Title
Understanding human autoimmunity risk within the IL2RA super enhancer

Author
Simeonov, Dimitre

Publication Date
2019

Understanding human autoimmunity risk within the IL2RA super enhancer

by
Dimitre R. Simeonov

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:
Mark Anderson
657D61B3C699406...
Mark Anderson
Chair

DocuSigned by:
Jeffrey Bluestone
Jeffrey Bluestone

DocuSigned by:
Chun Ye
Chun Ye

DocuSigned by:
Alexander Marson
7F25CDCE383C4A8...
Alexander Marson

Committee Members

For my family and friends

ACKNOWLEDGEMENTS

My early childhood years were spent on a farm in Bulgaria. I never imagined that one day I would earn a PhD in the technology capital of the world. There are many people in my life who were instrumental to this journey. Know that I am forever grateful for your influence. Here I would like to thank all of the people who were involved in my scientific training during the PhD program as well as in the time leading up to it.

First and foremost I would like to thank my mentor Dr. Alexander Marson. Alex embodies all of the qualities of a great scientific mentor – a brilliant scientific mind with infectious positivity, and a natural ability to work with others. I feel fortunate to have worked with Alex for so many years. His dedication to teaching and mentoring greatly accelerated my growth as a scientist, and provided me with the tools and skills to be independent and successful in science. Even now, as I prepare for another chapter of my life, Alex continues to provide valuable advice and mentorship.

I would like to thank my thesis committee, Drs. Mark Anderson, Jimmie Ye, and Jeffrey Bluestone. All three are super-stars in their fields of study and were invaluable during my PhD, providing support, inspiration, and motivation when needed. They were especially helpful in filling gaps of my immunology knowledge and always led me to think through the most exciting line of experimentation. I would also like to acknowledge Abul Abbas, who taught me the basics of immunology in under two weeks of lectures. I want to thank Dr. Nadav Ahituv for advising me early on in graduate school and getting me excited about non-coding DNA sequences. Finally, I would like to thank all of the great collaborators and researchers I have the pleasure of working with these last couple of years. In particular, I would like to mention Benjamin Gowen and Jacob Corn, wonderful collaborators and scientists with whom I published my first big paper.

I have been fortunate to work with great mentors who have left a lasting positive impact on my life. I want to thank Dr. Janet Paluh for taking a chance on me as an undergraduate and for

her mentorship, which ultimately led to my venture into science. A big thank you to Dr. William Gahl, who showed me the real life impact that science can have on people's lives. I must also thank Bill for many of the jokes I picked up from him over the years. I want to thank Drs. Thomas Markello and David R. Adams whom I worked closely with at the NIH for almost 3 years. Your training was invaluable in graduate school.

I want to thank all of my wonderful friends who have had to deal with my tardiness over the years because I am permanently on research time. Evan Markegard, James Salazar, Manabu Manandhar, Dylan Lowe, David Nguyen, Jessica Cortez, John Gagnon, and Zachary Lee - thank you for so many great memories and adventures over the years. I would also like to thank my significant other, Irene Li, for her support and continued love during the PhD, as well as teaching me that deadlines are negotiable.

To my family, thank for you for always being there for me. My grandfather, Mincho Karatchorov, you instilled in me from a young age the value of hard work. My sister, Teddy Simeonova, for always being supportive and staying calm, even when I was not. Finally, I would like to thank my mother, Milena Simeonova, who as a single mother raised two children in a foreign country, created opportunities out of nothing, and always pushed us to achieve. Words cannot express my gratitude for your sacrifices. Love you mom!

CONTRIBUTIONS

The research presented in this thesis was conducted under the supervision of Dr. Alexander Marson at the University of California, San Francisco. This work was supported by the Jeffrey G. Klein Family Diabetes Fellowship.

Chapter 1 is an introduction to CRISPR-based tools in immunity, published in *Annual Review of Immunology* in 2019 with the following authors: Dimitre R. Simeonov and Alexander Marson.

Chapter 2 describes work published in *Nature* in 2017 with the following authors: Dimitre R. Simeonov*, Benjamin G. Gowen*, Mandy Boontanrart, Theodore L. Roth, John D. Gagnon, Maxwell R. Mumbach, Ansuman T. Satpathy, Youjin Lee, Alice Y. Chan, Dmytro S. Lituiev, Michelle L. Nguyen, Rachel E. Gate, Meena Subramaniam, Zhongmei Li, Jonathan M. Woo, Therese Mitros, Graham J. Ray, Nicolas L. Bray, Gemma L. Curie, Nicki Naddaf, Julia S. Chu, Hong Ma, Eric Boyer, Frederic Van Gool, Kathrin Schumann, Mark J. Daly, Kyle K. Farh, K. Mark Ansel, Chun J. Ye, William J. Greenleaf, Mark S. Anderson, Jeffrey A. Bluestone, Howard Y. Chang, Jacob E. Corn†, Alexander Marson†. D.R.S., B.G.G, J.E.C. and A.M. designed the study and wrote the manuscript. B.G.G., M.B., N.L.B., T.M., G.J.R. and G.L.C. performed and analyzed CRISPRa screens. B.G.G., D.R.S., N.N., J.S.C. and H.M. performed luciferase reporter cloning and experiments. D.R.S., T.L.R., J.D.G., Y.L., A.C., M.L.N., Z.L., J.M.W, E.B., F.V.G, V.R.T, R.E.G., M.S. and K.S. contributed to functional experiments on CaRE4 and rs61839660. M.R.M., A.T.S., W.J.G. and H.Y.C. generated and analyzed Hi-ChIP data. D.S.L. and C.Y. performed ImmVar QTL analysis. H.H., R.L., K.K.F. and M.J.D. contributed to fine-mapping analysis of rs61839660 disease association. D.R.S, J.D.G. and K.M.A. contributed to T cell differentiation. M.S. and J.A.B. advised on functional studies in murine models. D.R.S. and B.G.G. are joint first authors. J.E.C. and A.M. are co-corresponding and co-senior authors.

Chapter 3 describes work published in Communications Biology in 2019 with the following authors: Dimitre R. Simeonov*, Alexander J. Brandt*, Alice Y. Chan, Jessica T. Cortez, Zhongmei Li, Jonathan M. Woo, Youjin Lee, Claudia M.B. Carvalho, Alyssa C. Indart, Theodore L. Roth, James Zou, Andrew P. May, James R. Lupski, Mark S. Anderson, F. William Buaas, Daniel S. Rokhsar†, Alexander Marson†. D.R.S., A.J.B., D.S.R., and A.M. designed study and wrote manuscript. D.R.S., A.J.B., A.C., Z.L., J.T.C., Y.L., J.M.W., A.C.I., and F.W.B. performed mouse experiments. M.S.A. and T.L.R. advised on murine studies. C.M.B.C., J.Z., A.P.M., and J.R.L. advised on DNA repair mechanisms of the bystander mutation.

Chapter 4 describes ongoing unpublished work with contributions from: Dimitre R. Simeonov, Harikesh Wong, Jessica T. Cortez, Jennifer Umhoefer, Alice Y. Chan, Alyssa C. Indart, Zhongmei Li, Barbara Celona, Jonathan M. Woo, Brian Black, Jeffrey A. Bluestone, Mark S. Anderson, Ronald Germain, and Alexander Marson. D.R.S. performed the bulk of experiments with help from J.T.C., J.U., and A.I.. Technical assistance was provided by Z.L. and J.M.W. Imaging experiments were performed in collaboration with H.W. and R.G. Experimental advice on murine studies was provided by J.A.B. and M.S.A. EMSA studies were performed by D.R.S. with the help of B.C. and B.B. The manuscript was written by D.R.S. and A.M.

Finally, Chapter 5 includes concluding remarks, discussion, and future directions.

Understanding human autoimmunity risk within the *IL2RA* super enhancer

Dimitre R. Simeonov

ABSTRACT

Human disease risk has been linked to hundreds of variants in our DNA. Understanding these unbiased genetic associations has long held the promise of revealing insights into disease mechanisms. However, efforts to study disease-associations have largely been frustrated as disease-associated variants overwhelmingly reside in non-coding sequences – long stretches of our chromosomes that we still know relatively little about. Here I develop new tools and methodologies to understand human genetic risk for disease by dissecting autoimmunity risk within the *IL2RA* locus.

Genetic variation within the *IL2RA* super-enhancer has been linked to at least 8 different autoimmune disorders. CRISPR/Cas9 genome engineering has revolutionized our ability to functionally interrogate genomic sequences. In Chapter 1, I introduce CRISPR/Cas9 technology and discuss how it is being used to understand the genetic underpinnings of immunity. In Chapter 2, we develop a CRISPR-activation (CRISPRa)-based high-throughput screening platform for the identification of functional enhancers. By using tiled CRISPRa we systematically map functional *IL2RA* enhancers within autoimmunity-associated non-coding sequences. Using a combination of CRISPR-engineered mouse models, population genetics, and molecular biology, I also demonstrate that an autoimmunity variant, rs61839660, resides within a stimulation-responsive enhancer that controls the kinetics of *IL2RA* induction on activated T cells. Here we mapped the regulatory landscape of a critical immune regulator and decoded the function of a causal autoimmunity variant. This work developed tools to study human autoimmunity risk.

How distinct *IL2RA* enhancers regulate immune cell function as well as shape risk of autoimmunity remain unknown. In Chapters 3 and 4, I functionally characterize two distinct *IL2RA* enhancers; a disease-associated *IL2RA* enhancer that controls the timing of IL2RA induction in pro-inflammatory immune cells and a novel maintenance enhancer that controls IL2RA expression in anti-inflammatory regulatory T cells. Having discovered enhancers that regulate IL2RA in different contexts I interrogated their effects in an in vivo model of autoimmune disease. Deletion of the conserved stimulation-responsive enhancer that harbors a human variant protective against T1D completely protected non-obese diabetic (NOD) mice from diabetes. Remarkably, the human genetics predicted this result in an animal model. Finally, I show that these *IL2RA* enhancers have opposing immunological functions by tuning IL2RA receptor expression in a cell type restricted manner. My studies develop an understanding of *IL2RA* enhancer regulation in the immune system and functionally link gene regulation to autoimmunity.

In Chapter 5, I discuss conclusions and future directions of these studies. This work lays the foundation for understanding a critical autoimmunity locus and establishes generalizable strategies for decoding human disease risk.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. CRISPR-Based Tools to Study Immunity

ABSTRACT

CRISPR technology has opened a new era of genome interrogation and genome engineering. Discovered in bacteria, where it protects against bacteriophage by cleaving foreign nucleic acid sequences, the CRISPR system has been repurposed as an adaptable tool for genome editing and multiple other applications. CRISPR's ease of use, precision, and versatility have led to its widespread adoption, accelerating biomedical research and discovery in human cells and model organisms. Here we review CRISPR-based tools and discuss how they are being applied to decode the genetic circuits that control immune function in health and disease. Genetic variation in immune cells can affect autoimmune disease risk, infectious disease pathogenesis, and cancer immunotherapies. CRISPR provides unprecedented opportunities for functional mechanistic studies of coding and noncoding genome sequence function in immunity. Finally, we discuss the potential of CRISPR technology to engineer synthetic cellular immunotherapies for a wide range of human diseases.

FUNCTIONAL GENETIC STUDIES OF IMMUNITY

The completion of the human genome reference sequence in the early 2000s marked a turning point for immunological research. The Human Genome Project identified the ~3.2 billion bases of our DNA, but we lacked understanding of their functions. Immune cells—which are critical for human health and can be studied ex vivo and in vivo in established animal models—have been a major focus of genomic exploration. Over the past two decades, diverse immune cells have been subjected to chromatin state and transcriptional profiling to map DNA elements and the genetic circuitry underlying immune cell types, states, and functions. However, key questions surrounding

DNA and immunity can only be answered through genetic perturbation. What are the functional sequences in our DNA and what is their biological importance? What is the genetic circuitry—encoded in genes, noncoding sequences, and trans-regulators—that wires specific cellular pathways and specialized functions in immune cells? How does variation in critical coding and noncoding sequences alter cellular function and contribute to risk of immune-mediated disease? Can we utilize our understanding of natural immune cell genetic circuits well enough to reprogram them for the next generation of engineered cellular therapies? The answers depend on new technologies to rewrite genomes in immune cells. CRISPR is one such technology, and with it immunologists are beginning to manipulate immune cell genomes to reveal the genetic underpinnings of immunity.

Here we review the tool set that has emerged rapidly for CRISPR-based genome engineering. CRISPR is a flexible system for targeted genome modifications. It has been used to knock out gene function or knock-in new genetic sequences in cell lines, primary human cells, and animal models. Beyond this, CRISPR has been adapted as a modular system to recruit diverse effector functions to specific sites in the genome in a programmable manner. CRISPR-based tools for transcriptional modulation, epigenetic modification, chromatin imaging and biochemistry, and targeted base-editing are introduced. Finally, we discuss how CRISPR targeting and large-scale CRISPR-based forward genetic screens are being deployed to reveal how immune cells are wired, how their circuits fail in disease states, and how they might be reprogrammed for new treatments.

CRISPR TOOL KIT

Gene Editing

The ability to induce double-stranded DNA breaks at specific sites in the genome of a cell can enable targeted genome modifications. Pioneering work showed that exogenous DNA could be

incorporated into a cell's genome through a process called homologous recombination[1]. However, the spontaneous efficiency of this process was low and the site of DNA incorporation was not precise. Induction of a double-stranded break (DSB) in genomic DNA could catalyze site-specific repair mechanisms and promote homology-directed repair (HDR) at the target site[2,3]. These findings sparked the beginnings of genome engineering, which began with restriction enzymes in yeast and moved to nucleases with longer recognition specificities, like meganucleases[4,5], and engineered DNA-specificities including zinc fingers[6,7], and transcription activator–like effector nucleases (TALENs) [4-6,8]. Due to the complexity of having to reengineer nuclease protein sequences to target different sites in the genome, the use of these tools remained largely restricted to labs and companies with specialized expertise. CRISPR has overcome this limitation by utilizing a highly predictable, RNA-programmable system. The ease of use and versatility of CRISPR has transformed genome engineering into a widely accessible and adaptable laboratory tool. The term CRISPR, or clustered regularly interspersed short palindromic repeats, originates in observations as far back as the 1980s that some bacteria harbored short repetitive DNA sequences in their genomes that surrounded short spacer sequences resembling viral DNA[9-14]. Decades later, we now understand that CRISPR evolved in some bacterial species as a DNA targeting system that cleaves foreign genomes[15-19]. Advances in our understanding of the basic mechanisms of this bacterial system enabled its widespread adoption for genome engineering. Different CRISPR systems continue to be identified today, but the best known and most widely used is the type II CRISPR system, in large part due to its simplicity. Whereas other CRISPR systems have multisubunit effector complexes that mediate nuclease activity, the type II CRISPR system uses a single DNA nuclease. Cas9 is the most widely known type II CRISPR nuclease and is the major focus of this review[18]. Cas9 is targeted to DNA sequences by a guide RNA (gRNA), which is made up of a trans-activating RNA (tracrRNA) and a CRISPR RNA (crRNA) in bacteria

(Figure 1.1a). The Cas9:gRNA complex scans DNA for sequences complementary to the crRNA that are appropriately spaced from a required protospacer adjacent motif (PAM) [20-23] (Figure 1.1a). Upon recognition, Cas9 cleaves the DNA to create a DSB between the third and fourth nucleotides upstream of the PAM site[18] (Figure 1.1a). By linking the crRNA and tracrRNA into a single guide RNA (sgRNA), Jinek et al. [24] reduced CRISPR into a two-component technology for DNA targeting. By varying RNA sequences in the crRNA region of the sgRNA, Cas9 could be reprogrammed to cut distinct DNA sequences[24]. This ability to introduce targeted DSBs at specific DNA sequences is fundamental for precise and efficient genome editing. Eukaryotic cells have evolved multiple mechanisms to repair DSBs, the most prominent being nonhomologous end joining (NHEJ) and HDR, which are differentially utilized and ultimately lead to different repair outcomes[25]. NHEJ is an error-prone mechanism that rejoins the two ends of a DSB with frequent small nucleotide insertions or deletions (indels). These errors in NHEJ repair can be exploited for gene ablation (knockout can be achieved by frameshift mutations) and sequence perturbation studies[26-28] (Figure 1.1b). In contrast, HDR relies on homologous DNA sequences to template repair of DSBs, which can be exploited to promote specific nucleotide sequence replacement. By adding exogenous DNA templates one can co-opt the cell's HDR pathway to deliver sequences at the site of a DSB[29] (Figure 1.1b). In 2013, the first applications of CRISPR for mammalian cell DNA editing were reported [26-28]. The ease with which the Cas9 nuclease could be reprogrammed to cut at different genomic sites by altering the gRNA made the system flexible to rapidly target sites throughout the genome. This also allowed for multiplexed editing by using multiple gRNAs to simultaneously target Cas9 to different parts of the genome[26]. For the first time biologists had a genome-engineering tool that could be deployed quickly and efficiently to edit diverse sequences in the genetic code. DNA editing with the CRISPR-Cas9 system is limited to sequences adjacent to PAM sites. The NGG PAM requirement for Streptococcus pyogenes Cas9 is not particularly

stringent, which has made this nuclease useful for most DNA-editing applications. However, for applications like therapeutic gene editing that require targeting of specific sequences the PAM requirement can be limiting. In addition, increasing the number of targetable sites can be useful for functional studies on noncoding sequences[30]. One approach to expanding the genome editing space of CRISPR has been to identify CRISPR systems from new microbial species that may have different PAM requirements[31-33]. An alternative approach has been to engineer Cas9 PAM specificities by structure-guided mutations [34-37] and directed evolution[38-40]. These efforts are expanding the CRISPR tool kit toward flexible targeting across the genome.

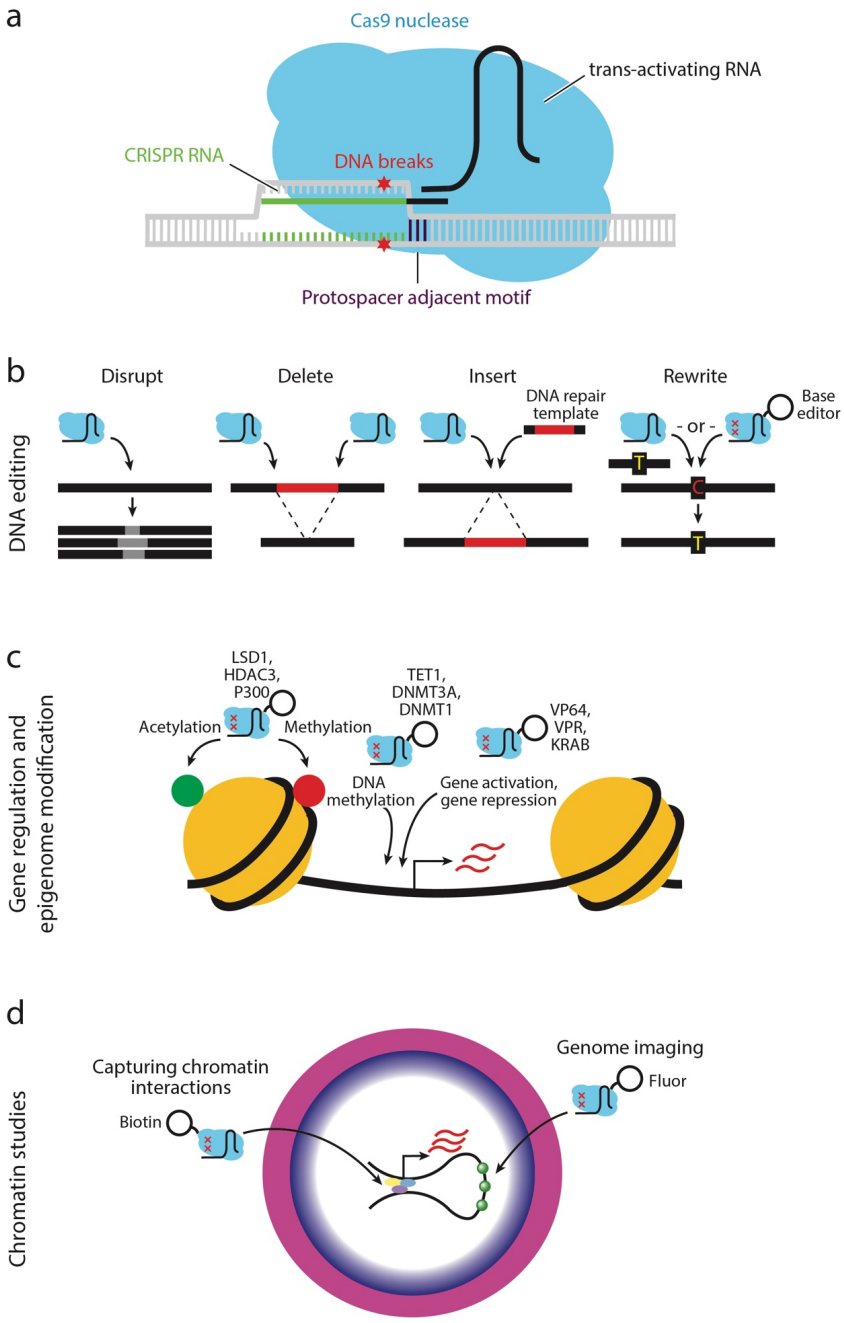


Figure 1.1. The CRISPR toolbox. (***a***) Schematic of CRISPR-Cas9 gene editing showing Cas9:gRNA complex at on-target genomic site inducing a double-stranded DNA break. CRISPR-based tools for (**b**) gene editing, (**c**) gene regulation and epigenome modification, and (**d**) chromatin studies.

Genetic Screens with CRISPR

Unbiased genetic screens have the potential to reveal unappreciated biological pathways and to identify new genetic circuits. CRISPR has facilitated large-scale genetic screens due to the ease, and relatively low cost, with which Cas9 can be reprogrammed to target different genomic sites simply by coupling the nuclease to varying gRNAs. CRISPR screens can be carried out in an arrayed or pooled fashion. Libraries of gRNAs can be used to generate pools of cells with CRISPR perturbations[41-43] (Figure 1.2a). Pooled screens to study the immune system have been performed in cell lines engineered to stably express Cas9[44,45] and primary immune cells derived from Cas9-expressing transgenic mice[46]. More recently, genome-scale pooled screens have also been performed in primary human cells[47,48]. Viral transduction of gRNA libraries with low multiplicity of infection ensures that the majority of transduced cells receive one gRNA, and therefore harbor a single genetic perturbation. In a large population of perturbed cells, cells with a phenotype of interest can be selected, and the causative perturbations can be mapped by sequencing the gRNAs in the selected population (Figure 1.2a). In this format, genomic integration of the gRNAs is necessary to link the phenotype to the perturbation caused by a particular gRNA. Pooled screening methods have enabled forward genetic screens at a genome-wide level. However, pooled screens generally are restricted to individual selectable phenotypes including cell survival/proliferation[41-43] or selectable protein markers[49].

In an arrayed format, different populations of cells are targeted with unique genetic perturbations[50] (Figure 1.2b). Arrayed screens generally are lower throughput than pooled approaches, but they allow for complex phenotypic readouts due to the homogeneity of the targeted population[50] (Figure 1.2b). Additionally, arrayed screens also are uniquely suited to study cell-cell interactions and assess cell non-autonomous effects of a genetic perturbation[50] (Figure 1.2b). Pooled and arrayed CRISPR screens serve complementary roles for functional genetic studies.

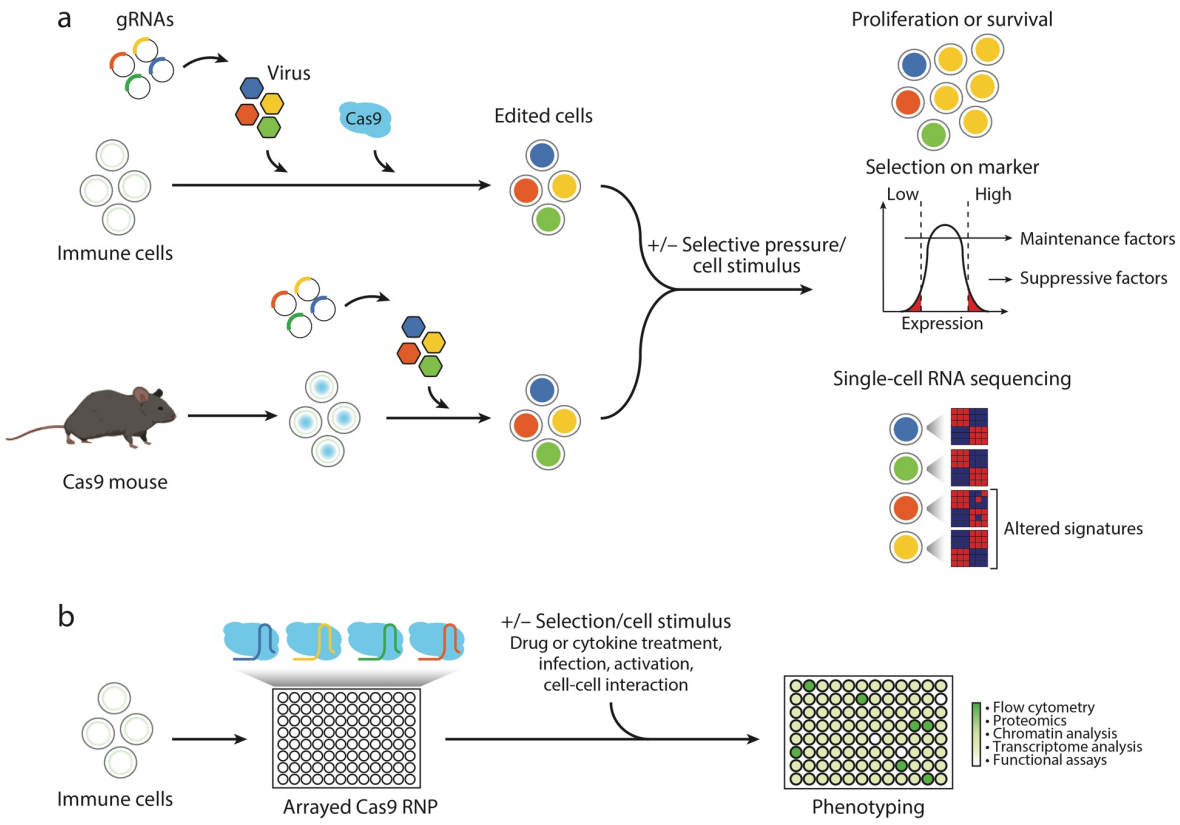


Figure 1.2. CRISPR genetic screens. (***a***) Pooled CRISPR screen workflow for immune cells. Generally, libraries of guide RNAs (gRNAs) are transduced into cells such that each cell, on average, receives a single gRNA that mediates a single genetic perturbation. Cas9 can be transduced or electroporated as protein to generate genetic perturbations. Genetic perturbations are made in a pool, and their effects on cellular proliferation and survival, protein expression, or other cell phenotypes can be assessed by deep sequencing gRNAs from cells with and without selection. These types of screens can be used to rapidly test large numbers of genetic perturbations. (***b***) Arrayed CRISPR screen workflow for immune cells. Single genetic perturbations are introduced to cells. Phenotypic effects can be measured for each genetic perturbation in edited cells. These screens are lower in throughput but allow for rich phenotypic readouts.

Specificity of Genome Editing

CRISPR-Cas9 gene editing can have unintended consequences in the genome. In screening approaches, off-target effects of individual guides can be handled by including multiple guides targeting each gene of interest. However, unintended edits are a significant concern for CRISPR-

based generation of animal models and human therapeutic applications. Early on it was recognized that Cas9 can cleave DNA sequences homologous to the on-target site despite complementarity mismatches[51]. The nucleotides immediately adjacent to the PAM, the seed sequence, are critical for targeting and generally do not tolerate mismatches. However, mismatches in the remainder of the gRNA can be tolerated and lead to DNA cleavage. Bioinformatic tools have been developed to minimize off-target effects and maximize on-target editing efficiency[51,52]. The algorithms serve as a general guide for gRNA design but do not account for all factors that govern success of CRISPR editing, including the local chromatin environment[53-55]. In addition, on-target DNA cleavage can cause unintended mutations in neighboring sequences due to DNA repair[56-58]. This has sparked efforts to engineer more specificity into the DNA-editing machinery and to develop tools to capture unintended mutations[58-63].

Several approaches have been developed to limit CRISPR-Cas9 off-target cleavage. Early efforts focused on limiting Cas9 nuclease activity and engineering Cas9 to require two neighboring targeting events to introduce a DSB[64,65]. Structure-guided studies have also had success in reducing off-target effects by mutating Cas9 residues that are not necessary for DNA binding or cleavage but interact with the DNA phosphate backbone[35-37]. A directed-evolution approach was successful in improving targeting specificity of Cas9[40]. Modifications to the gRNA length and sequence composition can also attenuate Cas9 off-target activity[66,67]. Finally, limiting Cas9 activity in targeted cells is emerging as an important factor. Several strategies have been used to do this, including delivery of Cas9 protein[66] or mRNA[68,69], either of which is degraded relatively rapidly; self-limiting circuits[70]; tunable systems[71-75]; and CRISPR inhibitors[76].

Targeted Genome Sequence Replacement

Targeted sequence replacement at endogenous genomic sites is a critical goal of genome engineering. Cas9:gRNA complexes are sufficient to disrupt genome sequences based on indels introduced during imperfect NHEJ repair or by excising DNA sequences when they are introduced in pairs (Figure 1.1b). However, cut-and-paste functionality for genome editing requires co-introduction of a DNA repair template (Figure 1.1b). DNA repair templates have been introduced into cell lines on plasmids that allow for long homology arms to the target sequence, which is important for HDR efficiency and specificity[6,77]. However, generating plasmids can be laborious, and delivery to primary cells can be inefficient and toxic. To overcome these technical barriers, short single-stranded DNA oligos (~200 bases) have been employed by multiple groups[78-80]. However, the short oligos limit the length of sequence replacement that is possible, especially when accounting for homology arms[81].

Many research and therapeutic goals depend on technologies to replace or insert larger sequences at endogenous sites. There have been concerted efforts to increase the size of the sequence payload that can be delivered while maintaining HDR efficiency and cell viability. One approach that has become widespread is combining Cas9, or other targeted nucleases, with adeno-associated virus (AAV) strains engineered to encode homology arms and the sequence to be introduced at the target site. The viral genome therefore serves as an HDR template and has been used to efficiently rewrite target sequences in human CD34+ hematopoietic stem cells[82,83], T cells[84-86], and B cells[87]. These methods are powerful, but they rely on viral production and transduction, which can be a bottleneck for both research and clinical applications. Recent efforts have revealed that Cas9 ribonucleoproteins (RNPs) can be coupled with long (>1 kilobase) DNA templates that are either double stranded (dsDNA) or single stranded (ssDNA) for HDR[88-90]. Efficient nonviral genome targeting can be achieved in primary human T cells by optimizing cell

culture conditions, Cas9 RNP concentrations, DNA template concentrations, and electroporation parameters[89]. This method is likely to be adaptable for other immune cell types as well. However, some cells—especially nondividing cells—may not be competent to undergo efficient HDR. Several groups have developed homology-independent targeted integration (HITI) [91] or microhomology-dependent precision integration into target chromosomes (PITCh) [92], which could help to expand the set of cells where targeted integrations can be achieved. Collectively, these technologies to knock-in large sequences will allow us to tag genes in their endogenous loci for biochemical and imaging studies. Furthermore, they will enable efforts to rewrite coding and noncoding sequences at specific sites in the genome to correct pathogenic mutations and reprogram immune cell functions.

Engineering Mouse Models

Mouse models are indispensable for functional genetic studies of the immune system. Gene targeting of embryonic stem cells by homologous recombination had been the method of choice for genome modifications[93]. However, this methodology is time consuming, expensive, and limited to certain genetic backgrounds for which established embryonic stem cell lines are available[94]. CRISPR-Cas9 overcomes these limitations and is now used routinely to engineer knockout and knock-in mice in a few months[95,96]. CRISPR modification of zygotes can be especially powerful for multiplex editing of multiple targets[96] or adding new modifications to existing mouse models that already carry multiple transgenic alleles. Gene targeting with CRISPR-Cas9 is possible on diverse murine backgrounds as long as gRNAs are properly designed based on the targeted genome. Recent work successfully introduced genetic modifications on the autoimmune prone non-obese diabetic genetic background, which is commonly used for human cell transfers and studies of autoimmune type 1 diabetes[97]. Cas9 nuclease can be microinjected

into single cell zygotes as DNA, RNA, or protein along with appropriate gRNAs. Exogenous DNA including plasmids or short, single-stranded oligonucleotides can be co-delivered for knock-ins by HDR at the Cas9 cut site[95,96]. Microinjection of the CRISPR machinery along with long ssDNA templates has proven useful for larger targeted modifications, including introduction of floxed alleles[98]. Electroporation of mouse zygotes with Cas9 RNP (CRISPR-EZ) eliminates the need for laborious zygote microinjections[99,100]. This method is efficient and simple and facilitates higher-throughput mouse generation. Although gene editing with CRISPR-EZ is highly efficient, further work is needed to enable HDR with larger repair templates. Taken together, these new tools are accelerating the production of engineered murine models.

CRISPR can also be used to engineer somatic murine cells as an alternative to germline editing. To facilitate CRISPR targeting of somatic cells and murine models for CRISPR screening, transgenic mice have been generated that express Cas9 constitutively[46], conditionally[101], or inducibly[102]. This facilitates genome editing in primary immune cells, where Cas9 delivery can be challenging. In Cas9 transgenic mice, cells can be modified simply by transducing sgRNA sequences individually or in pools. This has enabled studies of individual gene knockouts in somatic cells in addition to ex vivo and in vivo primary cell screens[46]. Finally, recent efforts have successfully used Cas9 RNP electroporation to directly edit primary immune cells isolated from mice[103]. These tools accelerate assessments of phenotypes arising from target perturbations in mature cells of the immune system.

Controlling Gene Expression

Beyond genome editing, CRISPR-Cas9 offers tremendous utility as a programmable scaffold to target effector molecules to DNA sequences. To transform Cas9 into a DNA-targeting scaffold, the nuclease domains were mutated to create a nuclease-deficient dead Cas9 (dCas9) that no longer

cuts DNA but could still target and bind to DNA sequences in a gRNA-programmable manner[104]. Targeting dCas9 to gene bodies could reduce gene expression through direct transcriptional interference with the RNA polymerase, without altering the genome sequence[104]. The utility of dCas9 was expanded further with the recognition that different effector molecules could be tethered to the inactivated enzyme to control gene expression. dCas9 tethered to a transcriptionally repressive domain (e.g., dCas9-KRAB) silenced gene expression when the molecule was targeted to gene promoters[105]. Alternatively, tethering a transcriptional activator (e.g., VP64) could increase gene expression[106-109]. CRISPR inhibition (CRISPRi) and activation (CRISPRa), as these systems are now known, have made it possible to toggle target gene expression in a controlled manner.

The effects of dCas9-VP64 on gene activation were often modest[108,110-112], but CRISPR-based control of transcriptional regulation continues to improve. Potent gene activation has been achieved by tiling multiple dCas9-VP64 molecules at a gene promoter[106,108] or by increasing the number of activation domains a single dCas9 can recruit[113]. Alternatively, tethering additional activation molecules to dCas9 could synergistically boost the effect of VP64, leading to more robust gene activation[114]. The gRNA sequence can also be engineered to recruit effector molecules[115]. Likewise, CRISPRi continues to be improved[116]. Improved ability to tune gene expression positively and negatively offers opportunities to test the effects of transcript levels on cell function and to perform large-scale complementary loss-of-function and gain-of-function screens[117]. Furthermore, the controlled expression of groups of genes by multiplexing gRNAs can be used for directing cellular differentiation and other genetic programs[118,119].

Epigenome Reprogramming

Epigenetic regulation of gene expression involves chromatin modifications that can be stably passed on to future generations of cells. Our understanding of epigenomic modifications has

increased with the advent of next-generation sequencing technologies that have profiled histone modifications and transcription factor binding by chromatin immunoprecipitation. A major undertaking in the field has been cataloging of epigenomic signatures across cell types and under different cell conditions, to understand the interplay among chromatin modifications, gene expression, and cell function. However, these data are correlative. A remaining fundamental question is which chromatin modifications have causal roles in gene expression and epigenetic memory. Engineered CRISPR systems are providing tools to recruit chromatin-modifying enzymes to specific genome sites and directly test the effects of epigenomic modifications.

RNA-programmable CRISPR systems have been to used recruit enzymes that modify histones or control DNA methylation to specific genomic loci. Fusing dCas9 to the catalytic domain of the DNA methyltransferase DMNT3A alone[120] or in combination with other factors[121] can specifically and efficiently methylate endogenous loci. Fusing dCas9 to TET1 can remove DNA methylation from silent loci, inducing expression of genes that were otherwise transcriptionally inactive[120,122-124]. These tools have been used to assess mechanistic consequences of site-specific DNA methylation. They are also potential therapeutic tools for diseases characterized by inappropriate methylation. Histone modifiers have also been engineered to allow precise control of locus-specific histone epigenetic marks. These include dCas9 fusions with histone demethylases (LSD1) [125] and methyltransferases (SMYD3, PRDM9 and DOT1L) [126,127], as well as deacetylases (HDAC3) [128] and acetyl transferases (p300) [129,130]. Chromatin-modifying versions of CRISPR are being used in pooled formats to assess the functions of site-specific chromatin marks that have been previously mapped[129,130].

One potential advantage of epigenome reprogramming over CRISPRi or CRISPRa approaches discussed above is that the consequences on cell function can be more stable. This has raised interest in hit-and-run epigenetic reprogramming with CRISPR[131]. With this method,

CRISPR is used to recruit multiple modifications to a target locus to induce stable gene silencing. This has potential for therapeutic manipulation of immune cells. Gene targets could be disabled with a transient treatment that does not alter any genetic sequences. Epigenome reprogramming of immune cells holds notable potential for adoptive cellular therapies.

Biochemistry and Imaging Studies of Chromatin

Gene regulation programs depend on physical interactions between transcriptional regulators and cis-regulatory elements, and complex three-dimensional interactions among chromatin sites. CRISPR tools are being developed to image chromatin sites and facilitate proteomic and genomic studies of chromatin interactions. Biotinylated dCas9 can be used to pull down endogenous genomic sequences and identify bound transcription factors in an unbiased manner by mass spectrometry and local 3D chromatin interactions by sequencing[132,133]. Fluorophore fusions to dCas9 have been used to visualize genomic loci in living cells in real time[134]. Using the SunTag system to recruit additional fluorophores to a single dCas9 molecule or similar methodologies to improve the signal may allow imaging of single genomic sites to address questions of nuclear organization and chromatin remodeling[113]. With the ability to write large exogenous sequences into the genome it is possible to begin to tag endogenous transcription factors[135,136]. High affinity tags can be fused to transcription factors to enable pulldown studies to map DNA binding sites and interacting partners[133,136,137]. Additionally, nuclear factors can be fused to fluorescent proteins or tags to study their localization in the cell and their dynamic interactions with DNA[135]. These emerging tools collectively should shed light on the physical interactions that contribute to immune cell circuitry.

Base Editing

Engineered effectors coupled to nuclease-deficient CRISPR systems now offer genetic reprogramming in addition to epigenetic reprogramming. Base editing is a new approach to rewriting genetic sequences. Base editors use cytidine deaminases to introduce nucleotide changes at the targeted genomic site[137]. This approach to genome editing has potential advantages over nuclease-dependent strategies, as DSBs are not introduced and desired editing outcomes may be achieved more homogeneously. Several iterations of this technology have been published that have improved the editing efficiency and flexibility of base editors[138-140]. Preclinical evidence suggests that this technology may be useful for therapeutic purposes to correct single-nucleotide mutations or introduce disease-causing mutations in cells to model disease. The mutagenic potential of cytidine deaminases does raise concern for off-target effects. Further modifications to base editor systems promise to deliver precise single-nucleotide edits at the on-target site. Base editing technology has also proven useful for targeted mutagenesis screens. Recent platforms were engineered to have mutagenic activity over larger DNA sequences (~100 bp), introducing distributed transition and transversion nucleotide changes giving rise to allelic diversity[141,142]. This functional diversification of DNA sequences at a target site, which could include loss-of-function and gain-of-function mutations, cannot be achieved readily with Cas9 cutting, which mostly introduces indels. Saturation mutagenesis screens using base editing across noncoding and coding DNA sequences promise insights into the grammar of our DNA.

Additional CRISPR Systems

New CRISPR systems continue to be discovered in bacteria and mined for new functions. For example, although CRISPR systems have largely been used to target DNA sequences, some have been discovered now that target RNA sequences[143]. Ongoing discovery of microbial systems is likely to provide an expanded CRISPR toolbox for genetic engineering of immune cells.

GENETIC VARIATION AND IMMUNE REGULATION

Mapping Genetic Circuits of Immunity

Immune homeostasis depends on complex coordination of cellular programs. Diverse cell types must traffic to appropriate sites, recognize antigenic targets, and respond effectively to threats without causing autoimmunity. These specialized, cell type–specific, and stimulus-responsive programs are governed by molecular circuits comprising signaling pathways, trans-regulators (e.g., transcription factors and epigenetic regulators) and networks of cis-regulatory elements and target genes. Correlative cellular measurements like transcriptome and chromatin analyses have been used extensively to infer elements of this circuitry. However, true understanding of circuit function depends on the ability to perturb the putative elements and observe the effects. CRISPR technology is well suited for systematic perturbation studies, and we can now directly test causal relationships between genome sequences and cellular phenotypes. CRISPR enables rapid high-throughput functional studies of immune cells to comprehensively map genetic regulators of cellular phenotypes.

Pooled CRISPR screens are emerging as a powerful approach to identify genes that control immune functions. A genome-wide CRISPR screen dissected the innate immune response of bone marrow–derived dendritic cells (BMDCs) [45]. BMDCs isolated from Cas9-expressing mice were transduced with gRNAs to introduce different genetic perturbations across the population. gRNA-transduced BMDCs were stimulated with LPS and then sorted based on the resulting levels of Tnf induction to identify the gRNAs that targeted key regulators. This unbiased screen identified known and unknown regulators of Tnf induction. Hits could be grouped into common biological pathways and known protein complexes illustrating the power of high-throughput CRISPR screens. This flow-based approach can be adapted to study diverse biological pathways or cellular phenotypes that can be selectively enriched by FACS.

