32.0) in R (v4.1.0). Comparisons were made within each cell type and stimulation condition using AAVS1s prepared in the same batch of samples. Gene annotation was performed using the gene with the nearest transcription start site to each region with the GenomicRanges[73] (v1.44.0) nearest function. Final bedgraph scaling was performed based on peak coverage across all samples and conditions using DESeq2 (v1.32.0) sizefactors. SEL120-34A and H2O treatment samples were compared as described for MED12 KO and AAVS1 KO samples, using the peak matrix from MED12 KO and AAVS1 KO samples to maximize detection of overlapping regions across datasets.

ChIP-seq

A portion of edited Teffs were re-stimulated with ImmunoCult™ Human CD3/CD28/CD2 T Cell Activator (STEMCELL Technologies, #10990) 10 days following isolation and collected 48 hours later. Up to 1-2 x 10^6 Teffs were cross-linked in PBS with 1% methanol-free formaldehyde (Thermo 28908) for 10 min at 18-22ºC followed by quenching in glycine at 125 mM final concentration. Cross-linked cell pellets were snap-frozen in liquid nitrogen and stored at –80ºC. Nuclei were isolated from thawed, cross-linked cells via sequential lysis in LB1 (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% IGEPAL CA-360, and 0.25% Triton X-100), LB2 (10 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM EDTA, and 0.5 mM EGTA), and LB3 (10 mM Tris-HCl pH 8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% sodium deoxycholate [NaDOC], and 0.5% N-laurylsarcosine) supplemented with 0.5 mM phenylmethylsulfonyl fluoride (PMSF, Sigma P7626) and 0.5X protease inhibitor cocktail (PIC, Sigma P8340). Chromatin was sheared on a Covaris E220 focused ultrasonicator using 1 mL milliTubes (Covaris 520128) with 140W peak incident power, 5% duty factor, 200 cycles per

burst, 6ºC temperature setpoint (minimum 3ºC, maximum 9ºC), fill level 10, and time 12-14 min to obtain target size 200-700 bp. Formaldehyde cross-linked, sheared mouse CD8+ T cell chromatin was spiked in at 2.5% of human Teff chromatin based on fluorometric (Qubit, Thermo Q33238) or OD260 (Nanodrop, Thermo 912A1099) quantification. Triton X-100 was added to a final concentration of 1% prior to immunoprecipitation for 16 h at 4ºC with 2-8 µg of indicated antibodies bound to a 1:1 mixture of protein A and protein G magnetic beads (Thermo 10001D and 10003D). Bead-bound antibody-chromatin complexes were sequentially washed three times with Wash Buffer 1 (20 mM Tris pH 8, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% sodium dodecyl sulfate [SDS], and 0.1% NaDOC), twice with Wash Buffer 2 (20 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, and 0.1% NaDOC), twice with Wash Buffer 3 (20 mM Tris-HCl pH 8, 250 mM LiCl, 1 mM EDTA, 0.5% IGEPAL CA-360, and 0.5% NaDOC), twice with TET (10 mM Tris-HCl pH 8, 1 mM EDTA, 0.2% Tween-20), and once with TE0.1 (10 mM Tris-HCl pH 8, 0.1 mM EDTA, 0.5 mM PMSF, and 0.5X PIC) supplemented with 0.5 mM PMSF and 0.5X PIC. Beads were resuspended in TT (10 mM Tris-HCl pH 8, 0.05% Tween-20) prior to on-bead library preparation using the NEBNext Ultra II DNA Library Prep Kit (NEB E7370L) as described previously[74]. ChIP-seq libraries were multiplexed for paired-end (2 x 50 bp) sequencing on an Illumina NextSeq 2000 instrument.

ChIP-seq analysis

Reads were trimmed to remove adapters and low-quality sequences and aligned to the hg38 and mm10 reference genome assemblies with bwa[75] (v0.7.17-r1188) before filtering to remove duplicates and low-quality alignments including problematic genomic regions[76] using the nf-core/ChIP-seq pipeline[77] (v2.0.0, doi: 10.5281/zenodo.3240506) with default parameters. Normalization to mouse spike-in chromatin was performed by scaling counts to the quotient of

the ratios of human:mouse ChIP reads and human:mouse input reads as described[78]. CXXC1 peaks for visualization were identified using bam files from all AAVS1 KO donors for MACS2 v2.2.6 [79] callpeak -q 0.05 with input samples used to define the background. High-confidence MED12 peaks were identified using bam files from all AAVS1 KO donors for MACS2 callpeak -q 0.05 with MED12 KO samples used to define the background. Utilization of high-confidence peaks generated from KO controls reduced potential false positive signals from the ChIP samples, providing a more rigorous assessment of MED12 binding[80,81]. ChIP-seq blacklist regions were removed from CXXC1 and MED12 peaks prior to analysis.

Polymerase pausing analysis

The polymerase pausing index was calculated as described in Wang et al. as (TSS coverage/TSS length)/(Gene body coverage/gene body length).[33] Gencode V43 gene structures were selected for APRIS genes and filtered to include only genes expressed in Teff bulk RNAseq data (defined from AAVS1 Teff RNAseq baseMean > 10). The TSS region of each gene was defined as 200 bp up- and downstream of the TSS. The gene body was defined as the region 400 bp downstream from the TSS plus 400 bp past the final exon of the gene. Rtracklayer[82] (v1.62.0) was used to import spike-in scaled RNA Pol II CTD bigwigs and GenomicAlignments (v1.38.2) summarizeOverlaps() was used to determine the coverage within the defined gene regions.

CUT&RUN and ChIP-seq visualization

Visualization of scaled tracks was performed with rtracklayer (v1.62.0) and ggplot2 (v3.5.1). APRIS gene structure was used for gene annotation with gggenes (v0.5.0). CD4+ Treg STAT5A ChIP-seq data was accessed from ChIP Atlas[83]- SRX212432, GSM1056923 and generated by Hoffmann et al[31]. Deeptools (v3.5.5) [84] was used to generate profile plots of ChIP-seq data using

computeMatrix scale-regions -b 3000 --regionBodyLength 5000 -a 3000 –skipZeros with scaled bigwigs and a bed file of all expressed genes (defined from AAVS1 Teff RNAseq baseMean > 10) as input with followed by plotProfile –perGroup.

MED12 CAR activation scoring

MED12 CAR RNAseq data from Freitas et al. was accessed from GEO, using the downloader to retrieve raw counts file GSE174279_raw_counts_GRCh38.p13_NCBI.tsv.gz. First, DESeq2 (v1.32.0) was used to identify differentially expressed genes between AAVS1 KO stimulated and resting samples. The top upregulated genes were defined using the following criteria: padj < 0.01, $\log_2$ fold change > 2, baseMean > 10. The resulting 797 genes were used to generate a gene signature of activation. Normalized counts for the MED12 KO and AAVS1 KO resting and stimulated samples were generated with DESeq2 vst and converted to a summarized experiment with SummarizedExperiment[85] (v1.22.0). The normalized count matrix and activation score were used as input for GSVA[86] (v1.40.1) using the gsva function with min.sz=10, max.sz=6000, kcdf="Poisson". Visualization of the resulting gene scores was performed with ggplot2(v3.4.1) and adjusted P values generated using rstatix (v0.7.2).

Activation-induced cell death (AICD) assays

Activation-induced cell death assays were performed using titrated amounts of ImmunoCult™ Human CD3/CD28/CD2 T Cell Activator (STEMCELL Technologies, #10990) in addition to 50 U/mL of IL-2. Active caspase-3/7 staining was performed 72 hours following addition of stimulus using the CellEvent™ Caspase-3/7 Green Flow Cytometry Assay Kit (Invitrogen, #C10427) according to manufacturer's protocol. Gating of apoptotic population was performed on the lymphocyte gate and defined as active caspase-3/7 positive and SYTOX nucleic acid stain

negative. FAS staining was performed using PE anti-human CD95 (Fas) Antibody (Biolegend, #305608, diluted 1:50).

Luminex assays

On day 12 following isolation for Teffs and day 8 following isolation for Tregs, cells were plated in 96 well plates in cytokine free media at a density of 2 x 10^5 cells per well. Cells were restimulated with ImmunoCult™ Human CD3/CD28/CD2 T Cell Activator (STEMCELL Technologies, #10990) and supernatant was collected after 24 hours. The supernatant was stored at -80 degrees until processing by EVE Technologies with the Luminex xMAP technology on the Luminex 200 system. After a serial titration to determine appropriate dilutions, samples were run in technical duplicate and Luminex 48 plex human panel A was run for Teffs (diluted 1:20) and Tregs (diluted 1:5). The multi species TGF 3 plex panel was also run for Tregs (undiluted). Technical replicates were averaged by EVE for each sgRNA and donor combination to determine protein concentration. Cytokines with more than one sample out of range were removed from the analysis to exclude low abundance proteins.

Suppression assays

Donor matched Teffs were isolated and frozen at -80 without activation until 24 hours prior to the assay. Teffs were thawed and cultured overnight at 2x10^6 cells/mL with 10 U/mL IL-2. On the day of the assay, Teffs were counted and stained with CellTrace Violet (Invitrogen, #C34557) according to the manufacturer's protocol using a 1:2000 dilution of dye. Assay plates were assembled with 1x10^5 Teffs per well in 96 well round bottom plates with titrated amounts of Tregs ranging from 1:1 to 8:1 Teffs:Tregs. One well per condition was also included of 1x10^5 Tregs and 5x10^4 Teffs (1:2 Teffs:Tregs) as well as resting and stimulated Tregs and Teffs individually as controls. Treg Suppression Inspector (Miltenyi Biotec, # 130-092-909) iMACS

particles were prepared and added to the appropriate wells according to the manufacture's recommendations. Assays were performed in technical triplicate for four donors and plates were incubated for 96 hours at 37°C. At the time of readout, cells were stained with Alexa Fluor® 647 anti-human IL2RA (CD25) (Biolegend, #302618), BV711 anti-human CD4 (Biolegend, #344648), and Ghost Dye™ Red 780 (Tonbo, #13-0865-T500) and analyzed on the Attune NxT flow cytometer (#A29004).

Analysis of flow data was performed in FlowJo (v10.8.1) with gating to select for lymphocytes, singlets, live cells (Ghost Dye negative), CD4+ T cells, and Teffs (CellTrace Violet+CD25low). A gate was then set for each donor using the non-stimulated Teff only control (CellTrace Violet high peak) to establish a proliferative Teff count. A gate was also set for iMACS beads by selecting non-lymphocytes, Beads using FSC-A and Ghost Dye. An absolute proliferating Teff count was then established using the formula (proliferative Teff count x input bead count)/(Beads) which adjusts for variations in stimulation and collection abnormalities. Percent suppression was calculated as (100 – (absolute proliferating Teff count / absolute proliferating Teff count of stimulated responder only condition)) x 100. The median of the technical replicate collection plates was used to calculate percent suppression and absolute proliferating Teff count per donor for visualization.

Data availability

IL2RA screens, CUT&RUN, ChIP-seq, Bulk RNAseq, and Perturb-CITE-seq data are accessible at NCBI Gene Expression Omnibus within GEO SuperSeries GSE271090. Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD056255. Publicly available data used in this study is accessible from the following sources: KLF2 KO differentially expressed genes are

available in Freimer et al[19]. MED12 CAR-T bulk RNAseq data is available on GEO: GSE174279. CD4+ Treg STAT5A ChIP-seq data is available on ChIP Atlas- SRX212432, GSM105692.

Code accessibility

Code for main figure generation and key analyses is available on Zenodo[87]:

https://doi.org/10.5281/zenodo.13924126

References

1. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251 (2013).
2. Ong, C. T. & Corces, V. G. Enhancer function: New insights into the regulation of tissue-specific gene expression. *Nat Rev Genet* **12**, 283–293 (2011).
3. Almeida, N. *et al.* Employing core regulatory circuits to define cell identity. *EMBO J* **40**, (2021).
4. Lee, W. & Lee, G. R. Transcriptional regulation and development of regulatory T cells. *Exp Mol Med* **50**, e456-10 (2018).
5. Ota, M. *et al.* Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell* **184**, 3006-3021.e17 (2021).
6. Simeonov, D. R. *et al.* Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549**, 111–115 (2017).
7. Calderon, D. *et al.* Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat Genet* **51**, 1494–1505 (2019).
8. Ohkura, N. & Sakaguchi, S. Transcriptional and epigenetic basis of Treg cell development and function: its genetic anomalies or variations in autoimmune diseases. *Cell Res* **30**, 465–474 (2020).
9. Kmieciak, M. *et al.* Human T cells express CD25 and Foxp3 upon activation and exhibit effector/memory phenotypes without any regulatory/suppressor function. *J Transl Med* **7**, 1–7 (2009).
10. Goudy, K. *et al.* Human IL2RA null mutation mediates immunodeficiency with lymphoproliferation and autoimmunity. *Clinical Immunology* **146**, 248–261 (2013).

11. Chinen, T. *et al.* An essential role for the IL-2 receptor in T reg cell function. *Nat Immunol* **17**, 1322–1333 (2016).
12. Field, A. & Adelman, K. Evaluating Enhancer Function and Transcription. *Annu Rev Biochem* **89**, 213–234 (2020).
13. Panigrahi, A. & O'Malley, B. W. Mechanisms of enhancer action: the known and the unknown. *Genome Biol* **22**, 1–30 (2021).
14. Propper, D. J. & Balkwill, F. R. Harnessing cytokines and chemokines for cancer therapy. *Nat Rev Clin Oncol* **19**, 237–253 (2022).
15. Hernandez, R., Põder, J., LaPorte, K. M. & Malek, T. R. Engineering IL-2 for immunotherapy of autoimmunity and cancer. *Nat Rev Immunol* (2022) doi:10.1038/s41577-022-00680-w.
16. Ribas, A. & Wolchok, J. D. Cancer immunotherapy using checkpoint blockade. *Science (1979)* **359**, 1350–1355 (2018).
17. Spolski, R., Li, P. & Leonard, W. J. Biology and regulation of IL-2: from molecular mechanisms to human therapy. *Nat Rev Immunol* **18**, 648–659 (2018).
18. Graßhoff, H. *et al.* Low-Dose IL-2 Therapy in Autoimmune and Rheumatic Diseases. *Front Immunol* **12**, (2021).
19. Freimer, J. W. *et al.* Systematic discovery and perturbation of regulatory genes in human T cells reveals the architecture of immune networks. *Nat Genet* **54**, 1133–1144 (2022).
20. Schumann, K. *et al.* Functional CRISPR dissection of gene networks controlling human regulatory T cell identity. *Nat Immunol* **21**, 1456–1466 (2020).

