Single-cell analysis of human B cell maturation predicts how antibody class switching shapes selection dynamics.

Hamish W King[1,2] *, Nara Orban[3], John C Riches[4,5], Andrew J Clear[4], Gary Warnes[6], Sarah A Teichmann[2,7], Louisa K James[1] *

[1] Centre for Immunobiology, Blizard Institute, Queen Mary University of London, London E1 2AT, UK
[2] Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK
[3] Barts Health Ear, Nose and Throat Service, The Royal London Hospital, London E1 1BB, UK
[4] Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London EC1M 6BQ, UK
[5] The Francis Crick Institute, London NW1 1AT, UK
[6] Flow Cytometry Core Facility, Blizard Institute, Queen Mary University of London, London E1 2AT, UK
[7] Theory of Condensed Matter, Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0EH, UK
* To whom correspondence can be addressed: h.king@qmul.ac.uk, drhamishking@gmail.com and louisa.james@qmul.ac.uk

Short title: A single-cell roadmap of human B cell maturation.

One Sentence Summary: Integrated single-cell transcriptomic and BCR analyses reveal how antibody class switching influences human B cell fate and function.

Abstract

Protective humoral memory forms in secondary lymphoid organs where B cells undergo affinity maturation and differentiation into memory or plasma cells. Here, we provide a comprehensive roadmap of human B cell maturation with single-cell transcriptomics matched with bulk and single-cell antibody repertoires to define gene expression, antibody repertoires and clonal sharing of B cell states at single-cell resolution, including memory B cell heterogeneity that reflects diverse functional and signalling states. We reconstruct gene expression dynamics during B cell activation to reveal a pre-germinal centre state primed to undergo class switch recombination and dissect how antibody class-dependent gene expression in germinal centre and memory B cells is linked with a unique transcriptional wiring with potential to influence their fate and function. Our analyses reveal the dynamic cellular states that shape human B cell-mediated immunity and highlight how antibody isotype may play a role during their antibody-based selection.

Introduction

Protective humoral immune responses require the formation of a functional antibody repertoire by B cells within secondary lymphoid organs (SLOs) such as the spleen, peripheral lymph nodes and tonsils. After antigen encounter, activated B cells either differentiate into short-lived plasma cells or following cognate interaction with antigen-specific T cells can form germinal centres (GCs) (*1*). GCs are transient structures where B cells undergo iterative cycles of clonal expansion and somatic hypermutation (SHM) in the variable regions of their immunoglobulin heavy (IgH) and light chain genes followed by affinity-based selection of antigen-specific clones. This dynamic process occurs in spatially distinct dark and light zones (DZ and LZ) under the regulation of a network of specialised T follicular helper cells, follicular dendritic cells and macrophages (*2*).

B cells differentiate and exit the GC either as antibody-secreting plasmablasts committed to the plasma cell lineage or memory B cells (MBC), which are long-lived quiescent cells capable of being reactivated upon antigen re-exposure (*3*). Antibody effector functions are broadly determined by their isotype, which is specified by the constant domain genes in the IgH locus. During maturation B cells may undergo class switch recombination (CSR), which involves deletional recombination of IgM and IgD constant domain genes and expression of a different downstream constant domain gene (IgG1-4, IgA1-2, IgE) (*4*). Specific hierarchies in isotype switching patterns have been reported in peripheral blood (*5*). Antibody repertoires and B cell-mediated immunity are therefore shaped by the combined influences of cell fate decisions, affinity maturation and CSR.

B cells express their antibody immunoglobulin genes as part of a membrane-bound complex termed the B cell receptor (BCR). During maturation in GCs, antigen-binding and downstream signalling of the BCR is a primary determinant of B cell survival and differentiation (*6, 7*). BCR activation thresholds and downstream signalling can differ due to isotype-specific differences in extracellular, transmembrane and intracellular domains of immunoglobulin proteins forming the BCR (*8-10*). Thus, as well as shaping the effector functions of subsequent antibody repertoires, CSR may influence B cell survival or fate

specification during affinity maturation. However, deconvolution of SHM, class switching and diverse cell states in the polyclonal context of human SLOs *in vivo* has proved a challenge (*2*).

To improve our understanding of the cellular, transcriptional and antibody repertoire dynamics during human B cell maturation *in vivo*, we performed unbiased single-cell transcriptomic and repertoire profiling of a widely studied model SLO, the human tonsil (https://www.tonsilimmune.org/). These integrated bulk and single-cell antibody repertoires paired with single-cell transcriptomics allowed us to define transcriptional B cell states, including a detailed exploration of MBC heterogeneity, reconstruct temporal gene expression dynamics and resolve unique functional capabilities of GC and MBCs linked to antibody class switching. This roadmap of human B cell maturation provides unprecedented resolution of the formation of B cell-mediated immunity and is a unique resource to study both normal and pathological B cell responses.

Results

Defining the antibody class switch recombination landscape of human tonsillar B cells

The majority of high-throughput antibody repertoire sequencing studies in human have examined peripheral blood B cell subsets, which significantly differ to dynamic repertoires in active SLOs. To provide a subclass-specific and quantitative analysis of the antibody repertoire landscape in human tonsils, we sequenced the antibody repertoires of four broadly defined tonsillar B cell subsets; naïve, GC, memory and plasmablasts using a unique molecular identifier-indexed protocol (Fig1A) (*5*). We analysed tonsillar B cells from both recurrent tonsillectomy and obstructive sleep apnoea patients and found good agreement between bulk FACS-sorted B cell repertoires and donor-matched unsorted single-cell antibody repertoires (FigS1, Fig2). As expected, naïve B cell repertoires were predominantly unswitched and unmutated. GC and MBCs comprised both switched and unswitched IgH sequences with elevated SHM and plasmablasts were nearly all switched and highly mutated (Fig1B-C).

Higher IgH SHM frequencies within the GC typically reflect higher affinity BCRs and are proposed to bias GC B cells towards the plasmablast rather than the MBC fate (*11, 12*). In keeping with this,

plasmablast-derived repertoires in tonsil generally contained higher SHM frequencies than those of MBCs (Fig1C). However, by resolving for antibody subclass, we found that SHM levels were broadly similar between different mature B cell subsets, with the exception of IgD and increased in frequency along the immunoglobulin locus (Fig1D). Comparison of the clonal diversity of subclass-specific MBCs and plasmablasts revealed unswitched and IgA^+ MBCs were less clonally expanded (as evidenced by higher diversity) than plasmablasts of the same isotype, while IgG^+ MBCs and plasmablasts appeared to have clonally expanded to similar degrees (Fig1E). This is not likely explained by differences in SHM frequencies (Fig1D), but may reflect differences in their selection or cell fate specification linked with the outcome of CSR. Indeed, when we examined antibody subclass frequencies in these broadly defined B cell populations we found that as well as an increased propensity to retain IgM expression (Fig1B), MBCs were 3.3- or 7.3-fold more likely than plasmablasts to express IgA1 or IgA2 respectively, while plasmablasts were more likely to express IgG1 (Fig1F). Intriguingly, these enrichments were linked with specific B cell fates even for expanded clones spanning different subsets (Fig1G, FigS2).

To explore how these subset-specific class switching patterns might arise, we reconstructed phylogenies for 28,845 expanded clones and calculated the likelihood that specific CSR events had occurred between closely related B cells compared to that expected by chance (Fig1H), similar to analyses performed in peripheral blood B cells (*5*). MBC clonal lineages exhibited greater likelihoods for switching of isotype pairs closely located in linear space along the IgH locus, compared to plasmablast clones which demonstrated a more eclectic pattern of isotype switching (Fig1H). Of note, both the antibody subclass frequencies and reconstructed class switch hierarchies of MBCs closely resembled those of GC cells, consistent with models that propose a stochastic exit of MBCs from the GC (*13, 14*), in contrast to active selection of the plasma cell fate. Together, these analyses provide evidence that antibody-based selection mechanisms differ for two major B cell fates and that this appears related to the outcomes of CSR earlier in maturation.

A single-cell atlas of tonsillar immune cells defines diverse B cell states during maturation

FACS-based strategies to study dynamic cell processes in human tissues may miss rare or unknown cell populations. Therefore, in parallel to our bulk antibody repertoires (Fig1), we performed single-cell RNA-seq (scRNA-seq) paired with single-cell immunoglobulin VDJ sequencing (scVDJ-seq) for unsorted tonsillar immune cells from the same samples (Fig2A-B). After stringent quality control, we retained the transcriptomes of 32,607 cells (n=7; median of 3142 and mean of 4658 cells per donor) from which we identified 30 distinct cell populations (Fig2C-D, TableS1). Although our primary focus was understanding B cell maturation, we annotated 11 T cell and 7 non-lymphoid populations (Fig2C, S3A-D, TableS2-3) as a valuable resource to study these cells. Importantly, we did not identify any populations unique to recurrent tonsillitis or obstructive sleep apnoea patients (FigS3E).

Tonsils are a model SLO capable of inducing both systemic and mucosal immunity. We found that major tonsillar immune cell states were broadly conserved in other SLOs such as lymph nodes and spleen (FigS4), similar to other analyses (*15*), although we observed enrichment of GC-associated cell states in paediatric tonsils that likely reflect differences in donor ages (FigS4B-D). We also observed differences in antibody features (FigS4E-F) that may reflect tissue-specific difference in local cytokine environments. The broad conservation of cell states across different SLOs supports the use of tonsils as a model system to study systemic immunity and highlights the importance of profiling immunologically-active tissues to understand GC-dependent B cell maturation.

