Integrated single-cell transcriptomics and epigenomics reveals strong germinal center-associated etiology of autoimmune risk loci

Hamish W King[1*§], Kristen L Wells[2*], Zohar Shipony[2*], Arwa S Kathiria[2], Lisa E Wagar[3,4], Caleb Lareau[2,5], Nara Orban[6], Robson Capasso[7], Mark M Davis[3,8,9], Lars M Steinmetz[2,10,11], Louisa K James[1], William J Greenleaf[2,12 §]

[1] Centre for Immunobiology, Blizard Institute, Queen Mary University of London, London, UK
[2] Department of Genetics, Stanford University, Stanford, CA, USA
[3] Department of Microbiology and Immunology, Stanford University, Stanford, CA, USA
[4] Department of Physiology and Biophysics, University of California Irvine, Irvine, CA, USA
[5] Department of Pathology, University of California Irvine, Irvine, CA, USA
[6] Barts Health Ear, Nose and Throat Service, The Royal London Hospital, London, UK
[7] Division of Sleep Surgery, Department of Otolaryngology–Head and Neck Surgery, Stanford University School of Medicine, Stanford, CA, USA
[8] Institute for Immunity, Transplantation and Infection, Stanford University, Stanford, CA, USA
[9] Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA
[10] Stanford Genome Technology Center, Stanford University, Stanford, CA, USA
[11] Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany
[12] Chan–Zuckerberg Biohub, San Francisco, CA, USA.
* These authors contributed equally.
§ To whom correspondence can be addressed: h.king@qmul.ac.uk, drhamishking@gmail.com and wjg@stanford.edu

One sentence summary: Single-cell ATAC sequencing maps the cell type-specific regulatory potential of transcription factors and autoimmune disease risk loci.

Abstract

The germinal center (GC) response is critical for both effective adaptive immunity and establishing peripheral tolerance by limiting autoreactive B cells. Dysfunction in these processes can lead to defective immune responses to infection or contribute to autoimmune disease. To understand the gene regulatory principles underlying the GC response, we generated a single-cell transcriptomic and epigenomic atlas of the human tonsil, a widely studied and representative lymphoid tissue. We characterize diverse immune cell subsets and build a trajectory of dynamic gene expression and transcription factor activity during B cell activation, GC formation, and plasma cell differentiation. We subsequently leverage cell type-specific transcriptomic and epigenomic maps to interpret potential regulatory impact of genetic variants implicated in autoimmunity, revealing that many exhibit their greatest regulatory potential in GC-associated cellular populations. These included gene loci linked with known roles in GC biology (*IL21*, *IL21R, IL4R, BCL6*) and transcription factors regulating B cell differentiation (*POU2AF1*, *HHEX*). Together, these analyses provide a powerful new cell type-resolved resource for the interpretation of cellular and genetic causes underpinning autoimmune disease.

Introduction

Autoimmune diseases result from a loss of tolerance to otherwise harmless endogenous or exogenous antigens, in part as a consequence of dysregulation in the selection, differentiation, or function of immune

cells. The propensity for such immune cell dysfunction can be potentiated by specific inherited genetic variants, as identified through genome wide association studies (GWAS). However, the majority of GWAS genetic variants reside in non-coding regions of the genome, and the identification of risk-associated genetic variants alone does not identify the cellular populations likely affected by the variant. Recent progress has been made linking autoimmune-associated genetic variants to immune cell type-specific gene regulation by examining functional epigenomic measures like chromatin accessibility, histone acetylation and/or chromatin topology, especially in activated immune cell states of immune subsets (*1-3*). However, such analysis remains incomplete due to limited mapping of important yet transient subpopulations of cells that exist in diverse immune organ contexts.

The development and commitment of different immune cell lineages occurs in primary lymphoid organs such as the bone marrow and thymus. Following lineage commitment and egress from these organs, adaptive immune cells can undergo additional maturation and differentiation in secondary lymphoid organs such as the spleen, lymph nodes and tonsils to generate T cell-mediated immunity and B cell-dependent antibody responses (*4*). The latter in particular is predominantly dependent on the formation of the germinal center (GC) response. This requires MHCII-dependent presentation of antigen-derived peptides by dendritic cells that can be recognized by naïve $CD4^+$ T cells, leading to their differentiation into T follicular helper (Tfh) cells. Tfh are vital to support activated B cells to form GC reactions, undergo somatic hypermutation and affinity maturation of their antibody genes before differentiating into plasma cells or memory B cells.

Mechanisms that ensure immune tolerance to self-antigen target autoreactive B cell clones during early development in the bone marrow (central tolerance) and *de novo* generation in GC responses in secondary lymphoid organs (peripheral tolerance). Autoantibodies are a feature of many systemic autoimmune diseases, and numerous studies have found that autoantibodies can bear somatic hypermutation and class switch recombination signatures indicative of GC-derived B cell populations (*5*), pointing to defects in peripheral tolerance. Because these tissues and GC-associated immune cell populations are directly involved in establishing both peripheral tolerance and forming effective adaptive immune responses, mapping the regulatory potential of autoimmune-associated genetic variants in these dynamic populations will enable the interpretation of how these variants may contribute to autoimmunity.

Here we apply single-cell transcriptomics (scRNA-seq), surface-protein profiling (scADT-seq), and epigenomics (scATAC-seq) to map the cellular states and gene regulatory networks of immune cells from the human tonsil, a model secondary lymphoid organ. By integrating gene expression and chromatin accessibility across 37 immune cell populations spanning bone marrow, peripheral blood, and tonsils, we

identify putative target genes of fine-mapped autoimmune-associated genetic variants and reveal extensive GC-specific regulatory potential, including at loci of major GC regulators such as *IL21*, *IL21R/IL4R* and *BCL6*, as well as two genes required for MBC fate commitment, *POU2AF1* and *HHEX*. Our integrative analyses ultimately provide original insights into the cellular and genetic etiology of autoimmune-associated genetic variants and generate a framework to functionally dissect their potential in the maintenance of peripheral tolerance and the generation of adaptive immunity.

Results

Single-cell transcriptomics and epigenomics of a model human secondary lymphoid organ to define immune cell states.

To map the diverse immune cell states of the adaptive immune response in human secondary lymphoid organs, and the gene regulatory elements active in these different populations, we performed high throughput single-cell RNA sequencing (scRNA-seq) coupled with antibody-derived tags (scADT-seq) for twelve surface protein markers on tonsillar immune cells obtained from pediatric patients undergoing routine tonsillectomy for obstructive sleep apnea or recurrent tonsillitis (Fig1A-C, S1, Data file S1; $n = 3$). In parallel, we performed single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) (*6*) to profile active chromatin regulatory elements in tonsillar immune cells (Fig1A-C, S1; Fig2 for more detailed analysis; $n = 7$). We first annotated 9 broad populations based on their surface protein and RNA levels of known markers (Fig1B) and observed good concurrence between RNA, surface protein expression, and chromatin accessibility of key marker genes and the frequency of different cell types (Fig1C, S1-2, Data file S1-2). We observed a relationship between patient age and the relative frequencies of B cells in our scRNA-seq datasets (FigS3A). CyTOF profiling of pediatric and adult tonsils revealed significantly fewer GC-specific B and T cell populations in older pediatric donors (>5 years old) and adults (FigS3B-D), consistent with reduced GC activity in older individuals (*7*). As the GC is a major site of many important cell fate decisions during adaptive immune responses, this demonstrates the need to profile pediatric and/or immunologically relevant (e.g. after vaccination or infection) lymphoid tissue, in contrast to peripheral blood-derived immune populations or lymphoid tissue from older individuals that lack these populations.

We next annotated B or T lymphocyte sub-populations at a higher resolution using our scRNA-seq dataset (Fig1D-G, S4, Data file S3-4). Within the T cell lineage, we identified naïve and central memory T (Tcm) cells, cytotoxic lymphocytes (CTL), NK cells, regulatory T cells (Treg) and two populations of Tfh cells, with one population expressing high levels of *CXCL13*, *CD200* and *IL21*, likely representing GC Tfh (*8*) (Fig1D-E). We also defined clusters with previously identified gene expression markers for many expected

B cell populations, including naïve, activated, memory, tissue-resident FCRL4+ memory, GC (light zone and dark zone) B cells, as well as plasmablasts (Fig1F-G) (*9*). A large population of proliferating B cells were predominantly dark zone GC B cells, as expected (FigS4C). We also found a small cluster of B cells expressing markers of type I interferon response genes such as *IFI44L*, *XAF1*, and *MX1* (Fig1F-G) that are known to be up-regulated after early stages of vaccination (*10*) and in patients with autoimmune diseases like lupus and Sjögren's syndrome (FigS5) (*11-13*). Importantly, all cellular populations, including this rare IFN-responsive state, were identified at consistent frequencies across all patient donors (FigS4D-E), and these annotations broadly agreed with recent single-cell studies of lymphocytes in pediatric tonsils and adult lymph node (*9, 14*).

Mapping chromatin accessibility and transcription factor activity in tonsillar immune subsets.

Our high-resolution annotation of immune cell populations by scRNA-seq (Fig1) allowed us to more comprehensively annotate our scATAC datasets (Fig2A; see Materials and Methods for details) (*15*). We limited our annotations of the chromatin accessibility maps to 14 cell populations to maximize coverage and representation of cell type-specific peaks in subsequent analyses. We identified naïve, activated, memory, FCRL4+ memory and GC (light zone and dark zone) B cell subsets, as well as plasmablasts, Tfh, Treg, naïve, central memory and cytotoxic T cells, and two smaller clusters representing a combination of monocytes, macrophages and dendritic cells (Fig2A). We found a strong correspondence between cluster identities and cell type-specific markers used in both scATAC-seq and scRNA-seq annotation of our datasets (FigS1-2). Cells at different stages of the cell cycle, such as proliferating dark zone GC B cells, were difficult to distinguish based on their chromatin accessibility profiles, as we and others have observed few qualitative differences in chromatin accessibility profiles between mitotic and interphase cells (*16, 17*). As in our scRNA-seq analysis, most scATAC-seq clusters were identified reproducibly in all tissue donors (FigS6A-B), although we did observe higher frequencies of activated and DZ GC B cells in two recurrent tonsillitis patients compared to sleep apnea patients. However, previous studies, including scRNA-seq analysis, have found no or few differences in the cellular phenotypes of immune cells between these two patient groups (*9, 18*). Overall, we provide a comprehensive resource of cell type-specific gene regulatory elements across 14 tonsillar immune cell populations in this model secondary lymphoid organ (FigS7A-B, Data file S5-8), including at the immunoglobulin heavy chain locus (FigS7C-D). We also report putative peak-to-gene linkages by correlating peak chromatin accessibility with scRNA-seq expression in our integrated analysis pipeline (see Materials and Methods for details) (FigS7B, Data file S7-8) (*15*), which, when paired with cell type-specific accessibility and gene expression, can provide insights into potential gene regulatory landscapes across these different immune cell populations.

Lymphocyte activation, maturation, and differentiation are underpinned by transcriptional networks controlled by sequence-specific transcription factors (TFs). To understand the regulatory potential of different TFs *in vivo* we correlated the expression of TFs with the chromatin accessibility of their target motif sequences in B and T lymphocyte populations (Fig2B-D). Specifically, we sought to identify TFs whose enrichment of their motif sequences in accessible chromatin was significantly and positively correlated with expression of that TF within a given cell type (as shown for all B cells in Fig2B) as a means to predict TFs most likely to regulate gene expression in those cells. This successfully identified enrichment of TFs known to be important for gene regulation in B and T cell subset-specific states, such as PAX5, EBF1, TCF7 and BATF (Fig2C). Our analysis also revealed shared regulatory TF activities between similar cell states, such as those active in naïve, activated and memory B cells (KLF2, BCL11A, ELF2, ETV6, ELK4) or GC B cells (EBF1, REST, POU2F1, PKNOX1) (Fig2C-D). We also found highly cell type-specific activities, such as for EOMES, IRF1/2 and RUNX1/3 in cytotoxic lymphocytes, and ID3, ASCL2, NFIA and TCF12 in Tfh cells (Fig2C-D).

While these analyses of defined cell types and states revealed putative transcriptional regulators specific to different populations, TFs also play major roles in shaping dynamic cell fate decisions during activation or differentiation of immune cells. B cell activation and subsequent participation in the GC reaction is essential for high quality B cell-dependent immune responses, yet the dynamics of different gene regulatory networks involved in this key process are poorly understood. We therefore performed a pseudotemporal reconstruction of a single-cell trajectory encompassing B cell activation, the GC reaction and plasmablast differentiation and identified modules of TF regulatory activity that corresponded with different stages of this trajectory (Fig2E-F, S7E). Intriguingly, the pseudotemporal ordering of activated B cells identified two distinct peaks of dynamic TF expression and chromatin accessibility at corresponding motif sequences before commitment to the GC state (Fig2E-F; Modules 2 and 3). This included early expression of NFκB family members (Module 2; REL, RELA, NFKB1, NFKB2), which was highly correlated with chromatin accessibility at their predicted binding sites genome-wide. We identified a NFκB/RELA binding site predicted to be disrupted by a rheumatoid arthritis (RA)-associated SNP (rs74405933; G→T), for which chromatin accessibility is strongly correlated with *CD83* expression (Fig2G), a key gene involved in B cell activation and maturation (*19*). In addition to this initial activation module, we identified a secondary activation state comprising several poorly understood TFs, including BHLHE40, CEBPE/Z, ZBTB33, and ZHX1 (Module 3). We also identified dynamic expression and chromatin activity in GC B cells, including one module that decreases through GC exit and plasma cell differentiation (Module 4; HNF1B, EBF1, SMAD2, POU2F1, MEF2B) and one module that is maintained or increases during commitment to the plasma fate (Module 5; NR2F6, FOXO4, JDP2, MSC). In contrast, a transcriptional regulatory module containing master plasma cell regulators such as IRF4, PRDM1 and XBP1 (Module 6)

exhibited reduced accessibility at target sites within GC B cells compared to both naïve and plasma populations, suggesting that these sites may be actively repressed to prevent inappropriate or premature commitment to the plasma fate during affinity maturation in the GC. Unfortunately, we were not able to reconstruct a trajectory for the memory B cell fate, perhaps due to the presence of both GC-derived and extra-follicular sources of memory B cells in tonsil tissue, the proposed stochastic nature of this cell fate decision (*20*), or limited number of B cells within our scATAC datasets.

Integration of secondary lymphoid organ datasets with bone marrow and peripheral blood single-cell transcriptome and epigenome atlases.

Other scRNA-seq analyses have recently demonstrated that tonsils are a transferable model tissue to study secondary lymphoid organs and adaptive immune responses more generally (*9, 14, 21*). In contrast to circulating or bone-marrow resident lymphocyte populations, immune cells within secondary lymphoid organs exist in a range of activation and maturation states, including GC-associated populations, that may reflect varied tissue niches, cell-cell communication and cytokine signaling. To examine the potential relevance of tissue-specific gene expression and chromatin-based regulatory activities, we integrated our tonsillar scRNA-seq and scATAC-seq datasets with those from publicly available bone marrow and peripheral blood immune cell atlases (*22*) to generate an overview of leukopoiesis comprising data for 60,639 and 91,510 high quality cells for scRNA-seq and scATAC-seq respectively (Fig3A, S8-9, Data file S9-12). As expected, activated B cells, GC-associated lymphocytes (GC B and Tfh cells) and tissue-resident macrophages were strongly enriched in secondary lymphoid organs, while progenitor populations like common lymphoid progenitors (CLP) and granulocyte-monocyte progenitors (GMP), and circulating monocytes were enriched in the bone marrow and peripheral blood respectively (Fig3B). In addition to differences in the frequency of immune cell subsets, we also examined if there might be differences between circulating or tissue resident B cells. We found significant differences in both the chromatin accessibility and gene expression of naïve and memory B cells in the tonsil compared to matched populations in the periphery (Fig3C-D, S10). In particular, chromatin accessibility profiles of tonsillar B cells were enriched with POU2F2 (also known as OCT2) motif sequences (Fig3E), a TF known to be important in the regulation of humoral B cell responses (*23*). These tissue-specific phenotypes likely reflect differences in cytokine exposure and microenvironment of the tonsil compared to circulating blood and highlight that it is essential to examine immune cell populations across varied tissue contexts, even for a single cell type.

Finally, we examined the cell type-specific expression of nine genes recently identified to be most commonly mutated within a sporadic primary immunodeficiency cohort (Fig3F) (*24*). Two genes, *TNFRSF13B* and *CTLA4* were relatively cell type-specific in their expression pattern. *TNFRSF13B*

(encoding TACI) was most highly expressed in memory B cells, particularly tonsillar FCRL4$^+$ memory B cells. Patients with immunodeficiency and *TNFRSF13B* mutations have fewer memory B cells expressing class-switched antibodies, although the mechanisms and penetrance of different coding *TNFRSF13B* mutations remain unclear given the prevalence of coding variants in healthy individuals (*25, 26*). *CTLA4* expression peaked in Tfh and Treg populations as expected. In contrast, *BTK*, *LRBA,* and the TF genes *STAT1, STAT3, NFKB1*, *NFKB2* and *IZKF1* were broadly expressed across varied subsets. We used our scATAC-seq data to examine the enrichment of their motif sequences in accessible chromatin to determine which cell type might be most sensitive to altered activity of these TF genes. This revealed that tonsillar myeloid cells (labelled here primarily as macrophages) had the highest activity of these immunodeficiency-associated TFs (Fig3F), although we observed enrichment of NFKB2 in activated B cells (Fig2F-G, 3F) and STAT1/STAT3 in circulating monocyte and T cells (Fig3F).

Identification of fine-mapped autoimmune GWAS variants in cell type-specific chromatin.

Our integrated scRNA-seq and scATAC-seq atlas of immune cell populations in bone marrow, peripheral blood and tonsils provided a unique opportunity to understand the regulatory potential and cell type-specificity of autoimmune-associated genetic variants across a broad diversity of immune cell types. By examining 12,902 statistically fine-mapped SNPs, of which 9,493 were significantly associated with disorders of the immune system (*1, 27*), we found that our single-cell accessibility profiles of immune cells were broadly enriched in immune-related genetic variants compared to non-immune related traits and background genetic variation (Fig4A, S11A-B). We found specific enrichment of disease-specific genetic variants in different immune cell lineages or subsets (Fig4B, S11C-D). For example, we found a strong enrichment of genetic variants associated with Kawasaki disease and systemic lupus erythematosus in chromatin accessibility maps of the B cell lineage, particularly tonsillar naïve and memory B cells, as well as enrichment of genetic variants associated with alopecia, autoimmune thyroiditis, systemic sclerosis and Behçets disease in cytotoxic lymphocyte regulatory elements (Fig4B, S11C-D). In contrast, genetic variants associated with multiple sclerosis were enriched in both B and T cell-specific chromatin, perhaps reflecting the multigenic nature and complex etiology of this disease (Fig4B, S11C-D).

Of the 1213 immune-related SNPs that overlapped with accessible chromatin peaks in our atlas (Data file S13), many were localized in cell type- or lineage-specific chromatin (Fig4C). Importantly, 342 (28.2%) of these SNPs fell within accessible chromatin only identified in tonsil-enriched immune subsets (Fig4D), demonstrating the value of our tonsillar immune cell atlas for interpretation of GWAS genetic variants. We next predicted the putative gene targets of these genetic variants by using our integrated scRNA-seq and scATAC-seq to identify highly correlated accessibility at chromatin regions to nearby gene expression (*15, 22*). This enabled us to examine 358 chromatin accessible regions (containing 460 unique immune-linked

SNPs) for which we identified significant peak-to-gene linkage correlations (Fig4E). These linkages revealed cell type-specific patterns of both the chromatin accessibility at autoimmune genetic variants and correlated expression of putative gene targets, providing a powerful resource to explore the potential regulatory mechanisms of these genetic variants and their relationship to autoimmune disease.

Chromatin regulatory activity at immune-associated genetic variants predicts importance of GC activity in autoimmunity.

Many studies examining the relationship between immune-associated genetic variants and their regulatory activity with functional genomics methods such as ATAC-seq or ChIP-seq have been limited to studying peripheral immune cell populations. This limitation is likely significant, given our knowledge that many lymphocyte maturation and antibody-based selection events occur in secondary lymphoid organs, and that GC-derived autoantibody production is a feature of many autoimmune diseases. Although we found examples of genetic variants in cell type-specific chromatin across diverse immune subsets (Fig4E, S12-S13; e.g. *GZMB/GZMH*, *NKX2-3*, *COTL1*/*KLHL36*, *KSR1*/*LGALS9*, *TNFRSF1A*/*LTBR*), we observed a striking enrichment of fine-mapped autoimmune variants in chromatin accessibility regions specific to GC-associated B and T populations, such as GC B cells and Tfh cells (Fig4E), including the *IL21*, *IL21R*/*IL4R*, *BCL6*/*LPP*, *CD80*, *PRAG1*, *SLC38A9*, *VAV3*/*SLC25A24*, *DLEU1*/*DLEU1*/*TRIM13* loci (Fig5-6, S14-S15).

We identified GC-specific regulatory elements at the *IL21* locus and the locus of its receptor *IL21R* (Fig5A-B, FigS16). Cytokine signaling by IL-21, primarily secreted by Tfh cells, is essential for B cells to form and participate in normal GC reactions. B cells respond to IL-21 through the IL-21 receptor (IL-21R). We identified several fine-mapped SNPs at the *IL21* locus highly correlated with both chromatin accessibility and gene expression at the *IL21* promoter (Fig5A). These SNPs exhibited Tfh-specific chromatin accessibility, although one SNP, rs13140464, was also highly accessible in several progenitor populations. These fine-mapped SNPs at *IL21* have been associated with alopecia (*1*), juvenile idiopathic arthritis or autoimmunity more generally (*27*), and some of these same SNPs are also significantly associated with celiac disease (rs7682241, rs6840978) (*28*), inflammatory bowel disease (rs7662182) (*29*), primary sclerosing cholangitis (rs13140464) (*30*) and lupus (rs13140464) (*31*). Conversely, we found two fine-mapped SNPs in strong linkage disequilibrium (rs6498021, rs6498019) located in close proximity to *IL21R* in B cell-specific chromatin accessibility regions that have been linked with allergy (*1*) and/or asthma (*32*) (Fig5B, S16). As well as significant correlations with *IL21R* expression, the chromatin accessibility of these two SNPs were also highly correlated with the nearby *IL4R* gene, encoding the IL-4 receptor (IL-4R), which, similar to *IL21R*, was most highly expressed in GC B cells and is vital for T cell-dependent maturation of B cells.

Autoimmune risk variants within a GC-specific locus control region.

Our analysis of genetic variants linked with autoimmunity identified a concentration of recently fine-mapped autoimmune-associated SNPs from the UKBB databank (*27*) in a GC-specific locus control region (LCR) (*33*) located between *BCL6* and *LPP* (Fig5C, S16). Of the genetic variants that fell within accessible chromatin across this locus, there were associations with celiac disease (rs11709472 (*34*), rs7628982 (UKBB), rs9834159 (*35*), rs4686484 (*1*)), allergy (rs56046601 and rs12639588 (*1*)), multiple sclerosis (rs4686953 (formerly rs66756607) (*36, 37*)), asthma (rs7640550 and rs7628982 (*38*)) and vitiligo (rs7628982 (*39*)). Many of these SNPs were present in chromatin accessible regions specific to GC B or Tfh cells, in which *BCL6*, *LPP* and the long non-coding RNA at the LCR (*LINC01991*) are most highly expressed (Fig5C). We report significant correlations in chromatin accessibility between many of these SNPs (and the LCR in general) with the expression of both *BCL6* and *LPP*, consistent with chromosome conformation interactions detected in GC B cells between this LCR and the *BCL6* promoter (*33*). Importantly, deletion of this LCR has been shown in mouse models to lead to defects in GC B cell formation (*33*), presumably through its transcriptional regulation of *BCL6*, one of the master regulatory TFs required for both GC B cells and Tfh cells. These observations suggest that association of this locus with autoimmunity is primarily driven through GC B and Tfh defects. However, some genetic variants (rs142486803, rs76288334, rs78146088) were accessible across many different immune lineages, as well as rs4686484 that was previously proposed to be located in a B cell-specific enhancer (*35*), revealing an additional layer of complexity to this autoimmune regulatory locus.

Autoimmune risk variants at the loci of transcriptional regulators *POU2AF1* and *HHEX*.

We identified cell type-specific chromatin accessibility at autoimmune risk variants across loci for many regulatory TFs or transcriptional regulators including *POU2AF1*, *HHEX*, *ETS1*, *STAT4*, *IKZF3*, *NKX2-3* and *IRF8* (Fig6, S12, S17-18), in addition to the GC master regulator *BCL6* (Fig5C). Of particular interest were *POU2AF1* and *HHEX*, which have recently been proposed to control memory B cell fate selection in the GC (*40, 41*). POU2AF1, also known as OCT binding factor 1 (OBF1), is a largely B cell-specific transcriptional coactivator with no intrinsic DNA binding activity that interacts with TFs POU2F1 (OCT1) and POU2F2 (OCT2). It is indispensable for formation of GCs and GC-dependent B cell maturation (*42*). We found two genetic variants associated with primary biliary cirrhosis/cholangitis (PBC) (rs4938541 and rs4393359 (*1, 43*)) within B cell-specific accessible chromatin and observed that *POU2AF1* expression peaks in GC B cells (Fig6A). Our analysis of B cell activation dynamics predicted POU2F1/POU2F2 as regulators in GC B cells (Fig2) and *POU2F2* is more highly expressed in tonsillar B cells compared to those circulating in peripheral blood (Fig3), suggesting that B cells within lymphoid tissues are likely to be most sensitive to altered *POU2AF1* levels.

HHEX has recently been reported to be an essential regulator of the memory B cell fate decision by GC B cells (*41*), although its potential mechanistic involvement in autoimmune disease is not known. Our integrated epigenomic and transcriptomic analyses identified three fine-mapped SNPs at the *HHEX* locus that fell within B cell specific-accessible chromatin, were implicated in the regulation of *HHEX* through peak-to-gene correlation analysis, and were associated with multiple sclerosis (MS) (rs11187144, rs4933736, rs10882106) (Fig6B). We also identified correlated peak-to-gene linkages between these SNPs and neighboring genes *KIF11* and *EXOC6* (FigS19). We note that rs4933736 falls within a predicted KLF TF binding site (Fig6B), providing a potential mechanism for disruption of *HHEX* expression.

Discussion

Here, we generated paired transcriptome and epigenome atlases of immune cell subsets in the human tonsil, a model system to study the GC reaction which is a major site for developing adaptive immunity to respond to infection and establishing peripheral tolerance to prevent autoimmunity. We defined gene expression and gene regulatory elements across dynamic immune cell states and examined the regulatory potential of transcription factors in these populations. We subsequently leveraged our single-cell resource to profile the cell type-specific chromatin accessibility at fine-mapped GWAS variants linked with autoimmune disorders to reveal that the chromatin of many such variants is most accessible in GC-associated cell types and this accessibility is highly correlated with cell type-specific expression of genes required for normal cytokine signaling or transcriptional regulation in the GC response.

