University of Cambridge

Clare College
Wellcome Sanger Institute

This dissertation is submitted for the degree of Doctor of Philosophy

Characterising the gene regulatory landscape of CD4+ T cells

Dafni Anna Glinos

External Examiner Julian C Knight
Nuffield Department of Medicine
Oxford University

Internal Examiner Klaus Okkenhaug
Department of Pathology
Cambridge University

Supervisors Gosia Trynka and Daniel Gaffney

September 2018

Characterising the gene regulatory landscape of CD4+ T cells

Dafni Anna Glinos

Despite the high prevalence of immune-mediated diseases, the molecular mechanisms by which they arise and the influence of genetic variation in the predisposition to disease are not well understood. Immune susceptibility loci identified by genome wide association studies (GWAS) overlap with active regulatory elements in CD4+ T cells, and particularly in regulatory T cells (Tregs). CD4+ T cells are the orchestrators of the adaptive immune response and their dysfunction has been associated with immune-mediated disorders through uncontrolled activation and resistance to downregulation, which is usually mediated by Tregs. T cell activation requires the combination of T cell receptor (TCR) recognition of an antigen and CD28 co-stimulation. The role of CD28 co-stimulation requirement in the activation of different T cell subsets has been understudied. Here, I assessed the role of immune disease variants in modulating pathways underlying T cell activation and Treg function. For that, I activated CD4+ T cells using different intensities of CD28 and TCR signals, followed by genome-wide transcriptome and chromatin profiling of naive and memory cells. I observed that CD28 plays a critical role in the expression of genes involved in effector functions, cell cycle regulation in memory T cells and in disease susceptibility. I profiled the gene expression regulatory landscape in Tregs using a combination of genomic assays. Due to the scarce Treg numbers in peripheral blood I first optimised the ChIPmentation (ChM) sequencing protocol to profile H3K4me3 and H3K27ac histone modifications in Tregs. I combined it with chromatin accessibility and gene expression profiling in resting and stimulated Tregs from ten donors. I observed cases of alternative transcription, such as alternative splicing and promoter, induced by stimulation, which could be predicted by changes in the chromatin landscape. Finally, I assessed how genetic variability impacts the function of Tregs and how this can lead to autoimmunity. I carried a quantitative trait locus (QTL) mapping using RNA-seq, ATAC-seq, H3K4me3 and H3K27ac ChM-seq data from Tregs isolated from 100 individuals. Additionally, I processed publicly available data from naive T cells to distinguish the Treg specific effects from generic CD4+ T cell signals. I recapitulated known colocalisations between QTLs and immune GWAS loci, and identified previously unknown Treg specific colocalisations. My findings highlight the value of carrying QTL studies in rare immune cell types relevant to the disease.

Declaration of Originality

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the introduction of each chapter. None of the contents in this dissertation have been submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit for the Biology Degree Committee.

Cambridge, September 2018

Dafni Anna Glinos

Acknowledgement

I am grateful for the inspiring working environment that is the Wellcome Genome Campus. Studying among world class scientists within state of the art facilities saw me benefit from a matrix conducive to thought provoking science.

I would like to express my deepest gratitude to my supervisor Gosia Trynka for shaping my scientific thought and for acting as a stabilising force on what has been a tumultuous journey. Her incisive approach to problems and contagious excitement for science has not only enabled me to flourish both personally and professionally but it has also made my PhD highly enjoyable. I am also appreciative of the members of the Trynka group - namely, Natalia, Lara, Blagoje and Eddie - all of whom have provided guidance and motivation when faced with scientific barriers. I thank my co-supervisor Dan Gaffney for challenging my analytical approach during our thesis committee meetings, and his PhD student, Kaur Alasoo, who readily shared his code with me and with whom I had fruitful discussions on a range of computational topics. I would like to thank my other co-supervisor, Sarah Teichmann, who was always at hand to provide me with invaluable feedback. I am also grateful to the other members of my thesis committee, John Todd and Tim Raine, for their insightful advice over the years. Finally, I would like to thank my various collaborators, Dave Sansom for introducing me to the world of co-stimulation, Rahul Roychoudhuri for igniting my interest in GARP, and Luke Jostins for his thoughtful comments on my results.

A PhD continues outside the institute, so I want to thank my PhD14 peers, my volleyball team and my Cambridge family for the support, distraction and the good times. I am most grateful to my actual family for their constant support, perspective and love. Finally, I want to thank Brian, for correcting my English and for always being around.

List of Abbreviations

actQTL Activity Quantitative Trait Locus

APC Antigen Presenting Cell

ATAC Assay for Transposase-Accessible Chromatin

caQTL Chromatin Accessibility Quantitative Trait Locus

CD Crohn's Disease

CEL Coeliac Disease

ChIP Chromatin Immunoprecipitation

CHO Chinese Hamster Ovary

DAR Differentially Accessible Region

DEG Differentially Expressed Gene

DHMR Differentially Modified Histone Region

DHS DNAse-Seq-Hypersensitivity Assay

eQTL Expression Quantitative Trait Locus

FACS Fluorescently Activated Cell Sorting

FDR False Discovery Rate

FRiP Fraction of Reads in Peaks

GWAS Genome Wide Association Study

HLA Human Leukocyte Antigen

IBD Inflammatory Bowel Disease

IPEX Immune Dysregulation Polyendocrinopathy, Enteropathy, X-linked Syndrome

JIA Juvenile Idiopathic Arthritis

LD Linkage Disequilibrium

MAF Minor Allele Frequency

MHC Major Histocompatibility Complex

MS Multiple Sclerosis

PBMCs Peripheral Blood Mononuclear Cells

PCA Principal Component Analysis

pLI Probability of Being Loss of Function Intolerant

PMA Phorbol 12-Myristate 13-Acetate

promQTL Promoter Quantitative Trait Locus

RA Rheumatoid Arthritis

SLE Systemic Lupus Erythematosus

SNP Single Nucleotide Polymorphism

SSc Systemic Sclerosis

T1D Type-1 Diabetes

Tcon Conventional T Cell

TCR T Cell Receptor

TF Transcription Factor

TFBS Transcription Factor Binding Site

Tfh Follicular Helper T Cell

Th Helper CD4+ T Cell

trQTL Transcript Ratio Quantitative Trait Locus

Treg Regulatory T Cell

TSS Transcription START SITE

UC Ulcerative Colitis

Contents

1 Introduction

Collaboration note

Parts of this chapter have been published as “Immunogenomic approaches to understand the function of immune disease variants” (Glinos et al., 2017). Many sections of the manuscript have directly been copied into this chapter.

It is estimated that around 20% of the population suffers from at least one autoimmune disease. Autoimmunity arises when the immune system fails to distinguish self from non-self, causing cells to respond against the antigens produced by the body. This can happen in specific tissues, such as the pancreas, where the immune system attacks the insulin-producing cells leading to the development of type-1 diabetes, or the lining of the joints, in the case of rheumatoid arthritis. Since the cells of the adaptive immune system are responsible for the recognition of different antigens, their study is of primordial importance for increasing our understanding of autoimmunity. Indeed, despite the high prevalence of the different disorders, the molecular mechanisms which predispose one to autoimmunity are not well understood. It is believed that, in the majority of cases, it is a combination of genetic and non-genetic factors that contribute to the risk of developing a disorder. Here, I will first describe the genetic architecture of autoimmune diseases, the complexity of which has hindered a deeper understanding of how diseases arise. I will then describe different immune cell types and pathways that are critical for the physiological functioning of the adaptive immune response. I will summarise the different genomic tools currently available for the study of gene expression and its regulation thereof. Finally, I will provide an overview of different studies that have attempted to link genetic variants associated to immune diseases with molecular measurements.

1.1 The problem: The complex genetic architecture of autoimmune diseases

1.1.1 Heritability of immune-mediated diseases

It is estimated that 1 in 5 people suffer from at least 1 of the 81 documented autoimmune diseases in the United States (Hayter and Cook, 2012). Despite the high prevalence, the molecular mechanisms which predispose to autoimmunity are not well understood. The clustering of autoimmune diseases in families has indicated a strong genetic component that underlies pathological processes driving many complex immune-mediated diseases. Indeed, siblings of an affected individual have significantly higher risk of developing an autoimmune disease compared to the general population. Interestingly, this is not disease specific since co-occurrence of autoimmune diseases in families is higher than expected by the population prevalence of the individual diseases (Eaton et al., 2007). Therefore, the clustering also indicates that the genetic component across different diseases is to a certain degree shared.

Most autoimmune diseases share a strong association to the major histocompatibility complex (MHC) region (Lenz et al., 2015). The MHC is one of the most polymorphic regions in the human genome, containing over 250 genes including the human leukocyte antigen (HLA) genes. The HLA genes encode for receptors expressed by antigen presenting cells (APCs). Early studies identified several loci in the MHC region with large effect sizes, including HLA-DQB1 associated with type-1 diabetes (T1D) (Todd et al., 1987), HLA-DQ2 and HLA-DQ8 associated with coeliac disease (CEL) (Sollid et al., 1989), and HLA-DR4 associated with rheumatoid arthritis (RA) (Nepom, 1998). It is now well appreciated that most of the autoimmune associations of the MHC region originate from these HLA genes (Matzaraki et al., 2017). Indeed, recent research highlights the complex genetic architecture of this locus, the non-additive effects between different HLA alleles, as well as the interactions between the different alleles modulating the risk to common autoimmune disorders (Hu et al., 2015a; Lenz et al., 2015).

However, the susceptible genetic background of the HLA alone is not sufficient to lead to the development of a complex immune-mediated disease. For example, T1D, RA, CEL and multiple sclerosis (MS), result from the combination of both the risk genotypes

of the HLA and non-HLA genes, as well as an environmental trigger. In order to study these effects before high-throughput genomic tools, such as whole genome genotyping and whole genome sequencing, became available, studies focussed on a few candidate genes within families of affected individuals. In T1D, a series of candidate gene studies identified *CTLA4*, which encodes for a protein receptor expressed on the surface of T cells, as a susceptibility gene (Nisticò et al., 1996). *CTLA4* was later also associated with patients with CEL (Djilali-Saiah et al., 1998) and RA (Plenge et al., 2007), highlighting the key roles of this T cell receptor in autoimmunity. That same study was the first that associated non-synonymous variants in *PTPN22* with RA, which was later shown to also contribute to the risk of T1D and systemic lupus erythematosus (SLE) (Criswell et al., 2005). While these represent examples of successful candidate gene studies, the majority of them resulted in non-reproducible results. This highlights the complexity underlying the genetics of autoimmune diseases and the human bias in choosing candidate genes for testing.

1.1.2 GWAS of immune-mediated diseases

The previously described limitations called for the development of an unbiased and comprehensive study approach, which emerged through genome wide association studies (GWAS). GWAS revealed that complex immune traits develop in consequence to an interplay between hundreds to thousands of common variants (Stahl et al., 2010), all with individually small effect sizes on the overall disease phenotype. Currently, more than 497 susceptibility loci for autoimmune disorders have been identified (Gutierrez-Arcelus et al., 2016). Despite the large number of mapped variants, the heritability explained by the non-HLA loci remains moderate. Even for the most successful examples such as MS, T1D or RA, where over a hundred risk variants have been mapped, the explained heritability varies between 20% for MS (Consortium and Others, 2013), 10% for T1D (Hu et al., 2015a) and about 5% for RA (Okada et al., 2014).

A picture emerging from GWAS is that immune-mediated diseases, to some extent, result from the dysregulation of the same biological pathways. For example, the initial T cell dysregulation observed in individuals with *CTLA4*-mutations was found to extend to an entire locus harbouring genes for receptors that control T cell activation, *CD28*,

ICOS and *CTLA4*, which is associated to CEL, RA and T1D. The sharing of genetic regions associated with immune diseases is widespread; of the 90 risk loci associated to T1D, RA, CEL and MS, 37% overlap between two or more diseases (Fortune et al., 2015). Despite the success in finding disease susceptibility loci, the molecular mechanisms by which these genetic variants hinder control of immune system and lead to autoimmunity have only been determined for a small number of variants. For example, missense mutations in the exon of *CTLA4*, induce severe immunodeficiency and a spectrum of autoimmune and autoinflammatory diseases (Schubert et al., 2014). These rare monogenic disorders have provided insights into the control of the immune system in the presence of the dysfunctional genes. However, the same genotype-to-function logic cannot be easily applied to all associations.

The challenges of understanding GWAS results for complex traits are both statistical and biological (Spain and Barrett, 2015). Among the statistical problems is that associated loci map to regions of the genome with extended linkage disequilibrium (LD). The LD blocks often comprise tens to hundreds of highly correlated single nucleotide polymorphisms (SNPs) that are co-inherited. Therefore, in a statistical test, the high correlation between SNPs results in equivalent strength of the association signal spread throughout all of the variants in LD. Practically, this renders the variants indistinguishable from one another and hinders the prioritisation of the causal variant based on the association statistics alone. Another problem is linking the associated SNPs to effector genes, as the causal variants may not necessarily affect the closest gene but instead act through long range genomic interactions. Finally, the majority of the associated variants localise to the non-coding regions of the genome, implicating that the disease variants are likely to act through dysregulation of gene expression. This poses a challenge because gene expression regulation can be highly cell type specific and therefore functional follow up studies have to be carried out in the cell types most relevant to the disease. However, for many immune diseases the exact pathological cell type is unknown.

1.2 The system: Immune cell types and their role in autoimmunity

1.2.1 CD4+ T cells within the adaptive immune system

Adaptive immunity is a branch of the immune system that is characterised by immunological memory and changes across an individual's lifespan depending on the pathogens to which they have been exposed. The adaptive immune system has two main players, B cells, which mature in the bone marrow, and T cells, which mature in the thymus. B cells and T cells sense their environment and communicate with each other via the expression of cytokines and chemokines, along with their receptors. These are small proteins whose concentrations within a cellular environment can determine the fate of the immune response.

T cells constitute the backbone of adaptive immunity and can be broadly divided into cytotoxic, which express CD8, and helper, which express CD4. Mature T cells that express T cell receptors (TCR+) are generated in the thymus, from where they are released into secondary lymphoid organs. Once in the secondary organs, the primary role of T cells is to recognise antigens derived from micro-organisms and orchestrate an immune response to fight them.

T cell maturation is a highly regulated process, in which the cells are screened for TCR reactivity to self-peptides bound to the MHC, and removed in case of high affinity. Cells with low affinity receive a weak TCR stimulation which contributes towards their maturation. The cells then undergo rounds of double positive selection through contact with MHC class I and II, and at the end of the process a cell only expresses either CD4+ (helper T cells (Th)) or CD8+ (cytotoxic T cells (Tc)) along with CD3 (Klein et al., 2014). Both CD4+ and CD8+ cells are characterised by the expression of CD3, which forms a complex with the TCR, and allows the cells to successfully establish contact with MHC presenting cells. When CD8+ T cells recognise an antigen and become activated they are able to induce apoptosis.

1.2.2 Signal transduction in T cells in response to stimulation

CD4+ T cell stimulation occurs in secondary lymphoid tissues where T cells interact with professional APCs, like dendritic cells, B cells and macrophages. Upon interaction, two coordinated signals are delivered to the T cell; the first is delivered via the TCR, which recognises antigen bound to MHC molecules, and the second is delivered via a co-stimulatory receptor. CD28 is the main co-stimulatory receptor expressed by T cells, and it interacts with CD80 and CD86 ligands on the APCs. The coordination of TCR and CD28 signals is important for T cell activation, proliferation, differentiation and survival (Figure 1.1). CD80 and CD86 can also be bound by CTLA-4 with higher affinity, as a means to dampen the immune response. CTLA-4 is expressed on the surface of all T cells upon stimulation and is the main pathway employed to downregulate the immune response upon completion. There are other co-stimulatory pathways, such as those mediated via ICOS, 4-1BB (encoded by *TNFRSF9*) and OX40 (encoded by *TNFRSF4*) receptors, all of which tend to be upregulated on the T cell surface upon stimulation. The combination of these signals critically affects the magnitude and fate of the T cell response. Therefore, co-stimulatory pathways provide a key checkpoint for controlling T cell responses, which is increasingly relevant therapeutically (Ford et al., 2014).

T cell activation promotes a number of signalling cascades that determine the fate of a cell. Activation is initiated via the phosphorylation of immunoreceptor tyrosine-based activation motifs (ITAMs) on the cytosolic side of CD3 by lymphocyte protein tyrosine kinase (Lck). Lck also interacts with CD28 C-terminal chain, and represents an important integration point for the two signals (Dobbins et al., 2016). Zeta-chain associated protein kinase (Zap-70) is then recruited to the TCR-CD3 complex where it becomes activated, promoting a series of phosphorylation events of adaptor and scaffold proteins. These result in the production of the second messengers diacylglycerol (DAG) and inositol trisphosphate (IP3). DAG activates protein kinase C (PKC) and the MAPK/Erk pathways, both promoting transcription factor (TF) NF-κB activation. IP3 indirectly promotes the entry of extracellular Ca2+ inside the cells, which in turn encourages *IL2* transcription by the NFAT and AP-1 TFs. IL-2 is a growth factor that initiates cell proliferation upon binding to the IL-2 receptor, which itself is composed of three chains, including the α chain, CD25, which is expressed on the surface of activated T cells. It is CD28 co-stimulation that regulates the expression and activity of NFAT and AP-1 TFs which in turn regulate

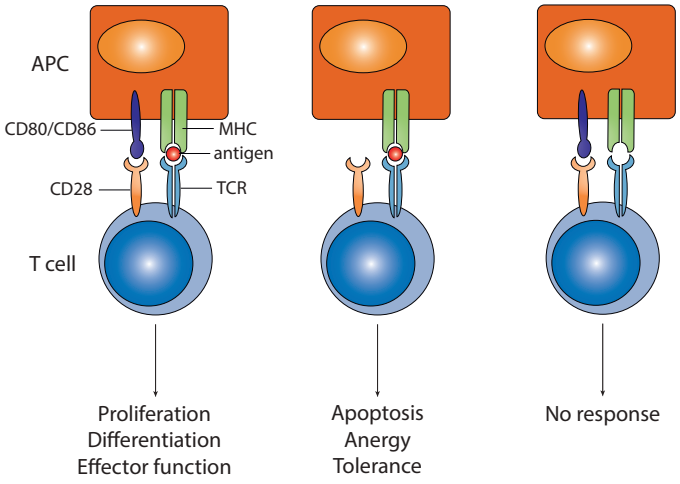

Figure 1.1: **T cell activation requires two signals to engage in cell proliferation and differentiation into effector functions.** In order to become activated, a CD4+ T cell needs an antigen presented by the major histocompatibility complex (MHC) to be recognised by the T cell receptor (TCR) and a B7 molecule (CD80 or CD86) to be recognised by the co-stimulatory receptor CD28. In the absence of a CD28 signal, T cells undergo apoptosis or become anergic. In the absence of the antigen a cell will not undergo any response.

the levels of IL-2 (Fraser et al., 1991; Shapiro et al., 1997). The Weiss laboratory identified a region in the *IL2* promoter that binds TFs in a CD28-dependent manner, therefore called CD28-response element (CD28RE).

Early investigations into TCR signalling relied on cloning techniques to identify the receptor responsible for the T cell identity (Hedrick et al., 1984; Yanagi et al., 1984). Subsequent signalling studies used cell line knock-outs and discovered they could stimulate T cells pharmacologically using phorbol esters, such as phorbol 12-myristate 13-acetate (PMA), and calcium ionophores, such as ionomycin (Weiss and Imboden, 1987). PMA is a small molecule that can diffuse through the cell membrane into the cytoplasm, where it can directly activate PKC and initiate the MAPK pathways, omitting the requirements for a surface receptor. Ionomycin complements PMA by triggering a calcium release which is necessary for NFAT signalling. This paved the way for a number of *in vitro* cell stimulation assays using PMA and ionomycin and antibodies against CD3. Once the important role of the CD28 co-stimulatory receptor in preventing cell anergy was identified (Figure 1.1 ; Jenkins et al., 1988), the most common stimulation method became a combination of anti-CD3 and anti-CD28 antibodies. These are commonly bound to magnetic beads in order to make their distribution and removal easier.

1.2.3 Helper T cell classification

Upon stimulation Th cells usually differentiate from the naive, that is a mature CD4+ T cell that has not encountered an antigen, into an effector state. Effector helper T cells are classified into Th1, Th2, Th17 and Tfh depending on the antigens present in the environment. Each subpopulation produces a different set of cytokines and promotes a different branch of the immune response (Zhu et al., 2010a). Th1 responses are critical for the defence against intracellular pathogens such as viruses and bacteria, and are characterised by the secretion of interferon-γ (IFN-γ). Th2 cells are important for controlling helminthic parasites and secrete interleukin-4 (IL-4), IL-5, and IL-13. Th17 cells regulate the host response against extracellular bacteria and fungi, and produce IL-17 (Zhu et al., 2010a). Tfh cells (follicular helper T cells) assist the B cells in the production of antibodies and express high levels of CD40 ligand (CD40L), IL-21 and IL-4. Another pathway is for CD4+ T cells to become T-regulatory (Treg) cells, which function in immune response homeostasis and downregulate the induction of effector T cells to avoid and reduce ongoing inflammation. CD4+ T cells can acquire a regulatory function either in the thymus, in which case they are referred to as natural Tregs (nTregs) or after activation in the periphery, referred to as induced Tregs (iTregs). Tregs exert their actions by producing the anti-inflammatory cytokines IL-10 and transforming growth factor-β (TGF-β) and uptaking the T cell growth factor IL-2 via high expression of its receptor, CD25, in order to limit its supply (Fontenot et al., 2005b). Tregs also express high levels of CTLA-4, which outstrips activating ligands from the surface of APCs, and is also necessary for the generation of Tregs (Read et al., 2000; Zheng et al., 2006).

Upon completion of the immune response, a small proportion of the responding effector cells survives to form antigen-specific memory T cells. Memory T cells retain elements of the previously induced state and are available to recreate a rapid defence in case of further antigenic challenge. Memory cells are characterised by the high expression of CD44 and CD45RO markers, while some of the cells are $CD62L^{high}$ and $CCR7^{high}$ (central memory T cells) and others are $CD62L^{low}$ $CCR7^{low}$ (effector memory T cells). Central memory T cells are found in the lymph nodes and the peripheral circulation and are capable of self-renewal, while effector memory T cells are found in the peripheral circulation and in the tissues and have a more specialised function. The molecular processes that define which cells will commit to become memory cells remain unknown. It has been

reported that increasing the strength of the TCR signal promotes an activated naive cell to become a memory cell (Williams et al., 2008). This is in contrast with other theories that support an asymmetric cell division model of naive cells while differentiating into memory and effector cells (Chang et al., 2007).

Given the importance of the CD28 pathway, a still largely unresolved question is the differential impact of CD28 co-stimulation at the transcriptional level in naive and memory T cells. Previous studies have suggested that memory T cells have lower co-stimulation thresholds or conversely, that naive cells have a greater requirement for co-stimulatory signals (Croft et al., 1994; Dubey et al., 1995; London et al., 2000). This has given rise to a widely perceived notion that, in contrast to naive, memory T cells do not require CD28 co-stimulation. However, there is evidence that this may not be the case in some settings, and that CD28 co-stimulation is important to numerous aspects of functional competence for previously primed T cells (Borowski et al., 2007; Linterman et al., 2014; Ndlovu et al., 2014; Fröhlich et al., 2016). These contrasting conclusions are likely influenced by the experimental systems used, as well as the nature and intensity of the TCR signal. Finally, different T cell processes may vary in their degree of dependence on CD28 co-stimulation, for example the requirement for CD28 to induce T cell proliferation appears lower than that for T follicular helper cell differentiation (Wang et al., 2015) which is strongly CD28-dependent.

1.2.4 T cell role in immune-mediated disease progression

The primary role of the immune system is to protect the host from infection by a variety of pathogens constantly present in the environment. As such, genetic defects that cause loss of the immune system's activity result in recurrent infections and severe immunodeficiency that is often life threatening. However, uncontrolled activation of immune cells may result in the response being targeted towards healthy cells causing chronic inflammation, tissue destruction, and eventually inflammatory diseases. Uncontrolled activation of the immune system can also occur when cells respond against the self-antigens and ultimately against the auto-antibodies, which would lead to autoimmunity. T cell related autoimmune and inflammatory diseases are characterised by an imbalance between effector T cells and functional Treg cells. Inadequate number of

Treg cells, defective Treg function, suppression resistant effector T cells or deficient T cell stimulation have all been associated with autoimmunity (Buckner, 2010).

The strongest evidence of the contribution of decreased Treg numbers to autoimmunity comes from patients with IPEX, who completely lack Treg cells due to different mutations in the *FOXP3* gene (Bennett et al., 2001). Furthermore, mice with Foxp3 mutations (*scrufy*) display systemic autoimmunity (Fontenot et al., 2005a). FoxP3 is the hallmark TF of Tregs and acts mostly as a repressor to downregulate genes involved in Treg cell activation (Fontenot et al., 2005b). However, in patients with common autoimmune diseases the number of circulating Tregs are more variable compared to IPEX patients, which makes them complex to study. This is exacerbated by the difficulty in isolating tissue specific Tregs, since FoxP3 is an intracellular protein. High levels of CD25 and low levels of CD127, the α receptor for IL-7, are typically used to isolate Tregs from a CD4+ cell population, to achieve a pure Treg cell population, where more than 95% will express FOXP3 (Liu et al., 2006). For the same reasons, as well as the absence of a reliable assay, Treg cell function is also complicated to study. Even when enough cells are obtained, the functional assays are set up *in vitro*, which might not accurately represent the processes *in vivo*, since Treg cells show decreased proliferation *in vitro* compared to *in vivo*. Treg function is usually assessed through suppression assays, first described 20 years ago (Takahashi et al., 1998), where the authors observed that co-culturing Tregs with CD4+CD25- conventional T cells (Tcons), led to the decreased proliferation of activated Tcons, measured through flow cytometry. Since then, numerous slight variations of the protocol have been proposed, such as measuring the cytokine milieu using ELISA (Nakamura et al., 2001) and changing the stimulation from anti-CD3 and anti-CD28 to different antibodies, such as glucocortoid induced TNF receptor (GITR) (Shimizu et al., 2002; McHugh et al., 2002). Using such assays a recently published study by the Todd group demonstrated that Tregs from SLE patients have lower levels of CD25, which affects their suppressive phenotype and decreases their survival (Ferreira et al., 2017b).

The functional defect at the source of an immune attack could also lie within the remaining T cell populations; naive, memory and effector. Different mechanisms by which T cells can become resistant to Treg suppression have been described. These vary between diseases and include modifications of intracellular signalling pathways, which can result in changes in the T cell activation threshold, and exposure to extracellular

signals such as a specific cytokine milieu or an activation signal. Early studies that provided strong co-stimulatory signals via the anti-CD28 antibody led to Tcon cells resisting Treg suppression (Takahashi et al., 1998) and highlighted co-stimulatory pathways as an important component of acquired resistance. Another co-stimulatory receptor, OX40, has also been associated with many immune-mediated diseases, including SLE, where its levels in memory cells were elevated and the cells were resistant to Treg suppression (Kshirsagar et al., 2013). Interestingly, polymorphisms in the gene encoding for its ligand (*TNFSF4* or OX40L) led to abnormal levels of its transcripts (Graham et al., 2007), which have been suggested to promote a Tfh profile (Jacquemin et al., 2015). The role of activated memory cells in immune diseases has also been observed in patients with juvenile idiopathic arthritis (JIA), where their synovial fluid had significantly elevated numbers of memory cells that were fully differentiated and active, which made them resistant to downregulation by Tregs (Haufe et al., 2011). Finally, autoimmune diseases are characterised by the overproduction of inflammatory cytokines, and a number of interleukins have been found to lead to Tcon resistance. The most notable example is IL6, which has elevated levels in JIA, RA, SLE, inflammatory bowel disease (IBD) and MS. Tcons from the peripheral blood of MS patients were observed to be highly proliferative and characterised by increased IL6 signalling through pSTAT3. Blocking of STAT3 phosphorylation led to increased suppression of these cells by Tregs (Schneider et al., 2013). Resistance acquirement can also be studied via suppression assays, and the development of carboxyfluorescein succinimidyl ester (CFSE) and CellTrace proliferation dyes marked an important step to be able to separately stain Tregs and Tcons within a co-culture to distinguish between the two proliferative profiles. In these assays it is difficult to distinguish between acquired resistance and loss of Treg function, which makes the establishment of appropriate controls crucial.

T cells continue to play a role in autoimmunity as the initial immune response escalates. After the self-antigen has initiated a reaction it is difficult to eradicate it since it leads to increased tissue damage and the emergence of new antigens. As cells are recruited to the affected site, they secrete cytokines inducing an inflammatory environment. In fact, targeting specific cytokines consists of one of the main strategies to control inflammation in immune-mediated diseases (Ishihara and Hirano, 2002; Taylor et al., 2009; Papoutsaki and Costanzo, 2013). Inflamed tissue is characterised by abnormal proportions of different immune cells when compared to ratios in their healthy coun-

terparts. For example, the inflamed skin of patients with psoriasis (PSO) has increased numbers of Tregs which however produce the pro-inflammatory cytokine IL-17 leading to increased inflammation (Bovenschen et al., 2011). In contrast, in lesions of SLE patients, Treg cells display increased sensitivity to cell death mediated by the death receptor CD95. This sensitivity leads to a decrease in their numbers the extent of which correlates with increased clinical severity of the flare (Miyara et al., 2005).

Tregs have the potential to resolve an autoimmune reaction. Effector cells produce IL-2 which leads to the activation and expansion of tissue resident Tregs and the formation of new Tregs (Knoechel et al., 2005). Depending on the reason for the development of the disease, the generation or maintenance of these Tregs might be defective, in which case there will be no disease resolution.

1.3 The tools: Genome wide assays for functional profiling of immune-disease loci

1.3.1 RNA sequencing

RNA sequencing (RNA-seq) is a widely used method to quantify the dynamic transcriptome of a cell in a genome wide manner. There are three steps involved in preparing a sample for RNA-seq. Firstly, RNA has to be isolated from the cells. In the case of T cells, this can be difficult as they are characterised by a small diameter and therefore reduced input material. Secondly, to obtain messenger RNA (mRNA) as the measure for the transcriptomic state of the cells isolated, the RNA has to be filtered. This step is necessary since the majority of all cellular RNA is ribosomal. In eukaryotes, filtering for mRNA is usually achieved either using a depletion method or poly-A selection. If this step is performed using poly-A selection, it can result in variation in read coverage across the gene body and introduce 3' bias (Lahens et al., 2014). Finally, the filtered mRNA is reverse transcribed to yield cDNA libraries. cDNA is amplified using polymerase chain reaction (PCR) to enrich for specific fragments and attain a desirable concentration necessary for loading onto the sequencing machine. PCR can also result in biases, since it preferentially amplifies sequences with high content of G and C nucleotids (Benjamini and Speed, 2012). In order to address different biases in RNA-seq, multiple

experimental and computational approaches have been developed to ensure that the generated libraries are of high complexity.

Conventionally, once the RNA-seq libraries have been sequenced, the individual reads are aligned to a reference genome using a splice aware algorithm, such as STAR (Dobin et al., 2013). The genomic coordinates of transcripts and different transcript isoforms are maintained in public databases such as GENCODE (Harrow et al., 2012). Mapped reads that overlap annotated genes are counted using tools such as featureCounts (Liao et al., 2014). More recently, pseudoalignment methods for gene quantification, such as Salmon (Patro et al., 2017), were developed. They bypass the requirement of a reference genome alignment step by quantifying expression levels directly based on the transcriptomic sequences, leading to a dramatic reduction in the computational time required. Furthermore, since these methods directly quantify reads, they can be used to assess the ratios of different transcripts present in a cell.

The majority of human genes express more than one transcript per gene, which increases their potential pool of proteins by 10-fold on average (Nilsen and Graveley, 2010). Transcripts can differ in their functionality and subcellular localisation, with the most notable example being the *PTPRC* gene encoding for the CD45 protein, a cell surface tyrosine phosphatase (Michie et al., 1992). Naive T cells express CD45RA while memory T cells express CD45RO, an isoform which lacks three exons.

Despite large efforts by databases such as Ensembl (Aken et al., 2017), a lot of transcripts remain unannotated along with the mechanisms by which they arise. Alternative transcripts can arise from alternative promoters, alternative splicing and alternative polyadenylation. To address this, one can use methods such as LeafCutter, which focuses on alternative transcription events directly, looking at the reads that span two exons, commonly referred to as exon-exon junctions (Li et al., 2018).

1.3.2 Chromatin state profiling

Gene expression regulation results from the interplay between gene enhancers and promoters. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) assesses DNA-protein interactions by pulling down DNA regions of the genome that are bound

by a protein of interest. ChIP-seq has been used to annotate the activity of non-coding regions of the genome through the presence of different post-translational histone modifications. Four core histones are used to pack the DNA of a cell within a structure called the nucleosome. Post-translational modifications of histones reflect changes in chromatin structure that is often coupled with accessibility of different proteins, such as TFs, and thus regulate gene expression. Histone H3 is the most modified histone. The fourth lysine of H3 can be mono- or tri-methylated (H3K4me1 or H3K4me3) denoting enhancers and promoters, accordingly. The 9th lysine, when trimethylated marks constitutively repressed genes while when acetylated it denotes actively transcribed promoters. The 27th lysine of H3 can be trimethylated (K27me3), which is a signal tagging inactivate enhancers and promoters, or acetylated (K27ac), which is a signal for active regulatory elements. Finally, H3K36me3 locates in the bodies of actively transcribed genes (ENCODE Project Consortium, 2012).

There have been large efforts led by international consortia to comprehensively annotate the non-coding sequences of the genome across a wide range of cell lines and primary cell types (ENCODE Project Consortium, 2012; Roadmap Epigenomics Consortium et al., 2015; Stunnenberg et al., 2016). Currently, more than 150 different cell types have been annotated using the above marks for the presence of promoters, enhancers and repressive sequences, and often are also complemented by the annotation of binding sites for specific TFs.

Similarly to RNA-seq data, ChIP-seq data is typically aligned to a reference genome using a standard aligner, such as bwa (Li and Durbin, 2009). The next step consists of identifying regions that show a higher pile-up of reads which would correspond to more protein binding or chromatin accessibility than the background. These are typically referred to as “peaks” and are identified using peak calling algorithms such as MACS2 (Zhang et al., 2008). Depending on the protein being assayed, the resulting peaks can be characterised as narrow, when spanning a few hundred base-pairs (e.g. a TF ChIP-seq), or broad, when spanning up to tens of thousands of base-pairs (e.g. H3K27me3 ChIP-seq), and a different algorithm has to be used when dealing with each.

However, ChIP-seq typically requires material in the order of millions of cells, an amount prohibitive for rare cell type population studies. These requirements can be overcome by

a recent modification of the standard ChIP-seq protocol, Chipmentation-seq (ChM-seq). In ChM a series of adapter ligation and purification steps during the library preparation stage is replaced by the more straight-forward transposase (Tn5) mediated sequencing adapters addition during the immunoprecipitation step (Schmidl et al., 2015). This new protocol results in reduced overall time, costs and input requirements and has been shown to yield reliable results with low cell numbers, as few as 10,000 cells for some histone modifications. Such low inputs have been shown to work in the cell line K562 (Schmidl et al., 2015) and innate lymphoid cells (Lim et al., 2017).

Tn5 has in fact already been used in the assay for transposable-accessible chromatin followed by sequencing (ATAC-seq) protocol, which identifies open chromatin regions (Buenrostro et al., 2013). Chromatin accessible regions determined by ATAC-seq replicated regions identified by a more established method, the DNAse-seq-Hypersensitivity (DHS) assay (Boyle et al., 2008). DHS is a challenging assay and for this reason it hasn't been applied at a large population scale. However, ATAC-seq has been successfully used in human population studies due to its easy implementation and low cell number requirement (Alasoo et al., 2018; Gate et al., 2018). The only initial limitation of ATAC-seq was the sheer number of resulting mitochondrial reads. A recent optimisation of the protocol led to decreased percentage of the mitochondrial reads captured in a sample, which has allowed to decrease the sequencing depth required (Sos et al., 2016).

1.4 The goal: Functional fine-mapping of disease variants

1.4.1 Importance of context specificity for the study of autoimmunity

The annotation of regulatory chromatin for enhancers and promoters has provided an opportunity to interpret the role of non-coding disease variants. The development of statistical approaches that integrate GWAS SNPs with histone marks allowed the prioritisation of disease relevant cell types (Trynka et al., 2013; Farh et al., 2015; Pickrell, 2014) (Figure 1.2). Enrichment of GWAS variants in cell-type-specific promoters marked by H3K4me3, confirmed the importance of CD4+ T cell subsets in a number of autoimmune disorders. CD4+ memory T cells and regulatory T (Treg) cells showed high enrichment for variants associated with a number of diseases, including RA, IBD and

CEL (Trynka et al., 2013). This observation converges with the previous immune studies pointing towards impaired function of Treg cells in autoimmune diseases (Buckner, 2010).

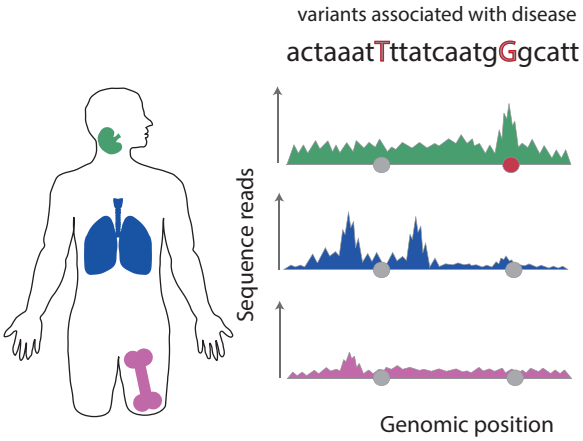

Figure 1.2: **Causal disease variants overlap cell type specific chromatin marks.** Chromatin marks from tissues of interest (left) can be obtained by chromatin immunoprecipitation followed by sequencing (ChIP-seq) and chromatin accessibility assays. These assays yield reads which are mapped against the reference genome to find the genomic location of the chromatin marks. Regions enriched in reads for the assayed chromatin marks can be recognised as 'peaks' in the genomic position plot of reads (right). Such genomic annotations generated from different tissues [lymph nodes (green), lungs (blue), femur (pink)] provide a valuable roadmap of cell type specific genome activity. The genome annotations can be overlapped with genetic variants associated with a phenotype of interest, such as an autoimmune disease, represented as grey circles. If a statistically significant proportion of associated variants overlaps with peaks specific to a cell type it can point towards disease-relevant tissue and prioritize causal variants. Here, we illustrate a single-associated locus where only one single nucleotide polymorphism (SNP; red circle) overlaps with a peak specific to the lymph nodes and absent from the other two tissues.

Since active histone marks tend to localise near the genes whose expression they control, gene expression measurements themselves can also be used to prioritise a specific cell type or condition (Slowikowski et al., 2014; Calderon et al., 2017). Using gene expression measurements across different conditions also allows the identification of potential key genes driving the enrichment in that cell type and could play a role in disease pathogenesis. Using this statistical approach, Hu and colleagues identified the critical cell types for different autoimmune diseases: transitional B cells for SLE, epithelial associated stimulated dendritic cells for Crohn's disease (CD) and CD4+ effector memory T cells for RA (Hu et al., 2011). Following this observation, they isolated high purity CD4+ T effector cells and measured relative cell abundance, as well as the gene expression of 215 genes located in RA loci, and cell proliferation capacity upon *in vitro*

stimulation with anti-CD28/anti-CD3 beads (Hu et al., 2014). They identified a group of genes whose basal level could predict the proliferative potential of CD4+ effector memory cells, along with a non-coding genetic variant that increased cell division capacity. Although, this study did not link RA risk loci with the modulation of CD4+ memory cell proliferation, it exemplifies how to link genotype to immune cell function.

1.4.2 Correlating disease variants with gene expression

Having identified the most relevant cell types, the effects of genetic variants on gene expression are often assessed through genotype correlation with gene expression levels measured across tens to thousands of individuals (expression quantitative trait loci, eQTL; Figure 1.3). Disease associated variants are enriched for eQTLs (Dimas et al., 2009; Nicolae et al., 2010). A relevant and easily accessible tissue for immune diseases is blood. As such, early studies that integrated GWAS SNPs with gene expression identified an enrichment of immune disease SNPs in whole blood or peripheral blood mononuclear cells (PBMCs) expression variants. For example, over 50% of CEL variants also had an eQTL effect (Dubois et al., 2010). Disease associated variants affecting gene expression can point towards a dysregulated specific pathway. For instance, five of the IBD risk variants were also eQTLs which had an increasing effect on the expression levels of the *ITGA*, *ITGAL*, *ICAM* and *ITGB8* genes, encoding for integrins, the proinflammatory cell surface proteins (Lange et al., 2017).

As the gene expression studies increased in sample size (Westra et al., 2013), and expanded across tissue types (GTEx Consortium, 2015) and cell states (Fairfax et al., 2014), it became evident that the eQTL effects are widespread. For instance, a study of whole blood gene expression from over 8,000 individuals identified that nearly 6,500 genes (44% of all tested genes) were under a genetic control (Westra et al., 2013).

Recent studies that have incorporated of disease associated loci defined by GWAS with eQTLs have highlighted that the initially observed enrichment of GWAS SNPs among eQTLs may have been overestimated. This is because of the confounding effects of the LD and cell type specific gene expression. The LD results in long distance correlations between tens to hundreds of variants. The GWAS and eQTL signals can overlap in a genetic location, however it is critical to determine whether the overlap is coincidental

or driven by the same functional variants. For that, simple overlap between the eQTL and the GWAS SNPs is not sufficient and more stringent methods that colocalise the LD variants between the two signals need to be applied instead (Guo et al., 2015; Chun et al., 2017). Guo et al. (Guo et al., 2015) assessed the colocalisation of 595 variants from 154 non-overlapping regions associated to ten different immune-mediated diseases with gene expression variants from primary resting and stimulated monocytes and resting B cells. Of the 1,414 genes mapping to these regions, 125 showed an eQTL effect that also overlapped with a disease SNP. However, there was only strong support of colocalising signals for six genes.

The early eQTL mapping studies had already recognised the importance of cell type specific gene regulation. Dimas et al. used lymphoblastoid cell lines (LCLs), primary fibroblasts and T cells from umbilical cords and reported that 69-80% of regulatory variants affected gene expression in a cell type specific context (Dimas et al., 2009). Another recent study found that only a small proportion of IBD variants overlapped with whole blood eQTL SNPs (8 of 76 IBD loci) (Huang et al., 2017). The lack of enrichment could be due to the presence of heterogeneous populations of immune cells in the whole blood. In fact, higher enrichment was observed with eQTLs from CD4+, ilium and CD14+ monocytes (an essential cell of the innate immune response), underscoring the importance of the cell type specific context when analysing the function of GWAS variants (Figure 1.3).

Regulatory variants can exhibit opposite effects across different cell types (Solovieff et al., 2013). For example, of over 7,000 eQTLs that were shared between monocytes and T cells, Raj et al. identified 42 to have inverse effects between the two cell types (Raj et al., 2014). One of these eQTLs affected the expression of CD52, a target for antibody therapy used in MS treatment (CAMMS223 Trial Investigators et al., 2008). Furthermore, even closely related cell types, such as CD4+ and CD8+ T cells showed a limited overlap between eQTLs and GWAS SNPs (21 SNPs affecting the expression of 133 genes) based on a cohort of 313 healthy individuals (Kasela et al., 2017). Improving our understanding of disease associated variants in a cell type specific context can therefore inform future drug development strategies by ensuring that a drug is targeting a protein in a specific disease pathological cell type.

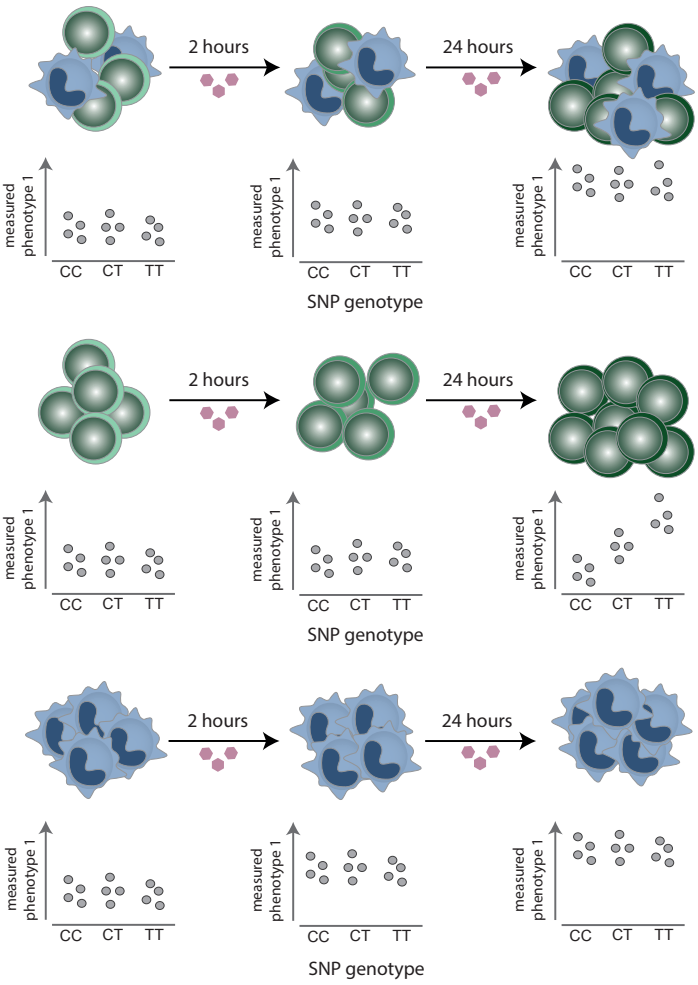

Figure 1.3: **Cell type specific and cell-state-specific expression quantitative trait loci.** Cellular phenotypes, such as gene expression, cytokine secretion or chromatin accessibility, might be affected by a genetic variant only in a specific cell type and under specific conditions, e.g. at a specific time-point following a stimulation. Here, a bulk of cell types (upper panel), as well as each cell type individually, were stimulated for 2 and 24 hr. In all scenarios, the measured phenotype, e.g. gene expression, increased upon stimulation, but the effect was only correlated with the genotypes in the green cell type (middle panel). The effect was missed when measured in the sample containing the mixed cell population (upper panel). The blue cell type (bottom panel) shows the strongest up-regulation in expression upon stimulation and largely drives the observed increase in expression in the bulk sample.

However, even if the relevant cell type is identified, the functional effect of a variant may not be detected unless the cell is challenged in an appropriate environment. Fairfax et al., (Fairfax et al., 2014) stimulated monocytes with lipopolysaccharide (LPS) and IFN-γ and demonstrated that 467 eQTLs overlapped with disease associated GWAS SNPs, 53% of which were stimulation specific. One of the eQTL-GWAS SNPs was an MS variant that

affected *IRF8* expression following two hours stimulation with LPS. Earlier studies that investigated the expression levels of *IRF8* in PBMCs using microarrays failed to identify a variant controlling the expression of this gene (De Jager et al., 2009), likely due to the cell type specific effect. The same locus is also associated with SLE but with an opposite effect. IRF8 is a regulatory factor of type-1 interferons, which are elevated in SLE patients while MS patients present low interferon levels (Chrabot et al., 2013). Importantly, the authors observed eQTL effects that differed between early and late stimulatory responses. The dynamic nature of gene expression regulation was also observed in dendritic cells stimulated with IFN–γ (Lee et al., 2014) and in CD4+ T cells stimulated with anti-CD3/anti-CD28 beads (Ye et al., 2014). Ye et al. identified 157 GWAS SNPs that overlapped with genetic variants that affected gene expression in CD4+ T cells in a cohort of 348 healthy individuals (Ye et al., 2014). Notably, an ulcerative colitis (UC) variant nearby *IL23R* and a variant nearby *IL2RA* associated to T1D, MS and vitiligo, only presented an effect on gene expression 48 hours following stimulation. Given that most effector functions of immune cells are performed following stimulation, it is not surprising that the majority of immune disease variants would only be functional in activated cells. Further studies are necessary to investigate different stimulation contexts in more detail.

1.4.3 Correlating disease variants with epigenetic marks

The correlation between histone marks and genotypes can also be directly analysed (Kumasaka et al., 2016). ChIP-seq assays result in pile-ups of sequence reads at genomic regions that map the interaction with proteins. In that respect, they produce a quantitative measurement and, just like gene expression, can be analysed in light of correlations with different alleles, as QTLs. Regulatory annotations are not only a good predictor of gene activity but can also suggest a mechanism of action, e.g. by implicating a specific TF binding site (McVicker et al., 2013), or when overlapping with eQTLs (Rosario et al., 2015). For example, H3K27ac QTLs in human lymphoblastoid cell lines are highly enriched for immune disease GWAS variants, particularly from MS. In a study of three major immune cell types; neutrophils, monocytes and CD4+ T cells, from 200 individuals, Chen et al. mapped coordinated genetic effects on the epigenome and transcriptome (Chen et al., 2016).

In addition to ChIP-seq, genome activity can also be measured through chromatin accessibility. Analysis of 349 tissues and cell types generated using DHS-seq found that over 75% of non-coding GWAS associated variants lie within a DHS. The localisation of SNPs in DHS was quantitative and cell specific, e.g. Th1 and Th17 CD4+ T cells were enriched for CD variants, indicating its potential future use for QTL mapping for mechanistic understanding of the disease. For example, around 25% of GWAS SNPs associated with autoimmune diseases within DHSs in immune cells (n=262) altered a TF binding motif for the IRF9 pathway (Maurano et al., 2012). The IRF9 network along with the Jak/Stat cascade are initiated in the presence of IFN-γ, indicating that this pathway might play an important role in autoimmunity.

Despite the observed enrichment of disease variants in regulatory non-coding sequences it is estimated that only 10-20% of 823 disease variants lay within TF binding motifs. This suggests that there are other mechanisms of gene expression regulation (Farh et al., 2015) such as mechanisms that affect spatial genomic interactions. To test this hypothesis, a promising approach is the integration of GWAS variants with Hi-C assays (Belton et al., 2012) that infer chromosome conformation by mapping interactions between genomic regions located nearby in the 3D space. The technique allows for both the high level mapping of chromatin loops as well as identification of precise interactions between enhancers and gene promoters. The latter can be directly used in mapping non-coding disease variants to target genes. For example, a recent study using Hi-C demonstrated that the non-coding region on chromosome 6 containing a variant that has been associated with RA and PSO, interacts not only with the promoter of *TNFAIP3*, the closest gene, but also with *IL2ORA*, which is 680kbp upstream from the variant (McGovern et al., 2016). A comprehensive examination of the promoter Hi-C interactions in 17 human primary blood cell types found an enrichment of autoimmune disease SNPs in lymphoid cells compared to myeloid cells (Javierre et al., 2016). The authors found that 76% of the identified genes had not been previously linked with immune-mediated diseases, since they were outside of the immune associated loci. Five of the newly identified genes, associated with RA and SLE, were under the control of an eQTL, e.g. *RASGRP1*, a gene that activates the Erk/MAP kinase cascade and regulates T and B cell development and differentiation.

Gene expression regulation is also reflected through DNA methylation. By combining

DNA methylation with RNA sequencing from 3,841 Dutch individuals Bonder and co-workers observed that trans methylation QTLs (meQTLs), where genetic variants affect distant rather than local methylation status, were enriched for immune associated GWAS traits (Bonder et al., 2017). While methylation has typically been linked to repressed regions, the authors found two distinct functions for methylation depending on its location. To investigate these findings further, they carried out TF ChIP-seq and identified 13 trans-meQTLs that influenced the TF binding site, including two SNPs on chromosome 4 associated with UC. They prioritised one of them as the causal SNP based on its association with lower methylation and the higher gene expression of *NFKB1*, the TF which controlled this binding motif. In addition, by incorporating the Hi-C assay with DNA methylation and RNA sequencing, Bonder et al., identified that interchromosomal contacts provide a mechanism by which some trans-meQTLs act. The 402 identified CpG islands overlapped with CTCF, RAD21 and SMC3 TF binding sites.

1.4.4 Correlating genetic variation with immunological readouts

To gain comprehensive insights into the effects of genetic variants on the immune system, genomic assays have to be complemented with traditional immunological methods. The network of cellular interactions can be assessed at both a global scale, for instance by looking at mechanisms that control cell frequency and cell proliferation, and at the cellular level, by measuring protein expression and identifying genes that drive the phenotype of interest. To accurately infer the causal relationships between genetic variants and cellular traits it is important to carry out the experiments in cells isolated from healthy individuals, limiting the effects of active disease or ongoing treatment.

The ratio between different immune cell subsets is heritable and could partly be a pathobiological disease mechanism (Hall et al., 2000; Brodin et al., 2015). Through an association study of genome wide SNPs with cell counts, two independent genetic signals were identified to control the CD4+ to CD8+ T cell ratio. Both signals mapped to MHC region, explaining 8% of the observed variance. Interestingly, one of these associations overlapped with a T1D protective variant where the T1D risk allele increases the number of CD4+ cells by decreasing their apoptosis rate (Ferreira et al., 2010). Past years have seen many systematic, large scale efforts aiming to associate genetic variants

with cell counts. Astle et al. identified over 2,500 variants associated with 36 different hematopoietic traits (Astle et al., 2016). They observed an overlap between asthma associated variants and the eosinophil counts, highlighting that the established positive association between eosinophil counts and asthma is genetically controlled. Additionally, variants associated within the MHC locus and nearby *COG6*, *SPRED2*, *RUNX1* and *ATXN2*/*SH2B3*/*BRAP* genes pointed towards a novel link between eosinophil function and RA. A study of immune cell frequencies and surface protein expression levels of 170 dizygotic and 75 monozygotic pairs of twins using seven distinct 14-plex antibody panels identified 151 independent heritable immune traits (Roederer et al., 2015). This study reported that one of the most heritable traits is the frequency of CD39+ Tregs. They identified a SNP that increases the level of CD39, and thus alters the proportion of CD39+ Tregs. CD39 along with CD73 are enzymes that degrade the proinflammatory ATP molecule to an anti-inflammatory adenosine (Antonioli et al., 2013). Dysregulation of this machinery has been observed in MS, RA and IBD patients, with multiple drugs targeting these two proteins. Additionally, the same variant had previously been identified in a study that measured counts of 95 different cell types from a cohort of 1,629 individuals from Sardinia (Orrù et al., 2013).

