Activated polyreactive B cells are clonally expanded in autoantibody positive and patients with recent-onset type 1 diabetes

Catherine A. Nicholas[1,2,3,6], Fatima A. Tensun[1], Spencer A. Evans[1], Kevin P. Toole[1], Jessica E. Prendergast[1], Hali Broncucia[1], Jay R. Hesselberth[2,3], Peter A. Gottlieb[1,4], Kristen L. Wells[1,4,6,*], Mia J. Smith[1,4,5,7,8,*]

[1]Barbara Davis Center for Diabetes, University of Colorado School of Medicine, Aurora, CO 80045, USA

[2]Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO 80045, USA

[3]Molecular Biology Graduate Program, University of Colorado School of Medicine, Aurora, CO 80045, USA

[4]Department of Pediatrics, University of Colorado School of Medicine, Aurora, CO 80045, USA

[5]Department of Immunology and Microbiology, University of Colorado School of Medicine, Aurora, CO 80045, USA

SUMMARY

Autoreactive B cells play an important but ill-defined role in autoimmune type 1 diabetes (T1D). We isolated pancreatic islet antigen-reactive B cells from the peripheral blood of non-diabetic autoantibody-negative first-degree relatives, autoantibody-positive, and recent-onset T1D donors. Single-cell RNA sequencing analysis revealed that islet antigen-reactive B cells from autoantibody-positive and T1D donors had altered gene expression in pathways associated with B cell signaling and inflammation. Additionally, BCR sequencing uncovered a similar shift in islet antigen-reactive B cell repertoires among autoantibody-positive and T1D donors where greater clonal expansion was also observed. Notably, a substantial fraction of islet antigen-reactive B cells in autoantibody-positive and T1D donors appeared to be polyreactive, which was corroborated

by analysis of recombinant monoclonal antibodies. These results expand our understanding of autoreactive B cell phenotypes during T1D and identify unique BCR repertoire changes that may serve as biomarkers for increased disease risk.

In brief

Nicholas et al. demonstrate that pancreatic islet antigen-reactive B cells from individuals with type 1 diabetes display a unique transcriptome and BCR repertoire compared with non-diabetic donors.

Graphical Abstract
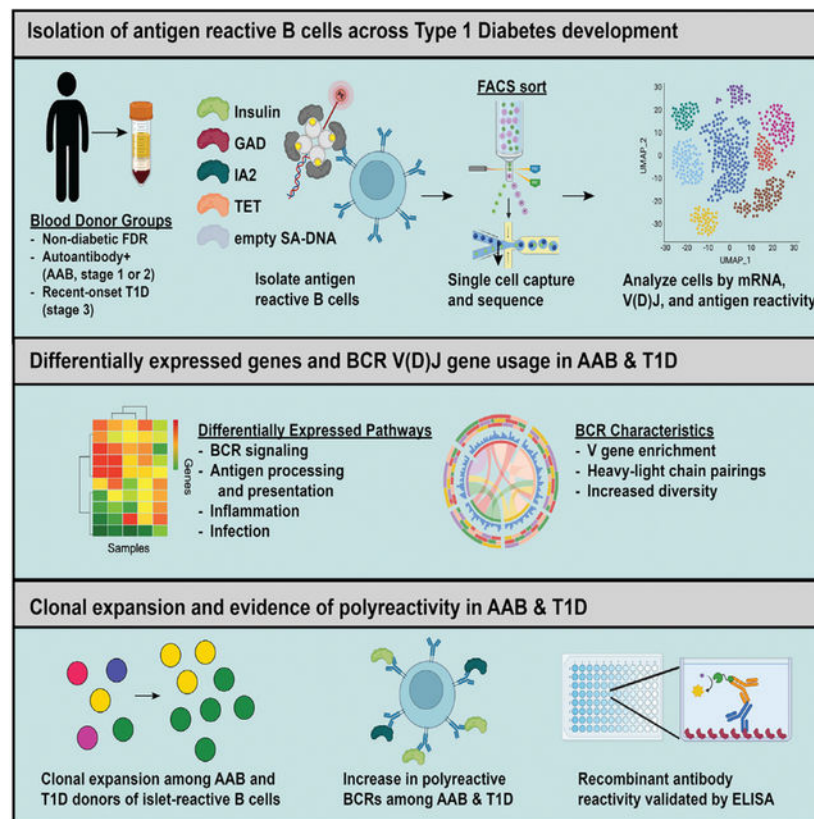
INTRODUCTION

Type 1 diabetes (T1D) develops as a consequence of a sustained autoimmune attack on the insulin-producing beta cells in the pancreas. T1D has historically been classified as a T cell-mediated disease due to the destruction of pancreatic islet beta cells by autoreactive T cells. However, previous experiments in the non-obese diabetic (NOD) mouse model have provided evidence for autoreactive B cell involvement with disease progression. This evidence includes demonstration of their essential role in antigen presentation to T cells, protection from diabetes progression in mice lacking B cells, and requirement for islet, i.e., insulin, reactive B cells to develop autoimmune diabetes.[1–9] Given the importance of B cells in the NOD mouse model, a phase 2 clinical trial was undertaken using the B cell depleting monoclonal antibody, rituximab, to target CD20+ B cells in recently diagnosed individuals

with T1D. The trial showed that patients treated with rituximab have preserved beta cell function 1 year post-treatment.[10,11] These benefits were largely lost 2 years after treatment when the B cell compartment had fully recovered.[12] Despite evidence for B cell involvement in T1D, few human B cell-focused studies have been completed, particularly those analyzing islet antigen-reactive (IAR) B cells. We previously analyzed insulin-binding B cells in the peripheral blood of subjects along a continuum of diabetes development and showed that anergic (unresponsive) insulin-binding B cells are lost in individuals with pre-clinical diabetes (autoantibody positive but not symptomatic) and individuals recently diagnosed with T1D.[13,14] Follow-up studies in young-onset T1D revealed an increase in activated B cells within the anergic insulin-binding B cell subset, suggesting they have lost tolerance.[15] But exactly how these B cells become activated and their role in disease progression remains unknown.

Autoantibodies produced by B cells reactive with pancreatic islet antigens, e.g., insulin (INS), glutamic acid decarboxylase 65 (GAD), insulinoma associated antigen 2 (IA2), and zinc transporter 8 (ZnT8), are found in the serum of individuals prior to onset of T1D, and are used as biomarkers to identify individuals with a high likelihood of progression to T1D.[16,17] Accumulation of multiple autoantibodies in individuals with pre-clinical diabetes (prediabetes) is strongly correlated with faster progression to T1D diagnosis.[18] Despite this, current dogma based on mouse studies suggests that autoantibodies in T1D are not pathogenic.[7] Instead, the role of B cells in T1D is likely through (auto)antigen presentation to T cells.[3,19,20]

It has been shown that up to 70% of newly generated B cells in the bone marrow are self-reactive.[21] Normally these cells are purged through central tolerance mechanisms of receptor editing or clonal deletion or by the peripheral tolerance mechanism of anergy.[22–25] Individuals with autoimmunity, including T1D, exhibit an increase in autoreactive B cells that escape the bone marrow and enter the periphery. Importantly, these cells tend to be polyreactive, binding to two or more of the following antigens: INS, DNA, or lipopolysaccharide.[13,23,26] Together these findings indicate that normal tolerance mechanisms are impaired and unregulated, autoreactive B cells play a role in the development of T1D.

Given how little is known about diabetogenic B cells, including their role in the pathogenesis of T1D and how their tolerance is broken, we sought to analyze the phenotype and BCR repertoire of islet-reactive B cells from the peripheral blood of subjects along a continuum of diabetes development. We designed a multiplexed single-cell RNA sequencing (scRNA-seq) method based on LIBRA-seq,[27] a method developed to characterized antigen-specific B cells, in order to study B cells reactive to three pancreatic islet antigens, INS, IA2, and GAD, as well as those reactive with the foreign antigen tetanus-toxoid (TET). While our lab has extensively studied the surface phenotype and functional properties of INS-reactive B cells using flow and mass cytometry, to our knowledge no such study has utilized scRNA-seq to study multiple IAR B cells in parallel. Our results demonstrate significant differences in the transcriptome and BCR repertoire of patients along a continuum of diabetes development compared with non-diabetic controls. These findings greatly expand our current knowledge regarding the phenotype of IAR B cells in T1D and suggest

some potential mechanisms by which self-reactive B cells may lose tolerance and become activated in T1D.

RESULTS

B cell subsets span all stages of B cell differentiation and are present in each donor group

To more deeply phenotype IAR B cells during T1D development and obtain information regarding their BCR repertoire, we performed single-cell RNA sequencing of B cells from first-degree relative non-diabetic (ND), autoantibody-positive prediabetic (AAB), and recently diagnosed (<100 days) T1D donors. We additionally employed the LIBRA-seq (linking B cell receptor to antigen specificity through sequencing)[27] method to further analyze B cells reactive to antigens of interest including INS, IA2, and GAD. Peripheral blood was collected from five ND (Age: 14.2 ± 8.9 years [mean $\pm$ SD]), six AAB (Age: 18.2 ± 11.8 years), and five T1D (Age: 15.8 ± 8.2 years) individuals and cryopreserved (Table S1). Autoantibody expression data were obtained from previous medical records and HLA testing, or HLA prediction based on mRNA expression was performed. All subjects in our study carried at least one high-risk T1D HLA allele, such as the DQ8 and/or DQ2 risk alleles.[28] PE-labeled antigen (INS, IA2, GAD, TET) tetramers, each carrying a unique barcode sequence, were freshly made on the day of each of the scRNA-seq experiments. In addition to these antigen tetramers, we included an "Empty" negative control tetramer that had its four binding sites occupied by biotin without protein antigen to detect any B cells that may be reactive to streptavidin, PE, biotin, or the DNA barcode. To minimize batch effects, four PBMC samples were thawed and processed in parallel, and eight samples were processed per day. Thawed PBMCs were incubated with the antigen and empty tetramers per LIBRA-seq,[27] and stained with various flow and CITE-seq antibodies. PE-reactive B cells were sorted via fluorescence-activated cell sorting (FACS). In addition, a small number of non-PE-binding B cells were also sorted and combined with the PE-binding B cells to increase the number of total cells recovered for each 10x Genomics capture (Figure 1A). FACS revealed a non-significant trend toward presence of more total antigen positive binding B cells as disease progressed through stages from ND (Mean $1.46\% \pm 0.25\%$) to AAB (Mean $1.63\% \pm 0.22\%$) to T1D (Mean $1.99\% \pm 0.36\%$) (Figure 1B).

To identify the B cell types present in our dataset irrespective of specificity, we completed standard pre-processing using the Seurat workflow[29] followed by sample integration and batch correction using the MNN batch correction method (Figures S1A and S1B).[30] Following pre-processing, we identified nine unique B cell clusters in various stages of differentiation including naive, early memory, age/autoimmune-associated B cells (ABC),[31] classical memory, and plasmablasts (Figures 1C and S2). The plasmablast cell type cluster may also contain plasma cells. Figure 1D confirms these cell subtypes by visualizing the expression of a selection of established genes for each cell type.[32–35] To determine whether there were differences in representation of B cell subsets between the donor groups, we plotted the frequency of each B cell subset as a percentage of total B cells that were found in ND, AAB, and T1D subjects (Figure 1E). While all B cell subsets could be found in each participant group, there was a non-significant increase in the frequency of ABC and plasmablasts in the ND group. Upon further analysis of the five ND individuals, the increase

in frequency of ABC was found only in subjects 107 and 108, who, interestingly, are the ND half-sister (107) and ND fraternal twin (108) of T1D subject 109. The increase in frequency of plasmablasts was solely attributable to ND subject 110, who had a large expansion of plasmablasts among their sorted B cells. While all subjects reported that they had not been sick in the 2 weeks prior to their blood draw, it is possible this individual was in the early stages of an illness or had been recently vaccinated, the latter of which was not an exclusion factor for entry into our study. Thus, these five major B cell subsets were identified in our dataset and were present across ND, AAB, and T1D donor groups.

IAR B cells are higher in frequency in AAB and T1D groups, occur in all five major B cell subsets, and tend to bind to more than one antigen

To determine the degree of IAR reactivity among the samples, we next defined a threshold of positive reactivity for each of the antigens. We calculated a binding score for each antigen using a quantile approach based on antigen binding to non-B cells, which would reflect non-specific background binding (Figure S3). A binding score greater than one indicates a positive binding event and was calculated for each individual antigen. As expected, the scRNA-seq readout revealed an increase in IAR B cells as a percent of all B cells in AAB and T1D donors, though this was not significant after adjusting for multiple testing (Figure 2A). Interestingly, while the proportion of INS-reactive B cells appeared stable across donor groups, we captured a greater percentage of GAD and IA2-reactive B cells from AAB and T1D compared with ND individuals. Importantly, reactivity to the empty tetramer was consistently low across all samples, indicating a low frequency of B cells reactive to any of the streptavidin, biotin, PE, or DNA components of the reagent. However, we were surprised to find that the largest share of IAR B cells was composed of B cells that appeared to bind one pancreatic islet antigen and at least one additional antigen in our assay, which we termed "Islet_Multi" reactive B cells. When we analyzed the antigen reactivities found among the Islet-Multi reactive B cells, we found B cells that bound to every possible combination of the five antigen tetramers, such as some that bound both INS and GAD and others that bound all five antigen tetramers (Figure S4). When we quantified the percent of Islet_Multi reactive B cells that bound to each antigen tetramer, we found that Islet_Multi reactive B cells bound to the islet antigens at an increased frequency compared with the empty and TET controls (Figure 2B).

To determine whether any of the B cell subsets (Figure 1) clustered together because of shared antigen reactivity, we plotted the antigen-reactive B cell groups onto the B cell subset UMAP. As seen in Figure 2C, antigen-reactive B cells could be found in each of the B cell subsets, indicating B cell subsets were not clustered based on their antigen reactivity. When we quantified the frequency of antigen-reactive B cells within the five major B cell subsets in each subject group, we found that the increased proportion of IAR B cells in AAB and T1D donors is most pronounced in the more antigen experienced ABC, memory, and plasmablast B cell subsets (Figure 2D). Given that IAR B cells were found at decreased but notable levels in the ND group compared with AAB and T1D donors, we next determined whether there was a difference in the level of antigen-tetramer binding in ND compared with AAB and T1D donors. Most strikingly, when the antigen binding score was summarized across donor groups for each antigen, the AAB and T1D donors

had higher average binding scores than ND donors (Figure 2E). This indicates that the amount of antigen-tetramer binding was higher in the B cells of those donors as compared with non-B cells, possibly reflecting increased affinity for antigen. The observed increased binding events for INS-reactive B cells was not due to differential insulin receptor (INSR) expression between donors (Figure S5). Overall, these results demonstrate that IAR B cells from AAB and T1D donors are increased in frequency, particularly in more mature B cell subsets, compared with ND donors, and tend to be polyreactive.