Our single-cell transcriptomic analysis identified a rare B cell population that expresses high levels of IFN-induced gene expression (Fig1). Unfortunately, we were unable to identify this rare B cell population in our scATAC profiling to explore how it may be linked to different autoimmune traits at the chromatin level. One of the genes most highly expressed by the IFN-responsive B cells was *IFI44L*. Splice and missense genetic variants at the *IFI44L* locus (rs1333973 and rs273259) have previously been linked with neutralizing antibody titers to the measles vaccine (*44*), and type I interferon-positive B cells have previously been implicated in the development of autoreactive B cells (*45*). Many of the genes uniquely expressed by this B cell state are also upregulated in the peripheral blood B cells of patients with lupus (FigS5) (*13*). These observations suggest this rare and poorly characterized B cell state may be involved in B cell-mediated antibody responses to vaccines and/or processes linked with autoimmunity.

The joint analysis of gene expression with chromatin accessibility landscapes allowed us to predict putative TF regulators in both steady state and dynamic immune cell populations, including temporally dynamic TFs during B cell activation and their participation in the GC reaction. As part of a dominant B cell

activation, maturation and plasma cell differentiation trajectory, we identified a secondary B cell activation state, after an initial NFκB-associated activation presumably linked with strong BCR activation and/or T cell help . One particularly interesting TF identified was BHLHE40, which has previously been shown to be required for the transition from an activated state prior to entry into the GC (*46, 47*) and is capable of binding key regulatory elements at the immunoglobulin heavy chain locus (*9*). Recent spatial epigenomic mapping of the human tonsil found BHLHE40 regulatory activity outside of the GC reaction, consistent with our pseudotemporal analyses (*48*). How this, and other putative regulators we identify in this secondary activation state (such as CEBPE/Z, ZBTB33, and ZHX1) may contribute to the transition from the activated B cell state to a GC-associated gene expression program will be an important question for future mechanistic studies. However, as the human tonsil represents a highly polyclonal source of B cells, which may arise from many different antigen sources, sub-tissue locations or clonal expansion events, it remains challenging to resolve potentially more complex B cell fate trajectories, such as whether the chromatin accessibility and transcription factor network dynamics in antigen-naïve or –experienced (memory) B cells vary during activation and the GC response.

The molecular mechanisms by which many GWAS-identified genetic polymorphisms contribute to autoimmune disease remain poorly understood. To address this, we and others have examined the relationships between non-coding SNPs and lineage- or cell type-specific expression of putative gene targets to predict the potential functional relevance of genetic variants (reviewed in *49*). For immune-associated GWAS variants, many resources have focused on gene expression or epigenomic profiles of cell types circulating in the peripheral blood or bone marrow (*1, 50*), although there is an emerging prioritization of activation or tissue-specific immune cell states (*2, 3*). Our analysis of chromatin accessibility and gene expression at GWAS loci in tonsillar immune cell states highlights the importance of examining cellular populations in secondary lymphoid tissues, especially of pediatric patients with highly active GC responses, to understand how regulatory activity at non-coding genetic variants in dynamic and tissue-specific populations might contribute to autoimmune disease. Specifically, we found that many autoimmune disease-associated genetic variants are localized within chromatin most accessible in GC B and T cell populations, including at the loci of genes with well-established roles in B cell activation (*CD83*, *CD80*), survival and participation in the GC (*IL21*, *IL21R*, *IL4R, BCL6*) and fate selection (*POU2AF1*, *HHEX*, *IRF8*). While our findings do not exclude dysregulation of autoimmune-associated loci in stromal cell populations which we did not profile here, or potential pleiotropic genetic effects from variants that are accessible across multiple immune cell lineages or tissues, they strongly implicate lymphocyte-intrinsic dysfunctional GC responses as a major feature in the genetic etiology of autoimmune disease.

Our integrated scRNA-seq and scATAC-seq resource maps the cell type-specific chromatin accessibility of autoimmune variant loci genome-wide and identifies highly correlated peak accessibility-gene expression relationships to identify gene targets that may be affected by those SNPs (*15*). Chromosome conformation capture methods such as Hi-C have also been used to predict putative gene targets of autoimmune GWAS variants in GC-associated cell populations (*33, 51*), but these experimental approaches can be limited in their ability to detect short range interactions (e.g. <10kb) and are challenging to perform at scale across many cell types at once or at single-cell resolution. While the inferred peak-to-gene relationships we report here do not provide direct evidence of physical interactions and will require experimental follow up in future studies, our integrated approach to predict gene targets has advantages over other co-accessibility models that link distal regulatory elements to promoters without taking account changes in gene expression, and our approach has successfully linked GWAS variants with putative targets in previous studies (*22, 52*).

To explain how individual non-coding genetic variants may contribute towards the development or pathology of autoimmune disease, it will be necessary to further understand their precise regulatory impact on gene expression. Our analyses do not predict whether specific polymorphisms might positively or negatively regulate gene expression of their putative gene targets. Expression quantitative trait loci (eQTL) analyses can be used to infer whether genetic variants are associated with loss or gain of gene expression (*53*). However, current eQTL databases have profiled either circulating immune cell subsets or whole tissues (e.g. spleen) from adult donors (GTEX median donor age is 50-59 years old). In both cases, these resources lack adequate representation of GC-associated gene expression to confidently dissect the directionality of many SNP-to-gene relationships we predict in our analyses. New advances in neural-network derived methods may prove useful to quantitatively model effects on gene expression in cell type-resolved chromatin accessibility maps (*17, 54*).

While at some loci we identified variants that appear likely to disrupt predicted TF binding sites, the highly context-dependent activating or repressive gene regulatory functions for many TFs remain poorly understood. This therefore makes it difficult to confidently predict whether the downstream gene targets are more likely to be activated or repressed. Inferring downstream targets of TFs without cell type-specific ChIP-seq datasets is likewise challenging, making prediction of the phenotypic impact of potentially altered TF expression at several loci we predict (*BCL6, HHEX*, *POU2AF1*, *ETS1*, *IKZF3*, *STAT4*, *IRF8*) difficult. Functional genomics, single-cell multi-omics and eQTL analyses in varied healthy and diseased immune organs and model systems will be essential to provide further mechanistic insights, as studies of healthy individuals lacking specific variants may miss gain-of-function mutations that create disease-specific regulatory elements *de novo (55)*. Although functional (epi)genomic editing of primary human immune

cells remains challenging, high throughput screening strategies are emerging as powerful new tools (*56*) to assign loss- or gain-of-function to GWAS variants linked with autoimmune disease. However, whichever method is employed to dissect mechanism of non-coding polymorphisms, the fact that many variants associated with disease are in linkage disequilibrium poses a significant challenge to confidently identify causal variants for any given locus.

While we are unable to confidently predict whether expression of a specific gene is enhanced or disrupted by autoimmune-associated genetic variants, either defective or enhanced GC phenotypes could contribute to the development of autoimmune disease by providing an opportunity for the expansion of self-reactive B cells that are normally inhibited in the periphery of healthy individuals (*57*). As a model example to illustrate this principle, we discuss here how altered signaling by IL-21 through IL-21R, for which we identified several autoimmune-associated genetic variants in Tfh- or GC B cell-enriched gene regulatory elements, could lead to altered cellular and immunological phenotypes that might contribute to autoimmunity. If at these loci, any of the genetic variants we characterize result in decreased *IL21* or *IL21R* expression, and subsequently reduced IL-21 signaling, even subtly, this could result in reduced B cell survival within the GC, and enhanced cell death would lead to high concentrations of nuclear autoantigens that might promote autoreactive B cells and loss of tolerance. Conversely, if *IL21* or *IL21R* gene expression was enhanced by genetic variation at distal regulatory elements, elevated autocrine IL-21 signaling by Tfh cells could result in Tfh expansion and proliferation that limit competition amongst GC B cells and lead to the survival of self-reactive B cells (*58, 59*). Indeed, B cell-specific depletion of IL-21R in a mouse model of lupus prevents the development of autoantibodies and disease (*60*), demonstrating that this pathway can play a major role in autoimmunity. While many of the precise molecular and immunological pathways involved in autoimmunity remain unclear, our genetic analyses provide a powerful resource to dissect the transcriptional and epigenetic landscapes of immune cells in secondary lymphoid organs of healthy individuals.

Finally, the development of transient GC-like lymphoid follicles in non-lymphoid tissue (termed ectopic GCs) has been associated with site-specific inflammation in autoimmune diseases and may contribute to loss of tolerance by promoting maturation of self-reactive B cell clones (*61*). Analysis of B cells from ectopic GCs in several autoimmune diseases provide evidence of site-specific clonal expansion and somatic hypermutation of antibody genes, and an absence of normal GC regulation (*62-64*). Single-cell analyses of “defective” and “ectopic” immune structures in different autoimmune diseases will be essential to understand how the regulatory and gene expression dysfunction we predict in the normal immune cell landscape may drive autoimmunity through altered GC response dynamics.

Materials and Methods

Study design

In this study we aimed to define the gene expression and accessible DNA landscapes of different immune cell populations found in the human tonsil, a model secondary lymphoid organ to study adaptive immune responses. This study used tonsil samples from pediatric patients undergoing routine tonsillectomy, numbers of samples per experiment are reported in Data file S1. We first looked at patients covering a wide range of ages and chose to focus for this study on patients ranging from age 3-7 where germinal center population were most abundant for subsequent analysis by scRNA-seq coupled with CITE-seq, and scATAC-seq, performed at Stanford University (n=3). During initial analysis, four additional tonsillar scATAC-seq datasets that had been generated with an identical protocol at Queen Mary University of London were integrated into the data analysis pipeline and used in all subsequent analyses. We used known gene expression markers to define different cell populations in the human tonsil scRNA-seq resource, before using this fine-scaled definition to annotate clusters in matched scATAC-seq datasets. Pseudotemporal ordering of single-cell chromatin accessibility profiles was performed to examine the dynamics of transcription factor activities between different B cell maturation stages. To understand cell type-specific regulatory potential of autoimmune genetic variants, we intersected published statistically fine-mapped GWAS variants with regions of cell-type-specific chromatin accessibility and examined the chromatin accessibility and gene expression of exemplar autoimmune gene loci.

Human ethics, tissue collection and preparation

Tonsil samples were collected from children and adults undergoing routine tonsillectomy. All participants provided written informed consent and the protocols were approved by Stanford University's Institutional Review Board (protocol numbers 30837 and 47690). Whole tonsils were collected in saline and processed within four hours of receipt. Tissues were treated with penicillin, streptomycin, and normocin for 30 minutes on ice and heavily clotted or cauterized areas of the tissue were removed. Tonsils were then dissected into small pieces (roughly 5-8 pieces per tonsil) before mechanical dissociation through a 100 µm cell strainer using a syringe plunger. Mononuclear cells were isolated by Ficoll density gradient centrifugation (GE Healthcare) and the buffy coats were collected. Cells were cryopreserved in 90% fetal bovine serum, 10% DMSO until use. Four additional cryopreserved tonsil samples at Queen Mary University of London included for scATAC-seq analyses were prepared as described previously (*9*) under approval from North West/Greater Manchester East Research Ethics Committee (17/NW/0664).

CyTOF staining and analysis

Cryopreserved samples were thawed in pre-warmed cell culture medium (RPMI1640 with 10 % FBS, non-essential amino acids, sodium pyruvate, antibiotics), washed, and rested for 1 hour at 37°C in culture medium supplemented with DNase (25 U/ml). Cells were then washed and resuspended in FACS buffer (PBS with 0.1% w/v bovine serum albumin, 2 mM EDTA, 0.05% v/v sodium azide). Individual donor samples were barcoded using a combination of metal-tagged CD45 antibodies, combined into barcoded pools, stained for surface antibody markers (Table S1), and treated with cisplatin for viability staining as described (*65*). Samples were then fixed overnight with 2% paraformaldehyde diluted in PBS. The next day, cells were permeabilized using a permeabilization buffer (eBioscience), stained with a DNA intercalator for 30 minutes, and washed. Just prior to CyTOF data collection, samples were washed three times with PBS, then three times with MilliQ water. Barcoded pools were run on a CyTOF2 instrument (Fluidigm) and fcs files were exported for analysis in FlowJo software. Live intact singlets were gated and samples were manually debarcoded using combinations of CD45 channels (5-choose-2 scheme) and individual donor samples were exported as separate fcs files before dimensionality reduction analyses.

Single-cell library preparation, sequencing and alignment.

Tonsillar immune cells were loaded on to the 10X Genomics Chromium according to the manufacturer's protocol using either the single-cell 3' kit (v3) or the single-cell ATAC kit (v1). Cell surface labelling for scADT-seq libraries was performed with 12 oligo-labelled TotalSeq™ antibodies (BioLegend; Table S2). Library preparation was performed according to the manufacturer's protocol prior to sequencing on either the Illumina NovaSeq 6000 or NextSeq 500 platforms. scRNA-seq libraries were sequenced with 28/10/10/90 bp cycles while scATAC-seq libraries were sequenced with 70/8/16/70 bp read configurations. BaseCall files were used to generate FASTQ files with either cellranger mkfastq (v3; 10X Genomics) or cellranger-atac (v1; 10X Genomics) prior to running cellranger count with the cellranger-GRCh38-3.0.0 reference or cellranger-atac count with the cellranger-atac-GRCh38-1.1.0 reference for scRNA-seq and scATAC-seq libraries respectively.

Quality control, integration and cell type annotation of tonsillar scRNA-seq

Gene expression count matrices from cellranger were processed with Seurat (v3.0.2) (*66, 67*) for genes detected in greater than 3 cells. Cell barcodes were filtered based on the number of genes per cell (between 200-7500), percentage of mitochondrial reads per cell (0-20 %) and the number of ADTs (less than 4000). Initial data quality control was performed separately on each biological sample. Data from technical replicate libraries were combined, normalized with SCTransform (*68*) before highly variable gene identification and PCA dimensionality reduction. Jackstraw plots were visually assessed to determine the number of principal components (PCs) for subsequent analysis: Tonsil1 = 11, Tonsil2 = 13, Tonsil3 = 12.

Preliminary clusters were identified (FindClusters; res = 0.8) before computing UMAP dimensionality reduction and identifying putative doublets with DoubletFinder (*69*) (sct=TRUE, expected_doublets=3.9%). Pre-processed Seurat objects were then merged, with SCTransform normalization and PCA computation repeated using all variable features (except for *IGKC*, *IGLC*, *IGLV*, *HLA*, and *IGH* genes). Batch correction was performed with harmony (*70*). UMAP dimensionality reduction and cluster identification were performed (27 PCs, res = 0.8). Broad cell type cluster frequencies (as in Fig1B) from an independent scRNA-seq analysis of human tonsils (*9*) were obtained to compare cell type frequencies between patients of different ages. For higher resolution analysis of B cells and T cells, data from B or T cells only were processed separately, with repeated variable gene identification (removing *IGKC*, *IGLC*, *IGLV*, *HLA*, and *IGH*) before repeated PCA, batch correction with Harmony, UMAP reduction and cluster identification (30 PCs, res = 0.6 for B cells; 20 PCs, res = 0.6 for T cells). Gene expression markers for clusters were identified (FindAllMarkers; log fold change > 1, adjusted p value < 0.05). Imputation of gene expression counts (for plotting only) was performed with MAGIC (*71*). Mean gene expression values per cell type per donor were used to calculate Spearman correlation coefficients between donors. Top 50 marker genes for the IFN_active B cell cluster were analyzed with the “Gene Set Query” function in the Autoimmune Disease Explorer (https://adex.genyo.es/) (*11*).

scATAC-seq quality control, batch correction and integration with scRNA-seq datasets

Mapped Tn5 insertion sites (fragments.tsv files) from cellranger were read into the ArchR (v0.9.4) package (*15*) retaining cell barcodes with at least 1000 fragments per cell and a TSS enrichment score > 4. Doublets were identified and filtered (addDoubletScores and filterDoublets, filter ratio = 1.4) before iterative LSI dimensionality reduction was computed (iterations = 2, res = 0.2, variable features = 25000, dim = 30). Sample batch correction was performed with harmony (*70*). Clustering was then performed on the harmony-corrected data (addClusters, res = 0.8) before UMAP dimensionality reduction (nNeighbors = 30, metric = cosine, minDist = 0.4). One cluster enriched for high doublet scores (cluster 7) was removed. A preliminary cell type annotation was performed using gene accessibility scores of known cell type markers. Tonsillar scRNA-seq gene expression and metadata were integrated with tonsillar scATAC data with ArchR as previously described (*15*). To improve cell type assignment of closely related cell types, we performed this step as a constrained integration, grouping GC B cell clusters, other B cell clusters and non-B cell clusters together during addGeneIntegrationMatrix. The most common predicted cell type from the integration with RNA expression in each previously identified ATAC-seq cluster was used to annotate scATAC cluster identity. The quality of mapping between the RNA and ATAC was confirmed by identifying marker gene scores in scATAC clusters using getMarkerFeatures. Additionally, cluster annotations derived from scATAC-only analysis were compared with annotations derived from scRNA-seq integration.

For high resolution clustering of B and T cell subsets (Fig2), scATAC clusters identified as B cells or T cells following scATAC/scRNA integration were subset, and use to recompute iterative LSI dimensionality reduction as described above, except 30 dimensions were used for B cell analysis. Batch correction, cluster identification and UMAP reduction were also performed as above, except that minDist = 0.1 (T cells) or 0.3 (B cells). Integration of B cell and T cell scATAC-seq datasets with gene expression and high resolution cluster annotations was performed using the T cell- or B cell-specific scRNA-seq Seurat objects as previously described with addGeneIntegrationMatrix in ArchR. Integration between assays were constrained with the following broad groups: B cell subgroups; plasmablasts, memory, naïve/activated and GC B cell clusters, T cells; CD8+/cytotoxic T cells and remaining T cell clusters. Mean peak accessibility scores per cell type per donor were used to calculate Spearman correlation coefficients between donors.

Peak calling and inference of transcription factor activity in scATAC-seq datasets.

Single-cell chromatin accessibility data were used to generate pseudobulk group coverages based on high resolution cluster identities of scATAC-seq datasets before peak calling with macs2 (*72*) using addReproduciblePeakSet in ArchR. A background peak set controlling for total accessibility and GC-content was generated using addBgdPeaks and used for TF motif enrichment analyses. Chromvar (*73*) was run with addDeviationsMatrix using the cisbp motif set to calculate enrichment of chromatin accessibility at different TF motif sequences in single cells. To identify correlations between the gene expression and transcription factor activity, RNA-expression projected into the ATAC subspace (GeneIntegrationMatrix) and the Chromvar deviations (MotifMatix) were correlated using correlateMatrices. A correlation of greater than 0.25 was used to determine if TF expression and activity were positively correlated, and the list of correlated TFs was further subset by only including TFs that were expressed in at least 25 percent of cells in one or more cell type cluster. To analyze transcription factor activity during B cell activation, GC entry and plasma differentiation, the harmony-corrected B cell ArchR object was subjected to “addTrajectory” from ArchR using the following user-defined trajectory as a guide: Naive→Activated→LZ GC→DZ GC→Plasmablasts. Gene expression and Chromvar deviation scores were correlated throughout pseudotime using correlateTrajectories (corCutOff = 0.25, varCutOff1 = 0.25, varCutOff2 = 0.25) and visualized using plotTrajectoryHeatmap. “Peak-to-gene links” were calculated using correlations between peak accessibility and integrated scRNA-seq expression data using addPeak2GeneLinks.

Integration of tonsil scATAC-seq and scRNA-seq with bone marrow and peripheral blood datasets

Published bone marrow and peripheral blood scRNA-seq and scATAC-seq (*22*) were aligned to the hg38 genome as described above. Additional hg38-aligned PBMC scATAC-seq datasets were downloaded from 10X Genomics (https://support.10xgenomics.com/single-cell-atac/datasets).

scRNA-seq

Cellranger gene expression matrices were used to sum and quantify mitochondrial gene expression before mitochondrial genes were removed from the gene expression matrices. Similarly, V, D and J gene counts from T cell and immunoglobulin receptors were summed and removed from matrices. Closely related IgH constant region genes were also summed and removed (IgG1-4, IgA1-2). Cell barcodes expressing >200 genes and genes detected in >3 cells were then processed in Seurat (*66, 67*), with doublet prediction using default settings with scrublet (*74*) (expected doublet frequency $8x10^{-6}$ X 1000 cells). Predicted doublets were removed, and cell barcodes with <750 or >30000 UMIs, <500 or >6000 genes detected, or >20% mitochondrial gene expression were also removed. Individual datasets were then merged together, before normalization and batch correction with SCTransform (3000 variable features) and scoring of cell cycle phase with Seurat. "IGLsum", "IGKsum", "IGHG", "IGHA", "IGHM", and "IGHD" were subsequently removed from highly variable gene list so they would not contribute to downstream dimensionality reductions. PCA was then computed before UMAP reduction (n.neighbors = 20, min.dist = 0.35, dims = 1:50), nearest neighbor identification (FindNeighbours; dims = 1:50) and cluster identification (FindClusters; res = 1.75). Some additional subclustering was performed to better match cell type annotations from previous tonsil analysis (this study) and peripheral blood/bone marrow analysis (*22*). In general, previous annotations were closely adhered to and confirmed by examination of known cell type-specific gene expression markers. Differential gene expression between clusters was performed with FindAllMarkers or FindMarkers, with *padj* < 0.05 and avg_logFC > 0.5. Imputation of gene expression counts (for plotting only) was performed with MAGIC (*71*).

scATAC-seq

Cellranger-derived fragments.tsv files of tonsil, peripheral blood and bone marrow samples were processed with ArchR (*15*) (createArrowFiles; filterTSS = 6, filterFrags = 1000, minFrags = 500, maxFrags = 1e+05). Doublets were identified (addDoubletScores; k=10) and removed with a filterRatio = 1.4, before additional filtering of cell barcodes to remove those with TSSEnrichment < 6, < $10^{3.25}$ or > 10^5 fragments per barcode, nucleosome ratio of > 2.5, ReadsInBlacklist > 800, or BlacklistRatio > 0.009. Preliminary LSI reduction was performed with addIterativeLSI (corCutOff = 0.25, varFeatures = 30000, dimsToUse = 1:40, selectionMethod = "var", LSIMethod = 1, iterations = 6, filterBias = FALSE, clusterParams = list(resolution = c(0.1,0.2,0.4,0.6,0.8,1), sampleCells = 10000, n.start = 10). To account for differences in sequencing coverage, Harmony batch correction (corCutOff = 0.25, lambda = 0.75, sigma = 0.2) was performed using library ID for tonsil samples, public 10X Genomics PBMC datasets and sample BMMC_D6T1, while remaining samples from Granja *et al* were treated as a single batch. Preliminary identification of clusters (addClusters; res = 1.5) identified two poor quality clusters enriched with doublets (C38, C7). These were

removed from subsequent analysis. Quality controlled datasets were then subjected to new LSI dimensionality reduction and Harmony batch correction with the same settings, before computing UMAP (RunUMAP; nNeighbors = 80, minDist = 0.45, seed = 1) and identifying cell type clusters with at least 80 cells (addClusters (method = "Seurat", res = 1.1 or 1.5, nOutlier = 80). Broad lineages were first annotated to help with integration and transfer of scRNA expression. Normalized, non-corrected scRNA expression counts and annotated cell types were transferred to nearest neighbor scATAC cells using addGeneIntegrationMatrix (sampleCellsATAC = 10000, nGenes (RNA) = 4000, sampleCellsRNA = 10000) with a constrained integration to the following groups: CD4T_cells, CD8T_cells, GC_PB, MBC_B_cells, Myeloid_cells, NaiveAct_B_cells, NK, Peripheral_B_cells, Progenitors. Accessibility gene scores and transferred RNA expression counts were imputed with addImputeWeights(corCutOff = 0.25). Cell type clusters were carefully annotated with a combination of pre-existing annotations from Granja *et al.* (*22*) and tonsil immune cell scATAC data (this study), transferred cell annotations from scRNA-seq and examination of known subset markers.

Pseudobulk group coverages of cell type clusters were calculated with addGroupCoverages and used for peak calling using macs2 (addReproduciblePeakSet in ArchR). A background peak set controlling for total accessibility and GC-content was generated using addBgdPeaks for TF enrichment analyses. Cell type-specific marker peaks were identified with getMarkerFeatures with the wilcoxon test and controlled for TSSEnrichment and fragment count. Peak accessibility was deemed significantly different between clusters if FDR < 0.05 and log2fc > 0.56. "Peak-to-gene links" were calculated using correlations between peak accessibility and integrated scRNA-seq expression data using addPeak2GeneLinks. Motif annotations and enrichment were calculated as described above with addMotifAnnotations and addDeviationsMatrix.

Analysis of fine-mapped GWAS variants

The results of two independent GWAS statistical fine-mapping studies (*1, 27*) (https://www.finucanelab.org/data) were combined. PICS SNPs from both immune and non-immune traits were included in analyses (*1*), while only SNPs from the study mapping the UK BioBank resource that were associated with a combined autoimmune disease trait (AID; labelled as AID_UKBB) were included (*27*). This provided a total of 12,902 non-redundant SNPs, of which 9,493 were significantly associated with disorders of the immune system. Fisher's exact test was used to calculate enrichment of immune trait-associated SNPs and non-immune trait-associated SNPs, against a background of common genetic variants (Common dbSnp153), in cell type-resolved peak sets or control background genomic intervals (either matched for GC content or distance to nearest TSS). Trait-specific enrichment analysis was performed using cell type-specific marker peaks (FDR < 0.05, log2FC > 0.25), with a background SNP set

comprising all fine-mapped SNPs across all traits. Cell type- and tissue- specificity of accessibility at SNPs was determined by presence or absence of a scATAC peak in each cell type, with cell type clusters regrouped based on enrichment in tonsils, peripheral blood or bone marrow. Of the immune-related SNPs that overlapped with accessible chromatin peaks (1213, 12.8%), we subsequently identified 460 unique immune-linked SNPs that fell within 358 chromatin accessible regions for which a significant Peak2Gene link had been identified to at least one gene (P2G_Correlation > 0.4; FDR < 0.01). Mean normalized chromatin accessibility counts (scATAC) and RNA expression counts for linked genes (scRNA) for each cell type cluster were calculated and used for heatmap visualization while pyGenomeTracks was used to visualize grouped scATAC pseudobulk tracks (*75*). Linkage disequilibrium scores of top candidate SNPs were calculated using LDlink across all populations (*76*).

Supplementary Materials

Figure S1. Single-cell library metadata, integration, batch correction and quality control.

Figure S2. Comparison of RNA expression, cell surface protein expression and chromatin accessibility of key marker genes.

Figure S3. Age-related changes in tonsillar immune cell populations by scRNA-seq and CyTOF.