Deeper measurements of the molecular phenotypes arising from each CRISPR perturbation would provide richer insights into gene circuity. Combining single-cell RNA sequencing (scRNA-seq) with pooled CRISPR libraries has allowed investigators to observe the effects of individual genetic perturbations on a cell's transcriptome. Three studies published simultaneously at the end of 2016 established this pooled CRISPR screening approach[144-146]. The power of this approach linked genetic perturbations to divergent cellular responses, developmental pathways, and faithful gene regulatory circuits. This required technical advances to obtain single-cell transcriptome data and simultaneously capture the gRNA found in each cell, as the gRNAs could not be directly captured by scRNA-seq due to their short length and absence of 3 poly A tails. Instead these studies sequenced the lentiviral constructs to pair individual gRNAs with a unique bar code in the 3 untranslated region of a fluorescent protein transcript. A variant of this protocol termed CROP-seq developed a vector in which gRNA sequences are captured directly by scRNA-seq[147]. The current costs associated with scRNA-seq can be prohibitive for large high-throughput studies. However, rapid advances in the technology, lower costs, and computational methods are beginning to make these studies feasible, which greatly accelerates our ability to map genetic circuits of immunity.

Noncoding Elements in Immune Circuitry and Disease Risk

CRISPR is also a powerful tool to probe noncoding elements in the genome. Deciphering the genetic underpinnings of common autoimmune diseases and other complex diseases of the immune system requires improved understanding of how noncoding sequence variation regulates gene expression and immune cell function. Most common autoimmune diseases are thought to have a complex genetic etiology that stems from the combined effects of common variants and environmental factors. Although individual autoimmunity variants have relatively small effects on disease risk, they mark important regions of our genome that are critical for normal immune function. Over the last two decades, genome-wide association studies (GWAS) have linked

common variation in hundreds of loci across the genome to risk of autoimmunity. Roughly 90% of the genetic variants implicated in autoimmune disease risk do not alter protein-coding sequences but rather fall in noncoding regions of the genome that remain relatively poorly understood[148,149]. Mapping functional noncoding sequences that harbor autoimmunity variants and identifying the biological programs they regulate will be critical in understanding how common variants predispose to autoimmunity.

Noncoding sequences harbor hundreds of thousands of putative enhancers—transcription factor docking sites that shape transcriptional programs in response to specific cellular signals. Consortia like ENCODE and the Roadmap Epigenomics Project have profiled transcription factor binding and epigenomic marks across diverse cell types and cell states to map putative enhancers in noncoding sequences[150,151]. These maps revealed that a majority of autoimmunity variants reside in sequences with features of immune enhancers. Taken together these data suggested that the dysregulation of transcriptional circuits plays a causal role in autoimmunity. Our ability to delete, paste, and rewrite the genetic code with CRISPR is now transforming our understanding of noncoding sequences and has enabled rapid functional testing of autoimmunity variants.

Sequence perturbation is the gold standard for identifying functional noncoding sequences. Given how little is known about the molecular grammar of sequences outside of amino acid–coding regions, CRISPR-based saturation mutagenesis screens of noncoding stretches of the genome have proven useful. These relatively unbiased functional screens use thousands of gRNAs to tile across entire loci, saturating them with Cas9-induced indels. Edited cells are then binned and sorted on target gene expression or downstream phenotype and sequenced to quantify the enrichment or depletion of gRNAs, a measure of the regulatory effects of the targeted genomic site. Several Cas9 cutting screens have been published showing the utility of this approach in identifying functional noncoding sequences[152,153]. Similarly, CRISPRi recruitment of the

transcriptional repressor dCas9-KRAB can also be used to map functional enhancers[154]. These loss-of-function strategies identify elements that are required for gene regulation in the particular context in which the screen is performed. However, many enhancers only contribute to gene regulation in the context of specific extracellular cues. We adapted CRISPRa to map functional enhancers where the activation domain was sufficient to induce a specific target gene. By recruiting dCas9-VP64 via CRISPRa to thousands of genomic sites in pooled experiments, immune enhancers were mapped across two key autoimmunity risk loci, *CD69* and *IL2RA*[155]. In these experiments, CRISPRa was able to identify stimulus-responsive enhancer elements even in unstimulated cells. Taken together, CRISPR-based approaches are revealing functional noncoding elements and linking them to their target genes.

CRISPR is also being used to fine-map critical sequences within individual enhancer elements. In enhancer bashing studies, Cas9 perturbations are targeted to every gRNA site within an enhancer. By sorting enhancer-edited cells on target gene expression or a downstream phenotype and correlating enhancer edits with gene expression, it is possible to footprint nucleotides required for optimal enhancer function[156]. This approach improves the resolution with which we can study noncoding sequences and variation within them. Ultimately, developing methods to efficiently introduce autoimmunity variants by HDR or base-editing on an isogenic background will allow for direct assessment of variant function. This fine-resolution functional mapping moves us toward an understanding of how single-nucleotide variants can tune gene programs.

The next frontier will be the mapping of functional enhancers across the genome and capturing their effects on specific transcriptional programs. Further resolution of functional sequences can be achieved by tiling regions with nucleases that have different PAM specificities. Screening with other dCas9-coupled effector molecules could discriminate classes of enhancers in

different functional chromatin states—for example, poised versus active enhancers. In addition, a recent study screened combinations of noncoding CRISPR perturbations with scRNA-seq to understand the regulatory logic of super-enhancer clusters[157]. Although technical and analytic challenges remain, these approaches are linking noncoding elements to downstream gene programs, providing biological insights into how noncoding variants can alter immune cell circuitry and contribute to disease risk.

Reverse Genetics of Pathogenic Sequence Variants

GWAS and genome sequencing of patients with monogenic immune dysregulation are identifying a growing number of candidate causal pathogenic sequences affecting the immune system. Genome editing is critical for testing which of these are truly causing disease phenotypes and for determining the underlying mechanism of pathology. One challenge is that natural variants are often inherited in combinations, making it very difficult to differentiate the ones that cause disease risk from neutral variants in the same in individuals. CRISPR provides a tool to create isogenic cells that differ only at a single targeted genomic site and to assess the in vitro and in vivo cellular effects of genetic variants.

Modeling conserved human variants provides an opportunity to pinpoint functional consequences of disease variants in an intact immune system. For example, CRISPR mouse engineering enabled functional assessment of a common coding single-nucleotide polymorphism (SNP) in PTPN22 implicated in risk of type 1 diabetes and other autoimmune diseases. As predicted from human population studies, non-obese diabetic (NOD) mice CRISPR-engineered with the human risk allele had increased incidence of autoimmune diabetes. CRISPR-engineered mice can also test the functional consequences of noncoding variants implicated in immune disease. Noncoding variants may only affect target gene regulation in particular cell types or in

response to particular stimuli. We recently used CRISPR mice to study a noncoding autoimmunity SNP found in an Il2ra intron[155]. The fine-mapped SNP explains the risk for Crohn disease at the IL2RA locus[158]. Interestingly, this same SNP is protective for type 1 diabetes, suggesting it might have context-restricted effects with divergent outcomes on disease[149]. We engineered SNP knock-in mice that differed in only the SNP nucleotide. We examined different T cell subsets under different states to map the context within which Il2ra might be dysregulated. We found that the SNP resides within a conserved stimulation-responsive Il2ra enhancer and delays the kinetics of Il2ra induction on naive T cells as they respond to stimulation[155]. Looking forward, CRISPR-generated animal models of human disease variants provide opportunities to localize disease risk to specific cellular compartments, test epistasis of multiple risk variants, and assess candidate pharmacological interventions.

Many genetic variants implicated in human disease are not conserved in mouse. Thus, genome editing in human cell lines and primary immune cells often is required to assess causal effects. Although Cas9 delivery to human cells was initially challenging, CRISPR can now be efficiently deployed for genome modification in primary human cell types. Purified recombinant Cas9 protein can be mixed in vitro with gRNAs to make Cas9 RNPs[66,159,160]. Cas9 RNPs can be electroporated in primary human immune cells to generate knockout and knock-in modifications[80,135,161]. This approach has been used to engineer DNA sequences in hematopoietic stem cells (HSCs), T cells, and B cells[80,87,103,161,162]. Genetic modification using Cas9 RNPs is likely to be successful in other primary immune cell types as well. Electroporation of Cas9 mRNA has also been employed for gene modifications in immune cells[161,163].

CRISPR now can be used to rapidly test variants, including variants in linkage disequilibrium with each other, to functionally fine-map genetic associations. For example, a recent study found a noncoding SNP in the CEBPA locus associated with altered basophil

counts[164]. CEBPA is a hematopoietic transcription factor, but it had not been previously linked to basophil abundance. In situ perturbation of the SNP site in human stem and progenitor cells revealed it resided within a CEBPA enhancer that influences basophil differentiation[164]. Remarkably, pursuing this natural genetic association revealed new biology about a cis-regulatory code underpinning basophil differentiation. Performing these studies in different cell types and states will be critical to assigning function to immune disease variants.

Therapeutic Correction of Monogenic Mutations in Immune Cells

CRISPR not only provides a means to confirm and characterize pathogenic variants but could also provide an avenue to correct the genetic causes of immune cell dysfunction. Therapeutic correction of causal mutations in the affected cell types or their progenitors can be curative for patients with immune-related disorders (Figure 1.3a). Recent work has focused on improving CRISPRCas9 editing efficiency in primary human immune cells to rewrite monogenic-disease-causing variants[135,165-167]. Diseases where the standard of care is currently allogeneic stem cell transplantation may eventually be treated with ex vivo stem cell correction and autologous transplantation. There are potential risks that must be explored, including unintended genome modifications[51,57,58,168-170] and perhaps altered cell programs that result from the CRISPR machinery[171,172], electroporation, or DSBs[173,174]. On the other hand, the risk of graft-versus-host disease would be reduced and treatments could become available for patients without donor matches. For example, severe combined immunodeficiency (SCID) is a rare genetic disorder that can be caused by diverse mutations. CRISPR with HDR-mediated correction could provide a flexible system for therapeutic correction of mutations in HSCs to restore gene function needed for healthy T cell generation. The base editor system is an alternative approach to therapeutic correction of T-to-C mutations, which does not require DSB formation or exogenous DNA for

repair. As protocols improve for differentiation of pluripotent stem cells, induced pluripotent stem cells could provide renewable resources of patient cells for experimental optimization and perhaps therapeutic gene correction[165].

Mutation correction in differentiated immune compartments is also being explored as a therapeutic strategy. For example, siblings in a family with varying autoimmune manifestations caused by recessive mutations in IL2RA have FOXP3+ regulatory T cell (Treg)-like cells that do not express appropriate levels of IL2RA and are dysfunctional. We demonstrated that nonviral CRISPR-based genome targeting could correct a pathogenic IL2RA mutation and rescue IL2RA expression in T cells from these patients[89]. Correction could be achieved in FOXP3+ cells, raising the possibility of autologous transfer of gene-corrected Treg therapy for the affected children. Tregs can enforce dominant tolerance, suggesting that a relatively small number of corrected Tregs might restore immune homeostasis to affected children. In the future, there may be more opportunities to tailor gene surgery approaches to specific cell populations that are impaired by a particular patient mutation. Ex vivo CRISPR gene correction is advancing rapidly toward the clinic. As delivery strategies for CRISPR continue to advance, there may eventually be opportunities for targeted in vivo editing of somatic cells to treat monogenic disease. These fields continue to evolve and much validation remains to be done, but there are concerted efforts to achieve safe and effective targeted genome surgery in immune cells.

MICROBIAL IMMUNITY

A primary function of the immune system is to recognize and eradicate pathogens. Understanding the genetic factors that regulate how immune cells interact with pathogens could reveal critical pathways co-opted by pathogens and open new therapeutic avenues. Almost 40 million people around the world are infected with human immunodeficiency virus (HIV). This virus selectively

infects CD4+ T cells and causes life-threatening immunodeficiency. The virus is well known to depend on host factors in human T cells at multiple stages of its life cycle. Several groups used RNA interference (RNAi) knockdown approaches to search for these host dependency factors in cell lines[175-177], but results were inconsistent. CRISPR, which tends to have fewer off-target effects than RNAi and can generate complete knockout cells, has renewed hopes for systematic identification of host factors that influence HIV infection.

Both pooled and arrayed CRISPR screens have been performed to identify host factors that influence HIV infection. A genome-wide CRISPR screen was performed in a human T cell line to identify genome modification that confers strong resistance to HIV[178]. This unbiased approach identified a remarkably restricted set of factors, including known entry receptors, that could be deleted to confer cell survival and resistance to infection upon challenge with HIV in vitro. The function of a novel gene pathway that post-translationally modifies the CCR5 coreceptor was validated by CRISPR modification of primary human T cells. Targeted arrayed screens have also been piloted in primary human T cells. Several candidate host factors implicated by protein-protein interaction studies[179] and the HIV literature were individually deleted with CRISPR. Knockout primary human CD4+ T cells for each gene were generated using three different Cas9 RNPs and tested for HIV infection[180]. This arrayed platform quantified rates of HIV infection in each cell population with high-throughput flow cytometry and identified both known and unknown host factors important for HIV infection. More broadly, this work established an arrayed CRISPR platform to test effects of genetic perturbations in primary human immune cells for studies of infectious diseases. Taken together, pooled and arrayed CRISPR screens are providing clues to critical functional interactions between immune cells and pathogens that infect them.

Various groups are considering how to translate CRISPR insights into HIV pathogenesis into new strategies for an HIV cure. Deleting host factors in human CD4+ T cells or HSCs to limit

HIV infection is one gene therapy approach for HIV therapy. Ablation of the HIV coreceptors CCR5 and CXCR4 can generate CD4+ T cells that are resistant to infection[181-183]. As above, CRISPR screens are identifying additional host factors that may also be modified to ensure viral resistance. Direct CRISPR targeting of the HIV genome has also been explored, although therapeutic delivery would be challenging. Recent work used gRNAs in the long terminal repeats of HIV that flank the viral genome to excise the virus from human T cells[184,185]. In a similar approach, conserved sequences of HIV were targeted by CRISPR-Cas9 to functionally ablate the virus[186]. In another approach pioneered with TALENs and meganucleases rather than Cas9, genome engineering was performed to knock-in an anti-HIV chimeric antigen receptor (CAR) sequence into the CCR5 locus[84]. Similar approaches could be extended to other pathogens to understand the genetics of their interactions with the immune system in addition to designing genome-engineering approaches for therapy[187-190].

CANCER IMMUNITY

Immunotherapy is offering new hope for previously untreatable cancers. Checkpoint inhibitors are reversing T cell dysfunction and causing productive anticancer immune responses in some patients[191]. In addition, adoptive transfer of tumor-infiltrating lymphocytes and genetically engineered T cells have demonstrated the potential for cellular therapies as a new drug class. Despite these advances in immunotherapy, a large fraction of malignancies remain incurable. CRISPR is being used to understand cancer immunity through unbiased genetic perturbation studies in immune cells and cancer cells. Unbiased CRISPR screens are rapidly revealing the genetic underpinnings of T cell responses and pointing toward new targets for pharmacological checkpoint blockade or genetic engineering in cell therapies.

CRISPR-Cas9 pooled screens can be used to rapidly map gene circuits that regulate cancer immunity. T cell exhaustion through activation of the PD-1/PD-L1 pathway is a major mechanism by which cancer cells evade the immune system. Given the remarkable clinical success of drugging this pathway, there has been great interest in understanding the regulation of these molecules. Genome-wide CRISPR screens using a FACS-based sorting strategy for PD-1 in T cells[178] or PD-L1 in cancer cells[192] identified factors that were critical for their expression. These studies identified novel trans-factors and implicated specific cellular pathways in the regulation of PD-1/PD-L1 expression. CRISPR has also helped to decode cis-regulatory circuitry of the PD1/PD-L1 pathway. Sen et al. [193] mapped chromatin accessible sites in acute and chronically activated murine CD8+ T cells to identify putative exhaustion enhancers in the PD-1 locus. A CRISPR-Cas9 pooled screening approach was used to saturate accessible sites with perturbations confirming required cis-regulatory sequences for eight of these putative enhancers[193]. Taken together, CRISPR screens are a powerful platform to map coding and noncoding sequences that regulate pathways for cancer immunotherapy.

Productive immune clearance of malignancies depends on genetic programs in cancer cells and immune cells. In vitro CRISPR experiments with coculture systems have been used to identify mutations in cancer cells that affect their survival in the presence of antigen-specific CD8+ T cells[194]. This work identified genes in antigen processing and presentation critical for T cell killing[194,195]. In vivo CRISPR screens have also been used to study the interaction between the immune system and transplantable tumors[196]. Due to the limit in numbers of cells that can be assayed in in vivo screens this study focused on genes that represent key functional pathways. Pools of edited cancer cells were transplanted into immunosufficient mice that were then given immunotherapies to identify genes that were important for resistance or susceptibility to these treatments. Immunodeficient animals were used to control for cancer cell–autonomous effects of

gene knockouts. TNF activation/NF-κB signaling, antigen processing and presentation, inhibition of kinase signaling, and ubiquitin proteasome pathway were all found to increase the efficacy of immunotherapy. In addition, PTPN2 knockout was found to sensitize cancer cells and promote immunotherapy by increasing antigen presentation and IFN-γ signaling. CRISPR-based functional studies in T cells can prioritize novel targets for immunotherapy drug development and improve the design of genetically reprogrammed adoptive cellular immunotherapies. We recently overcame challenges of genome-wide CRISPR screens in primary human immune cells to identify regulators of T cell stimulation and immunosuppression responses[47]. Together these approaches are revealing the complex network of genetic factors that mediate immune responses to cancer immunotherapy.

ENGINEERING CELLULAR THERAPIES

Cellular therapies are an emerging treatment class for human diseases. The US Food and Drug Administration has now approved two genetically modified T cell therapies for cancer immunotherapy. These products rely on nontargeted viral integrations to insert CARs into T cell genomes, programming the immune cells to recognize an antigen found on malignant cells (Figure 1.3b). As cell therapies gain momentum, CRISPR provides countless opportunities to modify endogenous immune cells' genome sequences to enhance therapeutic properties. For example, efforts are underway to engineer T cells to overcome the immunosuppressive tumor microenvironment. For example, several groups have demonstrated PD-1 can be ablated in engineered antigen-specific T cells[197]. The first CRISPR-Cas9 clinical trials are now enrolling patients for treatment with engineered TCR specificity (NY-ESO-1 TCR viral transduction with CRISPR-mediated TRAC and TRBC gene deletion) plus CRISPR PD-1 deletion. Numerous academic and commercial groups are pursuing additional gene targets that can be deleted to

enhance anticancer efficacy of immune cells[198] or to make off-the-shelf allogeneic CAR T cells that escape immune rejection[163].

The delivery of therapeutic transgenes to endogenous loci helps to prevent collateral damage to other genes and preserves endogenous regulation of the transgene, which can be important for engineered cellular function. For example, recent work with CRISPR editing plus AAV templates suggested that targeted site-specific integration of CARs can produce products with more homogenous CAR expression and improved efficacy in preclinical models[86]. We recently developed a strategy for nonviral integration of specific TCRα and TCRβ pairs into the endogenous TCRα locus of polyclonal T cell populations to generate functional cells with a desired antigen specificity[89] (Figure 1.3b). Thus, without the need for time-consuming virus-production steps, newly identified TCRs can be rapidly engineered into T cells for immunotherapy. More broadly, nonviral genome-targeting technology will enable us to delete, insert, or rewrite genetic sequences to replace genes, tune regulatory programs, and rewire immune cells to obtain desired functions.

CRISPR has been widely adopted to decode the fundamental circuitry of the immune system, as we have reviewed here. Beyond decoding, CRISPR also offers the opportunity to program new biology into immune cells (Figure 1.3c). As discussed above, immune cell specificity can already be written. Genome surgery will be attempted to correct pathogenic mutations that cause primary immune dysregulation syndromes. Likewise, CRISPR editing has the potential to strengthen Treg cell programs to suppress autoimmunity, graft-versus-host disease, and transplant rejection. In time, we are likely to have new tools to rewrite how cells sense extracellular signals and how they traffic, proliferate, and survive in the body, and to modify the effector programs they activate in specific settings[199].

We have discussed how CRISPR is providing insight into coding and noncoding gene programs that shape how immune cells contribute to autoimmunity, interact with pathogens, and participate in cancer immunotherapy. These CRISPR-generated functional maps will point us to the genomic sites that can be modified to alter these codes. With improving technology to rewrite nucleotides at those sites, synthetic biology approaches could be harnessed to confer complex functional logic into cells and tune their therapeutic properties. As we think about the next generation of engineered cellular therapies, it is important to consider how to manufacture these cells safely and ensure their beneficial effects in the human body.

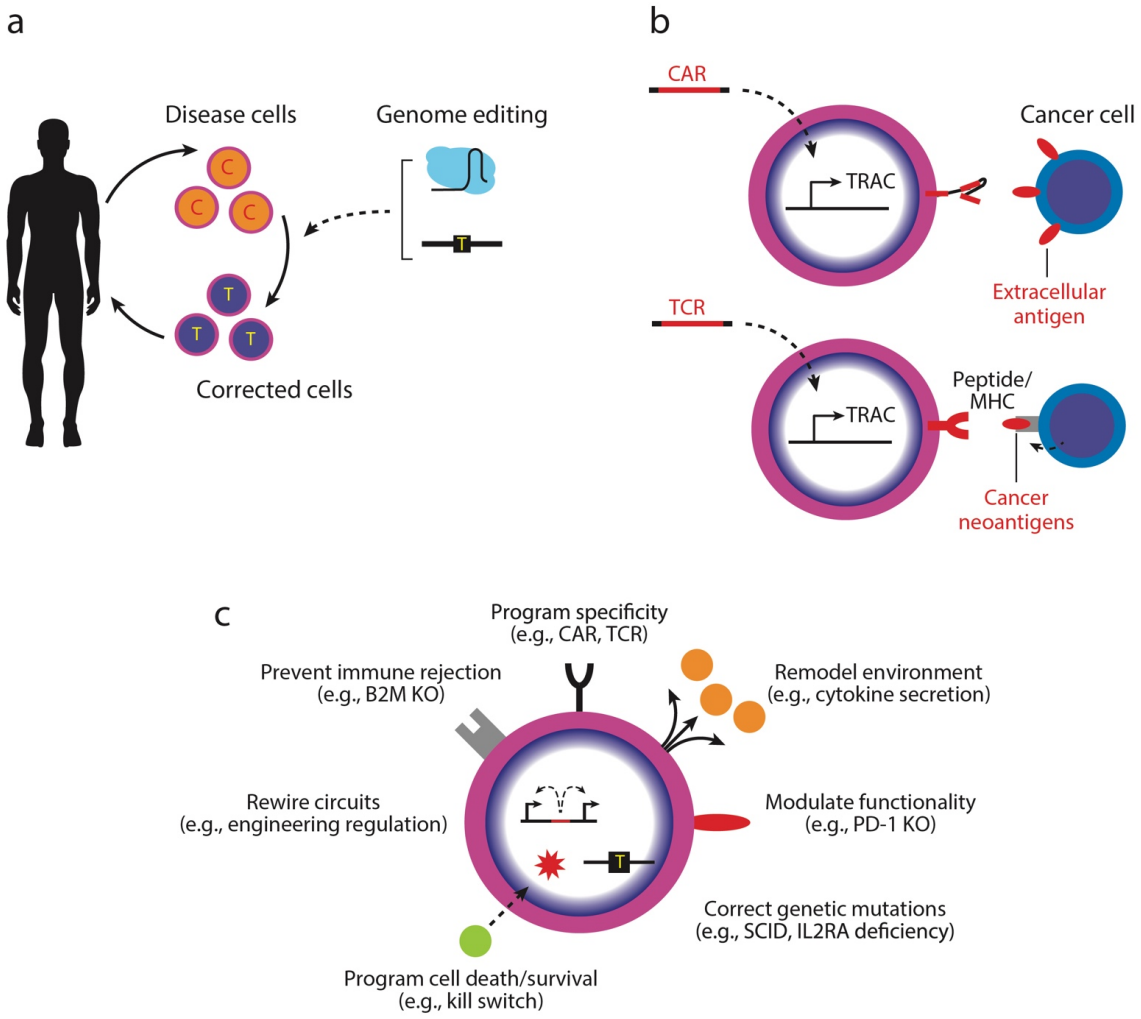


Figure 1.3. Using CRISPR to engineer immune cell therapies. (***a***) CRISPR-Cas9 gene editing of immune cells for correction of immune-related genetic disorders. (***b***) Engineering immune cell specificity through nonviral targeting of chimeric antigen receptors (CARs) or T cell receptors (TCRs) at endogenous loci. (***c***) Desired genetic modifications for cellular immunotherapy.

CONCLUSIONS AND FUTURE DIRECTIONS

The past decades have seen a revolution in reading genome sequences. An ever-increasing number of patient genomes are being sequenced and analyzed. Common variants that modulate the risk of immune dysregulation have been mapped, along with rare mutations that cause Mendelian forms

of immune dysregulation. However, major challenges remain to determine causal mutations, relevant genes, and affected cellular pathways. Moreover, the critical challenge remains to translate new genetic knowledge into new clinical interventions. These challenges require the ability to alter genome sequences and not merely the ability to read them. CRISPR has now brought the next revolution in writing genome sequences. In the post-CRISPR era, immunologists can query which genome sequences control specific immune functions. As adoptive immune therapies continue to advance, there are mounting opportunities to employ CRISPR-engineered cells for new targeted treatments. This could transform how genetic diseases of the immune system are treated in the future. CRISPR will help to validate causal mutations in affected cell types. As we develop tools needed for safe and effective genome surgery, impaired immune cells or hematopoietic stem cells could be corrected ex vivo and adoptively transferred into patients. Looking forward, decoding of immune cell programs with CRISPR will also enable more complex reprogramming of immune cells to make them more flexible and effective cellular drugs for a wide range of human diseases.

CHAPTER 2. Discovery of Stimulation-Responsive Immune Enhancers Using CRISPR Activation

The majority of genetic variants associated with common human diseases map to enhancers, non-coding elements that shape cell type-specific transcriptional programs and responses to extracellular cues[148,200,201]. Systematic mapping of functional enhancers and their biological contexts is required to understand the mechanisms by which non-coding genetic variation contributes to disease. Functional enhancers can be mapped by genomic sequence disruption[202-204], but this approach is limited to the subset of enhancers that are necessary in the particular cellular context being studied. We hypothesized that recruitment of a strong transcriptional activator to an enhancer would be sufficient to drive target gene expression, even if that enhancer is not currently active in the assayed cells. Here, we developed a discovery platform that can identify stimulus-responsive enhancers for a target gene independent of stimulus exposure. We used tiled CRISPR activation (CRISPRa) [205] to synthetically recruit a transcriptional activator to sites across large genomic regions (>100 kilobases) surrounding two key autoimmunity risk loci, *CD69* and *IL2RA* (interleukin-2 receptor alpha; *CD25*). We identified several CRISPRa Responsive Elements (CaREs) with chromatin features of stimulus-responsive enhancers, including an *IL2RA* enhancer that harbors an autoimmunity risk variant. Using engineered mouse models, we found that sequence perturbation of the disease-associated *IL2RA* enhancer did not entirely block *IL2RA* expression, but rather delayed the timing of gene activation in response to specific extracellular signals. Enhancer deletion skewed polarization of naïve T cells towards a pro-inflammatory Th17 state and away from a regulatory T cell state. This integrated approach identifies functional enhancers and reveals how non-coding variation associated with human immune dysfunction alters context-specific gene programs.

We adopted CRISPRa for high-throughput functional enhancer discovery with large libraries of guide RNAs (gRNAs) that tile genomic loci of interest (Figure 2.1). We first validated the CRISPRa method for enhancer discovery at the *CD69* locus, which contains a previously-characterized stimulation-responsive enhancer[206]. CD69 is a cell surface receptor that is rapidly induced on T cells in response to T cell receptor (TCR) stimulation[207]. We asked whether CRISPRa could identify *CD69 cis*-regulatory elements in resting cells, even in the absence of TCR stimulation. We transduced resting Jurkat T cells stably expressing dCas9-VP64 with a pooled lentiviral library of gRNAs (10,780 gRNAs) that targeted sites at all *S. pyogenes* Cas9 PAMs throughout a 135 kb region at the *CD69* locus, starting 100 kb upstream of the transcription start site [TSS] and extending through the gene body and 25 kb downstream (Figure 2.1). We sorted transduced cells into four bins of CD69 expression and measured the distribution of gRNAs in the sorted populations (Figures 2.1-2.3). As expected, the cells with high CD69 expression were enriched for gRNAs targeting the *CD69* transcriptional start site (TSS) (Figure 2.1)[117]. We also observed enrichment for gRNAs at three regions well outside the TSS-centric CRISPRa window. We refer to these sites as CRISPRa Responsive Elements (CaREs). One of the three CD69 CaREs represents a previously-characterized stimulation-responsive enhancer referred to as conserved non-coding sequence 2 (CNS2)[206]. Hence, tiling a transcriptional activator (dCas9-VP64) to non-coding sequences can identify stimulation-responsive enhancers in unstimulated cells.

We next applied our enhancer-discovery approach to the *IL2RA* locus. *IL2RA* encodes a subunit of the high affinity interleukin-2 (IL2) receptor also known as CD25. Genome-wide association studies (GWAS) have implicated non-coding variants in the *IL2RA* locus as risk factors for at least eight autoimmune disorders, underscoring the critical role of *IL2RA* regulation in human immune homeostasis[148]. However, the functional impact of *IL2RA* disease variants remains unclear because of the complex regulatory landscape at the *IL2RA* locus that is responsive to

multiple signals. In resting conventional T cells, *IL2RA* is not only induced by antigen stimulation via the TCR, but is also potently regulated by a number of other signals. Regulators of *IL2RA* expression include the cytokine IL2, which upregulates the receptor as part of a positive feedback loop[208,209]. *IL2RA* regulation is also dependent on cellular programming. FOXP3+ regulatory T cells (Tregs), which are required to suppress auto-reactive T cells and prevent the development of multi-organ autoimmunity, constitutively express high levels of *IL2RA* and depend on it for survival[210]. We hypothesized that multiple extracellular and cell-type specific signals are integrated to regulate gene expression through effects on distinct enhancer elements within the T cell super-enhancer at the *IL2RA* locus[211,212]. Whereas coding mutations in the gene affect all cell types that express IL2RA[213], disease-associated non-coding variants could selectively affect *IL2RA* induction in conventional T cells in response to a specific signal, or impair constitutive expression in Tregs. We sought to map functional *IL2RA* enhancer elements and determine how known disease risk variants affect enhancer function.

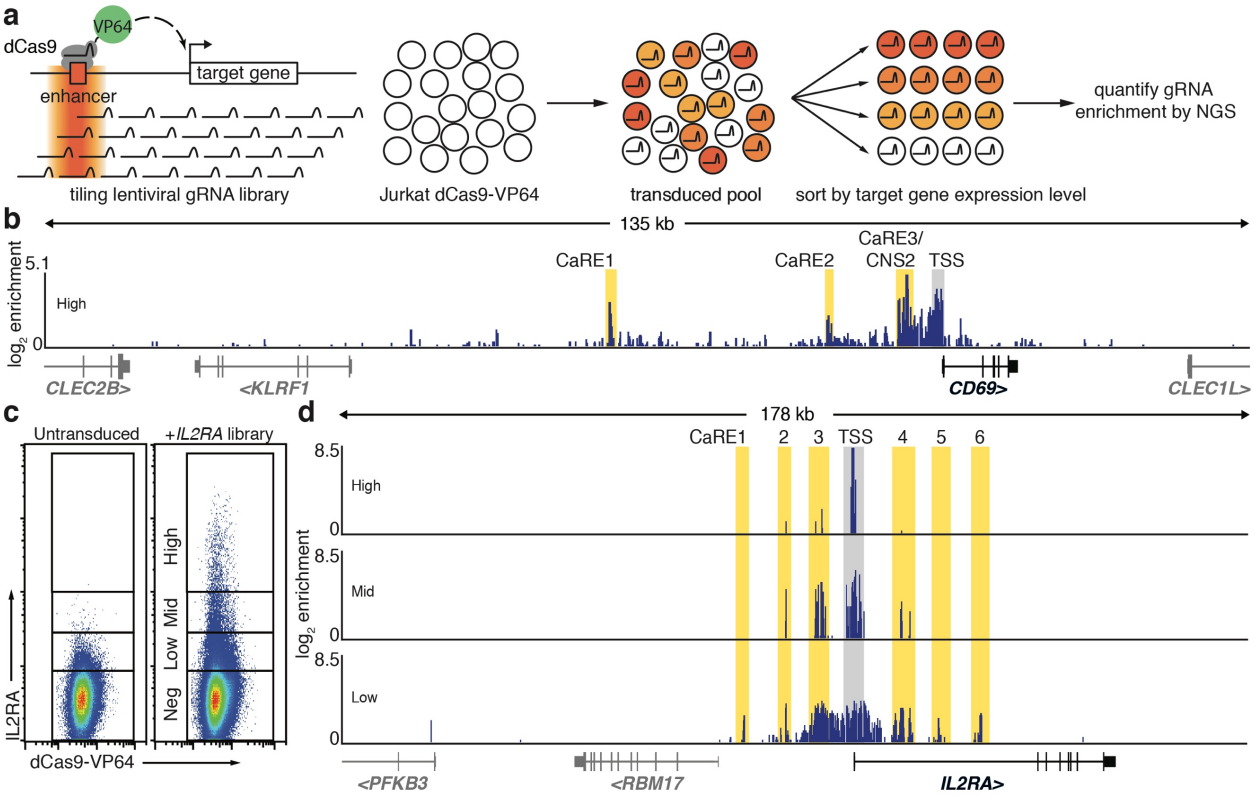


Figure 2.1. Discovery of putative enhancers with a tiling CRISPRa screen. (**a**) Schematic of the CRISPRa screen workflow. (**b**) Genomic coordinates of gRNAs plotted against enrichment into the "CD69 High" sorted population. Fold-enrichment over gRNA abundance in unsorted cells is plotted in 5-gRNA sliding windows. Peaks of guide activity are highlighted. (**c**) Flow cytometry distribution of IL2RA expression on Jurkat-dCas9-VP64 cells transduced with the IL2RA tiling gRNA library. (**d**) Genomic coordinates of gRNAs plotted against enrichment into the IL2RA "High", "Mid", and "Low" sorted population, plotted as in (b).

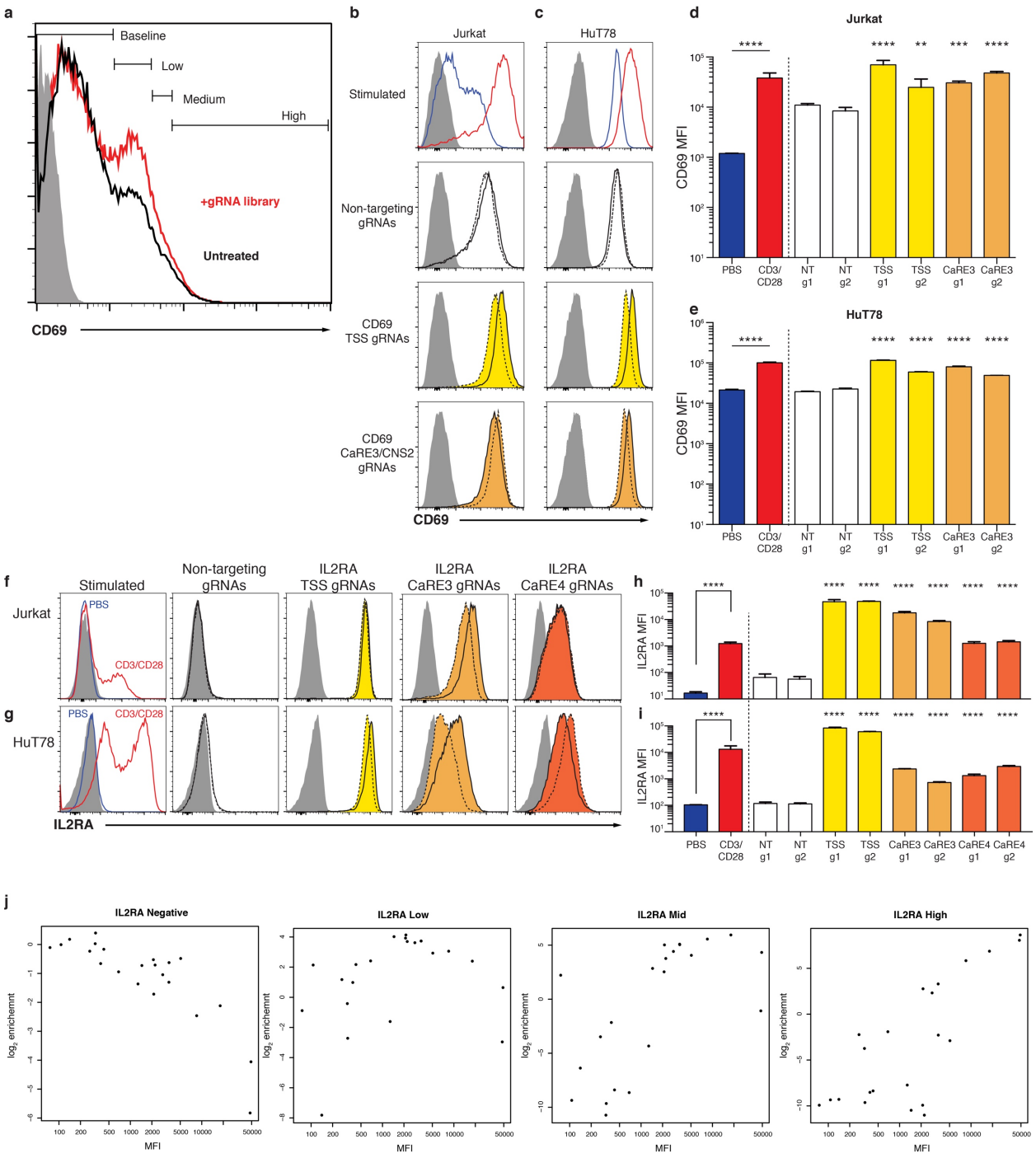


Figure 2.2. Upregulation of target gene expression on gRNA-expressing cells. (a) Distribution of CD69 expression on Jurkat-dCas9- VP64 cells transduced with the CD69 tiling gRNA library. (b,c) Representative flow cytometry plots of CD69 expression on Jurkat (b) or HuT78 cells (c) transduced with dCas9-VP64 and individual gRNAs. For each target region or control, solid black lines represent gRNA 1 and dashed black lines represent gRNA 2. Shaded gray histograms represent isotype control staining. Cells stimulated for 48 h with plate-bound anti-CD3/CD28 antibodies are shown for comparison. (d,e) Isotype-subtracted geometric mean fluorescence intensity (MFI) of data in (b) and (c). (f,g) Representative flow cytometry plots of IL2RA

expression on Jurkat (f) and HuT78 cells (g) as in (d). (h,i) Isotype-subtracted geometric mean fluorescence intensity (MFI) of data in (f) and (g). Statistical tests were performed on log-transformed MFI values. PBS and anti-CD3/CD28 treated samples were compared using an unpaired two-tailed student's t test. TSS and CaRE gRNA samples were compared to each non-targeting (NT) gRNA sample using one-way ANOVA followed by Sidak's multiple comparisons test. Data are presented as mean ± s.d., n=3 biological replicates. Data in (b-i) are representative of at least 2 independent experiments. **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$. (j) Jurkat dCas9-VP64 cells were transduced with individual gRNAs from the IL2RA library, and surface IL2RA expression was measured by flow cytometry. The isotype-subtracted geometric mean fluorescence intensity (MFI) of the transduced cells is plotted against gRNA enrichment in the indicated bin in the IL2RA screen.

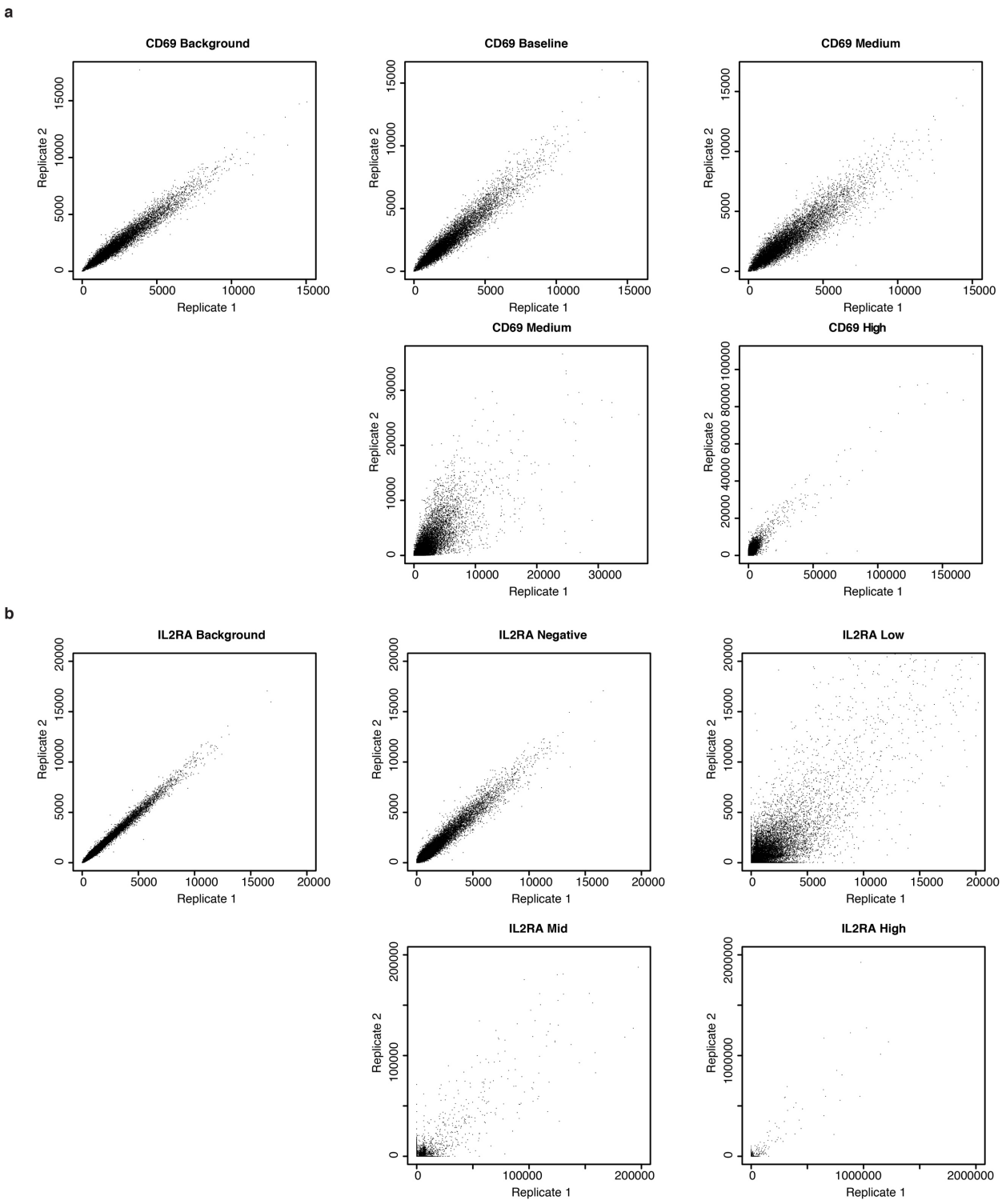


Figure 2.3. Correlation of results across CRISPRa screen replicates. Normalized read counts for gRNAs in the indicated cell populations are compared between biological replicates of the CD69 screen (a) and IL2RA screen (b).

