Title: Single-cell Landscape Analysis of the Human Circulating B Cell Pool under Selective Pressure of Allogeneic Stem Cell Transplantation

Authors: Jonathan C. Poe[1†], Jiyuan Fang[2†], Dadong Zhang[3†], Marissa R. Lee[2], Rachel A. DiCioccio[1], Hsuan Su[1], Xiaodi Qin[3], Jennifer Zhang[4], Jonathan Visentin[1,5,6], Sonali J. Bracken[7], Vincent T. Ho[8], Kathy S. Wang[8], Jeremy J. Rose[9], Steven Z. Pavletic[9], Frances T. Hakim[9], Wei Jia[1], Amy N. Suthers[1], Itaevia Curry-Chisolm[1], Mitchell E. Horwitz[1,3], David A. Rizzieri[1,3], William McManigle[1], Nelson J. Chao[1,3,10], Adela R. Cardones[4], Jichun Xie[2,3], Kouros Owzar[2,3], and Stefanie Sarantopoulos[1,3,10]*

Affiliations:

[1]Department of Medicine, Division of Hematological Malignancies and Cellular Therapy, Duke University Medical Center; Durham, NC, USA.

[2]Department of Biostatistics and Bioinformatics, Duke University Medical Center; Durham, NC, USA.

[3]Duke Cancer Institute, Duke University Medical Center; Durham, NC, USA.

[4]Department of Dermatology, Duke University Medical Center; Durham, NC, USA.

[5]Department of Immunology and Immunogenetics, Bordeaux University Hospital; Bordeaux, France.

[6]UMR CNRS 5164 ImmunoConcEpT, Bordeaux University; Bordeaux, France.

[7]Department of Medicine, Division of Rheumatology and Immunology, Duke University Medical Center; Durham, NC, USA.

[8]Division of Hematologic Malignancies and Department of Medical Oncology, Dana-Farber Cancer Institute; Boston, MA, USA.

[9]Experimental Transplantation and Immunology Branch, National Cancer Institute; Bethesda, MD, USA.

[10]Department of Immunology, Duke University Medical Center; Durham, NC, USA.

*Corresponding author. Email: stefanie.sarantopoulos@duke.edu

[†]Equal contribution to this work

ABSTRACT

A healthy immune state is achieved and maintained only if autoreactive B cells are eliminated or tolerized. After allogeneic hematopoietic stem cell transplantation (allo-HCT), donor-derived B cells develop under selective pressure from host alloantigens. Likewise, B cells play a substantiated role in the pathobiology of the allo-HCT-related autoimmune-like syndrome, chronic graft versus host disease (cGVHD). Allo-HCT thus represents a unique opportunity to study human peripheral B cell tolerance, since stem cell reconstitution occurs continuously in an alloantigen environment. Herein, we performed scRNA-Seq analysis on purified B cells from allo-HCT patients and resolved 10 clusters distinguishable by signature genes for maturation, activation and memory. Compared to single-cell RNA-Seq data from healthy or infected individuals, allo-HCT patients had striking transcriptional differences in the 'memory' B cell compartment that suggested a propensity toward autoreactivity. To identify B-cell intrinsic properties linked to overt loss of immune tolerance in allo-HCT, we then assessed all 10 clusters for differentially expressed genes (DEGs) between patients with vs. without autoimmune-like manifestations (Active cGVHD vs. No cGVHD). In Active cGVHD, notable DEGs occurred in both naïve and BCR-activated clusters, including a BCR-experienced cluster harboring a significantly expanded population of potentially pathologic, atypical/age-related memory B cells (ABCs). Pseudotime trajectory analyses plus use of a 'maturation' agent in vitro also indicated that aberrant B-cell activation and diversification states after allo-HCT are potentially 'plastic' and amenable to pharmacological manipulation. Our single-cell-level findings reveal potentially novel targets in cGVHD and provide new understanding of how alloreactivity may lead to human autoimmune diseases.

One Sentence Summary: We elucidate B cell subsets and DEGs at the single-cell level in humans receiving allo-HCT, when immune tolerance is lost or maintained.

Main Text:

INTRODUCTION

Operational B cell tolerance checkpoints normally function to silence the high proportion of potentially self-reactive peripheral B cells that recirculate and can mediate immune pathology (*1*). In the condition termed chronic graft-versus-host disease (cGVHD), a syndrome acquired by some, but not all patients undergoing allogeneic hematopoietic stem cell transplantation (allo-HCT), B cells have a substantial role (*2, 3*). Aberrant peripheral B cell activation, survival, and maturation signaling pathways are thus promising therapeutic targets for cGVHD patients (*4*). Regulatory B cells (Breg) also appear to be important in controlling the severity of cGVHD in both humans and mice (*5-7*). B cell recovery and homeostasis after allo-HCT are both altered in cGVHD, whereby extrinsic factors including BAFF (B cell-activating factor belonging to the tumor necrosis factor [TNF] family) contribute to a break in B cell tolerance (*8-13*).

B cells continually regenerate. After allo-HCT, B cells from donor stem cells and progenitor cells that are major histocompatibility antigen (HLA)-matched or mismatched replace the recipient (host) immune system. Prior to HCT, high dose chemotherapy and/or radiation conditioning creates "space" in the B cell compartment whereby excess BAFF promotes antigen-activated B cells (*12*) that regenerate daily. In a mouse model, alloantigen and BAFF appear to operate together to promote the loss of B cell tolerance that occurs in cGVHD (*9*). BAFF is also known to mediate B-cell driven autoimmune and fibrotic manifestations akin to findings in

cGVHD (*14*). After allo-HCT donor T cells recognize host polymorphic or alloantigens in a coordinated T-B response (*15*), and follicular helper T cells are required for B cell anti-host reactivity and cGVHD development (*16*). The peripheral B cell compartment in patients with active disease (Active cGVHD) becomes enriched for B-cell receptor (BCR)-stimulated and IgG-secreting populations (*2, 3, 8, 10, 17-20*). Alloreactivity has been shown to beget autoreactivity after allo-HCT in mice (*16*), but how constant exposure to alloantigens and high BAFF alter intrinsic B cell pathways in human allo-HCT remains unknown.

After allo-HCT, compared to patients without autoimmune manifestations, those with Active cGVHD have a circulating B cell compartment with a propensity toward activation (*11, 21*) and potential 'exhaustion' (*22*), yet intricately defining the pleomorphic outcomes on B cell subset diversity resulting from cGVHD genesis remains elusive. Human allo-HCT and the advent of single-cell RNA-sequencing (scRNA-Seq) now enables identification of potential intrinsic B cell pathways that potentially underpin pathological functions. We utilized scRNA-Seq to delineate transcriptional programs within 10 peripheral B cell clusters resolved in allo-HCT patients. We also identified important differentially expressed genes (DEGs) within these clusters that were associated with loss of B cell tolerance in Active cGVHD. Some DEGs were spread across these B cell populations, while others were restricted to certain B cell subsets. Homing and cell cycle genes including *GPR183* and *CKS2*, respectively, were overexpressed in circulating transitional, activated and memory subsets, suggesting potential roles in B cell dysfunction from the time B cells first emerge in the periphery, and then after alloantigen BCR stimulation. We found that after allo-HCT, the 'memory' cluster contained potentially-pathogenic 'atypical/age-related' B cell (ABC) populations found in both health and autoimmune disease (*23, 24*), and drivers of ABCs including *TBX21* (*TBET)* and *ZEB2* (*23, 25, 26*) were increased DEGs within this

cluster. We further confirmed significant expansion of ABC subsets in Active cGVHD using high-dimensional flow cytometry analysis. Finally, comparison of our allo-HCT scRNA-Seq data with a publically-available dataset on blood B cells from healthy donors and non-HCT patient groups highlighted distinct signature gene profiles within the ABC and typical memory subsets unique to the allo-HCT setting. Thus, our findings provide new insight into the diversity of the peripheral B cell compartment after allo-HCT. Our study also identifies potential molecular targets for future study that may effectively alleviate cGVHD and other B cell-mediated autoimmune diseases.

RESULTS

The circulating B cell compartment arising after allo-HCT comprises transitional, naïve, antigen-stimulated, and memory single-cell signature gene profiles

To assess molecular features of human B cells that regenerate and circulate in an alloantigen-rich environment when immune tolerance has been achieved or lost, we analyzed highly purified viable B cell samples (**Fig. 1A**) from 8 patients who were at least 10 months (mos) post allo-HCT. Four patients had Active cGVHD (median = 12.6 mos post-HCT) and four patients had no clinical history of cGVHD (No cGVHD, median = 11.5 mos post-HCT), **table S1**. Unsupervised clustering analysis of the 8-patient scRNA-Seq dataset identified 10 clusters of B cells (**Fig. 1B**). To confirm cell types, inferences were made by examining the relative expression of genes representing B cells (*PAX5*, *CD22*), and rare residual T cells (*CD3E*) and monocytes (*LYZ*) (**fig. S1**). All 10 B cell clusters were represented in both Active and No cGVHD patient groups (**Fig. 1C**). We numbered these clusters 1-10 from largest to smallest total number of cells in both patient groups combined. (**Fig. 1D**).

To validate known and potentially novel B cell subsets in allo-HCT patients, we first identified hallmark 'signature' (marker) genes that distinguish each B cell cluster from the others based on maturation or activation status. We then plotted the log2 normalized expression values for 16 major B cell development and function genes from all 8 transplant samples (**Fig. 1, E-H**), all validated using the function 'findMarkers' in the R package Seurat (*27*). Within clusters, all signature genes assessed were remarkably homogeneous among all 8 allo-HCT patients. Levels of the transitional B cell molecules *VPREB3*, *CD24* and *IRF8* were highest in Clusters 1 and 2 (**Fig. 1E**), with Cluster 1 primarily expressing kappa Ig light chain and Cluster 2 primarily expressing lambda Ig light chain (*IGKC* and *IGLC3*, respectively, **Fig. 1F**). High expression of *LTB* further supports that B cell Clusters 1 and 2 represent recent bone marrow (BM) emigrants because lymphotoxin beta is required for initial follicle development in the spleen (*28*) and expressed highest in late transitional B cells (immgen.org). Clusters 1, 2 and 3 expressed the highest levels of the IgM heavy chain gene (*IGHM*). Cluster 3 likewise showed high *IGKC* expression and intermediate *IGLC3* expression (**Fig. 1F**). Unlike Clusters 1 and 2, Cluster 3 had lower expression of *CD24* and *IRF8*, and higher *IRF4*, suggesting a more mature B cell stage. Clusters 9 and 10 expressed similarly high levels of microRNA-155 (*MIR155HG*, **Fig. 1G**), which is essential for B cell antibody isotype switching (*29*). Clusters 9 and 10 also showed similarly high expression of both *MYC* and *NFKB1* transcripts, suggesting these B cells may be poised to enter the cell cycle. Cluster 10 had uniquely high levels of a T cell-attracting chemokine produced by B cells, *CCL22*, indicating that Cluster 10 may harbor B cells poised to interact with T-follicular helper cells when localized to the secondary lymphoid organ (SLO) microenvironment. By contrast, Cluster 5 expressed high levels of the memory B cell marker *CD27*, along with Ig heavy chain genes for IgG1 and IgG3 (*IGHG1* and *IGHG3*, respectively), indicating an enrichment for isotype-switched

memory B cells (**Fig. 1H**). Cluster 5 also expressed the highest level of *ITGAX* (*CD11C*), a distinguishing marker of ABCs found in normal individuals and expanded during infections and in some autoimmune diseases (*30, 31*). In **Fig 1I**, we depict 3-dimensional (3-D) UMAP plots of the clustering to further highlight the spatial relationships among neighboring B cell clusters that reflect shared signature gene composition.

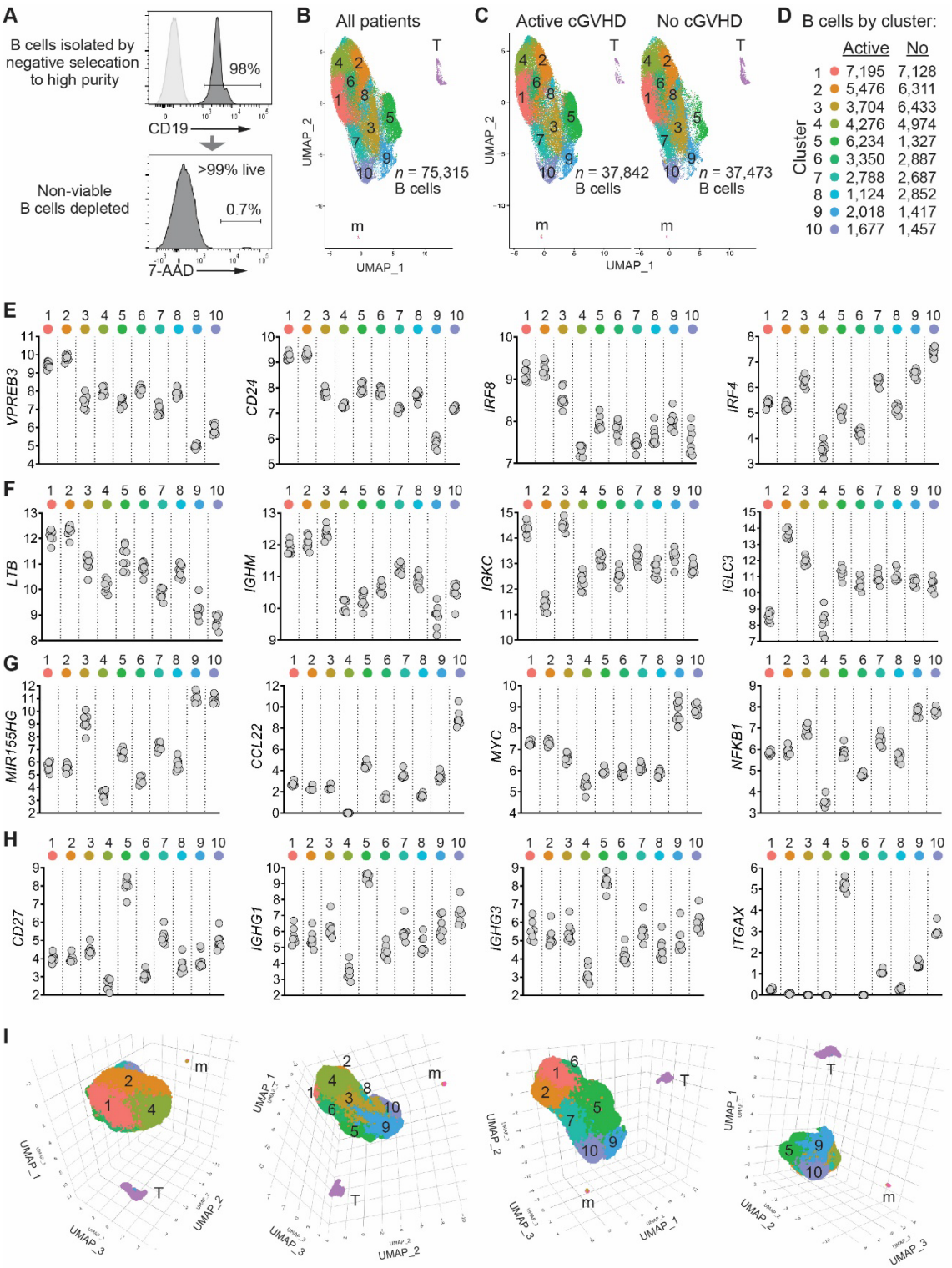


Fig. 1. Unsupervised clustering analysis of single-cell (sc)RNA-Seq data reveals multiple B cell subsets characterized by signature genes in allo-HCT patients. (**A**) Representative flow cytometry histograms showing B cell purity and efficiency of dead cell removal among PBMCs from an allo-HCT patient. 10,000 high-quality B cells per sample, isolated in the same manner from each of 8 allo-HCT patients, were targeted for 10X Genomics single-cell library construction

(*n*=4 No cGVHD, *n*=4 Active cGVHD). (**B**) Two-dimensional (2D) Uniform Manifold Approximation and Projection (UMAP) of expression profiles of all untreated cells from the 8 allo-HCT patients. Numbers indicate each of the 10 major clusters identified as B cells, which were distinct spatially from the small clusters of residual cells identified as T cells (T) and monocytes (m). Total high-quality B cells in the analysis are indicated. (**C**) UMAP as in panel (B) but with B cells partitioned by allo-HCT patient group. Total high-quality B cells per group are indicated. (**D**) Number of B cells mapping to each of the 10 B cell clusters by patient group. **(E-H)**, Log2 normalized expression of genes indicative of B cell maturity (E, F), along with activation, antibody production potential, and memory markers (G, H), shown by B cell cluster. Each symbol (gray circle) represents regularized log-transformed gene counts (*32*) aggregated to patient level (y-axes) from one of the 8 total allo-HCT patients assessed. **I**, Three-dimensional (3D) UMAP projections of B cells from all patients [as in (B)], viewed at different angles of rotation.