Immune cells exert their effector function by communicating with each other through the expression of receptors and receptor ligands, combined with the secretion of specific molecules, such as cytokines. Alteration of these cellular functions results in impaired immune response. The cytokine levels in the blood have been shown to be highly heritable (Brodin et al., 2015). Systematic characterisation of the cell response to different bacterial and fungal infections identified six cytokine QTLs (cQTLs) that explained the IL-6, IL-8, IL-10 and TNF-α levels (Li et al., 2016b). The genetic control of cytokine secretion in response to pathogens is relevant, as some of the identified cytokines have also been associated with autoimmune diseases and are being targeted by drugs, e.g. IL-6 with RA and PSO (Ishihara and Hirano, 2002). IL-6 pathways, along with IL-1β were identified as being mostly driven by genetics, compared to environmental factors and the microbiome. A larger study by the same group assessed five cytokine production responses of PBMCs, whole blood and macrophages from 500 individuals to a pathogenic stimulus compared to no stimulus. They identified cell type specific cQTLs, with monocyte specific cQTLs being associated with susceptibility to infectious diseases and T cell specific cQTLs being associated with autoimmune

diseases (Li et al., 2016a). In addition, colocalisation of genetic variants that control cytokine levels with those that contribute to susceptibility to autoimmune diseases linked the VEGF cascade with IBD and expression of the α subunit of the IL-2 receptor with CD and MS (Ahola-Olli et al., 2017). Together these data demonstrated that the levels of a number of pro- and anti- inflammatory cytokines are under genetic control. In the future, a large scale characterisation of molecular mechanisms that control cytokine levels in health and disease will provide further insights into disease pathology.

1.5 The product: Outline of the thesis

Over the last decade, hundreds of immune disease loci have been successfully mapped. Recent advances in genomics have enabled functional follow up studies that are starting to provide insights into the molecular mechanisms by which disease variants drive disease pathology. The aim of my thesis is to better understand the influence of genetic variation associated with common immune diseases on the function of human immune cells and identify pathways that are potentially perturbed. I focus on CD4+ T cells because GWAS variants are enriched in the active regulatory elements of these cells and they have an established role in autoimmune disorders.

Given the importance of studying the role of immune variants in the most relevant cellular context, I first investigate in Chapter 2 if the two major subsets of CD4+ T cells, naive and memory, operate the same gene expression programmes upon stimulation. I disentangle the differential requirement of CD28 in naive and memory CD4+ T cells using functional genomic assays. I show evidence that T helper differentiation cytokines and chemokines expression increases in response to CD28 signal intensity in both naive and memory cells. I observe that cell cycle and division are sensitive to CD28 in memory cells in contrast to the paradigm that memory cells are CD28-independent. Lastly, I show that CD28-sensitive genes, that is genes whose expression increases alongside the increase in the intensity of the CD28 stimulus, are enriched in autoimmune disease loci, pointing towards the role of memory cells and the regulation of T cell activation through CD28 in autoimmune disease development.

In Chapter 3 I optimise a newly published genomic protocol to investigate the regulatory landscape of resting and stimulated regulatory T cells, a critical cell type which has not been thoroughly studied due to its low abundance. I use ChM-seq assays to assess the promoter and enhancer landscape of resting and activated Tregs and link identified elements to putative changes in gene expression. I find that the majority of gene expression regulation changes are controlled by changing the levels of H3K27ac and not by altering the levels of H3K4me3. Furthermore, I identify differences in transcript ratio, which are indicative of differential transcript usage upon stimulation, including for key CD4+ T cell TFs, such as NFATC1 and YY1. Since I observe that many peaks remain unannotated from the target genes, I delve deeper into understanding other gene expression regulatory mechanisms and perform correlation between changes in the histone landscape and splicing patterns in this cell type.

Finally, in Chapter 4 I investigate the role of immune-mediated disease variants in Treg function. Using freshly isolated Tregs from 100 donors, I identify over 5,000 eQTLs and 2,000 transcript ratio (tr) QTLs, of which around 40% were not observed in naïve CD4+ T cells. I observe 117 genes colocalising with immune-mediated disease signals, many of which were not observed in naive T cells. Finally, I integrate this data with chromatin accessibility QTLs (caQTLs), promoter regions QTLs (promQTLs) and active regions QTLs (actQTL). This allows me to fine-map the association signal at 33 loci. For two of the immune associated genes I identify as being Treg specific, in relation to naive CD4+ T cells, but also the GTEx consortium, *MAP3K8* and *TNFRSF9*, I suggest one causal variant.

In summary, my thesis aims at addressing the role of immune disease variants in the context of three important components of the immune response: 1) differential requirement of stimulus for the activation of naive and memory T cells, 2) effect of activation on gene expression regulation in regulatory T cells, and 3) genetic variation on gene expression regulation in regulatory T cells.

2 The role of co-stimulation in naive and memory T cells

Collaboration note

The work described in this chapter is awaiting publication and is currently on biorxiv as "CD28 control of memory T cell proliferation, effector function and autoimmune susceptibility" (Glinos and Soskic et al. pending). Many sections of the manuscript have been directly copied into this chapter. I processed the samples to cell isolation, generated the RNA-seq, H3K27ac ChM-seq and ATAC-seq data, performed the flow cytometry and analysed all the data. Blagoje Soskic performed the cell stimulation, the proliferation assays and analysed the flow cytometry data, contributed to the ATAC-seq and H3K27ac ChM-seq data generation and the flow cytometry acquisition. Blagoje Soskic and Dave Sansom were involved in the experimental design and interpretation of the results. RNA-seq library construction and sequencing of all materials was done by DNA Pipelines core facility at Sanger.

2.1 Introduction

The ability of T cells to respond to pathogens whilst remaining tolerant to host antigens is critical for human health. T cell stimulation generally occurs in secondary lymphoid tissues where T cells interact with professional antigen presenting cells (APCs). Here, two coordinated signals are delivered: the first via T cell receptor (TCR) recognising antigen bound to MHC molecules and the second provided by APCs via upregulation of co-stimulatory ligands. In this regard, CD28 is the main co-stimulatory receptor expressed by T cells which interacts with CD80 and CD86 ligands on APCs. The coordination of TCR and CD28 signals is essential for T cell activation, proliferation, differentiation and survival. Therefore, the CD28 pathway provides a key checkpoint for controlling T cell responses (Rowshanravan et al., 2017), which is increasingly relevant therapeutically.

Early studies demonstrated that memory T cells have higher affinity for their antigen and consequently need lower concentrations of it in order to respond (Savage et al., 1999; Richer et al., 2013). Given that memory T cells have a lower activation threshold than naive cells it has been suggested that they can respond in the absence of CD28 triggering (Luqman and Bottomly, 1992; London et al., 2000). The concept that memory cells are CD28 independent has been challenged recently using *in vivo* mouse models and different methods to knock-out or block CD28 (Ndejembi et al., 2006, Borowski et al., 2007; Garidou et al., 2009; Teijaro et al., 2009; Ndlovu et al., 2014; Fröhlich et al., 2016). Experiments using CD28 deficient mice and mice with a Cre/lox induced deletion of CD28 prior to a secondary challenge demonstrated that CD28 is crucial for efficient defense against worms in both primary and memory responses (Ndlovu et al., 2014). Furthermore, Teijaro et al. demonstrated that treatment of mice with a co-stimulation blocker reduced both the expansion and accumulation of memory T cells in spleen and lungs following influenza infection (Teijaro et al., 2009). Finally, while it may be possible to trigger memory T cell activation without CD28 engagement, the long term survival of the cells and their abilities to assume effector functions are likely to be more demanding on CD28 involvement (Jenkins et al., 1991).

The level of CD28 co-stimulation is likely to vary considerably in different immunological settings. For example, the presence of regulatory T cells (Tregs) expressing CTLA-4, which degrades CD80 and CD86 ligands (Qureshi et al., 2011) will influence CD28 co-stimulation. Indeed, deficiency in CTLA-4 expression is associated with the development of profound autoimmune diseases (Tivol et al., 1995; Kuehn et al., 2014; Schubert et al., 2014; Lo et al., 2015) due to increased CD28 signalling (Tivol et al., 1997; Tai et al., 2007). In a pharmacological setting, severe adverse reactions were observed in reaction to a CD28 agonistic antibody, TGN1412, due to stimulation of a cytokine storm from effector memory T cells (Eastwood et al., 2010; Hünig, 2012). Furthermore, excessive activation of memory T cells is a hallmark of many common complex immune diseases, such as autoimmune arthritis and systemic lupus erythematosus (SLE) (Haufe et al., 2011; Kshirsagar et al., 2013). Genome-wide association studies (GWAS) have mapped numerous risk variants to loci encoding genes in T cell stimulatory pathways, including *CD28*, *CTLA4* and *ICOS* located at 2q33.2 (Fortune et al., 2015; Dubois et al., 2010; Okada et al., 2014; Onengut-Gumuscu et al., 2015). While the exact effects of the associated variants are unknown, their mapping to the non-coding regions of the genome suggests effects on regulation

of gene expression (Fairfax et al., 2014; Westra et al., 2013). Therefore, understanding how varying levels of TCR and CD28 co-stimulation impacts gene expression and the ensuing T cell response, has significant implications in understanding susceptibility to diseases, in particular autoimmunity and cancer.

The fact that T cell activation can be measured in different systems and using different assays, highlights the need for a new unbiased approach that would encompass more than one aspect of activation. Here, I designed an approach to address the requirement of TCR and CD28 in the activation of naive and memory human CD4+ T cells by profiling the transcriptome and epigenome of these cells in response to varying intensities of TCR and CD28 stimulation. I show that the major effector functions, such as T helper (Th) differentiation, expression of chemokine receptors and cytokines, are all strongly influenced by CD28 in both naive and memory T cells. Strikingly, cell division is markedly differently controlled between the two cells, which we find to be controlled by CD28 in memory cells whilst predominantly driven by TCR in naive cells. I identify genes that were sensitive to TCR or CD28 levels and map the promoter and enhancer landscape associated with gene expression regulation. Finally, I show that a proportion of loci associated to common immune diseases is enriched in CD28-sensitive genes, pointing towards the important role of this co-stimulatory pathway in disease pathogenesis.

2.2 Materials and Methods

2.2.1 Sample collection and DNA isolation

Blood samples were obtained from eight healthy adults, aged from 22 to 46 years. Peripheral blood mononuclear cells (PBMCs) were isolated using Ficoll-Paque PLUS (GE healthcare, Buckingham, UK) density gradient centrifugation. CD4+ T cells were isolated from PBMCs using EasySep® CD4+ enrichment kit (StemCell Technologies, Meylan, France) according to the manufacturer's instructions. DNA was isolated from live PBMCs using Qiagen DNeasy blood and tissue kit. All samples were obtained in accordance with commercial vendor's approved institutional review board protocols and their research use was approved by the Research Ethics Committee (reference number: 15/NW/0282).

2.2.2 Flow cytometry and cell sorting

CD4+ enriched cells were stained with the following antibodies for sorting: CD4+ (OKT4)-APC (Biolegend); CD25 (M-A251)-PE (Biolegend); CD127 (eBioRDR5)-FITC (eBioscience) and Live/Dead fixable blue dead cell stain. Live conventional T cells (Tcons, CD4+ CD25low CD127high) were isolated and cultured for 16 hours. Stimulated naive and memory cells were sorted based on the expression of CD25-PE, CD45RA- Alexa700 (Biolegend) and DAPI. Resting naive and memory cells I sorted for low expression of CD25 (proportion of CD25+ cells < 1%).

2.2.3 Cell culture and stimulation

Chinese hamster ovary (CHO) cells expressing CD86 or FcR (FcRγII, CD32) were cultured in DMEM (Life Technologies, Paisley, UK) supplemented with 10% v/v FBS (Sigma, Gillingham, UK), 50 U/ml penicillin and streptomycin (Life Technologies), and 200 µM l-glutamine (Life Technologies) and incubated at 37°C in a humidified atmosphere of 5% CO2. CHO cells expressing CD86 and FcR were generated as previously described (Qureshi et al., 2011). T cells were co-cultured with glutaraldehyde fixed CHO-CD86 to provide CD28 signal or CHO-FcR for 16 hours in RPMI 1640 supplemented with 10% v/v FBS, 50 U/ml penicillin and streptomycin, and 200 µM l-glutamine. Cultures stimulated with CHO-CD86 were treated with various concentrations of anti-CD3 (OKT3) while cultures stimulated with CHO-FcR were treated with 1 µg/ml of anti-CD3 (OKT3) or 1 µg/ml of anti-CD28 (9.3). For the full information of the samples and simulations used refer to Table 2.1 .

Table 2.1: **Metadata of blood donors processed and sample specifications. Ind. ID**: Identification number of blood donor; **Coll. time**: Time of the day the blood was collected; **Process Date**: Month and year the blood was processed; **Age**: Age of individuals at the time of the blood draw; **Sex**: Sex of the blood donors, F(emale) or M(ale); **Stimulation**: Stimulation used on this sample, l refers to low and h refers to high. When the two stimuli were combined lCD28 corresponds to 1 CHO cell for 25 T cells, hCD28 corresponds to 1 CHO cell to 2.5 T cells, lTCR corresponds to 0.01ng/µl and hTCR corresponds to 1ng/µl. When the stimuli are used alone hCD28 corresponds to 1ng/µl and hTCR corresponds to 1ng/µl; **Cell type**: Memory cells (M) are defined as CD4+CD25-CD127+CD45RA- and Naive (N) cells are defined as CD4+CD25-CD127+CD45RA+. Upon stimulation assayed cell are CD25+; **Cell num. (x1000)**: Number of cells used for RNA extraction; **Activated cells (%)**: Percentage of naive or memory T cells subpopulation that successfully underwent activation.

Ind. ID	Coll. time	Process date	Age	Sex	Stimulation	Cell type	Cell num. (x1000)	Activated cells (%)
25	1045	Feb-16	26	F	hCD28	M	301	65.98
25	1045	Feb-16	26	F	hTCR	M	309	55.98
25	1045	Feb-16	26	F	hTCR hCD28	M	300	55.41
25	1045	Feb-16	26	F	hTCR lCD28	M	250	25.12
25	1045	Feb-16	26	F	lTCR hCD28	M	250	25.78
25	1045	Feb-16	26	F	lTCR lCD28	M	150	12.46
25	1045	Feb-16	26	F	resting	M	248	0
25	1045	Feb-16	26	F	hCD28	N	750	60.05
25	1045	Feb-16	26	F	hTCR	N	1150	96.05
25	1045	Feb-16	26	F	hTCR hCD28	N	1550	88.65
25	1045	Feb-16	26	F	hTCR lCD28	N	886	35.78
25	1045	Feb-16	26	F	lTCR hCD28	N	1456	65.66
25	1045	Feb-16	26	F	lTCR lCD28	N	653	25.61
25	1045	Feb-16	26	F	resting	N	718	0
26	1205	Feb-16	22	M	hCD28	M	138	53.9
26	1205	Feb-16	22	M	hTCR	M	119	58.48
26	1205	Feb-16	22	M	resting	M	152	0
26	1205	Feb-16	22	M	hCD28	N	351	36.25
26	1205	Feb-16	22	M	hTCR	N	545	77.43
26	1205	Feb-16	22	M	resting	N	577	0
27	1300	Feb-16	31	M	hCD28	M	491	73.54

Continued on next page

Table 2.1 – *Continued from previous page*

Ind. ID	Coll. time	Process date	Age	Sex	Stimulation	Cell type	Cell num. (x1000)	Activated cells (%)
27	1300	Feb-16	31	M	hTCR	M	382	64.78
27	1300	Feb-16	31	M	hTCR hCD28	M	756	58.51
27	1300	Feb-16	31	M	hTCR lCD28	M	456	30.81
27	1300	Feb-16	31	M	lTCR hCD28	M	483	33.82
27	1300	Feb-16	31	M	lTCR lCD28	M	287	19.02
27	1300	Feb-16	31	M	resting	M	336	0
27	1300	Feb-16	31	M	hCD28	N	374	59.81
27	1300	Feb-16	31	M	hTCR	N	386	77.7
27	1300	Feb-16	31	M	hTCR hCD28	N	1200	90.73
27	1300	Feb-16	31	M	hTCR lCD28	N	679	49.75
27	1300	Feb-16	31	M	lTCR hCD28	N	969	66.24
27	1300	Feb-16	31	M	lTCR lCD28	N	434	29.9
27	1300	Feb-16	31	M	resting	N	354	0
29	1210	Feb-16	36	F	hCD28	M	248	59.67
29	1210	Feb-16	36	F	hTCR	M	104	31.6
29	1210	Feb-16	36	F	hCD28	N	75	47.03
29	1210	Feb-16	36	F	hTCR	N	63	53.36
30	1130	Feb-16	39	M	hCD28	M	212	NA
30	1130	Feb-16	39	M	hTCR	M	182	NA
31	1045	Feb-16	25	F	hCD28	M	355	NA
31	1045	Feb-16	25	F	hTCR	M	463	NA
31	1045	Feb-16	25	F	resting	M	500	NA
31	1045	Feb-16	25	F	hCD28	N	151	NA
31	1045	Feb-16	25	F	hTCR	N	366	NA
31	1045	Feb-16	25	F	resting	N	460	NA
42	1400	Apr-16	26	F	hCD28	M	210	37.59
42	1400	Apr-16	26	F	hTCR	M	300	47.09
42	1400	Apr-16	26	F	hTCR hCD28	M	315	19.57
42	1400	Apr-16	26	F	hTCR lCD28	M	220	11.97

Continued on next page

Table 2.1 – *Continued from previous page*

Ind. ID	Coll. time	Process date	Age	Sex	Stimulation	Cell type	Cell num. (x1000)	Activated cells (%)
42	1400	Apr-16	26	F	lTCR hCD28	M	187	13.03
42	1400	Apr-16	26	F	lTCR lCD28	M	146	7.2
42	1400	Apr-16	26	F	resting	M	250	0
42	1400	Apr-16	26	F	hCD28	N	124	15.76
42	1400	Apr-16	26	F	hTCR	N	757	100
42	1400	Apr-16	26	F	hTCR hCD28	N	2000	87.41
42	1400	Apr-16	26	F	hTCR lCD28	N	1300	52.24
42	1400	Apr-16	26	F	lTCR hCD28	N	1360	58.8
42	1400	Apr-16	26	F	lTCR lCD28	N	790	31.06
42	1400	Apr-16	26	F	resting	N	340	0
43	1235	Apr-16	46	M	hCD28	M	547	51.71
43	1235	Apr-16	46	M	hTCR	M	845	82.64
43	1235	Apr-16	46	M	hTCR hCD28	M	530	28.84
43	1235	Apr-16	46	M	hTCR lCD28	M	254	13.05
43	1235	Apr-16	46	M	lTCR hCD28	M	236	13.97
43	1235	Apr-16	46	M	lTCR lCD28	M	237	9.38
43	1235	Apr-16	46	M	resting	M	658	0
43	1235	Apr-16	46	M	hCD28	N	225	38.65
43	1235	Apr-16	46	M	hTCR	N	487	87.71
43	1235	Apr-16	46	M	hTCR hCD28	N	650	64.21
43	1235	Apr-16	46	M	hTCR lCD28	N	285	26.09
43	1235	Apr-16	46	M	lTCR hCD28	N	351	31.44
43	1235	Apr-16	46	M	lTCR lCD28	N	180	17.49
43	1235	Apr-16	46	M	resting	N	401	0

2.2.4 T cell proliferation assay

Prior to stimulation with CHO-FcR cells and anti-CD3 or anti-CD28, naive and memory T cells were labeled with CellTrace Violet dye (Life Technologies) according to the manufacturer's instructions. Five days following stimulation, T cell proliferation was

analyzed by flow cytometry and proliferation was modeled using the Flowjo proliferation platform. Total T cell numbers per sample were established relative to AccuCheck counting beads (Invitrogen).

2.2.5 FACS markers validation

PBMCs were isolated from whole blood, as described above. Naive and memory CD4+ T cells were isolated by negative selection using human CD4+ T Cell Enrichment Kits (EasySep™, STEMCELL Technologies). After six days of stimulation, as described above, cells were observed by flow cytometry using two directly conjugated antibody panels listed in Table 2.2 . Intracellular staining was done using the eBioscience FOXP3 staining buffers.

Table 2.2: Panel of antibodies used in the validation of genes that were TCR or CD28 sensitive.

Panel 1		**Panel 2**	
Epitope	**Fluorophore**	**Epitope**	**Fluorophore**
CD25 (2A3)	BV605 (BD)	CD25 (2A3)	BV605 (BD)
CD45RA (HI100)	PerCP-Cy5.5 (eBioscienc)	CD45RA (HI100)	PerCP-Cy5.5 (eBioscienc)
CD69 (FN50)	PE-Cy7 (BD)	ICOS (DX29)	BV711 (BD)
CD71 (M-A712)	AF700 (BD)	CD71 (M-A712)	AF700 (BD)
CD40L (24-31)	e450 (eBioscienc)	OX40 (ACT35)	PeCy7 (BD)
CTLA4 (BNI3)	PE (BD)	CD80 (L307.4)	PE (BD)
CD28 (CD28.2)	APC (eBioscienc)	PDL1 (MIH1)	FITC (BD)
		CD27 (L128)	BUV395 (BD)

2.2.6 RNA-seq

Naive and memory T cells were placed in 0.5 ml of Trizol (Invitrogen) and stored at -80°C. Samples were thawed at 37°C before adding 100 µl chloroform. After reaching equilibrium, samples were centrifuged for 15 minutes at 4°C at 10,000g. The collected aqueous phase was mixed 1:1 with 70% ethanol before proceeding with minElute columns (Qiagen) for purification, according to the manufacturer's protocol. RNA was quantified using Bioanalyzer (Agilent Technologies, USA). The RNA was sequenced in two separate batches. Libraries were prepared using Illumina TruSeq index tags and sequenced on the Illumina HiSeq 2500 platform using V4 chemistry and standard 75 bp paired-end. The first batch consisted of 56 samples that were multiplexed at equimolar concentrations and sequenced across 14 lanes, to yield on average 71.3 million reads per sample. The

second batch consisted of 18 samples that were multiplexed at equimolar concentrations and sequenced across 3 lanes to yield on average 61 million reads per sample.

2.2.7 RNA-seq data processing

Sequence reads were aligned to the GRCh38 human reference genome using STAR (v2.5.0c) (Dobin et al., 2013) and the Ensembl reference transcriptome (v83). Gene counts were estimated using featureCounts (v1.5.1) tools (Liao et al., 2014) from the subread package and only reads assigned to the transcripts were used for further processing (84-90% of reads were assigned).

To find genes that were upregulated upon stimulation, I first defined differentially expressed genes using DESeq2 (v1.14.1) (Love et al., 2014) by performing pairwise comparison of all conditions to the resting state, in a cell type specific manner, using Benjamini-Hochberg controlled false detection range (FDR) $\leq$ 5% (Benjamini and Hochberg, 1995) and an absolute fold-change $\geq$ 2. I then build a linear and a switch model of gene expression using a likelihood ratio test (LRT) separately for naive and memory cells. In the linear model, I assumed a linear increase of gene expression along with stimulus intensity (incremental fold-change $\geq$ 1.5). Genes that did not follow the linear model were tested for the switch model. Here, I assumed an "on-and-off" mode of expression, where a gene is significantly upregulated (fold-change $\geq$ 2) in response to the presence of either CD28 or TCR. In both of these models, I used all seven conditions, e.g. when testing for CD28-sensitive genes I grouped the TCR alone stimulation with the resting, since neither received CD28 signal. A gene was classified in one of the two categories without overlap and prioritised for the linear model.

To control for the different batches in which I processed the blood, which accounted for 12% of the observed variability, I performed batch correction prior to PCA using the Combat algorithm from the sva package (Leek et al., 2012). To estimate the percentage of the variance explained separately by each of the recorded variables, such as the stimulus, the cell type, the gender of the donors, I fitted a linear model with only the stimulus or the cell type as variables (method adapted from McCarthy et al., 2017).

I performed pathway enrichment analysis by testing whether different gene-sets were over-represented in particular hallmark pathways (Liberzon et al., 2015). I used the

Jaccard index to quantify the proportion of stimuli-specific genes present in a tested pathway and assessed the statistical significance of the over-representation using a permutation strategy. For that, within each cell type and condition, I randomly sampled 10,000 gene-sets of the same size as in the observed dataset, matching for the gene length and expression levels.

2.2.8 ChIPmentation-seq (ChM-seq)

The ChM protocol was performed according to the protocol presented in Schmidl et al., 2015 and adapted to work with the iDeal ChIP-seq Kit for Histones. After sorting the cells were resuspended in pre-warmed full medium (IMDM, 10% FCS) at 1-2 million cells per ml and allowed to recover in the incubator (37ºC, 5% CO2) for at least 30 minutes The cells were then fixed by addition of formaldehyde to the medium to reach a final concentration of 1% and were incubated for 5 minutes at 37ºC, followed by quenching at room temperature with glycine for 5 minutes at a final concentration of 125 mM. The cross-linked cells were subsequently washed twice with ice-cold PBS and snap-frozen by immersion in liquid nitrogen.

Five hundred thousand crosslinked cells were washed using 250 µl IL1 buffer and resuspended in 250 µl IL2 lysis buffer, both of which contained 1X protease inhibitor cocktail (PIC). Cells were left to lyse for 5 minutes at 4° C on a rotator and then centrifuged at 4° C (3000g) for 5 minutes. The pellets were resuspended in 250 µl IS1 buffer and sonicated using the Bioruptor®Pico (Diagenode) for 5 minutes for resting cells or 4 minutes for stimulated cells (Diagenode). We kept a portion of the chromatin from two samples aside, one naive and one memory stimulated with hTCR and hCD28, and used them as a ChM-seq input.

The chromatin was immunoprecipitated using protein-A coated IP beads. Twenty microliters of beads were washed four times using 40 µl IC1 buffer on the magnetic rack before being resuspended in 20 µl of IC1. The beads were mixed with 56 µl 5X IC1 buffer, 6 µl 50X BSA, 1.5 µl 200X PIC and 1 µg of H3K27ac (Cat. no. C15410196, Diagenode). We added to the mix 100 µl of chromatin (equivalent to 200,000 cells) and incubated the samples overnight at 4ºC at 10 rpm.

The beads were then washed on the magnet with 350 µl of iW1, iW2 and iW3 buffers and

a final wash with 2X 1000 µl 10 mM Tris pH 8. The beads with the chromatin, as well as the two input samples, were then resuspended in 29 µl ChM buffer (Tris pH 8 1M, MgCl2 1M, ChIP grade water) with 1 µl Tn5 and incubated for 10 minutes at 37ºC at 1500 rpm. The tagmentation was stopped with the addition of 2X 350 µl iW3. Finally, the beads were washed with 350 µl iW4. The chromatin from the beads was eluted using 96µl iE1 and incubated for 30 minutes at room temperature at 1500 rpm. The chromatin was reverse cross-linked overnight using 4 µl of iE2 buffer. The DNA was then eluted in 30 µl of water using the MinElute PCR CLeanup kit (QIAGEN). The DNA was the purified twice using SPRI beads at 1.6x ratio using a Zephyr G3 SPE Workstation. The libraries were amplified following the ATAC-seq library amplification protocol (Buenrostro et al., 2013), but using NEBNext® High-Fidelity 2X PCR Master Mix (New England Biolabs). Eighteen libraries were indexed and pooled in equimolar concentration and sequenced on three lanes using the Illumina HiSeq 2500 platform and V4 chemistry using standard 75 bp paired-end reads to yield on average 80 million reads per sample.

2.2.9 ATAC-seq

ATAC-seq was performed according to published protocol (Buenrostro et al., 2013), with the following modifications. Fifty thousand cells were washed with 1 ml of ice-cold PBS. The cells were then resuspended in the tagmentation buffer containing Tn5 transposase (Illumina Nextera) and 0.01% digitonin and incubated for 30 minutes at 37ºC before purifying the DNA using the MinElute PCR purification kit (QIAGEN). Sequencing libraries were prepared using Nextera primers as described in the ATAC-seq protocol (Buenrostro et al., 2013). Sixteen libraries were indexed and pooled in equimolar concentration and sequenced on three lanes using the Illumina HiSeq 2500 platform and V4 chemistry using standard 75 bp paired-end reads to yield on average 65 million reads per sample.

2.2.10 ChM and ATAC data processing

The quality of the sequence reads was assessed using the fastx toolkit and the adaptors were trimmed using skewer (v0.2.2) (Jiang et al., 2014). Reads were mapped to the human genome reference GRCh38 using the bwa mem algorithm (v0.7.9a) (Li and Durbin, 2009). I only kept uniquely mapped reads, removed PCR duplicated reads and for the ATAC I excluded mitochondrial reads using samtools (v0.1.9) (Li et al., 2009). I retained 83.3% of ATAC and 73.8% of ChM reads. Genome browser tracks were created using BEDTools

(v2.22.0) (Quinlan and Hall, 2010) and the UCSC binary utilities. Furthermore, I generated insert size distributions using PICARD tools (v2.6.0) CollectInsertSizeMetrics function which can be indicative of over-sonicated chromatin and excess of adapters in the data. The mapped reads were converted into bed files and chimeras were removed. Peaks were called using MACS2 (v2.1.1) (Zhang et al., 2008) setting the parameters to -q 0.05 –nomodel –extsize 200 –shift -100 for ATAC, and –broad –broad-cutoff 0.1 –nomodel –extsize 146 for H3K27ac ChM. For ChM, all samples were downsampled to to the same read number (21.6 million reads) prior to peak calling against the input.

I used the fraction of reads in peaks (FRiP), the proportion of peaks with signal value (fold-change compared to the background or the input) greater than 10, the insert size distribution and the genome tracks to investigate the quality of the data. The median FRiP score was 59.2% for ATAC and 73.5% for H3K27ac ChM. The proportion of peaks with fold-change > 10 was 22.9% for ATAC and 4.8% for H3K27ac ChM. I concluded that the quality of the data was high. I merged all the ATAC samples and called peaks again using the parameters described above. For the H3K27ac ChM samples, I first merged the donors within each cell type and condition and then randomly sampled 17 million reads from each into one sample to reach the same read number as in the input. Since I merged already QCed samples I used the –keep-dup flag when calling peaks with MACS2, as the PCR duplicated reads for individual samples were already removed and I expected to observe a small proportion of the same reads present in independent samples by chance. I also increased the -q value threshold to 0.1 for both assays. The resulting peak files were used as a reference to count the number of reads falling into peak regions using featureCounts (Liao et al., 2014), therefore generating a quantitative table of read counts specifically present across the different conditions and cell types.

To ensure the analysis focused on high confidence peaks I removed the bottom 10th percentile of peaks with the lowest read counts in each dataset and obtained a final count of 142,306 and 49,638 peaks for ATAC and ChM, accordingly. To define differentially accessible regions (DARs) and differentially modified histone regions (DMHRs) the dataset was processed using DESeq2. To find regions that were upregulated upon stimulation, I compared all conditions to the resting state and used Benjamini-Hochberg controlled FDR of 5% and an absolute fold-change ≥ 2.

2.2.11 Binding expression target analysis

To identify regions of the genome that changed upon stimulation, which could subsequently regulate gene expression, I used Binding Expression Target Analysis (BETA) in the plus mode (Wang et al., 2013b). I performed the analysis using both DMHRs and DARs, as well as the list of regions that required one or the other stimulus. For example, TCR specific DMHRs or DARs were defined as the regions that are present in TCR stimulation but not in CD28 stimulation. I used the differential gene expression output from the pairwise comparison between stimulated and resting states. The median distance of interaction between an enhancer and gene promoter is estimated at 150 kbp (Mumbach et al., 2017). I therefore used the same window size around the transcription start site (TSS) of the differentially expressed genes to define boundaries for the analysis, along with the transcription activation domains (TADs) identified in CD4+ T cells (Javierre et al., 2016). That is, if the extended 150 kbp region fell beyond the TAD boundary I considered the TAD coordinates as the boundary for testing predictive effects of DMHRs and DARs on gene expression. I relied on the ATAC-seq output for the transcription factor enrichment analysis (p-value $\leq$ 0.01) as it generates narrow peaks allowing for a more accurate estimation of the transcription factor binding sites (TFBS). I used the JASPAR database of transcription factors and the Cistrome method to calculate transcription factor enrichment (p-value $\leq$ 0.05).

2.2.12 Disease SNP enrichment for stimulus-sensitive genes

I collected the GWAS data for Crohn's disease (CD), ulcerative colitis (UC; Jostins et al., 2012), coeliac disease (CEL; Trynka et al., 2011), multiple sclerosis (MS; International Multiple Sclerosis Genetics Consortium (IMSGC) et al., 2013), rheumatoid arthritis (RA; Okada et al., 2014), psoriasis (PSO; Tsoi et al., 2012), systemic sclerosis (SSc; Bossini-Castillo et al., 2015) and type-1 diabetes (T1D; Onengut-Gumuscu et al., 2015). Additionally, I used Bone mineral density (BMD) as a control trait by searching in the GWAS catalog for "Bone Mineral Density". I excluded all variants that fell within the MHC locus and using a genome wide p-value threshold of < $5x10^{-8}$. I defined the disease loci by mapping all the single nucleotide polymorphisms (SNPs) in linkage disequilibrium (LD) with the reported index SNP, using R^2 > 0.8 calculated across the European populations present in the 1000 Genomes Project data, and extending the LD boundaries by 150 kbp on

each side, to account for the possibility of distant gene expression regulation between enhancers and gene promoters. This resulted in 234 unique regions associated to one of the 8 tested traits.

I tested whether the stimulus sensitive genes fell within the SNP loci boundaries more often than expected by chance using a permutation strategy. To build the null distribution I selected the same number of genes, matching for gene size and mean expression level. I repeated the process 10,000 times.

Finally, I examined whether any of the SNPs used to define the LD boundaries overlapped with an H3K27ac or an ATAC peak identified in the specific stimulatory condition; CD28 alone stimulation for CD28 sensitive genes and TCR alone stimulation for TCR sensitive genes. The disruption of TFBS by SNPs was assessed using the SNP2TFBS database (Kumar et al., 2017).

2.3 Results

2.3.1 Experimental approach

To study the influence of varying TCR and CD28 signals I sorted human blood CD4+ CD25- T cells and stimulated them with high (referred to as h) or low (referred to as l) concentrations of soluble anti-CD3 (referred to as TCR) and cells expressing the CD28 ligand, CD86 (referred to as CD28) (Figure 2.1 A). As previously described (Manzotti et al., 2006), Chinese hamster ovary (CHO)-CD86 cells provide a source of natural CD28 ligand in an otherwise irrelevant cell background. In this way CD28 engagement was mediated via the the CD86 ligand, which has been documented to be the main ligand driving T cell responses (Borriello et al., 1997). In order to deconvolute gene expression programmes controlled individually by TCR and CD28, I also stimulated T cells with either anti-CD3 (an antibody against the signalling component of the TCR complex) or anti-CD28 antibodies alone, in the presence of CHO cells expressing Fc-gamma Receptor II (CHO-FcR) to provide crosslinking. Since I used CHO cells expressing CD86 to provide CD28 signal, individual antibodies were also crosslinked on CHO cells to account for any CHO cell effects. Following a 16 hours stimulation, I sorted activated CD25+ cells into naive (CD45RA+) and memory (CD45RA-) subsets. I used CD25 as an early marker of

T cell activation, since cells had been depleted of regulatory T cells. To understand the differences in naive and memory sensitivity to TCR and CD28, I profiled gene expression with RNA-seq, chromatin accessibility with ATAC-seq and active enhancers and promoters with H3K27ac ChIPmentation-seq. As expected, across different stimulatory conditions I observed a variable proportion of activated T cells (11-78% of all cells) (Table 2.1). However, by sorting only activated T cells I ensured that the measured gene expression reflected cell activation state induced by different signal intensities, while not being confounded by the variable percentage of cells that had undergone activation. As a control, cells were cultured in the presence of fixed CHO cells expressing FcR in the absence of stimulating antibodies. In this condition cells were not activated and were sorted for low CD25 expression, henceforth they are referred to as resting T cells.

2.3.2 Naive and memory cells have cell type specific signatures

I first applied principal component analysis (PCA) to the RNA-seq data and observed that PC1 reflected cell stimulation while PC2 corresponded to cell type (Figure 2.1 B). Indeed, the majority of the gene expression variance was explained by the stimulation (47%) and by the cell type (10%) (Figure 2.1 C). The separation of naive and memory T cells by PC2 indicated clear differences in transcriptional responses between them. Furthermore, the different conditions clustered together, capturing the gradient of stimulation intensity in both cell types, with high levels of the combination of TCR with CD28 (hTCR hCD28) stimuli being the furthest separated from unstimulated cells on PC1 and intermediate intensity of stimuli mapping in between. Surprisingly, strong CD28 alone (hCD28) was amongst the lower responding conditions in naive cells yet it clustered with the more highly stimulated conditions in memory cells. Furthermore, naive cells stimulated with strong CD28 alone clustered towards the memory cells along PC2. Thus, the transcriptional program of cell activation is modulated by the intensity of TCR and CD28 signals, and CD28 stimulation enriches for the characteristics of memory T cells.

To confirm that I had successfully sorted naive and memory cells I performed differential gene expression analysis in the resting cells (fold-change ≥ 2 and false discovery rate (FDR) ≤ 0.05) (Figure 2.2 A). The 289 genes upregulated in memory cells included genes involved in the migration of T cells (chemokine receptors including *CXCR3*, *CCR6* and

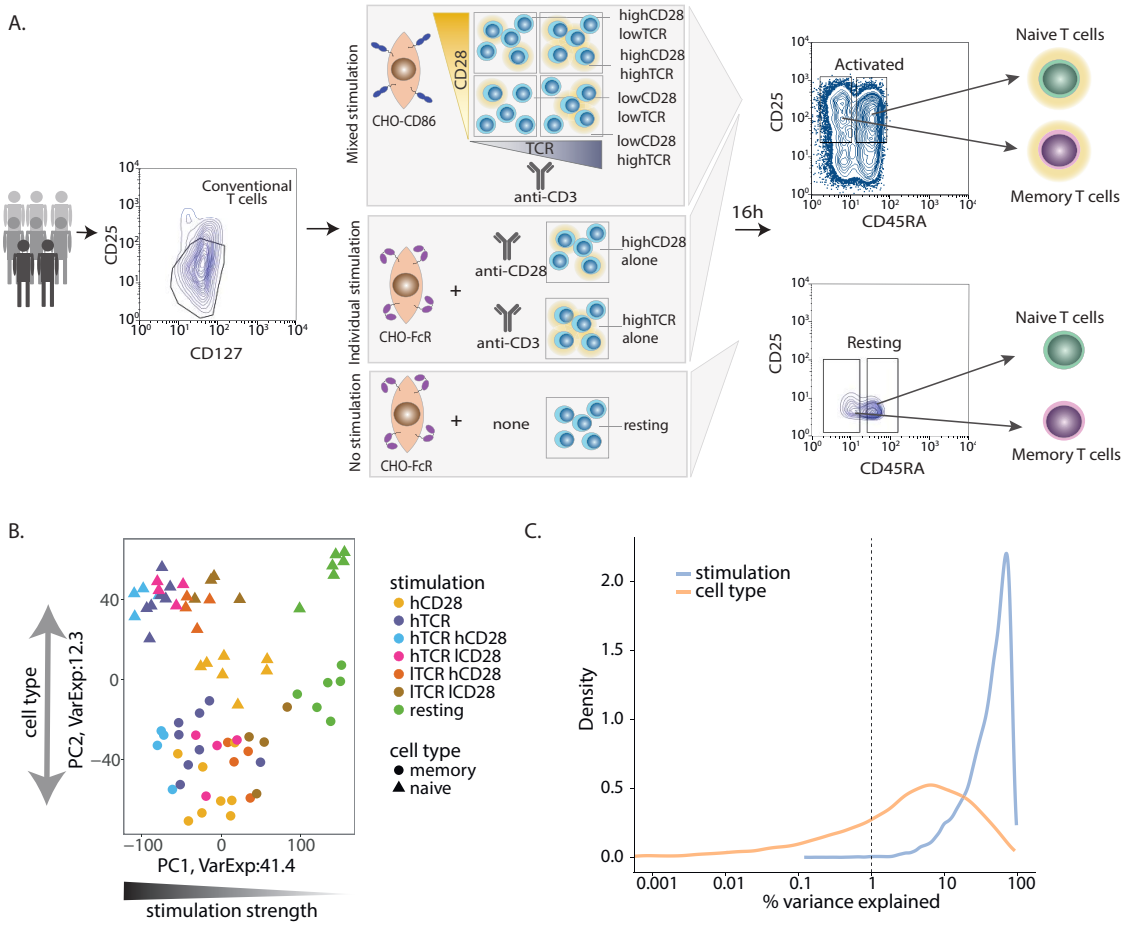

Figure 2.1: **Overview of study design and RNA-seq data. A.** Overview of the study design. CD4+ T cells were isolated from eight healthy individuals and cultured in six different stimulatory conditions, which included variable concentrations of anti-CD3 and anti-CD28 stimuli. In parallel, resting cells were cultured as a control. To ensure I measured cellular response to successful stimulation, I generated sequencing data from sorted activated naive and memory cells identified as CD4+CD25+CD45RAhigh and CD4+CD25+CD45RAlow, respectively. **B.** Principal component analysis using the expression of all genes. The first two components explain collectively 53.7% of the observed variability and correlate with stimulation strength and cell type. Each dot corresponds to an individual sample, colored by stimulation and shaped according to the cell type. **C.** The percentage of the total variance that can be explained by stimulation and cell type.

GPR1, cell adhesion molecules such as *CD58* (LFA-3) and B1 integrins), intracellular signalling (phosphatases, calcium signalling molecules, e.g. *SYT11*, *ITPRIPL1*, and kinases, e.g. *CDKN1A*), memory T cell survival and homeostasis (cytokine receptors (e.g. *IL1R1*, *IL2RB*, *IL12RB2* and *IL18RAP*), lectins and FAS) and transcription factors affecting T cell differentiation (e.g. *MAF*, *TBX21*, *RORC*, *BHLHE40* and *PRDM1*). *PECAM1* (CD31), a well known marker of a subset of naive T cells (Kimmig et al., 2002), was among the 33 observed genes upregulated in naive cells. Thus, the differential gene expression profile validated

my initial cell selection approach revealing clear and expected differences between naive and memory populations. The majority of the identified genes have already been documented to be differentially expressed between naive and memory cells, however, I also observed several genes that have not been reported before, including *STOM*, *AIM2* and *THBS1* in memory cells, and *FLT1* (VEGFR1) and *GNAI1* expressed at a higher level in naive cells. Recently, it has been suggested that THBS1 interaction with CD47 receptors determines regulatory T cells and long-lived memory T cells (Grimbert et al., 2006; Van et al., 2012). Importantly, I did not observe significant differences in gene expression variance between naive and memory cells across the eight donors (Wilcoxon rank sum test p-value = 0.48), indicating that the identified differentially expressed genes were not driven by individual samples and reflected consistent biological differences (Figure 2.2 B).

2.3.3 Naive and memory T cells operate different gene expression programmes upon activation

To understand cell type specific responses induced by the different stimuli, I compared gene expression profiles between resting and stimulated naive and memory T cells (false discovery rate (FDR) $\leq$ 0.05 and fold-change $\geq$ 2; Figure 2.3 A). As expected, the majority of upregulated genes was shared between the two cell types, however, naive cells displayed a larger number of differentially upregulated genes (DEGs) (mean: 1,240 genes) than memory cells (mean: 840 genes), with the exception of CD28 stimulation alone, where more genes were upregulated in memory cells. This likely reflects the greater changes in gene expression levels resulting from transitioning to activation from a deeper quiescent state in naive cells. Indeed, differential expression analysis between naive and memory cells in the resting state revealed a larger number of genes expressed highly in memory cells compared to naive.

When looking globally at the whole genome maps of H3K27ac and chromatin accessible regions (ATAC-seq), memory cells were characterised by more peaks in the resting state in both ATAC-seq (17.5% more peaks) and H3K27ac ChM-seq (10.6% more peaks). In order to gain a better understanding of the dynamic responses upon stimulation, I performed the same comparisons as with the RNA-seq data. At the mRNA level I observed 35% of the genes to be differentially expressed upon stimulation, however the

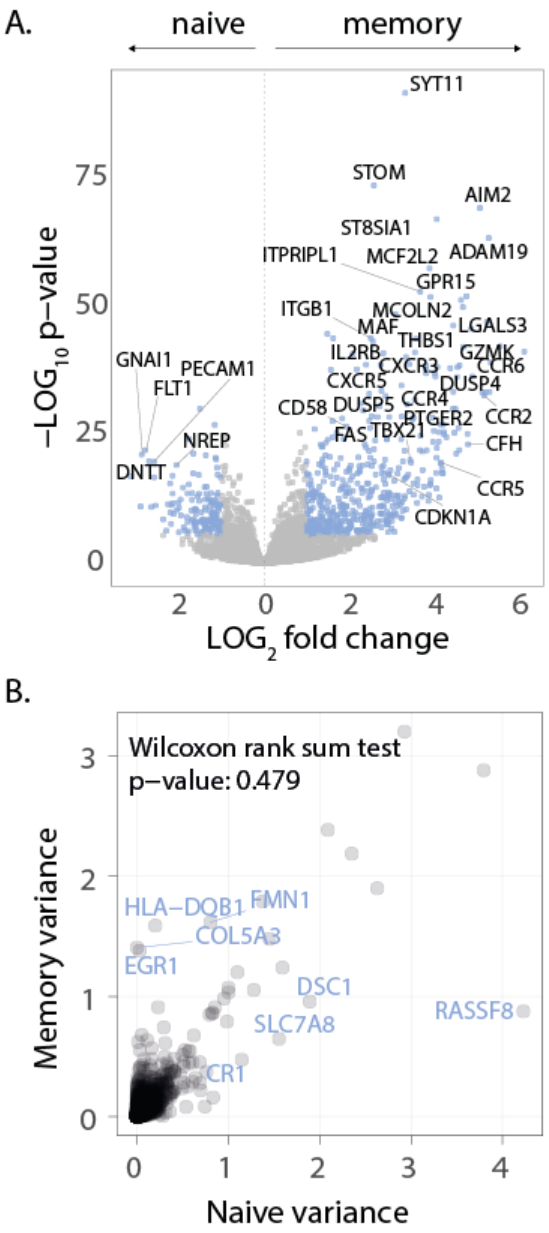

Figure 2.2: **Differential gene expression analysis between resting naive and memory CD4+ T cells. A.** Volcano plot of differential gene expression test between resting naive cells and resting memory cells. Genes colored in blue correspond to differentially expressed genes with log2 fold-change ≥ 1 and FDR $\leq 5\%$. Labelled are DEG with the lowest p-values. **B.** Wilcoxon rank sum test between the variance observed in memory and naive cells in the resting state.

chromatin regulatory landscape showed far fewer changes; only 8% of the chromatin changed in accessibility and 9% of the regions changed in histone acetylation. This is probably due to the fact that the majority of these genes are already expressed at low levels, which would be reflected by chromatin being already open and enhancers being marked by H3K27ac acetylation. Furthermore, similarly to the differential gene expression, the majority of differential histone modified regions (DMHRs; 54%) and

differentially accessible regions (DARs; 58%) were shared across responses to TCR, CD28, or both. However, as with the gene expression analysis, I observed a larger proportion of acetylation changes driven by CD28 alone in memory compared to naive cells (Fisher's exact test p-value RNA = $4.9x10^{-11}$; H3K27ac < $2.2x10^{-16}$), indicating that CD28, in line with my observations for the RNA, is also more potent in inducing chromatin changes in memory T cells. Conversely, in response to strong TCR stimulation alone, I detected a larger proportion of differentially acetylated H3K27 regions and chromatin accessible sites in naive cells compared to memory cells (Fisher's exact test p-value RNA < $2.2x10^{-16}$; H3K27ac < $2.2x10^{-16}$; ATAC < $2.2x10^{-16}$; Figure 2.3 B). As such, the observed high number of stimulus-specific differentially regulated regions suggests that there are unique chromatin remodelling changes acquired in both a cell type and stimulus-specific context.

Given the observation that TCR and CD28 induced a different number of changes in chromatin accessibility and histone modifications, I next assessed if there was a difference in the proportion of these chromatin activity changes that were also predictive of the observed differential gene expression (Wang et al., 2013b). I found that DARs and DMHRs were predictive of a large proportion of upregulated genes (mean 46.5%; Figure 2.4 A). Although, globally, TCR induced more changes in chromatin activity, the percentage of gene upregulation predicted by DMHRs and DARs was similar between TCR and CD28 (Figure 2.4 B). I observed that the majority of genes for which I had assigned predictive differentially regulated regions included peaks differentially regulated upon a specific stimulus (naive TCR 82.3%; memory TCR 70.6%; naive CD28 62.23%; memory CD28 50.2%). This indicates that each stimulus uniquely contributes to the gene regulatory landscape of a cell.

Whilst the majority of the genes were assigned a single differentially regulated region, a set of 69 genes displayed dramatic alterations in multiple regions of chromatin accessibility and histone modifications in response to one stimulus, indicating that they are highly sensitive to a particular signal (Figure 2.4 C). Among the genes with the highest number of differentially regulated regions (>10 regions) were *IRF4*, *DUSP5*, *IRF8*, *TNIP3* and *CD28*, all strongly modified by TCR alone. IRF4 protein abundance increases alongside TCR signal intensity increase, as it does at the RNA level, and programs the expansion of high-affinity T cell clones (Man et al., 2013). Other genes are known to be upregulated in naive cells upon TCR stimulation, but the mechanisms by which this

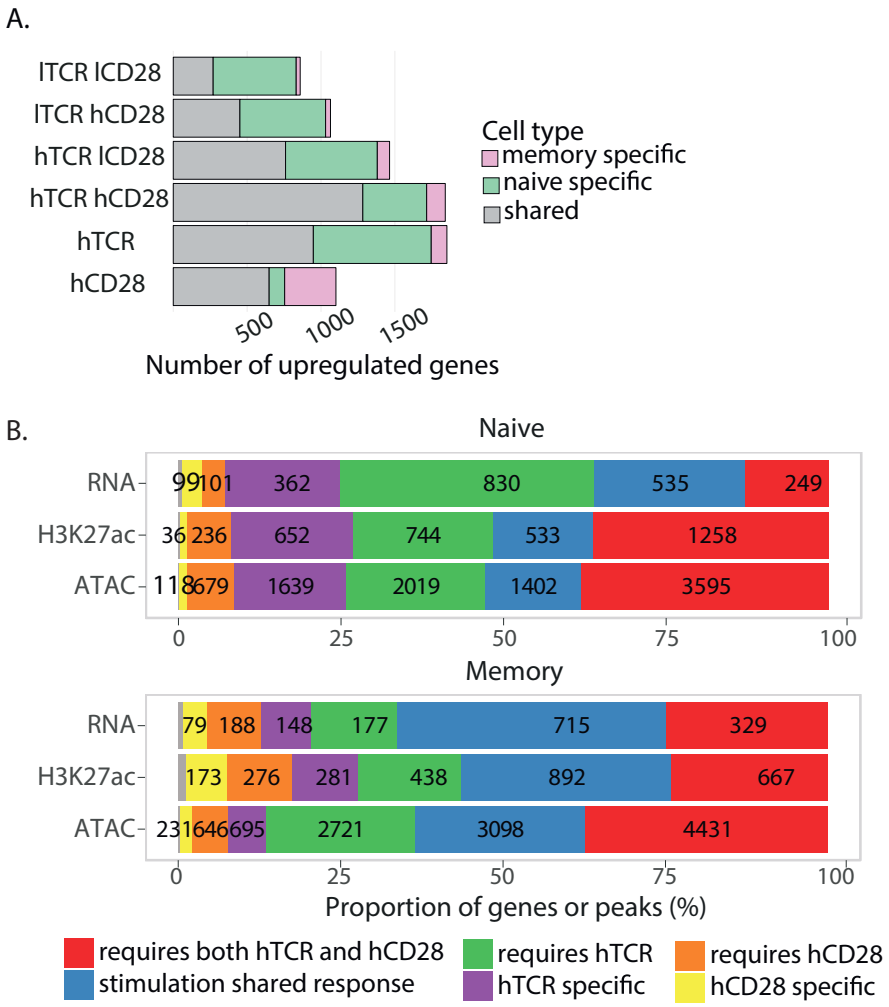

Figure 2.3: **Pairwise comparison between resting and stimulated states. A.** Number of upregulated genes upon stimulation defined from differential expression test between stimulated and resting cells (pairwise), with fold-change ≥ 2 and FDR $\leq 5\%$. **B.** Percentage barplot of differentially expressed genes (DEG), differentially accessible regions (DARs) and differentially histone modified regions (DHMRs) upon stimulation with both stimuli, strong TCR alone or strong CD28 alone. The percentage was calculated based on the total number of DEGs, DARs and DHMRs. Written inside the barplot is the corresponding number. The coloring corresponds to the overlap between the three stimulatory conditions.

occurs are unknown. For example it has been demonstrated that *DUSP5*-transgenic mice have impaired thymocyte positive selection by inhibiting ERK activation (Kovanen et al., 2008). Finally, for some of these genes it is the first time that a relationship between their levels and TCR has been suggested, such as *IRF8*. In comparison, the genes induced by strong CD28 alone revealed a smaller number of differentially regulated regions (>5 regions) but included interferon inducible chemokines *CXCL9*/*CXCL10*/*CXCL11* as well as the *IL13*/*IL5* locus, which encodes for classical Th2 cytokines. While hyperacetylation of the IL5 locus in response to co-stimulation, and the overall effect of CD28 in Th2 cell fate

had been documented (Avni et al., 2002), it is the first time that a similar effect has been observed in the CXCL locus. Therefore, despite the small percentage of regions changing upon activation, there is a subset of loci that undergo large chromatin configuration changes in order to regulate expression of nearby genes.