To discover *IL2RA* CaREs, we transduced Jurkat-dCas9-VP64 cells with a library of 20,412 gRNAs tiling 178 kb around the *IL2RA* locus (Figure 2.1). Transduced cells were sorted into four bins of expression ("Negative", "Low", "Mid", and "High"). Analysis of gRNAs enriched in each bin revealed six CaREs leading to different levels of IL2RA expression: three in the first intron and three upstream of the promoter (Figure 2.1). Recruitment of VP64 by individual gRNAs to CaRE3 and CaRE4 transactivated IL2RA to levels comparable to those resulting from T cell activation (Figure 2.2). In addition, RNA-seq of unstimulated HuT78 cells stably expressing dCas9-VP64 and gRNAs targeting the *IL2RA* TSS, CaRE3 or CaRE4 showed selective *IL2RA* induction, with transcripts originating at the TSS and no evidence of downstream intergenic or intronic transcripts (Figure 2.4). In sum, our unbiased transcriptional activation approach identified novel elements within the *IL2RA* super-enhancer where recruitment of a transcriptional activator is sufficient to induce IL2RA expression on resting cells.

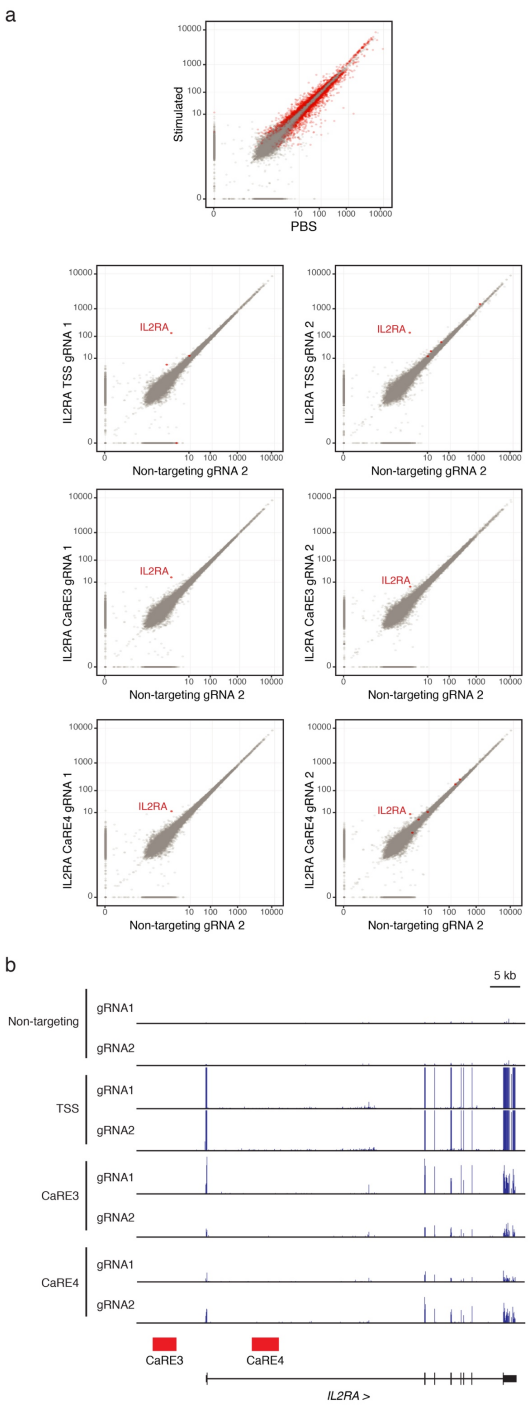


Figure 2.4. Activation of CaREs by CRISPRa specifically upregulates IL2RA. (a) Transcriptome comparison of HuT78 cells expressing dCas9-VP64 transduced with individual gRNAs targeting the IL2RA TSS, CaRE3 or CaRE4 versus a non-targeting sgRNA. Cells stimulated for 48 h with plate-bound CD3 and CD28 antibodies were also analyzed. Scatter plots show gene-level abundance estimates averaged over two replicates for each condition. Genes called as differentially expressed for each targeting guide, as described in Methods, are highlighted in red in their respective plot. For visualization purposes, transcripts per million (TPM) values

have been scaled by the transformation x -> x^(1/10). (b) RNA-Seq read coverage for IL2RA non-targeting, TSS, CaRE3 and CaRE4 gRNA samples. Tracks scaled to show maximum 55 reads.

We next investigated the biological significance of the *IL2RA* CaREs. Enhancers are often marked by signature histone modifications, chromatin accessibility and looping to promoters. We initially focused on chromatin state in human T cells because *IL2RA* regulation plays a critical role in T cells and because GWAS variants linked to autoimmunity reside within enhancers preferentially active in CD4+ T cells[148]. We analyzed data generated with HiChIP, a recently developed method that maps active enhancers based on H3K27 acetylation signature and simultaneously identifies long-range chromatin interactions[214]. All six *IL2RA* CaREs overlap with H3K27 acetylated elements that loop to the *IL2RA* TSS in primary human CD4+ T cell subsets, consistent with direct gene regulatory function (Fig. 2.5).

Although CaREs were mapped in a Jurkat T cell line, some corresponded with enhancer loops that are preferentially active in either Tregs or pro-inflammatory Th17 cells. Additional chromatin accessibility and histone modification mapping in human immune cell subsets corroborates that CaREs overlap with putative endogenous regulatory elements (Figures 2.5-2.7). We identified CaREs based on functional regulation of *IL2RA*, but chromatin conformation data suggests they also have the potential to regulate neighboring genes (Figure 2.8). Enhancer function could only be validated for a subset of CaREs with a heterologous reporter assay in Jurkats, suggesting that others may only exert gene regulatory functions in restricted cellular contexts (Fugrues 2.6, 2.7). Overall, we have shown that recruitment of VP64 identifies genetic elements with key, cell-type specific chromatin features of physiologic enhancers. While chromatin marks and three-dimensional configuration establish genomic features characteristically associated with potential enhancers, CRISPRa provides complementary functional data on the sufficiency of specific sites to serve as *cis*-regulatory elements for a target gene of interest.

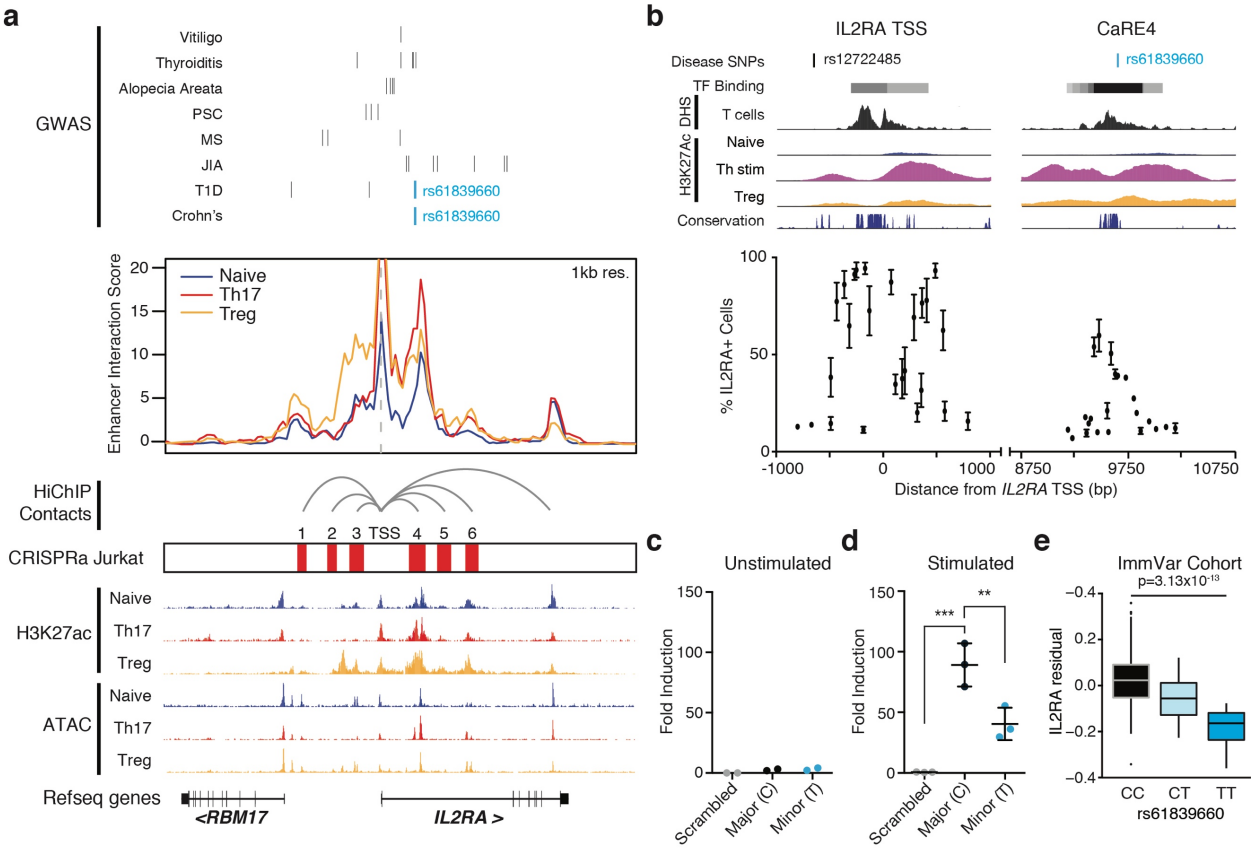


Figure 2.5. Identification of a stimulation-dependent disease-associated IL2RA enhancer. (**a**) The IL2RA locus showing fine mapped autoimmunity SNPs (T1D[149], Crohn's Disease[158], other associations[148]) from genome wide association studies (GWAS), HiChIP IL2RA promoter-enhancer loops and accessibility from primary human CD4+ T cells (naive, Th17 or Treg) and the overlap of these chromatin features with the IL2RA CRISPRa responsive elements (CaREs). (**b**) Zoomed-in view of the IL2RA transcriptional start site (TSS) and CaRE4 showing candidate autoimmunity SNPs in the region, DNAse hypersensitivity (DHS) and H3K27Ac from primary human T cells (Epigenome Roadmap) and vertebrate conservation (PhastCons 46-way). Sequences within these regions were targeted in Jurkat-dCas9-VP64 cells by nucleofecting gRNA expression plasmids. IL2RA expression on nucleofected (BFP+) cells was analyzed 48 hours post-nucleofection. (**c,d**) Jurkat cells were nucleofected with luciferase reporter constructs containing a minimal promoter downstream of the CaRE4 reference sequence, a scrambled sequence or CaRE4 with rs61839660 (SNP). Luciferase activity was measured 1 day later (c) or after 20 hours of anti-CD3/CD28 stimulation (**d**). All data is presented as mean +/- s.d. and are representative of at least two independent experiments. A one-way ANOVA with Holm Sidak's multiple comparisons test was used to compare scrambled and SNP enhancer sequences to reference sequence in the luciferase assays. **p≤0.01, and ***p≤0.001. (**e**) IL2RA QTL analysis on activated CD4+ T cells from 178 individuals in the ImmVar cohort reveals that the rs61839660 SNP is associated with reduced IL2RA expression. Shown here are the effects after conditioning on rs2476491.

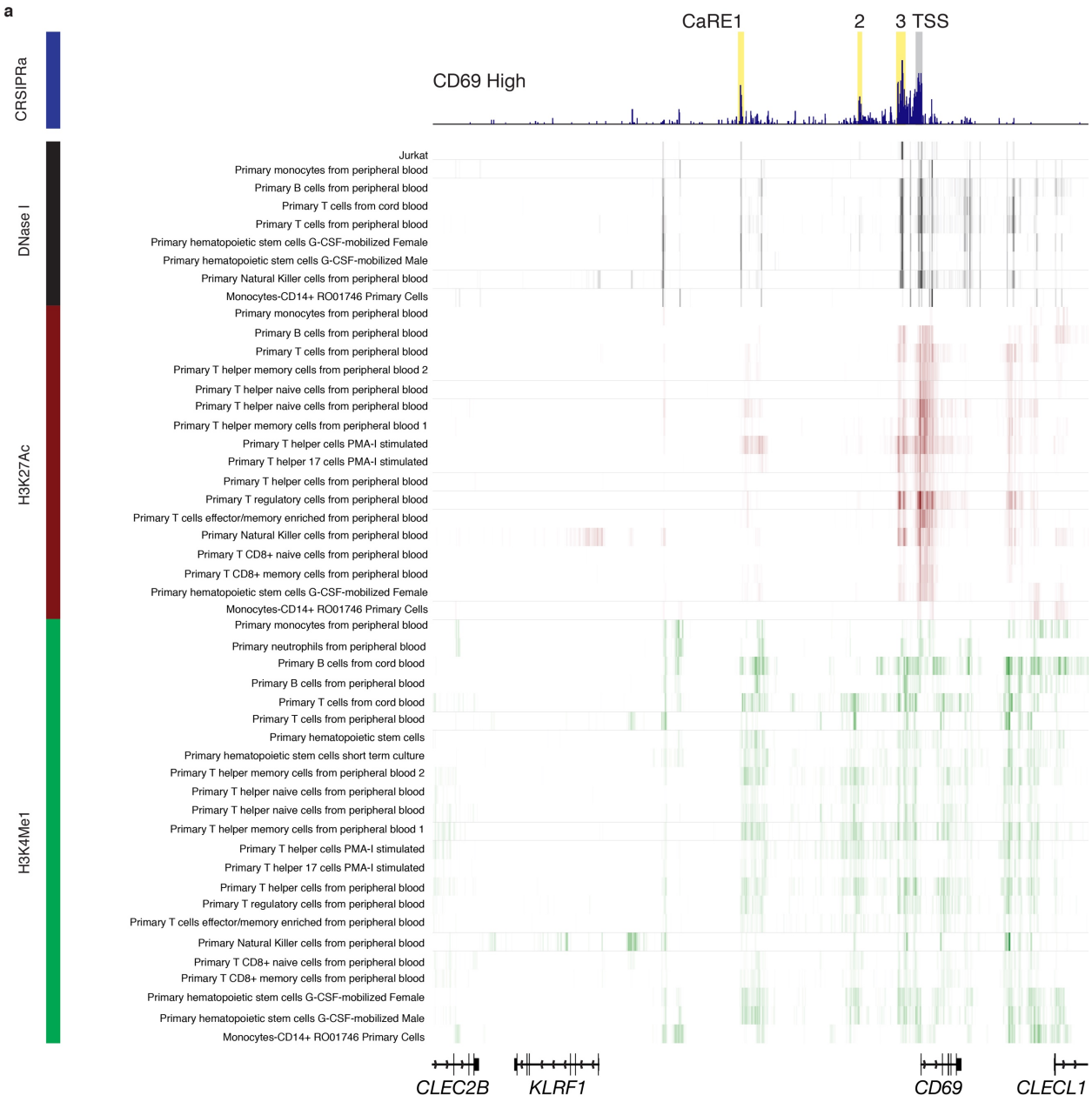

a
CRSIPRa
DNase I
H3K27Ac
H3K4Me1
CaRE1
2
3 TSS
CD69 High
Jurkat
Primary monocytes from peripheral blood
Primary B cells from peripheral blood
Primary T cells from cord blood
Primary T cells from peripheral blood
Primary hematopoietic stem cells G-CSF-mobilized Female
Primary hematopoietic stem cells G-CSF-mobilized Male
Primary Natural Killer cells from peripheral blood
Monocytes-CD14+ RO01746 Primary Cells
Primary monocytes from peripheral blood
Primary B cells from peripheral blood
Primary T cells from peripheral blood
Primary T helper memory cells from peripheral blood 2
Primary T helper naive cells from peripheral blood
Primary T helper naive cells from peripheral blood
Primary T helper memory cells from peripheral blood 1
Primary T helper cells PMA-I stimulated
Primary T helper 17 cells PMA-I stimulated
Primary T helper cells from peripheral blood
Primary T regulatory cells from peripheral blood
Primary T cells effector/memory enriched from peripheral blood
Primary Natural Killer cells from peripheral blood
Primary T CD8+ naive cells from peripheral blood
Primary T CD8+ memory cells from peripheral blood
Primary hematopoietic stem cells G-CSF-mobilized Female
Monocytes-CD14+ RO01746 Primary Cells
Primary monocytes from peripheral blood
Primary neutrophils from peripheral blood
Primary B cells from cord blood
Primary B cells from peripheral blood
Primary T cells from cord blood
Primary T cells from peripheral blood
Primary hematopoietic stem cells
Primary hematopoietic stem cells short term culture
Primary T helper memory cells from peripheral blood 2
Primary T helper naive cells from peripheral blood
Primary T helper naive cells from peripheral blood
Primary T helper memory cells from peripheral blood 1
Primary T helper cells PMA-I stimulated
Primary T helper 17 cells PMA-I stimulated
Primary T helper cells from peripheral blood
Primary T regulatory cells from peripheral blood
Primary T cells effector/memory enriched from peripheral blood
Primary Natural Killer cells from peripheral blood
Primary T CD8+ naive cells from peripheral blood
Primary T CD8+ memory cells from peripheral blood
Primary hematopoietic stem cells G-CSF-mobilized Female
Primary hematopoietic stem cells G-CSF-mobilized Male
Monocytes-CD14+ RO01746 Primary Cells
CLEC2B
KLRF1
CD69
CLECL1
b
CD69 CaREs
Normalized Luciferase Activity
30
20
10
0
****
****
Stim
Strand
Element
CaRE 1
CaRE 2
CaRE 3/CNS2

Figure 2.6. Chromatin features and enhancer activity of CD69 CaREs. (a) Results of the CD69 CRISPRa screen are overlapped with DNase I hypersensitivity, H3K27Ac, and H3K4me1 datasets from various primary human hematopoietic cell types. Data are shown for the indicated reference epigenomes from the Roadmap Epigenomics Project. Jurkat DNase HS data is from ENCODE. (b) Jurkat cells were nucleofected with luciferase reporter constructs containing sequences from CD69 CaREs upstream of a generic minimal promoter. 18 hours post-nucleofection, cells were split between a stimulation plate coated with anti-CD3/CD28 antibodies or a PBS control plate. Cells were lysed after 24 hours of stimulation, followed by measurement of luciferase activity. Data are presented as mean +/- s.d., n=4 biological replicates. Data are representative of two independent experiments. The dotted line represents the threshold of relevant luciferase activity defined as two times the value from a sequence-scrambled IL2RA CaRE4 control construct. Data were analyzed by one-way ANOVA followed by Dunnet's multiple comparisons test, comparing each construct this threshold. ****$p \leq 0.0001$

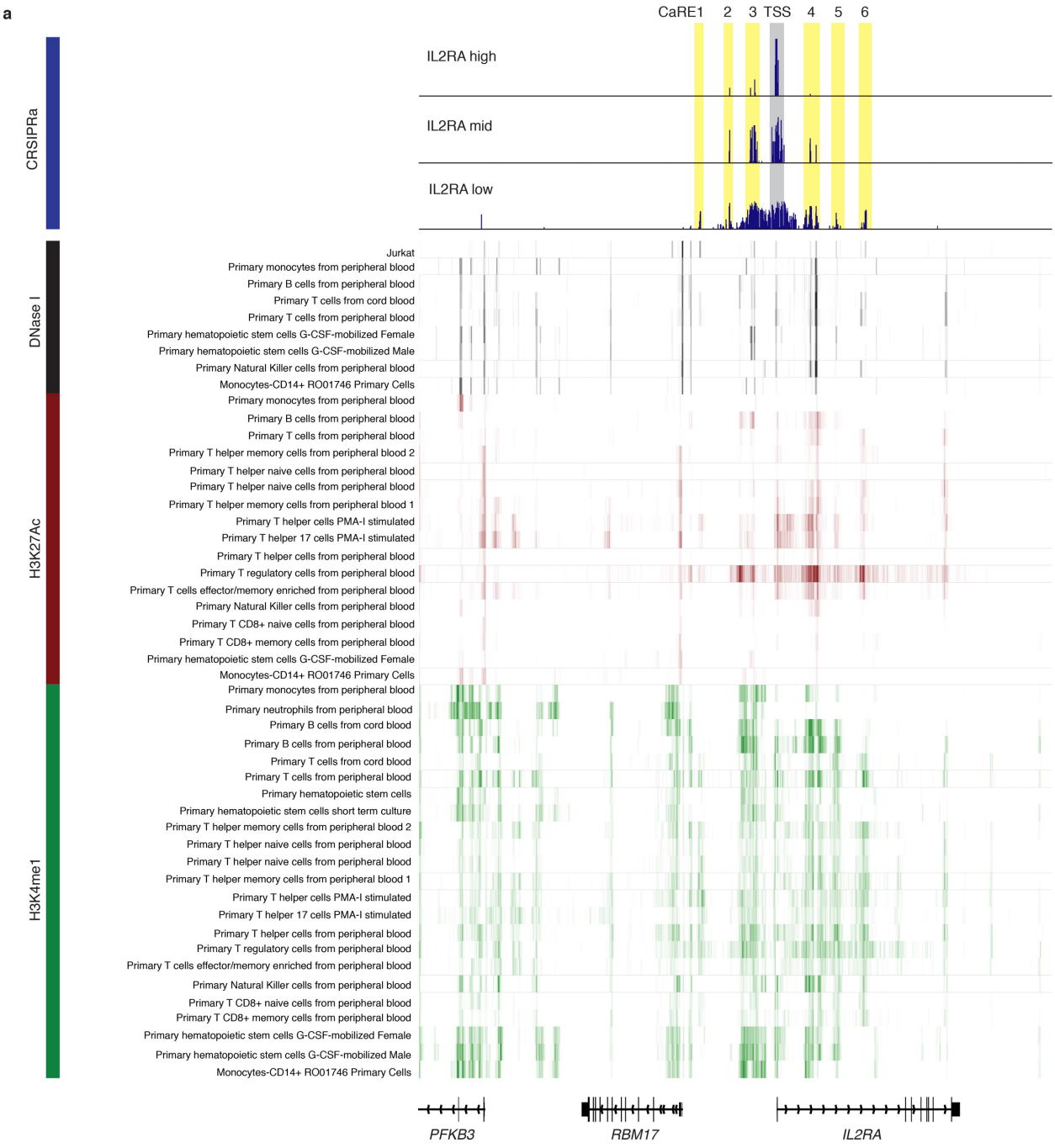

a
CaRE1
2
3
TSS
4
5
6
CRSIPRa
IL2RA high
IL2RA mid
IL2RA low
DNase I
H3K27Ac
H3K4me1
Jurkat
Primary monocytes from peripheral blood
Primary B cells from peripheral blood
Primary T cells from cord blood
Primary T cells from peripheral blood
Primary hematopoietic stem cells G-CSF-mobilized Female
Primary hematopoietic stem cells G-CSF-mobilized Male
Primary Natural Killer cells from peripheral blood
Monocytes-CD14+ RO01746 Primary Cells
Primary monocytes from peripheral blood
Primary B cells from peripheral blood
Primary T cells from peripheral blood
Primary T helper memory cells from peripheral blood 2
Primary T helper naive cells from peripheral blood
Primary T helper naive cells from peripheral blood
Primary T helper memory cells from peripheral blood 1
Primary T helper cells PMA-I stimulated
Primary T helper 17 cells PMA-I stimulated
Primary T helper cells from peripheral blood
Primary T regulatory cells from peripheral blood
Primary T cells effector/memory enriched from peripheral blood
Primary Natural Killer cells from peripheral blood
Primary T CD8+ naive cells from peripheral blood
Primary T CD8+ memory cells from peripheral blood
Primary hematopoietic stem cells G-CSF-mobilized Female
Monocytes-CD14+ RO01746 Primary Cells
Primary monocytes from peripheral blood
Primary neutrophils from peripheral blood
Primary B cells from cord blood
Primary B cells from peripheral blood
Primary T cells from cord blood
Primary T cells from peripheral blood
Primary hematopoietic stem cells
Primary hematopoietic stem cells short term culture
Primary T helper memory cells from peripheral blood 2
Primary T helper naive cells from peripheral blood
Primary T helper naive cells from peripheral blood
Primary T helper memory cells from peripheral blood 1
Primary T helper cells PMA-I stimulated
Primary T helper 17 cells PMA-I stimulated
Primary T helper cells from peripheral blood
Primary T regulatory cells from peripheral blood
Primary T cells effector/memory enriched from peripheral blood
Primary Natural Killer cells from peripheral blood
Primary T CD8+ naive cells from peripheral blood
Primary T CD8+ memory cells from peripheral blood
Primary hematopoietic stem cells G-CSF-mobilized Female
Primary hematopoietic stem cells G-CSF-mobilized Male
Monocytes-CD14+ RO01746 Primary Cells
PFKB3
RBM17
IL2RA
b
IL2RA CaREs
Normalized Luciferase Activity
80
60
40
20
0
****
****
Stim
Strand
Element
CaRE 1
CaRE 2
CaRE 3
CaRE 4
CaRE 5
CaRE 6

Figure 2.7. Chromatin features and enhancer activity of IL2RA CaREs. (a) Results of the IL2RA CRISPRa screen are overlapped with with DNase I hypersensitivity, H3K27Ac, and H3K4me1 datasets from primary human hematopoietic cell types. Data from the Roadmap Epigenomics Project. Jurkat DNase HS data is from ENCODE. (b) Jurkat cells were nucleofected with luciferase reporter constructs containing IL2RA CaRE sequences upstream of a generic minimal promoter. 18 hours post-nucleofection, cells were split to a stimulation plate coated with anti-CD3/CD28 or a PBS control plate. Cells were lysed 24 hours later and luciferase activity was measured. Data are presented as mean +/- s.d., n=4 biological replicates, representative of two independent experiments. Dotted line represents relevant luciferase activity defined as two times activity of sequence-scrambled IL2RA CaRE4 control. ****$p \leq 0.0001$ by one-way ANOVA followed by Dunnet's multiple comparisons test.

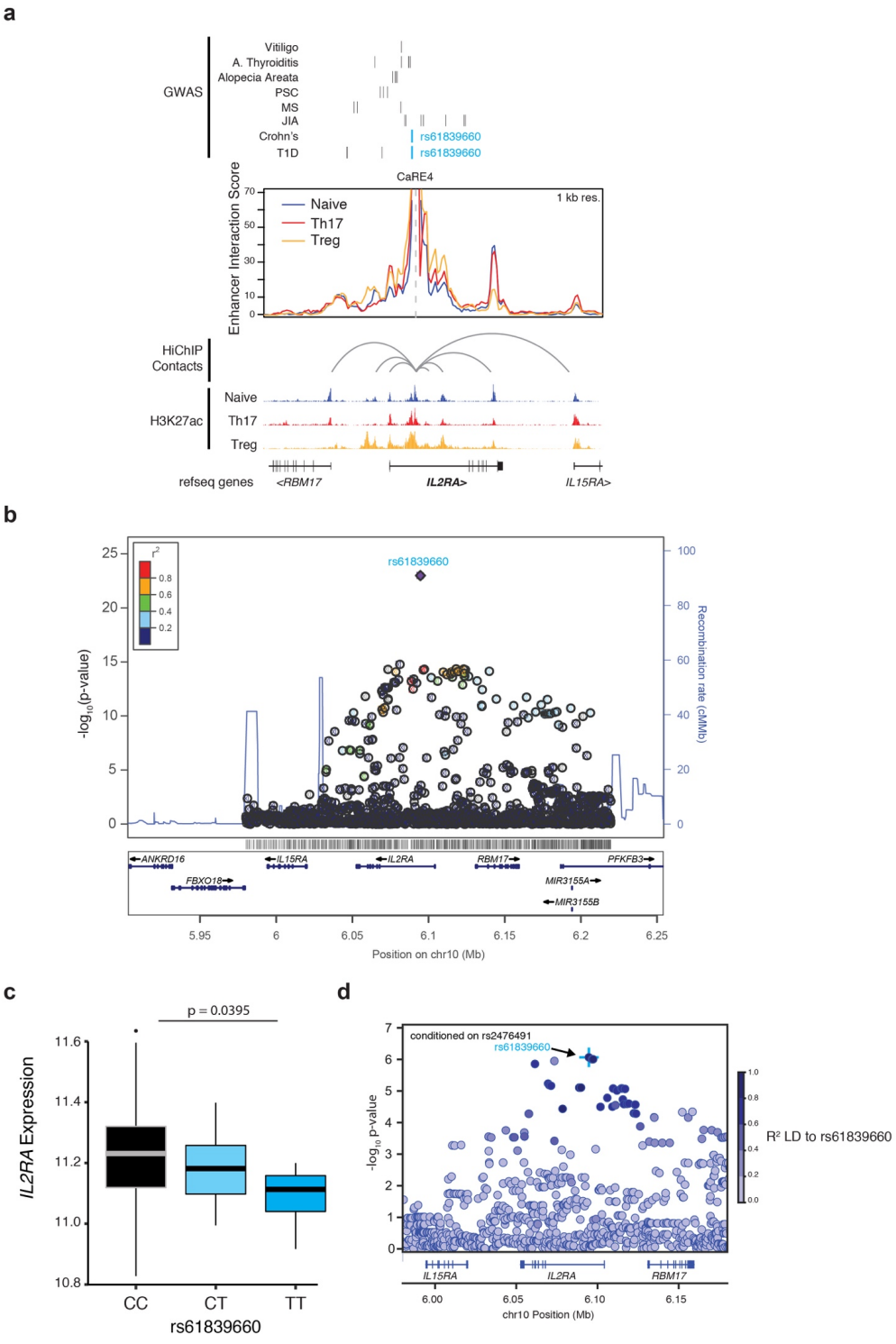


Figure 2.8. IL2RA CaRE4 harbors a risk variant linked to Crohn's Disease and reduced IL2RA expression in stimulated T cells. (a) HiChIP looping data anchored at IL2RA CaRE4 reveals that in addition to interacting with the IL2RA promoter, CaRE4 physically associates with other sites in the IL2RA locus as well as the promoters of IL15RA and RBM17. (b) IL2RA regional association plot. P-values of variants associated to Crohn's disease were taken from the inflammatory bowel diseases fine-mapping study (Huang et al., Nature, accepted), including all

SNPs and indels in the 1000 genomes phase 1 project. New SNPs and INDELs from the 1000 genomes phase 3 and the UK10K projects were not included in this figure, but none has high LD with rs61839660 that could explain the SNP association. Genes within 150kbp of IL2RA (from UCSC Genome Browser human GRCh37 assembly) were plotted. Figure generated using Locuszoom (http://locuszoom.org). (c) Reduced IL2RA levels in stimulated primary human T cells with the natural rs61839660 variant. The minor 'T' allele of rs61839660 is associated with reduced IL2RA levels in stimulated primary human T cells without conditioning. (d) rs61839660 is the mostly highly associated SNP with IL2RA levels in stimulated primary human T cells at 48hrs after conditioning on rs2476491.

Sequence variation in CaRE4 has been implicated in risk of human autoimmunity (Figure 2.5). The single nucleotide polymorphism (SNP) rs61839660, which resides in this element, has been convincingly statistically resolved to a single non-coding variant. This individual SNP accounts for the risk of inflammatory bowel disease (IBD) at the *IL2RA* locus[158,215] (Figure 2.8). Consistent with a critical and complex function in immune regulation, this same SNP paradoxically also contributes to protection from type 1 diabetes (T1D) [149,216].

Fine-mapping of CRISPRa responsiveness confirmed a functional role for the *cis*-regulatory sequence at the site of the rs61839660 autoimmunity variant. We expressed individual gRNAs surrounding the TSS and within CaRE4 to test their effects on *IL2RA* transactivation. The strongest transactivation in this region of CaRE4 was observed at a highly-conserved accessible region that harbors the autoimmunity SNP and is bound by multiple transcription factors between two peaks of H3K27Ac in stimulated T cells that harbors the autoimmunity SNP (Fugure 2.5). In an enhancer reporter assay, the conserved element within CaRE4 drove strong luciferase expression in Jurkat T cells, but only in response to stimulation (Figure 2.5). Introduction of rs61839660 diminished this stimulation-dependent enhancer function (Figure 2.5). These findings link this variant's role in disease to disruption of a stimulation-dependent *IL2RA* enhancer.

We next assessed if the rs61839660 SNP affects endogenous *IL2RA* gene regulation in primary human T cells. We analyzed transcript data collected from anti-CD3/CD28 stimulated CD4+ T cells from 178 genotyped people[217]. These data confirm that rs61839660 is a response

eQTL (reQTL), altering the transcriptional response to stimulation. Notably, the minor variant for rs61839660 is associated with reduced levels of *IL2RA* transcript in stimulated T cells (Figure 2.5, 2.8), confirming the functional effect of sequence variation in CaRE4 on human T cell gene regulation.

We then directly tested the *in vivo* effects of sequence variation in the enhancer in a mouse model. The *IL2RA* CaRE4 enhancer is highly conserved, which allowed us to use Cas9 genome editing to generate mice with the human autoimmune-associated SNP knocked in or with a 12 bp deletion at the site (12DEL) (Figure 2.9). Founders were backcrossed and bred to homozygosity. *In vivo* phenotyping of enhancer-edited mice revealed no evidence of overt immune dysregulation (Figure 2.10). T cell development was normal, with no differences in thymic cellularity or developmental stages (Figure 2.10). Furthermore, the enhancer does not appear to be required for IL2RA expression in Tregs at steady state as SNP and 12DEL Tregs had normal surface expression (Figures 2.9, 2.10).

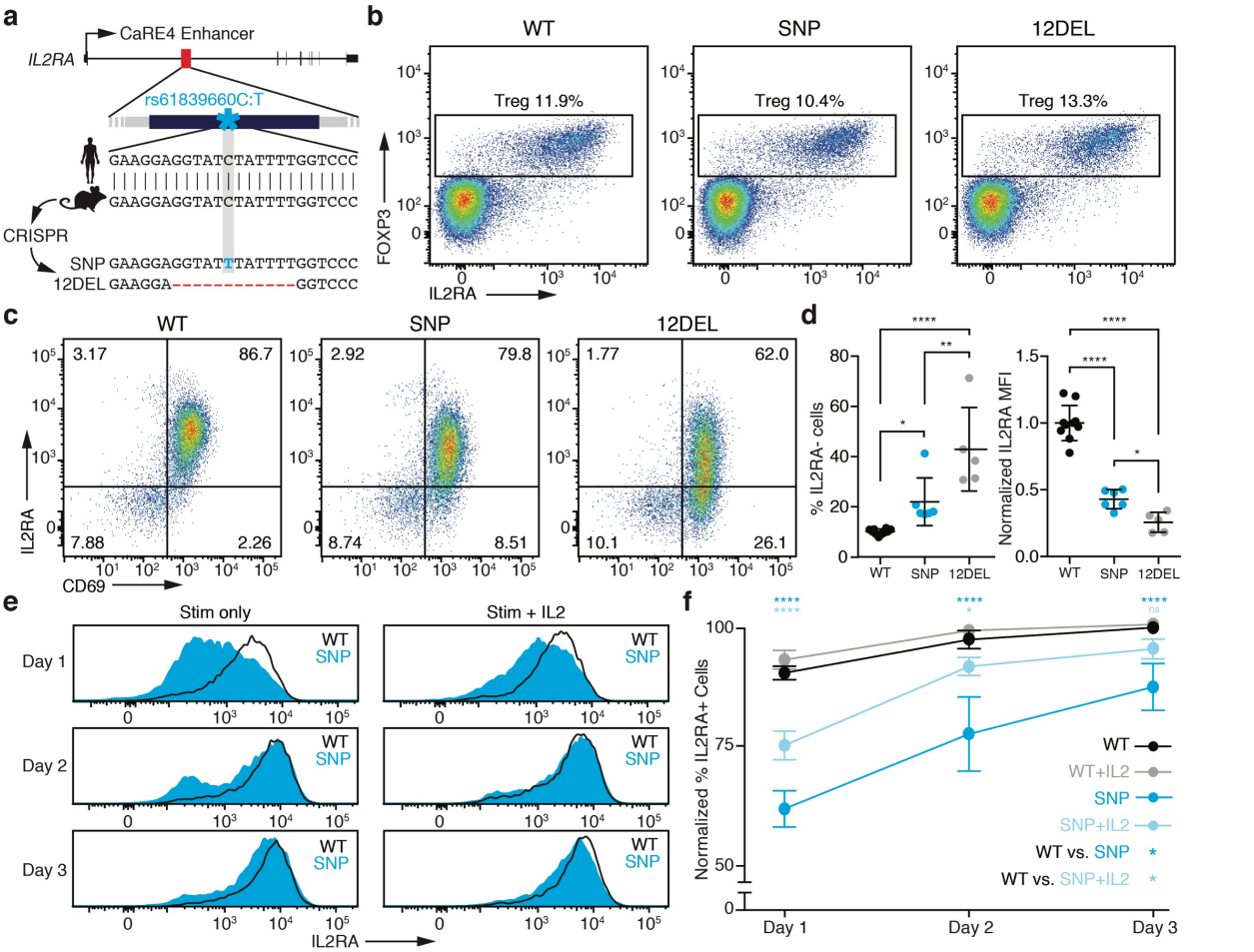


Figure 2.9. In vivo modeling of sequence variation in IL2RA enhancer. (**a**) Generation of rs61839660 knock-in (SNP) and 12 bp deletion (12DEL) B6 mice using CRISPR to edit the conserved IL2RA enhancer in zygotes. (**b**) Normal regulatory T cell (Treg) surface staining in peripheral lymph node (peri-LN) CD4+ T cells. (**c**) Surface staining of IL2RA and CD69 on naive T cells stimulated with plate-bound anti-CD3/CD28 antibodies for 1 day. (**d**) Quantification of percent IL2RA- cells and IL2RA MFI as in (c). Data in (d) derived from SNP (n=6), 12DEL (n=5) and WT littermate (n=10) mice. (**e**) 3-day time course of naive T cells isolated from spleen and peri-LN stimulated with plate-bound anti-CD3/CD28 antibodies alone (stim) or in combination with 50 U/ml IL2 (stim + IL2). (**f**) Quantification of flow cytometry data as in (e) normalized against Day 3 WT levels. Data in (f) derived from SNP (n=3) and WT littermate (n=3) mice. All data are presented as mean +/- s.d. and are representative of at least two independent experiments. *p≤0.05; **p≤0.01; ****p≤0.0001 by two-way ANOVA followed by Fisher's Least Significant Difference test.