Allo-HCT patient B cell clusters can be delineated by expression of signature genes involved in different B cell states and pathways

Based on the signature gene results in **Fig. 1E-I**, along with the data that is presented in **Fig. 2**, we reordered the 10 B cell clusters empirically according to maturation, activation and memory status. We started with the least mature/activated first, based on general knowledge about B cells as they enter the periphery and eventually encounter antigen/co-stimulatory signals. The following cluster order was thus used from **Fig. 2** onward: Cluster 1, 2, 4, 6, 8, 7, 3, 9, 10 and 5. In **Fig. 2A,C-H** the colored squares in the heat maps indicate that signature genes of interest (rows) reached significance (P_{adj} <0.05) in the B cell cluster indicated (columns), as weighted against all other clusters. Red hues indicate signature gene expression increase, and blue hues indicate signature gene expression decrease, with the intensity of each color representing relative expression level (log FC, indicated by scales). Notably, Clusters 1, 2, 4, 6 and 8 retained one or more hallmark genes of B cell immaturity, while Clusters 7, 3, 9, 10 and 5 expressed markers of antigen activation, cytoskeletal activity, or the capacity to produce antibodies (**Fig. 2A,B**). This ordering of the 10 B cell clusters was further validated and characterized by interrogating the Kyoto Encyclopedia of Genes and Genomes [KEGG] (*33*) and Gene Ontology [GO] (*34*) databases using signature genes

identified as significant in one or more clusters. As shown in the **Fig. 2C-H** heat maps, we grouped these signature genes into major molecular pathways. Notably, Clusters 9 and 10 showed high expression of a multitude of genes involved in B cell–T cell interactions, transcription, proliferation, survival, and metabolic processes across all heat maps (**Fig. 2A,C-H**). This corresponded with their high relative expression of *MYC* and *NFKB1* (**Fig. 1G**). Clusters 7 and 10 shared increases in multiple ribosomal genes (**Fig. 2G**), suggesting increased protein synthesis potential and a relative relatedness between these clusters. Overall, clusters 6 and 8 had relatively fewer gene level increases, suggesting relative quiescence, although Cluster 6 had strong expression of actin cytoskeleton organizer *CDC42SE1* and the proto-oncogene *MENT* (previously *C1orf56,* **Fig. 2F**). Interestingly, the less mature/less activated subsets (Clusters 1, 2, 4 and 6) tended to have a greater number of decreased signature genes ('Down', **Fig. 2I**), while the more activated/memory subsets (Clusters 3, 9, 10 and 5) tended to have a greater number of increased signature genes ('Up', **Fig. 2I**). Together these data define a hierarchy and relatedness among the peripheral B cell clusters identified in this unsupervised clustering analysis of allo-HCT patients.

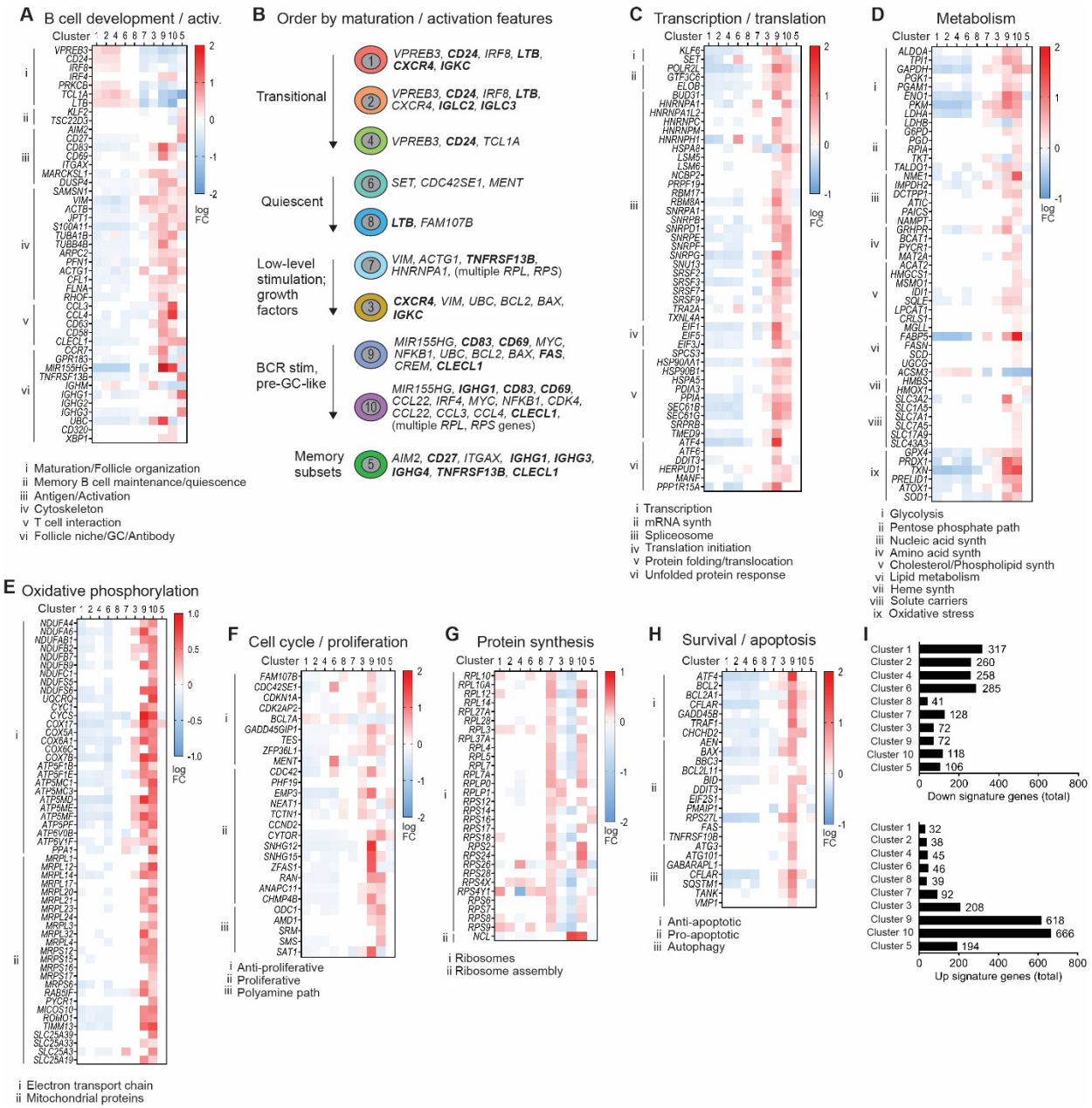


Fig. 2. Signature genes manually assigned to known biological pathways further highlight allo-HCT B cell cluster inter-relatedness and features of B cell development, activation, proliferative potential and survival. (**A**) Key signature (marker) genes related to B cell development and activation, assessed in an inter-cluster expression comparison of B cells from all 8 allo-HCT patients in the scRNA-Seq data set by manually interrogating the KEGG and GO databases. Each colored square in the heat maps represents a significant (P_{adj} <0.05) log fold change (log FC) value for the gene indicated (rows), in the B cell cluster indicated (columns). Red shading indicates increased signature gene expression within the cluster of interest, while blue shading indicates decreased signature gene expression within the cluster of interest, as weighted against all other clusters. Clusters were ordered empirically from the least differentiated (Cluster 1, Igκ$^+$ transitional) to most differentiated (Cluster 5, memory) as described in the Results section.

Genes were further subdivided to one of the specific pathways indicated by the roman numerals at left and corresponding key below. (**B**), Cartoon depiction of B cell clusters and their associated signature genes that help to further categorize and group these subsets (as labeled at left). Signature genes encoding surface proteins are highlighted in bold font. (**C-H**) As performed in (A), signature genes were assessed in the major biological pathways indicated above each heat map, and subdivided by the more specific pathways indicated by the roman numerals at left and corresponding keys below. (**I**) Total numbers of signature genes reaching significance within each cluster, being either decreased ('Down signature genes', top graph) or increased ('Up signature genes', bottom graph).

The BCR-experienced 'memory' Cluster 5 includes an extrafollicular-like ABC population that is expanded in Active cGVHD

Since antigen-experienced 'memory' B cells likely include pathogenic B cells that emerge in cGVHD, as occurs in other diseases such as lupus (*1, 35*), we interrogated the CD27 expressing 'memory' B cell compartment identified in Cluster 5. In addition to a CD27 expressing population, Cluster 5 also harbored B cells expressing *ITGAX* that encodes CD11c which broadly marks ABC populations that also lack CD21 ($CD11c^{+}CD21^{-}$). ABCs are proposed to have a pathogenic role in both autoimmunity and chronic inflammation, revealing a functional complexity of these 'memory' B cells (*30, 36, 37*). ABCs are often termed either 'atypical' or 'age-related' B cells. The latter term was originally given because ABCs tend to accumulate in aged mice (*38, 39*). ABCs are considered 'atypical' because they undergo activation and differentiation outside of the germinal center (GC), rapidly homing to the outer follicle or leaving the follicle as 'extrafollicular' B cells (*40-43*). To elucidate apparent 'memory' sub-populations in cGVHD, we first examined the pattern of *CD27* and *ITGAX* expression in Cluster 5. As shown in **Fig. 3A,** B cells expressing either *CD27* (left UMAPs) or *ITGAX* (right UMAPs) were largely spatially segregated within Cluster 5, suggesting most B cells don't co-express transcripts for these genes. To validate this finding, we performed flow cytometry analysis on blood B cells from an independent cohort of

allo-HCT patients. As shown in **Fig. 3B**, allo-HCT B cell subsets primarily expressed either CD27 or CD11c alone, with a smaller subset expressing both markers (double-positive, 'DP'). Importantly, CD11c^{+}CD27^{-} ABCs were significantly expanded in Active cGVHD patients compared to No cGVHD patients and healthy donors (HDs) (**Fig. 3B,C**). Interestingly, we found that the proportion of the CD11c^{+}CD27^{+} 'DP' subset was significantly lower in the allo-HCT environment compared to HDs (**Fig. 3B,C**). Overall these data are consistent with previous work showing that some CD11c^{+}CD27^{-} ABC subsets are expanded in cGVHD and other diseases, and can exhibit an 'exhausted' phenotype or be refractory to certain external stimuli (*22, 44-46*). To gain further insight into known ABC subsets, we also examined a subset of CD11c^{+}CD21^{-} ABCs that lacks both CD27 and IgD expression (CD11c^{+}CD21^{-}CD27^{-}IgD^{-}, called 'DN2'), and is prevalent in lupus and can produce disease-associated antibodies (*42, 47, 48*). The emergence of functional ABCs (including DN2) with pathogenic roles is driven by one or more costimulatory signals that cooperates with BCR stimulation (*42, 49*). We found DN2 ABCs were markedly expanded in the blood of Active cGVHD patients relative to No cGVHD patients (**Fig. 3D,E**). These data affirm our scRNA-seq data and advance previous work (*22, 42, 46-48*) suggesting that potentially pathologic ABC subsets arise under constant alloantigen exposure.

To assess B cell surface phenotypes more broadly in allo-HCT and potentially elucidate subsets particularly responsive to excess BAFF in cGVHD (*12*), we added CD24 and TACI (TNFRSF13B) in the context of CD27, IgD, CD11c, CD21 described above (see antibody panel, **Fig. 3F**). Using high-dimensional flow cytometry [PhenoGraph, (*50, 51*)] we identified discreet cell populations in an unbiased manner from concatenated allo-HCT sample groups. The PhenoGraph analysis resolved 15 clusters, designated by lower case letters 'a-o' (**Fig. 3G**). CD11c^{+} ABCs were primarily resolved by Cluster 'a' (**fig. S2**), which was significantly expanded

in Active cGVHD patients (**Fig. 3H,I**). By contrast, four clusters consisting of $IgD^{+}CD27^{-}$ naïve-like B cells (Clusters 'b, c, d and e', **fig. S2**) were significantly proportionally decreased in Active cGVHD patients (**Fig. 3H**). We found that ABC $CD11c^{+}CD27^{-}CD21^{-}CD24^{-}$ Cluster 'a' was $TACI^{+}$, with a mix of $IgD^{+/-}$ B cells (**fig. S2**). Six clusters expressed high levels of surface TACI (Clusters 'a, h, j, k, l and n', **Fig. 3J and fig. S2**). Cluster 'n' expressed substantially higher surface TACI in some Active cGVHD patients (**Fig. 3J**, arrow), and mapped adjacent to ABC Cluster 'a' in the PhenoGraph UMAP (**Fig. 3G).** Clusters 'j' and 'k' likewise had elevated TACI in some Active cGVHD patients (**Fig. 3J**, arrows), although to a lesser extent than Cluster 'n'. Notably, the frequency of B cells positive for *TNFRSF13B (TACI)* transcripts was also higher in Active cGVHD patients in the scRNA-Seq analysis, apparent within specific regions of the UMAP plots (**Fig. 3K**, boxed region II). Together these results support previous observations that BAFF is a likely driver of B cell hyper-responsiveness and alloantibody production in cGVHD (*9*).

To examine how this allo-HCT finding compared to the non-HCT setting, we performed a PhenoGraph analysis of HD blood B cells. The HD analysis also predicted 15 clusters, but HDs had a greater number of B cell clusters marking with CD27 (**fig. S3**, Clusters 'm,n,o') compared to allo-HCT patients (**fig. S2**, Clusters 'l,o' only). Furthermore, in allo-HCT patients Cluster 'o' represented mostly $CD27^{bright}$ B cells that lacked CD21, CD24, IgD and CD11c expression, suggestive of isotype-switched plasmablasts ("PB", **fig. S2**). These were among the rarest in both allo-HCT groups (**Fig. 3H**), and this cluster was notably absent in HDs (**fig. S3**). These data reveal that $CD27^{+}$ B subsets after allo-HCT are abnormal or lacking, extending previous observations (*44*). The relative reduction in $CD27^{+}$ B cells in allo-HCT with reciprocally-expanded ABCs in Active cGVHD is consistent with PhenoGraph findings recently reported in pediatric patients with CVID and related syndromes (*52*).

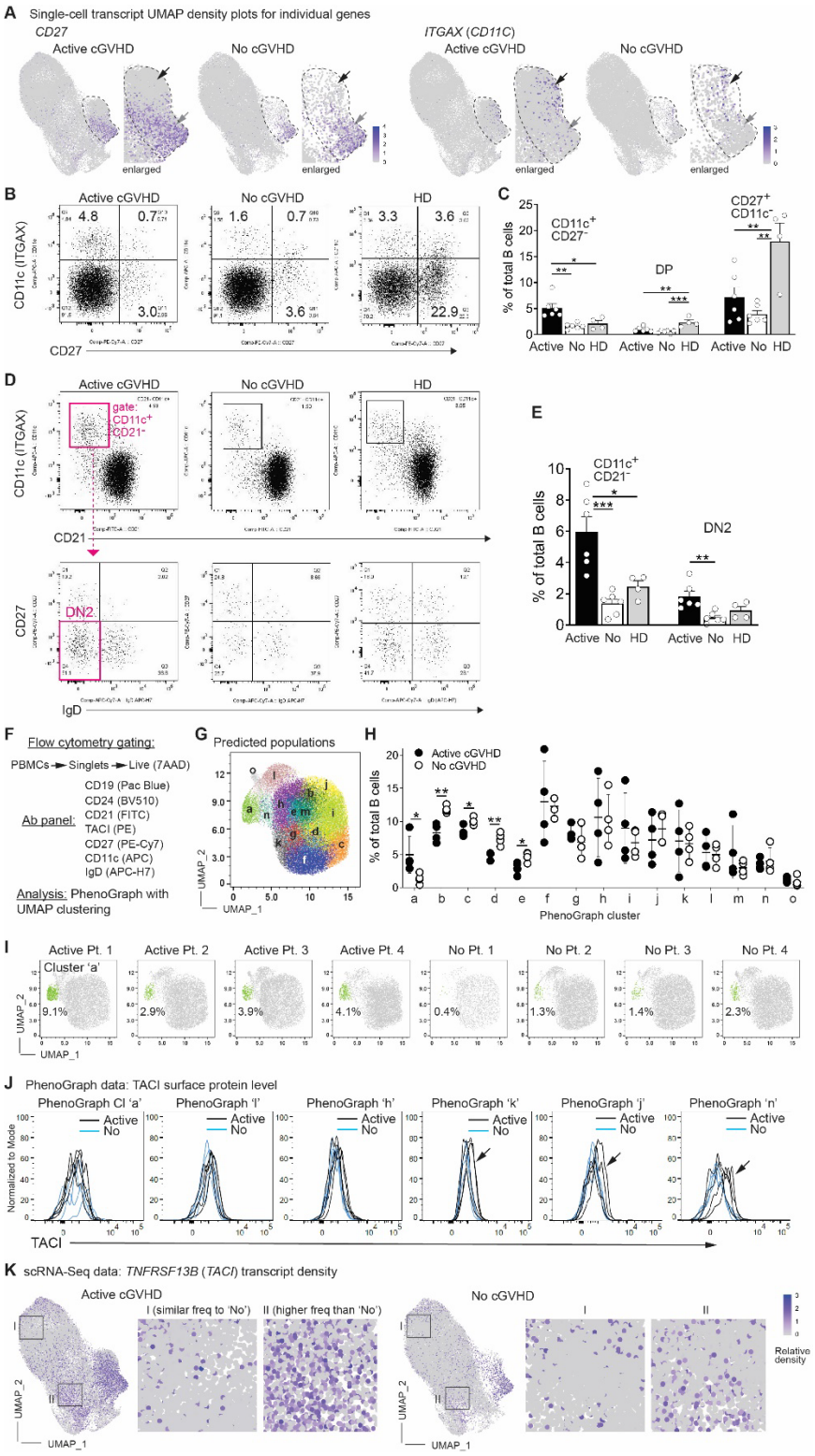


Fig. 3. Cluster 5 contains both CD27$^+$ 'memory' B cells and ABCs that are expanded in cGVHD. (**A**) B cells expressing *CD27* or *ITGAX* transcripts are enriched in Cluster 5. Shown are UMAP density plots from the scRNA-Seq dataset displaying relative transcript density for the gene indicated, at the single-cell level. The dashed line approximates the perimeter of Cluster 5