21. Cortez, J. T. *et al.* CRISPR screen in regulatory T cells reveals modulators of Foxp3. *Nature* **582**, 416–420 (2020).

22. Seo, H. *et al.* BATF and IRF4 cooperate to counter exhaustion in tumor-infiltrating CAR T cells. *Nat Immunol* **22**, 983–995 (2021).

23. Vasanthakumar, A. *et al.* The transcriptional regulators IRF4, BATF and IL-33 orchestrate development and maintenance of adipose tissue-resident regulatory T cells. *Nat Immunol* **16**, 276–285 (2015).

24. Khatun, A. *et al.* BATF is Required for Treg Homeostasis and Stability to Prevent Autoimmune Pathology. *Advanced Science* **10**, (2023).

25. Schmidt, R. *et al.* CRISPR activation and interference screens decode stimulation responses in primary human T cells. *Science (1979)* **375**, (2022).

26. Alon, U. Network motifs: Theory and experimental approaches. *Nature Reviews Genetics* vol. 8 450–461 Preprint at https://doi.org/10.1038/nrg2102 (2007).

27. Richter, W. F., Nayak, S., Iwasa, J. & Taatjes, D. J. The Mediator complex as a master regulator of transcription by RNA polymerase II. *Nat Rev Mol Cell Biol* **23**, 732–749 (2022).

28. Knuesel, M. T., Meyer, K. D., Bernecky, C. & Taatjes, D. J. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes Dev* **23**, 439–451 (2009).

29. El Khattabi, L. *et al.* A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. *Cell* **178**, 1145-1158.e20 (2019).

30. Raisner, R. *et al.* Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3K27 Acetylation. *Cell Rep* **24**, 1722–1729 (2018).

31. Hoffmann, P. *et al.* Loss of FOXP3 expression in natural human CD4+ CD25+ regulatory T cells upon repetitive in vitro stimulation. *Eur J Immunol* **39**, 1088–1097 (2009).

32. Policarpi, C., Munafò, M., Tsagkris, S., Carlini, V. & Hackett, J. A. Systematic epigenome editing captures the context-dependent instructive function of chromatin modifications. *Nat Genet* (2024) doi:10.1038/s41588-024-01706-w.

33. Wang, H. *et al.* H3K4me3 regulates RNA polymerase II promoter-proximal pause-release. *Nature* **615**, 339–348 (2023).

34. Hu, S. *et al.* H3K4me2/3 modulate the stability of RNA polymerase II pausing. *Cell Research* vol. 33 403–406 Preprint at https://doi.org/10.1038/s41422-023-00794-3 (2023).

35. Rahl, P. B. *et al.* C-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).

36. Freitas, K. A. *et al.* Enhanced T cell effector activity by targeting the Mediator kinase module. *Science (1979)* **378**, (2022).

37. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).

38. Aranda-Orgilles, B. *et al.* MED12 Regulates HSC-Specific Enhancers Independently of Mediator Kinase Activity to Control Hematopoiesis. *Cell Stem Cell* **19**, 784–799 (2016).

39. Weber, E. W. *et al.* Transient rest restores functionality in exhausted CAR-T cells through epigenetic remodeling. *Science (1979)* **372**, (2021).

40. Eyquem, J. *et al.* Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* **543**, 113–117 (2017).

41. Yamamoto, T. N. *et al.* T cells genetically engineered to overcome death signaling enhance adoptive cancer immunotherapy. *Journal of Clinical Investigation* **129**, 1551–1565 (2019).

42. Mowery, C. T. *et al.* Systematic decoding of cis gene regulation defines context-dependent control of the multi-gene costimulatory receptor locus in human T cells. *Nat Genet* **56**, 1156–1167 (2024).

43. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15**, 554 (2014).

44. Clement, K. *et al.* CRISPResso2 provides accurate and rapof genome editing analysis. *Nat Biotechnol* **37**, 220–224 (2019).

45. Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* **9**, 1–15 (2018).

46. Pacalin, N. M. *et al.* Bidirectional epigenetic editing reveals hierarchies in gene regulation. *Nat Biotechnol* (2024) doi:10.1038/s41587-024-02213-3.

47. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, (2017).

48. Heaton, H. *et al.* Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat Methods* **17**, 615–620 (2020).

49. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420 (2018).

50. Steinhart, Z. Code repository for: “CRISPR activation and interference screens decode stimulation responses in primary human T cells”. *Zenodo* (2022).

51. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289–1296 (2019).

52. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 1–21 (2014).

53. Salavaty, A., Ramialison, M. & Currie, P. D. Integrated Value of Influence: An Integrative Method for the Identification of the Most Influential Nodes within Networks. *Patterns* **1**, (2020).

54. Pedersen T. ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. Preprint at (2024).

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

56. Smith, T., Heger, A. & Sudbery, I. UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491–499 (2017).

57. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Preprint at (2010).

58. Wang, L., Wang, S. & Li, W. RSeQC: Quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).

59. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

60. Ulgen, E., Ozisik, O. & Sezerman, O. U. PathfindR: An R package for comprehensive identification of enriched pathways in omics data through active subnetworks. *Front Genet* **10**, (2019).

61. Shannon, P. *et al.* Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504 (2003).

62. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

63. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367–1372 (2008).

64. Teo, G. *et al.* SAINTexpress: Improvements and additional features in Significance Analysis of INTeractome software. *J Proteomics* **100**, 37–43 (2014).

65. Choi, M. *et al.* MSstats: An R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524–2526 (2014).

66. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**, D638–D646 (2023).

67. Arang, N. *et al.* Whole-genome CRISPR screening identifies PI3K/AKT as a downstream component of the oncogenic GNAQ–focal adhesion kinase signaling circuitry. *Journal of Biological Chemistry* **299**, (2023).

68. Zheng Y. No Title. *CUTTag_tutorial* https://yezhengstat.github.io/CUTTag_tutorial/.

69. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).

70. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).

71. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

72. Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 1–11 (2019).

73. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* **9**, (2013).

74. Texari, L. *et al.* An optimized protocol for rapid, sensitive and robust on-bead ChIP-seq from primary cells. *STAR Protoc* **2**, (2021).

75. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

76. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, (2019).

77. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* **38**, 276–278 (2020).

78. Orlando, D. A. *et al.* Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep* **9**, 1163–1170 (2014).

79. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, (2008).

80. Krebs, W. *et al.* Optimization of transcription factor binding map accuracy utilizing knockout-mouse models. *Nucleic Acids Res* **42**, 13051–13060 (2014).

81. Jain, D., Baldi, S., Zabel, A., Straub, T. & Becker, P. B. Active promoters give rise to false positive 'Phantom Peaks' in ChIP-seq experiments. *Nucleic Acids Res* **43**, 6959–6968 (2015).

82. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: An R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).

83. Zou, Z., Ohta, T. & Oki, S. ChIP-Atlas 3.0: a data-mining suite to explore chromosome architecture together with large-scale regulome data. *Nucleic Acids Res* (2024) doi:10.1093/nar/gkae358.

84. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, (2014).

85. Morgan M, Obenchain V, Hester J & Pagès H. SummarizedExperiment: SummarizedExperiment container. Preprint at (2024).

86. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* **14**, (2013).

87. Arce, M. Code repository for Central control of dynamic gene circuits governs T cell rest and activation. Zenodo (2024).

Chapter 3 Inborn error of immunity trans-regulatory gene network

Abstract

The effects of genetic variation on complex traits act mainly through changes in gene regulation. Although many genetic variants have been linked to target genes in *cis*, the *trans*-regulatory cascade mediating their effects remains largely uncharacterized. Mapping *trans*-regulators based on natural genetic variation has been challenging due to small effects, but experimental perturbations offer a complementary approach. Using CRISPR, we knocked out 84 genes in primary $CD4^+$ T cells, targeting inborn error of immunity (IEI) disease transcription factors (TFs) and TFs without immune disease association. We developed a novel gene network inference method called linear latent causal Bayes (LLCB) to estimate the network from perturbation data and observed 211 regulatory connections between genes. We characterized programs affected by the TFs, which we associated with immune genome-wide association study (GWAS) genes, finding that JAK-STAT family members are regulated by *KMT2A*, an epigenetic regulator. These analyses reveal the *trans*-regulatory cascades linking GWAS genes to signaling pathways.

Introduction

A primary mission of human genetics is to discover genetic variation that is associated with disease. Genome-wide association studies (GWAS) have identified thousands of variant-disease pairs in recent years, spanning disease, behavioral, and molecular phenotypes. Functional analyses of GWAS loci have revealed that most GWAS single-nucleotide polymorphisms (SNPs) are non-coding, demonstrating that the effects of genetic variation on complex traits largely manifest through regulatory variation[1,2]. However, the identification of the molecular consequences of non-coding SNPs has proven challenging. Recent efforts have catalogued expression quantitative trait loci (eQTLs) across diverse tissues and contexts[3–6]. These eQTL studies have been very successful in identifying genetic variation that associates with expression variation of nearby genes in *cis*. However, except for a small number of examples, the *trans*-regulatory cascade beyond the associated locus of these *cis*-acting genetic variants remains largely unknown. Recent analyses of the genetic architecture of complex traits have shown that the bulk (60-90%) of expression heritability is mediated through a constellation of *trans* effects which typically have small effects individually but have a large contribution in aggregate[7–9]. These *trans* effects are difficult to discover with natural genetic variation because their effect sizes are small and may only exist in contexts that are missed in bulk-tissue steady state models of gene expression[10–13]. Thus, alternative approaches are needed to map the trans-regulatory effects of *cis*-acting eQTLs.

We previously mapped the trans-regulators of key autoimmune disease genes, including *IL2RA*, *IL2*, and *CTLA4* in primary human CD4+ T-cells using CRISPR knock-outs (KOs)[14,15]. In contrast to natural genetic variation, experimental perturbations enable the manipulation of gene expression in ways that are unlikely to be permitted by natural selection[16]. We therefore sought to apply this approach to inborn errors of immunity (IEI) genes, which are associated with

monogenic immune disease spanning regulation and function[17]. Although hundreds of these genes have been reported, the transcriptional consequences of their loss of function remain largely uncharacterized. We selected 30 IEI transcription factors (TFs) for CRISPR ablation in human CD4+ T cells both to characterize their function and to construct a regulatory network. CD4+ T cells have previously been implicated as a causal cell type in the pathology of many autoimmune traits, including rheumatoid arthritis, multiple sclerosis, type 1 diabetes, among others[18–20]. To enable characterization of the properties of the IEI TFs as a whole, we selected 30 background TFs that are matched to the IEI genes in terms of the constraint metric pLI (probability of loss-of-function intolerance[21]) and expression level in CD4+ T-cells but have not been implicated in GWAS of immune phenotypes. We also included 24 upstream regulators of *IL2RA* which we had previously perturbed using the same protocol[14], because these genes are likely enriched for master regulators of CD4+ GRNs. In total, we perturbed 84 genes from three gene sets which we used to construct a high-fidelity gene network relevant to immune disease.

Building on recent advances in the causal inference literature[22,23], we developed a novel statistical method for estimating causal gene-regulatory networks (GRNs) from perturbation data. In contrast to differential expression or correlation analyses, incorporation of causal inference approaches enables the estimation of both direct and indirect regulatory effects, where edges are interpreted as direct effects. We emphasize that in this work the term ‘direct effect’ is used to convey that the effect of one gene on another is adjusted for confounding pathways among other perturbed genes, rather than a claim of physical interaction. Direct effects are useful because they facilitate a coherent interpretation of gene networks as directed probabilistic graphical models. Our approach differs from many other gene networks in two key ways: 1) because our network is derived from experimental perturbations, the edges are much more likely to be causal than the edges in a network estimated from observational co-expression data, where the constituent variation is often of an unknown genesis; 2) our method

enables estimation of possibly cyclic graphs, rather than the common restriction to directed acyclic graphs (DAGs)[22,24–26]. Human genetics has identified several examples of cyclic regulatory behavior[27], so the restriction of GRNs to DAGs represents an artificial constraint that we circumvent with appropriate statistical technology.

We report the causal, cyclic GRN derived from applying our novel statistical method to the 84 CRISPR KOs. Because this method is a Bayesian modification of the Linear Latent Causal (LLC) algorithm, we refer to our method as LLC Bayes (LLCB). Using our network we show that all the genes, including IEI and background TFs, form a highly interconnected network, rather than distinct communities of disease and background genes. We then identified nine coherent gene programs among the 84 KOs and their downstream genes, which we characterized using enrichment analyses to identify points of functional convergence in T cell biology. In addition to downstream characterization, we used GWAS summary statistic heritability analyses to estimate the contribution of gene program linked SNPs to immune trait heritability. This profiling highlighted the importance of a module comprised of key JAK-STAT-IL2 signaling regulators and *KMT2A*, a global epigenetic regulator that we observed to be upstream of classic IL2 signaling TFs and receptors, including *IRF4*, *STAT5B*, and *IL2RA*.

In summary, we perturbed a diverse set of genes to characterize the immune regulatory landscape and develop novel statistical methodology to characterize the CD4+ T cell network centered around immune disease genes. Our network reveals the entire trans-regulatory cascade of these gene programs and elucidates the transcriptional logic of immune GWAS loci.

Results

Perturbation of IEI TFs and matched background TFs

To construct a network enriched for genes relevant to immune disease in CD4+ T cells we perturbed 30 TFs from the IEI genes implicated in Mendelian forms of immune disease[17]. We also included 30 background TFs that were not annotated for immune function but were matched on gene constraint and expression to the IEI TFs in order to characterize the properties that distinguish IEI TFs (**Figure 3.1A-B**). Lastly, to expand the breadth of our network, we integrated data from 24 previously mapped IL2RA regulators[14]. (**Figure 3.1C**). We used CRISPR Cas9 ribonucleoproteins (RNPs) to perform arrayed perturbations in CD4+ T cells from three donors as described in Freimer et al.[14] We validated the efficiency of our CRISPR editing by genotyping the 60 additional targeted loci, which indicated a high editing efficiency (**Supplemental Figure 3.1A-B**).