We next examined B cell states within our single-cell transcriptomic atlas. To improve the power and accuracy of our single-cell antibody repertoire analyses, we integrated single-cell and bulk repertoires for each donor (Fig2E). We identified 12 distinct B cell populations based on unbiased clustering of gene expression (Fig2F, S5, Table S4) and quantified their antibody isotype frequencies (Fig2G, S6A), SHM levels (Fig2H, S6B), clonal diversity (Fig2I-J) and relationships with other B cell subsets (Fig2K-L). All populations were reproducibly observed across patients, regardless of tonsillitis history (Fig2M, S3E). We identified all major stages of B cell maturation in human tonsils, including naïve, activated, GC (including both LZ and non-proliferating DZ cells), MBCs, tissue-resident FCRL4$^+$ MBCs,

plasmablasts and a cycling population consisting mostly of DZ GC cells (Fig2H). We also annotated a class-switched, hypermutated and clonally-expanded GC B cell subset expressing plasmablast-specific transcription factors (*PRDM1*, *XBP1*) (Fig2F-J, S5), consistent with a pre-plasmablast (prePB) population reported recently in tonsillar GC B cells (FigS5C) (*16*). We found that prePB GC cells uniquely expressed several interesting gene markers, including signalling molecules like *FRZB* and *BTNL9* (Fig2F, S5A), although the functional relevance of this is unknown.

Our unbiased analysis of B cell states revealed two additional populations of particular interest. First, we define a naïve "preGC" B cell state expressing unmutated IgM and IgD that transcriptionally share markers with both naïve and LZ GC populations (FigS5A-B), but have yet to acquire features consistent with B cell maturation in the GC such as *CD27* and *CD38* expression, hypermutated antibody genes or clonal expansion (Fig2F-J, S5B). Flow cytometry of several marker genes expressed by these cells (CD23 (*FCER2*), CD108 (*SEMA7A*) and CD58) confirmed the enrichment of a similar state in naïve B cell populations (FigS7). Secondly, we discover a transcriptionally distinct and clonally-expanded IgM^{+} B cell population in the GC with elevated expression of genes associated with inhibitory BCR signalling, such as *FCRL3*, *FCRL2*, *SAMSN1*, and *SIGLEC10* (Fig2D,F-J), that we refer to as $FCRL3^{high}$ GC B cells. While these cells also expressed high levels of the proposed pre-MBC marker CCR6 (*12*), they typically had inconsistent expression of other recently defined pre-MBC markers (FigS5C) (*16*). As this cluster of unswitched cells was defined by very high *FCRL2* and *FCRL3* expression (Fig3F), we examined these markers by flow cytometry of tonsillar B cells and found a strong relationship between surface expression of FCRL3, but not FCRL2, and IgM expression in both GC and MBC populations (FigS8A-F). We also identified a rare population of $FCRL3^{high}$ GC B cells by immunohistochemistry (FigS8G-H). Although the precise ontology of these cells remains unclear, scVDJ-based analyses found that $FCRL3^{high}$ GC B cells are part of highly expanded GC-derived clones that contain both MBCs or plasmablasts (Fig2K-L, FigS9), indicating that they arise through productive GC reactions and are unlikely to be derived from a separate or committed lineage.

Pseudotemporal reconstruction of human B cell activation and GC formation

During activation, B cells acquire antigen, either in soluble form or displayed on the surface of follicular antigen-presenting cells (APCs), allowing them to migrate to the T cell zone where they can participate in GC reactions. We reasoned that our single-cell atlas of B cell maturation would contain a full spectrum of activation states and allow us to reconstruct temporal gene expression and antibody repertoire dynamics during B cell activation and formation of the GC response. Our scRNA-seq analysis identified a B cell cluster composed of both antigen-inexperienced naïve B cells and antigen-experienced MBCs (FigS10) that had elevated expression of activation marker genes (Fig3A) and high frequency of predicted cell-cell interactions with APCs (Fig3B). This activated B cell state appeared capable of communicating with both APCs and T cells through ICAM1-ITGAL1 (LFA-1) and/or IL6, respectively (Fig3C). Many of these predicted cell-cell interactions were also detected in preGC and LZ GC B cell states (Fig3C) before being lost in other GC B cell populations, suggesting that preGC B cells might exist as a transitional state between initial B cell activation and GC formation.

To better understand the transcriptional dynamics of different B cell populations, particularly the preGC B cell state, we performed RNA velocity analysis (*17, 18*) to model transcriptional kinetics of individual B cells and infer their directionality and relationship with other B cell populations. This confirmed that preGC cells exhibit a strong directionality towards the LZ GC state (Fig3D), fitting with their transcriptional relatedness (FigS5A) and independent partition-based graph analysis (FigS11). Velocity-based pseudotemporal ordering of naïve, activated, preGC and LZ GC B cells placed preGC cells between activated and LZ GC B cells and revealed the full continuum of gene expression from early activation events to *bona fide* GC B cells (Fig3E-F). This included the dynamic expression of key signalling molecules and transcription factors such as *CD40*, *EBI3*, *MIF*, *BATF*, *BHLHE40* and *MYC* (FigS12A). Crucially, we found that the activated B cell cluster (including both activated naïve and MBC sub-populations) (FigS10) placed early in the trajectory and were enriched for an experimentally-derived gene signature associated with acute (1hr) B cell stimulation, while preGC B cells more closely resembled intermediate level B cell stimulation (3-6hr) (Fig3G) (*19*). This confirms that these B cells represent a secondary activation state following antigen encounter and before formation of GCs. Our

reconstruction of B cell activation therefore provides a valuable framework to study these dynamic B cell states *in vivo*.

While it was long held that CSR occurs exclusively within the GC, several studies in mice have demonstrated that CSR can occur prior to the GC response (*20-22*). However, the precise B cell state in which this occurs has proved elusive in humans. Our reconstruction of *in vivo* human B cell activation revealed a surprising enrichment of genes associated with CSR in preGC B cells (Fig3H-J), with similar or higher enrichment compared to LZ GC B cells which are canonically the site of CSR. One notable example was *APEX1*/APE1 which is required for CSR and was expressed mostly by non-GC B cells, with the exception of a small number of GC B cells (*22*) (FigS12B). Although expression levels of *AICDA* were low (Fig2H), our analysis of the preGC and LZ GC transition revealed other genes mechanistically linked with CSR (*23-28*), including those capable of binding switch region sequences within the IgH locus, interacting with CSR machinery or regulating *AICDA*/AICDA stability (Fig3K). Other genes of interest included *miR155HG* and transcription factors *BATF* and *IRF4*, known B cell-intrinsic regulators of GC formation in mice (*29-31*) and the poorly understood transcription factor BHLHE40 that is capable of binding to the major regulatory regions at the IgH locus (FigS13). Although we observed very few switched preGC B cells (Fig2G,3L) and did not observe any preGC B cells concurrently expressing multiple isotypes with identical VDJ sequences based on analysis of high-confidence scVDJ sequences, we found that expression of IgH germline transcripts (GLTs), which is essential for CSR to take place at the IgH locus, peaks in preGC cells (Fig3M). We could also detect some coding *IGHG* scRNA-seq expression in unswitched preGC B cells (FigS12C), although this appeared to be at too low a level in single cells to enable reconstruction of high-quality scVDJ contigs. Intriguingly, these observations in human tonsil are consistent with GLT transcription peaking and CSR occurring prior to GC formation in mouse models (*22*). This suggests the possibility that the preGC B cell state may be primed to undergo CSR before formation of the GC response, although our observations do not preclude CSR occurring within the GC as well.

Antibody-based selection of germinal centre B cells at single-cell resolution

CSR has the potential to influence the antibody-based selection of B cells within GC reactions as a consequence of differential signalling through the membrane-bound immunoglobulin BCR, but studying these dynamics is challenging *in vivo*. We therefore used our paired single-cell transcriptomes and VDJ antibody datasets to dissect the gene expression linked with antibody-based selection during the GC response.

Although GC B cells cycle between physically distinct light and dark zones we found that many GC B cells exist in a continuum between these two states (Fig4A-B), similar to previous studies (*16, 32*), except for FCRL3^{high} and prePB cells that are transcriptionally distinct (Fig2F, 4C). These sub-populations of GC B cells also exhibit different class switching patterns, with prePB cells more frequently expressing class-switched isotypes and FCRL3^{high} GC cells retaining expression of IgM/IgD (Fig2G, 4D, S8). Intriguingly, expanded B cell clones containing FCRL3^{high} GC cells were almost exclusively IgM^{+} (Fig4E, FigS9), suggesting that antibody class may be linked with a specific gene expression program and cell fate, or vice versa.