IAR B cells display distinct mRNA expression patterns across T1D development

To further characterize the differences between B cells between status groups, we assessed mRNA gene expression patterns of IAR B cells. Differential gene expression between AAB and ND donors and T1D and ND donors uncovered expression changes as stages of disease progress (Figure 3A). Gene set enrichment analysis identified enrichment of genes in many cell-signaling pathways in AAB and T1D donors when compared with ND. We clustered a selection of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways significantly altered in our data and grouped them into six broad signaling themes: Proinflammatory, BCR and General Signaling, Antigen Processing and Presentation, Bacterial Infection, Viral Infection, and Autoimmunity (Figure 3B). As highlighted in Figure 3C, BCR signaling pathway gene expression was significantly increased. These included increases of *BLNK, PTPRC (CD45), SKAP2, CD79B, BANK1*, and *SYK* in AAB and T1D donors—all promoting B cell activation upon antigen binding to the BCR.[36] Downregulated genes that can contribute to inhibition of BCR signaling included *LYN* and the phosphatases *DUSP1, DUSP2,* and *DUSP5*.[37,38] Additionally, IAR B cells in autoimmune donors showed increased expression of markers related to antigen presentation including *CD40, CIITA*, and numerous *HLA* genes, supporting their role as active participants in an autoreactive immune response beyond autoantibody production. KEGG pathways for proinflammatory, bacterial, and viral infection response contained some genes that were common among these pathways, suggesting IAR B cells from AAB and T1D donors exhibit an overall proinflammatory phenotype compared with ND donors.

Based on the differential expression analysis, donors 107, 108, and 109 appeared to have exceptionally similar gene expression patterns irrespective of disease status. These donors are siblings, with donors 108 and 109 being fraternal twins. In returning to the pre-batch corrected and normalized data, we found that those three individuals were part of an obvious batch effect when compared with all other donors (Figure S6A). The twins clustered together away from all other donor's cells with their half-sister, donor 107, clustering between the twins and other donors. Differential gene expression analysis between the twin sisters, among the three sisters, and among all donors confirmed high overlap of genes between the sisters and all donors with few genes unique to the sister comparisons (Figure S6B). These findings reveal that although these siblings cluster separately from the other donors, the differential gene expression patterns observed across all donors are not affected by unique gene expression due to a batch effect among these siblings.

AAB and T1D B cells have greater BCR repertoire diversity than ND B cells

Next, we wanted to delve into the BCR repertoires of IAR B cells between donor groups. Interrogation of IAR BCR isotypes across cell types and antigen reactivities revealed that most cells in our study expressed IgM based on BCR sequencing (Figure 4A). Since we sorted all PE-binding B cells, irrespective of their differentiation state (i.e., naive vs. memory), these results are consistent with previous reports that ~60%–70% of peripheral blood B cells are naive and co-express immunoglobulin (Ig) M and IgD.[39] It has been demonstrated previously that the CDR3 regions of autoreactive BCRs tend to be longer and contain more positively charged amino acids.[40,41] We found a modest increase in CDR3 amino acid length and positive charges in the IAR B cells (Figures 4B and 4C). There was no difference in level of somatic hypermutation (SHM) in the heavy or light chains of IAR B cells among donors, likely due to the high level of IgM isotype of B cells in the samples (Figures 4D and 4E). We next calculated a Shannon diversity score to measure the diversity of species in a community, where a larger score indicates greater diversity. Interestingly, we discovered significantly higher diversity scores in the autoimmune AAB and T1D donors compared with ND individuals across cell types and donor groups in total B cells (Figure 4F). This trend was largely maintained in more mature cell populations of ABC, memory, and plasmablasts. The increased diversity of BCRs among autoimmune donors may reflect defects in central and peripheral B cell tolerance in individuals with T1D,[23,26] allowing more autoreactive B cells to escape into the periphery.

AAB and T1D donors exhibit a unique BCR repertoire with significant enrichment of islet-reactive heavy and light chain gene pairs

To determine whether there is any T1D-associated BCR gene usage, we next analyzed the overall heavy and light chain V gene repertoire across ND, AAB, and T1D B cells, irrespective of their antigen specificity (Figures 5A and S6A). We found significant differences in the frequency of heavy and light V genes identified in T1D and AAB individuals compared with ND individuals. An odds ratio test revealed *IGHV4–4, IGHV1–3, IGHV3–64, IGHV3–66, IGHV4–61,* and *IGHV1–18* gene usage was significantly greater in AAB and T1D donors than in ND (Figure 5B). Light chain V genes used more frequently in both AAB and T1D were *IGLV9–49* and *IGKV1–8* (Figure S7B). Conversely, several heavy and light chain genes were expressed less among AAB and T1D donors; these included *IGHV3–30, IGHV1–2, IGHV1–69D, IGHV3–11, IGHV4–30-4, IGHV2–70D, IGHV1–69, IGLV2–23, IGLV3–1, IGLV3–19, IGLV2–14,* and *IGKV1–33* (Figures 5B and S7B).

We next determined whether expression of any unique heavy and light chain gene pairings was increased in the autoimmune donors compared with ND donors. Multiple heavy and light chain gene pairs found included pairs that used the significantly increased heavy chain genes *IGHV4–4, IGHV1–3, IGHV4–61,* and *IGHV1–18*, as well as the light chain gene *IGKV1–8* (Figure 5C). Interestingly, most of these pairings were observed in the AAB group. When we analyzed antigen reactivity of these paired heavy and light chains, we found that many bound islet antigens and were among the Islet_Multi group (Figure 5D). Altogether, we found unique BCR repertoire features in AAB and T1D donors compared with ND. These included significant increases in heavy-light chain gene pairs that exhibit islet reactivity.

Clonally expanded IAR B cells are predominantly found among AAB and T1D donors

We next examined whether any IAR B cells were clonally expanded in the AAB and/or T1D donors compared with ND donors. While there was relatively equal representation of negative, empty, and TET reactive clones in ND, AAB, and T1D donors, we found that clonally expanded IAR B cells were almost exclusively found among the autoimmune donors, particularly in the AAB group (Figures 6A and 6B, Data S1). Given that previous studies have identified public (i.e., common) TCRs that associate with T1D disease,[42,43] we searched for BCR sequences in our dataset that were enriched across donors. We identified 89 public B cell clonal sequences across 16 donors with diverse V gene pairings (Figure 6C). These public sequences varied in their representation among individuals and their donor group status but consisted of predominantly autoimmune AAB and T1D donors (Figure 6D). When we analyzed the antigen reactivity of these public BCR sequences, we found that many were islet antigen-reactive (Data S1). Overall, these findings demonstrate that clonal expansions of both private and public IAR BCR sequences occur and are found predominantly in AAB and T1D donors.

Recombinant antibodies validate the antigen reactivity of BCR sequences

To validate the antigen reactivity of the BCR sequences identified using our quantile approach, we selected 27 paired heavy and light chain V(D)J sequences (Data S2) for recombinant antibody production. Sequences were selected from the private clones in AAB and T1D donors that were found to bind each of the islet antigens, tetanus, and some that were in the Islet_Multi group (Table S2). As shown in Figure 7, results demonstrated that the recombinant antibodies were positive for each of the antigens implicated in cell binding studies, although some only became positive at the highest concentration of antibody added (Figures 7A–7D). These results indicate that while the antigen reactivities confirmed our quantile scoring method, many of these antibodies had relatively low affinity/reactivity for their antigen. This is not surprising given all but two were IgM isotype and had not undergone extensive somatic hypermutations. We next tested the reactivity of all of the antibodies including those that were identified to have bound to two or more antigens in the Islet_Multi group against each of the islet antigens, tetanus, DNA, and PE. Four of the nine Islet_Multi sequences tested were from IgG class switched cells. We found that nine of nine B cells identified as Islet_Multi were indeed polyreactive, although a few were only lowly reactive to antigens. In addition, we found one Islet_Multi antibody was largely PE-reactive (Figure 7E). Polyreactivity was defined as binding two or more antigens at a level above two standard deviations above the mean baseline (no antigen added) OD value. Additionally, among the antibodies classified as single antigen reactive, we found that the majority maintained high specificity to their antigen identified by sequencing and only had low reactivity to other antigens at the highest concentration tested of 40 μg/mL. Based on our definition of polyreactivity, 12 of 17 single-reactive antibodies were polyreactive. However, most of these were only reactive to two antigens, while the majority of the Islet_Multi reactive antibodies bound at least three or more antigens, indicating a greater breadth of polyreactivity. One of the antibodies that we identified as solely TET reactive was reactive to all antigens tested including PE. Finally, while our quantile scoring approach closely agreed with the previous method (LIBRA-seq) to identify antigen-reactive B cells via scRNA-seq of single-reactive B cells, we found that the two methods did not agree

well for cells that were called Islet_Multi (Figure S8). Since each of the tested Islet_Multi antibodies were reactive to at least one antigen, and most were confirmed to be polyreactive (Figure 7E), we found that our quantile scoring method was much more accurate than the LIBRA-seq method for identifying polyreactive B cells.

DISCUSSION

Although most studies of T1D have focused on T cells, B cells are essential to the pathogenesis of T1D. Here we performed the first scRNA-seq study that focused on the phenotypes and BCR repertoire of peripheral blood B cells in both AAB and T1D donors. Moreover, by studying INS, GAD, and IA-2 reactive B cells, we incorporated analyses of the pertinent antigen-reactive B cells thought to contribute to the pathogenesis of T1D. Our findings revealed that IAR B cells from AAB and recent-onset T1D donors display a unique phenotype including altered gene expression in pathways related to cell signaling, antigen presentation, and inflammation compared with ND donors. In addition, IAR B cells from AAB and T1D donors are clonally expanded, tend to be polyreactive, and exhibit a distinct BCR repertoire compared with ND donors. These findings provide an excellent backdrop and resource to pursue further studies on the unique features of IAR B cells during progression to T1D. They will additionally aid in development of more tailored B cell-targeted therapies and to monitor B cell response to therapy.

Our analysis of AAB patients begins to unlock early disease phenotypes specific to IAR B cells. AAB individuals have phenotypic patterns similar to T1D donors yet remain distinct with increased unique V(D)J gene usage over ND donors and greater representation of clonally expanded IAR B cells compared with T1D donors. These findings suggest that altered B cell phenotypes may be most pronounced prior to clinical diagnosis of T1D. By the time of clinical diagnosis, B cells are likely less actively participating in direct autoimmune attack, and therefore, the unique gene signatures that we identified may be waning. Future studies will be aimed at extending these analyses to more ND autoantibody-negative first-degree relatives to determine whether a B cell gene signature exists prior to seroconversion, and therefore could be used as the earliest biomarkers of disease described to date. An addition of healthy controls without detectable autoantibodies, non-HLA-matched, and no family history of autoimmunity would provide useful information to determine whether a B cell biomarker such as expansion of a specific B cell phenotype or clone, is present in high-risk, non-diabetic autoantibody-negative individuals, and therefore, precedes development of autoantibodies.

During our investigation into differentially expressed genes in IAR B cells in AAB and T1D donors compared with ND donors, we found several disease-related markers. Significant differences in mRNA encoding BCR signaling proteins including *BLNK, PTPN6 (SHP-1), LYN, CD79B, BANK1*, and *SYK* were observed in AAB and T1D samples compared with ND donors. Many of these genes function as activating or inhibiting signals upon antigen binding to the BCR, depending on the context. *LYN*, an Src tyrosine kinase known for dual roles in BCR signaling including both activation and inhibition, was downregulated in AAB and T1D donors. Studies in mice have demonstrated that the inhibitory role of *LYN* is important in maintaining B cell tolerance to self-antigens.[44,45] Perhaps the decreased

expression of *LYN* in AAB and T1D donors contributes to the rogue activation of IAR B cells. Our data also revealed that IAR B cells may function as antigen-presenting cells, as evidenced by increased expression of genes related to this pathway in AAB and T1D donors. While mouse studies have suggested the role of B cells in T1D involves antigen presentation to T cells,[3,19,20] this has never been formally demonstrated in humans. Differential gene expression showed IAR B cells from AAB and T1D donors are enriched in genes related to inflammation, activation, and infection. Taken together, IAR B cells in AAB and T1D donors display signs of activation and loss of B cell anergy, which is consistent with previous reports.[13,15] We compared our differentially expressed genes with a T1D GWAS study with 80 associated genes and found that there was an overlap of 32 differentially expressed genes between the two.[46] Using a hypergeometric test, we found that this is a 2-fold enrichment above what would be expected with a *p* value of 1.65e-5, indicating that our differentially expressed genes are associated with T1D risk.

Intriguingly, we detected a substantial number of polyreactive IAR B cells in our samples. Polyreactivity among early transitional and mature B cells has previously been established as a feature of multiple autoimmune diseases, including T1D.[41,47] Strikingly, the AAB donors in our study had the most Islet_Multi polyreactive expanded clones, even more than T1D donors. We hypothesize that these findings reflect increased immune activation in AAB donors during the earlier stages of disease. The increase in polyreactivity and clonal expansion of IAR B cells may indicate B cells could be initially activated by an antigen other than islet antigens, such as DNA, or potentially by a foreign antigen, such as a virus. Then, through molecular mimicry and/or affinity maturation toward islet reactivity, those B cells begin targeting islet antigens.[48–51] Environmental triggers, such as viral infection, have been proposed as precipitating events during the pathogenesis of T1D.[52,53] While none of our subjects reported illness at the time of blood draw, infection can lead to loss of B cell tolerance, activation, and autoantibody production.[54–56] Given that Epstein-Barr virus (EBV) infects B cells, we tested 14 of 16 donors for evidence of recent or prior infection with EBV. We found that three of five T1D, three of six AAB, and one of three ND donors were positive for anti-EBV IgG (Table S1). With this evidence along with the polyreactive nature of many of our antibodies, we chose eight to test for reactivity to EBV antigens by ELISA. One of those antibodies did react with EBV antigens and had been identified from AAB donor 117 who also was positive for EBV IgG in his serum (Figure S9). These results are in line with the hypothesis that molecular mimicry and B cell polyreactivity could have functional importance in the pathogenesis of T1D. Future studies in the lab will address this possibility in greater detail. Furthermore, it is important to note that these polyreactive B cells were identified using our quantile scoring approach and not via the previously published LIBRA-seq approach.[27] The LIBRA-seq method combines CLR normalization with a *Z* score to define a cell as antigen binding. This method is consequently biased against the possibility of finding cells reactive to more than a single antigen. This is different from our quantile method where we determined a cutoff for antigen reactivity on a per-sample basis using the non-B cells captured in the experiment to set a threshold for positivity, which was better able to identify the polyreactive B cells.