(also enlarged for better clarity). Gray arrows indicate a region within Cluster 5 enriched for *CD27* expressing B cells, and black arrows indicate a region within Cluster 5 enriched for *ITGAX* expressing B cells. (**B,C**), Flow cytometric analysis for the surface proteins shown to identify ABCs, $CD27^+$ memory B cells, and a small population of $CD11c^+CD27^+$ memory B cells in viably cryopreserved PBMC samples from a cohort of allo-HCT patients with either Active cGVHD (*n*=6) or No cGVHD (*n*=6), or from HDs (*n*=4). PBMCs were pre-gated on live (7-AAD^-) $CD19^+$ B cells. Dot plots in (B) show representative individuals analyzed from each group. In (C), results from all groups for the $CD11c^+CD27^-$ population, $CD11c^+CD27^+$ (DP) population, and $CD11c^-$ $CD27^+$ population (as gated in (B)) are represented. Statistical comparison was performed using a one-way ANOVA with Tukey's multiple comparisons test (GraphPad Prism 9 software; *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$). (**D,E**), DN2 B cells are expanded in allo-HCT patients with Active cGVHD. (D) Representative FACS plots from PBMCs of allo-HCT patients with Active or No cGVHD, and from a HD. Plots were generated after first gating on viable (7-AAD^-), $CD19^+$ B cells, with $CD11c^+CD21^-$ B cells (upper panel gate) and then DN2 B cells ($CD11c^+CD21^-CD27^-$ IgD^-) identified in all groups as depicted in magenta for the Active cGVHD sample. (E) Statistical comparison between groups (Active cGVHD, n=6; No cGVHD n=6, HDs, n=4) was performed using a one-way ANOVA with Tukey's multiple comparisons test (GraphPad Prism 9 software; **, $p<0.01$). (**F-J**) Flow cytometry plus PhenoGraph analysis performed on blood B cells from Active cGVHD (*n*=4) or No cGVHD (*n*=4) patients. Shown in (F) is the gating strategy to distinguish B cells (top line), the B cell antibody panel used (middle), and the analysis platform (bottom). Shown in (G) is the concatenated PhenoGraph UMAP plot from all 8 patients, with 15 clusters identified (letters). The graph in (H) indicates B cell frequency within each of the 15 clusters by patient group. Each symbol represents results from one of the 8 patients assessed. Statistical comparisons were performed using a two-tailed, unpaired t-test (GraphPad Prism 9 software; *, $p<0.05$; **, $p<0.01$). The PhenoGraph UMAP plots in (I) show the position (green) and frequency of B cells in Cluster 'a', which represents a population of ABCs expanded in Active cGVHD patients that are primarily $CD21^-$, $CD24^-$, $IgD^{+/-}$, $CD27^-$, $CD11c^+$ and TACI bright ($TACI^{Br}$, **fig. S2**). In (J), histogram overlays for TACI surface protein expression on B cells in the six clusters from the PhenoGraph assessment described as being designated $TACI^{Br}$ (**fig. S2**), from each of the 4 allo-HCT patients assessed. Arrows in the PhenoGraph plots for Clusters 'k', 'j' and 'n' highlight the elevated TACI expression levels observed in some Active cGVHD patients. (**K**) B cells positive for *TNFRSF13B* (*TACI*) transcripts appear elevated in Active cGVHD patients. Normalized expression of *TNFRSF13B* in UMAP space are shown, separated by patient disease group. Boxed areas in representative regions (I, II) are enlarged to show detail.

The transcriptional profile of the circulating 'memory' B cell compartment in allo-HCT patients is distinct from HDs and other non-HCT patients

Given that the post-HCT 'memory' B cell compartment had a reduction of DP ($CD27^+CD11c^+$) B cells (**Fig. 3B,C**), we next asked how memory B cell subset gene profiles following allo-HCT compared to healthy individuals, or to individuals with non-HCT chronic diseases. For these analyses, we leveraged a similar, publicly-available scRNA-Seq dataset generated from blood B cells of healthy donors (HDs, *n*=3) and non-HCT patients chronically exposed to HIV (*n*=3) or malaria (*n*=3), published by Holla et al. (*30*). We first identified total $CD27^+ITGAX^-$, $CD27^-ITGAX^+$, and $CD27^+ITGAX^+$ B cells from all patient and HD groups (described in the Methods section) to match the same subsets defined by flow cytometry (**Fig. 3B,C**). We then assessed the frequency of B cells within these Cluster 5 subsets that expressed other genes associated with the function or additional phenotype(s) of memory B cells (*53-56*). By displaying the frequency of positive B cells for all genes as a heat map (**fig. S4**), we found that for most genes, B cell frequencies were similar between all 5 groups. However, certain genes were remarkably distinct in allo-HCT patients, irrespective of cGVHD status, when compared to either HDs, HIV patients or malaria patients (**Fig. 4**). B cells expressing *ADA* [most frequently associated with a form of SCID with disrupted B cell tolerance (*57*)], the T cell-attracting chemokine *CCL22* (*58*), and 3 G-protein coupled receptors (GPCRs) important for B cell exit from the BM and/or homing to SLO niches, *CCR7*, *S1PR1*, and *S1PR2* (*59-64*), were increased in frequency across *CD27*/*ITGAX* subsets in allo-HCT patients. Interestingly, B cells expressing the adhesion molecule *ITGB2*, which pairs with various integrin alpha chains to mediate migration, and is a potential marker of ABCs (*30*), were markedly proportionally decreased in allo-HCT patients. *CD27*/*ITGAX* subsets in allo-HCT patients also had broadly increased proportions of B cells expressing survival

regulators with a known (or potential) role in memory B cell maintenance including *BCL2* (*65*) and *TRADD* (*66*). By contrast, in allo-HCT patients there was a paucity in B cells expressing *FOS*, a gene that lies at a nexus of B cell survival and proliferation during the GC response (*67, 68*) and terminal differentiation to plasma cells (*69*). This observation is consistent with our previous finding that NOTCH2 signaling represses *FOS* expression induced during BCR stimulation of Active cGVHD patient B cells, with a 'NOTCH2-BCR axis' mediating enhanced cGVHD B cell survival and proliferation (*10*). Together these data reveal that following allo-HCT the pool of memory- and ABC-like B cells possess unique intrinsic molecular properties with hallmarks suggesting altered tolerance, migration, SLO homing, and survival.

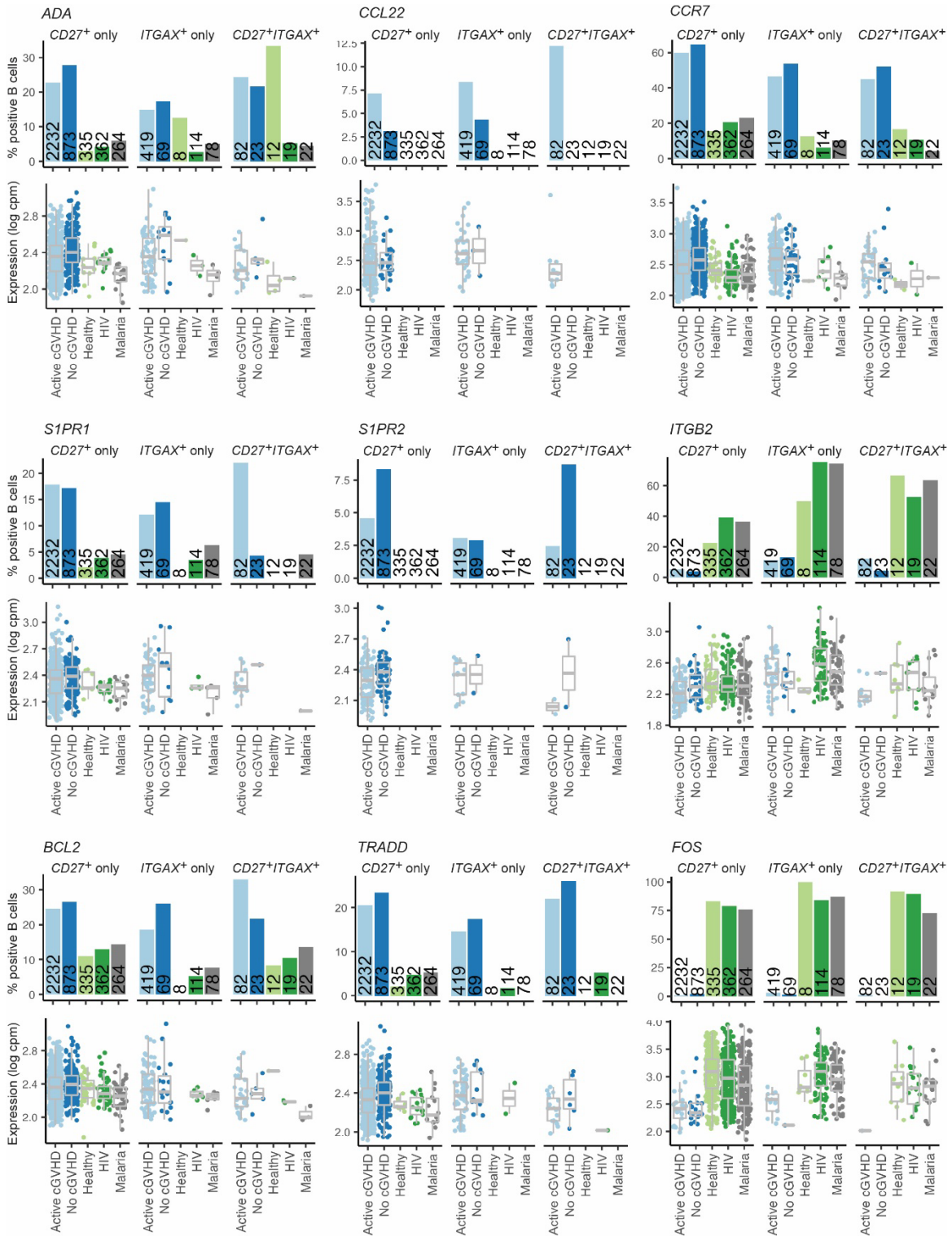


Fig. 4. 'Memory' *CD27* and *ITGAX* B cell subsets are transcriptionally unique in allo-HCT patients compared to healthy individuals and non-HCT patient populations. Shown are results for 9 'memory' B cell genes of particular interest that display differences in the frequency of positive B cells between allo-HCT patients compared to HD (Healthy) and non-HCT patient

groups (HIV, Malaria). Bar graphs at top represent the % positive B cells for the population indicated. Numbers embedded within the bars indicate the total number of B cells identified within the subset indicated, for the patient or HD group. In the gene expression graphs below each bar graph, each point represents an individual B cell, with the data normalized to show expression values for the gene of interest per million total mapped reads in the same B cell, expressed on a log scale (log CPM).

Genes critical for B cell fate and function are differentially expressed in Active cGVHD B cell subsets

Having determined common signature gene features of the allo-HCT B cell compartment, we next focused on differentially expressed genes (DEGs) between allo-HCT patient groups (Active cGVHD vs. No cGVHD). In **Fig. 5**, bar graphs represent total DEGs by cluster, either increased (Up, **Fig. 5A**) or decreased (Down, **Fig. 5B**) in Active cGVHD. Cluster 9, previously inferred to be the most metabolically activated subset after allo-HCT (**Fig. 2**) had the greatest number of Up DEGs in Active cGVHD (193). Cluster 5, previously inferred to be the memory/ABC subset (**Fig. 1H**, **Fig. 2A**), had the greatest number of Down DEGs (335). From the total list of DEGs (**table S3**), we first focused on genes with known functional roles in B cells ('Select' DEGs). As shown in the heatmaps in **Fig. 5C,D**, we identified a relatively large number of select DEGs between allo-HCT patient groups in at least one B cell cluster. We further categorized these DEGs according to their known primary roles in B cell function, including lymphoid follicle and GC organization, tolerance and survival, BCR signaling and activation, cell cycle/proliferation, isotype switching/antibody generation, or ABC development and function (**Fig. 5E**). We found some DEGs in Active cGVHD patient B cells suggested aberrant migration and homing. For example *CXCR4*, an important regulator of B cell migration from the BM to the periphery (*70*), was increased in the transitional Active cGVHD B cell clusters (Clusters 1, 2 and 4, **Fig. 5C,E**). Active cGVHD B cells also had Up DEGs for multiple other GPCRs known to

regulate B cell homing and retention within essential SLO niches. These included *GPR183* (i.e., *EBI2*), *P2RY8*, *S1PR2*, and *CXCR3* (*61, 71-75*). Of these GPCRs, *GPR183/EBI2* was notably increased across five B cell clusters in Active cGVHD (**Fig. 5C,E**), including the transitional-like subsets (Clusters 1 and 2) and the memory/ABC subset (Cluster 5). This was a potentially important finding because EBI2 has an intricate role in regulating B cell movement within the follicle between the B cell zone, the B-T cell border, into the GC reaction, and eventually to the extrafollicular space within SLOs before emerging as plasmablasts or memory B cells (*72, 75*). *P2RY8* and *S1PR2* were both Down DEGs in activated Cluster 9 and/or 'memory' Cluster 5 (**Fig. 5D,E**). This was intriguing since these molecules are also pivotal in B cell homing and confinement to follicular niches or the GC reaction (*61, 74, 76*), and a recent study strongly linked decreased *P2RY8* expression to pathogenic antibody production and expansion of both plasma cells and ABCs in humans with lupus or antiphospholipid syndrome, and in a mouse model of lupus (*77*). Thus, our DEG data suggest altered follicle and GC movement of B cells during Active cGVHD in patients, consistent with some findings in cGVHD mouse studies (*3, 18, 78*).

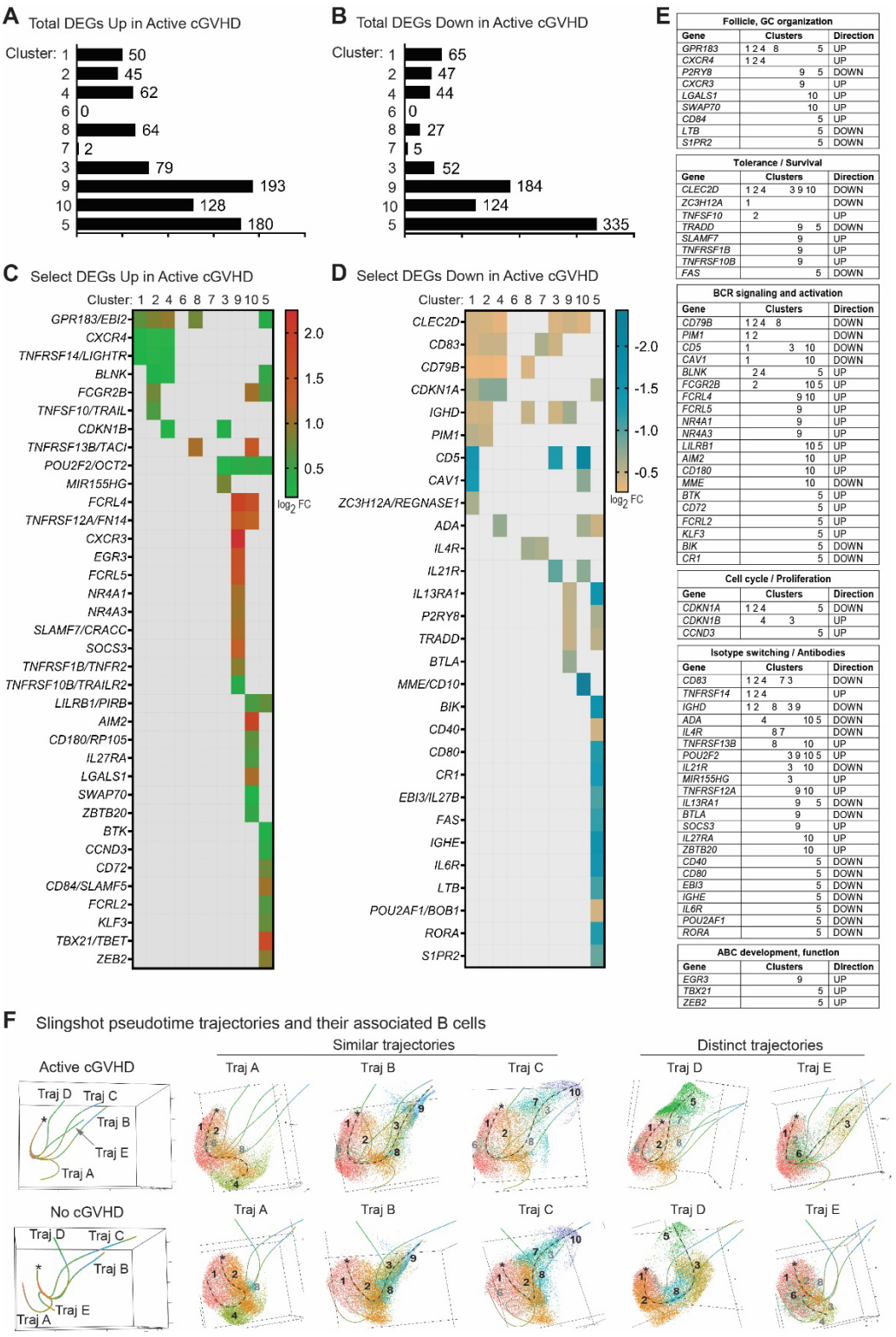
Other notable Down DEGs in Active cGVHD B cells (**Fig. 5D,E**) included *CD83*, known for discriminating light and dark zone B cells during antigen selection within GCs (*79*). *IL21R* expression was also Down in Clusters 3 and 10, potentially significant because of the role interleukin (IL)-21 plays in the both the humoral immune response and the development of regulatory B cells (*80*), whose numbers are significantly reduced in Active cGVHD patients (*5*). Furthermore, we found decreased *FAS* expression in Cluster 5, along with decreased *TRADD* in both Cluster 5 and metabolically active Cluster 9. This indicates disruption of important mediators of apoptosis and potentially impaired elimination of autoreactive B cells (*81*). *TRADD* and *S1PR2* as Down DEGs in Cluster 5 (**Fig. 5D,E**) was also notable given the overall increased frequency of

memory B cells expressing these genes in allo-HCT patients relative to HDs, and HIV and malaria patients (**Fig. 4**). *SLAMF7* (*CRACC*, *CD319*) was also notable since it can positively regulate B cell proliferation and cytokine production (*82*), being an Up DEG in Cluster 9 (**Fig. 5C,E**).