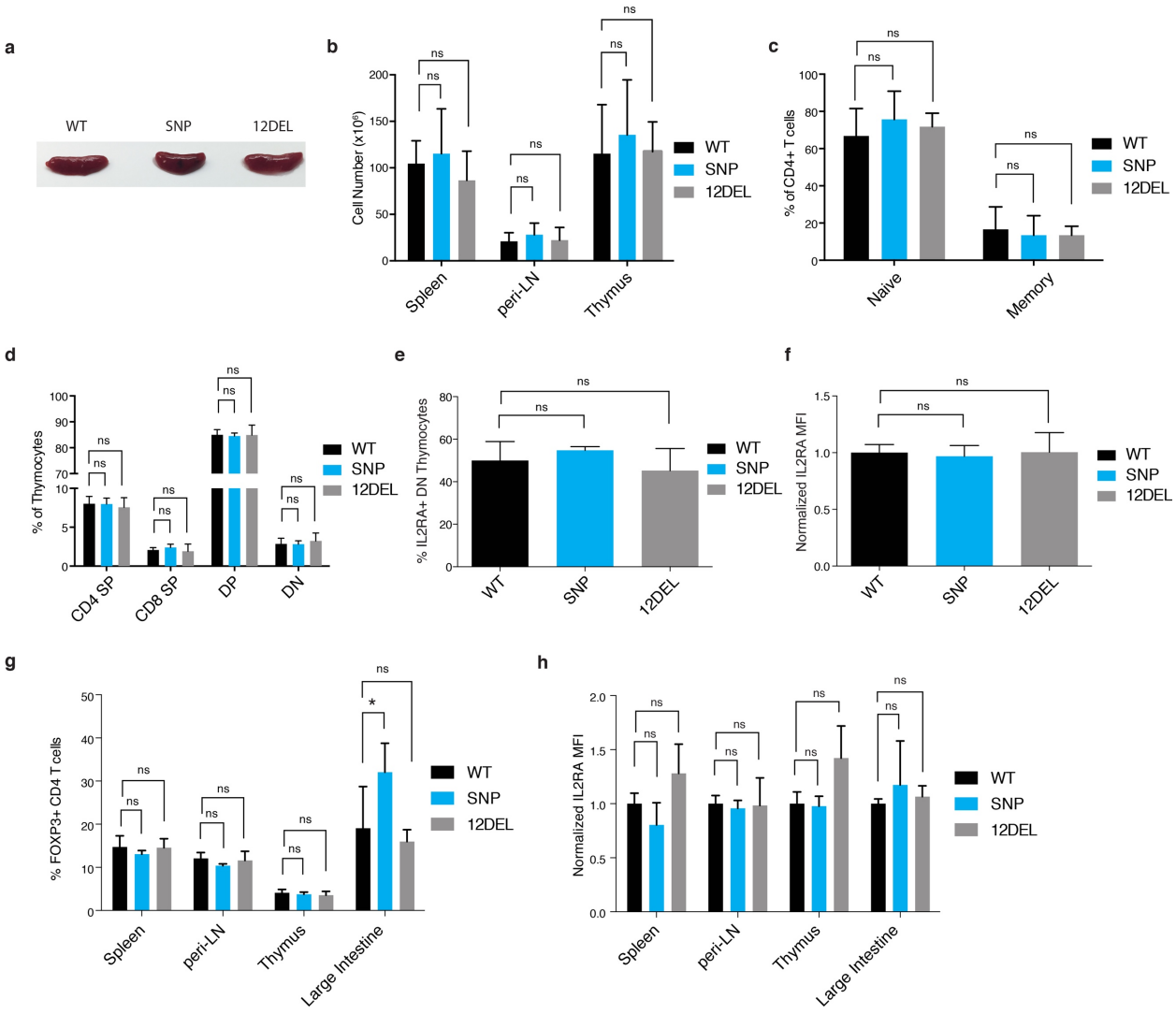


Figure 2.10. IL2RA enhancer-edited mice show no steady-state immune dysfunction. Enhancer-edited mice and littermate controls were immunophenotyped at 2-4 months of age. (a) Spleens from WT, SNP and 12DEL mice. (b) Total number of cells in spleen, peripheral lymph nodes (peri-LNs) and thymus. (c) Percentage of naive (CD4+CD62L+CD44-) and memory (CD4+CD62L-CD44+) CD4+ T cells isolated from spleen and peri-LNs. (d) Percentage of thymocytes in T cell developmental stages from the thymus. Data shown for CD4/CD8 single-positive (SP), double-positive (DP), and double-negative (DN) populations. (e,f) Quantification of percent IL2RA+ DN thymocytes and IL2RA MFI on IL2RA+ DN thymocytes. (g) Percent Tregs of CD4+ T cells in tissues of enhancer edited mice and littermate controls at 2-4 months of age. (h) Quantification of IL2RA surface staining (geometric mean fluorescence intensity, MFI) on FOXP3+ cells. All data are presented as mean +/- s.d. and are representative of at least two independent experiments. (a-c) Data are biological replicates of WT (n=7), SNP (n=4), and 12DEL (n=5) mice. (d-f) Data are biological replicates of WT (n=7), SNP (n=4), and 12DEL (n=5) mice. (g,h) Data are biological replicates of WT (n=6), SNP (n=4), and 12DEL (n=5) mice. A non-parametric one-way ANOVA (significance level 0.05) followed by Dunn's multiple comparison test was used to compare enhancer-edited mice to WT controls. *$p \leq 0.05$

Given the stimulation-dependent enhancer activity in human cells *in vitro*, we reasoned that the CaRE4 enhancer might regulate *IL2RA* induction on naïve CD4+ T cells following stimulation. We isolated naïve T cells (CD4+CD62L+CD44-) from edited and wild type (WT) mice and activated them *in vitro* with anti-CD3/CD28 antibodies. Remarkably, naïve T cells from both SNP and 12DEL mice had significantly reduced IL2RA surface expression compared to WT mice 24 hours post-activation (Figure 2.9). The deficit was more pronounced in 12DEL cells, but the SNP alone resulted in significant reduction of IL2RA levels (~50% of WT IL2RA, $p \leq 0.0001$ by one-way ANOVA and Fisher's LSD) (Figure 2.9). Reduced IL2RA levels were not due to a general defect in response to stimulation, as CD69 expression was induced to levels comparable to WT cells 24 hours post-stimulation (Figure 2.11).

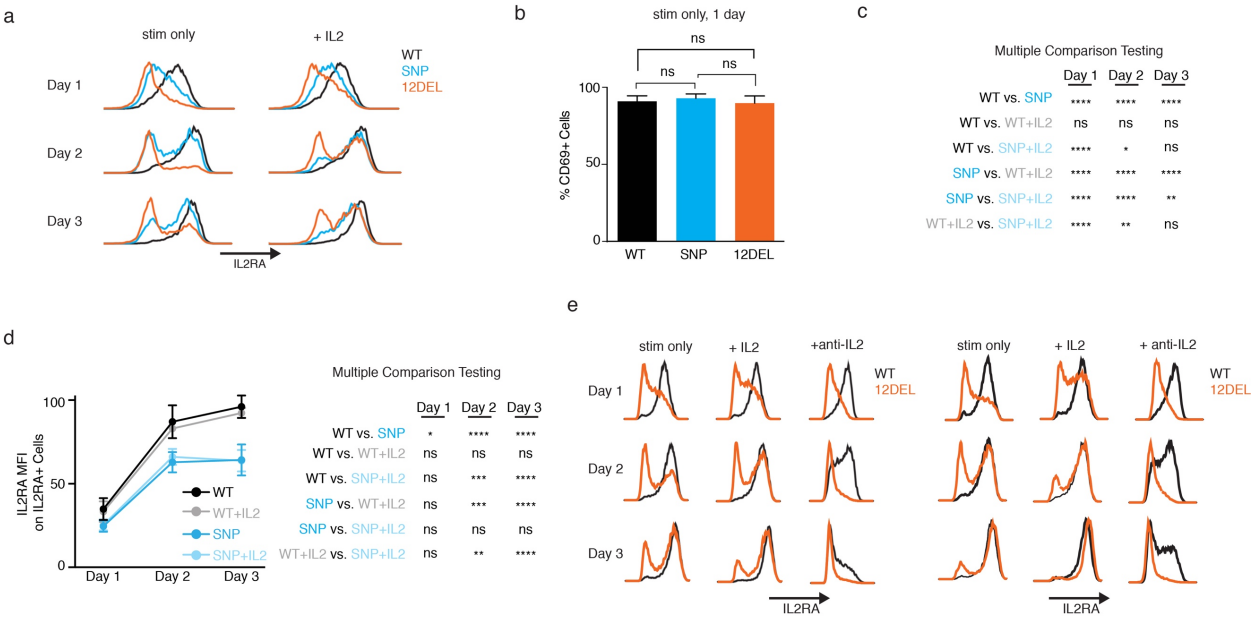


Figure 2.11. IL2RA induction in stimulated SNP and 12DEL T cells. (a) WT, SNP, and 12DEL cells with stim (anti-CD3/CD28) or stim + 50 U/ml IL2 over 3 days. (b) Percent CD69+ cells by surface levels on WT and enhancer edited cells 1 day after stimulation. (c) Statistical analysis using Fisher's Least Significant Difference (LSD) at each day of stimulation time course comparing WT and SNP naive T cells, with or without IL2. (d) IL2RA MFI on IL2RA+ T cells with stim or stim + 50 U/mL IL2 over 3 days. Table shows the Fisher's LSD statiscal analysis at each day of T cell stimulation time course. (e) IL2RA MFI on 12DEL naive T cells with stim, with stim + IL2 or

10 µg/mL anti-IL2 blocking antibody. Data displayed in (d) and (e) are each representative of two independent experiments. Data in (d) come from WT (n=3) and SNP (n=3) gender matched litter mate controls. All data are normalized to IL2RA MFI on WT stim only cells at Day 3. A two-way ANOVA with multiple comparisons testing followed by Fisher's Least Significant Difference test was used for statistical analysis. Data in (e) come from WT (n=2) and 12DEL (n=2) littermate controls. *p≤0.05, **p≤0.01, ***p≤0.001, ****p≤0.0001

Since disruption of CaRE4 did not ablate IL2RA expression, we asked whether mutant T cells were able to eventually recover IL2RA levels after stimulation. Indeed, three days post-stimulation, the percent of cells expressing IL2RA from SNP mice was much closer to that of WT cells than it had been at day one, although measurable defects in the percent of cells expressing IL2RA and intensity of expression persisted (Figures 2.9, 2.11). The disease-associated single nucleotide change within CaRE4 has subtle effects on final levels of IL2RA, but exerts a pronounced effect on the timing of induction.

Multiple extracellular signals can induce *IL2RA*. Having shown that the intronic enhancer alters response to T cell stimulation, we next tested if IL2, another critical signal that regulates *IL2RA*, could compensate for the disease variant's effect. IL2 promoted induction of IL2RA on the surface of cells from the SNP mice (Figure 2.9). Blocking IL2 in the culture using an anti-IL2 antibody had the opposite effect and severely impaired IL2RA expression on enhancer edited cells (Figure 2.11). Our results suggest that the human rs61839660 disease-associated variant impairs the function of an intronic enhancer that regulates IL2RA induction in response to anti-CD3/CD28 stimulation of conventional T cells. IL2 signals partially compensate for mutations in the stimulus-responsive enhancer, suggesting IL2 response is mediated through additional *cis*-regulatory elements.

We sought to determine the *in vivo* response of enhancer mutant cells to acute stimulation. We deleted the entire *IL2RA* enhancer on the autoimmune-prone NOD background to dissect subtle phenotypes that may be missed with the SNP and 12DEL mutations. This enhancer deletion

(EDEL) strain also had no obvious T cell phenotypes at steady state (Figure 2.12, 2.13). When EDEL mice were treated with anti-CD3 monoclonal antibody to stimulate T cells *in vivo*, we observed significant differences in IL2RA induction on conventional T cells (Figures 2.12, 2.13). In contrast, Tregs continued to express WT levels of IL2RA, although subtle changes in Treg contribution to the CD4+ population in the spleen were observed (Figure 2.12, 2.13). These findings confirm *in vivo* that the intronic enhancer controls the acute induction of IL2RA in conventional T cells by the TCR pathway.

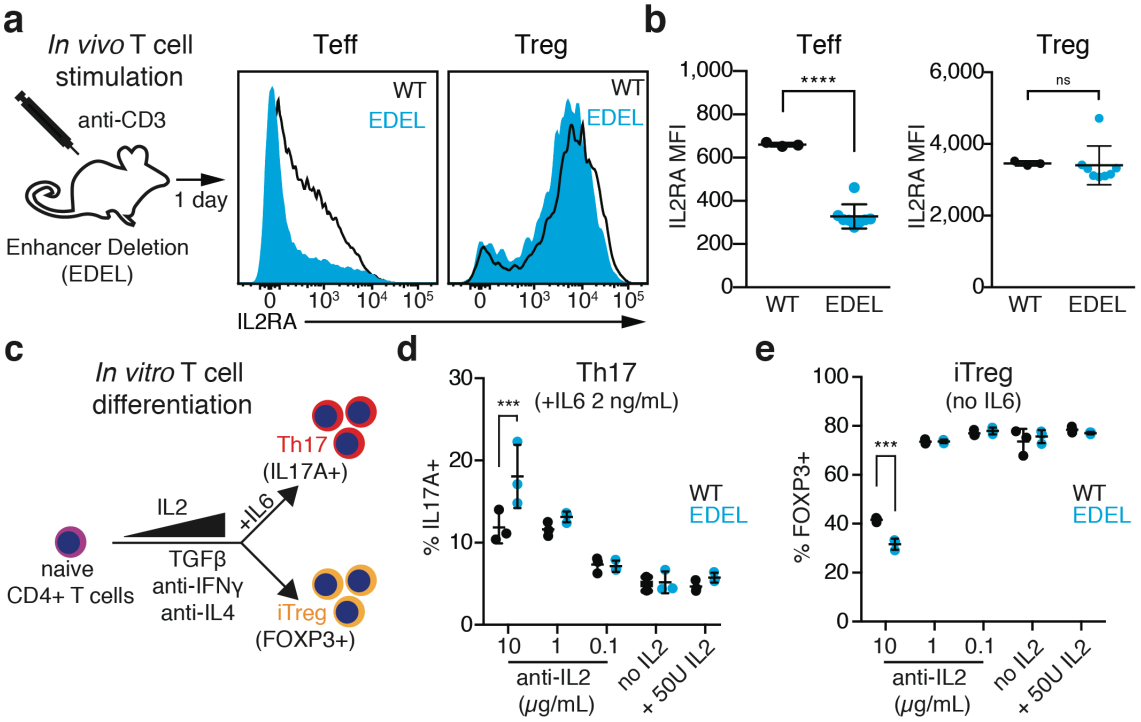


Figure 2.12. IL2RA enhancer controls IL2RA induction in response to TCR stimulation in vivo and skews T cell differentiation in vitro. (**a**) Wildtype (WT) and homozygous enhancer deletion (EDEL) non-obese diabetic mice were dosed with 50ug anti-CD3 antibody (Clone 145-2C11). IL2RA surface expression was assessed on CD4+FOXP3- (Teff) and CD4+FOXP3+ (Treg) T cells from spleen 1 day after treatment. (**b**) Quantification of IL2RA MFI on Teff and Treg from WT (n=3) and EDEL (n=8) littermate mice as in (a). Data are representative of two independent experiments. (**c**) Naive T cells from WT (n=3) and EDEL (n=3) mice were differentiated into Th17 and induced Tregs (iTreg) under various cytokine conditions. The outcome of the differentiation was assessed by measuring percent (**d**) IL17+ cells and (**e**) FOXP3+ cells. All data are presented as mean +/- s.d. ***p≤0.001; ****p≤0.0001 by one-way ANOVA followed by Holm Sidak's multiple comparisons test.

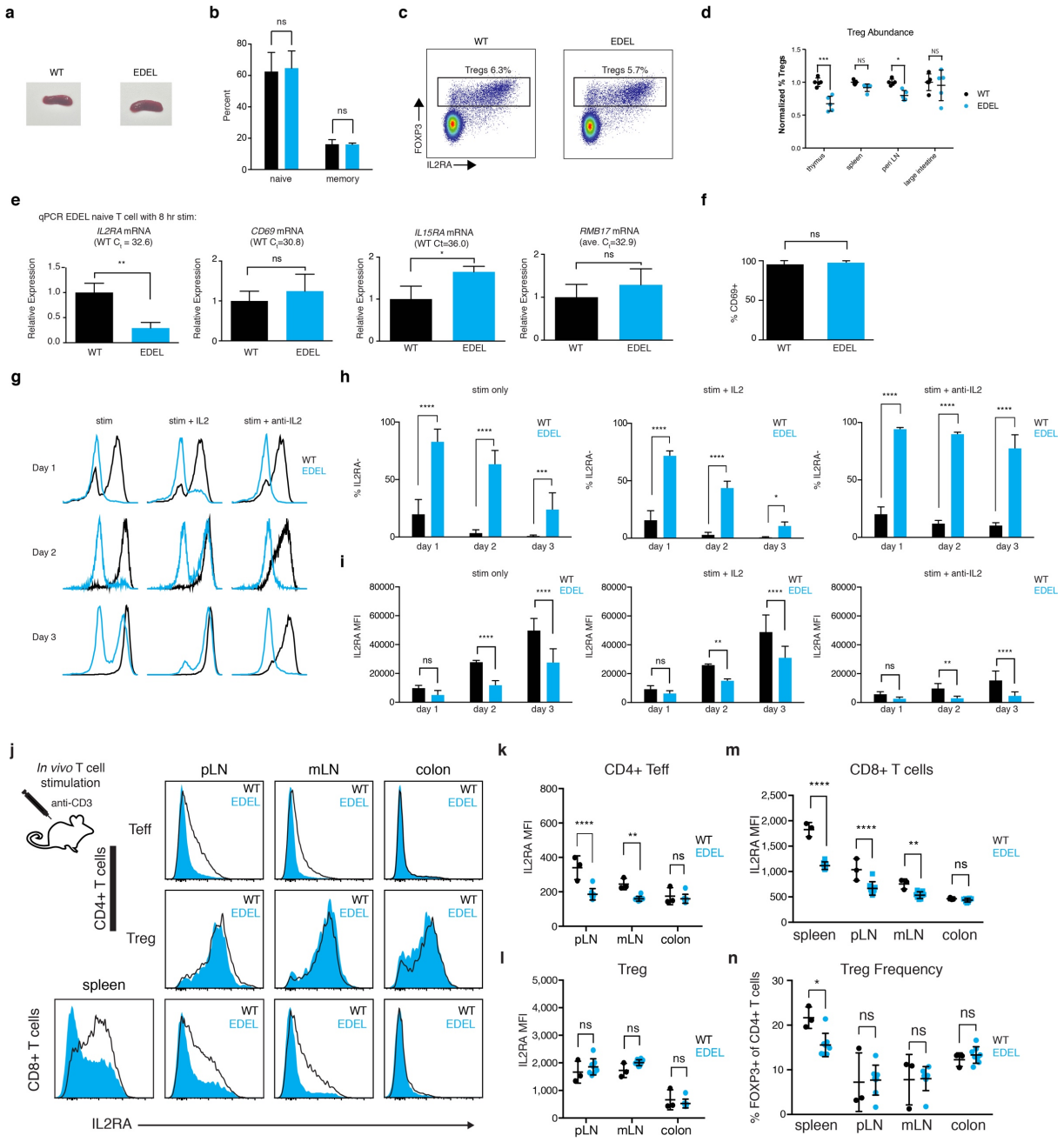


Figure 2.13. Characterization of IL2RA enhancer deletion (EDEL) on the NOD background. (a) Representative spleens from wild type (WT) and EDEL mice. (b) Naive (CD62L+CD44-) and memory (CD44+CD62L-) compositions of CD4+ T cells. (c) Representative lymph node staining showing Treg (CD4+FOXP3+) and Teff (CD4+FOXP3-) compartments. (d) Quantification of Treg abundance across multiple different tissues. (e) Since we did not uncover defects in steady state T cells, we isolated naive T cells and activated them in vitro with anti-CD3/CD28 antibodies. qPCR on naive T cells from WT or EDEL mice 8 hr after stimulation. Relative transcript levels for IL2RA, CD69 (control), IL15RA (adjacent gene), and RBM17 (adjacent gene) are shown. The average Ct value for each transcript on WT cells is shown. Data normalized to GAPDH expression. (f) CD69 protein surface expression on WT and EDEL naive T cells 1 day after stimulation with antibodies. (g) Representative flow plot of 3 day time course with naive T cells stim only (anti-

CD3/CD28), stim + 50 U/mL IL2 or stim + 10 µg/mL anti-IL2. (h) Quantification of percent IL2RA- cells in the time course. (i) Quantification of IL2RA MFI on IL2RA+ cells. Data were generated from two independent experiments with WT (n=6 and EDEL (n=6) mice. EDEL and WT mice were treated with 50 µg anti-CD3 to assess the in vivo T cell response to stimulation. Mice were sacrificed 1 day after treatment and IL2RA surface levels were checked by flow cytometry on T cells from spleen, peripheral lymph nodes (pLN), mesenteric lymph nodes (mLN) and colon. (j) Representative IL2RA MFI histograms on CD4+ and CD8+ T cells from various tissues. (k-m) Quantification of IL2RA MFI on CD4+FOXP3- Teff, CD4+FOXP3+ Treg and CD8+ T cells from different tissues. (n) Abundance of regulatory T cells in tissues following acute stimulation with anti-CD3 antibody. Data is representative of two experiments. EDEL (n=8) and WT (n=3) littermate mice were used for experiments. A two-way ANOVA with Holm-Sidak multiple comparisons test was used for statistical analysis.*$p \leq 0.05$,**$p \leq 0.01$,****$p \leq 0.0001$

We asked how impaired IL2RA induction could lead to cellular phenotypes associated with autoimmune disease. As naïve CD4+ T cells polarize in response to stimulation, IL2 provides a critical signal that restrains secretion of the pro-inflammatory IL17 in Th17 cells and promotes induction of Tregs[218]. We hypothesized that impaired *IL2RA* activation in the context of enhancer mutation could reduce IL2 signals to skew cells towards IL17 secretion and away from Treg induction. In Th17 and Treg polarizing cytokine conditions, we assessed IL17 secretion in cells treated with a range of IL2 concentrations or antibodies to inhibit IL2 signaling (Figure 2.12 and Figure 2.14). Strikingly, only under conditions where IL2 signals were limited by antibody, we found that *IL2RA* enhancer deletion increased the percentage of IL17 secreting cells, a hallmark of Crohn's disease pathology[219] (Figures 2.12, 2.14). In the absence of IL6, the enhancer deletion caused reduced induced Tregs to form only when IL2 concentrations were limited by antibody (Figures 2.12, 2.14).

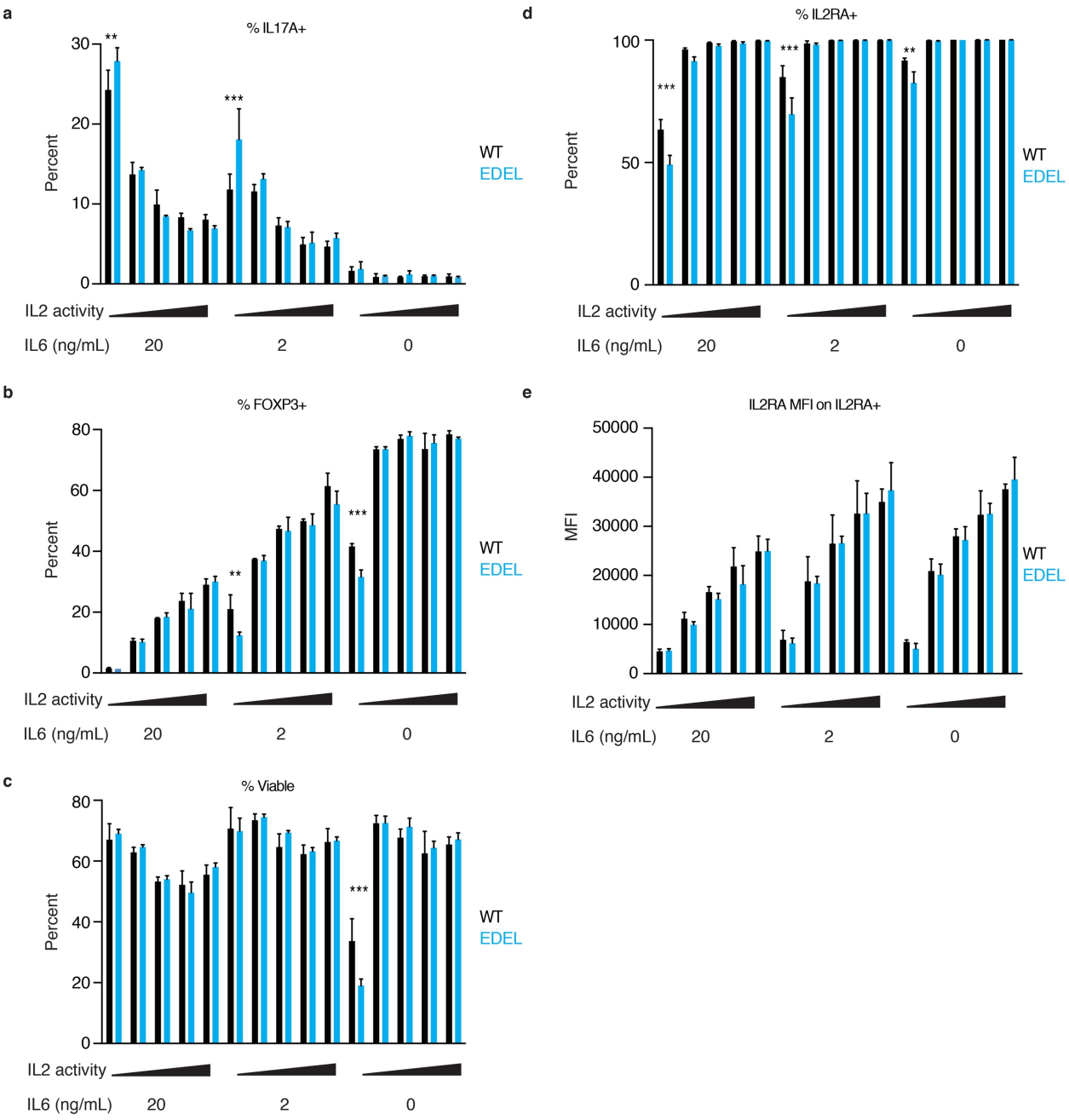


Figure 2.14. IL2RA enhancer deletion promotes Th17 and inhibits iTreg CD4+ T cell differentiation in IL2-limiting conditions. Naive CD4+ T cells were activated with anti-CD3/anti-CD28 and differentiated in the presence of TGFB, anti-IL4, anti-IFNg with high (20 ng/mL), medium (2 ng/mL) or no IL6. The IL2 activity was varied within each IL6 concentration by adding IL2 blocking antibody (10 ng/mL, 1 ng/mL or 0.01 ng/mL), no IL2 or 50 U/mL IL2. Five days after initial activation flow cytometry was used to assess (a) IL17A for Th17 differentiation, (b) FOXP3 for iTreg differentiation, (c) viability, and (d-e) IL2RA induction. Experiments were carried out with WT(n=3) and EDEL (n=3) age matched and sex matched littermate controls. A two-way ANOVA with Holm-Sidak method for multiple comparisons testing was used for statistcal analysis.**$p \leq 0.01$,***$p \leq 0.001$

Tregs are highly dependent on IL2 signaling for survival[210,220]. In Treg polarizing conditions where IL2 was limited, we found evidence of decreased cell viability, which was exacerbated by the enhancer mutation (Figure 2.14). Viability differences were not observed in other conditions tested, consistent with a selective effect on differentiating induced Tregs (Figure 2.14). Taken together, enhancer mutations that impair induction of the high affinity IL2RA receptor can disrupt critical IL2 signals and shift CD4+ T cell polarization towards a pro-inflammatory state.

Here we show that CRISPRa is a powerful approach for unbiased enhancer discovery at a target locus that can rapidly map functional enhancers without prior knowledge of their exact biological contexts. While we focused on immune-related genes, we anticipate this approach will have general utility as an enhancer discovery platform and can be used for functional annotation of the vast non-coding genomic space. Our functional enhancer mapping approach complements publicly available chromatin maps and enabled us to discover a disease-associated enhancer that controls the timing of gene expression. Some enhancers may be missed with CRISPRa, perhaps especially very distal regulatory elements, and further investigation will be required to determine the limits of the method. Candidate regulatory regions identified by CRISPRa should be validated with genome editing in addition to chromatin data.

Our findings reveal that human non-coding disease variants can shape the kinetics of genetically encoded responses. Human immune homeostasis depends not only on the level of IL2RA expression in Tregs, but also on proper dynamics of IL2RA induction in conventional T cells. Further study is needed to determine the full set of cell types and responses that are altered by the rs61839660 SNP. Ongoing clinical trials are testing the use of anti-CD3 antibodies and IL2 to treat various autoimmune and inflammatory conditions[221-223]. Understanding how genetic variation interacts with exogenous signals to regulate IL2RA induction may provide mechanistic

insights relevant to such therapies and inform patient stratification decisions. The data presented here critically identifies a functional context for a genetic autoimmunity risk factor, and suggests a new model of how common non-coding genetic variants control stimulation-responsive temporal gene regulation in health and disease.

METHODS

Cell culture

Cell culture was performed at 37℃ in a humidified atmosphere containing 5% CO_2. Jurkat cells (Clone E6-1) were obtained from the Berkeley Cell Culture Facility for CRISPRa experiments. HuT78 cells were a gift from Art Weiss (UCSF, San Francisco, CA). Jurkat and HuT78 cells were cultured is RPMI-1640 medium (Gibco) supplemented with 10% fetal bovine serum (FBS), 100 U/mL penicillin (Gibco), 100 μg/mL streptomycin (Gibco), and 1 mM sodium pyruvate (Gibco). Cell line identity for the CRISPRa screen and arrayed VP64 experiments was authenticated by short tandem repeat (STR) analysis and verified mycoplasma free using the MycoAlert Mycoplasma Detection Kit (Lonza).

Generation of dCas9-VP64 cells

Jurkat and HuT78 cells were transduced with a lentiviral dCas9-VP64-2A-GFP expression vector (Addgene 61422). Single GFP+ cells were sorted by FACS into the wells of a 96-well plate, and clones with bright uniform GFP expression were selected for use in future experiments.

Antibodies

All antibodies used in this study are listed in Supplementary File 1.

Primers

All primers for this study are listed in Supplementary File 2.

Guide RNAs

All gRNAs for this study are listed in Supplementary File 3.

Tiling gRNA library generation

For each gene of interest, the window of tiling gRNA libraries extended from 100 kb upstream of the transcription start site through 25 kb downstream of the end of the gene. The hg19 coordinates of the *CD69* library window were chr12:9,880,082-10,013,497. The hg19 coordinates of the *IL2RA* library window were chr10:6,027,657- 6,204,333.

gRNAs were designed against all NGG PAMs in the window, excluding sequences containing BstXI or BlpI/Bpu1102l cut sites. Each library contained 2,244 negative control gRNAs taken from the genome-scale CRISPRi/a libraries described in Gilbert et al[117]. Protospacer sequences flanked by restriction enzyme sites and PCR adaptors were synthesized by as pooled oligonucleotides by Agilent Technologies (Santa Clara, CA). Pooled gRNA libraries were then cloned into the lentiviral expression vector "pCRISPRia-v2" (Addgene #84832) as described in Horlbeck et al[224].

Tiling transcriptional activation screen

Protocols for the pooled lentiviral CRISPRa screens were adapted from Gilbert et al[117]. Lentivirus was produced by transfecting HEK293T with standard packaging vectors using *Trans*IT-LTI Transfection Reagent (Mirus, MIR 2306). Viral supernatant was harvested 48–72 hr following transfection, filtered through a 0.45 μm PES syringe filter, snap-frozen, and stored at -80 °C for future use.

Jurkat-dCas9-VP64 cells were infected with lentiviral gRNA libraries by resuspending cells at $2x10^6$ cells/mL in fresh media containing titered lentivirus and 4 μg/mL polybrene. Cells were spin-infected for 2 hours at 1000 xg, 33 °C, followed by resuspension in fresh media at 0.25-$0.5x10^6$ cells/mL. To limit the number of cells expressing multiple gRNAs, lentivirus was titered to infect only 10-20% of cells. Cells were cultured in media containing 0.75 or 1.5 μg/mL puromycin for days 2-5 post-infection to remove uninfected cells. The number of initially infected cells was at least 500x the number of gRNAs in the library, and at least this many cells were maintained throughout the course of the experiment.

7-10 days post-infection cells were sorted based on IL2RA or CD69 expression. Briefly, cells were resuspended in sterile sort buffer (PBS + 2% FBS) containing either IL2RA-PE or CD69-PE antibody at a 1:25 dilution. Cells were stained for 30 minutes on ice, washed twice with sort buffer, and passed through a 70 μm mesh. Cells were sorted into 4 bins based on IL2RA or CD69 expression using a BD Influx cell sorter. The total number of cells collected was at least 500x the number of gRNAs in the library. Additional unsorted cells totaling 500x the number of gRNAs in the library were collected at this time. Duplicate infections and sorts were performed for each library. Collected cells were centrifuged at 500 xg for 5 minutes, and cell pellets were stored at -80 °C until genomic DNA was isolated.

Genomic DNA was isolated from sorted cells using NucleoSpin Blood kits (Macherey-Nagel), or by Proteinase K digestion and isopropanol precipitation for samples with fewer than 10^6 cells. PCR was used to amplify gRNA cassettes with Illumina sequencing adapters and indexes as described in Kampmann, et al. [225]. Genomic DNA samples containing less than 10 μg of gDNA were loaded directly into PCR. For genomic DNA samples containing more than 10 μg of DNA, samples were

first digested for 18 hours with SbfI-HF (NEB) to liberate a ~500 bp fragment containing the gRNA cassette. The gRNA cassette was isolated by gel electrophoresis as described in Kampmann, et al. [225], and the DNA was then used for PCR. Custom PCR primers are listed in Supplementary File 2. Indexed samples were pooled and sequenced on an Illumina HiSeq-2500 with the custom sequencing primer 5'-gtgtgttttgagactataagtatcccttggagaaccaccttgttg-3'. Sequencing libraries were pooled proportional to the number of sorted cells in each sample. The target sequencing depth was 2,000 reads/gRNA in the library for unsorted "background" samples, and 10 reads/cell in sorted samples.

Screen data analysis

Sequence files were processed to remove low quality reads and reads lacking the gRNA constant region. Reads were then trimmed for the common sequence using the cutadapt script and the command "cutadapt -a GTTTAAGAGCTAAGCTG"[225]. Trimmed reads were then aligned against a database of the guide sequences using bowtie2 with option –norc[226]. gRNAs with fewer than 50 reads in either of the background samples were excluded from all downstream processing and data analysis. Read counts for each sample were then normalized to the total number of gRNA read counts in that sample. A pseudo-count of 1 was added to all normalized guide counts. gRNA enrichment was calculated as follows:

mean(log2(IL2RA_gate_rep1/IL2RA_background_rep1),
log2(IL2RA_gate_rep2/IL2RA_background_rep2))

The mean gRNA enrichment score was then calculated using a 5-gRNA sliding window and visualized with the Integrative Genomics Viewer (Broad Institute). Non-targeting control gRNAs and gRNAs that map perfectly to multiple sequences within the gRNA library window were

excluded from visualization. Raw and processed data for CRISPRa screen are included in Supplementary File 4. For each sorted cell population, normalized read counts for a given gRNA were well correlated between the two replicates of the screen (Figure 2.3).

Screen validation

For screen validation using individual gRNAs, gRNAs were cloned into the same expression plasmid used for the gRNA library. Lentivirus was produced as described above and used to infect Jurkat and HuT78 cells expressing dCas9-VP64. Expression of IL2RA and CD69 on infected cells was analyzed by flow cytometry. A complete list of gRNAs used in CRISPRa follow-up experiments is provided in Supplementary File 3.

Transient gRNA expression

For the transient gRNA expression experiment shown in Fig. 2b, gRNAs were cloned into the same expression plasmid used for the gRNA library. $2x10^5$ Jurkat-dCas9-VP64 cells were nucleofected with 1 μg of sgRNA plasmid with a 4-D Nucleofector (Lonza) using 20 μL of Nucleofector Buffer SE and nucleofection program CL-120. IL2RA expression on nucleofected cells was analyzed by flow cytometry 48 hours post-nucleofection.

HiChip and ATAC-seq Experiments

Human Subjects

This study was approved by the Stanford University Administrative Panels on Human Subjects in Medical Research, and written informed consent was obtained from all participants.

Primary T Cell Isolation

Normal donor human peripheral blood cells were obtained fresh from AllCells. CD4+ T cells were enriched from peripheral blood using the RosetteSep Human $CD4^+$ T Cell Enrichment Cocktail (StemCell Technology). For CD4+ T helper cell subtypes, Naïve T cells were sorted as $CD4^+IL2RA^-CD45RA^+$, Th17 cells were sorted as $CD4^+IL2RA^-CD45RA^-CCR6^+CXCR5^-$, and T_{reg} cells were sorted as $CD4^+IL2RA^+CD127^{lo}$. For HiChIP experiments, 500,000 - 1 million cells were sorted into RPMI + 10% FCS. For ATAC-seq experiments, 55,000 cells were sorted into RPMI + 10% FCS. Post-sort purities of > 95% were confirmed by flow cytometry for each sample.

HiChIP Protocol

The HiChIP protocol was performed as previously described (Mumbach et al., 2016) with the following modifications. For approximately 500 thousand to one million cells per T cell subtype per replicate, we performed two minutes of sonication, no Protein A bead preclearing, used 4 µg of H3K27ac antibody (Abcam ab4729), and captured the chromatin-antibody complex with 34 uL of Protein A beads (Thermo Fisher). Qubit quantification post ChIP ranged from 5 – 25 ng depending on the cell type and amount of starting material. The amount of Tn5 used and PCR cycles performed were based on the post ChIP Qubit amounts, as previously described (Mumbach et al., 2016). HiChIP samples were size selected by PAGE purification (300-700 bp) for effective paired-end tag mapping, and therefore were removed of all primer contamination which would contribute to recently reported "index switching" on the Illumina HiSeq 4000 sequencer[227].

HiChIP Data Processing and Virtual 4C Visualization

HiChIP paired-end reads were aligned to the hg19 genome using the HiC-Pro pipeline (Servant et al., 2015). Default settings were used to remove duplicates, assign reads to MboI restriction

fragments, filter for valid interactions, and generate binned interaction matrices. Virtual 4C profiles were generated from 1 kilobase resolution HiChIP interaction matrices by filtering the matrix for all bin-pairs in which one bin matched a single anchor bin of interest. Depth-normalization was achieved by scaling counts by the total number of filtered reads in each experiment. WashU Epigenome Browser sessions contained publically available H3K27ac ChIP-seq and ChromHMM data from the Roadmap Epigenome Project[151]. Browser shots from WashU track sessions were then included in virtual 4C representations. HiChIP data for the *IL2RA* promoter and *IL2RA* CaREs are provided in Supplementary File 5.

ATAC-seq

Cells were isolated and subjected to ATAC-seq as previously described[16]. Briefly, 55,000 cells were pelleted, resuspended in 50 µL lysis buffer (10mM Tris-HCl, pH 7.4, 3mM MgCl2, 10mM NaCl, 0.1% NP-40 (Igepal CA-630)), and immediately centrifuged at 500g for 10 min at 4^0C. The nuclei pellets were resuspended in 50 µL transposition buffer (25 µl 2X TD buffer, 22.5 µL dH_20, 2.5 µL Illumina Tn5 transposase), and incubated at 37^0C for 30 min. Transposed DNA was purified with MinElute PCR Purification Kit (Qiagen), and eluted in 10 µL EB buffer.

Luciferase Assays

CaRE sequences were synthesized as gBlock Gene Fragments (IDT, Coralville, IA, USA) or PCR amplified from Jurkat cell genomic DNA, then cloned into the Firefly Luciferase (Fluc) reporter vector pGL4.23 (Promega, Madison, WI, USA), upstream of a generic minimal promoter. The details of each construct are listed in Supplementary File 6. All plasmids have been deposited with Addgene (Addgene IDs 91835-91852).

For the experiments shown in Figure 2.5, each FLuc construct (700 ng) was electroporated with a Renilla luciferase plasmid (pGL4.74, 70 ng) into $5x10^5$ Jurkat cells using the 4-D Nucleofector, 20 µL Nucleofection buffer SE and nucleofection program CL-120. Cells were rested overnight and then activated using plate bound anti-CD3 (clone UCHT1, 10 µg/mL, TONBO Biosciences) and anti-CD28 (clone CD28.2, 10 µg/mL, TONBO Biosciences) antibodies for 22 hours. Luciferase expression was assessed using the Dual-Glo Luciferase Assay (Promega, Madison, WI, USA) on a 96 well plate luminometer. FLuc activity was normalized to Renilla activity and is reported as fold induction over empty pGL4.23 vector.

Experiments shown in Figures 2.6 and 2.7 were performed similarly, with the following modifications. Equimolar amounts of each FLuc construct (~600 ng/sample) were electroporated with a Renilla luciferase plasmid (pGL4.74, 150 ng) into $5x10^5$ Jurkat cells using the 4-D Nucleofector with 96-well shuttle, 20 µL Nucleofection buffer SE and nucleofection program CL-120. Cells were rested for 18 hours. Cells were then split between an antibody coated stimulation plate (as above) and a PBS control plate. Luciferase expression was assessed using the Dual-Glo Luciferase Assay (Promega, Madison, WI, USA) on a 96 well plate luminometer after 24 hours of stimulation. FLuc activity was then normalized to Renilla luciferase activity for each well.

RNA-Seq

HuT78 cells expressing dCas9-VP64 and individual gRNAs were grown in normal media. HuT78 cells expressing dCas9-VP64 but no gRNA were stimulated with plate-bound anti-CD3 and anti-CD28 or PBS control plates for 48 hours. Conditions for coating tissue culture plates with antibody were identical to those used for the luciferase reporter experiments. Cells were harvested and total RNA was isolated from samples using the RNeasy Mini Kit (QIAGEN, Cat. #74104) according to

the manufacturer's instructions with the following options: Cells were pelleted and re-suspended in RLT buffer with β-mercaptoethanol and homogenized using QIAshredder (QIAGEN, Cat. #79654). On-column DNase digestion was performed with the RNase-Free DNase Set (QIAGEN, Cat. #79254). RNA samples were analyzed with a NanoDrop spectrophotometer and all samples had a 260/280 and 260/230 ratio of 1.80 or higher. RNA integrity was measured with the AATI Fragment Analyzer (Advanced Analytical Technologies, Ankeny, Iowa), and all samples had an RNA Quality score (RQN) of 10.0. RNA concentration was measured using the Qubit RNA BR Assay (ThermoFisher Scientific, Cat. #Q10210). RNA-Seq libraries from biological duplicate samples were prepared using the TruSeq RNA Library Preparation Kit v2, Set A (Illumina, Cat. # RS-122-2001) and Set B (Cat. # RS-122-2002), following the Illumina TruSeq sample Preparation v2 Guide: Low Sample (LS) Protocol. The samples were pooled and sequenced with the Illumina HiSeq4000.

RNA-seq sequencing data was analyzed using kallisto[228] with reference to version 80 of the Ensembl annotation of the human genome, and taking 30 bootstrap samples to estimate inferential variance of the abundance estimates. The resulting abundance estimate data was then analyzed with sleuth[229] using a model with covariates indicating presence or absence of stimulation, each targeting guide, and a generic 'any guide' covariate to capture the non-specific effects of lentiviral transduction and gRNA expression (Supplementary File 7). Gene-level abundance estimates were computed by summing transcripts per million (TPM) estimates for transcripts for each gene. Wald tests were then performed at a gene level for each covariate with hits being called at a false-discovery rate of 10%. Data from these experiments are included in Supplementary Files 8 and 9.

To test stringently for expression from cryptic promoters induced by the CRISPRa treatment, reads in each sequencing sample were aligned to a repeat-masked version of the IL2RA region using HISAT2[230]. The resulting alignments were then tested with bedtools[231] for overlap with 1 kb regions centered around the IL2RA CaRE3 and CaRE4 guides, but no such reads were detected in any of CaRE-targeting samples.