Figure S4. Tonsil scRNA-seq marker gene heatmaps, annotation of GC B cells and reproducibility of cell types across donors.

Figure S5. Differential expression in autoimmune disease of top scRNA-seq marker genes for the IFN_active B cell cluster.

Figure S6. Reproducibility of scATAC-seq cluster frequencies and correlation of peak accessibilities between donors.

Figure S7. Differential peak analysis of scATAC-seq clusters, peak-to-gene predictions and alternative pseudotime analysis.

Figure S8. Batch correction and quality control of integrated bone marrow, blood and tonsil immune scRNA-seq and scATAC-seq.

Figure S9. Integrated bone marrow, blood and tonsil scRNA-seq and scATAC-seq markers.

Figure S10. Tonsil B cell-enriched gene expression markers compared to peripheral blood B cells.

Figure S11. Enrichment of fine-mapped autoimmune variants in immune cell subsets.

Figure S12. Genome snapshots of fine-mapped autoimmune variants at *GZMB*/*GZMH*, *NKX2-3* and *COTL1*/*KHLH36* loci.

Figure S13. Genome snapshots of fine-mapped autoimmune variants at *KSR1*/*LGALS9* and *TNFRSF1A*/*LTBR* loci.

Figure S14. Genome snapshots of germinal center-associated cell type-specific regulatory activity at fine-mapped autoimmune variants at *CD80*, *PRAG1* and *SLC38A9*/*DDX4* loci.

Figure S15. Genome snapshots of germinal center-associated cell type-specific regulatory activity at fine-mapped autoimmune variants at *VAV3* and *DLEU2* loci.
Figure S16. Linkage disequilibrium scores for variants at *IL21*, *IL21R* and *BCL6* loci.
Figure S17. Genome snapshots of fine-mapped autoimmune variants at *ETS1* and *IKZF3* loci.
Figure S18. Genome snapshots of fine-mapped autoimmune variants at *STAT4* and *IRF8* loci.
Figure S19. Genomic landscape at *HHEX* and expression of KLF family transcription factors.

Table S1. CyTOF phenotyping antibody panel.
Table S2. CITE-seq antibody details.

Data file S1. Tonsillectomy patient donor details, experimental study design and cell type frequencies across donors.
Data file S2. Broad resolution of tonsil immune cell subset scRNA-seq gene expression markers.
Data file S3. High resolution B cell subset scRNA-seq gene expression markers.
Data file S4. High resolution T cell subset scRNA-seq gene expression markers.
Data file S5. Differential chromatin accessibility peaks from high resolution annotation of tonsil immune cell scATAC-seq.
Data file S6. Differential chromatin accessibility gene scores from high resolution annotation of tonsil immune cell scATAC-seq.
Data file S7. Peak2gene predictions for tonsil scATAC-seq and scRNA-seq analysis.
Data file S8. Loop coordinates to visualize predicted tonsil immune peak2gene interactions.
Data file S9. Gene expression markers for integrated tonsil, peripheral blood and bone marrow immune cell populations from scRNA-seq.
Data file S10. Differential chromatin accessibility peak markers for integrated tonsil, peripheral blood and bone marrow immune cell populations from scATAC-seq.
Data file S11. Peak2gene predictions for integrated bone marrow, blood and tonsil scATAC-seq and scRNA-seq analysis.
Data file S12. Loop coordinates to visualize predicted integrated bone marrow, blood and tonsil immune peak2gene interactions.
Data file S13. Peak2gene linkage annotation of fine-mapped SNPs found in chromatin accessibility peaks from integrated tonsil, peripheral blood and bone marrow scATAC-seq datasets.
Data file S14. Raw Data file.

References

1. K. K.-H. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shoresh, H. Whitton, R. J. H. Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos, P. L. De Jager, V. K. Kuchroo, C. B.

Epstein, M. J. Daly, D. A. Hafler, B. E. Bernstein, Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-343 (2015).
2. D. Calderon, M. L. T. Nguyen, A. Mezger, A. Kathiria, F. Müller, V. Nguyen, N. Lescano, B. Wu, J. Trombetta, J. V. Ribado, D. A. Knowles, Z. Gao, F. Blaeschke, A. V. Parent, T. D. Burt, M. S. Anderson, L. A. Criswell, W. J. Greenleaf, A. Marson, J. K. Pritchard, Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat Genet* **51**, 1494-1505 (2019).
3. B. Soskic, E. Cano-Gamez, D. J. Smyth, W. C. Rowan, N. Nakic, J. Esparza-Gordillo, L. Bossini-Castillo, D. F. Tough, C. G. C. Larminie, P. G. Bronson, D. Willé, G. Trynka, Chromatin activity at GWAS loci identifies T cell states driving complex immune diseases. *Nat Genet* **51**, 1486-1493 (2019).
4. N. H. Ruddle, E. M. Akirav, Secondary lymphoid organs: responding to genetic and environmental cues in ontogeny and the immune response. *J Immunol* **183**, 2205-2212 (2009).
5. J. Suurmond, B. Diamond, Autoantibodies in systemic autoimmune diseases: specificity and pathogenicity. *J Clin Invest* **125**, 2194-2202 (2015).
6. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).
7. J. Lee, D.-Y. Chang, S.-W. Kim, Y. S. Choi, S.-Y. Jeon, V. Racanelli, D. W. Kim, E.-C. Shin, Age-related differences in human palatine tonsillar B cell subsets and immunoglobulin isotypes. *Clin Exp Med* **16**, 81-87 (2016).
8. S. Crotty, T follicular helper cell differentiation, function, and roles in disease. *Immunity* **41**, 529-542 (2014).
9. H. W. King, N. Orban, J. C. Riches, A. J. Clear, G. Warnes, S. A. Teichmann, L. K. James, Single-cell analysis of human B cell maturation predicts how antibody class switching shapes selection dynamics. *Sci Immunol* **6**, (2021).
10. J. Zhang, J. Shao, X. Wu, Q. Mao, Y. Wang, F. Gao, W. Kong, Z. Liang, Type I interferon related genes are common genes on the early stage after vaccination by meta-analysis of microarray data. *Hum Vaccin Immunother* **11**, 739-745 (2015).
11. J. Martorell-Marugán, R. López-Domínguez, A. García-Moreno, D. Toro-Domínguez, J. A. Villatoro-García, G. Barturen, A. Martín-Gómez, K. Troule, G. Gómez-López, F. Al-Shahrour, V. González-Rumayor, M. Peña-Chilet, J. Dopazo, J. Sáez-Rodríguez, M. E. Alarcón-Riquelme, P. Carmona-Sáez, A comprehensive database for integrated analysis of omics data in autoimmune diseases. *BMC Bioinformatics* **22**, 343 (2021).
12. J. Hutcheson, J. C. Scatizzi, A. M. Siddiqui, G. K. Haines, III, T. Wu, Q.-Z. Li, L. S. Davis, C. Mohan, H. Perlman, Combined Deficiency of Proapoptotic Regulators Bim and Fas Results in the Early Onset of Systemic Autoimmunity. *Immunity* **28**, 206-217 (2008).
13. A. M. Becker, K. H. Dao, B. K. Han, R. Kornu, S. Lakhanpal, A. B. Mobley, Q.-Z. Li, Y. Lian, T. Wu, A. M. Reimold, N. J. Olsen, D. R. Karp, F. Z. Chowdhury, J. D. Farrar, A. B. Satterthwaite, C. Mohan, P. E. Lipsky, E. K. Wakeland, L. S. Davis, SLE Peripheral Blood B Cell, T Cell and Myeloid Cell Transcriptomes Display Unique Profiles and Each Subset Contributes to the Interferon Signature. *PLoS ONE* **8**, e67003 (2013).
14. V. Kleshchevnikov, A. Shmatko, E. Dann, A. Aivazidis, H. W. King, T. Li, A. Lomakin, V. Kedlian, M. S. Jain, J. S. Park, L. Ramona, E. Tuck, A. Arutyunyan, R. Vento-Tormo, M. Gerstung, L. James, O. Stegle, O. A. Bayraktar, Comprehensive mapping of tissue cell architecture via integrated single cell and spatial transcriptomics. *bioRxiv*, 10.1101/2020.1111.1115.378125 (2020).
15. J. M. Granja, M. R. Corces, S. E. Pierce, S. T. Bagdatli, H. Choudhry, H. Y. Chang, W. J. Greenleaf, ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet*, (2021).
16. C. C. S. Hsiung, C. S. Morrissey, M. Udugama, C. L. Frank, C. A. Keller, S. Baek, B. Giardine, G. E. Crawford, M.-H. Sung, R. C. Hardison, G. A. Blobel, Genome accessibility is widely preserved and locally modulated during mitosis. *Genome Res* **25**, 213-225 (2015).
17. A. E. Trevino, F. Muller, J. Andersen, L. Sundaram, A. Kathiria, A. Shcherbina, K. Farh, H. Y. Chang, A. M. Pasca, A. Kundaje, S. P. Pasca, W. J. Greenleaf, Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *bioRxiv*, 10.1101/2020.1112.1129.424636 (2020).
18. J. M. Dan, C. Havenar-Daughton, K. Kendric, R. Al-Kolla, K. Kaushik, S. L. Rosales, E. L. Anderson, C. N. LaRock, P. Vijayanand, G. Seumois, D. Layfield, R. I. Cutress, C. H. Ottensmeier, C. S. Lindestam Arlehamn, A. Sette, V. Nizet, M. Bothwell, M. Brigger, S. Crotty, Recurrent group A Streptococcus tonsillitis is an immunosusceptibility disease involving antibody deficiency and aberrant TFH cells. *Sci Transl Med* **11**, (2019).
19. M. Breloer, B. Fleischer, CD83 regulates lymphocyte maturation, activation and homeostasis. *Trends Immunol* **29**, 186-194 (2008).
20. K. R. Duffy, C. J. Wellard, J. F. Markham, J. H. S. Zhou, R. Holmberg, E. D. Hawkins, J. Hasbold, M. R. Dowling, P. D. Hodgkin, Activation-induced B cell fates are selected by intracellular stochastic competition. *Science* **335**, 338-341 (2012).
21. L. E. Wagar, A. Salahudeen, C. M. Constantz, B. S. Wendel, M. M. Lyons, V. Mallajosyula, L. P. Jatt, J. Z. Adamska, L. K. Blum, N. Gupta, K. J. L. Jackson, F. Yang, K. Röltgen, K. M. Roskin, K. M. Blaine, K. D. Meister, I. N. Ahmad, M. Cortese, E. G. Dora, S. N. Tucker, A. I. Sperling, A. Jain, D. H. Davies, P. L. Felgner, G. B. Hammer, P. S. Kim, W. H. Robinson, S. D. Boyd, C. J. Kuo, M. M. Davis, Modeling human adaptive immune responses with tonsil organoids. *Nat Med* **27**, 125-135 (2021).
22. J. M. Granja, S. Klemm, L. M. McGinnis, A. S. Kathiria, A. Mezger, M. R. Corces, B. Parks, E. Gars, M. Liedtke, G. X. Y. Zheng, H. Y. Chang, R. Majeti, W. J. Greenleaf, Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol* **37**, 1458-1465 (2019).
23. L. Corcoran, D. Emslie, T. Kratina, W. Shi, S. Hirsch, N. Taubenheim, S. Chevrier, Oct2 and Obf1 as Facilitators of B:T Cell Collaboration during a Humoral Immune Response. *Front Immunol* **5**, 108 (2014).

24. J. E. D. Thaventhiran, H. Lango Allen, O. S. Burren, W. Rae, D. Greene, E. Staples, Z. Zhang, J. H. R. Farmery, I. Simeoni, E. Rivers, J. Maimaris, C. J. Penkett, J. Stephens, S. V. V. Deevi, A. Sanchis-Juan, N. S. Gleadall, M. J. Thomas, R. B. Sargur, P. Gordins, H. E. Baxendale, M. Brown, P. Tuijnenburg, A. Worth, S. Hanson, R. J. Linger, M. S. Buckland, P. J. Rayner-Matthews, K. C. Gilmour, C. Samarghitean, S. L. Seneviratne, D. M. Sansom, A. G. Lynch, K. Megy, E. Ellinghaus, D. Ellinghaus, S. F. Jorgensen, T. H. Karlsen, K. E. Stirrups, A. J. Cutler, D. S. Kumararatne, A. Chandra, J. D. M. Edgar, A. Herwadkar, N. Cooper, S. Grigoriadou, A. P. Huissoon, S. Goddard, S. Jolles, C. Schuetz, F. Boschann, N. B. Primary Immunodeficiency Consortium for the, P. A. Lyons, M. E. Hurles, S. Savic, S. O. Burns, T. W. Kuijpers, E. Turro, W. H. Ouwehand, A. J. Thrasher, K. G. C. Smith, Whole-genome sequencing of a sporadic primary immunodeficiency cohort. *Nature* **583**, 90-95 (2020).
25. U. Salzer, H. M. Chapel, A. D. B. Webster, Q. Pan-Hammarström, A. Schmitt-Graeff, M. Schlesier, H. H. Peter, J. K. Rockstroh, P. Schneider, A. A. Schäffer, L. Hammarström, B. Grimbacher, Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nat Genet* **37**, 820-828 (2005).
26. Q. Pan-Hammarström, U. Salzer, L. Du, J. Björkander, C. Cunningham-Rundles, D. L. Nelson, C. Bacchelli, H. B. Gaspar, S. Offer, T. W. Behrens, B. Grimbacher, L. Hammarström, Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA deficiency. *Nat Genet* **39**, 429-430 (2007).
27. E. M. Weeks, J. C. Ulirsch, N. Y. Cheng, B. L. Trippe, R. S. Fine, J. Miao, T. A. Patwardhan, M. Kanai, J. Nasser, C. P. Fulco, K. C. Tashman, F. Aguet, T. Li, J. Ordovas-Montanes, C. S. Smillie, M. Biton, A. K. Shalek, A. N. Ananthakrishnan, R. J. Xavier, A. Regev, R. M. Gupta, K. Lage, K. G. Ardlie, J. N. Hirschhorn, E. S. Lander, J. M. Engreitz, H. K. Finucane, Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *medRxiv*, 10.1101/2020.1109.1108.20190561 (2020).
28. D. A. van Heel, L. Franke, K. A. Hunt, R. Gwilliam, A. Zhernakova, M. Inouye, M. C. Wapenaar, M. C. N. M. Barnardo, G. Bethel, G. K. T. Holmes, C. Feighery, D. Jewell, D. Kelleher, P. Kumar, S. Travis, J. R. F. Walters, D. S. Sanders, P. Howdle, J. Swift, R. J. Playford, W. M. McLaren, M. L. Mearin, C. J. Mulder, R. McManus, R. McGinnis, L. R. Cardon, P. Deloukas, C. Wijmenga, A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* **39**, 827-829 (2007).
29. C. C. Robertson, J. R. J. Inshaw, S. Onengut-Gumuscu, W.-M. Chen, D. F. Santa Cruz, H. Yang, A. J. Cutler, D. J. M. Crouch, E. Farber, S. L. Bridges, J. C. Edberg, R. P. Kimberly, J. H. Buckner, P. Deloukas, J. Divers, D. Dabelea, J. M. Lawrence, S. Marcovina, A. S. Shah, C. J. Greenbaum, M. A. Atkinson, P. K. Gregersen, J. R. Oksenberg, F. Pociot, M. J. Rewers, A. K. Steck, D. B. Dunger, L. S. Wicker, P. Concannon, J. A. Todd, S. S. Rich, C. Type 1 Diabetes Genetics, Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes. *Nat Genet*, (2021).
30. J. Z. Liu, J. R. Hov, T. Folseraas, E. Ellinghaus, S. M. Rushbrook, N. T. Doncheva, O. A. Andreassen, R. K. Weersma, T. J. Weismüller, B. Eksteen, P. Invernizzi, G. M. Hirschfield, D. N. Gotthardt, A. Pares, D. Ellinghaus, T. Shah, B. D. Juran, P. Milkiewicz, C. Rust, C. Schramm, T. Müller, B. Srivastava, G. Dalekos, M. M. Nöthen, S. Herms, J. Winkelmann, M. Mitrovic, F. Braun, C. Y. Ponsioen, P. J. P. Croucher, M. Sterneck, A. Teufel, A. L. Mason, J. Saarela, V. Leppa, R. Dorfman, D. Alvaro, A. Floreani, S. Onengut-Gumuscu, S. S. Rich, W. K. Thompson, A. J. Schork, S. Næss, I. Thomsen, G. Mayr, I. R. König, K. Hveem, I. Cleynen, J. Gutierrez-Achury, I. Ricaño-Ponce, D. van Heel, E. Björnsson, R. N. Sandford, P. R. Durie, E. Melum, M. H. Vatn, M. S. Silverberg, R. H. Duerr, L. Padyukov, S. Brand, M. Sans, V. Annese, J.-P. Achkar, K. M. Boberg, H.-U. Marschall, O. Chazouillères, C. L. Bowlus, C. Wijmenga, E. Schrumpf, S. Vermeire, M. Albrecht, U.-P. Consortium, J. D. Rioux, G. Alexander, A. Bergquist, J. Cho, S. Schreiber, M. P. Manns, M. Färkkilä, A. M. Dale, R. W. Chapman, K. N. Lazaridis, I. P. S. Group, A. Franke, C. A. Anderson, T. H. Karlsen, I. I. G. Consortium, Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat Genet* **45**, 670-675 (2013).
31. T. Hughes, X. Kim-Howard, J. A. Kelly, K. M. Kaufman, C. D. Langefeld, J. Ziegler, E. Sanchez, R. P. Kimberly, J. C. Edberg, R. Ramsey-Goldman, M. Petri, J. D. Reveille, J. Martín, E. E. Brown, L. M. Vilá, G. S. Alarcón, J. A. James, G. S. Gilkeson, K. L. Moser, P. M. Gaffney, J. T. Merrill, T. J. Vyse, M. E. Alarcón-Riquelme, B. Network, S. K. Nath, J. B. Harley, A. H. Sawalha, Fine-mapping and transethnic genotyping establish IL2/IL21 genetic association with lupus and localize this genetic effect to IL21. *Arthritis Rheum* **63**, 1689-1697 (2011).
32. T. A. Olafsdottir, F. Theodors, K. Bjarnadottir, U. S. Bjornsdottir, A. B. Agustsdottir, O. A. Stefansson, E. V. Ivarsdottir, J. K. Sigurdsson, S. Benonisdottir, G. I. Eyjolfsson, D. Gislason, T. Gislason, S. Guðmundsdóttir, A. Gylfason, B. V. Halldorsson, G. H. Halldorsson, T. Juliusdottir, A. M. Kristinsdottir, D. Ludviksdottir, B. R. Ludviksson, G. Masson, K. Norland, P. T. Onundarson, I. Olafsson, O. Sigurdardottir, L. Stefansdottir, G. Sveinbjornsson, V. Tragante, D. F. Gudbjartsson, G. Thorleifsson, P. Sulem, U. Thorsteinsdottir, G. L. Norddahl, I. Jonsdottir, K. Stefansson, Eighty-eight variants highlight the role of T cell regulation and airway remodeling in asthma pathogenesis. *Nat Commun* **11**, 393 (2020).
33. K. L. Bunting, T. D. Soong, R. Singh, Y. Jiang, W. Béguelin, D. W. Poloway, B. L. Swed, K. Hatzi, W. Reisacher, M. Teater, O. Elemento, A. M. Melnick, Multi-tiered Reorganization of the Genome during B Cell Affinity Maturation Anchored by a Germinal Center-Specific Locus Control Region. *Immunity* **45**, 497-512 (2016).
34. A. Sharma, X. Liu, D. Hadley, W. Hagopian, E. Liu, W.-M. Chen, S. Onengut-Gumuscu, V. Simell, M. Rewers, A.-G. Ziegler, Å. Lernmark, O. Simell, J. Toppari, J. P. Krischer, B. Akolkar, S. S. Rich, D. Agardh, J.-X. She, T. S. Group, Identification of Non-HLA Genes Associated with Celiac Disease and Country-Specific Differences in a Large, International Pediatric Cohort. *PLoS ONE* **11**, e0152476 (2016).
35. R. Almeida, I. Ricaño-Ponce, V. Kumar, P. Deelen, A. Szperl, G. Trynka, J. Gutierrez-Achury, A. Kanterakis, H.-J. Westra, L. Franke, M. A. Swertz, M. Platteel, J. R. Bilbao, D. Barisani, L. Greco, L. Mearin, V. M. Wolters, C. Mulder, M. C. Mazzilli, A. Sood, B. Cukrowska, C. Núñez, R. Pratesi, S. Withoff, C. Wijmenga, Fine mapping of the celiac disease-associated LPP locus reveals a potential functional variant. *Hum Mol Genet* **23**, 2481-2489 (2014).

36. C. M. Lill, F. Luessi, A. Alcina, E. A. Sokolova, N. Ugidos, B. de la Hera, L. Guillot-Noël, S. Malhotra, E. Reinthaler, B.-M. M. Schjeide, J. Y. Mescheriakova, A. Mashychev, I. Wohlers, D. A. Akkad, O. Aktas, I. Alloza, A. Antigüedad, R. Arroyo, I. Astobiza, P. Blaschke, A. N. Boyko, M. Buttmann, A. Chan, T. Dörner, J. T. Epplen, O. O. Favorova, M. Fedetz, O. Fernández, A. García-Martínez, L.-A. Gerdes, C. Graetz, H.-P. Hartung, S. Hoffjan, G. Izquierdo, D. S. Korobko, A. Kroner, C. Kubisch, T. Kümpfel, L. Leyva, P. Lohse, N. A. Malkova, X. Montalban, E. V. Popova, P. Rieckmann, A. S. Rozhdestvenskii, C. Schmied, I. V. Smagina, E. Y. Tsareva, A. Winkelmann, U. K. Zettl, H. Binder, I. Cournu-Rebeix, R. Hintzen, A. Zimprich, M. Comabella, B. Fontaine, E. Urcelay, K. Vandenbroeck, M. Filipenko, F. Matesanz, F. Zipp, L. Bertram, Genome-wide significant association with seven novel multiple sclerosis risk loci. *J Med Genet* **52**, 848-855 (2015).
37. C. International Multiple Sclerosis Genetics, A. H. Beecham, N. A. Patsopoulos, D. K. Xifara, M. F. Davis, A. Kemppinen, C. Cotsapas, T. S. Shah, C. Spencer, D. Booth, A. Goris, A. Oturai, J. Saarela, B. Fontaine, B. Hemmer, C. Martin, F. Zipp, S. D'Alfonso, F. Martinelli-Boneschi, B. Taylor, H. F. Harbo, I. Kockum, J. Hillert, T. Olsson, M. Ban, J. R. Oksenberg, R. Hintzen, L. F. Barcellos, C. Wellcome Trust Case Control, I. B. D. G. C. International, C. Agliardi, L. Alfredsson, M. Alizadeh, C. Anderson, R. Andrews, H. B. Søndergaard, A. Baker, G. Band, S. E. Baranzini, N. Barizzone, J. Barrett, C. Bellenguez, L. Bergamaschi, L. Bernardinelli, A. Berthele, V. Biberacher, T. M. C. Binder, H. Blackburn, I. L. Bomfim, P. Brambilla, S. Broadley, B. Brochet, L. Brundin, D. Buck, H. Butzkueven, S. J. Caillier, W. Camu, W. Carpentier, P. Cavalla, E. G. Celius, I. Coman, G. Comi, L. Corrado, L. Cosemans, I. Cournu-Rebeix, B. A. C. Cree, D. Cusi, V. Damotte, G. Defer, S. R. Delgado, P. Deloukas, A. di Sapio, A. T. Dilthey, P. Donnelly, B. Dubois, M. Duddy, S. Edkins, I. Elovaara, F. Esposito, N. Evangelou, B. Fiddes, J. Field, A. Franke, C. Freeman, I. Y. Frohlich, D. Galimberti, C. Gieger, P.-A. Gourraud, C. Graetz, A. Graham, V. Grummel, C. Guaschino, A. Hadjixenofontos, H. Hakonarson, C. Halfpenny, G. Hall, P. Hall, A. Hamsten, J. Harley, T. Harrower, C. Hawkins, G. Hellenthal, C. Hillier, J. Hobart, M. Hoshi, S. E. Hunt, M. Jagodic, I. Jelčić, A. Jochim, B. Kendall, A. Kermode, T. Kilpatrick, K. Koivisto, I. Konidari, T. Korn, H. Kronsbein, C. Langford, M. Larsson, M. Lathrop, C. Lebrun-Frenay, J. Lechner-Scott, M. H. Lee, M. A. Leone, V. Leppä, G. Liberatore, B. A. Lie, C. M. Lill, M. Lindén, J. Link, F. Luessi, J. Lycke, F. Macciardi, S. Männistö, C. P. Manrique, R. Martin, V. Martinelli, D. Mason, G. Mazibrada, C. McCabe, I.-L. Mero, J. Mescheriakova, L. Moutsianas, K.-M. Myhr, G. Nagels, R. Nicholas, P. Nilsson, F. Piehl, M. Pirinen, S. E. Price, H. Quach, M. Reunanen, W. Robberecht, N. P. Robertson, M. Rodegher, D. Rog, M. Salvetti, N. C. Schnetz-Boutaud, F. Sellebjerg, R. C. Selter, C. Schaefer, S. Shaunak, L. Shen, S. Shields, V. Siffrin, M. Slee, P. S. Sorensen, M. Sorosina, M. Sospedra, A. Spurkland, A. Strange, E. Sundqvist, V. Thijs, J. Thorpe, A. Ticca, P. Tienari, C. van Duijn, E. M. Visser, S. Vucic, H. Westerlind, J. S. Wiley, A. Wilkins, J. F. Wilson, J. Winkelmann, J. Zajicek, E. Zindler, J. L. Haines, M. A. Pericak-Vance, A. J. Ivinson, G. Stewart, D. Hafler, S. L. Hauser, A. Compston, G. McVean, P. De Jager, S. J. Sawcer, J. L. McCauley, Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* **45**, 1353-1360 (2013).
38. Å. Johansson, M. Rask-Andersen, T. Karlsson, W. E. Ek, Genome-wide association analysis of 350 000 Caucasians from the UK Biobank identifies novel loci for asthma, hay fever and eczema. *Hum Mol Genet* **28**, 4022-4041 (2019).
39. P. K. Singh, P. R. van den Berg, M. D. Long, A. Vreugdenhil, L. Grieshober, H. M. Ochs-Balcom, J. Wang, S. Delcambre, S. Heikkinen, C. Carlberg, M. J. Campbell, L. E. Sucheston-Campbell, Integration of VDR genome wide binding and GWAS genetic variation data reveals co-occurrence of VDR and NF-κB binding that is linked to immune phenotypes. *BMC Genomics* **18**, 132 (2017).
40. M. J. Levels, C. M. Fehres, L. G. M. van Baarsen, N. O. P. van Uden, K. Germar, T. G. O'Toole, I. C. J. Blijdorp, J. F. Semmelink, M. E. Doorenspleet, A. Q. Bakker, M. Krasavin, A. Tomilin, S. Brouard, H. Spits, D. L. P. Baeten, N. G. Yeremenko, BOB.1 controls memory B-cell fate in the germinal center reaction. *J Autoimmun* **101**, 131-144 (2019).
41. B. J. Laidlaw, L. Duan, Y. Xu, S. E. Vazquez, J. G. Cyster, The transcription factor Hhex cooperates with the corepressor Tle3 to promote memory B cell development. *Nat Immunol* **21**, 1082-1093 (2020).
42. D. B. Schubart, A. Rolink, M. H. Kosco-Vilbois, F. Botteri, P. Matthias, B-cell-specific coactivator OBF-1/OCA-B/Bob1 required for immune response and germinal centre formation. *Nature* **383**, 538-542 (1996).
43. M. Nakamura, N. Nishida, M. Kawashima, Y. Aiba, A. Tanaka, M. Yasunami, H. Nakamura, A. Komori, M. Nakamuta, M. Zeniya, E. Hashimoto, H. Ohira, K. Yamamoto, M. Onji, S. Kaneko, M. Honda, S. Yamagiwa, K. Nakao, T. Ichida, H. Takikawa, M. Seike, T. Umemura, Y. Ueno, S. Sakisaka, K. Kikuchi, H. Ebinuma, N. Yamashiki, S. Tamura, Y. Sugawara, A. Mori, S. Yagi, K. Shirabe, A. Taketomi, K. Arai, K. Monoe, T. Ichikawa, M. Taniai, Y. Miyake, T. Kumagi, M. Abe, K. Yoshizawa, S. Joshita, S. Shimoda, K. Honda, H. Takahashi, K. Hirano, Y. Takeyama, K. Harada, K. Migita, M. Ito, H. Yatsuhashi, N. Fukushima, H. Ota, T. Komatsu, T. Saoshiro, J. Ishida, H. Kouno, H. Kouno, M. Yagura, M. Kobayashi, T. Muro, N. Masaki, K. Hirata, Y. Watanabe, Y. Nakamura, M. Shimada, N. Hirashima, T. Komeda, K. Sugi, M. Koga, K. Ario, E. Takesaki, Y. Maehara, S. Uemoto, N. Kokudo, H. Tsubouchi, M. Mizokami, Y. Nakanuma, K. Tokunaga, H. Ishibashi, Genome-wide association study identifies TNFSF15 and POU2AF1 as susceptibility loci for primary biliary cirrhosis in the Japanese population. *Am J Hum Genet* **91**, 721-728 (2012).
44. I. H. Haralambieva, I. G. Ovsyannikova, R. B. Kennedy, B. R. Larrabee, M. T. Zimmermann, D. E. Grill, D. J. Schaid, G. A. Poland, Genome-wide associations of CD46 and IFI44L genetic variants with neutralizing antibody response to measles vaccine. *Hum Genet* **136**, 421-435 (2017).
45. P. P. Domeier, S. B. Chodisetti, S. L. Schell, Y. I. Kawasawa, M. J. Fasnacht, C. Soni, Z. S. M. Rahman, B-Cell-Intrinsic Type 1 Interferon Signaling Is Crucial for Loss of Tolerance and the Development of Autoreactive B Cells. *Cell Rep* **24**, 406-418 (2018).
46. A. Camponeschi, L. Todi, C. Cristofoletti, C. Lazzeri, M. Carbonari, M. Mitrevski, R. Marrapodi, M. Del Padre, M. Fiorilli, M. Casato, M. Visentini, DEC1/STRA13 is a key negative regulator of activation-induced proliferation of human B cells highly expressed in anergic cells. *Immunol Lett* **198**, 7-11 (2018).