We next leveraged our paired single-cell VDJ and transcriptomic datasets to stratify all non-cycling GC B cells (and excluding prePB and FCRL3^{high} populations) based on their antibody SHM frequencies as a proxy for affinity (Fig4F). GC B cells with high SHM were significantly enriched with gene signatures of high affinity antigen-binding B cells (*11*), higher expression of the B cell maturation marker CD27 and larger clone sizes (Fig4F), reflecting increased expansion and maturation based on BCR affinity. We detected very few B cells expressing non-functional IgH sequences (Fig4F) and apoptotic cells failed to generate sufficiently high quality transcriptomic data or were removed due to elevated mitochondrial content. GC B cells with high or low SHM had many differences in gene expression (Fig4G, TableS5), consistent with high and low affinity binding events differentially regulating GC B cells (*11*). However, these SHM gene signatures were very similar to gene signatures for switched or unswitched GC B cells (Fig4G, TableS6), as predicted given lower average SHM frequencies of IgM^{+} GC B cells (Fig1D).

To overcome the confounding influence of maturation state on examining isotype-specific gene expression, we compared GC B cells with matched SHM levels expressing different antibody classes. This revealed that switched GC B cells were enriched for genes required for cell survival, BCR signalling, antigen presentation, immune responses and metabolism, as well as other important signalling pathways (Fig4H-I, TableS7). Switched GC B cells also had evidence of increased signalling in response to T cell-derived cytokines such as IL4, TGFβ, IFNγ and CD40LG, or signalling through different toll-like receptors (TLR) (Fig4H), consistent with them receiving more T cell help.

Several genes involved with GC confinement or B cell niche homing were up-regulated in IgG^+ and IgA^+ GC B cells, such as genes required for CXCL12-mediated migration to GCs (*LCP1, MYO1E*) (*33, 34*) and GC confinement receptor *P2RY8* (*35*), indicating that switched B cells may be more likely to remain in the GC for longer. This could explain their higher rates of SHM. Most gene expression differences were comparable between IgG^+ and IgA^+ B cells (Fig4I, S14) and we identified few significant or meaningful differences for subclass-specific B cells (FigS14, TableS8). One interesting exception was the specific expression of *CLEC16A* in IgA^+ GC B cells given that this gene is associated with a selective IgA immunodeficiency (*36*). Finally, we observed many differences in the expression of transcription factors between class-specific GC B cells (Fig4I). IgG^+ B cells expressed higher levels of *BCL6*, *XBP1* and *ID3* that may increase their ability to reside in the GC or differentiate to plasma cells, while IgM^+ cells had higher levels of *BACH2* that represses plasma cell differentiation (*11, 37-39*). We also found differential expression of other factors (*LMO2*, *TOX*, *BCL11A, CUX1)*, raising the question of their role in the unique transcriptional wiring of switched and unswitched B cells within the GC.

Our single-cell resolution of the GC response allowed us to uncouple antibody affinity and class and dissect differential contributions of these two critical arms of the B cell repertoire in shaping B cell fate and function in the GC. These analyses suggest varying abilities of switched and unswitched B cells to survive and reside in the GC and establish that in addition to BCR affinity, antibody-based selection in the GC can be shaped by whether a B cell has undergone CSR.

Diverse MBC states in human tonsils

Maturation state and antibody class were linked to specific gene expression by GC B cells, potentially through membrane-specific isoforms of immunoglobulin as part of the BCR. In contrast to plasmablasts, MBCs retain BCR expression of which the isotype may influence the phenotypic properties of different MBC subsets (*40*), the full diversity of which remains unknown. We therefore sought to better define the heterogeneity within the MBC pool in human tonsils and determine whether antibody class expression by MBCs might be linked with different functional abilities or subsets.

A significant proportion of MBCs in paediatric human tonsils are unswitched (Fig1B, 2G). To examine potential differential gene expression across class-switched MBC subsets we generated paired single-cell transcriptomics and VDJ repertoires for all (n=2) or IgM-depleted (n=4) MBCs (Table S1). Dataset integration and quality control provided 21,595 high-quality MBC single-cell transcriptomes that we annotated with 11 clusters reflecting different MBC subsets and states (Fig5A-C), all of which lacked marker gene expression for naïve or GC B cells (FigS15A). In addition to tissue-resident FCRL4^{+} MBCs previously identified (Fig2) (*41*), we annotated two rare *CR2*/CD21^{low} MBC subsets that appeared to resemble atypical MBC populations described in varied tissues and disease states (*42-44*) (Fig5A-C, S15B). The CD21^{low} 2 cluster shared several features with atypical FCRL5^{high} MBCs (*44*), such as high expression of inhibitory receptors (*CD22*) and genes involved in antigen presentation (*CIITA*) (Fig5C). Otherwise, the majority of MBC diversity within human tonsils reflected differences in cellular state or activity rather than distinct cell types, such as heat shock protein (HSP)-related and IFN-related gene activity (Fig5A-C, S15C). We also identified multiple activation states (Activated1/2), including MBCs with similar gene expression to naïve preGC cells and an enrichment for CSR genes (Fig5C, S15D-E), that we label as preGC MBCs. We also identified an FCRL3^{high} MBC state similar to the FCRL3^{high} GC population (Fig5C, S15D) and similar MBC populations could also be detected in lymph nodes and spleen, albeit at varying frequencies (FigS15F). This suggests that these annotated MBC populations may be widely shared functional states spanning multiple B cell fates and organ systems.

We next examined whether class-switched and unswitched MBCs exhibit different gene expression networks that might reflect unique functional abilities (*45*). We found little evidence that antibody class contributed towards the likelihood of an MBC to exist in a given state, with the exception of FCRL3^{high} MBCs which, similar to FCRL3^{high} GC cells, were predominantly IgM^{+} by both scVDJ and flow cytometry analysis (Fig5A-B, S8, S16A). Intriguingly, an FCRL3^{high} MBC gene signature was enriched in IgM^{+} cells across all MBC clusters (FigS16B), further supporting a close relationship between IgM expression and this cell state.

Transcriptional wiring of unswitched MBCs underlies predisposition to form secondary GCs

Comparison of switched and unswitched MBCs with equivalent SHM levels, as MBCs expressing switched isotypes tend to have higher SHM frequencies (FigS16C), confirmed widespread gene expression differences that were independent of MBC subset or state (Fig5D-F). This indicates a specific transcriptional wiring of MBCs dependent upon their antibody class. Unswitched MBCs were broadly enriched for genes involved in cytokine-mediated signalling, activation, antigen presentation and migration (Fig5E-F), which may reflect the increased capacity of IgM^{+} MBCs to re-initiate GC reactions as part of a recall memory response (*45-47*), an essential component of long-term B cell-mediated immunity. Counterintuitively, IgM^{+} MBCs were also enriched for anti-proliferative and apoptotic gene pathways (Fig5E), as well as many genes proposed to regulate or inhibit B cell activation, such as *FCRLA*, *FCRL2*, *FCRL3*, *CBLB*, *CD72* and *SIGLEC10* (*48-52*). This suggests a fine regulatory balance at the transcriptional level may control the activation threshold of unswitched MBCs.

Unswitched MBCs expressed higher levels of *POU2F2* (OCT2) and *FOXP1* than class-switched MBCs (Fig5F), which coordinate the capacity of B cells to respond normally to antigen receptor signals and directly repress key regulators of plasma cell differentiation respectively (*53, 54*). This is consistent with switched IgG^{+} MBCs being more likely to differentiate into plasma cells, while unswitched IgM^{+} MBCs are more likely to re-enter or form secondary GC responses to gain higher affinity (*45-47*). Indeed, unswitched MBCs in the preGC cluster were significantly enriched for gene signatures linked with GC entry and CSR compared to switched MBCs in the same cluster (Fig5G-H). This supports a model

where unswitched MBCs are more likely to form secondary GC responses than switched MBCs, and that these unswitched MBCs are primed to undergo class switching during this process.

Discussion

High affinity antibodies are generated in SLOs, but the dynamic nature of this response has presented significant challenges to understanding human B cell-mediated immunity *in vivo*. By combining bulk antibody repertoire analysis with single-cell transcriptomics we have generated a detailed resource of human B cell maturation and the GC reaction. This enabled us to explore diverse B cell states, reconstruct B cell activation dynamics, dissect the relationship between antibody class and functional capability of GC and MBC subsets, and reveal an unappreciated heterogeneity in tonsillar MBC population. These integrated analyses highlight how the outcome of CSR can influence B cell fate and function, and provide a detailed framework through which to view B cell-mediated immune responses.

The humoral immune response exhibits compartmentalisation that divides the gut-associated lymphoid tissue from blood-rich tissues such as spleen, bone marrow and lung (*55*). Tonsils are positioned in the upper airway and are highly exposed to antigen relative to encapsulated lymph nodes. Memory responses generated in tonsillar GCs seed the airway mucosa (*56*). Tonsils therefore form an important component of both systemic and mucosal immunity. Apart from known tonsillar-specific B cell populations such as tissue-resident FCRL4^{+} MBC and IgD^{+} plasma cells (Fig2, S4), we found that B cell populations in the human tonsil are transcriptionally highly similar to those in lymph nodes, although we cannot exclude the possibility that functional differences or unique populations may exist between tissues. Nonetheless, tonsils provide a useful and largely representative model for studying the dynamic GC responses of SLOs.