Even though all donors in this study shared *HLA* T1D risk alleles[28] and all ND donors were first-degree relatives of someone with T1D, we discovered BCR repertoire differences

between our donor groups. In particular, *IGHV4–4, IGHV1–3,* and *IGHV3–64* were significantly enriched in individuals experiencing islet autoimmunity compared with ND donors. Interestingly, *IGHV1–3* has previously been shown to be reactive to the EBV protein LMP-1.[57] In addition, *IGHV4–4* gene usage was enriched in antibodies found to target both rheumatoid arthritis (RA)-associated antigens and those from bacterial species.[58] These results may support an infectious environmental instigator of T1D hypothesis. Interestingly, AAB and T1D donor B cells showed a significantly decreased usage of *IGHV1–69,* which has previously been enriched in RA and multiple sclerosis (MS).[59–61] *IGHV1–69* polymorphisms have also been linked with varying responses to both natural infection and vaccine response.[62] Despite the fact that *IGHV4–34* gene usage is over-represented in many autoimmune diseases, including systemic lupus erythematosus (SLE), MS, and RA,[40,63] we did not find it to be enriched in our AAB and T1D donors. However, we did find significant enrichment for light chain *IGKV1–8* in AAB and T1D donors compared with ND, and this paired with *IGHV4–34* was found to be represented at a greater frequency in the autoimmune samples in this study. Moreover, there is 98.38% sequence similarity between *IGHV4–4* and *IGHV4–34,* which may warrant further exploration. Taken together, our findings revealed significant differences in the BCR repertoire of AAB and T1D donors compared with ND. These results represent a foundation from which to consider new therapeutics to target only the pathogenic B cells, while preserving non-pathogenic B cells, as well as the possibility of using the BCR repertoire as a biomarker of increased risk for disease development and/or to monitor response to therapy.

Limitations of the study

While this study has expanded our understanding of the role of B cells and B cell phenotypes during development of T1D, it has a few limitations. We sequenced peripheral blood B cells from five to six donors per group, which is a relatively small number of subjects. Despite this, we found many significant differences in gene expression and BCR repertoire between donor groups, and adding more individuals would further improve our statistical power. In addition, by sequencing more individuals in the future, we hope to compare B cell phenotypes of young (<12 years) to adult (>18 years) AAB and T1D donors, given previous studies have demonstrated B cells likely play a more important role in individuals diagnosed with T1D at a young age.[64–66] By focusing this study on peripheral blood B cells, we may have identified phenotypes that are not reflective of pathogenic B cells localized to the target tissue, such as the pancreas or pancreatic lymph nodes. Studies in our laboratory are currently under way to address this possibility. In addition, a large portion of the B cells that we captured were naive or IgM-expressing B cells. Recombinant monoclonal antibodies demonstrated that the affinities of the B cells were largely low affinity/reactivity, which is not unexpected given the low level of somatic hypermutations found in naive, non-class-switched B cells. However, autoreactive IgM+ cells have been found to still contribute to autoimmune pathogenesis in RA, autoimmune hemolytic anemia, and pemphigus or autoimmune neuropathy.[67] Future studies in the lab also aim to sequence more memory-like B cells to further characterize these B cell subsets and their differences in T1D vs. ND donors. Finally, while the ELISAs that were performed to determine the reactivity of the recombinant antibodies were done in triplicate, some ELISA assays had to

be repeated, which was not able to be completed for all 27 recombinant antibodies due to sample depletion.

In summary, this study revealed significant differences in gene expression and the BCR repertoire of peripheral blood IAR B cells in individuals with prediabetes and T1D compared with ND donors. These findings provide insight into how autoreactive B cells can be solicited to participate in an autoimmune response, demonstrate the breath of polyreactivity that is present among self-reactive B cells, and suggest pathways and/or targets that could be manipulated for future therapeutics to treat or prevent T1D.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to Mia J. Smith (mia.smith@cuanschutz.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Single-cell RNA-seq fastq and process files have been deposited at NCBI GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. All scripts and the analysis pipeline written in 'snakemake' v6.3.0 to replicate this analysis are publicly available on github (https://github.com/CUAnschutzBDC/smith_nicholas_sc_antigen). All R analysis, scar, dropkick, and T1K were run through docker containers that are available on dockerhub (https://hub.docker.com/repositories/kwellswrasman - smith_2024_scar, smith_2024_dropkick, smith_2024_r_docker, smith_2024_t1k). Full links to these images are listed in the key resources table. Any additional information required to reanalyze the data reported in this paper is available from Kristen L. Wells (kristen.wells-wrasman@cuanschutz.edu) upon request.

STAR★METHODS

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study design: This study sought to determine whether differences in the transcriptional phenotype and BCR repertoire of IAR B cells exist during development of autoimmunity. We generated uniquely barcoded PE-labeled antigen tetramers for INS, GAD, IA-2, and TET to identify antigen-reactive B cells using scRNA-seq. PBMCs from five ND first-degree relatives, six AAB subjects, and five recent-onset T1D patients were incubated with the various tetramers. Then, PE-binding B cells were sorted and loaded onto the 10X Genomics platform to be analyzed for gene expression and BCR repertoire differences. Recombinant antibodies were generated to confirm antigen reactivity of some of the identified BCR sequences.

Study participants: All participants were enrolled under study protocols approved by the Colorado Multiple Institutional Review Board (COMIRB #01-384). Written informed consent was obtained from each adult participant, or from parents or guardians of participants less than 18 years old. Assent was obtained from participants over the age of 7 years old who were cognitively able to consent. Participants were compensated for completing study procedures. All procedures were performed in accordance with COMIRB guidelines and regulations. Demographics and other data about research participants can be found in Table S1. Participants were recruited at the Barbara Davis Center at the University of Colorado Anschutz Medical Campus. Eligible subjects were male or female who had been recently diagnosed (<100 days) with T1D, Stage 1 or Stage 2 autoantibody positive prediabetic, or first-degree relatives of someone diagnosed with T1D. Stage 1 and 2 T1D subjects and first-degree relatives were recruited through TrialNet as local prospective collections under Ancillary Study #57 (PI Smith). Inclusion criteria for the first-degree relative group included: 1) autoantibody negative and 2) no history of autoimmune disease. All participants were required to have no history of illness within the last two weeks. HLA haplotypes and autoantibody levels against INS, IA-2, GAD65, and ZnT8 were determined at the CLIA certified Autoantibody/HLA Service Center at the Barbara Davis Center for Diabetes.

METHOD DETAILS

Blood sample processing: Between 15 and 45 mL of blood was drawn into sodium heparin tubes. The blood was rocked for no more than 2 h at room temperature prior to PBMC isolation using Ficoll density centrifugation. DNA for HLA typing was extracted from granulocytes using the DNA midi kit (Qiagen, Cat#51185). Diluted (1:1) plasma was collected from the Ficoll layer and frozen. PBMCs were cryo-frozen prior to use in the described experiments.

IA2 protein expression: A gene block was ordered from IDT containing the reverse transcription DNA sequence of most likely codons taken from the amino acid sequence of residues 605–979 of IA2 (i.e., PTPRN, which is known to bind IA2 specific autoantibodies in T1D individuals) and the appropriate flanking sequences for expression in our system (Data S3). 100 ng of the IA2 gene block was combined with PDonor 221 plasmid, BP clonase (Invitrogen, Cat#56481), and DH5-alpha competent cells were transformed and grown. Next, the IA2 transformed plasmid DNA was isolated and combined with Pet53dest plasmid, LR clonase (Invitrogen, Cat#56484), and DH5-alpha competent cells were transformed and grown. After this, the IA2 geneblock within the Pet53dest plasmid was isolated and mixed with BL21DE3 competent cells and grown prior to IPTG (Millipore Sigma, Cat#367–93-1) and EtOH induction of gene expression. The His-tagged IA2 was then purified with a NiNTA Agarose bead (Qiagen, Cat#1018244) packed column fraction collection. Pure IA2 fractions were determined by gel electrophoresis and proper folding and IA2 epitope accessibility was confirmed by the Barbara Davis Center Autoantibody Core. All cells and materials used in IA2 production were donated by Jay Hesselberth.

Generation of antigen tetramers: Human recombinant insulin (Millipore Sigma, Cat#19278), GAD65 (Diamyd Medical, Cat#45–08029-01), IA2 (described above), and

tetanus-toxoid (Colorado Serum Company, Cat# 11415) were biotinylated using the Thermo EZ-Link Sulfo-NHS-LC-Biotin Kit (Cat#A39257) with a 20 M ratio of biotin to protein. Excess biotin was removed using 7K Zeba columns for IA2, GAD, and TET (Thermo, Cat#89882). To ensure no loss of biotinylated INS, an Amicon 3kDa molecular weight cutoff centrifugal filter column was used to remove excess biotin and the top fraction was recovered (Millipore Sigma, Cat#UFC500396). Next, one mole of TotalSeqC PE Streptavidin was added to 4 mol of biotinylated antigen as follows: 0.5 μg of TotalSeqC PE Streptavidin was added to 1.54e-10 mol of biotinylated protein for 15 min on ice in the dark. This was repeated three times until a total of 2 μg (2 μg = 3.85e-11 mol) of TotalSeqC PE Streptavidin had been added. For samples 107, 108, 109, and 112, biotinylated insulin was combined with TotalSeq-C0961 PE Streptavidin (BioLegend, Cat#405155). All other samples were incubated with biotinylated insulin combined with TotalSeq-C0951 PE Streptavidin (BioLegend, Cat#405261). All samples were incubated with biotinylated tetramerized antigens as follows: GAD was combined with TotalSeq-C0954 PE Streptavidin (BioLegend, Cat#405267), IA2 was combined with TotalSeq-C0953 PE Streptavidin (BioLegend, Cat#405265), TET was combined with TotalSeq-C0952 PE Streptavidin (BioLegend, Cat#405263). For the "Empty" tetramer, TotalSeq-C0955 PE Streptavidin (BioLegend, Cat#405269) was quenched with biotin and excess biotin was removed using 7K Zebra columns (Thermo, Cat#89882). Validation of individual tetramers and appropriate dilutions to add to PBMCs was confirmed using flow cytometry. Tetramers were made fresh on the day of sample preparation for FACS sorting and 10X Genomics capture.

Sample preparation for FACS sorting: Four cryopreserved PBMC samples were thawed and processed together, and eight samples were processed per day to minimize batch effects. Cells were thawed in a 37 C water bath, washed twice with 10 mL 1X PBS, and kept on ice for cell staining. Cells were stained with LIVE/DEAD Fixable Near-IR Dead Cell Stain Kit at a 1:1500 dilution (ThermoFisher, Cat#L34975) in FACS buffer (1% BSA in 1X PBS) for 15 min on ice in the dark. Then, 2 μg of INS tetramer, 1 μg of IA2 and TET tetramers, and 0.75 μg of GAD and Empty tetramers in addition to 1:200 anti-human CD19-BV421 (BioLegend, Cat#363017) and 1:100 anti-human CD3-FITC (BioLegend, Cat#317305) were added to ~30e6 PBMCs in 1 mL total volume. Cells were stained on ice for 20 min in the dark. Finally, cells were washed twice in FACS buffer (400 × g for 5 min) prior to analysis on the cell sorter.

Sorting of PE-reactive B cells using FACS: After tetramer and antibody staining, cells were immediately brought to the CU Cancer Center Flow Cytometry Core. PE-binding B cells were sorted as Live/singlet/lymphocyte/CD19+/CD3−/PE + cells into 100 μL of FACS buffer. Up to 20,000 B cells were collected, and in some instances, PE-negative binding B cells were also sorted to increase the overall yield of B cells to be captured.

CITE-seq antibody staining: After the cells were sorted, they were stained with 1 μL each of the following TotalSeqC surface antibodies: anti-human IgM (BioLegend, Cat#314547), anti-human IgD (BioLegend, Cat#348245), anti-human CD27 (BioLegend, Cat#302853), anti-human CD21 (BioLegend, Cat#354923), and anti-human CXCR5

(BioLegend, Cat#356939). Cells were stained for 15 min on ice in the dark, washed once (400 × g for 5 min), then loaded onto the 10X Genomics system for cell capture at the CU Genomics Core. This data is unreported in the manuscript but present within the publicly available sequencing data and processed files.

5′ single cell capture and library prep: The 10X Genomics Chromium system was used for 5′ mRNA capture with chip K according to user guide CG000330_ChromiumNextGEMSingleCell5-v2_CellSurfaceProtein_UserGuide_RevF.pdf. The deviations from this protocol included: 1.) Overloading the chip with up to 20,000 cells per well depending on the sample, and 2.) Using the Feature Barcode Kit Primer 4 (Cat# 2000277) to detect and amplify TotalSeq C barcodes. Otherwise, the protocol was followed as written to produce gene expression, BCR V(D)J, and Feature Barcode Cell Surface Protein libraries for sequencing analysis. Library quality and quantity was determined using Agilent High Sensitivity D1000 TapeStation Screentape (Cat# 5067–5584), Reagents (Cat# 5067–5585), and Qubit dsDNA Quantification Assay (Cat# Q32851).

Sequencing: The libraries produced for each sample were sequenced at 5,000 read pairs per cell for V(D)J and Cell Surface Protein Dual Index libraries and 20,000 read pairs per cell for the 5′ Gene Expression Dual Index library. They were sequenced paired end dual index on an Illumina NovaSeq6000 and demultiplexed all by the CU Genomics Core Staff.

Recombinant monoclonal antibody production: Representative V(D)J sequences from AAB and T1D subjects identified by our quantile scoring approach that were part of an expanded IAR clone were chosen for antibody synthesis. B cells designated Islet_Multi reactive were chosen and validated for polyreactivity. Most monoclonal antibodies were synthesized by Twist Bioscience. Briefly, sequences identified in these experiments were cloned into the company's human IgG1, human kappa chain, or human lambda chain expression vectors. Heavy and light chain genes were co-transfected into HEK293 cells. Secreted monoclonal antibodies were then purified from the supernatant using protein A agarose beads. The HA-2B05 antibody was a gift from Jenna Guthmiller who described generation of this reagent in Guthmiller et al. 2021.[81] The cardiolipin and APOH antibodies were gifts from Elena Hsieh who obtained them from Creative Diagnostics.