47. R. Rauschmeier, A. Reinhardt, C. Gustafsson, V. Glaros, A. V. Artemov, R. Taneja, I. Adameyko, R. Månsson, M. Busslinger, T. Kreslavsky, Cell-intrinsic functions of the transcription factor Bhlhe40 in activated B cells and T follicular helper cells restrain the germinal center reaction and prevent lymphomagenesis. *bioRxiv*, 2021.2003.2012.435122 (2021).
48. Y. Deng, M. Bartosovic, S. Ma, D. Zhang, Y. Liu, X. Qin, G. Su, M. L. Xu, S. Halene, J. E. Craft, G. Castelo-Branco, R. Fan, Spatial-ATAC-seq: spatially resolved chromatin accessibility profiling of tissues at genome scale and cellular level. *bioRxiv*, 2021.2006.2006.447244 (2021).
49. E. Cano-Gamez, G. Trynka, From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front Genet* **11**, (2020).
50. J. C. Ulirsch, C. A. Lareau, E. L. Bao, L. S. Ludwig, M. H. Guo, C. Benner, A. T. Satpathy, V. K. Kartha, R. M. Salem, J. N. Hirschhorn, H. K. Finucane, M. J. Aryee, J. D. Buenrostro, V. G. Sankaran, Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat Genet* **51**, 683-693 (2019).
51. C. Su, M. E. Johnson, A. Torres, R. M. Thomas, E. Manduchi, P. Sharma, P. Mehra, C. Le Coz, M. E. Leonard, S. Lu, K. M. Hodge, A. Chesi, J. Pippin, N. Romberg, S. F. A. Grant, A. D. Wells, Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular helper T cells. *Nat Commun* **11**, 3294 (2020).
52. M. R. Corces, J. M. Granja, S. Shams, B. H. Louie, J. A. Seoane, W. Zhou, T. C. Silva, C. Groeneveld, C. K. Wong, S. W. Cho, A. T. Satpathy, M. R. Mumbach, K. A. Hoadley, A. G. Robertson, N. C. Sheffield, I. Felau, M. A. A. Castro, B. P. Berman, L. M. Staudt, J. C. Zenklusen, P. W. Laird, C. Curtis, W. J. Greenleaf, H. Y. Chang, The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).
53. F. Aguet, A. A. Brown, S. E. Castel, J. R. Davis, Y. He, B. Jo, P. Mohammadi, Y. Park, P. Parsana, A. V. Segrè, B. J. Strober, Z. Zappala, B. B. Cummings, E. T. Gelfand, K. Hadley, K. H. Huang, M. Lek, X. Li, J. L. Nedzel, D. Y. Nguyen, M. S. Noble, T. J. Sullivan, T. Tukiainen, D. G. MacArthur, G. Getz, A. Addington, P. Guan, S. Koester, A. R. Little, N. C. Lockhart, H. M. Moore, A. Rao, J. P. Struewing, S. Volpi, L. E. Brigham, R. Hasz, M. Hunter, C. Johns, M. Johnson, G. Kopen, W. F. Leinweber, J. T. Lonsdale, A. McDonald, B. Mestichelli, K. Myer, B. Roe, M. Salvatore, S. Shad, J. A. Thomas, G. Walters, M. Washington, J. Wheeler, J. Bridge, B. A. Foster, B. M. Gillard, E. Karasik, R. Kumar, M. Miklos, M. T. Moser, S. D. Jewell, R. G. Montroy, D. C. Rohrer, D. Valley, D. C. Mash, D. A. Davis, L. Sobin, M. E. Barcus, P. A. Branton, N. S. Abell, B. Balliu, O. Delaneau, L. Frésard, E. R. Gamazon, D. Garrido-Martín, A. D. H. Gewirtz, G. Gliner, M. J. Gloudemans, B. Han, A. Z. He, F. Hormozdiari, X. Li, B. Liu, E. Y. Kang, I. C. McDowell, H. Ongen, J. J. Palowitch, C. B. Peterson, G. Quon, S. Ripke, A. Saha, A. A. Shabalin, T. C. Shimko, J. H. Sul, N. A. Teran, E. K. Tsang, H. Zhang, Y.-H. Zhou, C. D. Bustamante, N. J. Cox, R. Guigó, M. Kellis, M. I. McCarthy, D. F. Conrad, E. Eskin, G. Li, A. B. Nobel, C. Sabatti, B. E. Stranger, X. Wen, F. A. Wright, K. G. Ardlie, E. T. Dermitzakis, T. Lappalainen, F. Aguet, K. G. Ardlie, B. B. Cummings, E. T. Gelfand, G. Getz, K. Hadley, R. E. Handsaker, K. H. Huang, S. Kashin, K. J. Karczewski, M. Lek, X. Li, D. G. MacArthur, J. L. Nedzel, D. T. Nguyen, M. S. Noble, A. V. Segrè, C. A. Trowbridge, T. Tukiainen, N. S. Abell, B. Balliu, R. Barshir, O. Basha, A. Battle, G. K. Bogu, A. Brown, C. D. Brown, S. E. Castel, L. S. Chen, C. Chiang, D. F. Conrad, N. J. Cox, F. N. Damani, J. R. Davis, O. Delaneau, E. T. Dermitzakis, B. E. Engelhardt, E. Eskin, P. G. Ferreira, L. Frésard, E. R. Gamazon, D. Garrido-Martín, A. D. H. Gewirtz, G. Gliner, M. J. Gloudemans, R. Guigo, I. M. Hall, B. Han, Y. He, F. Hormozdiari, C. Howald, H. Kyung Im, B. Jo, E. Yong Kang, Y. Kim, S. Kim-Hellmuth, T. Lappalainen, G. Li, X. Li, B. Liu, S. Mangul, M. I. McCarthy, I. C. McDowell, P. Mohammadi, J. Monlong, S. B. Montgomery, M. Muñoz-Aguirre, A. W. Ndungu, D. L. Nicolae, A. B. Nobel, M. Oliva, H. Ongen, J. J. Palowitch, N. Panousis, P. Papasaikas, Y. Park, P. Parsana, A. J. Payne, C. B. Peterson, J. Quan, F. Reverter, C. Sabatti, A. Saha, M. Sammeth, A. J. Scott, A. A. Shabalin, R. Sodaei, M. Stephens, B. E. Stranger, B. J. Strober, J. H. Sul, E. K. Tsang, S. Urbut, M. van de Bunt, G. Wang, X. Wen, F. A. Wright, H. S. Xi, E. Yeger-Lotem, Z. Zappala, J. B. Zaugg, Y.-H. Zhou, J. M. Akey, D. Bates, J. Chan, L. S. Chen, M. Claussnitzer, K. Demanelis, M. Diegel, J. A. Doherty, A. P. Feinberg, M. S. Fernando, J. Halow, K. D. Hansen, E. Haugen, P. F. Hickey, L. Hou, F. Jasmine, R. Jian, L. Jiang, A. Johnson, R. Kaul, M. Kellis, M. G. Kibriya, K. Lee, J. Billy Li, Q. Li, X. Li, J. Lin, S. Lin, S. Linder, C. Linke, Y. Liu, M. T. Maurano, B. Molinie, S. B. Montgomery, J. Nelson, F. J. Neri, M. Oliva, Y. Park, B. L. Pierce, N. J. Rinaldi, L. F. Rizzardi, R. Sandstrom, A. Skol, K. S. Smith, M. P. Snyder, J. Stamatoyannopoulos, B. E. Stranger, H. Tang, E. K. Tsang, L. Wang, M. Wang, N. Van Wittenberghe, F. Wu, R. Zhang, C. R. Nierras, P. A. Branton, L. J. Carithers, P. Guan, H. M. Moore, A. Rao, J. B. Vaught, S. E. Gould, N. C. Lockart, C. Martin, J. P. Struewing, S. Volpi, A. M. Addington, S. E. Koester, A. R. Little, G. Consortium, Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).
54. Ž. Avsec, M. Weilert, A. Shrikumar, S. Krueger, A. Alexandari, K. Dalal, R. Fropf, C. McAnany, J. Gagneur, A. Kundaje, J. Zeitlinger, Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat Genet* **53**, 354-366 (2021).
55. Z. Mu, W. Wei, B. Fair, J. Miao, P. Zhu, Y. I. Li, The impact of cell type and context-dependent regulatory variants on human immune traits. *Genome Biol* **22**, 122 (2021).
56. J. W. Freimer, O. Shaked, S. Naqvi, N. Sinnott-Armstrong, A. Kathiria, A. F. Chen, J. T. Cortez, W. J. Greenleaf, J. K. Pritchard, A. Marson, Systematic discovery and perturbation of regulatory genes in human T cells reveals the architecture of immune networks. *bioRxiv*, 2021.2004.2018.440363 (2021).
57. C. G. Vinuesa, I. Sanz, M. C. Cook, Dysregulation of germinal centres in autoimmune disease. *Nat Rev Immunol* **9**, 845-857 (2009).
58. A. Pratama, C. G. Vinuesa, Control of TFH cell numbers: why and how? *Immunol Cell Biol* **92**, 40-48 (2014).
59. M. A. Linterman, L. Beaton, D. Yu, R. R. Ramiscal, M. Srivastava, J. J. Hogan, N. K. Verma, M. J. Smyth, R. J. Rigby, C. G. Vinuesa, IL-21 acts directly on B cells to regulate Bcl-6 expression and germinal center responses. *J Exp Med* **207**, 353-363 (2010).

60. C. G. McPhee, J. A. Bubier, T. J. Sproule, G. Park, M. P. Steinbuck, W. H. Schott, G. J. Christianson, H. C. Morse, D. C. Roopenian, IL-21 is a double-edged sword in the systemic lupus erythematosus-like disease of BXSB.Yaa mice. *J Immunol* **191**, 4581-4588 (2013).
61. P. P. Domeier, S. L. Schell, Z. S. M. Rahman, Spontaneous germinal centers and autoimmunity. *Autoimmunity* **50**, 4-18 (2017).
62. Y. Qin, P. Duquette, Y. Zhang, P. Talbot, R. Poole, J. Antel, Clonal expansion and somatic hypermutation of V(H) genes of B cells from cerebrospinal fluid in multiple sclerosis. *J Clin Invest* **102**, 1045-1050 (1998).
63. H. J. Kim, V. Krenn, G. Steinhauser, C. Berek, Plasma cell development in synovial germinal centers in patients with rheumatoid and reactive arthritis. *J Immunol* **162**, 3053-3062 (1999).
64. N. Amft, S. J. Curnow, D. Scheel-Toellner, A. Devadas, J. Oates, J. Crocker, J. Hamburger, J. Ainsworth, J. Mathews, M. Salmon, S. J. Bowman, C. D. Buckley, Ectopic expression of the B cell-attracting chemokine BCA-1 (CXCL13) on endothelial cells and within lymphoid follicles contributes to the establishment of germinal center-like structures in Sjögren's syndrome. *Arthritis Rheum* **44**, 2633-2641 (2001).
65. L. E. Wagar, Live cell barcoding for efficient analysis of small samples by mass cytometry. *Methods Mol Biol* **1989**, 125-135 (2019).
66. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018).
67. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e1821 (2019).
68. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 296 (2019).
69. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337.e324 (2019).
70. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296 (2019).
71. D. van Dijk, R. Sharma, J. Nainys, K. Yim, P. Kathail, A. J. Carr, C. Burdziak, K. R. Moon, C. L. Chaffer, D. Pattabiraman, B. Bierie, L. Mazutis, G. Wolf, S. Krishnaswamy, D. Pe'er, Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, 716-729.e727 (2018).
72. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
73. A. N. Schep, B. Wu, J. D. Buenrostro, W. J. Greenleaf, chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978 (2017).
74. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291.e289 (2019).
75. L. Lopez-Delisle, L. Rabbani, J. Wolff, V. Bhardwaj, R. Backofen, B. Grüning, F. Ramírez, T. Manke, pyGenomeTracks: reproducible plots for multivariate genomic data sets. *Bioinformatics*, (2020).
76. M. J. Machiela, S. J. Chanock, LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555-3557 (2015).

Acknowledgements

We thank all members of the Greenleaf and James laboratories for helpful comments and advice. We thank the QMUL Genome Centre for sequencing support and A. Orantes for administrative support.

Funding

This work was supported by funding from the Rita Allen Foundation (W.J.G.), the Human Frontiers Science (RGY006S) (W.J.G) and the Wellcome Trust (213555/Z/18/Z) (H.W.K). Z.S is supported by EMBO Long-Term Fellowship (EMBO ALTF 1119-2016) and by Human Frontier Science Program Long-Term Fellowship (HFSP LT 000835/2017-L). K.L.W is supported by a National Science Foundation GRFP award (DGE-1656518). W.J.G. is a Chan Zuckerberg Biohub investigator and acknowledges grants 2017-174468 and 2018-182817 from the Chan Zuckerberg Initiative.

Author contributions

H.W.K., K.L.W., Z.S. and W.J.G. conceived the project and designed experiments. H.W.K., Z.S., A.S.K. and L.E.W. processed samples for single cell experiments. L.E.W performed CyTOF experiments and data analysis. H.W.K., K.L.W., and Z.S. performed scRNA and scATAC data analysis. C.L. provided guidance for analysis and interpretation of GWAS variants. R.C. and N.O. provided tonsillectomy surgical samples. H.W.K., K.L.W., Z.S. and W.J.G. wrote the manuscript with input from all authors. M.M.D, L.M.S., L.K.J. and W.J.G supervised the work.

Competing interests

WJG is a consultant for 10x Genomics and Guardant Health, and is named as an inventor on patents describing ATAC-seq methods.

Data and materials availability

Raw and processed data for this study are available at Gene Expression Omnibus under accession GSE165860. All other data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials. All code and scripts necessary to repeat analysis in this manuscript is available upon request.

Figure Legends

Figure 1. Single-cell mapping of immune cell subsets in human tonsils.

A) Experimental strategy for single-cell transcriptomics, surface marker expression, and chromatin accessibility of immune cells from pediatric tonsils.
B) UMAP of tonsillar immune scRNA-seq data (left; 3 donors) and scATAC-seq data (right; 7 donors).
C) Heatmap comparing gene expression, surface protein, and chromatin accessibility across immune cell types.
D) UMAP of T cell sub-populations in the tonsillar immune scRNA-seq data in B). NK = natural killer, CTL = cytotoxic lymphocyte, Treg = regulatory T cell, TfH = T follicular helper cell, Tcm = T central memory.
E) Mean expression of key marker genes for T cell sub-populations by scRNA-seq. Frequency of cells for which each gene is detected is denoted by size of the dots.
F) UMAP of B cell sub-populations in the tonsillar immune scRNA-seq data in B). MBC = memory B cell, LZ GC = light zone germinal center, DZ GC = dark zone germinal center; IFN = interferon.
G) Mean expression of key marker genes for B cell sub-populations by scRNA-seq. Frequency of cells for which each gene is detected is denoted by size of the dots.

Figure 2. Tonsillar immune cell type-specific transcription factor regulatory activity.

A) UMAP of tonsillar immune scATAC-seq with high resolution annotation of immune cell types.
B) Correlation of TF motif deviation (enrichment) scores with TF expression (x axis) compared to TF motif deviation scores (y axis) to predict positive TF regulators across B cell populations.
C) Motif deviation scores (top panels) and RNA expression (bottom) for exemplar TFs.
D) Motif deviation scores for transcription factors (expressed in >25% cells in at least one cell type cluster). Mean gene expression is depicted by dot size.
E) Pseudotemporal reconstruction of B cell activation, GC entry and plasmablast differentiation using scATAC-seq. Dotted lines highlight major transition points between cell types. Top; TF motif deviations. Bottom; TF gene expression.
F) Grouped patterns of TF motif deviations (left) and TF gene expression (right) through B cell pseudotemporal reconstruction shown in (e). Colored line represents mean of all TFs per group (listed on right).
G) Genomic snapshot of tonsillar immune cell scATAC-seq tracks at *CD83* locus, highlighting rheumatoid arthritis-associated SNPs rs74405933 and rs12529514 and correlated peak2gene linkages. rs74405933 falls within an NFKB2 predicted binding site (G→T). scRNA-seq expression of *CD83* and *NFKB2* are shown to the right.

Figure 3. Integrated single-cell transcriptomics and epigenomics of human bone marrow, peripheral blood and tonsillar immune cell states.

A) UMAP of integrated scATAC-seq and scRNA-seq for human bone marrow, peripheral blood and tonsils. CLP: common lymphoid progenitors. GMP: granulocyte-monocyte progenitors. CM: central memory. EM: effector memory. CTL: cytotoxic lymphocyte.
B) Relative frequency of cell type clusters in A) across different tissues.
C) Differential scATAC-seq peak analysis of tonsillar compared to peripheral blood/bone marrow-enriched naïve and memory B cell (MBC) clusters. FCRL4+ MBC cluster was compared to peripheral blood-enriched MBC cluster.
D) Differential gene expression analysis of tonsillar compared to peripheral blood/bone marrow-enriched naïve and MBC clusters in integrated scRNA-seq dataset. Selected genes are annotated.
E) Ranking of TF motif deviation enrichment within tissue-enriched (red, upper) or tissue-depleted (blue, lower) peaks naïve and MBCs.
F) Expression of top genes identified to be mutated by whole genome sequencing in a sporadic immunodeficiency cohort (*24*). For TFs, motif deviation scores are also provided.

Figure 4. Autoimmune-associated genetic variants enriched in immune cell chromatin accessibility maps.

A) Fisher enrichment test of immune-associated fine mapped genetic variants, compared to common genetic variants, for chromatin accessibility scATAC peaks across 37 immune cell populations. Results for non-immune traits and background control peaks are shown. Dot size conveys significance (-log10(*p* value)).
B) Fisher enrichment test for trait-specific SNPs, compared to the complete fine-mapped SNP set, within cell type-specific chromatin accessibility peaks. Dot size conveys enrichment (odds ratio) and color denotes significance of enrichment.
C) Frequency histogram of immune-associated SNPs that fall within chromatin accessibility peaks across 37 immune cell types.
D) Tissue-specificity of chromatin accessibility peaks overlapping autoimmune SNPs.
E) Chromatin accessibility of peaks containing >1 immune-associated SNP (scATAC; left) for which at least one significant peak2gene correlation is identified. Expression of linked genes (scRNA; right) is also plotted. Accessibility or expression counts are scaled by peak or gene respectively.

Figure 5. Chromatin regulatory landscapes of GC-specific autoimmune risk variants.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants that localize to accessible chromatin in the integrated human bone marrow, peripheral blood and tonsil scATAC-seq atlas. High resolution of individual SNP loci and larger view of the *IL21* locus are shown, with significantly correlated peak2gene linkages colored in red and significant links

between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in the scATAC tracks. AID: autoimmune disease. IBD: inflammatory bowel disease. Juv Idio Arthritis: Juvenile idiopathic arthritis. Scl cholangitis: Primary sclerosing cholangitis.

B) Same as A), at the *IL4R*/*IL21R* locus.

C) Same as A), at the *BCL6*/*LPP* locus. A germinal center (GC)-specific locus control region (LCR) is highlighted in green. MS: multiple sclerosis.

Figure 6. Autoimmune risk variants at transcription regulator genes *POU2AF1* and *HHEX*.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants at the *POU2AF1* locus that localize to accessible chromatin in the integrated human bone marrow, peripheral blood and tonsil scATAC-seq atlas. Significantly correlated peak2gene linkages colored in red and significant links between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in scATAC tracks. PBC: primary biliary cirrhosis.

B) Same as A), at the *HHEX* locus. MS: multiple sclerosis.