Although first histologically observed over a century ago, many questions remain about how B cells enter, experience and exit the GC reaction (*1, 2, 7*). Understanding the early events that facilitate GC entry by human B cells could provide new targets for adjuvants during vaccination or other immunotherapies. We mapped gene expression dynamics of the early stages of B cell activation that

correspond to antigen-dependent signalling through the BCR and the subsequent transition to a transcriptionally-distinct preGC state, the latter of which is presumably under the regulation of cognate antigen-specific T helper cells. Recent spatial mapping of this specific preGC B cell state in the human lymph node has predicted that these cells are distinct from existing GCs (*15*), suggesting that they are likely forming new GC structures. This preGC state has many features suggesting it is primed to undergo class switching (Fig3), although we were not able to directly detect switch events themselves possibly due to the rarity of this occurrence. Alternatively, CSR may be linked with a rapid change from preGC to the LZ GC transcriptional state that is very dynamic and difficult to observe. Nonetheless, this B cell state appears to agree with observations in mouse models that B cells can undergo class switching before formation of GCs (*20-22*). If class switching is indeed capable of occurring before GC formation, this could then shape antibody-based selection dynamics in the GC (Fig4).

Within the GC, B cell survival and selection is dependent on antigen binding to the BCR and its downstream signalling pathways (*6*). By using single-cell transcriptomics paired with BCR sequence analysis we uncouple antibody class, SHM frequency and B cell phenotype. We show that switched GC B cells, in contrast to IgM^+ cells, have gene expression patterns consistent with increased BCR signalling and a greater capacity to remain within the GC and acquire T cell help to undergo additional SHM and increase their affinity. If CSR does indeed occur prior to GC entry, as we and others suggest (*20-22*), these data support that the ability of a B cell to acquire high affinity is dictated by the outcome of a specific CSR “checkpoint” at the preGC stage. This would explain our observation that isotype-matched MBCs have comparable SHM frequencies to plasmablasts, in contrast to the prevailing paradigm that higher affinity GC B cells preferentially differentiate towards the plasmablast fate whereas lower affinity clones seed the memory compartment (*3, 11, 57*). While we were not able to directly quantify antigen-specificity or antibody affinity given the highly polyclonal nature of tonsillar B cells, a recent study in mice found comparable antibody affinities between antigen-specific MBC and plasmablasts following influenza infection (*58*). The differences in affinity between these populations may instead be explained by the likelihood of whether they retained IgM expression prior to entering affinity maturation in the GC. This is also relevant during the secondary activation of MBCs, as IgM^+

MBCs appear more primed for CSR and GC entry, which may be important to provide a higher affinity secondary response. Finally, our discovery of a specific unswitched gene expression signature across different B cell states, exemplified by the FCRL3^{high} B cell state, raises interesting questions about how this unique transcriptional wiring is regulated and how it may contribute to their function. It will also be of interest to determine the relevance of the distinct lack of class switching across FCRL3^{high} containing clones, for example whether they arise as a consequence of prolonged antigen exposure and/or in response to particular antigens.

Although the direct mechanisms shaping the class-specific gene expression differences remain to be elucidated, variations in the immunoglobulin tail tyrosine (ITT), linker flexibility or glycosylation sites between IgM and other antibody isotypes may all contribute to differential BCR signalling (*8-10*). The enhanced BCR signalling capacity conferred by the ITT has been linked to the propensity of IgG+ memory B cells to differentiate into plasmablasts (*46*). Indeed, memory B cells were more enriched for IgA (which lacks an ITT) than PBs, although those IgA MBC clones were less clonally expanded than comparable PB clones. Although our study lacks spatial information about these different B cell populations, it is interesting to speculate that these repertoire-based differences between MBC and PBs might be linked to different spatial locations within tonsillar tissue that could be under greater or less influence of mucosal cytokines that drive IgA class switching. A limitation of our study, like others in human subjects, was that we could not compare multiple tissues in the same individual. We observed differences in isotype frequencies and SHM in tonsil compared to mesenteric lymph nodes which may relate to how different tissue environments shape the local antibody repertoire (e.g. pathogen exposure and cytokine production), but could also represent enrichment of antigen-experienced cells due to the relatively advanced age of lymph node patient donors. Of note, despite differences in the relative frequencies and clonal diversities of IgA and IgG in MBC versus plasmablasts in tonsillar repertoires, we did not identify many differences in gene expression between IgG$^+$ and IgA$^+$ B cells, or between subclass-specific B cells, which may reflect the need to increase the power of future studies to identify potentially subtle gene expression differences between the less abundant isotypes. While human tonsils are a useful tissue with which to examine active GC responses, it will be of interest to compare and

contrast the class-specific gene expression differences we observe with those that might exist in other tissues in both health and disease to investigate how the local cellular and cytokine environment under the influence of different patterns of microbial exposure can influence antibody-based selection in different contexts or diseases states.

Finally, our profiling of human tonsillar MBCs revealed diverse states reflecting different activation, signalling and functional potential. The upstream regulators of these different transcriptional states remains unclear from our analyses, but these unique gene expression patterns are likely to reflect the combined influence of extrinsic microenvironmental cues and cell intrinsic identity. We found that switched MBCs expressed genes consistent with being primed to undergo plasma cell differentiation whereas the more abundant unswitched MBC exhibited gene expression making them more likely to re-enter GC reactions. Given an emerging appreciation for heterogeneity within both human and mouse MBC populations (*14*) and building on evidence for functionally distinct MBC populations enriched for different FCRL expression patterns (*41, 44*), some of which we observe here, our single-cell characterisation of different MBC populations and activation states provides a valuable dataset to interrogate the potential relevance for such diverse populations in mediating humoral immunity and immune-related disorders.

Materials and Methods

Study Design

This study aimed to characterise antibody repertoires of human tonsillar B cell states with bulk antibody repertoire sequencing of sorted B cell populations combined with an unbiased characterisation of the gene expression and repertoires using single-cell transcriptomics and VDJ sequencing. The study used tonsil samples from paediatric patients (>3yo) who were undergoing routine tonsillectomy, numbers of samples per experiment are reported in figure legends. Bulk and single-cell antibody repertoires were analysed together to examine subset-specific features of antibody repertoires and increase the power to identify clonally-related B cells in the single-cell VDJ-seq assay. Immune cell populations were identified by the unbiased clustering of scRNA-seq datasets and annotation with known and novel gene

445 expression markers and antibody repertoire features for B cell subsets. Novel populations were
446 examined further by flow cytometry. Pseudotemporal ordering was performed to examine the
447 relationship between different B cell maturation stages. To test the hypothesis that antibody class is
448 linked with B cell fate and function, we compared single-cell transcriptomes of antibody class-specific
449 B cells with matched somatic hypermutation frequencies (as a proxy for affinity) and performed
450 differential gene expression and pathway enrichment analyses.
451
452 **Human ethics and tissue collection**
453 Paediatric tonsillectomy patients at Royal London Hospital were consented with approval from North
454 West/Greater Manchester East Research Ethics Committee (17/NW/0664). Tonsillar tissue was
455 homogenised with the gentleMACS™ Dissociator and mononuclear lymphocytes purified using Ficoll-
456 Paque™ gradients. Cells were processed immediately for single-cell library preparation (after ensuring
457 high cell viability by TrypanBlue staining) or cryopreserved in FCS with 10% DMSO at -70°C.
458
459 **Sorting B cell subsets**
460 Cells were stained with Zombie NIR™ Fixable Viability Kit (BioLegend) to label dead cells, blocked with
461 human FcR Blocking Reagent (Miltenyi Biotec) and stained with CD19-APC, CD38-PE-Cy7, CD27-
462 PacificBlue™, IgD-PerCP-Cy5.5, and IgM-FITC (TableS10). For bulk repertoires, two replicates of
463 250,000 live B cells were sorted using a BD FACSAria™ IIIu before RNA extraction: total ($CD19^+$), naïve
464 ($CD19^+IgD^+CD38^-$), germinal centre ($CD19^+IgD^-CD38^+$), memory ($CD19^+IgD^-CD38^-$), and plasmablasts
465 ($CD19^+IgD^-CD38^{++}$). For single-cell RNA-seq of sorted MBCs, live $CD19^+IgD^-CD38^-$ (n=2) or live
466 $CD19^+IgD^-CD38^-IgM^-$ (n=4) cells were sorted.
467
468 **Bulk BCR repertoire sequencing**
469 RNA was isolated using the RNAqueous™-Micro Total RNA Isolation Kit (ThermoScientific)
470 supplemented with β-mercaptoethanol. Immunoglobulin heavy (IgH) chain sequences were amplified
471 as previously described (*5*) (Supplementary Methods). Briefly, 50-100 ng RNA was annealed to pooled
472 IgH constant region primers containing unique molecular identifiers (UMIs) of 10 or 12 nucleotides at

72°C for 5 min before incubation on ice for 2 min. First-strand cDNA synthesis was performed using SuperScript IV reverse transcriptase (ThermoFisher Scientific) before second-strand cDNA synthesis with Phusion® High-Fidelity DNA Polymerase (NEB) and UMI-containing variable region primers. Double-stranded cDNA was purified using Ampure XP beads (BeckmanCoulter) before library amplification with NEBNext UltraII Q5 Master Mix (NEB). Libraries were quantified by Qubit™ dsDNA HS Assay Kit prior to Illumina MiSeq sequencing with paired-end 301bp reads.