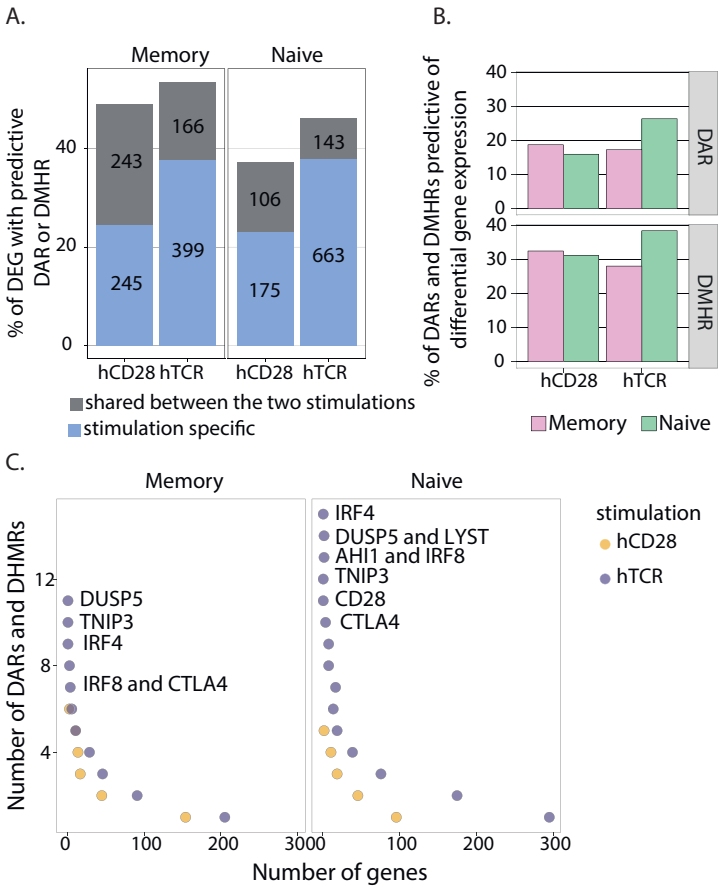

Figure 2.4: **Integration across assays of pairwise comparison between resting and stimulated states. A.** Proportion of differentially expressed genes that have at least one differentially regulated region nearby (< 150 kbp) as detected by ATAC-seq and ChM-seq on H3K27ac. The analysis was repeated using only the stimulus specific peaks. **B.** Proportion of differential accessible regions (DARs) and differentially modified histone regions (DMHRs) that regulate at least one differentially expressed gene (< 150 kbp away from TSS). **C.** Number or predictive DARs and DHMRs discovered per genes. Marked are the genes with the highest number of predictive regulatory elements.

Among the genes with predictive chromatin changes I identified a 26.4 kbp region that overlapped with the transcription start site (TSS) of *TBX21*, the gene that encodes for T-bet transcription factor. The acetylation changes were specific to naive cells and shared across all three stimuli (Figure 2.5 A). This revealed an increase in H3K27ac that was driven by both TCR and CD28 while memory cells already displayed some acetylated regions even in the resting state. In contrast, an example of a DAR detected in naive cells

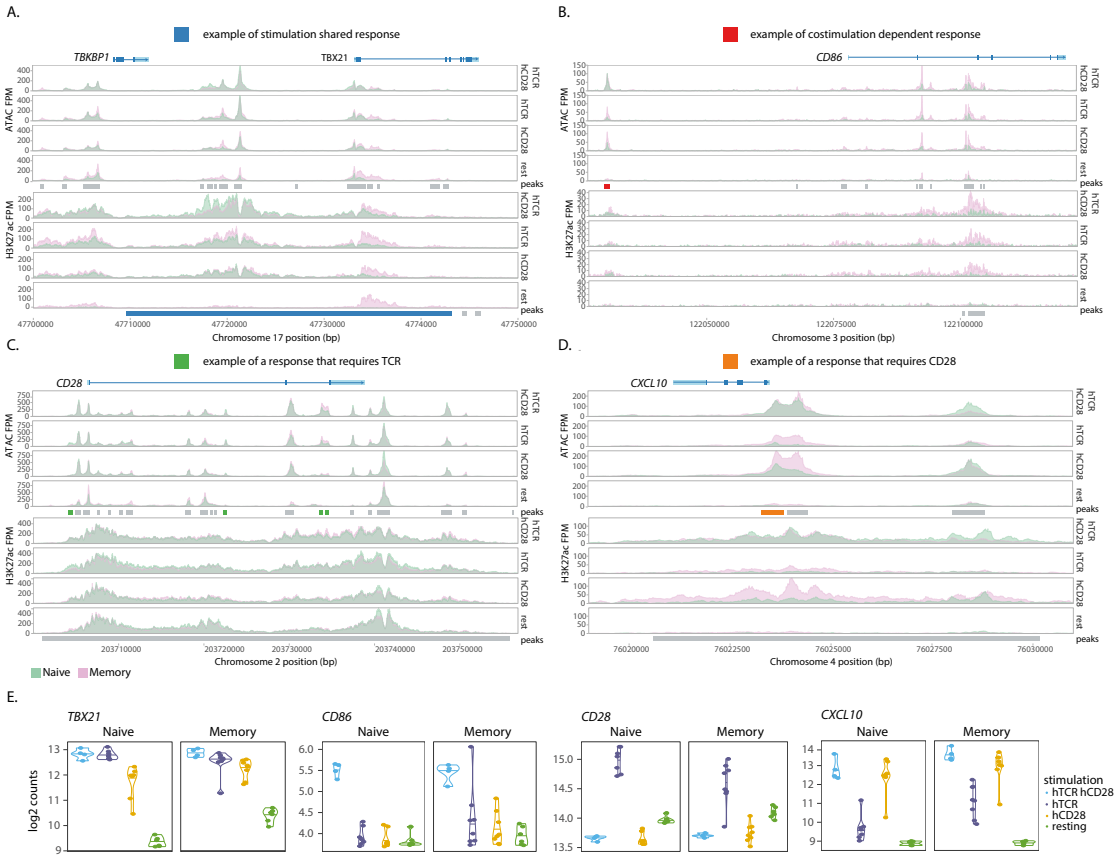

Figure 2.5: **Examples of a gene expression predictive DHMRs and DARs.** For the purposes of plotting the two donors for each assay were combined and averaged **A.** A predictive DHMR in naive cells that is a stimulation shared response (shared between hTCR hCD28, hTCR alone and hCD28 alone) in the transcription start site (TSS) of *TBX21*, the gene that encodes for T-bet. **B.** A DAR in naive cells that requires both hTCR and hCD28 in proximity to the promoter of *CD86*. **C.** Four DARs in naive cells that require TCR around the *CD28* gene. **D.** A DAR in naive cells that requires CD28 in the promoters of *CXCL10* gene. **E.** Gene expression profiles for the four genes examined above in four of the seven conditions.

that required both TCR and CD28 signals spanned 1 kbp and was located in proximity to the promoter of the *CD86* co-stimulatory molecule (Figure 2.5 B). Notably, this region was strongly co-stimulation dependent in naïve cells, but had a more relaxed response to stimulation in memory cells, where TCR, CD28 or both could trigger an opening in chromatin. The *CD28* gene had four DARs around its TSS and within its introns, and the chromatin accessibility change appeared contingent on the presence of TCR specifically in naïve cells (Figure 2.5 C). Lastly, an example of a DAR that required CD28 spanned 560 bp and localised on the TSS of *CXCL10* gene (Figure 2.5 D). Here, CD28 alone in memory cells and CD28 alone or as a co-stimulus in naive cells induced chromatin

changes, whereas TCR alone did not have an effect. All of these genes (Figure 2.5 E) are examples that showcase the stimulus-specific predictability of histone modifications and chromatin accessibility.

Finally, I investigated the expression and regulation of *IL2*, a gene that has been widely demonstrated to be co-stimulation dependent in memory cells (Ndejembi et al., 2006). *IL2* gene expression was upregulated in all but memory cells stimulated with low doses of CD28 and TCR (Figure 2.6 A). The biggest upregulation was observed in response to a combination of the two stimuli in high dose in both cell types, and in response to hCD28 only in memory cells. There were four chromatin accessible regions detected by ATAC-seq near the *IL2* gene, all of which were more open in memory compared to naive cells (Figure 2.6 B). Similarly, the H3K27ac profile around the gene highlighted greater activity levels in memory compared to naive cells. Interestingly, in memory cells both the chromatin activity and the accessibility decrease upon stimulation, despite the increase observed in gene expression.

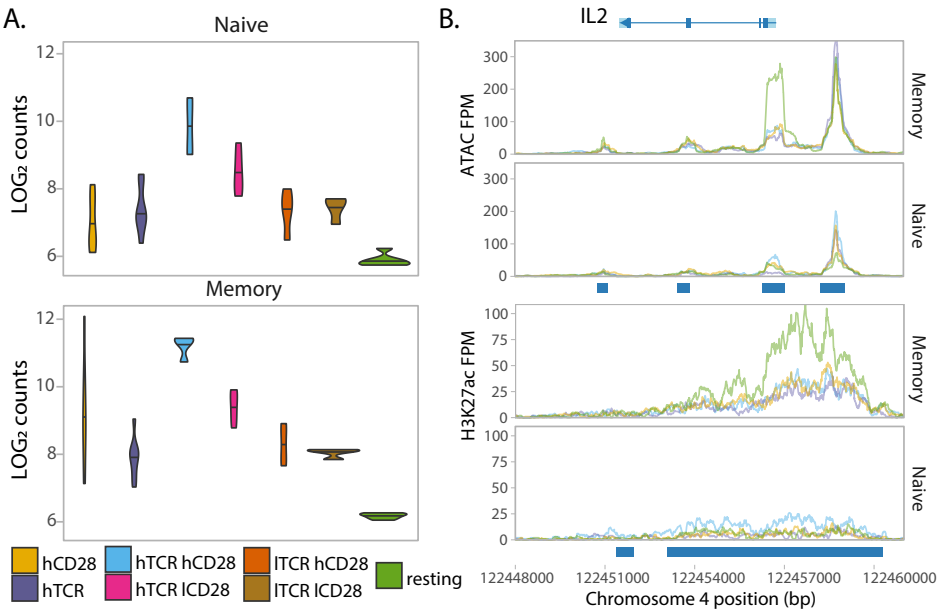

Figure 2.6: **Gene expression and chromatin regulation of *IL2*. A.** *IL2* gene expression per condition. **B.** ATAC-seq and H3K27ac ChM-seq profiles in the chr4:122448000-122461000 region surrounding the *IL2* gene. The read counts of two donors for each assay were combined and averaged for plotting purposes.

2.3.4 T cell effector functions are predominantly controlled by CD28

Based on the above observations I sought to better understand the role of each stimulatory signal in gene upregulation in the two cell types. To classify genes as either CD28 or TCR-sensitive, I used two models (linear and switch modes) of gene expression where the linear model reflects changes in stimulation intensity whereas the switch model reflects a digital on/off state (Figure 2.7 A). In both cell types together, I was able to assign a unique stimulus sensitivity to 1,567 genes, meaning that the expression of these genes was either sensitive to TCR or CD28 intensity. I observed that the majority of stimulus-sensitive genes (88%) followed the linear model (Figure 2.7 B).

I next assessed if naive and memory cells differed in sensitivity to the two stimuli. I observed that most of the stimulus-sensitive genes in naive T cells were TCR-sensitive (1,057; Figure 2.7 C), whereas a smaller number of genes was CD28-sensitive (n=363). However, a larger number of genes was CD28-sensitive (n=351) than TCR-sensitive (n=299) in memory cells. As such, a both TCR and CD28 sensitive genes were unevenly distributed between the two cell types, with a shift towards naive cells for TCR genes (Fisher's exact test p-value < $2.2x10^{-16}$) and a shift towards memory cells for CD28 genes (Fisher's exact test p-value < $2.14x10^{-5}$). Based on the pairwise comparisons across the six conditions against the resting state, I defined a group of genes that was upregulated upon stimulation. Of these, I observed that the expression of 1,228 genes in naive cells (55%) and only 490 genes in memory cells (29%) was sensitive to a single stimulus (Figure 2.7 D). This indicated that the majority of the upregulated genes in memory cells either responded to both stimuli (i.e. TCR or CD28 were both capable of driving the response) or they were truly CD28 co-stimulation dependent, requiring TCR and CD28 together.

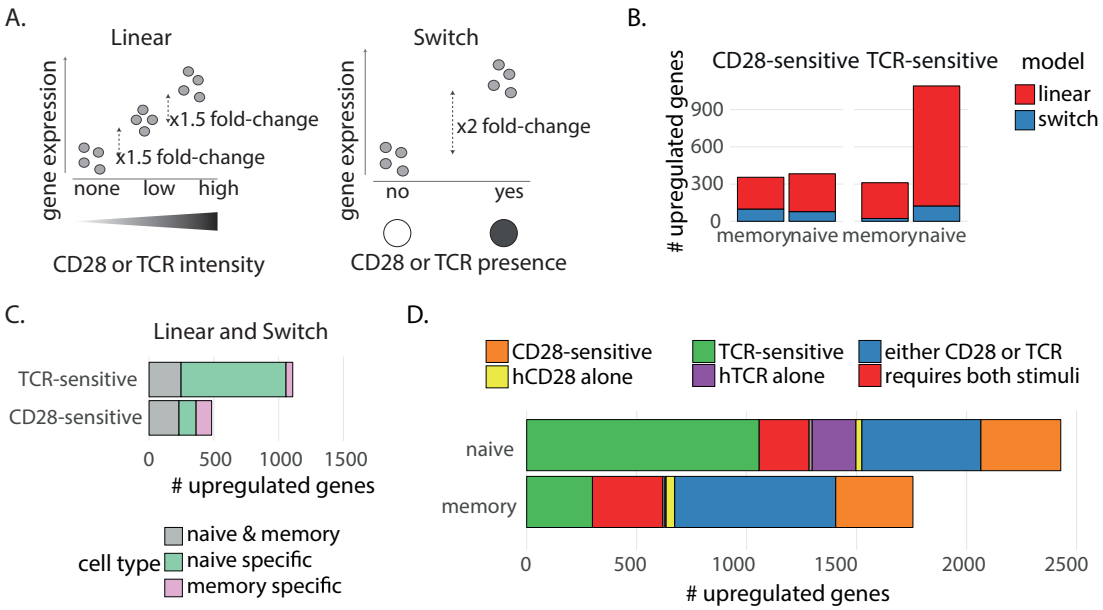

Figure 2.7: **Models of gene expression upregulation alongside stimulus intensity increase. A.** Classification of genes as CD28- or TCR-sensitive using different models of gene expression. In the linear model, I required a linear increase of gene expression along with stimulus intensity (incremental gene expression fold-change ≥ 1.5), separately evaluating naive and memory cells. Genes that did not follow the linear model were tested for the switch model. Here, I assumed an "on-and-off" mode of expression where a gene is significantly upregulated (fold-change ≥ 2) in response to the presence of either CD28 or TCR. In both of these models, I used all seven conditions, e.g. when testing for CD28-sensitive genes I grouped the TCR alone stimulation with the resting, since neither received a CD28 signal. A gene was classified in one of the two categories without overlap and prioritised for the linear model. **B.** Number of TCR and CD28 sensitive genes identified by the linear and the switch model in the two cell types. **C.** Comparison of the number of genes in naive and in memory cells under TCR or CD28 control. **D.** Number of upregulated genes upon stimulation. The colour represents different stimulatory dependencies. Grey indicates a small number of peaks that were only shared between hTCR alone and hCD28 alone stimulations.

Since the majority of human T cell stimulation experiments use both TCR and CD28 to activate cells, it is unclear which cell functions are controlled by TCR and which by CD28, and how they differ between naive and memory cells. Using this approach, I found that many classical T cell activation markers such as *EGR2*, *EGR3* and *CTLA4* were TCR-sensitive in both cell types, while *TNFRSF8* and *CD69* were TCR-sensitive in naive cells only (Figure 2.8 A), possibly explained by the fact that they were already expressed at high levels in resting memory cells (resting memory and naive cells *TNFRSF8* log2FC = 1.19 and *CD69* log2FC = 1) and thus not detected as differentially expressed upon stimulation of memory cells. In contrast, I found that the majority of cytokines and chemokines were CD28-sensitive (Figure 2.8 B and C). Among the chemokines, I observed that *CXCL9*, *CXCL10*, *CXCL11* and the *CCL25* CC chemokine were CD28-sensitive in both cell types. The

expression of cytokines essential for the differentiation of the major Th subsets were also under CD28 control, including *IFNG* (Th1), *IL4* and *IL13* (Th2) and *IL17A*, *IL17F* and *IL22* (Th17). In addition, I observed that the expression of the Treg transcription factor, *FOXP3*, was also CD28-sensitive. Taken together, these results demonstrate that a number of pathways associated with the effector functions of CD4+ T cells are predominantly controlled by the CD28 pathway in both naive and memory cells.

Different co-stimulatory and co-inhibitory receptors along with their ligands function at different stages of the T cell activation timeline, which might explain a different requirement on TCR or CD28 for their expression (Chen and Flies, 2013). I therefore investigated the control of expression of co-stimulatory ligands and receptors, since they play a key role in T cell activation (Figure 2.8 D). The expression of *CD28* itself was only upregulated upon strong TCR stimulation alone, suggesting that TCR signalling makes cells more receptive to CD28 engagement. However, the presence of CD28 stimulus impeded an increase in its expression, highlighting that this is not a self-reinforcing process. Other co-stimulatory receptors involved in the co-regulation of T cells were CD28-sensitive, in both naive and memory T cells. These included, *CD27-CD70* co-stimulatory pair and *CD274* (PD-L1). On the other hand, *CD80* and *ICOS* were CD28-sensitive specifically in memory cells. Thus, I was able to detect different modes of regulation for *CD28*, *CTLA4* and *ICOS*, despite the fact that they are all encoded within the same 260 kbp locus. This implies complex mechanisms of gene expression regulation in the two cell types and a fine-tuned control highlighting that co-located genes can be regulated by different modes of stimulation.

In order to confirm whether the observed linear changes at the RNA level were also observed at the protein level, I stimulated naive and memory cells using the same stimulation conditions and carried out flow cytometry analysis six days post-stimulation. I used two antibody panels; one containing selected antibodies for genes that had been determined to be TCR sensitive (CD25, CD28, CD40L, CD69 and CTLA4), and one with antibodies for genes that were found to be CD28 sensitive (CD25, CD27, CD80, ICOS, OX40 and PD-L1). I was able to replicate the linear relationship for all of the TCR sensitive genes in both cell types (Figure 2.8 E). On the other hand, I was not able to demonstrate a linear relationship at the protein level between any of the selected CD28 sensitive

genes. This suggested that CD28 sensitive genes might be subject to more complex post-transcriptional regulatory events than TCR sensitive genes.

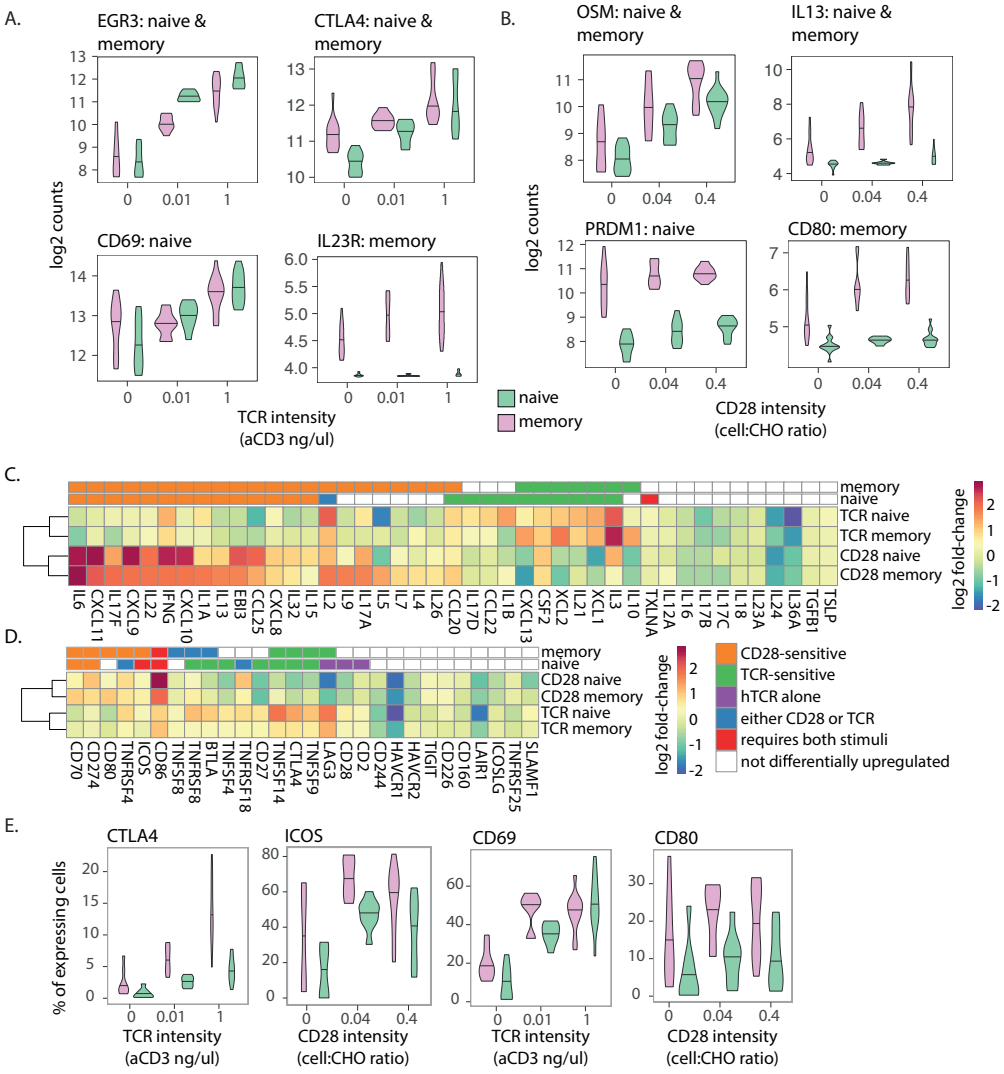

Figure 2.8: **Examples of TCR and CD28 sensitive genes in naive and memory cells. A+B.** Selected examples of TCR- or CD28-sensitive genes. The x-axis corresponds (**A**) to the level of TCR (anti-CD3 antibody in µg/ml) or (**B**) to the level of CD28 (proportion of T cells to CHO-CD86 cells) and the y-axis corresponds to the log2 counts of gene expression. Labels next to the gene name indicate the cell type in which stimulus sensitivity is observed. **C.** Cytokines and chemokines, and **D.** co-stimulatory ligands and receptors that are expressed in the dataset. Colouring represents the log2 fold change of gene expression **E.** Selected examples of protein levels for genes that were found to be TCR or CD28 sensitive by flow cytometry.

2.3.5 DNA replication and proliferation are driven by different stimuli in naive and memory T cells

To characterise whether genes sensitive to CD28 or TCR regulate the same cellular processes in naive and memory T cells I tested if these genes were over-represented in hallmark functional pathways (Liberzon et al., 2015) (Figure 2.9 A). I observed seven shared pathways enriched in both cell types. For example, IL2 signalling through STAT5 and TNF-α signalling via NF-κB was significantly enriched in both naive and memory cells and sensitive to both TCR and CD28. However, the expression of the majority of genes broadly classified as immune cell pathways, such as IL6 signalling through the Jak/Stat3 and interferon α and γ response, were CD28-sensitive in both cell types. In contrast, genes that are targets of Myc and E2F transcription factors were controlled by TCR in naive cells, which is concordant with *MYC* and *E2F6* genes being TCR-sensitive as well as with previous studies in mice (Allison et al., 2016).

Interestingly, I observed that some pathways were differentially sensitive to TCR and CD28 in the two cell types. Most notably, the G2M checkpoint, which marks DNA replication and cell division, was CD28-sensitive in memory cells but TCR-sensitive in naive cells, suggesting that commitment to cell division is more dependent on TCR in naive cells but driven by CD28 in memory cells. To functionally test this observation I stimulated CellTrace Violet (CTV) labelled naive and memory T cells with either anti-CD3 or anti-CD28 crosslinked by CHO cells expressing FcR (Figure 2.9 B). Five days following stimulation, T cell division was measured by flow cytometry. In accordance with the results from gene expression, naive T cells proliferated extensively following TCR crosslinking but mounted poor responses to CD28. In contrast, cross-linking CD28 was sufficient to induce division in memory cells whereas TCR stimulation was clearly much less effective. Thus, although combined TCR and CD28 co-stimulation is generally utilised to trigger T cell proliferation, this data indicates a division of labour between these stimuli where control of cell cycle in naive cells is generally TCR-sensitive, whilst in memory cells it is more dependent on CD28.

I noticed three genes (*CDC6*, *CDC20* and *CHEK1*) driving the enrichment of the G2M pathway, which were TCR sensitive in naive cells but switched to CD28 sensitivity in memory cells. I therefore sought to identify if there were more “switching” genes present in

the dataset, i.e. genes sensitive to the different stimuli between the two cell types. I identified a group of 18 genes that were TCR-sensitive in naive cells and changed to CD28 sensitivity in memory cells. Among others, I identified transferrin receptor *TFRC* (CD71) and DNA replication initiation factor *MCM10* (Figure 2.9 C and Table 2.3). Together these data highlight the enrichment of cell cycle/DNA replication pathway as targets for CD28 in memory T cells which, in contrast, are controlled by TCR in naive cells.

Table 2.3: **List of switcher genes.** N: Naive cell; M: Memory cell

Gene	LOG2FC M CD28	FDR M CD28	LOG2FC N TCR	FDR N TCR
TRNP1	0.635	$9.08x10^{-5}$	0.627	$2.06x10^{-3}$
CDC20	0.636	$1.56x10^{-4}$	0.88	$2.03x10^{-9}$
AK4	0.756	0.0351	1.407	$7.36x10^{-6}$
DTL	1.153	$8.09x10^{-8}$	0.807	$1.32x10^{-3}$
MTHFD2	1.149	0.0134	0.923	$1.04x10^{-4}$
CCL20	1.075	$3.34x10^{-4}$	0.971	$1.87x10^{-3}$
TFRC	0.691	$2.47x10^{-4}$	0.91	$5.16x10^{-8}$
STC2	3.051	$2.85x10^{-3}$	1.534	$4.81x10^{-4}$
EPB41L4B	0.771	$7.18x10^{-4}$	0.968	$3.82x10^{-4}$
TXN	0.68	$2.69x10^{-4}$	0.683	$1.86x10^{-4}$
MCM10	1.173	$1.62x10^{-8}$	1.107	$2.83x10^{-5}$
DNAJC12	1.011	$2.18x10^{-3}$	1.015	$3.72x10^{-5}$
DGAT2	1.07	$5.53x10^{-3}$	0.75	$2.99x10^{-6}$
CHEK1	0.727	$2.31x10^{-9}$	0.603	$1.57x10^{-8}$
NETO2	0.807	$1.45x10^{-4}$	1.355	$1.59x10^{-5}$
NLN	0.635	$6.08x10^{-3}$	0.919	$5.06x10^{-6}$
CDC6	0.658	$1.74x10^{-4}$	0.734	$4.65x10^{-7}$
NME1-NME2	1.072	0.033	1.054	$5.67x10^{-7}$

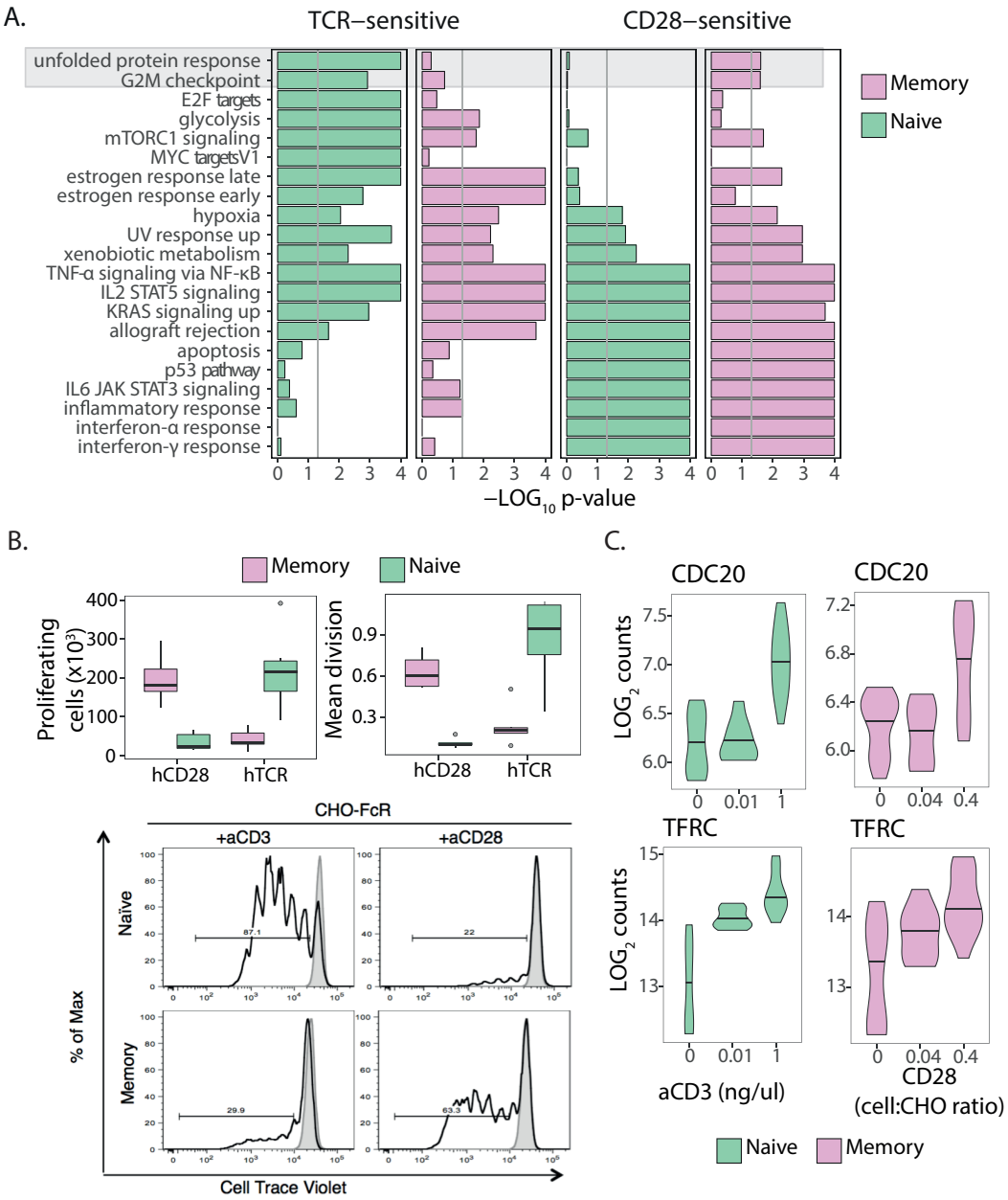

Figure 2.9: **TCR and CD28 sensitive genes enrichment across hallmark gene pathways highlights different stimulus sensitivity in naive and memory cells. A.** Measure of similarity between the genes that are TCR-sensitive or CD28-sensitive and a selection of the 50 hallmark gene-sets. The similarity coefficient was calculated using the Jaccard index and I used a permutation strategy to assess the likelihood of this being observed by chance (p-value < 0.05). **B.** Number of proliferating cells and mean division cycle upon TCR or CD28 stimulation separately for naive and memory cells. **C.** Switcher genes examples that play a role in cell cycle.

2.3.6 AP1 initiated transcriptional cascade is co-stimulation dependent in memory cells

To understand how TCR and CD28 stimuli exert effects on gene expression I tested for enrichment of transcription factor binding sites (TFBSs) in DARs identified by ATAC-seq that were concordant with gene expression (Figure 2.10 A). I tested 242 transcription factors with detectable levels of gene expression in the dataset.

I identified 5 different motifs enriched in naive cells and 20 motifs in memory cells (p-value $\leq$ 0.01). In naive T cells I observed an enrichment for Blimp-1, a transcriptional repressor that maintains T cell homeostasis (Martins et al., 2006), combined with interferon response elements, IRF2, IRF3 and IRF8. Increases in TCR signalling have been shown to induce the IRF4 transcription factor, which mediates Blimp-1 abundance in mice (Man et al., 2013). As expected, the enrichment was driven by TCR, with little enhancement in response to CD28 co-stimulation. Blimp-1 and the identified IRFs recognize a similar motif (Figure 2.10 B), and antagonize each others binding *in vitro* (Doody et al., 2010).

In memory T cells I observed that the profile driven by TCR was similar to naive cells, consisting of a combination of Blimp-1 with interferon response elements. However, in marked contrast, a more robust response was observed in the presence of CD28 co-stimulation. Notably, this was characterised by the enrichment of AP1 transcription factors (Figure 2.10 A). Specifically, only IRF4, JunD and BATF transcription factor binding sites co-occurred in regions that changed in response to strong TCR alone, whereas c-Fos, FOSL1, c-jun, jun-B and JDP2 all required the presence of CD28. C-jun and c-Fos constitute the backbone of AP1, a transcription factor that plays an important role in the induction of the immune response. Thus my observations are consistent with previous work which suggested that CD28 regulates the expression and activity of AP1 transcription factors (Shapiro et al., 1997; Fraser et al., 1991; Edmead et al., 1996). Interestingly, BATF and c-jun transcription factors can also form a heterodimer and cooperate with IRF4 and recognise AP1–IRF composite elements (AICEs) in pre-activated CD4+ T cells (Figure 2.10 B) (Li et al., 2012). These transcription factors play a crucial role in the initiation of transcriptional programs specific to T cell activation and cell division.

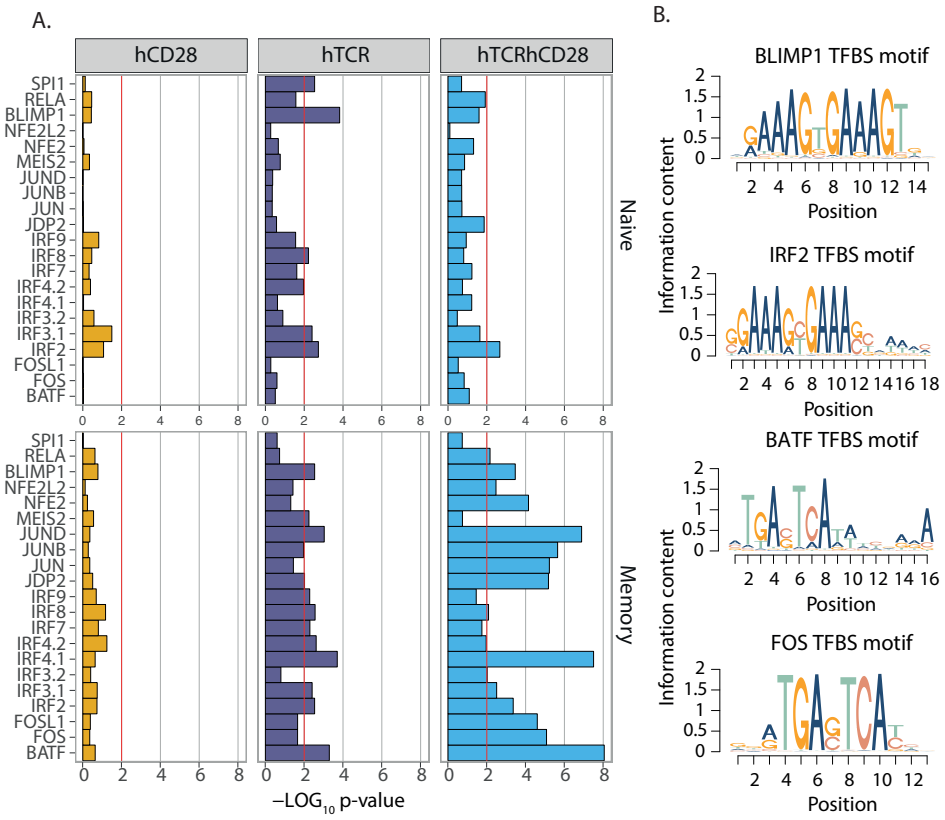

Figure 2.10: **Transcription factor enrichment in TCR, CD28 and TCR+CD28 induced peaks. A.** Transcription factor binding sites (TFBS) enriched in the differentially regulated ATAC-seq peaks that were assigned to predict gene expression changes. **B.** Position weight matrices (PWMs) for the enriched TFBS motifs.

2.3.7 Immune GWAS loci are enriched for CD28 sensitive genes

The role of T cell activation in the development of immune-mediated diseases is well established and single nucleotide polymorphisms (SNPs) nearby genes relevant to T cell activation, differentiation and trafficking have been implicated in autoimmune diseases through GWAS (Trynka et al., 2013; Farh et al., 2015; Hu et al., 2014; Okada et al., 2014). Therefore, I sought to investigate if immune disease associated loci were enriched for the genes identified as TCR or CD28-sensitive, thereby implicating the involvement of these stimulatory pathways in disease pathogenesis.

In the enrichment analysis I tested eight immune-mediated conditions, Crohn's disease (CD), ulcerative colitis (UC), celiac disease (CEL), rheumatoid arthritis (RA), type-1 diabetes (T1D), systemic sclerosis (SSc), multiple sclerosis (MS) and psoriasis (PSO). I used bone mineral density (BMD) as a negative control as I would not expect to observe significant enrichment among BMD loci. The majority of the tested immune diseases were

more strongly enriched (permuted p-value < 0.01) for CD28-sensitive than TCR-sensitive genes. An exception was T1D where genes sensitive to TCR showed a higher enrichment (Figure 2.11 A). In addition, I observed that TCR-sensitive genes were enriched in CEL loci in both cell types (permuted p-value < 0.0097) and in RA (permuted p-value = 0.0096) and MS (permuted p-value = 0.0054) specifically in memory cells. Taken together, this suggests that the variants associated with common immune-mediated diseases could regulate the expression of genes sensitive to CD28 and through this pathway modulate the outcome of T cell activation.

The majority of immune disease associated genetic variants fall in the non-coding regions of the genome and previous studies showed that disease associated variants are enriched in regions highlighting active enhancers (Trynka et al., 2013; Trynka et al., 2015). I therefore investigated if disease SNPs map within active or open chromatin regions as defined by H3K27ac or ATAC peaks near the genes driving the enrichment. Given that the majority of peaks were shared between the cell types and stimuli, these chromatin marks were uninformative in discriminating if disease SNPs were more significantly enriched in specific conditions. However, I observed that, on average across traits, 73% of the genes driving the enrichment also had at least one disease associated variant overlapping an active promoter or enhancer, identified based on the presence of H3K27ac. Among the GWAS loci with the highest number of SNPs falling in regulatory regions I identified a single locus that contributed to the enrichment observed for CD28 sensitive genes in both cell types for CEL, CD, UC and RA. The locus overlaps with *STAT1* and *MYO1B*. *STAT1* encodes for a transcription factor that responds to IFN-γ signalling and plays an important role in cell survival upon a pathogenic attack (Krause et al., 2006). In contrast, the TCR-sensitive *CTLA4* gene fell within a GWAS locus with the highest number of SNPs falling in regulatory regions in both cell types, and contributing to the enrichment observed in CEl, T1D and RA.

I then tested whether any of these SNPs disrupted a TFBS, limiting the analysis to the TFs that had previously been identified as enriched. I found 13 unique SNPs that disrupted a binding motif (Table 2.4). Two SNPs, associated with MS and in high LD with the reported index variant rs1021156 (R^2>0.8), disrupted the IRF TFBS within the *ZC2HC1A*/*IL7* locus (Figure 2.11 B). The first one, rs3808619 localised in the promoter of *ZC2HC1A* and the risk allele led to decreased binding by IRF family of transcription factors (Figure 2.11 B).

The second variant, rs60486739, located in intron 3 of *IL7*, and the minor allele also led to increased binding by IRF transcription factors (Figure 2.11 B). Both genes were sensitive to CD28 stimulation in memory cells and the H3K27ac peak that contained the rs60486739 variant was only present in stimulated memory cells, suggesting a potential functional role of the variant in modulating TF binding in this enhancer and affecting the expression levels of the gene.

Table 2.4: **Transcription factor binding sites in open chromatin or H3K27ac peaks that are disrupted by immune-disease associated variants. SNP**: Disease SNP residing within an ATAC peak and disrupting TFBS; **All**: SNP alleles; **Gene**: Gene that is stimulus-sensitive and maps to the disease locus; **Stim.**: Stimulus to which the gene is sensitive; **Cell**: Cell type in which disease gene shows stimulus sensitivity, M corresponds to memory and N to naive; **TF**: Transcription factor recognising the binding site; $\Delta_{\mathbf{PWM}}$: Difference in PWM score between the reference and the alternative allele; **Trait**: Disease for which the association was reported (T1D - type-1 diabetes, CEL - celiac disease, CD - Crohn's disease, UC - ulcerative colitis, RA - rheumatoid arthritis, MS - multiple sclerosis, PSO - psoriasis, SSc - systemic sclerosis).

SNP	All	Gene	Stim.	Cell	TF	$\Delta_{\mathbf{PWM}}$	Trait
rs2852151	G/A	SEH1L	TCR	N	RELA	-132	T1D, CEL
rs3181365	A/T	TNFSF15	CD28	M	BLIMP1	-249	CD, UC
rs6715826	T/G	MYO1B, STAT1	CD28	M	BLIMP1	-55	CEL, RA, CD, UC, MS
rs2548530	T/C	ERAP2	CD28	N	IRFs	-159	CD, UC
rs10061936	T/C	ERAP2	CD28	N	IRFs	-50	CD, UC
rs9392504	G/A	IRF4	TCR	N, M	IRFs	93	CEL, RA
rs4679081	T/C	TRIM71	TCR	M	MEF2C, IRFs	-32, -14	CEL, MS
rs2115592	T/C	REL	TCR	M	IRFs, BLIMP1	382, 14	CEL, RA, MS
186498:20:00	C/CT	SOCS1	CD28	N, M	IRFs	-24	MS
rs60486739	G/A	IL7, ZC2HC1A	CD28	M	IRFs	311	MS
rs3808619	A/C	IL7, ZC2HC1A	CD28	M	IRFs	-71	MS
rs11249219	C/T	CLIC4	CD28	M	IRFs	-24	CEL, PSO
rs3784789	C/G	CYP1A1, CYP1A2	CD28	N	SPI1	-172	CEL, SSc

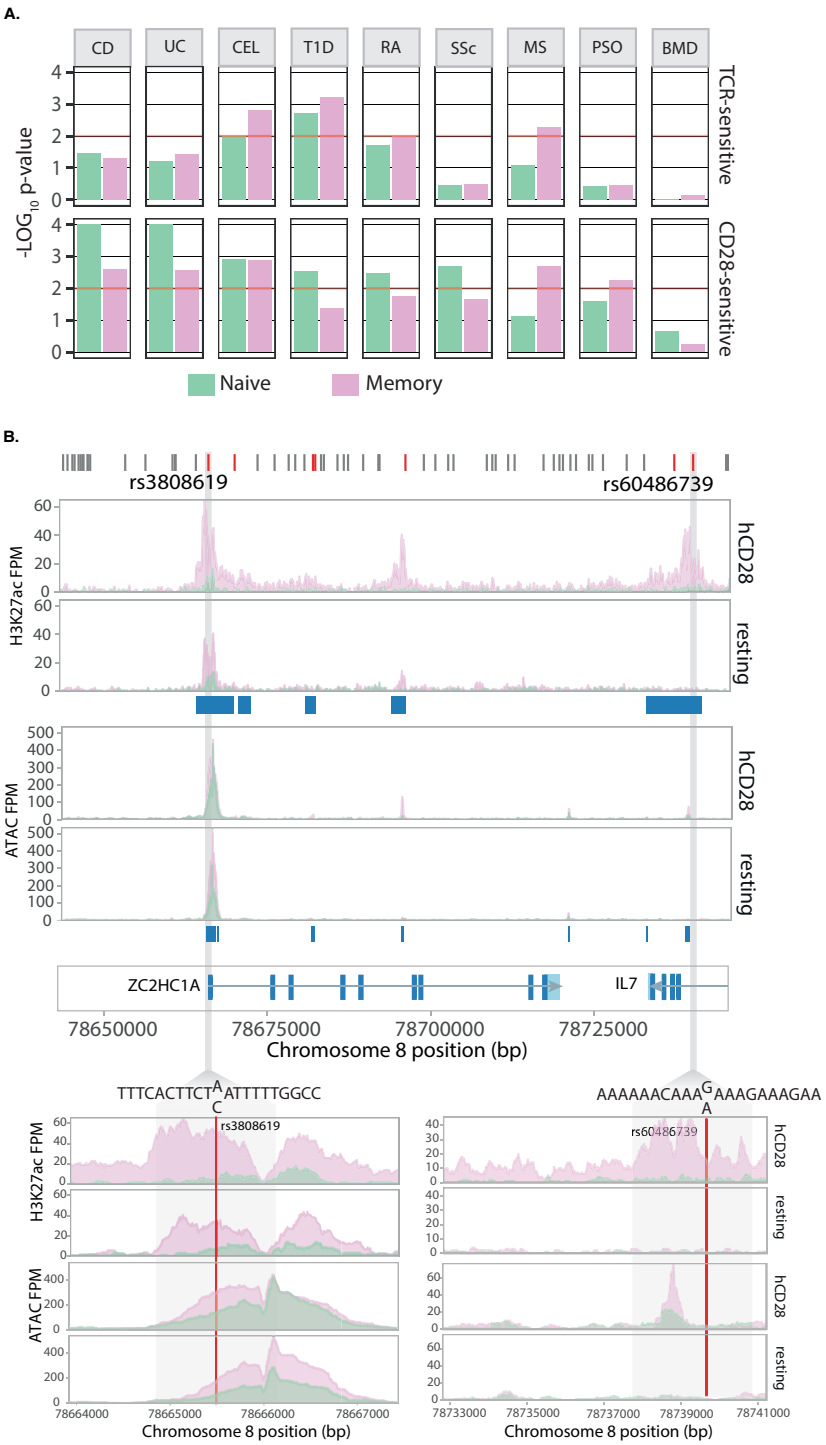

Figure 2.11: **TCR and CD28 sensitive genes enrichment in immune GWAS loci. A.** Enrichment of TCR- and CD28-sensitive genes in immune-mediated disease loci. Bone mineral density (BMD) is used as a negative control. (Crohn's disease - CD, ulcerative colitis - UC, celiac disease - CeD, type-1 diabetes - T1D, rheumatoid arthritis - RA, systemic sclerosis - SSc, multiple sclerosis - MS, psoriasis - PSO). **B.** MS associated locus containing two genes, ZC2HC1A and IL7, that are CD28-sensitive. In the upper panel in red are indicated all the SNPs in LD with previously reported GWAS index variant, rs1021156. Of these, the two highlighted variants, rs3808619 and rs60486739, overlap CD28-upregulated H3K27ac peaks and are predicted to disrupt and IRF binding site.

2.4 Discussion

Productive T cell activation is thought to involve recognition of antigen by the TCR in association with co-stimulatory signals via receptors such as CD28 (Chen and Flies, 2013). However, the relative quantities of both signals are likely to be highly variable depending on the setting of T cell activation. There is evidence to suggest that naive and memory cells have different requirements for CD28 co-stimulation, and a widely held view is that CD28 co-stimulation is less required for activation of memory than naive T cells (Dubey et al., 1995; London et al., 2000; Croft et al., 1994; Luqman and Bottomly, 1992). Understanding the requirements for CD28 co-stimulation during immune responses is important for many therapeutic approaches including immune suppression in autoimmunity and transplantation, as well as cancer immunotherapy.

Previous genomic studies have solely focussed on the additional effects of CD28 as a co-stimulus, and not on varying the levels of CD28 itself, and were carried out in populations of mixed cells which includes cells that have successfully undergone activation as well as the resting cells (Diehn et al., 2002; Wakamatsu et al., 2013; Allison et al., 2016). The data presented here reveal that CD28 has an important impact on memory T cells, particularly their proliferation. I was able to reach this conclusion thanks to three key components in the experimental design: (i) I specifically isolated and profiled only stimulated cells, therefore reducing the confounding effect of variable cell activation across different cell cultures, (ii) I separately assessed naive and memory cells, and (iii) I provided the first genome-wide perspective of gene expression regulation through mapping RNA and chromatin changes induced by strong CD28 stimulation in the absence of TCR. Thus, although providing CD28 on its own might not be physiological, this system allowed me for the first time to disentangle genes that are TCR or CD28 sensitive. I did not assay the CD25- cells within each experimental condition as I expected those to mostly represent resting cells. However, future time-course RNA-seq experiments could provide further insights into the dynamic nature of CD28 stimulus on memory cells and determine whether the totality of cells eventually gets activated. Recent research in CD8 naive T cells suggests that a weaker stimulus would indeed not affect the type of the response, but the time it takes to be initiated (Mehlhop-Williams and Bevan, 2014; Richard et al., 2018).

I observed a significant upregulation of genes in memory T cells in response to stimulation with CD28 alone that were related to cell cycle, suggesting a role for CD28 in memory cell proliferation. Indeed, I validated this directly by stimulating T cells *in vitro* and demonstrating proliferation in memory T cells following activation with CD28 alone. Furthermore, I observed a significant upregulation of effector cytokines and chemokines in response to CD28, indicating that CD28 is also key to cell effector functions. However, the experiment was performed on a mixed memory cell population of effector and central Th cells, which differ in their levels of expression of CD28 (Koch et al., 2008). The differences in the expression levels of CD28 might have an impact in the immune response initiated, given that influenza infected mice treated with CTLA4Ig shift their memory T cell pool to one predominantly composed of central memory cells, while influenza specific memory cells in untreated mice were predominantly effector cells (Ndejembi et al., 2006). This could explain why activated effector memory T cells produce cytokines more efficiently than central memory T cells (Barski et al., 2017). Finally, the variability in expression of the CD28 receptor between young and old is only significant in effector memory T cells (Koch et al., 2008), suggesting that the ability of memory T cells to get activated is directly concordant with their ability to sense CD28, and as I have here demonstrated, initiate proliferation. Therefore, the lack of CD28 signal might contribute towards immunosenescence.

The data presented here provide a new context for the interpretation of several previous observations on CD28 function. Firstly, the ill-fated CD28 superagonist antibody (TGN1412) trial for B-cell chronic lymphocytic leukemia and RA, which tested the ability of CD28 co-stimulation to specifically expand and activate Tregs, revealed a powerful effector response to CD28 stimulation driven by effector memory T cells (Hünig, 2012). Secondly, recent data indicated that human Treg cells can be expanded by utilising CD28 antibodies alone (He et al., 2017). This is in line with my observations, given that Tregs predominantly consist of memory cells and that CD28 stimulation upregulates *FOXP3*. Thirdly, there is now increasing evidence using conditional deletion of *CD28* in mice that memory T cell responses are dependent on CD28 stimulation (Linterman et al., 2014; Ndlovu et al., 2014; Fröhlich et al., 2016). Finally, I showed that a strong TCR signal alone is sufficient to induce the expression of key drivers of cell division and consequently trigger proliferation of naive T cells, but surprisingly had smaller effect on the proliferation of memory T cells. As such, I conclude that memory cells are not

simply more sensitive to activation signals, but that there is a true difference in the TCR and CD28 requirement between the two cell types. This is further supported by a recent study which demonstrated that naive CD8 T cells proliferate in response to low antigen concentrations, in contrast to memory T cells, which while detect the antigen, fail to engage the TCR signalling cascade (Mehlhop-Williams and Bevan, 2014). Taken together, this data provides a genomic explanation for the requirement of CD28 in the activation and effector functions of memory T cell.

My findings have further implications for understanding susceptibility to complex immune-mediated diseases, where T cell activation is one of the hallmark pathobiological processes. GWAS of immune diseases have mapped hundreds of associated risk loci, many of which harbour genes of immune function. However, the specific role of the identified genes in T cell activation processes is unclear. By examining T cell gene expression sensitivity in response to specific stimuli I demonstrated that GWAS loci are enriched for CD28-sensitive genes, rather than TCR-sensitive genes, thereby increasing support for the role of T cell activation via CD28 co-stimulation in susceptibility of immune-mediated diseases. For example, a recent study identified that cytokine oncostatin M (*OSM*) is expressed at higher levels in inflamed intestinal tissues from IBD patients compared to healthy controls (West et al., 2017). In the dataset presented here *OSM* is CD28-sensitive and is among the genes driving the enrichment of CD28-sensitive genes in IBD. The importance of CD28 co-stimulation in immune-mediated diseases is further supported by data from the CTLA-4 field (Kuehn et al., 2014; Lo et al., 2015). Loss of CTLA-4 in mice and heterozygous mutations in humans reveal profound autoimmunity where enteropathy is a consistent feature (Tivol et al., 1995; Schubert et al., 2014; Kuehn et al., 2014). The fact that the CTLA-4 pathway directly regulates CD28 stimulation by competing for the same ligands strongly suggests that CD28 plays a key role in susceptibility to immune-mediated diseases and that increased CD28 co-stimulation may be sufficient for effector T cells to support their survival, proliferation and secretion of proinflammatory cytokines (Khattri et al., 1999; Qureshi et al., 2011). The data shows that memory T cells are highly sensitive to CD28 stimulation, which is consistent with these cells being under the control of CTLA-4.

Lastly, the concept that CD28 is involved in the proliferation of memory T cells is intriguing in the light of recent data related to checkpoint blockade for cancer treatment.

It has been suggested that PD-1 blockade, which is important to the reinvigoration of exhausted effector T cells, requires CD28 signalling (Hui et al., 2017; Kamphorst et al., 2017). Again, this aligns well with the data presented here and supports the concept that differentiated memory T cells in tumours utilise CD28 and that CTLA-4 and the PD1 blockade are known to trigger autoimmunity (Nasr et al., 2017; Bertrand et al., 2015).

Taken together, this chapter provides new insights into the role of TCR and CD28 co-stimulation in the activation and proliferation of human naive and memory CD4+ T cells, and the influence of these stimuli on immune disease susceptibility.

3 Regulation of gene expression in regulatory T cells in response to cell stimulation

Collaboration note

The data acquisition described in this chapter was performed in collaboration with Natalia Kunowska, who was a Senior Research Assistant in Gosia Trynka's lab at the time. Natalia also optimised the ChM-seq protocol. I performed the flow cytometry validation and Blagoje Soskic analysed the flow cytometry data. I did all the genomic data analysis. RNA-seq library construction and sequencing of all materials was done by DNA Pipelines core facility at Sanger.