Using bulk RNA-seq, we detected ~13,000 genes that were expressed highly enough for analysis (Methods). As our data were generated in two batches, we performed stringent quality control of the RNA-seq data. We performed alignment and gene count quantification using one pipeline on the 84 samples and performed principal component analysis (PCA) of the normalized expression data. Pathway enrichment analysis revealed that the first four PCs were associated with very broad biological phenomena including cell cycle regulation and ribosome activity. Because the PCs also captured batch effects, we included the first four PCs as covariates in downstream analyses. Regressing out PCs has previously been shown to improve inference of gene networks[28].

Perturbed genes form a highly interconnected network

Next, we developed a statistical method to estimate the GRN among the 84 genes. We extended the linear-latent-causal (LLC) method introduced by Hyttinen et al.[23] by recasting the statistical estimand in a Bayesian framework, which enabled the incorporation of prior knowledge about the properties of biological networks. The theory, construction and validation of this statistical method, which we termed LLCB, is detailed in the associated publication Weinstock et al. We used LLCB to estimate the causal CD4+ GRN among the 84 genes, observing a highly connected constellation of factors (**Figure 3.2**).

Next, we expanded our network analyses to include all 12,803 other genes that were expressed highly enough for analysis, which we refer to as non-perturbed genes. We estimated the effects of the 84 perturbed genes on the non-perturbed genes using two methods. First, we used a traditional differential expression approach using DESeq2[29], where we regressed the normalized expression of each gene against a design matrix that included an indicator for the perturbation status of the sample, the donor identity, and the first four expression PCs. Next, we used mashr[30] to perform statistical shrinkage of the differential expression estimates. We refer to these results as DEG-mashr estimates. To model the effects of multiple upstream TFs at the same time, we developed a novel statistical estimator of the bipartite graph (BG), which models the effects of the 84 perturbed genes on the 12,803 non-perturbed genes jointly in a single linear model. In contrast to a differential expression approach the BG model is less likely to detect redundant causal pathways.

Among the non-perturbed genes, 7,299 (57%) had an incoming edge from at least one KO. Among the non-perturbed genes with at least one incoming edge, the median number of incoming edges was 5. The median number of downstream effects from the BG model was

251.5, ranging from 52 (*EGR3*) to 2,634 (*MED12*). Estimates from both the DEG-mashr and BG approaches revealed the striking enrichment of *IL2RA* regulators among the genes with the largest number of downstream connections (**Figure 3.3A**). We observed that *MED12* and *CBFB* regulated more genes than any canonical T-cell transcription factor. *MED12* is a sub-unit of the mediator complex, which transmits signals from enhancer bound TFs to RNA-polymerase II bound at the promoter[31,32]. Despite its large effects, *MED12* has never been reported in any autoimmune GWAS, nor does it have a known cis-eQTL in CD4+ T-cells[6], underscoring the value of perturbations for characterizing its function. The centrality of MED12 in the network of perturbed TFs (**Figure 3.2**) and its widespread effects across the transcriptome support our observations in chapter 2 that MED12 is a critical regulator of T cell function that operates at the top of a hierarchy of key immune regulatory genes.

Next, we asked which properties of the 12,803 non-perturbed genes were associated with regulation from the three gene groups. We performed a series of negative-binomial regressions of the incoming connections to non-perturbed genes, including six gene annotations as covariates (**Figure 3.3B**). We observed non-perturbed autoimmune GWAS genes were much more likely to be enriched for regulation from IEI TFs (~20% enrichment) and *IL2RA* regulators (~30% enrichment). S_{het} was negatively associated with incoming connections in three of the four regressions, consistent with our prior observation that gene constraint is more strongly associated with the number of outgoing connections from a gene than the number of incoming connections to the gene. We also observed that eQTL trans-eGenes were strongly enriched for incoming connections in each regression, suggesting that trans-eGenes reside in the periphery of the network with many incoming connections. Using GTEx, we also identified genes that

were only expressed in whole blood and asked whether regulation of blood specific genes varied by the three gene groups. We observed that blood-specific genes were much more likely to be regulated by IEI TFs (~20% enrichment) and *IL2RA* regulators (~40% enrichment) than background TFs. Collectively, these observations highlight that although background TFs have similar graph centrality to IEI TFs, they are much less likely to disrupt cell type-specific transcriptional pathways.

Gene modules link groups of genes to shared function

Next, we asked whether there were groups of the 84 perturbed genes with similar effects on downstream pathways among the 12,803 non-perturbed genes. Hierarchical clustering of the DEG-mashr results revealed the presence of nine gene modules (**Figure 3.4**), which we also grouped into a coarser set of super-modules. We remark that although the perturbed genes within each of these modules are mutually exclusive, the non-perturbed genes may overlap. To identify pathways that were regulated by these gene modules, we performed systematic enrichment analyses using KEGG genetic, signaling, and immune pathways[33] (**Figure 3.5A**, **Supplemental Figures 3.2-4**).

Module 1 contained the largest number of perturbed genes; 14 IEI TFs, 19 background TFs, and two *IL2RA* regulators (*RELA* and *YY1*). The remaining modules had less members and modules 3-4 were primarily *IL2RA* regulators. We observed that module 1A was enriched for disruption of MAPK and p53 signaling. Module 1B included T-bet (*TBX21*), a transcription factor that is required for interferon-gamma production and the Th1 phenotype[34], and three members of the Rel family (*NFAT5*, *RELB,* and *REL*), sub-units of NF-κb, a transcription factor complex that plays a role in T-cell activation[35]. Surprisingly, this cluster also included four background TFs without any annotated immune function (*ZNF329*, *ZNF791*, *ZBTB14*, and *ZKSCAN1*).

ZBTB7B has been observed to be required for CD4+ commitment, and interacts with NF-κB[36], but many other members of the ZBTB family, including *ZBTB14*, remain relatively uncharacterized. The high proportion of shared effects between *ZBTB14*, *T-bet*, and the Rel family proteins suggests that *ZBTB14* may have similar function to *ZBTB7B*.

Genes in super-module 2 were enriched for effects on cell cycle regulation and apoptosis. Modules 3-4 were much more strongly enriched for *IL2RA* regulators than clusters 1-2. Consistent with their annotation, every gene in module 3-4 had downstream effects on the JAK-STAT and chemokine signaling pathways. Surprisingly, *KMT2A*, a methylation writer clustered in the same module as *JAK3, STAT5A, STAT5B, IRF4,* and *IL2RA*. Although translocations of *KMT2A* have been shown to cause lymphoid malignancy[37], it has no annotated function in non-mutated cells in the JAK-STAT pathway[38]. We then examined the structure of module 4 (**Figure 3.5B**), observing that *KMT2A* is upstream of *IRF4*, *STAT5A*, and *IL2RA*, and directly regulates several downstream effector cytokines through pathways not mediated by the other perturbed genes.

Several modules were strongly enriched for cell cycle and proliferation pathways. To determine if there was a uniform effect on *in vitro* expansion within any of the modules, we quantified the number of live cells per KO compared to cells where the guide RNA targeted the safe harbor locus AAVS1 from the respective donor. Nearly all members of module 2A, which was enriched for cell cycle effects, showed a mean increase in cell counts across three donors as the result of the perturbation. Collectively, the module had a 1.14-fold increase in live cells when KO'd compared to the controls, suggesting that genes in 2A function as proliferation repressors (**Figure 3.5C**). Concordant with these observations, a recent report described the proliferation promoting effects of disruption of a module 2A member, *TET2*, in CAR-T cells[39]. Our analyses suggest that other members of 2A may have similar properties to *TET2* and thus may represent a group of genes that could be perturbed to alter engineered T-cell function. Several upstream

members of 2A upregulated three of four CDKN genes which inhibit cyclin dependent kinases and potentially lead to reduced cycling (**Supplemental Figure 3.5**). Taken together, our inference of gene modules recapitulates known regulators of immune signaling pathways and identifies novel members of these modules

Identification of KMT2A as a member of the JAK-STAT pathway

Given the large effects of module 4A on T cell differentiation, and the unexpected interaction within the module between KMT2A and canonical JAK-STAT family members, we integrated ChIP-seq and ATAC readouts to elucidate the regulatory structure of the module. We observed that *KMT2A* was a positive regulator of IL17F and IL21 expression, two Th17 secreted factors (**Figure 3.6A**). Notably, IL17F had a striking decrease in expression (-5.9 log2 fold change) upon *KMT2A* KO. We also observed concordant decreases in chromatin accessibility near (5.7 kb and 40 kb upstream of TSS) IL17F and IL21 upon KO of *KMT2A* via ATAC-seq. We then intersected the differentially accessible chromatin regions from the *KMT2A* KO condition with each of the KOs within module 4 and observed that *STAT5B* shared several differentially expressed sites, including possible distal enhancer regions upstream of IL17F (**Supplemental Figure 3.6**). An additional Th17 secreted factor, IL22, also had a shared region between the two conditions, although the transcript was only differentially expressed in the *STAT5B* KO. The *STAT5B* KO also abrogated chromatin accessibility 5.7 Kb upstream of the IL17F promoter, in a region bound by *STAT5B* in CD4+ T cell ChIP-seq (**Figure 3.6B**). Because *KMT2A* is a methyltransferase that deposits activating methylation marks on H3K4, we then asked whether H3K4me3 was present at the same locus in activated Th17s, finding broad H3K4me3 within the region (**Figure 3.6B**). These observations led us to suggest the following mechanism for the regulatory logic of module 4: *KMT2A*, a global epigenetic regulator of transcription, collaborates with downstream factors, including members of JAK-STAT, to positively regulate IL17F through modulation of a putative IL17F specific enhancer.

Discussion

Human genetics has been remarkably productive in discovering complex-trait associated SNPs. There are now several resources to map the effects of these SNPs to molecular phenotypes in *cis,* however, the development of maps of the regulatory cascades of these SNPs has progressed much more slowly. Enabled by recent innovations in large-scale perturbation technologies, we are now able to systematically perturb large numbers of genes in primary human cell contexts. These perturbations complement natural genetic variation approaches to mapping trans-regulators as they facilitate the examination of biological variance that is unlikely to be observed in healthy cells. After network inference with LLCB, we observed 211 trans-regulatory causal connections in our upstream GRN, none of which were reported in the largest catalogue of CD4+ eQTLs performed to date[6].

We developed LLCB to infer the gene network which builds upon recent advances in the structure learning literature to estimate a graph with edge weights that are interpretable as direct effects. This stands in contrast to the majority of effect estimates reported in the functional genomics literature, which primarily report estimates from differential expression analyses performed separately in each perturbed gene. These estimates confer results that are difficult to interpret because they do not attempt to adjust for confounding pathways in the GRN, which are known to be highly abundant in biological networks. We use LLCB to estimate the topology and effect size of these confounding pathways. We found that direct effects were generally much larger than indirect effects in magnitude, and that the largest indirect effects were mediated by local feedback cycles.

Using experimental perturbations, we investigated the properties of IEI TFs which are infrequently mutated in natural genetic variation. We performed a series of systematic analyses

that delineate the commonalities and differences among the IEI TFs, background TFs, and *IL2RA* regulators. Consistent with our previous report[14], we found that the *IL2RA* regulators were potent regulators of downstream effects. Both the IEI TFs and *IL2RA* regulators were enriched for being upstream and were much more likely than background TFs to disrupt autoimmune GWAS loci and whole blood specific genes even after adjustment for gene constraint. We also observed that the topology of the regulatory network is strongly associated with selective constraint. S_{het} was among the best predictors of the topological properties of the perturbed genes: S_{het} was strongly associated with the number of outgoing connections of a gene, but not the number of incoming connections. This is reflected in the dense downstream network identified for the *IL2RA* regulators with overall high levels of constraint, compared to the other TF groups. Overall, the difference in enrichment based on S_{het} suggests that the centrality of genes is best expressed as a multi-dimensional construct. This further highlights the value of estimating GRNs with directed edges, as opposed to estimating undirected graphs from observational co-expression data, as the richer graphical structure enables much more granular topological analyses.

Utilizing the novel connections in the GRN, we report several observations that improve annotation of canonical immune pathways. We observed that three of the background TFs (*DR1, BPTF,* and *YBX1*) regulated more downstream genes than any of the 30 IEI TFs, including *TBX21,* a master regulator of Th1 differentiation. After identifying gene modules and their downstream pathways, we observed multiple novel members of canonical gene modules, including *KMT2A* in the JAK-STAT pathway. We observed that *KMT2A*, a methyltransferase that deposits activating methylation marks, modulated the expression of canonical IL-2 signaling TFs. *KMT2A* collaborated with these TFs to upregulate IL17F, a pro-inflammatory cytokine that is secreted by Th17 cells, indicating that *KMT2A* is an under-appreciated regulator of the IL2-JAK-STAT axis and Th17 regulation. Meta-analysis of biobank autoimmune GWAS revealed a

novel risk locus in a Th17 enhancer upstream of *KMT2A*. Given the success of JAK inhibitors in treating a subset of patients with autoimmunity[40,41], our pathway analysis could offer an expanded set of candidate drug targets.

Although we have demonstrated that our regulatory network is useful for discovery of novel immune pathway biology and that it is validated by orthogonal data modalities, our study is not without limitations. While CD4+ T cells play a role in many immune pathologies, the construction of networks in more cell types and cellular contexts would undoubtedly result in increased discovery, as would the inclusion of additional perturbations. The restriction to transcriptional regulation also inhibits the interrogation of post-translational regulation, which makes the interpretation of edges from genes where post-translational regulation important challenging. For example, STAT proteins, which are known to be sensitive to phosphorylation, may regulate more genes than is estimated in our transcriptional network. The use of a bulk expression read-out, although more sensitive to genes with low expression than single cell assay transcriptome analysis, also precludes the analysis of more granular cell types and contexts which are easier to assess in parallel using single-cell profiling methodologies including Perturb-seq and single-cell eQTL studies.

Conclusion

In conclusion, we describe the gene regulatory network of key CD4+ T cell regulators. This network enabled both the broad characterization of the properties of immune disease genes and the discovery of novel regulatory connections between TFs and signaling pathways that modulate immune disease genes. We anticipate that our approach can be applied in other cell types and contexts to generate maps of the molecular consequences of regulatory variation of disease genes.