Single-cell library preparation, sequencing and alignment

Tonsillar immune cells (*n*=7) or FACS-sorted MBCs (*n*=6) single-cell libraries were generated with the 10x Genomics Chromium Single Cell 3' (v2; n=1) or 5' and V(D)J (v1; n = 6) assays (TableS1) prior to sequencing on the Illumina NextSeq500 with 26/8/134 bp (scRNA) or 155/8/155 bp (scVDJ) read configurations. scRNA-seq basecall files were processed with CellRanger (v3.0.0) to provide FASTQ files for mapping to GRCh38 (release 92) to produce gene-by-cell expression matrices. scVDJ datasets were processed with CellRanger vdj using the refdata-cellranger-vdj-GRCh38-alts-ensembl-2.0.0 reference. Incomplete or low quality IgH contigs, or those with insufficient coverage of constant regions to ensure accurate isotype assignment between closely related subclasses, were filtered from subsequent analyses.

Integrated repertoire analysis

Bulk VDJ sequencing reads were processed with pRESTO (v0.5.10) (*59*) and combined with scVDJ-derived IgH sequences (Supplementary Methods). Briefly, reads with mean Phred scores >25 underwent UMI alignment using MUSCLE (v3.8.31) (*60*) and UMIs with >3 unique reads were used to assemble consensus VDJ sequences. Duplicate sequences within each biological sample were collapsed before integrating with filtered scVDJ contigs from cellranger. All IgH sequences were annotated with AssignGenes.py (ChangeO v0.4.5) (*61*) and IgBLAST (v1.12.0) (*62*) before correction of ambiguous V gene assignments using TIgGER (v0.3.1) (*63*). Clonally-related sequences were identified using DefineClones.py with nearest neighbour distance threshold determined by distToNearest (Shazam v0.1.11) (*61*). Germline sequences were inferred using CreateGermlines.py

and SHM frequencies calculated with observedMutations. SHM frequencies >0.02 were annotated as "High", 0-0.02 as "Low" and 0 as "None". For bulk BCR repertoire analysis, scVDJ sequences were excluded, providing ~1.5 million high-confidence unique IgH sequences. Lineage trees of expanded clones were constructed via maximum parsimony using buildPhylipLineage (Alakazam v0.2.11) (*61*). To quantify class switch hierarchies, the observed frequency of direct edges between unique sequences of different isotypes were counted and normalised to the frequency of such transitions expected by chance (calculated by iterative (k=100) sampling of isotype frequencies). scVDJ sequences were integrated with scRNA-seq datasets as described previously (*64*). Diversity analyses were performed with Alakazam. Mean clonal diversity scores, SHM and isotype frequencies for each donor were compared using Student's t test.

Single-cell RNA-seq processing and analysis

Cellranger expression matrices were used to quantify mitochondrial percentages and generate summed Ig and TCR VDJ counts before these genes were removed before processing with Seurat (v3.0.3) (*65*) with log transformation, normalisation, cell cycle prediction and variable gene identification (Supplementary Methods). A preliminary integration of unsorted immune samples or sorted MBC samples was performed using FindIntegrationAnchors and IntegrateData (3000 genes) regressing out cell cycle scores and mitochondrial gene expression, principle component analysis (PCA) and preliminary clustering. One preliminary cluster enriched with high frequency of predicted doublets from DoubletFinder (v2.0.1) (*66*) was removed. Following this quality control, all samples were integrated using the unsorted immune samples as reference with 4000 highly variable genes before regressing cell cycle and mitochondrial gene expression, PCA and identifying broad cell type lineages (B cell, T cell and non-lymphoid cells) which were then reclustered separately. Clusters were manually annotated using differential gene markers from FindAllMarkers and scVDJ antibody features where relevant. Uniform Manifold Approximation and Projection (UMAP) was used for visualisation.

Integration with lymph node and spleen datasets

Cellranger gene expression matrices from human lymph nodes (*n*=7) and spleen (*n*=1) (*64, 67*) were processed as above, before integration with the unsorted tonsillar immune cell gene expression objects using FindIntegrationAnchors and IntegrateData (Seurat) with the unsorted tonsillar immune samples as reference with 4000 highly variable genes before regressing cell cycle and mitochondrial gene expression, PCA and unbiased clustering. Clusters were annotated based on a consensus of previous cell type annotations and confirmed by differential gene expression analysis. Processed scVDJ metadata from mesenteric lymph nodes (*n*=2) were obtained from (*64*). Lymph node and spleen memory B cells (B_MBC and B_MBC_FCRL4) were annotated with high resolution tonsillar memory B cell subsets using FindTransferAnchors and TransferData with 3000 variable features (Seurat).

Gene expression and signature enrichment analysis

Differential gene expression for antibody class-specific or somatic hypermutation frequency for GC B cells or class-switched MBCs was performed using FindAllMarkers with Benjamini-Hochberg false discovery rate (FDR) correction. Genes were deemed significantly different if FDR <0.05, average log fold change >0.1 and the gene was detected in >20% of cells in that group. Ingenuity Pathway Analysis (Qiagen) was performed using avg_logFC values of all genes significantly enriched in at least one class. Gene ontology enrichment analyses were otherwise performed with Metascape (*68*). Single-cell gene signature scoring was performed with AUCell (v1.5.5) (*69*), including a manually curated shortlist of CSR-related genes (*4*), high and low affinity GC B cells (GSE73729) (*11*) and anti-IgM stimulation of B cells (GSE41176) (*19*). Affinity signatures were derived by quantifying RNA-seq transcript counts using Salmon (v1.0.0) (*70*), collapsing protein-coding transcripts with tximport (v1.10.1) (*71*), identifying significant gene expression differences using DESeq2 (v1.22.2) (*72*) with a threshold of fold change > 1.5 and FDR < 0.05 and converting mouse gene IDs to human with bioMart (v2.38.0) (*73*). Genes up-regulated following α-IgM treatment were identified from microarray data with Geo2R (*74*). preGC and FCRL3^{high} signatures used the top 50 most significantly enriched genes per cluster. Unless indicated otherwise, Wilcoxon Ranked Signed Sum test was used to test for significant differences.

Cell-cell communication using CellPhoneDB

To examine the expression of ligand-receptor pairs between different scRNA-seq clusters raw count matrices were analysed with "statistical_analysis" option of CellPhoneDB (v2.0.6) (*75*). The number of unique significant ligand-receptor co-expression pairs between each cell type were determined and interactions of interest visualised by the means of average expression of gene 1 in cell type 1 and gene 2 in cell type 2 indicated by colour and *p* values indicated by circle size.

RNA velocity and pseudotemporal ordering

Spliced and unspliced transcripts were quantified for tonsil immune samples with velocyto (v0.17.10) (*17*) and integrated with raw counts of annotated B cell scRNA-seq in Scanpy (v1.4) (*76*) using scVelo (v0.1.23) (*18*). Velocyto-derived counts were processed, filtered and normalised prior to velocity estimation using a dynamical model with scVelo. Velocities were projected and visualised onto UMAP embedding. Velocity-based pseudotime reconstruction was performed using latent time recovery of single-cell velocities with tl.recover_latent_time and tl.velocity_pseudotime of the Naïve, Activated, preGC and LZ GC B cell clusters. Dynamic gene expression changes were examined using tl.rank_velocity_genes for velocity-based sub-clustering (res=1) and top 200 genes per sub-cluster (ribosomal genes removed) were collapsed to unique genes for heatmap visualisation with smoothed scores. Gene expression or signature scores were quantified across pseudotime as smoothed normalised counts with geom_smooth() ±95% CI. For pseudotemporal analysis of the continuum of GC B cell states (Fig4), all GC B cell clusters were analysed with diffusion-based pseudotime (tl.dpt, Scanpy) independent of RNA velocity using default settings. Partition-based graph abstraction (PAGA) analysis was performed with Scanpy using default settings (threshold = 0.05) with all clusters except for MBCs due to their close transcriptional similarity with naïve B cells.

Quantitation of IgH germline transcripts

All reads mapped to the IgH locus (chr14:105540180-105879151) were quantified with dropEst (v0.8.6) (77) against a custom GTF file containing I promoter germline transcript coordinates. Counts were log10-normalised and scaled in Seurat before class-specific counts were summed.

Flow cytometry identification of preGC and FCRL3^{high} B cells

Cryopreserved cells were thawed, washed and blocked with human FcR Blocking Reagent (Miltenyi Biotec) then stained with a panel of fluorophore conjugated antibodies (Table S10) and DAPI (Sigma) to discriminate live and dead cells. Samples were run on a Cytek Aurora spectral flow cytometer using SpectroFlo (Cytek) and unmixed before analysis. Data were analysed using FlowJo v10 (Treestar) and gates were set using fluorescence minus one (FMO) controls.