3.1 Introduction

Regulatory T cells (Tregs) play an essential role in the homeostasis of the immune system and the downregulation of inflammation by suppressing the proliferation and effector function of other T cells. Therefore, Treg cells constitute an essential component in the prevention of autoimmunity and the maintenance of self-tolerance. Tregs are characterised by high expression of the forkhead box transcription factor (TF) FoxP3, which is essential for their development and function (Fontenot et al., 2005b). Mutations of the *FOXP3* gene in humans leads to immune dysregulation polyendocrinopathy, enteropathy, X-linked syndrome (IPEX), an autoimmune disease characterised by very low Treg cell numbers (Bennett et al., 2001).

This has established FoxP3 as the master regulator of Treg cell fate. However, genetic and epigenetic studies have highlighted that the complex regulation of Treg cell identity does not depend on FoxP3 alone. Differential gene expression analysis between Tregs and CD4+ T helper cells in resting and activated states have shown that the Treg gene signature is one distinct from other T cells. A number of molecules form the Treg

gene signature, including the phosphatase *DUSP4*, the membrane protein *LRRC32*, the receptor *CTLA4* and the TFs *IKZF2* and *IKZF4* (Pfoertner et al., 2006; Birzele et al., 2011). It has been shown that the development of Treg cells is contingent on the establishment of a Treg specific DNA hypomethylation pattern which gets established in the thymus (Ohkura et al., 2012). Hypomethylation of *CTLA4*, *IL2RA* and *IKZF4* therefore precedes FoxP3 expression, but requires TCR stimulation and is necessary for lineage stability and suppressive capacity. Furthermore, the enhancers of *CTLA4*, *IL2RA* and *IKZF2* are active, as assessed by H3K27 acetylation from a pre-Treg stage (Kitagawa et al., 2017). This places a number of different genes at the epicentre of Treg identity, not all of which have been thoroughly studied in different cellular settings.

Given the importance of regulatory T cells (Tregs) in the downregulation of the immune system and the implications of their role in the pathogenesis of autoimmune diseases (Buckner, 2010), a more comprehensive understanding of Treg cell regulatory processes is necessary. Specifically, it is unclear how Tregs gene expression regulation profile changes upon activation. T cells are typically activated using a combination of anti-CD3 and anti-CD28 antibodies or PMA combined with ionomycin, which mimics a strong TCR signal. PMA diffuses into the cell and directly stimulates protein kinase C which initiates the MAPK cascade, while ionomycin triggers a Ca2+ influx into the cell (Weiss and Imboden, 1987). Combined, they can activate the three TF pathways, NFAT, NF-κB and AP1. Since Treg cells constitute a small percentage of CD4+ T cells, the implementation of high-throughput assays has been difficult due to limitations in obtaining sufficient amount of cellular material. For this reason, there are scarce resources of data available and the majority of the work has been carried out in mice.

In this study, I adapted the Chipmentation (ChM)-seq protocol, a recently published modification to the widely used ChIP-seq assay, to generate high quality data. For the optimisations I used the Jurkat cell line followed by primary human Treg cells. Comparison with the traditional ChIP-seq assay showed high correlation between the two methods, both in the cell line and in primary cells. ChM-seq was then used to study the gene regulatory landscape of Treg cells upon stimulation. I performed ChM-seq of H3K4me3 to assess promoters and of H3K27ac to assess active elements on ten human donors in resting and stimulated state. I also acquired open chromatin profiles using ATAC-seq and gene expression profiles using RNA-seq. Using this multi-layer

dataset, I identified gene regulatory programs initiated upon Treg cell stimulation at the chromatin level. Furthermore, I investigated the dynamics of gene transcription in Treg activation by assessing gene expression levels globally, as well as the alternative splicing and the impact on the usage of different gene isoforms by comparing transcript ratios.

3.2 Materials and Methods

3.2.1 Sample collection

Whole blood samples were obtained from healthy adults and CD4+ cells were isolated as described in Chapter 2. CD4+ enriched cells were stained with the following antibodies for sorting: CD4+ (OKT4)-APC (Biolegend); CD25 (M-A251)-PE (Biolegend); CD127 (eBioRDR5)-FITC (eBioscience) and Live/Dead fixable blue dead cell stain. Live regulatory T cells (CD4+ CD25high CD127low) were sorted. Treg cells used for the optimisation of ChIP and ChM assays were obtained from the NHS Blood and Transfusion in Cambridgeshire. The ten samples used for the comparison between resting and stimulated Tregs were obtained from a commercial vendor with commercial vendor's approved institutional review board protocols. Research use was approved by the Research Ethics Committee (reference number: 15/NW/0282).

3.2.2 Cell culture

Human Tregs and Jurkat human T-lymphoblast cells were grown in Iscove's Modified Dulbecco's Media (IMDM) (Life Technologies, Paisley, UK), supplemented with 10% human serum (HS), 50 U/ml penicillin and streptomycin (Life Technologies) and 100 U/ml recombinant human IL-2 and incubated at 37°C in a humidified atmosphere of 5% CO_2. Cells were activated using PMA (5-10 ng/µl) with ionomycin (200 ng/µl) (Sigma-Aldrich) overnight (18 hours). See Table 3.1 for full specifications.

3.2.3 ChIP-seq and ChIPmentation-seq

In order to reliably compare the ChIP with the ChM protocol, I sonicated 500,000 Jurkat cells, and equally divided the samples to continue with the ChIP and ChM protocols. Cells were sonicated for 40 minutes (T regulatory cells) or 30 minutes (Jurkat cells) in

250 µl using the Bioruptor®Pico (Diagenode, Belgium) sonicator and DNA LoBind tubes (Eppendorf, Germany) to obtain 200 to 3000 bp fragments. Resting and stimulated Tregs were sonicated for 6 minutes and 5 minutes, accordingly, using Diagenode tubes. I tested sonicated chromatin fragment sizes using gel electrophoresis. I used 100,000 cells of each for qPCR confirmation and proceeded with sequencing the remaining 100,000 cells. Once the protocol was deemed reliable I repeated the same steps using isolated T regulatory cells.

Both ChIP-seq and Chipmentation were performed using the Low SDS iDeal histone ChIP kit (Diagenode), as described in Chapter 2. The immunoprecipitation was done using 1µg of each antibody; H3K27ac (Cat. no. C15410196, Diagenode), H3K27me3(Abcam), H3K4me1(Active Motif) and H3K4me3 (Catalog No: 39915, Active Motif). The only difference to the protocol described in Chapter 2 was that all the washes with iW1, iW2, iW3 (iDeal ChIP-seq Kit for Histones, Diagenode) were performed using an Agilent Bravo Automated Liquid Handling Platform (Agilent, Santa Clara, U.S.). ChIP-seq libraries were performed using Illumina TruSeq index tags, while ChM-seq libraries were performed using the Nextera dual index tags.

The samples used for protocol optimisation were indexed and polled in equimolar concentration and sequenced on the MiSeq 2500 platform as such: 7 ChIP samples and 7 ChM samples produced from Jurkat cells (5 antibodies, ATAC-seq and input) were sequenced across 3 lanes to generate 62 million reads per sample; 5 ChIP samples and 5 ChM samples from 2 donors (4 antibodies and input) were sequenced across two lanes each to generate 27 million reads per sample; 13 samples, including five ChM samples from one donor were sequenced across two lanes to generate 32 million reads per sample. Thirty-one (because it was combined with different experiments) ChM libraries were indexed and pooled in equimolar concentration and sequenced on eight lanes using the Illumina HiSeq 2500 platform and V4 chemistry using standard 75 bp paired-end reads. Sequencing yielded on average for the samples in this study 62 million reads per sample.

3.2.4 ATAC-seq

ATAC-seq was performed according to protocol (Buenrostro et al., 2013), with the following modifications. After sorting, T cells were washed with ice-cold PBS and resuspended

in sucrose buffer (10 mM Tris pH 8, 3 mM CaCl2, 2 mM MgOAc, 1 mM DTT, 0.32 M sucrose, 0.5 mM EDTA, 0.25% TritonX-100), followed by 5 minutes incubation on ice to isolate the nuclei. Isolated nuclei were washed once with 1x TD buffer (Tagment DNA Buffer, Nextera DNA Library Prep Kit, Illumina, U.S) and resuspended in 50 µl 1x TD buffer containing 2.5 µl of Tn5 enzyme (TDE1, Nextera). The reaction was carried out at 37ºC, mixing and then stopped by addition of 250 µl of buffer PB (MinElute PCR Purification Kit, QIAgen, Hilden, Germany). The DNA was then purified on MinElute columns according to the manufacturer's instructions and eluted in 10 µl sterile double distilled water. The libraries were amplified using the Nextera PCR Master Mix from Nextera DNA Library Prep Kit and Index adapters i7 and i5 (Nextera Index Kit, Illumina, U.S), according to the manufacturer's instructions. The number of amplification PCR cycles for each sample was determined individually by performing a qPRC reaction of 7.5 µl aliquot of the mix with an addition of the EvaGreen dye (Biotium, Fremont, U.S.). The amplified libraries were SPRI purified (upper cut 0.5x, lower cut 1.8 x) on a Zephyr G3 SPE Workstation (PerkinElmer, Waltham, U.S.), multiplexed the libraries and did 75 bp sequencing using an Illumina HiSeq V4 to yield on average 147 million reads per sample.

Table 3.1: Donors specifications and culture conditions.

Ind. ID	Batch	Age	Sex	PMA (ng/µl)	Cells/mL	FoxP3+ (%)	RNA	ATAC	H3K27ac	H3K4me3
20	1	NA	M	10	1	83.5	Yes	Yes	No	No
21	1	NA	F	10	1	78	Yes	Yes	No	No
22	1	NA	F	10	1	68.6	Yes	Yes	No	No
25	2	26	F	5	1	NA	Yes	Yes	Yes	Yes
26	2	22	M	5	1	NA	Yes	Yes	Yes	Yes
27	2	31	M	5	1	NA	Yes	Yes	Rest.	Yes
28	3	22	M	5	2	NA	Yes	Yes	Yes	Yes
29	3	36	F	5	2	NA	Yes	Yes	Yes	Yes
30	4	39	M	5	2	80.7	Yes	Yes	Yes	Yes
31	4	25	F	5	2	78.5	Yes	Yes	Yes	Yes

3.2.5 ChM and ATAC data processing

The quality of the sequence reads was assessed using the fastx toolkit and the adaptors were trimmed using skewer (Jiang et al., 2014). Reads were mapped to the human genome reference GRCh38 using the bwa mem algorithm (Li and Durbin, 2009). I only kept uniquely mapped reads, removed PCR duplicated reads and for the ATAC I excluded mitochondrial reads using samtools (Li et al., 2009). Genome browser tracks were cre-

ated using BEDTools (Quinlan and Hall, 2010) and the UCSC binary utilities. Furthermore, I generated insert size distributions using PICARD tools CollectInsertSizeMetrics function, since these can be indicative of over-sonicated chromatin and excess of adapters in the data.

The ATAC mapped reads were converted into bed files and chimeras were removed. Peaks were called using MACS2 (Zhang et al., 2008) setting the parameters to -f BED -q 0.05 –nomodel –extsize 50 –shift -25 for ATAC. Peaks were called directly from the bam files for the ChM samples, using the parameters -q 0.01 for H3K4me3 and H3K4me1 and –broad –broad-cutoff 0.1 –nomodel –extsize 146 for H3K27ac and –extsize 73 H3K27me3. For ChM, all samples were downsampled to to the same read number prior to peak calling against the input. I used the fraction of reads in peaks (FRiP), the insert size distribution and the genome tracks to investigate the quality of our data. I excluded one H3K27me3 ChIP-seq sample from Tregs and one Treg ATAC sample in the stimulated state for having FRiP < 5%.

For the comparison between ChIP and ChM, I merged all samples from the same cell type, antibody and method, and downsampled to the smallest read number, which was 19 million reads for ChIP and 41 million reads for ChM. I called peaks in H3K4me1 and H3K4me3 using a -q value threshold of 0.001, of 0.1 for H3K27ac and a p-value of 0.001 for H3K27me3.

For the comparison between resting and stimulated Tregs, I merged all ATAC samples and called peaks again using the parameters described above. For the H3K27ac and H3K4me3 ChM samples, I first merged the donors within each condition (resting and stimulated) and then randomly sampled 17 million reads from each into one sample to reach the same read number as in the input. Since I merged QCed samples I used the –keep-dup flag when calling peaks with MACS2, as the PCR duplicated reads for individual samples were already removed and I expected to observe a small proportion of the same reads present in independent samples by chance. I also increased the -q value threshold to 0.1 for both assays. The resulting peak files were used as a reference to count the number of reads falling into peak regions using featureCounts (Liao et al., 2014). I only kept regions that had at least 20 reads in at least 3 samples to get a final count of 54,811 in ATAC, 20,363 in H3K4me3 ChM and 44,820 in H3K27ac ChM.

3.2.6 Global ChIP, ChM and ATAC data overview

To correlate the samples I divided the genome into 10 kbp windows and counted the number of reads that fell within each, normalised by the total number of reads. I then calculated the Pearson correlation coefficient. The overlap among different samples, the distance from the transcription start site (TSS) of a gene and the permutations for the enrichment of the different chromatin marks in promoter capture Hi-C regions mapped in CD4+ T cells (Javierre et al., 2016) was calculated using BEDtools (Quinlan and Hall, 2010).

3.2.7 Differential regulatory regions analysis

To carry out differential regulation analysis between the resting and stimulated samples I used DESeq2 (v1.14.1) (Love et al., 2014). To find regions that were changed upon stimulation, I compared the stimulated samples to the resting state, using the blood processing batch as a covariate, and used Benjamini-Hochberg controlled false detection range (FDR) of 10% (Benjamini and Hochberg, 1995) and an absolute fold-change $\geq$ 2. I overlapped the differentially defined regions of one assay with the total regions of another assay using ChIPpeakAnno's findOverlapsOfPeaks and plotted their distribution around the TSS using the function binOverFeature from the same package (Zhu et al., 2010b).

3.2.8 RNA-seq and initial processing

Sorted regulatory T cells were placed in 0.7ml of QIAzol Lysis Reagent (Qiagen) and stored at -80°C. RNA was isolated following the Qiagen miRNeasy Micro kit manufacturer's protocol. RNA was quantitated using Bioanalyzer (Agilent Technologies, USA). Libraries were prepared using Illumina TruSeq index tags and sequenced on the Illumina HiSeq 2500 platform using V4 chemistry and standard 75 bp paired-end. Twenty samples were multiplexed at equimolar concentrations and sequenced across five lanes, to yield on average 72.5 million reads per sample.

Reads were aligned to the GRCh38 human reference genome using STAR (Dobin et al., 2013) and the Ensembl reference transcriptome (version 87). Gene counts was performed using featureCounts tools from the subread package v1.5.1 (Liao et al., 2014) and only

assigned reads were used for further processing (58.5-70.3% of reads were assigned). Sequence reads were also aligned using Salmon (Patro et al., 2017). I used the options –seqBias –gcBias to adjust for sample specific fragment bias and GC bias accordingly.

3.2.9 Differential gene expression analysis

To identify differentially expressed genes (DEGs), I used DESeq2 (R version 3.3.3, DESeq2 version 1.14.1) (Love et al., 2014). For downstream analysis I only considered 17,225 genes that had at least 20 copies in at least 3 samples. I compared resting to stimulated cells using Benjamini-Hochberg controlled FDR (Benjamini and Hochberg, 1995) of 5% and an absolute fold-change ≥ 2 to find activation induced genes.

For the clustering of DEGs, raw counts were normalised using DESeq2 rlog function. To control for the different batches in which we processed the blood, which accounted for 14.6% of the observed variability, I performed batch correction using the combat algorithm as implemented by the sva package (Leek et al., 2012). Finally, the values were quantile normalised. I did hierarchical clustering on all DEGs using the euclidean distance between samples and the optimal number of clusters was defined using dynamicTreeCut (Langfelder et al., 2008) (minimum cluster size was set to 50). I used g:Profiler (Reimand et al., 2016) R package to identify the top three Reactome pathways (Croft et al., 2014) enriched in each cluster.

3.2.10 Primary transcript usage

I used tximport to obtain the transcript level abundances, and only used the transcripts that had at least one transcript per million (TPM) in at least four samples and whose gene expression passed the filtering threshold used with the STAR output. I calculated transcript ratios (tr) based on the total expression per gene. The final table was composed of 67,239 transcripts, which corresponded to 12,100 genes. I calculated the difference in ratios for each transcript between the resting and stimulated state and derived the mean across the ten donors. Finally, I extracted the extreme 1% tail of each side of the distribution and required a gene to have a transcript in both tails in order to be included.

3.2.11 Binding expression target analysis

I identified regions of the genome that changed upon stimulation as described in Chapter 2. Based on the distribution of the peaks around the TSS, I used a window of 3 kbp around the TSS for active promoters and 100 kbp for active enhancers. If active enhancers were less than 3 kbp apart they were considered to be a single enhancer.

3.2.12 LRRC32 flow cytometry validation

The level of expression of LRRC32 was quantified in ten donors in resting cells and cells stimulated with anti-CD3/anti-CD28 beads (1 bead:4 cells). CD4+ cells were isolated and cultured for 18 hours with or without stimulant. After 18 hours of stimulation cells were observed by flow cytometry using a directly conjugated antibodies CD4+ (RP4-T4)-Alexa700 (eBioscience), GARP (G14D4)-PE (eBioscience), CD25 (B96C)-PeCy7 (eBioscience), CD69 (FN50)-FITC (Biolegend), CD28 (CD28.2)-APC (eBiosience), CD45RA (HI100)-BV785 (Biolegend). Intracellular staining for *FOXP3* (206D)-BV421 (Biolegend) was done using the eBioscience FOXP3 staining buffers.

3.2.13 Differential intron excision analysis

I used LeafCutter's annotation free approach (Li et al., 2018) to identify 26,342 clusters of intron excision events corresponding to 96,809 alternatively excised introns. I limited the analysis to the genes that had been detected by gene counts. In each sample, I counted the number of reads supporting each intron excision event in a cluster as well as the total number of reads in a cluster. I removed clusters with less than 2 samples with coverage > 20, clusters that had more than 10 introns, clusters with less than 2 introns used in more than 4 samples. I identified a total of 6,861 clusters corresponding to 4,307 genes that were differentially spliced upon stimulation (using Benjamini-Hochberg controlled FDR (Benjamini and Hochberg, 1995) of 1% and an absolute fold-change ≥ 2).

To calculate the conservation score across the junctions I used the conservation scoring PhyloP phylogenetic values obtained for multiple alignments of 99 vertebrate genomes to the human genome (Pollard et al., 2010). I used both junction ends for the novel annotated pairs and the unanchored junction, but only the 3' end for the cryptic 3' sites and the 5' end for the cryptic 5' sites. I compared the result to a randomly sampled

distribution of base pairs of the same size.

To provide functional support for some of the novel annotated pairs and cryptic 5' junctions, I overlapped the 5' of these clusters with differentially active promoters using bedtools. I only kept promoters that had a peaks value $\geq$ 100 and did not overlap an annotated TSS.

3.3 Results

3.3.1 Optimising ChIPmentation-seq in Tregs

To compare the performance of ChIP and ChM I assayed in parallel four different marks (H3K4me3, H3K4me1, H3K27ac and H3K27me3) using material from the same cell culture for Jurkat cells and the same donors for Tregs. I observed a higher percentage of mapped reads and non-PCR duplicates across all ChM experiments. After calling ChIP and ChM-seq peaks, i.e. genomic regions enriched for read pile-ups compared to the background read distribution, I observed that ChM displayed more peaks regardless of the antibody used and showed a higher fraction of reads in peaks (FRiP), which resulted in a higher fold-change enrichment across the called peaks (Figure 3.1 A and B). This suggested that ChM-seq is more sensitive than ChIP-seq.

In order to assess how the different histone marks and cell types related to each other, I calculated the Pearson correlation coefficient (PCC) between all the assays performed. In this analysis I included ATAC assays to identify potential Tn5 biases in the ChM-seq assay. I observed a high PCC between assays performed with the same antibody and between ATAC-seq from different donors (Figure 3.1 C). Hierarchical clustering of the samples showed two distinct clusters, the first one was composed of H3K27me3, a repressive mark, while the second one encompassed the rest of the assays, which are all active marks. Within the active marks I identified some substructure depending on the assay performed. Therefore, samples separate according to chromatin activity.

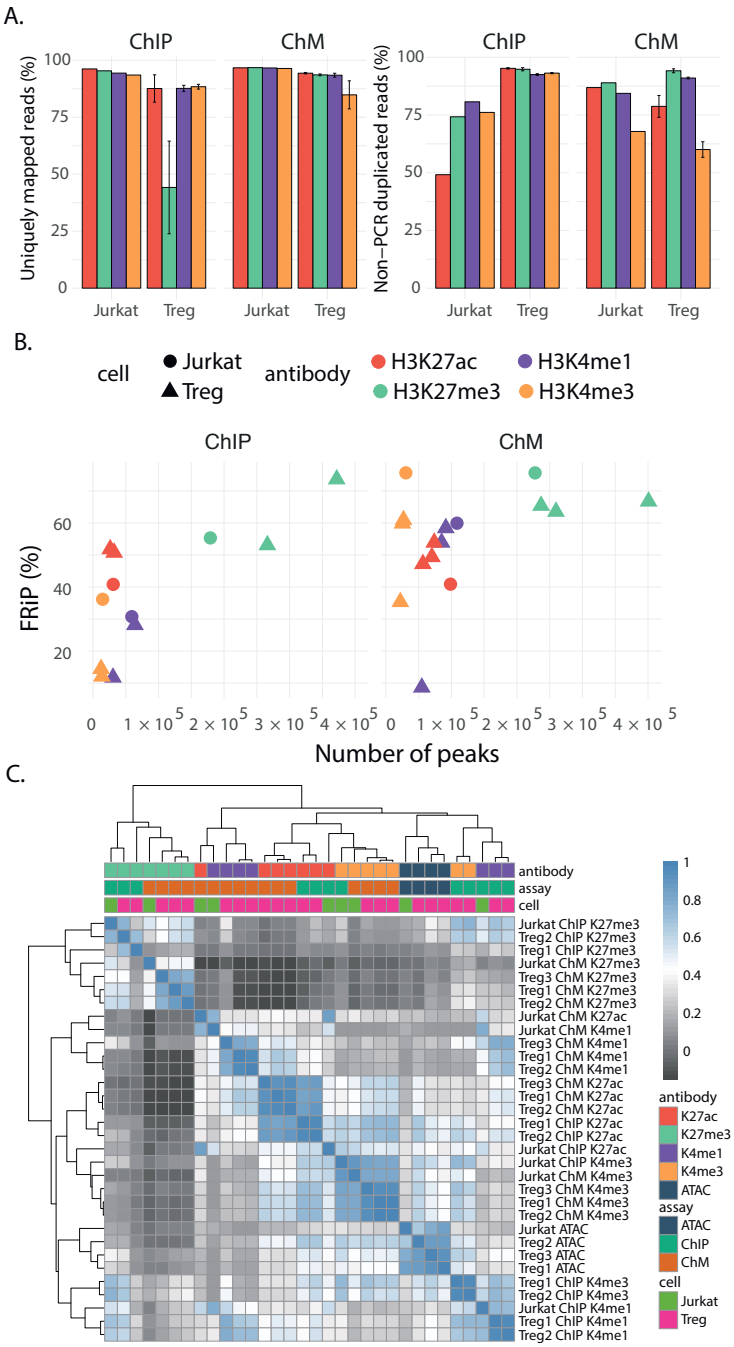

Figure 3.1: **Qualitative control shows that ChM-seq is a more sensitive method than ChIP-seq.** **A.** Percentage of uniquely mapped reads (left panel) and non-PCR duplicate fragments (right panel) for ChIP and ChM quantified for the input and the four antibodies against histone modifications assayed in the Jurkat cell line and in T regulatory cells. **B.** Fraction of reads in peaks (FRiP) and number of peaks called from ChIP and ChM data for all sequenced libraries. **C.** Genome-wide Pearson correlation heatmap for all histone marks and cell types. Correlation was derived based on reads in 10,000 bp bins across the whole genome.

In order to increase the sensitivity of peak calling, I recalled peaks for each histone mark by merging the files from the different T regulatory cells donors. This was possible because of the high correlation observed between the same histone marks within an individual method (ChM or ChIP-seq) across all replicates. For H3K27ac, H3K4me1 and H3K4me3 antibodies, ChM-seq captured over 95% of the peaks called using ChIP-seq (Figure 3.2 A). The overlap was smaller for H3K27me3 (70%), which marks broader regions of the genome. The mean peak length for this mark was 3.85 kbp, compared to 2.57 kbp for H3K27ac, 1.39 kbp for H3K4me1 and 1.55 kbp for H3K4me3. Since ChM-seq resulted in a higher number of method specific peaks, I examined the characteristics of these regions in relation to all peaks. Method specific peaks, that is peaks called using only one method, had similar p-value and peak length distribution as the total peak-set. However, the largest peaks with the smallest q-values were shared between the two methods (Figure 3.2 B). Therefore the peaks identified by ChM-seq are very similar to the ones found by ChIP-seq.

I annotated the peaks based on the nearest transcription start site (TSS) of a gene across the genome. As expected, H3K4me3 and H3K27ac were enriched in gene promoters, H3K4me1 and H3K27ac in enhancers and unannotated regions of the genome were marked by H3K27me3 (Figure 3.2 C). Because ChM-seq resulted on average in an increased number of peaks compared to ChIP-seq, I further assessed if the distribution of method specific peaks was skewed towards specific genomic regions. For H3K4me3, ChIP-seq specific peaks (37 in total) were mostly located away from the gene body, while ChM-seq specific peaks (15,798 in total) localised to promoter regions (up to 10 kbp upstream from the TSS) and introns. To further validate that ChM specific peaks mapped to biologically relevant gene regulatory regions, I tested if the peaks colocalised with chromatin contacts mapped with promoter capture Hi-C (PCHi-C) in CD4+ cells (Javierre et al., 2016). Both ChIP-seq and ChM-seq active marks were enriched in PCHi-C chromatin contacts, while the H3K27me3 silencing mark was depleted from those regions (Figure 3.2 D). However, when repeating this on method specific peaks, only ChM-seq peaks remained significantly colocalised with PCHi-C mapped regions. This suggested that ChM-seq is a more sensitive and specific method than ChIP-seq since it identified a higher number of peaks in relevant regulatory regions.

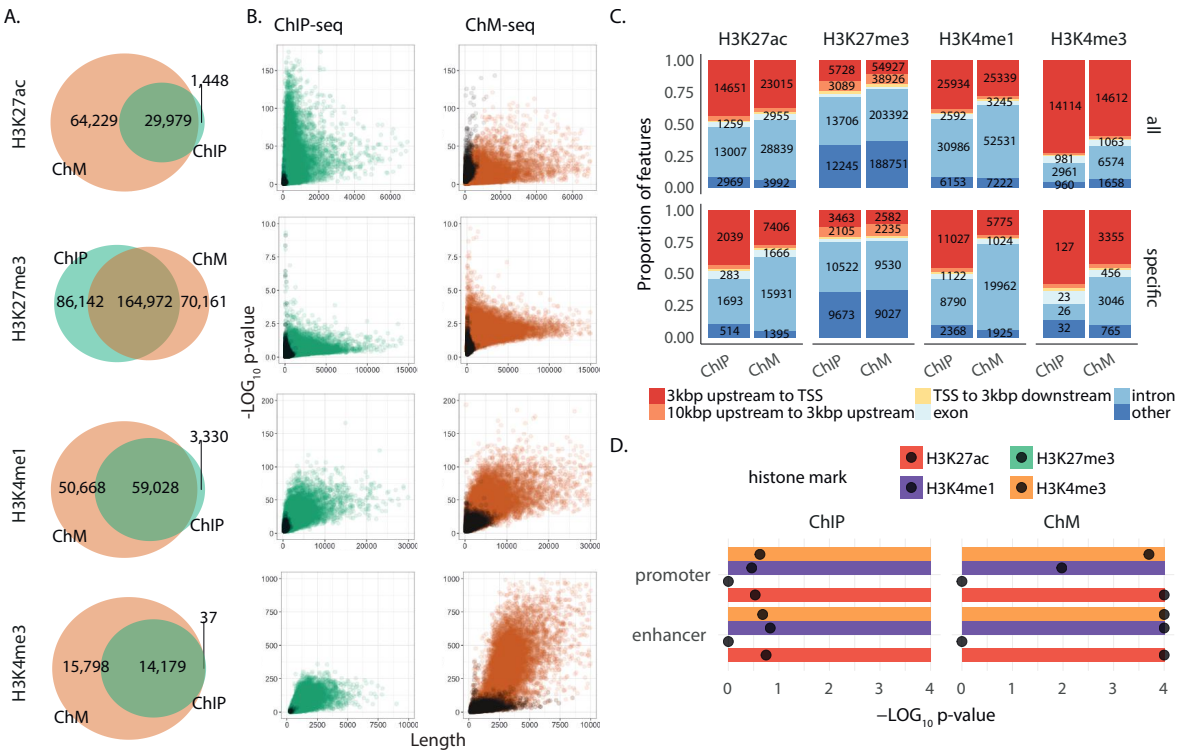

Figure 3.2: **Global comparison of ChIP-seq and ChM-seq in Tregs highlights increased sensitivity of ChM-seq. A.** Overlap between ChIP-seq and ChM-seq peak-sets. **B.** Peak length and significance (darker colors correspond to the results of the analysis using method specific peaks). **C.** Peak annotation for ChIP and ChM-seq peaks based on the distance to the transcription start site of the closest gene. **D.** Significance of enrichment of ChIP and ChM-seq peaks in PCHi-C regions mapped in CD4+ T cells (black dots correspond to the results of the analysis using method specific peaks).

3.3.2 Identification of the optimal conditions for Treg stimulation

Having validated the applicability of ChM in producing robust and reproducible results in primary Tregs, I applied the protocol to profile the chromatin landscape of this rare cell population upon stimulation. I used ChM-seq for the H3K27ac and the H3K4me3 antibodies, as well as ATAC-seq and RNA-seq to generate the first comprehensive map of stimulation induced changes in Tregs isolated from ten individuals.

I isolated Tregs by fluorescently activated cell sorting (FACS), selecting for CD4+CD25high CD127low. Cells were cultured for 18 hours with or without PMA and ionomycin to provide the stimulation. I initially stimulated Tregs with beads coated with antibodies against CD3 and CD28. However, I found that the interaction between cells and beads was extremely strong and washing cells off the beads resulted in a significant loss of

cellular material. I reasoned this loss could be reduced by substituting beads with PMA/ionomycin. Therefore, I tested how different the gene expression changes induced by PMA/ionomycin and anti-CD3/anti-CD28 beads were. To assess this, I used CD4+ conventional T cells (CD4+CD25lowCD127high) isolated from two donors. I observed 93% correlation of log2 fold change in gene expression between resting and stimulated cells in each of the two conditions (Figure 3.3 A). This indicated that the differences between the two stimuli were negligible and that PMA/ionomycin could be reliably used as a relevant stimulus for T cell activation. To ensure that a large proportion of Tregs was successfully stimulated in culture, I quantified cell activation through CD69 staining, a cell surface marker of early activation (Figure 3.3 B). On average, 100% of cells responded to activation. Additionally, to measure the purity of the Treg isolation protocol I performed intracellular staining for FoxP3 (mean FoxP3+ cells = 83.5%) (Figure 3.3 C). In conclusion, using flow cytometry I confirmed both the successful isolation of Tregs and that cells were stimulated, therefore allowing me to proceed with genomic protocols.

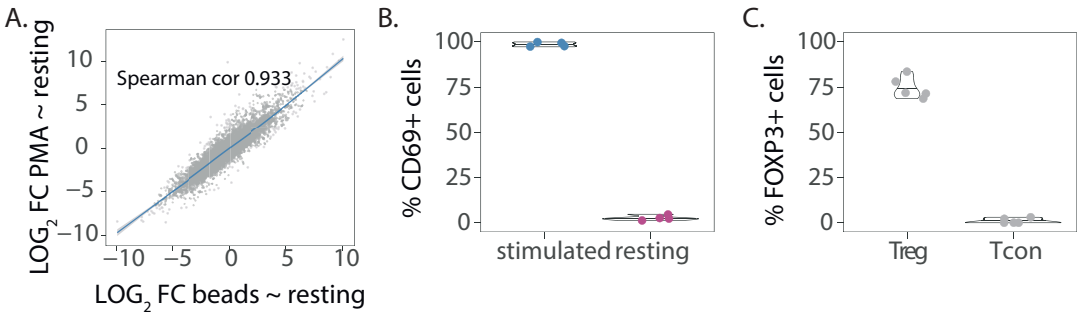

Figure 3.3: **Optimisation of Treg isolation and stimulation by PMA/ionomycin. A.** Correlation of log2 fold changes in gene expression between resting and anti-CD3/CD28 beads stimulated (x-axis) and resting and PMA stimulated conventional T cells (y-axis). The Spearman correlation was calculated using all genes. **B.** Percentage of Treg cells expressing CD69 activation marker after 18 hours of cell culture either with PMA/ionomycin or without, as measured by flow cytometry. **C.** Percentage of cells expressing FoxP3 as measured by flow cytometry.

3.3.3 Stimulation induces specific changes in promoter and enhancer activity

I firstly looked at the epigenetic data from ten donors to quantify stimulation induced changes. I used MACS2 to call peaks per donor for each of the three assays individually. Globally, upon stimulation Tregs displayed a higher number of ATAC and H3K27ac peaks (paired t-test p-value < 0.05), while no difference was observed for H3K4me3

(Figure 3.4 A). This suggested that stimulation increased chromatin accessibility and upregulated active regulatory elements, but did not result in the formation of new promoters. Next, I defined a union set of peaks in each assay from stimulated and resting cells by merging the peaks across different samples. I counted the number of reads that fell within these genomic coordinates per sample. Using these tables, I calculated the Euclidean distance between different samples per assay and created a heatmap (Figure 3.4 B). I observed that resting and stimulated samples clustered separately. To further understand if the observed chromatin remodelling was driven by increased promoter or enhancer activity, I overlapped the peaks marked by the three assays and defined four groups of regulatory elements: active promoters (AP; H3K4me3+H3K27ac), non-active promoters (NAP; H3K4me3-H3K27ac), active enhancers (AE; H3K27ac-H3K4me3) and open chromatin (OC; ATAC-H3K27ac-H3K4me3) (Figure 3.4 C).

In order to uncover the genetic mechanisms of Treg activation, I used the three assays to compare the gene regulatory landscape of resting and stimulated cells to define the differential peaks (FDR $\leq$ 0.1 and fold-change $\geq$ 2). For each assay, I then overlapped the set of differential regulatory regions with the full peak-sets from the remaining two assays to define stimulation induced changes in active promoters, non-active promoters, active enhancers and open chromatin. I considered a regulatory element to undergo a stimulation induced change if at least one chromatin mark peak showed a significant differential change (Figure 3.5 A). The greatest number of changes occurred in active regions of the genome, where more than 10,000 regions were altered upon stimulation (Figure 3.5 B). The open chromatin and the non-active promoters changed very little with stimulation (less than 2,000 regions). A comparable number of active enhancers and promoters were upregulated (approx. 3,200). I examined the distribution of the differentially up and down regulated regions with respect to the TSS and compared it with the global distribution of H3K27ac ChM-seq for active enhancers, H3K4me3 ChM-seq for active and non-active promoters, and ATAC-seq of open chromatin. I observed a depletion of differentially active enhancers near the TSS, while the distribution of differentially active promoters remained largely unchanged and concentrated around the TSS. Therefore stimulation induced changes in gene expression via the regulation of distal enhancers. Interestingly, there was a slight shift towards the gene body in upregulated, but not downregulated, active promoters compared to the background distribution of all H3K4me3 ChM-seq peaks (Figure 3.5 C). The shift towards the gene

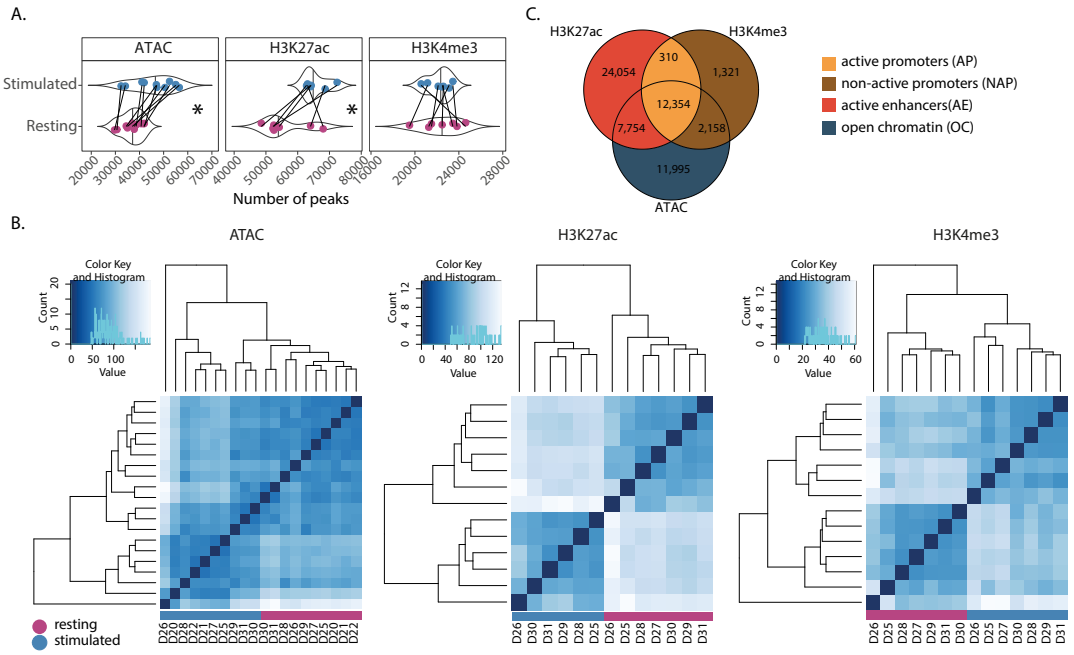

Figure 3.4: **Stimulation induces changes in H3K4 tri-methylation, H3K27 acetylation and chromatin accessibility in Tregs. A.** Distribution of the peak number detected across samples in resting and stimulated regulatory T cells for the three assayed chromatin marks. Asterisk indicates significantly different numbers of peaks detected between the two cell states, ATAC (p-value = 2.95×10^{-3}), H3K27ac ChM (p-value = 4.82×10^{-2}) and H3K4me3 ChM (p-value = 0.83). **B.** Heatmap of Euclidean distance between the samples acquired from each assay. **C.** Overlap of peaks between the ATAC, H3K27ac and H3K4me3 assays. A peak from an individual assay can overlap more than one peak from another assay, I have written in the Venn diagram the smallest number to not inflate the values. For example, if one H3K27ac ChM-seq peak overlaps two H3K4me3 ChM-seq peaks and three ATAC-seq peak, I would only count this as a single overlap.

body in active promoters could suggest an additional gene regulation by promoter switching, and the induction of expression of other isoforms.

3.3.4 Gene expression signatures of activated regulatory T cells

In order to better understand the association between changes in chromatin marks and gene expression regulation I integrated RNA-seq data from the same ten individuals with the chromatin modification and accessibility results. I first assessed global gene expression changes triggered by stimulation, before combining the RNA-seq differential expression results with the changes in chromatin marks and accessibility. I performed differential gene expression analysis (fold-change ≥ 2 and a false discovery rate (FDR) ≤ 0.05) between resting and activated Treg cells and detected over 2,300 upregulated and 3,000 downregulated genes upon stimulation. Among the upregulated genes I

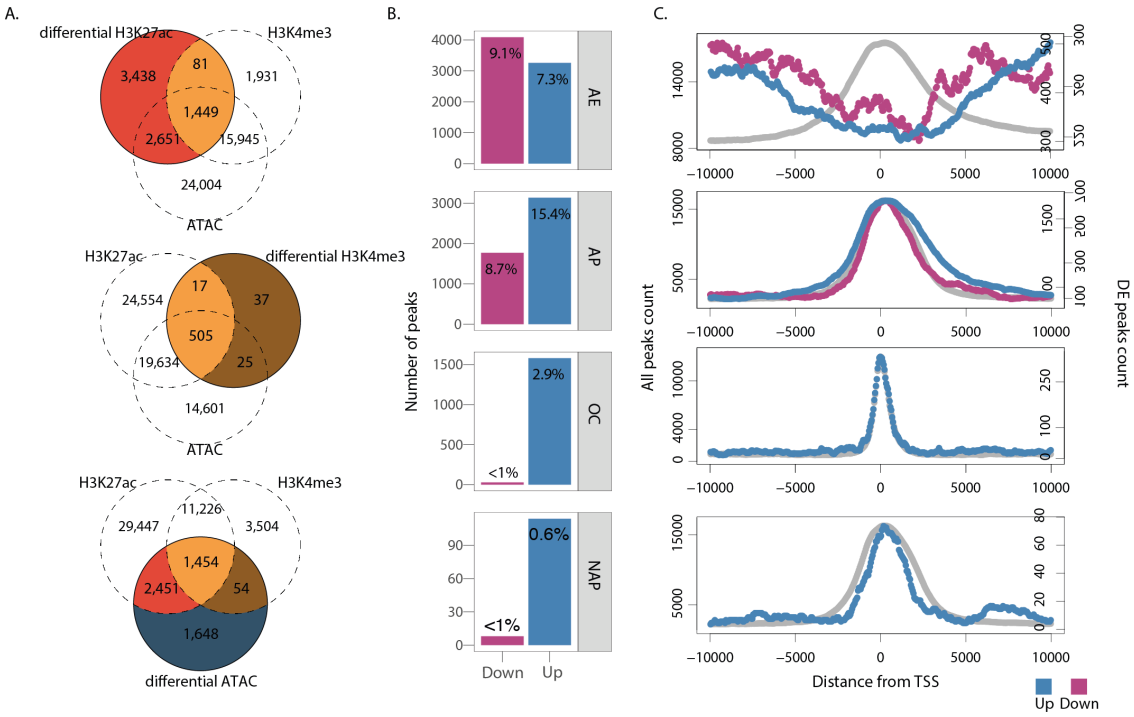

Figure 3.5: **Differential epigenetic regulation upon stimulation in Tregs occurs primarily in active regions of the genome. A.** Number of peaks with significant stimulation induced changes in individual assays profiling chromatin activity (H3K27ac, H3K4me and ATAC). The Venn diagrams illustrate the overlap between the differential peaks of each assay and the total number of peaks of the other two assays. **B.** Number of differentially regulated peaks per defined region (active promoters (AP), non-active promoters (NAP), active enhancers (AE) and open chromatin (OC)) across the three assays. **C.** Distribution of the peaks relative to the TSS. Differentially upregulated (blue) and downregulated (grey) regions are plotted against the distribution of all peaks (black) using H3K27ac ChM-seq (for active enhancers), H3K4me3 ChM-seq (for active and non-active promoters) and ATAC-seq (for open chromatin).

found multiple classical T cell activation markers (e.g. *CD69*, *CD200*, *CTLA4*, *IL2RA* and *TNFRSF9*), transcription factors (TFs) related to the immune response (e.g. *EGR1*, *EGR2*, *EGR3*, *NR4A2*, *FOSL1*, *IRF4*, *IRF8* and *TBX21*) and cytokines (*IL1A*, *IL2*, *IL3*, *IL6*, *IL10*, *IL17A*, *IL21*, *IL23A*, *EBI3* and *IFNG*) (Figure 3.6 A). Among the down regulated genes I found genes important for Treg cell identity, such as *IKZF4* (Figure 3.6 A).

To identify clusters of co-regulated genes, I grouped genes that followed similar expression trajectories in response to stimulation. I identified four clusters of co-upregulated genes and nine clusters of co-downregulated genes (Figure 3.6 B). Interestingly, cluster 1, which contained the majority of upregulated genes (72%) and cluster 2, which contained the majority of downregulated genes (40%), were characterised by limited variability compared to the remaining clusters. This meant that the majority of up and downregulated genes were not variable across donors, and there is a uniform response

to stimulus. On the other hand, the median gene dispersion in clusters 4, 9, 10, 12 and 13 was more than five times greater than the gene dispersion of non-differentially expressed genes (Figure 3.6 C). These clusters contain genes whose expression in response to stimulation varies across donors and might be influenced by the genetic background or are artifacts, in the case of the smallest clusters 12 and 13.

To assess if the co-regulated genes were enriched for specific pathways I used the Reactome database (Croft et al., 2014). While all upregulated clusters were significantly enriched for relevant pathways, only three of the nine downregulated clusters showed significant enrichments. This suggests that downregulated genes are more randomly distributed across different pathways. Cluster 1 was highly enriched for processes related to G2M progression (p-value = $9.8x10^{-20}$), RNA metabolism (p-value = $7.7x10^{-30}$) and NF-κB signaling (p-value=$5x10^{-20}$), indicating that cells were successfully stimulated, initiated gene expression programmes to trigger cell proliferation and invoke an immune response. This analysis provided insights into the dynamic cellular processes in Tregs upon stimulation. For example, I observed that immune pathways, including cytokine production were generally upregulated (cluster 4 and 8), while at the same time they were downregulated (clusters 2 and 6), mostly driven by the interferon signalling response (Figure 3.7).

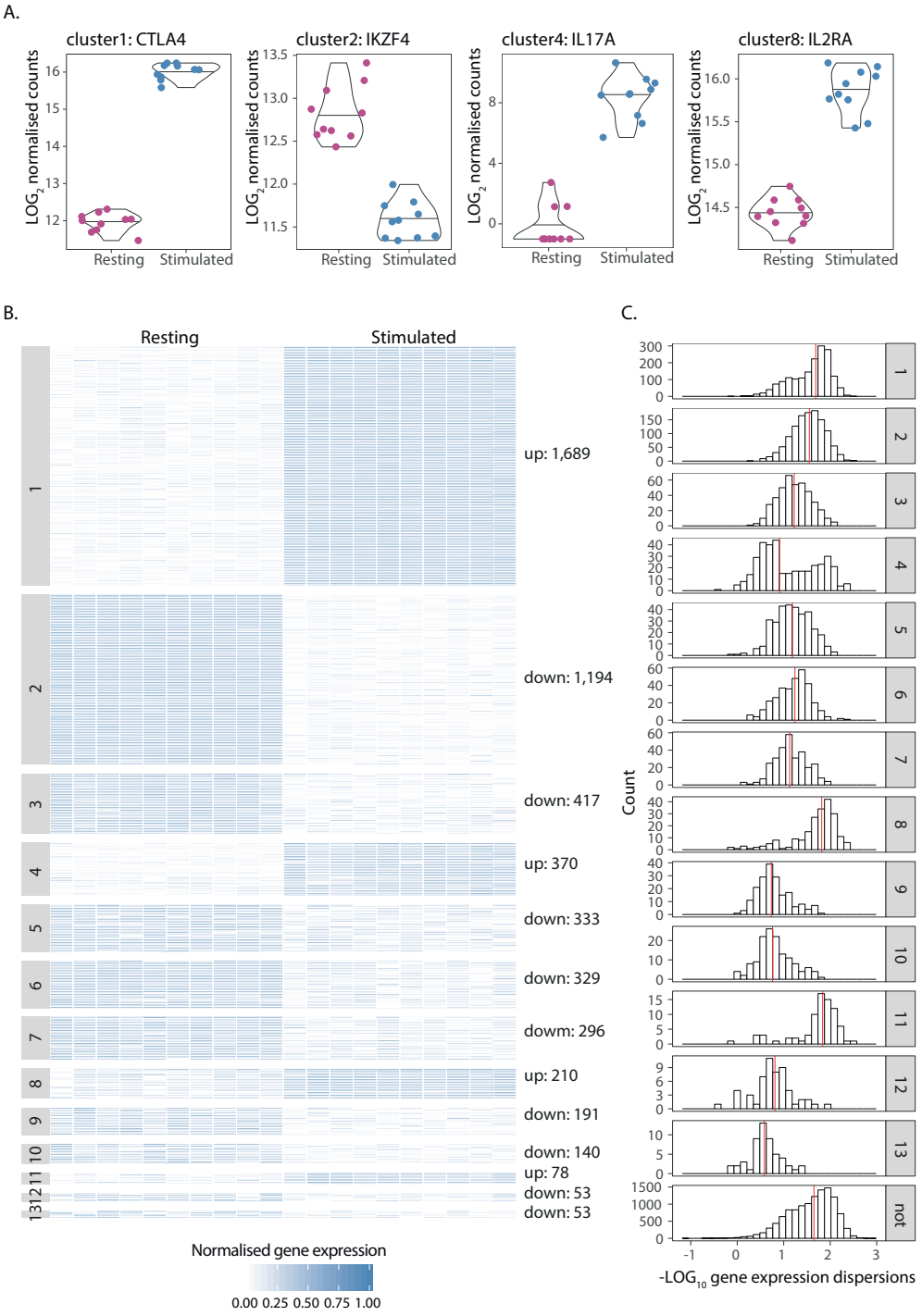

Figure 3.6: **Differential gene expression analysis between resting and stimulated Tregs. A.** Gene expression levels for selected differentially expressed genes of resting and stimulated Tregs. **B.** Heatmap of differentially expressed genes that follow a similar co-expression dynamics and form 13 co-regulated gene clusters. On the x-axis are the donors, on the y-axis is a random selection of a 10% of the genes present in each cluster and the color corresponds to the quantile normalised gene expression value. **C.** Distribution of gene expression dispersion estimates calculated per batch and condition, and stratified per cluster. In red is shown the median distribution of each cluster.

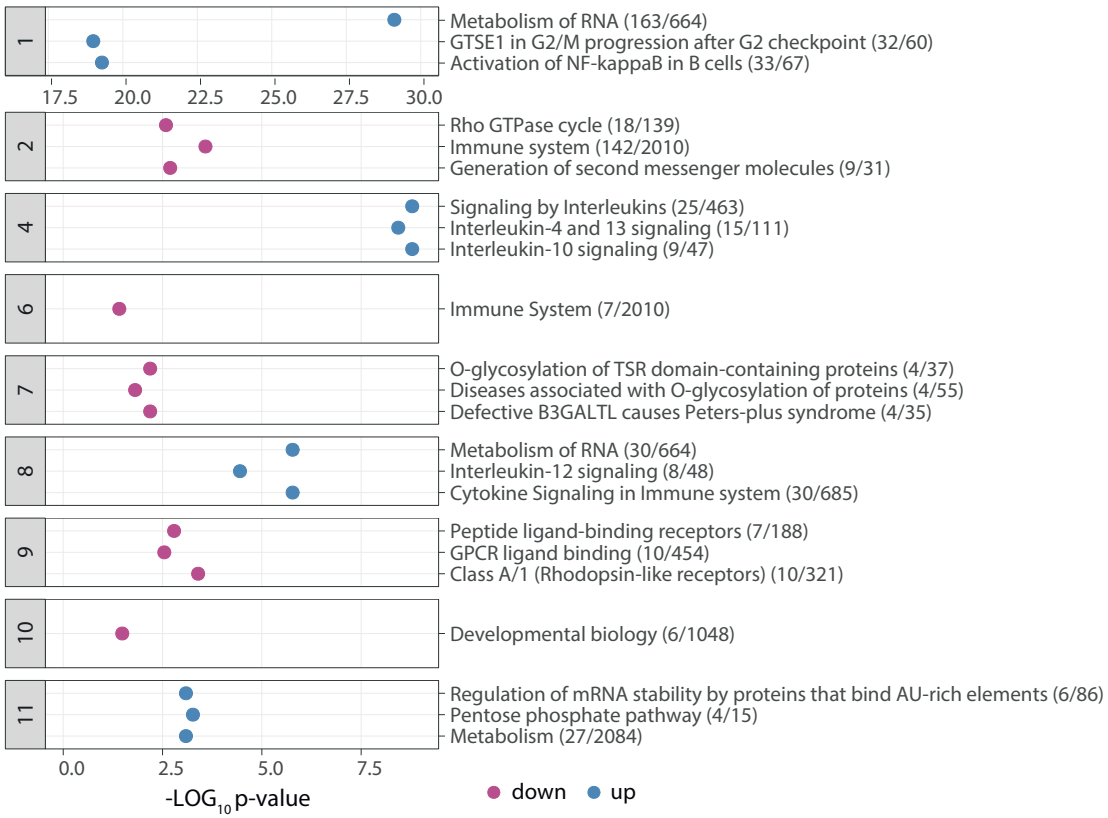

Figure 3.7: **Pathways enriched in differentially regulated clusters.** Top three results of the pathway enrichment analysis per each cluster of co-regulated genes using the Reactome database. Numbers in brackets correspond to the genes contributing to the observed enrichment.

The above analyses collapses sequence read counts to gene counts. I additionally wanted to investigate to what extend Treg stimulation would have an impact on alternative transcripts. For that, I quantified the relative transcript ratios and found that the majority of genes expressed more than one transcript (82.7%) (Figure 3.8 A). However, most of the genes did not alter their relative transcript ratios, therefore I further focussed on a subset of genes that switched their primary transcript upon stimulation, (Figure 3.8 B). Of the 9,727 genes with more than one transcript, 274 showed different dominant splice isoforms in resting cells and in stimulated cells (Figure 3.8 C).