ELISA assay: High-protein binding microtiter plates (Corning, Cat#07–200-35) were coated overnight at 4 C with 10 μg/mL of antigen (INS, IA2, GAD, TET, DNA, or PE (Invitrogen, Cat#P801)) in 1X PBS. We included DNA of the same length and sequence as the TotalSeqC tags and PE as antigens to represent potential binding to the DNA barcode or PE fluorophore components of the tetramer probes. Plates were washed three times with 1X PBS 0.05% Tween and blocked with 1X PBS 2% BSA blocking buffer for 1 h at room temperature. The plates were washed as before, and monoclonal antibodies were titrated down 2-fold onto antigen coated wells from 40–0.625 μg/mL in triplicate. Antibodies were incubated for 2 h at room temperature before washing. To detect antibody binding, goat anti human IgG (H + L) HRP (Invitrogen, Cat#31412) was added to wells at a 1:10,000 dilution for 1 h at room temperature in the dark. Plates were washed and developed using TMB solution (Life Technologies, Cat#002023). The reaction was stopped with addition of 1N

HCl and the plate was read at OD450nm using a BioTek Synergy H1 Microplate Reader. OD450 values from the blank wells included on each plate were subtracted to remove background signal from each test well prior to plotting with GraphPad Prism v9.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistics: Statistical analysis was performed using R version 4.2.3. Comparisons between two groups were conducted using paired or unpaired Student's t tests when comparing between samples given the small sample size. Bonferroni correction was always used when more than one comparison was run. Statistical analysis of sequencing data is described in depth below. Briefly, differential expression was run using MAST with multiple testing correction, gene set enrichment used a hypergeometric test, clustering of gene pathways was run based on jaccard distances and hierarchical clustering, differences in V gene usage was analyzed using an odds ratio test, and any differences between status groups using all cells were determined using a Wilcoxon rank-sum test.

Bioinformatic analysis

Pre-processing: Sequences from scRNA-seq were processed using 10x Genomics Cellranger v 7.1.0 software[82] using the human 10x genomics refdata-gex-GRCh38–2020-A reference downloaded from 10x genomics (https://www.10xgenomics.com/support/software/cell-ranger/downloads). V(D)J data was aligned using the 10x genomics reference refdata-cellranger-vdj-GRCh38-alts-ensembl-5.0.0 downloaded from 10x genomics (https://www.10xgenomics.com/support/software/cell-ranger/downloads). The antibody reference used can be found on our github https://github.com/CUAnschutzBDC/smith_nicholas_sc_antigen/blob/main/files/antibodies.csv.

Raw data generated by Cellranger were then read into R v4.2.3 using the Seurat[82] v4.3.0 R package with at least 200 genes per cell and at least 3 cells. Cells were further filtered based on the number of genes per cell and the percent of mitochondrial reads per cell by finding outliers using 'perCellQCMetrics' and 'perCellQCFilters' from 'scuttle' based on total RNA and ADT reads and features and percent mitochondria.[83] The data were normalized by using 'NormalizeData'. For each sample, variable genes were found by using 'FindVariableFeatures' followed by removal of the immunoglobulin genes from the set and data was scaled using 'ScaleData'.

Doublet removal: Doublets were removed using 'DoubletFinder'[68] using the default values except for pK, nExp, and PCs. The pK was identified using the pK associated with the maximum BCmetric value after running 'find.pK' from doublet finder. All samples used PCs 1:20. After doublet removal, 2057 (110), 4240 (116), 4353 (108), 4466 (107), 2848 (113), 9058 (114), 8815 (118), 4911 (106), 7949 (117), 8497 (115), 9265 (105), 10547 (111), 9750 (102), 8609 (112), 11114 (119), 5135 (109) were used for downstream analysis.

Dimensionality reduction and clustering: Dimensionality reduction and clustering were performed using 'RunPCA', 'FindNeighbors', 'FindClusters', and 'RunUMAP'. 'RunPCA' was run using the default values except all variable features (minus the immunoglobulin genes) were used for the features argument. 'FindNeighbors' was run with

default parameters except for the dims argument (different for each sample but ranging between 25 and 35). 'FindClusters' was run with default parameters except for the resolution argument (different for each sample but ranging between 0.6 and 1). 'RunUMAP' was run with default parameters except for the dims and metric arguments (the same number of dims as for 'FindNeighbors', metric = "correlation").

Cell type identification: We named clusters using existing PBMC datasets[68] as references and determined cluster identity using 'clustifyr'.[69] For each reference, we used our previously identified variable genes passed to 'query_genes' in the 'clustify' function. The top correlated cell type from any reference to each cluster was used as the cell type.

Analysis of tetramers

Scar ambient background removal: Because we noticed high amounts of tetramer expression in empty droplets, we performed scar[70] (single cell ambient remover) on each sample individually. To run scar, we followed the recommended steps (https://scar-tutorials.readthedocs.io/en/latest/tutorials/scAR_tutorial_ambient_profile.html). First, we read in the raw and filtered ADT matrices. We then ran 'scar.setup_anndata' to identify the ambient profile which we accessed with 'filtered_object.uns["ambient_profile_Antibody Capture"]'. Using this ambient profile and the raw data, we then ran 'scar.model', 'scar.train', and 'scar.inference' using the recommended arguments to pull out the ambient corrected ADT values.

Libra seq normalization: Libra seq normalization was performed on each sample individually as previously described.[70] We read in the corrected tetramer matrices from running scar ambient background removal. We then used the following formula

$$clrMat < -log2\left(\frac{(tetMat + 1)}{rowMeans(tetMat)}\right)$$

$$libraScore < -scale(clrMat, center = TRUE, scale = TRUE)$$

Any cells that had a value greater than 1 for the libra score for a tetramer were called positive for that tetramer.

Quantile normalization: Because we also had non-B cells captured in our sample, we used these to generate a second "quantile" tetramer score. Within each cell type individually, we used the cell types determined above and subset to only the non-B cells. We then read in the corrected tetramer matrices from running scar ambient background removal and identified the cutoff based on the 95th quantile. We generated our score by dividing each tetramer count for each cell by the cutoff determined for the given tetramer. In this way, like the Libra score, any score greater than 1 is considered positive for a given tetramer while any score less than 1 is considered negative.

Sample merging

Merging: The 16 samples were merged using the 'merge' function from Seurat. The merged data was then normalized and scaled using 'NormalizeData', 'FindVariableFeatures' followed by removal of the immunoglobulin genes, and 'ScaleData' as described above. Dimensionality reduction and clustering were performed using 'RunPCA', 'FindNeighbors', 'FindClusters', and 'RunUMAP'. 'RunPCA' was run using the default values except all variable features were used for the features argument (except for immunoglobulin genes). 'FindNeighbors' was run with default parameters except for the dims argument (dims = 1:30). 'FindClusters' was run with default parameters except for the resolution argument (resolution = 0.8). 'RunUMAP' was run with default parameters except for the dims and metric arguments (dims = 1:30, metric = "correlation").

Batch correction: We tested both 'harmony'[71] (using the default values except theta = 5) and 'mnn'[27,29] (using default values) batch correction methods on our data. After a thorough comparison between the two and exploration into how well cells that had been previously assigned as the same cell types overlapped, we chose to use the 'mnn' batch correction for all downstream analysis.

Cell type identification

Clustifyr: We made a first pass at naming clusters existing PBMC datasets[68] as references and determined cluster identity using 'clustifyr'.[69] For each reference, we used our previously identified variable genes passed to 'query_genes' in the 'clustify' function. The top correlated cell type from any reference to each cluster was used as the first determination of cell type.

Cluster markers and pathways: To identify markers of each cluster, 'FindAllMarkers' was run using default settings except we set only.pos to TRUE. Genes were called differentially expressed if the adjusted p-value was less than 0.05 and the log fold change was greater than 0.5. We inspected these marker genes to determine if the cell types determined above were correct.

V(D)J analysis

V(D)J statistics: V(D)J data was read in using 'import_vdj' from 'djvdj'[72] using the defaults except 'filter_paired = FALSE' and 'include_mutations = TRUE'. The values read in by 'djvdj' were used to determine V and J usage, CDR3 length, CDR3 charge, and percent somatic hypermutation. To calculate CDR3 charge, we used the 'charge' function from 'alakazam'.[73] Shannon diversity was calculated using 'calc_diversity' from 'djvdj' using the 'shannon' method from 'abdiv'.[72,84] Odds ratios and p-values were determined by finding the total count of each V gene for each comparison (T1D vs. ND or AAB vs. ND) and the total count of all cells for each comparison. The Odds ratio was calculated using the following formula

$$or = \frac{a * d}{(b * c)}$$

And the fishers exact test was run using

$$mat < - matrix(c(a, c, b, d), ncol = 2)$$

$$res < - fisher \cdot test(x = mat)$$

Where

$$a = total\ count\ of\ v\ gene\ \times\ in\ status\ 1$$

$$b = total\ count\ of\ v\ gene\ \times\ in\ status\ 2$$

$$total\ A = total\ v\ genes\ in\ status\ 1$$

$$total\ B = total\ v\ genes\ in\ status\ 2$$

$$c = total\ A - a$$

$$d = total\ B - b$$

Clonal analysis: Clones were identified using the 'AssignGenes.py', 'MakeDb.py' and 'DefineClones.py' from the Change-O package in the immcantation suite.[73] Clones were found both within and between samples. These clones were found based on CDR3 similarity using a hamming distance of 0.15 (85% similarity). We further refined these clones to have the same V and J call in addition to the CDR3 similarity calculated by 'DefineClones.py'

Differential expression: Differential expression analysis between each of the different status groups was performed using 'FindMarkers' from Seurat on each set of pairwise identities. Instead of using the default Wilcoxon test, we instead used the 'MAST' implementation. We performed DE using the same parameters on the twin sisters and the full family.

To find enriched pathways, we ran gProfiler2[74] through the 'run_gost' function from the 'scAnalysisR' package.[75,76] We then selected only the kegg pathways and clustered them using a custom function. This function first finds all genes associated with a particular keg term using 'keggLink' and 'keggConv' from the 'KEGGREST'[77] package and 'mapIds' from the 'AnnotationDbi'[79] package. Then, a matrix of jaccard distances for all pairwise gene sets is calculated. The jaccard distance was calculated with the following formula

$$intersection = length(intersect(a; b))$$

$$union = length(a) + length(b) - intersection$$

$$jaccard = 1 - \frac{intersection}{union}$$

Where

$$a = gene\ list\ 1$$

$$b = gene\ list\ 2$$

Finally, this matrix is converted into a distance matrix using the 'as.dist' function and clustered using 'hclust'. We used this clustering to narrow down the kegg pathways into discrete groups as displayed.

HLA typing: We identified HLA type using the 'T1K'[78] package. We performed HLA prediction using the example for 10x genomics data in their tutorial https://github.com/mourisl/T1K. Briefly, we used the R2 fastq as in input file, the R1 fastq as the barcode file, 0 15 + as the barcode range, the whitelist provided by 10x genomics as the barcodeWhitelist, and the hlaidx rna seq fasta file provided by the T1K package for the reference file. We used the output HLA alleles to determine the HLA types.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Most figures were created using the 'scAnalysisR' package available on github (https://github.com/CUAnschutzBDC/scAnalysisR). Figure 1A was created using BioRender, Figure 1B was created using FlowJo software v10.9, and Figure 7 was created using GraphPad Prism v9.

REFERENCES

1. Hulbert C, Riseili B, Rojas M, and Thomas JW (2001). B cell specificity contributes to the outcome of diabetes in nonobese diabetic mice. J. Immunol. 167, 5535–5538. 10.4049/jimmunol.167.10.5535. [PubMed: 11698422]

2. Noorchashm H, Lieu YK, Noorchashm N, Rostami SY, Greeley SA, Schlachterman A, Song HK, Noto LE, Jevnikar AM, Barker CF, and Naji A (1999). I-Ag7-mediated antigen presentation by B lymphocytes is critical in overcoming a checkpoint in T cell tolerance to islet beta cells of nonobese diabetic mice. J. Immunol. 163, 743–750. [PubMed: 10395666]

3. Mariño E, Tan B, Binge L, Mackay CR, and Grey ST (2012). B-cell cross-presentation of autologous antigen precipitates diabetes. Diabetes 61, 2893–2905. 10.2337/db12-0006. [PubMed: 22829452]

4. Serreze DV, Chapman HD, Varnum DS, Hanson MS, Reifsnyder PC, Richard SD, Fleming SA, Leiter EH, and Shultz LD (1996). B lymphocytes are essential for the initiation of T cell-mediated autoimmune diabetes: analysis of a new "speed congenic" stock of NOD.Ig mu null mice. J. Exp. Med. 184, 2049–2053. 10.1084/jem.184.5.2049. [PubMed: 8920894]

5. Noorchashm H, Greeley SA, and Naji A (2003). The role of t/b lymphocyte collaboration in the regulation of autoimmune and alloimmune responses. Immunol. Res. 27, 443–450. 10.1385/IR:27:2-3:443. [PubMed: 12857987]

6. Tian J, Zekzer D, Hanssen L, Lu Y, Olcott A, and Kaufman DL (2001). Lipopolysaccharide-activated B cells down-regulate Th1 immunity and prevent autoimmune diabetes in nonobese diabetic mice. J. Immunol. 167, 1081–1089. 10.4049/jimmunol.167.2.1081. [PubMed: 11441119]

7. Wong FS, Wen L, Tang M, Ramanathan M, Visintin I, Daugherty J, Hannum LG, Janeway CA Jr., and Shlomchik MJ (2004). Investigation of the role of B-cells in type 1 diabetes in the NOD mouse. Diabetes 53, 2581–2587. 10.2337/diabetes.53.10.2581. [PubMed: 15448087]

8. Mallone R, and Brezar V (2011). To B or not to B: (anti)bodies of evidence on the crime scene of type 1 diabetes? Diabetes 60, 2020–2022. 10.2337/db11-0700. [PubMed: 21788580]

9. Silva DG, Daley SR, Hogan J, Lee SK, Teh CE, Hu DY, Lam K-P, Goodnow CC, and Vinuesa CG (2011). Anti-islet autoantibodies trigger autoimmune diabetes in the presence of an increased frequency of islet-reactive CD4 T cells. Diabetes 60, 2102–2111. 10.2337/db10-1344. [PubMed: 21788582]

10. Pescovitz MD, Greenbaum CJ, Krause-Steinrauf H, Becker DJ, Gitelman SE, Goland R, Gottlieb PA, Marks JB, McGee PF, Moran AM, et al. (2009). Rituximab, B-lymphocyte depletion, and preservation of beta-cell function. N. Engl. J. Med. 361, 2143–2152. 10.1056/NEJMoa0904452. [PubMed: 19940299]