We are aware of the recently reported "index switching" on the Illumina HiSeq 4000 sequencer and analysis of the non-targeted condition showed low to no *IL2RA* expression suggesting that this was not a problem[227].

ChIP-Seq and DNase-Seq data

ChIP-seq and DNase-Seq data were obtained from the Roadmap Epigenomics Project and ENCODE. Details of the samples plotted are in Supplementary File 10.

Generation of CRISPR Mouse Models

12DEL Mouse

12DEL mice were generated by the UCSF Mouse Genetic core (San Francisco, CA, USA) by microinjection of Cas9 ribonucleoprotein (PNA Bio, Newbury Park, CA, USA) into C57BL/6 zygotes. Briefly, Cas9 (50 ng/μL), mIL2RA-CaRE4 gRNA-1 (25 ng/μL), and ssDNA HDR template (50 ng/μL) were mixed in injection buffer (10 mM Tris, 0.1 mM EDTA) and incubated on ice for 10 minutes, as per the manufacturer's instructions. The mixture was microinjected into the cytoplasm of C57BL/6 single-cell zygotes isolated from super-ovulated females. We did not observe knock-in of the SNP in the progeny, but a single founder carried a 12 bp deletion in the *IL2RA* intronic enhancer. The 12DEL mouse line was established by backcrossing this founder for

at least one generation before breeding to homozygosity. gRNA sequence identical to gRNA used for SNP mouse generation, listed in Supplementary File 2.

SNP Mouse

SNP knock-in mice were generated by the Jackson Laboratory (Bar Harbor, ME, USA) by microinjection of gRNA and Cas9 mRNA. Briefly, Cas9 mRNA (100 ng/μL), mIL2RA-CaRE4 gRNA-1 (50 ng/μL), and ssDNA HDR template (100 ng/μL) were mixed and injected into C57BL/6 zygotes. Three founders with the knock-in SNP were identified by PCR amplicon sequencing and confirmed by sequencing of TOPO-cloned PCR products. The SNP mouse lines were established by backcrossing founders for at least one generation before breeding to homozygosity. gRNA and HDR template sequences were identical to those used to generate the 12DEL mouse line and are listed in Supplementary File 2.

EDEL Mouse

Enhancer deletion mice were generated by the Jackson Laboratory (Bar Harbor, ME, USA) by microinjection of gRNA and Cas9 mRNA. Briefly, Cas9 mRNA (100 ng/μL), mIL2RA-CaRE4 gRNAs 2-5 (50 ng/μL) were mixed and injected into NOD/ShiLtJ zygotes. Three founders with the enhancer deletion were identified by PCR amplicon size and confirmed by sequencing of TOPO-cloned PCR products. Immunophenotyping showed consistent phenotypes across three founders (not shown). Data shown is from one founder. The EDEL mouse lines were established by backcrossing founders for at least one generation before breeding to homozygosity. gRNAs are listed in Supplementary File 3.

Mouse Genotyping

All mice were genotyped by Sanger sequencing genomic DNA from proteinase K digested tail tissue. PCR amplification of the CaRE4 enhancer was carried out using HotStart Taq (Bioline USA Inc, Taunton, MA, USA) and primers (mIL2RA-CaRE4-F, mIL2RA-CaRE4-R) that span the edited site. PCR amplicons were then sequenced with the mIL2RA-CaRE4-F primer.

Mouse Experiments

Cell Preparation

Mice were maintained in the UCSF specific pathogen-free animal facility in accordance with guidelines established by the Institutional Animal Care and Use Committee and Laboratory Animal Resource Center. The experiments were not done in a blinded fashion and no randomization was performed. Experiments were done with animals aged between 2 to 4 months, unless otherwise noted. Wild type littermate controls were used in all experiments. The number of mice used for these experiments was sufficient given the consistency and magnitude of the observed phenotypes. Spleen, peripheral lymph nodes (peri-LNs), thymus and large intestine was collected from each mouse. Spleen, peri-LNs, and thymus were dissociated in 1x PBS with 2% FBS and 1 mM EDTA. The mixture was then passed through a 70 μm filter. ACK lysis was used to deplete red blood cells from splenocytes.

Lamina propria lymphocytes (LPLs) were isolated from the large using the lamina propria dissociation kit (Miltenyi Biotec, San Diego, CA, USA; Cat#130-097-410) and a gentleMACS dissociator (Miltenyi Biotec, San Diego, CA, USA).

Staining

All antibody stains were performed at a 1:100 dilution in 30 µL of 1x PBS. To pellet the cells centrifugation was done at 400xg for 5 minutes. For immunophenotyping, two million cells were stained per tissue sample. Cells were first stained with a viability dye at a 1:1000 dilution in 1x PBS for 20 minutes at 4 °C, then washed with EasySep Buffer (1x PBS, 2% FBS, 1 mM EDTA). Cells were then resuspended in the appropriate surface staining antibody cocktail and incubated for 30 minutes at 4 °C, then washed with 1x PBS. Cells were then fixed, permeabilized, and stained for transcription factors using the FOXP3 staining kit (eBioscience, Cat#00-5523-00) according to the manufacturer's instructions. Antibody staining panels are listed in Supplementary File 1.

IL2RA induction in anti-CD3/anti-CD28 stimulated naïve T cell

Naive T cells were isolated from spleen and lymph nodes with CD4+ negative selection (StemCell Technologies) followed by fluorescence activated cell sorting for CD4+IL2RA-CD44-CD62L+ cells. 80,000 to 100,000 cells we activated per well of a 96 well plate coated with 2ug/mL anti-CD3 and anti-CD28. For some conditions 10 µg/mL anti-IL2 blocking antibody or 50 U/mL IL2 was added. Cell analysis by flow cytometry was performed every day for 3 days.

***In Vivo* T cell Stimulation with anti-mouse CD3**

NOD EDEL (n=8) and WT littermate controls (n=3) between 3-4 months of age were used for these experiments. All mice were checked for diabetes and were found to be normoglycemic. Animals were anesthetized with isofluorane. 50 µg anti-CD3 was injected retro-orbitally in 100 µL volume. Mice were sacrificed 24 hrs later for isolation of spleen, peripheral lymph nodes (pLNs), mesenteric lymph nodes (mLN) and large intestine. Large intestine was processed using

a lamina dissociation kit described above. Cells were stained with antibodies for flow cytometry (Supplementary File 1).

Mouse T cell Differentiation

Naïve CD4+ T cells were isolated from spleen and lymph nodes of NOD EDEL mice and WT littermate controls with a CD4+ negative selection kit (StemCell Technologies, Vancouver, BC, CA; Cat#19752) followed by fluorescence-activated cell sorting of CD4+IL2RA-CD44-CD62L+ cells. WT (n=3) and EDEL (n=3) were matched by age and sex. All mice were normoglycemic. Naïve CD4+ T cells were activated in 96-well plates coated with anti-CD3 (1 μg/mL) and anti-CD28 (0.5 μg/mL) monoclonal antibodies at 37°C for ~72 h at an initial density of 0.75×10^6 cells/well and then were allowed to rest for 48-72 h in medium. For T helper polarizing conditions, 0 ng/mL (Treg), 2 ng/mL (TH17 low) or 20 ng/mL (TH17 high) recombinant mouse IL-6 (PeproTech), 5 ng/mL recombinant human TGF-β (PeproTech), 10 μg/mL anti-IL-4 (11B11; BioXCell), and 10 μg/mL anti-IFN-γ (XMG1.2; BioXCell) were added to the culture throughout the 5 days. IL2 concentration was additionally titrated across all polarizing conditions: 10 μg/mL, 1 μg/mL, 0.1 ug/mL anti-IL2 (S4B6; eBioscience) or 20 ng/mL recombinant IL2 (National Cancer Institute). All cultures used DMEM high glucose media supplemented with 10% FCS, pyruvate, nonessential amino acids, MEM vitamins, L-arginine, L-asparagine, L-glutamine, folic acid, β-mercaptoethanol, penicillin and streptomycin. After 5 days, cells were stimulated for 4 hours with a cocktail of PMA and Ionomycin, in the presence of Brefeldin A. Cells were then stained for CD4, IL2RA, IL17A, RORgt and FOXP3 (Supplementary File 1).

Analysis of gene regulatory effects of naturally occurring rs61839660 variant in human T-cells

Nanostring expression data from the ImmVar cohort was reanalyzed and normalized as described in (Ye et al., Science 2014)[216]. The data was residualized against biological covariates and the first 6 principal components. As the SNP is found in appreciable frequencies only in individuals of European American background, only Caucasian samples (n=178) were analyzed. Linear regression was applied to the data using a modified scikit-learn f_regression function. The data was residualized against the primary associated SNP and second round of regression was performed in the same manner. The FDR was calculated using R qvalue package. Visualization was performed with matplotlib, seaborn, and ggplot2 packages. Figure 2.5 shows the SNP effects conditioned on rs2476491. The effects without conditioning are shown in Figure 2.8.

Data Availability

Data from the CRISPRa screen (accession number: GSE98178; Fig. 2.1) and RNA-sequencing (accession number: GSE98178; Figure 2.4) are available at NCBI GEO. HiChIP interaction matrices for the *IL2RA* promoter and *IL2RA* CaREs (Figure 2.5, 2.8) are available in Supplementary File 5.

CHAPTER 3. A Large CRISPR-Induced Bystander Mutation in *Il2ra* Causes Immune Dysregulation

A persistent concern with CRISPR-Cas9 gene editing has been the potential to generate mutations at off-target genomic sites. While CRISPR engineering mice to delete a ~360 bp intronic enhancer, here we discovered a founder line that had marked immune dysregulation caused by a 24 kb tandem duplication of the sequence adjacent to the on-target deletion. Our results suggest unintended repair of on-target genomic cuts can cause pathogenic "bystander" mutations that escape detection by routine targeted genotyping assays.

INTRODUCTION

CRISPR-Cas9 genome engineering is employed widely to generate targeted *in vitro* and *in vivo* genetic modifications[232]. The Cas9 nuclease can be programmed to target specific genome sequences via a short guide RNA. Although unintended genome alterations have been mitigated by recent technical advances[26,35,52,233,234], they remain a concern, especially for therapeutic applications of CRISPR. To date, attention has been focused on "off target" editing in which Cas9 nuclease activity is directed towards genomic sites, other than the target, with varying degrees of homology to the guide RNA. Here we demonstrate that "bystander" mutations – unintended mutations neighboring the "on-target" cut site - must also be considered.

RESULTS

CRISPR-Cas9 deletion of *Il2ra* enhancer

One advantage of genome editing over RNA knock-down approaches is that non-coding sequences can be modified, which enables studies of non-coding variants commonly associated with human

disease risk. We recently identified a conserved autoimmunity-associated *IL2RA* intronic enhancer that controls the timing of gene expression in response to T cell stimulation[235]. To study its *in vivo* function, we used CRISPR to engineer non-obese diabetic (NOD) mice with deletion of this enhancer (EDEL). We successfully generated EDEL founder lines by targeting Cas9 to cut on either side of the ~360 bp enhancer (Figure 3.1). Genomic PCR and targeted Sanger sequencing confirmed that approximately 360-370 bp was deleted at the enhancer site in multiple founders (Figure 3.1, 3.2). Three of the founders were backcrossed to wildtype NOD animals at least one generation before breeding the enhancer deletion to homozygosity for experimentation.

Surprisingly, immunophenotyping revealed a marked systemic difference in one line of mice. Unlike the other two characterized lines, homozygous EDEL progeny from the third founder line had hallmark features of a lymphoproliferative disorder, including variable splenomegaly, increased cellularity and higher percentages of memory T cells (Figure 3.1, 3.2). Despite the phenotypic differences among the lines, the on-target enhancer deletion only differed by a few nucleotides at the margins of the deletion (Figure 3.2). The evolutionarily conserved DNA sequence at the site was deleted in all three lines and the line with more severe phenotype had a slightly smaller deletion, suggesting that the genotyped sequence differences directly at the deletion site did not explain observed differences in immune regulation (Figure 3.2). The more severe immune phenotype persisted in progeny with the enhancer deletion from the affected line, even after an additional round of backcrossing and multiple generations of breeding, suggesting a mutation in close genomic proximity to the on-target deletion site rather than an unlinked off-target effect. Taken together our data suggested the presence of an additional mutation linked to the *Il2ra* enhancer deletion in this immune dysregulated founder line (IDFL).

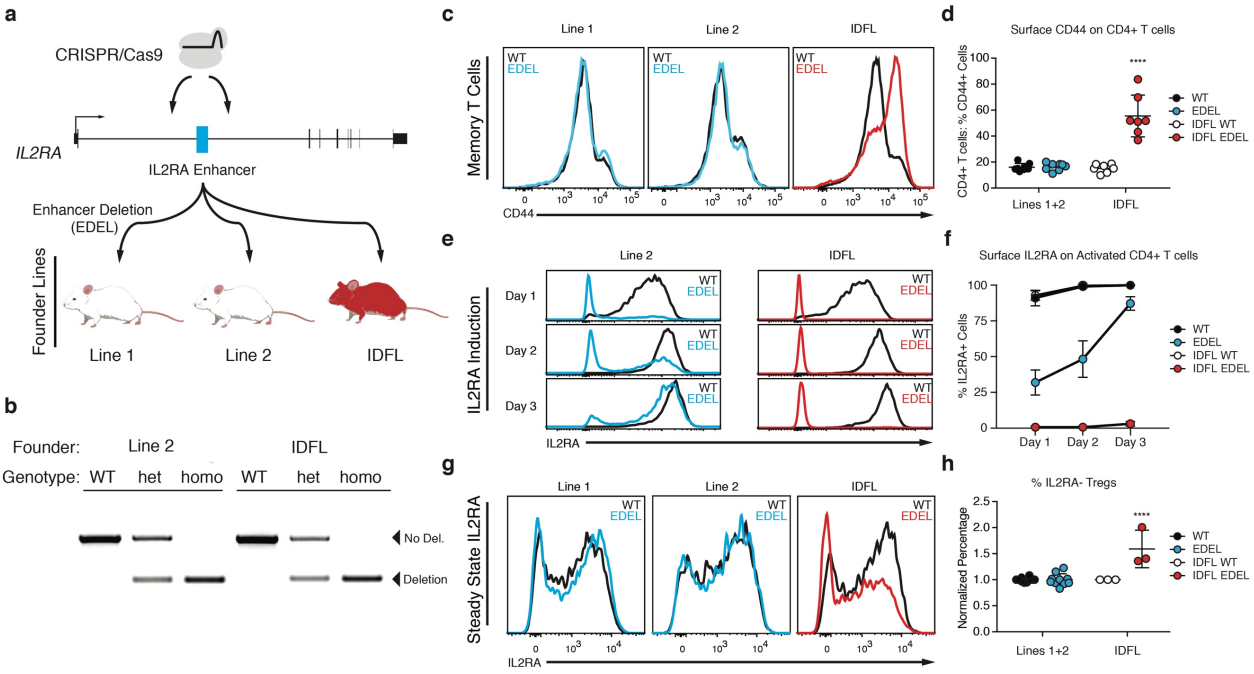


Figure 3.1. Immune dysregulation in a founder line of CRISPR-engineered *Il2ra* enhancer deletion mice. a, CRISPR-engineered *Il2ra* enhancer deletion (EDEL) founder lines that were bred for immunophenotyping. **b**, Genomic DNA PCR to genotype the *Il2ra* enhancer deletion in animals from Line 2 and the immune dysregulated founder line (IDFL). **c**, Representative CD44 surface staining on CD4+ T cells isolated from spleens of wild-type (WT) and EDEL mice from different founder lines. **d**, Quantification of percent CD44+ cells from (c) (Lines 1 and 2: WT n=8, EDEL n=7; IDFL: WT n=8, EDEL n=7). **e**, Representative induction of IL2RA surface expression on naïve CD4+ T cells (CD4+IL2RA-CD44-) activated with anti-CD3/CD28 antibodies. **f**, Quantification of percent IL2RA+ cells from (e) (Line 2: WT n=4, EDEL n=4; IDFL: WT n=4, EDEL n=4). **g**, Representative IL2RA surface expression on FOXP3+CD4+ T cells (Tregs) from spleen of different founders. **h**, Quantification of percent IL2RA- Tregs from (g) (Lines 1 and 2: WT n=10, EDEL n=10; IDFL: WT n=3, EDEL n=3). Panels (d) and (h) include data from Line 1 and 2 animals previously published[235]. All data are presented as mean +/- s.d. and are representative of at least two independent experiments.****P ≤ 0.001 by two-way ANOVA with Dunnett's multiple comparisons test.

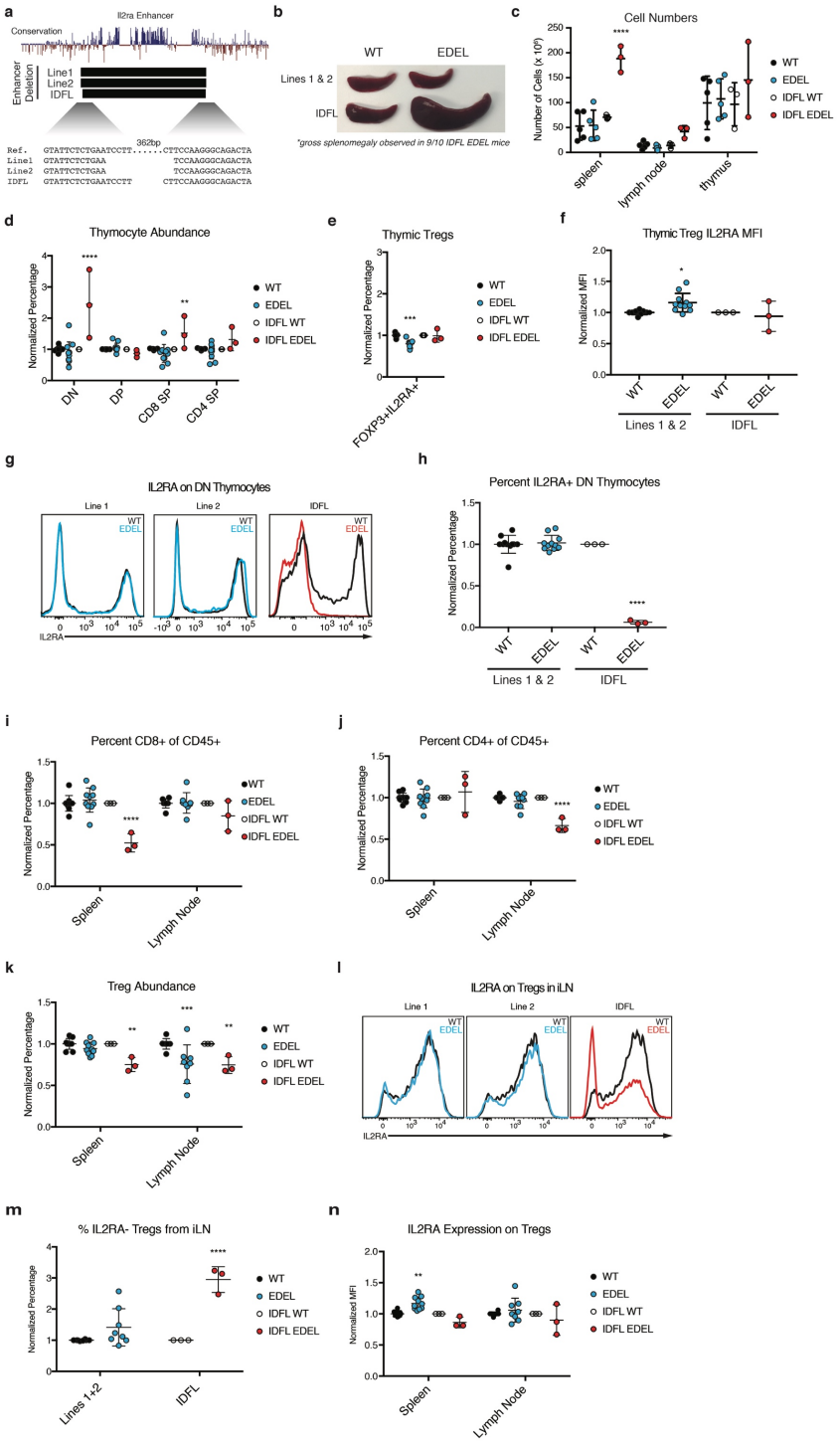


Figure 3.2. Characterization of enhancer deletion founder lines. a, Genetic deletion of the conserved Il2ra enhancer showing the nucleotide sequence at the deletion breakpoints. b, Representative spleens from WT and EDEL mice derived from different founder lines. c, Cell counts from spleen, inguinal lymph nodes and thymus. d, Normalized percentage of double negative (DN), double positive (DP), CD8 SP and CD4 SP thymocytes of live CD45+ thymocytes (Lines 1 and 2: WT n=11, EDEL n=11; IDFL: WT n=3, EDEL n=3). e, Normalized percentage of

FOXP3+IL2RA+ mature regulatory T cells (Tregs) of CD4 SP thymocytes (Lines 1 and 2: WT n=11, EDEL n=11; IDFL: WT n=3, EDEL n=3). f, Normalized IL2RA MFI on mature FOXP3+IL2RA+ Tregs (Lines 1 and 2: WT n=11, EDEL n=11; IDFL: WT n=3, EDEL n=3). g, Representative IL2RA surface expression on DN thymocytes from different founders. h, Normalized percentage of IL2RA+ DN cells of live CD45+ thymocytes. i, Normalized CD8+ percentage of live CD45+ cells in spleen and inguinal lymph nodes. j, Normalized CD4+ percentage of live CD45+ cells in peripheral lymphoid organs. k, Normalized percentage of FOXP3+ Tregs of CD4+ T cells in peripheral lymphoid organs. l, Representative IL2RA surface expression of FOXP3+ Tregs. m, Normalized percentage of IL2RA- FOXP3+ Tregs in peripheral lymphoid organs. n, Normalized IL2RA MFI on FOXP3+IL2RA+ Tregs in peripheral lymphoid organs. Data for Lines 1 and 2 includes animals for which immunophenotyping was previously published (Simeonov et al. Nature. 2017). Data in (c) derived from Lines 1 and 2: WT n=2, EDEL n=2 and IDFL: WT n=3, EDEL n=3. Data in (d,e,f,h) derived from Lines 1 and 2: WT n=11, EDEL n=11 and IDFL: WT n=3, EDEL n=3. Data in (m,n) derived from Lines 1 and 2: WT n=10, EDEL n=8 and IDFL: WT n=3, EDEL n=3. All data are presented as mean +/- s.d. and are representative of at least two independent experiments. *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.001 by two-way ANOVA with Dunnett's multiple comparisons test.

To determine the molecular and cellular effects of the linked mutation in the IDFL mice, we analyzed IL2RA expression. Double negative (DN) thymocytes from IDFL mice had marked loss of IL2RA expression, whereas DN thymocytes from EDEL mice of other founder lines had normal IL2RA expression (Figure 3.2). Mature CD4+ effector T cells (Teffs) normally upregulate IL2RA to their surface after activation. Strikingly, *in vitro* stimulated IDFL Teffs failed to express IL2RA on their surface (Figures 3.1, 3.3). This was in contrast to the other EDEL lines, which showed delayed but not ablated induction of IL2RA following stimulation of naïve T cells[235]. We also examined FOXP3+ regulatory T cells (Tregs), which constitutively express high levels of IL2RA and require it for their survival. Across lymphoid tissues there was an increased percentage of FOXP3+IL2RA- Tregs in IDFL mice compared to other EDEL lines (Figures 3.1, 3.2). *In vitro* and *in vivo* regulatory T cell differentiation were impaired (Figures 3.4). Interestingly, a subset of T cells, including some Tregs, did express high levels of IL2RA. An *Il2ra* null mutation would be expected to ablate expression across cell types. Instead we find that the linked mutation has effects on IL2RA expression that vary among cells, with a subset of T cells selectively maintaining IL2RA expression.

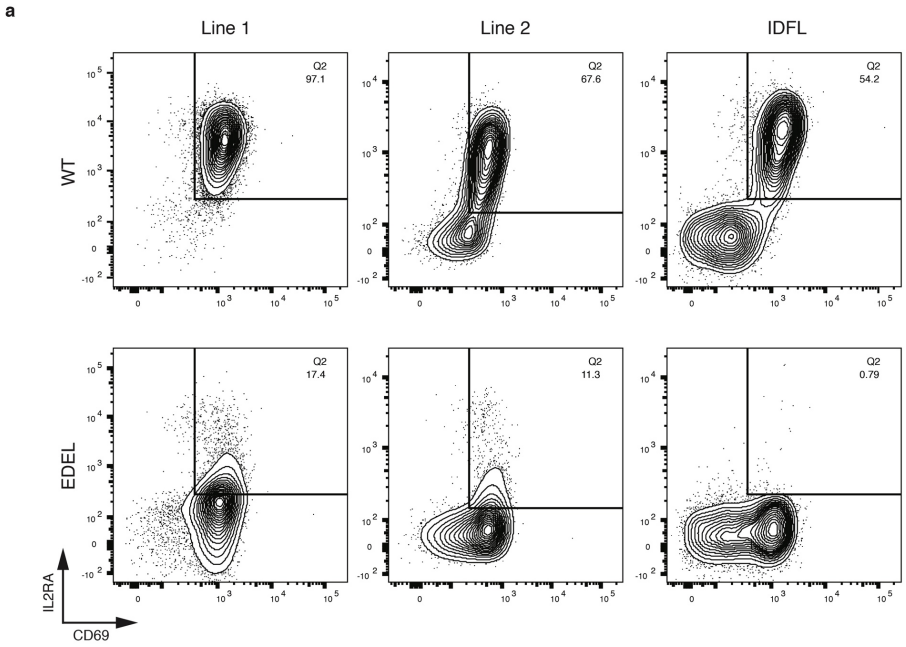


Figure 3.3. IL2RA Surface Expression on Activated Naïve CD4+ T cells. a, IL2RA and CD69 surface expression on naïve CD4+ T cells (CD4+IL2RA-CD62L+CD44-) 24 hours after stimulation with anti-CD3/CD28 plate bound antibodies shown.

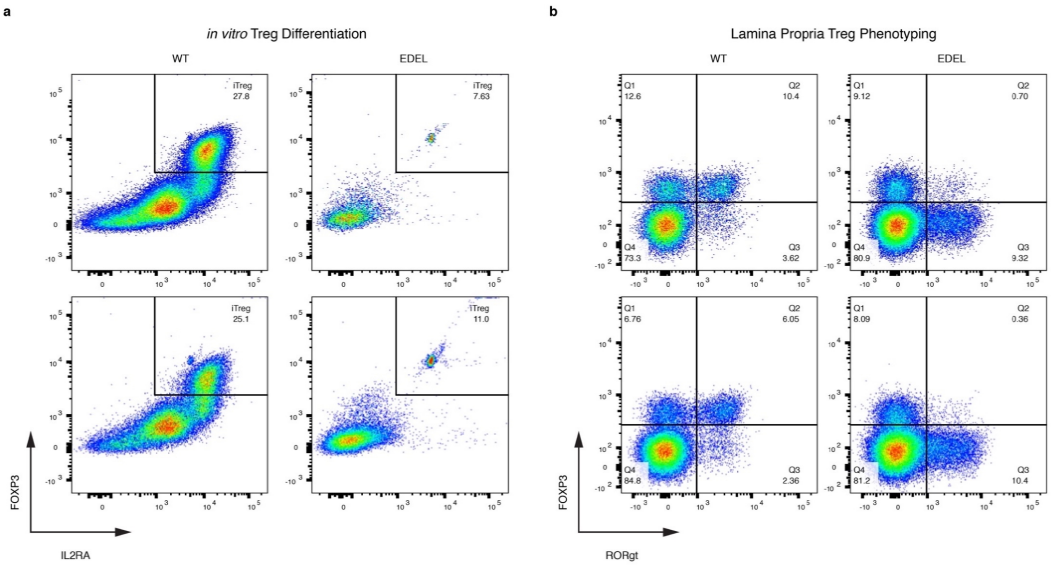


Figure 3.4. Regulatory T cell characterization in vitro and in vivo. a, Expression of FOXP3 and IL2RA in IDFL naïve CD4+ T cells (CD4+IL2RA-CD62L+CD44-) in vitro differentiated to become regulatory T cell (Treg) for 3 days. b, FOXP3 and RORgt staining of IDFL CD4+ T cells (live CD45+TCRb+CD4+) from large intestine lamina propria. Data shown is from independent biological replicates and is derived from a single experiment.

Identification of a bystander mutation

To identify the mutation causing marked immune dysregulation, we sequenced the whole genomes of EDEL mice from the immune dysregulated founder line and from one of the other founder lines (Figure 3.5). We looked for a causative IDFL mutation both at the *IL2ra* locus and throughout the genome. Consistent with the observed genetic linkage with the enhancer deletion, we discovered a large structural mutation in the *Il2ra* locus that was unique to the IDFL genome. Careful analysis of the read pileups revealed a 24 kb block of DNA with elevated coverage in the IDFL genome compared to adjacent sequences, consistent with an increase in copy number (Figure 3.5). Paired end reads at the breakpoint implied a tandem duplication, which we confirmed by genomic PCR and Sanger sequencing (Figure 3.5). The duplicated sequence starts immediately downstream of the deleted *Il2ra* enhancer and spans the remainder of the first intron, the second *Il2ra* exon and most of the second intron. This unexpected structural mutation tightly linked to the intended on-target edit is a "bystander" mutation that causes marked immune dysregulation.

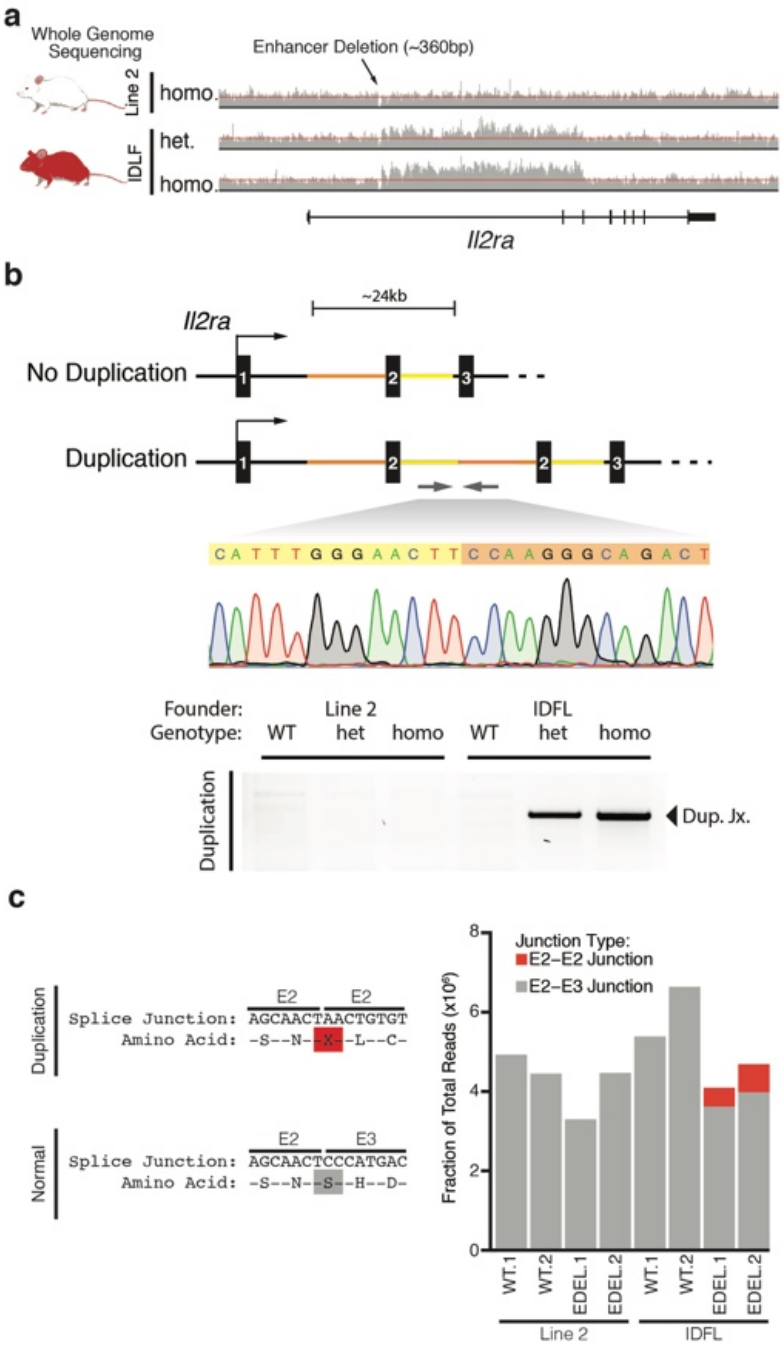
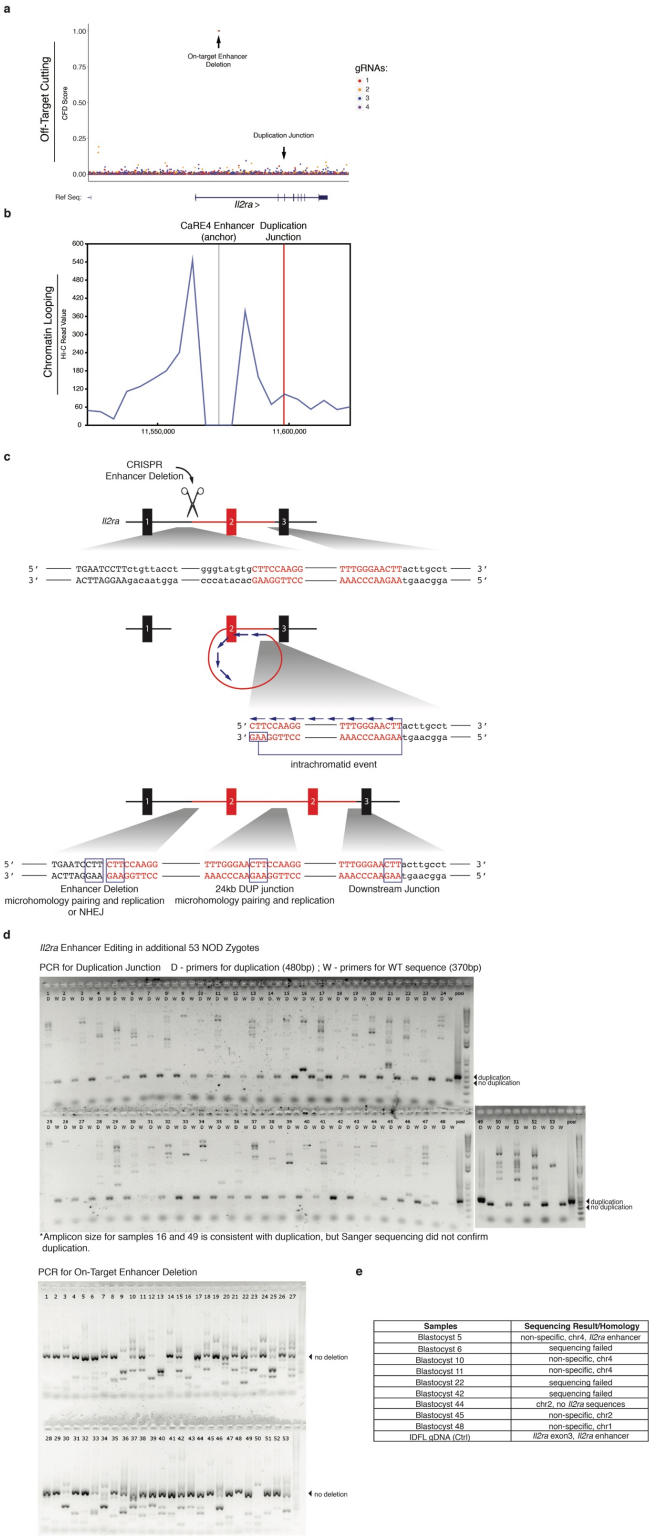


Figure 3.5. Identifying a large tandem duplication in the Il2ra locus. a, Read pileups at the *Il2ra* locus from genome sequencing of a homozygous enhancer deletion (EDEL) mouse (Line 2) and homozygous and heterozygous EDEL mice from the immune dysregulated founder line (IDFL). Red lines were added to highlight the elevated read counts in IDFL mice. **b**, Schematic of the *Il2ra* locus with the large tandem duplication in IDFL mice. PCR and Sanger sequencing across the novel junction sequence created by the duplication. **c**, Read counts from RNA sequencing of IL2RA+ CD4+ T cells showing reads that span the aberrant exon 2-exon 2 junction in red and the normal exon 2-exon 3 junction in grey. Data in (c) are from biological replicates derived from two independent experiments.

We next interrogated how the duplication formed. Previous work showed that paired CRISPR-induced DNA breaks can result in tandem duplication of the intervening sequence[236,237], however no predicted off-target cutting sequences were identified in the *Il2ra* locus to explain this duplication event (Figure 3.6). Although we cannot rule out spontaneous DNA breaks from DNA replication, our data suggests that the duplication is more likely an unintended product of repair from on-target editing. Sequence homology could contribute to a duplication event. We did not find extended sequence homology between the cut site and the duplication junction, but we did discover microhomology at the breakpoint junction (Figure 3.6). However, three nucleotides of microhomology are found commonly in the genome, raising a question of why this distal site may have been used for repair. Chromatin looping can bring distal genomic sites into close proximity. Indeed, published high-resolution chromatin conformation capture data suggested three-dimensional proximity between the *Il2ra* enhancer and the site of the duplication junction (Figure 3.6)[238], although the statistical significance of this putative looping interaction was not determined. We then tested whether microhomology and looping were sufficient to drive recurrent duplications at this site (Figure 3.6). We repeated CRISPR microinjections into single cell NOD/ShiLtJ zygotes, cultured them in vitro and analyzed >50 blastocysts and failed to observe the recurrence of this particular duplication event despite efficient deletion (~80%) of the enhancer at the on-target site. Note, there were two blastocysts with PCR bands of roughly the expected duplication size, but the sequence could not be confirmed. Taken together our analysis of the duplication is consistent with a complex unintended repair consequence that occurs less frequently than the intended enhancer deletion.



Samples	Sequencing Result/Homology
Blastocyst 5	non-specific, chr4, *Il2ra* enhancer
Blastocyst 6	sequencing failed
Blastocyst 10	non-specific, chr4
Blastocyst 11	non-specific, chr4
Blastocyst 22	sequencing failed
Blastocyst 42	sequencing failed
Blastocyst 44	chr2, no *Il2ra* sequences
Blastocyst 45	non-specific, chr2
Blastocyst 48	non-specific, chr1
IDFL gDNA (Ctrl)	*Il2ra* exon3, *Il2ra* enhancer

Figure 3.6. Characterization of Il2ra bystander duplication. a, Computational prediction of off-target sites for the Il2ra enhancer gRNAs throughout the Il2ra gene body assessed by Continuous Frequency Determination (CFD) scoring (Doench and Fusi et al. 2016. Nat. Biotech.). We highlight the position of the duplication breakpoint. b, Overlap of the Il2ra locus with 5kb-resolution Hi-C data from mouse embryonic stem cells anchored at the Il2ra enhancer (Bonev et

al. 2017. Cell). c, Schematic of proposed microhomology-mediated repair at the Il2ra locus that could generate the observed duplication (created with BioRender). d, PCR genotyping for Il2ra enhancer editing experiment in additional 53 microinjected NOD zygotes. Top, PCR results for the duplication breakpoint. Amplification was carried out for the duplication junction (D) or WT sequence (W). Bottom, PCR results for the on-target enhancer deletion. e, Summary of sequencing results from nested PCR to genotype IDFL duplication junction in enhancer-targeted NOD blastocysts.

The duplication of exon 2 is predicted to generate a novel splice junction that would result in a premature stop codon in the *Il2ra* mRNA (Figure 3.5). While the expected effect of the homozygous duplication would be to ablate protein expression in every cell, we were nevertheless able to detect CD4+ T cells with near normal levels of IL2RA expression. To understand how these cells expressed IL2RA despite the predicted premature stop codon generated by the duplication we performed RNA-seq on IL2RA+ T cells from spleen. As expected, we could identify reads with the aberrant exon2-exon2 splice junction only in the IFDL EDEL cells (Figure 3.5, 3.7). However, IL2RA+ IDFL cells had ~10-fold more reads that contained the wild-type exon2-exon3 junction than the aberrant exon2-exon2 junction (Figure 3.5). PCR and Sanger sequencing of *Il2ra* cDNA confirmed that these cells predominantly generate *Il2ra* transcript with a single exon 2 between exons 1 and 3 (Figure 3.7). In contrast, we did not detect transcripts with a single exon2 in IL2RA- T cells induced to express *Il2ra* (Figure 3.7). These isoform differences suggest that some T cells, including a subset of Foxp3+ Tregs, are able to correctly splice the aberrant *Il2ra* genetic structure and productively translate IL2RA.