In order to uncover regulatory processes unrelated to the total changes in transcript levels, I focussed on the genes that were not differentially expressed but switched splicing isoforms between the two cell states. I identified five genes which encoded for TFs; *NFATC1*, *ELF4*, *RUNX1*, *HLF* and *YY1* (Figure 3.9), all of which showed alternative promoter usage upon stimulation. I focussed on TFs because changes in parts of the gene body

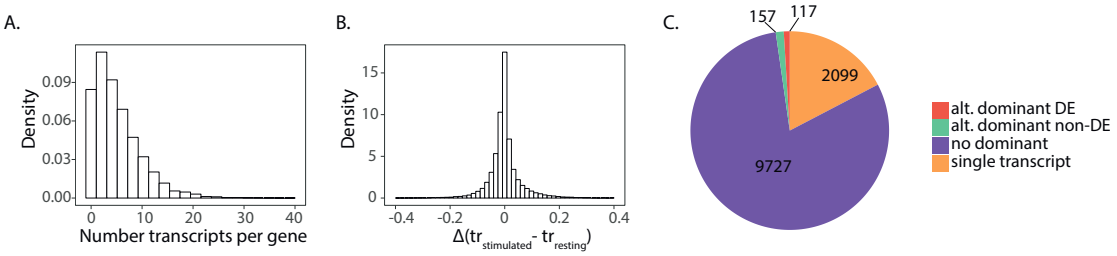

Figure 3.8: **Five transcription factors have a different transcript dominant upon stimulation in Tregs. A.** Distribution of the number of transcripts per genes. **B.** Distribution of the delta between the transcript ratio of the stimulated sample and the resting sample. **C.** Pie chart illustration of the number of genes that encode for a single transcript, that either have the same transcript being dominant in resting and stimulated cells, or that have an alternative transcript being dominant and whether it is differentially expressed or not.

which encode for functional domains can have direct implications in the downstream processes they regulate. With the exception of *RUNX1* and *HLF*, the other TFs in this list have been associated with Treg function, even if the exact role of each alternative transcript remains unknown. In resting Treg cells, the main *NFATC1* transcripts expressed were ENST00000318065 (10 coding exons) and ENST00000329101 (10 coding exons). Conversely, upon stimulation the main transcript was ENST00000253506 (10 coding exons), which uses an alternative upstream promoter (Figure 3.9). *ELF4* main transcript in resting cells was ENST00000335997 (8 coding exons), while in stimulated cells the main transcript was ENST00000308167 (8 coding exons) (Figure 3.9). *RUNX1* has many isoforms, but the most highly expressed one in resting cells was ENST00000437180 (8 coding exons) and in stimulated cells it was the shorter transcript ENST00000344691 (6 coding exons), which lacks the first two exons of ENST00000437180. *HLF* had three transcripts with detectable levels in Tregs, ENST00000226067 (4 coding exons) was the main transcript in resting cells, while upon stimulation this switched to the other two transcripts, ENST00000573945 (3 coding exons) and ENST00000575345 (3 coding exons), which lacked the first exon. Finally, the main transcript for *YY1* in resting cells was ENST00000262238 (5 coding exons), which was also highly expressed in stimulated cells, and additionally the expression of ENST00000554804 (4 coding exons) isoform increased. ENST00000554804 uses an alternative promoter and lacks the 3' UTR region of the gene (Figure 3.9). Alternative transcripts with different exon composition could explain differences in the genes function upon stimulation in Tregs. Indeed, the main

YY1 transcript expressed by resting Tregs results in the coding protein P25490, while ENST00000554804 results in the coding protein H0YJV7.

3.3.5 Integration of epigenetic and gene expression data points towards the regulation of immune genes

I next sought to determine the chromatin changes that contribute towards the regulation of transcriptional programs in response to cell stimulation. First, to identify chromatin regions that could control gene expression, I assessed if differential active promoters and enhancers correlated with differential gene expression (Wang et al., 2013b). Using a window of 3 kbp around the TSS I found that active promoters induced upon stimulation were predictive of 16% of upregulated genes (p-value = $1.2x10^{-9}$) while the downregulation of promoters predicted 9% of downregulated genes (p-value = $2.7x10^{-6}$). On the other hand, using a window of 100 kbp around the TSS I observed that upregulated active enhancers were predictive of 27% of upregulated genes (p-value = $3.9x10^{-11}$), while downregulated active enhancers were predictive of 25% of downregulated genes (p-value = $1.4x10^{-11}$) (Figure 3.10 A). Open chromatin regions and non-active promoters were not predictive of changes in gene expression (p-value > 0.05). The majority of genes (96%) were assigned a single differentially active promoter, with the exception of a small set of 27 genes that had two or three promoters. However many genes (37%) had more than one enhancer, highlighting the interplay between multiple enhancers in complex gene regulatory processes (Figure 3.10 B). Amongst the genes with four or more enhancers I identified interleukin receptors (*IL1RN*, *IL23R*, *IL12RB2* and *IL36RN*) and TFs (*REL*, *EGR2* and *IRF8*). These genes play an important role in cell signalling and transcriptional responses, which could explain why multiple enhancer interactions would regulate their expression. The median distance between any two enhancers regulating the same gene was 20.4 kbp, indicating potential enhancer clusters (Figure 3.10 C). Finally, I was interested to see whether any of the chromatin marks were more informative than others in predicting gene expression changes. I found that the different assays equally contributed to the predictability of gene expression (Figure 3.10 D), highlighting the importance of performing them in tandem.

Next, I assessed if genes with predictive regulatory elements were enriched for any pathways. I found that both genes with upregulated enhancers (763 genes) and pro-

Figure 3.9: **Transcription factor genes employing alternative promoters upon stimulation. A.** Transcript expression levels of resting and stimulated Tregs for selected transcription factors that are not differentially expressed but show an alternative transcript being dominant between the resting and stimulated conditions. **B.** RNA-seq read pile-ups across the intron and exons of the genes in resting and stimulated cells, indicating different usage of transcript isoforms.

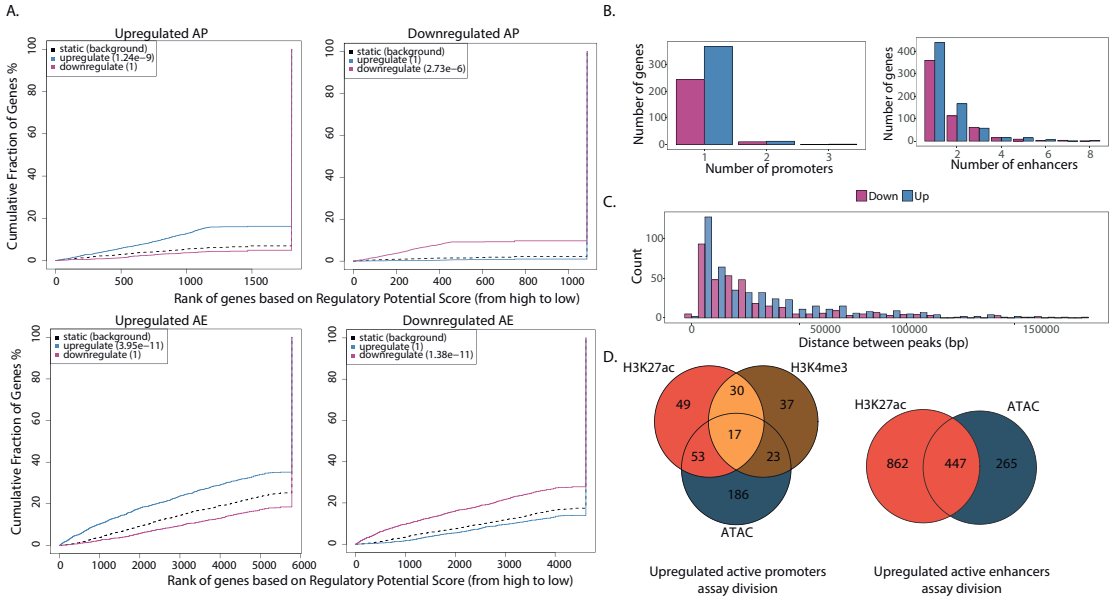

Figure 3.10: **Differentially regulated active promoters and enhancers provide insights into mechanisms of gene expression regulation upon stimulation. A.** Predictive function of the differentially active promoters (AP) and differentially active enhancers (AE) to regulate gene expression. **B.** Number of differential active promoters and enhancers predicted to regulate gene expression. **C.** Distance between the enhancers for the genes with multiple predictive differentially active enhancers. **D.** Regulatory elements predictive of gene expression changes split per chromatin mark that was differentially regulated.

moters (383 genes) were related to the immune response (Figure 3.11). Conversely, no enrichment was found for the downregulated genes with associated active elements. The genes driving the enrichment in immune related pathways included key Treg genes induced upon activation. For example, *TNFSF9*, a gene that encodes for the 4-1BB ligand that binds the TNFRSF9 co-stimulatory molecule that is upregulated upon stimulation in a FoxP3 dependent manner (Marson et al., 2007) (Figure 3.12 A) and *LRRC32*, a gene that encodes for GARP, and is upregulated on the surface of activated Tregs (Wang et al., 2009) (Figure 3.12 B). To validate if the observed gene expression changes were recapitulated at the protein level, I performed flow cytometry analysis to measure the protein expression of GARP. Tregs from ten donors were stimulated overnight with anti-CD3/anti-CD28 beads and stained with a fluorophore labeled antibody against GARP 16 hours after activation. Indeed, the upregulation of GARP upon stimulation was also significant at the protein level (p-value < 10^{-5}) (Figure 3.12 B).

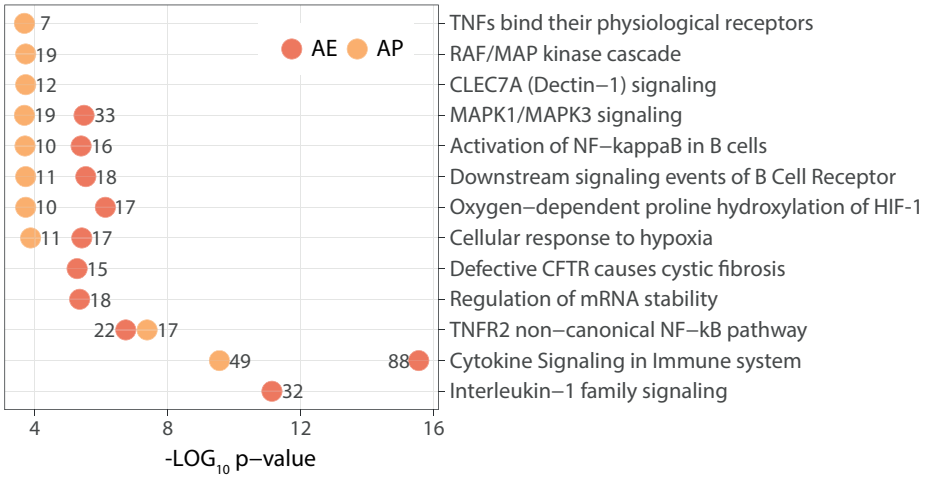

Figure 3.11: **Pathway enrichment analysis for the upregulated genes that have assigned predictive regulatory elements in the chromatin.** The numbers correspond to the number of genes contributing towards the observed enrichment.

Interestingly, both genes, *TNFSF9* and *LRRC32*, were characterised by changes in the levels of deposition of H3K4 trimethylation only, while the levels of H3K27 acetylation and chromatin accessibility remained unchanged. This indicates a mechanism in which the promoter is primed for activity by high levels of H3K27ac and the gene expression is regulated by the modulation of H3K4me3 levels. An additional 37 genes followed a similar expression regulation pattern (Table 3.2), including 8 TFs (e.g. *IRF1*, *NFKBIA*, *EGR1*, *TBX21*, *FOSL2* and *TRAF4*). Therefore some active promoters of genes that play an important role in Treg functions, were upregulated upon activation of Tregs.

Finally, I looked at the enrichment of TF binding sites (TFBSs) in the differential active promoters and enhancers with concordant effects with gene expression (Figure 3.13 A). I identified six enriched motifs in upregulated active enhancers, including the SMAD family of TFs, which are required for TGF-β signalling, an essential Treg cytokine (Figure 3.13 B). Downregulated active enhancers were enriched for four motives, including FoxP3 (Figure 3.13 B), the hallmark TF for Tregs. A different set of TFs was enriched in active promoters. Here, I identified five enriched motifs in upregulated promoters, including IRF4 and RARA (Figure 3.13 B). The only enriched motif in downregulated promoters was TP53 (Figure 3.13 B). Interestingly, upon stimulation I observed the enrichment of TFs that are able to polarise cells towards different fates. The SMAD family of TFs positively regulates the generation of Th17 cells from Tregs, RARA is known to inhibit Th17 fate while promoting Th1 responses (Brown et al., 2015) and IRF4 regulates Th2

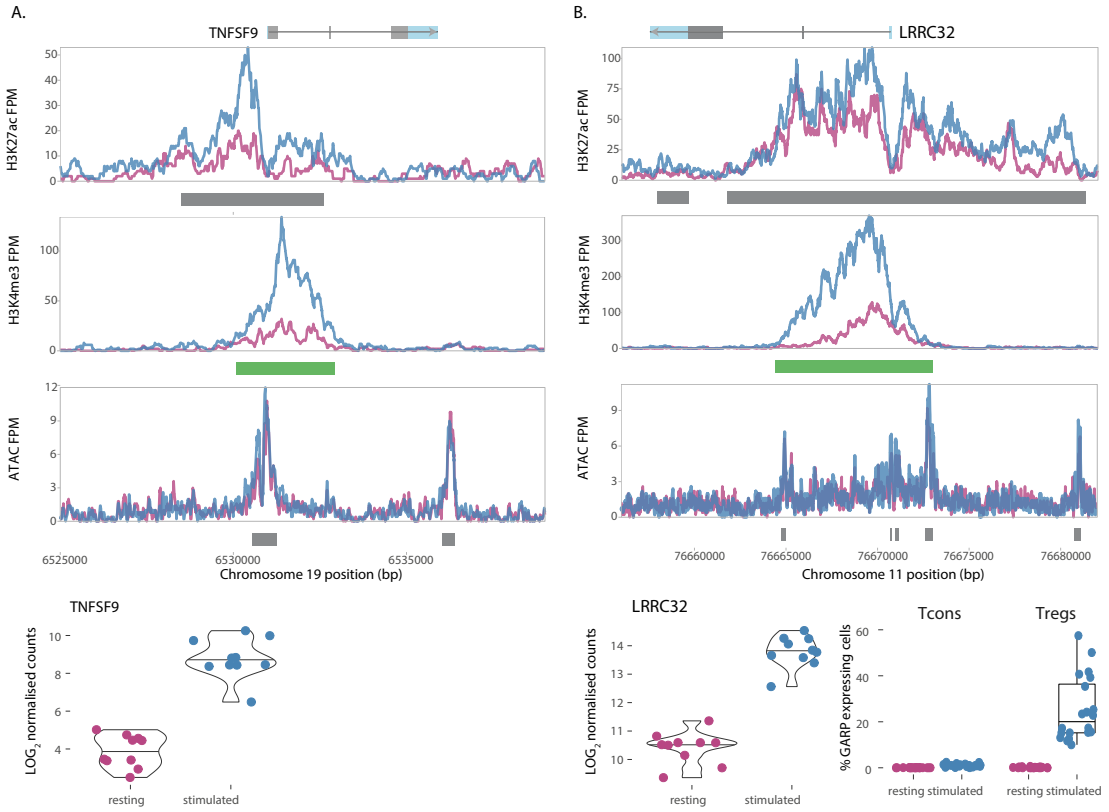

Figure 3.12: **Expression of genes related to the immune response is modulated through active promoters and enhancers differentially regulated upon stimulation. A.** Gene expression and genome track sequence pile-ups for H3K27ac, H3K4me3 and ATAC for *TNFSF9*. **B.** Gene expression and genome track sequence pile-ups for H3K27ac, H3K4me3 and ATAC for *LRRC32*. Also shown is the FACS plot of expression of GARP in resting and stimulated conventional and regulatory T cells.

responses (Zheng et al., 2009). This finding could implicate that upon stimulation Tregs recruit different TFs to promote T cell activation and ultimately arrive to an immune response resolution.

Table 3.2: Upregulated genes that are only regulated at the level of H3K4me3.

Symbol	Reg. Score	Symbol	Reg. Score	Symbol	Reg. Score
SESN2	0.067	*TNFSF9*	0.309	*MAP2K3*	0.081
AIM2	0.067	*PPAN*	0.413	*C8orf33*	0.12
LINC01353	0.094	*PPP1R15A*	0.024	*CHRNA6*	0.153
SH2D2A	0.092	*DUSP2*	0.138	*PYCR1*	0.368
NTRK1	0.156	*FOSL2*	0.08	SNHG15	0.078
NPM3	0.319	*SNHG17*	0.041	TRAF4	0.14
PPIF	0.117	*CENPM*	0.226	*TRIP6*	0.585
CCDC86	0.042	*HMOX1*	0.108	*AKAP7*	0.086
LRRC32	0.105	*RRP9*	0.287	*C16orf91*	0.221
IL23A	0.272	*EGR1*	0.132	*CHSY1*	0.035
LINC00944	0.062	*IRF1*	0.028	*CDKN1A*	0.08
NFKBIA	0.083	*LUCAT1*	0.033	*IPO4*	0.166
TBX21	0.037				

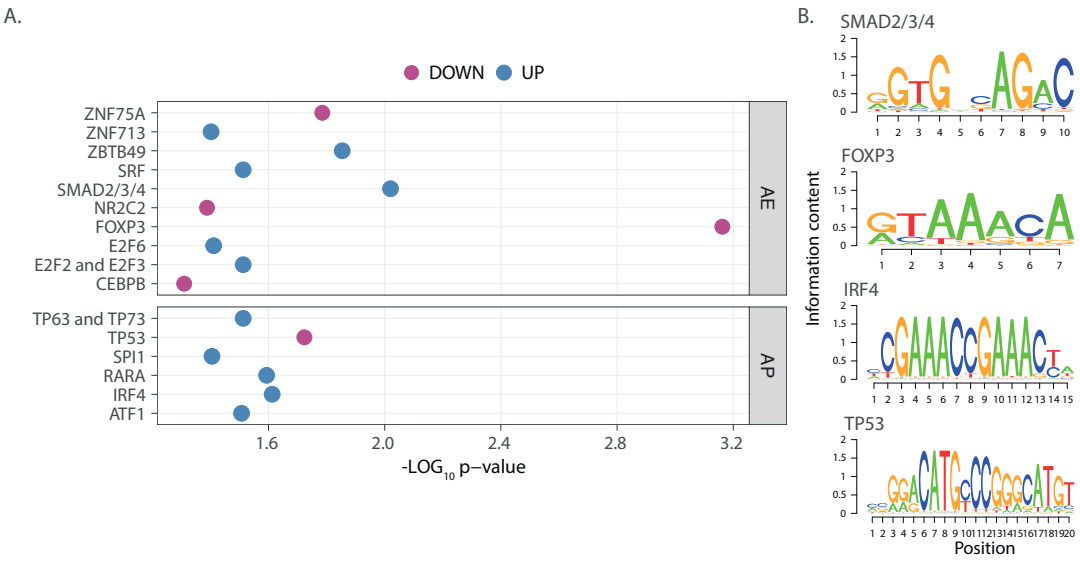

Figure 3.13: **Transcription factor enrichment in differential active promoters and enhancers. A.** Transcription factor binding sites (TFBS) enriched in the differentially regulated active enhancers (AE) and active promoters (AP) that were predictive of gene expression changes. **B.** Position weight matrices (PWMs) for the enriched TFBS motifs.

3.3.6 Differential splicing upon stimulation

Of 5,018 differential peaks there were 4,367 that were not assigned to regulate the expression of a gene nearby. This could implicate a long range enhancer-promoter interactions, not captured by the 3 kbp window applied in my analysis. Another explanation could be that a proportion of these peaks are associated with alternative splicing or promoter switching. Previous studies did not report concordant changes between H3K4me3 and gene expression upon T cell stimulation (Barski et al., 2009; Ni et al., 2016). However, using the combination of different chromatin marks I was able to establish an association between histone modifications and gene expression levels. Therefore, I hypothesise that similar effects could also be extended to differential splicing events.

I used LeafCutter (Li et al., 2018) to identify and quantify the relative excision ratios of 26,342 alternative introns. LeafCutter only maps reads across junctions, so an excision ratio is estimated based on the relative number of reads spanning a specific junction in relation to the total number of reads that fall within that cluster. I performed differential intron excision analysis between resting and activated Treg cells and detected 6,861 differentially spliced clusters corresponding to 4,279 genes (FDR < 0.01). Only a quarter of these genes had been detected as differentially expressed when considering total gene counts, the majority (72%) of which were upregulated (Figure 3.14 A). For example, 276 TFs (35% of all the TFs present in the dataset) had at least one differentially spliced cluster, of which only 67 were differentially expressed as a total gene count. This set included TFs essential to Treg cell identity, such as *FOXP3* (Figure 3.14 B) and *IKZF2*, which encodes for Helios, as well as *BACH2*, which is important in the NF-κB signalling pathway and *PRDM1*, which controls the expression of IL-10 (Martins et al., 2006). Therefore splicing analysis can give new insights into Treg stimulation induced gene regulatory processes that might be missed at the gene counts level.

Since the LeafCutter algorithm has been reported to annotate up to 37% of novel junctions (Li et al., 2018), I expected to find previously unreported splicing events in Tregs. I investigated whether any of the differentially spliced junctions identified were novel based on the GENCODE intron database (Harrow et al., 2012). I annotated 34.5% of junctions as cryptic (Figure 3.14 C), due to either a new 5' splice site, a 3' splice site, two new splice sites or a new connecting junction. In order to assess the likelihood

of the newly identified introns being functional I measured the conservation of their splicing patterns across different species using the PhyloP score (Pollard et al., 2010). I found that the annotated splice junctions along with the novel junctions had a PhyloP score shifted towards higher conservation compared to a random selection of base pairs (t-test p-value < 10^{-16}, mean PhyloP 3 and 3.4 accordingly). Cryptic 5' and 3' splice sites also had a shifted score towards high conservation (t-test p-value < 10^{-16}), but less than the annotated junctions (mean PhyloP 0.4 and 0.9 accordingly). The score of unanchored sites differed only slightly from the background (Figure 3.14 D).

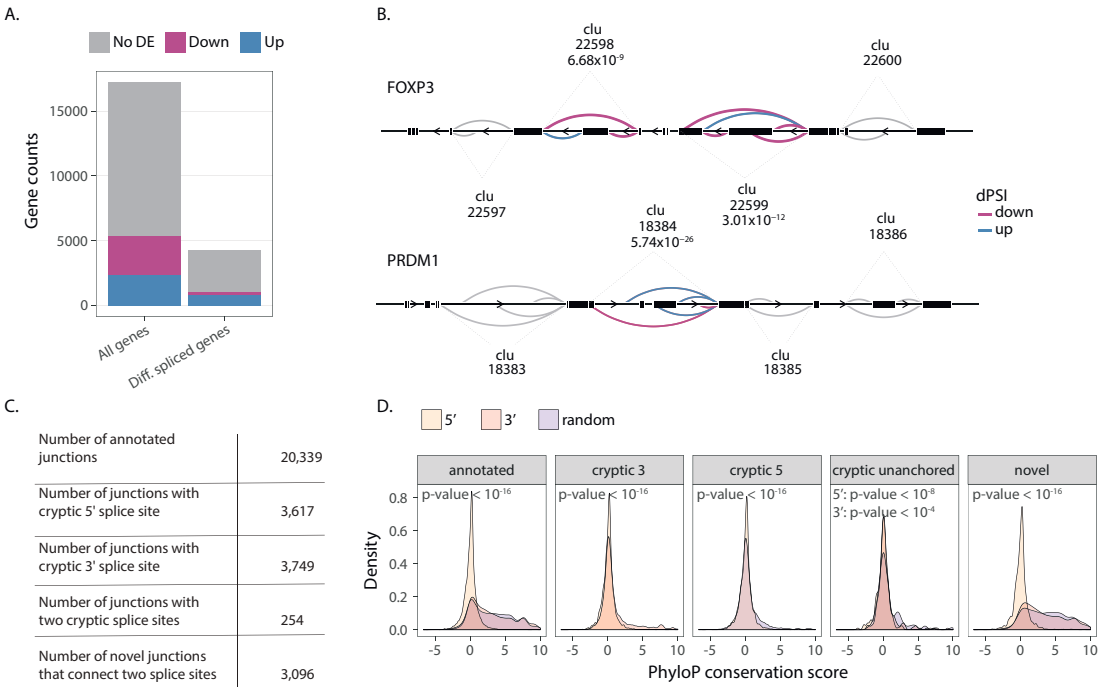

Figure 3.14: **Treg stimulation induces alternative splicing events that are not captured by gene counts analysis. A.** Differential expression status of all genes compared to differentially spliced genes. **B.** The hallmark Treg TF, FoxP3, displayed two differentially spliced clusters in response to stimulation. **C.** The number of differently annotated discovered junctions. **D.** PhyloP conservation score of the differentially spliced annotated junctions.

After having built confidence in the reliability of the novel splice junctions, I investigated whether any of the identified events could result from a novel TSS. In order to answer this, I overlapped the introns with differentially up and down regulated active promoters. An additional 18% of upregulated active promoters (corresponding to 592 regions) to the 12% annotated previously overlapped with differentially spliced sites. The overlap

was smaller in downregulated regions, with only 9% of peaks overlapping. The overlap was equally distributed across the five different junction annotations (Figure 3.15 A), where of the 6,861 differential clusters 536 overlapped with an active promoter. The majority of the genes overlapped with more than one differentially active promoter, suggesting indeed more than one TSS for many differentially spliced genes (Figure 3.15 B).

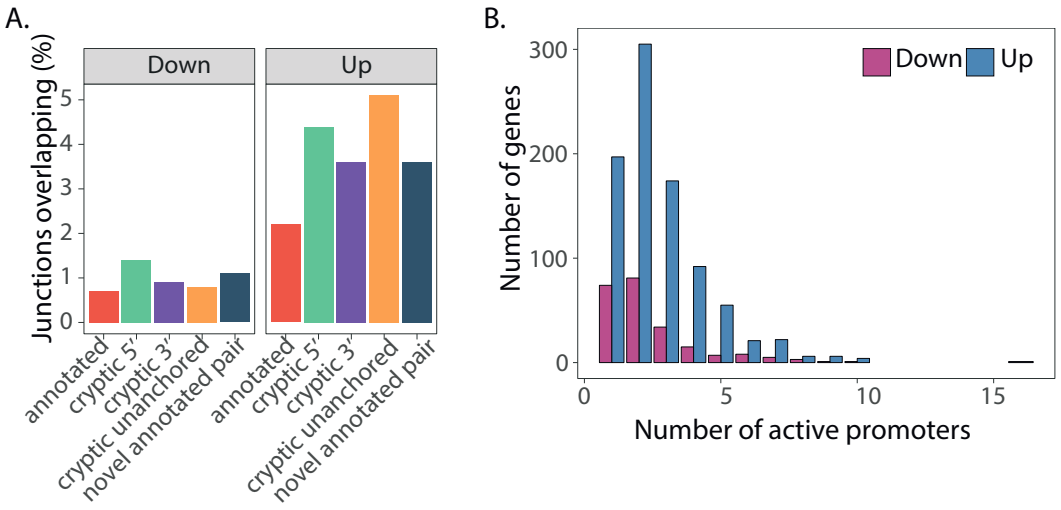

Figure 3.15: **Differential splicing analysis combined with differential active promoters identifies unannotated promoters. A.** Proportion of splice junctions that overlap with a differentially active promoter, stratified by the same direction of regulation. **B.** Number of differentially active promoters per differentially spliced gene.

To examine this hypothesis, I focussed on the genes that had more than one peak overlapping a differentially spliced site. I found 137 genes which had an intron cluster that overlapped with more than one upregulated active promoter, of which 21 had at least one non-annotated site identified through alternative splicing. For example, *BACH2* was characterised by a hidden promoter near exon 5 which has not been previously reported as a TSS (Figure 3.16). I manually curated all of the results into a confident list of putative novel TSS discovered in activated Tregs (Table 3.3).

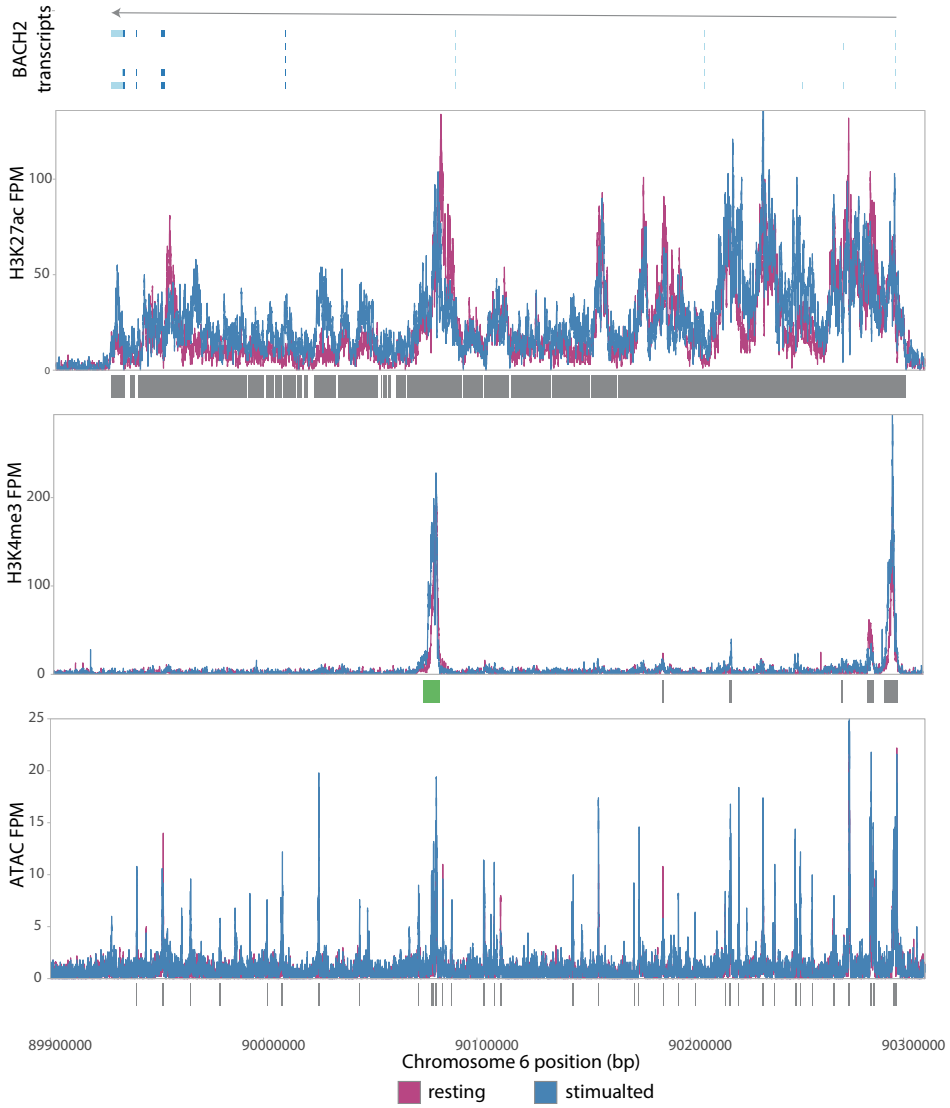

Figure 3.16: **BACH2 read pile ups across the different assays.** BACH2 is an example of a differentially spliced gene with a hidden promoter near exon 5.

Table 3.3: Curated table of gene clusters with cryptic 5' as identified with LeafCutter which overlap with a differential active promoter.

Gene	Chr	Cluster start	Cluster end	Psi	Peak start	Peak end	Reg.
THEM4	chr1	151895194	151907321	-0.035	151904370	151906291	down
RABGAP1L	chr1	174982905	174989849	-0.061	174989781	174991559	up
DNMT3A	chr2	25314161	25314239	-0.013	25312413	25314978	up
ACOX3	chr4	8416535	8435923	-0.003	8434792	8436758	up
RFC1	chr4	39351476	39353287	-0.032	39352048	39354813	up
GPRIN3	chr4	89250233	89292972	0.013	89287934	89291450	up
CAMK2D	chr4	113661772	113677508	0.02	113677125	113677605	up
RP11-223C24.1	chr4	142556334	142583452	0.074	142556250	142558001	up
EXOC2	chr6	497489	498736	-0.025	496434	501111	up
BACH2	chr6	90008856	90079634	-0.017	90074556	90082215	up
PRKAR1B	chr7	584568	585196	-0.053	582705	585353	up
PRKAR1B	chr7	596304	601453	-0.068	600745	605710	up
MAD1L1	chr7	2014642	2066490	0.112	2065942	2066709	down
NCOA2	chr8	70141399	70141479	-0.703	70137278	70145939	up
NCOA2	chr8	70216764	70216997	-0.169	70216687	70217838	up
TTC39B	chr9	15267948	15299350	0.037	15294520	15297400	up
TTC39B	chr9	15267948	15299350	0.037	15298564	15300490	up
ETS1	chr11	128462576	128480191	0	128478615	128481236	up
ATP10A	chr15	25781223	25861782	0.071	25849286	25851564	up
CFDP1	chr16	75305182	75391380	0.011	75380122	75381111	down
RAI1	chr17	17681793	17792933	-0.002	17781594	17784555	down
RAI1	chr17	17681793	17792933	-0.002	17792699	17793267	down
SIRPG	chr20	1649408	1686332	-0.048	1681827	1687206	up
TIAM1	chr21	31195305	31195406	-0.351	31185659	31196858	up
RAC2	chr22	37226803	37241587	0	37237226	37239494	up

3.4 Discussion

Regulatory T cells are necessary to maintain the immune system in balance, and both changes in Treg cell numbers and Treg functional defects can result in autoimmunity (Buckner, 2010). The importance of Tregs is greatly appreciated by the immunology community and there is an abundance of studies characterising Tregs through FACS (Roncador et al., 2005; Duhen et al., 2012). However, these studies rely on a limited number of protein markers, and a more global approach to identify Treg specific genes and transcriptional programs is needed. Previous studies suggest that there are large changes in gene expression in response to stimulation, which together compose a Treg specific signature. However these studies have either been carried out in mice (Stubbington et al., 2015), using microarray technology (Pfoertner et al., 2006) or in a limited sample size (Birzele et al., 2011; Bhairavabhotla et al., 2016). Understanding the changes induced by stimulation at the transcriptional level and integrating this with chromatin data is important to advance our understanding of how the immune response is orchestrated, and how that might increase the risk of autoimmunity.

The lack of genomic resources for Tregs can be partly attributed to the fact that they are scarce. Therefore, to study rare cell populations it is important to develop robust genomic protocols that are sensitive enough to be applied to small amounts of cellular material. In this chapter I presented a methodical comparison between an established protocol for chromatin profiling, ChIP-seq, and its recent modification, ChM-seq (Schmidl et al., 2015). ChM has been successfully used on cell lines (Schmidl et al., 2015), which are easier to handle than primary cells, and on innate lymphoid cells (Lim et al., 2017), but the robustness of the protocol across other cell types remains to be shown. Here, I demonstrated that ChM-seq can be robustly used for profiling primary Tregs, a cell type that is rare and has low viability *ex vivo*. ChM-seq was faster to perform than established ChIP-seq protocols, captured 100% of the peaks mapped with ChIP-seq and achieved an overall higher significance for the same peaks. I therefore applied ChM-seq in a larger scale genomic analysis to profile stimulation induced changes in Treg regulatory landscape.

Through the integration of profiles from three different chromatin marks, H3K27ac ChM-seq, H3K4me3 ChM-seq and ATAC-seq, I was able to gain valuable insights into the

dynamic process underlying complex gene regulatory processes in Tregs and define gene promoters and enhancers. The majority of histone changes upon stimulation were induced at the level of acetylation of H3K27 and chromatin accessibility, while the gene promoters, as marked by the tri-methylation of H3K4 remained unchanged. H3K4me3 is deposited at the 5'-end of a genes and is known to mark actively transcribed genes. The strength of H3K4me3 signal at promoters is strongly correlated with the expression of genes (Okitsu et al., 2010). By combining chromatin data with gene expression I was able to demonstrate that changes in the chromatin landscape were predictive of the expression of 1,706 genes. Of these, 35 genes were regulated at the level of H3K4me3 exclusively, without changing the status of H3K27ac. This implies the precise action of methyltransferases independently of acetyltransferases, which to my knowledge has yet to be reported in Tregs. Among the 35 genes was *LRRC32*, which encodes for GARP, whose expression on the cell surface correlated with the increase in gene expression. GARP has been suggested to be a specific marker for activated Tregs, with levels of GARP expression correlating with the cells' suppressive capacity (Wang et al., 2009). It would therefore be interesting to study the remaining genes on this list in the context of Treg biology to better understand their regulatory mechanism and function. In fact, many of these are already known to be highly expressed in Tregs upon stimulation, such as *DUSP2* and *FOSL2* (Birzele et al., 2011), while others have never been reported before, such as *MAP2K3* and *TNFSF9*. *TNFSF9* encodes for the ligand of 4-1BB, which has been extensively studied in Tregs (Marson et al., 2007). Recent data suggests that 4-1BB ligand is expressed on T cells in order to prevent effector T cell development and maintain a favourable Treg to Tcon ratio (Eun et al., 2015). Along with our data demonstrating the upregulation of *TNFSF9* and the formation of a new promoter upon stimulation, this suggest that expression of 4-1BB ligand might be another mechanism employed by Tregs to maintain the immune balance.

The confident definition of active promoters through the intersection of the chromatin annotation layers also gave me the opportunity to examine unannotated genes TSS. I used RNA-seq annotation-free approaches to quantify splicing events induced upon stimulation. I overlapped the newly defined exon junctions with the differentially active promoters to determine whether some of these could mark unknown transcripts. I found 21 genes that had unannotated promoters spanning their gene body. These included the transcription repressor *BACH2* which promotes the differentiation of Tregs

(Roychoudhuri et al., 2013) and has been implicated in immune-mediated diseases such as coeliac disease and inflammatory bowel disease. Another gene with a putative cryptic TSS is the DNA methyltransferase *DNMT3A*. *DNMT3A* is expressed at lower levels in Tregs of patients with rheumatoid arthritis (RA) than healthy controls as measured by qPCR (Kennedy et al., 2014), even if deletion of the gene in mice does not seem to impair the expression of *FOXP3* (Wang et al., 2013a). This discrepancy could be due to the fact that RA patients express a different isoform of *DNMT3A* which might be missed by conventional primers designed for the primary isoform. Experimental validation of the expression of different transcripts and determining whether they are coding would increase our understanding of Treg cell biology.

The clustering of genes by expression pattern across the two conditions provided insights on the processes of gene regulation upon stimulation. The majority of genes related to the immune response were upregulated upon stimulation. These included previously identified Treg specific genes such as *IL2RA*, *CTLA4*, *LRRC32*, *IL13*, *IL10* and *TNIP3* (Birzele et al., 2011) and non-Treg specific genes such as *SCD* and *IL22*. In addition, I found that some genes previously believed to be specific to naive T cells, were also upregulated in Tregs, such as *FOS*, *CXCL11* and *NR4A2*. These discrepancies could come from different Treg isolation protocols (use of commercially available kit instead of flow cytometry sorting), different stimulation approaches (anti-CD3/anti-CD28 antibodies instead of PMA/ionomycin) and differences in sample size. Integration of different datasets and combined analysis could help determine the source of these differences and instruct the design of future RNA-seq experiments.

By examining different isoforms in Tregs I was able to find a set of genes that employ alternative promoters upon stimulation. Interestingly, amongst these genes I identified five transcription factors (TFs), three of which have documented functions in T cells. For example, *ELF4* has been shown to facilitate *FOXP3* expression in thymic Tregs (Rudra et al., 2012) and to be downregulated upon activation in naive CD4+ cells (Yamada et al., 2010). Alternative transcription regulation represents a potential regulatory method for the expression levels of this gene. Previous studies on alternative transcription of *NFATC1* have mostly focused on the lack of the C-terminal domain, which is specific to effector T cells and not Tregs (Vaeth and Feske, 2018). The long transcript of *NFATC1* plays an important role in Treg differentiation and function, where it can directly bind the

CNS2 of the *FOXP3* gene to maintain its stable expression (Li et al., 2014). However there are no studies that show a specific function for the alternative promoter I identified here. Finally, *YY1* TF levels are lower in resting Tregs compared to other T helper cells whereby it is essential for the regulation of Th2 cell fate (Hwang et al., 2013). YY1 overexpression anti-correlates with the level of FoxP3, since YY1 inhibits SMAD3/4 binding at the *FOXP3* locus, resulting in the loss of the cells' suppressive capacity (Hwang et al., 2016). In fact, in natural settings *YY1* expression is increased in Treg cells under inflammatory conditions. Here I observed a different isoform being upregulated upon stimulation, which results in a different protein coding product, shedding light into the potential regulation of this gene. If the alternative transcript of a TF lacks the DNA binding domain, this could have substantial effects on the downstream processes the TF regulates. Further validation studies on the potentially different functions of the two *YY1* isoforms are needed to better understand Treg regulation.

This chapter provides a resource of gene regulatory events detected in Tregs upon stimulation. It will take follow up studies to assess the extent to which these events are specific to Tregs or whether they reflect more global changes in gene expression in T cells upon stimulation.

4

Linking genetic effects of molecular phenotypes of regulatory T cells to immune disease associated variants

Collaboration note

The data acquisition was performed in collaboration with Natalia Kunowska, Gosia Golda and Claire Cattermole, all of whom were Research Assistants in Gosia Trynka's lab. I did the RNA-seq data analysis and the integration with GWAS variants. Lara Bossini-Castillo performed the ATAC-seq, H3K4me3 and H3K27ac ChM-seq data analysis, and was involved in all discussions around this project. RNA-seq library construction and sequencing was undertaken by the DNA Pipelines core facility at Sanger. I thank Kaur Alasoo for the helpful discussions on the best QTL analysis approaches, Alice Mann for sharing the GWAS summary statistics and David Roberts' lab for providing us with half of the processed lymphocyte cones.

4.1 Introduction

Genome-wide association studies (GWAS) have revealed hundreds of genetic variants associated to common immune diseases, such as inflammatory bowel disease (IBD) and rheumatoid arthritis (RA). The vast majority of disease associated variants reside outside gene coding regions, which means that their downstream effects remain unknown. Furthermore, the majority of disease variants map to regions of strong linkage disequilibrium (LD), which can include up to hundreds of highly correlated single nucleotide polymorphisms (SNPs), making the causal variant statistically indistinguishable from others (Spain and Barrett, 2015). Finally, the associated loci harbour multiple genes, and without further experiments it is impossible to determine which gene is affected by the associated variant. It is important to address these limitations, since linking associated disease variants to target genes, pathways and cellular functions, is instrumental for understanding disease processes and the development of new therapies.

In order to link disease variants to target genes, an easily implemented approach is to measure the genome-wide gene expression levels to identify genes whose expression is correlated in a linear fashion with the allele frequency of the nearby variants, referred to as expression quantitative trait loci (eQTLs). Statistical approaches have shown that a small proportion of the total number of eQTLs (2%) is expected to colocalise with the associated variants (Guo et al., 2015). Indeed, coeliac disease variants were found to be enriched in peripheral blood mononuclear cells (PBMCs) eQTLs (Dubois et al., 2010; Trynka et al., 2011), and eQTLs detected in CD4+ memory T cells were enriched for RA and type-1 diabetes (T1D) variants (Hu et al., 2014).

However, in order to achieve reliable eQTL and subsequent colocalisation results, it is important to carry out gene expression assays in an isolated cell type, since gene expression is cell type specific. Different statistical approaches have been developed to determine the most relevant cell types for the study of different complex diseases. These studies leverage the fact that disease variants localise in the non-coding regions of the genome, and assess their enrichment in epigenetic marks. Such approaches identified that across different immune diseases the associated variants were enriched in active chromatin marks and accessible chromatin sites of CD4+ T cells (Maurano et al., 2012; Trynka et al., 2015, 2013; Finucane et al., 2015). CD4+ T cells are a heterogenous population of cells, with each subset characterised by a specialised function, and genetic effects specific to a rare subset of cells would be missed when assaying total CD4+ cells. For example, while an eQTL study in total CD4+ T cells from over 300 individuals found 166 genes colocalising with five immune traits, it was difficult to deconvolute the role of the different genes in each T cell subset (Kasela et al., 2017). In fact, previous studies showed that enrichment of disease variants was observed in rare cell subpopulations, such as regulatory T cells (Trynka et al., 2013), which only constitute 5% of CD4+ cells. Tregs were more highly enriched than other CD4+ T cell types for T1D and RA risk variants, which would imply that a proportion of the disease variants modulate gene expression by affecting the function of enhancers and promoters that are specific to Tregs. The observations made were based on genetic evidence, but the role of Tregs in immune diseases is well documented in the immunology field (Buckner, 2010) and many differences from naive cells have been reported using comparative RNA-seq approaches (Birzele et al., 2011; Bhairavabhotla et al., 2016).

Although correlation of disease associated variants with gene expression through eQTL studies is informative, most of the time it is insufficient to determine the exact causal variant within a locus of statistically correlated polymorphisms. This challenge can be addressed by carrying out RNA-seq in conjunction with assays that annotate the non-coding portion of the genome, such as profiling histone modifications through ChIP-seq or open chromatin regions with ATAC-seq. By performing QTL mapping across different molecular layers one is able to refine the allelic effects propagation from chromatin to gene expression. Additionally, assaying chromatin profiles gives the advantage that the variants can be functionally prioritised by assessing their location in the regulome. A variant within a peak will have a higher probability of being functional compared to a variant outside of a peak. Colocalisation across multiple layers of gene expression regulation would therefore provide a mechanism of action for a specific variant within the locus, such as the disruption of a transcription factor binding site. Of course, such approaches are expensive and if carried out would require a commensally large effort. It has been estimated that approximately 30% of variants affecting gene expression in CD4+ T cells act by affecting the chromatin conformation nearby a gene (Chen et al., 2016; Gate et al., 2018).

In this chapter I assess the role of genetic variants associated with immune-mediated diseases on gene expression (eQTLs) and transcript ratios (trQTLs) in naive and regulatory CD4+ T cells from 169 and 100 individuals, accordingly. I identify thousands of eQTLs and trQTLs and find that they are largely independent from each other. By comparing eQTLs and trQTLs between Tregs and naive CD4+ T cells, I pinpoint to Treg specific effects, and find that 58% of the effects are shared between these two closely related cell types. I colocalise eQTLs and trQTLs with GWAS in Tregs and naive T cells, and find that the majority of colocalisation associations occur with immune-mediated diseases. While the majority of colocalising signals were shared across the two cell types, there were many that were Treg specific, implicating important differences in the biology of the two cells. Finally, the same individuals from which I had Treg gene expression profile were assayed for ATAC-seq, H3K4me3 ChM-seq and H3K27ac ChM-seq, allowing me to map QTLs. By using the five different QTL maps I was able to refine 33 GWAS association signals to prioritise functional variants for *MAP3K8* and *TNFRSF9*/*PARK7*.

4.2 Materials and Methods

4.2.1 Cell culture and sample collection

Lymphocyte cones were obtained with informed consent from donors at the NHS Blood and Transplant, Cambridge (REC 15/NW/0282) and from the NHS Blood and Transplant, Oxford (REC 15/NS/0060).

All donors were healthy adults of Caucasian origin. PBMCs were isolated using Lympholyte-H (Cedarlane Labs, Burlington, Canada) density gradient centrifugation. CD4+ T cells fraction of the PBMCs was obtained by negative selection using EasySep® Human CD4+ T Cell Enrichment Kit (Cat. no. 19052, StemCell Technologies, Vancouver, Canada), following the manufacturer's instructions but using half the recommended volumes of the CD4+ T Cell Enrichment Cocktail and the D Magnetic Particles. CD4+ cells were resuspended in the FACS staining buffer (2 mM EDTA and 0.5% FCS in PBS) at 10^8 cells per ml. The cells were stained with the following antibody cocktail: anti-CD4+-APC (30 µl/ml final volume, clone OKT4, Cat. no. 317416, BioLegend, San Diego, U.S.), anti-CD127-FITC and (30 µl/ml, clone eBioRDR5, Cat. no.11-1278-42, Thermo Fisher Scientific, Waltham, U. S.) and anti-CD25-PE (80 µl/ml, clone M-A251, Cat. no. 356104, BioLegend) for at least 30 minutes at room temperature in the darkness. The cells were washed copiously with FACS buffer and resuspended at 10^8 cells per ml in full medium (IMDM, 10% FCS) and kept overnight at 4ºC. Before sorting, the cells were stained with DAPI, to discriminate live and dead cells. The CD4+, CD25high, CD127neg population corresponding to Treg lymphocytes was used for the downstream assays.

4.2.2 FACS staining

To define the proportions of memory and naive cells in the CD4+ population, an aliquot of 10^6 cells CD4+ enriched cells were resuspended in 100 µl FACS buffer and stained with a cocktail of anti-CD4+-APC and anti-CD127-FITC antibodies (3 µl each), anti-CD25-PE (8 µl) and anti-CD4+5RA-BV785 (4 µl, clone HI100, Cat. no. 304140, BioLegend), incubated at room temperature in the dark for at least 30 minutes, washed copiously with FACS buffer and analysed on BD Fortessa.

To verify the FoxP3 expression in the sorted Treg populations, after the sorting the cells were stained for expression of CD4+, CD25 and CD127 surface markers, and then stained with anti-FOXP3-BV421 antibody (5 µl per 10^6 cells, clone 206D, BioLegend) using the eBioscience™ Foxp3 / Transcription Factor Staining Buffer Set (Thermo Fisher Scientific), according to the manufacturer's instructions.

4.2.3 RNA-seq

For RNA-seq experiments, 0.5 x 10^6 sorted Treg cells were washed with ice-cold PBS and resuspended in TRIzol (Thermo Fisher Scientific). After standard phenol/chloroform isolation step, the total RNA contained in the upper, aqueous phase was further purified with RNeasy Mini Kit (QIAgen, Hilden, Germany), according to the manufacturer's instructions. The RNA libraries were constructed using KAPA RNA HyperPrep Kit (Roche, Basel, Switzerland), following a standard automated protocol. The libraries were multiplexed and sequenced at 75 bp paired-end on an Illumina HiSeq V4 to yield on average 57 million reads per sample.

4.2.4 ATAC-seq and ChM-seq

ATAC-seq was performed as described in Chapter 3 to yield on average 112 million reads per sample. ChM-seq was performed as described in Chapter 2 and Chapter 3 to yield 70 million reads per samples for H3K27ac and 79 million reads per samples for H3K4em3.

4.2.5 Polymorphism genotyping and imputation

A total of 551,839 genetic markers were genotyped using the Infinium® CoreExome-24 v1.1 BeadChip by Illumina. After quality control per individual, the total genotyping rate reached 0.99869. 243,820 variants passed several per polymorphism filtering steps (minor allele frequency (MAF) > 10%; SNP call rate > 95%, Hardy-Weinberg equilibrium (HWE) p-value < 0.001). We then performed imputation using BEAGLE 4.1 software (Browning et al., 2018) with a reference panel comprising the individuals included 1000 Genomes Phase 3 and the UK10K projects (model scale parameter = 2). We applied stringent

post-imputation quality filtering (Allelic R-Squared $\geq$ 0.8, HWE p-value < 0.001, MAF > 10% in the analysed cohort and in the reference panel) and 4,647,308 variants remained. Of those, 512,320 were insertion-deletions and 608 were multiallelic polymorphisms.

European origin was confirmed by principal component analysis (PCA) including the 1000 Genomes project and UK10K project individuals. Identity by state was calculated for all individuals and relatives and replicates were removed (pi hat $>$ 0.2). Duplicate donors were identified.

4.2.6 RNA-seq data processing

Reads were aligned to the GRCh38 human reference using STAR and quantified using featureCounts and using Salmon, as described in Chapter 3. I used VerifyBamID v1.0.0 (Jun et al., 2012) to detect and correct sample swaps and cross-contamination across the donors. I excluded short RNAs and pseudogenes from the analysis. I quantile normalised the gene expression values using the CQN method (Hansen et al., 2012) for downstream analysis. I collapsed the transcript expression levels to genes using tximport and excluded short RNAs and pseudogenes from the analysis. I only kept genes with mean expression greater than 1 transcript per million (TPM). This resulted in 12,227 genes using the STAR alignment approach and 13,773 genes using the Salmon approach.

In addition to collapsing the transcripts to genes, I used the Salmon transcript output to calculate transcript ratios based on the total expression per gene. I determined the transcript ratios of 125,879 transcripts corresponding to 14,885 genes. I used the STAR+featureCounts output to determine the euclidean distance between all samples in the dataset. Finally, I used Leafcutter (Li et al., 2018) to identify intron clusters and determine alternatively spliced junctions within these. I used these measurements to calculate the abundance ratios for the different junctions within a cluster. I determined the junction ratios of 162,251 junctions corresponding to 39,265 clusters.

4.2.7 Gene expression QTL mapping and analysis

Before carrying out a QTL study I removed the major histocompatibility complex locus and the X and Y chromosomes of each sample. I used linear regression implemented in the QTLtools (Delaneau et al., 2017) software to map cis eQTLs within a +/-500 kbp window around the gene. I used the "–permute 10000" within a 100 kb window option to obtain permutation p-values for each gene. For each mapping approach I used as covariates the number of PCs that would lead me to explain down to 1% of the observed variance, which was 16 PCs for STAR and Salmon, 19 PCs for transcript ratios and 37 PCs for splice ratios. I picked the top most significantly associated variant for each gene and used false discovery rate (FDR) correction to identify genes with at least one significant eQTL at 5% FDR level. I processed in exactly the same way the naive CD4+ T cell dataset after downloading it from the EGA archive using the study code EGAD00001002671. I included 14 PCs for STAR alignment and 8 PCs for transcript ratios using Ensembl.

I identified genes that were eQTLs using one method or one cell type at 5% FDR and used their lead variants in the other method or cell type to calculate the linkage disequilibrium (LD) between the two variants.

I defined an eQTL as cell type specific when comparing naive and regulatory T cells using three criteria: (i) the gene was expressed in one cell type only, (ii) the gene had a significant eQTL effect in one cell type (FDR $\leq$ 0.05) and not in the other (FDR > 0.2) and (iii) one cell-type's eQTL SNP was unlinked from any of the eQTL SNPs reported in the other cell type. I correlated the regression slopes of the top significant variant of each cell type with the same gene-variant combination of the other cell type, even if the gene-variant were not eQTLs. I defined genes with opposite effect if the absolute regression coefficient was greater than 0.5 in both cell types. In trQTL, I defined cell type specific effects using the same first two criteria, but the third one was: (iii) a different transcript's ratio was disturbed between the two cell types. I used g:Profiler (Reimand et al., 2016) R package to identify the KEGG pathways enriched for cell type specific eQTL and trQTL genes.

I used the fdensity function from the QTLtools package (Delaneau et al., 2017) to calculate the enrichment of regulatory marks near lead eQTL and trQTL variants.