Given that BAFF likely plays a prominent role in the genesis and progression of cGVHD (*8, 9, 11, 12, 83*) along with alloantigens, it was notable that the BAFF receptor *TNFRSF13B* (*TACI*) was increased in Clusters 8 and 10 in Active cGVHD (**Fig. 5C,E**). Other members of the TNF superfamily (TNFSF) or TNF-receptor superfamily (TNFRSF) involved in key B cell pathways were likewise Up DEGs in Active cGVHD B cells, including *TNFRSF14* (*LIGHTR*, *CD270*), *TNFSF10* (*TRAIL*, *CD253*), *TNFRSF12A* (*FN14*, *TWEAKR*, *CD266*), *TNFRSF1B* (*TNFR2*, *CD120b*), and *TNFRSF10B* (*TRAILR2*, *CD262*), **Fig. 5C,E**. Interestingly, *GPR183* (*EBI2*) was overexpressed in Cluster 8 along with *TNFRSF13B* (**Fig. 5C,E**; **fig. S5A**), possibly reflecting a positive influence by BAFF on EBI2 expression, as reported in a B cell leukemia cell line (*84*). Co-expression of *GPR183* and *TNFRSF13B* indeed occurred in some B cells, which were expanded 4-fold in Active cGVHD patients (**fig. S5B,C**). Indeed EBI2, like TACI, was expressed at a high level on the surface of CD11c^{+} ABCs by flow cytometry (**fig. S5D**). By comparison, follicular-like naïve B cell subsets were generally low/negative for EBI2 and TACI expression (**fig. S5E**). Thus our data add new insight to the recent finding that *TNFRSF13B* and *GPR183* are major DEGs defining a B cell branch point that immediately precedes ABCs, leading to ABC expansion in adults chronically exposed to malaria antigens (*30*).

Other notable DEGs such as *BTK*, increased in Cluster 5 in Active cGVHD (**Fig. 5C,E**), suggested constitutive BCR activation and autoantibody-production potential. Consistent with work showing BAFF-dependent SYK protein increases in B cells from Active cGVHD patients and mice (*9, 17*), *BLNK*, the adapter molecule immediately downstream of SYK in BCR signaling,

was Up in Clusters 2, 4 and 5 (**Fig. 5C,E**). The transcription factor *ZBTB20* was increased in pre-GC-like Cluster 10, and this zinc finger protein was recently shown to be essential for the conversion of BCR-activated B cells to antibody-secreting cells (ASCs) and for the maintenance of long-lived ASCs (*85, 86*). *CAV1* (*CAVEOLIN-1*), a molecule that controls BCR compartmentalization on the plasma membrane (*87*), was decreased in Clusters 1 and 10 (**Fig. 5D,E**). Accordingly, *Cav1* deficiency in mice leads to altered nanoscale organization of IgM-BCRs and a skewed Ig repertoire with features of poly-reactivity and auto-reactivity (*87*). Molecules shown to be functionally important in ABC expansion or function were also significantly affected. *ZC3H12A*, which encodes the RNA-binding protein REGNASE-1, was decreased in Cluster 1 (**Fig. 5D,E**), and Regnase-1 deficiency at earlier stages of B cell development in mice produces severe autoimmune immunopathology and early mortality, accompanied by ABC expansion (*88*). *TBX21* (encoding *TBET*), *ZEB2* and *EGR3* are important and potential master regulators of ABC development and/or function in mice and humans (*23, 25, 26, 49*), and each were Up DEGs in either 'memory' Cluster 5 (*TBX21*, *ZEB2*) or activated Cluster 9 (*EGR3*), **Fig. 5C,E**. Cluster 5 was enriched for ABCs (**Fig. 3A**), suggesting that dysregulated *TBET*, *ZEB2* and *EGR3* are potential drivers of the observed ABC expansion in Active cGVHD patients (**Fig. 3B-I**). Finally, a set of genes associated with cell cycle were among the observed DEGs in Active cGVHD. These included *CDKN1B* and *CCND3* (Up, **Fig. 5C,E**), and *CDKN1A* (Down, **Fig. 5D,E**). This was consistent with our previous observations that Active cGVHD B cells demonstrate enhanced proliferation to some important stimuli, including low-dose surrogate antigen and NOTCH ligand (*9, 10*). Together these data define new aspects of B cell subsets following allo-HCT, whereby distinct maturation, activation, survival and homing pathway alterations occur when B cell tolerance is not achieved in the form of cGVHD.



E

Follicle, GC organization		
Gene	Clusters	Direction
GPR183	1 2 4 8 5	UP
CXCR4	1 2 4	UP
P2RY8	9 5	DOWN
CXCR3	9	UP
LGALS1	10	UP
SWAP70	10	UP
CD84	5	UP
LTB	5	DOWN
S1PR2	5	DOWN

Tolerance / Survival		
Gene	Clusters	Direction
CLEC2D	1 2 4 3 9 10	DOWN
ZC3H12A	1	DOWN
TNFSF10	2	UP
TRADD	9 5	DOWN
SLAMF7	9	UP
TNFRSF1B	9	UP
TNFRSF10B	9	UP
FAS	5	DOWN

BCR signaling and activation		
Gene	Clusters	Direction
CD79B	1 2 4 8	DOWN
PIM1	1 2	DOWN
CD5	1 3 10	DOWN
CAV1	1 10	DOWN
BLNK	2 4 5	UP
FCGR2B	2 10 5	UP
FCRL4	9 10	UP
FCRL5	9	UP
NR4A1	9	UP
NR4A3	9	UP
LILRB1	10 5	UP
AIM2	10	UP
CD180	10	UP
MME	10	DOWN
BTK	5	UP
CD72	5	UP
FCRL2	5	UP
KLF3	5	UP
BIK	5	DOWN
CR1	5	DOWN

Cell cycle / Proliferation		
Gene	Clusters	Direction
CDKN1A	1 2 4 5	DOWN
CDKN1B	4 3	UP
CCND3	5	UP

Isotype switching / Antibodies		
Gene	Clusters	Direction
CD83	1 2 4 7 3	DOWN
TNFRSF14	1 2 4	UP
IGHD	1 2 8 3 9	DOWN
ADA	4 10 5	DOWN
IL4R	8 7	DOWN
TNFRSF13B	8 10	UP
POU2F2	3 9 10 5	UP
IL21R	3 10	DOWN
MIR155HG	3	UP
TNFRSF12A	9 10	UP
IL13RA1	9 5	DOWN
BTLA	9	DOWN
SOCS3	9	UP
IL27RA	10	UP
ZBTB20	10	UP
CD40	5	DOWN
CD80	5	DOWN
EBI3	5	DOWN
IGHE	5	DOWN
IL6R	5	DOWN
POU2AF1	5	DOWN
RORA	5	DOWN

ABC development, function		
Gene	Clusters	Direction
EGR3	9	UP
TBX21	5	UP
ZEB2	5	UP

Fig. 5. Molecules critical to B cell function are altered within clusters in Active cGVHD as revealed by DEG analysis. (**A-D**) DEG analysis within the 10 B cell clusters in the scRNA-Seq dataset between allo-HCT patient groups based on disease status (as described in **Fig. 1**: Active

cGVHD, *n*=4; No cGVHD, *n*=4). Bar graphs indicate the total number of DEGs by cluster Up in Active cGVHD (A) or Down in Active cGVHD (B), as compared to No cGVHD. The heat maps in (C) and (D) depict DEGs selected from the entire dataset (**table S3**), known to be critical for various aspects of B cell function ('Select' DEGs). Colored squares represent significant (P_{adj} <0.05) log2 fold change (log2 FC) values for DEGs shown (rows) within the cluster(s) indicated (columns), either Up (C) or Down (D) in Active cGVHD B cells. (**E**) All genes from (C) and (D) were grouped into the major categories shown based on their known roles in B cell function. The clusters in which each gene was significantly altered are summarized in the middle column, with the direction of change in Active cGVHD B cells relative to No cGVHD B cells indicated in the right column. (**F**) The panels at left show Slingshot pseudotime trajectory predictions (Traj A-E) for untreated B cells from Active cGVHD patients (top) and No cGVHD patients (bottom). The origin of the pseudotime analysis (denoted by asterisks) was set as Cluster 1 based on our knowledge of its transitional, Igκ signature gene profile (**Figs. 1E,F** and **2A,B**), suggesting it is the earliest peripheral B cell population emerging from the bone marrow. Panels at right for each patient group represent each trajectory in isolation (dashed line for reference), along with all of its associated B cells. B cells clusters are colored and numbered per original unbiased clustering (**Fig 1B,C,I**). Black numbers indicate major clusters that lie along each trajectory, while clusters present but having only a small number of B cells represented are indicated in gray.

Lesional skin B cells in Active cGVHD share characteristics of hyper-activity with some circulating B cell subsets

Certain DEGs aberrantly activated in B cells that circulate in Active cGVHD patients suggest that these cells may home to sites of disease involvement. Thus, we performed scRNA-Seq on a punch biopsy of lesional skin (dermis) from a patient with sclerodermatous cGVHD and compared it with a biopsy of normal dermal skin from a healthy donor undergoing abdominoplasty. Dermal B cells were detected in both samples, and were distinguishable by a single UMAP cell cluster plus signature gene profile (**table S4**), which indicated the presence of Igκ (*IGKC*), mostly isotype-switched (*IGHG1*, *IGHG3*, *IGHG4*, *IGHA1*) B cells. Other notable signature genes included *GPR183* (*EBI2*) and receptors for BAFF (*TNFRSF13B* [*TACI*] and *TNFRSF13C* [*BAFFR*]). Remarkably, multiple DEGs increased in subsets of blood B cells in Active cGVHD patients (**Fig. 5C,E**) were also Up DEGs in B cells from Active cGVHD sclerodermatous skin (**table S5**). *POU2F2* (*OCT2*), important for generation of antibody-producing plasma cells(*89*), *SWAP70*, a

molecule phosphorylated by SYK that influences B cell migration and ABC expansion (*90, 91*), and *AIM2*, which has a positive role in the expansion of autoreactive B cells in a mouse lupus model (*92*), were all increased in both blood and skin B cells in Active cGVHD. Not surprisingly, most DEGs observed in sclerodermatous skin that were also DEGs in blood B cells of Active cGVHD patients occurred in the most activated and differentiated blood B cell clusters, primarily Clusters 3, 9, 10 and 5 (**Fig. 5C,E**, **table S3**, **table S5**).

Trajectories for B cell subset diversification in Active cGVHD patients reflect maturation defects that can potentially be reversed

Our current findings are consistent with past data (*10*) suggesting that altered maturation is linked to aberrant activation signaling in circulating Active cGVHD B cells. Thus, we performed pseudotime trajectory analysis of our scRNA-Seq dataset using Slingshot (*93*). Pseudotime trajectories may provide insight into the relatedness of distinct cell subsets within the same lineage as they mature, diversify, or reach various states of activation. As shown in **Fig. 5F** (far left panels), our analysis predicted 5 pseudotime trajectories (designated Traj A-E) for each allo-HCT patient group. Asterisks indicate the origin (starting point) for the trajectories based on 'biological knowledge.' We assigned the origin to Cluster 1 because of its transitional-like, *IGKC* (Igκ$^+$) phenotype, its proximity to Cluster 2 [the *IGLC3* (Igλ$^+$) transitional-like subset], and its presence in all five trajectories. Traj A, B and C were remarkably similar between allo-HCT patient groups (**Fig. 5F**, 'Similar trajectories'). Traj A consisted primarily of B cell Clusters 1, 2, and 4 only, possibly indicating termination at Cluster 4 that had generally low transcripts for signature genes important for B cell differentiation such as *IRF8* and *IRF4* consistent with a state of B cell anergy, (**Fig. 1E**). Traj B proceeded through Clusters 1, 2, 8, 3, and 9. Traj C proceeded through Clusters

1, 2, 8, 7, and 10, although in Active cGVHD there were notably fewer B cells in Cluster 8 (apparent minor contribution indicated by gray font). Traj D and E were markedly different between patients with Active cGVHD and No cGVHD (**Fig. 5F**, 'Distinct trajectories'). In No cGVHD, Traj D had clear progression through Clusters 1, 2, 8, 3 and 5. By contrast in Active cGVHD, Traj D completely lacked Cluster 3 and had minimal Cluster 8 B cells, proceeding primarily through Clusters 1 and 2, then directly to Cluster 5. Traj E had major contributions from Clusters 1 and 6 in both groups, while Cluster 3 was uniquely prominent in Active cGVHD relative to No cGVHD where few Cluster 3 B cells were apparent. These observations suggest that the generation of Cluster 5 B cells ('memory' B cells, ABCs) in Active cGVHD patients can potentially occur along a diversification pathway that bypasses tolerance checkpoints. These tolerance checkpoints are represented by Cluster 8 and Cluster 3 B cells in the No cGVHD setting where tolerance is achieved.

The remarkable set of DEGs and altered trajectories described above expand on our previous study demonstrating a maturation block in circulating B cells in Active cGVHD patients (*10*). Previously, we utilized all-*trans* retinoic acid (ATRA) as a tool *in vitro* to 'mature' Active cGVHD B cells by restoring a normal IRF4/IRF8 ratio, attenuating hyper-responsiveness to BCR and NOTCH co-stimulation (*10*). Thus, we assessed the effects of ATRA on B cell clustering, DEGs, and trajectory inferences. A total of 10,000 ATRA-treated B cells per sample were targeted for scRNA-Seq analysis, from the same 8 allo-HCT donor patients described in **Fig. 1** and **table S1**. The effect of ATRA treatment was validated by the robust upregulation of known ATRA-responsive genes across multiple B cell clusters including *PLAAT4* (i.e., *RIG1*, *Retinoid-Inducible Gene 1*), as well as *ASB2* which is a major ATRA target in the treatment of acute promyelocytic leukemia patients (*94*) [**table S6**]. ATRA treatment resulted in four predicted pseudotime

trajectories (designated A-D) in both allo-HCT patient groups (**Fig. 6A,B**), compared to the five trajectories for untreated B cells (**Fig. 5F**), suggesting a potential narrowing of transcriptional diversity. Supporting this, Cluster 8 was notably absent from the post-ATRA UMAP clusters, while the nine remaining clusters were present (**Fig. 6C,D**). Most remarkably, the trajectory leading to Cluster 5 in ATRA-treated Active cGVHD B cells (Traj D, **Fig. 6A**) included Clusters 1, 2, 7, 3, 10 and 5, which was a markedly different profile from the trajectory leading to Cluster 5 in untreated Active cGVHD B cells (Traj D, **Fig. 5F** top panel). This suggests that ATRA, to some extent, normalized the trajectory leading to ABCs and memory B cells when cGVHD was clinically apparent.

Additional nuances in pseudotime trajectories for ATRA-treated B cells were the inclusion of Cluster 10 in Traj D for Active cGVHD B cells (**Fig. 6A**). By contrast, Cluster 10 appeared in Traj C in No cGVHD B cells (**Fig. 6B**). In all scenarios with ATRA treatment, Cluster 3 preceded Clusters 9, 10 and 5. This finding along with the observation that major apoptosis regulatory genes *BCL2* and *BAX* were Cluster 3 signature genes (**Fig. 2H**) implicates this subset as a pivotal precursor memory checkpoint population.

ATRA affected the distribution of some B cell populations differently between patient groups. For example, the ratio of B cells mapping to Clusters 1 and 2 was much lower for No cGVHD B cells (0.38 and 0.43, respectively) compared to Active cGVHD B cells (0.84 and 0.79, respectively) following ATRA treatment (**Fig. 6C,D**). Additionally, the ratios of B cells mapping to Clusters 7 and 3 were reciprocally changed between groups by ATRA, decreasing in No cGVHD (0.76 and 0.93, respectively) and increasing in Active cGVHD (1.23 and 1.51, respectively). Together these differences may be explained, at least in part, by the apparent greater loss of total viable B cells with ATRA treatment in No cGVHD compared to Active cGVHD (ATRA/Untr

total B cell ratio of 0.70 for No cGVHD vs. ATRA/Untr total B cell ratio of 0.93 for Active cGVHD, **Fig. 6C,D**). Thus, distinct responses to ATRA between the Active vs. No cGVHD peripheral B cell compartments revealed additional potential survival and maturation checkpoints.