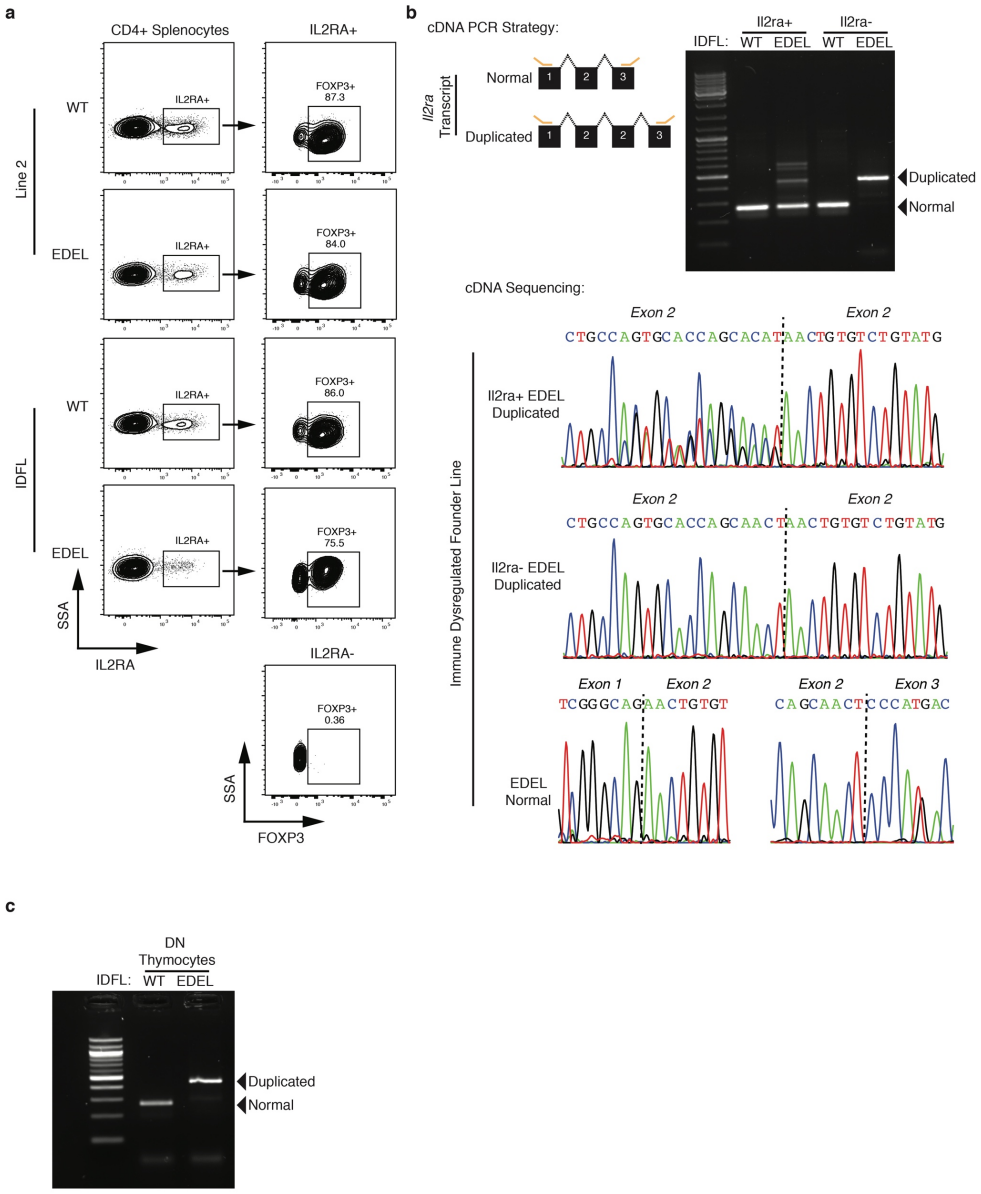


Figure 3.7. Il2ra splicing analysis. a, Gating scheme used to sort IL2RA+ and IL2RA- cells. Foxp3 staining in sorted IL2RA+ cells. b, Il2ra exon 2 splicing analysis on IL2RA+ and stimulated IL2RA-CD44- cells in WT and EDEL CD4+ T cells from the immune dysregulated founder line (IDFL). PCR amplicons across exon 2 were generated from cDNA and run on a 1% agarose gel. The lower ("normal") band is consistent a single exon 2 Il2ra isoform, whereas the higher ("duplicated") band is consistent with an Il2ra isoform with two exon 2s. The amplicons were isolated and Sanger sequenced for verification. We observed that IL2RA+ IDFL EDEL cells showed variable larger amplicons of unknown significance. Sequencing these larger amplicons showed Il2ra transcript with two exon 2s, but did not explain the variable sizes observed. In these cells, we noted mixed Sanger sequencing peaks adjacent to the exon2-exon2 junction potentially consistent with an aberrant splice isoform or isoforms for which we do not know the significance. c, Il2ra exon 2 splicing analysis of IDFL WT and EDEL double negative (DN) thymocytes. The amplicons were isolated and Sanger sequenced for verification. We observed variable smaller

amplicons of unknown significance. Sequencing these amplicons showed Il2ra transcript with normal splicing (WT) or two exon 2s (EDEL), but did not explain the variable sizes observed.

DISCUSSION

This report links a CRISPR-induced bystander mutation to *in vivo* pathology. We build on previous work that showed CRISPR editing can cause bystander deletions and complex rearrangements in neighboring on-target sequences[58]. Further work is needed to understand the frequency with which such mutations occur, as well as the DNA repair rules that underlie these events. Identification of the bystander mutation depended on having multiple independent founder pedigrees to demonstrate an aberrant phenotype in one line and genetically link it to the on-target edit. New methods and analytical tools are needed to detect both unintended CRISPR-induced bystander and off-target mutations. Genome engineering not only allows gene knockout, but also permits targeted alterations of non-coding *cis*-regulatory sequences for mechanistic study of human variants and for cell therapies. The bystander mutation allele observed here was introduced by murine zygote editing by Cas9 mRNA and gRNA microinjection. The marked immune phenotype was revealed by breeding the rare allele to homozygosity. The functional consequences, if any, of rare unintended alleles in a population of human primary somatic cells edited by various CRISPR delivery strategies remain largely untested. Bystander editing effects – which can be easily missed with conventional genotyping methods – must be carefully assessed for research and clinical CRISPR applications, especially for mounting therapeutic efforts to fine tune gene regulatory programs.

METHODS

Mouse Generation

Il2ra enhancer deletion (EDEL) mice were generated by the Jackson Laboratory (Bar Harbor, ME, USA) by microinjection of gRNA and Cas9 mRNA. Briefly, Cas9 mRNA (100 ng μl^{-1}) and mIL-2Ra-CaRE4 gRNAs (50 ng μl^{-1}) were mixed and injected into NOD/ShiLtJ zygotes. Four founders with the enhancer deletion were identified by PCR amplicon size and confirmed by sequencing of TOPO-cloned PCR products. We immunophenotyped three founders. The EDEL mouse lines were established by backcrossing founders for at least one generation before breeding to homozygosity. Protospacer sequence for gRNAs used in the production of the founder lines in this study are:

mIl2ra-EDEL-up1 TGCTCTTTGAAGGTAACAGA

mIl2ra-EDEL-up2 GTTACCTTCAAAGAGCAGCC

mIl2ra-EDEL-down1 AAGATGGGTATGTGCTTCCA

mIl2ra-EDEL-down2 AGATGGGTATGTGCTTCCAA

Mouse Genotyping

The enhancer deletion was initially genotyped in all founders by Sanger sequencing genomic DNA from proteinase K digested tail tissue. PCR amplification of the CaRE4 enhancer was carried out using HotStart Taq (Bioline USA Inc.) and primers (mIl2ra-EDEL-F, mIl2ra-EDEL-R) that span the edited site. PCR amplicons were then sequenced with the mIl2ra-EDEL-F primer. The duplication junction was confirmed by PCR amplification of the junction followed by gel electrophoresis and Sanger sequencing. The primers used are listed in Primer Section of Methods.

Mouse Experiments and Data Analysis

All mice were maintained in the UCSF specific-pathogen-free animal facility in accordance with guidelines established by the Institutional Animal Care and Use Committee and Laboratory Animal Resource Center. Experiments were done with animals aged between 1 to 4 months. Wildtype littermate mice were used as controls for all immunophenotyping experiments. All mice used in this study were normoglycemic. No data was excluded from analysis. Power calculations were not performed and data was assumed to be normally distributed. Experiments were done without blinding or randomization. All data is derived from at least two independent experiments unless otherwise stated. Both male and female mice were used for experiments.

Mouse Immunophenotyping

Briefly, cells from spleen, peripheral lymph nodes (peri-LNs) and thymus were collected from each mouse. Spleen, peri-LNs, and thymus were dissociated in 1x PBS with 2% FBS and 1 mM EDTA. The mixture was then passed through a 70-μm filter. ACK lysis was used to deplete red blood cells from splenocytes. All antibody stains were performed at a 1:100 dilution in 30-50 μl of 1x PBS. To pellet the cells, centrifugation was performed at 300g for 5 min. For immunophenotyping, approximately 2 million cells were stained per tissue sample. Cells were first stained with a viability dye at a 1:1,000 dilution in 1x PBS for 20 min at 4 °C, then washed with EasySep Buffer (1x PBS, 2% FBS, 1 mM EDTA). Cells were then resuspended in the appropriate surface staining antibody cocktail and incubated for 30 min at 4 °C, then washed with 1x PBS. Cells were then fixed, permeabilized, and stained for transcription factors using the Foxp3 staining kit (eBioscience) according to the manufacturer's instructions.

Antibodies and Staining Panels

PANEL 1 - mouse peripheral LNs, spleen and LPL-CD4+

Viability Dye:

LIVE/DEAD® Fixable Blue Dead Cell Stain Kit (Thermo Fisher Scientific, Cat# L23105)

Surface stain:

mCD45-BUV395 (Clone 30-F11, BD, Cat#564279)

mCD4-BV605 (Clone RM4-5, Biolegend, Cat#100548)

mCD8-PerCP-Cy5.5 (Clone 53-6.7, Biolegend, Cat#100734)

mCD25-PE-Cy7 (Clone PC61.5, Affymetrix, Cat#25-0251-82)

mTCRB-PB (Clone H57-597, Biolegend, Cat#109226)

mCD69-APC (Clone H1.2F3, Biolegend, Cat#104514)

Intracellular:

mRORgt-PE-CF594 (Clone Q31-378, BD, Cat#562684)

mHelios-PE (Clone 22F6, Biolegend, Cat#137216)

mFOXP3-FITC (Clone FJK-16s, Biolegend, Cat#320112)

PANEL 2 - mouse thymus

Viability Dye:

LIVE/DEAD Fixable Blue Dead Cell Stain Kit (Cat# L23105, Thermo Fisher Scientific)

Surface stain:

mCD45-BUV395 (Clone 30-F11, BD, Cat#564279)

mCD4-BV605 (Clone RM4-5, Biolegend, Cat#100548)

mCD8-PerCP-Cy5.5 (Clone 53-6.7, Biolegend, Cat#100734)

mCD25-PE-Cy7 (Clone PC61.5, Affymetrix, Cat#25-0251-82)

mTCRB-PB (Clone H57-597, Biolegend, Cat#109226)

mCD44-APC (Clone IM7, Biolegend, Cat#103012)

Intracellular:

mHelios-PE (Clone 22F6, Biolegend, Cat#137216)

mFOXP3-FITC (Clone FJK-16s, Biolegend, Cat#320112)

PANEL 3 – mouse CD4+ T cell sort

mCD4-PB (GK1.5, Biolegend, Cat#100428)

mCD62L-PE-Cy7 (MEL-14, Biolegend, Cat#104418)

mCD44-APC (Clone IM7, Biolegend, Cat#103012)

mCD25-AF488 (Clone PC61, Biolegend, Cat#B220118)

PANEL 4– mouse activated naïve T cells

Viability Dye:

Ghost Dye Violet 510 (Tonbo Biosciences, Cat#13-0870) or LIVE/DEAD Fixable Blue Dead Cell Stain Kit (Cat# L23105, Thermo Fisher Scientific)

Surface Stain:

mCD25-AF488 (Clone PC61, Biolegend, Cat#B220118)

mCD69-APC (Clone H1.2F3, Biolegend, Cat#104514)

PANEL 5 - stimulation of T cells with plate-bound antibody

mCD3e (Clone 145-2C11, Biolegend, Cat#100302)

mCD28 (Clone 37.51, Biolegend, Cat#102102)

IL2RA Induction on Stimulated Naive T Cells

Naive T cells were isolated from spleen and lymph nodes with CD4+ negative selection (Stemcell Technologies) followed by fluorescence activated cell sorting for CD4+IL-2Ra−CD44−. 100,000 cells we activated per well of a 96-well plate coated with 2 μg ml^{-1} anti-CD3 and anti-CD28. Cell analysis by flow cytometry was performed every day for 3 days.

***In vitro* Treg Differentiation**

Spleens were dissociated in EasySep Buffer and splenocytes were enriched for CD4+ T cells using the Mouse CD4 Negative Selection Kit as described above. Naïve CD4+CD62L+CD44- T cells were sorted. 100,000 cells were stimulated with plate bound anti-CD3/CD28 antibodies in a 96 well plate in the presence of 2ng/ml TGF-b (Miltenyi Biotec) and 200U/ml IL2 (Miltenyi Biotec) for 3 days. Treg differentiation was assessed by surface staining for IL2RA and intracellular staining for Foxp3 without a viability dye.

Genome Sequencing Sample Preparation

DNA was isolated from kidney tissue by phenol-chloroform extraction. PCR-free whole genome libraries were constructed by the Genome Technologies Core at The Jackson Laboratory using the KAPA Hyper Prep Kit (KAPA Biosystems), targeting an insert size of 400 base pairs. Libraries were checked for quality and concentration using the Bioanalyzer High Sensitivity DNA Assay (Agilent), Qubit dsDNA BR Assay (ThermoFisher), and quantitative PCR (KAPA Biosystems), according to the manufacturers' instructions. Libraries were sequenced at Novogene, 150 base pairs paired-end on the HiSeq X (Illumina) to a target mean coverage depth of 30X.

Genome Sequencing Alignment and Variant Calls

Bwa mem was used for alignment to the mouse mm10 reference sequence. Reads at the identified tandem duplication junctions were then assembled using Velvet[239] in order to confirm the exact base pair sequence as well as assist in picking primers for confirmation of the duplication via PCR. The Picard Software Suite and GATK 4.0 pipeline[240] with default settings was used for variant analysis. Base recalibration was performed with NOD specific variants (both SNPs and INDELs) obtained from the Wellcome Sanger Institute Mouse Genome Project.

CRISPR Off-target Analysis

We first performed a biased off-target analysis looking for variants 5bp on either side of predicted off-target cut sites. This was done for the top 49 predicted off-target sites for each gRNA that was used to make the enhancer deletion mouse lines. In total 6 variants were found with this analysis, all of which were present in the NOD background variant panel. We also performed an unbiased variant analysis to examine potentially confounding mutations that fit a likely inheritance model. The cohort variant call file was subset for biallelic SNPs and INDELs where all individuals were assigned a genotype with a sufficient average coverage (>= 10 reads). Variants that were unique or in excess in the immune dysregulated mouse as compared to the other two mice in the cohort. The resulting alleles were further subset by removing NOD specific SNP and INDEL variants and selecting only variants that fell within exonic regions. This revealed 2407 variants. The remaining variants are likely specific to the NOD mice used to generate our founder lines. Other than the 24kb duplication, we found no evidence of coding or splicing mutations near the enhancer deletion that might contribute to the observed phenotypes. Identifiers GT_05102 and GT_05105 refer to heterozygous and homozygous EDEL mice from the IDFL founder line. GT05111 is homozygous

EDEL mouse from Line 2. Finally, we performed computational prediction of off-target cutting for the *Il2ra* enhancer gRNAs throughout the *Il2ra* gene body (Figure 3.6).

RNA Sequencing

Briefly, approximately 500,000 CD4+IL2RA+ cells were sorted from CD4-enriched splenocytes and total RNA was isolated from samples using the RNeasy Micro Kit (QIAGEN) according to the manufacturer's instructions with the following options: cells were pelleted and re-suspended in RLT buffer with β-mercaptoethanol and homogenized using QIAshredder (QIAGEN). DNA removal was performed with gDNA Eliminator Columns (QIAGEN). RNA samples were analyzed with a NanoDrop spectrophotometer and all samples had a 260/280 ratio of 1.80 or higher. RNA integrity was measured by Bioanalyzer and all samples had an RNA Integrity score (RIN) of 8.0 or more. RNA-seq libraries were prepared by the Functional Genomics Laboratory at Berkeley. RNA samples were poly-A selected and then converted into sequencing libraries with the ultra-low input SMART-seq kit. The samples were pooled and sequenced on one lane of the Illumina HiSeq4000. *Il2ra* isoform analysis was done using the UNIX grep command to identify reads in raw fastq that contained sequences for the E2-E2 junction (ACCAGCAACTAACTGTGTCT) or E2-E3 junction (ACCAGCAACTCCCATGACAA). Read counts were normalized to the total number of reads for a given sample. Short reads were also aligned with STAR to the mouse mm10 reference. Differential expression analysis was performed using EdgeR from Bioconductor Package for R[241].

***Il2ra* cDNA Isoform Analysis**

500,000 IL2RA+ (CD4+IL2RA+) and IL2RA- (CD4+IL2RA+CD44-) CD4+ T cells were sorted from CD4-enriched splenocytes. IL2RA- CD4+ T cells were stimulated *in vitro* for 10hrs with

2ug/ml plate bound anti-CD3/CD28 antibodies (Biolegend). 430,000 DN thymocytes (Live CD45+CD4-CD8-) were sorted from thymus. RNA was extracted using the RNA Micro Kit (QIAGEN) as described above. RNA was reverse transcribed into cDNA using SuperScript VILO MasterMix as per manufacturer's protocol (Thermo Scientific). PCR of the cDNA to assess exon 2 splicing was performed with forward and reverse primers that sit in *Il2ra* exon 1 and exon 3, respectively (Primers Section). PCR was carried out with Bioline Taq 2x MasterMix as per manufacturer's protocol. Amplicons were cut out of the agarose gel and purified using QIAGEN's Gel Extraction Kit. Amplicons were sequenced in the forward and reverse directions using the primers from the initial PCR amplification.

Zygote *Il2ra* Enhancer Editing

NOD/ShiLtJ zygotes were microinjected with gRNAs and Cas9 mRNA, identical to the generation of *Il2ra* EDEL mice. PCR amplification and Sanger sequencing were used to check gDNA for the duplication junction and WT *Il2ra* sequence upstream of *Il2ra* exon 3, the genomic site of the IDFL duplication junction (Figure 3.6). Blastocyst gDNA was also checked for on-target enhancer deletion by PCR amplification and Sanger sequencing (Figure 3.6). A second test for the duplication junction was performed using a nested PCR. Briefly, duplication junction PCR samples were diluted (1ul sample in 10ul water) and 1ul of dilution was used as the template. A non-specific band was observed in a majority of the blastocysts (data not shown). Nine amplicons were extracted and sent for Sanger sequencing, of which six were successfully sequenced. The primers used for these assays are listed in Primers Section. The reconstructed sequences are assembled in Figure 3.8. Takara PrimeSTAR polymerase was used for amplification with 30 second extension for 35 cycles. PCR products were run on agarose gel for size separation and visualization.

Primers for Assay

Il2raNstdF 5′ GCCATTTCTCATGCCTGTCT 3′
Il2raNstdR 5′ CTCAGCCCTTAGCTTGGGTA 3′

Positive Control

>IDFL gDNA (chr2:11Mb, *Il2ra* x3 and *Il2ra* enhancer; forward and reverse)
GCCATTTCTCATGCCTGTCTGGTAACAAGGTCCCTCATGCCCTACACAGGAGCGGTCTTTTGCTTATCTTTGCACACCCAGCACAGATATGTCTCTG
GGCACCAGAGAGGTGTCCCAGGGTAAATGACCGCATAGACAGCTCTTTCTATTCTGAACATTTGGGAACTTCCAAGGGCAGACTAAGAAGCCACTTA
CAGGTGAGTGTGACAGGTGAGAGGAACCGTGGCAAGGAACACCCTACCCTGGTGGCAGAGCTTCTCGATATTTGTATGTCTGCTGCTTTGCTTCTCA
GTTCATTCATGATCTATTGAGCAACTTCACGCAAGGTTAGAGAGGCAGAGATGAATAAGACCCAACTCTTGGCCTCCAGGAGCTTCTTGCTTATGCG
GAGAGAATACCCAAGCTAAGGGCTGAG

CRISPR-Targeted Blastocysts

>Blastocyst 5 (chr4:42Mb and chr2:11Mb; forward and reverse seq.)
TTCAAATGAAGATGTAGAACCCTCAGAGCTCCTTCTGAACCATGCCTGTCTGGATGCTGCCATGTTCTTGCCTTGGTGATAATGGATGACCTCTGAA
CCTGTAAGCCAGCCCCAGGTTAAATGTTGTCTTTTATAAGACTTGCATTGGTCATGGTGTCTGTTCACAGCAGTAAAACCCTAACTAAGACAGCTGC
TGAGCATGTTAATAGAAGACGGGTCTGTGTATCCCCCAAGCCAAGTAGAGTCTCCGTTTGCTAAACTATTTCCTATGCTTCCCTACCTCACTGTGTG
TGTGTGTGTGTGTGTGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTGGTAATAAAAAACAAAAGACAAAGATGGGTAT
GTGCAAGGGCAGACTAAGAAGCCACTTACAGGTGAGTGTGACAGGTGAGAGGAACCGTGGCAAGGAACACCCTACCCTGGTGGCAGAGCTTCTCGAT
ATTTGTATGTCTGCTGCTTTGCTTCTCAGTTCATTCATGATCTATTGAGCAACTTCACGCAAGGTTAGAGAGGCAGAGATGAATAAGACCCAACTCT
TGGCCTCCAGGAGCTTCTTGCTTATGCGGAGAGAATACCCAAGCTAAGGGCTGAG

>Blastocyst 10 (chr4:42M; forward only seq.)
TTCAAATGAAGATGTAGAACCCTCAGAGCTCCTTCTGAACCATGCCTGTCTGGATGCTGCCATGTTCTTGCCTTGCTTGGTGATAATGGATGAACCT
CTGAACCTGTAAGCCAGCCCCAGTTAAATGTTGTCTTTTATAAGACTTGCATTGGTCATGGTGTCTGTTCACAGCAGTAAAACCCTAACTAAGACAG
CTGCTGAGCATGTTAATAGAAGACGGGTCTGTGTATCCCCCAAGCTAAGGGCTGAG

>Blastocyst 11 (chr4:42M; forward only seq.)
AGAACCCTCAGAGCTCCTTCTGAACCATGCCTGTCTGGATGCTGCCATGTTCTTGCCTTGGTGATAATGGATGAACCTCTGACCTGTAAGCCAGCCC
CAGGTTAAATGTTGTCTTTTATAAGACTTGCATTGGTCATGGTGTCTGTTCACAGCAGTAAAACCCTAACTAAGACAGCTGCTGAGCATGTTAATAG
AAGACGGGTCTGTGTATCCCCCAAGCTAAGGGCTGAG
!
>Blastocyst 44 (chr2:31M; reverse only seq.)
AGGGGAGGGGTGTCCCAGAGAGCACACACCATGTAGTGGCATGGAGCCAAGTTTGTCCCTGCTTGGGTACAGGTGTCCCAGGTGTCCCAGCACGTAG
CTCTGGCCCAGGATGTCTGGCAGGGCAACTTCCCCTTGCCTTGAGGTGCCTTCCTGAATGTAGTGTGGGTGTGTGTGGGTGTCCAGGCATGGGAAAG
CTGGAGAAAAAGCCCTGTACTCATGTGCCCGCCCCTGTACCTCCGACAGCTTGGCCCCTCCTTACCCGGGGCTGGAGCCCTGCCGGTAGGTGGTAGG
ACAGGGCGTGTCTACACACTGGAAGCTGCCTCGGGTGTTGAAGCACATCTGGCCAGGCCCACACTCGATGCCATCTTCCTCACACTCGTTGATGTCT
GGGCGGTAGGGTGCGGTAGGAGATGGGAAGATAGAACAGGGGAGAGGGAGACCAGAGGAGGCCGTGTGAAAGAGACCAGCT
!
>Blastocyst 45 (chr2:130M, forward and reverse seq.)
ATTTGCCATTGTTTTCCAGGTTGGCTTGCTGGCCTGTGAGCCTGAGGTATCTGCCTGTCTTCCCCAGGGTTGGGGGTTACAGGCGTGTTCCAGTACA
CTCAGCCTTTGTGTGGAAGTTAGGTGTCTGAACTCAGATGTTCGAGATTATACATCAAATACTTGGCGCACCAACCCCTCTCCCAAGCTCAGACACA
CTTTCTTCTGGGCCTGTGCCAAGGTGAGGAAGCAGATGGGTGAGGACCAGGGCAGTAAGGAGGAAGCAGCTTCCAGGAACACGGGTGCAGCCTCTTG
TCTGCCACACACAGGTGGGAAATACATCGTGTCACTAGCTCCCACCCTGCTGAGCCCTGTCTATGTGCTGGCACGACTCTTCAGTCCTCTATCTTCC
CTATCTCTGAAATTTTAGCTGCATACATCTTAAGACAAATTACCCAAGAGTTGAAAATAAGACCAAGAACAGAGGGGCTGTATCGTCTTAGGGGCT
GTGGAAGT

>Blastocyst 48 (chr1:64M; forward and reverse seq.)
AGGCAGACCTATTGCTCTTGGCTCGTTCTGCTCATGCCTCTCTGGGTCAGCTTTGCTCTAGGCTGGGTGGGGCTTCTGCAGTTTCCTTGGGGAAGCT
TTGCCTTCACACATCTCTCATCCCTTCCTGGAACCTTGGTTTCTTCTGGACATGACACATTCATCTCATGGAAAGAACAGAGATAGATACACGGAAG
AAAGCCAATCATAGCATGTTTTCTAGCCTGTACTTGTAGCATGTCTGTTAAAATCCTACTGGCCAAAGCAGGTGTTGCCATGAGAGCTAATGGGTGA
GGAGTATTCTCGACTCTTGTTGAAAGTGTGACACCTTCTATGGGCTTACTGACATAAAGAGGAATGAAAAATGCAGGTACTGGTGCACTCTGCTGCA
GCCTCCAGGCAGCAAGCCTTCTTGGTATCAGCTACTTTTTGTGCCCATCCCTTGAGATACAGGCCTTTCAAAGCCTGGTGTTCTCTCTGTTTAGGGT
TCTAAGCATGCAGGACAAACCAGACCAGGCAGAGAAGTGGCTACCCTCGCCAGTGACTTCAGCACAAAGGCTACACTTTGGCCTGCCTCCTAGGGTA
AACAGAGAAGGCTGGCAAGCCAGCACTGGCTCACGCCTCCTGGCTCTGGCTCCCTCCCCAGCTACCCCCACCCTAACAGCTTCTTAGAGTGTTAGCC
ACTGCTCTAGTGAGAGAAGAGTAATTGAGGGTGTAATTAAAAAAAAAAGAAGAAGAAGAAGGAGGAGGAGGAGAAGGAGAAGGAGAAGGAGAAGGAG
AAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGAAGAAGAAGACATTAG
AGAGCAATAGACCAGATACCCCAGCTCAGGGCAGCTCTAAACCAAAAGACTAAGGCAG

Figure 3.8. Nested PCR of enhancer-targeted blastocysts to detect IDFL duplication junction. Reconstructed amplicon sequences using Sanger sequencing and blast results in mouse genome (mm10). Nested primers for the IDFL duplication junction were used to genotype

blastocysts. Underlined sequences show putative primer binding sites. Colored sequences represent genomic regions identified in Sanger sequencing. Orange - sequence adjacent to Il2ra exon 3; red - sequence adjacent to Il2ra enhancer; Additional colors – other genomic sequences.

Primers

Pimer Name	Sequence	Purpose
mIl2ra-EDEL-F	TCCTCAGGACCCTGCTAGTC	gDNA PCR to genotype enhancer deletion
mIl2ra-EDEL-R	GAGAAGCAAAGCAGCAGACA	gDNA PCR to genotype enhancer deletion
Dup_jx_F	CTGAGAAGCAAAGCAGCAGA	gDNA PCR to confirm duplication junction
Dup_jx_R	TGGCTGATGGCTAAGGGATA	gDNA PCR to confirm duplication junction
mIl2ra-cDNA-E1-F	CCAGTTGTCGGGCAGAAC	mIl2ra cDNA PCR to check splicing of exon 2
mIl2ra-cDNA-E3-R	TGCATGTCTGTTGTGGTTTG	mIl2ra cDNA PCR to check splicing of exon 2
Il2ra-comF	CCGGGATTTAAGCTCATTCA	zygote editing, duplication breakpoint genotyping
Il2ra-dupR	GTGGAGTGTGTGTCCACCAG	zygote editing, duplication breakpoint genotyping
Il2ra-wtR	GGCTAGAGGATGGTTGCTGA	zygote editing, duplication breakpoint genotyping
Il2ra-tgt-f	cctcttgctctcccagacag	zygote editing, enhancer deletion genotyping
Il2ra-tgt-r	aaccttgcgtgaagttgctc	zygote editing, enhancer deletion genotyping
Il2ra-NstdF	GCCATTTCTCATGCCTGTCT	nested duplication breakpoint genotyping
Il2ra-NstdR	CTCAGCCCTTAGCTTGGGTA	nested duplication breakpoint genotyping

Table 3.1. Primers used in *Il2ra* bystander mutation study.

Data Analysis

The statistical tests and sample sizes used for data analysis are included in figure legends.

Data Availability

Whole genome sequencing and RNA sequencing data have been uploaded to NCBI Sequencing Read Archive (SRA Accession: PRJNA510427)

CHAPTER 4. Functionally restricted *IL2RA* enhancers control divergent cellular phenotypes and disease outcomes

Immune enhancers maintain the delicate balance between autoimmunity and tolerance by tuning transcriptional programs in response to specific extracellular cues. The high affinity alpha subunit of the IL-2 receptor (IL2RA) is a critical immune regulator that provides survival signals to immune cells that express it[242], yet its complex genetic regulation in the immune system is incompletely understood. We previously tiled CRISPR-activation across the super-enhancer at the *IL2RA* locus to systematically map functional *IL2RA* enhancers, including a stimulation-responsive enhancer that harbors an autoimmune disease risk variant[235]. Here, we undertook a genetic perturbation approach to dissect how distinct enhancers within the *Il2ra* super-enhancer regulate immune cell function as well as shape risk of autoimmunity *in vivo*. Using CRISPR-engineered enhancer deletion mice we identified a novel *Il2ra* maintenance enhancer that controls IL2RA expression in FOXP3+ regulatory T cells, distinct from the disease-associated *Il2ra* enhancer that controls an *Il2ra* induction program in stimulated T and B cells. Having discovered enhancers that regulate in IL2RA in different contexts we interrogated their effects in an *in vivo* model of autoimmune disease. Deletion of the conserved stimulation-responsive enhancer that harbors a human SNP protective against T1D completely protected non-obese diabetic (NOD) mice from diabetes. Our work has implication for decoding genetic circuits in a critical autoimmunity locus and builds on our understanding of enhancer regulation in the immune system.

Results

To understand how *IL2RA* enhancers control immune responses, we searched for enhancers with activities in immune cells that have opposing immunological functions. We previously identified

an intronic *IL2RA* enhancer that controls IL2RA induction on pro-inflammatory conventional T cells (Tconv), termed CaRE4 enhancer, for which we generated enhancer deletion (EDEL) mice on the NOD background[235]. We now sought to identify an *IL2RA* enhancer that is functional in anti-inflammatory immune cells, such as Tregs. Using published Hi-C chromatin immunoprecipitation (Hi-ChIP) data[235] as well as histone 3 lysine 27 acetylation (H3K27Ac) from subsets of primary human T cell, we identified a conserved ~200bp sequence upstream of the IL2RA promoter that was predicted to be preferentially active in Tregs (Figure 4.1). We named this site CaRE3 enhancer due to its genomic location and overlap with another CaRE. Distinct chromatin signatures suggested that we have identified two enhancers predicted to control *Il2ra* expression in cell types with opposing immunological functions.

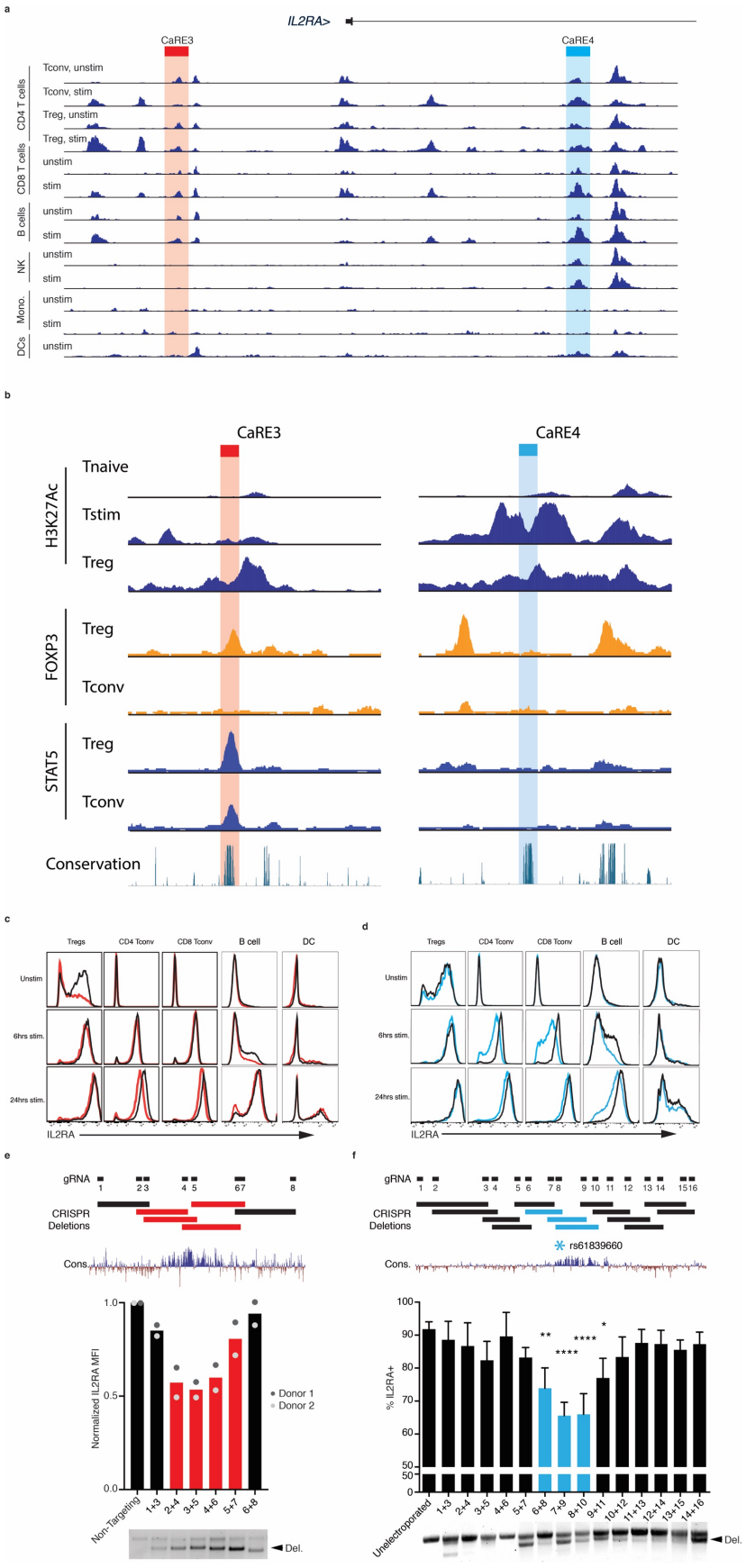

a
IL2RA>
CaRE3
CaRE4
CD4 T cells
Tconv, unstim
Tconv, stim
Treg, unstim
Treg, stim
CD8 T cells
unstim
stim
B cells
unstim
stim
NK
unstim
stim
Mono.
unstim
stim
DCs
unstim
b
CaRE3
CaRE4
H3K27Ac
Tnaive
Tstim
Treg
FOXP3
Treg
Tconv
STAT5
Treg
Tconv
Conservation
c
Tregs
CD4 Tconv
CD8 Tconv
B cell
DC
Unstim
6hrs stim.
24hrs stim.
IL2RA
d
Tregs
CD4 Tconv
CD8 Tconv
B cell
DC
Unstim
6hrs stim.
24hrs stim.
IL2RA
e
gRNA
1
2 3
4 5
6 7
8
CRISPR Deletions
Cons.
Normalized IL2RA MFI
1.0
0.5
0.0
Donor 1
Donor 2
Non-Targeting
1+3
2+4
3+5
4+6
5+7
6+8
Del.
f
gRNA
1 2
3 4
5 6
7 8
9 10
11
12
13 14
15 16
CRISPR Deletions
rs61839660
Cons.
% IL2RA+
100
90
80
70
60
50
50
0
**
****
****
*
Unelectroporated
1+3
2+4
3+5
4+6
5+7
6+8
7+9
8+10
9+11
10+12
11+13
12+14
13+15
14+16
Del.

Figure 4.1. Mapping the functional immune contexts of IL2RA CaRE3 and CaRE4 enhancers. a, *IL2RA* CaRE3 and CaRE4 enhancers overlapped with chromatin accessibility from diverse primary human immune cell types (Calderon et al. unpublished[243]). **b**, Zoomed in view of *IL2RA* CaRE3 and CaRE4 enhancers showing H3K27Ac, STAT5, and FOXP3 transcription factor ChIP-seq from primary human immune T cells. **c** and **d,** IL2RA surface expression time course on immune cell subsets stimulated in splenocyte cultures from NOD CaRE3 or CaRE4 mice. B cell data is from anti-CD3/CD28 treated splenocytes due to stronger B cell activation response. Anti-IgM/CD40 data is consistent, but not shown here. **e,f**, Deletion tiling of the *IL2RA* CaRE3 and CaRE4 enhancers in human T cells using CRISPR.

To study the *in vivo* role of the CaRE3 enhancer element, we deleted the conserved site in mice. We first examined T cell development. Although the abundance of major thymocyte populations was normal, double negative thymocytes had impaired Il2ra expression (Figures 4.2, 4.3). We then assessed Treg development which was impaired. We observed frequency changes in Treg progenitor populations and reduced frequencies of mature Tregs (FOXP3+CD25+) (Figures 4.2, 4.3). Consistent with these findings, Il2ra surface expression on CaRE3 EDEL Tregs was impaired in all lymphoid tissues we immunophenotyped (Figures 4.2, 4.3). Normal Il2ra expression is necessary for the survival of mature Tregs in the periphery[210], yet the abundance of Tregs in secondary lymphoid organs of CaRE3 EDEL mice was normal (Figures 4.2, 4.3). We found that the Il2ra+ subset of CaRE3 EDEL Tregs are more highly proliferative by Ki67 staining in thymus and inguinal lymph nodes, perhaps compensating for the cells with impaired Il2ra expression (Figure 4.3). Finally, we find that CaRE3 EDEL mice have higher frequencies of activated T conventional cells suggesting a breakdown in tolerance in these animals (Figure 4.3). In summary, the CaRE3 enhancer is required for normal constitutive expression of Il2ra in the regulatory T cell lineage.

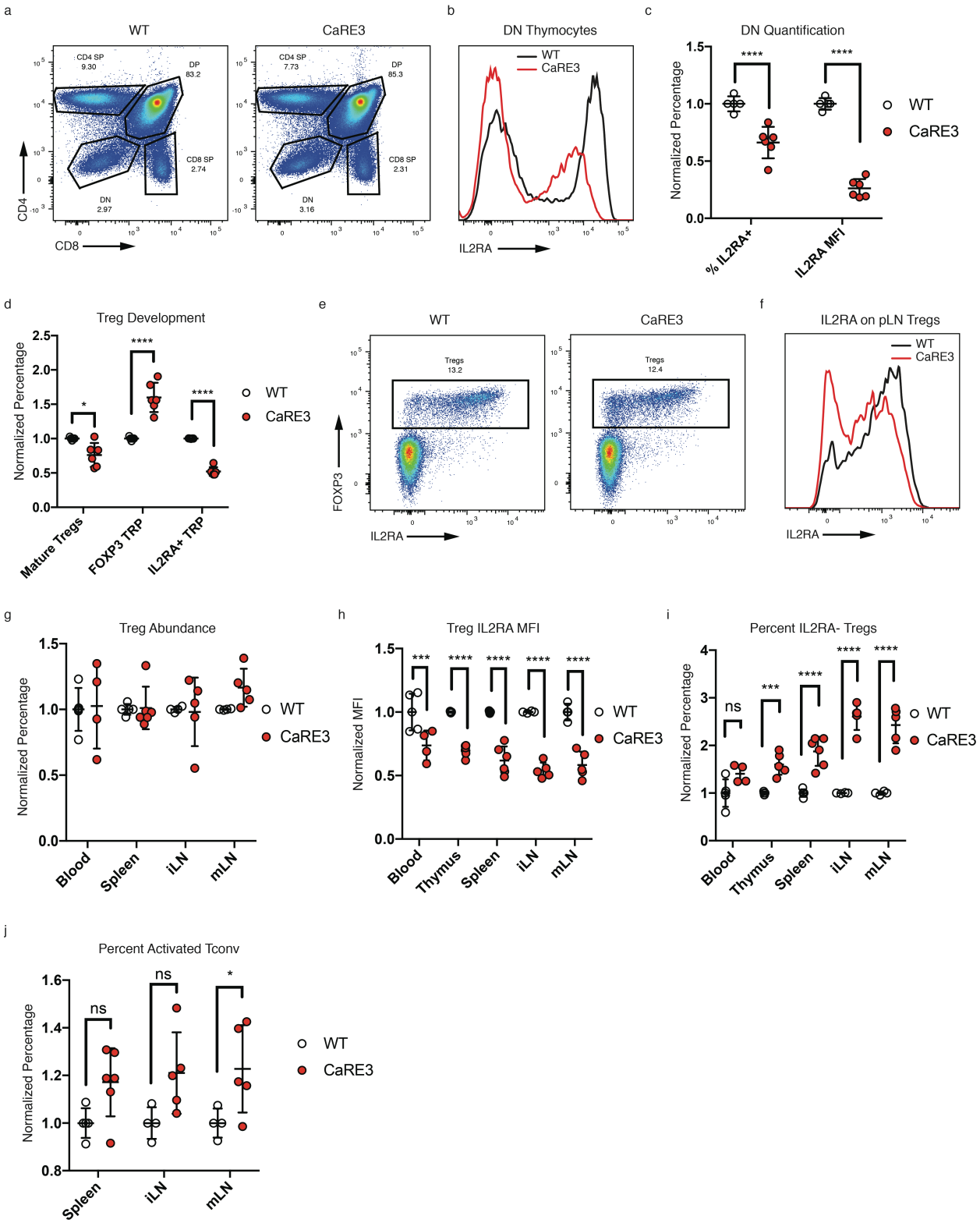


Figure 4.2. Immunophenotyping of Il2ra CaRE3 enhancer deletion on C57BL6 background. **a**, Representative staining of thymocyte populations gated on Live CD45+ cells in thymus. **b**, Histogram of Il2ra surface expression on double negative (CD4-CD8-) thymocytes. **c**,

Quantification of Il2ra surface expression on DN thymocytes. **d**, Quantification of Treg progenitors and mature Tregs in thymus. **e**, Representative Treg staining in CD4+ T cell compartment from inguinal LN. **f**, Histogram showing Il2ra surface expression on FOXP3+CD4+ Tregs in inguinal LN. **g**, Quantification of Treg percentages of CD4+ T cells in peripheral tissues, including blood, spleen, inguinal LN (iLN), and mesenteric LN (mLN). **h**, Quantification of Il2ra surface expression on Tregs in peripheral tissues. **i**, Quantification of percent Il2ra- Tregs in peripheral tissues. **j**, Quantification of activated Tconv cells (CD4+FOXP3-CD69+) in peripheral tissues. WT and CaRE3 littermates were used for experiments. None of these animals had elevated blood glucose levels. All data are presented as mean +/- s.d. and are representative of at least two independent experiments. *P $\leq$ 0.05, **P $\leq$ 0.01, ***P $\leq$ 0.001, ****P $\leq$ 0.001 by two-way ANOVA with multiple comparisons test.