4.2.8 Colocalisation with immune disease traits

I used coloc v2.3–1 (Giambartolomei et al., 2014) to test for colocalisation between QTLs and GWAS hits listed in Table 4.1 . I ran coloc on a 400 kbp region centered on each lead expression (e)QTL, transcript ratio (tr)QTL, chromatin accessibility (ca)QTL, activity (act)QTL and promoter (prom)QTL variant that was less than 100 kbp away from a GWAS variant with a nominal p-value $< 10^{-5}$. I only kept the colocalisations between molecular QTLs and GWAS that had more than 50 SNPs tested, and were therefore well powered, and where the sum of the probabilities of having a significant signal in both studies was over 0.8 (H3+H4 $\geq$ 0.8). I finally required the probability of the signal being due to a single shared variant (H4) to represent 0.8 of the probability of having a significant signal in both studies (H4/(H3+H4) $\geq$ 0.8). I calculated the R^2 between all colocalising lead immune GWAS variants to determine immune disease loci.

Table 4.1: GWAS summary statistics used in the colocalisation analysis.

Abb.	**Trait**	**Reference**	**Category**
ALL	Allergies	Ferreira et al., 2017a	Autoimmune
AST	Asthma	Demenais et al., 2018	Autoimmune
CD	Crohn's disease	Lange et al., 2017	Autoimmune
CEL	Celiac disease	Trynka et al., 2011	Autoimmune
IBD	Inflammatory bowel disease	Lange et al., 2017	Autoimmune
MS	Multiple sclerosis	International Multiple Sclerosis Genetics Consortium (IMSGC) et al., 2013	Autoimmune
PBC	Primary biliary cirrhosis	Cordell et al., 2015	Autoimmune
PS	Psoriasis	Tsoi et al., 2012	Autoimmune
RA	Rheumatoid arthritis	Okada et al., 2014	Autoimmune
SLE	Systemic lupus erythromatosus	Bentham et al., 2015	Autoimmune
T1D	Type-1 diabetes	Onengut-Gumuscu et al., 2015	Autoimmune
UC	Ulcerative colitis	Lange et al., 2017	Autoimmune
MCH	Mean corpuscular hemoglobin	Astle et al., 2016	Metabolic

CAD	Coronary artery disease	Nelson et al., 2017	Metabolic
LDL	Low density lipoprotein	Astle et al., 2016	Metabolic
T2D	Type-2 diabetes	Morris et al., 2012	Metabolic
AD	Alzheimer's disease (late onset)	Lambert et al., 2013	Other
DEP	Depression (broad)	Howard et al., 2018	Psychiatric
INT	Intelligence	Sniekers et al., 2017	Psychiatric
SCZ	Schizophrenia	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014	Psychiatric

4.3 Results

4.3.1 eQTL mapping and comparison of alignment methods

I collected 117 RNA samples from 100 genotyped donors of European ancestry over the course of 10 months. I used two alignment methods to quantify gene expression levels; the first one, STAR (Dobin et al., 2013), relied on a reference genome, while the second approach, Salmon (Patro et al., 2017), used a pseudo-alignment algorithm which is reference free. Additionally, the Salmon algorithm allowed me to quantify the transcript ratios (referred to as Ensembl) and the LeafCutter algorithm allowed to assess for differential splicing events across all samples (referred to as leafcutter) (Li et al., 2018).

I first verified whether there were any sample swaps, and found a single swap between two donors that was rectified. I examined how much of the gene expression variability was explained by any of the recorded experimental procedures using the STAR alignments and featureCounts quantification, and found that there was a large batch effect, based on when the blood samples were processed, explaining 37.7% of the variability. Using PCA I determined that the first 16 PCs collectively explained down to 1% of the variability. Since the Treg cohort was collected over the course of 10 months I wanted to test if that would have an effect on the gene expression profiles. For that I used the 17 donors who were sampled at two independent time points. I observed that the Euclidean distance between the gene profiles of two replicates was smaller than between

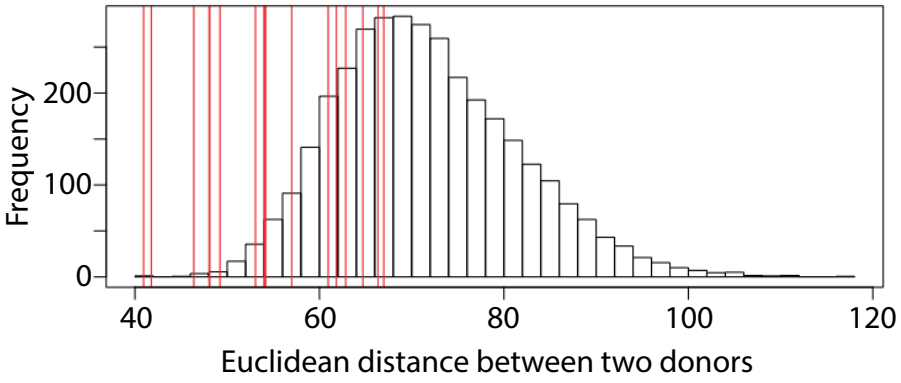

Figure 4.1: **Euclidean distance between all pairs of donors.** Histogram of the Euclidean distance between any two donors in the dataset. The red lines mark the location of the 17 donors who had blood drawn on two occasions.

two random samples (Kolmogorov-Smirnov test p-value = 1.33×10^{-6}) (Figure 4.1).

I used a standard linear regression model to map QTLs (Delaneau et al., 2017) using the gene expression and transcript ratios measurements, within a 500 kbp cis-window around the gene, transcript or intron cluster. I included the first 16 PCs as covariates for both quantification methods. I detected genetic effects on over 4,000 genes and 2,000 transcript ratios at 5% FDR (10,000 permutations; Figure 4.2). At the gene level I observed 23% more eQTLs using STAR combined with featureCounts than using Salmon quantification while at the transcript ratio level (trQTLs) there was no difference in the number of trQTLs detected. After conditioning on the lead variant I found at least one additional effect for 32%-48% of the genes, depending on the alignment method used.

In order to compare the different QTL mapping approaches I tested if the lead variants for the same gene-sets (or transcript-sets) were concordant. Specifically, I took all lead variants at 5% FDR from one method and compared them to the lead variants of the same genes (or transcripts) from the second method. I then calculated the fraction of lead variant pairs that were in high ($R^2 \geq 0.8$), medium ($0.2 \leq R^2 < 0.8$) or low LD ($R^2 < 0.2$). At the gene level, a small proportion of genes that were eQTLs did not pass the filtering thresholds using STAR, and were therefore annotated as specific (Figure 4.3). Nevertheless, 46% of the eQTLs detected using either of the two mapping approaches were in high LD with each other. For the remaining eQTL analysis I used the STAR quantification since it led to a higher discovery of eQTLs, yet discarding the eQTLs for the genes with low counts. At the transcript ratio level, 45% of Ensembl trQTLs were

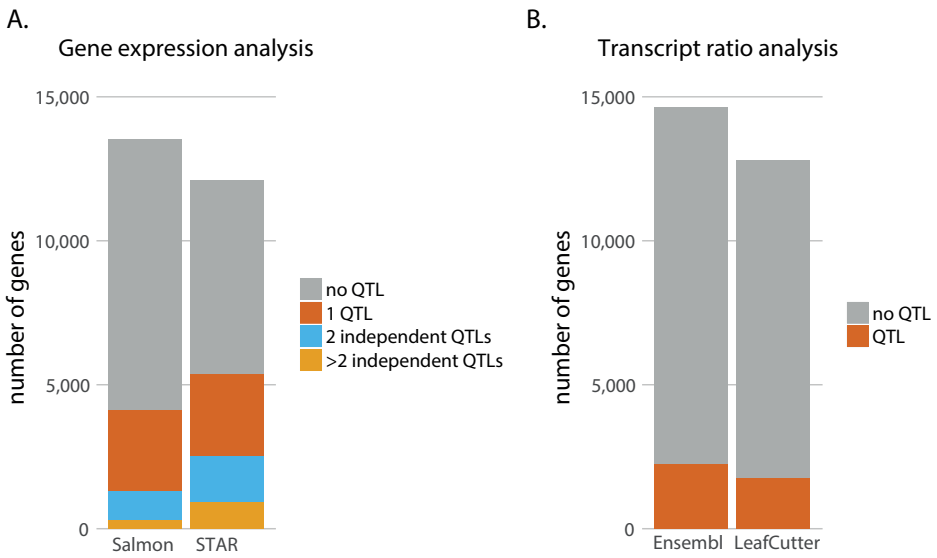

Figure 4.2: **Number of Treg genes whose expression is under control of a common genetic variant.** Number of gene eQTLs (**A**) and trQTLs (**B**) detected by four different mapping and expression quantification approaches. The colour corresponds to the number of independent variants influencing the expression of each gene.

concordant with the leafcutter trQTLs, while 37% of leafcutter QTLs were concordant with Ensembl trQTLs. Since I discovered more trQTLs using the Ensembl annotation and they resulted in higher concordance with leafcutter, I only used Ensembl for any further trQTL analysis. Finally, I compared the concordance between eQTLs and trQTLs. Interestingly, I observed that the two were largely independent of each other, with only 25% of eQTLs also affecting transcript ratios and 29% of trQTLs also being eQTLs. This suggests that the majority of eQTL variants are independent of splicing, and that genetic variants that affect transcript ratios might be undetected when collapsing read counts to gene levels.

4.3.2 Cell type specificity of gene expression

Next, to estimate the proportion of eQTLs specific to Tregs, I compared eQTLs discovered in Tregs with eQTLs called in CD4+ naive T cells. I chose this dataset because these two cell types are closely related, the naive CD4+ T cell dataset is of similar size (169 individuals) to the Treg dataset and all the individuals are of British origin (Chen et al., 2016). I assessed if the same set of genes was under genetic control in the two cell types. To call an eQTL gene cell type specific I required FDR $\leq$ 0.05 in one cell type and FDR > 0.2 in the second cell type. Using this criterion, I observed that the majority

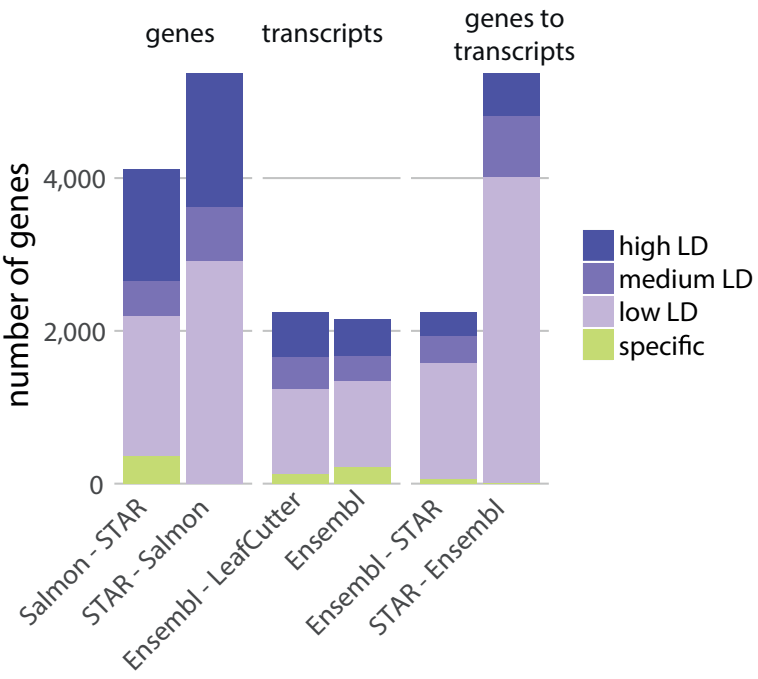

Figure 4.3: **Concordance between the lead QTL variants detected by different methods.** Sum of genes with different LD scores between the top variants identified by two quantification methods. The LD categories are: high LD: $R^2 \geq 0.8$; medium LD: $0.2 \leq R^2 < 0.8$; low LD: $R^2 < 0.2$.

of Treg eQTL genes were shared with naive T cells (77.4%) (Figure 4.4 A). To further understand if the genes that were called eQTLs in both cell types were regulated by the same genetic variants, I calculated the strength of LD between all of the independent associated genetic variants reported per gene between the two cell types. I only kept the strongest LD score per gene for all downstream analysis. I found that amongst the shared genes, the majority of the variants were in high LD with each other, but a substantial proportion (33%) were caused by independent signals. Therefore, to define a gene as a cell type specific eQTL I used the following three categories: (i) the gene was expressed in one cell type only (2.9% of eQTLs), (ii) the gene had a significant eQTL effect in one cell type (FDR $\leq$ 0.05) and not in the other (FDR > 0.2; 15.7% of eQTLs) and (iii) the Treg eQTL SNP was unlinked from any of the eQTL SNPs in naive T cells (23.9% of eQTLs). Amongst the second category I found many genes related to Treg function, including *CD28* co-stimulation receptor (FDR=2.38×10^{-8}; regression slope=0.155) and *MAP3K8* which induces the NF-κB immune response (FDR=5.11×10^{-6}; regression slope=-0.272) (Figure 4.4 C). In total, I found 2,284 Treg specific genes and 4,419 naive T cell specific genes. The higher number of naive eQTL genes can be partly explained by the larger sample size resulting in a larger starting number of eQTLs. Interestingly,

both naive and regulatory T cell specific eQTL genes were enriched in T cell receptor signalling pathways (Figure 4.4 E; p-value naive = $4.1x10^{-3}$; p-value Treg = $4.9x10^{-3}$). These genes can help us better understand the differences in biology between these two cell types, since they are involved in essential T cell processes.

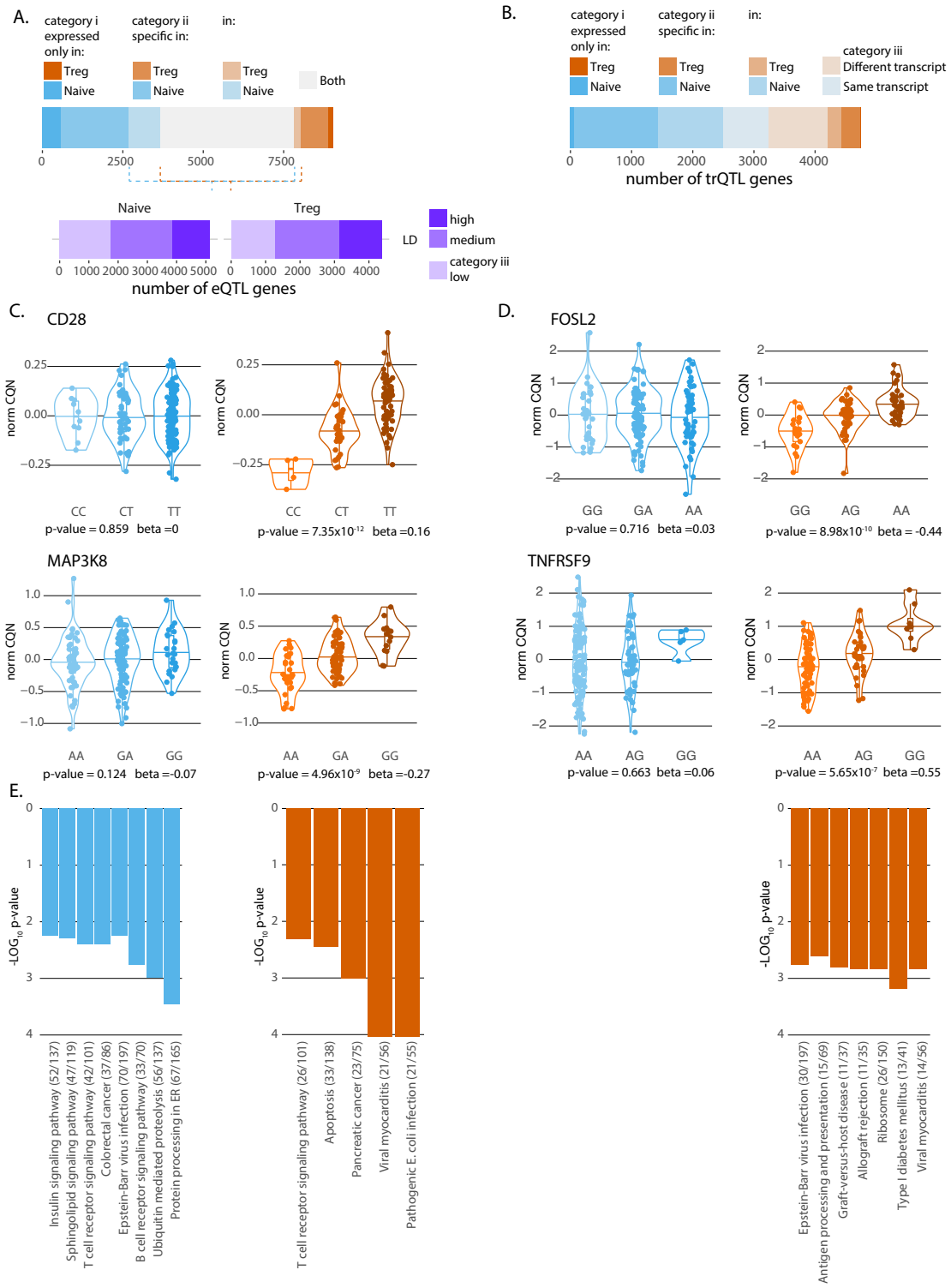

Figure 4.4: **Comparison of eQTLs and trQTLs called in CD4+ naive and regulatory T cells.** Classification of cell type specific QTLs discovered in naive and regulatory T cells for eQTL (**A**) and trQTL (**B**). Examples of Treg specific **C.** eQTLs and **D.** trQTLs. **E.** Results of the pathway enrichment analysis for the cell type specific eQTLs and trQTLs using the KEGG database.

I repeated the same analysis using the trQTLs (Figure 4.4 B). While the overall pattern looked similar to the eQTL analysis, I observed a lesser degree of overlap between the two cells. Additionally, I found that when examining transcript ratios, even when the same gene in naive and regulatory T cells was affected by a trQTL effect, in 42.7% of cases it was a different transcript's ratio that was disturbed by a genetic variant. I defined Treg/naive T cell specific trQTLs using the same first two categories as with the cell type specific eQTL genes, that is (i) the gene was expressed in one cell type only and (ii) the gene was a significant trQTL in one cell type (FDR $\leq$ 0.05) and not in the other (FDR > 0.2), which collectively represented 13.2% of trQTLs. In addition, I included a third category (iii) that a different transcript's ratio was disturbed between naive and regulatory T cells. Amongst the Treg specific trQTLs there were many genes related to the immune response, such as *FOSL2*, a component of the AP1 transcription factor machinery (FDR=1.3x10^{-5}; regression slope=-0.442) and *TNFRSF9*, which encodes for the 4-1BB co-stimulatory ligand (FDR=2.8x10^{-3}; regression slope=0.55) (Figure 4.4 D). No pathways were enriched in naive T cell specific trQTLs.

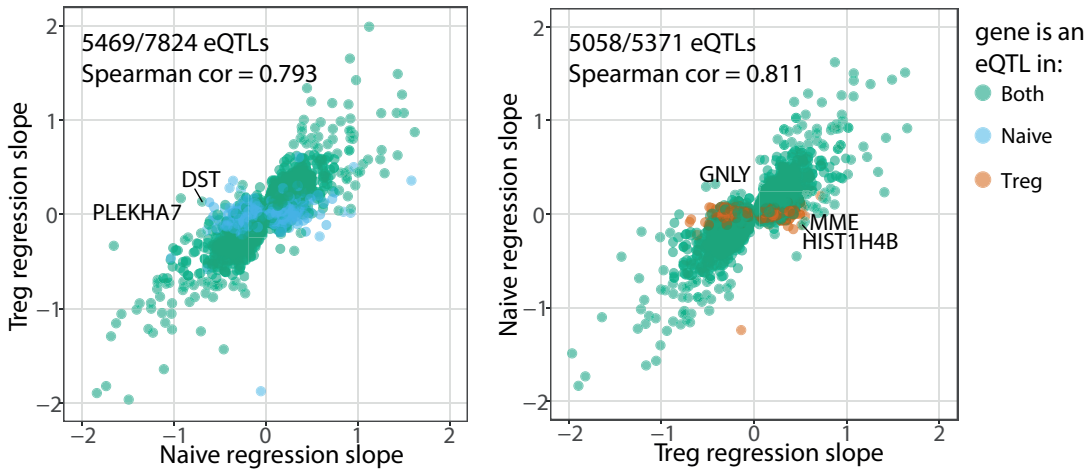

Figure 4.5: **Correlation between the regression slopes for the top eQTL variants discovered in CD4+ naive and regulatory T cells.** The regression slope of the top eQTL variant in naive T cells plotted against the slope for the same variant-gene pair in Tregs (left panel) and vice versa (right panel).

I observed that the sizes and the direction of the effects of the mapped eQTLs were highly correlated between the two cell types (Figure 4.5 ; Spearman correlation 0.793 in naive and 0.811 in regulatory T cells). There were only twelve genes with opposite directions and only five of these were eQTLs in both cell types; *DST* and *PLEKHA7*, which

are adhesion junction plaque proteins, were also eQTLs in Tregs and *GNLY*, *MME* and *HIST1H4B* were also eQTLs in naive T cells. None of these genes have a different function in the two cell types. The number of eQTLs with opposite effects was therefore very small, which is consistent with the close relation of these cell types.

4.3.3 Colocalisation of eQTLs and trQTLs with disease associated variants

I used a statistical colocalisation approach to determine whether any of the observed eQTLs and trQTLs colocalised with immune disease associations (Giambartolomei et al., 2014). Such colocalisation of disease and molecular signals would implicate the same underlying causal variant and could provide insights in the molecular mechanisms for some of the immune associated GWAS loci. I tested five hypothesis for colocalisation. Hypothesis zero (H0) tests whether there is any association at all, H1 and H2 test whether there is an association with just one or the other study, H3 tests whether the signal from GWAS and QTL is due to two independent SNPs, and H4, that the association between GWAS and QTL is due to one shared SNP. I used 0.5 as a probability threshold for H4 and calculated the colocalisation power (sum of H3 and H4), on which I used a 0.8 threshold. To report a significant colocalisation I used a 0.8 threshold of H4/power (Figure 4.6). By doing so I ensured that I focused my downstream analysis on signals at loci with colocalising associations predominantly driven by H4 (Guo et al., 2015).

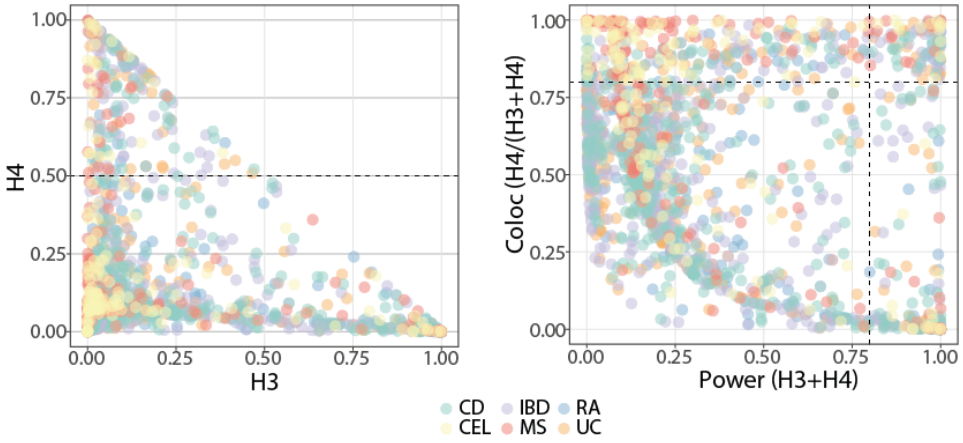

Figure 4.6: **Thresholds used for colocalisation between eQTLs and GWAS traits. A.** Probability of the signal being significant in both assays due to two independent causal SNPs (H3) or the same SNP (H4). **B.** Power (H3+H4) versus the coloc score (H4/power). I used a selection of six immune traits with a high number of associated SNPs for illustration purposes.

I observed tens of colocalisations between GWAS traits and eQTLs in both naive and regulatory T cells (Figure 4.7). The highest number of colocalisations was achieved for inflammatory bowel diseases (IBD), ulcerative colitis (UC) and Crohn's disease (CD), but also allergies (ALL), schizophrenia (SCZ) and rheumatoid arthritis (RA). I used metabolic disorders as a control, because I would expect to observe less colocalising signals between Tregs and metabolic traits compared to immune disorders. Indeed, the median number of colocalising signals across the ten most enriched immune diseases was ten, while it was only two for the ten most enriched metabolic traits. The number of colocalisations correlated with the number of regions tested, and in consequence I observed more colocalisations in naive T cells. Interestingly I observed a large number of colocalisations with SCZ, which has a known immune component. I additionally observed a substantial number of colocalisations with other psychiatric traits, such as depression (DEP) and intelligence (INT), since psychiatric disorders are characterised by a large degree of sharing across their loci (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013). I observed a similar number of hits in trQTLs and in eQTLs, despite testing approximately 25% fewer regions for trQTLs. This implicates that in addition to regulating gene expression, a proportion of immune disease variants will function through altering transcript ratios.

To examine whether among the colocalising signals there were some that were cell type specific, I focussed on ten immune diseases. There was a clear separation between the Treg colocalising signals, the naive colocalising signals and the colocalisations shared between the two cell types. A proportion of signals that colocalised in Tregs was accounted for by the Treg specific eQTL or trQTL (Figure 4.8 A). Those genes are thereby referred to as Treg exclusive colocalisations. The colocalising signals shared across the two cell types had similar regression slopes (R^2 = 0.88 for eQTLs and R^2 = 0.78 for trQTLs (Figure 4.8 B). The only exception was *ACO2*, which showed opposite effects in naive and regulatory T cells.

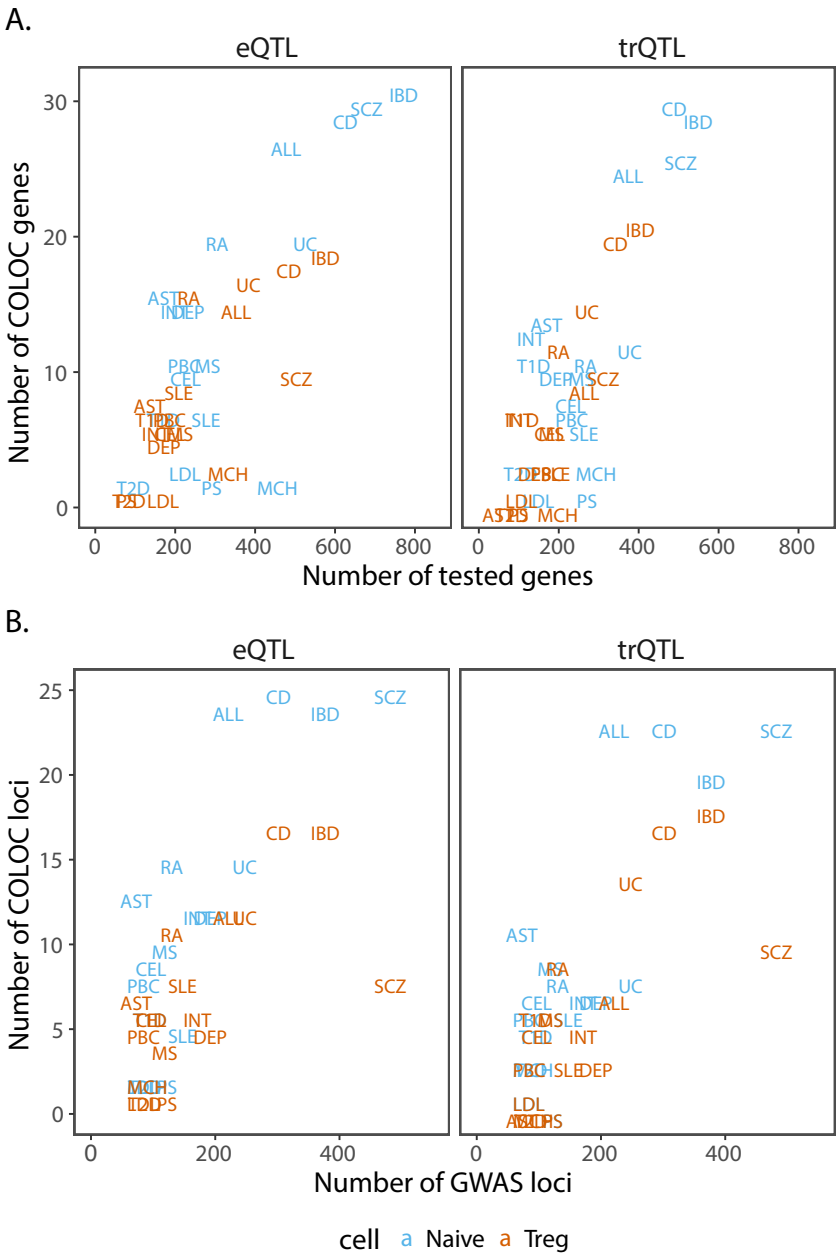

Figure 4.7: **Number of colocalising QTL genes and GWAS loci in relation to the number of tested QTL genes and the disease loci. A.** Number of tested eQTL and trQTL genes (FDR $\leq$ 0.05) per trait and number of genes colocalising in naive and regulatory T cells. **B.** Number of all GWAS loci with p-value < 10^{-5} versus the number of colocalising GWAS loci. AD: Alzheimer's disease; ALL: allergic disease (asthma, hay fever and eczema); AST: asthma; CAD: coronary artery disease; CD: Crohn's disease; CEL: celiac disease; DEP: Broad depression; IBD: inflammatory bowel disease; INT: intelligence; LDL: low density lipoprotein; MCH: mean corpuscular hemoglobin; MS: multiple sclerosis; PBC: primary biliary cirrhosis; PS: psoriasis; RA: rheumatoid arthritis; SCZ: schizophrenia; SLE: systemic lupus erythematosus; T1D: type-1 diabetes; T2D: type-2 diabetes; UC: ulcerative colitis.

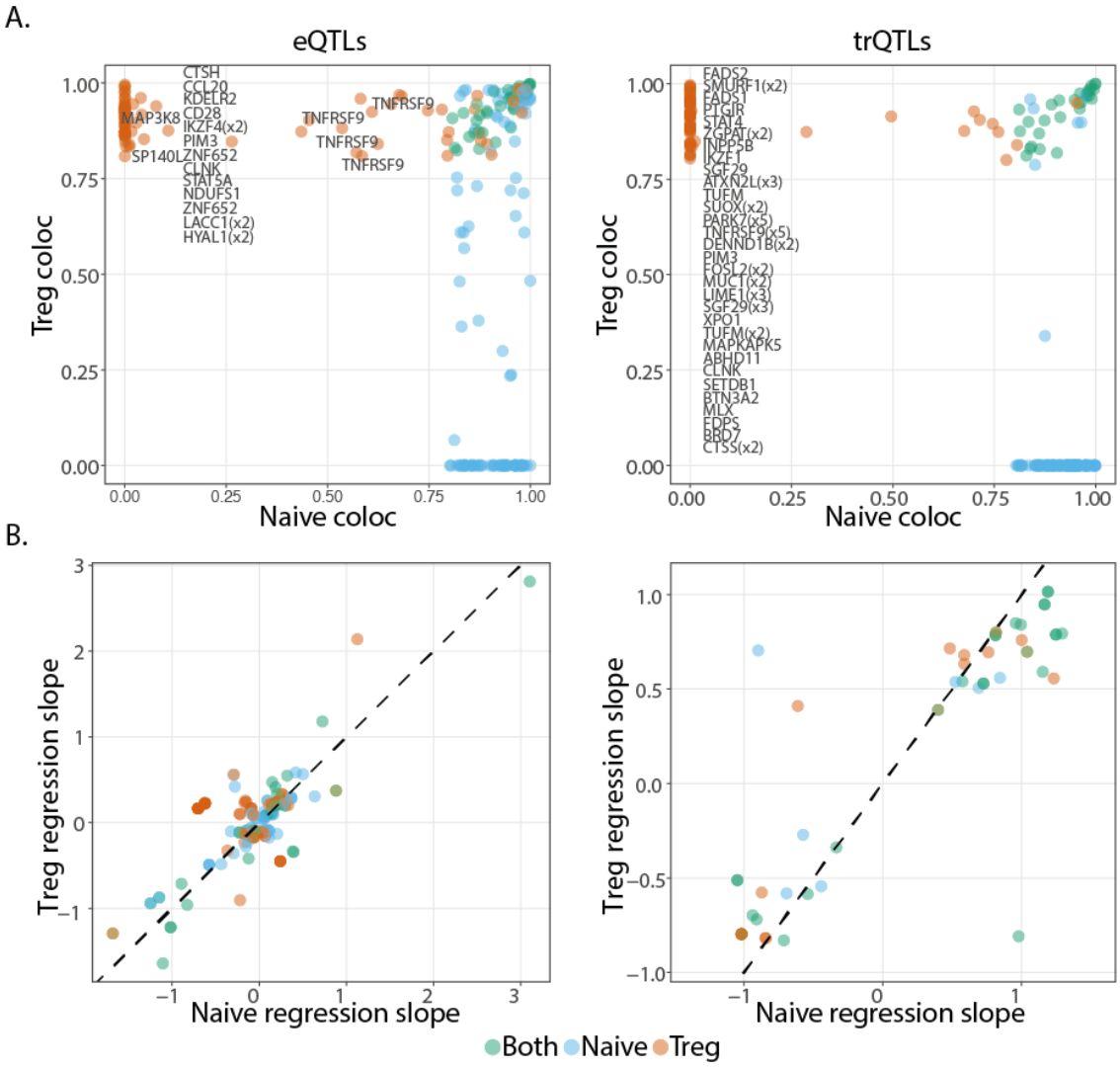

Figure 4.8: **Shared and specific colocalising signals in naive and regulatory T cells. A.** Coloc values in naive cells and Tregs for eQTLs and trQTLs. Genes have been assigned a color based on whether they colocalise in one or both cell types. Additionally, genes that are eQTLs or trQTLs specifically in Tregs are labelled. **B.** Correlation between the regression slopes of naive cells and Tregs for the same genes.

4.3.4 Coloc genes have higher probability of being loss of function intolerant

I wanted to further investigate if colocalising genes were truly biologically more relevant to the studied diseases. If that would be the case I would expect them to be less tolerant to loss of function mutations. I therefore assessed the identified eQTL and colocalising genes in the context of protein-coding studies and human knock-out variants (Lek et al., 2016). In this catalog each gene is assigned a probability of being loss of function intolerant (pLI) by counting the number of observed loss of functions alleles in 50 thousand exomes and extrapolating that to the probability of a gene being intolerant to being

knocked-out, therefore labelling it as essential. This means that a gene with an assigned high score is likely to be indispensable to the cell's core pathways. In their study the authors reported that eQTL genes are more tolerant to loss of function mutations. When assessing 10,747 genes that were expressed in Tregs and had calculated pLIs, I observed that the median pLI was 0.087, which decreased to 0.039 when only considering genes with eQTL effects in Tregs (Figure 4.9 A). This is concordant with the fact that genes with eQTLs are more likely to withstand genetic variability. When restricting the analysis to eQTL genes in immune GWAS loci, I observed a similarly low median pLI of 0.058. However, for the eQTL genes that colocalised with immune disease signals the median pLI increased to 0.2, indicating that disease colocalising Treg eQTL signals are enriched for genes that are more likely to be disease causal.

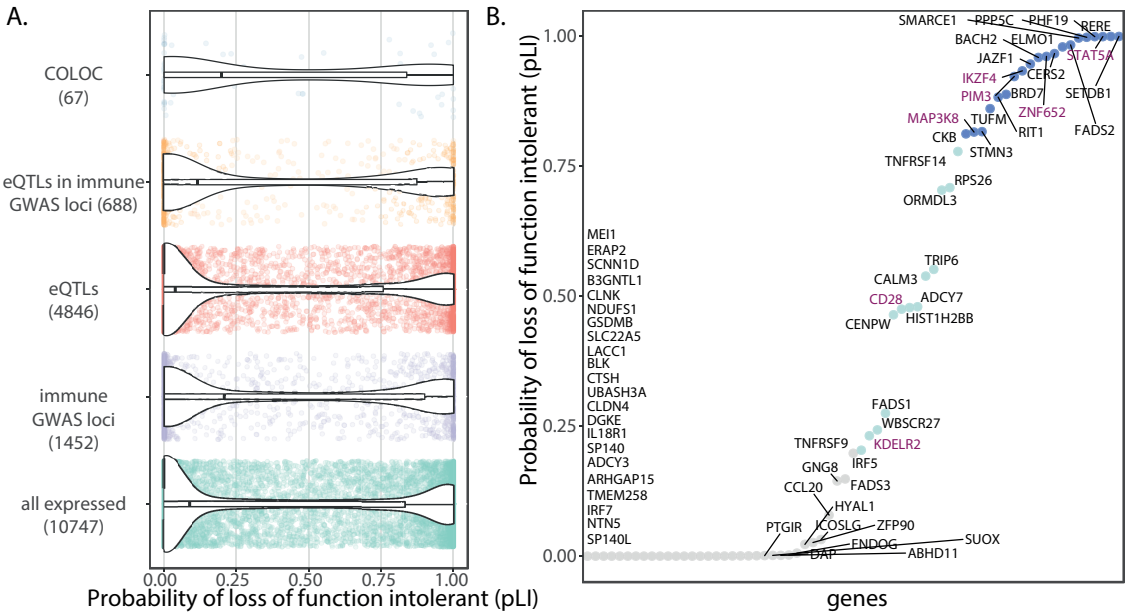

Figure 4.9: **Probability of coloc genes being loss function intolerant (pLI). A.** Distribution of probability of genes being loss function intolerant (pLI) grouped by whether (i) they are expressed in the dataset (ii) they fall within GWAS loci, defined by the lead reported variants and LD $R^2 \geq 0.8$. (iii) they are eQTLs in the dataset, (iv) they are eQTLs and they fall in immune GWAS loci, and (v) they colocalise with immune GWAS signals. **B.** Distribution of COLOC gene's pLIs.

Using a combination of colocalisation of eQTL and immune disease GWAS signals and restricting the list to the genes with pLI scores greater than the median, that is pLI > 0.2, I obtained a list of 33 genes with higher confidence of being truly disease relevant. Interestingly, I observed that many colocalising genes involved in the signalling of the

immune response (*CD28*, *TNFRSF14* and *MAP3K8*) and essential transcription factors (*ELMO1*, *BACH2* and *IKZF4*) were characterised by a pLI higher than the median 0.2 (Figure 4.9 B). Furthermore, for loci where a GWAS signal colocalised with a variant regulating the expression of more than one gene, this analysis allowed me to prioritise the most relevant gene. For example, *ORMDL3* and GSDMB both colocalised with a signal from RA, T1D, ALL, UC, CD, IBD, asthma (AST) and systemic lupus erythematosus (SLE). However, the pLI of *ORMDL3* (0.7) was much higher than that of *GSDMB* (0.0), which would indicate that *ORMDL3* is more essential for the cell's function, and therefore variation in its expression might have a higher phenotypic impact. Of the 33 genes, 7 are likely to play a role in disease biology through specifically affecting Treg function, since these were not eQTLs in naive cells (*CD28*, *KDELR2*, *MAP3K8*, *IKZF4*, *STAT5A*, *ZNF652* and *PIM3*).

4.3.5 eQTLs are enriched in active chromatin marks

Amongst the mechanisms by which eQTLs and trQTLs might exert their effects are alterations of transcription factor binding sites, leading to differential binding between the two alleles, switching on/off alternative promoters and altering the interactions between enhancers and promoters. These events are reflected in regulatory regions of the genome mapped by histone modifications and chromatin accessibility. I therefore tested if regulatory annotations were enriched in the vicinity of the eQTL and trQTL variants. I measured the density of functional annotations in a one million bp window around the lead associated variant reported for each eQTL or trQTL gene. I observed that both eQTLs and trQTLs were enriched in ATAC-seq, H3K27ac, H3K4me1 and H3K4me3 ChM-seq defined peaks (Figure 4.10). On the other hand, they were depleted from H3K27me3 ChM-seq peaks. Furthermore the observed enrichment was higher than in non-QTL genes. Having confirmed that both eQTLs and trQTLs were enriched in active regulatory regions, I proceeded to perform a QTL analysis using ATAC-seq, H3K27ac and H3K4me3 ChM-seq annotation layers.

4.3.6 Coordinated influence of gene expression and regulatory QTLs on GWAS loci

For the same individuals for whom I mapped eQTL and trQTLs I obtained QTL results for H3K27ac ChM-seq (91 individuals), H3K4me3 ChM-seq (88 individuals) and ATAC-seq

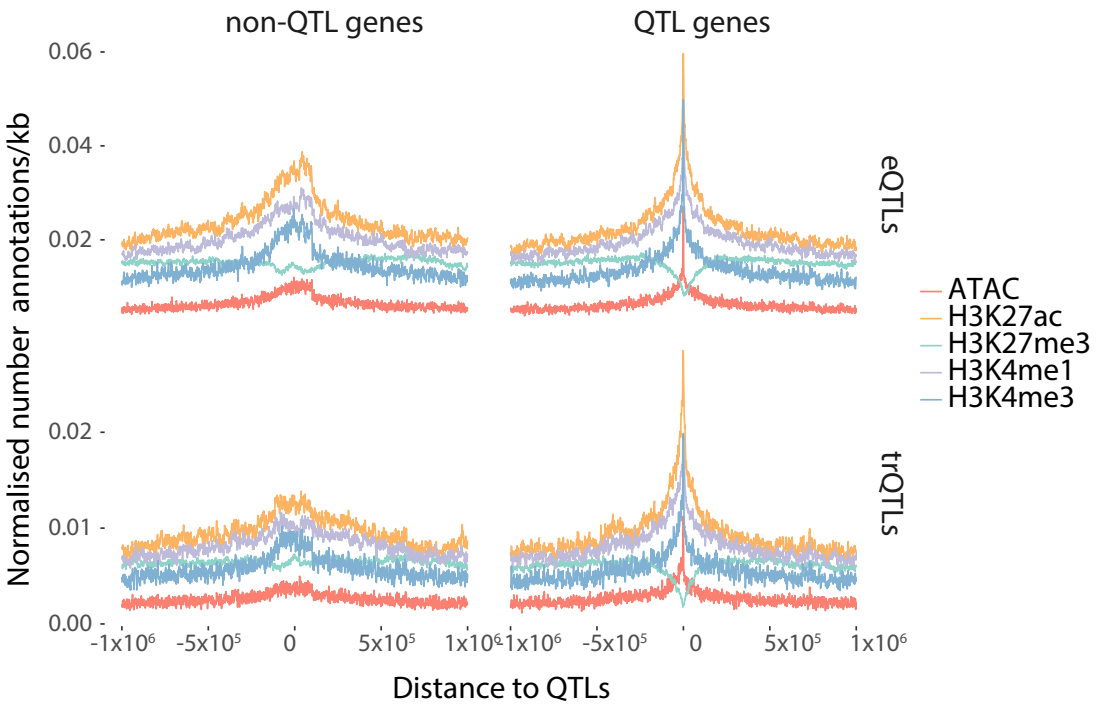

Figure 4.10: **Density of functional genomic annotations around Treg eQTLs and trQTLs.** The number of peaks called for the different chromatin assays was counted in 1,000 bp bins around 1,000 kbp of the positions of the top eQTL and trQTL for the significant genes.

(73 individuals). There were 9,179 activity QTLs (actQTLs called from H3K27ac ChM-seq), 5,992 promoter QTLs (promQTLs called from H3K4me3 ChM-seq) and 1,253 chromatin accessibility QTLs (caQTLs called from ATAC-seq) detected. Given that I observed that immune disease variants colocalised with eQTL variants, and active regulatory regions were enriched in proximity to eQTL and trQTL genes, I wanted to further investigate if there were instances where the chromatin regulatory signals (regQTLs) would also colocalise with disease variants. For that, I ran the colocalisation method, as previously described, but instead of eQTL and trQTL summary statistics I used regQTLs with GWAS signals. In immune diseases, I detected between 0 (psoriasis (PS)) and 48 (IBD) signals colocalising with QTLs mapped for all three chromatin marks (Figure 4.11 A), and H3K27ac showed the highest number of colocalising signals. In order to further quantify this, I calculated the relationship between all the lead GWAS variants which colocalised with the five QTL studies, and defined LD blocks based on $R^2 \geq 0.5$ (Figure 4.11 B).

I focussed on the GWAS variants that in addition to an eQTL or trQTL also colocalised with a caQTL, promQTL or actQTL (colocalization between eQTLs and actQTLs shown

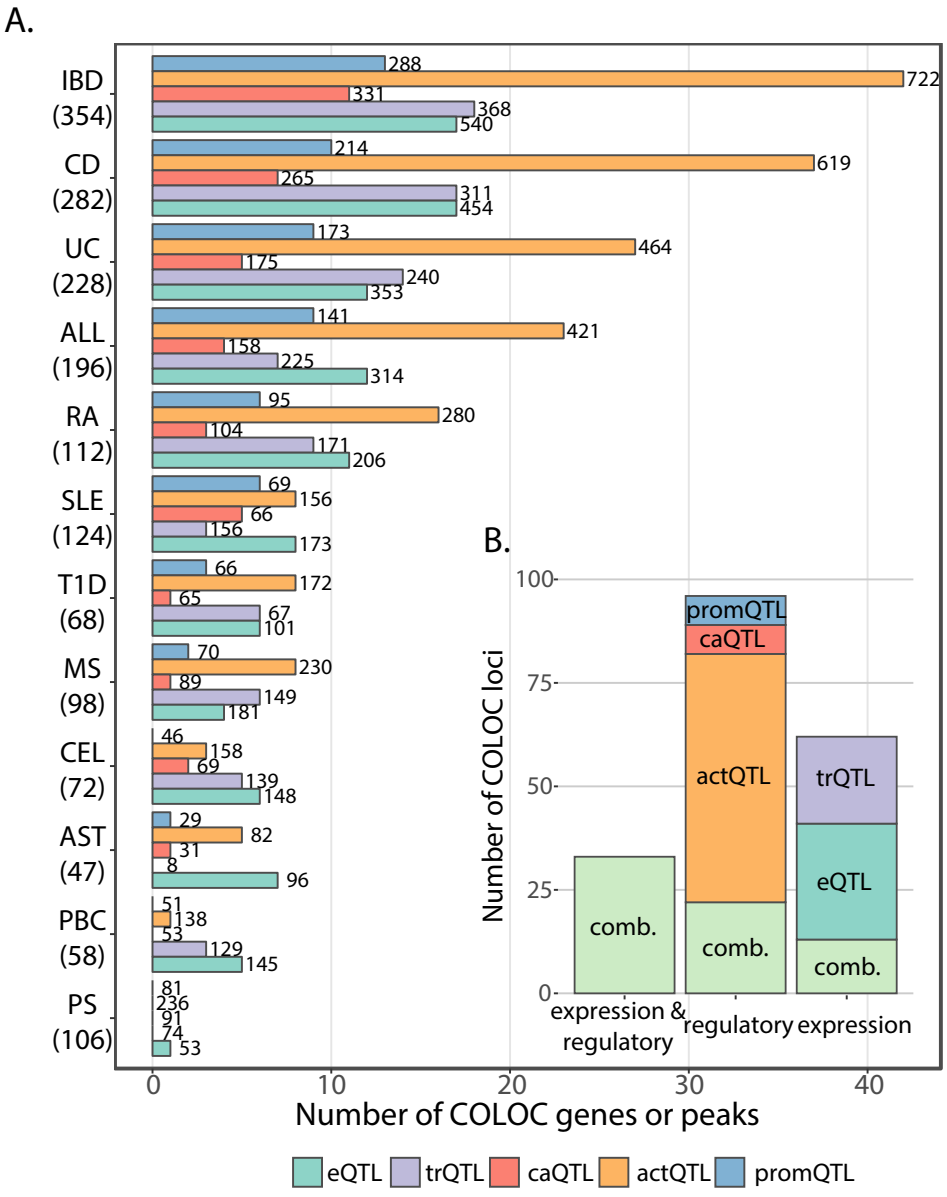

Figure 4.11: **Numbers of gene expression QTLs and regQTL colocalising with disease associated loci. A.** On the x-axis are the number of colocalising genes or peaks per assay and on the y-axis are the different immune traits tested. Below the traits, in parenthesis are the numbers of independent loci associated to the tested trait. The numbers at the end of the bars correspond to the total number of features (genes or peaks) tested for colocalisation. **B.** Number of colocalising loci determined by calculating the R^2 between the lead GWAS variants for the different traits. They have been categorised based on what type of assay they colocalised with.

on Table 4.2). Loci that overlap with QTLs for different chromatin marks and gene expression could be particularly informative for functional fine-mapping of disease loci and could provide a potential explanation for the observed effect on the expression of the gene. I found 47 lead GWAS SNPs which colocalised with regQTLs, the majority overlapped with an actQTL (42/47). Amongst these genes there were eight genes that I previously determined to be Treg exclusive colocalisations; *CCL20*, *CLNK*, *CTSH*, *MAP3K8*, *PIM3*, *SP140L*, *STAT5A* and *TNFRSF9*. Using the GTEx database (GTEx Consortium, 2015), I noticed that two genes, *CTSH* and *SP140L*, were ubiquitous eQTLs (Figure 4.12). I did not further investigate whether these eQTL effects were caused by the same signal as the one observed in Tregs. The remaining genes were eQTLs in a restricted number of tissues, which would suggest that these genes are only functional in specific cell types or tissues. For example, *MAP3K8* was only affected in skin from sun exposed areas while *TNFRSF9* was only affected in the stomach (nominal p-value < 10^{-6}).

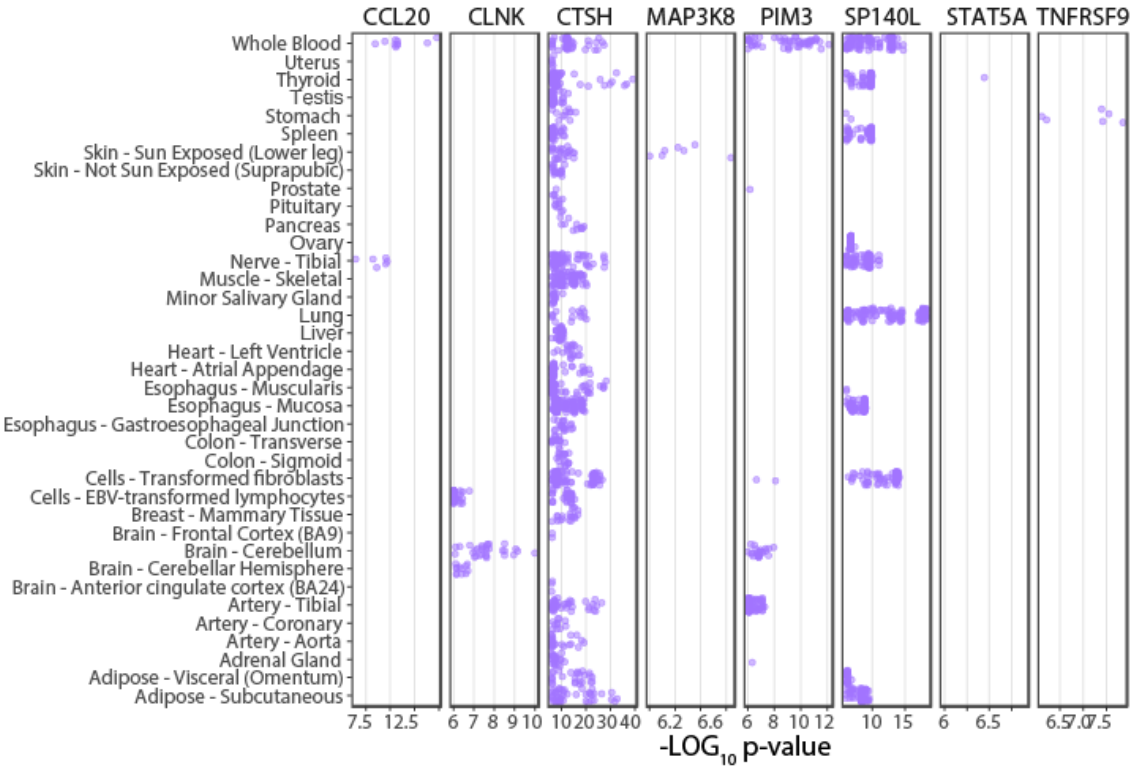

Figure 4.12: **eQTL effects in tissues assayed in GTEx for the immune disease loci with Treg exclusive eQTLs and colocalisations with regQTLs.** Included are only the eQTL signals with p-value $\leq 10^{-6}$.

Table 4.2: **Treg eQTL colocalisations with actQTLs and immune disease GWAS. SNP ID**: Lead GWAS variant colocalising rsid; **Trait**: GWAS immune trait (ALL: allergic disease (asthma, hay fever and eczema); AST: asthma; CD: Crohn's disease; CEL: celiac disease; IBD: inflammatory bowel disease; MS: multiple sclerosis; RA: rheumatoid arthritis; SLE: systemic lupus erythematosus; T1D: type-1 diabetes; UC: ulcerative colitis); **eQTL gene**: symbols of eQTL genes; **actQTL peak co-ords**: peak co-ordinates for actQTLs.

