


RESEARCH

Open Access



3DFAACTS-SNP: using regulatory T cell-specific epigenomics data to uncover candidate mechanisms of type 1 diabetes (T1D) risk

Ning Liu^{1,2,3†}, Timothy Sadlon^{2,4†}, Ying Y. Wong^{2,4}, Stephen Pederson³, James Breen^{1,2,3,5,6*†}  and Simon C. Barry^{2,4†}

Abstract

Background: Genome-wide association studies (GWAS) have enabled the discovery of single nucleotide polymorphisms (SNPs) that are significantly associated with many autoimmune diseases including type 1 diabetes (T1D). However, many of the identified variants lie in non-coding regions, limiting the identification of mechanisms that contribute to autoimmune disease progression. To address this problem, we developed a variant filtering workflow called 3DFAACTS-SNP to link genetic variants to target genes in a cell-specific manner. Here, we use 3DFAACTS-SNP to identify candidate SNPs and target genes associated with the loss of immune tolerance in regulatory T cells (Treg) in T1D.

Results: Using 3DFAACTS-SNP, we identified from a list of 1228 previously fine-mapped variants, 36 SNPs with plausible Treg-specific mechanisms of action. The integration of cell type-specific chromosome conformation capture data in 3DFAACTS-SNP identified 266 regulatory regions and 47 candidate target genes that interact with these variant-containing regions in Treg cells. We further demonstrated the utility of the workflow by applying it to three other SNP autoimmune datasets, identifying 16 Treg-centric candidate variants and 60 interacting genes. Finally, we demonstrate the broad utility of 3DFAACTS-SNP for functional annotation of all known common (> 10% allele frequency) variants from the Genome Aggregation Database (gnomAD). We identified 9376 candidate variants and 4968 candidate target genes, generating a list of potential sites for future T1D or other autoimmune disease research.

Conclusions: We demonstrate that it is possible to further prioritise variants that contribute to T1D based on regulatory function, and illustrate the power of using cell type-specific multi-omics datasets to determine disease mechanisms. Our workflow can be customised to any cell type for which the individual datasets for functional annotation have been generated, giving broad applicability and utility.

Keywords: Type 1 diabetes, Autoimmune disease, Regulatory T cells, Hi-C, Transcription factor binding, Functional annotation

Background

Autoimmune diseases are chronic inflammatory disorders caused by a breakdown of immunological tolerance to self-antigens, which results in an imbalance between multiple immune cells, including conventional T cells (Tconv) and regulatory T cells (Tregs) [1]. The imbalance of immune cell function can lead to the destruction of host tissues, such as is observed in multiple autoimmune

[†]Ning Liu and Timothy Sadlon contributed equally to this work

[†]James Breen and Simon C. Barry jointly directed this work

*Correspondence: jimmy.breen@telethonkids.org.au

¹ South Australian Health and Medical Research Institute, Adelaide, Australia
Full list of author information is available at the end of the article



diseases, including rheumatoid arthritis (RA), multiple sclerosis (MS) and inflammatory bowel disease (IBD). In the case of type 1 diabetes (T1D), a reduction of Treg cell function contributes to unrestrained immune destruction of the insulin-generating pancreatic beta cells, resulting in the loss of control of blood sugar levels [2].

Regulatory T cell function is mediated by expression of the Foxhead Box Protein 3 (FOXP3) transcription factor (TF) as evidenced by severe autoimmune diseases observed in FOXP3-deficient scurfy mice [3] and IPEX in humans [4–6]. RNA sequencing and chromatin immunoprecipitation (ChIP) studies have uncovered an extensive FOXP3-dependent molecular program involved in Treg cell development and stability [7, 8]. Functional fitness of Treg is dependent on stable robust expression of FOXP3, such that reduced FOXP3 expression is linked to reduced Treg function. For example, in a small T1D cohort study, we have shown that there is a decrease in FOXP3 expression in the Treg of children over the first 9 months post diagnosis [9]. However, since FOXP3 itself is not mutated in autoimmune diseases other than IPEX, the loss of FOXP3 levels and functional fitness is likely caused by perturbation of the Treg gene regulatory network. Hence, by decoding the regulatory network of FOXP3, and mapping the genetic risk to the key functional genes it impacts, we will gain a better understanding of how autoimmune diseases like T1D could be countered.

T1D has a strong pattern of inheritance [10]. Genome-Wide Association Studies (GWAS) have identified over 50 loci that are strongly associated with T1D, based on the genotyping of a total of 9934 cases and 16,956 controls from multiple cohorts and resources [11]. In addition, fine-mapping of immune-disease associated loci represented on the Immunochip Array [12] followed by a Bayesian approach identified 44 significant T1D-associated Loci and over 1,000 credible SNPs [13]. Although GWAS and fine-mapping studies have revealed significant associations between genetic variants and T1D, the vast majority of the sampled single nucleotide polymorphisms (SNPs) are located in non-coding regions that do not alter the amino acid sequence in a protein, making it difficult to assign direct biological functions to variants [14–16]. Non-coding variants can be linked to direct changes in gene expression by identifying expression quantitative trait loci (eQTL) that aim to associate allelic changes to cis (within 1Mbp of the associated gene) and trans (>1Mbp) changes in gene expression [17]. This additional direct gene expression association however still fails to identify direct mechanisms by which a specific genetic variant can change gene expression. In addition, usage of eQTLs to establish direct changes from GWAS variants is somewhat limited to local, or cis-eQTLs [18], whereas mounting evidence shows that

long-range regulatory connections, driven by three-dimensional chromatin interactions [19, 20], can mediate these changes in expression.

With the increasing affordability and availability of high-throughput sequencing techniques and various epigenomics sequencing data protocols, the impact of genome organisation and accessibility can now be added to the functional annotation of genetic risk [1]. Chromatin immunoprecipitation sequencing (ChIP-seq) allows us to identify the binding sites of a transcription factor; assay for transposase-accessible chromatin sequencing (ATAC-seq) data offers the ability to identify accessible regions of the genome; and high resolution chromosome conformation capture sequencing (Hi-C) data can facilitate the investigation of the three-dimensional structure of the genome. Since it is believed that the mechanisms by which non-coding SNPs contribute to diseases are mostly via changes to the function of regulatory elements [16], we believe that combining multiple genomics and epigenetics sequencing data can further reveal the relationship between GWAS SNPs and disease pathways.

While alterations in either the effector or regulatory arms of the immune system can result in loss of tolerance and autoimmune disease, we have used a Treg-centric view of loss of tolerance. This is based on the observation that defects in Treg function have been reported in autoimmune diseases including T1D and MS [21, 22] and that experimental deletion of FOXP3 or reduced Treg function results in autoimmune disease in many model systems [23, 24]. Our hypothesis is that the genetic variation that specifically alters Treg function will reside in open chromatin in Treg cells that is bound by FOXP3 and the genes controlled by these regulatory regions can be identified by chromosome conformation capture approaches. Therefore, in this paper, we describe a filtering workflow using multiple sequencing datasets from human Tregs, aiming to identify plausible immunomodulatory mechanisms and potentially find previously unknown connections between causative variant SNPs significantly associated with T1D and the genes they impact. Only by doing this will a new wave of personalised medicines to prevent T1D be discovered.

Results

Post-GWAS filtering using Treg-specific epigenomic datasets prioritises functionally relevant genetic variants contributing to T1D

As T1D is partly a consequence of Treg dysfunction, we infer that variants contained within active regulatory regions of Treg cells are likely to contribute to disease progression by impacting Treg function. A view

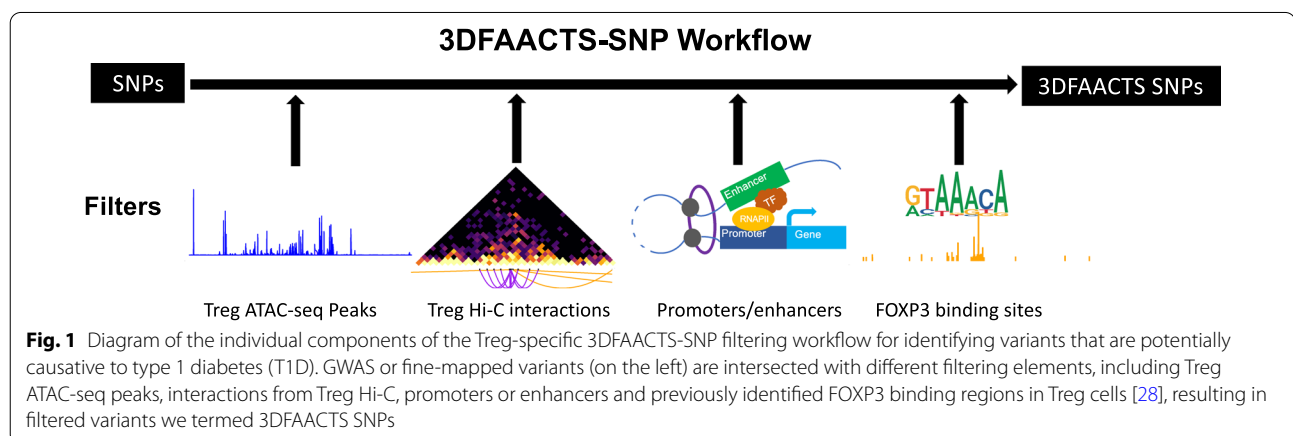
supported by the finding that T1D-associated SNPs are enriched at Treg-specific regulatory regions [25]. Therefore, starting with published T1D GWAS variant information, we designed a filtering workflow (Fig. 1) using multiple human Treg-specific epigenomic data to identify perturbations within defined “regulatory T cell active regions”.

In order to obtain highly accessible chromatin regions in Treg, we performed transposase-accessible chromatin using sequencing (ATAC-seq) on resting and stimulated Treg cells from three donors and sequenced to an average of 37.1 million read pairs (± 4 million) per sample, with TSS enrichment score from 8.13 to 25.7 (Additional file 1: Fig. S1 and Table S1). From the ATAC-seq data, we identified $\sim 577,866$ ATAC-seq peaks on average, with over 100,000 and 200,000 peaks shared by the replicates in resting and stimulated Tregs, respectively (Additional file 1: Fig. S2). These ATAC-seq peaks were then merged into 683,954 non-redundant peaks and used to screen for variants located in accessible regions in regulatory T cells as the first filtering step of the 3DFAACTS-SNP pipeline (Fig. 1).

Numerous studies have shown that three-dimensional (3D) interactions play important roles in gene regulation, mediated by DNA looping bringing enhancers and promoters together at transcriptional hubs [26–28]. As a result, distant loci which physically interact with disease-associated regulatory regions can be potentially impacted by these regions. To identify 3D interaction in Treg cells, we generated and sequenced an in situ Hi-C library of Tregs to 1.3 billion read pairs and identified 345,259 statistically significant Hi-C interactions at 5-kb bin resolution after a series of stringent filtering processes (see “Methods”) (Additional file 1: Table S2). These Hi-C interactions were then used to connect ATAC-seq peaks containing disease-associated variants to other regions in Tregs.

To assign potential function to these variant associated ATAC-seq peaks and Hi-C interacting regions, we next determined the overlap of these regions using enhancer and promoter annotations. This included 113,369 enhancers (mean size of 698 bp) identified by the Functional Annotation of the Mammalian Genome (FANTOM5) project [29] and promoter regions ($n = 73,171$) associated with GRCh38/hg38 UCSC known transcripts. Promoters were defined by extending the sequence 2 kb upstream of transcription start sites (TSS). Additionally, we extended the list of regulatory regions using the 15-state chromHMM model for CD4+CD25+CD127– primary Treg cells from the Roadmap Epigenomics Project [30]. We defined chromHMM states *EnhG*, *Enh* and *EnhBiv* as enhancers and *TssA*, *TssAFlnk*, *TssBiv* and *BivFlnk* as promoters. FANTOM5 enhancers and defined promoters and chromHMM enhancers/promoters states were then merged, respectively, to represent all possible genetic regulatory elements, covering 7.49% of the genome (Additional file 2).

The transcription factor FOXP3 is critical for Treg function and orchestrating immunological tolerance, and stable high FOXP3 expression levels are observed specifically in Tregs [3, 31]. Therefore, by intersecting filtered SNPs with significant human FOXP3-binding signals, we can largely constrain SNPs within regulatory regions to FOXP3 controlled Treg-specific gene networks. In the pipeline, we used 8,304 FOXP3 ChIP-chip peaks (mean size = 1317 bp) from our previous study [31] to specify FOXP3 binding in human Tregs. Of interest, by searching the Gene and Autoimmune Disease Association Database (GAAD) [32], we obtained 245 annotated genes that are associated with T1D, and found a significant enrichment of FOXP3 binding sites in T1D-associated genes (Fisher exact test: P -value = 4.519×10^{-9}), suggesting a strong association between T1D risk and FOXP3 controlled Treg function. Taken together, FOXP3 binding,



physical interaction, regulatory element and open chromatin regions offer a large subset of regions to use for GWAS variant prioritisation and functional annotation experiments.