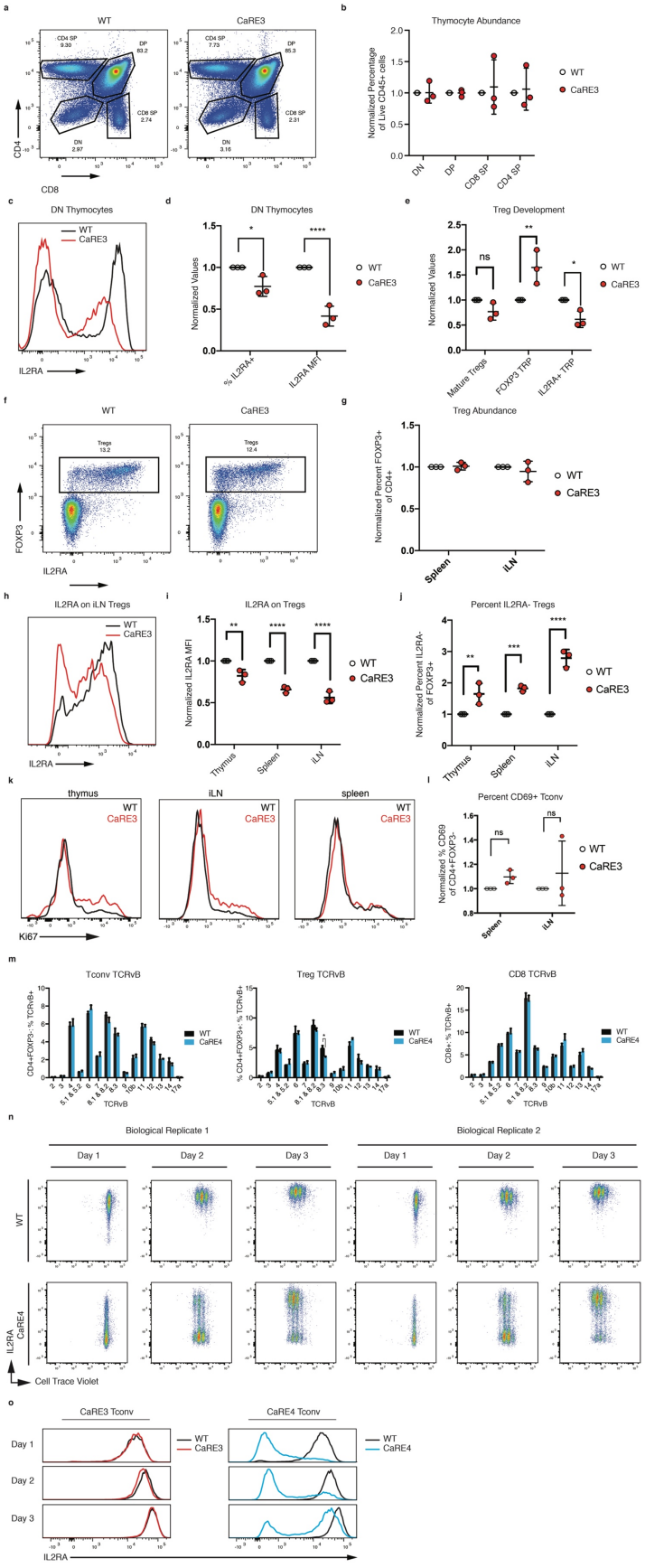

a
WT
CaRE3
CD4 SP 9.30
DP 83.2
CD8 SP 2.74
DN 2.97
CD4 SP 7.73
DP 85.3
CD8 SP 2.31
DN 3.16
CD4
CD8
b
Thymocyte Abundance
Normalized Percentage of Live CD45+ cells
DN
DP
CD8 SP
CD4 SP
WT
CaRE3
c
DN Thymocytes
IL2RA
d
DN Thymocytes
Normalized Values
% IL2RA+
IL2RA MFI
e
Treg Development
Mature Tregs
FOXP3 TRP
IL2RA+ TRP
f
Tregs 13.2
Tregs 12.4
FOXP3
g
Treg Abundance
Normalized Percent FOXP3+ of CD4+
Spleen
iLN
h
IL2RA on iLN Tregs
i
IL2RA on Tregs
Normalized IL2RA MFI
Thymus
Spleen
iLN
j
Percent IL2RA- Tregs
Normalized Percent IL2RA- of FOXP3+
k
thymus
iLN
spleen
Ki67
l
Percent CD69+ Tconv
Normalized % CD69 of CD4+FOXP3-
m
Tconv TCRvB
Treg TCRvB
CD8 TCRvB
CD4+FOXP3- % TCRvB+
% CD4+FOXP3+ % TCRvB+
CD8+ % TCRvB+
TCRvB
CaRE4
n
Biological Replicate 1
Biological Replicate 2
Day 1
Day 2
Day 3
IL2RA
Cell Trace Violet
o
CaRE3 Tconv
CaRE4 Tconv
IL2RA

Figure 4.3. Immunophenotyping of Il2ra CaRE3 and CaRE4 enhancer deletion mice on non-obese diabetic background. a, Representative thymocyte populations stained for CD4 and CD8 of Live CD45+ cells in CaRE3 enhancer deletion mice. **b,** Quantification of percentage of major thymocyte populations in a. **c,** Representative Il2ra surface expression on DN thymocytes from CaRE3 enhancer deletion mice. **d,** Quantification of percent Il2ra- DN thymocytes and Il2ra levels on Il2ra+ DN thymocytes on CaRE3 animals. **e,** Quantification of abundance of Treg progenitor and mature Treg populations in thymus from CaRE3 animals. **f,** Representative FOXP3 and Il2ra staining on CD4+ T cells from inguinal LN of CaRE3 animals. **g,** Quantification of Treg percentages in peripheral lymphoid organs from CaRE3 animals. **h,** Il2ra surface expression on Tregs from inguinal LN of CaRE3 animals. **i,** Quantification of Il2ra levels on Il2ra+FOXP3+ Tregs in CaRE3 animals. **j,** Quantification of percent Il2ra- Tregs in CaRE3 animals. **k,** Ki67 staining on Il2ra+ mature CD4+FOXP3+ Tregs from CaRE3 (n=1) and WT (n=1) littermate control. **l,** Quantification of % CD69+ cells in CD4+FOXP3- Tconv. **m,** Quantification of TCRvB repertoire on Tconv (CD4+ or CD8+) as well as CD4+FOXP3+ Tregs from spleen of CaRE4 (n=3) and WT (n=3) littermate pre-diabetic female mice. **n,** Il2ra activation versus proliferation time course of CaRE4 enhancer deletion naïve CD4+ T cells. **o,** Il2ra activation time course in CaRE3 and CaRE4 naïve CD4+ Tconv cells. Data in a-j and l from CaRE3 (n=3) and WT (n=3) littermate controls. All data are presented as mean +/- s.d. and are representative of at least two independent experiments. *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.001 by two-way ANOVA with multiple comparisons test.

We next wanted to map the relevant functional contexts for the IL2RA enhancers we had identified. To do this we examined chromatin accessibility and epigenetic profiling data across diverse primary human immune cells, with and without stimulation (Human Immune Atlas[243]; Epigenome Roadmap Project[151]). We found that the CaRE3 enhancer was accessible in multiple immune cell types, including T cells, and the accessibility at the site did not change in response to stimulation. In contrast, the CaRE4 enhancer showed increased accessibility upon stimulation in T and B cells, suggesting a shared stimulation response program. To further define the function of these enhancers we also looked at transcription factors (TFs) that bound to the enhancers in primary human T cells[244]. We found FOXP3 and STAT5 selectively bound to CaRE3, and not to CaRE4. Both of these factors are critical for Treg function and fitness. Motif analysis within CaRE4 predicted binding of the stimulation-responsive MEF2 TF (Figure 4.4). The predicted MEF2 binding site overlapped with the position of a disease-associated SNP, rs61839660 (Figure 4.4). Using electric mobility shift assay we find that MEF2 transcription factors bind the CaRE4 enhancer and that the SNP impairs binding (Figure 4.4). Whereas the disease-associated CaRE4

enhancer controls IL2RA induction in immune cells, the CaRE3 enhancer appears to control IL2RA levels on Tregs by integrating at least two different signals central to the function and fitness of this cell type.

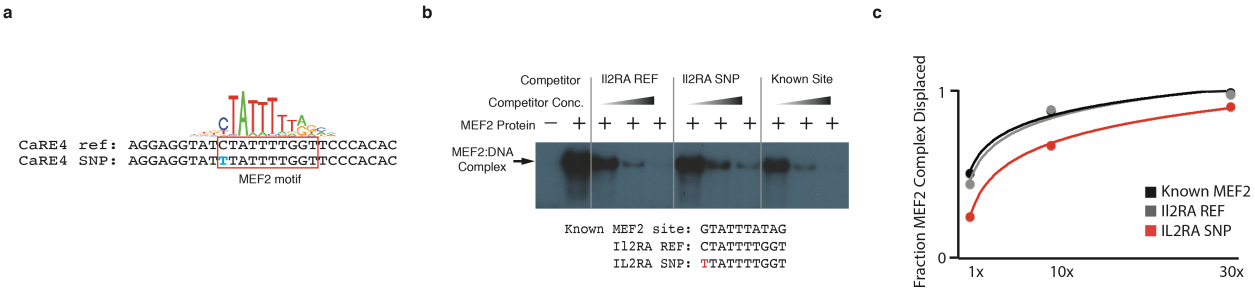


Figure 4.4. MEF2 binding at CaRE4 enhancer impaired by autoimmunity risk SNP, rs61839660. a, Identification of a predicted MEF2 binding site within the CaRE4 enhancer. **b,** Electric mobility shift assay of MEF2 protein complexed with radiolabeled WT CaRE4 sequence only or with increasing concentrations of unlabeled oligos CaRE4 reference (IL2RA REF), SNP (IL2RA SNP), or known MEF2 binding site sequence. **c,** Quantification of fraction of MEF2 complex displaced by competition with known MEF2 binding site, IL2RA reference, and IL2RA SNP oligos.

We next tested the functional cellular contexts of the conserved enhancers. Splenocytes from NOD CaRE3 and CaRE4 enhancer deletion mice were treated with anti-CD3/CD28, anti-IgM/CD40 or lipopolysaccharide to activate T cells, B cells, and dendritic cells. We measured IL2RA surface expression in the splenocyte culture by flow cytometry at different time points after stimulation. In CaRE4 EDEL cells, we found significantly impaired Il2ra levels on conventional T and B cells early after activation as compared to WT control cells. Unstimulated CaRE4 Tregs showed no obvious changes in Il2ra expression. In contrast, unstimulated CaRE3 Tregs exhibited impaired Il2ra expression similar to our *in vivo* findings. Upon stimulation, surface Il2ra on Tregs was rescued back to WT levels. Other cell types showed subtle changes in Il2ra expression when compared to WT littermate controls. Our data suggest that the CaRE4 enhancer is part of a shared stimulation-response program controlling the kinetics of Il2ra expression in Tconv and B lymphocytes. In contrast, the CaRE3 enhancer maintains Il2ra expression on steady state Tregs.

To validate our findings in mouse cells we also functionally tested enhancer function in human cells. Using pairs of Cas9 ribonucleoproteins (RNPs) we tiled deletions across the enhancers in Treg and Tconv cells (Figure 4.1, 4.5). We confirmed efficient editing by generating PCR fragments across the enhancer and identifying smaller amplicons consistent with the intended deletion size (Figures 4.1,4.5). Disruption of the conserved sequences within the CaRE3 enhancer showed reduced IL2RA levels in Tregs (Figures 4.1, 4.5). As expected disruption of adjacent sequences outside of the conserved site had no effect (Figure 4.1). Interestingly, when we tiled deletions in Tconv cells we observed similar effects on IL2RA expression which could largely be rescued by stimulating the cells (Figure 4.5. These effects may be a result of FOXP3 expression in human Tconv after stimulation, which does not happen in mouse T cells. To test CaRE4 function we used a human T cell line to generate deletions in unstimulated cells and then look for the effects on IL2RA expression following stimulation. Cells with deletions overlapping the conserved enhancer site showed impaired ability to turn on surface IL2RA (Figure 4.1). Again, as expected, deletions tiling adjacent sequences outside of the enhancer had no effect on IL2RA expression (Figure 4.1). Ongoing work is focusing on functionally testing the CaRE4 element in primary human T cells. However, these experiments are challenging due to the need to stimulate cells to achieve high-efficiency editing as well as continuous IL-2 cytokine supplementation of the culture. Taken together our data confirms the functional role for these conserved enhancers in human T cells, consistent with our findings in mouse cells.

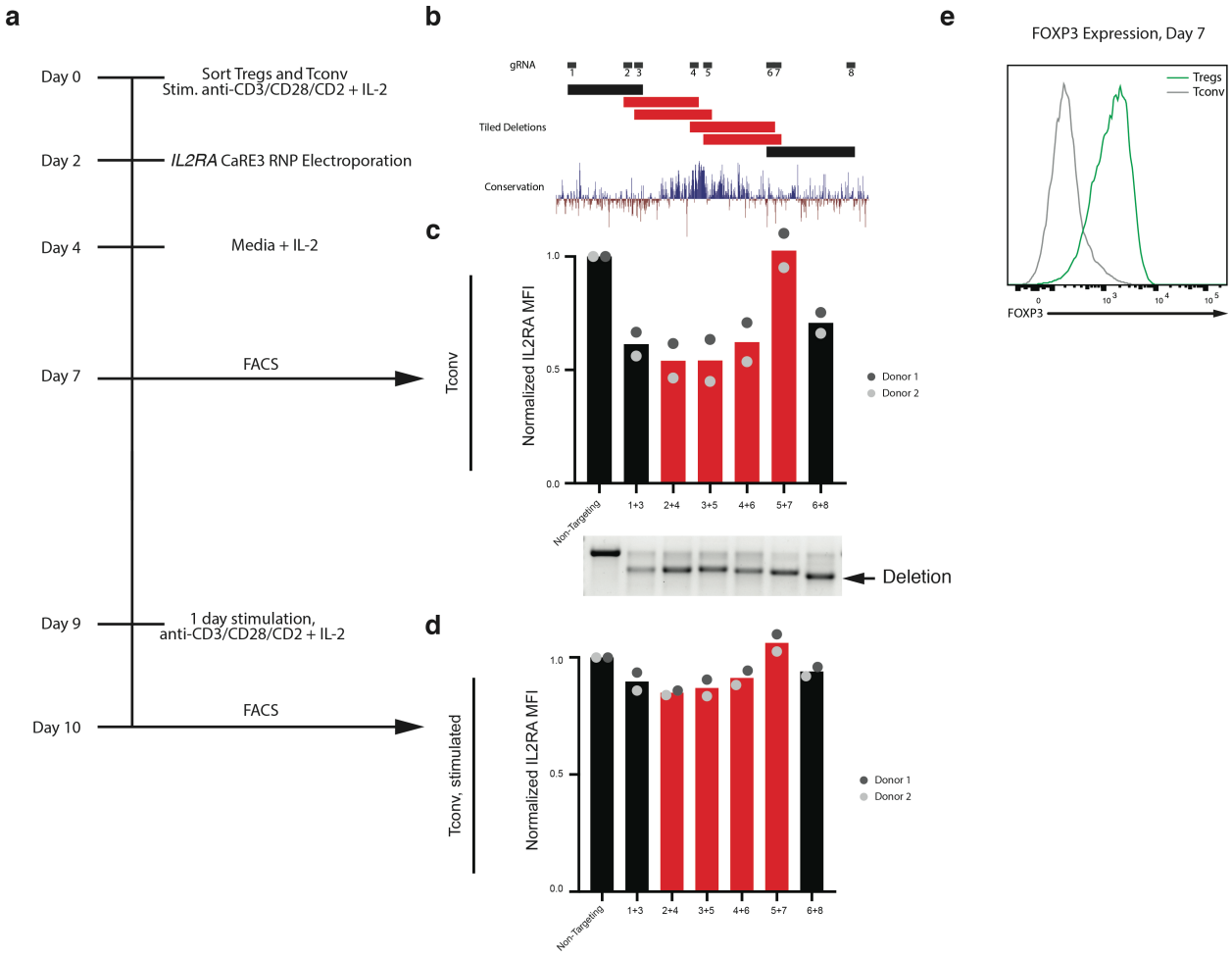


Figure 4.5. IL2RA CaRE3 deletion tiling using CRISPR in primary human T cells. a, Experimental workflow for primary human T cell editing. **b,** CRISPR/Cas9 RNP deletion tiling of CaRE3 enhancer. **c,** IL2RA surface expression on resting Tconv cells after tiling deletions within CaRE3 enhancer. PCR across enhancer was performed to assess editing and deletion efficiency. **d,** IL2RA surface expression on Tconv cells 24 hours after activation. **e**, FOXP3 staining in cultured Tconv and Tregs on Day 7.

Having mapped distinct cellular contexts for the Il2ra enhancers, we were interested in understanding their role in disease. NOD mice develop spontaneous autoimmune diabetes and are a commonly used model of human T1D. We aged CaRE3 and CaRE4 mice monitoring their blood glucose levels weekly to check for diabetes incidence. We found that homozygous CaRE4 EDEL animals were completely protected against diabetes (Figure 4.6). Remarkably, the human genetics predicted this result in an animal model as the CaRE4 enhancer harbors a T1D-protective SNP that

impairs enhancer function. Consistent with the idea that we had faithfully modeled a human T1D association we also observed the predicted dosage effect with heterozygous CaRE4 animals showing reduced incidence (Figure 4.6). Histology revealed homozygous CaRE4 EDEL mice had little or no islet immune infiltration (Figure 4.7). This was in stark contrast to islets in WT and heterozygous CaRE4 EDEL mice, which were highly infiltrated (Figure 4.7). Similar studies with CaRE3 EDEL mice are ongoing, with the expectation that these animals will have accelerated diabetes incidence. To attempt to break tolerance and induce diabetes in NOD CaRE4 EDEL mice we dosed animals with anti-PD1 antibodies[245]. Whereas all WT and most heterozygous CaRE4 EDEL mice developed diabetes by day 10 of treatment, none of the homozygous CaRE4 EDEL mice developed diabetes for the 30-day duration of the study (Figure 4.8). These studies establish the biological relevance of a human disease-associated enhancer in mice and directly link the kinetics of *Il2ra* expression on T and B cells to autoimmune diabetes.

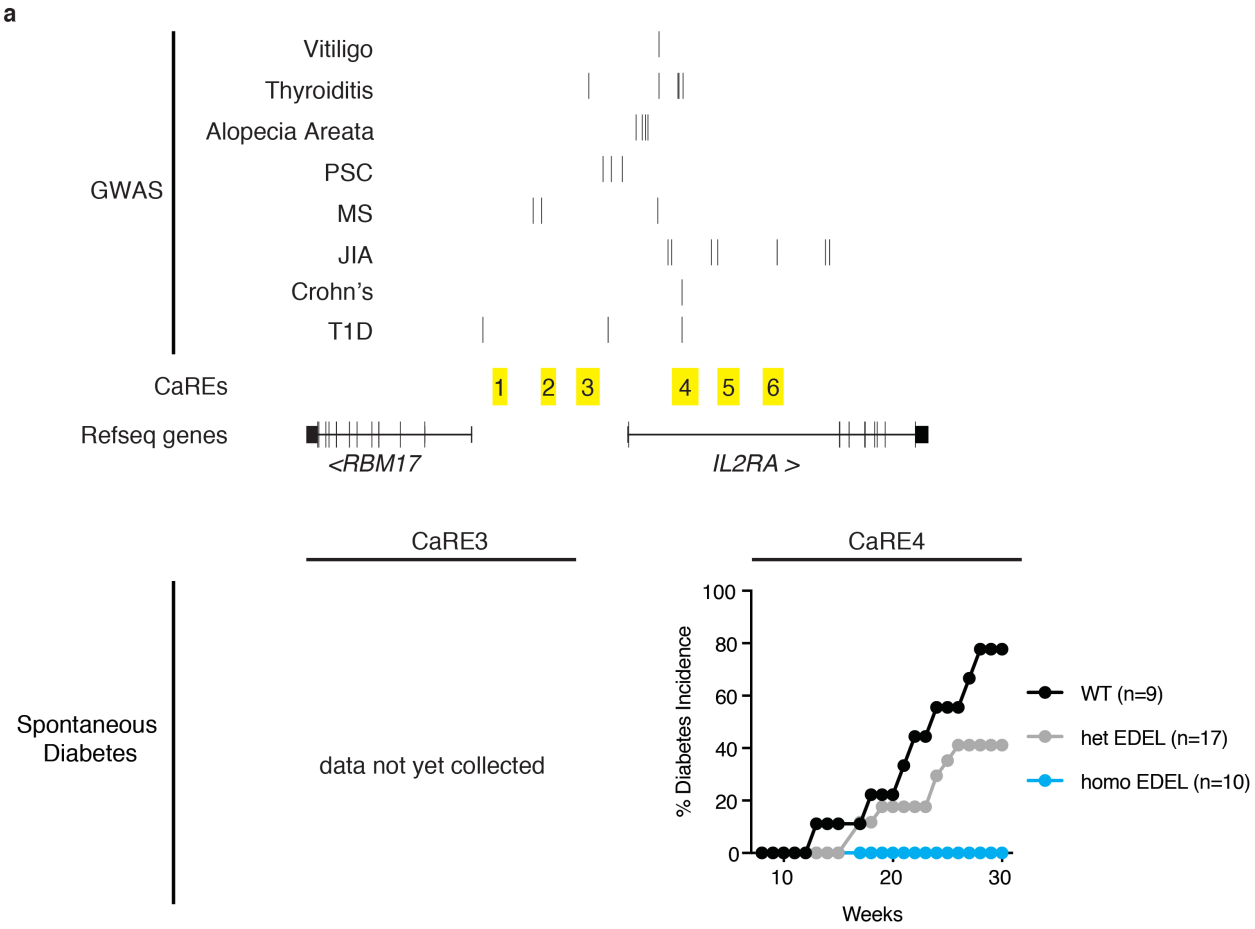


Figure 4.6. Il2ra CaRE4 enhancer protects against autoimmune diabetes. a, Fine mapped GWAS associations at the *IL2RA* locus (T1D[149], Crohn's Disease[158], and others[148]). **b**, Spontaneous diabetes incidence in NOD *Il2ra* CaRE3 and CaRE4 enhancer deletion mice. Numbers of animals followed are indicated in the figure.

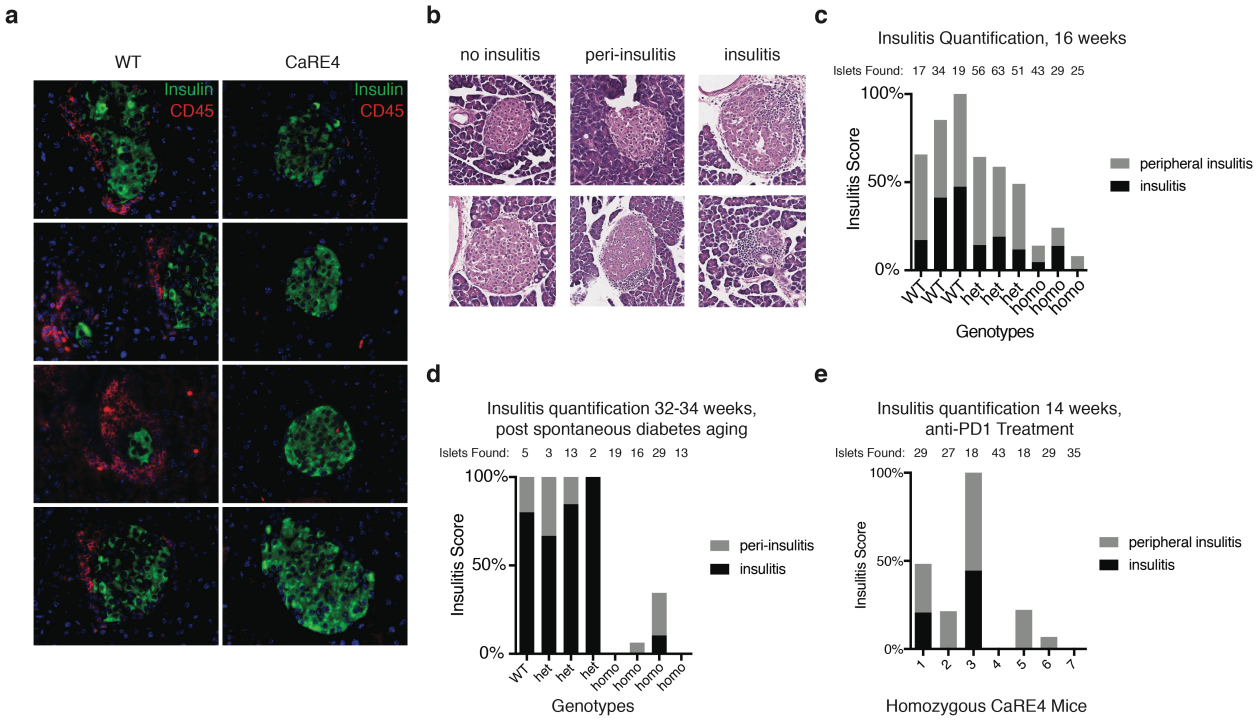


Figure 4.7. Islet characterization of NOD CaRE4 animals. **a,** Representative immunofluorescence staining for insulin (green) and CD45 (red) from CaRE4 EDEL (n=1) and WT littermate control (n=1) mice. **b,** Representative images showing islet scoring scale for islet infiltration quantification. **c,** Insulitis quantification in 16-week female WT and CaRE4 EDEL NOD mice. **d,** Insulitis quantification in 32-34-week female WT and CaRE4 EDEL NOD mice. **e,** Insulitis quantification of NOD CaRE4 EDEL mice after anti-PD1 treatment.

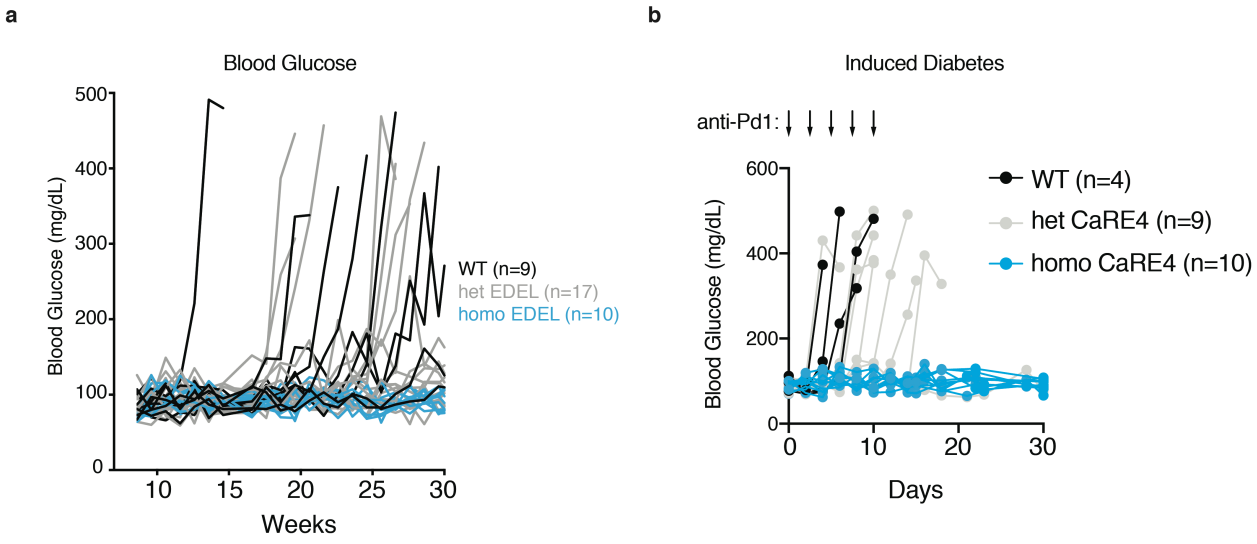


Figure 4.8. Spontaneous and induced diabetes in NOD CaRE4 EDEL mice. a, Once a week blood glucose measurements for CaRE4 spontaneous diabetes cohort. **b,** Blood glucose measurements every two days following anti-Pd1 treatment. Mice aged 8-10 weeks were used in

this study. All mice used in these studies were female and had normal blood glucose levels at the beginning of the studies.

The competition for IL-2 between CD4+ Tconvs and Tregs plays a central role in determining the immunological outcome of T cell activation. We examined how cell-type specific *Il2ra* expression defects caused by the deletion of Il2ra enhancers affect T cell competition for IL-2. To measure IL-2 competitiveness we stained for phosphorylated STAT5 (pSTAT5), a direct consequence of IL-2 signaling in pancreatic lymph (pLN) nodes of 6-week-old female NOD mice, an age well before these animals develop diabetes (Figure 4.9, 4.10). Even at these early time points we found evidence of dysregulated T cell competition for IL-2. In the WT setting, the percentage of pSTAT5+ Tregs outnumber pSTAT5+ Tconv cells consistent with Tregs outcompeting Tconv for IL-2 and tolerance (Figure 4.9, 4.10). In homozygous CaRE4 EDEL pLN, this balance is further shifted in favor of Tregs, with increased numbers of pSTAT5+ Tregs and fewer pSTAT5+ Tconv (Figure 4.9, 4.10). The opposite trend is observed in pLN from CaRE3 homozygous EDEL mice, where the number of pSTAT5+ Tconv cells are increased and pSTAT5+ Tregs are decreased as compared to WT littermate controls (Figure 4.9, 4.10). Our data is consistent with *Il2ra* CaRE3 and CaRE4 enhancers controlling immune responses during the early stages of T cell activation by tuning the expression of the high affinity IL-2 receptor.

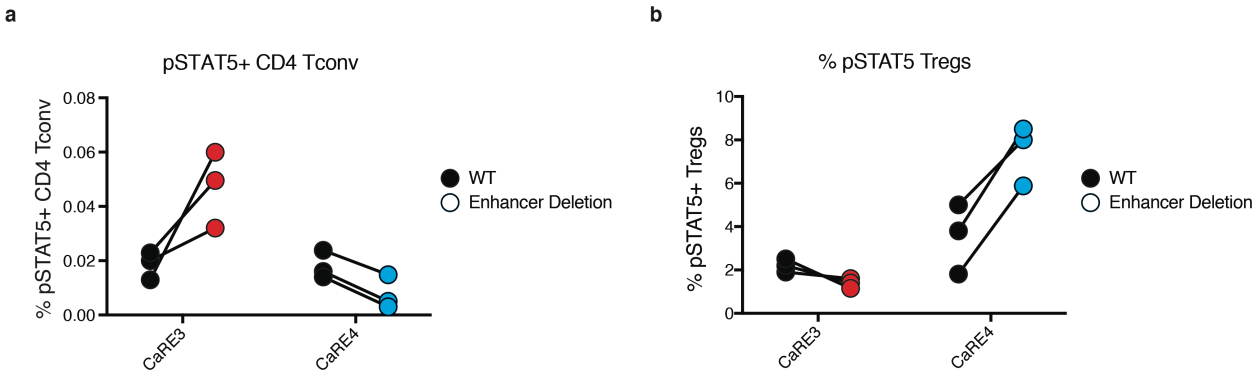


Figure 4.9. Il2ra CaRE3 and CaRE4 enhancers control T cell competition for IL-2. Quantification of pSTAT5 on **a,** Tconv (CD3+CD4+FOXP3-) and **b**, Treg (CD3+CD4+FOXP3+) cells in pLN from 6-week old pre-diabetic, female, NOD enhancer deletion mice. pLNs were

compared between EDEL (n=3) and WT (n=3) littermate controls. Data from two independent experiments.

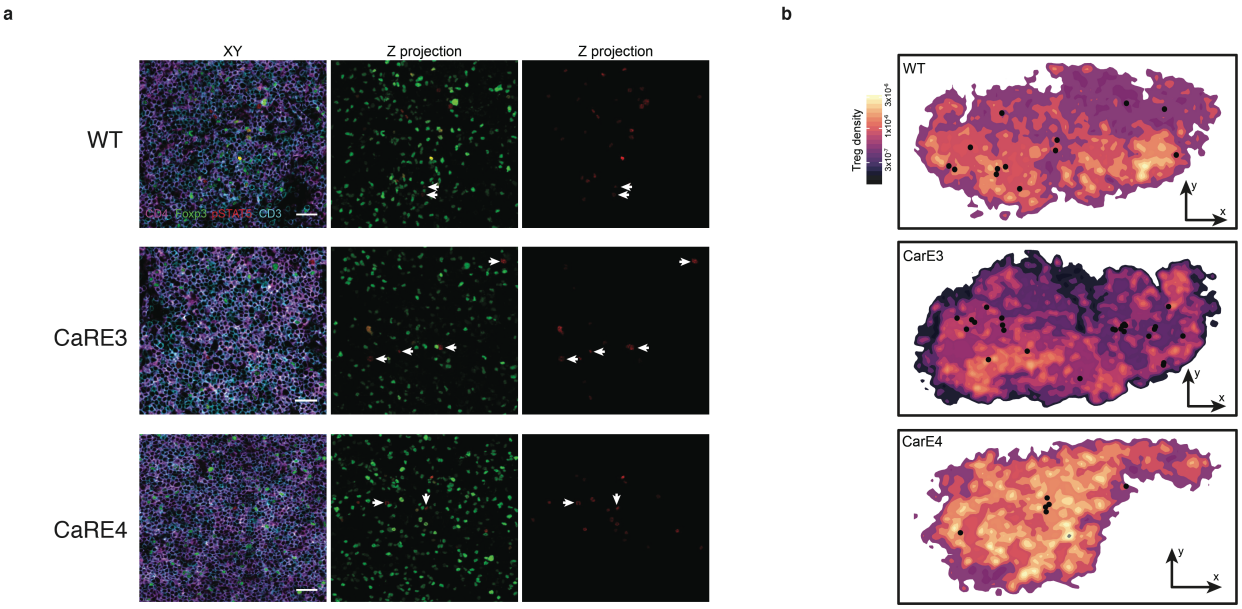


Figure 4.10. Characterization of pancreatic lymph node T cells in pre-diabetic female NOD mice. a, Immunofluorescent staining of pancreatic lymph node sections (pLN) from 6-week female NOD Il2ra enhancer deletion mice for CD3, CD4, FOXP3 and pSTAT5. **b**, Treg densities in pLNs from *Il2ra* CaRE3, CaRE4 and WT littermate control mice.

Discussion

Non-coding sequences in the *Il2ra* locus have been linked to at least 8 different autoimmunity disorders[148] reaffirming the importance of dissecting the gene regulatory circuits controlling this critical immune regulator. Here we show that two distinct *IL2RA* enhancers have opposing immunological functions due to cell type-restricted activities. The *IL2RA* CaRE3 enhancer maintains steady state IL2RA levels on anti-inflammatory Tregs, whereas the *IL2RA* CaRE4 enhancer controls the kinetics of IL2RA induction in response to stimulation. These enhancers fine tune the ability of CD4+ Tconvs and Tregs to compete for IL-2 thereby maintaining a balance between the proinflammatory and anti-inflammatory arms of the immune system.

Interestingly, we find IL-2-related cellular phenotypes in draining lymph nodes well before disease onset suggesting that earlier diagnosis of autoimmunity may be possible. We speculate that these enhancers are also important in determining immune function in the context of infection.

Here we successfully model a human T1D association in a small cohort of animals by amplifying the effects of a causal SNP through deletion of the disease-associated enhancer. Our data suggests that this enhancer controls a shared immune stimulation-response program to induce IL2RA expression. Further work is needed to identify the full set of immune cells within which this enhancer is active. Genetic risk for spontaneous T1D is also implicated in induced autoimmune diabetes. Consistent with this finding, CaRE4 mice, a model of a protective T1D association, were protected from anti-PD1 induced autoimmune diabetes. The characteristic histologic finding in these animals is little or no autoimmune infiltration in pancreatic islets. Our *in vivo* findings suggest that pre-existing islet infiltration may predict development of autoimmune diabetes.

Finally, our work supports ongoing clinical efforts in autoimmune and inflammatory disorders attempting to rebalance the immune system by increasing the number of patient Tregs. In general, mechanisms to limit IL-2 signals in Tconv cells may be therapeutic, including engineering improved Treg competition for IL-2 through overexpression of IL2RA or a higher affinity IL2RA receptor. Our findings also lay the foundation to rationally engineer non-coding circuits to control steady state and inducible gene expression for the next generation of safer and more effective immune cell therapies.

Methods

Mouse Generation

CaRE3 enhancer deletion mice were generated by electroporating Cas9 RNPs into mouse blastocysts. Briefly, equal volumes of 160uM CaRE3 1-4 crRNA (IDT) were mixed. An equal

volume of 160uM tracrRNA was then added and the gRNA components were allowed to hybridize by incubating at 37C for 10 minutes. We then added an equal volume of 40uM Cas9 protein (QB3 Macrolab) and incubated the solution at 37C for 10 minutes to allow the Cas9 RNP to complex. Mouse lines were established by backcrossing founders to wild type animals at least one generation before performing experiments. Genotyping was done by PCR to look for the expected enhancer deletion. In total four founder lines were established, one on the NOD background and three on the C57BL6 background. All founders were immunophenotyped and phenotypes were found to be consistent. CaRE4 mouse generation was previously reported[235].

crRNA	**Sequence**
crCaRE3-1	TACTCCCCTAATAAAGTACT
crCaRE3-2	TCCAGAGGGGCCCATCTGAC
crCaRE3-3	ACTTGAGATTCTGCCCTGTT
crCaRE3-4	ACTTGAGATTCTGCCCTGTT

Table 4.1. gRNAs used for NOD CaRE3 mouse generation.

Mouse experiments and data analysis

All mice were maintained in the UCSF-specific pathogen-free animal facility in accordance with guidelines established by the Institutional Animal Care and Use Committee and Laboratory Animal Resource Center. Experiments were done with animals aged between 1 and 4 months. Wildtype littermate mice were used as controls for all immunophenotyping experiments. All mice used for immunophenotyping were normoglycemic. No data were excluded from analysis. Power calculations were not performed and data were assumed to be normally distributed. Experiments were done without blinding or randomization. All data are derived from at least two independent experiments unless otherwise stated. Both male and female mice were used for experiments.

Mouse Immunophenotyping

Staining

Staining protocol of mouse cells were previously described[235]. The following staining panels were used:

Thymus
Surface and Viability:
Live/Dead-UVB
CD45-BUV395
CD8-PerCP-Cy5.5
CD4-BV605
CD25-PE-Cy7
CD69-APC
TCRB-PB
CD73-PE
Intracellular:
FOXP3-FITC

Spleen and inguinal LN
Surface and Viability:
Live/Dead-UVB
CD45-BUV395
CD8-PerCP-Cy5.5
CD4-BV605
CD25-PE-Cy7
CD69-APC
Intracellular:
FOXP3-FITC
Helios-PE
Rorgt-PB

Ki67-BV421 (Clone P56, BD) also used by replacing TCRB and RORgt in the respective panels.

Mouse anti-TCRvB screening kit (BD, Cat: 557004) was used to assess TCRvB usage in T cell subsets from splenocytes of 10 week male and female CaRE4 mice. For experiments that required culture of cells complete RPMI with 10% FBS buffered with HEPES. We also added sodium pyruvate, non-essential amino acids, penicillin streptomycin, and beta-mercaptoethanol to the media. Proliferation of Tconv cells was assessed using cell trace violet as per manufacturer's

protocol and stimulating approximately 100,000 cells with plate bound anti-CD3/CD28 antibodies. Il2ra time course on sorted naïve CD4+Il2ra-CD62L+CD44- Tconv cells was performed using 2ug/ml plate bound anti-CD3/CD28. Cells were stimulated for three days and surface Il2ra expression was measured by flow cytometry every day.

Splenocyte Stimulations

Splenocytes were plated at approximately 4 million cells per 500ul of complete RPMI in a well of a 48 well plate. Cells were activated using 1ug/ml anti-CD3/CD28, 20ug/ml anti-IgM 10ug/ml anti-CD40, and 100ng/ml LPS. Cells were stained at the end of the time course using the following antibody panels:

Antibodies for all panels:

Live/Dead-UVB

CD45-BUV395

TCRB-BV421

B220-APC-Cy7

CD25-PE-Cy7

Add for CD3/CD28 (T cell):

CD4-BV605

CD8-PerCP-Cy5.5

FOXP3-FITC

Add for LPS (DC):

CD11c-PerCP-Cy5.5

I-Ad-AF648 (APC)

Spontaneous Diabetes

Litters of female NOD mice from CaRE3 and CaRE4 mice set aside for the diabetes studies. Blood glucose was measured once a week from a drop of blood at the end of the tail. We measured blood glucose in animals starting at 8-10 weeks of age until 30 weeks of age. Once animals entered the study they were not removed for any reason. The diabetes endpoint was defined as two consecutive measurements of blood glucose over 200mmol/L. Diabetic animals were euthanized according to institutional standards.