Finally, assessment of DEGs following ATRA treatment revealed altered expression of a multitude of genes compared to untreated B cells from all 8 allo-HCT patients (**table S6**). Notably, these DEG changes were statistically indistinguishable between allo-HCT patient groups based on cGVHD disease status. Total decreased or increased DEGs after ATRA vs. untreated B cells are represented numerically in **Fig. 6E** and **6G** pie charts, respectively. Cluster 7 was most affected by ATRA, with 1,673 Down DEGs compared to untreated B cells (**Fig. 6E, table S6**). Interestingly, some notable DEGs in untreated B cells from Active cGVHD patients (Up or Down compared to No cGVHD, **Fig. 5C-E**) were significantly changed in the opposite direction by ATRA (**Fig. 6F,H**), hinting at potential corrective effects by agents that promote B cell maturation. Strikingly, both *GPR183* (*EBI2*) and *TNFRSF13B* (*TACI*) were decreased upon ATRA treatment across multiple clusters, representing both 'transitional' Clusters 1 and 2, and 'memory' Cluster 5 (**Fig. 6F**). Together these data suggest that ATRA both reshaped the distribution of B cells among clusters and impacted the transcription of some DEGs observed in Active cGVHD.

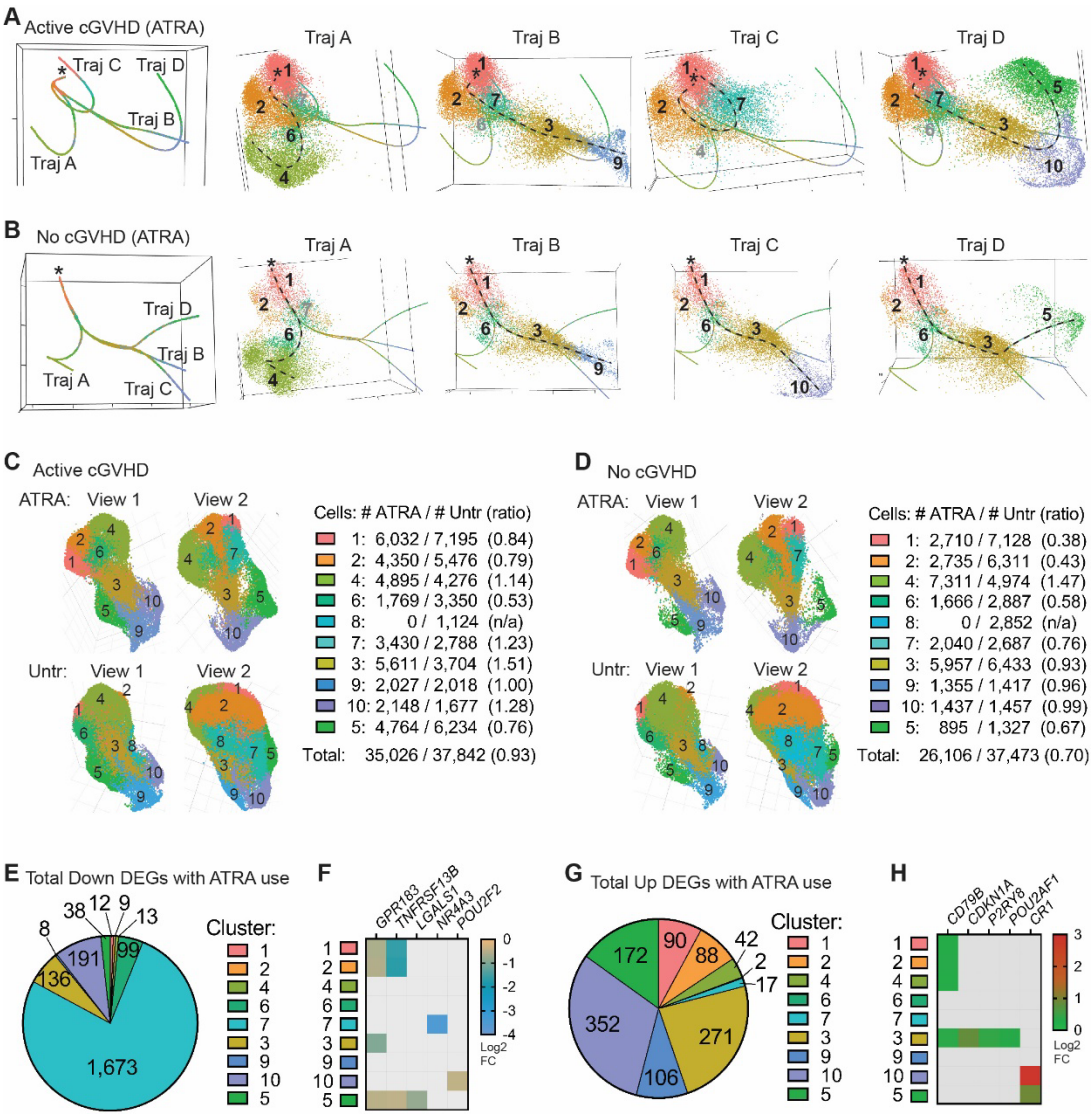


Fig. 6. An inducer of cell differentiation, all-*trans* retinoic acid (ATRA), influences B cell pseudotime trajectories, cluster distribution, and some cGVHD-related DEGs relative to untreated B cells. (**A,B**) Panels at left show trajectory estimates (Traj A-D) for B cells from Active cGVHD patients (A) and No cGVHD patients (B) after treatment with ATRA for 18 h. The origin of the pseudotime analysis (denoted by asterisks) was set as Cluster 1, as in **Fig. 5F**. Panels at right for each patient group represent each trajectory in isolation (dashed line for reference), along with all of its associated B cells. B cells are colored and numbered by the cluster to which they belong. Black numbers indicate major clusters that lie along each trajectory, while clusters present but having only a small number of B cells represented are indicated in gray. (**C,D**) Effect of ATRA on B cell distribution among clusters relative to untreated B cells. UMAP cluster projection and B cell distribution per cluster among ATRA-treated and untreated (Untr) Active cGVHD samples (C), and among ATRA-treated and untreated (Untr) No cGVHD samples (D). Total B cell numbers within each cluster for each treatment group and patient group are indicated. Cluster 8 was absent from the UMAP clustering predictions for ATRA-treated B cells, whereby the 9 clusters shown were identified as corresponding to the same 9 clusters in the untreated groups based on signature

gene profiles. Ratios represent the number of B cells within each cluster for ATRA treatment divided by the numbers of B cells within each cluster for untreated cells (#ATRA / #Untr). (**E-H**) DEGs induced by ATRA in B cells from all 8 allo-HCT patients in the single-cell RNA-Seq dataset, compared to untreated B cells from all 8 patients. Pie charts indicate the total number of DEGs significantly decreased (E) or increased (G) after ATRA treatment in the cluster indicated. Heat maps (F and H) represent DEGs from **Fig. 5** that were significantly altered in the opposite direction following ATRA treatment. Colored squares represent significant (P_{adj} <0.05) log2 FC value for the gene and cluster indicated, for ATRA-treated B cells (all 8 samples) compared to untreated B cells (all 8 samples).

Some genes are broadly dysregulated across B cell populations in patients with Active cGVHD

Having found that blood B cells from Active cGVHD patients are hyper-proliferative to surrogate antigen (*10, 17*), we hypothesized that some DEGs occur broadly in this disease. Identifying DEGs across multiple clusters might suggest genes capable of affecting B cells at multiple stages of peripheral diversification. In **Fig. 7A** we show Up DEGs (left) or Down DEGs (right) in the scRNA-Seq dataset that occurred in at least 4 B cell clusters in untreated Active cGVHD B cells. DEGs were ranked based on occurrence in the most clusters, and then by their first occurrence in the least mature/activated clusters. Twenty-nine DEGs were increased and 34 DEGs were decreased in 4 or more clusters in Active cGVHD. Several of these DEGs were already depicted in **Fig. 5C-E** because of their known role in B cell function (bold font in **Fig. 7A**). To corroborate broad expression of these genes we subjected the scRNA-Seq dataset to a 'bulk-like' DEG analysis of all B cells without clustering (**Fig. 7B**). Indeed, we found that numerous DEGs from **Fig. 7A** were also found to have significant broad dysregulation per the bulk-like analysis (**Fig. 7B**, asterisks). To further highlight the potential function of these broadly expressed DEGs, we grouped the DEGs from **Fig. 7B** empirically into GO-annotated cellular pathways (**table S7**). Based on the direction of change in B cells from Active cGVHD patients compared to No cGVHD

patients, along with the relevant published literature, the potential regulatory impact of these DEGs on B cell function (positive or negative) was inferred.

ARRDC3 was among the most broadly Down DEGs across B cell clusters (**Fig 7A,B**). Although the function of this α-arrestin in B cells remains to be determined, this finding highlights the potential for dysregulation of GPCRs, with numerous being identified as DEGs in Active cGVHD. Also notable, *NFKBIA* (encoding IκBα, the inhibitor of nuclear factor (NF)κB signaling), known to negatively regulate B cell activation and keep humoral immune responses in check, was decreased in eight clusters in cGVHD B cells. By contrast, *CKS2* stood out among the most broadly increased DEGs in Active cGVHD B cells, reaching significance in eight B cell clusters (**Fig. 7A**) and in the bulk-like analysis (**Fig. 7B**).

CKS2 has an important yet incompletely-defined and potentially paradoxical role in regulating cell cycle progression and survival. First, CKS2 is required for cell cycle transition from the G0-G1 phases, serving as an essential co-activator of cyclin-dependent kinases (CDKs) (*95*). Alternatively, CKS2 can prevent or delay cell cycle progression by protecting the CDK inhibitor $P27^{KIP1}$ from proteasomal degradation by competing with proteins that recruit ubiquitin ligases to $P27^{KIP1}$ (*96*). This coordinated protection of $P27^{KIP1}$ from degradation by CKS2 averts premature cell cycle entry, thus preventing apoptosis (*97*). UMAP plot comparisons (**Fig. 7C**) clearly depict the widespread increase in *CKS2*-positive B cells in Active cGVHD relative to No cGVHD. Displayed as normalized expression values (**Fig. 7D**), *CKS2* was uniformly increased in Active cGVHD samples in the eight clusters reaching significance (Clusters 1,2,4,8,3,9,10&5). We further validated this *CKS2* finding by qPCR analysis on purified B cells from an independent cohort of allo-HCT patients (**Fig. 7E**, **table S2**).

Given our previous studies showing that non-manipulated B cells from cGVHD patients have an enhanced rate of survival in culture (*8*), the observation that *CKS2* was broadly increased in Active cGVHD B cells hinted that $P27^{KIP1}$ may be protected from degradation. If so, $P27^{KIP1}$ would likely exhibit elevated phosphorylation at a regulatory site (T198) required for recruitment of ubiquitin ligase complexes (*96*). Employing unbiased protein phosphoarrays and whole cell lysates from purified B cells of an independent cohort of allo-HCT patients, phospho-T198 was detected at a 6- to 10-fold greater level in B cells from Active cGVHD patients compared to No cGVHD patients (**Fig. 7F,G** and **fig. S6**). By contrast, active-site phosphorylation on isoforms of the kinases AMPK and RSK, shown previously to phosphorylate $P27^{KIP1}$ T198 (*98, 99*), were similar between allo-HCT patient B cells, suggesting that increased kinase activity was not responsible for enhanced phospho-T198 in Active cGVHD. Analysis of $P27^{KIP1}$ total protein expression supported the concept that increased phospho-T198 was protective against $P27^{KIP1}$ loss in Active cGVHD B cells, with protein levels remaining significantly higher compared to No cGVHD B cells (**Fig. 7H**). A potential model for potential $P27^{KIP1}$ regulation in Active cGVHD B cells is shown in **Fig. 7I**, whereby broadly increased CKS2 expression, in conjunction with increased *CDKN1B* (*P27KIP1*) transcription in some subsets (**Fig. 5C,E**), may lead to $P27^{KIP1}$ accumulation. Thus, our dataset allows us to begin to determine functional contributors to enhanced survival in alloantigen-rich microenvironments where B cell tolerance is impaired. These findings support prior work showing that cGVHD B cells are primed for proliferation and protected from apoptosis (*8, 10, 17*).

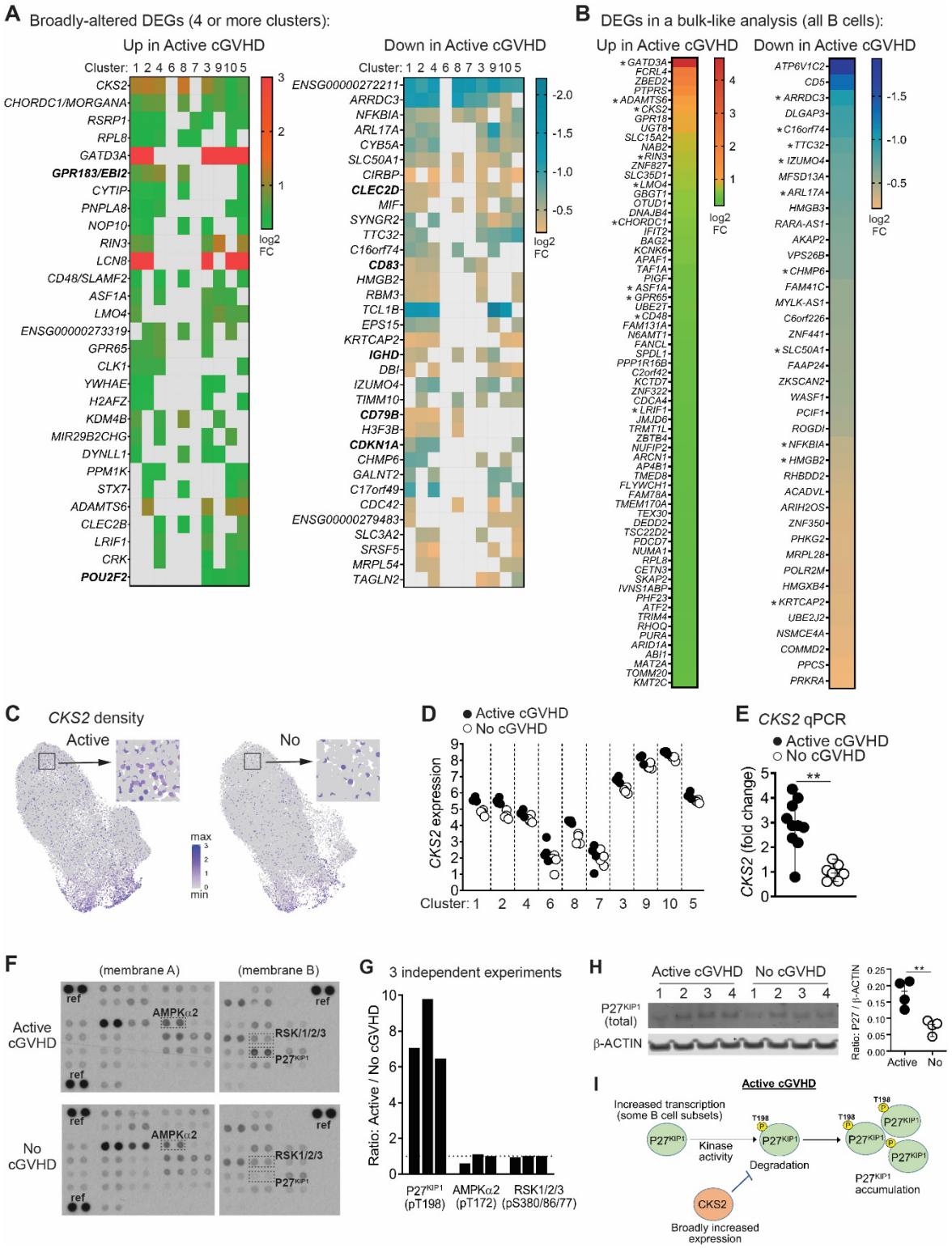


Fig. 7. Assessment of the most broadly-altered DEGs may provide additional insight into altered B cell effector functions in Active cGVHD. A, DEGs in the scRNA-Seq dataset reaching significance (P_{adj} <0.05) in four or more of the 10 B cell clusters described in **Fig. 5**, either Up in

Active cGVHD (left heat map) or Down in Active cGVHD (right heat map). Colored squares in the heat maps indicate that the gene reached significance in that cluster, with log2 FC values as indicated. No DEGs mapped to Cluster 6. Genes that were also depicted in **Fig. 5C-E** are shown in bold font. **B,** DEG analysis between disease groups performed on total untreated B cells. Heat maps show log2 FC values for annotated genes with a statistically different (P_{adj} <0.05) expression between allo-HCT patient groups, representing the difference in Active cGVHD B cells compared to No cGVHD B cells (Up or Down in Active cGVHD, as indicated). **C-E,** Validation of *CKS2* transcript overexpression in Active cGVHD B cells within the scRNA-Seq dataset and for a different cohort of allo-HCT patients. In (C), normalized expression density UMAP plots for *CKS2* from the single-cell RNA-Seq dataset in the Active cGVHD (Active) and No cGVHD (No) groups are shown. Representative regions depicted by the boxes were chosen randomly and enlarged (arrows) to visualize single B cells more easily. In (D), normalized *CKS2* expression values across all 10 B cell clusters for all 8 allo-HCT patients are shown, separated by disease group. In (E), qPCR analysis of *CKS2* was performed on freshly isolated, untreated B cells from a different allo-HCT patient cohort having Active cGVHD (*n*=10) or No cGVHD (*n*=7). Results indicate the fold change in *CKS2* expression based on the mean value in the No cGVHD group normalized to 1. *ACTB* (β-ACTIN) was the housekeeping gene in the qPCR analysis. Statistical comparison was performed using a two-tailed Mann-Whitney test (GraphPad Prism 9 software; **, *p*<0.01). (**F**) Representative phosphoprotein capture arrays for detection of various intracellular signaling molecules phosphorylated on key sites involved in their regulatory activity, performed on whole cell lysates of purified, untreated B cells isolated from Active cGVHD (*n*=3) and No cGVHD (*n*=3) patient blood samples (see also **fig. S6**). Dashed boxes and protein IDs indicate the location and assay results for duplicate spots of capture antibodies against $P27^{KIP1}$ (phospho-T198), AMPKα2 (phospho-T172), and RSK1/2/3 (phospho-S380/S386/S377, respectively). Reference control spots on the arrays are indicated ('ref'). (**G**) Combined density results from the 3 independent phosphoprotein array assays shown in (F) and **fig. S6**. Each bar indicates the results from one experiment and represents the ratio of the average dual spot intensity for Active cGVHD B cells over No cGVHD B cells for the protein indicated (dashed line represents a ratio of 1 as a guide). (**H**) Western blot analysis of total $P27^{KIP1}$ protein levels relative to β-ACTIN in whole cell lysates of B cells isolated from Active cGVHD (*n*=4) and No cGVHD (*n*=4) patient blood samples. Statistical comparisons were performed using a two-tailed, unpaired t-test (GraphPad Prism 9 software; **, *p*<0.01). (**I**) Cartoon depiction of how CKS2 expression in Active cGVHD B cells may function via $P27^{KIP1}$ maintenance and accumulation.