Figure 1

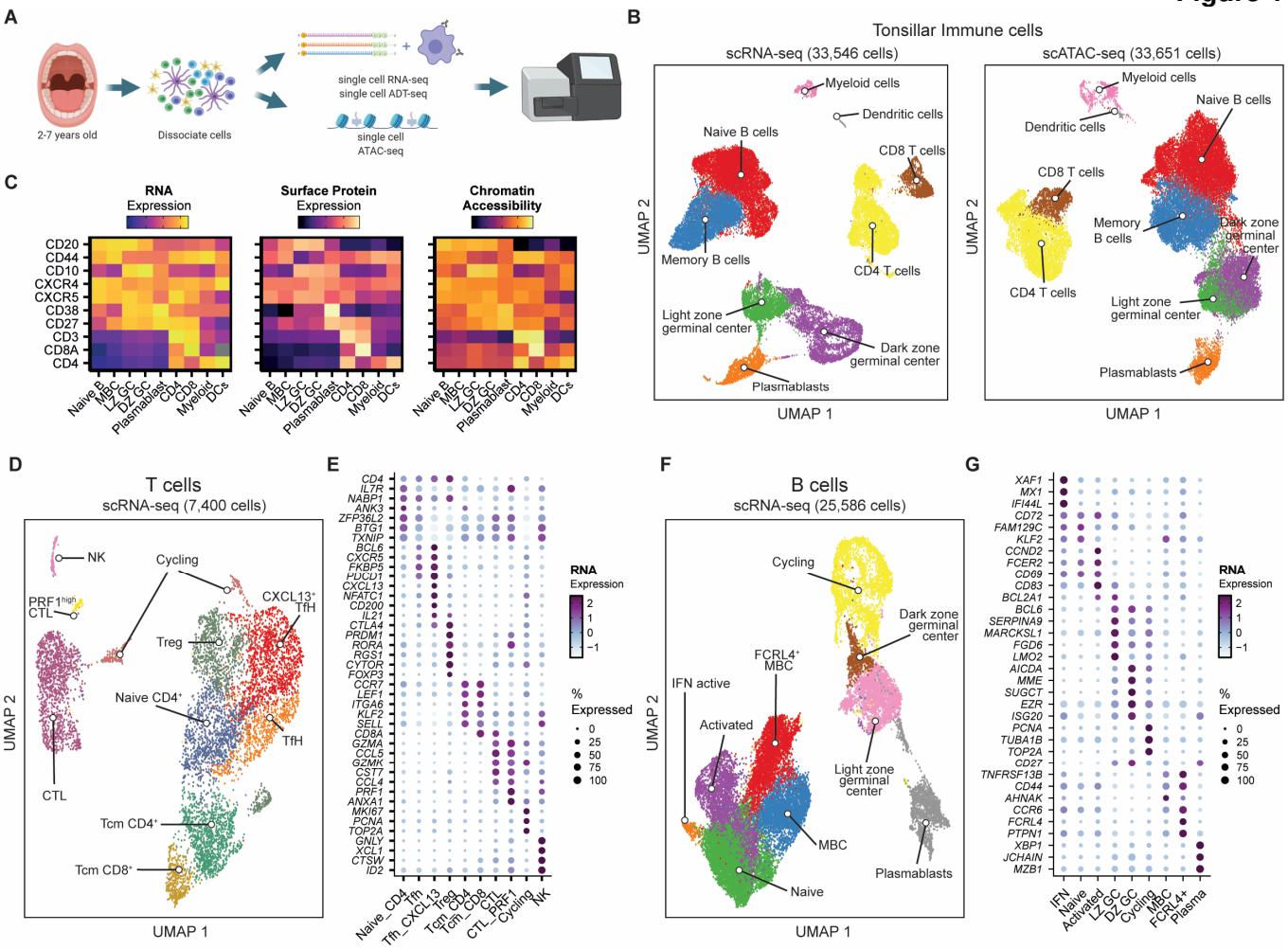

A
2-7 years old
Dissociate cells
single cell RNA-seq
single cell ADT-seq
single cell ATAC-seq
B
Tonsillar Immune cells
scRNA-seq (33,546 cells)
scATAC-seq (33,651 cells)
Myeloid cells
Dendritic cells
Naive B cells
CD8 T cells
Memory B cells
CD4 T cells
Light zone germinal center
Dark zone germinal center
Plasmablasts
UMAP 1
UMAP 2
C
RNA Expression
Surface Protein Expression
Chromatin Accessibility
CD20
CD44
CD10
CXCR4
CXCR5
CD38
CD27
CD3
CD8A
CD4
Naive B
MBC
LZ GC
DZ GC
Plasmablast
CD4
CD8
Myeloid
DCs
D
T cells
scRNA-seq (7,400 cells)
NK
Cycling
PRF1high CTL
CXCL13+ TfH
Treg
Naive CD4+
TfH
CTL
Tcm CD4+
Tcm CD8+
UMAP 1
UMAP 2
E
CD4
IL7R
NABP1
ANK3
ZFP36L2
BTG1
TXNIP
BCL6
CXCR5
FKBP5
PDCD1
CXCL13
NFATC1
CD200
IL21
CTLA4
PRDM1
RORA
RGS1
CYTOR
FOXP3
CCR7
LEF1
ITGA6
KLF2
SELL
CD8A
GZMK
CST7
CCL4
PRF1
ANXA1
MKI67
PCNA
TOP2A
GNLY
XCL1
CTSW
ID2
Naive_CD4
Tfh
Tfh_CXCL13
Treg
Tcm_CD4
Tcm_CD8
CTL
CTL_PRF1
Cycling
NK
RNA Expression
2
1
0
−1
% Expressed
0
25
50
75
100
F
B cells
scRNA-seq (25,586 cells)
Cycling
Dark zone germinal center
FCRL4+ MBC
IFN active
Activated
Light zone germinal center
MBC
Naive
Plasmablasts
UMAP 1
UMAP 2
G
XAF1
MX1
IFI44L
CD72
FAM129C
KLF2
CCND2
FCER2
CD69
CD83
BCL2A1
BCL6
SERPINA9
MARCKSL1
FGD6
LMO2
AICDA
MME
SUGCT
EZR
ISG20
PCNA
TUBA1B
TOP2A
CD27
TNFRSF13B
CD44
AHNAK
CCR6
FCRL4
PTPN1
XBP1
JCHAIN
MZB1
IFN
Naive
Activated
LZ GC
DZ GC
Cycling
MBC
FCRL4+
Plasma
RNA Expression
2
1
0
−1
% Expressed
0
25
50
75
100

Figure 2

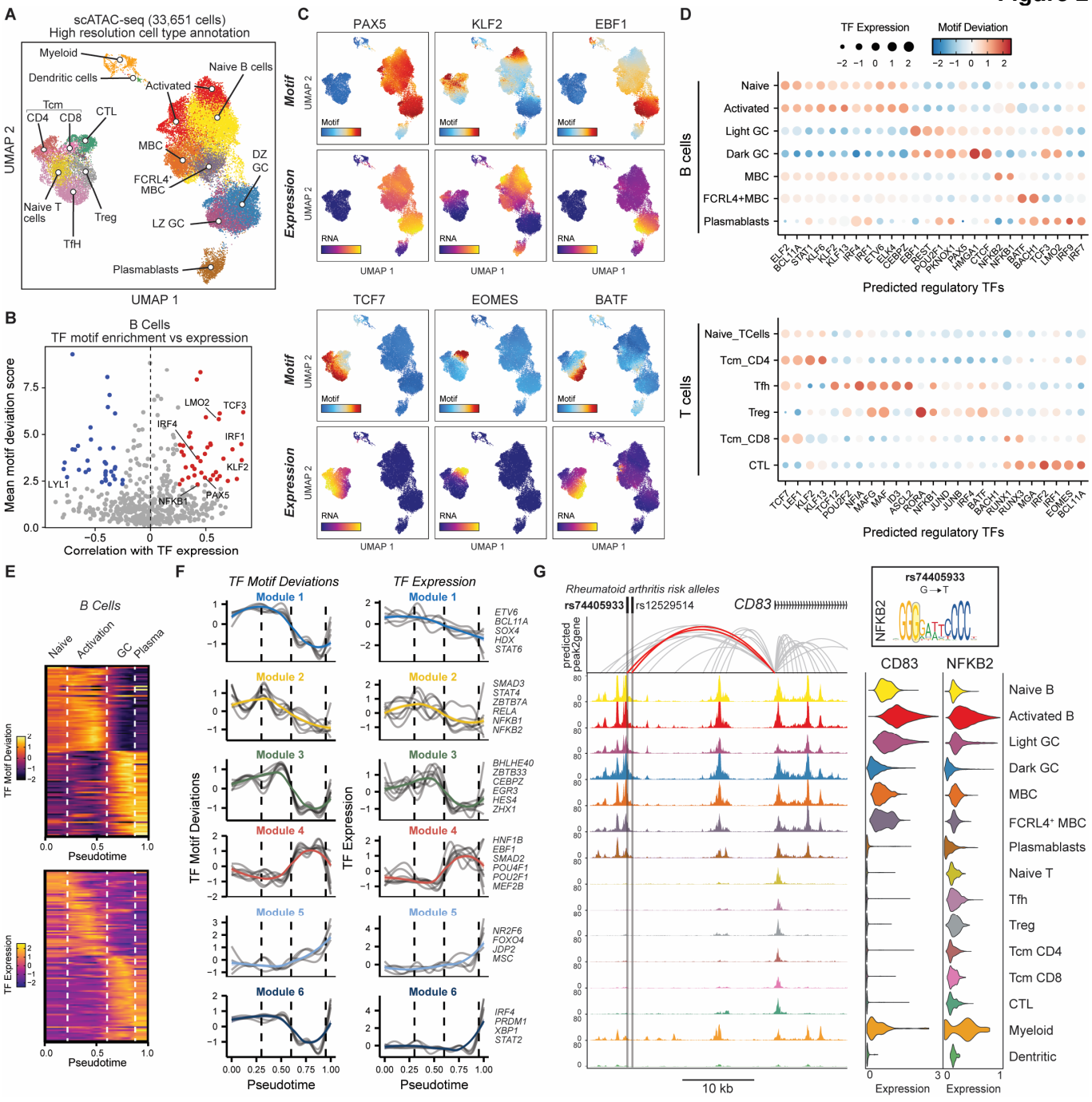

A
scATAC-seq (33,651 cells)
High resolution cell type annotation
Myeloid
Dendritic cells
Activated
Naive B cells
Tcm
CD4
CD8
CTL
MBC
DZ GC
FCRL4+ MBC
Naive T cells
Treg
LZ GC
TfH
Plasmablasts
UMAP 2
UMAP 1
B
B Cells
TF motif enrichment vs expression
Mean motif deviation score
Correlation with TF expression
LMO2
TCF3
IRF4
IRF1
KLF2
LYL1
NFKB1
PAX5
C
PAX5
KLF2
EBF1
TCF7
EOMES
BATF
Motif
Expression
RNA
UMAP 1
UMAP 2
D
TF Expression
Motif Deviation
B cells
Naive
Activated
Light GC
Dark GC
MBC
FCRL4+MBC
Plasmablasts
Predicted regulatory TFs
T cells
Naive_TCells
Tcm_CD4
Tfh
Treg
Tcm_CD8
CTL
E
B Cells
Naive
Activation
GC
Plasma
TF Motif Deviation
TF Expression
Pseudotime
F
TF Motif Deviations
TF Expression
Module 1
ETV6
BCL11A
SOX4
HDX
STAT6
Module 2
SMAD3
STAT4
ZBTB7A
RELA
NFKB1
NFKB2
Module 3
BHLHE40
ZBTB33
CEBPZ
EGR3
HES4
ZHX1
Module 4
HNF1B
EBF1
SMAD2
POU4F1
POU2F1
MEF2B
Module 5
NR2F6
FOXO4
JDP2
MSC
Module 6
IRF4
PRDM1
XBP1
STAT2
Pseudotime
G
Rheumatoid arthritis risk alleles
rs74405933
rs12529514
CD83
predicted peak2gene
10 kb
NFKB2
CD83
NFKB2
Naive B
Activated B
Light GC
Dark GC
MBC
FCRL4+ MBC
Plasmablasts
Naive T
Tfh
Treg
Tcm CD4
Tcm CD8
CTL
Myeloid
Dentritic
Expression

Figure 3

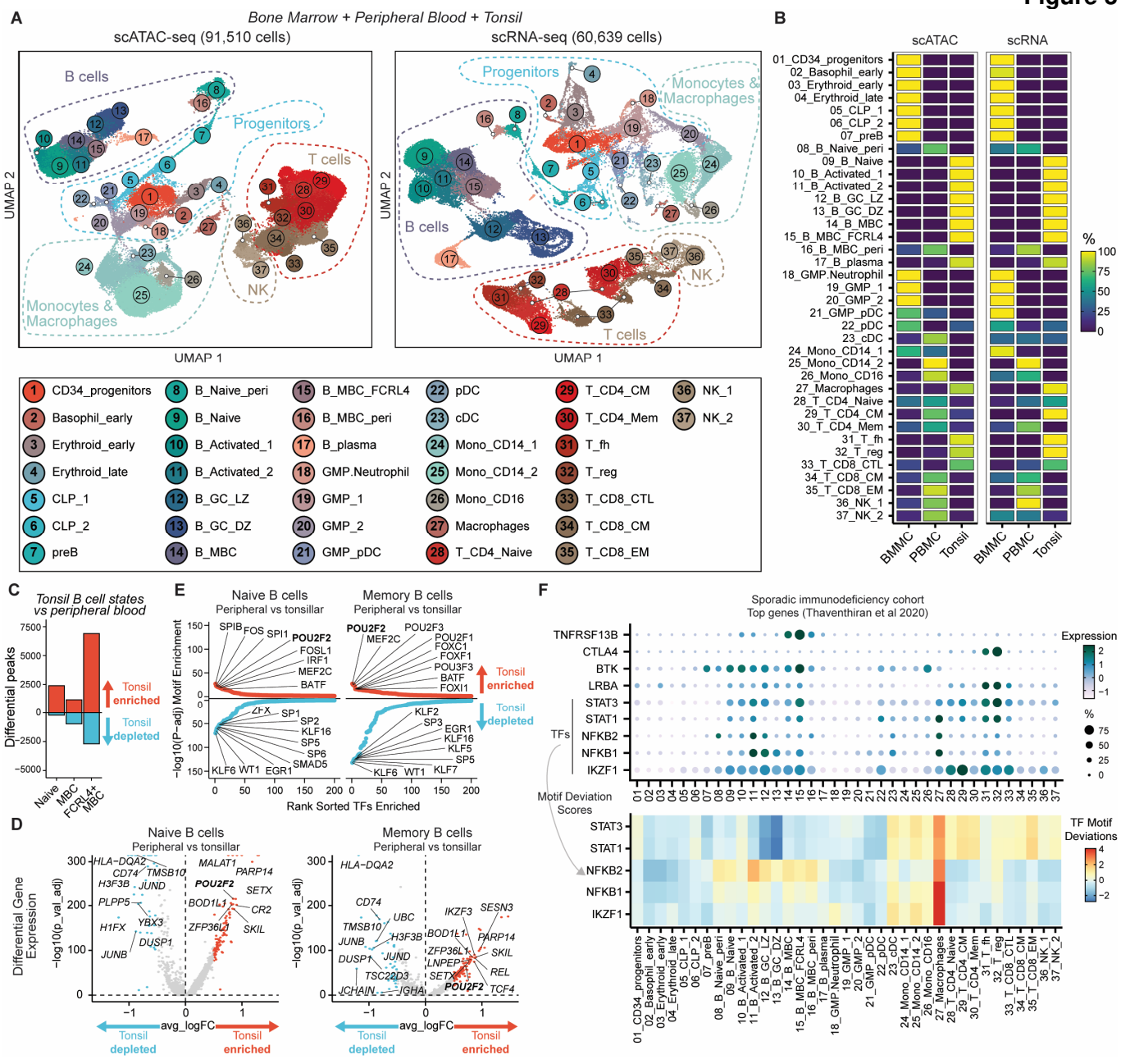
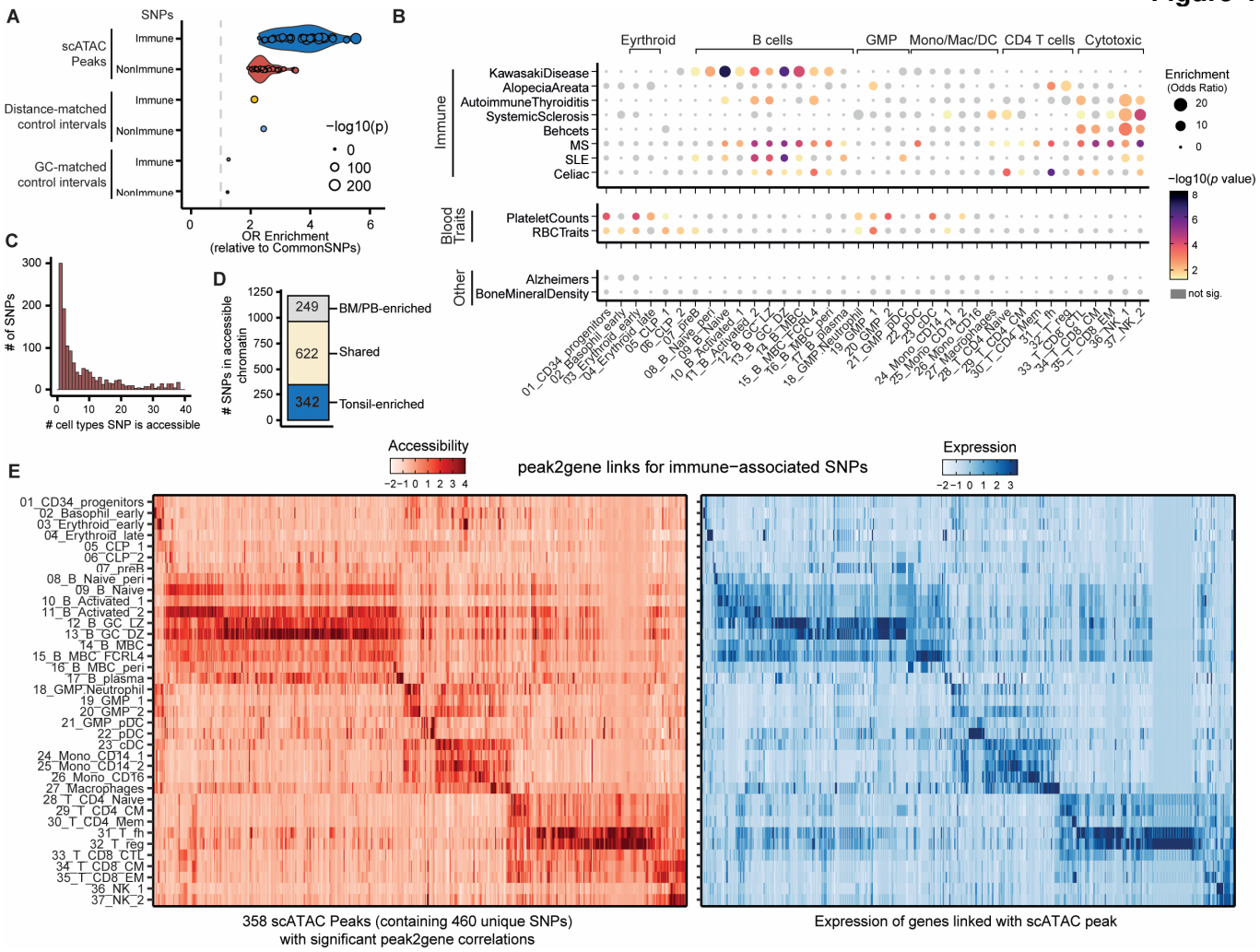
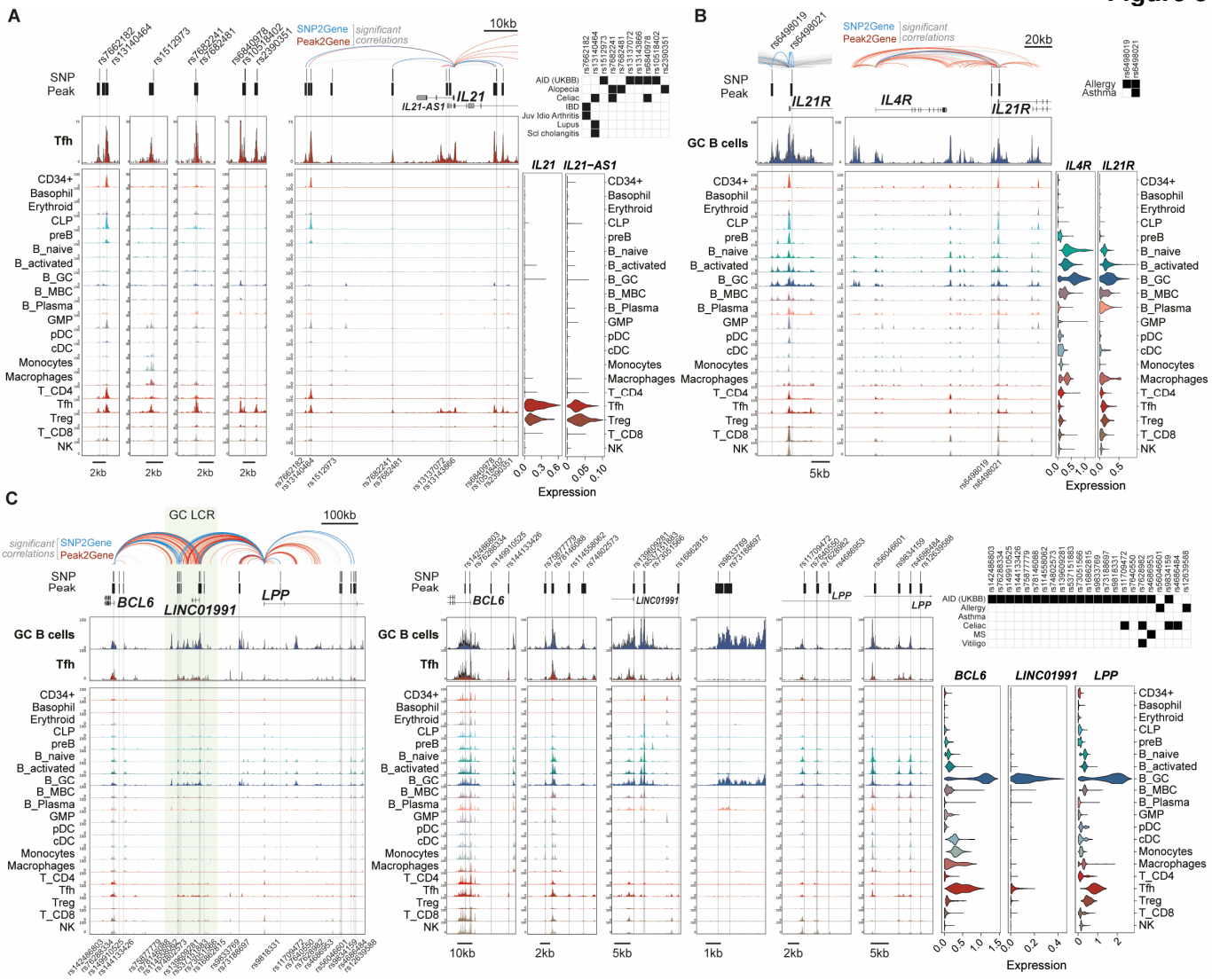

A
Bone Marrow + Peripheral Blood + Tonsil
scATAC-seq (91,510 cells)
scRNA-seq (60,639 cells)
B cells
Progenitors
T cells
Monocytes & Macrophages
NK
UMAP 1
UMAP 2
1 CD34_progenitors
2 Basophil_early
3 Erythroid_early
4 Erythroid_late
5 CLP_1
6 CLP_2
7 preB
8 B_Naive_peri
9 B_Naive
10 B_Activated_1
11 B_Activated_2
12 B_GC_LZ
13 B_GC_DZ
14 B_MBC
15 B_MBC_FCRL4
16 B_MBC_peri
17 B_plasma
18 GMP.Neutrophil
19 GMP_1
20 GMP_2
21 GMP_pDC
22 pDC
23 cDC
24 Mono_CD14_1
25 Mono_CD14_2
26 Mono_CD16
27 Macrophages
28 T_CD4_Naive
29 T_CD4_CM
30 T_CD4_Mem
31 T_fh
32 T_reg
33 T_CD8_CTL
34 T_CD8_CM
35 T_CD8_EM
36 NK_1
37 NK_2
B
scATAC
scRNA
BMMC
PBMC
Tonsil
%
C
Tonsil B cell states vs peripheral blood
Differential peaks
Tonsil enriched
Tonsil depleted
Naive
MBC
FCRL4+ MBC
D
Naive B cells
Peripheral vs tonsillar
Memory B cells
Peripheral vs tonsillar
Differential Gene Expression
−log10(p_val_adj)
avg_logFC
Tonsil depleted
Tonsil enriched
E
Naive B cells
Peripheral vs tonsillar
Memory B cells
Peripheral vs tonsillar
−log10(P−adj) Motif Enrichment
Rank Sorted TFs Enriched
F
Sporadic immunodeficiency cohort
Top genes (Thaventhiran et al 2020)
Expression
TFs
Motif Deviation Scores
TF Motif Deviations

Figure 4


A
SNPs
scATAC Peaks
Immune
NonImmune
Distance-matched control intervals
GC-matched control intervals
−log10(p)
0
100
200
OR Enrichment (relative to CommonSNPs)
B
Eyrthroid
B cells
GMP
Mono/Mac/DC
CD4 T cells
Cytotoxic
Immune
KawasakiDisease
AlopeciaAreata
AutoimmuneThyroiditis
SystemicSclerosis
Behcets
MS
SLE
Celiac
Blood Traits
PlateletCounts
RBCTraits
Other
Alzheimers
BoneMineralDensity
Enrichment (Odds Ratio)
−log10(p value)
not sig.
C
# of SNPs
cell types SNP is accessible
D
SNPs in accessible chromatin
249
622
342
BM/PB-enriched
Shared
Tonsil-enriched
E
Accessibility
peak2gene links for immune−associated SNPs
Expression
358 scATAC Peaks (containing 460 unique SNPs) with significant peak2gene correlations
Expression of genes linked with scATAC peak

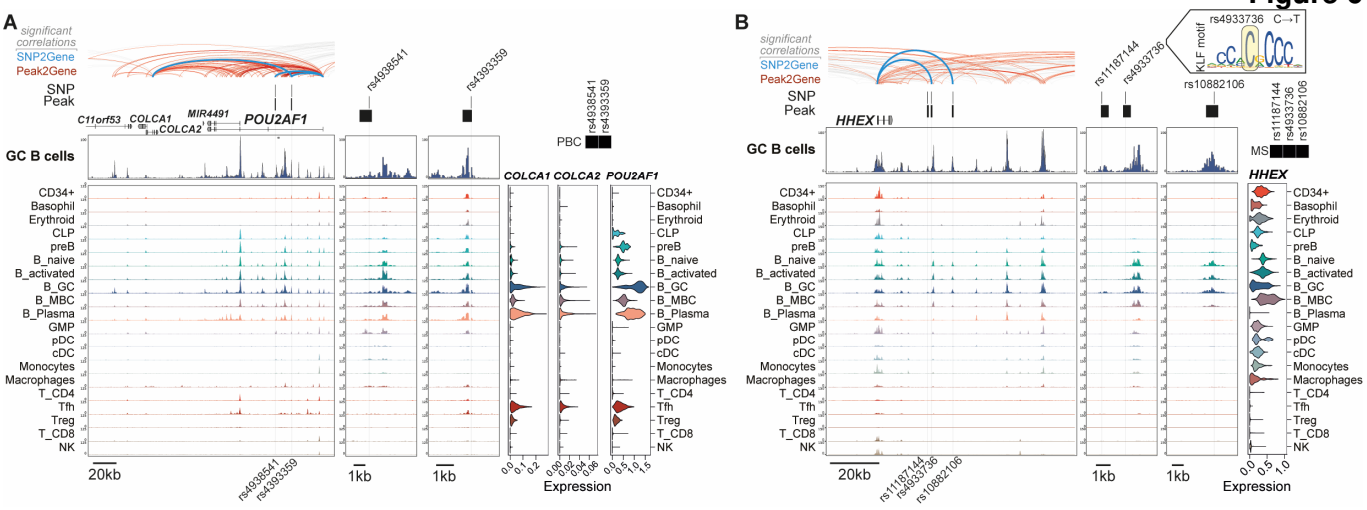
1139

Figure 5



1140

Figure 6


A
significant correlations
SNP2Gene
Peak2Gene
SNP
Peak
C11orf53
COLCA1
COLCA2
MIR4491
POU2AF1
rs4938541
rs4393359
PBC
GC B cells
COLCA1 COLCA2 POU2AF1
CD34+
Basophil
Erythroid
CLP
preB
B_naive
B_activated
B_GC
B_MBC
B_Plasma
GMP
pDC
cDC
Monocytes
Macrophages
T_CD4
Tfh
Treg
T_CD8
NK
20kb
1kb
1kb
Expression
B
rs4933736 C→T
KLF motif
rs11187144
rs10882106
MS
HHEX
20kb
1kb
1kb
Expression

A

Tonsil	PatientID	Age	Sex	Clinical indication	Site	scRNA	scADT	scATAC
1	IMD030	2	F	OSA	Stanford	X	X	X
2	IMD109	5	F	OSA	Stanford	X	X	X
3	IMD117	7	M	OSA	Stanford	X	X	X
4	BCP003	6	F	OSA	QMUL			X
5	BCP004	6	M	RT	QMUL			X
6	BCP005	3	M	RT	QMUL			X
7	BCP006	5	F	OSA+RT	QMUL			X

OSA; Obstructive sleep apnea
RT; recurrent tonsillitis

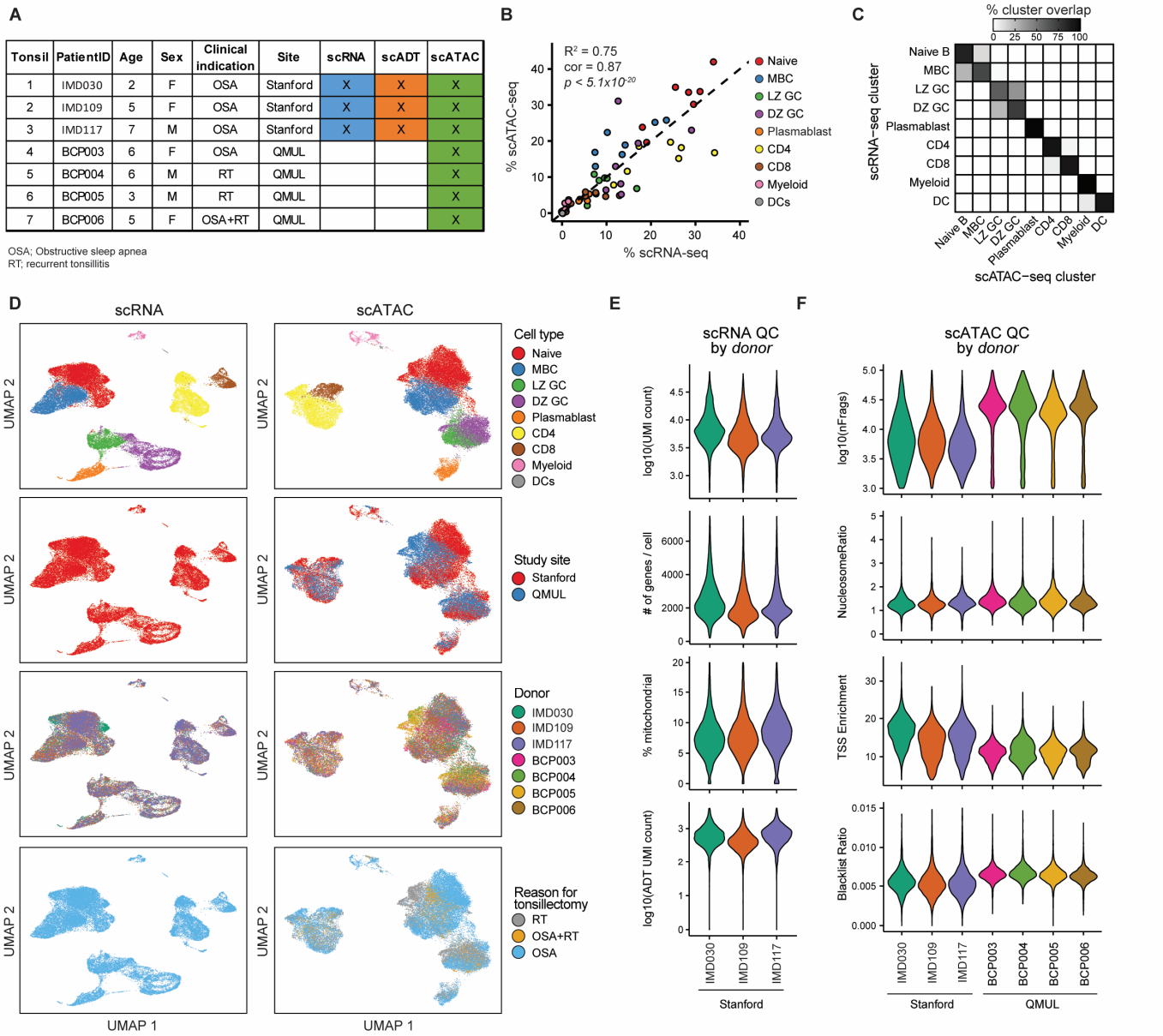


Figure S1. Single-cell library metadata, integration, batch correction and quality control.