Figures

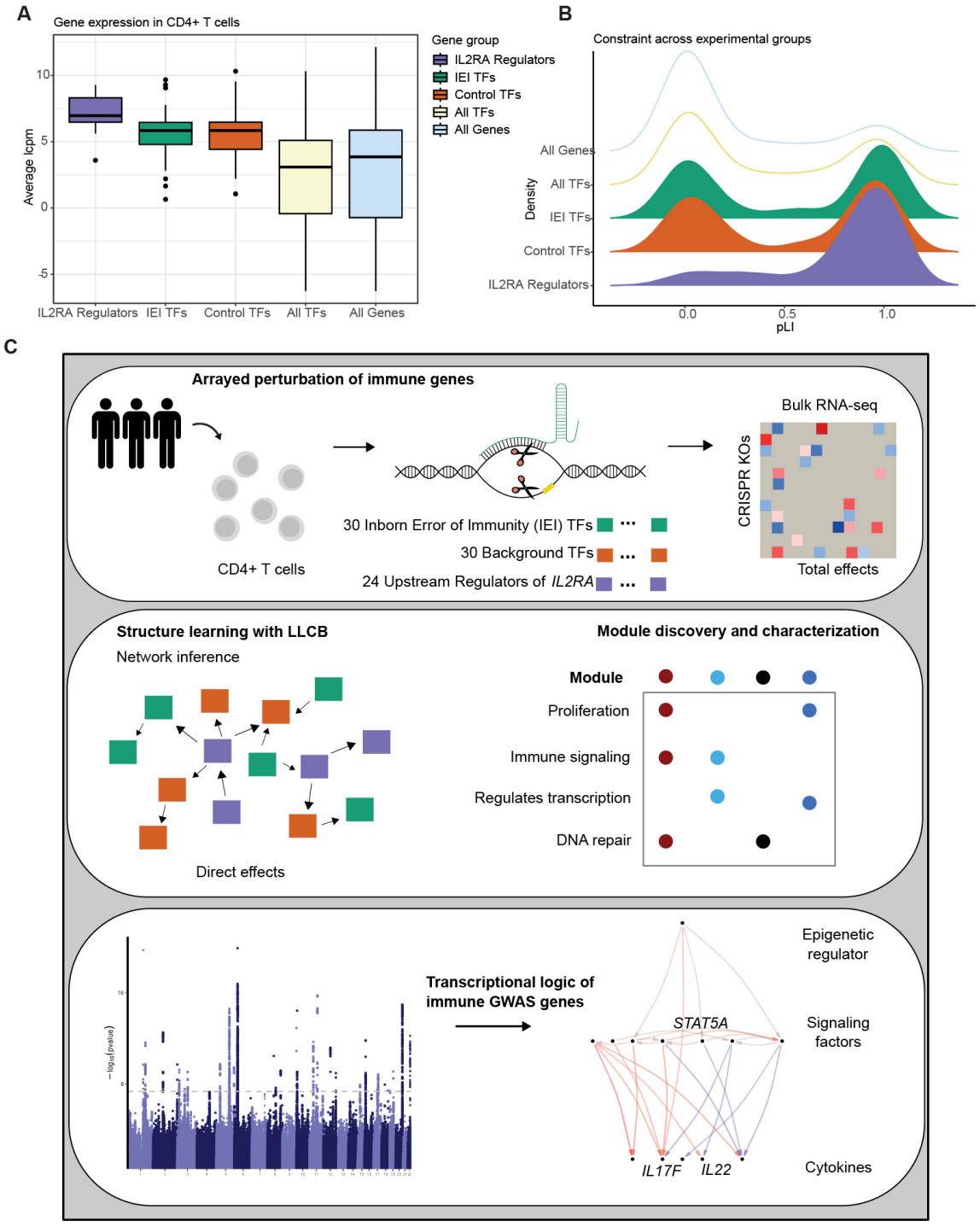


Figure 3.1 Construction of immune-centric gene regulatory network via CRISPR perturbation

A. The expression level of each gene group included in the construction of the gene regulatory network. The expression level of IEI TFs and Control TFs was matched in CD4+ T cells. **B.** The constraint metric (pLI score) for each gene group included in the construction of the gene regulatory network. The pLI of IEI TFs and Control TFs was matched in CD4+ T cells. **C.** Experimental design and analysis overview used to define key immune paths and regulators.

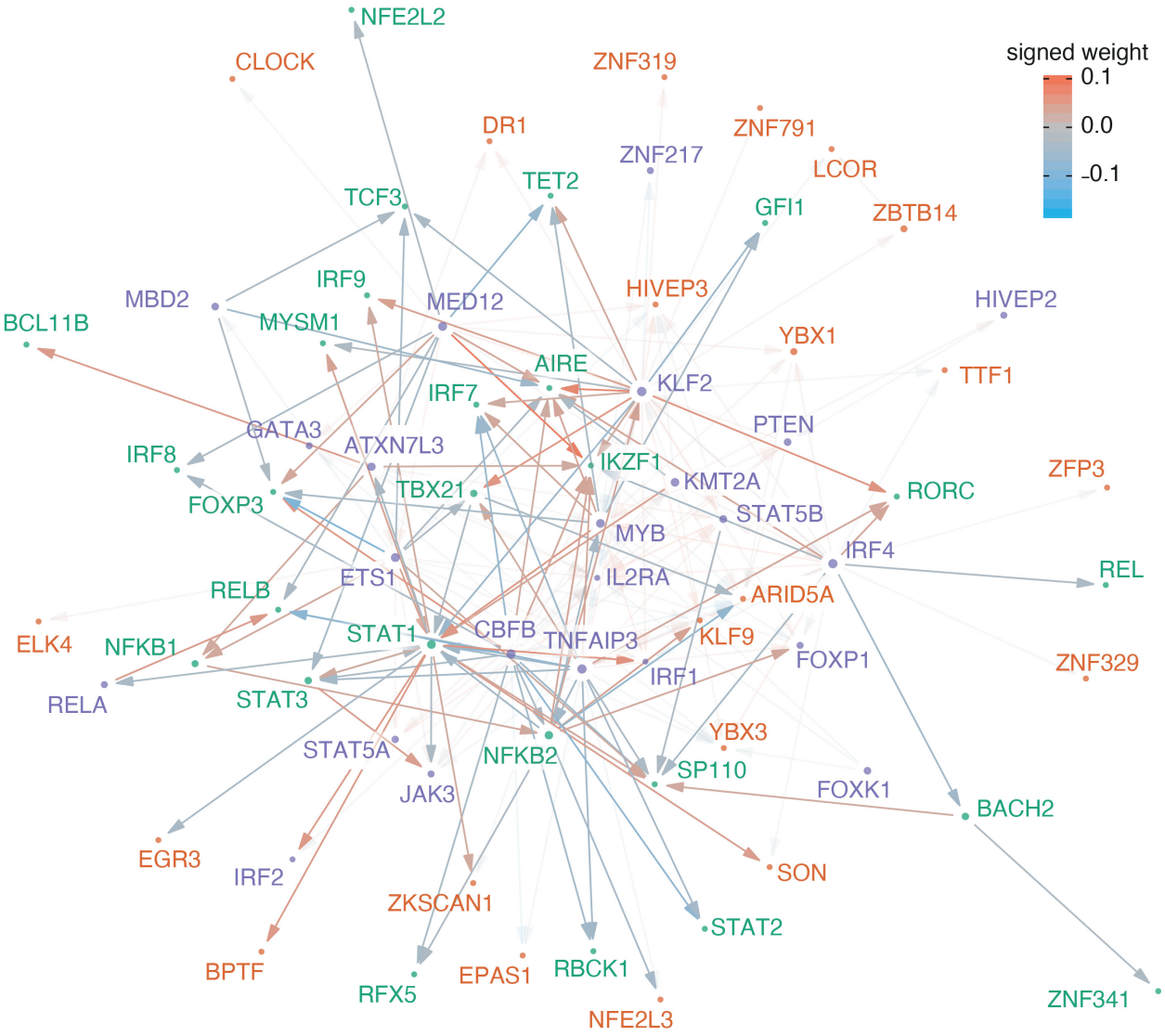


Figure 3.2 The gene network of the 84 perturbed genes
Estimate of the directed network that describes how the 84 perturbed genes interact. The radius of each point is proportional to the degree of that gene. Arrows are used to indicate directionality of the edges, such that an arrow pointing into a gene indicates that it is being regulated by another gene. For emphasis, the opacity of the edges from or to Inborn Error of Immunity transcription factors is increased, and all other edges are displayed with greater transparency. Positive values in the color scale indicate that the parent gene is a positive regulator of the child gene.

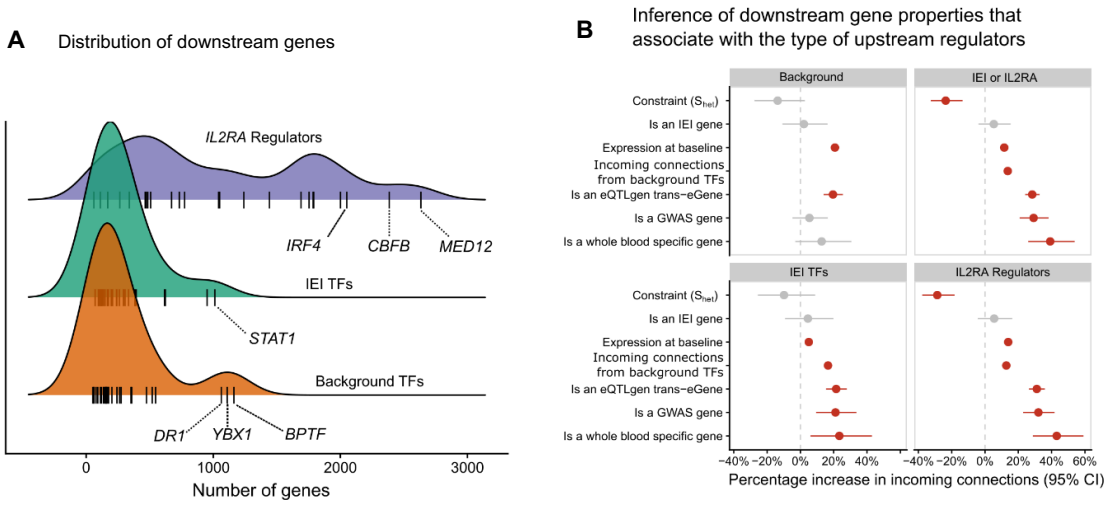


Figure 3.3 The landscape of downstream effects

A. The distribution of the number of downstream effects for each of the 84 genes, stratified by gene group. Genes that are outliers with respect to their gene group distribution are labeled. **B.** Association between the properties of downstream genes and the gene-set of the upstream regulators. Coefficients are estimated with negative binomial regressions of the gene-set specific indegree. Downstream gene annotations are indicated on the y-axis and the facets are used to indicate the gene-set of the upstream regulator.

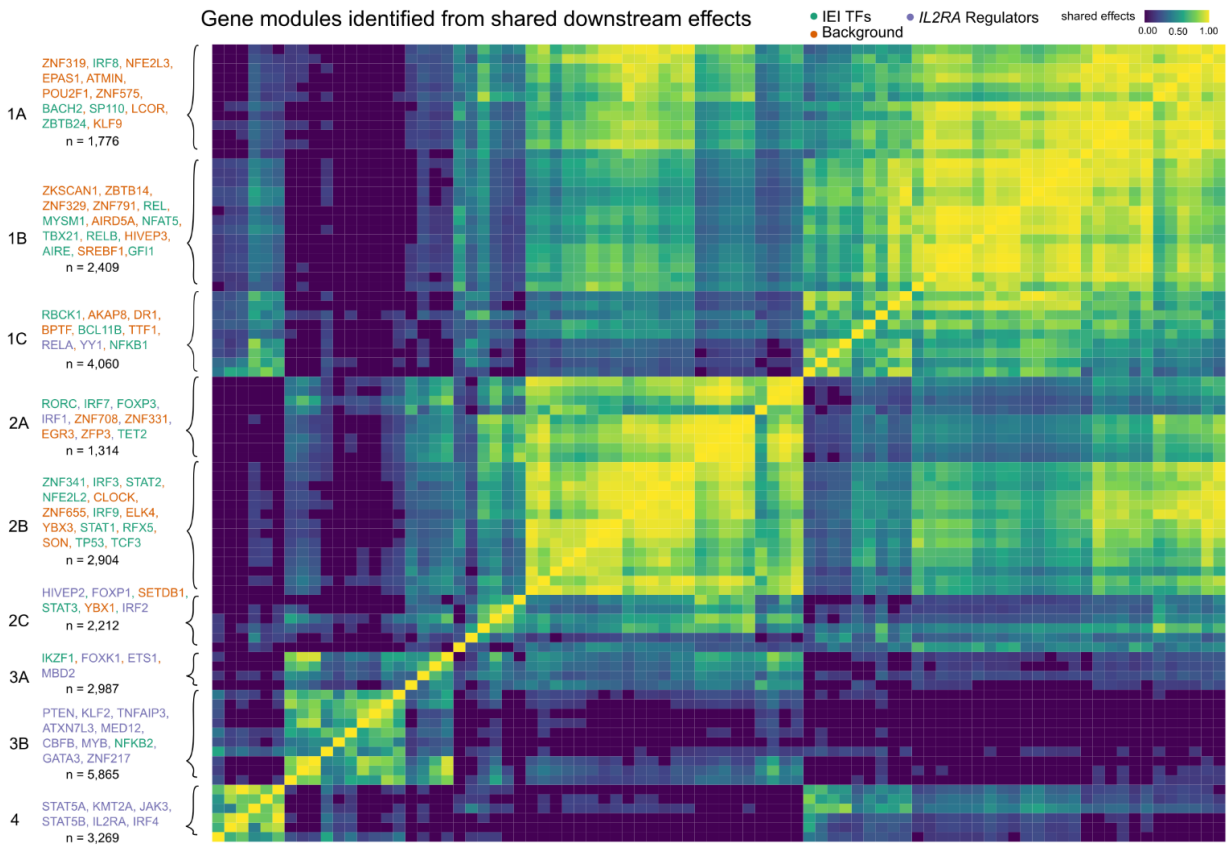


Figure 3.4 The discovery of gene modules
Hierarchical clustering is used to identify clusters of shared downstream effects. The upstream gene members within each module are labeled in the left-handed margin of the plot, and the gene group of each gene is indicated by the text color. The total number of genes in the module, including both upstream and downstream effects, is included under the list of genes.