Linking fine-mapped T1D-associated variants to their targets via chromatin interactions

Genetic studies have identified over 50 candidate gene regions that contain potentially causative SNPs that impact T1D [11]. Recently, 1228 putative causal variants associated with T1D (99% credible set) were identified from a study by using ImmunoChIP and a Bayesian fine-mapping method [13]. We used our workflow to further prioritise variants from this fine-mapped set to investigate potentially causative SNPs that contribute to T1D via affecting promoter/enhancer interaction in human Treg cells.

From the 1228 fine-mapped T1D-associated SNPs, we identified 36 variants that meet our filtering criteria: a variant must be in open chromatin region, where the chosen Hi-C interactions are contacting, overlapping with either a promoter or an enhancer region and bound by FOXP3 binding sites, in this study we will refer to them as T1D 3DFAACTS SNPs. To demonstrate the biological relevance of the process and the identified T1D 3DFAACTS SNPs, we performed 100 permutations with the workflow to confirm that the identification of 36 T1D-specific SNPs was significantly higher than by random chance (Fisher's exact test, average P -value $7.22e-08$) (Additional file 1: Fig. S3). These variants are located at 14 different chromosomal loci and distally interact with a further 266 regions three-dimensionally in Tregs (Table 1 and Additional file 3: Table S3). Most variants (71.4%, 25 out of 35 SNPs) were in enhancer regions rather than promoters while one variant, rs614120 is located in both the *TssAFlnk* chromHMM state and T cell-specific enhancers from FANTOM5. Given that a *TssAFlnk* state can either indicate a promoter or enhancer [33], combining with the identified FANTOM enhancer information, we believe that rs614120 is more likely to be located within an enhancer region.

Of the 14 loci identified, 8 contained more than two plausible variants across the loci. For example, variants located near the CD69 gene on chromosome 12 had the highest number of filtered variants, with 9 variants located in regulatory regions around the gene. In order to annotate the filtered variants to nearby genes, we took two approaches: annotated genes that were located in proximity to the SNPs using linear, chromosomal distances, and genes identified by their interaction with variant-containing regulatory regions via Treg Hi-C interactions (Table 1). Genes proximal to the identified 36 T1D variants include CTLA4, CCR5, IL7R, BACH2, IKZF1, IL2RA, CD69, RASGRP1, CCR3, CCR2,

CLEC16A, HORMAD2 and PTPN2. These genes have previously been linked to T1D due to their proximity [13] and in addition are associated with other autoimmune disorders such as Multiple Sclerosis (MS), Rheumatoid Arthritis (RA), Crohn's Disease (CD) and Inflammatory Bowel Disease (IBD) [13, 34–37]. Additionally, we annotated the filtered variants using significant cis eQTL data across all tissues from the Genotype-Tissue Expression (GTEx) project [38]. We found that 16 filtered SNPs are annotated as the eQTL to their nearest loci while 15 filtered SNPs are annotated as the eQTL to their 3D interacting genes (Additional file 3: Table S3). These data confirmed the ability of 3DFAACTS-SNP to identify potential disease-associated regulatory region–target gene networks in a cell type-specific manner.

In addition to the annotation of the 36 T1D SNPs to 14 genes in closest linear proximity, 3DFAACTS-SNP identified 266 interacting regions and a further 47 genes that interact with the variant-containing regulatory regions via Treg Hi-C (Table 1 and Additional file 3: Table S3). We next used the 15 states regulatory model for CD4+CD25+CD127– Treg primary cells from the Roadmap Epigenomics Project [30] to annotate the interacting regions. Of 266 regions, 145 of them overlap with the chromHMM states that associate with transcription and gene regulation, such as transcription (4_Tx) and weak transcription states (5_TxWk), which overlaps with 71 and 124 interacting regions, respectively, and enhancers states (7_Enh), which overlaps with 88 interacting regions. We then used the Tregs chromHMM states, induced Treg super-enhancers from SEDb [39] and Tregs expressed genes from T1D patients [40] to annotate the 3D interacting genes. Of these 47 interacting genes, 38 overlapped with a Treg active chromHMM state, 14 overlapped with Treg super-enhancers and 25/47 are differentially expressed in Tregs in T1D patients [40] (Additional file 3: Table S4). Of these 25 genes 8 (CTLA4, ICOS, CCR5, BACH2, IL2RA, RASGRP1, CLEC16A and PTPN2) have been shown to be significantly associated with T1D previously [32] while our analysis has identified a further 17 new candidate genes that may be disrupted in a Treg in T1D. These data indicate that distal interacting regions contain regulatory regions and genes important for Treg function and are consistent with a model in which the variant-containing regulatory regions may contribute to T1D by disrupting the regulation of these distal interacting genes.

The topological neighbourhood surrounding filtered T1D variants

We next investigated the topological neighbourhood, i.e. the presence of topologically associated or frequently interacting domains, in which regulatory regions harbouring the filtered T1D variants reside. By establishing

Table 1 T1D 3DFAACTS SNPs identified using the 3DFAACTS-SNP filtering workflow from T1D fine-mapping SNPs

Chromosome	Position	SNP id	Nearest locus (linear distance)	Located within regulatory regions		Interacting genes (3D)
				Treg ChromHMM	FANTOM5 expressed enhancers	
chr2	204700689	rs12990970	CTLA4	TssAFlnk		CD28,CTLA4,ICOS
	204732714	rs231775		TssAFlnk		RAPH1,CD28,ICOS
	204738919	rs3087243		EnhG		RAPH1
chr3	46327588	rs11718385	CCR3	Enh		XCR1,CCR2,CCR5AS,CCR5
	46391390	rs6441972	CCR2	TssAFlnk		CCR3
	46401032	rs3138042		Enh		
	46411661	rs2856758	CCR5	Enh		SLC6A20,FYCO1,CXCR6,XCR1, LOC105377067,CCR3,CCR1,CCR2
	46412259	rs1799988		TssAFlnk		SLC6A20, FYCO1,CXCR6, XCR1, LOC105377067,CCR3,CCR1,CCR2
chr5	35852311	rs6890853	IL7R	TssAFlnk		SPEF2,LOC105374724,CAPSL
chr6	90948476	rs62408222	BACH2	Enh		BACH2,LOC105377891
	90983850	rs905671		Enh	✓	BACH2
	90984035	rs943689		Enh	✓	BACH2
	90995980	rs614120		TssAFlnk	✓	BACH2,LOC105377891
chr7	50462418	rs10216316	IKZF1	EnhG		SPATA48
	50462498	rs10215297		EnhG		SPATA48
	50465206	rs55981617		EnhG		SPATA48,ZPBP,IKZF1
	50465654	rs12670555		EnhG		SPATA48,ZPBP,IKZF1,
						RMB17
chr10	6088743	rs12722508	IL2RA	TssAFlnk		IL2RA,RMB17,PFKFB3,LINC02649
	6094697	rs61839660		TssAFlnk		IL2RA,RMB17,PFKFB3,LINC02649
	6096667	rs12722496			✓	IL2RA,RMB17
	6107534	rs11597367		Enh		LOC374443,CLEC2D
chr12	9910720	rs3176793	CD69	TssA		LOC374443,CLEC2D
	9912182	rs2160086		TssA		LOC374443,CLEC2D
	9912730	rs3176789		TssA		LOC374443,CLEC2D,LOC105369728
	9916640	rs3136559		Enh		LOC374443,CLEC2D,LOC105369728
	9925758	rs1029992		Enh		
	9926064	rs1029991		Enh		
	9926397	rs1029990		Enh	✓	
	9926624	rs10844749		Enh		
	9926784	rs1540356		Enh		
chr15	38903672	rs16967112	RASGRP1	Enh	✓	FAM98B,RASGRP1,LOC105370775, LOC105370780,LINC02694,FSIP1
	38903884	rs56249992		Enh		RASGRP1,FAM98B, LOC105370775, LOC105370780,LINC02694,FSIP1
chr16	11188949	rs71136618	CLEC16A	Enh		CLEC16A,LOC107984859, LOC105371082,RMI2,SOCS1
chr17	38755665	rs11656173	SMARCE1	Enh	✓	CCR7
chr18	12838767	rs17657058	PTPN2	Enh		LINC01882,PTPN2,SEH1L
chr22	30581722	rs5753037	HORMAD2	Enh		HORMAD2,LIF-AS1,LIF

The nearest locus indicates the closest gene to the variants in linear distance, while 3D interacting genes are genes in contact with the variants via Treg Hi-C interactions. Overlapped regulatory elements of each 3DFAACTS SNPs are displayed, including chromatin states from a 15-state model and expressed enhancers from FANTOM5. Detailed SNP and interaction information is contained in Additional file 3: Table S3

*Genes in bold indicate novel 3D interacting genes of the identified SNPs

putative boundaries of each 3D structural domain, we are then able to characterise the coordination of contacts within a locus and how they act to control gene expression. We called topologically associated domains (TADs) using Treg Hi-C interactions of 20-kb resolution (Additional file 3: Table S5) and integrated with publicly available super-enhancer, chromHMM data of T cell lineages and Treg expression data [41]. All data were overlapped across each locus and displayed in Figs. 2, 3, 5 and Additional file 1: Fig. S4–S12.

Chromatin interactions between genes and enhancer regions was detected within the variant-containing TADs, with these interactions both confirming previously identified SNP–target combinations and indicating potential new targets for investigation. For example, 3DFAACTS-SNP identified rs12990970 (chr2:203,835,966) as a potential causative T1D SNP. In Treg cells, rs12990970 is found in a flanking active TSS (TssAFlnk) state and it is located within a Treg super-enhancer (Fig. 2). This variant is located in a non-coding region between gene CTLA4 and CD28 and in past studies, and it has been associated with CTLA4 as it is an eQTL for CTLA4 expression in testis, although not in T lymphocytes or whole blood (Additional file 3: Table S3) [11, 13, 38, 42]. While Hi-C interaction signals indicate that the rs12990970-containing region interacts with the CTLA4 promoter in Treg, additional Hi-C interactions indicates that this region also form interactions with the promoter and enhancer regions connected to the costimulatory receptor CD28 gene and the inducible costimulatory ICOS gene as well (Table 1 and Fig. 2). CD28 is a family member known to play a critical role in Treg homeostasis and function [43] and ICOS is believed to play an essential part in the suppressive function of Tregs [44], suggesting CD28 and ICOS are potential novel functional targets for this variant in Treg in T1D. Finally, additional contacts between the regulatory regions harbouring variants rs231775 and rs3087243, respectively, and the RAPH1 gene were identified. In mice the adaptor protein RAPH1 has recently been shown to play an important Treg-specific role in integrin activation, Treg-suppressive function and Treg homing to the gut in a mouse model of colitis [45] suggesting changes in RAPH1 expression associated with regulatory regions harbouring variants rs231775 and rs3087243 may contribute to Treg defects in humans.

Another example of novel T1D-linked functional annotation is on chromosome 3, where Hi-C interactions indicated that the chemokine receptor genes CCR1, CCR2, CCR3, CCR5 and CCR9 (Fig. 3) are extensively linked in a set of continuous TADs, indicating that these genes may be coordinately regulated. This is supported by previous RNA Pol-II ChIA-PET work [46] that detected

interactions between chemokine gene clusters during immune responses including an increase in interactions amongst the CCR1, CCR2, CCR3, CCR5 and CCR9 genes during TNF stimulation of primary human endothelial cells [46] (Additional file 1: Fig. S13). In support of this we found extensive Hi-C contacts between variant associated regulatory regions and CCR1, CCR2, CCR3 and CCR5 genes in Treg cells. Recently, CCR2, CCR3 and CCR5 have been shown to have additional chemotaxis-independent effects on Treg cells with individual studies, reporting positive roles for individual chemokine receptors on CD25, STAT5, and FOXP3 expression and Treg potency [47–49], highlighting the importance of multiple genes at this locus on Treg function. The chemokine receptor XCR1 gene also contacted by regulatory regions harbouring T1D associated variants rs11718385, rs2856758 and rs1799988 has also been implicated in Treg defects in human allergic asthma, with reduced XCR1 expression on CD4+CD25highCD-127low/– regulatory T cell (Treg) shown to be associated with impaired regulatory function [50].

Filtered T1D variants are enriched at lineage-specific T cell super-enhancers

SEs usually consist of a cluster of closely spaced enhancers that are defined by their exceptionally high level of transcription co-factor binding and enhancer-associated histone modifications (i.e. H3K27ac) compared to all other active enhancers within a specific cell type [51]. SEs are also linked to the control of important processes such as cell lineage commitment, development and function [52]. Analysing T cell SE information annotated in the Super-Enhancer Database [39] (SEdb), 8 out of the 14 variant-containing loci were found to contain filtered T1D variants located in SEs formed in various T cell lineages including Treg cells consistent with the enrichment of autoimmune-disease associated variants within T cell super enhancers reported previously [52] (Fig. 4A). The loci containing the CTLA4 and CLEC16A genes were the only loci that overlapped with Treg-specific SEs. The existence of a Treg SE is consistent with the different regulation of CTLA4 in Treg cells compared with other T cell lineages [53] and a recent report linking T1D risk variants to altered CLEC16A expression in Treg [40]. Five other SNPs are located within SEs in multiple T cell types including induced Treg (iTreg) suggesting the gene controlled by these SE play a broad role in T cell function. While no Treg SEs are detectable at the CD69 locus the T1D associated variants in this region overlapped with SEs formed in other T subsets. No T cell-associated SEs are found in the loci containing the CCR1/2/3/5, PTPN2, RASGRP1 and HORMAD2 genes (Fig. 4A).