A) Metadata of tonsillectomy patients analyzed by single-cell genomics in this study, including age, sex, reason for tonsillectomy (RT; recurrent tonsillitis, OSA; obstructive sleep apnea) and site of study.

B) Correlation of relative cluster frequencies for scRNA-seq and scATAC-seq within each donor.

C) Confusion matrix depicting overlap between transferred scRNA-seq cluster identities to scATAC-seq clusters.

D) UMAP visualization of major cell type clusters, site of study, donor and clinical indication for tonsillectomy for scRNA-seq and scATAC-seq datasets

E) Quality control metrics for scRNA-seq datasets by donor, including unique molecular identifier (UMI) counts per cell barcode, number of genes detected per cell barcode, percentage mitochondrial gene expression and cell surface ADT UMI counts.

F) Quality control metrics for scATAC-seq datasets by donor, including number of unique fragments per cell barcode, ratio of nucleosomal to non-nucleosomal fragment sizes, transcription start site enrichment score, and ratio of fragments in genomic blacklist regions (see ArchR for details).

Figure S2. Comparison of RNA expression, cell surface protein expression and chromatin accessibility of key marker genes.

scRNA-seq gene expression (top rows), CITE-seq surface protein expression (middle rows) and chromatin accessibility gene scores (bottom rows) for key marker genes. Gene expression and surface protein expression are visualized on the scRNA-seq UMAP manifold and chromatin accessibility scores are visualized on the scATAC-seq UMAP manifold (see Figure 1b).

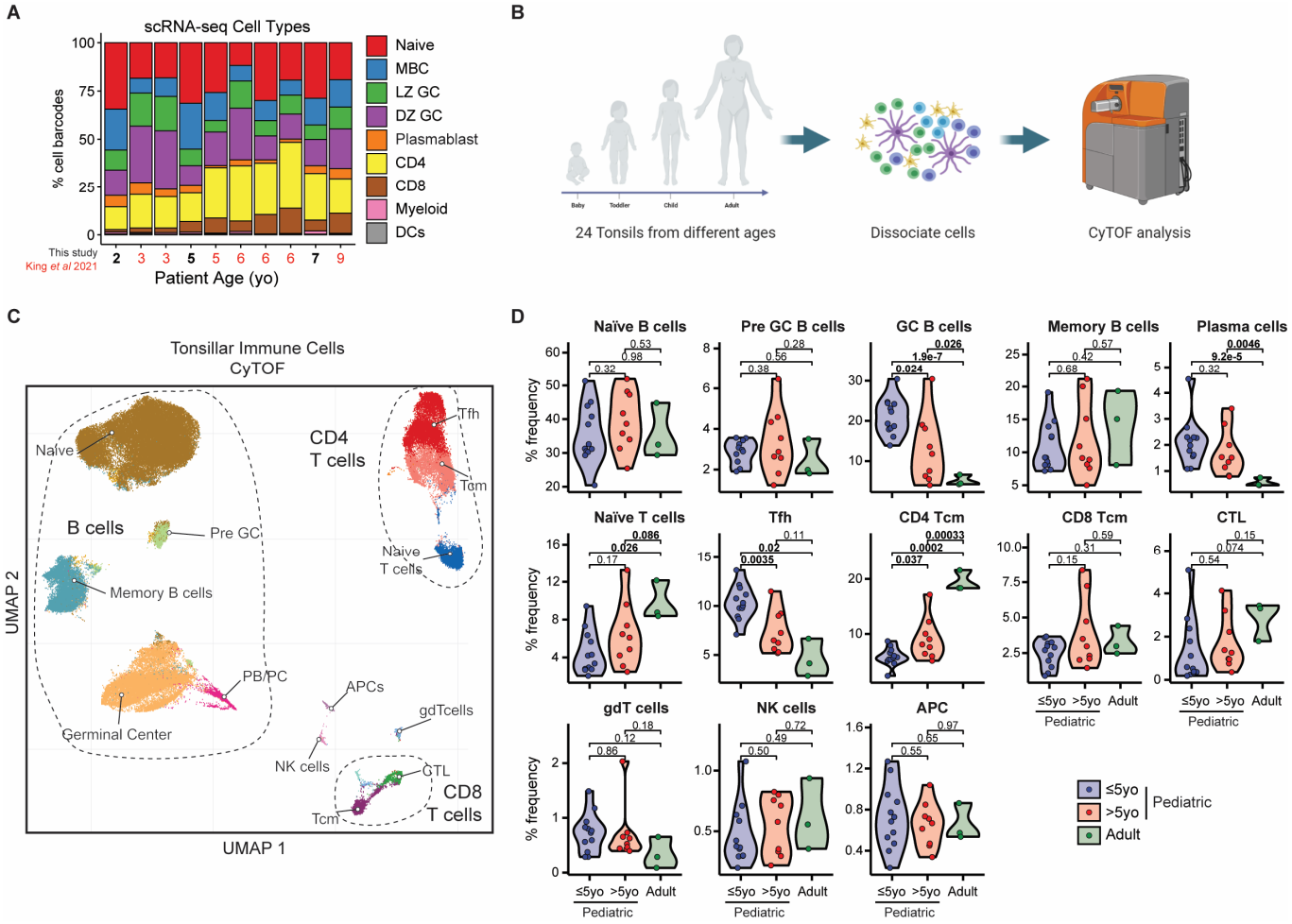


Figure S3. Age-related changes in tonsillar immune cell populations by scRNA-seq and CyTOF.

A) Relative scRNA-seq cluster frequencies of different donors ordered by patient age. Additional tonsillar immune cell samples from King *et al.* (2021) are included (see red labels).

B) Schematic of CyTOF analyses for age-related differences in tonsillar immune cell populations.

C) UMAP visualization of CyTOF analysis of tonsillar immune cells with major immune cell populations (n=24).

D) Quantitation of relative frequencies of immune cell subsets separated by patient age. p values denote results from Student's t test.

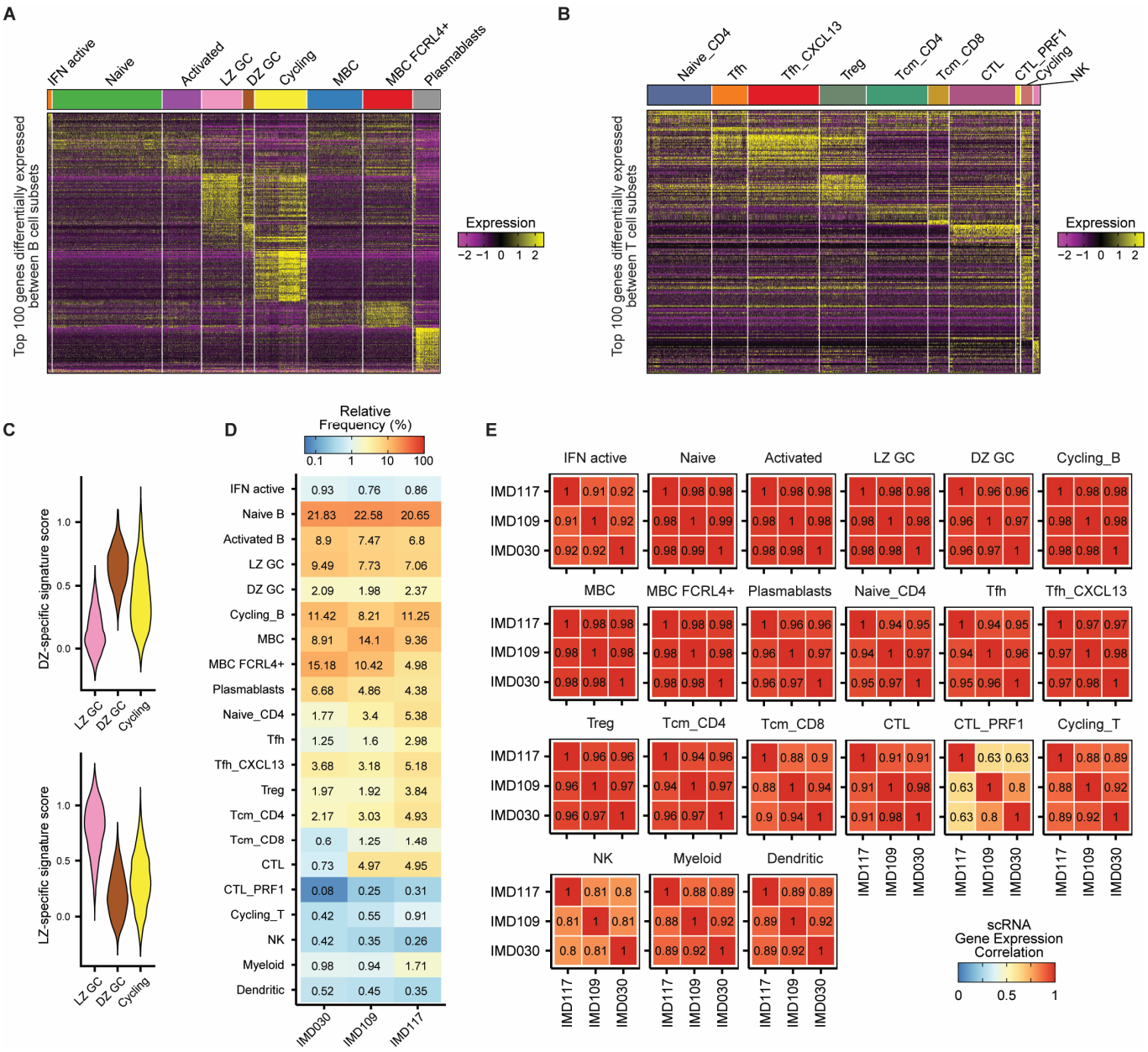


Figure S4. Tonsil scRNA-seq marker gene heatmaps, annotation of GC B cells and reproducibility of cell types across donors.

A) scRNA-seq heatmap of top 100 most differentially expressed genes for high resolution B cell clusters (as in Fig1F-G).
B) scRNA-seq heatmap of top 100 most differentially expressed genes for high resolution T cell clusters (as in Fig1D-E).
C) Gene signature scores for DZ-specific and LZ-specific marker genes in LZ GC, DZ GC and cycling B cell clusters.
D) Relative frequency of high resolution scRNA-seq cell type clusters across the three patient donors.
E) Spearman correlation coefficients of mean gene expression per cell type cluster between patient donors.

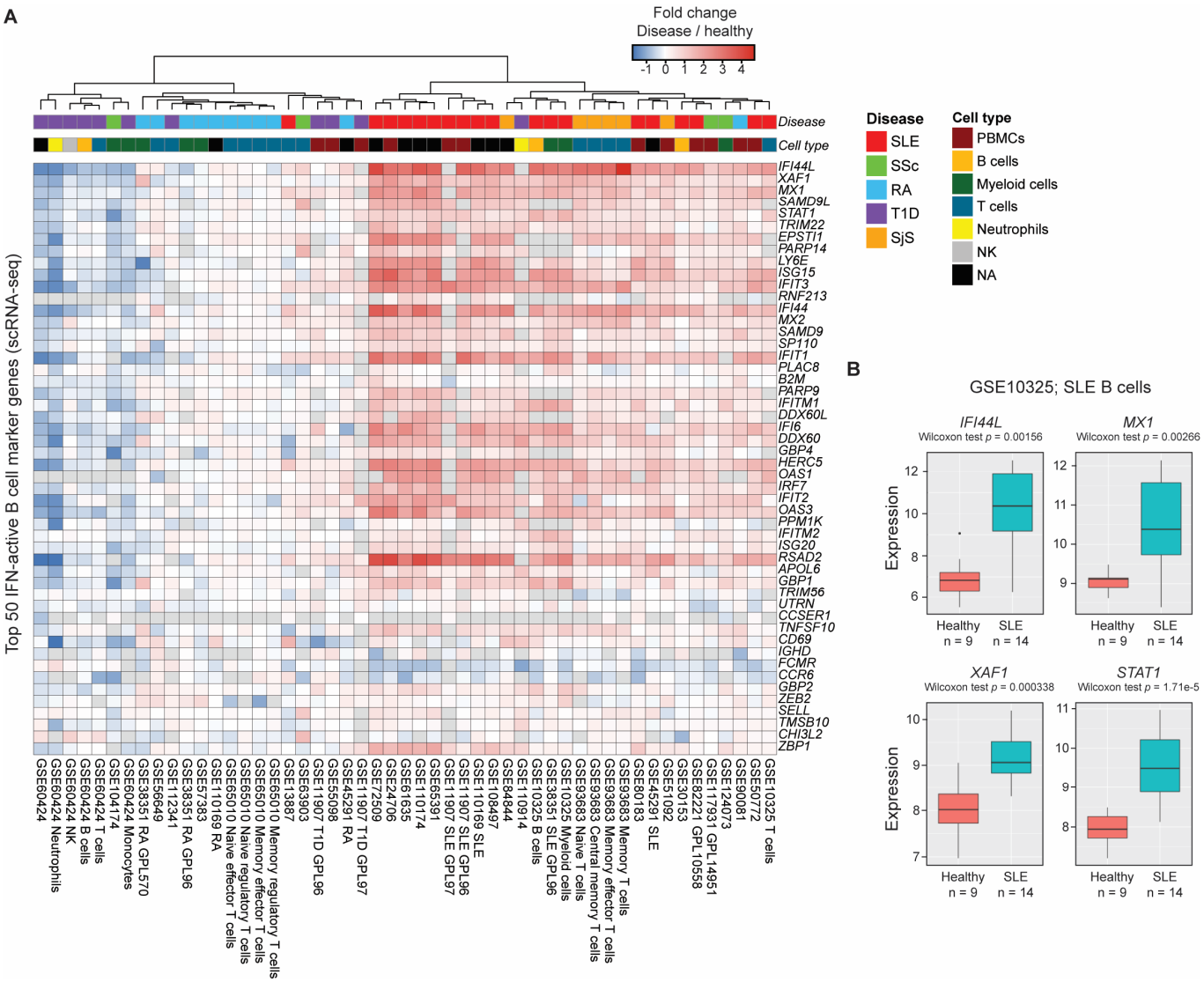


Figure S5. Differential expression in autoimmune disease of top scRNA-seq marker genes for the IFN_active B cell cluster.

A) Differential expression of top 50 marker genes for IFN-active B cell cluster (as in Fig1F-G) between control and autoimmune patient gene expression studies, generated from ADEx: Autoimmune Diseases Explorer (https://adex.genyo.es/; (*11*)). SLE; systemic lupus erythematosus. SSc; systemic sclerosis. RA; rheumatoid arthritis. T1D; type I diabetes. SjS; Sjögren's syndrome.

B) Expression of *IFI44L*, *MX1*, *XAF1* and *STAT1* in peripheral blood-derived B cells healthy and systemic lupus erythematosus (SLE) patients (*12*).

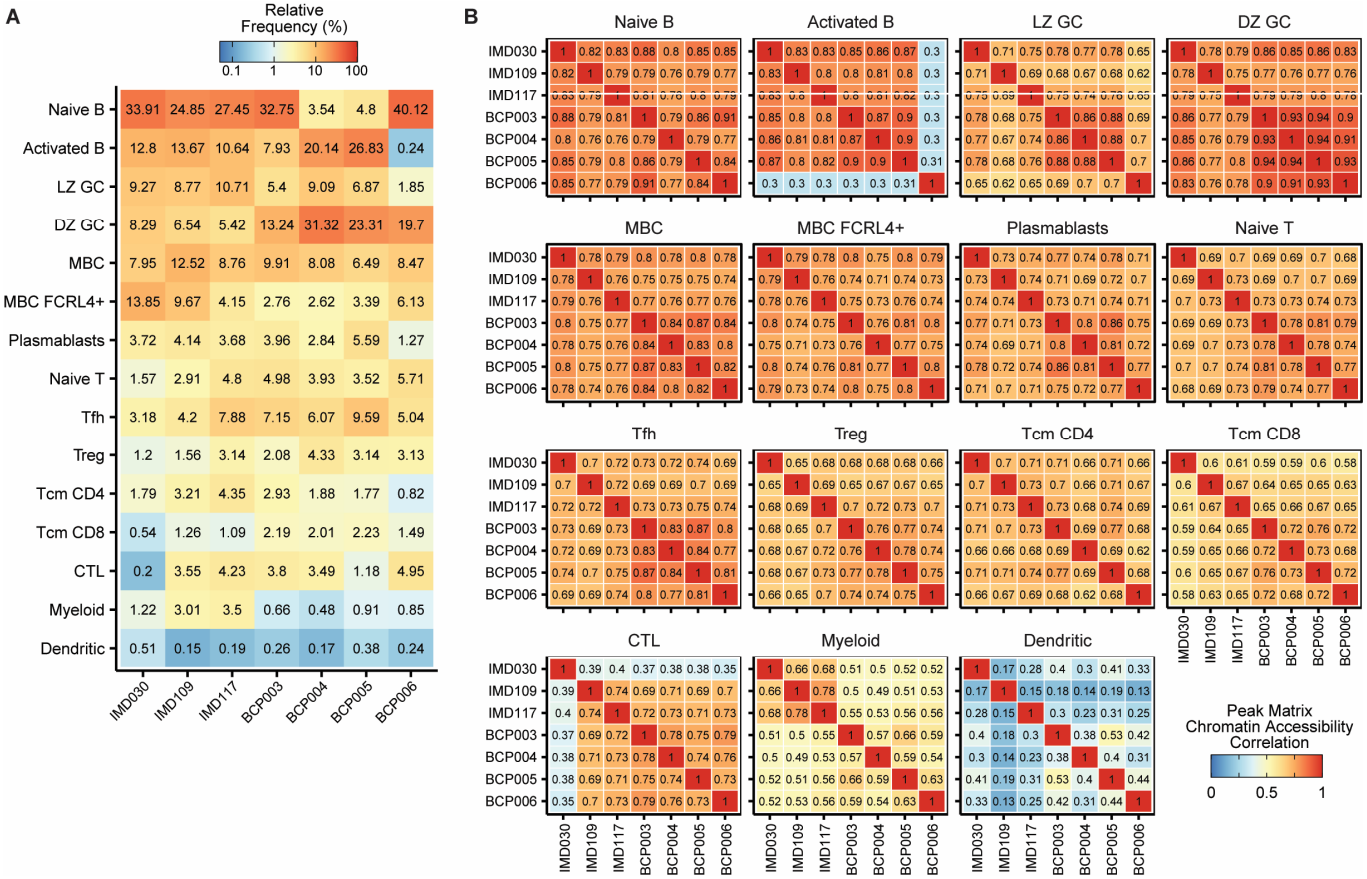


Figure S6. Reproducibility of scATAC-seq cluster frequencies and correlation of peak accessibilities between donors.

A) Relative frequency of high resolution scATAC-seq cell type clusters across patient donors.

B) Spearman correlation coefficients of mean peak accessibility per cell type cluster between patient donors.

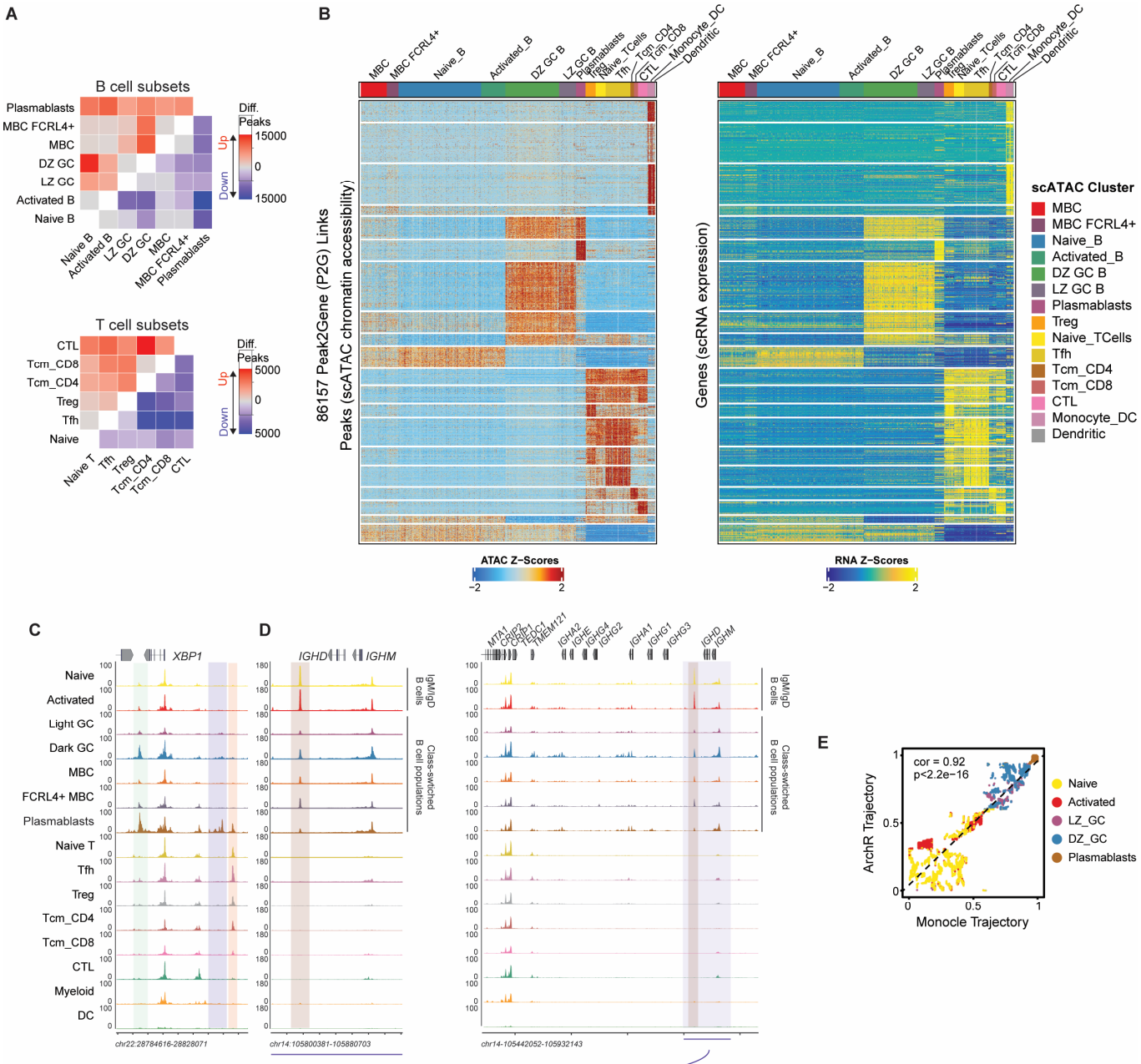


Figure S7. Differential peak analysis of scATAC-seq clusters, peak-to-gene predictions and alternative pseudotime analysis.

A) Differential peak analysis of B (top) and T (bottom) cell subsets comparing number of up-regulated and down-regulated chromatin accessibility regions.

B) Single-cell heatmaps of marker peak accessibility (scATAC) and gene expression (scRNA) for integrative peak2gene predictions in tonsil immune cell type clusters (as in Fig2A).

C) Example genome snapshot of *XBP1* regulatory landscape. B cell-, plasmablast- and T cell-specific regulatory elements are highlighted.

D) Genome snapshot of immunoglobulin heavy chain locus, including closer resolution of regulatory element downstream of *IGHD* and *IGHM* that is lost during class switch recombination (i.e. through deletional recombination).

E) Correlation between two independent pseudotemporal ordering methods (ArchR and Monocle) for naïve, activated, GC and plasmablast B cell lineage. Correlation coefficient and *p* value denotes result from Pearson correlation.