Immunohistochemistry

Formalin-fixed paraffin-embedded tonsil samples were deparaffinized in xylene and rehydrated through a series of ethanols to water. Endogenous peroxidase was blocked with 3% hydrogen peroxide before heat-mediated antigen retrieval with a citrate-based unmasking buffer (Vector Labs) at 120°C. 3 µm sections were incubated for 40 min at RT with anti-APE1 (1:1000; HPA002564; Sigma) or anti-FCRL3 (1:100; HPA048022; Sigma) before using the Super-sensitive–Polymer HRP system (Biogenex) with purple chromogen VIP (Vector Labs) and hematoxylin as a nuclear counterstain. Slides were scanned (Pannoramic250 Flash) before soaking in xylene to de-coverslip before rehydration through ethanol to water. De-staining and stripping of primary antibodies and heat-labile chromogen was achieved by a subsequent round of heat-mediated antigen retrieval. Anti-CD20 (1:500; M0755; Dako) was incubated for 40 min at RT, followed by detection, visualization and scanning as before. Negative controls were performed by treating sequential sections as above but without anti-CD20 staining to confirm complete stripping. Images were prepared using CaseViewer (3DHistTech).

Acknowledgements

We thank the QMUL Genome Centre for sequencing support, Marilena Crescente for help with pathway analyses, Mirjana Efremova for help with CellPhoneDB, and all members of the James and Teichmann labs for their support. Finally, we thank Neil McCarthy, Kylie James, Kerstin Meyer, Lou Herman and Jo Spencer for reviewing the manuscript.

Funding

Supported by funding from the Wellcome Trust to L.K.J. (208961/Z/17/Z), S.A.T. (206194), J.C.R. (110020/Z/15/Z) and H.W.K. (213555/Z/18/Z).

Author contributions

H.W.K. designed and performed experiments, analysed data and wrote the manuscript. N.O. performed tonsillectomy tissue collection. J.C.R. and A.J.C. designed and performed immunohistochemistry experiments. G.W. performed FACS sorting. S.A.T. supervised analysis. L.K.J. designed and supervised experiments and analysis and wrote the manuscript.

Competing interests

In the past three years, S.A.T has worked as a consultant for Genentech, Biogen and Roche, and is a remunerated member of the Foresite Labs Scientific Advisory Board.

Data and materials availability

Raw and processed data for this study are available at ArrayExpress (accession numbers: E-MTAB-8999, E-MTAB-9003 and E-MTAB-9005) and https://www.tonsilimmune.org/. All other data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials.

Supplementary Material

Supplementary Methods

Figure S1. Comparison of FACS-based and scRNA-seq analysis of major B cell subsets.

Figure S2. Clonal overlap of expanded B cell clones between GC, MBC and PB subsets.

Figure S3. Analysis of T cell and non-lymphoid cell populations in human tonsils by scRNA-seq.

Figure S4. Integration of tonsillar scRNA-seq datasets with lymph node and spleen scRNA-seq.

Figure S5. Marker gene analysis of human B cell subsets from tonsils by single-cell RNA-seq.

Figure S6. UMAP visualisation of B cell scVDJ antibody isotype and SHM frequency.

638 Figure S7. Flow cytometry analysis of CD23, CD58 and CD108 expression in naïve and GC B cells.

639 Figure S8. Flow cytometry and immunohistochemistry analysis of FCRL3^{+} B cells.

640 Figure S9. Example lineage trees of expanded FCRL3^{high} GC B cell clones.

641 Figure S10. Sub-clustering of naïve, memory and activated B cell scRNA-seq clusters.

642 Figure S11. PAGA-based trajectory analysis of B cell scRNA-seq clusters.

643 Figure S12. Dynamic gene expression during B cell activation and GC entry/formation.

644 Figure S13. Transcription factor binding at the immunoglobulin heavy chain locus.

645 Figure S14. Class- and subclass-specific gene expression analyses of high SHM GC B cells.

646 Figure S15. Characterisation of MBC states identified by scRNA-seq.

647 Figure S16. Antibody and gene expression features of switched and unswitched MBCs.

648 Table S1. Sample metadata for scRNA-seq libraries and frequencies of immune cell clusters in scRNA-
649 seq datasets.

650 Table S2. Differential scRNA-seq gene expression tables for human tonsillar T cell clusters.

651 Table S3. Differential scRNA-seq gene expression tables for human tonsillar non-lymphoid cell clusters.

652 Table S4. Differential scRNA-seq gene expression tables for human tonsillar B cell clusters.

653 Table S5. Differential scRNA-seq gene expression tables for non-cycling germinal centre B cells based
654 on somatic hypermutation frequency levels.

655 Table S6. Differential scRNA-seq gene expression tables for non-cycling germinal centre B cells based
656 on scVDJ antibody class expression (switched vs unswitched).

657 Table S7. Differential scRNA-seq gene expression tables for non-cycling, high somatic hypermutation
658 GC B cells based on scVDJ antibody class.

659 Table S8. Differential scRNA-seq gene expression tables for non-cycling, high somatic hypermutation
660 GC B cells based on scVDJ antibody subclass.

661 Table S9. Differential scRNA-seq gene expression tables for human tonsillar memory B cell clusters.

662 Table S10. Antibody panels used for cell sorting and flow cytometry analysis.

663 Table S11. Raw data file.

References

1. J. G. Cyster, C. D. C. Allen, B Cell Responses: Cell Interaction Dynamics and Decisions. *Cell* **177**, 524-540 (2019).
2. L. Mesin, J. Ersching, G. D. Victora, Germinal Center B Cell Dynamics. *Immunity* **45**, 471-482 (2016).
3. D. Suan, C. Sundling, R. Brink, Plasma cell and memory B cell differentiation from the germinal center. *Curr Opin Immunol* **45**, 97-102 (2017).
4. J. Stavnezer, C. E. Schrader, IgH Chain Class Switch Recombination: Mechanism and Regulation. *J Immunol* **193**, 5370-5378 (2014).
5. F. Horns, C. Vollmers, D. Croote, S. F. Mackey, G. E. Swan, C. L. Dekker, M. M. Davis, S. R. Quake, Lineage tracing of human B cells reveals the in vivo landscape of human antibody class switching. *eLife* **5**, e16578 (2016).
6. K. Kwak, M. Akkaya, S. K. Pierce, B cell signaling in context. *Nat Immunol* **20**, 963-969 (2019).
7. M. J. Shlomchik, W. Luo, F. Weisel, Linking signaling and selection in the germinal center. *Immunol Rev* **288**, 49-63 (2019).
8. S. W. Martin, C. C. Goodnow, Burst-enhancing role of the IgG membrane tail as a molecular determinant of memory. *Nat Immunol* **3**, 182-188 (2002).
9. N. Engels, L. M. Konig, W. Schulze, D. Radtke, K. Vanshylla, J. Lutz, T. H. Winkler, L. Nitschke, J. Wienands, The immunoglobulin tail tyrosine motif upgrades memory-type BCRs by incorporating a Grb2-Btk signalling module. *Nat Commun* **5**, 5456 (2014).
10. Y. Xu, L. Xu, M. Zhao, C. Xu, Y. Fan, S. K. Pierce, W. Liu, No receptor stands alone: IgG B-cell receptor intrinsic and extrinsic mechanisms contribute to antibody memory. *Cell Research* **24**, 651-664 (2014).
11. R. Shinnakasu, T. Inoue, K. Kometani, S. Moriyama, Y. Adachi, M. Nakayama, Y. Takahashi, H. Fukuyama, T. Okada, T. Kurosaki, Regulated selection of germinal-center cells into the memory B cell compartment. *Nat Immunol* **17**, 861-869 (2016).
12. D. Suan, N. J. Kräutler, J. L. V. Maag, D. Butt, K. Bourne, J. R. Hermes, D. T. Avery, C. Young, A. Statham, M. Elliott, M. E. Dinger, A. Basten, S. G. Tangye, R. Brink, CCR6 Defines Memory B Cell Precursors in Mouse and Human Germinal Centers, Revealing Light-Zone Location and Predominant Low Antigen Affinity. *Immunity* **47**, 1142-1153.e1144 (2017).
13. K. R. Duffy, C. J. Wellard, J. F. Markham, J. H. Zhou, R. Holmberg, E. D. Hawkins, J. Hasbold, M. R. Dowling, P. D. Hodgkin, Activation-induced B cell fates are selected by intracellular stochastic competition. *Science* **335**, 338-341 (2012).
14. K. L. Good-Jacobson, M. J. Shlomchik, Plasticity and Heterogeneity in the Generation of Memory B Cells and Long-Lived Plasma Cells: The Influence of Germinal Center Interactions and Dynamics. *J Immunol* **185**, 3117-3125 (2010).
15. V. Kleshchevnikov, A. Shmatko, E. Dann, A. Aivazidis, H. W. King, T. Li, A. Lomakin, V. Kedlian, M. S. Jain, J. S. Park, L. Ramona, E. Tuck, A. Arutyunyan, R. Vento-Tormo, M. Gerstung, L. James, O. Stegle, O. A. Bayraktar, Comprehensive mapping of tissue cell architecture via integrated single cell and spatial transcriptomics. *bioRxiv*, 2020.2011.2015.378125 (2020).
16. A. B. Holmes, C. Corinaldesi, Q. Shen, R. Kumar, N. Compagno, Z. Wang, M. Nitzan, E. Grunstein, L. Pasqualucci, R. Dalla-Favera, K. Basso, Single-cell analysis of germinal-center B cells informs on lymphoma cell of origin and outcome. *J Exp Med* **217**, (2020).
17. G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, P. V. Kharchenko, RNA velocity of single cells. *Nature* **560**, 494-498 (2018).
18. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol*, 10.1038/s41587-41020-40591-41583 (2020).
19. H. Shinohara, M. Behar, K. Inoue, M. Hiroshima, T. Yasuda, T. Nagashima, S. Kimura, H. Sanjo, S. Maeda, N. Yumoto, S. Ki, S. Akira, Y. Sako, A. Hoffmann, T. Kurosaki, M. Okada-Hatakeyama, Positive Feedback Within a Kinase Signaling Complex Functions as a Switch Mechanism for NF-κB Activation. *Science* **344**, 760-764 (2014).
20. K. M. Toellner, A. Gulbranson-Judge, D. R. Taylor, D. M. Sze, I. C. MacLennan, Immunoglobulin switch transcript production in vivo related to the site and time of antigen-specific B cell activation. *J Exp Med* **183**, 2303-2312 (1996).
21. K. A. Pape, V. Kouskoff , D. Nemazee , H. L. Tang , J. G. Cyster , L. E. Tze , K. L. Hippen , T. W. Behrens , M. K. Jenkins Visualization of the Genesis and Fate of Isotype-switched B Cells during a Primary Immune Response. *J Exp Med* **197**, 1677-1687 (2003).
22. J. A. Roco, L. Mesin, S. C. Binder, C. Nefzger, P. Gonzalez-Figueroa, P. F. Canete, J. Ellyard, Q. Shen, P. A. Robert, J. Cappello, H. Vohra, Y. Zhang, C. R. Nowosad, A. Schiepers, L. M. Corcoran, K. M. Toellner, J. M. Polo, M. Meyer-Hermann, G. D. Victora, C. G. Vinuesa, Class-Switch Recombination Occurs Infrequently in Germinal Centers. *Immunity* **51**, 337-350.e337 (2019).
23. T. Borggrefe, M. Wabl, A. T. Akhmedov, R. Jessberger, A B-cell-specific DNA Recombination Complex. *J Biol Chem* **273**, 17025-17035 (1998).
24. L. A. Hanakahi, L. A. Dempsey, M.-J. Li, N. Maizels, Nucleolin is one component of the B cell-specific transcription factor and switch region binding protein, LR1. *PNAS* **94**, 3605-3610 (1997).
25. S. Mondal, N. A. Begum, W. Hu, T. Honjo, Functional requirements of AID's higher order structures and their interaction with RNA-binding proteins. *PNAS* **113**, E1545-E1554 (2016).
26. A. Orthwein, A.-M. Patenaude, E. B. Affar, A. Lamarre, J. C. Young, J. M. Di Noia, Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. *J Exp Med* **207**, 2751-2765 (2010).
27. E. K. S. McRae, E. P. Booy, A. Moya-Torres, P. Ezzati, J. Stetefeld, S. A. McKenna, Human DDX21 binds and unwinds RNA guanine quadruplexes. *Nucleic Acids Res* **45**, 6656-6668 (2017).
28. S. Zheng, A. Kusnadi, J. E. Choi, B. Q. Vuong, D. Rhodes, J. Chaudhuri, NME proteins regulate class switch recombination. *FEBS Lett* **593**, 80-87 (2019).