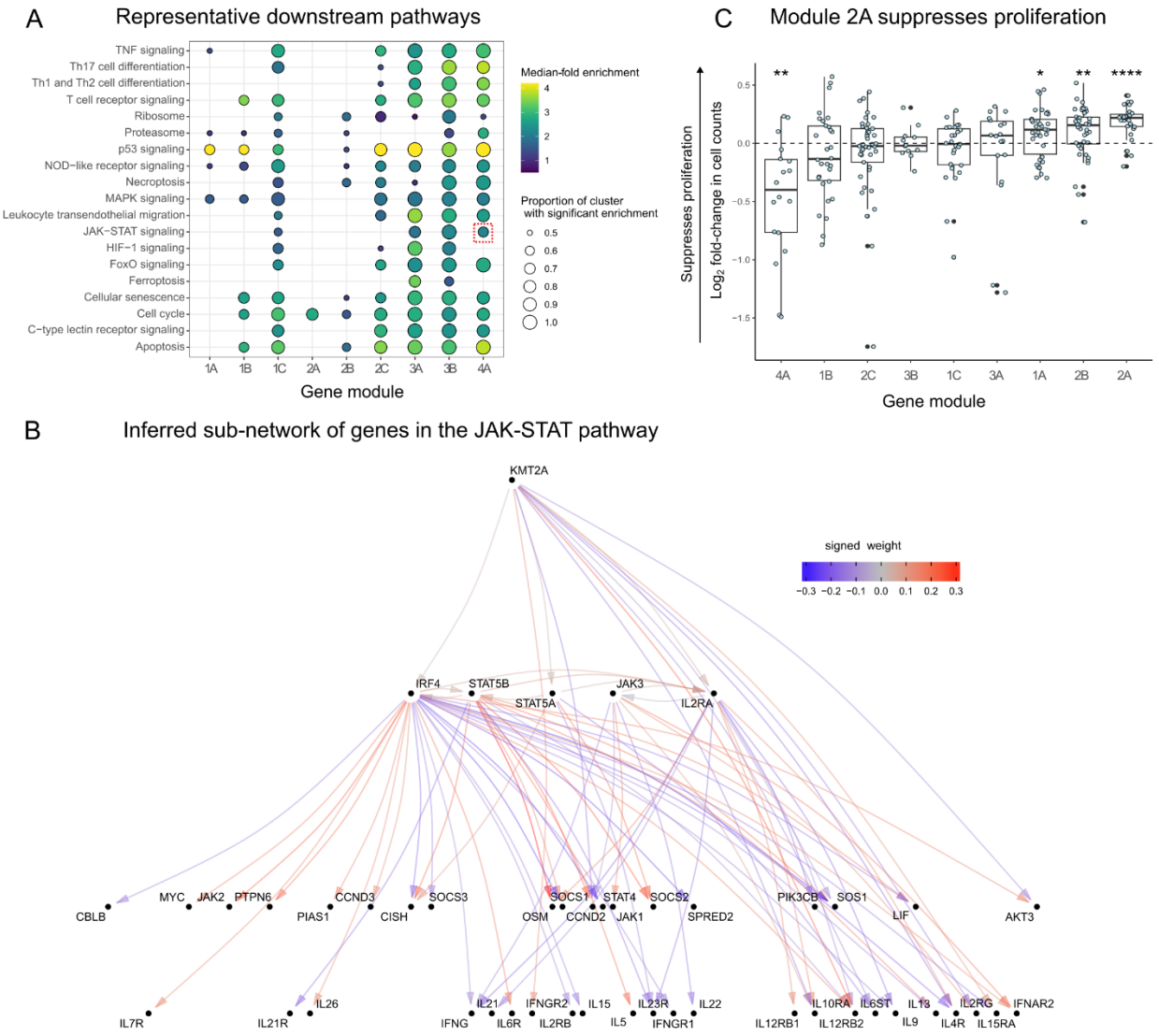


Figure 3.5 Gene module characterization

A Enrichment analyses of KEGG genetic, immune, and signaling pathways for each of the 84 perturbed genes, stratified by gene module. The JAK-STAT pathway is highlighted with a dashed-red box. The color bar maximum is set to 4. **B** The JAK-STAT sub-network, which is organized such that cytokine genes are at the bottom and upstream regulators are at the top. **C** Effects of knock outs in the gene modules on a proliferation assay. Each point represents an individual gene perturbation sample plotted as the log2 fold change sample count as compared to AAVS1 KO control samples from the same donor. (*: p-value < 0.05, **** p-value < 0.001; n=3 donors per KO, the number of KOs per cluster is reflected in Figure 3.4).

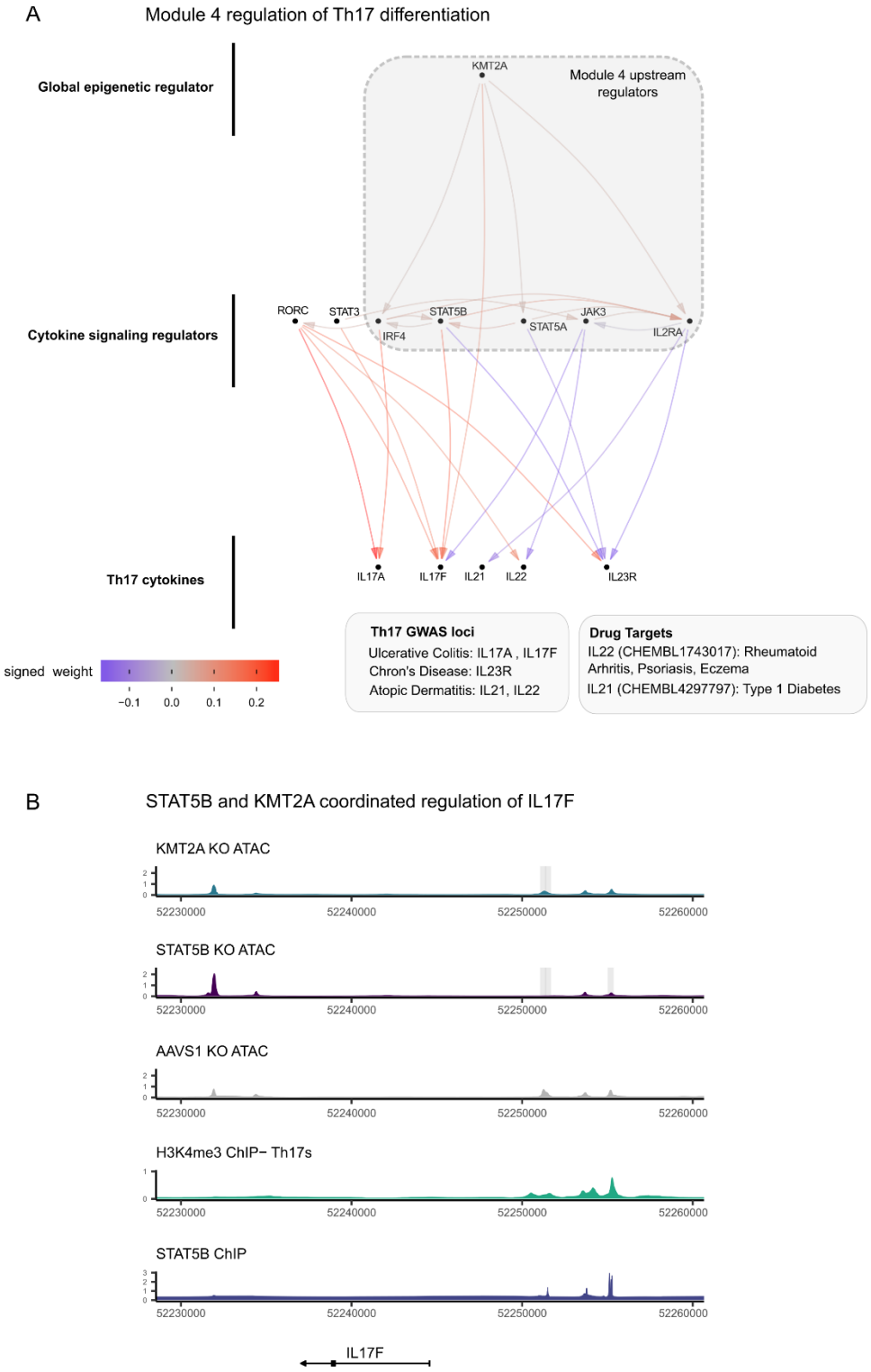
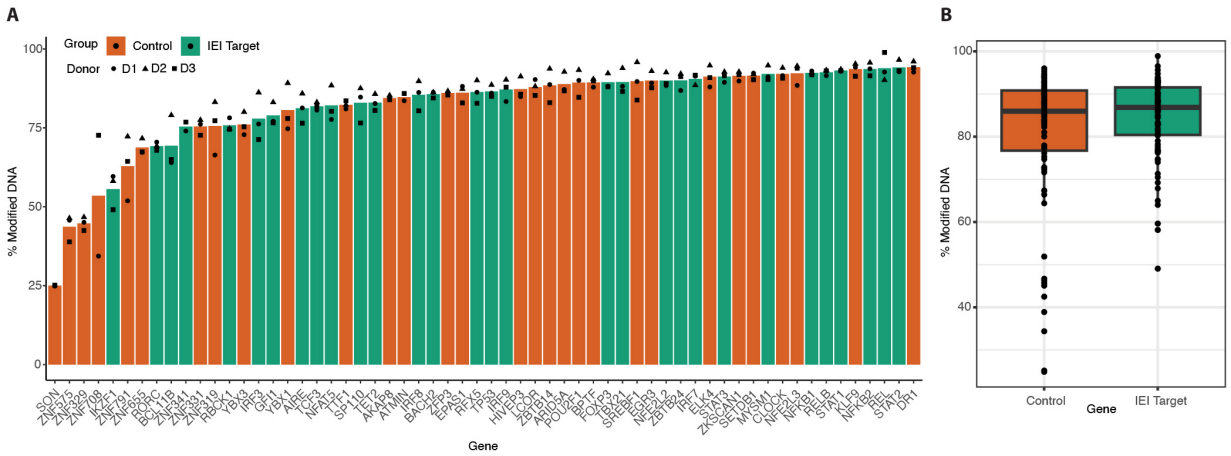
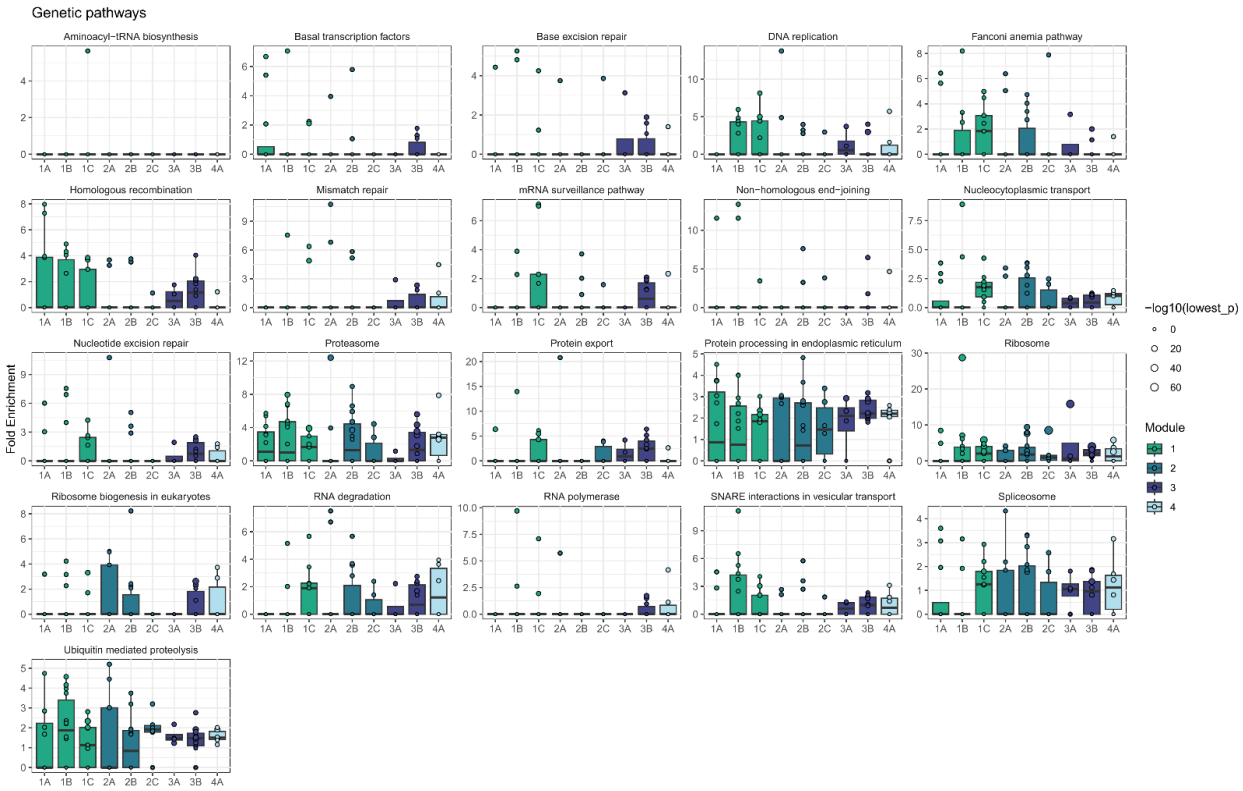


Figure 3.6 The transcriptional logic linking module 4 to GWAS loci
A. The sub-network of module 4 and Th17 cytokines. **B.** locus plot including tracks describing the functional characteristics of the region. Each track is constructed from publicly available ChIPseq data (STAR Methods) or ATAC-seq data from Freimer et al. Grey boxes indicate significantly different regions between the respective KO and AAVS1 control KO ATAC data (padj < 0.05, n = 3 donors per KO). The Y-axis displays normalized counts.