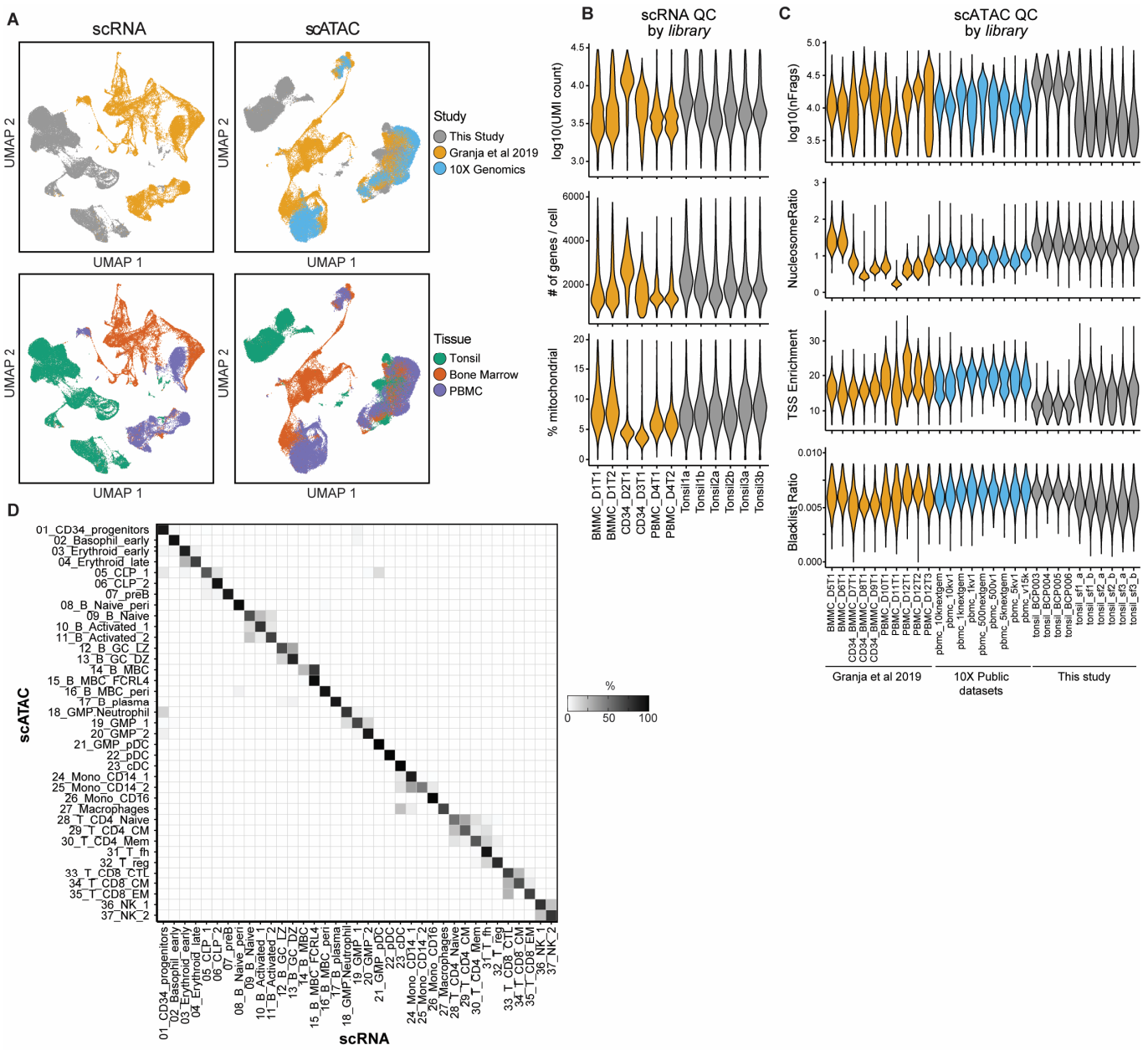


Figure S8. Batch correction and quality control of integrated bone marrow, blood and tonsil immune scRNA-seq and scATAC-seq.

A) UMAP visualization of batch and tissue for tonsil, peripheral blood and bone marrow scRNA-seq and scATAC-seq datasets.

B) Quality control metrics for scRNA-seq datasets by donor, including unique molecular identifier (UMI) counts per cell barcode, number of genes detected per cell barcode, and percentage of mitochondrial gene expression.

C) Quality control metrics for scATAC-seq datasets by donor, including number of unique fragments per cell barcode, ratio of nucleosomal to non-nucleosomal fragment sizes, transcription start site enrichment score, and ratio of fragments in genomic blacklist regions (see ArchR for details).

D) Confusion matrix depicting overlap between transferred scRNA-seq cluster identities to scATAC-seq clusters.

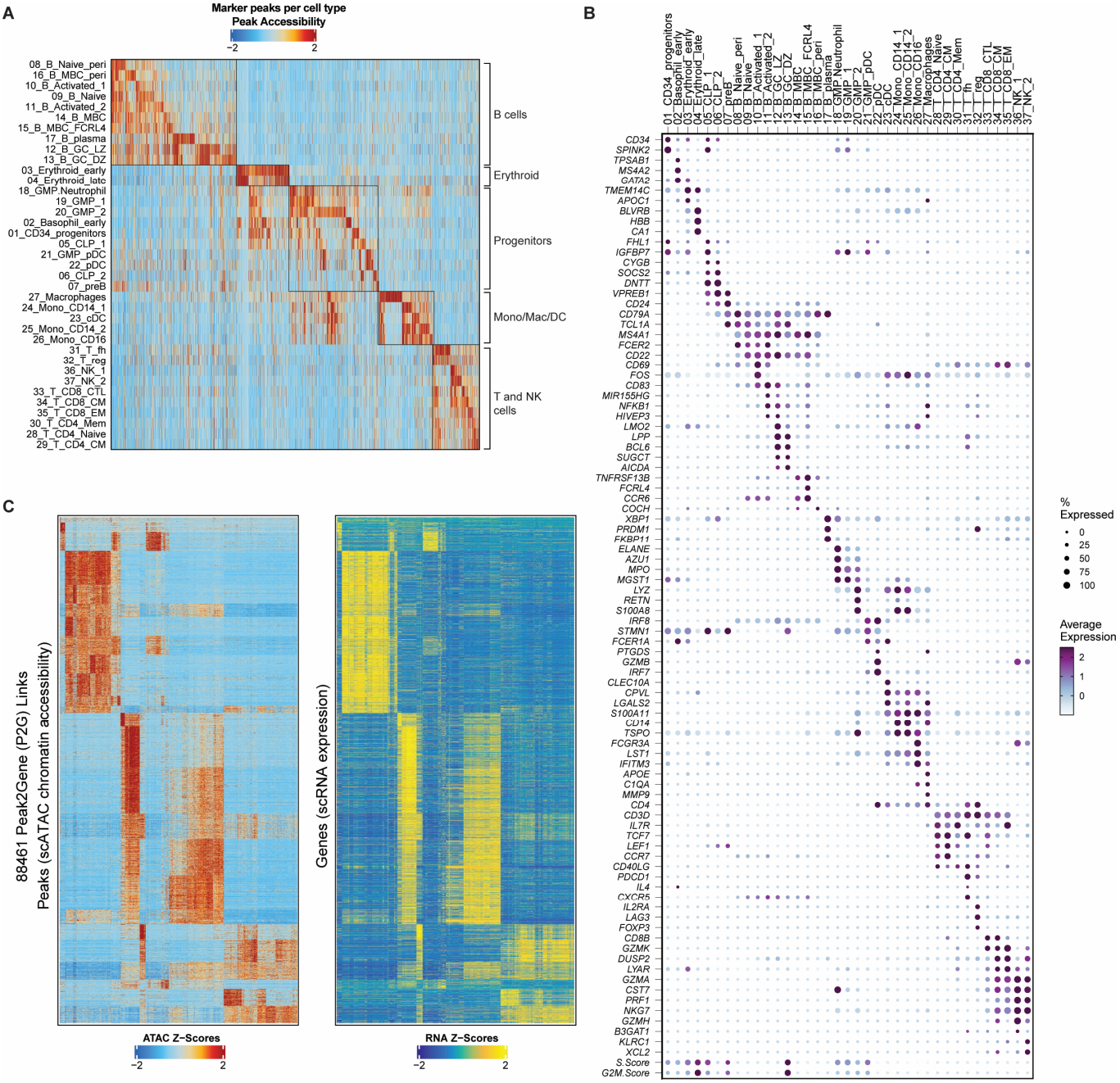


Figure S9. Integrated bone marrow, blood and tonsil scRNA-seq and scATAC-seq markers.

A) Expression of top marker genes for scRNA-seq clusters of integrated bone marrow, blood and tonsil dataset.

B) Chromatin accessibility at cluster-specific peaks for scATAC-seq clusters of integrated bone marrow, blood and tonsil dataset.

C) Peak accessibility (scATAC) and gene expression (scRNA) for integrative peak2gene predictions in tonsil, peripheral blood and bone marrow immune cell type clusters (as in Fig3A).

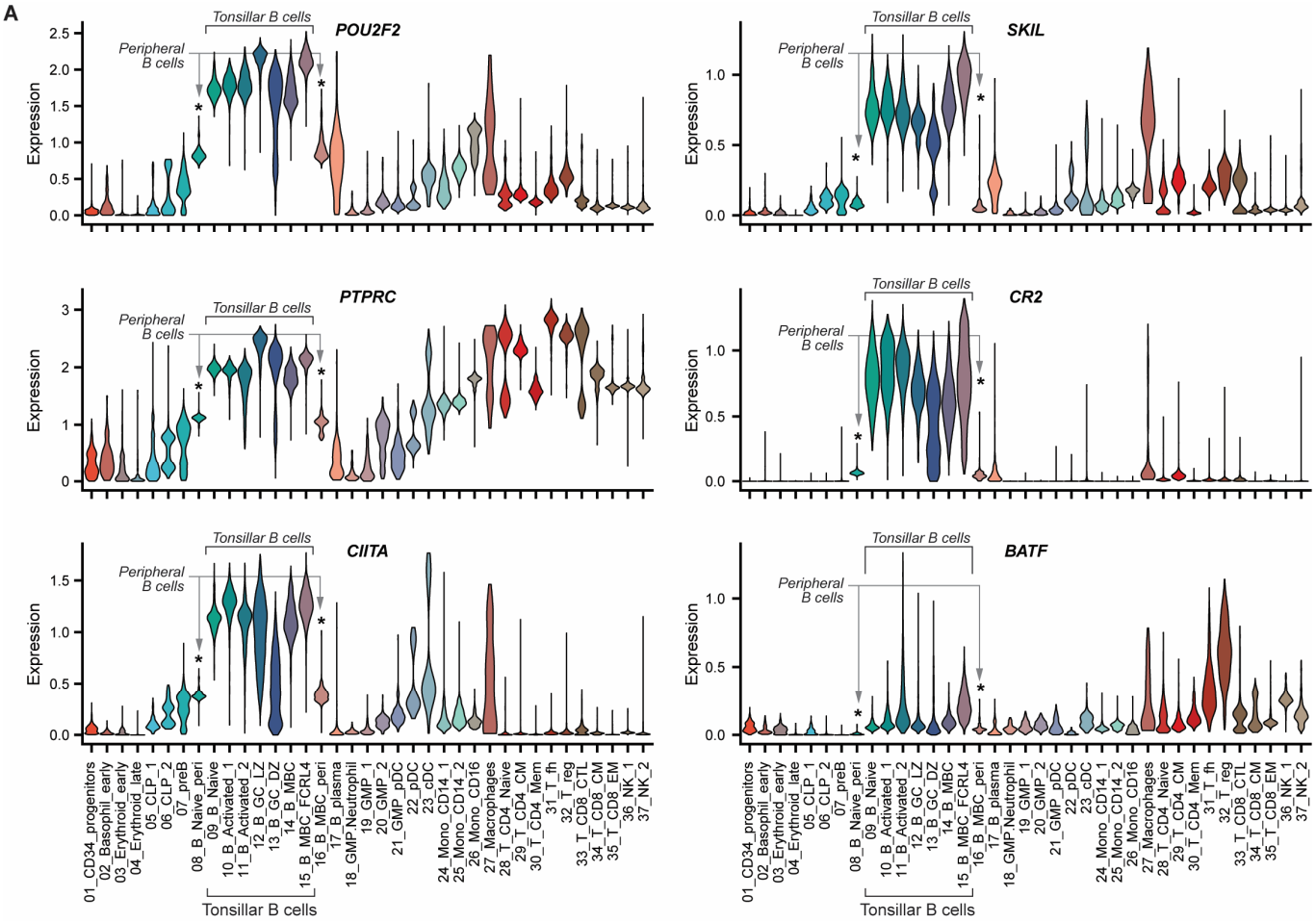


Figure S10. Tonsil B cell-enriched gene expression markers compared to peripheral blood B cells.

A) Expression of genes significantly differentially expressed between tonsil-specific naïve or memory B cell clusters compared to peripheral blood naïve or memory B cell clusters.

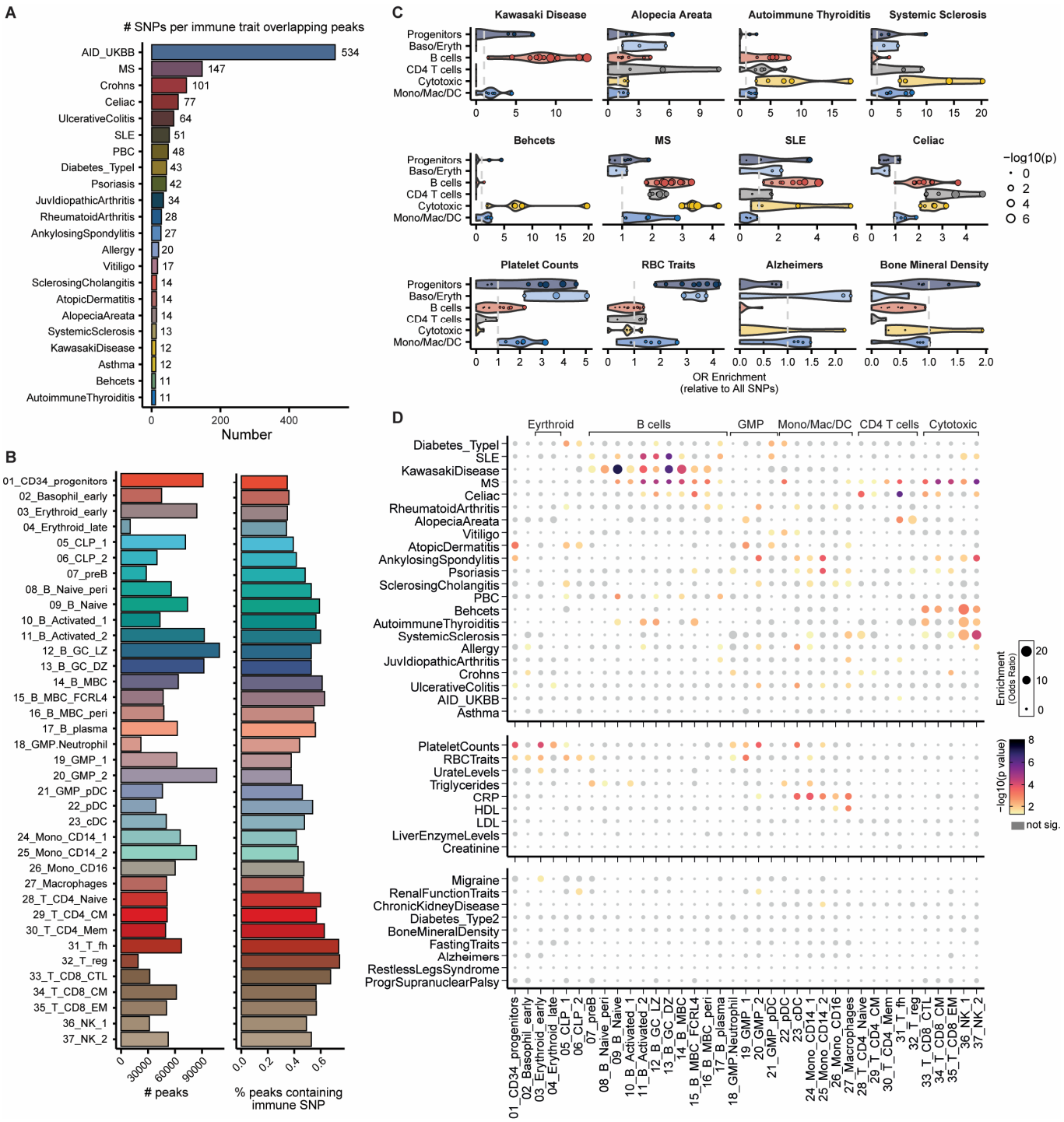


Figure S11. Enrichment of fine-mapped autoimmune variants in immune cell subsets.

A) Number of fine-mapped SNPs per autoimmune trait that overlap with a chromatin accessibility peak in the integrated human bone marrow, peripheral blood and tonsil scATAC-seq atlas. AID_UKBB represents variants identified from fine-mapping a combination of datasets from diverse autoimmune traits. AID; autoimmune disease. MS; multiple sclerosis. SLE; systemic lupus erythematosus. PBC; primary biliary cirrhosis.

B) Number of peaks identified in each scATAC-seq cell type cluster (left) and the percentage of those peaks that overlap with an autoimmune-associated SNP.

C) Fisher enrichment test results for variants specific to selected traits in cell type-specific chromatin. Individual points represent single cell type clusters, separated into five broad lineages. Dot size reflects level of significance of enrichment.

D) Fisher enrichment test results for trait-specific variants in cell type-specific chromatin across all traits in fine-mapped resources analyzed. Dot size conveys enrichment (odds ratio) and color denotes significance of enrichment.

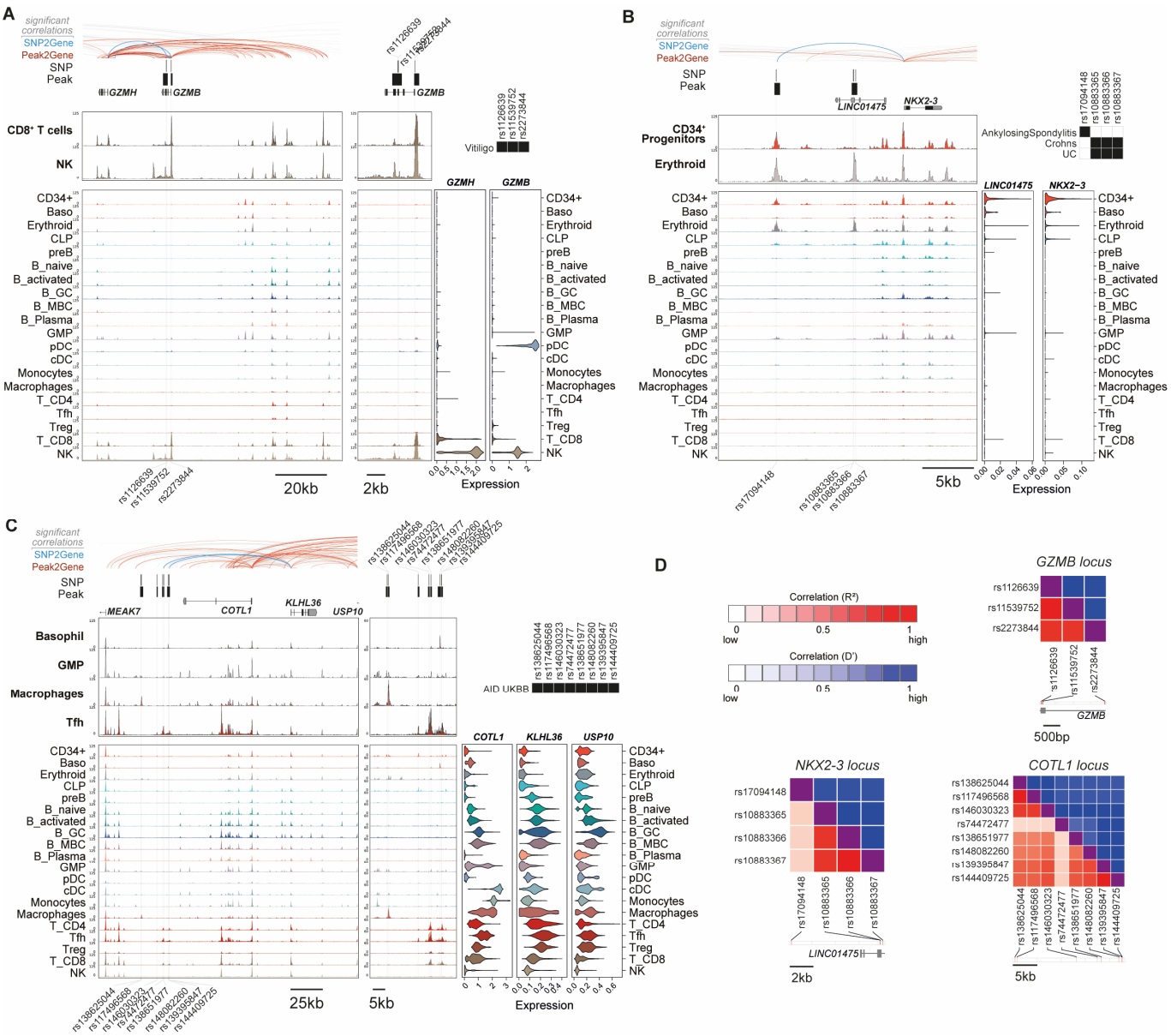


Figure S12. Genome snapshots of fine-mapped autoimmune variants at *GZMB*/*GZMH*, *NKX2-3* and *COTL1*/*KHLH36* loci.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants at the *GZMB*/*GZMH* locus. Significantly correlated peak2gene linkages colored in red and significant links between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in scATAC tracks.

B) Same as A), at the *NKX2-3* locus. UC; ulcerative colitis.

C) Same as A), at the *COTL1*/*KHLH36* locus. AID; autoimmune disease.

D) Linkage disequilibrium heatmaps for SNPs at loci depicted in A-C. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation.

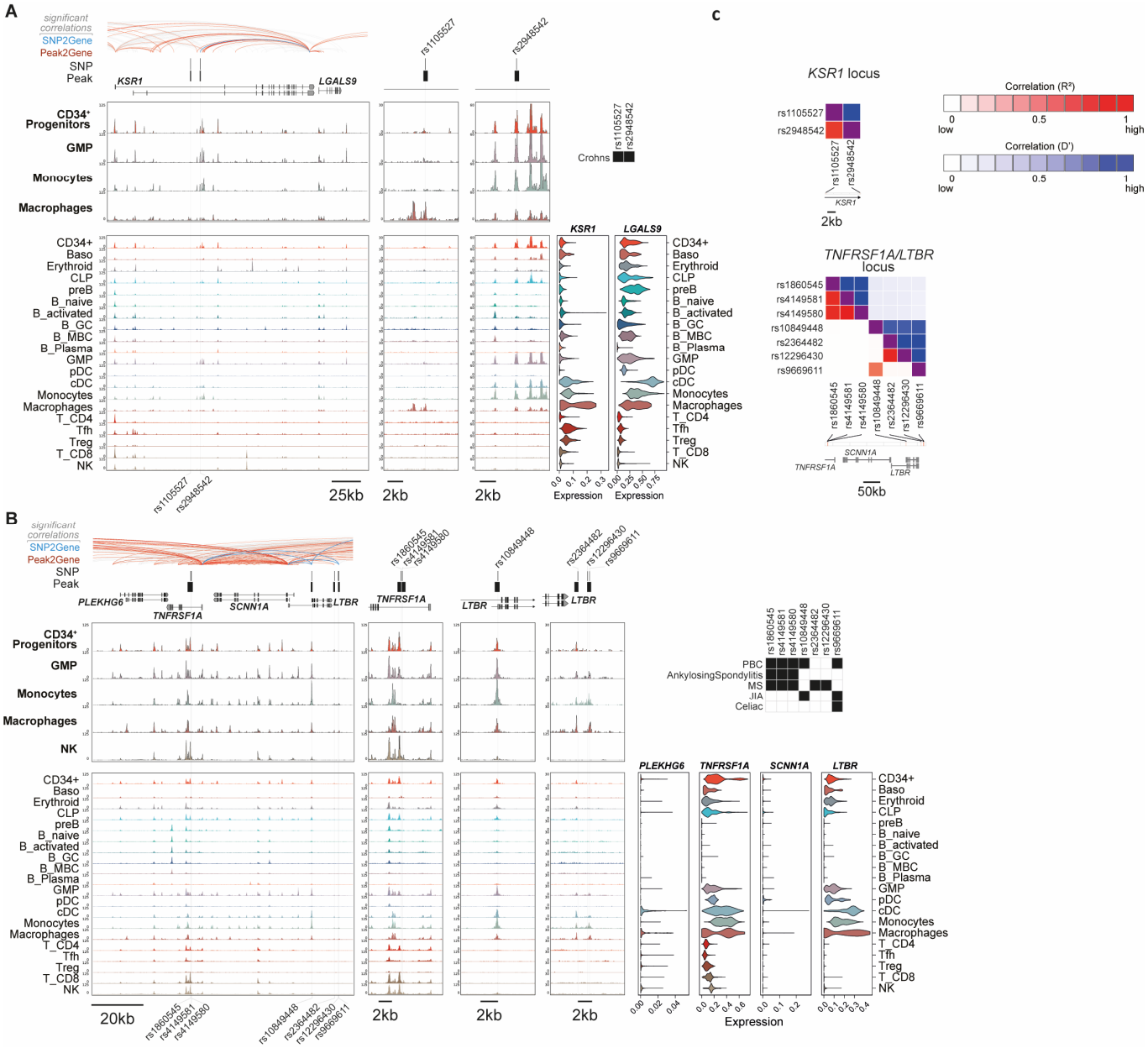


Figure S13. Genome snapshots of fine-mapped autoimmune variants at *KSR1*/*LGALS9* and *TNFRSF1A*/*LTBR* loci.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants at the *KSR1*/*LGALS9* locus. Significantly correlated peak2gene linkages colored in red and significant links between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in scATAC tracks.

B) Same as A), at the *TNFRSF1A/LTBR* locus. PBC; primary biliary cirrhosis. MS; multiple sclerosis. JIA; juvenile idiopathic arthritis.

C) Linkage disequilibrium heatmaps for SNPs at loci depicted in A and B. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation.