Anti-PD1 Induced Diabetes

For this study we followed a previously published dosing regimen[245]. Briefly, anti-PD1 (Clone: RMP1-14, BioXCell, Cat: BE0146) was administered intraperitoneally to 8-10 week old female NOD mice. Animals were dosed with 500ug of antibody on the first day, and then an additional four injections of 250ug every other day until day 10. Mice were followed for a total of 30 days. The diabetes endpoint was defined as two consecutive measurements of blood glucose over 200mmol/L. Diabetic animals were euthanized according to institutional standards.

Pancreas Histology for Immune Infiltration

Briefly, animals were perfused with 1x PBS, the pancreas was isolated and placed into zinc formaldehyde fixative overnight at 4C. The next day, the pancreas was washed with water three times with gentle rocking for five minutes each. Three more washes with 1x PBS were done similar in a similar manner. Tissue was transferred to 70% EtOH and stored at 4C before being sent to

HistoWiz for processing and sectioning. H&E staining on sections from 32-week-old animals was done internally. Sections from 16-week-old animals were H&E stained at HistoWiz. Immune infiltration in pancreatic islets was scored as being normal, peri-insulitis, and insulitis. Examples of this are shown in Figure 4.7.

Lymph Node Imaging

Briefly, pancreatic lymph nodes (pLNs) were isolated from 6-week old female NOD mice. pLNs were fixed for 12-15hrs at 4C in 1% BD Fix/Perm Buffer and 1x PBS. pLNs were washed three times in 1ml PBS for 10 minutes at RT with gentle shaking. pLNs were shipping overnight for processing. After processing and section, pLNs were stained with anti-CD3, CD4, FOXP3, and pSTAT5.

Human T Cell Editing

CaRE4 enhancer deletion tiling was done in HuT78 cells. To generate tiled deletions we first combined equal volumes of two 160uM crRNAs. Then an equal volume of 160uM tracrRNA was then added. After incubating at 37C for 10 minutes, an equal volume of 40uM Cas9 protein (QB3 Macrolab) was added. To 7ul of RNP, 1ul of electroporation enhancer was added. 7ul of this Cas9 RNP mixture was added to approximately 400,000 cells in 18ul SE Buffer (Lonza) and electroporated using the Amaxa 4D (Lonza). Cells were rested in complete RPMI for 3 days. Cells were then stimulated with PMA (1:50000) and ionomycin (1:10000) or anti-CD3/CD28 antibodies (data not shown) overnight. One day later IL2RA surface expression was measured using flow cytometry. Enhancer editing was confirmed by PCR using hIL2RA-CaRE4DEL-F and hIL2RA-CaRE4DEL-R primers. Amplification of the endogenous enhancer was performed in edited cells looking for smaller fragments that were indicative of sequence deletion.

CaRE3 deletion tiling was performed in primary human Tconv (CD4+CD25-CD127+) and Tregs (CD4+CD25highCD127low) that were sorted from negative-selected CD4+ T cells from residuals. Cells were cultured in X-Vivo medium. On Day 0, approximately 1-4 million cells were plated in 6ml media in T25 flask. Tconv were supplemented with 30U/ml IL-2 whereas Tregs were supplemented with 300U/ml IL-2 (UCSF Pharmacy). Immunocult (CD3/CD28/CD2) was added at 25ul per 1 million cells to activate the cells. On day 2, approximately 500,000 cells in 18ul of P3 buffer were electroporated with 6ul of RNP and 1ul of electroporation enhancer using Amaxa 4D (Lonza), program EH115. On day 4, cells were split and the media and IL-2 were replaced. Tregs were cultured as per published Bluestone laboratory densities. Tconv cells were cultured at 1 million per milliliter for the rest of the experiment. On day 7, we assessed IL2RA surface expression by flow cytometry on Tregs and Tconv. We also checked FOXP3 expression using the eBioscience FOXP3 staining kit as per manufacturer's instructions. Finally, we assessed editing by PCR across the enhancer to look for smaller fragments indicative of deletions. On day 9, Tconv cells were split and the media and IL-2 were replaced. The cells were reactivated with immunocult overnight and IL2RA surface expression was measured by flow cytometry. Enhancer editing was confirmed by PCR using hCaRE3-PCR-F and hCaRE3-PCR-R primers. Amplification of the endogenous enhancer was performed in edited cells looking for smaller fragments that were indicative of sequence deletion.

crRNA	Sequence
hCaRE4-1	GAGATGTGCTGGGTTTTGCA
hCaRE4-2	AAGTTTAAAGAAGGAGTCGA
hCaRE4-3	GGAATGAACAAACCAATAGC
hCaRE4-4	GAGACAACCTCTTGTGTGGT
hCaRE4-5	AGTTGTTGCTACAGCAGTAC
hCaRE4-6	GCGCGTTTTCTCCCCTGGGA
hCaRE4-7	GCTGAGAGTACAGAAAGCAG
hCaRE4-8	TCTGAAGGAGGTATCTATTT
hCaRE4-9	GTATGGTGACTCACGCCCGG
hCaRE4-10	CCCCAGCAGCCACAAAACAG
hCaRE4-11	AGATGCCCCCAAGCTGCTCT
hCaRE4-12	CCTTGGGGTCTGGGTTCTCC
hCaRE4-13	GCAGGTTGAGGAGTGCCACG
hCaRE4-14	AGCCCCAGGTACATGCAGTG
hCaRE4-15	AAATAACTTCCCTGCTCACA
hCaRE4-16	CTGCTAAAGGTCAAAGACAA
sgCtrl-1	GGTTCTTGACTACCGTAATT
sgCtrl-2	TCGGATGTAAATTATGCCGT
hCaRE3-1	AGGTGGATCACTGGCTCTTG
hCaRE3-2	CCACCTGTGGGAATTTGGAT
hCaRE3-3	TTTGCTTGTTAAGTGGATGG
hCaRE3-4	TGGATAGATGTTGCTGAGAT
hCaRE3-5	GCTAATTCTACGAAATCTAC
hCaRE3-6	CAACTCAGTGCCGCAATAAC
hCaRE3-7	CACAACCCACAAGTCAACAG
hCaRE3-8	TTGTGCCACCCAGAAAACAC
hCaRE3-9	TGGAGGACGTGTTTCCTGGG
hCaRE3-10	GGGATAGCAGCAAAGCTCCC

Table 4.2. gRNAs used for Cas9 RNP deletion tiling of *IL2RA* CaRE3 and CaRE4 enhancers in human T cells.

Primer	Sequence
hIL2RA-CaRE4DEL-F	TCCTCTGTCTACCTGAAGTTCAAA
hIL2RA-CaRE4DEL-R	TGCCTGGTGAACTGTGGTTA
hCaRE3-PCR-F	ACTTCGCTCATGCACTCTTG
hCaRE3-PCR-R	CACAGACCACTGAGCTCCAA

Table 4.3. Primers for deletion efficiency check of *IL2RA* CaRE3 and CaRE4 enhancers.

CHAPTER 5. Concluding Remarks and Discussion

The work described here led to the development of new tools and methodologies to understand human genetic disease risk. My studies focused on the *IL2RA* locus as a paradigm for studying genetic risk for autoimmunity. Broadly, my work developed an understanding of *IL2RA* gene regulatory circuits and their role in autoimmune disease. I developed a high throughput CRISPR screening platform for the identification of functional enhancers. I generated a comprehensive functional enhancer map within the disease-associated *IL2RA* locus. Using CRISPR-engineered mice and human immune cells I characterized the function of two distinct enhancers at the IL2RA locus, one of which harbors a casual risk variant. Broadly, I found that the activity of these IL2RA enhancers are restricted to immune cells with opposing immunological functions. *In vivo* modeling showed that these enhancers tune competition for IL-2 in T cell subsets thereby playing a central role in immune tolerance. This work decodes a critical autoimmunity association, develops a cis-regulatory framework at the *IL2RA* locus, and causally links *IL2RA* gene regulation to autoimmunity. The tools and strategies developed in these studies can generally be applied to decode disease-associated loci in the human genome. The results of these studies raise interesting questions and future directions, which are discussed below.

Decoding Autoimmunity Associations

Disease associations identified by unbiased genetic studies have long held the promise to reveal disease mechanisms. Two notable examples illustrate this. Studies of a non-coding obesity association at the *FTO* locus lead to the identification of a new obesity gene by identifying a long-range chromatin enhancer-promoter loop[246,247]. Similarly, associations for fetal hemoglobin levels at the *BCL11A* locus revealed an enhancer that controls the switch between fetal and adult hemoglobin[153], which is now a therapeutic genome editing target for beta-thalassemia and other

hemoglobinopathies[156]. Although these studies were transformational to their respective fields, hundreds of genetic disease-associations remain that we do not understand.

Efforts to decode autoimmunity-associations have largely been frustrated because of the challenges in 1) identifying causal variants and 2) understanding their function. In recent years, larger cohort studies and the development of improved statistical algorithms have overcome the first challenge, fine-mapping disease associations to single causal variants for some loci[148,149,158]. However, understanding the function of disease-associated variants has been a much more difficult problem because the majority of these variants are in non-coding sequences – regions of the genome we still know relatively little about[148]. In my work I focused on developing methods to identify functional non-coding sequences within disease-associated loci and on strategies to understand their biological functions with human cells and mouse models.

Mapping Functional Enhancers

With the completion of the reference human genome, the focus shifted from assembling and cataloging our genetic sequences to understanding their function. In particular, large scale efforts were directed towards understanding non-coding sequences, which make up 98% the genome and harbor the majority of disease-associated variants. Chromatin profiling studies for epigenetic marks and transcription factors in diverse cell types and cell states led to the identification of hundreds of thousands of putative enhancers – transcription factor docking sites that loop to gene promoters and tune gene expression. These data provided a basic logical framework to think about non-coding sequences and disease-associated variants[248]. The next steps were to 1) identify functional enhancers, 2) connect them with a target gene, and 3) understand their biological activities.

Sequence perturbation is the gold standard for identifying functional enhancers. With the advent of CRISPR, high-throughput sequence perturbation could be used to saturate entire genomic loci to identify functional enhancers for a target gene. CRISPR cutting and inhibition were both shown to effectively identify functional enhancers[152,154]. However, these loss-of-function approaches were limited to the identification of active enhancers in the particular cell type or state that was being tested. Inactive enhancers are missed by this approach.

My work adapted the CRISPR-activation system for high-throughput functional enhancer screens[155]. Importantly, by "turning on" enhancers through recruitment of an artificial activation signal we were able to identify enhancers independent of their functional biological contexts. Ongoing work is expanding the repertoire of dCas9 fusion proteins for the identification of enhancers[129]. These CRISPR-based screening methodologies are now being used in combination with single cell RNA-sequencing, to map functional sequences, pair them with a target gene, and measure their effects on cellular transcriptional programs[144-146,249].

Autoimmunity Associations at the IL2RA Locus

IL2RA is a critical regulator of the immune system. Rare coding mutations in *IL2RA* cause immunodeficiency and variable autoimmunity due to consequences in cells that express the receptor, in particular regulatory T cells, which require IL2RA expression for fitness and function. However, much less is known about the biological effects of non-coding variants within the locus predicted to have subtle cell-type specific effects on gene expression.

Variants within the IL2RA locus have been linked to at least 8 different autoimmune disorders[148]. Remarkably, all of the autoimmunity variants are located in non-coding sequences. This led to the idea that autoimmunity variants at the IL2RA locus might land in enhancers that tune expression of the receptor in specific biological contexts. Consistent with this idea noncoding

single nucleotide polymorphisms (SNPs) have been shown to affect levels of IL2RA transcript[250,251], soluble IL2RA receptor[252], expression of IL2RA in specific immune cell types, and the abundance of IL-2RA–expressing memory T cells in the blood[253].

Our work focused on a causal autoimmunity SNP, rs61839660, which carries risk for Crohn's Disease[158] and simultaneously protects against T1D[149]. I showed that this SNP disrupts a stimulation-responsive enhancer that controls the timing of IL2RA induction on stimulated immune cells. Previous work suggested that autoimmunity risk was linked to stimulation-responsive transcriptional circuits. Our functional studies demonstrate that an autoimmunity variant dysregulates the kinetics of stimulation, having pleiotropic effects on cells in the immune system. Furthermore, I was able to model the protective human T1D association in an *in vivo* diabetes mouse model. My work has laid the foundation for mechanistic studies to mechanistically understand the protective human T1D association.

Defining Functional Contexts for Enhancers

Recent advances in genome engineering have enabled rapid functional testing of genetic sequences in animal models. Today, mouse generation is quick, on the order of months, requires little technical expertise, and can be done on virtually any genetic background[99]. This has greatly accelerated the pace of immunogenetics research, as the effects of genetic perturbations on the immune system can be quickly tested in animal models.

Sequence conservation is a prerequisite for using animal models to study human genetics. The *IL2RA* enhancers I studied were highly conserved between human and mouse. In addition, the position of the autoimmunity variant within the CaRE4 enhancer was also conserved in the mouse. This allowed me to generate multiple models – including the variant knock-in – to test the function

of these sequences. However, this is the exception as most disease-associated sequences lack conservation. This has spurred efforts to genetically engineer human immune cells.

Primary human immune cells are resistant to genetic manipulation. However, in 2015, we and others showed that electroporation of Cas9 ribonucleoprotein could be used to efficiently generate knock-out and knock-in primary human T cells[80,161]. This method has now been extended to additional immune cell types. This work promises to be transformational for human disease research. The function of disease-associated sequences can now be directly tested in different human cells and under different conditions. New methods are being developed to improve editing efficiency and allow for introduction of desired sequences of ever-increasing size through homology directed repair mechanisms. Beyond studying cells in a dish, we can also do functional studies on the human immune system by reconstituting animal models with human hematopoietic stem cells. Genetic manipulation of enhancer sequences in these cells will give us insights into effects on immune cell development and function.

Refining the Definition of CRISPR-activation Responsive Element

The definition of a CRISPR-activation responsive element was largely based on a best guess of the CRISPRa screen data. This is because at the time of the study we were unable to identify a peak calling algorithm to use with our dataset. The CaREs are broad regions of ~2-3kb of genetic sequence. Multiple sites of CRISPRa transactivation are evident within CaRE elements suggesting that some CaREs are composed of multiple enhancers. As such the regulation of the *IL2RA* locus is almost certainly more complex than we currently appreciate.

In my work, I identified enhancer function in *in vitro* enhancer reporter assays for 2 of 6 CaREs. The function of the remaining CaREs remains an open question. Interestingly, we also observed that CRISPRa did not transactivate gene expression from some sites that looped to the gene promoter, suggesting selectivity in the enhancers that were being identified. I looked at CTCF

binding as one potential explanation for this, but found no correlation. Perhaps additional transcription factors dictate this specificity. Future work will focus on mapping the functional contexts for additional enhancer elements at the *IL2RA* locus. In addition, testing their effects on nearby genes may reveal new functional contexts for these sequences.

Mapping *IL2RA* Trans-Regulators

In these studies I largely focused on understanding the cis-regulators of gene regulation by mapping enhancers for target genes and defining their functional contexts. However, transcription factors and protein regulators are key components of these regulatory circuits. To identify these important trans-regulators of gene expression I adapted a CRISPR cutting screen to screen ~1300 TFs for their effects on IL2RA expression. I sorted edited cells based on IL2RA expression and sequenced the guide RNAs from these cells to identify suppressive and maintenance factors of IL2RA expression in human T cells. This work laid the foundation for genome wide screens that can generate comprehensive genetic circuit maps for any gene. This approach is valuable for understanding gene regulation, disease risk, and identifying targets for engineering transcriptional programs.

Engineering Non-Coding Circuits

The focus of genome engineering has largely been on editing protein coding DNA sequences. Engineering non-coding circuits to obtain desired cellular phenotypes and functions remains a new and exciting opportunity. Numerous studies have identified functional non-coding sequences and characterized their molecular and cellular effects. One can leverage this information for therapeutic genome engineering. For example, disrupting the CaRE4 enhancer in adoptive Treg therapies may protect against autoimmunity risk from the activation of destabilized Tregs. In addition to

knocking-out sequences, we can also write them in. Knocking-in the *IL2RA* enhancer sequences may allow for exquisite control over gene expression – steady state or kinetics of induction in response to cell stimulation.

Final Thoughts

It is an exciting time for biomedical science. We are learning more about our biology than ever before, and rapidly translating that knowledge into clinical therapies. The ability to make desired changes to our DNA has revolutionized the field in the span of a few years. If the past is any indication, there is much to look forward to.

REFERENCES

1. Thomas, K. R., Folger, K. R. & Capecchi, M. R. High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44,** 419–428 (1986).
2. Jasin, M. & Berg, P. Homologous integration in mammalian cells without target gene selection. *Genes & Development* **2,** 1353–1363 (1988).
3. Orr-Weaver, T. L., Szostak, J. W. & Rothstein, R. J. Yeast transformation: a model system for the study of recombination. *Proceedings of the National Academy of Sciences* **78,** 6354–6358 (1981).
4. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences* **91,** 6064–6068 (1994).
5. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and Cellular Biology* **14,** 8096–8106 (1994).
6. Urnov, F. D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435,** 646–651 (2005).
7. Bibikova, M. *et al.* Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Molecular and Cellular Biology* **21,** 289–297 (2001).
8. Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326,** 1501–1501 (2009).
9. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **169,** 5429–5433 (1987).

10. Jansen, R., Embden, J. D. A. V., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43,** 1565–1575 (2002).

11. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, Engl.)* **151,** 2551–2561 (2005).

12. Mojica, F. J., Díez-Villaseñor, C., Soria, E. & Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* **36,** 244–246 (2000).

13. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60,** 174–182 (2005).

14. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading, Engl.)* **151,** 653–663 (2005).

15. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1,** 7 (2006).

16. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).

17. Hale, C., Kleppe, K., Terns, R. M. & Terns, M. P. Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *RNA* **14,** 2572–2579 (2008).

18. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).

19. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).

20. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513,** 569–573 (2014).

21. Szczelkun, M. D. *et al.* Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 9798–9803 (2014).

22. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **109,** E2579–86 (2012).

23. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).

24. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337,** 816–821 (2012).

25. Jasin, M. & Haber, J. E. The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. *DNA Repair (Amst.)* **44,** 6–16 (2016).

26. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).

27. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2,** e00471 (2013).

28. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826 (2013).

29. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8,** 2281–2308 (2013).

30. Vierstra, J. *et al.* Functional footprinting of regulatory DNA. *Nature Methods* **12,** 927–930 (2015).

31. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods* **10,** 1116–1121 (2013).

32. Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular Cell* **60,** 385–397 (2015).

33. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Research* **42,** 6091–6105 (2014).

34. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–949 (2014).

35. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351,** 84–88 (2016).

36. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529,** 490–495 (2016).

37. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550,** 407–410 (2017).

38. Kleinstiver, B. P. *et al.* Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol* **33,** 1293–1298 (2015).

39. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523,** 481–485 (2015).

40. Casini, A. *et al.* A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol* **36,** 265–271 (2018).

41. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* **32,** 267–273 (2014).

42. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343,** 84–87 (2014).

43. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* **343,** 80–84 (2014).

44. Schmid-Burgk, J. L. *et al.* A Genome-wide CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) Screen Identifies NEK7 as an Essential Component of NLRP3 Inflammasome Activation. *J. Biol. Chem.* **291,** 103–109 (2016).

45. Parnas, O. *et al.* A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* **162,** 675–686 (2015).

46. Platt, R. J. *et al.* CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159,** 440–455 (2014).

47. Shifrut, E. *et al.* Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function. *Cell* (2018). doi:10.1016/j.cell.2018.10.024

48. Ting, P. Y. *et al.* Guide Swap enables genome-scale pooled CRISPR-Cas9 screening in human primary cells. *Nature Methods* **15,** 941–946 (2018).

49. Parnas, O. *et al.* A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* **162,** 675–686 (2015).

50. Agrotis, A. & Ketteler, R. A new age in functional genomics using CRISPR/Cas9 in arrayed library screening. *Front Genet* **6,** 300 (2015).

51. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31,** 827–832 (2013).

52. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34,** 184–191 (2016).

53. Horlbeck, M. A. *et al.* Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *Elife* **5,** 2767 (2016).

54. Isaac, R. S. *et al.* Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *Elife* **5,** 1 (2016).

55. Knight, S. C. *et al.* Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **350,** 823–826 (2015).

56. Shin, H. Y. *et al.* CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nature Communications* **8,** 15464 (2017).

57. Li, J. *et al.* Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *J Mol Cell Biol* **7,** 284–298 (2015).

58. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* (2018). doi:10.1038/nbt.4192

59. Tsai, S. Q. *et al.* CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nature Methods* **14,** 607–614 (2017).

60. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33,** 187–197 (2015).

61. Chailleux, C. *et al.* Quantifying DNA double-strand breaks induced by site-specific endonucleases in living cells by ligation-mediated purification. *Nat Protoc* **9,** 517–528 (2014).

62. Giannoukos, G. *et al.* UDiTaS™, a genome editing detection method for indels and genome rearrangements. *BMC Genomics* **19,** 212 (2018).

63. Crosetto, N. *et al.* Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nature Methods* **10,** 361–365 (2013).

64. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).

65. Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* **32,** 577–582 (2014).

66. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24,** 1012–1019 (2014).

67. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32,** 279–284 (2014).

68. Liang, X. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of Biotechnology* 1–10 (2015). doi:10.1016/j.jbiotec.2015.04.024

69. Yin, H. *et al.* Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. *Nat Biotechnol* **34,** 328–333 (2016).

70. Petris, G. *et al.* Hit and go CAS9 delivered through a lentiviral based self-limiting circuit. *Nature Communications* **8,** 15334 (2017).

71. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat Chem Biol* **11,** 316–318 (2015).

72. Polstein, L. R. & Gersbach, C. A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat Chem Biol* **11,** 198–200 (2015).

73. Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H. & Sato, M. CRISPR-Cas9-based photoactivatable transcription system. *Chem. Biol.* **22,** 169–174 (2015).

74. Truong, D.-J. J. *et al.* Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Research* **43,** 6450–6458 (2015).

75. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nature Publishing Group* **33,** 139–142 (2015).

76. Harrington, L. B. *et al.* A Broad-Spectrum Inhibitor of CRISPR-Cas9. *Cell* **170,** 1224–1233.e15 (2017).

77. Moehle, E. A. *et al.* Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proceedings of the National Academy of Sciences* **104,** 3055–3060 (2007).

78. Campbell, C. R., Keown, W., Lowe, L., Kirschling, D. & Kucherlapati, R. Homologous recombination involving small single-stranded oligonucleotides in human cells. *New Biol.* **1,** 223–227 (1989).

79. Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Meth* **8,** 753–755 (2011).

80. Schumann, K. *et al.* Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 10437–10442 (2015).

81. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* 1–7 (2016). doi:10.1038/nbt.3481

82. Dever, D. P. *et al.* CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. *Nature* **539,** 384–389 (2016).

83. Wang, J. *et al.* Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. *Nat Biotechnol* **33,** 1256–1263 (2015).

84. Sather, B. D. *et al.* Efficient modification of CCR5 in primary human hematopoietic cells using a megaTAL nuclease and AAV donor template. *Sci Transl Med* **7,** 307ra156–307ra156 (2015).

85. Wang, J. *et al.* Highly efficient homology-driven genome editing in human T cells by combining zinc-finger nuclease mRNA and AAV6 donor delivery. *Nucleic Acids Research* **44,** e30–e30 (2016).

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

87. Hung, K. L. *et al.* Engineering Protein-Secreting Plasma Cells by Homology-Directed Repair in Primary Human B Cells. *Molecular Therapy* **26,** 456–467 (2018).

88. Leonetti, M. D., Sekine, S., Kamiyama, D., Weissman, J. S. & Huang, B. A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc. Natl. Acad. Sci. U.S.A.* **113,** E3501–8 (2016).

89. Roth, T. L. *et al.* Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature* **559,** 405–409 (2018).

90. Li, H. *et al.* Design and specificity of long ssDNA donors for CRISPR-based knock-in. *bioRxiv* 178905 (2017). doi:10.1101/178905

91. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540,** 144–149 (2016).

92. Nakade, S. *et al.* Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nature Communications* **5,** 5560 (2014).

93. Capecchi, M. R. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* **6,** 507–512 (2005).

94. Pelletier, S., Gingras, S. & Green, D. R. Mouse genome engineering via CRISPR-Cas9 for study of immune function. *Immunity* **42,** 18–27 (2015).

95. Yang, H. *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154,** 1370–1379 (2013).

96. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153,** 910–918 (2013).

97. Li, F. *et al.* Efficient genetic manipulation of the NOD-Rag1-/-IL2RgammaC-null mouse by combining in vitro fertilization and CRISPR/Cas9 technology. *Sci. Rep.* **4,** 5290 (2014).

98. Miyasaka, Y. *et al.* CLICK: one-step generation of conditional knockout mice. *BMC Genomics* **19,** 318 (2018).

99. Modzelewski, A. J. *et al.* Efficient mouse genome engineering by CRISPR-EZ technology. *Nat Protoc* **13,** 1253–1274 (2018).

100. Chen, S., Lee, B., Lee, A. Y.-F., Modzelewski, A. J. & He, L. Highly Efficient Mouse Genome Editing by CRISPR Ribonucleoprotein Electroporation of Zygotes. *J. Biol. Chem.* **291,** 14457–14467 (2016).

101. Chen, J., Du, Y., He, X., Huang, X. & Shi, Y. S. A Convenient Cas9-based Conditional Knockout Strategy for Simultaneously Targeting Multiple Genes in Mouse. *Sci. Rep.* **7,** 517 (2017).

102. Katigbak, A., Robert, F., Paquet, M. & Pelletier, J. Inducible Genome Editing with Conditional CRISPR/Cas9 Mice. *G3 (Bethesda)* **8,** 1627–1635 (2018).

103. Seki, A. & Rutz, S. Optimized RNP transfection for highly efficient CRISPR/Cas9-mediated gene knockout in primary T cells. *J. Exp. Med.* **215,** 985–997 (2018).

104. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152,** 1173–1183 (2013).

105. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–451 (2013).

106. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nature Methods* **10,** 977–979 (2013).

107. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods* **10,** 973–976 (2013).

108. Cheng, A. W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23,** 1163–1171 (2013).

109. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500,** 472–476 (2013).

110. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods* **10,** 973–976 (2013).

111. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31,** 833–838 (2013).

112. Hu, J. *et al.* Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. *Nucleic Acids Research* **42,** 4375–4390 (2014).

113. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159,** 635–646 (2014).

114. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nature Methods* **12,** 326–328 (2015).

115. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583–588 (2015).

116. Yeo, N. C. *et al.* An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nature Methods* **15,** 611–616 (2018).

117. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159,** 647–661 (2014).

118. Balboa, D. *et al.* Conditionally Stabilized dCas9 Activator for Controlling Gene Expression in Human Cell Reprogramming and Differentiation. *Stem Cell Reports* **5,** 448–459 (2015).

119. Black, J. B. *et al.* Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. *Cell Stem Cell* **19,** 406–414 (2016).

120. Liu, X. S. *et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167,** 233–247.e17 (2016).

121. Stepper, P. *et al.* Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. *Nucleic Acids Research* **45,** 1703–1713 (2017).

122. Xu, X. *et al.* A CRISPR-based approach for targeted DNA demethylation. *Cell Discov* **2,** 16009 (2016).

123. Choudhury, S. R., Cui, Y., Lubecka, K., Stefanska, B. & Irudayaraj, J. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* **7,** 46545–46556 (2016).

124. Morita, S. *et al.* Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat Biotechnol* **34,** 1060–1065 (2016).

125. Kearns, N. A. *et al.* Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nature Methods* **12,** 401–403 (2015).

126. Kim, J.-M. *et al.* Cooperation between SMYD3 and PC4 drives a distinct transcriptional program in cancer cells. *Nucleic Acids Research* **43,** 8868–8883 (2015).

127. Cano-Rodriguez, D. *et al.* Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. *Nature Communications* **7,** 12284 (2016).

128. Kwon, D. Y., Zhao, Y.-T., Lamonica, J. M. & Zhou, Z. Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nature Communications* **8,** 15315 (2017).

129. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* **33,** 510–517 (2015).

130. Klann, T. S. *et al.* CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol* **35,** 561–568 (2017).

131. Amabile, A. *et al.* Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell* **167,** 219–232.e14 (2016).

132. Fujita, T. *et al.* Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). *Sci. Rep.* **3,** 3171 (2013).

133. Liu, X. *et al.* In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell* **170,** 1028–1043.e19 (2017).

134. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155,** 1479–1491 (2013).

135. Roth, T. L. *et al.* Reprogramming human T cell function and specificity with non-viral genome targeting. *bioRxiv* 1–75 (2017). doi:10.1101/183418

136. Lackner, D. H. *et al.* A generic strategy for CRISPR-Cas9-mediated gene tagging. *Nature Communications* **6,** 10237 (2015).

137. Dalvai, M. *et al.* A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. *Cell Rep* **13,** 621–633 (2015).

138. Kuscu, C. *et al.* CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nature Methods* **14,** 710–712 (2017).

139. Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353,** aaf8729–aaf8729 (2016).

140. Komor, A. C. *et al.* Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv* **3,** eaao4774 (2017).

141. Hess, G. T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nature Methods* **13,** 1036–1042 (2016).

142. Ma, Y. *et al.* Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nature Methods* **13,** 1029–1035 (2016).

143. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. *Nature* **550,** 280–284 (2017).

144. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167,** 1853–1866.e17 (2016).

145. Jaitin, D. A. *et al.* Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167,** 1883–1896.e15 (2016).

146. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167,** 1867–1882.e21 (2016).

147. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nature Methods* **14,** 297–301 (2017).

148. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518,** 337–343 (2015).

149. Onengut-Gumuscu, S. *et al.* Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nature* **47,** 381–386 (2015).

150. Mouse ENCODE Consortium *et al.* An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol.* **13,** 418 (2012).

151. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).

152. Sanjana, N. E. *et al.* High-resolution interrogation of functional elements in the noncoding genome. *Science* **353,** 1545–1549 (2016).

153. Bauer, D. E. *et al.* An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* **342,** 253–257 (2013).

154. Fulco, C. P. *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354,** 769–773 (2016).

155. Simeonov, D. R. *et al.* Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549,** 111–115 (2017).

156. Canver, M. C. *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527,** 192–197 (2015).

157. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Molecular Cell* **66,** 285–299.e5 (2017).

158. Huang, H. *et al.* Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547,** 173–178 (2017).

159. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3,** e04766 (2014).

160. Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol* **33,** 73–80 (2015).

161. Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol* (2015). doi:10.1038/nbt.3290

162. Wu, C.-A. M. *et al.* Genetic engineering in primary human B cells with CRISPR-Cas9 ribonucleoproteins. *J. Immunol. Methods* **457,** 33–40 (2018).

163. Ren, J. *et al.* Multiplex Genome Editing to Generate Universal CAR T Cells Resistant to PD1 Inhibition. *Clin. Cancer Res.* **23,** 2255–2266 (2017).

164. Guo, M. H. *et al.* Comprehensive population-based genome sequencing provides insight into hematopoietic regulatory mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* **114,** E327–E336 (2017).

165. Chang, C.-W. *et al.* Modeling Human Severe Combined Immunodeficiency and Correction by CRISPR/Cas9-Enhanced Gene Targeting. *Cell Rep* **12,** 1668–1677 (2015).

166. De Ravin, S. S. *et al.* CRISPR-Cas9 gene repair of hematopoietic stem cells from patients with X-linked chronic granulomatous disease. *Sci Transl Med* **9,** eaah3480 (2017).

167. Flynn, R. *et al.* CRISPR-mediated genotypic and phenotypic correction of a chronic granulomatous disease mutation in human iPS cells. *Exp. Hematol.* **43,** 838–848.e3 (2015).

168. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31,** 839–843 (2013).

169. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31,** 822–826 (2013).

170. Boroviak, K., Fu, B., Yang, F., Doe, B. & Bradley, A. Revealing hidden complexities of genomic rearrangements generated with Cas9. *Sci. Rep.* **7,** 12867 (2017).

171. Wienert, B., Shin, J., Zelin, E., Pestal, K. & Corn, J. E. In vitro-transcribed guide RNAs trigger an innate immune response via the RIG-I pathway. *PLoS Biol.* **16,** e2005840 (2018).

172. Kim, S. *et al.* CRISPR RNAs trigger innate immune responses in human cells. *Genome Res.* **28,** 367–373 (2018).

173. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med* **24,** 927–930 (2018).

174. Ihry, R. J. *et al.* p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat Med* **24,** 939–946 (2018).

175. König, R. *et al.* Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* **135,** 49–60 (2008).

176. Brass, A. L. *et al.* Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **319,** 921–926 (2008).

177. Zhou, H. *et al.* Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* **4,** 495–504 (2008).

178. Park, R. J. *et al.* A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. *Nature* **49,** 193–203 (2017).

179. Jäger, S. *et al.* Global landscape of HIV-human protein complexes. *Nature* **481,** 365–370 (2011).

180. Hultquist, J. F. *et al.* A Cas9 Ribonucleoprotein Platform for Functional Genetic Studies of HIV-Host Interactions in Primary Human T Cells. *Cell Rep* **17,** 1438–1452 (2016).

181. Tebas, P. *et al.* Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med* **370,** 901–910 (2014).

182. Hütter, G. *et al.* Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N Engl J Med* **360,** 692–698 (2009).

183. Didigu, C. A. *et al.* Simultaneous zinc-finger nuclease editing of the HIV coreceptors ccr5 and cxcr4 protects CD4+ T cells from HIV-1 infection. *Blood* **123,** 61–69 (2014).

184. Ebina, H., Misawa, N., Kanemura, Y. & Koyanagi, Y. Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Sci. Rep.* **3,** 2510 (2013).

185. Hu, W. *et al.* RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 11461–11466 (2014).

186. Ophinni, Y., Inoue, M., Kotaki, T. & Kameoka, M. CRISPR/Cas9 system targeting regulatory genes of HIV-1 inhibits viral replication in infected T-cell cultures. *Sci. Rep.* **8,** 7784 (2018).

187. Marceau, C. D. *et al.* Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. *Nature* **535,** 159–163 (2016).

188. Ma, H. *et al.* A CRISPR-Based Screen Identifies Genes Essential for West-Nile-Virus-Induced Cell Death. *Cell Rep* **12,** 673–683 (2015).

189. Zhang, R. *et al.* A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. *Nature* **535,** 164–168 (2016).

190. Kim, H. S. *et al.* Arrayed CRISPR screen with image-based assay reliably uncovers host genes required for coxsackievirus infection. *Genome Res.* **28,** 859–868 (2018).

191. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* **12,** 252–264 (2012).

192. Burr, M. L. *et al.* CMTM6 maintains the expression of PD-L1 and regulates anti-tumour immunity. *Nature* **549,** 101–105 (2017).

193. Sen, D. R. *et al.* The epigenetic landscape of T cell exhaustion. *Science* aae0491 (2016). doi:10.1126/science.aae0491

194. Patel, S. J. *et al.* Identification of essential genes for cancer immunotherapy. *Nature* **548,** 537–542 (2017).

195. Pan, D. *et al.* A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. *Science* **359,** 770–775 (2018).

196. Manguso, R. T. *et al.* In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. *Nature* **547,** 413–418 (2017).

197. Rupp, L. J. *et al.* CRISPR/Cas9-mediated PD-1 disruption enhances anti-tumor efficacy of human chimeric antigen receptor T cells. *Sci. Rep.* **7,** 737 (2017).

198. Mariathasan, S. *et al.* TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554,** 544–548 (2018).

199. Roybal, K. T. & Lim, W. A. Synthetic Immunology: Hacking Immune Cells to Expand Their Therapeutic Capabilities. *Annu. Rev. Immunol.* **35,** 229–253 (2017).

200. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337,** 1190–1195 (2012).

201. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).

202. Canver, M. C. *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527,** 192–197 (2015).

203. Korkmaz, G. *et al.* Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol* **34,** 192–198 (2016).

204. Rajagopal, N. *et al.* High-throughput mapping of regulatory DNA. *Nat Biotechnol* **34,** 167–174 (2016).

205. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168,** 20–36 (2017).

206. Laguna, T. *et al.* New insights on the transcriptional regulation of CD69 gene through a potent enhancer located in the conserved non-coding sequence 2. *Mol. Immunol.* **66,** 171–179 (2015).

207. Ziegler, S. F., Ramsdell, F. & Alderson, M. R. The activation antigen CD69. *Stem Cells* **12,** 456–465 (1994).

208. Leonard, W. J., Krönke, M., Peffer, N. J., Depper, J. M. & Greene, W. C. Interleukin 2 receptor gene expression in normal human T lymphocytes. *Proceedings of the National Academy of Sciences* **82,** 6281–6285 (1985).

209. Kim, H. P., Imbert, J. & Leonard, W. J. Both integrated and differential regulation of components of the IL-2/IL-2 receptor system. *Cytokine Growth Factor Rev.* **17,** 349–366 (2006).

210. Fontenot, J. D., Rasmussen, J. P., Gavin, M. A. & Rudensky, A. Y. A function for interleukin 2 in Foxp3-expressing regulatory T cells. *Nat Immunol* **6,** 1142–1151 (2005).

211. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** 934–947 (2013).

212. Hnisz, D. *et al.* Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super- Enhancers. *Molecular Cell* 1–10 (2015). doi:10.1016/j.molcel.2015.02.014

213. Goudy, K. *et al.* Human IL2RA null mutation mediates immunodeficiency with lymphoproliferation and autoimmunity. *Clin. Immunol.* **146,** 248–261 (2013).

214. Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nature Methods* **13,** 919–922 (2016).

215. Huang, J., Ellinghaus, D., Franke, A., Howie, B. & Li, Y. 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. *Eur. J. Hum. Genet.* **20,** 801–805 (2012).

216. Ye, C. J. *et al.* Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* **345,** 1254665–1254665 (2014).

217. Laurence, A. *et al.* Interleukin-2 signaling via STAT5 constrains T helper 17 cell generation. *Immunity* **26,** 371–381 (2007).

218. Fujino, S. *et al.* Increased expression of interleukin 17 in inflammatory bowel disease. *Gut* **52,** 65–70 (2003).

219. Furtado, G. C., Curotto de Lafaille, M. A., Kutchukhidze, N. & Lafaille, J. J. Interleukin 2 signaling is required for CD4(+) regulatory T cell function. *Journal of Experimental Medicine* **196,** 851–857 (2002).

220. Chatenoud, L. & Bluestone, J. A. CD3-specific antibodies: a portal to the treatment of autoimmunity. *Nat. Rev. Immunol.* **7,** 622–632 (2007).

221. Kuhn, C. & Weiner, H. L. Therapeutic anti-CD3 monoclonal antibodies: from bench to bedside. *Immunotherapy* **8,** 889–906 (2016).

222. Klatzmann, D. & Abbas, A. K. The promise of low-dose interleukin-2 therapy for autoimmune and inflammatory diseases. *Nat. Rev. Immunol.* (2015).

223. Horlbeck, M. A. *et al.* Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *Elife* **5,** 2767 (2016).

224. Kampmann, M., Bassik, M. C. & Weissman, J. S. Functional genomics platform for pooled screening and generation of mammalian genetic interaction maps. *Nat Protoc* **9,** 1825–1847 (2014).

225. Marcel, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).

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

227. Sinha, R. *et al.* Index Switching Causes 'Spreading-Of-Signal' Among Multiplexed Samples In Illumina HiSeq 4000 DNA Sequencing. *bioRxiv* 125724 (2017). doi:10.1101/125724

228. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34,** 525–527 (2016).

229. Pimentel, H. J., Bray, N., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-Seq incorporating quantification uncertainty. *bioRxiv* 058164 (2016). doi:10.1101/058164

230. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12,** 357–360 (2015).

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

232. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32,** 347–355 (2014).

233. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529,** 490–495 (2016).

234. Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* **32,** 577–582 (2014).

235. Simeonov, D. R. *et al.* Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549,** 111–115 (2017).

236. Boroviak, K., Fu, B., Yang, F., Doe, B. & Bradley, A. Revealing hidden complexities of genomic rearrangements generated with Cas9. *Sci. Rep.* **7,** 12867 (2017).

237. Li, J. *et al.* Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *J Mol Cell Biol* **7,** 284–298 (2015).

238. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171,** 557–572.e24 (2017).

239. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18,** 821–829 (2008).

240. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).

241. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).

242. Liao, W., Lin, J.-X. & Leonard, W. J. Interleukin-2 at the crossroads of effector responses, tolerance, and immunotherapy. *Immunity* **38,** 13–25 (2013).

243. Calderon, D. *et al.* Landscape of stimulation-responsive chromatin across diverse human immune cells. *bioRxiv* 1–28 (2018). doi:10.1101/409722

244. Schmidl, C. *et al.* The enhancer and promoter landscape of human regulatory and conventional T-cell subpopulations. *Blood* **123,** e68–78 (2014).

245. Ansari, M. J. I. *et al.* The programmed death-1 (PD-1) pathway regulates autoimmune diabetes in nonobese diabetic (NOD) mice. *Journal of Experimental Medicine* **198,** 63–69 (2003).

246. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med* **373,** 895–907 (2015).

247. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507,** 371–375 (2014).

248. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337,** 1190–1195 (2012).

249. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Molecular Cell* **66,** 285–299.e5 (2017).

250. Qu, H.-Q. *et al.* A cis-acting regulatory variant in the IL2RA locus. *J. Immunol.* **183,** 5158–5162 (2009).

251. Dendrou, C. A. *et al.* Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource. *Nature* **41,** 1011–1015 (2009).

252. Maier, L. M. *et al.* IL2RA Genetic Heterogeneity in Multiple Sclerosis and Type 1 Diabetes Susceptibility and Soluble Interleukin-2 Receptor Production. *PLoS Genet* **5,** e1000322–8 (2009).

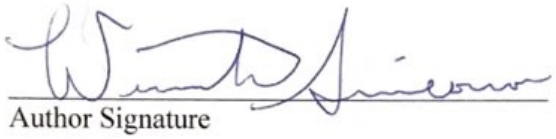
253. Orrù, V. *et al.* Genetic variants regulating immune cell levels in health and disease. *Cell* **155,** 242–256 (2013).

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

3/8/2019

Date