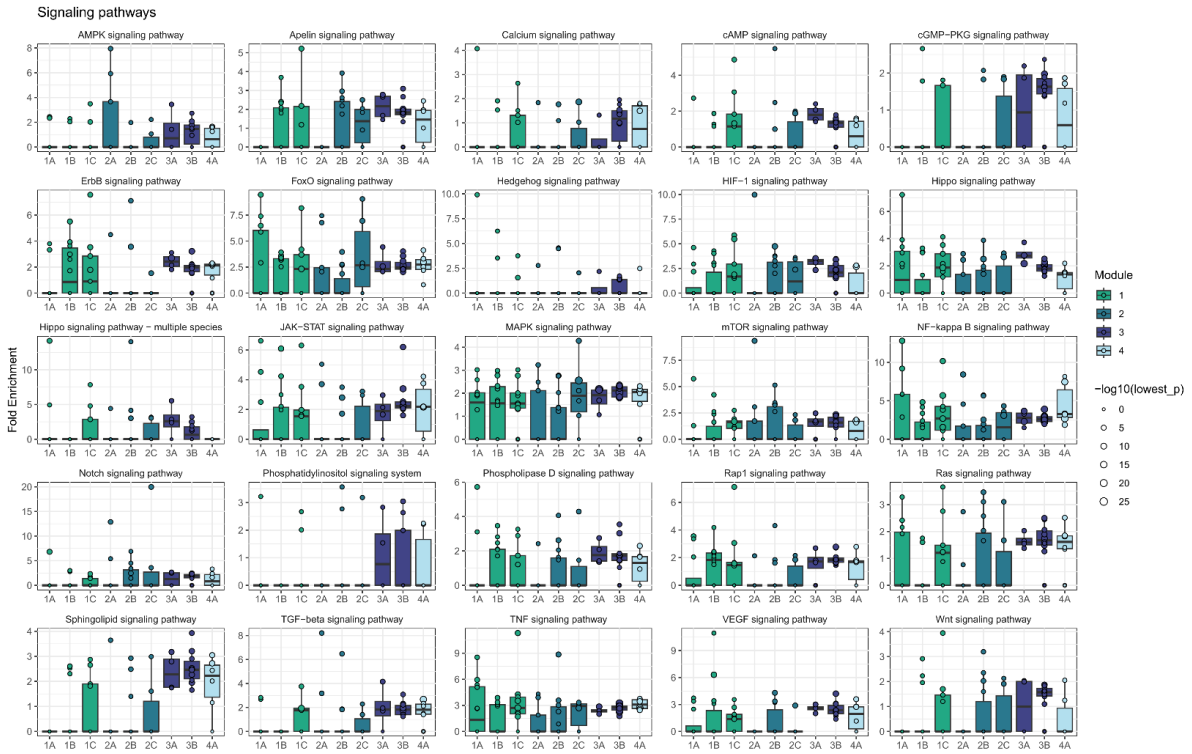
Supplemental Figures



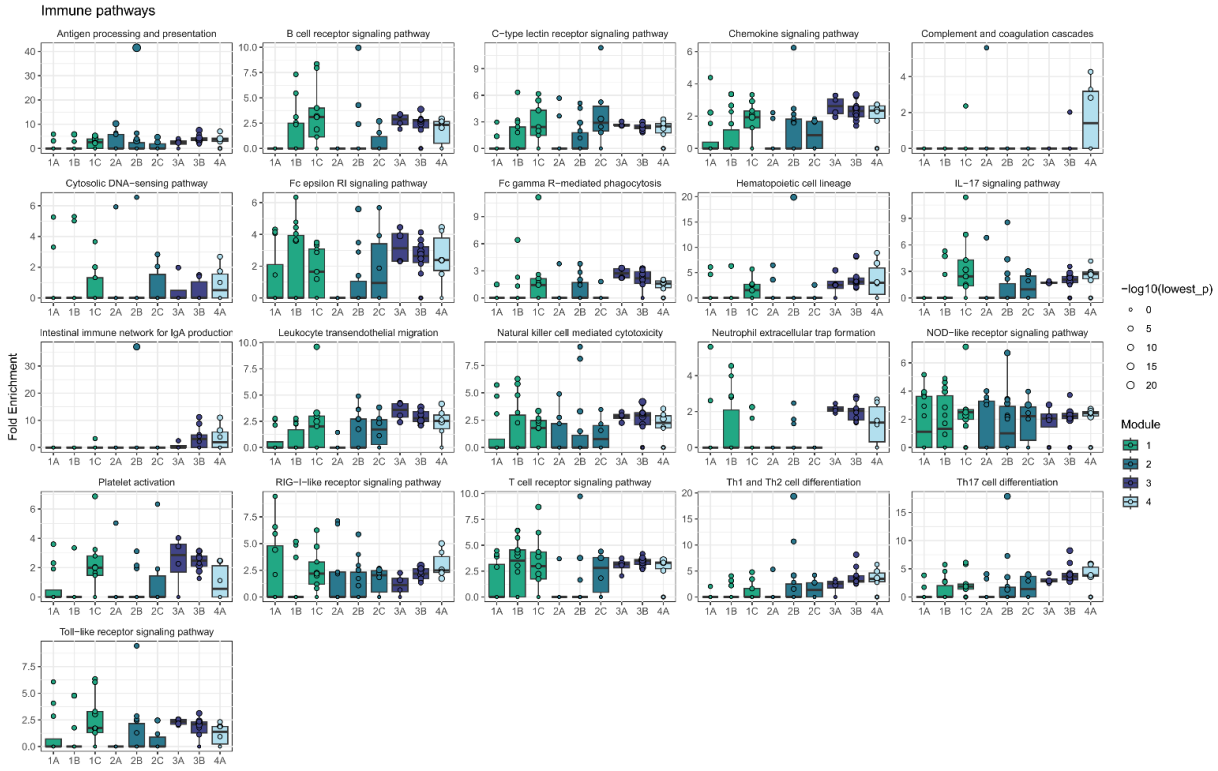
Supplemental Figure 3.1 CRISPR editing efficiency by gene group
A Percent of reads with indels, stratified by individual gene. **B** Percent of reads with indels, aggregating by gene group.



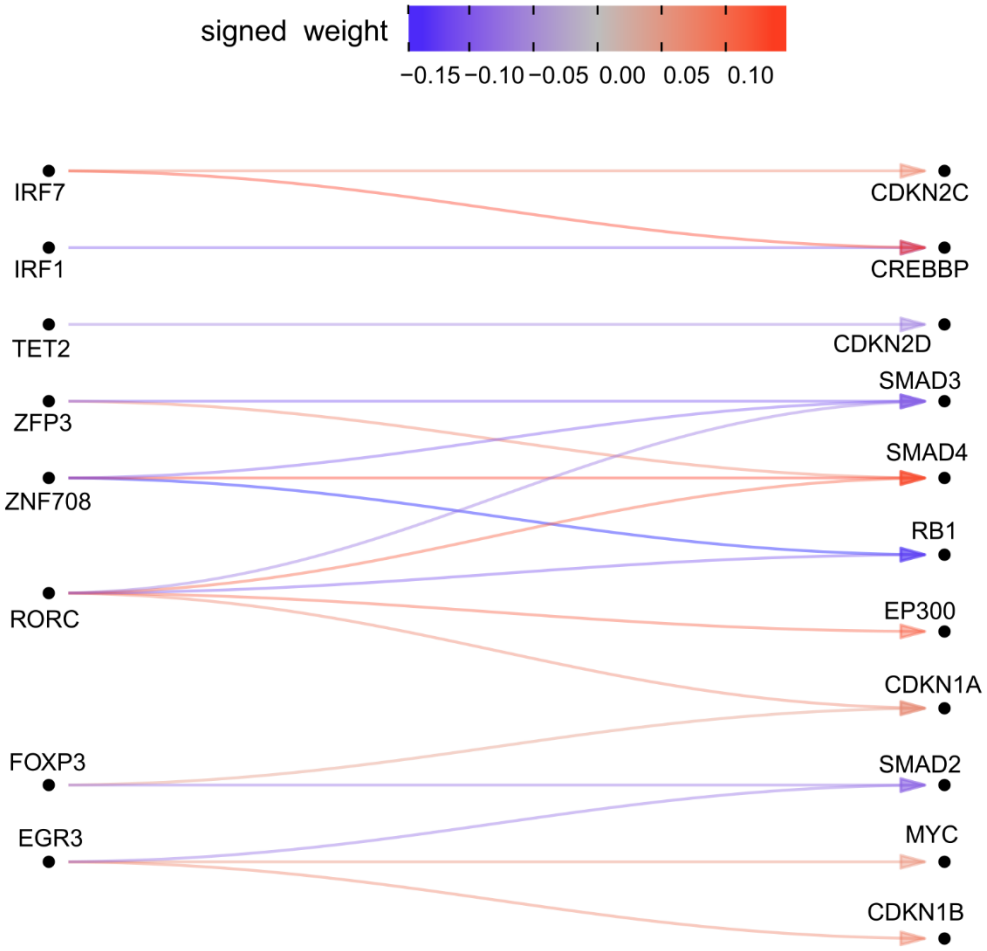
Supplemental Figure 3.2 KEGG genetic pathways module enrichment
Enrichment analyses were performed with pathfindR.



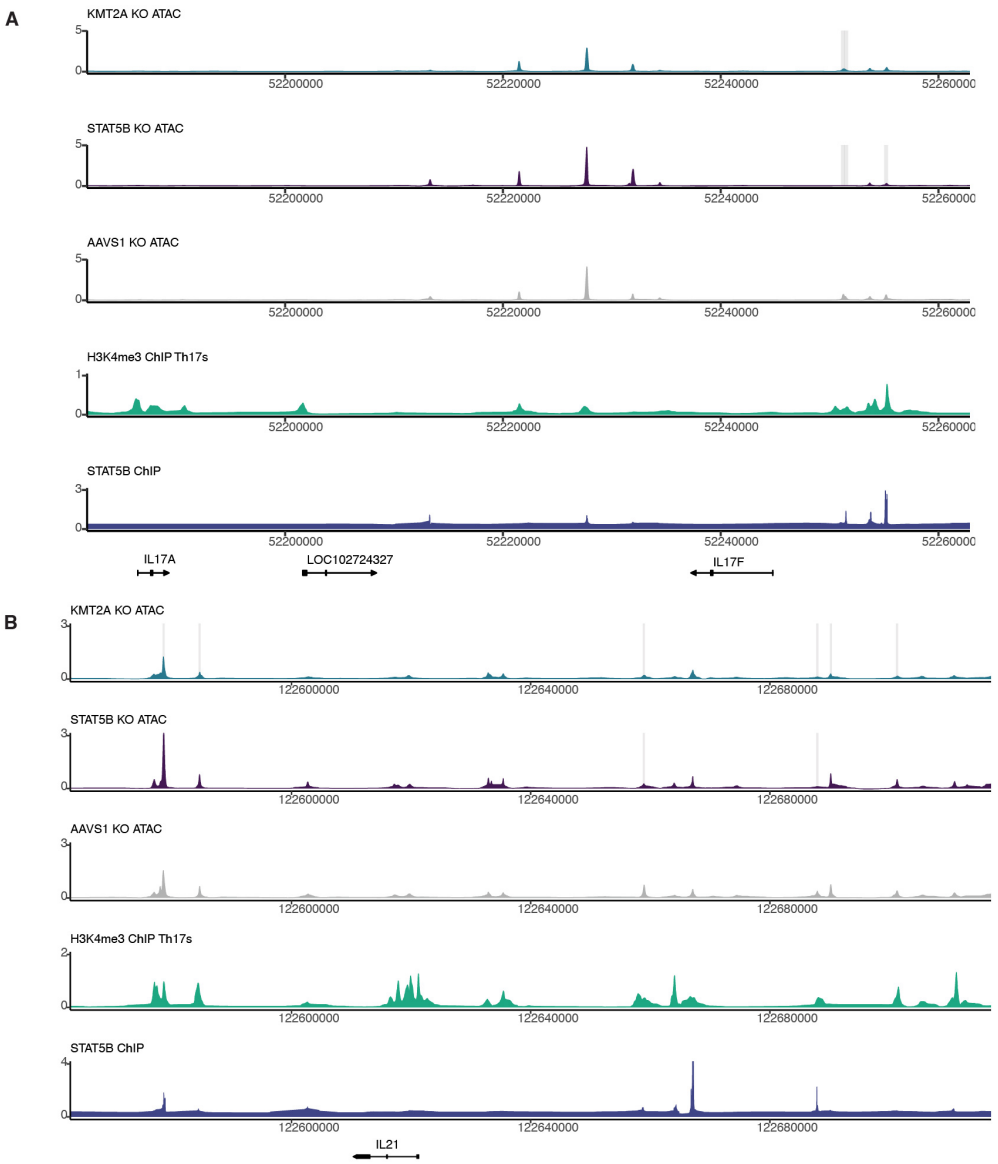
Supplemental Figure 3.3 KEGG signaling pathways module enrichment
Enrichment analyses were performed with pathfindR.



Supplemental Figure 3.4 KEGG immune pathways signaling enrichment
Enrichment analyses were performed with pathfindR.



Supplemental Figure 3.5 Network plot demonstrates the effect of the cluster 2A upstream regulators on cell-cycle genes
The network using edges estimated from the BG model are plotted. Colors indicate the effect size and arrows indicate the direction of effect. The genes on the left-hand side are among the 84 KO'd genes, and the genes on the right are genes that are listed among the KEGG cell cycle pathway genes.



Supplemental Figure 3.6 KMT2A and STAT5B jointly regulate chromatin accessibility at the IL17F locus (A) and IL21 locus (B)
For A and B, locus plot including tracks describing the functional characteristics of the region. Each track is constructed from publicly available ChIPseq data (methods) or ATAC-seq data from Freimer et al. Grey boxes indicate significantly different regions between the respective KO and AAVS1 control KO ATAC data (padj < 0.05, n = 3 donors per KO). The Y-axis displays normalized counts.