11. Herold KC, Pescovitz MD, McGee P, Krause-Steinrauf H, Spain LM, Bourcier K, Asare A, Liu Z, Lachin JM, and Dosch HM; Type 1 Diabetes TrialNet Anti-CD20 Study Group (2011). Increased T cell proliferative responses to islet antigens identify clinical responders to anti-CD20 monoclonal antibody (rituximab) therapy in type 1 diabetes. J. Immunol. 187, 1998–2005. 10.4049/jimmunol.1100539. [PubMed: 21775681]

12. Pescovitz MD, Greenbaum CJ, Bundy B, Becker DJ, Gitelman SE, Goland R, Gottlieb PA, Marks JB, Moran A, Raskin P, et al. (2014). B-lymphocyte depletion with rituximab and β-cell function: two-year results. Diabetes Care 37, 453–459. 10.2337/dc13-0626. [PubMed: 24026563]

13. Smith MJ, Packard TA, O'Neill SK, Henry Dunand CJ, Huang M, Fitzgerald-Miller L, Stowell D, Hinman RM, Wilson PC, Gottlieb PA, and Cambier JC (2015). Loss of anergic B cells in

prediabetic and new-onset type 1 diabetic patients. Diabetes 64, 1703–1712. 10.2337/db13-1798. [PubMed: 25524915]

14. Smith MJ, Rihanek M, Wasserfall C, Mathews CE, Atkinson MA, Gottlieb PA, and Cambier JC (2018). Loss of B-Cell Anergy in Type 1 Diabetes Is Associated With High-Risk HLA and Non-HLA Disease Susceptibility Alleles. Diabetes 67, 697–703. 10.2337/db17-0937. [PubMed: 29343548]

15. Stensland ZC, Magera CA, Broncucia H, Gomez BD, Rios-Guzman NM, Wells KL, Nicholas CA, Rihanek M, Hunter MJ, Toole KP, et al. (2023). Identification of an anergic BND cell-derived activated B cell population (BND2) in young-onset type 1 diabetes patients. J. Exp. Med. 220, e20221604. 10.1084/jem.20221604. [PubMed: 37184563]

16. Pociot F, and Lernmark Å (2016). Genetic risk factors for type 1 diabetes. Lancet 387, 2331–2339. 10.1016/S0140-6736(16)30582-7. [PubMed: 27302272]

17. Sosenko JM, Skyler JS, Palmer JP, Krischer JP, Yu L, Mahon J, Beam CA, Boulware DC, Rafkin L, Schatz D, et al. (2013). The prediction of type 1 diabetes by multiple autoantibody levels and their incorporation into an autoantibody risk score in relatives of type 1 diabetic patients. Diabetes Care 36, 2615–2620. 10.2337/dc13-0425. [PubMed: 23818528]

18. Ziegler AG, Rewers M, Simell O, Simell T, Lempainen J, Steck A, Winkler C, Ilonen J, Veijola R, Knip M, et al. (2013). Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. JAMA 309, 2473–2479. 10.1001/jama.2013.6285. [PubMed: 23780460]

19. Silveira PA, Johnson E, Chapman HD, Bui T, Tisch RM, and Serreze DV (2002). The preferential ability of B lymphocytes to act as diabetogenic APC in NOD mice depends on expression of self-antigen-specific immunoglobulin receptors. Eur. J. Immunol. 32, 3657–3666. 10.1002/1521-4141(200212)32:12<3657::AID-IMMU3657>3.0.CO;2-E. [PubMed: 12516557]

20. Serreze DV, Fleming SA, Chapman HD, Richard SD, Leiter EH, and Tisch RM (1998). B lymphocytes are critical antigen-presenting cells for the initiation of T cell-mediated autoimmune diabetes in nonobese diabetic mice. J. Immunol. 161, 3912–3918. [PubMed: 9780157]

21. Grandien A, Fucs R, Nobrega A, Andersson J, and Coutinho A (1994). Negative selection of multireactive B cell clones in normal adult mice. Eur. J. Immunol. 24, 1345–1352. 10.1002/eji.1830240616. [PubMed: 8206094]

22. Halverson R, Torres RM, and Pelanda R (2004). Receptor editing is the main mechanism of B cell tolerance toward membrane antigens. Nat. Immunol. 5, 645–650. 10.1038/ni1076. [PubMed: 15156139]

23. Wardemann H, Yurasov S, Schaefer A, Young JW, Meffre E, and Nussenzweig MC (2003). Predominant autoantibody production by early human B cell precursors. Science 301, 1374–1377. 10.1126/science.1086907. [PubMed: 12920303]

24. Pelanda R, and Torres RM (2012). Central B-cell tolerance: where selection begins. Cold Spring Harb. Perspect. Biol. 4, a007146. 10.1101/cshperspect.a007146. [PubMed: 22378602]

25. Gauld SB, Merrell KT, and Cambier JC (2006). Silencing of autoreactive B cells by anergy: a fresh perspective. Curr. Opin. Immunol. 18, 292–297. 10.1016/j.coi.2006.03.015. [PubMed: 16616480]

26. Meffre E, Schaefer A, Wardemann H, Wilson P, Davis E, and Nussenzweig MC (2004). Surrogate light chain expressing human peripheral B cells produce self-reactive antibodies. J. Exp. Med. 199, 145–150. 10.1084/jem.20031550. [PubMed: 14699083]

27. Setliff I, Shiakolas AR, Pilewski KA, Murji AA, Mapengo RE, Janowska K, Richardson S, Oosthuysen C, Raju N, Ronsard L, et al. (2019). High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity. Cell 179, 1636–1646.e15. 10.1016/j.cell.2019.11.003. [PubMed: 31787378]

28. Noble JA, Valdes AM, Cook M, Klitz W, Thomson G, and Erlich HA (1996). The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. Am. J. Hum. Genet. 59, 1134–1148. [PubMed: 8900244]

29. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. (2021). Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587.e29. 10.1016/j.cell.2021.04.048. [PubMed: 34062119]

30. Haghverdi L, Lun ATL, Morgan MD, and Marioni JC (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. 36, 421–427. 10.1038/nbt.4091. [PubMed: 29608177]

31. Mouat IC, Goldberg E, and Horwitz MS (2022). Age-associated B cells in autoimmune diseases. Cell. Mol. Life Sci. 79, 402. 10.1007/s00018-022-04433-9. [PubMed: 35798993]

32. King HW, Wells KL, Shipony Z, Kathiria AS, Wagar LE, Lareau C, Orban N, Capasso R, Davis MM, Steinmetz LM, et al. (2021). Integrated single-cell transcriptomics and epigenomics reveals strong germinal center-associated etiology of autoimmune risk loci. Sci. Immunol. 6, eabh3768. 10.1126/sciimmunol.abh3768. [PubMed: 34623901]

33. Stewart A, Ng JC-F, Wallis G, Tsioligka V, Fraternali F, and Dunn-Walters DK (2021). Single-Cell Transcriptomic Analyses Define Distinct Peripheral B Cell Subsets and Discrete Development Pathways. Front. Immunol. 12, 602539. 10.3389/fimmu.2021.602539. [PubMed: 33815362]

34. Franzén O, Gan L-M, and Björkegren JLM (2019). PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database (Oxford). 10.1093/database/baz046.

35. Jourdan M, Caraux A, De Vos J, Fiol G, Larroque M, Cognot C, Bret C, Duperray C, Hose D, and Klein B (2009). An in vitro model of differentiation of memory B cells into plasmablasts and plasma cells including detailed phenotypic and molecular characterization. Blood 114, 5173–5181. 10.1182/blood-2009-07-235960. [PubMed: 19846886]

36. Efremov DG, Turkalj S, and Laurenti L (2020). Mechanisms of B Cell Receptor Activation and Responses to B Cell Receptor Inhibitors in B Cell Malignancies. Cancers 12, 1396. 10.3390/cancers12061396. [PubMed: 32481736]

37. Getahun A (2022). Role of inhibitory signaling in peripheral B cell tolerance. Immunol. Rev. 307, 27–42. 10.1111/imr.13070. [PubMed: 35128676]

38. Tsubata T, and Wienands J (2001). B cell signaling. Introduction. Int. Rev. Immunol. 20, 675–678. 10.3109/08830180109045584. [PubMed: 11913944]

39. Perez-Andres M, Paiva B, Nieto WG, Caraux A, Schmitz A, Almeida J, Vogt RF Jr., Marti GE, Rawstron AC, Van Zelm MC, et al. (2010). Human peripheral blood B-cell compartments: a crossroad in B-cell traffic. Cytometry B Clin. Cytom. 78, S47–S60. 10.1002/cyto.b.20547. [PubMed: 20839338]

40. Bashford-Rogers RJM, Smith KGC, and Thomas DC (2018). Antibody repertoire analysis in polygenic autoimmune diseases. Immunology 155, 3–17. 10.1111/imm.12927. [PubMed: 29574826]

41. Yurasov S, Wardemann H, Hammersen J, Tsuiji M, Meffre E, Pascual V, and Nussenzweig MC (2005). Defective B cell tolerance checkpoints in systemic lupus erythematosus. J. Exp. Med. 201, 703–711. 10.1084/jem.20042251. [PubMed: 15738055]

42. Linsley PS, Barahmand-Pour-Whitman F, Balmas E, DeBerg HA, Flynn KJ, Hu AK, Rosasco MG, Chen J, O'Rourke C, Serti E, et al. (2021). Autoreactive T cell receptors with shared germline-like α chains in type 1 diabetes. JCI Insight 6, e151349. 10.1172/jci.insight.151349. [PubMed: 34806648]

43. Nakayama M, and Michels AW (2021). Using the T Cell Receptor as a Biomarker in Type 1 Diabetes. Front. Immunol. 12, 777788. 10.3389/fimmu.2021.777788. [PubMed: 34868047]

44. Lamagna C, Hu Y, DeFranco AL, and Lowell CA (2014). B cell-specific loss of Lyn kinase leads to autoimmunity. J. Immunol. 192, 919–928. 10.4049/jimmunol.1301979. [PubMed: 24376269]

45. Getahun A, Beavers NA, Larson SR, Shlomchik MJ, and Cambier JC (2016). Continuous inhibitory signaling by both SHP-1 and SHIP-1 pathways is required to maintain unresponsiveness of anergic B cells. J. Exp. Med. 213, 751–769. 10.1084/jem.20150537. [PubMed: 27114609]

46. Chiou J, Geusz RJ, Okino M-L, Han JY, Miller M, Melton R, Beebe E, Benaglio P, Huang S, Korgaonkar K, et al. (2021). Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. Nature 594, 398–402. 10.1038/s41586-021-03552-w. [PubMed: 34012112]

47. Meffre E, and O'Connor KC (2019). Impaired B-cell tolerance checkpoints promote the development of autoimmune diseases and pathogenic autoantibodies. Immunol. Rev. 292, 90–101. 10.1111/imr.12821. [PubMed: 31721234]

48. Oldstone MB (1998). Molecular mimicry and immune-mediated diseases. FASEB J. 12, 1255–1265. 10.1096/fasebj.12.13.1255. [PubMed: 9761770]

49. Suliman BA (2024). Potential clinical implications of molecular mimicry-induced autoimmunity. Immun. Inflamm. Dis. 12, e1178. 10.1002/iid3.1178. [PubMed: 38415936]

50. Sundaresan B, Shirafkan F, Ripperger K, and Rattay K (2023). The Role of Viral Infections in the Onset of Autoimmune Diseases. Viruses 15, 782. 10.3390/v15030782. [PubMed: 36992490]

51. Trier NH, and Houen G (2023). Antibody Cross-Reactivity in Auto-Immune Diseases. Int. J. Mol. Sci. 24, 13609. 10.3390/ijms241713609. [PubMed: 37686415]

52. Coppieters KT, Boettler T, and von Herrath M (2012). Virus infections in type 1 diabetes. Cold Spring Harb. Perspect. Med. 2, a007682. 10.1101/cshperspect.a007682. [PubMed: 22315719]

53. Isaacs SR, Foskett DB, Maxwell AJ, Ward EJ, Faulkner CL, Luo JYX, Rawlinson WD, Craig ME, and Kim KW (2021). Viruses and Type 1 Diabetes: From Enteroviruses to the Virome. Microorganisms 9, 1519. 10.3390/microorganisms9071519. [PubMed: 34361954]

54. Castleman MJ, Stumpf MM, Therrien NR, Smith MJ, Lesteberg KE, Palmer BE, Maloney JP, Janssen WJ, Mould KJ, Beckham JD, et al. (2022). SARS-CoV-2 infection relaxes peripheral B cell tolerance. J. Exp. Med. 219, e20212553. 10.1084/jem.20212553. [PubMed: 35420627]

55. Soulas P, Woods A, Jaulhac B, Knapp A-M, Pasquali J-L, Martin T, and Korganow A-S (2005). Autoantigen, innate immunity, and T cells cooperate to break B cell tolerance during bacterial infection. J. Clin. Investig. 115, 2257–2267. 10.1172/JCI24646. [PubMed: 16041408]

56. Hansen KE, Arnason J, and Bridges AJ (1998). Autoantibodies and common viral illnesses. Semin. Arthritis Rheum. 27, 263–271. 10.1016/s0049-0172(98)80047-4. [PubMed: 9572708]

57. Lomakin YA, Zakharova MY, Stepanov AV, Dronina MA, Smirnov IV, Bobik TV, Pyrkov AY, Tikunova NV, Sharanova SN, Boitsov VM, et al. (2014). Heavy–light chain interrelations of MS-associated immunoglobulins probed by deep sequencing and rational variation. Mol. Immunol. 62, 305–314. 10.1016/j.molimm.2014.01.013. [PubMed: 24534716]

58. Chriswell ME, Lefferts AR, Clay MR, Hsu AR, Seifert J, Feser ML, Rims C, Bloom MS, Bemis EA, Liu S, et al. (2022). Clonal IgA and IgG autoantibodies from individuals at risk for rheumatoid arthritis identify an arthritogenic strain of Subdoligranulum. Sci. Transl. Med. 14, eabn5166. 10.1126/scitranslmed.abn5166. [PubMed: 36288282]

59. Vencovský J, Zd'árský E, Moyes SP, Hajeer A, Ruzicková S, Cimburek Z, Ollier WE, Maini RN, and Mageed RA (2002). Polymorphism in the immunoglobulin VH gene V1–69 affects susceptibility to rheumatoid arthritis in subjects lacking the HLA-DRB1 shared epitope. Rheumatology 41, 401–410. 10.1093/rheumatology/41.4.401. [PubMed: 11961170]

60. Edwards JCW, Szczepanski L, Szechinski J, Filipowicz-Sosnowska A, Emery P, Close DR, Stevens RM, and Shaw T (2004). Efficacy of B-cell-targeted therapy with rituximab in patients with rheumatoid arthritis. N. Engl. J. Med. 350, 2572–2581. 10.1056/NEJMoa032534. [PubMed: 15201414]