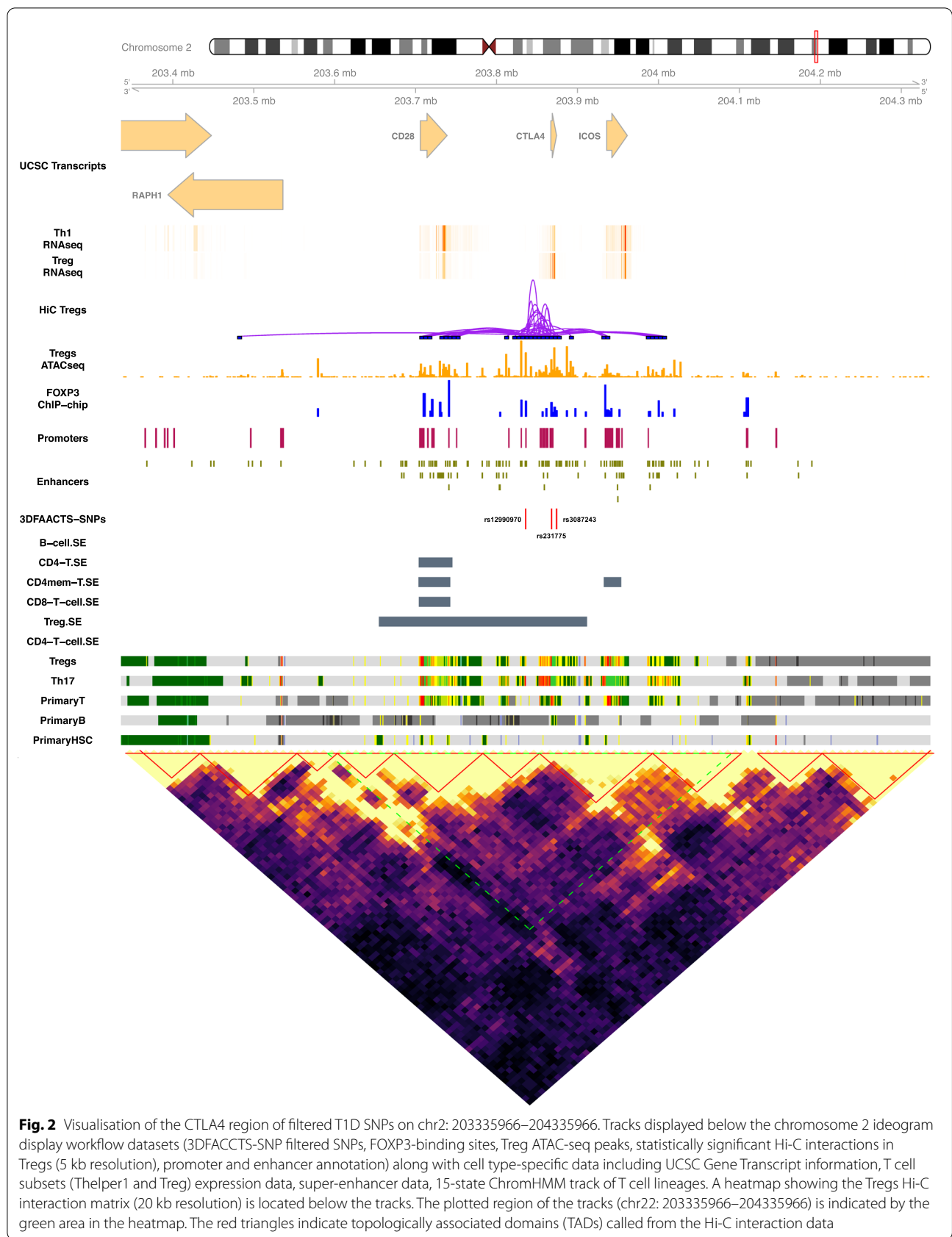


Fig. 2 Visualisation of the CTLA4 region of filtered T1D SNPs on chr2: 203335966–204335966. Tracks displayed below the chromosome 2 ideogram display workflow datasets (3DFACCTS-SNP filtered SNPs, FOXP3-binding sites, Treg ATAC-seq peaks, statistically significant Hi-C interactions in Tregs (5 kb resolution), promoter and enhancer annotation) along with cell type-specific data including UCSC Gene Transcript information, T cell subsets (Thelper1 and Treg) expression data, super-enhancer data, 15-state ChromHMM track of T cell lineages. A heatmap showing the Tregs Hi-C interaction matrix (20 kb resolution) is located below the tracks. The plotted region of the tracks (chr22: 203335966–204335966) is indicated by the green area in the heatmap. The red triangles indicate topologically associated domains (TADs) called from the Hi-C interaction data

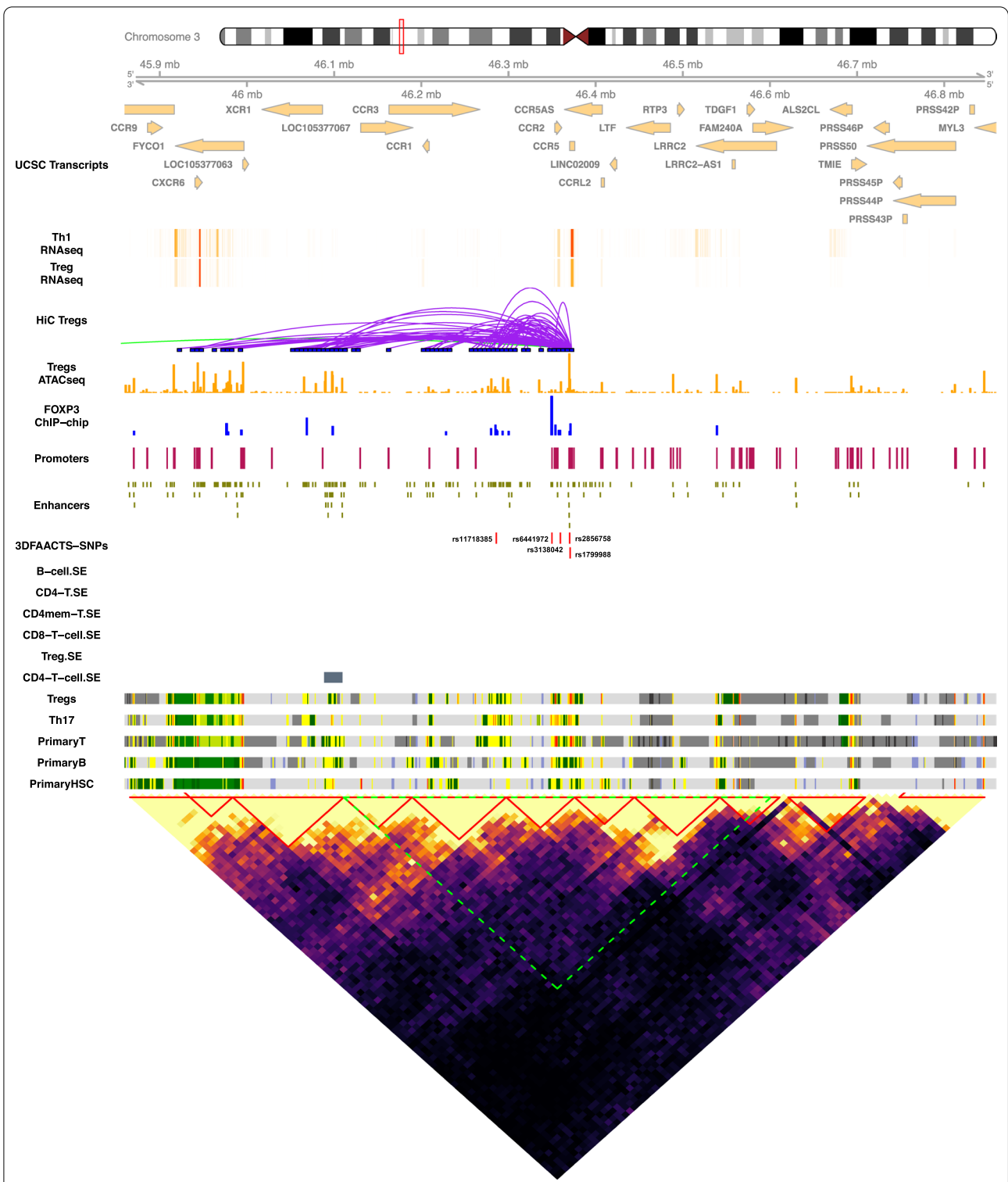


Fig. 3 Visualisation of the CCR3/2/5 region of filtered T1D SNPs on chr3: 45859541–46859541. Tracks displayed below the chromosome 3 ideogram display workflow datasets (3DFACTS-SNP filtered SNPs, FOXP3-binding sites, Treg ATAC-seq peaks, statistically significant Hi-C interactions of Tregs (5 kb resolution), promoter and enhancer) annotation along with cell type-specific data including UCSC Gene Transcript information, T cell subsets (Thelper1 and Treg) expression data, super-enhancer data, 15-state ChromHMM track of T cell lineages. A heatmap showing the Tregs Hi-C interaction matrix (20 kb resolution) is shown below the tracks. The plotted region of the tracks (chr3: 45859541–46859541) are indicated by the green area in the heatmap. The red triangles indicate the topologically associated domains (TADs) called from the Hi-C interaction data

We then investigated the level of active enhancer marks (normalised H3K27ac-binding) and chromatin accessibility (normalised ATAC-seq peak coverage) overlapping each variant from Table 1 (Fig. 4B). A range of tissue restriction patterns of chromatin states were observed using the NIH Epigenomics Roadmap data with enhancers displaying in general a more cell type-restricted pattern of H3K27ac signal compared to promoters. No variant was found to be located in a regulatory region that

was exclusively active in Treg cells although rs12990970, rs231775 (CTLA4), rs11597367, rs12722508 (IL2RA) and rs5753037 (HORMAD2) are associated with a restricted H3K27ac pattern that included Treg. The absence of Treg-specific enhancers is consistent with FOXP3 binding data where FOXP3 binds many enhancer regions active in other T cell lineages to modify their activity in Treg cells [54]. Evidence suggests FOXP3 cooperates with other T helper-lineage specifying transcription factors

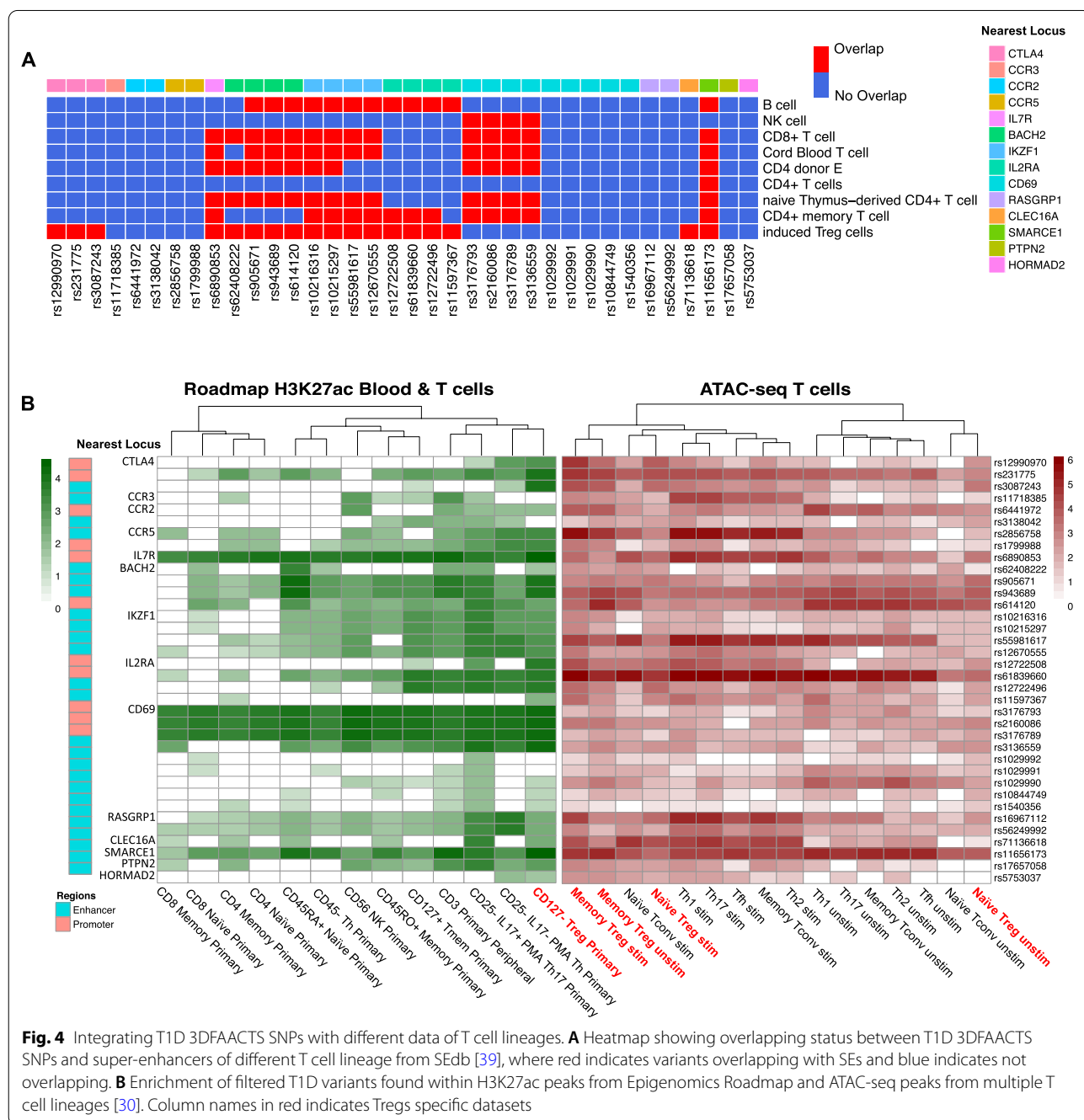


Fig. 4 Integrating T1D 3DFAACTS SNPs with different data of T cell lineages. **A** Heatmap showing overlapping status between T1D 3DFAACTS SNPs and super-enhancers of different T cell lineage from SEdb [39], where red indicates variants overlapping with SEs and blue indicates not overlapping. **B** Enrichment of filtered T1D variants found within H3K27ac peaks from Epigenomics Roadmap and ATAC-seq peaks from multiple T cell lineages [30]. Column names in red indicates Tregs specific datasets