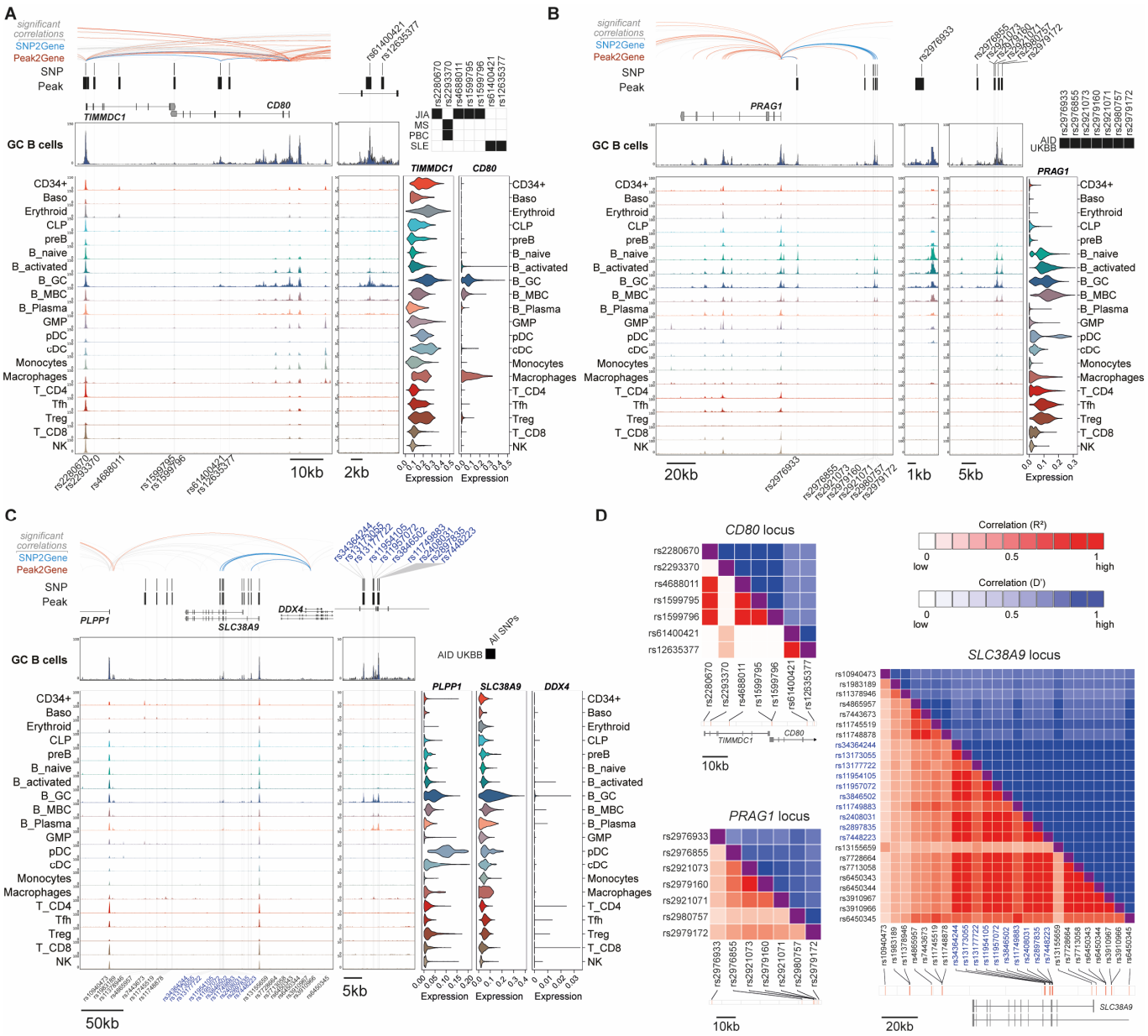


Figure S14. Genome snapshots of germinal center-associated cell type-specific regulatory activity at fine-mapped autoimmune variants at *CD80*, *PRAG1* and *SLC38A9*/*DDX4* loci.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants at the *CD80* locus. Significantly correlated peak2gene linkages colored in red and significant links between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in scATAC tracks. PBC; primary biliary cirrhosis. MS; multiple sclerosis. JIA; juvenile idiopathic arthritis. SLE; systemic lupus erythematosus.

B) Same as A), at the *PRAG1* locus. AID; autoimmune disease.

C) Same as A), at the *SLC38A9*/*DDX4* locus.

D) Linkage disequilibrium heatmaps for SNPs at loci depicted in A and B. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation.

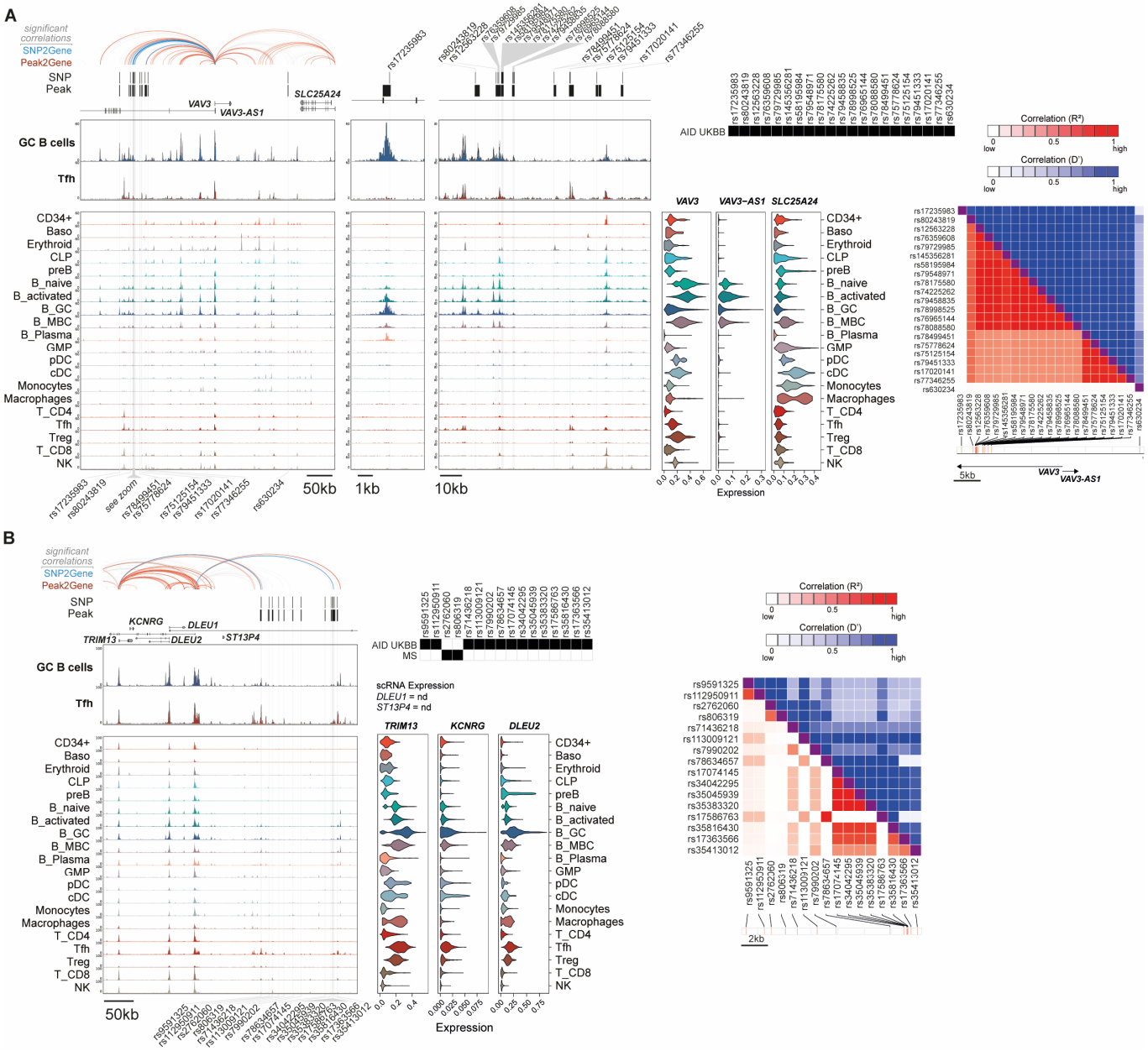


Figure S15. Genome snapshots of germinal center-associated cell type-specific regulatory activity at fine-mapped autoimmune variants at *VAV3* and *DLEU2* loci.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants at the *VAV3* locus. Significantly correlated peak2gene linkages colored in red and significant links between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in scATAC tracks. Linkage disequilibrium heatmaps are also shown separately. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation. AID; autoimmune disease.

B) Same as A), at the *DLEU2* locus. MS; multiple sclerosis.

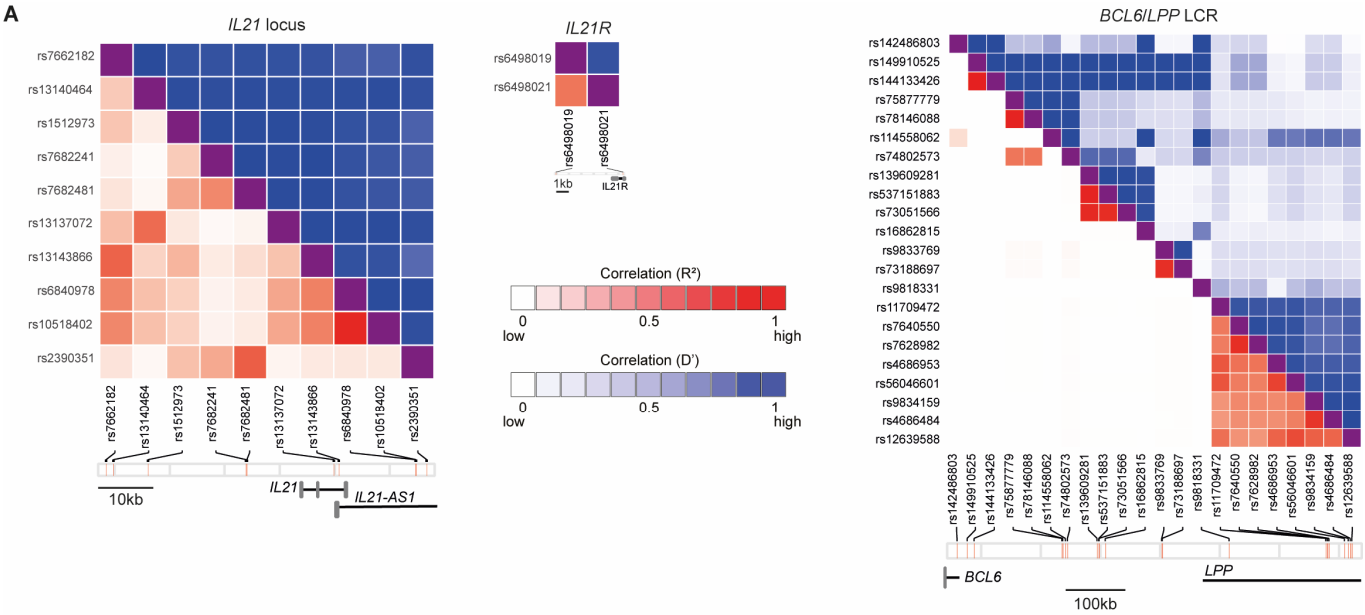


Figure S16. Linkage disequilibrium scores for variants at *IL21*, *IL21R* and *BCL6* loci.

A) Linkage disequilibrium heatmaps for SNPs at *IL21*, *IL21R* and *BCL6*/*LPP* loci depicted in Fig5. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation.

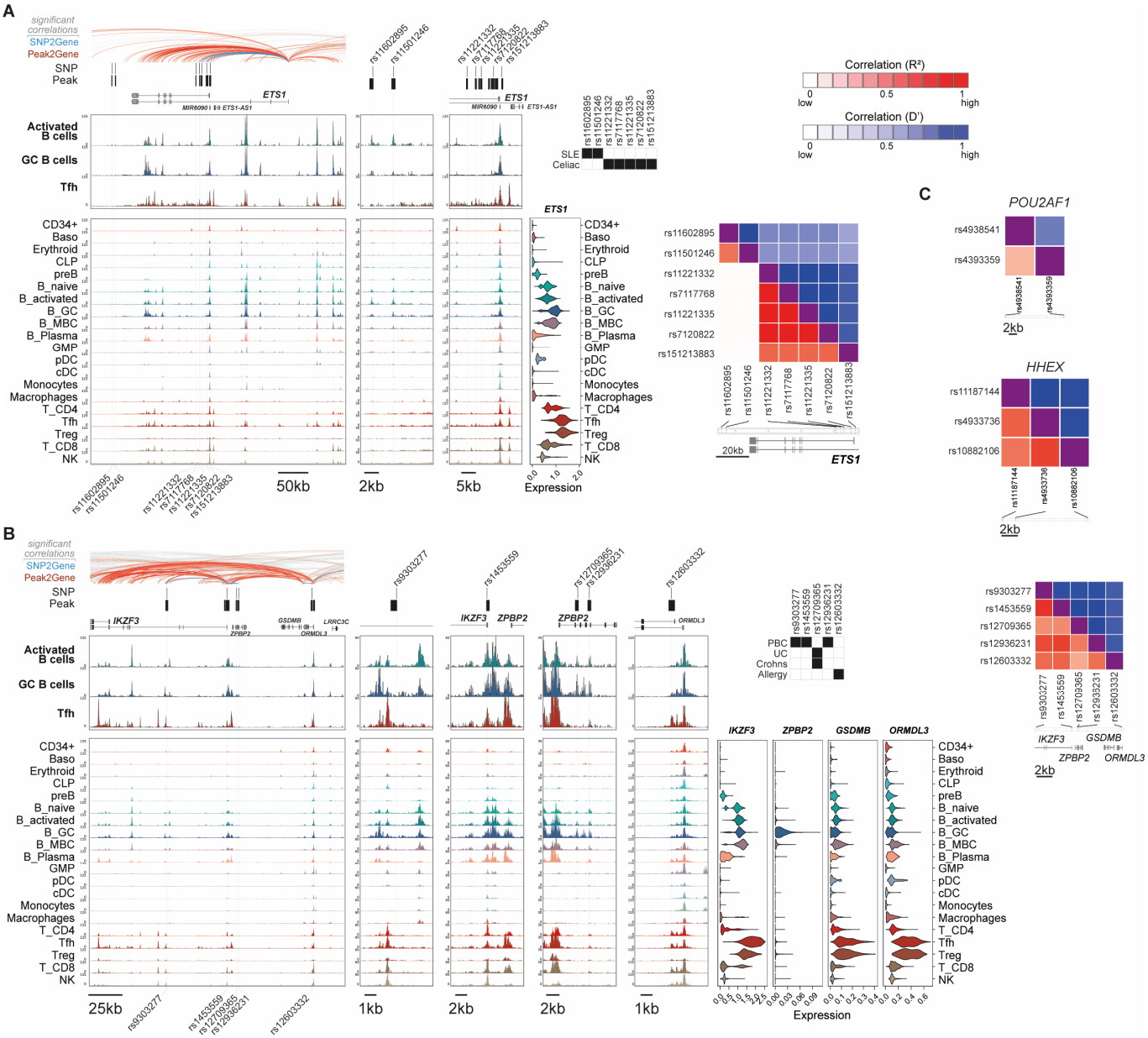


Figure S17. Genome snapshots of fine-mapped autoimmune variants at *ETS1* and *IKZF3* loci.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants at the *ETS1* locus. Significantly correlated peak2gene linkages colored in red and significant links between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in scATAC tracks. Linkage disequilibrium heatmap is also shown. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation. SLE; systemic lupus erythematosus.

B) Same as A), at the *IKZF3* locus. PBC; primary biliary cirrhosis. UC; ulcerative colitis.

C) Linkage disequilibrium heatmaps for SNPs at *POU2AF1* and *HHEX* loci depicted in Fig6. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation.

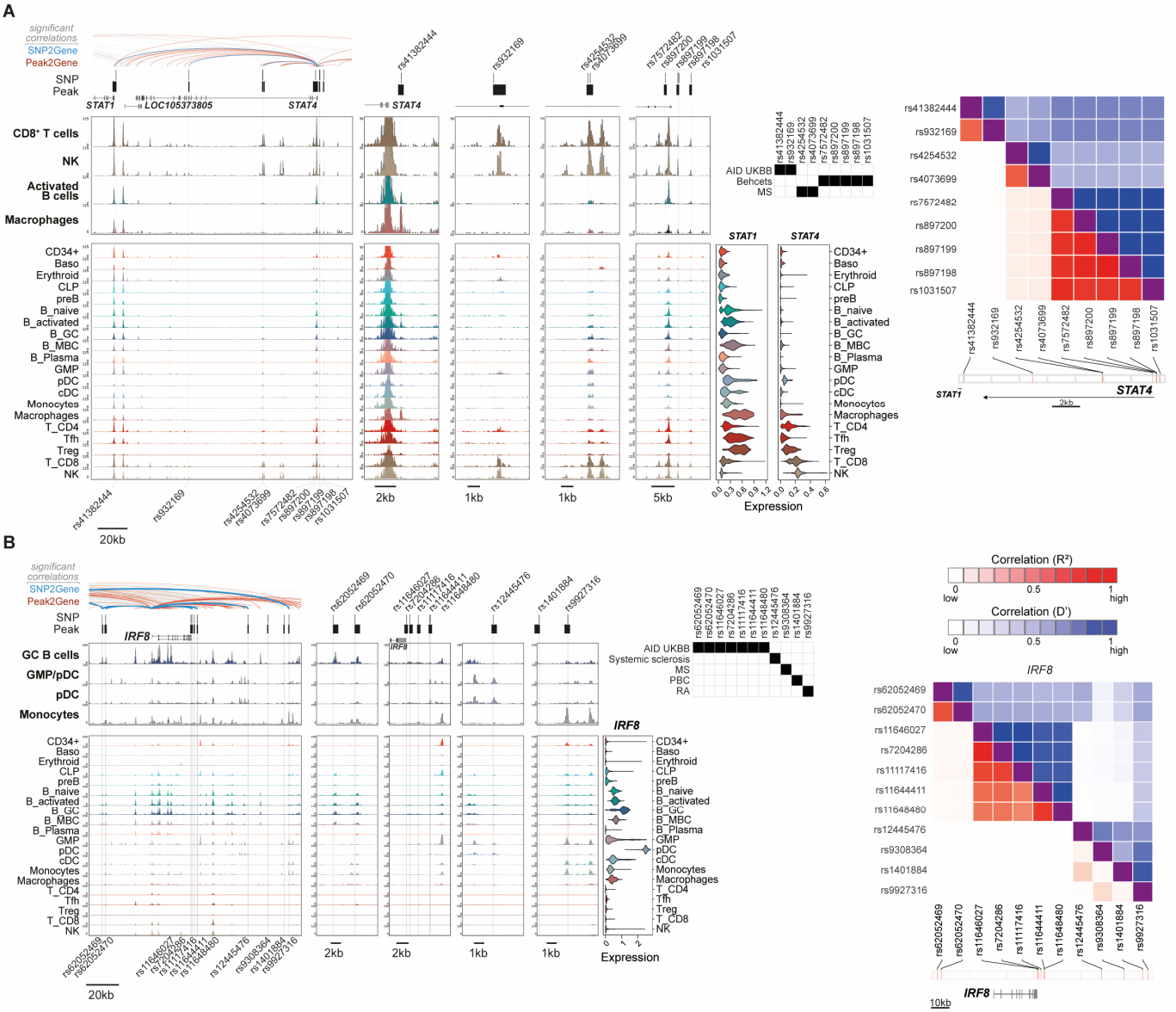


Figure S18. Genome snapshots of fine-mapped autoimmune variants at *STAT4* and *IRF8* loci.

A) Genomic snapshot of fine-mapped autoimmune-associated GWAS variants at the *STAT4* locus. Significantly correlated peak2gene linkages colored in red and significant links between SNPs and gene promoters (SNP2gene) in blue and bold. Significant associations between individual SNPs and autoimmune diseases are shown in black boxes and gene expression is shown as violin plots for matched populations in scATAC tracks. Linkage disequilibrium heatmap is also shown. D′ denotes normalized linkage disequilibrium; R^2 denotes Pearson coefficient of correlation. AID; autoimmune disease. MS; multiple sclerosis.

B) Same as A), at the *IRF8* locus. PBC; primary biliary cirrhosis. RA; rheumatoid arthritis.

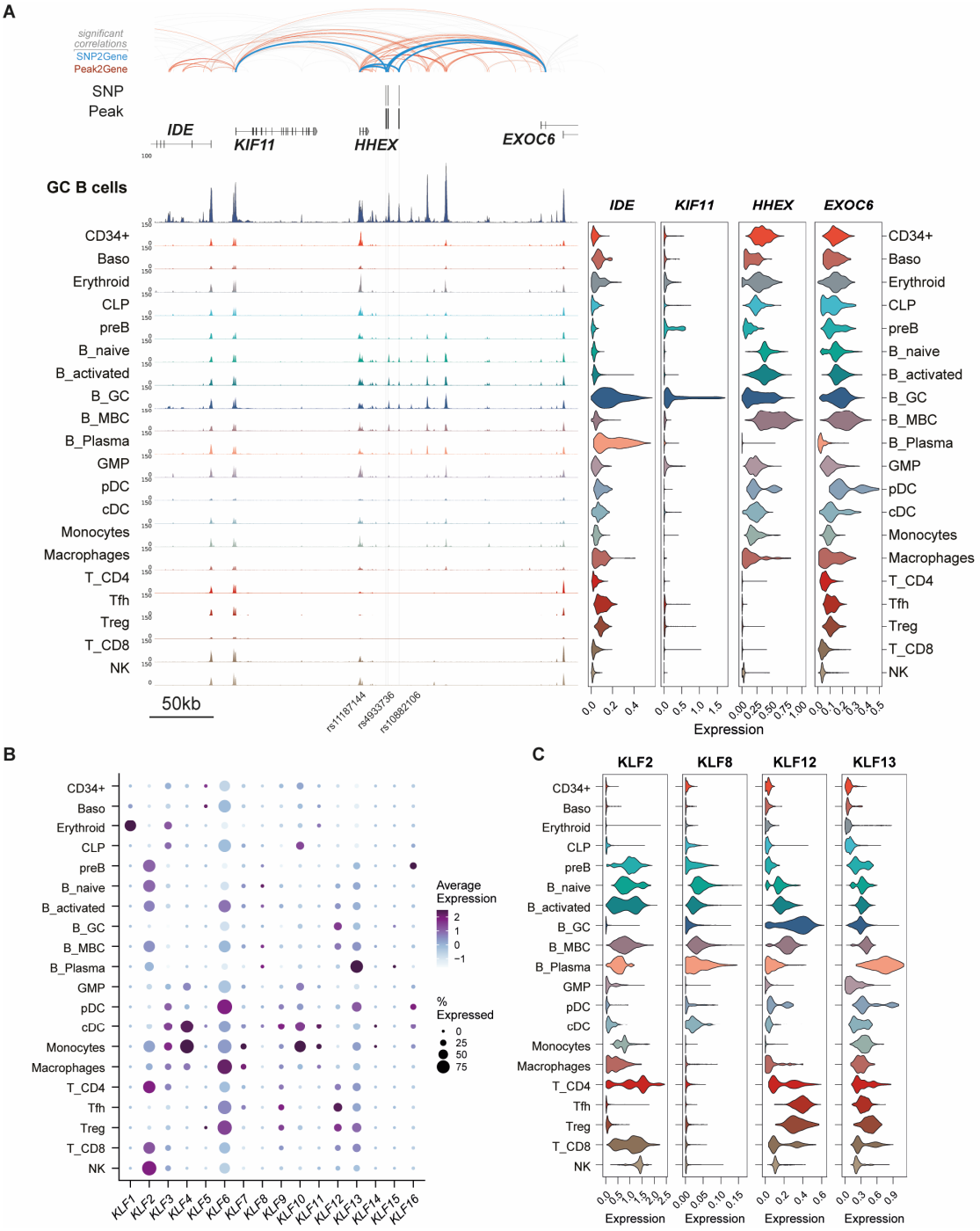


Figure S19. Genomic landscape at *HHEX* and expression of KLF family transcription factors.

A) Broader view of the *HHEX* locus (see Fig6b), showing significantly correlated peak2gene links and gene expression of neighboring genes *KIF11* and *EXOC6*.

B) Mean expression of all KLF family transcription factors detected in scRNA-seq dataset. Dot size denotes percent of cluster in which gene is detected.

C) Single-cell expression of *KLF2*, *KLF8, KLF12* and *KLF13*, with highest expression in B cell subsets.

Table S1. CyTOF phenotyping antibody panel.

metal	marker	clone name	clone source
Y	CD45(1)	HI30	DVS
Pd	CD45(2)	HI30	Biolegend
Pd	CD45(3)	HI30	Biolegend
Pd	CD45(4)	HI30	Biolegend
In	CD45(5)	HI30	Biolegend
In	CD57	HCD57	Biolegend
La	CD16	3G8	Biolegend
Pr			
Nd	CD45RO	UCHL1	Biolegend
Nd			
Nd			
Nd	CD33	WM53	Biolegend
Nd	CD14	M5E2	Biolegend
Nd			
Sm	CD20	2H7	Biolegend
Nd			
Sm	CD25	2A3	Fluidigm
Nd	**CD138**	**DL-101**	**Fluidigm**
Eu			
Sm	CD103	B-Ly7	eBioscience
Eu			
Sm	CD27	LG.7F9	eBioscience
Gd	CD56	NCAM16.2	BD
Gd	CD3	UCHT1	Biolegend
Gd	CD19	HIB19	Biolegend
Gd	CD45RA	HI100	Biolegend
Tb	CXCR5	RF8B2	BD
Gd	CD28	CD28.2	Biolegend
Dy	CD38	HIT2	Biolegend
Dy	**CD8**	**RPA-T8**	**Fluidigm**
Dy			
Dy			
Ho	**CD40**	**5C3**	**Fluidigm**
Er			
Er	int b7	FIB504	Biolegend
Er	CCR7	G043H7	Biolegend
Tm	IgD	IA6-2	Biolegend
Er	IgA	G18-1	BD
Yb	HLA-DR	L243	Biolegend
Yb	**IgM**	**MHM-88**	**Fluidigm**
Yb	TCRgd	5A6.E9	in-house
Yb	**CD4**	**SK3**	**Fluidigm**
Lu	PD-1	EH12.2H7	Fluidigm
Yb	CD127	A019D5	Biolegend
Ir	DNA1	-	Fluidigm
Ir	DNA2	-	Fluidigm
Pt	cisplatin l/d	-	Fluidigm

Table S2. CITE-seq antibody details.

Target	Antibody	Clone	Catalog#
CD3	TotalSeq™-A0034 anti-human CD3 Antibody	UCHT1	300475
CD4	TotalSeq™-A0072 anti-human CD4 Antibody	RPA-T4	300563
CD8a	TotalSeq™-A0080 anti-human CD8a Antibody	RPA-T8	301067
CD20	TotalSeq™-A0100 anti-human CD20 Antibody	2H7	302359
CD27	TotalSeq™-A0154 anti-human CD27 Antibody	O323	302847
CD38	TotalSeq™-A0389 anti-human CD38 Antibody	HIT2	303541
CD10	TotalSeq™-A0062 anti-human CD10 Antibody	HI10a	312231
CXCR4	TotalSeq™-A0366 anti-human CD184 (CXCR4) Antibody	12G5	306531
CXCR5	TotalSeq™-A0144 anti-human CD185 (CXCR5) Antibody	J252D4	356937
CD44	TotalSeq™-A0125 anti-human CD44 Antibody	BJ18	338825
IgD	TotalSeq™-A0384 anti-human IgD Antibody	IA6-2	348243
IgM	TotalSeq™-A0136 anti-human IgM Antibody	MHM-88	314541