The 'plastic' state of peripheral B cell selection and activation following allo-HCT can be modeled based on our scRNA-seq data

Our previous work has suggested a non-static, 'plastic' molecular and functional state of peripheral B cells in allo-HCT, with maturation defects that associate with cGVHD (*10*). Our first-of-its kind scRNA-Seq dataset provides new insight into the potential B-cell plasticity state of the peripheral

B cell compartment in the allo-HCT setting. Signature gene profiles defined 10 distinct B cell subsets (**Fig. 8, left**) that continually develop under constant exposure to alloantigens and circulate after allo-HCT. Allo-HCT patient 'memory' B cell subsets also had some remarkably distinct gene features, regardless of cGVHD status, compared to the same subsets from healthy individuals or patients infected and chronically exposed to microbial antigens ('non-HCT' setting). This suggests a unique "Allo-HCT Plasticity" involving altered migration and survival signals in the allo-HCT environment that may function to shuttle B cells more freely and aberrantly among the most diversified clusters, Clusters 5, 9 and 10 (**Fig. 8**, dashed 2-headed arrows). Thus, the peripheral B cell pool in allo-HCT patients is potentially 'malleable', and as such is possibly amenable to novel therapeutic strategies to prevent or treat cGVHD, while maintaining immunity against pathogens and malignant cells.

Our data also reveal gene alterations important for emergence and/or persistence of aberrantly activated B cells in allo-HCT, where cGVHD manifests. DEGs in Active cGVHD could be subdivided into two major categories (**Fig. 8**, right) that comprise intrinsic differences from a B-cell tolerant state (No cGVHD). The first category of DEGs emerge in B cells in early maturation states that are not constitutively activated ("Early Subsets", Clusters 1, 2, 4, 6, and 8). These genes are largely maintained as DEGs throughout B cell diversification ('Broadly-Altered'). The second category of DEGs seem to emerge only after some degree of BCR stimulation ("BCR-Activated", Clusters 3, 7, 9, 10 and 5). The observed expansion of potentially pathogenic ABCs in Active cGVHD patients ("ABCs (including DN2) Expanded") is thus likely a consequence of gene alterations before and after BCR-engagement occurring in a cumulative way.

Altogether, new insight gained in our analyses herein enable new opportunities for exploitation of the identified molecular pathways to further understand tolerance loss in B cells, in

allo-HCT and beyond. These molecular programs are possibly driven by extrinsic factors that alter trajectories for B cell diversification, impairing tolerance and expanding ABCs. The identification of these highly novel DEG profiles in cGVHD provides context for new avenues of exploration into mechanisms of B cell tolerance loss.

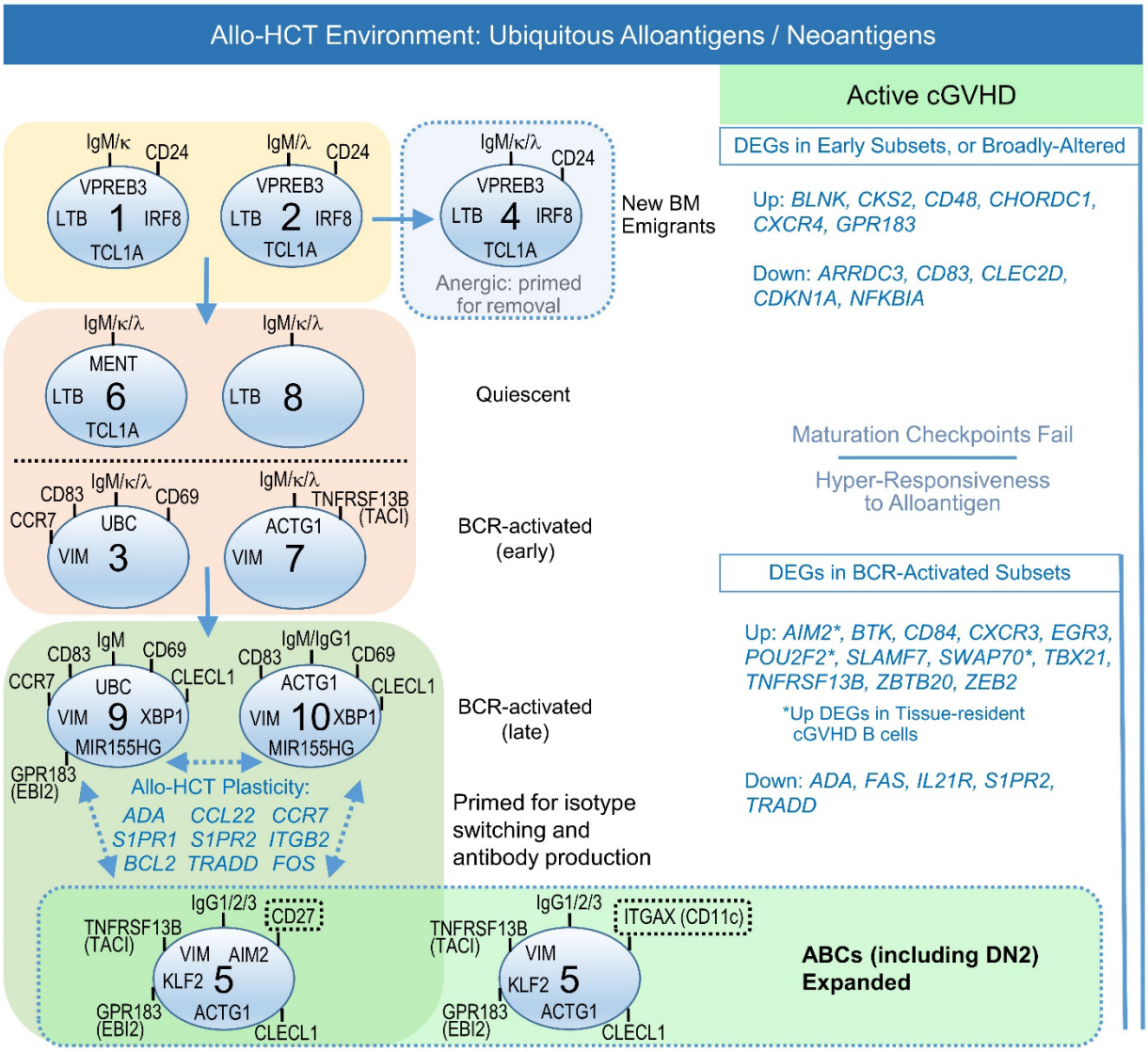


Fig. 8. Working model for extrinsic activation signals and intrinsic gene changes influencing B cell maturation, survival, and proliferative capacity that lead to the expansion of pathologic 'memory' B cells in Active cGVHD, including ABCs. Cluster-defining 'signature' genes (left side) and trajectory analyses (**Fig. 5F**) enabled the assignment of the 10 clusters along a logical pathway of B cell diversification in the periphery of allo-HCT patients, from newly-emigrated populations, through proposed checkpoint steps where survival decisions occur, and finally to BCR-activated and 'memory' states. As a culmination of extrinsic activation signals, including ubiquitous alloantigens and soluble factors including BAFF, the allo-HCT environment becomes amenable to gene changes that influence allo-HCT patients generally relative to non-HCT individuals ('Allo-HCT Plasticity'), and DEGs responsible for eventual pathogenesis occur early, broadly or exclusively after BCR stimulation, to promote Active cGVHD (right side).

DISCUSSION

Our scRNA-seq analysis reveals and details the breadth of B cell subset and gene expression abnormalities in human cGVHD. While immediate BCR hyper-responsiveness and aberrant upstream signaling in B cells from patients who manifest cGVHD after allo-HCT has been shown (*8, 10, 17*), comprehensive analyses of the peripheral B cell compartment has been lacking. The clarity provided by the identification of signature genes that define individual B cell subpopulations and DEGs associated with Active cGVHD may support the development of improved targeted therapies for patients. Taken together, our data demonstrate early defects in the expression of some key genes will almost certainly alter the normal trajectory of B cells as they diversify, and be reflected in additional DEGs that lead to tolerance loss and host reactivity.

Our in-depth analyses also now elucidate transcriptional programs that manifest and may lead to pathological functional changes in peripheral 'memory' B cell populations including ABCs. ABCs have emerged in the literature as important in both normal immunity and autoimmunity (*25, 26, 30, 31*). Indeed, transcriptional changes present in the earliest B cell subsets to inhabit the periphery are likely to influence the diversity and functional capacity of more mature or activated subsets, especially ABCs (including DN2) and other memory B cells. These subsets likely directly contribute to pathogenesis by producing alloantibodies after migrating to tissue sites where cGVHD manifests (*12*).

Examples of unique transcriptional changes observed in our scRNA-Seq data included *CKS2* and *GPR183* (*EBI2*), each overexpressed in Active cGVHD from the earliest transitional/naïve B cell subsets to the most differentiated subset, Cluster 5, enriched for ABCs and $CD27^+$ memory B cells. Such DEGs occurring from (at least) the time B cells first enter the circulation may cause early epigenetic changes that alter subsequent tolerance checkpoints and

potentially affect clonal diversity. Furthermore, by leveraging healthy individual and non-HCT patient B cell data, we found evidence that the allo-HCT microenvironment tips the scales toward hyper-responsiveness, with expansion in Active cGVHD of the extrafollicular-like and potentially-pathogenic DN2 ABC subset.

The emergence of functional ABCs (including DN2) with pathogenic roles in some disease setting is driven by one or more costimulatory signals that cooperate with BCR stimulation, including TLR7, IL-21 and IFNγ (*100-103*). Indeed, B cells isolated from Active cGVHD patients are also hyper-responsive to minimal BCR ligation in synergy with co-stimulatory molecules including NOTCH2 and BAFF (*10*). These processes are likely driven in part by the increased BAFF-to-B cell ratio that occurs in cGVHD patients (*11, 12, 83, 104*), and by niche production of BAFF in the SLO, as recently confirmed in a mouse cGVHD model (*9*). BAFF influence may be additionally propagated by increases in BAFF receptors such as TACI (encoded by the *TNFRSF13B* gene), as published (*12*) and herein described in new detail for specific B cell subsets in our scRNA-Seq and flow cytometry data. It is also noteworthy that multiple other TNF superfamily and TNF-receptor superfamily members were increased DEGs in multiple cGVHD B cell clusters, including the metabolically active Clusters 9 and 10. Some of these molecules (or their cognate receptors/ligands) are being evaluated as targets in clinical trials for various autoimmune and inflammatory diseases, supporting the potential assessment of these pathways in clinical trials for cGVHD as well. Thus, extrinsic factors in microenvironmental niches may be the initial drivers of DEGs that lead to early mechanisms of loss in peripheral B cell tolerance that can be elucidated through studies of circulating B cells.

Direct comparison of our scRNA-Seq dataset from allo-HCT patients with a publicly-available dataset on blood B cells isolated from HDs, or individuals with chronic infections (HIV, malaria), provided new insight into the unique aspects of the B cell pool that persists following immune reconstitution with allo-HCT therapy, regardless of whether have patients developed cGVHD or not. It is thus likely that some of these striking differences in the frequencies of B cells expressing molecules that orchestrate exit from the BM and homing to/establishment in SLO niches, as well as molecules key to survival, impact the uniqueness of the clinical manifestations of cGVHD once it occurs. These observations may be an additional reflection of the apparent maturation block that leads to B cell hyper-responsiveness in cGVHD patients (*10*), and may also at least partially explain the lower frequency of $CD27^{+}CD11c^{-}$ follicular memory B cells in allo-HCT patients compared to HDs. Together these aspects of allo-HCT B cells may indicate a unique opportunity for therapeutic targeting before cGVHD manifests.

We have now identified potentially targetable molecules dysregulated in subsets of Active cGVHD B cells that are novel, particularly molecules shown to regulate B cell movement and positioning within lymphoid follicles and during immune responses, including the GC reaction. The identification of *GPR183* (*EBI2*), *P2RY8*, and *S1PR2* as each being dysregulated in cGVHD in one or more clusters of B cells is particularly remarkable, given the roles these molecules play in follicular movement and GC positioning of B cells (*61, 71-75*). The dysregulation of *GPR183*/*EBI2* across five B cell clusters, including the newly recirculating transitional-like subsets, suggests that strictly regulated tolerance mechanisms that are initiated as B cells first recirculate and enter the SLO microenvironment may be disrupted in cGVHD patients. This finding is consistent with the notion that altered intrinsic functions of lymphocytes and stromal cells within SLOs disrupt normal architecture during follicle organization and GC development,

potentially leading to premature displacement of B cells to extrafollicular spaces before they re-enter the circulation and home to peripheral tissue sites (*78*). Remarkably, *GPR183* (*EBI2*), *P2RY8*, and *S1PR2* were each DEGs in Cluster 5, which is enriched for ABCs and other memory B cells. *P2RY8* expression was additionally decreased in metabolically active Cluster 9, also intriguing given the recent evidence that decreased B cell *P2RY8* expression propagates pathogenic antibody-producing plasma cells and ABCs in humans with lupus or antiphospholipid syndrome (*77*). The ligands for each of these molecules are known, and in some cases inhibitors and agonists have been described, permitting future studies in mouse models of cGVHD and eventually clinical trials. This concept is supported by an observation in a mouse lung transplant model of bronchiolitis obliterans (BO), a condition that frequently causes lung damage in cGVHD patients, where inhibition of EBI2 with a highly selective inhibitor reduced pulmonary lymphoid lesions and lung damage (*105*). Elucidating B cell transcriptional changes that help incite cGVHD development, versus transcriptional changes that occur after cGVHD genesis and nevertheless contribute to ongoing pathogenesis in an unrelenting cycle, will remain an area of future investigation.

While the full implications of the findings herein are currently limited and remain to be realized in the clinical setting, our data provide a foundation for future assessment of novel, potentially targetable pathways that alter B cell homeostasis during the diversification of B cells in the periphery when alloantigens or autoantigens are prevalent. Mouse studies can now be employed to elucidate the hierarchy of importance for some of these molecules in initiating and propagating loss of B cell tolerance. The ultimate benefit of this type of single-cell level research will be the ability to design tailored therapies that eliminate or prevent the emergence of pathogenic B cells without disrupting normal humoral immune responses required for long-term health.

MATERIALS AND METHODS

Design

Single-cell RNA-Seq was performed on peripheral B cells from allo-HCT patients chosen at random, with the desire to gain new knowledge about the molecular mechanisms that emerge in the allo-HCT setting to drive B cell hyper-responsiveness and cGVHD development in some patients (loss of tolerance), whereas other allo-HCT patients do not develop cGVHD and B cells function normally (maintenance of tolerance). Beyond some basic inclusion and exclusion criteria given to clinicians for obtaining PBMC samples for study (including cGVHD status), samples were obtained and chosen for use in a blinded fashion. For all supporting experiments, allo-HCT patient samples were likewise obtained and chosen for use in a blinded fashion.