to diversify Treg cells into subsets that mirror the different Th-lineages [55–57]. Most regions associated with the variants show an increase in chromatin accessibility upon stimulation in Treg and T helper subsets consistent with increased enhancer activity upon T cell activation, however, in a few instances variants are in regions that decrease in accessibility in stimulated Treg and T helper subsets compared with their matched unstimulated counterpart. Notably these include the variants rs905671, rs943689 and rs614120 associated with BACH2. This is consistent with the reduction in BACH2 expression in CD4 T cells as they mature, and alteration to this repression is linked to proinflammatory effector function [58]. Together these data are consistent with a model in which causal variants alter the output of enhancers that respond to environmental cues [59].

Filtered variants disrupt transcription factor binding sites (TFBS) including a FOXP3-like binding site

Fundamental to understanding the function of specific disease-associated variants is the identification of the potential impact of these non-coding variants on transcription factor binding. Analysis of ATAC-seq datasets with HINT-ATAC [60], identified over 5 million active TF footprints in chromatin accessibility profiles from stimulated and resting Treg populations (Additional file 4). By imposing the additional FOXP3 binding annotation to the footprint dataset, we identified 7 T1D-associated variants that have the potential to alter the binding of 9 TFs, suggesting the molecular mechanisms by which these variants could impact Treg function (Additional file 3: Table S6). Of these 7 SNPs, one SNP rs3176789 is located in an active TSS chromHMM state region, while the others are located either in enhancers or flanking active TSS that are associated with active enhancers, suggesting these variants might interrupt the binding of TFs to affect enhancer functions, with the potential for a network effect on multiple genes.

We then used GWAS4D [61], which computes log-odds of probabilities of the reference and alternative alleles of a variant for each selected TF motif to calculate binding affinity, to predict the regulatory effect of each variant (Additional file 3: Table S7). Several of the variants are predicted to alter the binding of transcription factors with known roles in Treg and other T cell lineages including nuclear activator of T cells (NFATC2 and NFATC3, rs1029991) [62], interferon regulatory transcription factor (IRF, rs3176789) [63], myocyte enhancer factor 2 (MEF2, rs6441972 and rs3176789) and FOX (Forkhead box, rs614120) family members. In addition, variant (rs1029991) has the potential to alter the binding of YY1, recently identified as an essential looping factor involved in promoter–enhancer interactions [64]. Other

variants (rs1136618 and rs3176789) potentially alter the binding of the zinc finger protein ZNF384. Although expressed in T cells, the importance of ZNF384 in T cell biology has not yet been explored.

Of note, rs614120 is predicted to decrease the binding affinity of FOXA2 in this enhancer region (Additional file 3: Table S8). As FOXA2 is not expressed in the immune compartment, this SNP may interfere with the binding of another member of the forkhead class of DNA-binding proteins, e.g. FOXP3, which is localised to this region based on our FOXP3 ChIP (Fig. 5). This suggests that a model in which rs614120 impacts the expression level of BACH2 and/or AFG1L is by altered binding of a FOX protein to this enhancer.

Filtered variant rs1029991 is predicted to alter the binding of NFAT family members and/or YY1 to the enhancer region, but we have been unable to link the associated regulatory region which harbours this variant to any gene. Filtered variant rs3176789 is predicted to alter IRF and/or MEF2 binding, linking these transcription factors to the regulation of the CLEC family member CLEC2D and two additional genes LOC374443 (C-Type Lectin Domain Family 2 Member D Pseudogene) and LOC105369728 a putative lncRNA class gene. Examination of gene expression data from Treg indicate that only CLEC2D is expressed in a Treg (Additional file 3: Tables S3 and S4). The CD69 and CLEC2D genes have previously been associated with T1D by GWAS, however we could not detect any significant interaction of regulatory regions harbouring 3DFAACTS-SNP filtered SNPs and the CD69 gene. Filtered variant rs6441972 is also predicted to influence the binding of MEF2 to a regulatory region in proximity to the promoter of CCR2. Consistent with this variant disrupting CCR2 expression, CCR2 is a target gene for eQTL rs6441972, indicating that rs6441972 may result in altered CCR2 expression in a Treg in T1D by interfering with MEF2 binding.

Filtered Treg variants identified in other autoimmune diseases

The primary rationale of our filtering workflow is that autoimmune diseases like T1D are mediated by altered Treg functions. Hence, using GWAS data for other autoimmune diseases, we aimed to discover variants which potentially act by disrupting 3D gene regulation in Tregs. Like filtering fine-mapped T1D-associated SNPs, here we used the 3DFAACTS-SNP filtering workflow to process variants identified by Immunochip fine-mapping experiments and meta-analysis from three studies for a broad range of autoimmune and inflammatory diseases. SNPs associated with 10 autoimmune diseases were identified, representing 221 fine-mapped SNPs associated with multiple sclerosis (MS) [65]; 69 SNPs identified by the

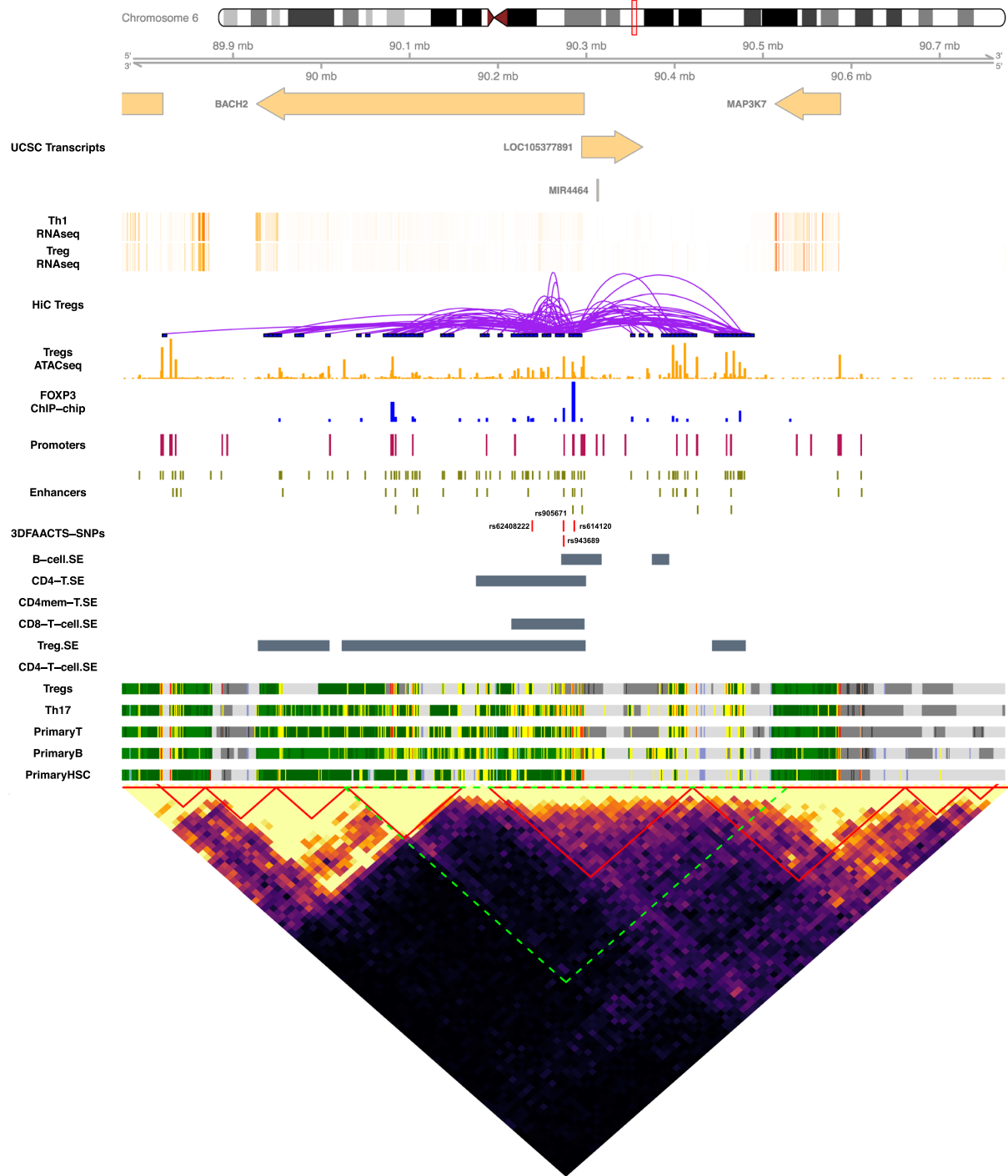


Fig. 5 A visualisation of the BACH2 region of filtered T1D SNPs on chr6: 89774131–90774131. Tracks displayed below the chromosome 6 ideogram display workflow datasets [3DFAACTS-SNP filtered SNPs, FOXP3-binding sites, Treg ATAC-seq peaks, statistically significant Hi-C interactions in Tregs (5 kb resolution), promoters and enhancers along with cell type-specific data including UCSC Gene Transcript information, T cell subsets (Thelper1 and Treg) expression data, super-enhancer data, 15-state ChromHMM track of T cell lineages. A heatmap showing the Tregs Hi-C interaction matrix (20 kb resolution) is shown below the tracks. The plotted region of the tracks (chr6: 89774131–90774131) are indicated by the green area in the heatmap. The red triangles indicate the topologically associated domains (TADs) called from the Hi-C data

meta-analysis of celiac disease (CeD), rheumatoid arthritis (RA), systemic sclerosis (SSc), and T1D [66] (which we refer to the 4AI dataset); and 244 SNPs identified by the meta-analysis of GWAS datasets for ankylosing spondylitis (AS), Crohn's disease (CD), psoriasis (PS), primary sclerosing cholangitis (PSC) and ulcerative colitis (UC) [67] (which we refer as 5ID dataset). Applying the 3DFAACTS-SNP pipeline we identified 9, 3 and 6 filtered variants from the MS, 4AI and 5ID datasets, respectively (Additional file 3: Table S9). We identified putative target genes for these disease associated variants by Hi-C interactions resulting in 34, 8 and 23 genes (60 unique genes) linked to MS, 4AI and 5ID, respectively. Of these 34/60 are differentially expressed in Tregs isolated from T1D patients compared to healthy controls highlighting the potential of 3DFAACTS-SNP to identify candidate SNP-target gene interactions in disease (Additional file 3: Table S9). Many of these genes have either known roles in Treg differentiation, stability and function (CD48, IL6ST, EZR, ELMO1, IL18R1) [68–72], or altered expression in human Treg in autoimmune disease (SLAMF1, ANKRD55, TAGAP) [73–75].

Of the variants identified by 3DFAACTS-SNP, one variant (rs60600003) located at a locus on chromosome 7 was found to be associated with several diseases, including MS [65], celiac and systemic sclerosis [66], suggesting its interacting gene, ELMO1, may contribute to a common Treg defect in these diseases (Additional file 3: Table S9). When compared with the 36 variants identified from our T1D dataset analysis, two variants, rs61839660 on chromosome 10 and rs3087243 on chromosome 2 were also prioritised by 3DFAACTS-SNP analysis of the 5ID and 4AI datasets, respectively, implicating their interacting genes IL2RA, RBM17, PFKFB3 and LINC02649 (rs61839660), RAPH1 (rs3087243) are functionally implicated in the development of these diseases. While different variants were identified in the analysis of the various disease datasets, the regulatory elements in which these variants reside can be linked by Hi-C data to common candidate target gene such as PFKFB3 (rs12722496 from T1D and rs947474 from 4AI) and RAPH1 (rs231775 from T1D and rs3087243 from 4AI). This is consistent with the view that common mechanistic pathways underlie some autoimmune diseases, although the specific risk allele within a locus can be disease-specific [76].