29. E. Vigorito, K. L. Perks, C. Abreu-Goodger, S. Bunting, Z. Xiang, S. Kohlhaas, P. P. Das, E. A. Miska, A. Rodriguez, A. Bradley, K. G. C. Smith, C. Rada, A. J. Enright, K.-M. Toellner, I. C. M. MacLennan, M. Turner, microRNA-155 Regulates the Generation of Immunoglobulin Class-Switched Plasma Cells. *Immunity* **27**, 847-859 (2007).
30. R. E. Morman, P. G. Schweickert, S. F. Konieczny, E. J. Taparowsky, BATF regulates the expression of Nfil3, Wnt10a and miR155hg for efficient induction of antibody class switch recombination in mice. *Eur J Immunol* **48**, 1492-1505 (2018).
31. S. N. Willis, K. L. Good-Jacobson, J. Curtis, A. Light, J. Tellier, W. Shi, G. K. Smyth, D. M. Tarlinton, G. T. Belz, L. M. Corcoran, A. Kallies, S. L. Nutt, Transcription factor IRF4 regulates germinal center cell formation through a B cell-intrinsic mechanism. *J Immunol* **192**, 3200-3206 (2014).
32. P. Milpied, I. Cervera-Marzal, M.-L. Mollichella, B. Tesson, G. Brisou, A. Traverse-Glehen, G. Salles, L. Spinelli, B. Nadel, Human germinal center transcriptional programs are de-synchronized in B cell lymphoma. *Nat Immunol* **19**, 1013-1024 (2018).
33. J. A. Dubovsky, D. L. Chappell, B. K. Harrington, K. Agrawal, L. A. Andritsos, J. M. Flynn, J. A. Jones, M. E. Paulaitis, B. Bolon, A. J. Johnson, J. C. Byrd, N. Muthusamy, Lymphocyte cytosolic protein 1 is a chronic lymphocytic leukemia membrane-associated antigen critical to niche homing. *Blood* **122**, 3308-3316 (2013).
34. D. A. Girón-Pérez, E. Vadillo, M. Schnoor, L. Santos-Argumedo, Myo1e modulates the recruitment of activated B cells to inguinal lymph nodes. *J Cell Sci* **133**, jcs235275 (2020).
35. J. R. Muppidi, R. Schmitz, J. A. Green, W. Xiao, A. B. Larsen, S. E. Braun, J. An, Y. Xu, A. Rosenwald, G. Ott, R. D. Gascoyne, L. M. Rimsza, E. Campo, E. S. Jaffe, J. Delabie, E. B. Smeland, R. M. Braziel, R. R. Tubbs, J. R. Cook, D. D. Weisenburger, W. C. Chan, N. Vaidehi, L. M. Staudt, J. G. Cyster, Loss of signalling via Galpha13 in germinal centre B-cell-derived lymphoma. *Nature* **516**, 254-258 (2014).
36. R. C. Ferreira, Q. Pan-Hammarström, R. R. Graham, V. Gateva, G. Fontán, A. T. Lee, W. Ortmann, E. Urcelay, M. Fernández-Arquero, C. Núñez, G. Jorgensen, B. R. Ludviksson, S. Koskinen, K. Haimila, H. F. Clark, L. Klareskog, P. K. Gregersen, T. W. Behrens, L. Hammarström, Association of IFIH1 and other autoimmunity risk alleles with selective IgA deficiency. *Nat Genet* **42**, 777-780 (2010).
37. D. J. Todd, L. J. McHeyzer-Williams, C. Kowal, A. H. Lee, B. T. Volpe, B. Diamond, M. G. McHeyzer-Williams, L. H. Glimcher, XBP1 governs late events in plasma cell differentiation and is not required for antigen-specific memory B cell development. *J Exp Med* **206**, 2151-2159 (2009).
38. R. Gloury, D. Zotos, M. Zuidscherwoude, F. Masson, Y. Liao, J. Hasbold, L. M. Corcoran, P. D. Hodgkin, G. T. Belz, W. Shi, S. L. Nutt, D. M. Tarlinton, A. Kallies, Dynamic changes in Id3 and E-protein activity orchestrate germinal center and plasma cell development. *J Exp Med* **213**, 1095-1111 (2016).
39. C. Huang, H. Geng, I. Boss, L. Wang, A. Melnick, Cooperative transcriptional repression by BCL6 and BACH2 in germinal center B-cell differentiation. *Blood* **123**, 1012-1020 (2014).
40. N. Engels, J. Wienands, Memory control by the B cell antigen receptor. *Immunol Rev* **283**, 150-160 (2018).
41. G. R. Ehrhardt, J. T. Hsu, L. Gartland, C. M. Leu, S. Zhang, R. S. Davis, M. D. Cooper, Expression of the immunoregulatory molecule FcRH4 defines a distinctive tissue-based population of memory B cells. *J Exp Med* **202**, 783-791 (2005).
42. D. Lau, L. Y.-L. Lan, S. F. Andrews, C. Henry, K. T. Rojas, K. E. Neu, M. Huang, Y. Huang, B. DeKosky, A.-K. E. Palm, G. C. Ippolito, G. Georgiou, P. C. Wilson, Low CD21 expression defines a population of recent germinal center graduates primed for plasma cell differentiation. *Sci Immunol* **2**, eaai8153 (2017).
43. K. Thorarinsdottir, A. Camponeschi, N. Cavallini, O. Grimsholm, L. Jacobsson, I. Gjertsson, I. L. Mårtensson, CD21(-/low) B cells in human blood are memory cells. *Clin Exp Immunol* **185**, 252-262 (2016).
44. H. Li, F. Borrego, S. Nagata, M. Tolnay, Fc Receptor–like 5 Expression Distinguishes Two Distinct Subsets of Human Circulating Tissue–like Memory B Cells. *J Immunol* **196**, 4064-4074 (2016).
45. I. Dogan, B. Bertocci, V. Vilmont, F. Delbos, J. Megret, S. Storck, C. A. Reynaud, J. C. Weill, Multiple layers of B cell memory with different effector functions. *Nat Immunol* **10**, 1292-1299 (2009).
46. J. Lutz, K. Dittmann, M. R. Bosl, T. H. Winkler, J. Wienands, N. Engels, Reactivation of IgG-switched memory B cells by BCR-intrinsic signal amplification promotes IgG antibody production. *Nat Commun* **6**, 8575 (2015).
47. M. Seifert, M. Przekopowitz, S. Taudien, A. Lollies, V. Ronge, B. Drees, M. Lindemann, U. Hillen, H. Engler, B. B. Singer, R. Kuppers, Functional capacities of human IgM memory B cells in early inflammatory responses and secondary germinal center reactions. *PNAS* **112**, E546-555 (2015).
48. H. J. Wu, S. Bondada, CD72, a coreceptor with both positive and negative effects on B lymphocyte development and function. *J Clin Immunol* **29**, 12-21 (2009).
49. H. W. Sohn, H. Gu, S. K. Pierce, Cbl-b negatively regulates B cell antigen receptor signaling in mature B cells through ubiquitination of the tyrosine kinase Syk. *J Exp Med* **197**, 1511-1524 (2003).
50. S. J. Meyer, A. T. Linder, C. Brandl, L. Nitschke, B Cell Siglecs-News on Signaling and Its Interplay With Ligand Binding. *Front Immunol* **9**, 2820-2820 (2018).
51. M. Shabani, A. A. Bayat, M. Jeddi-Tehrani, H. Rabbani, M. Hojjat-Farsangi, C. Ulivieri, Z. Amirghofran, C. T. Baldari, F. Shokri, Ligation of human Fc receptor like-2 by monoclonal antibodies down-regulates B-cell receptor-mediated signalling. *Immunology* **143**, 341-353 (2014).
52. Y. Kochi, K. Myouzen, R. Yamada, A. Suzuki, T. Kurosaki, Y. Nakamura, K. Yamamoto, FCRL3, an Autoimmune Susceptibility Gene, Has Inhibitory Potential on B-Cell Receptor-Mediated Signaling. *J Immunol* **183**, 5502-5510 (2009).
53. M. van Keimpema, L. J. Grüneberg, M. Mokry, R. van Boxtel, M. C. van Zelm, P. Coffer, S. T. Pals, M. Spaargaren, The forkhead transcription factor FOXP1 represses human plasma cell differentiation. *Blood* **126**, 2098-2109 (2015).
54. L. Corcoran, D. Emslie, T. Kratina, W. Shi, S. Hirsch, N. Taubenheim, S. Chevrier, Oct2 and Obf1 as Facilitators of B:T Cell Collaboration during a Humoral Immune Response. *Front Immunol* **5**, 108 (2014).
55. W. Meng, B. Zhang, G. W. Schwartz, A. M. Rosenfeld, D. Ren, J. J. C. Thome, D. J. Carpenter, N. Matsuoka, H. Lerner, A. L. Friedman, T. Granot, D. L. Farber, M. J. Shlomchik, U. Hershberg, E. T. Luning Prak, An atlas of B-cell clonal distribution in the human body. *Nat Biotechnol* **35**, 879-884 (2017).