SNP ID	Trait	eQTL gene	actQTL peak co-ords
rs697693	CD	*TNFRSF9*	1:7921757-7945632
rs9658012	CEL	*TNFRSF9*	1:7921757-7945632
rs10746475	IBD	*TNFRSF9*	1:7921757-7945632
rs7523335	UC	*TNFRSF9*	1:7921757-7945632
rs301802	ALL	*RERE*	1:8361298-8433466
rs10826797	IBD	*MAP3K8*	10:30432917-30439043
rs968567	RA	*FADS1*, *FADS2*, *FADS3*, *TMEM258*	11:61831659-61836649
rs58688157	SLE	*IRF7*	11:599606-617445
rs12148472	T1D	*CTSH*	15:78941912-78945542
rs9934775	IBD	*BRD7*, *ADCY7*	16:50262778-50371480
rs59716545	RA	*GSDMB*, *ORMDL3*	17:39912458-39929022
rs12936409	CD, UC, IBD	*GSDMB*, *ORMDL3*	17:39912458-39929022
rs8067378	AST	*GSDMB*, *ORMDL3*	17:39912458-39929022
rs12453507	T1D	*GSDMB*, *ORMDL3*	17:39912458-39929022
17:54863502	IBD	*DGKE*	17:56866409-56871045
rs7207591	ALL	*STAT5A*	17:42219755-42299818
rs4803937	CD	*PPP5C*	19:46345737-46359367
rs11667255	UC	*PTGIR*, *GNG8*, *CALM3*	19:46599219-46627182
rs10175070	UC	*CCL20*	2:227804673-227819641
rs7563433	CD	*SP140*	2:230323771-230338021
rs9989735	MS	*SP140L*, SP140	2:230323771-230338021
rs4343432	CD	*ADCY3*	2:24899368-24923784
rs76286777	IBD	*ADCY3*	2:24870549-24891422
rs137845	UC	*PIM3*	22:49975365-49977811
rs2581828	CD	*RP11-894J14.5*	3:53095151-53133630
rs7660626	RA	*CLNK*	4:10654765-10658912

Continued on next page

Table 4.2 – *Continued from previous page*

SNP ID	Trait	eQTL gene	actQTL peak co-ords
rs6894249	AST	*SLC22A5*	5:132369046-132371617
6:90947340	CD	*BACH2*	6:90264695-90268560
rs57585717	RA	*JAZF1*	7:28168193-28183051
rs4722758	ALL	*JAZF1*	7:28168193-28183051
rs917116	MS	*JAZF1*	7:28120760-28131482
rs11981405	UC	*WBSCR27*, *CLDN4*, *ABHD11*	7:73846994-73854377

I further investigated two of the Treg exclusive colocalising signals to gain a better understanding of how the gene expression and regulation QTL analysis coalesce with the GWAS signals. The approach of combining the colocalisation of GWAS variants with different types of QTL analyses allowed me to prioritise genes and suggest functional variants. An IBD GWAS signal, tagged by the top variant 10:30690376, colocalised with a *MAP3K8* eQTL signal and a 6.13 kbp peak marking the transcription start site of *MAP3K8* actQTL signal (Figure 4.13). The whole locus was characterised by 5 genetic variants in high LD ($R^2 \geq 0.8$) with the top IBD variant and 69 genetic variants in medium LD ($R^2 \geq 0.5$) that were shared between the GWAS summary statistics and the Treg summary statistics. Because not all of the variants from the Treg QTL studies were present in the IBD GWAS summary statistics, I used the LD information, precomputed using the hundred individuals of the Treg cohort, to determine the closest proxy SNPs. The eleven variants that overlapped the actQTL (p-value < 0.001), formed two blocks whereby the first block was in strong LD with the reported GWAS top variant (6 SNPs; LD 0.75) and the second block was in medium LD (5 SNPs; LD 0.5). All the SNPs had p-value < 10^{-11} in the IBD GWAS study. I tested whether any of the SNPs overlapped with H3K4me ChM and ATAC peaks and found that two SNPs overlapped an ATAC peak. Of these only one had p-value < 10^{-6} in both the eQTL (nominal p-value 2.25×10^{-8}) and the actQTL (nominal p-value 1.46×10^{-7}) analysis, 10:30722908. This variant is common (MAF=0.354), and the minor allele C is the protective allele (regression slope = 0.09). It results in decreased levels of acetylation (regression slope = -0.19) and consequently decreased levels of *MAP3K8* expression (regression slope = -0.25).

Another GWAS signal for IBD, UC, CD and CEL colocalised with an eQTL and a trQTL affecting the expression of *TNFRSF9*, and a trQTL of *PARK7*, an adjacent gene encoded on the opposite strand. Additionally, the signal colocalised with an actQTL marking the promoter of *TNFRSF9* and overlapping an H3K4me3 peak (Figure 4.14). There were five SNPs in the actQTL peak that were significant in the IBD GWAS study (p-value < $5x10^{-8}$), in the actQTL study (p-value < 0.0005) and the eQTL study (p-value < 0.001). One of the five variants mapped onto the H3K4me3 peak, 1:7997183. This SNP is also in a target region that interacts with the *PARK7* promoter (bait) detected by promoter capture Hi-C in CD34+ cells and the GM12878 cell line (Mifsud et al., 2015). The minor allele of this SNP, A, was the risk allele (regression slope = 0.12) and resulted in decreased levels of acetylation of the *TNFRSF9* enhancer, and subsequently decreased levels of *TNFRSF9* and *PARK7* expression.

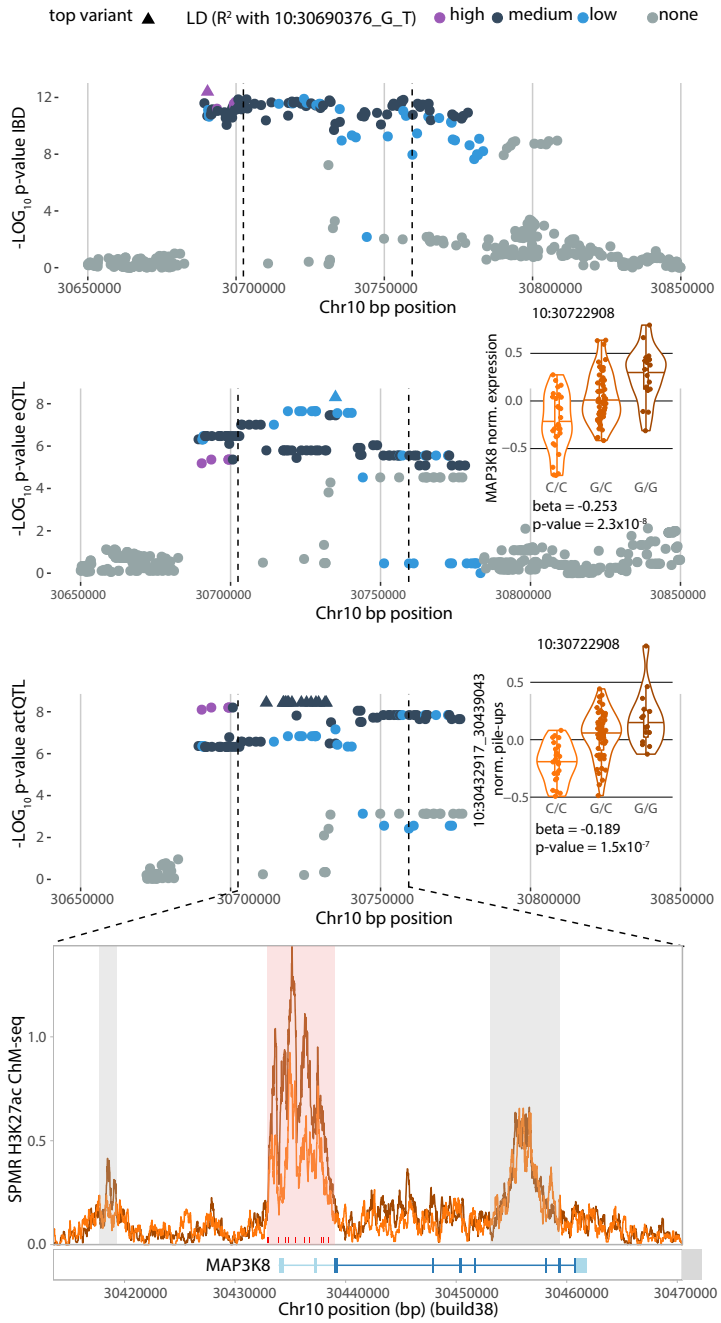

Figure 4.13: **Fine-mapping of a colocalisation signal between an IBD GWAS variant, a *MAP3K8* eQTL and a *MAP3K8* promoter actQTL.** The three colocalising signals are plotted using the same coordinates on the x-axis and the significance for the different variants on the y-axis. Each variant is represented by a dot, colored based on the LD relationship with the top GWAS variant 10:30690376. The lower panel is a focus on the actQTL peak and eQTL gene, there are 11 prioritised SNPs. Of these one variant, 10:30722908 also overlapped with an ATAC peak and was therefore used to plot the expression of the gene and the red pile-ups in the peak. The actQTL peak is highlighted in red, while the remaining called peaks are in grey.

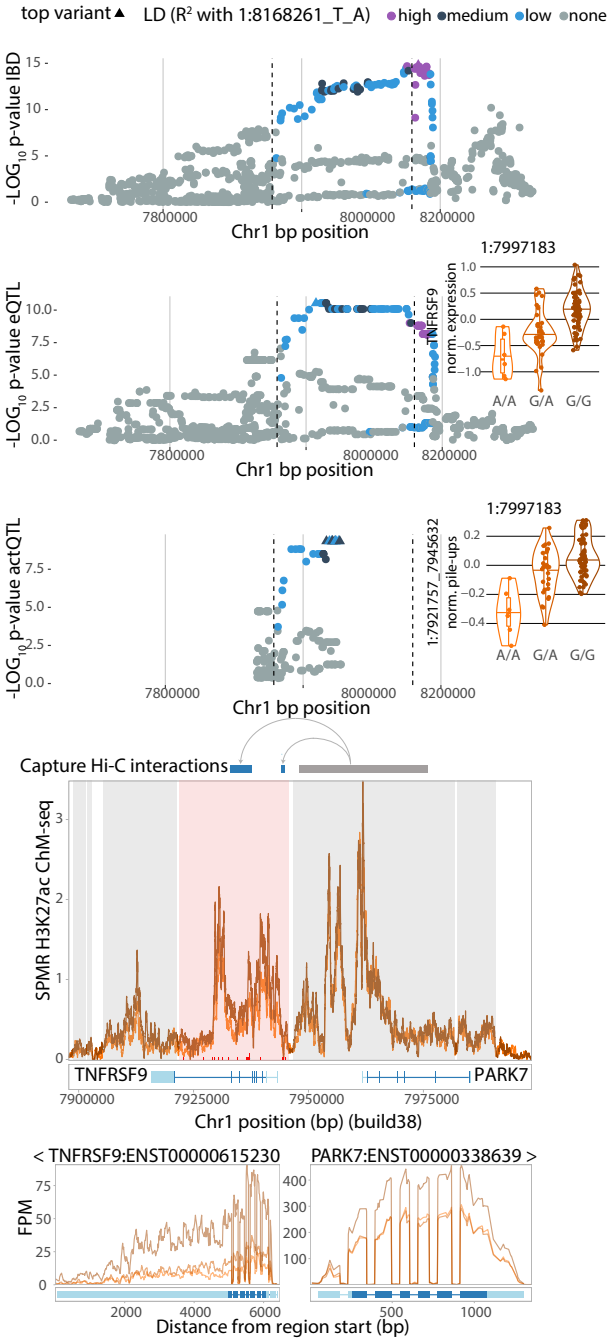

Figure 4.14: **Fine-mapping of a colocalisation signal between the IBD GWAS variants, a *TNFRSF9* eQTL, a *TNFRSF9* and *PARK7* trQTLs and a *TNFRSF9* enhancer actQTL.** The three colocalising signals are plotted using the same coordinates on the x-axis and the significance for the different variants on the y-axis. Each variant is represented by a dot, colored based on the LD relationship with the top GWAS variant 1:8168261. The lower panel is a zoom in on the actQTL peak and the eQTL gene, which prioritised five SNPs. The actQTL peak is highlighted in red, while the remaining called peaks are in grey. The promoter Hi-C interactions between the *PARK7* promoter and regions of *TNFRSF9* are highlighted. The panel below shows the effect of the trQTLs on *TNFRSF9* and *PARK7*. Of the five variants I prioritised a single variant, 1:7997183. I used the alleles of this variant to plot the expression of *TNFRSF9* and the read pile-ups of the actQTL peak.

4.3.7 regQTLs without gene expression QTLs are indicative of condition specific effects

Importantly, in addition to the loci where I was able to link disease associated variants to regQTLs and ultimately gene expression effects, I observed that the highest number of colocalising signals was between the disease variants and actQTLs, without an affected gene. I found that of the 191 LD blocks associated with immune diseases that colocalised with any of the 5 Treg QTLs, 62 were supported by a gene expression effect, while 96 were supported by a QTL in a regulatory mark, and only 33 were supported by both chromatin mark and gene expression (Figure 4.11 B). Since H3K27ac mark is known to be highly cell type specific, this could be indicative that while I assayed the disease relevant cell type, the gene expression could only be manifested in a specific cell state, e.g. upon cell stimulation. In order to assess whether this was the case, I investigated how many of the actQTL peaks had a differentially regulated gene, as identified in Chapter 3, within a 150 kbp window. Of the 78 actQTL loci, 34 contained at least one upregulated gene and 32 contained at least one downregulated gene, with 15 of the loci containing both (Figure 4.15 A). Amongst the upregulated genes in actQTL loci colocalising with immune GWAS, I found many immune relevant genes such as *CTLA4*, *PRDM1*, *LIF* and *LRRC32*.

The actQTL enhancer in the proximity of *LRRC32* has been shown to interact with the promoter of *LRRC32* in the GM12878 cell line and in primary endothelial precursor cells, B cells, macrophages and monocytes (Javierre et al., 2016). This locus colocalises with a GWAS signal from IBD, UC, CD, ALL and AST. In total, 28 SNPs overlapped with the actQTL peak, located 57 kbp downstream of *LRRC32*, of which 10 were significant actQTLs (p-value < 0.0003) (Figure 4.15 B). The lead IBD GWAS variant was located within the peak, and it was an insertion, where the major allele G became GT and correlated with an increase in IBD risk (regression slope = 0.149). Presence of the GT haplotype led to decreased levels of acetylation (regression slope = -0.094) which would probably correspond to decreased levels of *LRRC32* upon stimulation.

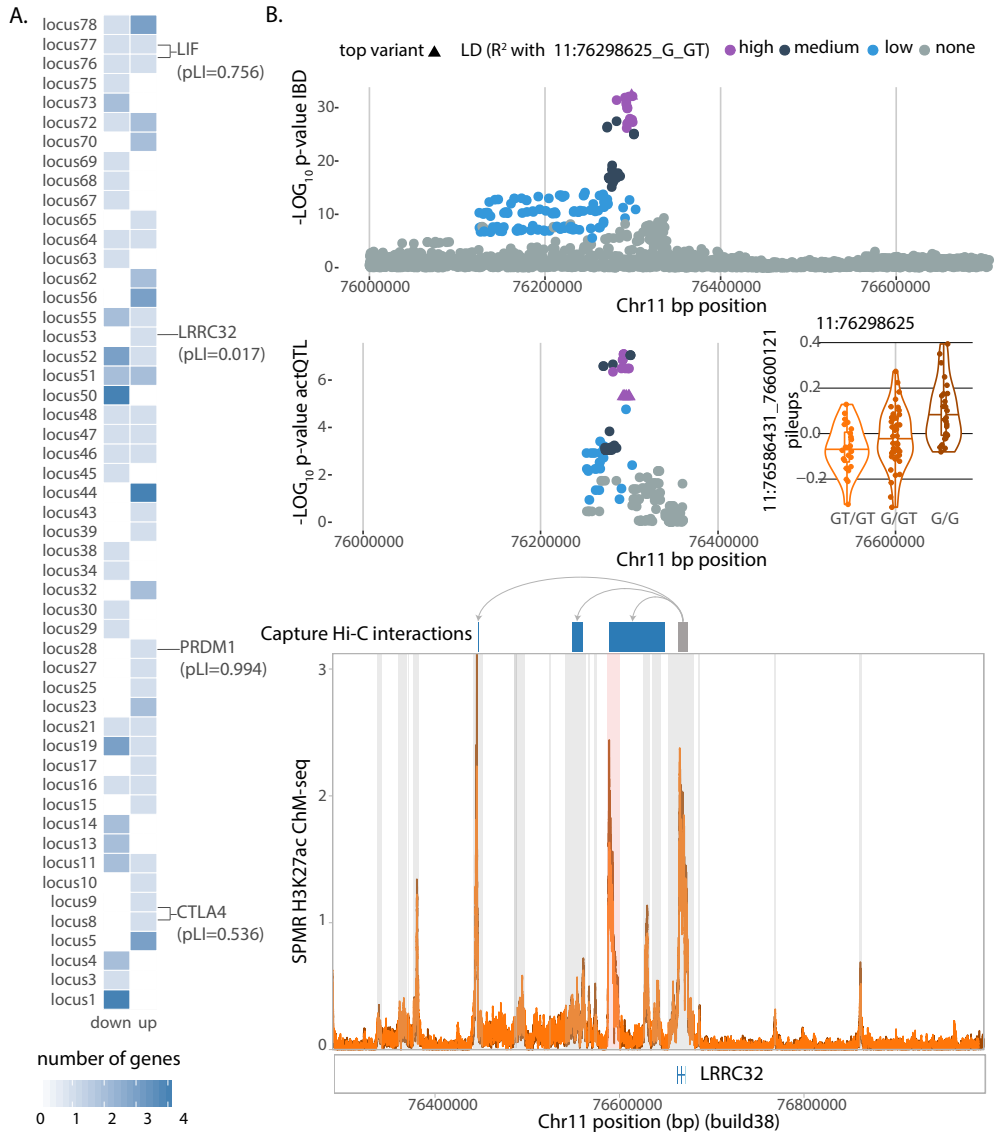

Figure 4.15: **Regulatory QTLs without gene expression QTLs are indicative of context specificity.** **A.** Number of up and downregulated regulated genes per actQTL locus defined by taking at 150kbp window around the start and end co-ordinates of the peak. Highlighted are a few gene examples along with their pLIs. **B.** Fine-mapping of a colocalisation signal between IBD GWAS variants and a *LRRC32* enhancer actQTL. The two colocalising signals are plotted using the same coordinates on the x-axis and the significance for the different variants on the y-axis. Each variant is represented by a dot, coloured based on the LD relationship with the top GWAS variant 11:76298625. The lower panel is the actQTL peak highlighted in red, while the remaining called peaks are in grey. The promoter Hi-C interactions between the *LRRC32* promoter and nearby enhancers are highlighted. I used the alleles of 11:76298625 to plot the read pile-ups of the actQTL peak.

4.4 Discussion

In this chapter I integrated eQTLs and trQTLs mapped in CD4+ naive and regulatory T cells to reveal multiple effects on the regulation of expression of genes associated with immune diseases. Despite the close nature of the two cell types, I observed hundreds of cell type specific effects, many of which play an important role in T cell activation pathways. Given the different roles of naive and regulatory T cells in regulating the immune response, whereby the naive cells exert an effector function upon stimulation by proliferating and secreting pro-inflammatory cytokines while the second returns the system to homeostasis, these differences could be essential to better understand CD4+ T cell biology. For example, I was able to recapitulate a *CTLA4* eQTL effect that was previously reported in whole CD4+ and CD8+ cells (Kasela et al., 2017) specifically in naive T cells. On the other hand, the expression of the *CD28* gene, which is 130 kbp upstream of *CTLA4*, was an eQTL specifically in Tregs. *CTLA4* is a hallmark Treg gene, and is known for its inhibitory role in T cell mediated immune responses by outcompeting CD28 for ligand binding (Read et al., 2000; Zheng et al., 2006). CD28 is the main co-stimulatory receptor found on the surface of all T cells, and its engagement is essential for a successful T cell activation event. Both of these genes sit within the 2q33.2 locus which has been associated to CEL and RA (Trynka et al., 2011; Okada et al., 2014), but the lead variants affecting their expression are not in LD (Raychaudhuri et al., 2009). Interestingly, only the *CD28* eQTL colocalised with a disease variant associated with CEL, where the minor allele T (MAF=0.23) was the risk allele and resulted in decreased levels of *CD28* expression. In fact, while the majority of the observed colocalisations were shared between the two cell types, there were many that were cell type exclusive. This highlights the value of carrying out genomic analysis in isolated rare cell populations.

By performing expression QTL analysis in conjunction with transcript ratio QTL, I was able to determine instances of colocalisation with immune traits in which the observed variability in gene expression could be attributed to an affected transcript. However, these only represented a minority (13/191) of eQTL genes colocalising with GWAS. In my effort to understand how the remaining eQTL effects might arise, I observed that eQTL signals were enriched in active chromatin regions (Gaffney et al., 2012; Pelikan et al., 2018), thereby providing a potential gene expression regulation mechanism for their action. I found that for 33/191 of cases the effect could be attributed to the disruption

of a promoter, enhancer or accessible chromatin, as marked by H3K27ac, H3K4me3 and ATAC-seq. This number is in the same range as previously reported for naive CD4+ T cells (26%), neutrophils (18%) and monocytes (27%), using H3K27ac, H3K4me1 and DNA methylation (Chen et al., 2016). I observed that in 96/191 of cases there was an epigenetic QTL without an immediate effect on the expression of nearby gene. This might be indicative of either long-range interactions which are missed by the current cis-QTL window or that the effect would be manifested in a specific condition. Similar observations have been previously noted for other cells, such as monocytes and induced pluripotent stem cell derived macrophages (Fairfax et al., 2014; Alasoo et al., 2018). Therefore, while the promoters and enhancers are primed for expression, they require an additional stimulus to have an effect on the expression of the gene. In fact, it has been reported that 53% of eQTLs are stimulation specific in monocytes, which is in the same range of the regQTLs reported here without a gene (Fairfax et al., 2014). For example, while no *CTLA4* eQTL colocalisation was observed in naive cells, an actQTL peak colocalised with a lead GWAS variant, 2:204738919, that has been associated with T1D and has been suggested to affect the gene expression levels of *CTLA4* via the disruption of an enhancer (Chen et al., 2016; Westra et al., 2018). Amongst the regQTLs that colocalised with immune disease variants, I also identified an actQTL nearby *LRRC32* which colocalised with variants associated with IBD, allergy and asthma (Lange et al., 2017; Ferreira et al., 2017a; Demenais et al., 2018). *LRRC32* encodes for GARP protein, which binds the latent form of TGF-β and its blockade results in loss of Treg suppressive capacity (Konopacki et al., 2015). *LRRC32* gene expression and protein levels are upregulated upon stimulation (Marson et al., 2007), therefore, it is likely that the disease variants in the enhancer could affect the gene expression of *LRRC32* upon stimulation.

Finally, as an example of successful functional fine-mapping I presented two loci, *TNFRSF9*/*PARK7* and *MAP3K8*, where I observed colocalisation of disease associated variants with gene expression QTLs that were further supported by chromatin regulatory QTLs. *TNFRSF9* gene expression and protein levels (4-1BB or CD137) increase specifically in activated Tregs and not conventional T cells (Marson et al., 2007; Nagar et al., 2010). The expression of CD137 has been reported to correlate with a Treg phenotype by displaying increased suppressor function on effector T cell proliferation (Schoenbrunn et al., 2012), increased FoxP3 expression and epigenetic Treg identity through the demethylation of *FOXP3* and *CTLA4* amongst other Treg genes (Nowak et al., 2018). Therefore, a

QTL that would lead to decreased *TNFRSF9* expression might lead to a decrease in Treg suppressor capacity. *PARK7* also has a documented role in Treg development, where one study demonstrated that *PARK7* knockout mice had dysfunctional induced Tregs, with regards to their proliferative ability, while natural Tregs remained normal (Singh et al., 2015).

While the correlation between the expression of *TNFRSF9* and Treg phenotype are well established (Schoenbrunn et al., 2012), the effect of *MAP3K8* expression in Treg remains a matter of controversy. *MAP3K8* encodes for a serine/threonine kinase, referred to as Tpl-2 or COT, which plays an important function in the processing and signal transduction of the inflammatory cytokine TNF-α (Dumitru et al., 2000). Both a positive and a negative regulation of FoxP3 have been suggested for Tpl-2. One study that used a mouse strain prone to develop intestinal adenoma found that Tpl-2 ablation resulted in increased inflammation-induced intestinal tumorigenesis, which correlated with decreased levels of IL-10 and Treg generation (Serebrennikova et al., 2012). On the other hand, a screen using a luciferase-based reporter system to identify which of 192 kinases could modulate the DNA binding activity of FoxP3 highlighted Tpl-2 as an inducer of Treg instability (Guo et al., 2014).

This chapter provides the first QTL study of Tregs, a rare cell type with an increasingly important role in therapeutic settings. Some of the genes identified and discussed here are already investigated as potential new drug targets. This study could therefore serve a valuable resource for suggesting additional targets for future studies in the context of immune-mediated diseases.

5 Discussion

Collaboration note

Parts of the discussion in this chapter have been published as “Immunogenomic approaches to understand the function of immune disease variants” (Glinos et al., 2017). Many sections of the manuscript have been directly copied into this chapter.

The overarching theme of my research over the past three years has been understanding the gene expression regulation of CD4+ cells. Starting from a fundamental immunological question, the differential requirement of co-stimulation in naive and memory T cells, I set out to identify groups of genes that displayed increased sensitivity to a specific stimulus. I then applied what I had learned in Tregs, but this time I focussed on the mechanisms of gene expression regulation by assessing the effects of stimulation across different gene regulatory layers, as well as their interplay. Finally, I reached a population-scale problem, and set out to study the impact of genetic variation on molecular traits in Tregs. I focussed on the overlap of the molecular genetic effects to gain an understanding of the disturbed immunological pathways and their impact on the development of immune-mediated diseases.

5.1 Mapping QTL effects in rare immune cell types

Over the past few years there have been a number of studies investigating the correlation of quantitative immunogenomic phenotypes with disease associated loci (Nica et al., 2010; Westra et al., 2013). These approaches have allowed to prioritise causal variants (Hormozdiari et al., 2016), explained a proportion of the missing heritability (Gamazon et al., 2018) and suggested disrupted gene pathways and cellular functions. Although the majority of individual allelic effects have a minuscule effect on the overall phenotype, such as a specific disease or height, the effects can be higher on the molecular or cellular levels. Hence, some of the current efforts are focused on identifying the critical disease cell types in which the associated variants are functional.[0.5cm]

A challenge in GWAS functional follow up studies is that often the causal cell types are unknown. Typically, for immune-mediated diseases, cell types can be identified through immunology studies, where a limited set of specific cell markers, both intra- and extracellular, used to characterise different cell populations between disease cases and healthy control subjects. An alternative method that is independent from a predefined set of markers, is single cell RNA sequencing (scRNA-seq). This approach can identify previously unknown heterogeneity in a sample based on gene expression measured at the individual cell level. There are now major international efforts that aim to use scRNA-seq to characterise all human cells, the Human Cell Atlas (Regev and Others, 2016). This resource will provide the most comprehensive annotation of gene expression in the human body which, when integrated with GWAS variants, has the potential to improve our understanding of immune-mediated diseases by carrying single cell QTL studies (Wills et al., 2013; Wijst et al., 2018). Since scRNA-seq outputs gene expression measurements per cell, this approach will allow to group different cells together and carry eQTL analsyses, thereby improve on understanding cell type specific effects.

In Chapter 4 I carried out a QTL study using a rare cell type, regulatory T cells, which only represent 1-2% of total lymphocytes and need to be isolated using reliable cell markers and cell sorting. Tregs, as well as other isolated cell populations, are difficult to culture *in vitro*, rendering the identification of condition specific effects problematic. In addition, in order to reach sufficient statistical power to detect eQTL effects, it is important to obtain blood from a high enough number of individuals. This is not only a laborious task, but also increases the number of variables included in the repetition of blood drawing and processing. Novel systems that allow cell culture and stimulation of cells by limiting the batch effect as well as the establishment of standardised differentiation protocols are necessary for future studies investigating rare cell types. Such approaches are currently being developed by international consortia, such as the Milieu Interieur, who by standardizing the blood culture of healthy volunteers and using flow cytometry analysis have successfully deconvoluted complex immune response signatures and provided valuable insights into the natural variation of immune cells (Urrutia et al., 2016; Patin et al., 2018). Furthermore, as I showed in Chapter 3, the development of reliable genomic protocols that work with low cell numbers is essential when scaling-up to large cohorts (Buenrostro et al., 2013; Schmidl et al., 2015).

The establishment of these standards might facilitate the expansion of QTL studies in patient consortia. These efforts have so far been limited due to the difficulty in discerning the etiology of a disease from the consequences. Furthermore, it is often not possible to find patients in the early stages of the disease that have not been subjected to a drug treatment, an effect that is not well understood. Clinical trials are promising cohorts to study the effects of a drug on gene expression, and the interaction between genotypes and drug effect. There are already a few studies that veer towards this direction. One study investigated the effect of an anti-IL-6 drug in SLE patients (Davenport et al., 2018) and another study looked at the genomic determinants of variation in sepsis patients (Davenport et al., 2016).

5.2 Predicting gene expression in CD4+ T cells

Gene expression regulation is complex and arises from the tight interplay of histone modifications, transcription factor binding and post-transcriptional processes such as small RNAs inhibiting translation. In Chapter 3 I assessed changes in gene expression, transcript ratios, splicing junctions, H3K4me3, H3K27ac and ATAC measurements induced by stimulation in Treg cells. I observed that many of the individual transcript ratios affected by the presence of a stimulus, as well as the differentially spliced introns, were not observed when only examining whole gene counts. This is concordant with past research which showed that 60% of genes express different alternative spliced isoforms in T cells (Ergun et al., 2013). While I did not examine alternative transcription events initiated by CD28 in Chapter 2, this could be an interesting analysis to carry. Past research has shown that changes in alternative transcription events are mostly mediated through co-stimulation (Butte et al., 2012), with CD28 signalling directly affecting the levels of the splicing regulator *hnRNPLL*. The role of histone marks in the regulation of alternative transcription is not well understood, but it has been shown that alternative promoter usage is marked by a change in the levels of H3K4me3 (Luco et al., 2010), but it is important to find a way of reliably validating these findings using spatiotemproal approaches approaches such as RNA fluorescent in-situ hybridization or data driven approaches such as shotgun proteomics (Hu et al., 2015b).

It is also especially interesting to study transcript ratios, in addition to gene expression

measurements, in the context of diseases. In Chapter 4 I observed that eQTLs and trQTLs were largely independent of each other, a finding that has also been shown in past studies (Lappalainen et al., 2013; Alasoo et al., 2018). The colocalisation analysis indicated that trQTLs and eQTLs had a similar enrichment in GWAS loci, despite testing less genes in trQTLs, which has also been previously shown (Li et al., 2016c). When examining patient studies, significant differences in alternative isoforms were detected in multiple sclerosis patients that harboured a SNP in exon 4 of the *PTPRC* gene that encodes for CD45 (Jacobsen et al., 2000) and a switch of the primary alternative transcript used in human islets in response to cytokines (Eizirik et al., 2012). Together these findings highlight the importance of studying alternative transcription events in complex diseases.

The relationship between gene expression and chromatin marks is well documented (Roadmap Epigenomics Consortium et al., 2015). I was able to annotate many genes to differentially active enhancers or promoters in naive, memory and regulatory T cells. I only annotated differentially active elements, however the degree to which different histone marks change upon stimulation is not as well studied as gene expression. This is because histone marks currently lack a reference against which peaks can be called and the combinatorial presence of different modifications within the same locus is not well understood. The establishment of computational protocols would help in the standardisation of these analyses and in the integration of different datasets. The concordance between histone marks and gene expression was reinforced in Chapter 4, where I observed that a third of GWAS colocalising eQTLs could be explained by changes in regQTLs. However, the opposite was not true, and many regQTLs were not assigned an affected gene. By using the differential gene expression analysis from Chapter 3, I was able to deduce a subset of genes that might only be eQTLs upon stimulation. Given the high costs of sequencing, the labour invested in isolating cell types and the wide range of possible stimuli, this provides a potential framework for the identification of a few optimal conditions to carry condition specific eQTL analyses.

QTL effects have in fact mostly been annotated in the context of expression assays, however, a proportion of GWAS variants may not act through gene regulation measured by bulk gene expression assays, calling for the assessment of QTL effects using different assays for gene expression regulation. In Chapter 4 I carried out gene expression

profiling in parallel with multiple assays to understand gene expression regulation. A promising technology is chromatin interaction analysis performed using Hi-C (Belton et al., 2012) and coupled with chromatin immunoprecipitation (Mumbach et al., 2017). This approach has recently been applied on CD4+ T cells, and the authors observed that the majority of H3K27ac peaks do not necessarily interact with the closest gene. This was the case of the *LRRC32* gene that I observed to be regulated by an enhancer located downstream. Furthermore, this method provides a direct mechanism for a genetic variant to act, which might not be obvious when assaying H3K27ac alone. Therefore, results from chromatin conformation assays can be used to help functionally fine map disease associated variants.

5.3 Regulation of co-stimulatory pathways in immune diseases

A recurrent theme across this thesis has been the role of co-stimulation in the initiation of the immune response, and how its perturbation can contribute towards the predisposition to immune diseases. The main co-stimulator is CD28, which assists T cell receptors in the conduction of a successful activation event. CD28 is encoded on the same locus as its main competitor, CTLA-4, which has higher affinity for the same ligands. While CD28 is constitutively expressed on the surface of T cells, CTLA-4 is only expressed upon stimulation, outstripping the activating ligands from the surface of other cells, and allowing the T cell to regain equilibrium. Interestingly, in Chapter 4 I observed that both of these genes are under genetic control. *CD28* was an eQTL in Tregs and *CTLA4* was an eQTLS in naive T cells. This is especially intriguing in the context of my findings in Chapter 2, that memory T cells are more sensitive to the levels of CD28 cross-linking, and the fact that Tregs have many characteristics of memory T cells. This could provide an explanation as to why the *CD28* eQTL is specific to Tregs; since these are these cells are sensitive to its level. Naive T cells on the other hand mostly depend on TCR signalling. As a future experiment, it would be important to carry out proliferation assays in cells isolated from homozygous carriers for each allele affecting the levels of *CD28* and *CTLA4* in order to confirm this hypothesis.

The experimental setting devised in Chapter 2 could be used to assess the impact of genetic variability in co-stimulation sensitivity. This would be especially important

for the design of future therapies that rely on targeting co-stimulatory pathways. The central role of co-stimulation in T cell function has indeed rendered it an attractive target for drug development. Strategies using CD28 and CTLA-4 blockade in animal models have achieved success, paving the way to a number of clinical trials and safe drugs (Ford et al., 2014; Lo et al., 2015). The most notable example is abatacept, a soluble CTLA4-binding domain linked to an Ig region that binds the ligands CD80 and CD86, used for the treatment of rheumatoid arthritis (RA) (Genovese et al., 2005). There are currently more than 60 active clinical trials using CTLA4Ig drugs for treating a range of immune diseases (clinicaltrials.gov). However, targeting CTLA-4 has mainly been successful in immune diseases characterised by an increased Th1 response, probably due to the fact that these cells display increased sensitivity to CD28 blockade (Ford et al., 2014). This highlights the need to increase our understanding of alternative co-stimulatory pathways in a cell type specific context.

It is not only the balance between CD28 and CTLA-4 signalling that determines the activation status of a cell, there are many other co-stimulatory molecules that contribute towards tipping the balance towards a specific response. While I performed the titration experiments by initiating the CD28 pathways, it would be interesting to apply the same framework using a range of stimulants in order to deconvolute the effect of each co-stimulus from the one of TCR presence. However, if performed across all stimulants and titrations, it would be difficult to reach a sufficiently high number of cells per individual and to carry out the assays in parallel. In order to address the first issue, a better measurement might be scRNA-seq, which would also indicate commitment to specific cell fates. In order to address the second problem, a scheme to prioritise co-stimulants might be necessary. In Chapter 4 I observed that the gene expression of more than twenty co-stimulatory molecules and receptors were eQTLs in naive and regulatory T cells. Among them was *CD40*, which is the target for a number of current drugs being developed to attenuate a self-reactive immune response (Ford et al., 2014). Other genes in this list would therefore also present interesting targets for clinical trials with patients that are stratified by genotype. Among the co-stimulatory molecules that were eQTLs in Tregs there were both *TNFRSF9* (encodes for 4-1BB) and its ligand *TNFSF9* (encodes for 4-1BBL). 4-1BB signalling has been shown to be CD28-independent in conventional T cells cells and is able to induce cell division and proliferation (DeBenedette et al., 1997; Cannons et al., 2001). 4-1BBL can expand Tregs *ex vivo* (Elpek et al., 2007) but

also induce resistance to Treg-mediated suppression in conventional T cells (Robertson et al., 2008). There is therefore evidence to support that 4-1BB blockade might be an interesting target for future clinical studies.

5.4 Bridging immunology with genomics - time for proteomics?

However, it remains unclear to what degree the observed correlations between genotype and a quantitative genomic phenotype result in alterations of protein levels. I demonstrated in Chapter 2 a poor replicability of CD28 sensitive genes defined by RNA-seq at the protein level. Furthermore, recent evidence suggests that less than 10% of all mRNA variants could be validated at the protein level, using 29 healthy tissues (Wang et al., 2019). This research was conducted as part of the Human Protein Atlas Project, which aims to map all of the human proteins in cells, tissues and organs using a combination of imaging, mass spectrometry-based proteomics and transcriptomic approaches. If performed across hundreds of individuals this will prove an important resource to carry out protein QTL (pQTL) studies. There are already a number of studies looking at cytokine levels in combination with gene expression (Li et al., 2016b; Ahola-Olli et al., 2017; Bakker et al., 2018) and two pQTL studies, one in human LCLs (Battle et al., 2015) and one in mice (Chick et al., 2016). Bakker and colleagues reaffirmed the importance of context specificity when assessing protein levels, even more so than when looking at RNA. To obviate the need for an exhaustive stimulus assessment for a cytokine QTL study, one can use differential gene expression analysis to prioritise a few conditions that are more likely to result in drastic changes upon stimulation. These approaches will help to understand how immune variants translate to dysfunctional protein products and shed light on the unexplored post-translational mechanisms involved in immune-mediated diseases, which have been suggested to be more important than the RNA levels (Battle et al., 2015). This can be achieved by assessing the phosphoproteome of immune cells, since phosphorylation is especially important in signalling cascades of CD4+ T cells.

Even when a correlation between the relative allele frequency and the level of a protein is established, it is important to place the affected protein levels within a network. The immune system is a highly dynamic network, with complex interactions taking place between the different cell types and states. It has been demonstrated using protein

protein interaction (PPI) assays that proteins encoded by genes in regions associated to RA and Crohn's disease form a network more closely connected compared to a random set of proteins (Rossin et al., 2011). As such, it is becoming increasingly appreciated that disease pathology associated with immune diseases is induced by the disruption of the whole network. Therefore, an interesting future study would be to define PPI networks based on the list of colocalising genes identified in Chapter 4 using purified Tregs for which a reliable bait can be designed.

Mass cytometry has indicated that there is more heterogeneity than previously thought in PBMCs and in different regions of the gut (Unen et al., 2016), which would imply novel undocumented interactions between the different cellular subsets. For example, Rieckmann and colleagues performed an extensive characterization of the proteome and secretome of 28 human immune cell types (Rieckmann et al., 2017) and discovered new signalling events between cells, such as a signalling event induced by IL-34 in CD4+ T cells. Therefore, even if a specific cell type is more relevant for the study of genetic variation, this does not mean that other cell types within the affected tissue would not display differences when compared to healthy tissues. This is why the proportion of different immune subpopulations vary depending on the type of inflammatory diseases (Unen et al., 2016), and while this can be a disease phenotype, it can also be a direct result of a genetic variant. Elucidating mechanistic details of network disruption in autoimmune diseases will contribute towards understanding how hundreds of genetic variants affect different cellular processes and lead to a disease.

5.5 Concluding remarks

GWAS have discovered hundreds of associations with complex immune diseases, yet deciphering the causal variants and molecular mechanisms which give rise to disease has proven to be challenging. Through the accumulation of genetic, functional and immunological data, a picture of how CD4+ T cell regulation is disrupted in immune diseases is starting to emerge. I hope that in this thesis I have clarified some pixels of this image.

Bibliography

Ahola-Olli, Ari V, Peter Würtz, Aki S Havulinna, et al. (2017). “Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors”. en. In: *Am. J. Hum. Genet.* 100.1, pp. 40–50 (quoted on pages 24, 143).

Aken, Bronwen L, Premanand Achuthan, Wasiu Akanni, et al. (2017). “Ensembl 2017”. In: *Nucleic Acids Res.* 45.D1, pp. D635–D642 (quoted on page 13).

Alasoo, Kaur, Julia Rodrigues, Subhankar Mukhopadhyay, et al. (2018). “Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response”. en. In: *Nat. Genet.* 50.3, pp. 424–431 (quoted on pages 15, 135, 140).

Allison, Karmel A, Eniko Sajti, Jana G Collier, et al. (2016). “Affinity and dose of TCR engagement yield proportional enhancer and gene activity in CD4+ T cells”. en. In: *Elife* 5 (quoted on pages 54, 62).

Antonioli, Luca, Pál Pacher, E Sylvester Vizi, and György Haskó (2013). “CD39 and CD73 in immunity and inflammation”. en. In: *Trends Mol. Med.* 19.6, pp. 355–367 (quoted on page 23).

Astle, William J, Heather Elding, Tao Jiang, et al. (2016). “The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease”. en. In: *Cell* 167.5, 1415–1429.e19 (quoted on pages 23, 110, 111).

Avni, Orly, Dong Lee, Fernando Macian, et al. (2002). “T(H) cell differentiation is accompanied by dynamic changes in histone acetylation of cytokine genes”. en. In: *Nat. Immunol.* 3.7, pp. 643–651 (quoted on page 47).

Bakker, Olivier B, Raul Aguirre-Gamboa, Serena Sanna, et al. (2018). “Integration of multi-omics data and deep phenotyping enables prediction of cytokine responses”. In: *Nat. Immunol.* 19.7, pp. 776–786 (quoted on page 143).

Barski, Artem, Raja Jothi, Suresh Cuddapah, et al. (2009). “Chromatin poises miRNA- and protein-coding genes for expression”. en. In: *Genome Res.* 19.10, pp. 1742–1751 (quoted on page 94).

Barski, Artem, Suresh Cuddapah, Andrey V Kartashov, et al. (2017). “Rapid Recall Ability of Memory T cells is Encoded in their Epigenome”. en. In: *Sci. Rep.* 7, p. 39785 (quoted on page 63).

Battle, Alexis, Zia Khan, Sidney H Wang, et al. (2015). “Impact of regulatory variation from RNA to protein”. en. In: *Science* 347.6222, pp. 664–667 (quoted on page 143).

Belton, Jon-Matthew, Rachel Patton McCord, Johan Harmen Gibcus, et al. (2012). “Hi-C: a comprehensive technique to capture the conformation of genomes”. en. In: *Methods* 58.3, pp. 268–276 (quoted on pages 21, 141).

Benjamini, Yoav and Yosef Hochberg (1995). “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”. In: *J. R. Stat. Soc. Series B Stat. Methodol.* 57.1, pp. 289–300 (quoted on pages 35, 73–75).

Benjamini, Yuval and Terence P Speed (2012). “Summarizing and correcting the GC content bias in high-throughput sequencing”. en. In: *Nucleic Acids Res.* 40.10, e72 (quoted on page 12).

Bennett, C L, J Christie, F Ramsdell, et al. (2001). “The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3”. en. In: *Nat. Genet.* 27.1, pp. 20–21 (quoted on pages 10, 67).

Bentham, James, David L Morris, Deborah S Cunninghame Graham, et al. (2015). “Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus”. en. In: *Nat. Genet.* 47.12, pp. 1457–1464 (quoted on page 110).

Bertrand, Anne, Marie Kostine, Thomas Barnetche, Marie-Elise Truchetet, and Thierry Schaeverbeke (2015). “Immune related adverse events associated with anti-CTLA-4 antibodies: systematic review and meta-analysis”. en. In: *BMC Med.* 13, p. 211 (quoted on page 65).

Bhairavabhotla, Ravikiran, Yong C Kim, Deborah D Glass, et al. (2016). “Transcriptome profiling of human FoxP3+ regulatory T cells”. en. In: *Hum. Immunol.* 77.2, pp. 201–213 (quoted on pages 99, 104).

Birzele, Fabian, Tanja Fauti, Heiko Stahl, et al. (2011). “Next-generation insights into regulatory T cells: expression profiling and FoxP3 occupancy in Human”. en. In: *Nucleic Acids Res.* 39.18, pp. 7946–7960 (quoted on pages 68, 99–101, 104).

Bonder, Marc Jan, René Luijk, Daria V Zhernakova, et al. (2017). “Disease variants alter transcription factor levels and methylation of their binding sites”. en. In: *Nat. Genet.* 49.1, pp. 131–138 (quoted on page 22).

Borowski, Annie B, Alina C Boesteanu, Yvonne M Mueller, et al. (2007). “Memory CD8+ T cells require CD28 costimulation”. en. In: *J. Immunol.* 179.10, pp. 6494–6503 (quoted on pages 9, 28).

Borriello, F, M P Sethna, S D Boyd, et al. (1997). “B7-1 and B7-2 have overlapping, critical roles in immunoglobulin class switching and germinal center formation”. en. In: *Immunity* 6.3, pp. 303–313 (quoted on page 40).

Bossini-Castillo, Lara, Elena López-Isac, Maureen D Mayes, and Javier Martín (2015). “Genetics of systemic sclerosis”. en. In: *Semin. Immunopathol.* 37.5, pp. 443–451 (quoted on page 39).

Bovenschen, H Jorn, Peter C van de Kerkhof, Piet E van Erp, et al. (2011). “Foxp3+ regulatory T cells of psoriasis patients easily differentiate into IL-17A-producing cells and are found in lesional skin”. en. In: *J. Invest. Dermatol.* 131.9, pp. 1853–1860 (quoted on page 12).

Boyle, Alan P, Sean Davis, Hennady P Shulha, et al. (2008). “High-resolution mapping and characterization of open chromatin across the genome”. en. In: *Cell* 132.2, pp. 311–322 (quoted on page 15).

Brodin, Petter, Vladimir Jojic, Tianxiang Gao, et al. (2015). “Variation in the human immune system is largely driven by non-heritable influences”. en. In: *Cell* 160.1-2, pp. 37–47 (quoted on pages 22, 23).

Brown, Chrysothemis C, Daria Esterhazy, Aurelien Sarde, et al. (2015). “Retinoic acid is essential for Th1 cell lineage stability and prevents transition to a Th17 cell program”. en. In: *Immunity* 42.3, pp. 499–511 (quoted on page 91).

Browning, Brian L, Ying Zhou, and Sharon R Browning (2018). "A One-Penny Imputed Genome from Next-Generation Reference Panels". en. In: *Am. J. Hum. Genet.* 103.3, pp. 338–348 (quoted on page 107).

Buckner, Jane Hoyt (2010). "Mechanisms of impaired regulation by CD4+ CD25+ FOXP3+ regulatory T cells in human autoimmune diseases". In: *Nat. Rev. Immunol.* 10.12, pp. 849–859 (quoted on pages 10, 16, 68, 99, 104).

Buenrostro, Jason D, Paul G Giresi, Lisa C Zaba, Howard Y Chang, and William J Greenleaf (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position". en. In: *Nat. Methods* 10.12, pp. 1213–1218 (quoted on pages 15, 37, 70, 138).

Butte, Manish J, Sun Jung Lee, Jonathan Jesneck, et al. (2012). "CD28 Costimulation Regulates Genome-Wide Effects on Alternative Splicing". In: *PLoS One* 7.6, e40032 (quoted on page 139).

Calderon, Diego, Anand Bhaskar, David A Knowles, et al. (2017). "Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression". en. In: *Am. J. Hum. Genet.* 101.5, pp. 686–699 (quoted on page 16).

CAMMS223 Trial Investigators, Alasdair J Coles, D Alastair S Compston, et al. (2008). "Alemtuzumab vs. interferon beta-1a in early multiple sclerosis". en. In: *N. Engl. J. Med.* 359.17, pp. 1786–1801 (quoted on page 18).

Cannons, J L, P Lau, B Ghumman, et al. (2001). "4-1BB ligand induces cell division, sustains survival, and enhances effector function of CD4 and CD8 T cells with similar efficacy". en. In: *J. Immunol.* 167.3, pp. 1313–1324 (quoted on page 142).

Chang, John T, Vikram R Palanivel, Ichiko Kinjyo, et al. (2007). "Asymmetric T lymphocyte division in the initiation of adaptive immune responses". en. In: *Science* 315.5819, pp. 1687–1691 (quoted on page 9).

Chen, Lieping and Dallas B Flies (2013). "Molecular mechanisms of T cell co-stimulation and co-inhibition". en. In: *Nat. Rev. Immunol.* 13.4, pp. 227–242 (quoted on pages 52, 62).

Chen, Lu, Bing Ge, Francesco Paolo Casale, et al. (2016). "Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells". en. In: *Cell* 167.5, 1398–1414.e24 (quoted on pages 20, 105, 113, 135).

Chick, Joel M, Steven C Munger, Petr Simecek, et al. (2016). "Defining the consequences of genetic variation on a proteome-wide scale". en. In: *Nature* 534.7608, pp. 500–505 (quoted on page 143).

Chrabot, B S, S N Kariuki, M I Zervou, et al. (2013). "Genetic variation near IRF8 is associated with serologic and cytokine profiles in systemic lupus erythematosus and multiple sclerosis". en. In: *Genes Immun.* 14.8, pp. 471–478 (quoted on page 20).

Chun, Sung, Alexandra Casparino, Nikolaos A Patsopoulos, et al. (2017). "Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types". en. In: *Nat. Genet.* 49.4, pp. 600–605 (quoted on page 18).

Consortium, International Multiple Sclerosis Genetics and Others (2013). "Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis". In: *Nat. Genet.* 45.11, pp. 1353–1360 (quoted on page 3).

Cordell, Heather J, Younghun Han, George F Mells, et al. (2015). "International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways". en. In: *Nat. Commun.* 6, p. 8019 (quoted on page 110).

Criswell, Lindsey A, Kirsten A Pfeiffer, Raymond F Lum, et al. (2005). “Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes”. en. In: *Am. J. Hum. Genet.* 76.4, pp. 561–571 (quoted on page 3).

Croft, David, Antonio Fabregat Mundo, Robin Haw, et al. (2014). “The Reactome pathway knowledgebase”. en. In: *Nucleic Acids Res.* 42.Database issue, pp. D472–7 (quoted on pages 74, 84).

Croft, M, L M Bradley, and S L Swain (1994). “Naive versus memory CD4 T cell response to antigen. Memory cells are less dependent on accessory cell costimulation and can respond to many antigen-presenting cell types including resting B cells”. en. In: *J. Immunol.* 152.6, pp. 2675–2685 (quoted on pages 9, 62).

Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). “Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis”. en. In: *Lancet* 381.9875, pp. 1371–1379 (quoted on page 119).

Davenport, Emma E, Katie L Burnham, Jayachandran Radhakrishnan, et al. (2016). “Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study”. en. In: *Lancet Respir Med* 4.4, pp. 259–271 (quoted on page 139).

Davenport, Emma E, Tiffany Amariuta, Maria Gutierrez-Arcelus, et al. (2018). “Discovering in vivo cytokine-eQTL interactions from a lupus clinical trial”. en. In: *Genome Biol.* 19.1, p. 168 (quoted on page 139).

De Jager, Philip L, Xiaoming Jia, Joanne Wang, et al. (2009). “Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci”. en. In: *Nat. Genet.* 41.7, pp. 776–782 (quoted on page 20).

DeBenedette, M A, A Shahinian, T W Mak, and T H Watts (1997). “Costimulation of CD28- T lymphocytes by 4-1BB ligand”. en. In: *J. Immunol.* 158.2, pp. 551–559 (quoted on page 142).

Delaneau, Olivier, Halit Ongen, Andrew A Brown, et al. (2017). “A complete tool set for molecular QTL discovery and analysis”. In: *Nat. Commun.* 8, p. 15452 (quoted on pages 109, 112).

Demenais, Florence, Patricia Margaritte-Jeannin, Kathleen C Barnes, et al. (2018). “Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks”. en. In: *Nat. Genet.* 50.1, pp. 42–53 (quoted on pages 110, 135).

Diehn, Maximilian, Ash A Alizadeh, Oliver J Rando, et al. (2002). “Genomic expression programs and the integration of the CD28 costimulatory signal in T cell activation”. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 99.18, pp. 11796–11801 (quoted on page 62).

Dimas, Antigone S, Samuel Deutsch, Barbara E Stranger, et al. (2009). “Common regulatory variation impacts gene expression in a cell type-dependent manner”. en. In: *Science* 325.5945, pp. 1246–1250 (quoted on pages 17, 18).

Djilali-Saiah, I, J Schmitz, E Harfouch-Hammoud, et al. (1998). “CTLA-4 gene polymorphism is associated with predisposition to coeliac disease”. en. In: *Gut* 43.2, pp. 187–189 (quoted on page 3).

Dobbins, Jessica, Etienne Gagnon, Jernej Godec, et al. (2016). “Binding of the cytoplasmic domain of CD28 to the plasma membrane inhibits Lck recruitment and signaling”. en. In: *Sci. Signal.* 9.438, ra75–ra75 (quoted on page 6).

Dobin, Alexander, Carrie A Davis, Felix Schlesinger, et al. (2013). “STAR: ultrafast universal RNA-seq aligner”. en. In: *Bioinformatics* 29.1, pp. 15–21 (quoted on pages 13, 35, 73, 111).

Doody, Gina M, Matthew A Care, Nicholas J Burgoyne, et al. (2010). “An extended set of PRDM1/BLIMP1 target genes links binding motif type to dynamic repression”. en. In: *Nucleic Acids Res.* 38.16, pp. 5336–5350 (quoted on page 57).

Dubey, C, M Croft, and S L Swain (1995). “Costimulatory requirements of naive CD4+ T cells. ICAM-1 or B7-1 can costimulate naive CD4 T cell activation but both are required for optimum response”. en. In: *J. Immunol.* 155.1, pp. 45–57 (quoted on pages 9, 62).

Dubois, Patrick C A, Gosia Trynka, Lude Franke, et al. (2010). “Multiple common variants for celiac disease influencing immune gene expression”. en. In: *Nat. Genet.* 42.4, pp. 295–302 (quoted on pages 17, 28, 104).

Duhen, Thomas, Rebekka Duhen, Antonio Lanzavecchia, Federica Sallusto, and Daniel J Campbell (2012). “Functionally distinct subsets of human FOXP3+ Treg cells that phenotypically mirror effector Th cells”. en. In: *Blood* 119.19, pp. 4430–4440 (quoted on page 99).

Dumitru, C D, J D Ceci, C Tsatsanis, et al. (2000). “TNF-alpha induction by LPS is regulated posttranscriptionally via a Tpl2/ERK-dependent pathway”. en. In: *Cell* 103.7, pp. 1071–1083 (quoted on page 136).

Eastwood, D, L Findlay, S Poole, et al. (2010). “Monoclonal antibody TGN1412 trial failure explained by species differences in CD28 expression on CD4+ effector memory T-cells”. en. In: *Br. J. Pharmacol.* 161.3, pp. 512–526 (quoted on page 28).