Identifying new variants that are candidates for impacting autoimmune disease

Most variants identified by GWAS have small effect sizes that together only represent a fraction of the heritability predicted by phenotype correlations between relatives [77]. To account for this missing heritability, various models have been proposed including a highly polygenic

architecture with small effect sizes of the causal variants [78, 79], rare variants with large effect size [80, 81] and epistatic mechanisms including gene–gene and gene–environment interactions [82, 83]. As a consequence many causal variants with small effect sizes are unlikely to reach genome wide significance in current GWAS, whereas rare variants are often under-represented on SNP arrays [84]. Lastly the preponderance of studies utilise populations of European descent which can result in a bias for SNPs with a higher minor allele frequencies in Europeans compared to other populations, potentially limiting the relevance of these SNPs to the associated traits in non-Europeans [85]. As an alternative approach, to identify novel putative autoimmune disease-associated SNPs independently of association studies, we sampled 5,888,594 common variants (MAF > 0.1) from the Genome Aggregation Database (gnomAD) (version 3.0) [86] as inputs to our filtering workflow, identifying a total of 9376 gnomAD-3DFAACT SNPs (Additional file 3: Table S10).

In order to characterise the SNPs, we used *GIGGLE* [87] to compare the regions in which filtered SNPs reside against 15 predicted chromHMM genomic states across 127 cell types and tissues from Epigenomic Roadmap [30] (Fig. 6 left and Additional file 1: Fig. S14), calculating positive and negative enrichment scores according to overlapping sets. Interestingly, there was strong positive enrichment signal in active TSS (*TssA*), flanking active TSS (*TssAFlnk*) and enhancers (*Enh*) states in thymus, HSC, B- and T- cell groups, while only enrichment of *TssA* was observed across all cell types, and enrichment of *TssAFlnk* and *Enh* are seen in only immune-related cell types. This suggest that the enhancer regions related to our identified SNPs are highly specific to immune cells while the promoter regions and by extension their target genes are broadly expressed (Additional file 1: Fig. S14). Additionally, negative enrichment of the quiescent (*Quies*) states was observed in all cell types whereas heterochromatin (*Het*) exhibited a negative enrichment in some specific cell types (foreskin fibroblast primary cells) and cell lines (IMR90, HVEC and HMEC).

Treg Hi-C data were used to explore the FOXP3-associated regulatory networks that include these SNPs in a Treg. For the regions identified to interact with the 9,376 variants located in FOXP3 binding regions by Hi-C, we observed positive enrichments of regulatory states such as *TssA*, *TssAFlnk*, *Tx*, *Txwk*, *EnhG* and *Enh* in blood, HSC, B and T cells, supporting a regulatory role for these interacting regions (Fig. 6 right and Additional file 1: Fig. S15). 4,968 genes are found to interact with the identified filtered gnomAD-3DFAACTS SNPs via the Treg Hi-C interactions (Additional file 3, Table S10). Gene set enrichment analysis of these genes was performed using

the genes they affect. Open chromatin regions identified by ATAC-seq and regulatory regions identified by epigenetic marks such as histone H3K27ac can number in the tens of thousands in a specific cell type [40, 97], we therefore initially focused on regulatory regions bound by the Treg-specific transcription factor FOXP3 given the essential role of FOXP3 in the Treg functional phenotype we hypothesised that candidate variants that are found within open, FOXP3-bound regions are likely to alter immunological tolerance. In addition, as different autoimmune diseases share genetic risk regions [13], we speculated that by identifying specific genetic variants that may contribute to T1D through the dysregulation of regulatory T cell functional fitness, this could be via mechanisms consistent across many autoimmune diseases [1, 98, 99].

The design and implementation of the 3DFAACTS-SNPs workflow champions a new data-centric view of functional genomics analysis, with the development of cell type-specific epigenomic and 3D datasets enabling researchers to narrow down on molecular changes at a fine-scale resolution. However, the results of this study suggests that cell type-specific viewpoints can be broadened to a much more lineage (T cell) or immune (e.g. innate or adaptive) system-specific level. While we focused on Treg cells and expected to identify Treg-specific enhancer-controlled targets, based on the criteria of inclusion of FOXP3 binding data, no functional variant was uniquely accessible in only Tregs, nor were they specifically enriched with Treg-exclusive H3K27ac modified regions (Fig. 4B). This likely reflects the propensity of FOXP3 to bind to enhancers active in multiple CD4+ T cell lineages [54] (Fig. 4) to modify their output in a Treg-specific manner and therefore we cannot currently discern whether these filtered variants act predominantly in Tregs or on other CD4+ T cell subsets. The incorporation of context- and CD4+ T cell subset-specific gene expression [100] and epigenomic [94, 101] data into the 3DFAACTS-SNPs workflow may help resolve this. Although we have focused here on using FOXP3-binding as a filtering criteria, it is known that other FOXP3-independent pathways are important for Treg function and the 3DFAACTS-SNPs workflow could be modified to incorporate other TFs or other epigenetic profiles such as CpG-demethylated regions [102] to further explore the relationship between disease-associated variants and these pathways.

In total using the 3DFAACTS-SNPs workflow we identified 47 novel candidate genes connected to variants in 12 T1D risk loci that could plausibly function in a Treg whereas we could not define plausible candidate Treg-specific activity at the other T1D risk regions that met all our filtering criteria. This may indicate that these other

risk-regions are active in immune cell types other than a Treg or they impact genes and regulatory elements within a Treg that are not dependent upon FOXP3. As an example of how the 3DFAACTS-SNPs workflow can lead to testable insights into the molecular mechanisms of non-coding variants, the SNP rs614120 was found to be located in a FANTOM5 annotated T cell-specific enhancer region in the first intron of the BACH2 gene, and is predicted to disrupt the binding of Forkhead Transcription factor family member FOXA2 (Fig. 5 and Additional file 3: Table S6). However, FOXA2 is not expressed in T cells, indicating that rs614120 might disrupt the binding of other Forkhead family members which bind to very similar DNA sequences, such as FOXP3, which is known to bind in this region (Fig. 5). The 3DFAACTS-SNPs workflow further indicates that this enhancer region containing rs614120 interacts with the promoter of BACH2, forming a distal promoter–enhancer interactions, suggesting that rs614120 may disrupt FOXP3 binding to the enhancer leading to the dysregulation of BACH2 expression. It has been recently shown that Bach2 plays roles in the regulation of T cell receptor signalling in Tregs, including averting premature differentiation and assisting peripherally induced Treg development [103]. Therefore, we suggested that this single variant may regulate BACH2 expression and ultimately may affect the progression of T1D, and this requires further experiments to verify. This can further aid the development of novel therapeutic approaches to restore function in Treg of patients with this genotype. This finding also suggests that variants can contribute to the causal mechanisms of disease by altering the efficacy/stability of TF binding in important regions such as enhancers or SEs.

The power of 3DFAACTS-SNPs is its ability to incorporate chromosome organisation in 3D and identify long-range interactions involving variant-containing regulatory regions leading to the identification of target genes that have not previously been associated with these diseases associated risk regions. This is illustrated by the finding that the majority (39/46) of the genes that interact with the T1D variants are not the closest gene in linear proximity and of these interacting genes some have not been previously associated with any autoimmune disease.

The idea that high-order nuclear organisation coordinates transcription in times of immune challenge or tolerance was recently shown in a study demonstrating that 3D chromatin looping topology is important for a subset of long non-coding RNAs (lncRNAs), termed immune gene-priming lncRNAs (IPLs), to be correctly positioned at the promoters of innate genes [46]. This positioning of the IPLs then allows for the recruitment of the WDR5–mixed lineage leukaemia protein 1 (MLL1) complex to these promoters to facilitate their H3K4me3 epigenetic

priming [46]. An example of long-range enhancer gene interactions in conveying autoimmune-disease risk in Treg cells has also recently been published [104]. In this work a distal enhancer at the 11q13.5 locus associated with multiple autoimmune-disease risk, including T1D was found to participate in long-range interactions with the LRRC32 gene exclusively in Treg. Deletion of this enhancer in mice resulted in the specific loss of *Lrcc32* expression in Treg cells and the inability of Treg to control gut-inflammation in an adoptive transfer colitis model. Furthermore CRISPR-activation experiments in human Tregs identified a regulatory element located in proximity to a risk variant rs11236797 that is capable of influencing LRRC32 expression. This data together highlights the mechanistic basis of how non-coding variants may function to interfere with Treg activity in disease. This interaction was present in our Hi-C dataset, but it was filtered out as the enhancer is not bound by FOXP3, showing the cell type-specific filtering power, as well as conserved connectivity in T cells. Coordinated genome topology has also been shown in immune cell lineage commitment, both at a locus [105, 106] and compartment level [107], consistent with the concept of immune transcriptional “factories” where genes congregate in regions of the nucleus to undergo coordinated transcriptional activation [108].

Although a shared genetic aetiology between T1D and other immune-mediated diseases has been proposed, we did not find a large overlap between the variants or interacting genes identified by 3DFAACTS SNP in T1D and other autoimmune disease datasets. The reason for this is not clear, but may be a result of the relatively low number of input SNPs for the other autoimmune diseases. Irrespective of this, two candidate causal SNPs and genes including rs3087243 (*RAPH1*) and rs61839660 (*IL2RA*, *RBM17*, *PFKFB3*, *LINC02649*) were found to be common between T1D and other autoimmune diseases. Several of these genes such as *IL2RA* and *PFKFB3* have previously been implicated in the development of autoimmune diseases or play a role in critical T cell pathways, suggesting these genes are likely targets that explain the molecular function of the risk variants. *PFKFB3* is involved in both the synthesis and degradation of fructose-2,6-bisphosphate, a regulatory molecule that controls glycolysis in eukaryotes. Regulation of glycolysis has increasingly been implicated in shaping immune responses [109] and *PFKFB3* has been associated with multiple autoimmune diseases [110]. Importantly, reduced *PFKFB3* enzyme activity leading to redox imbalance and apoptosis has been reported in CD4+ T from RA patients [111] directly linking the *PFKFB3* gene to the disease. This provides molecular insight into environmental pressure on immune cell function driving loss of tolerance.

A highly polygenic architecture with small effect sizes of many causal variants [78, 79] has been proposed to account for missing heritability associated with phenotypic traits. Most of these small effect size variants have yet to be identified. Here we have begun to investigate whether common genetic variation found within populations could contribute to autoimmune diseases by altering gene-expression by altering enhancer and promoter output. In this study we illustrate this potential by accessing large population-scale variant resources in the gnomAD database, identifying 9376 filtered common variants that have the potential to impact Treg function. Based on the search of discovered associations of autoimmune diseases (EFO_0005140) from the GWAS Catalog [112], over half of the variants surveyed here have not been used in large-scale autoimmune disease GWAS [11, 67, 113–119], precluding their assessment for potential disease risk in sampled disease/control populations. While filtered variants identified here are biased towards the inclusion of FOXP3-binding within the workflow, their potential immune response impact is highlighted by the finding that their interacting regions are positively enriched for transcription and enhancer-associated chromatin states (Fig. 6, Additional file 1: Fig. S14 and S15). This accessibility of regulatory variants among a population could potentially explain additional variation in effector responses in T cell activation [120], relevant not only to autoimmune disease, but also to broader immune responses for example to SARS-CoV-2.

In conclusion, while we initially restricted the application of 3DFAACTS-SNP to Treg-centric genome-wide interaction frequency profiles to give functional annotation in T1D data, we have demonstrated that valid interacting pairs from Hi-C dataset can be functionally mapped with high confidence from multiple disease datasets as well as whole genome variant datasets, which presents a valuable resource in establishing cell type-specific interactomes. Coupled with cell type-specific genomic data available from public repositories, such as the NIH Roadmap [30], Blueprint [121] and ENCODE [122] projects, this workflow provides a useful mechanism to identify potential mechanisms by which non-coding variants regulate disease causing genes, and identifies new targets for therapeutic modulation to treat or prevent disease.

Conclusion

Based on Treg ATAC-seq, Hi-C data, promoter and enhancer annotation and FOXP3 binding site annotation, we have developed a variant filtering workflow named 3DFAACTS-SNP to identify potential causative SNPs and their 3D interacting genes for T1D from GWAS fine-mapped variants. Our workflow can easily be used with

variants associated with other autoimmune diseases or even large population-scale variants.

Methods

Cell preparation

Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood obtained from healthy human donors with informed consent at the Women's and Children's Hospital, Adelaide (ethics approval and consent see "Declarations" section). Cells were labelled with the following fluorochrome conjugated anti-human monoclonal antibodies: anti-CD4 (BD Biosciences, BUV395 Mouse Anti-Human), anti-CD25 (BD Biosciences, BV421), anti-CD127 (BD Biosciences, PE-CF594) and viability dye (BD Biosciences, BD Horizon Fixable Viability Stain 700) for FACS analysis by surface expression staining. Regulatory T (Treg) cells were sorted as CD4⁺CD25^{hi}CD127^{dim} population (>90% purity). Following cell sorting Treg cells were plated at 100,000 cells per well in a 96-well U-bottom plate and maintained in complete X-VIVO 15 culture media (X-VIVO 15 Serum-free media supplemented with 2 mM HEPES pH 7.8, 2 mM L-glutamine and 5% heat inactivated human serum) in 400 U/mL rIL-2 for 2 h at 37 °C in a humidified 5% CO₂ incubator prior to cell preparation for ATAC-seq experiment.