61. Baranzini SE, Jeong MC, Butunoi C, Murray RS, Bernard CC, and Oksenberg JR (1999). B cell repertoire diversity and clonal expansion in multiple sclerosis brain lesions. J. Immunol. 163, 5133–5144. [PubMed: 10528220]

62. Mikocziova I, Greiff V, and Sollid LM (2021). Immunoglobulin germline gene variation and its impact on human disease. Genes Immun. 22, 205–217. 10.1038/s41435-021-00145-5. [PubMed: 34175903]

63. Yuuki H, Itamiya T, Nagafuchi Y, Ota M, and Fujio K (2024). B cell receptor repertoire abnormalities in autoimmune disease. Front. Immunol. 15, 1326823. 10.3389/fimmu.2024.1326823. [PubMed: 38361948]

64. Smith MJ, Cambier JC, and Gottlieb PA (2020). Endotypes in T1D: B lymphocytes and early onset. Curr. Opin. Endocrinol. Diabetes Obes. 27, 225–230. 10.1097/MED.0000000000000547. [PubMed: 32618634]

65. Hinman RM, and Cambier JC (2014). Role of B lymphocytes in the pathogenesis of type 1 diabetes. Curr. Diab. Rep. 14, 543. 10.1007/s11892-014-0543-8. [PubMed: 25189436]

66. Redondo MJ, and Morgan NG (2023). Heterogeneity and endotypes in type 1 diabetes mellitus. Nat. Rev. Endocrinol. 19, 542–554. 10.1038/s41574-023-00853-0. [PubMed: 37337007]

67. Amendt T, and Yu P (2023). TLR7 and IgM: Dangerous Partners in Autoimmunity. Antibodies (Basel) 12, 4. 10.3390/antib12010004. [PubMed: 36648888]

68. McGinnis CS, Murrow LM, and Gartner ZJ (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. 8, 329–337.e4. 10.1016/j.cels.2019.03.003. [PubMed: 30954475]

69. Fu R, Gillen AE, Sheridan RM, Tian C, Daya M, Hao Y, Hesselberth JR, and Riemondy KA (2020). clustifyr: an R package for automated single-cell RNA sequencing cluster classification. F1000Res. 9, 223. 10.12688/f1000research.22969.2. [PubMed: 32765839]

70. Sheng C, Lopes R, Li G, Schuierer S, Waldt A, Cuttat R, Dimitrieva S, Kauffmann A, Durand E, Galli GG, et al. (2022). Probabilistic machine learning ensures accurate ambient denoising in droplet-based single-cell omics. Preprint at: bioRxiv. 10.1101/2022.01.14.476312

71. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh P-R, and Raychaudhuri S (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296. 10.1038/s41592-019-0619-0. [PubMed: 31740819]

72. An R package to analyze single-cell V(D)J data (Github). https://github.com/rnabioco/djvdj.

73. Gupta NT, Vander Heiden JA, Uduman M, Gadala-Maria D, Yaari G, and Kleinstein SH (2015). Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. Bioinformatics 31, 3356–3358. 10.1093/bioinformatics/btv359. [PubMed: 26069265]

74. Kolberg L, Raudvere U, Kuzmin I, Vilo J, and Peterson H (2020). gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res. 9, ELIXIR-709. 10.12688/f1000research.24956.2.

75. scAnalysisR: A package to analyze single cell data in R. This package is under active development (Github). https://github.com/CUAnschutzBDC/scAnalysisR

76. Wells KL, Miller CN, Gschwind AR, Wei W, Phipps JD, Anderson MS, and Steinmetz LM (2020). Combined transient ablation and single-cell RNA-sequencing reveals the development of medullary thymic epithelial cells. Elife 9, e60188. 10.7554/eLife.60188. [PubMed: 33226342]

77. KEGGREST Bioconductor. http://bioconductor.org/packages/KEGGREST.

78. Song L, Bai G, Liu XS, Li B, and Li H (2023). Efficient and accurate KIR and HLA genotyping with massively parallel sequencing data. Genome Res. 33, 923–931. 10.1101/gr.277585.122. [PubMed: 37169596]

79. AnnotationDbi Bioconductor. http://bioconductor.org/packages/AnnotationDbi.

80. Koster J, and Rahmann S (2012). Snakemake—a scalable bioinformatics workflow. Bioinformatics 28, 2520–2522. 10.1093/bioinformatics/bts480. [PubMed: 22908215]

81. Guthmiller JJ, Han J, Li L, Freyn AW, Liu STH, Stovicek O, Stamper CT, Dugan HL, Tepora ME, Utset HA, et al. (2021). First exposure to the pandemic H1N1 virus induced broadly neutralizing antibodies targeting hemagglutinin head epitopes. Sci. Transl. Med. 13, eabg4535. 10.1126/scitranslmed.abg4535. [PubMed: 34078743]

82. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049. 10.1038/ncomms14049. [PubMed: 28091601]

83. McCarthy DJ, Campbell KR, Lun ATL, and Wills QF (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics 33, 1179–1186. 10.1093/bioinformatics/btw777. [PubMed: 28088763]

84. Bittinger K (2020). Alpha and Beta Diversity Measures [R package abdiv version]. https://github.com/kylebittinger/abdiv.

Highlights

- Aab+ and T1D donors have distinct gene expression changes

- BCR V gene enrichment is observed in Aab+ and T1D B cells

- Clonal expansions are greater among Aab+ and T1D donors

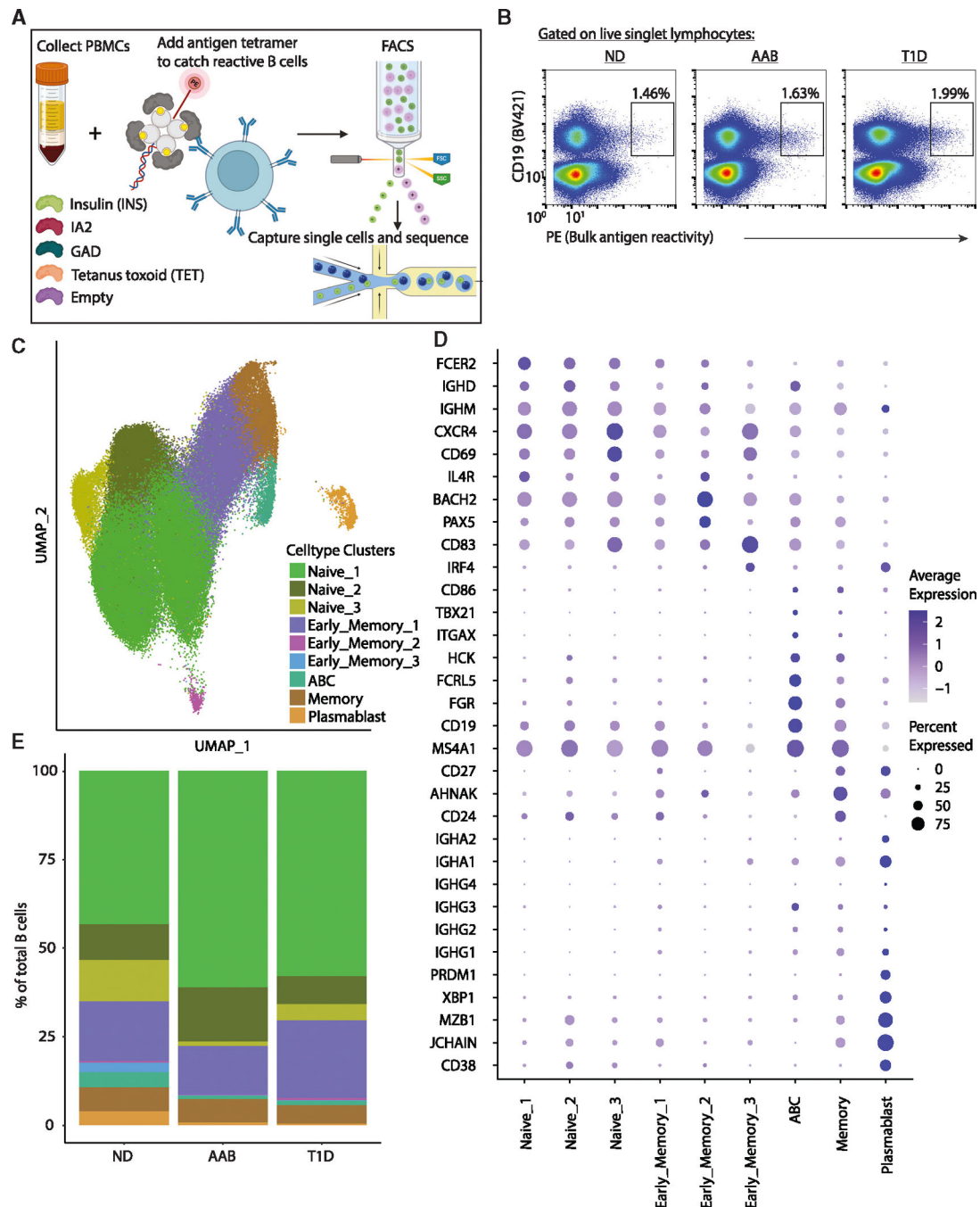- Islet antigen-reactive cells appear polyreactive

Figure 1. Sorted antigen-reactive B cells constitute nine B cell clusters among all donors

(A) Schematic overview of antigen-tetramer labeling of reactive B cells, FACS, and single-cell sequencing.

(B) Representative flow cytometry plots of how total antigen-reactive (PE+) B cells from ND, AAB, and T1D donors were sorted.

(C) UMAP plot of total B cells collected across 16 donors clustered into nine subpopulations: three naive, three early memory, ABC, classical memory, and plasmablasts.

(D) Dot plot of select gene expression across B cell clusters to confirm B cell subtype identity.

(E) Stacked bar chart of B cell subtype distribution as a percent of total B cells across donor status groups. See also Figures S1 and S2.
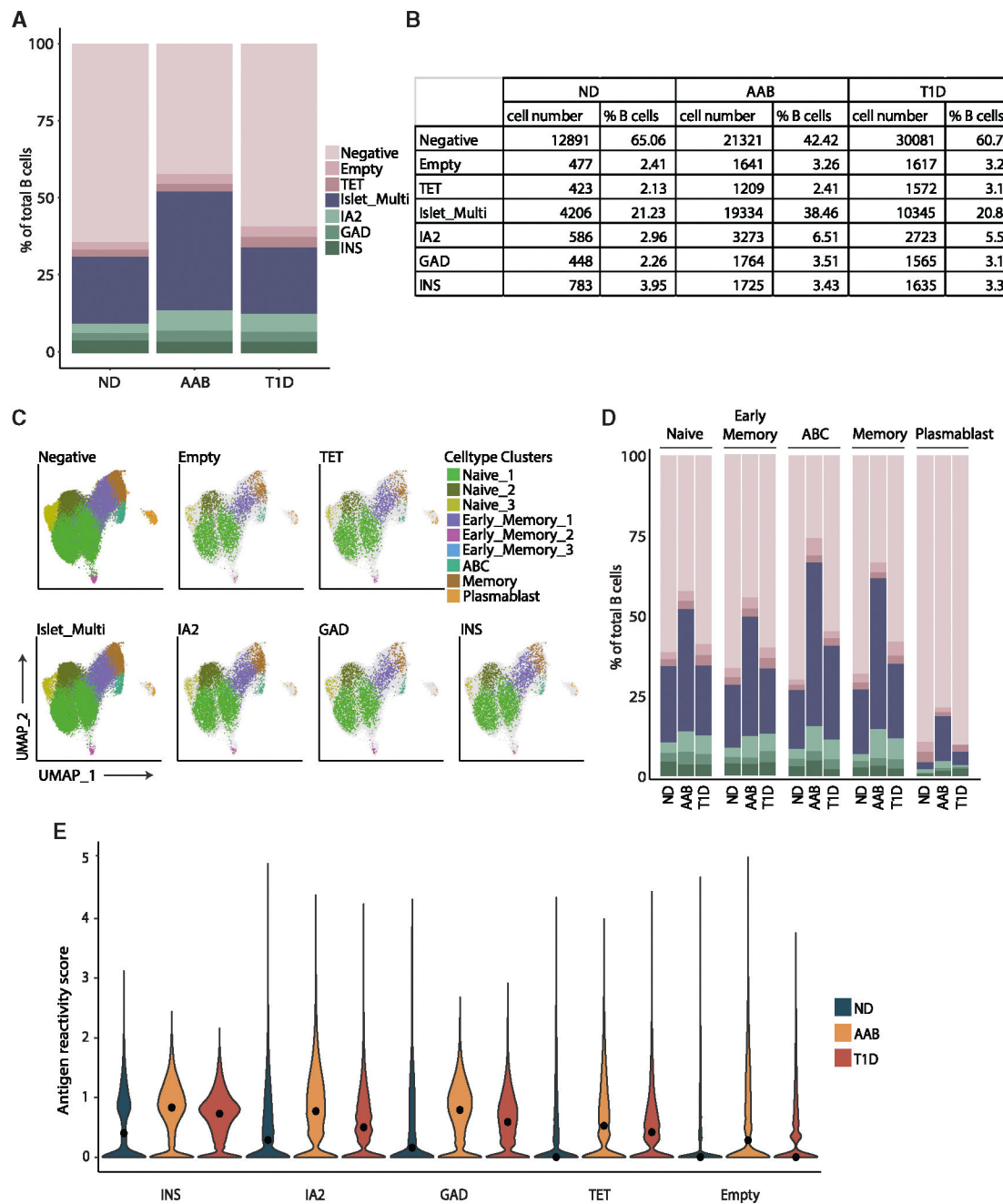
A

B

	ND		AAB		T1D	
	cell number	% B cells	cell number	% B cells	cell number	% B cells
Negative	12891	65.06	21321	42.42	30081	60.72
Empty	477	2.41	1641	3.26	1617	3.26
TET	423	2.13	1209	2.41	1572	3.17
Islet_Multi	4206	21.23	19334	38.46	10345	20.88
IA2	586	2.96	3273	6.51	2723	5.50
GAD	448	2.26	1764	3.51	1565	3.16
INS	783	3.95	1725	3.43	1635	3.30

C

D

E

Figure 2. AAB and T1D donors have increased IAR B cells compared with ND donors

(A) Stacked bar chart of antigen reactivity as a percent of total B cells across ND, AAB, and T1D donors.

(B) Table listing the frequency of cells that are antigen reactive and their percentage of the whole B cell population for each donor group.

(C) UMAP projections of all cells colored by antigen reactivity identity.