Patient and healthy donor blood samples

De-identified whole blood samples or apheresis samples were obtained from allo-HCT patients under IRB protocols from Duke University, the National Cancer Institute (NIH), or the Dana-Farber Cancer Institute. Anonymous healthy donor samples were buffy coats purchased from Gulf Coast Regional Blood Center (Houston, TX). Viably frozen PBMCs from allo-HCT patients or healthy donors were prepared by Ficoll-Paque® PLUS (GE Healthcare) gradient separation prior to storage in a liquid nitrogen freezer. For some experiments, purified B cells were additionally isolated immediately following PBMC gradient separation using a Human B Cell Isolation Kit II (Miltenyi Biotech), washed in cold Dulbecco's phosphate-buffered saline, and stored as B cell pellets at -80°C. The characteristics of allo-HCT patients who provided the samples used for

single-cell RNA-Seq analysis and the samples used for additional experiments are described in **tables S1** and **S2,** respectively.

B cell culture, single-cell RNA-Seq library preparation, and library sequencing

B cells were purified from viably cryopreserved allo-HCT patient PBMCs using the Human B Cell Isolation Kit II (Miltenyi Biotech). B cells were then cultured for 18 h at 5% CO_2 in RPMI-1640 medium containing 10% fetal bovine serum (FBS) and 55 μM 2-Mercaptoethanol (Gibco). In some wells, cells were treated with 100 nM ATRA (MP Biomedicals, cat# 190269) for the entire culture period. Tissue culture-grade dimethyl sulfoxide (DMSO) alone (Sigma Chemical Co) was used as the vehicle control in wells not treated with ATRA. The cells were then harvested, and dead cells were removed using an EasySep™ Dead Cell Removal (Annexin V) Kit (STEMCELL Technologies, Inc.). Live B cells (in HBSS containing 2% FBS) were immediately transferred on ice to the Duke Molecular Genomics Core for 10X Genomics Chromium™ Single-Cell 3' library generation according to the manufacturer's instructions and the Core Facility's standard methods, targeting 10,000 B cells per sample. 10X Genomics libraries were then transferred directly to the Duke Center for Genomic and Computational Biology, Sequencing and Genomic Technologies Shared Resource, for sequencing. All individual 10X Genomics libraries (16 total) were pooled and then applied evenly across all 4 lanes of a NovaSeq S4 Flow Cell (Illumina) with 100-bp paired-end sequencing at a depth of ~700 million reads per sample (~70,000 reads per cell).

Single-cell RNA-Seq data analysis on purified B cells from allo-HCT patients

The raw data in BCL format were de-multiplexed into FASTQs by Illumina's bcl2fastq (https://support.illumina.com). The extraction of cellular barcodes and unique molecular identifiers (UMI sequences) and genome alignment of the biological sequences were performed using the cellranger pipeline (version 2.1.1, https://support.10xgenomics.com) to generate a filtered sparse matrix of gene read counts per cell. The 16 individual matrices were integrated using Stuart et al.'s method to remove batch effects (*27*). Any cells in which fewer than 500 genes were detected and those with greater than 5% mitochondrial gene counts were considered poor quality and excluded from the analysis. Any genes detected in fewer than 150 cells were likewise excluded from the analysis.

Clustering analysis was conducted by using R statistical environment (*106*) and its extension package Seurat v3.1.4 (*27*). DEG analysis with respect to disease status (Active cGVHD vs. No cGVHD) within each cluster was performed based on the read counts across all cells within each sample using the Bioconductor package DESeq2 (version 1.26.0) (*107*). Global interaction between disease status and the B cell clusters also was analyzed. Multiple testing was accounted for within the framework of control of the false discovery rate (FDR) using the Q-value approach (version 2.14.1) (*108*). Visualizations of single-cell-level transcription profiles in UMAP space using color was performed using R package Seurat v3.1.4 (*27*). Trajectory analysis was performed to infer the relative relationship among cell clusters over 'pseudotime' using the package Slingshot as described previously (*93*).

Comparison of allo-HCT patient single-cell RNA-Seq data on purified B cells to a publicly available dataset on purified B cells from HDs, HIV patients, and malaria-exposed individuals

To compare scRNA data from this study to data published by Holla *et al.* (*30*), both datasets were filtered for cell quality using the same thresholds; i.e., cells with more than 5% mitochondrial gene counts, counts from fewer than 200 genes, or counts from more than 2,500 genes were excluded. Cells that passed quality filtering were then filtered to retain only B cells since non-B cells were considered contaminants. Cell type was inferred based on the tumor microenvironment signature genes from CellAssign (*109*) and expression of the union of those signature genes and the top 2000 most variable genes (following variance stabilizing transformation (*107*)) using the R package scSorter (*110*). To normalize the allo-HCT and Holla *et al*. datasets, gene counts for the remaining cells were transformed to counts per million reads mapped (CPM) using the Seurat R package. For a visual comparison of gene expression, data from the present study and *Holla et al.*(*30*) were integrated following the steps recommended by the developers of Seurat (*111*). A total of five cohorts were integrated to account for batch effects, specifically: the data from Holla *et al*. (*30*) were split by sample type (HD vs. HIV vs. Malaria), and data from the allo-HCT patients were split by the two sequencing runs. After integration, cells were clustered and plotted in the UMAP space to illustrate variation between B cell expression across studies and sample types. To investigate the expression of biologically-relevant genes in *CD27*$^{+}$*ITGAX*$^{-}$, *CD27*$^{-}$*ITGAX*$^{+}$ and *CD27*$^{+}$*ITGAX*$^{+}$ B cells across studies and sample types, we generated a heatmap with the following attributes: First, we identified each subset based on a detection threshold of 50 CPM for *ITGAX* and *CD27* expression. We then scaled and centered the CPM values for the genes to highlight variation across cells. To illustrate the distribution of gene expression across cell populations, values were plotted lowest to highest from left to right for each gene and within each cell type (3 levels: *CD27* single-positive, *ITGAX* single-positive, *CD27/ITGAX* double-positive), and study and sample type (2 levels from this study: Active cGVHD, No cGVHD; 3 levels from Holla, *et*

al.(*30*): Healthy, HIV, Malaria). Cell populations with more than 100 cells were randomly down-sampled to 100 cells. A subset of 9 genes from the heatmap were selected to show variation in expression and summarize the percentage of cells per population with expression greater than 50 CPM. Data analyses and plot generation were carried out in the R statistical environment (*112*) with packages from the comprehensive R archive network (CRAN; https://cran.r-project.org/) and the Bioconductor project (*113*). Code needed to replicate the analyses and figures follows principles of reproducible analysis, leveraging knitr (*114*) dynamic reports and git source code management (*115*).

Data and Code availability

All blood B cell scRNA-Seq library data from the 8 allo-HCT patients (16 samples) are available through the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) under accession code GSE161343. The R scripts to reproduce the analyses of these scRNA-Seq data are available at this site: https://gitlab.oit.duke.edu/dcibioinformatics/pubs/sarantopoulos-10x-cgvhd/.

Dermal skin cell isolation and single-cell RNA-Seq analysis

A skin punch biopsy was collected from an adult allo-HCT patient with Active cGVHD in accordance with an IRB protocol approved by Duke University. Likewise, a skin punch biopsy from surgically discarded abdominoplasty tissue from a healthy individual was obtained in accordance with an IRB protocol approved by Duke University. The freshly collected skin biopsies were washed with cold DPBS and digested with 1X dispase (ThermoFisher Scientific) overnight

at 4ºC. The dermis was mechanically separated from the epidermis and then digested at 37ºC and 5% CO2 with 0.25% trypsin/EDTA for 1.5 hr. Dermal cell suspensions were then passed through 70 µm and 40 µm cell strainers; the cells were pelleted by centrifugation and resuspended in Keratinocyte-SFM (ThermoFisher Scientific) containing 0.4% BSA. Single-cell suspensions were then subjected to 10X Genomics Chromium library generation and Illumina sequencing as described above. Signature gene analysis was performed using the R package Seurat to identify B cells within dermal cell clusters. DE gene analysis was performed using the R package DESeq2, as described for B cells above. The skin cell scRNA-Seq library data will be made available through the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) following peer review.

Quantitative PCR (qPCR)

Frozen (-80°C) B cell pellets were thawed on ice, and total RNA was isolated using the RNeasy Plus Mini Kit (Qiagen) and quantified with a Qubit® 3.0 Fluorometer (ThermoFisher). RNA (500 ng) was reverse-transcribed with an iScript cDNA Synthesis Kit (BioRad). *CKS2* gene expression was assessed by qPCR using *ACTB* (*β-ACTIN*) as the housekeeping gene. Primers were designed with the NCBI primer BLAST program, with the following sequences: *CKS2*, forward 5'-ACGAGTACCGGCATGTTATGT-3' and reverse 5'-GCCTAGACTCTGTTGGACACC-3'; *ACTB*: forward, 5'-GCTGTGCTACGTCGCCCT-3' and reverse, 5'-AAGGTAGTTTCGTGGATGCC-3'. Amplification was performed using the iTaq Universal SYBR® Green Supermix (Bio-Rad) with an annealing temperature of 60°C on an ABI StepOne Plus™ Real-time PCR System (ThermoFisher Scientific). Each 20-µl reaction contained 2.5 ng of cDNA, with each primer used at a concentration of 250 nM. Data analysis was performed using

Step One Software v2.3 software, and relative quantitation of gene expression was achieved using the standard-curve method. Fold change in expression in each sample was calculated using the average gene expression in patients with No cGVHD.

Flow cytometry and PhenoGraph analysis

Viably frozen PBMCs from allo-HCT patients were surface stained using antibodies against CD19, CD11c (ITGAX), CD21, CD24, CD27, EBI2 (GPR183), and TACI, along with their corresponding isotype control antibodies. 7-AAD (BioLegend) was used to assess cell viability. PhenoGraph cluster analysis was then performed on the remaining 6 markers after first pre-gating on viable B cells ($CD19^{+}$ $7AAD^{-}$), using this available function in Flowjo in combination with R. A detailed summary of all antibodies used is as follows: CD19 Pacific blue (clone J3-119, Beckman Coulter, cat#A86355); CD24 BV510 (clone ML5, BioLegend, cat#311126); CD21 FITC (Bu32, BioLegend, cat#354910); TACI PE (clone 1A1, BioLegend, cat#311906), EBI2 (GPR183) PE (clone SA313E4, BioLegend, cat#368912); CD11c APC (clone 3.9, BioLegend, cat#301614); CD27 PE-Cy7 (clone O323, eBioscience, cat#25-0279-42); IgD APC-H7 (clone IA6-2, BD Biosciences, cat#561305). Isotype control antibodies used for flow cytometry were: Mouse IgG2a BV510 (clone MOPC-173, BioLegend, cat#400268); Mouse IgG1 FITC (clone MOPC-21, BioLegend, cat#400110); Rat IgG2a PE (clone R35-95, BD Biosciences, cat#553930); Mouse IgG2a PE (clone MOPC-173, BioLegend, cat#400214); Mouse IgG1 APC (clone MOPC-21, BioLegend, cat#400120); Mouse IgG1 PE-Cy7 (clone MOPC-21, BioLegend, cat#400126); Mouse IgG2a APC-H7 (clone G155-178, BD Biosciences, cat#560897).

Phosphoarray analysis

Proteome Profiler™ Human Phospho-Kinase Array Kits (R&D Systems) were utilized according to the manufacturer's instructions to analyze whole cell lysates of purified, untreated B cell samples from patients with Active cGVHD (*n*=3) or No cGVHD (*n*=3). Spot densities on the arrays were quantified using ImageJ software (https://imagej.nih.gov/ij/).

Western blot analysis

Frozen (-80°C) pellets of purified B cells from allo-HCT patients with Active cGVHD (*n*=4) or No cGVHD (*n*=4) were lysed using a commercially available whole cell lysis buffer (Roche Complete Lysis-M) with freshly added protease inhibitors (Roche). Protein lysates were then denatured and loaded into wells of a 4–12% Bis-Tris gel (ThermoFisher) for electrophoresis under reducing conditions. Proteins were transferred from the gels to a nitrocellulose membrane using dry electroblotting (20 V for 9 min). Membranes were blocked in 2% fish gelatin buffer for 75 min and then incubated with rabbit anti-human $P27^{KIP1}$ polyclonal antibody (C-19, Santa Cruz Biotechnology, cat#sc-528, lot#KO413) overnight at 4°C. The membrane was thoroughly washed in Tris-buffered saline containing Tween 20 (TBST) before incubation for 1 h with donkey anti-rabbit polyclonal secondary antibody labeled with a near-infrared fluorescent dye (IRDye® 680, LI-COR, cat#926-68073, lot#C50821-05). The membrane was then washed thoroughly in TBST, and fluorescently labeled proteins were detected using an LI-COR Odyssey CLX Imaging System.

Supplementary Materials

figs. S1 to S6

tables S1 to S7

References and Notes

1. H. Wardemann *et al.*, Predominant autoantibody production by early human B cell precursors. *Science* **301**, 1374-1377 (2003).
2. H. Jin *et al.*, Antibodies from donor B cells perpetuate cutaneous chronic graft-versus-host disease in mice. *Blood*, (2016).
3. M. Srinivasan *et al.*, Donor B-cell alloantibody deposition and germinal center formation are required for the development of murine chronic GVHD and bronchiolitis obliterans. *Blood* **119**, 1570-1580 (2012).
4. R. Zeiser, S. Sarantopoulos, B. R. Blazar, B-cell targeting in chronic graft-versus-host disease. *Blood* **131**, 1399-1405 (2018).
5. A. de Masson *et al.*, CD24(hi)CD27(+) and plasmablast-like regulatory B cells in human chronic graft-versus-host disease. *Blood* **125**, 1830-1839 (2015).
6. D. Le Huu *et al.*, FTY720 ameliorates murine sclerodermatous chronic graft-versus-host disease by promoting expansion of splenic regulatory cells and inhibiting immune cell infiltration into skin. *Arthritis Rheum.* **65**, 1624-1635 (2013).
7. D. Le Huu *et al.*, Donor-derived regulatory B cells are important for suppression of murine sclerodermatous chronic graft-versus-host disease. *Blood* **121**, 3274-3283 (2013).
8. J. L. Allen *et al.*, B cells from patients with chronic GVHD are activated and primed for survival via BAFF-mediated pathways. *Blood* **120**, 2529-2536 (2012).
9. W. Jia *et al.*, BAFF Promotes Heightened BCR Responsiveness and Manifestations of Chronic GVHD after Allogeneic Stem Cell Transplantation. *Blood*, (2021).
10. J. C. Poe *et al.*, An aberrant NOTCH2-BCR signaling axis in B cells from patients with chronic GVHD. *Blood* **130**, 2131-2145 (2017).
11. S. Sarantopoulos, J. Ritz, Aberrant B-cell homeostasis in chronic GVHD. *Blood* **125**, 1703-1707 (2015).
12. S. Sarantopoulos *et al.*, Altered B-cell homeostasis and excess BAFF in human chronic graft-versus-host disease. *Blood* **113**, 3865-3874 (2009).
13. A. N. Suthers, S. Sarantopoulos, TLR7/TLR9- and B Cell Receptor-Signaling Crosstalk: Promotion of Potentially Dangerous B Cells. *Front. Immunol.* **8**, 775 (2017).
14. T. Matsushita *et al.*, BAFF inhibition attenuates fibrosis in scleroderma by modulating the regulatory and effector B cell balance. *Sci Adv* **4**, eaas9944 (2018).
15. E. Zorn *et al.*, Reduced frequency of FOXP3+ CD4+CD25+ regulatory T cells in patients with chronic graft-versus-host disease. *Blood* **106**, 2903-2911 (2005).
16. E. Tivol, R. Komorowski, W. R. Drobyski, Emergent autoimmunity in graft-versus-host disease. *Blood* **105**, 4885-4891 (2005).
17. J. L. Allen *et al.*, Increased BCR responsiveness in B cells from patients with chronic GVHD. *Blood* **123**, 2108-2115 (2014).
18. R. Flynn *et al.*, Increased T follicular helper cells and germinal center B cells are required for cGVHD and bronchiolitis obliterans. *Blood* **123**, 3988-3998 (2014).