ATAC-seq library preparation and high-throughput sequencing

Treg cells were rested for 2-h post-sort and then were either left untreated or stimulated with beads conjugated with anti-CD3 and anti-CD28 antibodies (Dynabeads Human T-Expander CD3/CD28, Gibco no. 11141D, Life Technologies) in complete X-VIVO 15 culture in 400 U/mL rIL-2 at a cell/bead ratio of 1:1 for 48 h. After 48 h Dynabeads were removed from culture medium by magnetic separation. Omni ATAC-seq was then performed as described previously [154] with minor modifications. Briefly, cells with 5–15% dead cells were pre-treated with 200 U/μL DNase (Worthington) for 30 min at 37 °C prior to ATAC-seq experiments. Treg cells (50,000) were lysed in 50 μL of cold resuspension buffer (RSB: 10 mM Tris-HCl pH 7.4, 10 mM NaCl, and 3 mM MgCl₂) containing 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin on ice for 3 min. The reaction was then washed with 1 mL of ATAC-seq RSB containing 0.1% Tween-20 by centrifugation at 500 xg for 10 min at 4 °C and the nuclei were resuspended in 50 μL of transposition mix (30 μL 2 × TD buffer, 3.0 μL Tn5 transposase, 16.5 μL PBS, 0.5 μL 1% digitonin and 0.5 μL 10% Tween-20) (Illumina Inc). The transposition reaction was incubated at 37 °C for 45 min in a thermomixer with 1000 rpm mixing. The reaction was purified using a Zymo DNA Clean and Concentrator-5 (D4014) kit. All libraries were amplified for a total

of 9 PCR cycles and size selection was carried out to enrich for a fragment size window of 200–900 bp prior to sequencing. Libraries were quantified by PCR using a KAPA Library Quantification Kit for NGS (KAPA Biosystems, Roche Sequencing). Barcoded libraries were pooled and sequenced on a paired-end 75-cycle Illumina Next-Seq 550 High-Output platform (Illumina) to an average read depth of 37.1 million read pairs (±4 million) per sample.

Treg sample preparation, Hi-C library production and high-throughput sequencing

Cord blood was obtained with informed consent at the Women's and the Children's Hospital, Adelaide (HREC1596; WCHN Research Ethics Committee). Mononuclear cells were isolated from cord blood postpartum as previously described [123]. Briefly, cord blood CD4⁺CD25⁺(Treg) were isolated from purified mononuclear cells using a Regulatory CD4⁺CD25⁺T Cell Kit (Dynabeads; Invitrogen, Carlsbad, CA). Ex vivo expansion of isolated T cell populations (1 × 10⁶ cells per well in a 24-well plate) were performed in X-Vivo 15 media supplemented with 5% human AB serum (Lonza, Walkersville, MD), 20 mM HEPES (pH 7.4), 2 mM L-glutamine, and 500 U/mL recombinant human IL-2 (R&D Systems, Minneapolis, MN) in the presence of CD3/CD28 T cell expander beads (Dynabeads; Invitrogen; catalogue no. 111-41D) at a bead-to-cell ratio of 3:1. Cell harvesting, formaldehyde cross-linking (2%) and nuclei isolation was per [156,157]. Treg cell nuclei were frozen in aliquots of 1 × 10⁷. The in situ Hi-C procedure was carried out as per Rao et al., (2014) [124] with the following modifications MboI digestion was carried out in CutSmart[®] Buffer (NEB) and biotin-14-dCTP (Invitrogen; catalogue no. 19518018) replaced biotin-14-dATP in the reaction to end-fill MboI overhangs. To generate DNA suitable for library construction ligated DNA in TE buffer (10 mM Tris-HCl, pH8.0 and 0.1 mM EDTA, pH 8.0) was sheared to an average size of 300–500 bp using a Covaris S220 (Covaris, Woburn, MA) instrument with the following parameters; 130ul in a microTube AFA fibre, 140 peak incidence power, 10% Duty cycle 10%, 200 cycles per burst for 55 s. Sheared fragment ends were made suitable for adapter ligation with a NEBNext[®] Ultra II End Repair/dA-Tailing Module (NEB #E7546). For adapter ligation the End Prep reaction was split into two and appropriately diluted NEBNext Adapter ligated to fragment ends using the NEBNext Ultra II Ligation module. Library size distribution was determined using an Experion DNA 1K kit and library concentration estimated by real-time qPCR using a Kapa universal Library quantitation kit (Roche Sequencing Solutions; 07960140001). Hi-C libraries were sequenced on an Illumina NovaSeq[™] 6000 Sequencing System (2 × 150 bp).

ATAC-seq data analysis

The sequencing data quality was determined using *FastQC* (ver. 0.11.7) [159] followed by trimming of Nextera adapters using *cutadapt* (ver. 1.14) [125]. Trimmed reads were aligned to the human hg19 (hs37d5) reference genome using *Bowtie2* (ver. 2.2.9) [126] with ‘-X 2000’ setting. For each sample quality trimming was performed with option ‘-q 10’ with unmapped and non-primary mapped reads filtered with option ‘-F 2828’ using *Samtools* (ver. 1.3.1) [127]. PCR duplicates were then removed from Uniquely mapped paired reads using *Picard* (ver. 2.2.4). Mitochondrial reads, reads mapping to ENCODE hg19 blacklisted regions and mitochondrial blacklisted regions were filtered out using *BEDTools* (ver. 2.25.0). The TSS enrichment score for each replicate was determined using *ATACseqQC* [128]. For peak calling the read start sites were adjusted to represent the centre of Tn5 transposase binding event. Peaks were called from ATAC-seq data using *MACS2* (ver. 2.1.2) [129] with parameters ‘-atac-seq -paired-end -organism=hg19 -P 0.05’ [130, 131] and *HINT-ATAC* [60] was used to call footprints from the ATAC-seq peaks with parameters ‘-atac-seq -paired-end’.

The peak summits from resting and stimulated Treg were concatenated and sorted by chromosome and then by position. The sorted peak summits were then handled using an in-house Python script *ATACseqCollapsing.py*, which adapted a peak processing approach described by Corces et al. [132] to generate a list of non-redundant peaks. Briefly, through an iterative procedure, the peak summits are extended by 249 bp upstream and 250 bp downstream to a final width of 500 bp. Any adjacent peak that overlaps with the most significant peak (significance value defined by *MACS2*) within the interval is removed. This process iterates to the next peak interval resulting in a list of non-redundant significant peaks. Finally, to be consistent with other annotations in the 3DFAACTS-SNP workflow, the ATAC peaks were lifted over to the hg38 genome.

Hi-C data analysis

The raw sequencing read files were first aligned to the human hg38 genome using *BWA-mem* with setting ‘-SP5M’. The mapped read pairs with mapping quality (MAPQ) over 30 were selected to process with *pairtools* [133] to identify Hi-C interactions. The interactions are then mapped to genomic bins of 5 kb and 20 kb resolution using *cooler* [134]. The contacts of 5 kb bins were further processed to identified intra-chromosomal statistically significant (BH adjusted *P*-value < 0.05) interactions over a background model using *MaxHiC* [135]. Finally, a count filter (at least 5 sequencing read pairs mapped to the interaction) and a distance filter (distance between two interacting bins must larger than 5 kb) were

applied to the identified significant interactions to select Tregs specific chromatin interactions, which were used in the 3DFAACT-SNP workflow.

RNA-seq data processing

The raw sequencing data were first trimmed using *AdapterRemoval* (ver 2.2.1a) [136] with default parameters to remove sequencing adapters. Trimmed reads were then aligned to hg38 using *STAR* (ver 2.7.0d) [137]. The resulting BAM files were converted into bedgraph files using *bamCoverage* from *deepTools* [138] with count normalised using counts per million mapped reads (CPM).

Topologically associated domain identification

Hi-C interactions of Tregs were mapped to equal-size bins (20 kb) of the hg38 genome and normalised using *ICE* [32], resulting in a normalised interaction matrix. The matrix was then used as input to identify topologically associated domains (TADs) via *TopDom* [139].

Visualisation and downstream analyses

Gene set enrichment analysis (GSEA) was performed using a function *enrichr* from the R package *clusterProfiler* [140] with the hallmark gene sets and C5 gene sets (gene ontology terms) from Molecular Signatures Database (MSigDB). Adjusted *P*-value (Benjamini–Hochberg adjusted) of 0.05 is set as the significant threshold. Visualisation of normalised Hi-C interaction matrices (Figs. 2, 3, and 5, Additional file1: Fig. S4–S12) was performed on 40 kb resolution using an in-house R function *hicHeatmap*. The visualisations of individual filtered T1D-associated SNP loci (Figs. 2, 3, 5 and Additional file 1: Fig. S4–S13) were constructed using the R packages *Gviz* [141], *GenomicInteractions* [142] and *coMET* [143].

Abbreviations

GWAS: Genome-wide association study; SNP: Single nucleotide polymorphism; T1D: Type 1 diabetes; Treg: Regulatory T cells; Tconv: Conventional T cells; FOXP3: Foxhead box protein 3; TF: Transcription factor; eQTL: Expression quantitative trait loci; ChIP-seq: Chromatin immunoprecipitation sequencing; ATAC-seq: Assay for transposase-accessible chromatin sequencing; Hi-C: High-resolution chromosome conformation capture sequencing; TAD: Topologically associated domain; SE: Super-enhancer.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13072-022-00456-5>.

Additional file 1. Figures. S1–S14, and Tables S1, S2.

Additional file 2. Promoters and enhancers used in the 3DFAACTS-SNP workflow in this study.

Additional file 3. Table S3–S10.

Additional file 4. Transcription factor footprint identified from rest and stimulated Tregs ATAC-seq data using HINT-ATAC.

Acknowledgements

We are thankful for the access to publicly available resources, and grateful for discussions and comments from members of the Barry lab. We thank Dr R Gross for cell sorting expertise.

Author contributions

All authors designed the study design. NL, TS, SB and JB wrote the manuscript, with additional help from YYW and SMP. NL developed and implemented all computational workflows and HiC-seq analyses. YYW analysed the ATAC-seq datasets and JB analysed public ChIP-seq and ATAC-seq data. TS and YYW constructed the ATAC-seq and Hi-C libraries. All authors read and approved the final manuscript.

Funding

This research was supported by a 2017 National Health and Medical Research Council (NHMRC) project Grant (#399123, #565314 and 1120543). Dr Sadlon and Professor Barry are supported by a Women's and Children's Hospital Foundation Grant. Dr James Breen is supported by the James & Diana Ramsay Foundation.

Availability of data and materials

The Treg ATAC-seq and Hi-C datasets analysed during the current study are available in the European Nucleotide Archive (ENA) repository (PRJEB39882). Published data:

FOXP3 ChIP-chip data used during the current study are available on Gene Expression Omnibus (accession no. GSE20995) [31]. Treg and Th1 RNA-seq data used during the current study are available on European Nucleotide Archive (accession no. ERR1198158 and ERR1198159) [41].

Database used in the current study:

NIH Roadmap Epigenomics Project: <http://www.roadmapepigenomics.org/> [30].

SEdb: <http://www.licpathway.net/sedb/> [39].

DICE: <https://dice-database.org/landing> [42].

gnomAD: <https://gnomad.broadinstitute.org/> [86].

FANTOM5: <https://fantom.gsc.riken.jp/5/> [29].

Source code for 3DFAACTS-SNP workflow and related in-house scripts are available in GitHub (<https://github.com/ningbioinfostruggling/3DFAACTS-SNP>).

Declarations

Ethics approval and consent to participate

Cord blood used in this study was obtained with approval from the donor and the Children's, Youth and Women's Health Service Research Ethics Committee (HREC1596 and HREC 19/wchn/65 from the Women's and Children's Health Network Human Ethics committee). Buffy Coats were obtained from the Australian Red Cross (Material Supply Deed 19-035A-02).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹South Australian Health and Medical Research Institute, Adelaide, Australia. ²Robinson Research Institute, University of Adelaide, Adelaide, Australia. ³Bioinformatics Hub, School of Biological Sciences, Faculty of Health and Medical Sciences, University of Adelaide, Adelaide, Australia. ⁴Women's and Children's Health Network, Women's and Children's Hospital, Adelaide, Australia. ⁵Black Ochre Data Labs, Indigenous Genomics, Telethon Kids Institute, Adelaide, Australia. ⁶John Curtin School of Medical Research, Australian National University, Canberra, Australia.