Tables

Table 3.1 CRISPR KO sgRNAs
CRISPR sgRNAs from Brunello library used for arrayed perturbations

Sequence	sgrna
GTGGCGGTACCAAAGCCTTG	AIRE_1
GCTGTCCTGAGGCCTACAAG	ARID5A_1
GGGGAGTGATGAGCTACGTG	IRF3_1
ACTGTGTGCACTGGAATGAG	ZNF319_1
TAAACTTGGACCACAACAGG	TBX21_1
TGGAGGAAGACGGAACCTGA	AAVS1_9
GCAAATGTTACAGGTCCTGG	YBX1_1
CAGGCCTGATAAAATCCATG	NFAT5_1
GTGGCGAGATGTACCCACAT	HIVEP3_1
CAAGCTCAGAAATTACACTG	MYSM1_1
TAGCCACACAGACTATTGTG	EPAS1_1
GCAGAGCCAGGAACCCCTGT	AAVS1_10
TCTGCCCTGAGGTATGCGAT	TET2_1
GTATCCTCTTCAGCACACAA	CLOCK_1
CCGCCATGGACTTCTCGCGG	BCL11B_1
CAACGATGCTCGAGAGCTGG	DR1_1
AATTTAAGGAGGTTCCTGAG	IRF9_1
GCTGATACCGTCGGCGTTGG	AAVS1_11
TCCGTGTGCTAAGCGGTGCG	ZNF575_1
GATGGCTTGGTCCATCACGT	IKZF1_1
ACATCAGGGAGATACAACAG	ZFP3_1
ATTGGGTTCGAGACAACAGG	REL_1
CGACACAGAGGAATCAGCAG	ATMIN_1
GAAGCAGGGGAGCTGGGTTT	AAVS1_12
GAGACCGAGGTGTATCACCA	STAT3_1
GATGAGCATAAGTTGCACAA	ZNF708_1
GTCGTCTGGGATCCACTACG	RORC_1
TCTGGGTGAGAACTCAGTGG	ELK4_1
GAACTATGATCCCAAGACCG	ZBTB24_1
CCTCTAAGGTTTGCTTACGA	AAVS1_4
GGACATCATAGAGCGCTCGT	SON_1
CATACCGTCTAAATCAACAG	NFE2L2_1
GCCTTCGATCAAGTGCACTG	BPTF_1

Sequence	sgrna
GATGCACTCACCTTGCACCG	IRF7_1
AAAGACAAGCAGATCCACCC	EGR3_1
GGACGCACCATTCTCACAAA	AAVS1_2
TGGGCACATTGACTTCTCCG	GFI1_1
GATGGCGCCATGATGGAAGA	TTF1_1
GTTATTGCTTGAGTGATCCG	TCF3_1
GGTCAAAGGTCTGTCCCACG	YBX3_1
CAACAAATACTCCACCCCTG	ZNF341_1
GTCACCAATCCTGTCCCTAG	AAVS1_8
CCTTACCTGAATCAATACTG	SETDB1_1
CCCACCCACAGGGATCAACG	FOXP3_1
TGAACACTTAAATCTAACAG	ZNF655_1
GTACTTACGAAATGGTACCT	RFX5_1
TGATGATGTAGAGGAAATCG	ZBTB14_1
ACAGTGGGGCCACTAGGGAC	AAVS1_7
ATTGAGCGGAAGATTCAACT	RELB_1
GAAGCAGTTCCAACTTTACG	AKAP8_1
TTTGATCCCTGTATCCAGAG	SP110_1
AGCTGGAGGACAGATAACTG	POU2F1_1
ACTCGACTACGGCGTCACCG	NFKB2_1
TGTCCCTAGTGGCCCCACTG	AAVS1_6
GCACAAGAGAATTCATAGTG	ZNF331_1
CCCAGAATAAGAATGCACCG	BACH2_1
TTTCTGACAGTAAGGTCCAG	LCOR_1
CTTCTGTGGACGATTACATG	IRF8_1
CCTGTAGAGAAGCCTCCCGG	SREBF1_1
GTCCCCTCCACCCCACAGTG	AAVS1_5
GGAGACCCTGCACTCCCATG	RBCK1_1
CTTTCCCACAGCCACTGTAG	KLF9_1
ACGGTCAAATATACCTACCA	STAT2_1
CTTCGTCCCTCGCACCAGCG	NFE2L3_1
AGAACACGAGACCAATGGTG	STAT1_1
TAAGCAAACCTTAGAGGTTC	AAVS1_3
AGAACCAGGAGGGACACTTG	ZNF329_1
CCATTGTTCAATATCGTCCG	TP53_1
GACCTGAGATGCTCGCAAGG	ZKSCAN1_1
TTGTCTATGAACATCTGTGG	NFKB1_1
GTGTGTTAGAACACTTGTGT	ZNF791_1

Sequence	sgrna
GGGGCCACTAGGGACAGGAT	AAVS1_1

Methods

Data and code availability

Bulk RNA-seq data have been deposited at GEO accessions GSE271788 and GSE171737 and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Genotyping data and cell counts are available in the supplemental tables. All original code has been deposited at github and is publicly available as of the date of publication. Code used to generate the main figures can be found here 10.5281/zenodo.12807946. LLCB code is available at 10.5281/zenodo.12807979 and https://github.com/weinstockj/LLCB .

Cell Isolation and expansion

Primary CD25-CD4+ effector T cells were isolated from fresh Human Peripheral Blood Leukopaks (STEMCELL Technologies, #70500) from healthy donors, after institutional review board–approved informed written consent (STEMCELL Technologies). Peripheral blood mononuclear cells (PBMCs) were washed twice with a 1X volume of EasySep buffer (DPBS, 2% fetal Bovine Serum (FBS), 1mM pH 8.0 EDTA). The washed PBMCs were resuspended at 200E6 cells/mL in EasySep buffer and isolated with the EasySep™ Human CD4+CD127lowCD25+ Regulatory T Cell Isolation Kit (STEMCELL Technologies, #18063), according to the manufacturer's protocol. Cells were seeded at $1x10^6$ cells/mL in complete RPMI-1640 supplemented with 10% FCS, 2 mM L-Glutamine (Fisher Scientific #25030081), 10 mM HEPES (Sigma, #H0887-100ML), 1X MEM Non-essential Amino Acids (Fisher, #11140050), 1 mM Sodium Pyruvate (Fisher Scientific #11360070), 100 U/mL Penicillin-Streptomycin (Sigma, #P4333-100ML), and 50 U/mL IL-2 (Amerisource Bergen, #10101641) and stimulated with 6.25 uL/mL ImmunoCult™ Human CD3/CD28/CD2 T Cell Activator

(STEMCELL Technologies, #10990). Following activation and electroporation, cells were split 1:2 every other day to maintain an approximate density of $1x10^6$ cells/mL.

Cas9 RNP preparation and delivery

Custom crRNAs (Dharmacon) and Dharmacon Edit-R CRISPR-Cas9 Synthetic tracrRNA (Dharmacon, #U-002005-20) were resuspended in Nuclease Free Duplex Buffer (IDT, #11-01-03-01) at 160uM stock concentration. In a 96 well plate, each crRNA was combined with tracrRNA at a 1:1 molar ratio and incubated at 37°C for 30 minutes. Custom crRNA sequences are included in **Table 3.1**. Single-stranded donor oligonucleotides (ssODN; sequence: TTAGCTCTGTTTACGTCCCAGCGGGCATGAGAGTAACAAGAGGGTGTGGTAATATTACGGT ACCGAGCACTATCGATACAATATGTGTCATACGGACACG, 100uM stock) was added to the complex at a 1:1 molar ratio and incubated at 37°C for 5 minutes. Finally, Cas9 protein (MacroLab, Berkeley, 40 μM stock) was added at a 1:2 molar ratio and incubated at 37°C for 15 minutes. The resulting RNPs were frozen at -80°C until the day of electroporation. 48 hours following effector T cell activation, the cells were pelleted at 100x g for 10 minutes and resuspended in room temperature Lonza P3 buffer (Lonza, catalog no. V4XP-3032) at $1.5x10^6$ cells per 20 ul P3. The cells were combined with 5 ul aliquots of the thawed RNPs, transferred to a 96-well electroporation cuvette plate (Lonza, #VVPA-1002) and nucleofected with pulse code EH-115. Immediately following electroporation, the cells were gently resuspended in 90 ul warmed complete RPMI with IL-2 and incubated at 37 C for 15 minutes. After recovery, the cells were cultured in 96 well plates at $1x10^6$ cells/mL for the duration of the experiment. To prevent edge effects, the guides were randomly distributed across each plate and the first and last column of each plate was excluded, being filled instead with PBS to prevent evaporation.

RNA isolation and library preparation:

8 days after T cell isolation and activation, the cells were pelleted and resuspended at $1x10^6$ cells per 300 ul of RNA lysis buffer (Zymo, #R1060-1-100). Cells were pipette mixed and frozen at -80 until RNA isolation was performed. RNA was isolated using the Zymo-Quick RNA micro prep kit (#R1051) according to the manufacturer's protocol with the following modifications: After thawing the samples, each tube was vortexed vigorously to ensure complete lysis prior to loading into the extraction columns. In lieu of the kit provided DNAse, RNA was eluted from the isolation column after the recommended washes and digested with Turbo-DNAse (Fisher Scientific, AM2238) at 37 C for 20 minutes. Following digestion, RNA was purified using the RNA Clean & Concentrator-5 kit (Zymo, #R1016) according to the manufacturer's protocol. The resulting purified RNA was submitted to the UC Davis DNA Technologies and Expression Analysis Core to generate 3′ Tag-seq libraries with unique molecular indices (UMIs). Barcoded sequencing libraries were prepared using the QuantSeq FWD kit (Lexogen) for multiplexed sequencing on an Hiseq 4000 (Illumina).

Genotyping of arrayed KOs

On the day of cell collection for RNAseq, genomic DNA was isolated using DNA QuickExtract (Lucigen, Cat #QE09050) according to the manufacturer's protocol. Primers were designed to flank each sgRNA cut site. Amplicons of the region were generated by adding 1.25 µL each of forward and reverse primer at 10uM to 5 µL of sample in QuickExtract, 12.5 µL of NEBNext Ultra II Q5 master mix (NEB, Cat #M0544L), and H2O to a total 25 µL reaction volume. Touchdown PCR was used with the following cycling conditions: 98°C for 3 minutes, 15 cycles of 94°C for 20 seconds followed by 65°C-57.5°C for 20 seconds (0.5°C incremental decreases per cycle), and 72°C for 1 minute, and a subsequent 20 cycles at 94°C for 20 seconds, 58°C for

20 seconds and 72°C for 1 minute, and a final 10 minute extension at 72°C. Amplicons were diluted 1:200 and Illumina sequencing adapters were then added in a second PCR reaction. Indexing reactions included 1 µL of the diluted PCR1 sample, 2.5 µL of each the forward and reverse Illumina TruSeq indexing primers at 10 µM each, 12.5 µL of NEB Q5 master mix, and H2O to a total 25 µL reaction volume. The following PCR cycling conditions were used: 98°C for 30 seconds, followed by 98°C for 10 seconds, 60°C for 30 seconds, and 72°C for 30 seconds for 12 cycles, and a final extension period at 72°C for 2 minutes. Samples were pooled at an equivolume ratio and SPRI purified prior to sequencing on an Illumina MiSeq with PE 150 reads. Analysis was performed with CRISPResso2[42] (v2.2.7) CRISPRessoBatch --skip_failed --n_processes 4 --exclude_bp_from_left 0 --exclude_bp_from_right 0 --plot_window_size 10.

Cell proliferation quantification

One replica plate of cells from each donor was run on the Attune NxT Flow Cytometer (Thermo Fisher) within 24 hours of cell lysis for RNA extraction. The lymphocyte count was collected for each well using an equi-volume amount of sample. Counts were normalized to the mean AAVS1 lymphocyte count for the respective donor and experiment. Samples with a total Lymphocyte_count < 1000 were excluded from the analysis, removing one donor sample from SP110, EPAS1, and ZBTB14 KOs.

RNA-seq alignment and gene count quantification

Adapters were trimmed from fastq files with cutadapt[43]. Low-quality bases from reads were trimmed using the Phred algorithm implemented in seqtk[44]. Reads were then aligned with STAR[45] and mapped to GRCh38. Gene counts from deduplicated reads were quantified using featureCounts[46]. Sample quality control reports were generated with Fastqc[47], rseqc[48], and Multiqc[49].

Gene filtering and PCA analysis

Genes were first filtered to those with at least 10 counts in at least five samples. PCA was then performed on the variance stabilizing transformed[29] (vst) counts of the 500 most variable genes. Three outlier samples were excluded and then the above process was repeated. The PCs were then assessed for association with batch effects and very broad cellular pathways. PCs 1-2 associated with batch effects, and PCs 3-4 were associated with cell cycle state, suggesting that PCs 1-4 should be included as covariates or otherwise adjusted for in downstream analysis.

Differential expression analysis

Differential expression analysis was performed using DESeq2[29], including donor identity, PCs 1-4, and the KO as predictors of the response. Donor identity and PCs 1-4 were included as covariates to mitigate their confounding effects on gene expression. We emphasize that the statistical estimand in this analysis the total effect of the perturbation of a given gene on the readout gene. This effect may include several indirect paths between the perturbed gene and the readout gene.

We used mashr[30] to perform shrinkage of the effect sizes of the differentially expressed genes. This yielded an estimate of the local false-sign rate (LFSR), which is the posterior probability that the true effect has a different sign (positive or negative) than the sign that is most compatible with the posterior distribution. We used a threshold of LFSR < 5×10^{-3} as a significance threshold.

Pathway analysis

Downstream enriched pathways were identified for each perturbation using pathfindR (v1.6.4)[50]. For each upstream gene perturbed, outgoing edges within the BG model were used as input for pathfinder, with a significance threshold of LFSR < 5 x 10^{-3}. Gene sets were limited to KEGG[33], Reactome[38], and GO-BP[51] and the minimum gene set size and enrichment threshold were set to 10 and 0.05 respectively. Pathways were prioritized for visualization based on the number of genes within the module with enrichment for the pathway, median fold enrichment across all members of the module, and relevance to T cell biology.

ATAC and ChIPseq data visualization

Bigwigs for each of the tracks were downloaded from ChIP-Atlas. ATAC bigwigs and differentially expressed regions were procured from Freimer et al. and a representative donor was used for visualization of each perturbation effect at the IL17F locus. Visualization was performed with rtracklayer (v1.52.1) and ggplot2 (v3.4.1). APRIS gene structure was used for gene annotation with gggenes (v0.5.0). Differentially accessible chromatin regions were defined in Freimer at al. [14], Supplementary Data 2.

Bigwig files were obtained for visualization from the following ChIP-Atlas sources: STAT5B KO ATAC- SRX10558086, KMT2A KO ATAC- SRX10558079, AAVS1 KO ATAC- SRX10558063 (all ATAC samples from CD4+ T cells treated with IL-2), H3K4me3 ChIP- activated Th17 ChIP[52] (stimulated with anti-CD3/CD28 beads and IL-2)- SRX16500373 (GSM6376841), STAT5B ChIP[53] (treated with IL-2)- SRX041293 (GSM671402)

LLCB and Bipartite graph model construction

The statistical details and theory of the models used in this work are detailed in Weinstock et al[54].