19. J. C. Poe *et al.*, SYK inhibitor entospletinib prevents ocular and skin GVHD in mice. *JCI Insight* **3**, (2018).
20. C. Zhang *et al.*, Donor CD4+ T and B cells in transplants induce chronic graft-versus-host disease with autoimmune manifestations. *Blood* **107**, 2993-3001 (2006).
21. B. R. Blazar, W. J. Murphy, M. Abedi, Advances in graft-versus-host disease biology and therapy. *Nat. Rev. Immunol.* **12**, 443-458 (2012).
22. A. Khoder *et al.*, Evidence for B Cell Exhaustion in Chronic Graft-versus-Host Disease. *Front. Immunol.* **8**, 1937 (2017).
23. X. Gao *et al.*, ZEB2 regulates the development of CD11c+ atypical B cells. *bioRxiv*, 2022.2009.2001.506173 (2022).
24. R. Yang *et al.*, Human T-bet governs the generation of a distinct subset of CD11c(high)CD21(low) B cells. *Sci Immunol* **7**, eabq3277 (2022).
25. B. Keller *et al.*, The expansion of human T-bet(high)CD21(low) B cells is T cell dependent. *Sci Immunol* **6**, eabh0891 (2021).
26. N. Obeng-Adjei *et al.*, Malaria-induced interferon-gamma drives the expansion of Tbethi atypical memory B cells. *PLoS Pathog.* **13**, e1006576 (2017).
27. T. Stuart *et al.*, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).
28. A. Tumanov *et al.*, Distinct role of surface lymphotoxin expressed by B cells in the organization of secondary lymphoid tissues. *Immunity* **17**, 239-250 (2002).
29. E. Vigorito *et al.*, microRNA-155 regulates the generation of immunoglobulin class-switched plasma cells. *Immunity* **27**, 847-859 (2007).
30. P. Holla *et al.*, Shared transcriptional profiles of atypical B cells suggest common drivers of expansion and function in malaria, HIV, and autoimmunity. *Sci Adv* **7**, (2021).
31. H. J. Sutton *et al.*, Atypical B cells are part of an alternative lineage of B cells that participates in responses to vaccination and infection in humans. *Cell Rep.* **34**, 108684 (2021).
32. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
33. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27-30 (2000).
34. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25-29 (2000).
35. P. E. Lipsky, Systemic lupus erythematosus: an autoimmune disease of B cell hyperactivity. *Nat. Immunol.* **2**, 764-766 (2001).
36. H. Rincon-Arevalo *et al.*, Deep Phenotyping of CD11c(+) B Cells in Systemic Autoimmunity and Controls. *Front. Immunol.* **12**, 635615 (2021).
37. A. V. Rubtsov *et al.*, CD11c-Expressing B Cells Are Located at the T Cell/B Cell Border in Spleen and Are Potent APCs. *J. Immunol.* **195**, 71-79 (2015).
38. Y. Hao, P. O'Neill, M. S. Naradikian, J. L. Scholz, M. P. Cancro, A B-cell subset uniquely responsive to innate stimuli accumulates in aged mice. *Blood* **118**, 1294-1304 (2011).
39. I. Sanz *et al.*, Challenges and Opportunities for Consistent Classification of Human B Cell and Plasma Cell Populations. *Front. Immunol.* **10**, 2458 (2019).
40. G. J. Brown *et al.*, TLR7 gain-of-function genetic variation causes human lupus. *Nature* **605**, 349-356 (2022).

41. S. A. Jenks, K. S. Cashman, M. C. Woodruff, F. E. Lee, I. Sanz, Extrafollicular responses in humans and SLE. *Immunol. Rev.* **288**, 136-148 (2019).
42. S. A. Jenks *et al.*, Distinct Effector B Cells Induced by Unregulated Toll-like Receptor 7 Contribute to Pathogenic Responses in Systemic Lupus Erythematosus. *Immunity* **49**, 725-739 e726 (2018).
43. M. C. Woodruff *et al.*, Extrafollicular B cell responses correlate with neutralizing antibodies and morbidity in COVID-19. *Nat. Immunol.* **21**, 1506-1516 (2020).
44. H. T. Greinix *et al.*, Elevated numbers of immature/transitional CD21- B lymphocytes and deficiency of memory CD27+ B cells identify patients with active chronic graft-versus-host disease. *Biol. Blood Marrow Transplant.* **14**, 208-219 (2008).
45. I. Isnardi *et al.*, Complement receptor 2/CD21- human naive B cells contain mostly autoreactive unresponsive clones. *Blood* **115**, 5026-5036.
46. S. Moir, A. S. Fauci, B-cell exhaustion in HIV infection: the role of immune activation. *Curr. Opin. HIV AIDS* **9**, 472-477 (2014).
47. Y. Li, Z. Li, F. Hu, Double-negative (DN) B cells: an under-recognized effector memory B cell subset in autoimmunity. *Clin. Exp. Immunol.* **205**, 119-127 (2021).
48. C. Wei *et al.*, A new population of cells lacking expression of CD27 represents a notable component of the B cell memory compartment in systemic lupus erythematosus. *J. Immunol.* **178**, 6624-6633 (2007).
49. E. Masle-Farquhar *et al.*, Uncontrolled CD21(low) age-associated and B1 B cell accumulation caused by failure of an EGR2/3 tolerance checkpoint. *Cell Rep.* **38**, 110259 (2022).
50. J. Brummelman *et al.*, Development, application and computational analysis of high-dimensional fluorescent antibody panels for single-cell flow cytometry. *Nat. Protoc.* **14**, 1946-1969 (2019).
51. X. Liu *et al.*, A comparison framework and guideline of clustering methods for mass cytometry data. *Genome Biol.* **20**, 297 (2019).
52. F. Corrente *et al.*, CD21(-) CD27(-) Atypical B Cells in a Pediatric Cohort Study: An Extensive Single Center Flow Cytometric Analysis. *Front Pediatr* **10**, 822400 (2022).
53. A. Arazi *et al.*, The immune cell landscape in kidneys of patients with lupus nephritis. *Nat. Immunol.* **20**, 902-914 (2019).
54. D. R. Glass *et al.*, An Integrated Multi-omic Single-Cell Atlas of Human B Cell Identity. *Immunity* **53**, 217-232 e215 (2020).
55. D. Nehar-Belaid *et al.*, Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nat. Immunol.* **21**, 1094-1106 (2020).
56. F. Zhang *et al.*, Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat. Immunol.* **20**, 928-942 (2019).
57. A. V. Sauer *et al.*, Defective B cell tolerance in adenosine deaminase deficiency is corrected by gene therapy. *J. Clin. Invest.* **122**, 2141-2152 (2012).
58. C. Schaniel *et al.*, Activated murine B lymphocytes and dendritic cells produce a novel CC chemokine which acts selectively on activated T cells. *J. Exp. Med.* **188**, 451-463 (1998).
59. M. L. Allende *et al.*, S1P1 receptor directs the release of immature B cells from bone marrow into blood. *J. Exp. Med.* **207**, 1113-1124 (2010).
60. T. I. Arnon, R. M. Horton, I. L. Grigorova, J. G. Cyster, Visualization of splenic marginal zone B-cell shuttling and follicular B-cell egress. *Nature* **493**, 684-688 (2013).

61. J. A. Green *et al.*, The sphingosine 1-phosphate receptor S1P(2) maintains the homeostasis of germinal center B cells and promotes niche confinement. *Nat. Immunol.* **12**, 672-680 (2011).
62. C. Huang *et al.*, The BCL6 RD2 domain governs commitment of activated B cells to form germinal centers. *Cell Rep.* **8**, 1497-1508 (2014).
63. I. Y. Hwang, C. Park, K. Harrison, J. H. Kehrl, Biased S1PR1 Signaling in B Cells Subverts Responses to Homeostatic Chemokines, Severely Disorganizing Lymphoid Organ Architecture. *J. Immunol.* **203**, 2401-2414 (2019).
64. G. Muller, M. Lipp, Shaping up adaptive immunity: the impact of CCR7 and CXCR5 on lymphocyte trafficking. *Microcirculation* **10**, 325-334 (2003).
65. G. Nunez, D. Hockenbery, T. J. McDonnell, C. M. Sorensen, S. J. Korsmeyer, Bcl-2 maintains B cell memory. *Nature* **353**, 71-73 (1991).
66. A. Kieser, Pursuing different 'TRADDes': TRADD signaling induced by TNF-receptor 1 and the Epstein-Barr virus oncoprotein LMP1. *Biol. Chem.* **389**, 1261-1271 (2008).
67. K. Inada *et al.*, c-Fos induces apoptosis in germinal center B cells. *J. Immunol.* **161**, 3853-3861 (1998).
68. K. Kobayashi *et al.*, Overexpression of c-fos inhibits down-regulation of a cyclin-dependent kinase-2 inhibitor p27Kip1 in splenic B cells activated by surface Ig cross-linking. *J. Immunol.* **158**, 2050-2056 (1997).
69. B. Grotsch *et al.*, The AP-1 transcription factor Fra1 inhibits follicular B cell differentiation into plasma cells. *J. Exp. Med.* **211**, 2199-2212 (2014).
70. M. Mandal *et al.*, CXCR4 signaling directs Igk recombination and the molecular mechanisms of late B lymphopoiesis. *Nat. Immunol.* **20**, 1393-1403 (2019).
71. C. D. Allen *et al.*, Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5. *Nat. Immunol.* **5**, 943-952 (2004).
72. D. Gatto, D. Paus, A. Basten, C. R. Mackay, R. Brink, Guidance of B cells by the orphan G protein-coupled receptor EBI2 shapes humoral immune responses. *Immunity* **31**, 259-269 (2009).
73. A. E. Hauser *et al.*, Chemotactic responsiveness toward ligands for CXCR3 and CXCR4 is regulated on plasma blasts during the time course of a memory immune response. *J. Immunol.* **169**, 1277-1282 (2002).
74. J. R. Muppidi, E. Lu, J. G. Cyster, The G protein-coupled receptor P2RY8 and follicular dendritic cells promote germinal center confinement of B cells, whereas S1PR3 can contribute to their dissemination. *J. Exp. Med.* **212**, 2213-2222 (2015).
75. J. P. Pereira, L. M. Kelly, Y. Xu, J. G. Cyster, EBI2 mediates B cell segregation between the outer and centre follicle. *Nature* **460**, 1122-1126 (2009).
76. E. Lu, F. D. Wolfreys, J. R. Muppidi, Y. Xu, J. G. Cyster, S-Geranylgeranyl-L-glutathione is a ligand for human B cell-confinement receptor P2RY8. *Nature* **567**, 244-248 (2019).
77. Y. He *et al.*, P2RY8 variants in lupus patients uncover a role for the receptor in immunological tolerance. *J. Exp. Med.* **219**, (2022).
78. R. Deng *et al.*, Extrafollicular CD4(+) T-B interactions are sufficient for inducing autoimmune-like chronic graft-versus-host disease. *Nat Commun* **8**, 978 (2017).
79. D. E. Kennedy *et al.*, Novel specialized cell state and spatial compartments within the germinal center. *Nat. Immunol.* **21**, 660-670 (2020).
80. A. Yoshizaki *et al.*, Regulatory B cells control T-cell autoimmunity through IL-21-dependent cognate interactions. *Nature* **491**, 264-268 (2012).

81. G. Koncz, A. O. Hueber, The Fas/CD95 Receptor Regulates the Death of Autoreactive B Cells and the Selection of Antigen-Specific B Cells. *Front. Immunol.* **3**, 207 (2012).
82. J. K. Lee, S. O. Mathew, S. V. Vaidya, P. R. Kumaresan, P. A. Mathew, CS1 (CRACC, CD319) induces proliferation and autocrine cytokine expression on human B lymphocytes. *J. Immunol.* **179**, 4672-4678 (2007).
83. S. Sarantopoulos *et al.*, High levels of B-cell activating factor in patients with active chronic graft-versus-host disease. *Clin. Cancer Res.* **13**, 6107-6114 (2007).
84. Y. Saito *et al.*, B-cell-activating factor inhibits CD20-mediated and B-cell receptor-mediated apoptosis in human B cells. *Immunology* **125**, 570-590 (2008).
85. S. Chevrier *et al.*, The BTB-ZF transcription factor Zbtb20 is driven by Irf4 to promote plasma cell differentiation and longevity. *J. Exp. Med.* **211**, 827-840 (2014).
86. Y. Wang, D. Bhattacharya, Adjuvant-specific regulation of long-term antibody responses by ZBTB20. *J. Exp. Med.* **211**, 841-856 (2014).
87. S. Minguet *et al.*, Caveolin-1-dependent nanoscale organization of the BCR regulates B cell tolerance. *Nat. Immunol.* **18**, 1150-1159 (2017).
88. N. Bhat *et al.*, Regnase-1 is essential for B cell homeostasis to prevent immunopathology. *J. Exp. Med.* **218**, (2021).
89. D. J. Hodson *et al.*, Regulation of normal B-cell differentiation and malignant B-cell survival by OCT2. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E2039-2046 (2016).
90. G. Pearce *et al.*, Signaling protein SWAP-70 is required for efficient B cell homing to lymphoid organs. *Nat. Immunol.* **7**, 827-834 (2006).
91. G. Pearce, T. Audzevich, R. Jessberger, SYK regulates B-cell migration by phosphorylation of the F-actin interacting protein SWAP-70. *Blood* **117**, 1574-1584 (2011).
92. M. Yang *et al.*, AIM2 deficiency in B cells ameliorates systemic lupus erythematosus by regulating Blimp-1-Bcl-6 axis-mediated B-cell differentiation. *Signal Transduct Target Ther* **6**, 341 (2021).
93. K. Street *et al.*, Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
94. F. C. Guibal *et al.*, ASB-2 inhibits growth and promotes commitment in myeloid leukemia cells. *J. Biol. Chem.* **277**, 218-224 (2002).
95. H. E. Parge, A. S. Arvai, D. J. Murtari, S. I. Reed, J. A. Tainer, Human CksHs2 atomic structure: a role for its hexameric assembly in cell cycle control. *Science* **262**, 387-395 (1993).
96. M. Frontini *et al.*, The CDK subunit CKS2 counteracts CKS1 to control cyclin A/CDK2 activity in maintaining replicative fidelity and neurodevelopment. *Dev. Cell* **23**, 356-370 (2012).
97. K. Hiromura, J. W. Pippin, M. L. Fero, J. M. Roberts, S. J. Shankland, Modulation of apoptosis by the cyclin-dependent kinase inhibitor p27(Kip1). *J. Clin. Invest.* **103**, 597-604 (1999).
98. M. D. Larrea *et al.*, RSK1 drives p27Kip1 phosphorylation at T198 to promote RhoA inhibition and increase cell motility. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9268-9273 (2009).
99. J. D. Short *et al.*, AMPK-mediated phosphorylation of murine p27 at T197 promotes binding of 14-3-3 proteins and increases p27 stability. *Mol. Carcinog.* **49**, 429-439 (2010).

100. A. V. Rubtsov *et al.*, Toll-like receptor 7 (TLR7)-driven accumulation of a novel CD11c(+) B-cell population is important for the development of autoimmunity. *Blood* **118**, 1305-1315 (2011).
101. C. D. Scharer *et al.*, Epigenetic programming underpins B cell dysfunction in human SLE. *Nat. Immunol.* **20**, 1071-1082 (2019).
102. S. Wang *et al.*, IL-21 drives expansion and plasma cell differentiation of autoreactive CD11c(hi)T-bet(+) B cells in SLE. *Nat Commun* **9**, 1758 (2018).
103. E. Zumaquero *et al.*, IFNgamma induces epigenetic programming of human T-bet(hi) B cells and promotes TLR7/8 and IL-21 induced differentiation. *Elife* **8**, (2019).
104. S. Sarantopoulos *et al.*, Recovery of B-cell homeostasis after rituximab in chronic graft-versus-host disease. *Blood* **117**, 2275-2283 (2011).
105. N. F. Smirnova *et al.*, Inhibition of B cell-dependent lymphoid follicle formation prevents lymphocytic bronchiolitis after lung transplantation. *JCI Insight* **4**, (2019).
106. R Core Team. (R Foundation for Statistical Computing, Vienna, Austria, 2019).
107. S. Anders, W. Huber, Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
108. J. Bass, A. Swcf, A. Dabney, D. Robinson, qvalue: Q-value estimation for false discovery rate control. *R package version 2.8.0, http://github.com/jdstorey/qvalue*, (2015).
109. A. W. Zhang *et al.*, Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. *Nat Methods* **16**, 1007-1015 (2019).
110. H. Guo, J. Li, scSorter: assigning cells to known cell types according to marker genes. *Genome Biol.* **22**, 69 (2021).
111. S. L. a. C. Paul Hoffman. (2022), vol. 2022.
112. R. C. Team. (R Foundation for Statistical Computing, Vienna, Austria, 2021).
113. R. Gentleman *et al.*, Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**, R80 (2004).
114. Y. Xie, *Dynamic documents with R and knitr*. (CRC Press/Taylor and Francis, Boca Raton, ed. Second edition., 2015), pp. xxvii, 266.
115. K. Ram, Git can facilitate greater reproducibility and increased transparency in science. *Source Code Biol. Med.* **8**, 7 (2013).

Acknowledgments:

The authors wish to thank Laney Weber, Certified Editor in the Life Sciences, for the critical review of this manuscript and assistance with formatting. The authors also wish to thank Lauren S. Riley and Serach Patterson for assistance with proofreading.

Funding:

National Institutes of Health (NHLBI) grant R01HL129061 (SS)

Leukemia & Lymphoma Society grant 6497-16 (SS)

National Institutes of Health (NIAID) grant R38AI140297 (SJB)

European Union's Horizon 2020 Research and Innovation Program under the Marie Sklodowska-Curie grant agreement No 888743 (JV, SS)

Author contributions:

Conceptualization: JCP, SS

Data curation: JF, MRL, KO

Formal analysis: JF, DZ, MRL, XQ, JX, KO

Funding acquisition: SS, SJB, JV

Investigation: JCP, RAD, HS, JZ, ANS, ICC

Methodology: JF, DZ, MRL, RAD, XQ, JV, JX, KO

Resources: VTH, KSW, JJR, SZP, FTH, MEH, DAR, ARC

Software: JF, DZ, MRL, XQ, JX, KO

Supervision: SS, KO

Validation: JF, MRL, XQ, KO

Visualization: JCP, JF, JV, KO, SS

Writing – original draft: JCP, SS, KO

Writing – review & editing: JF, DZ, MRL, RAD, XQ, JZ, JV, SJB, WJ, WM, NJC, ARC, JX

Competing interests: The authors report no conflict of interest related to this study.

Data and materials availability: All scRNA-Seq library data will be made available through the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) following peer review. The R scripts to reproduce the analyses of these scRNA-Seq data will be made available following peer review.