Received: 15 December 2021 Accepted: 6 June 2022

Published online: 30 June 2022

References

- Long AS, Buckner JH. CD4+FOXP3+ T regulatory cells in human autoimmunity: more than a numbers game. *J Immunol*. 2011;187:2061–6.
- Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *Lancet*. 2014;383:69–82.
- Fontenot JD, Gavin MA, Rudensky AY. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat Immunol*. 2003;4:ni904.
- Bennett CL, Christie J, Ramsdell F, Brunkow ME, Ferguson PJ, Whitesell L, Kelly TE, Saulsbury FT, Chance PF, Ochs HD. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat Genet*. 2001;27:20–1.
- Wildin RS, Ramsdell F, Peake J, Faravelli F, Casanova JL, Buist N, Levy-Lahad E, Mazzella M, Goulet O, Perroni L, et al. X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. *Nat Genet*. 2001;27:18–20.
- Fontenot JD, Rudensky AY. A well adapted regulatory contrivance: regulatory T cell development and the forkhead family transcription factor Foxp3. *Nat Immunol*. 2005;6:ni1179.
- Ono M. Control of regulatory T-cell differentiation and function by T-cell receptor signalling and Foxp3 transcription factor complexes. *Immunology*. 2020;160:24–37.
- Sadlon T, Brown CY, Bandara V, Hope CM, Schjenken JE, Pederson SM, Breen J, Forrest A, Beyer M, Robertson S, Barry SC. Unravelling the molecular basis for regulatory T-cell plasticity and loss of function in disease. *Clin Transl Immunology*. 2018;7: e1011.
- Hope CM, Welch J, Mohandas A, Pederson S, Hill D, Gundsambuu B, Eastaff-Leung N, Grosse R, Bresatz S, Ang G, et al. Peptidase inhibitor 16 identifies a human regulatory T-cell subset with reduced FOXP3 expression over the first year of recent onset type 1 diabetes. *Eur J Immunol*. 2019;29:1057.
- Redondo MJ, Yu L, Hawa M, Mackenzie T, Pyke DA, Eisenbarth GS, Leslie RD. Heterogeneity of type 1 diabetes: analysis of monozygotic twins in Great Britain and the United States. *Diabetologia*. 2001;44:354–62.
- Bradfield JP, Qu H-QQ, Wang K, Zhang H, Sleiman PM, Kim CE, Mentch FD, Qiu H, Glessner JT, Thomas KA, et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet*. 2011;7:e1002293.
- Cortes A, Brown MA. Promise and pitfalls of the Immunochip. *Arthritis Res Ther*. 2011;13:101.
- Suna O-G, Chen W-M, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, Farber E, Bonnie JK, Szpak M, Schofield E, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet*. 2015;47:381–6.
- Bush WS, Moore JH. Chapter 11: genome-wide association studies. *PLoS Comput Biol*. 2012;8:e1002822.
- Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 Years of GWAS discovery: biology, function, and translation. *Am J Hum Genet*. 2017;101:5–22.
- Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res*. 2012;22:1748–59.
- Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*. 2010;6: e1000888.
- Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013;45:1238–43.
- Qian Y, Zhang L, Cai M, Li H, Xu H, Yang H, Zhao Z, Rhie SK, Farnham PJ, Shi J, Lu W. The prostate cancer risk variant rs55958994 regulates multiple gene expression through extreme long-range chromatin interaction to control tumor progression. *Sci Adv*. 2019;5:eaaw6710.
- Jung I, Schmitt A, Diao Y, Lee AJ, Liu T, Yang D, Tan C, Eom J, Chan M, Chee S, et al. A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat Genet*. 2019;51:1442–9.
- Cerosaletti K, Schneider A, Schwedhelm K, Frank I, Tatum M, Wei S, Whalen E, Greenbaum C, Kita M, Buckner J, Long SA. Multiple

- autoimmune-associated variants confer decreased IL-2R signaling in CD4+ CD25(hi) T cells of type 1 diabetic and multiple sclerosis patients. *PLoS ONE*. 2013;8: e83811.
22. Viglietta V, Baecher-Allan C, Weiner HL, Hafler DA. Loss of functional suppression by CD4+CD25+ regulatory T cells in patients with multiple sclerosis. *J Exp Med*. 2004;199:971–9.
 23. Kim JM, Rasmussen JP, Rudensky AY. Regulatory T cells prevent catastrophic autoimmunity throughout the lifespan of mice. *Nat Immunol*. 2007;8:191–7.
 24. Feuerer M, Shen Y, Littman DR, Benoist C, Mathis D. How punctual ablation of regulatory T cells unleashes an autoimmune lesion within the pancreatic islets. *Immunity*. 2009;31:654–64.
 25. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, Raychaudhuri S. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet*. 2013;45:124–30.
 26. Krijger PHL, de Laat W. Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol*. 2016;17:771–82.
 27. Dekker J, Mirny L. The 3D genome as moderator of chromosomal communication. *Cell*. 2016;164:1110–21.
 28. Gorkin DU, Leung D, Ren B. The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell*. 2014;14:762–75.
 29. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, Abugessaisa I, Fukuda S, Hori F, Ishikawa-Kato S, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol*. 2015;16:22.
 30. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317–30.
 31. Sadlon TJ, Wilkinson BG, Pederson S, Brown CY, Bresatz S, Gargett T, Melville EL, Peng K, D'Andrea RJ, Glonek GG, et al. Genome-wide identification of human FOXP3 target genes in natural regulatory T cells. *J Immunol*. 2010;185:1071–81.
 32. Lu G, Hao X, Chen W-H, Mu S. GAAD: a gene and autoimmune disease association database. *Genom Proteom Bioinform*. 2018;16:252–61.
 33. Zacher B, Michel M, Schwalb B, Cramer P, Tresch A, Gagneur J. Accurate promoter and enhancer identification in 127 ENCODE and roadmap epigenomics cell types and tissues by GenoSTAN. *PLoS ONE*. 2017;12: e0169249.
 34. Lowe CE, Cooper JD, Brusko T, Walker NM, Smyth DJ, Bailey R, Bourget K, Plagnol V, Field S, Atkinson M, et al. Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nat Genet*. 2007;39:1074–82.
 35. Mlynarski WM, Placha GP, Wolkow PP, Bochenski JP, Warram JH, Krolewski AS. Risk of diabetic nephropathy in type 1 diabetes is associated with functional polymorphisms in RANTES receptor gene (CCR5): a sex-specific effect. *Diabetes*. 2005;54:3331–5.
 36. Baker C, Chang L, Elsegood KA, Bishop AJ, Gannon DH, Narendran P, Leech NJ, Dayan CM. Activated T cell subsets in human type 1 diabetes: evidence for expansion of the DR+ CD30+ subpopulation in new-onset disease. *Clin Exp Immunol*. 2007;147:472–82.
 37. Marroquí L, Santin I, Dos Santos RS, Marselli L, Marchetti P, Eizirik DL. BACH2, a candidate risk gene for type 1 diabetes, regulates apoptosis in pancreatic β -cells via JNK1 modulation and crosstalk with the candidate gene PTPN2. *Diabetes*. 2014;63:2516–27.
 38. Consortium GT, Laboratory DA, Coordinating Center —Analysis Working G, Statistical Methods groups—Analysis Working G, Enhancing Gg, Fund NIH, Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550:204–13.
 39. Jiang Y, Qian F, Bai X, Liu Y, Wang Q, Ai B, Han X, Shi S, Zhang J, Li X, et al. SEdb: a comprehensive human super-enhancer database. *Nucleic Acids Res*. 2019;47:D235–43.
 40. Gao P, Uzun Y, He B, Salamati SE, Coffey JKM, Tsalikian E, Tan K. Risk variants disrupting enhancers of TH1 and TREG cells in type 1 diabetes. *Proc National Acad Sci*. 2019;116:201815336.
 41. De Simone M, Arrigoni A, Rossetti G, Guarini P, Ranzani V, Politano C, Bonnal RJP, Provasi E, Sarnicola ML, Panzeri I, et al. Transcriptional landscape of human tissue lymphocytes unveils uniqueness of tumor-infiltrating T regulatory cells. *Immunity*. 2016;45:1135–47.
 42. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Jose Z-G, Ha B, Altay G, Greenbaum JA, Graham M, et al. Impact of genetic polymorphisms on human immune cell gene expression. *Cell*. 2018;175:1701–15.
 43. Bour-Jordan H, Bluestone JA. Regulating the regulators: costimulatory signals control the homeostasis and function of regulatory T cells. *Immunol Rev*. 2009;229:41–66.
 44. Chen Q, Mo L, Cai X, Wei L, Xie Z, Li H, Li J, Hu Z. ICOS signal facilitates Foxp3 transcription to favor suppressive function of regulatory T cells. *Int J Med Sci*. 2018;15:666–73.
 45. Sun H, Lagarrigue F, Wang H, Fan Z, Lopez-Ramirez MA, Chang JT, Ginsberg MH. Distinct integrin activation pathways for effector and regulatory T cell trafficking and function. *J Exp Med*. 2021; 218.
 46. Fanucchi S, Fok ET, Dalla E, Shibayama Y, Börner K, Chang EY, Stoychev S, Imakaev M, Grimm D, Wang KC, et al. Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments. *Nat Genet*. 2019;51:138–50.
 47. Zhan Y, Wang N, Vasanthakumar A, Zhang Y, Chopin M, Nutt SL, Kallies A, Lew AM. CCR2 enhances CD25 expression by FoxP3+ regulatory T cells and regulates their abundance independently of chemotaxis and CCR2+ myeloid cells. *Cell Mol Immunol*. 2020;17:123–32.
 48. Soler DC, Sugiyama H, Young AB, Massari JV, McCormick TS, Cooper KD. Psoriasis patients exhibit impairment of the high potency CCR5+ T regulatory cell subset. *Clin Immunol*. 2013;149:111–8.
 49. Wang R, Huang K. CCL11 increases the proportion of CD4+CD25+Foxp3+ Treg cells and the production of IL-2 and TGF- β by CD4+ T cells via the STAT5 signaling pathway. *Mol Med Rep*. 2020;21:2522–32.
 50. Nguyen KD, Fohner A, Booker JD, Dong C, Krensky AM, Nadeau KC. XCL1 enhances regulatory activities of CD4+ CD25(high) CD127(low/-) T cells in human allergic asthma. *J Immunol*. 2008;181:5386–95.
 51. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee T, Young RA. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153:307–19.
 52. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, Young RA. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155:934–47.
 53. Walker L. Treg and CTLA-4: two intertwining pathways to immune tolerance. *J Autoimmun*. 2013;45:49–57.
 54. Samstein RM, Arvey A, Josefowicz SZ, Peng X, Reynolds A, Sandstrom R, Neph S, Sabo P, Kim JM, Liao W, et al. Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell*. 2012;151:153–66.
 55. Chaudhry A, Rudra D, Treuting P, Samstein RM, Liang Y, Kas A, Rudensky AY. CD4+ regulatory T cells control TH17 responses in a Stat3-dependent manner. *Science*. 2009;326:986–91.
 56. Zheng Y, Chaudhry A, Kas A, deRoos P, Kim JM, Chu T-T, Corcoran L, Treuting P, Klein U, Rudensky AY. Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control T(H)2 responses. *Nature*. 2009;458:351–6.
 57. Duhén T, Duhén R, Lanzavecchia A, Sallusto F, Campbell DJ. Functionally distinct subsets of human FOXP3+ Treg cells that phenotypically mirror effector Th cells. *Blood*. 2012;119:4430–40.
 58. Tsukumo S-I, Unno M, Muto A, Takeuchi A, Kometani K, Kurosaki T, Igarashi K, Saito T. Bach2 maintains T cells in a naive state by suppressing effector memory-related genes. *Proc Natl Acad Sci U S A*. 2013;110:10735–40.
 59. Farh KK-H, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shores N, Whitton H, Ryan RJH, Shishkin AA, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015;518:337–43.
 60. Li Z, Schulz MH, Look T, Begemann M, Zenke M, Costa IG. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol*. 2019;20:45.
 61. Huang D, Yi X, Zhang S, Zheng Z, Wang P, Xuan C, Sham PC, Wang J, Li MJ. GWAS4D: multidimensional analysis of context-specific regulatory variant for human complex diseases and traits. *Nucleic Acids Res*. 2018;46:W114–20.
 62. Wu Y, Borde M, Heissmeyer V, Feuerer M, Lapan AD, Stroud JC, Bates DL, Guo L, Han A, Ziegler SF, et al. FOXP3 controls regulatory T cell function through cooperation with NFAT. *Cell*. 2006;126:375–87.