References

1. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science (New York, N.Y.) *337*, 1190. https://doi.org/10.1126/science.1222794.

2. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet *47*, 1228–1235. https://doi.org/10.1038/ng.3404.

3. Aguet, F., Brown, A.A., Castel, S.E., Davis, J.R., He, Y., Jo, B., Mohammadi, P., Park, Y., Parsana, P., Segrè, A.V., et al. (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213. https://doi.org/10.1038/nature24277.

4. Consortium, T.Gte. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.

5. Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., et al. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet *53*, 1300–1310. https://doi.org/10.1038/s41588-021-00913-z.

6. Yazar, S., Alquicira-Hernandez, J., Wing, K., Senabouth, A., Gordon, M.G., Andersen, S., Lu, Q., Rowson, A., Taylor, T.R.P., Clarke, L., et al. Single-cell eQTL mapping identifies cell type–specific genetic control of autoimmune disease. Science *376*, eabf3041. https://doi.org/10.1126/science.abf3041.

7. Liu, X., Li, Y.I., and Pritchard, J.K. (2019). Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell *177*, 1022-1034.e6. https://doi.org/10.1016/j.cell.2019.04.014.

8. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell *169*, 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038.

9. Price, A.L., Helgason, A., Thorleifsson, G., McCarroll, S.A., Kong, A., and Stefansson, K. (2011). Single-Tissue and Cross-Tissue Heritability of Gene Expression Via Identity-by-Descent in Related or Unrelated Individuals. PLOS Genetics *7*, e1001317. https://doi.org/10.1371/journal.pgen.1001317.

10. Connally, N.J., Nazeen, S., Lee, D., Shi, H., Stamatoyannopoulos, J., Chun, S., Cotsapas, C., Cassa, C.A., and Sunyaev, S.R. (2022). The missing link between genetic association and regulatory function. eLife *11*, e74970. https://doi.org/10.7554/eLife.74970.

11. Elorbany, R., Popp, J.M., Rhodes, K., Strober, B.J., Barr, K., Qi, G., Gilad, Y., and Battle, A. (2022). Single-cell sequencing reveals lineage-specific dynamic genetic regulation of gene expression during human cardiomyocyte differentiation. PLOS Genetics *18*, e1009666. https://doi.org/10.1371/journal.pgen.1009666.

12. Strober, B.J., Elorbany, R., Rhodes, K., Krishnan, N., Tayeb, K., Battle, A., and Gilad, Y. (2019). Dynamic genetic regulation of gene expression during cellular differentiation. Science *364*, 1287–1290. https://doi.org/10.1126/science.aaw0040.

13. Nathan, A., Asgari, S., Ishigaki, K., Valencia, C., Amariuta, T., Luo, Y., Beynor, J.I., Baglaenko, Y., Suliman, S., Price, A.L., et al. (2022). Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. Nature, 1–9. https://doi.org/10.1038/s41586-022-04713-1.

14. Freimer, J.W., Shaked, O., Naqvi, S., Sinnott-Armstrong, N., Kathiria, A., Garrido, C.M., Chen, A.F., Cortez, J.T., Greenleaf, W.J., Pritchard, J.K., et al. (2022). Systematic discovery and perturbation of regulatory genes in human T cells reveals the architecture of immune networks. Nat Genet, 1–12. https://doi.org/10.1038/s41588-022-01106-y.

15. Mowery, C.T., Freimer, J.W., Chen, Z., Casaní-Galdón, S., Umhoefer, J.M., Arce, M.M., Gjoni, K., Daniel, B., Sandor, K., Gowen, B.G., et al. (2024). Systematic decoding of cis gene regulation defines context-dependent control of the multi-gene costimulatory receptor locus in human T cells. Nat Genet, 1–12. https://doi.org/10.1038/s41588-024-01743-5.

16. Mostafavi, H., Spence, J.P., Naqvi, S., and Pritchard, J.K. (2022). Limited overlap of eQTLs and GWAS hits due to systematic differences in discovery. Preprint at bioRxiv, https://doi.org/10.1101/2022.05.07.491045 https://doi.org/10.1101/2022.05.07.491045.

17. Bousfiha, A., Moundir, A., Tangye, S.G., Picard, C., Jeddane, L., Al-Herz, W., Rundles, C.C., Franco, J.L., Holland, S.M., Klein, C., et al. (2022). The 2022 Update of IUIS Phenotypical Classification for Human Inborn Errors of Immunity. J Clin Immunol. https://doi.org/10.1007/s10875-022-01352-z.

18. Marrack, P., Kappler, J., and Kotzin, B.L. (2001). Autoimmune disease: why and where it occurs. Nat Med *7*, 899–905. https://doi.org/10.1038/90935.

19. Attfield, K.E., Jensen, L.T., Kaufmann, M., Friese, M.A., and Fugger, L. (2022). The immunology of multiple sclerosis. Nat Rev Immunol, 1–17. https://doi.org/10.1038/s41577-022-00718-z.

20. Sun, L., Su, Y., Jiao, A., Wang, X., and Zhang, B. (2023). T cells in health and disease. Sig Transduct Target Ther *8*, 1–50. https://doi.org/10.1038/s41392-023-01471-y.

21. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.

22. Zheng, X., Aragam, B., Ravikumar, P., and Xing, E.P. (2018). DAGs with NO TEARS: Continuous Optimization for Structure Learning. arXiv:1803.01422 [cs, stat].

23. Hyttinen, A., Eberhardt, F., and Hoyer, P.O. (2012). Learning Linear Cyclic Causal Models with Latent Variables. Journal of Machine Learning Research *13*, 3387–3439.

24. Friedman, N., and Koller, D. (2003). Being Bayesian About Network Structure. A Bayesian Approach to Structure Discovery in Bayesian Networks. Machine Learning *50*, 95–125. https://doi.org/10.1023/A:1020249912095.

25. Battle, A., Jonikas, M.C., Walter, P., Weissman, J.S., and Koller, D. (2010). Automated identification of pathways from quantitative genetic interaction data. Molecular Systems Biology *6*, 379. https://doi.org/10.1038/msb.2010.27.

26. Agrawal, R., Uhler, C., and Broderick, T. (2018). Minimal I-MAP MCMC for Scalable Structure Discovery in Causal DAG Models. In Proceedings of the 35th International Conference on Machine Learning (PMLR), pp. 89–98.

27. Harris, S.L., and Levine, A.J. (2005). The p53 pathway: positive and negative feedback loops. Oncogene *24*, 2899–2908. https://doi.org/10.1038/sj.onc.1208615.

28. Parsana, P., Ruberman, C., Jaffe, A.E., Schatz, M.C., Battle, A., and Leek, J.T. (2019). Addressing confounding artifacts in reconstruction of gene co-expression networks. Genome Biol *20*, 94. https://doi.org/10.1186/s13059-019-1700-9.

29. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

30. Urbut, S.M., Wang, G., Carbonetto, P., and Stephens, M. (2019). Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. Nat Genet *51*, 187–195. https://doi.org/10.1038/s41588-018-0268-8.

31. Soutourina, J. (2018). Transcription regulation by the Mediator complex. Nat Rev Mol Cell Biol *19*, 262–274. https://doi.org/10.1038/nrm.2017.115.

32. Richter, W.F., Nayak, S., Iwasa, J., and Taatjes, D.J. (2022). The Mediator complex as a master regulator of transcription by RNA polymerase II. Nat Rev Mol Cell Biol, 1–18. https://doi.org/10.1038/s41580-022-00498-3.

33. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., and Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res *51*, D587–D592. https://doi.org/10.1093/nar/gkac963.

34. Zhu, J., Yamane, H., and Paul, W.E. (2010). Differentiation of Effector CD4 T Cell Populations. Annu Rev Immunol *28*, 445–489. https://doi.org/10.1146/annurev-immunol-030409-101212.

35. Oh, H., and Ghosh, S. (2013). NF-κB: Roles and Regulation In Different CD4+ T cell subsets. Immunol Rev *252*, 41–51. https://doi.org/10.1111/imr.12033.

36. Wang, L., Wildt, K.F., Castro, E., Xiong, Y., Feigenbaum, L., Tessarollo, L., and Bosselut, R. (2008). The zinc finger transcription factor Zbtb7b represses CD8-lineage gene expression

in peripheral CD4+ T cells. Immunity *29*, 876–887. https://doi.org/10.1016/j.immuni.2008.09.019.

37. Baffa, R., Negrini, M., Schichman, S.A., Huebner, K., and Croce, C.M. (1995). Involvement of the ALL-1 gene in a solid tumor. Proc Natl Acad Sci U S A *92*, 4922–4926.

38. Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., et al. (2022). The reactome pathway knowledgebase 2022. Nucleic Acids Research *50*, D687–D692. https://doi.org/10.1093/nar/gkab1028.

39. Jain, N., Zhao, Z., Feucht, J., Koche, R., Iyer, A., Dobrin, A., Mansilla-Soto, J., Yang, J., Zhan, Y., Lopez, M., et al. (2023). TET2 guards against unchecked BATF3-induced CAR T cell expansion. Nature *615*, 315–322. https://doi.org/10.1038/s41586-022-05692-z.

40. Benucci, M., Bernardini, P., Coccia, C., De Luca, R., Levani, J., Economou, A., Damiani, A., Russo, E., Amedei, A., Guiducci, S., et al. (2023). JAK inhibitors and autoimmune rheumatic diseases. Autoimmunity Reviews *22*, 103276. https://doi.org/10.1016/j.autrev.2023.103276.

41. Kotyla, P., Gumkowska-Sroka, O., Wnuk, B., and Kotyla, K. (2022). Jak Inhibitors for Treatment of Autoimmune Diseases: Lessons from Systemic Sclerosis and Systemic Lupus Erythematosus. Pharmaceuticals *15*, 936. https://doi.org/10.3390/ph15080936.

42. Clement, K. *et al.* CRISPResso2 provides accurate and rapof genome editing analysis. *Nat Biotechnol* **37**, 220–224 (2019).64.

43. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.

44. Li, H. (2023). lh3/seqtk.

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

46. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

47. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

48. Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184–2185. https://doi.org/10.1093/bioinformatics/bts356.

49. Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354.

50. Ulgen, E., Ozisik, O., and Sezerman, O.U. (2019). pathfindR: An R Package for Comprehensive Identification of Enriched Pathways in Omics Data Through Active Subnetworks. Frontiers in Genetics *10*.

51. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat Genet *25*, 25–29. https://doi.org/10.1038/75556.

52. An Integrated Encyclopedia of DNA Elements in the Human Genome (2012). Nature *489*, 57–74. https://doi.org/10.1038/nature11247.

53. Liao, W., Lin, J.-X., Wang, L., Li, P., and Leonard, W.J. (2011). Modulation of cytokine receptors by IL-2 broadly regulates differentiation into helper T cell lineages. Nat Immunol *12*, 551–559. https://doi.org/10.1038/ni.2030.

54. Weinstock J.S., Arce M.M., Freimer J.W., Ota M., Marson A., Battle A., Pritchard JK. (2024) Gene regulatory network inference from CRISPR perturbations in primary CD4+ T cells elucidates the genomic basis of immune disease. Cell Genomics https://doi.org/10.1016/j.xgen.2024.100671

Chapter 4 Conclusions

The projects described in **chapter 2** and **chapter 3** of this work were performed in an effort to better understand the gene regulatory mechanisms that enable CD4+ T cell function. Although these projects had differing approaches, shared themes emerged from the analyses performed. Gene regulatory networks in both chapters centralized around one gene, MED12, which consistently had strong and broad effects on gene expression across contexts. In **chapter 2** we characterized MED12 as a regulator of both T cell rest and activation, clarifying a critical function for the protein within the T cell compartment. While less functional analysis was performed for individual regulators in **chapter 3**, we did observe a striking lack of regulators upstream of MED12, despite the inclusion of 84 genes within the broader network. This observation suggests that MED12 is in fact at the top of an extensive hierarchy of important immune genes, including regulators of IL2RA, but also several inborn error of immunity genes and other factors associated with immune disease. These conclusions then lead to questions about upstream regulators of MED12 itself, which we did not identify in this work, and more broadly the role of other Mediator complex subunits in the immune system which we were not able to assess with as great detail.

The role of epigenetic regulation in immune signaling also emerged as a motif across projects. While neither the screens or arrayed perturbations performed in chapter 3 were designed to specifically study this subset of trans regulatory proteins, chromatin and histone modifiers emerged naturally as the stories evolved. In **chapter 2**, the discovery that MED12 interacts to some degree with members of SETD1A/COMPASS, an H3K4me1-3 depositing enzyme with additional functions in protein ubiquitylation, reformed our understanding of the role of Mediator proteins in chromatin level regulation. In **chapter 3**, we observed shared downstream networks between KMT2A, another histone methylating enzyme, and canonical members of the JAK-

STAT pathway, resulting in a novel connection between KMT2A and CD4+ T cell subset specific cytokine regulation. These occurrences serve as a reminder that trans regulation cannot be fully understood without considering each layer of regulation. These layers include transcription factors with DNA sequence specificity, but also cofactors like MED12 with ubiquitous expression, and epigenetic regulators of the chromatin state. With these discoveries in hindsight, future screens to study trans-regulation of gene expression might be designed with more inclusivity to members of each of these regulatory compartments that are required for coordinated gene expression.

Ultimately, the work performed in **chapters 2** and **3** has led to improved understanding of gene regulation in the immune system. It has also opened up new questions about the mechanisms that control dynamic gene expression within the T cell compartment. While we made progress in understanding the role MED12 in T cell specific gene expression, parts of its mechanism of action remain unclear as we observed mixed shared effects with other components of Mediator, the Mediator kinase, and other regulatory proteins that interact with MED12. Further investigation of the role of MED12 and its interaction partners both in T cells and other cell types will lead to a clearer understanding of this system of dynamic gene regulation. In particular, recent advances in CRISPR technology that enable the investigation of base level mutations as well as epigenetic modifications, may enable more finite dissection of the role of cofactors like the Mediator proteins and associated chromatin regulators.

Finally, a substantial amount of sequencing and proteomics data was generated as part of the projects described here which has since been deposited to public repositories. These cell type and state specific data sets will hopefully serve as a useful resource for future discovery as perturbation data across conditions and rare cell subsets such as Tregs remains difficult to access and costly to generate.

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:
Mayalree
AC1D519AEA43407...
Author Signature

2/28/2025
Date