Eaton, William W, Noel R Rose, Amanda Kalaydjian, Marianne G Pedersen, and Preben Bo Mortensen (2007). “Epidemiology of autoimmune diseases in Denmark”. en. In: *J. Autoimmun.* 29.1, pp. 1–9 (quoted on page 2).

Edmead, C E, Y I Patel, A Wilson, et al. (1996). “Induction of activator protein (AP)-1 and nuclear factor-kappaB by CD28 stimulation involves both phosphatidylinositol 3-kinase and acidic sphingomyelinase signals”. en. In: *J. Immunol.* 157.8, pp. 3290–3297 (quoted on page 57).

Eizirik, Décio L, Michael Sammeth, Thomas Bouckenooghe, et al. (2012). “The Human Pancreatic Islet Transcriptome: Expression of Candidate Genes for Type 1 Diabetes and the Impact of Pro-Inflammatory Cytokines”. In: *PLoS Genet.* 8.3, e1002552 (quoted on page 140).

Elpek, Kutlu G, Esma S Yolcu, Deanna D H Franke, et al. (2007). “Ex vivo expansion of CD4+CD25+FoxP3+ T regulatory cells based on synergy between IL-2 and 4-1BB signaling”. en. In: *J. Immunol.* 179.11, pp. 7295–7304 (quoted on page 142).

ENCODE Project Consortium (2012). “An integrated encyclopedia of DNA elements in the human genome”. en. In: *Nature* 489.7414, pp. 57–74 (quoted on page 14).

Ergun, Ayla, Graeme Doran, James C Costello, et al. (2013). “Differential splicing across immune system lineages”. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 110.35, pp. 14324–14329 (quoted on page 139).

Eun, So-Young, Seung-Woo Lee, Yanfei Xu, and Michael Croft (2015). “4-1BB ligand signaling to T cells limits T cell activation”. en. In: *J. Immunol.* 194.1, pp. 134–141 (quoted on page 100).

Fairfax, Benjamin P, Peter Humburg, Seiko Makino, et al. (2014). “Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression”. en. In: *Science* 343.6175, p. 1246949 (quoted on pages 17, 19, 29, 135).

Farh, Kyle Kai-How, Alexander Marson, Jiang Zhu, et al. (2015). “Genetic and epigenetic fine mapping of causal autoimmune disease variants”. en. In: *Nature* 518.7539, pp. 337–343 (quoted on pages 15, 21, 58).

Ferreira, Manuel A, Judith M Vonk, Hansjörg Baurecht, et al. (2017a). “Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology”. In: *Nat. Genet.* 49, p. 1752 (quoted on pages 110, 135).

Ferreira, Manuel A R, Massimo Mangino, Chanson J Brumme, et al. (2010). “Quantitative trait loci for CD4:CD8 lymphocyte ratio are associated with risk of type 1 diabetes and HIV-1 immune control”. en. In: *Am. J. Hum. Genet.* 86.1, pp. 88–92 (quoted on page 22).

Ferreira, Ricardo C, Daniel B Rainbow, Arcadio Rubio García, et al. (2017b). “Human IL-6R hi TIGIT- CD4+ CD127 low CD25+ T cells display potent in vitro suppressive capacity and a distinct Th17 profile”. In: *Clin. Immunol.* (Quoted on page 10).

Finucane, Hilary K, Brendan Bulik-Sullivan, Alexander Gusev, et al. (2015). “Partitioning heritability by functional annotation using genome-wide association summary statistics”. In: *Nat. Genet.* 47, p. 1228 (quoted on page 104).

Fontenot, Jason D, Jeffrey P Rasmussen, Marc A Gavin, and Alexander Y Rudensky (2005a). “A function for interleukin 2 in Foxp3-expressing regulatory T cells”. en. In: *Nat. Immunol.* 6.11, pp. 1142–1151 (quoted on page 10).

Fontenot, Jason D, Jeffrey P Rasmussen, Luke M Williams, et al. (2005b). “Regulatory T cell lineage specification by the forkhead transcription factor foxp3”. en. In: *Immunity* 22.3, pp. 329–341 (quoted on pages 8, 10, 67).

Ford, Mandy L, Andrew B Adams, and Thomas C Pearson (2014). “Targeting co-stimulatory pathways: transplantation and autoimmunity”. en. In: *Nat. Rev. Nephrol.* 10.1, pp. 14–24 (quoted on pages 6, 142).

Fortune, Mary D, Hui Guo, Oliver Burren, et al. (2015). “Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls”. en. In: *Nat. Genet.* 47.7, pp. 839–846 (quoted on pages 4, 28).

Fraser, J D, B A Irving, G R Crabtree, and A Weiss (1991). “Regulation of interleukin-2 gene enhancer activity by the T cell accessory molecule CD28”. en. In: *Science* 251.4991, pp. 313–316 (quoted on pages 7, 57).

Fröhlich, Monika, Tea Gogishvili, Daniela Langenhorst, Fred Lühder, and Thomas Hünig (2016). “Interrupting CD28 costimulation before antigen rechallenge affects CD8(+) T-cell expansion and effector functions during secondary response in mice”. en. In: *Eur. J. Immunol.* 46.7, pp. 1644–1655 (quoted on pages 9, 28, 63).

Gaffney, Daniel J, Jean-Baptiste Veyrieras, Jacob F Degner, et al. (2012). “Dissecting the regulatory architecture of gene expression QTLs”. en. In: *Genome Biol.* 13.1, R7 (quoted on page 134).

Gamazon, Eric R, Ayellet V Segrè, Martijn van de Bunt, et al. (2018). “Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation”. In: *Nat. Genet.* 50.7, pp. 956–967 (quoted on page 137).

Garidou, Lucile, Sara Heydari, Phi Truong, David G Brooks, and Dorian B McGavern (2009). “Therapeutic Memory T Cells Require Costimulation for Effective Clearance of a Persistent Viral Infection”. en. In: *J. Virol.* 83.17, pp. 8905–8915 (quoted on page 28).

Gate, Rachel E, Christine S Cheng, Aviva P Aiden, et al. (2018). “Genetic determinants of co-accessible chromatin regions in activated T cells across humans”. en. In: *Nat. Genet.* 50.8, pp. 1140–1150 (quoted on pages 15, 105).

Genovese, Mark C, Jean-Claude Becker, Michael Schiff, et al. (2005). “Abatacept for rheumatoid arthritis refractory to tumor necrosis factor alpha inhibition”. en. In: *N. Engl. J. Med.* 353.11, pp. 1114–1123 (quoted on page 142).

Giambartolomei, Claudia, Damjan Vukcevic, Eric E Schadt, et al. (2014). “Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics”. In: *PLoS Genet.* 10.5, e1004383 (quoted on pages 110, 118).

Glinos, Dafni A, Blagoje Soskic, and Gosia Trynka (2017). “Immunogenomic approaches to understand the function of immune disease variants”. en. In: *Immunology* 152.4, pp. 527–535 (quoted on pages 1, 137).

Graham, Deborah S Cunninghame, Robert R Graham, Harinder Manku, et al. (2007). “Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus”. In: *Nat. Genet.* 40, p. 83 (quoted on page 11).

Grimbert, Philippe, Salim Bouguermouh, Nobuyasu Baba, et al. (2006). “Thrombospondin/CD47 interaction: a pathway to generate regulatory T cells from human CD4+ CD25- T cells in response to inflammation”. en. In: *J. Immunol.* 177.6, pp. 3534–3541 (quoted on page 43).

GTEx Consortium (2015). “Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans”. en. In: *Science* 348.6235, pp. 648–660 (quoted on pages 17, 126).

Guo, Hui, Mary D Fortune, Oliver S Burren, et al. (2015). “Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases”. en. In: *Hum. Mol. Genet.* 24.12, pp. 3305–3313 (quoted on pages 18, 104, 118).

Guo, Jitao, Jianhua Zhang, Xuejie Zhang, et al. (2014). “Constitutive activation of MEK1 promotes Treg cell instability in vivo”. en. In: *J. Biol. Chem.* 289.51, pp. 35139–35148 (quoted on page 136).

Gutierrez-Arcelus, Maria, Stephen S Rich, and Soumya Raychaudhuri (2016). “Autoimmune diseases - connecting risk alleles with molecular traits of the immune system”. en. In: *Nat. Rev. Genet.* 17.3, pp. 160–174 (quoted on page 3).

Hall, M A, K R Ahmadi, P Norman, et al. (2000). “Genetic influence on peripheral blood T lymphocyte levels”. en. In: *Genes Immun.* 1.7, pp. 423–427 (quoted on page 22).

Hansen, Kasper D, Rafael A Irizarry, and Zhijin Wu (2012). “Removing technical variability in RNA-seq data using conditional quantile normalization”. en. In: *Biostatistics* 13.2, pp. 204–216 (quoted on page 108).

Harrow, Jennifer, Adam Frankish, Jose M Gonzalez, et al. (2012). “GENCODE: the reference human genome annotation for The ENCODE Project”. en. In: *Genome Res.* 22.9, pp. 1760–1774 (quoted on pages 13, 94).

Haufe, Susanne, Markus Haug, Carsten Schepp, et al. (2011). “Impaired suppression of synovial fluid CD4+CD25- T cells from patients with juvenile idiopathic arthritis by CD4+CD25+ Treg cells”. en. In: *Arthritis Rheum.* 63.10, pp. 3153–3162 (quoted on pages 11, 28).

Hayter, Scott M and Matthew C Cook (2012). “Updated assessment of the prevalence, spectrum and case definition of autoimmune disease”. en. In: *Autoimmun. Rev.* 11.10, pp. 754–765 (quoted on page 2).

He, Xuehui, Ruben L Smeets, Esther van Rijssen, et al. (2017). “Single CD28 stimulation induces stable and polyclonal expansion of human regulatory T cells”. en. In: *Sci. Rep.* 7, p. 43003 (quoted on page 63).

Hedrick, S M, D I Cohen, E A Nielsen, and M M Davis (1984). “Isolation of cDNA clones encoding T cell-specific membrane-associated proteins”. en. In: *Nature* 308.5955, pp. 149–153 (quoted on page 7).

Hormozdiari, Farhad, Martijn van de Bunt, Ayellet V Segrè, et al. (2016). “Colocalization of GWAS and eQTL Signals Detects Target Genes”. en. In: *Am. J. Hum. Genet.* 99.6, pp. 1245–1260 (quoted on page 137).

Howard, David M, Mark J Adams, Masoud Shirali, et al. (2018). “Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways”. en. In: *Nat. Commun.* 9.1, p. 1470 (quoted on page 111).

Hu, Xinli, Hyun Kim, Eli Stahl, et al. (2011). “Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets”. en. In: *Am. J. Hum. Genet.* 89.4, pp. 496–506 (quoted on page 16).

Hu, Xinli, Hyun Kim, Towfique Raj, et al. (2014). “Regulation of gene expression in autoimmune disease loci and the genetic basis of proliferation in CD4+ effector memory T cells”. en. In: *PLoS Genet.* 10.6, e1004404 (quoted on pages 17, 58, 104).

Hu, Xinli, Aaron J Deutsch, Tobias L Lenz, et al. (2015a). “Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk”. en. In: *Nat. Genet.* 47.8, pp. 898–905 (quoted on pages 2, 3).

Hu, Zhiqiang, Hamish S Scott, Guangrong Qin, et al. (2015b). “Revealing Missing Human Protein Isoforms Based on Ab Initio Prediction, RNA-seq and Proteomics”. In: *Sci. Rep.* 5, p. 10940 (quoted on page 139).

Huang, Hailiang, Ming Fang, Luke Jostins, et al. (2017). “Fine-mapping inflammatory bowel disease loci to single-variant resolution”. en. In: *Nature* 547.7662, pp. 173–178 (quoted on page 18).

Hui, Enfu, Jeanne Cheung, Jing Zhu, et al. (2017). “T cell costimulatory receptor CD28 is a primary target for PD-1-mediated inhibition”. en. In: *Science* 355.6332, pp. 1428–1433 (quoted on page 65).

Hünig, Thomas (2012). “The storm has cleared: lessons from the CD28 superagonist TGN1412 trial”. en. In: *Nat. Rev. Immunol.* 12.5, pp. 317–318 (quoted on pages 28, 63).

Hwang, Soo Seok, Young Uk Kim, Sumin Lee, et al. (2013). “Transcription factor YY1 is essential for regulation of the Th2 cytokine locus and for Th2 cell differentiation”. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 110.1, pp. 276–281 (quoted on page 102).

Hwang, Soo Seok, Sung Woong Jang, Min Kyung Kim, et al. (2016). “YY1 inhibits differentiation and function of regulatory T cells by blocking Foxp3 expression and activity”. In: *Nat. Commun.* 7, p. 10789 (quoted on page 102).

International Multiple Sclerosis Genetics Consortium (IMSGC), Ashley H Beecham, Nikolaos A Patsopoulos, et al. (2013). “Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis”. en. In: *Nat. Genet.* 45.11, pp. 1353–1360 (quoted on pages 39, 110).

Ishihara, Katsuhiko and Toshio Hirano (2002). “IL-6 in autoimmune disease and chronic inflammatory proliferative disease”. en. In: *Cytokine Growth Factor Rev.* 13.4-5, pp. 357–368 (quoted on pages 11, 23).

Jacobsen, M, D Schweer, A Ziegler, et al. (2000). “A point mutation in PTPRC is associated with the development of multiple sclerosis”. en. In: *Nat. Genet.* 26.4, pp. 495–499 (quoted on page 140).

Jacquemin, Clément, Nathalie Schmitt, Cécile Contin-Bordes, et al. (2015). “OX40 Ligand Contributes to Human Lupus Pathogenesis by Promoting T Follicular Helper Response”. en. In: *Immunity* 42.6, pp. 1159–1170 (quoted on page 11).

Javierre, Biola M, Oliver S Burren, Steven P Wilder, et al. (2016). “Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters”. en. In: *Cell* 167.5, 1369–1384.e19 (quoted on pages 21, 39, 73, 78, 132).

Jenkins, M K, J D Ashwell, and R H Schwartz (1988). “Allogeneic non-T spleen cells restore the responsiveness of normal T cell clones stimulated with antigen and chemically modified antigen-presenting cells”. en. In: *J. Immunol.* 140.10, pp. 3324–3330 (quoted on page 7).

Jenkins, M K, P S Taylor, S D Norton, and K B Urdahl (1991). “CD28 delivers a costimulatory signal involved in antigen-specific IL-2 production by human T cells”. en. In: *J. Immunol.* 147.8, pp. 2461–2466 (quoted on page 28).

Jiang, Hongshan, Rong Lei, Shou-Wei Ding, and Shuifang Zhu (2014). “Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads”. en. In: *BMC Bioinformatics* 15, p. 182 (quoted on pages 37, 71).

Jostins, Luke, Stephan Ripke, Rinse K Weersma, et al. (2012). “Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease”. en. In: *Nature* 491.7422, pp. 119–124 (quoted on page 39).

Jun, Goo, Matthew Flickinger, Kurt N Hetrick, et al. (2012). “Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data”. en. In: *Am. J. Hum. Genet.* 91.5, pp. 839–848 (quoted on page 108).

Kamphorst, Alice O, Andreas Wieland, Tahseen Nasti, et al. (2017). “Rescue of exhausted CD8 T cells by PD-1-targeted therapies is CD28-dependent”. en. In: *Science* 355.6332, pp. 1423–1427 (quoted on page 65).

Kasela, Silva, Kai Kisand, Liina Tserel, et al. (2017). “Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells”. en. In: *PLoS Genet.* 13.3, e1006643 (quoted on pages 18, 104, 134).

Kennedy, Alan, Emily M Schmidt, Adam P Cribbs, et al. (2014). “A novel upstream enhancer of FOXP3, sensitive to methylation-induced silencing, exhibits dysregulated methylation in rheumatoid arthritis Treg cells”. en. In: *Eur. J. Immunol.* 44.10, pp. 2968–2978 (quoted on page 101).

Khattri, R, J A Auger, M D Griffin, A H Sharpe, and J A Bluestone (1999). “Lymphoproliferative disorder in CTLA-4 knockout mice is characterized by CD28-regulated activation of Th2 responses”. en. In: *J. Immunol.* 162.10, pp. 5784–5791 (quoted on page 64).

Kimmig, Sonja, Grzegorz K Przybylski, Christian A Schmidt, et al. (2002). “Two subsets of naive T helper cells with distinct T cell receptor excision circle content in human adult peripheral blood”. en. In: *J. Exp. Med.* 195.6, pp. 789–794 (quoted on page 42).

Kitagawa, Yohko, Naganari Ohkura, Yujiro Kidani, et al. (2017). “Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment”. en. In: *Nat. Immunol.* 18.2, pp. 173–183 (quoted on page 68).

Klein, Ludger, Bruno Kyewski, Paul M Allen, and Kristin A Hogquist (2014). “Positive and negative selection of the T cell repertoire: what thymocytes see (and don’t see)”. en. In: *Nat. Rev. Immunol.* 14.6, pp. 377–391 (quoted on page 5).

Knoechel, Birgit, Jens Lohr, Estelle Kahn, Jeffrey A Bluestone, and Abul K Abbas (2005). “Sequential development of interleukin 2-dependent effector and regulatory T cells in response to endogenous systemic antigen”. en. In: *J. Exp. Med.* 202.10, pp. 1375–1386 (quoted on page 12).

Koch, Sven, Anis Larbi, Evelyna Derhovanessian, et al. (2008). “Multiparameter flow cytometric analysis of CD4 and CD8 T cell subsets in young and old people”. In: *Immun. Ageing* 5.1, p. 6 (quoted on page 63).

Konopacki, Catherine, George Plitas, and Alexander Rudensky (2015). “Reigning in regulatory T-cell function”. en. In: *Nat. Biotechnol.* 33.7, pp. 718–719 (quoted on page 135).

Kovanen, Panu E, Jérôme Bernard, Amin Al-Shami, et al. (2008). “T-cell development and function are modulated by dual specificity phosphatase DUSP5”. en. In: *J. Biol. Chem.* 283.25, pp. 17362–17369 (quoted on page 46).

Krause, Christopher D, Natasha Lavnikova, Junxia Xie, et al. (2006). “Preassembly and ligand-induced restructuring of the chains of the IFN-gamma receptor complex: the roles of Jak kinases, Stat1 and the receptor chains”. en. In: *Cell Res.* 16.1, pp. 55–69 (quoted on page 59).

Kshirsagar, Sudhir, Elisabeth Binder, Magdalena Riedl, et al. (2013). “Enhanced Activity of Akt in Teff Cells From Children With Lupus Nephritis Is Associated With Reduced Induction of Tumor Necrosis Factor Receptor-Associated Factor 6 and Increased OX40 Expression: Akt Activity in Lupus T Cells”. In: *Arthritis & Rheumatism* 65.11, pp. 2996–3006 (quoted on pages 11, 28).

Kuehn, Hye Sun, Weiming Ouyang, Bernice Lo, et al. (2014). “Immune dysregulation in human subjects with heterozygous germline mutations in CTLA4”. en. In: *Science* 345.6204, pp. 1623–1627 (quoted on pages 28, 64).

Kumar, Sunil, Giovanna Ambrosini, and Philipp Bucher (2017). “SNP2TFBS - a database of regulatory SNPs affecting predicted transcription factor binding site affinity”. en. In: *Nucleic Acids Res.* 45.D1, pp. D139–D144 (quoted on page 40).

Kumasaka, Natsuhiko, Andrew J Knights, and Daniel J Gaffney (2016). “Fine-mapping cellular QTLs with RASQUAL and ATAC-seq”. en. In: *Nat. Genet.* 48.2, pp. 206–213 (quoted on page 20).

Lahens, Nicholas F, Ibrahim Halil Kavakli, Ray Zhang, et al. (2014). “IVT-seq reveals extreme bias in RNA sequencing”. In: *Genome Biol.* 15.6, R86 (quoted on page 12).

Lambert, J C, C A Ibrahim-Verbaas, D Harold, et al. (2013). “Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease”. en. In: *Nat. Genet.* 45.12, pp. 1452–1458 (quoted on page 111).

Lange, Katrina M de, Loukas Moutsianas, James C Lee, et al. (2017). “Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease”. en. In: *Nat. Genet.* 49.2, pp. 256–261 (quoted on pages 17, 110, 135).

Langfelder, Peter, Bin Zhang, and Steve Horvath (2008). “Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R”. en. In: *Bioinformatics* 24.5, pp. 719–720 (quoted on page 74).

Lappalainen, Tuuli, Michael Sammeth, Marc R Friedländer, et al. (2013). “Transcriptome and genome sequencing uncovers functional variation in humans”. In: *Nature* 501, p. 506 (quoted on page 140).

Lee, Mark N, Chun Ye, Alexandra-Chloé Villani, et al. (2014). “Common genetic variants modulate pathogen-sensing responses in human dendritic cells”. en. In: *Science* 343.6175, p. 1246980 (quoted on page 20).

Leek, Jeffrey T, W Evan Johnson, Hilary S Parker, Andrew E Jaffe, and John D Storey (2012). “The sva package for removing batch effects and other unwanted variation in high-throughput experiments”. en. In: *Bioinformatics* 28.6, pp. 882–883 (quoted on pages 35, 74).

Lek, Monkol, Konrad J Karczewski, Eric V Minikel, et al. (2016). “Analysis of protein-coding genetic variation in 60,706 humans”. en. In: *Nature* 536.7616, pp. 285–291 (quoted on page 121).

Lenz, Tobias L, Aaron J Deutsch, Buhm Han, et al. (2015). “Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases”. en. In: *Nat. Genet.* 47.9, pp. 1085–1090 (quoted on page 2).

Li, Heng and Richard Durbin (2009). “Fast and accurate short read alignment with Burrows-Wheeler transform”. en. In: *Bioinformatics* 25.14, pp. 1754–1760 (quoted on pages 14, 37, 71).

Li, Heng, Bob Handsaker, Alec Wysoker, et al. (2009). “The Sequence Alignment/Map format and SAMtools”. en. In: *Bioinformatics* 25.16, pp. 2078–2079 (quoted on pages 37, 71).

Li, Peng, Rosanne Spolski, Wei Liao, et al. (2012). “BATF-JUN is critical for IRF4-mediated transcription in T cells”. en. In: *Nature* 490.7421, pp. 543–546 (quoted on page 57).

Li, Xudong, Yuqiong Liang, Mathias LeBlanc, Chris Benner, and Ye Zheng (2014). “Function of a Foxp3 cis-element in protecting regulatory T cell identity”. en. In: *Cell* 158.4, pp. 734–748 (quoted on page 102).

Li, Yang, Marije Oosting, Sanne P Smeekens, et al. (2016a). “A Functional Genomics Approach to Understand Variation in Cytokine Production in Humans”. en. In: *Cell* 167.4, 1099–1110.e14 (quoted on page 24).

Li, Yang, Marije Oosting, Patrick Deelen, et al. (2016b). “Corrigendum: Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi”. en. In: *Nat. Med.* 22.10, p. 1192 (quoted on pages 23, 143).

Li, Yang I, Bryce van de Geijn, Anil Raj, et al. (2016c). “RNA splicing is a primary link between genetic variation and disease”. en. In: *Science* 352.6285, pp. 600–604 (quoted on page 140).

Li, Yang I, David A Knowles, Jack Humphrey, et al. (2018). “Annotation-free quantification of RNA splicing using LeafCutter”. In: *Nat. Genet.* 50.1, pp. 151–158 (quoted on pages 13, 75, 94, 108, 111).

Liao, Yang, Gordon K Smyth, and Wei Shi (2014). “featureCounts: an efficient general purpose program for assigning sequence reads to genomic features”. en. In: *Bioinformatics* 30.7, pp. 923–930 (quoted on pages 13, 35, 38, 72, 73).

Liberzon, Arthur, Chet Birger, Helga Thorvaldsdóttir, et al. (2015). “The Molecular Signatures Database (MSigDB) hallmark gene set collection”. en. In: *Cell Syst* 1.6, pp. 417–425 (quoted on pages 35, 54).

Lim, Ai Ing, Yan Li, Silvia Lopez-Lastra, et al. (2017). “Systemic Human ILC Precursors Provide a Substrate for Tissue ILC Differentiation”. en. In: *Cell* 168.6, 1086–1100.e10 (quoted on pages 15, 99).

Linterman, Michelle A, Alice E Denton, Devina P Divekar, et al. (2014). “CD28 expression is required after T cell priming for helper T cell responses and protective immunity to infection”. en. In: *Elife* 3 (quoted on pages 9, 63).

Liu, Weihong, Amy L Putnam, Zhou Xu-yu, et al. (2006). “CD127 expression inversely correlates with FoxP3 and suppressive function of human CD4+ T reg cells”. en. In: *J. Exp. Med.* 203.7, pp. 1701–1711 (quoted on page 10).

Lo, Bernice, Kejian Zhang, Wei Lu, et al. (2015). “AUTOIMMUNE DISEASE. Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy”. en. In: *Science* 349.6246, pp. 436–440 (quoted on pages 28, 64, 142).

London, C A, M P Lodge, and A K Abbas (2000). “Functional responses and costimulator dependence of memory CD4+ T cells”. en. In: *J. Immunol.* 164.1, pp. 265–272 (quoted on pages 9, 28, 62).

Love, Michael I, Wolfgang Huber, and Simon Anders (2014). “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. en. In: *Genome Biol.* 15.12, p. 550 (quoted on pages 35, 73, 74).

Luco, Reini F, Qun Pan, Kaoru Tominaga, et al. (2010). “Regulation of alternative splicing by histone modifications”. en. In: *Science* 327.5968, pp. 996–1000 (quoted on page 139).

Luqman, M and K Bottomly (1992). “Activation requirements for CD4+ T cells differing in CD45R expression”. en. In: *J. Immunol.* 149.7, pp. 2300–2306 (quoted on pages 28, 62).

Man, Kevin, Maria Miasari, Wei Shi, et al. (2013). “The transcription factor IRF4 is essential for TCR affinity-mediated metabolic programming and clonal expansion of T cells”. en. In: *Nat. Immunol.* 14.11, pp. 1155–1165 (quoted on pages 45, 57).

Manzotti, Claire N, Michael K P Liu, Fiona Burke, et al. (2006). “Integration of CD28 and CTLA-4 function results in differential responses of T cells to CD80 and CD86”. en. In: *Eur. J. Immunol.* 36.6, pp. 1413–1422 (quoted on page 40).

Marson, Alexander, Karsten Kretschmer, Garrett M Frampton, et al. (2007). “Foxp3 occupancy and regulation of key target genes during T-cell stimulation”. en. In: *Nature* 445.7130, pp. 931–935 (quoted on pages 90, 100, 135).

Martins, Gislâine A, Luisa Cimmino, Miriam Shapiro-Shelef, et al. (2006). “Transcriptional repressor Blimp-1 regulates T cell homeostasis and function”. en. In: *Nat. Immunol.* 7.5, pp. 457–465 (quoted on pages 57, 94).

Matzaraki, Vasiliki, Vinod Kumar, Cisca Wijmenga, and Alexandra Zhernakova (2017). “The MHC locus and genetic susceptibility to autoimmune and infectious diseases”. en. In: *Genome Biol.* 18.1, p. 76 (quoted on page 2).

Maurano, Matthew T, Richard Humbert, Eric Rynes, et al. (2012). “Systematic localization of common disease-associated variation in regulatory DNA”. en. In: *Science* 337.6099, pp. 1190–1195 (quoted on pages 21, 104).

McCarthy, Davis J, Kieran R Campbell, Aaron T L Lun, and Quin F Wills (2017). “Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R”. en. In: *Bioinformatics* 33.8, pp. 1179–1186 (quoted on page 35).

McGovern, Amanda, Stefan Schoenfelder, Paul Martin, et al. (2016). “Capture Hi-C identifies a novel causal gene, IL20RA, in the pan-autoimmune genetic susceptibility region 6q23”. en. In: *Genome Biol.* 17.1, p. 212 (quoted on page 21).

McHugh, Rebecca S, Matthew J Whitters, Ciriaco A Piccirillo, et al. (2002). “CD4(+)CD25(+) immunoregulatory T cells: gene expression analysis reveals a functional role for the glucocorticoid-induced TNF receptor”. en. In: *Immunity* 16.2, pp. 311–323 (quoted on page 10).

McVicker, Graham, Bryce van de Geijn, Jacob F Degner, et al. (2013). “Identification of genetic variants that affect histone modifications in human cells”. en. In: *Science* 342.6159, pp. 747–749 (quoted on page 20).

Mehlhop-Williams, Erin R. and Michael J. Bevan (2014). “Memory CD8+ T cells exhibit increased antigen threshold requirements for recall proliferation”. In: *Journal of Experimental Medicine* 211.2, pp. 345–356 (quoted on pages 62, 64).

Michie, C A, A McLean, C Alcock, and P C Beverley (1992). “Lifespan of human lymphocyte subsets defined by CD45 isoforms”. en. In: *Nature* 360.6401, pp. 264–265 (quoted on page 13).

Mifsud, Borbala, Filipe Tavares-Cadete, Alice N Young, et al. (2015). “Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C”. en. In: *Nat. Genet.* 47.6, pp. 598–606 (quoted on page 129).

Miyara, Makoto, Zahir Amoura, Christophe Parizot, et al. (2005). “Global natural regulatory T cell depletion in active systemic lupus erythematosus”. en. In: *J. Immunol.* 175.12, pp. 8392–8400 (quoted on page 12).

Morris, Andrew P, Benjamin F Voight, Tanya M Teslovich, et al. (2012). “Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes”. en. In: *Nat. Genet.* 44.9, pp. 981–990 (quoted on page 111).

Mumbach, Maxwell R, Ansuman T Satpathy, Evan A Boyle, et al. (2017). “Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements”. en. In: *Nat. Genet.* 49.11, pp. 1602–1612 (quoted on pages 39, 141).

Nagar, Meital, Jasmine Jacob-Hirsch, Helly Vernitsky, et al. (2010). “TNF activates a NF-kappaB-regulated cellular program in human CD45RA- regulatory T cells that modulates their suppressive function”. en. In: *J. Immunol.* 184.7, pp. 3570–3581 (quoted on page 135).

Nakamura, K, A Kitani, and W Strober (2001). “Cell contact-dependent immunosuppression by CD4(+)CD25(+) regulatory T cells is mediated by cell surface-bound transforming growth factor beta”. en. In: *J. Exp. Med.* 194.5, pp. 629–644 (quoted on page 10).

Nasr, M B, S Tezza, F D’addio, C Mameli, and others (2017). “PD-L1 genetic overexpression or pharmacological restoration in hematopoietic stem and progenitor cells reverses autoimmune diabetes”. In: *Sci. Transl. Med.* (Quoted on page 65).

Ndejembi, Modesta P, John R Teijaro, Deepa S Patke, et al. (2006). “Control of memory CD4 T cell recall by the CD28/B7 costimulatory pathway”. en. In: *J. Immunol.* 177.11, pp. 7698–7706 (quoted on pages 28, 49, 63).

Ndlovu, Hlumani, Mathew Darby, Monika Froelich, et al. (2014). “Inducible Deletion of CD28 Prior to Secondary Nippostrongylus brasiliensis Infection Impairs Worm Expulsion and Recall of Protective Memory CD4+ T Cell Responses”. In: *PLoS Pathog.* 10.2, e1003906 (quoted on pages 9, 28, 63).

Nelson, Christopher P, Anuj Goel, Adam S Butterworth, et al. (2017). “Association analyses based on false discovery rate implicate new loci for coronary artery disease”. en. In: *Nat. Genet.* 49.9, pp. 1385–1391 (quoted on page 111).

Nepom, G T (1998). “Major histocompatibility complex-directed susceptibility to rheumatoid arthritis”. en. In: *Adv. Immunol.* 68, pp. 315–332 (quoted on page 2).

Ni, Ting, Wenjing Yang, Miao Han, et al. (2016). “Global intron retention mediated gene regulation during CD4+ T cell activation”. en. In: *Nucleic Acids Res.* 44.14, pp. 6817–6829 (quoted on page 94).

Nica, Alexandra C, Stephen B Montgomery, Antigone S Dimas, et al. (2010). “Candidate Causal Regulatory Effects by Integration of Expression QTLs with Complex Trait Genetic Associations”. In: *PLoS Genet.* 6.4, e1000895 (quoted on page 137).

Nicolae, Dan L, Eric Gamazon, Wei Zhang, et al. (2010). “Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS”. In: *PLoS Genet.* 6.4, e1000888 (quoted on page 17).

Nilsen, Timothy W and Brenton R Graveley (2010). “Expansion of the eukaryotic proteome by alternative splicing”. In: *Nature* 463, p. 457 (quoted on page 13).

Nisticò, L, R Buzzetti, L E Pritchard, et al. (1996). “The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry”. en. In: *Hum. Mol. Genet.* 5.7, pp. 1075–1080 (quoted on page 3).

Nowak, Anna, Dominik Lock, Petra Bacher, et al. (2018). “CD137+CD154- Expression As a Regulatory T Cell (Treg)-Specific Activation Signature for Identification and Sorting of Stable Human Tregs from In Vitro Expansion Cultures”. en. In: *Front. Immunol.* 9, p. 199 (quoted on page 135).

Ohkura, Naganari, Masahide Hamaguchi, Hiromasa Morikawa, et al. (2012). “T cell receptor stimulation-induced epigenetic changes and Foxp3 expression are independent and complementary events required for Treg cell development”. en. In: *Immunity* 37.5, pp. 785–799 (quoted on page 68).

Okada, Yukinori, Di Wu, Gosia Trynka, et al. (2014). “Genetics of rheumatoid arthritis contributes to biology and drug discovery”. en. In: *Nature* 506.7488, pp. 376–381 (quoted on pages 3, 28, 39, 58, 110, 134).

Okitsu, Cindy Yen, John Cheng Feng Hsieh, and Chih-Lin Hsieh (2010). “Transcriptional activity affects the H3K4me3 level and distribution in the coding region”. en. In: *Mol. Cell. Biol.* 30.12, pp. 2933–2946 (quoted on page 100).

Onengut-Gumuscu, Suna, Wei-Min Chen, Oliver Burren, et al. (2015). “Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers”. en. In: *Nat. Genet.* 47.4, pp. 381–386 (quoted on pages 28, 39, 110).

Orrù, Valeria, Maristella Steri, Gabriella Sole, et al. (2013). “Genetic variants regulating immune cell levels in health and disease”. en. In: *Cell* 155.1, pp. 242–256 (quoted on page 23).

Papoutsaki, Marina and Antonio Costanzo (2013). “Treatment of psoriasis and psoriatic arthritis”. en. In: *BioDrugs* 27 Suppl 1, pp. 3–12 (quoted on page 11).

Patin, Etienne, Milena Hasan, Jacob Bergstedt, et al. (2018). “Natural variation in the parameters of innate immune cells is preferentially driven by genetic factors”. In: *Nat. Immunol.* 19.3, pp. 302–314 (quoted on page 138).

Patro, Rob, Geet Duggal, Michael I Love, Rafael A Irizarry, and Carl Kingsford (2017). “Salmon provides fast and bias-aware quantification of transcript expression”. en. In: *Nat. Methods* 14.4, pp. 417–419 (quoted on pages 13, 74, 111).

Pelikan, Richard C, Jennifer A Kelly, Yao Fu, et al. (2018). “Enhancer histone-QTLs are enriched on autoimmune risk haplotypes and influence gene expression within chromatin networks”. en. In: *Nat. Commun.* 9.1, p. 2905 (quoted on page 134).

Pfoertner, Susanne, Andreas Jeron, Michael Probst-Kepper, et al. (2006). “Signatures of human regulatory T cells: an encounter with old friends and new players”. en. In: *Genome Biol.* 7.7, R54 (quoted on pages 68, 99).

Pickrell, Joseph K (2014). “Joint analysis of functional genomic data and genome-wide association studies of 18 human traits”. en. In: *Am. J. Hum. Genet.* 94.4, pp. 559–573 (quoted on page 15).

Plenge, Robert M, Chris Cotsapas, Leela Davies, et al. (2007). “Two independent alleles at 6q23 associated with risk of rheumatoid arthritis”. en. In: *Nat. Genet.* 39.12, pp. 1477–1482 (quoted on page 3).

Pollard, Katherine S, Melissa J Hubisz, Kate R Rosenbloom, and Adam Siepel (2010). “Detection of nonneutral substitution rates on mammalian phylogenies”. en. In: *Genome Res.* 20.1, pp. 110–121 (quoted on pages 75, 95).

Quinlan, Aaron R and Ira M Hall (2010). “BEDTools: a flexible suite of utilities for comparing genomic features”. en. In: *Bioinformatics* 26.6, pp. 841–842 (quoted on pages 38, 72, 73).

Qureshi, Omar S, Yong Zheng, Kyoko Nakamura, et al. (2011). “Trans-endocytosis of CD80 and CD86: a molecular basis for the cell-extrinsic function of CTLA-4”. en. In: *Science* 332.6029, pp. 600–603 (quoted on pages 28, 30, 64).

Raj, Towfique, Katie Rothamel, Sara Mostafavi, et al. (2014). “Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes”. en. In: *Science* 344.6183, pp. 519–523 (quoted on page 18).

Raychaudhuri, Soumya, Brian P Thomson, Elaine F Remmers, et al. (2009). “Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk”. en. In: *Nat. Genet.* 41.12, pp. 1313–1318 (quoted on page 134).

Read, S, V Malmström, and F Powrie (2000). “Cytotoxic T lymphocyte-associated antigen 4 plays an essential role in the function of CD25(+)CD4(+) regulatory cells that control intestinal inflammation”. en. In: *J. Exp. Med.* 192.2, pp. 295–302 (quoted on pages 8, 134).

Regev, Aviv and Others (2016). *The Human Cell Atlas* (quoted on page 138).

Reimand, Jüri, Tambet Arak, Priit Adler, et al. (2016). “g:Profiler-a web server for functional interpretation of gene lists (2016 update)”. en. In: *Nucleic Acids Res.* 44.W1, W83–9 (quoted on pages 74, 109).

Richard, Arianne C, Aaron T L Lun, Winnie W Y Lau, et al. (2018). “T cell cytolytic capacity is independent of initial stimulation strength”. In: *Nat. Immunol.* 19.8, pp. 849–858 (quoted on page 62).

Richer, Martin J, Jeffrey C Nolz, and John T Harty (2013). “Pathogen-specific inflammatory milieux tune the antigen sensitivity of CD8(+) T cells by enhancing T cell receptor signaling”. en. In: *Immunity* 38.1, pp. 140–152 (quoted on page 28).

Rieckmann, Jan C, Roger Geiger, Daniel Hornburg, et al. (2017). “Social network architecture of human immune cells unveiled by quantitative proteomics”. en. In: *Nat. Immunol.* 18.5, pp. 583–593 (quoted on page 144).

Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, et al. (2015). “Integrative analysis of 111 reference human epigenomes”. en. In: *Nature* 518.7539, pp. 317–330 (quoted on pages 14, 140).

Robertson, Shelly J, Ronald J Messer, Aaron B Carmody, et al. (2008). “CD137 costimulation of CD8+ T cells confers resistance to suppression by virus-induced regulatory T cells”. en. In: *J. Immunol.* 180.8, pp. 5267–5274 (quoted on page 143).

Roederer, Mario, Lydia Quaye, Massimo Mangino, et al. (2015). “The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis”. en. In: *Cell* 161.2, pp. 387–403 (quoted on page 23).

Roncador, Giovanna, Philip J Brown, Lorena Maestre, et al. (2005). “Analysis of FOXP3 protein expression in human CD4+CD25+ regulatory T cells at the single-cell level”. en. In: *Eur. J. Immunol.* 35.6, pp. 1681–1691 (quoted on page 99).

Rosario, Ricardo Cruz-Herrera del, Jeremie Poschmann, Sigrid Laure Rouam, et al. (2015). “Sensitive detection of chromatin-altering polymorphisms reveals autoimmune disease mechanisms”. en. In: *Nat. Methods* 12.5, pp. 458–464 (quoted on page 20).

Rossin, Elizabeth J, Kasper Lage, Soumya Raychaudhuri, et al. (2011). “Proteins Encoded in Genomic Regions Associated with Immune-Mediated Disease Physically Interact and Suggest Underlying Biology”. In: *PLoS Genet.* 7.1, e1001273 (quoted on page 144).

Rowshanravan, Behzad, Neil Halliday, and David M Sansom (2017). “CTLA-4: a moving target in immunotherapy”. en. In: *Blood*, blood–2017–06–741033 (quoted on page 27).

Roychoudhuri, Rahul, Kiyoshi Hirahara, Kambiz Mousavi, et al. (2013). “BACH2 represses effector programs to stabilize T(reg)-mediated immune homeostasis”. en. In: *Nature* 498.7455, pp. 506–510 (quoted on page 101).

Rudra, Dipayan, Paul deRoos, Ashutosh Chaudhry, et al. (2012). “Transcription factor Foxp3 and its protein partners form a complex regulatory network”. en. In: *Nat. Immunol.* 13.10, pp. 1010–1019 (quoted on page 101).

Savage, Peter A, J Jay Boniface, and Mark M Davis (1999). “A Kinetic Basis For T Cell Receptor Repertoire Selection during an Immune Response”. en. In: *Immunity* 10.4, pp. 485–492 (quoted on page 28).

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). “Biological insights from 108 schizophrenia-associated genetic loci”. en. In: *Nature* 511.7510, pp. 421–427 (quoted on page 111).

Schmidl, Christian, André F Rendeiro, Nathan C Sheffield, and Christoph Bock (2015). “ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors”. In: *Nat. Methods* 12, p. 963 (quoted on pages 15, 36, 99, 138).

Schneider, Anya, Sarah Alice Long, Karen Cerosaletti, et al. (2013). “In active relapsing-remitting multiple sclerosis, effector T cell resistance to adaptive T(regs) involves IL-6-mediated signaling”. en. In: *Sci. Transl. Med.* 5.170, 170ra15 (quoted on page 11).

Schoenbrunn, Anne, Marco Frentsch, Siegfried Kohler, et al. (2012). “A converse 4-1BB and CD40 ligand expression pattern delineates activated regulatory T cells (Treg) and conventional T cells enabling direct isolation of alloantigen-reactive natural Foxp3+ Treg”. en. In: *J. Immunol.* 189.12, pp. 5985–5994 (quoted on pages 135, 136).

Schubert, Desirée, Claudia Bode, Rupert Kenefeck, et al. (2014). “Autosomal dominant immune dysregulation syndrome in humans with CTLA4 mutations”. en. In: *Nat. Med.* 20.12, pp. 1410–1416 (quoted on pages 4, 28, 64).

Serebrennikova, Oksana B, Christos Tsatsanis, Changchuin Mao, et al. (2012). “Tpl2 ablation promotes intestinal inflammation and tumorigenesis in Apcmin mice by inhibiting IL-10 secretion and regulatory T-cell generation”. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 109.18, E1082–91 (quoted on page 136).

Shapiro, V S, K E Truitt, J B Imboden, and A Weiss (1997). “CD28 mediates transcriptional upregulation of the interleukin-2 (IL-2) promoter through a composite element containing the CD28RE and NF-IL-2B AP-1 sites”. en. In: *Mol. Cell. Biol.* 17.7, pp. 4051–4058 (quoted on pages 7, 57).

Shimizu, Jun, Sayuri Yamazaki, Takeshi Takahashi, Yasumasa Ishida, and Shimon Sakaguchi (2002). “Stimulation of CD25+CD4+ regulatory T cells through GITR breaks immunological self-tolerance”. In: *Nat. Immunol.* 3, p. 135 (quoted on page 10).

Singh, Yogesh, Hong Chen, Yuetao Zhou, et al. (2015). “Differential effect of DJ-1/PARK7 on development of natural and induced regulatory T cells”. en. In: *Sci. Rep.* 5, p. 17723 (quoted on page 136).

Slowikowski, Kamil, Xinli Hu, and Soumya Raychaudhuri (2014). “SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci”. en. In: *Bioinformatics* 30.17, pp. 2496–2497 (quoted on page 16).

Sniekers, Suzanne, Sven Stringer, Kyoko Watanabe, et al. (2017). “Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence”. In: *Nat. Genet.* 49, p. 1107 (quoted on page 111).

Sollid, L M, G Markussen, J Ek, et al. (1989). “Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer”. en. In: *J. Exp. Med.* 169.1, pp. 345–350 (quoted on page 2).

Solovieff, Nadia, Chris Cotsapas, Phil H Lee, Shaun M Purcell, and Jordan W Smoller (2013). “Pleiotropy in complex traits: challenges and strategies”. en. In: *Nat. Rev. Genet.* 14.7, pp. 483–495 (quoted on page 18).

Sos, Brandon Chin, Ho-Lim Fung, Derek Rui Gao, et al. (2016). “Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay”. en. In: *Genome Biol.* 17, p. 20 (quoted on page 15).

Spain, Sarah L and Jeffrey C Barrett (2015). “Strategies for fine-mapping complex traits”. en. In: *Hum. Mol. Genet.* 24.R1, R111–9 (quoted on pages 4, 103).

Stahl, Eli A, Soumya Raychaudhuri, Elaine F Remmers, et al. (2010). “Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci”. en. In: *Nat. Genet.* 42.6, pp. 508–514 (quoted on page 3).

Stubbington, Michael Jt, Bidesh Mahata, Valentine Svensson, et al. (2015). “An atlas of mouse CD4(+) T cell transcriptomes”. en. In: *Biol. Direct* 10, p. 14 (quoted on page 99).

Stunnenberg, Hendrik G, International Human Epigenome Consortium, and Martin Hirst (2016). “The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery”. en. In: *Cell* 167.7, p. 1897 (quoted on page 14).

Tai, Xuguang, Francois Van Laethem, Arlene H Sharpe, and Alfred Singer (2007). “Induction of autoimmune disease in CTLA-4-/- mice depends on a specific CD28 motif that is required for in vivo costimulation”. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 104.34, pp. 13756–13761 (quoted on page 28).

Takahashi, T, Y Kuniyasu, M Toda, et al. (1998). “Immunologic self-tolerance maintained by CD25+CD4+ naturally anergic and suppressive T cells: induction of autoimmune disease by breaking their anergic/suppressive state”. en. In: *Int. Immunol.* 10.12, pp. 1969–1980 (quoted on pages 10, 11).

Taylor, Alison, Johan Verhagen, Tunç Akkoç, et al. (2009). “IL-10 suppresses CD2-mediated T cell activation via SHP-1”. en. In: *Mol. Immunol.* 46.4, pp. 622–629 (quoted on page 11).

Teijaro, John R, Modesta N Njau, David Verhoeven, et al. (2009). “Costimulation modulation uncouples protection from immunopathology in memory T cell responses to influenza virus”. en. In: *J. Immunol.* 182.11, pp. 6834–6843 (quoted on page 28).

Tivol, E A, F Borriello, A N Schweitzer, et al. (1995). “Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4”. en. In: *Immunity* 3.5, pp. 541–547 (quoted on pages 28, 64).

Tivol, E A, S D Boyd, S McKeon, et al. (1997). “CTLA4Ig prevents lymphoproliferation and fatal multiorgan tissue destruction in CTLA-4-deficient mice”. en. In: *J. Immunol.* 158.11, pp. 5091–5094 (quoted on page 28).

Todd, J A, J I Bell, and H O McDevitt (1987). “HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus”. en. In: *Nature* 329.6140, pp. 599–604 (quoted on page 2).

Trynka, Gosia, Karen A Hunt, Nicholas A Bockett, et al. (2011). “Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease”. en. In: *Nat. Genet.* 43.12, pp. 1193–1201 (quoted on pages 39, 104, 110, 134).

Trynka, Gosia, Cynthia Sandor, Buhm Han, et al. (2013). “Chromatin marks identify critical cell types for fine mapping complex trait variants”. en. In: *Nat. Genet.* 45.2, pp. 124–130 (quoted on pages 15, 16, 58, 59, 104).

Trynka, Gosia, Harm-Jan Westra, Kamil Slowikowski, et al. (2015). “Disentangling the Effects of Colocalizing Genomic Annotations to Functionally Prioritize Non-coding Variants within Complex-Trait Loci”. en. In: *Am. J. Hum. Genet.* 97.1, pp. 139–152 (quoted on pages 59, 104).

Tsoi, Lam C, Sarah L Spain, Jo Knight, et al. (2012). “Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity”. en. In: *Nat. Genet.* 44.12, pp. 1341–1348 (quoted on pages 39, 110).

Unen, Vincent van, Na Li, Ilse Molendijk, et al. (2016). “Mass Cytometry of the Human Mucosal Immune System Identifies Tissue- and Disease-Associated Immune Subsets”. en. In: *Immunity* 44.5, pp. 1227–1239 (quoted on page 144).

Urrutia, Alejandra, Darragh Duffy, Vincent Rouilly, et al. (2016). “Standardized Whole-Blood Transcriptional Profiling Enables the Deconvolution of Complex Induced Immune Responses”. en. In: *Cell Rep.* 16.10, pp. 2777–2791 (quoted on page 138).

Vaeth, Martin and Stefan Feske (2018). “NFAT control of immune function: New Frontiers for an Abiding Trooper”. en. In: *F1000Res.* 7, p. 260 (quoted on page 101).

Van, Vu Quang, Marianne Raymond, Nobuyasu Baba, et al. (2012). “CD47(high) expression on CD4 effectors identifies functional long-lived memory T cell progenitors”. en. In: *J. Immunol.* 188.9, pp. 4249–4255 (quoted on page 43).

Wakamatsu, Ei, Diane Mathis, and Christophe Benoist (2013). “Convergent and divergent effects of costimulatory molecules in conventional and regulatory CD4+ T cells”. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 110.3, pp. 1023–1028 (quoted on page 62).

Wang, Chun Jing, Frank Heuts, Vitalijs Ovcinnikovs, et al. (2015). “CTLA-4 controls follicular helper T-cell differentiation by regulating the strength of CD28 engagement”. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 112.2, pp. 524–529 (quoted on page 9).

Wang, Dongxue, Basak Eraslan, Thomas Wieland, et al. (2019). "A deep proteome and transcriptome abundance atlas of 29 healthy human tissues". en. In: *Mol. Syst. Biol.* 15.2, e8503 (quoted on page 143).

Wang, Liqing, Yujie Liu, Ulf H Beier, et al. (2013a). "Foxp3+ T-regulatory cells require DNA methyltransferase 1 expression to prevent development of lethal autoimmunity". en. In: *Blood* 121.18, pp. 3631–3639 (quoted on page 101).

Wang, Rui, Lina Kozhaya, Frances Mercer, et al. (2009). "Expression of GARP selectively identifies activated human FOXP3+ regulatory T cells". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 106.32, pp. 13439–13444 (quoted on pages 90, 100).

Wang, Su, Hanfei Sun, Jian Ma, et al. (2013b). "Target analysis by integration of transcriptome and ChIP-seq data with BETA". en. In: *Nat. Protoc.* 8.12, pp. 2502–2515 (quoted on pages 39, 45, 88).

Weiss, A and J B Imboden (1987). "Cell surface molecules and early events involved in human T lymphocyte activation". en. In: *Adv. Immunol.* 41, pp. 1–38 (quoted on pages 7, 68).

West, Nathaniel R, Ahmed N Hegazy, Benjamin M J Owens, et al. (2017). "Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor-neutralizing therapy in patients with inflammatory bowel disease". en. In: *Nat. Med.* 23.5, pp. 579–589 (quoted on page 64).

Westra, Harm-Jan, Marjolein J Peters, Tõnu Esko, et al. (2013). "Systematic identification of trans eQTLs as putative drivers of known disease associations". en. In: *Nat. Genet.* 45.10, pp. 1238–1243 (quoted on pages 17, 29, 137).

Westra, Harm-Jan, Marta Martínez-Bonet, Suna Onengut-Gumuscu, et al. (2018). "Fine-mapping and functional studies highlight potential causal variants for rheumatoid arthritis and type 1 diabetes". en. In: *Nat. Genet.* 50.10, pp. 1366–1374 (quoted on page 135).

Wijst, Monique G P van der, Harm Brugge, Dylan H de Vries, et al. (2018). "Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs". In: *Nat. Genet.* 50.4, pp. 493–497 (quoted on page 138).

Williams, Matthew A, Eugene V Ravkov, and Michael J Bevan (2008). "Rapid culling of the CD4+ T cell repertoire in the transition from effector to memory". en. In: *Immunity* 28.4, pp. 533–545 (quoted on page 9).

Wills, Quin F, Kenneth J Livak, Alex J Tipping, et al. (2013). "Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments". In: *Nat. Biotechnol.* 31, p. 748 (quoted on page 138).

Yamada, Takeshi, Kirsten Gierach, Ping-Hsien Lee, Xiaohong Wang, and H Daniel Lacorazza (2010). "Cutting edge: Expression of the transcription factor E74-like factor 4 is regulated by the mammalian target of rapamycin pathway in CD8+ T cells". en. In: *J. Immunol.* 185.7, pp. 3824–3828 (quoted on page 101).

Yanagi, Yusuke, Yasunobu Yoshikai, Kathleen Leggett, et al. (1984). "A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains". In: *Nature* 308, p. 145 (quoted on page 7).

Ye, Chun Jimmie, Ting Feng, Ho-Keun Kwon, et al. (2014). "Intersection of population variation and autoimmunity genetics in human T cell activation". en. In: *Science* 345.6202, p. 1254665 (quoted on page 20).

Zhang, Yong, Tao Liu, Clifford A Meyer, et al. (2008). “Model-based Analysis of ChIP-Seq (MACS)”. In: *Genome Biol.* 9.9, R137 (quoted on pages 14, 38, 72).

Zheng, Song Guo, Ju Hua Wang, William Stohl, et al. (2006). “TGF-beta requires CTLA-4 early after T cell activation to induce FoxP3 and generate adaptive CD4+CD25+ regulatory cells”. en. In: *J. Immunol.* 176.6, pp. 3321–3329 (quoted on pages 8, 134).

Zheng, Ye, Ashutosh Chaudhry, Arnold Kas, et al. (2009). “Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control T(H)2 responses”. en. In: *Nature* 458.7236, pp. 351–356 (quoted on page 92).

Zhu, Jinfang, Hidehiro Yamane, and William E Paul (2010a). “Differentiation of effector CD4 T cell populations (*)”. en. In: *Annu. Rev. Immunol.* 28, pp. 445–489 (quoted on page 8).

Zhu, Lihua J, Claude Gazin, Nathan D Lawson, et al. (2010b). “ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data”. In: *BMC Bioinformatics* 11.1, p. 237 (quoted on page 73).

List of Figures

List of Tables