(D) Stacked bar chart of antigen reactivity as a percentage of total B cells separated by B cell subtype and donor group.

(E) Violin plots of relative antigen reactivity score using the quantile approach, separated by donor group. See also Figures S3 and S4.
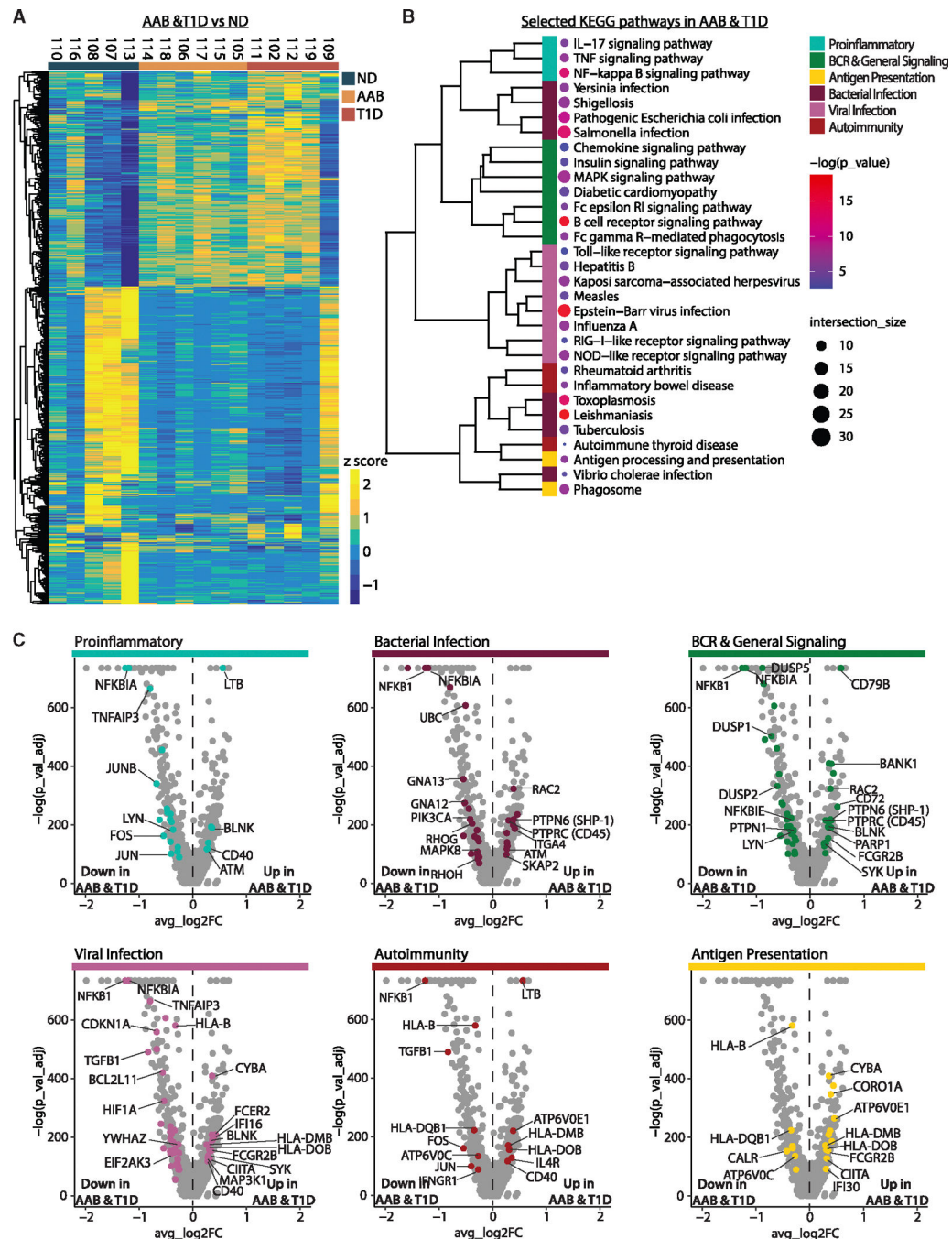
Figure 3. Differential mRNA gene expression patterns are found in IAR B cells in AAB and T1D donors

(A) Heatmap showing differentially expressed gene patterns of IAR B cells in AAB and T1D donors compared with ND donors.

(B) Selected significantly enriched KEGG pathways for AAB and T1D donors grouped into six broad categories by general theme: Proinflammatory, BCR & General Signaling, Antigen Processing and Presentation, Bacterial Infection, Viral Infection, and Autoimmunity. Similarity between individual pathways, determined by jaccard distance, is shown by clustering dendrogram on the left side.

(C). Volcano plots highlighting selected differentially expressed genes that are found in each of the KEGG pathways from (B). See also Figures S5 and S6.
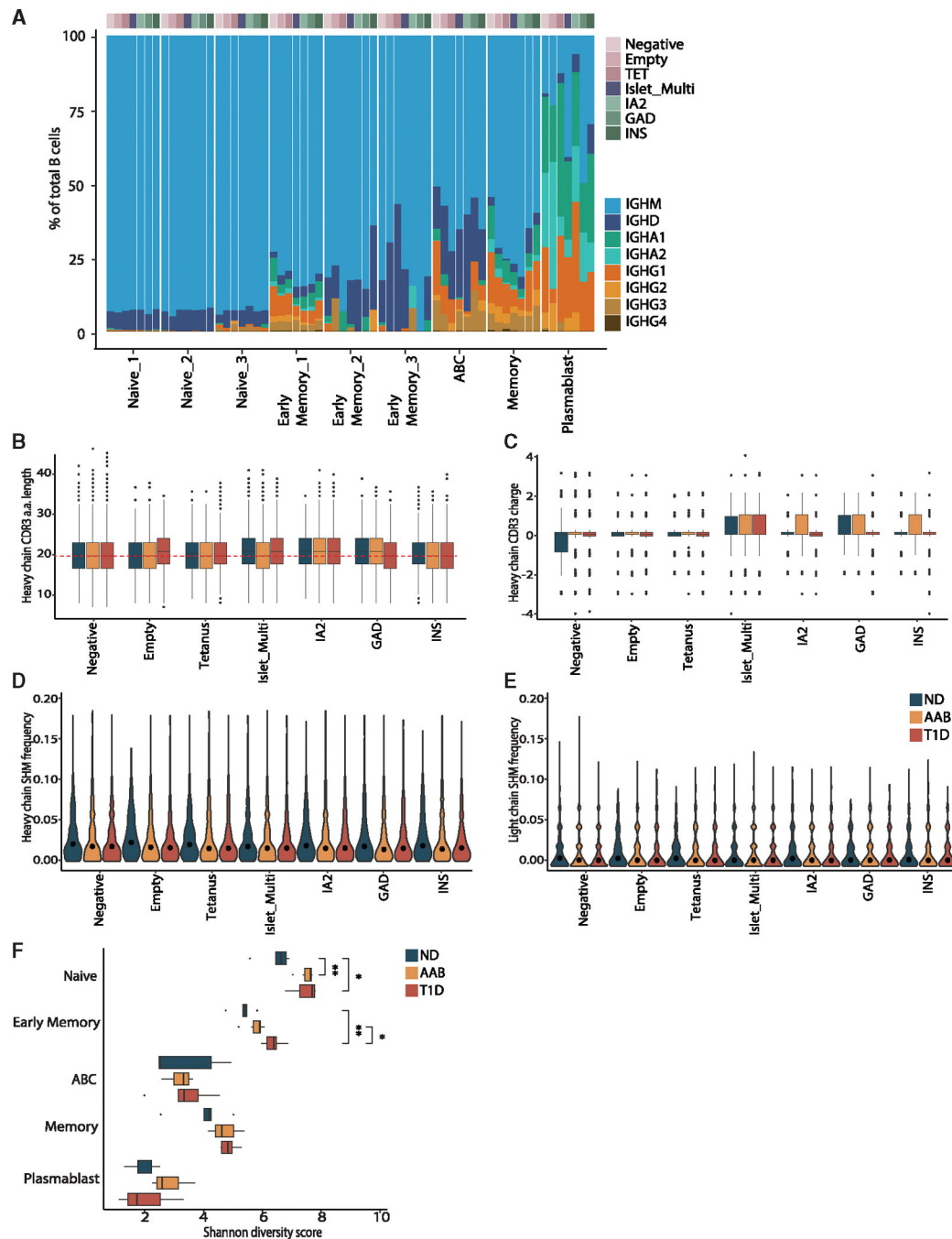
Figure 4. B cell receptor analysis reveals increased diversity in IAR B cells of AAB and T1D donors

(A) Stacked bar plot of BCR isotype usage determined by V(D)J analysis as a percent of all B cells. Cells are divided by subtype and antigen reactivity.

(B) Boxplots of amino acid length of heavy chain CDR3s for each antigen and donor group.

(C) Boxplots of heavy chain CDR3 net charge for each antigen and donor group.

(D and E) Violin plots of somatic hypermutation (SHM) frequency for each antigen and donor group for the heavy chain (D) and light chain (E).

(F) Boxplots of Shannon diversity scores for each B cell subtype within each donor group. A t test was calculated for each donor group comparison. Significant differences are shown with $*p < 0.05$ and $**p < 0.01$.
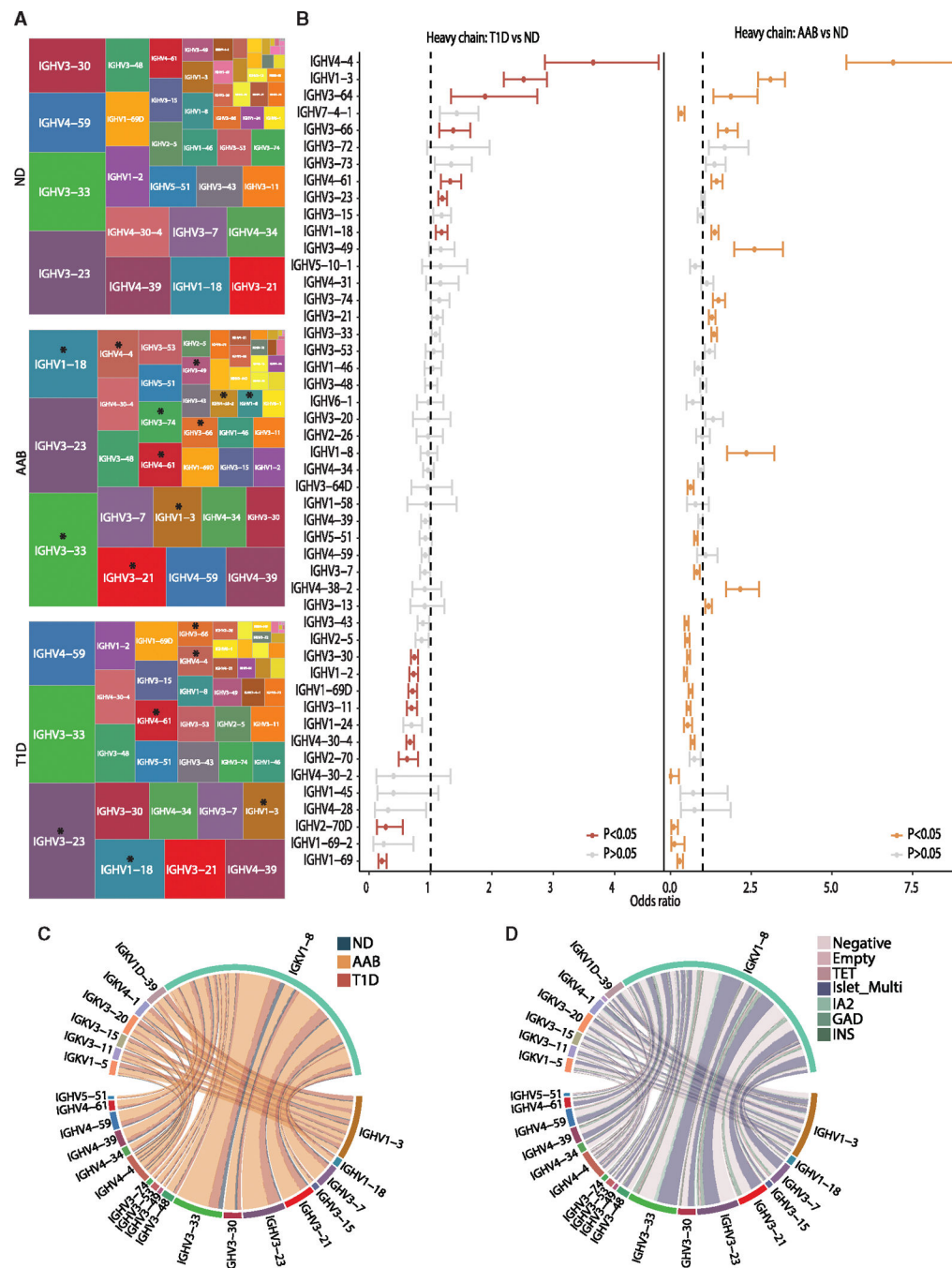
Figure 5. AAB and T1D donors have significant enrichment for particular heavy chain V gene usages

(A) Total B cells for each donor group were assessed for heavy chain V gene usage and plotted such that the size of each box is proportional to the frequency with which that gene was found among donors. Asterisks mark genes that were found to be significantly upregulated in AAB and T1D donors compared with ND.

(B) Odds ratios were performed for heavy chain V gene usage between T1D and ND donors and AAB and ND donors. *p* value is indicated by color where red indicates *p* < 0.05 for T1D vs. ND and orange indicates *p* < 0.05 for AAB vs. ND.

(C) Circos plot of significantly enriched heavy and light chain gene pairings in T1D and AAB vs. ND. Heavy and light pairs are connected by lines colored by donor group and are thickened by the number of cells that used each segment.