63. Fragale A, Gabriele L, Stellacci E, Borghi P, Perrotti E, Ilari R, Lanciotti A, Remoli AL, Venditti M, Belardelli F, Battistini A. IFN regulatory factor-1 negatively regulates CD4+ CD25+ regulatory T cell differentiation by repressing Foxp3 expression. *J Immunol*. 2008;181:1673–82.
64. Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, Abraham BJ, Cohen MA, Nabet B, Buckley DL, et al. YY1 is a structural regulator of enhancer-promoter loops. *Cell*. 2017;171:1573–1588.e1528.
65. International Multiple Sclerosis Genetics C. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science*. 2019; 365.
66. Márquez A, Kerick M, Zhernakova A, Gutierrez-Achury J, Chen W-M, Onengut-Gumuscu S, González-Álvoro I, Rodríguez-Rodríguez L, Rios-Fernández R, González-Gay MA, et al. Meta-analysis of Immunochip data of four autoimmune diseases reveals novel single-disease and cross-phenotype associations. *Genome Med*. 2018;10:97.
67. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, Park YR, Raychaudhuri S, Pouget JG, Hübenthal M, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet*. 2016;48:510–8.
68. Wang Z, He L, Li W, Xu C, Zhang J, Wang D, Dou K, Zhuang R, Jin B, Zhang W, et al. GDF15 induces immunosuppression via CD48 on regulatory T cells in hepatocellular carcinoma. *J Immunother Cancer*. 2021;9:e002787.
69. Gao W, Thompson L, Zhou Q, Putheti P, Fahmy TM, Strom TB, Metcalfe SM. Treg versus Th17 lymphocyte lineages are cross-regulated by LIF versus IL-6. *Cell Cycle*. 2009;8:1444–50.
70. Shaffer MH, Dupree RS, Zhu P, Saotome I, Schmidt RF, McClatchey AI, Freedman BD, Burkhardt JK. Ezrin and Moesin function together to promote T cell activation. *J Immunol*. 2009;182:1021–32.
71. Pezolesi MG, Katavetin P, Kure M, Poznik GD, Skupien J, Mychaleckyj JC, Rich SS, Warram JH, Krolewski AS. Confirmation of genetic associations at ELMO1 in the GoKinD collection supports its role as a susceptibility gene in diabetic nephropathy. *Diabetes*. 2009;58:2698–702.
72. Peligero-Cruz C, Givony T, Sebe-Pedros A, Dobes J, Kadouri N, Nevo S, Roncato F, Alon R, Goldfarb Y, Abramson J. IL18 signaling promotes homing of mature Tregs into the thymus. *Elife*. 2020; 9.
73. Vitales-Noyola M, Ramos-Levi AM, Serrano-Somavilla A, Martinez-Hernandez R, Sampedro-Nunez M, Di Pasquale C, Gonzalez-Amaro R, Marazuela M. Expression and function of the costimulatory receptor SLAMF1 is altered in lymphocytes from patients with autoimmune thyroiditis. *J Clin Endocrinol Metab*. 2017;102:672–80.
74. Mena J, Alloza I, Tulloch Navarro R, Aldekoa A, Diez Garcia J, Villanueva Etxebarria A, Lindskog C, Antiguada A, Boyero S, Mendibe-Bilbao MDM, et al. Genomic multiple sclerosis risk variants modulate the expression of the ANKRD55-IL6ST gene region in immature dendritic cells. *Front Immunol*. 2021;12: 816930.
75. Tamehiro N, Nishida K, Yanobu-Takanashi R, Goto M, Okamura T, Suzuki H. T-cell activation RhoGTPase-activating protein plays an important role in TH17-cell differentiation. *Immunol Cell Biol*. 2017;95:729–35.
76. Arakelyan A, Nersisyan L, Poghosyan D, Khondkaryan L, Hakobyan A, Löffler-Wirth H, Melanitou E, Binder H. Autoimmunity and autoinflammation: a systems view on signaling pathway dysregulation profiles. *PLoS ONE*. 2017;12: e0187572.
77. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747–53.
78. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42:565–9.
79. Boyle EA, Li Yi, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169:1177–86.
80. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet*. 2008;40:695–701.
81. Saint Pierre A, Génin E. How important are rare variants in common disease? *Brief Funct Genomics*. 2014;13:353–61.
82. Dempfle A, Scherag A, Hein R, Beckmann L, Chang-Claude J, Schäfer H. Gene-environment interactions for complex traits: definitions, methodological requirements and challenges. *Eur J Hum Genet*. 2008;16:1164–72.
83. Wei W-H, Hemani G, Haley CS. Detecting epistasis in human complex traits. *Nat Rev Genet*. 2014;15:722–33.
84. Auer PL, Lettre G. Rare variant association studies: considerations, challenges and opportunities. *Genome Med*. 2015;7:16.
85. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet*. 2017;100:635–49.
86. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*. 2019.
87. Layer RM, Pedersen BS, DiSera T, Marth GT, Gertz J, Quinlan AR. GIG-GLUE: a search engine for large-scale integrated genome analysis. *Nat Methods*. 2018;15:123–6.
88. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1:417–25.
89. Tak YG, Farnham PJ. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin*. 2015;8:57.
90. Hou L, Zhao H. A review of post-GWAS prioritization approaches. *Front Genet*. 2013;4:280.
91. Lee PH, Lee C, Li X, Wee B, Dwivedi T, Daly M. Principles and methods of in-silico prioritization of non-coding regulatory variants. *Hum Genet*. 2018;137:15–30.
92. Cano-Gamez E, Trynka G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front Genet*. 2020;11:424.
93. Jeng MY, Mumbach MR, Granja JM, Satpathy AT, Chang HY, Chang ALS. Enhancer connectome nominates target genes of inherited risk variants from inflammatory skin disorders. *J Invest Dermatol*. 2019;139:605–14.
94. Mumbach MR, Satpathy AT, Boyle EA, Dai C, Gowen BG, Cho S, Nguyen ML, Rubin AJ, Granja JM, Kazane KR, et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat Genet*. 2017;49:1602–12.
95. Fletcher JM, Lonergan R, Costelloe L, Kinsella K, Moran B, O'Farrelly C, Tubridy N, Mills KHG. CD39+Foxp3+ regulatory T Cells suppress pathogenic Th17 cells and are impaired in multiple sclerosis. *J Immunol*. 2009;183:7602–10.
96. Lindley S, Dayan CM, Bishop A, Roep BO, Peakman M, Tree TIM. Defective suppressor function in CD4+ CD25+ T-cells from patients with type 1 diabetes. *Diabetes*. 2005;54:92–9.
97. Qu K, Zaba LC, Giresi PG, Li R, Longmire M, Kim YH, Greenleaf WJ, Chang HY. Individuality and variation of personal regulomes in primary human T cells. *Cell Syst*. 2015;1:51–61.
98. Cvetanovich GL, Hafler DA. Human regulatory T cells in autoimmune diseases. *Curr Opin Immunol*. 2010;22:753–60.
99. Dominguez-Villar M, Hafler DA. Regulatory T cells in autoimmune disease. *Nat Immunol*. 2018;19:665–73.
100. Höllbacher B, Duhén T, Motley S, Klicznik MM, Gratz IK, Campbell DJ. Transcriptomic profiling of human effector and regulatory T cell subsets identifies predictive population signatures. *Immunohorizons*. 2020;571:265.
101. Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, Olsen BN, Mumbach MR, Pierce SE, Corces MR, et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol*. 2019;37:925–36.
102. Ohkura N, Yasumizu Y, Kitagawa Y, Tanaka A, Nakamura Y, Motooka D, Nakamura S, Okada Y, Sakaguchi S. Regulatory T cell-specific epigenomic region variants are a key determinant of susceptibility to common autoimmune diseases. *Immunity*. 2020;52:1119–1132.e1114.
103. Sidwell T, Liao Y, Garnham AL, Vasanthakumar A, Gloury R, Blume J, Teh PP, Chisanga D, Thelemann C, de Labastida Rivera F, et al. Attenuation of TCR-induced transcription by Bach2 controls regulatory T cell differentiation and homeostasis. *Nat Commun*. 2020; 11.
104. Nasrallah R, Imianowski CJ, Bossini-Castillo L, Grant FM, Dogan M, Placek L, Kozhaya L, Kuo P, Sadiyeh F, Whiteside SK, et al. A distal enhancer at

- risk locus 11q13.5 promotes suppression of colitis by Treg cells. *Nature*. 2020;583:447–52.
105. Ing-Simmons E, Seitan VC, Faure AJ, Flicek P, Carroll T, Dekker J, Fisher AG, Lenhard B, Merckenschlager M. Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome Res*. 2015;25:504–13.
 106. Duan J, Shi J, Fiorentino A, Leites C, Chen X, Moy W, Chen J, Alexandrov BS, Usheva A, He D, et al. A rare functional noncoding variant at the GWAS-implicated MIR137/MIR2682 locus might confer risk to schizophrenia and bipolar disorder. *Am J Hum Genet*. 2014;95:744–53.
 107. Isoda T, Moore AJ, He Z, Chandra V, Aida M, Denholtz M, Piet van Hamburg J, Fisch KM, Chang AN, Fahl SP, et al. Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T cell fate. *Cell*. 2017;171:103–19.
 108. Papanonis A, Kohro T, Baboo S, Larkin JD, Deng B, Short P, Tsutsumi S, Taylor S, Kanki Y, Kobayashi M, et al. TNF α signals through specialized factories where responsive coding and miRNA genes are transcribed. *EMBO J*. 2012;31:4404–14.
 109. Ganeshan K, Chawla A. Metabolic regulation of immune responses. *Annu Rev Immunol*. 2014;32:609–34.
 110. Carvalho-Silva D, Pierleoni A, Pignatelli M, Ong C, Fumis L, Karamanis N, Carmona M, Faulconbridge A, Hercules A, McAuley E, et al. Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res*. 2019;47:D1056–65.
 111. Yang Z, Fujii H, Mohan SV, Goronzy JJ, Weyand CM. Phosphofruktokinase deficiency impairs ATP generation, autophagy, and redox balance in rheumatoid arthritis T cells. *J Exp Med*. 2013;210:2119–34.
 112. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47:D1005–12.
 113. Grant SFA, Qu H-Q, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Taback SP, Frackelton EC, Eckert AW, et al. Follow-up analysis of genome-wide association data identifies novel loci for type 1 diabetes. *Diabetes*. 2009;58:290–5.
 114. Zhu M, Xu K, Chen Y, Gu Y, Zhang M, Luo F, Liu Y, Gu W, Hu J, Xu H, et al. Identification of novel T1D risk loci and their association with age and islet function at diagnosis in autoantibody-positive T1D individuals: based on a two-stage genome-wide association study. *Diabetes Care*. 2019;42:1414–21.
 115. International Multiple Sclerosis Genetics C, Beecham AH, Patsoopoulos NA, Xifara DK, Davis MF, Kempainen A, Cotsapas C, Shah TS, Spencer C, Booth D, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet*. 2013;45:1353–60.
 116. Andlauer TFM, Buck D, Antony G, Bayas A, Bechmann L, Berthele A, Chan A, Gasperi C, Gold R, Graetz C, et al. Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. *Sci Adv*. 2016;2: e1501678.
 117. Huang C, Haritunians T, Okou DT, Cutler DJ, Zwick ME, Taylor KD, Datta LW, Maranville JC, Liu Z, Ellis S, et al. Characterization of genetic loci that affect susceptibility to inflammatory bowel diseases in African Americans. *Gastroenterology*. 2015;149:1575–86.
 118. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, Jostins L, Rice DL, Gutierrez-Achury J, Ji S-G, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet*. 2017;49:256–61.
 119. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, Bakker SF, Bardella MT, Bhaw-Rosun L, Castillejo G, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet*. 2011;43:1193–201.
 120. Cano-Gamez E, Soskic B, Roumeliotis TI, So E, Smyth DJ, Baldridge M, Wille D, Nakic N, Esparza-Gordillo J, Larmine CGC, et al. Single-cell transcriptomics identifies an effectness gradient shaping the response of CD4(+) T cells to cytokines. *Nat Commun*. 2020;11:1801.
 121. Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, Bock C, Boehm B, Campo E, Caricasole A, et al. BLUEPRINT to decode the epigenetic signature written in blood. *Nat Biotechnol*. 2012;30:224–6.
 122. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018;46:D794–801.
 123. Bresatz S, Sadlon T, Millard D, Zola H, Barry SC. Isolation, propagation and characterization of cord blood derived CD4+ CD25+ regulatory T cells. *J Immunol Methods*. 2007;327:53–62.
 124. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden E. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159:1665–80.
 125. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*. 2011;17:10–2.
 126. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
 127. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
 128. Ou J, Liu H, Yu J, Kelliher MA, Castilla LH, Lawson ND, Zhu LJ. ATACseqQC: a bioconductor package for post-alignment quality assessment of ATAC-seq data. *BMC Genomics*. 2018;19:169.
 129. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137.
 130. Yang C, Ma L, Xiao D, Ying Z, Jiang X, Lin Y. Integration of ATAC-Seq and RNA-seq identifies key genes in light-induced primordia formation of *Sparassis latifolia*. *Int J Mol Sci*. 2019;21:185.
 131. Miko H, Qiu Y, Gaertner B, Sander M, Ohler U. Inferring time series chromatin states for promoter-enhancer pairs based on Hi-C data. *BMC Genomics*. 2021;22:84.
 132. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods*. 2017;14:959–62.
 133. Dekker J, Belmont AS, Guttman M, Leshy VO, Lis JT, Lomvardas S, Mirny LA, O'Shea CC, Park PJ, Ren B, et al. The 4D nucleome project. *Nature*. 2017;549:219–26.
 134. Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*. 2020;36:311–6.
 135. Alinejad-Rokny H, Ghavami R, Rabiee HR, Rezaei N, et al. MaxHiC: robust estimation of chromatin interaction frequency in Hi-C and capture Hi-C experiments. *bioRxiv*. 2020;8:15454.
 136. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*. 2016;9:88.
 137. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
 138. Ramirez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42:W187–191.
 139. Shin H, Shi Y, Dai C, Tjong H, Gong K, Alber F, Zhou X. TopDom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Res*. 2016;44: e70.
 140. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284–7.
 141. Hahne F, Ivanek R. Visualizing genomic data using gviz and bioconductor. *Methods Mol Biol*. 2016;1418:335–51.
 142. Harmston N, Ing-Simmons E, Perry M, Barešić A, Lenhard B. GenomicInteractions: an R/Bioconductor package for manipulating and investigating chromatin interaction data. *BMC Genomics*. 2015;16:963.
 143. Martin TC, Yet I, Tsai P-C, Bell JT. coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. *BMC Bioinform*. 2015;16:131.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.