(D) As in (C), but pairs are colored by antigen reactivity. See also Figure S7.
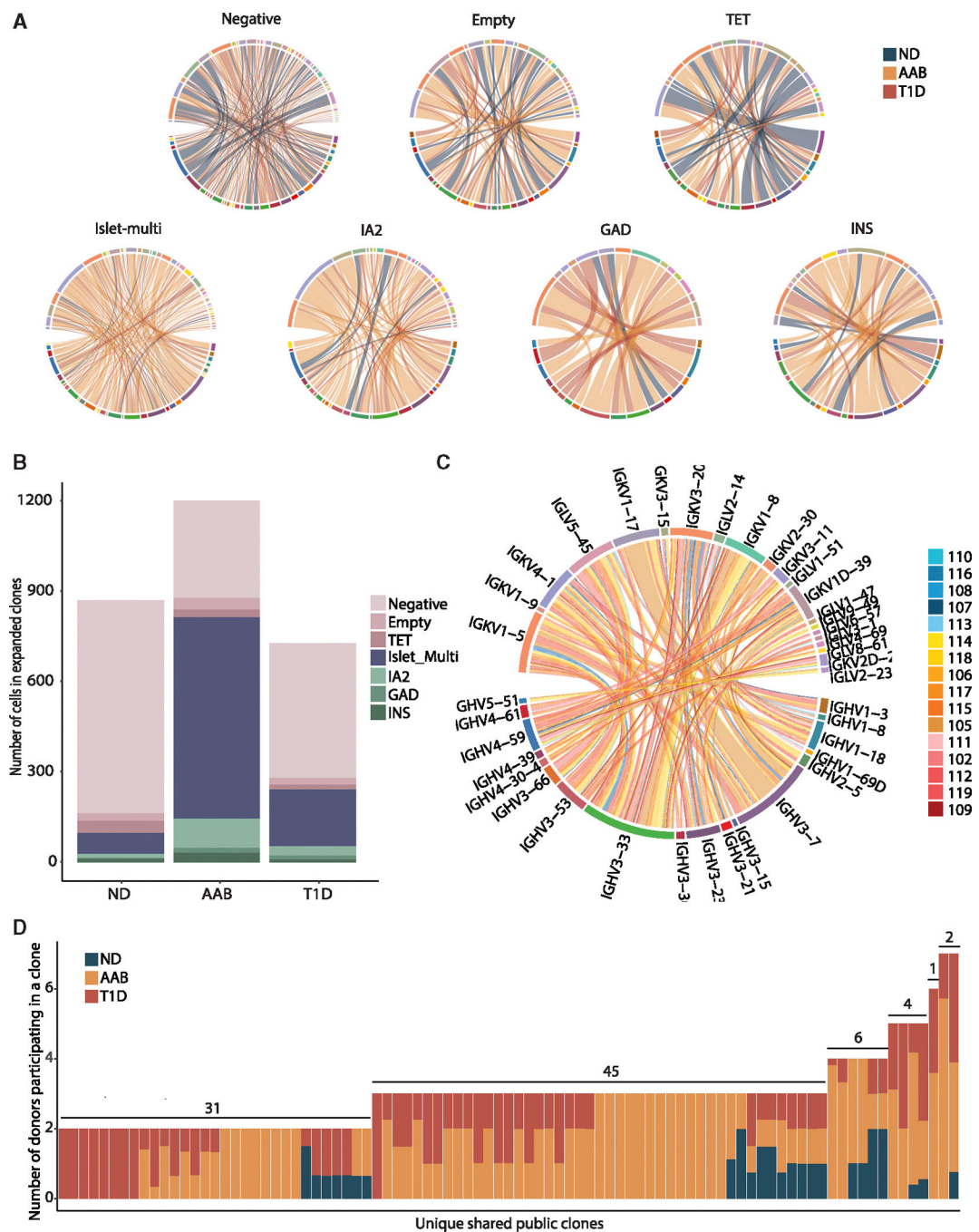
Figure 6. Clonal expansions of IAR B cells are increased in AAB and T1D donors

(A) Circos plots of heavy and light V gene pairings from private (within a single individual donor) clonally expanded cells. Segment connections are colored by donor status groups and plots are separated by antigen reactivity.

(B) Stacked bar plot of the absolute number of private clonally expanded cells grouped by donor type and colored by antigen reactivity.

(C) Circos plot of V gene pairs from public (shared among at least two donors) clonally expanded cells. Segment connections are colored by individual donor ID. Shades of blue represent ND, shades of yellow represent AAB, and shades of red represent T1D donors.

(D) Stacked bar plot with the number of individual donors participating in each shared public clone ($n = 89$) colored by donor status.
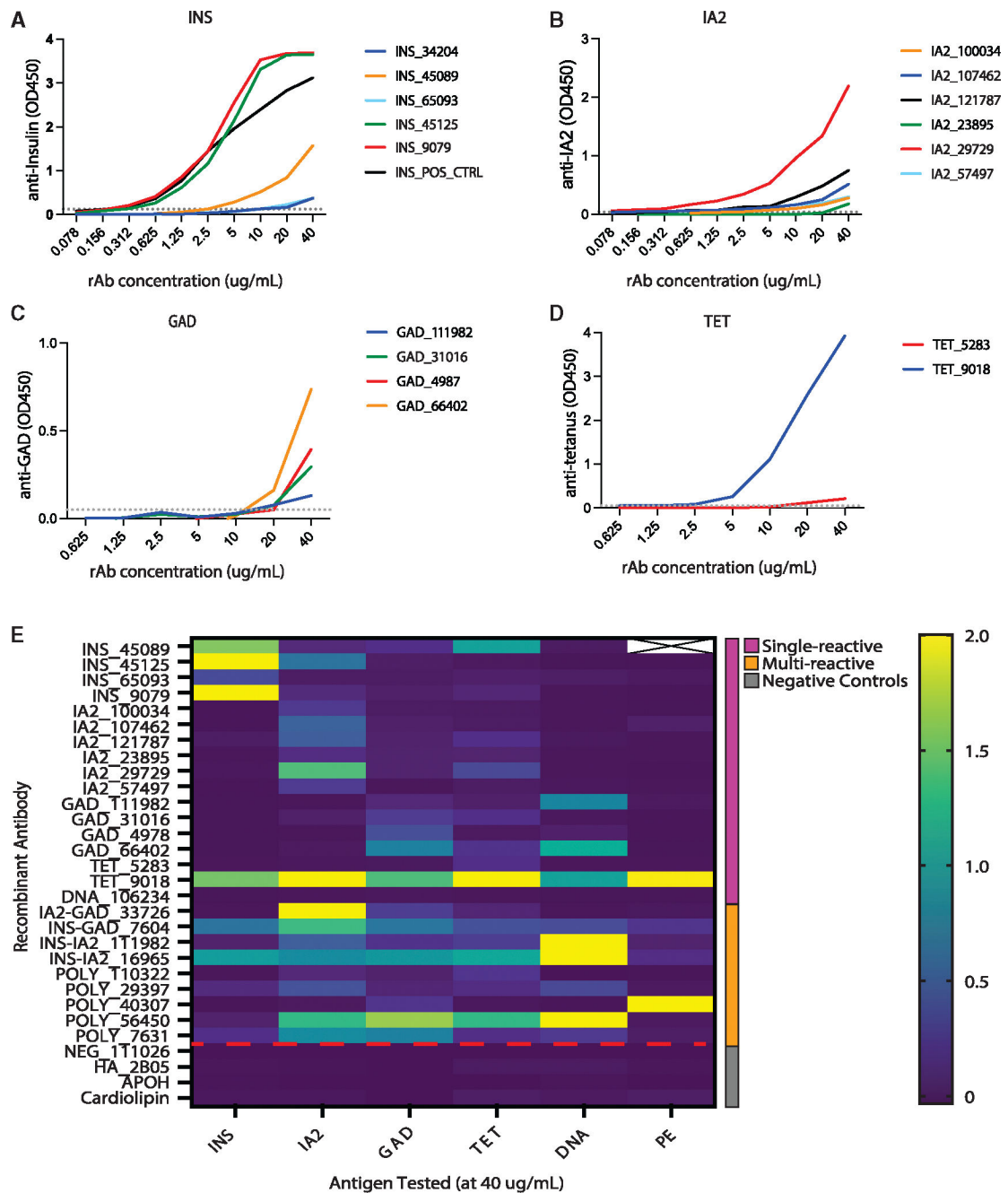
Figure 7. Antigen-reactive BCR V(D)J sequences were confirmed by ELISA

(A–D) Sequences identified from IAR B cells in AAB and T1D donors were used to make recombinant monoclonal antibodies, which were tested by ELISA against their identified antigen. One positive control INS-reactive sequence was tested alongside our other experimentally indicated INS-reactive sequences. Plots show OD450 values of three technical replicates at each antibody concentration. A gray dashed line indicates the level of assay background for each antigen tested.

(E) Along with the single-reactive antibodies tested in (A)–(D), an additional nine sequences identified from Multi-islet antigen-reactive B cells in AAB and T1D donors had

recombinant antibodies made. These were tested by ELISA against all experimental antigens present, including the PE and DNA molecules that were present in the tetramer probes. Negative control antibodies are included beneath the red dashed line. An antibody sequence negative for any antigen binding from a donor in the experimental cohort was tested along with three additional monoclonal antibodies with known specificities for influenza HA, human APOH, and human cardiolipin. This heatmap represents OD450 values for each antibody's reactivity to each given antigen. OD450 values 2.0 are yellow. Molarity value conversions are as follows: 40 μg/mL = 267 μM, 20 mg/mL = 133.5 μM, 10 μg/mL = 66.75 μM, 5 μg/mL = 33.38 μM, 2.5 μg/mL = 16.69 μM, 1.25 μg/mL = 8.34 μM, 0.625 μg/mL = 4.17 μM, 0.312 μg/mL = 2.09 μM, 0.156 μg/mL = 1.04 μM, and 0.078 μg/mL = 0.52 μM. See also Figures S7 and S8.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-human CD19-BV421	BioLegend	Cat#363017; RRID:AB_2564226
anti-human CD3-FITC	BioLegend	Cat#317305; RRID:AB_571906
anti-human IgM	BioLegend	Cat#314547; RRID:AB_2800835
anti-human IgD	BioLegend	Cat#348245; RRID:AB_2810553
anti-human CD27	BioLegend	Cat#302853; RRID:AB_2800747
anti-human CD21	BioLegend	Cat#354923; RRID:AB_2800953
anti-human CXCR5	BioLegend	Cat#356939; RRID:AB_2800968
goat anti human IgG (H + L) HRP	Invitrogen	Cat#31412; RRID:AB_228265
Biological samples		
PBMCs from autoantibody positive non-diabetic, new-onsetT1D, and autoantibody negative first-degree relative controls	Subjects from the Barbara Davis Center	N/A
Chemicals, peptides, and recombinant proteins		
BP Clonase	Invitrogen	Cat#56481
LR Clonase	Invitrogen	Cat#56484
IPTG	Millipore Sigma	Cat#367-93-1
NiNTA Agarose Beads	Qiagen	Cat#1018244
Human recombinant insulin	Millipore Sigma	Cat#19278
Human recombinant GAD65	Diamyd Medical	Cat#45-08029-01
Tetanus Toxoid	Colorado Serum Company	Cat#11415
PE	Invitrogen	Cat#P801
TMB Solution	Life Technologies	Cat#002023
TotalSeq™-C0961 PE Streptavidin	BioLegend	Cat#405155
TotalSeq™-C0951 PE Streptavidin	BioLegend	Cat#405261
TotalSeq™-C0954 PE Streptavidin	BioLegend	Cat#405267
TotalSeq™-C09653 PE Streptavidin	BioLegend	Cat#405265
TotalSeq™-C0952 PE Streptavidin	BioLegend	Cat#405263
TotalSeq™-C0955 PE Streptavidin	BioLegend	Cat#405269
Human Anti-cardiolipin monoclonal antibody, clone D24G	Creative Diagnostics	Cat#CABT-L6468Z

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse Anti-Human APOH monoclonal antibody, clone I320	Creative Diagnostics	Cat# CABT-L6552Z
Critical commercial assays		
DNA Midi Kit	Qiagen	Cat#51185
EZ-Link Sulfo-NHS-LC-Biotin Kit	ThermoFisher	Cat#A39257
7K Zeba Column	ThermoFisher	Cat#89882
Amicon 3kDa column	Millipore Sigma	Cat#UFC500396
LIVE/DEAD™ Fixable Near-IR Dead Cell Stain Kit	ThermoFisher	Cat#L34975
Chromium Next GEM Single Cell 5′ Kit v2	10X Genomics	Cat#1000263
Library Construction Kit	10X Genomics	Cat#1000190
Chromium Single Cell Human BCR Amplification Kit	10X Genomics	Cat#1000253
Chromium Next GEM Chip K Single Cell Kit	10X Genomics	Cat#1000287
Feature Barcode Kit Primer 4	10X Genomics	Cat# 2000277
Qubit dsDNA Quantification Assay	ThermoFisher	Cat# Q32851
High Sensitivity D1000 TapeStation Screentape	Agilent	Cat# 5067-5584
High Sensitivity D1000 TapeStation Reagents	Agilent	Cat# 5067-5585
Epstein-Barr virus/VCA IgG ELISA Kit	QED Bioscience	Cat#ESR1361G
Deposited data		
Sequencing files	This paper	GEO: GSE270142
Software and algorithms		
Original code	This paper	https://github.com/CUAnschutzBDC/smith_nicholas_sc_antigen
Rv4.2.3	CRAN	https://cran.r-project.org/
10x Genomics Cellranger v 7.1.0	10X Genomics	https://www.10xgenomics.com/support/software/cell-ranger/downloads
Seurat v4.3.0	Hao et al.[29]	https://satijalab.org/seurat/
DoubletFinder	McGinnis et al.[68]	https://github.com/chris-mcginnis-ucsf/DoubletFinder
Clustifyr	Fu et al.[69]	https://github.com/rnabioco/clustifyr
Scar	Sheng et al.[70]	https://github.com/Novartis/scAR
Harmony	Korsunsky et al.[71]	https://github.com/immunogenomics/harmony
Scran (mm)	Haghverdi et al.[30]	https://bioconductor.org/packages/release/bioc/html/scran.html
djvdj	Sheridan[72]	https://rnabioco.github.io/djvdj/
Change-O	Gupta et al.[73]	https://changeo.readthedocs.io/en/stable/

REAGENT or RESOURCE	SOURCE	IDENTIFIER
gProfiler	Kolberg et al.[74]	https://biit.cs.ut.ee/gprofiler/gost
scAnalysisR	Wells[75,76]	https://github.com/CUAnschutzBDC/scAnalysisR
KEGGREST	Tenenbaum[77]	https://bioconductor.org/packages/KEGGREST
T1K	Song[78]	https://github.com/mourisl/T1K
AnnotationDbi	Pagès[79]	https://bioconductor.org/packages/AnnotationDbi
Snakemake v. 6.0.3	Koster et al.[80]	https://snakemake.readthedocs.io/en/stable/
Docker container for running scar	This paper	https://hub.docker.com/r/kwellswrasman/smith_2024_scar
Docker container for running dropkick	This paper	https://hub.docker.com/r/kwellswrasman/smith_2024_dropkick
Docker container for running T1K	This paper	https://hub.docker.com/r/kwellswrasman/smith_2024_t1k
Docker container for running all R analysis	This paper	https://hub.docker.com/r/kwellswrasman/smith_2024_t1k