56. F.-E. Johansen, E. S. Baekkevold, H. S. Carlsen, I. N. Farstad, D. Soler, P. Brandtzaeg, Regional induction of adhesion molecules and chemokine receptors explains disparate homing of human B cells to systemic and mucosal effector sites: dispersion from tonsils. *Blood* **106**, 593-600 (2005).
57. T. G. Phan, D. Paus, T. D. Chan, M. L. Turner, S. L. Nutt, A. Basten, R. Brink, High affinity germinal center B cells are actively selected into the plasma cell compartment. *J Exp Med* **203**, 2419-2424 (2006).
58. N. R. Mathew, J. K. Jayanthan, I. Smirnov, J. L. Robinson, H. Axelsson, S. S. Nakka, A. Emmanouilidi, P. Czarnewski, W. T. Yewdell, C. Lebrero-Fernández, V. Bernasconi, A. M. Harandi, N. Lycke, N. Borcherding, J. W. Yewdell, V. Greiff, M. Bemark, D. Angeletti, Single cell BCR and transcriptome analysis after respiratory virus infection reveals spatiotemporal dynamics of antigen-specific B cell responses. *bioRxiv*, 2020.2008.2024.264069 (2020).
59. J. A. Vander Heiden, G. Yaari, M. Uduman, J. N. H. Stern, K. C. O'Connor, D. A. Hafler, F. Vigneault, S. H. Kleinstein, pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930-1932 (2014).
60. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).
61. N. T. Gupta, J. A. Vander Heiden, M. Uduman, D. Gadala-Maria, G. Yaari, S. H. Kleinstein, Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* **31**, 3356-3358 (2015).
62. J. Ye, N. Ma, T. L. Madden, J. M. Ostell, IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res* **41**, W34-W40 (2013).
63. D. Gadala-Maria, G. Yaari, M. Uduman, S. H. Kleinstein, Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. *PNAS* **112**, E862-870 (2015).
64. K. R. James, T. Gomes, R. Elmentaite, N. Kumar, E. L. Gulliver, H. W. King, M. D. Stares, B. R. Bareham, J. R. Ferdinand, V. N. Petrova, K. Polański, S. C. Forster, L. B. Jarvis, O. Suchanek, S. Howlett, L. K. James, J. L. Jones, K. B. Meyer, M. R. Clatworthy, K. Saeb-Parsy, T. D. Lawley, S. A. Teichmann, Distinct microbial and immune niches of the human colon. *Nat Immunol* **21**, 343-353 (2020).
65. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018).
66. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Systems* **8**, 329-337.e324 (2019).
67. J.-E. Park, R. A. Botting, C. Domínguez Conde, D.-M. Popescu, M. Lavaert, D. J. Kunz, I. Goh, E. Stephenson, R. Ragazzini, E. Tuck, A. Wilbrey-Clark, K. Roberts, V. R. Kedlian, J. R. Ferdinand, X. He, S. Webb, D. Maunder, N. Vandamme, K. T. Mahbubani, K. Polanski, L. Mamanova, L. Bolt, D. Crossland, F. de Rita, A. Fuller, A. Filby, G. Reynolds, D. Dixon, K. Saeb-Parsy, S. Lisgo, D. Henderson, R. Vento-Tormo, O. A. Bayraktar, R. A. Barker, K. B. Meyer, Y. Saeys, P. Bonfanti, S. Behjati, M. R. Clatworthy, T. Taghon, M. Haniffa, S. A. Teichmann, A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, eaay3224 (2020).
68. Y. Zhou, B. Zhou, L. Pache, M. Chang, A. H. Khodabakhshi, O. Tanaseichuk, C. Benner, S. K. Chanda, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).
69. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086 (2017).
70. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419 (2017).
71. C. Soneson, M. Love, M. Robinson, Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 2; peer review: 2 approved]. *F1000Research* **4**, (2016).
72. 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).
73. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184-1191 (2009).
74. T. Barrett, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C. L. Robertson, N. Serova, S. Davis, A. Soboleva, NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* **41**, D991-D995 (2012).
75. M. Efremova, M. Vento-Tormo, S. A. Teichmann, R. Vento-Tormo, CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc* **15**, 1484–1506 (2020).
76. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018).
77. V. Petukhov, J. Guo, N. Baryawno, N. Severe, D. T. Scadden, M. G. Samsonova, P. V. Kharchenko, dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. *Genome Biol* **19**, 78 (2018).

Figure Legends

Figure 1. Subclass- and subset-specific features of human tonsillar B cell repertoires.

A) Sorting strategy to isolate naïve, germinal centre (GC), memory (MBC) and plasmablasts (PB) from live CD19+ human tonsillar B cells for bulk repertoire sequencing.

B) Mean frequency of antibody subclasses within each B cell subset across donors (*n*=8 donors).

C) Mean frequency of antibody SHM levels within each B cell subset across donors.

D) Mean SHM frequencies per donor for subclass-specific antibody sequences within each B cell subset. Isotypes are sorted in order of the IgH locus. p values denote result of Student's T test.

E) Mean clonal diversity scores per donor of subclass- and subset-specific B cell clones. p values denote result of Student's T test.

F) Frequencies of switched subclass sequences per donor within each B cell subset. p values denote result of Student's T test.

G) Mean frequency of antibody subclasses for expanded clones spanning different B cell populations. For each class of clone, subset-specific members are examined (groups at top).

H) Observed/expected frequencies for isotype pairs detected in reconstructed phylogenies of clonally-related sequences for all B cell clones (left) or subset-specific sequences. Antibody subclasses are ordered according to the IgH locus. nd = not detected.

Figure 2. A single-cell atlas of human tonsillar immune cells to understand B cell maturation.

A) Schematic of tonsillar immune cell isolation followed by single-cell transcriptomics and antibody repertoire sequencing.

B) UMAP projection of tonsillar immune scRNA-seq and scVDJ data (32,607 cells; 7 donors).

C) UMAP projections of tonsillar T cell (8,753 cells) and non-lymphoid cell (363 cells) scRNA-seq clusters.

D) UMAP projection of tonsillar B cell (25,728 cells) scRNA-seq clusters.

E) Schematic of scVDJ and bulk repertoire integration analysis strategy.

F) Mean marker gene expression for B cell subsets. Dot size depicts frequency of cells a gene is detected in.

G) Relative scVDJ-derived antibody subclass frequencies in different B cell states.

H) SHM frequencies of scVDJ-derived antibody genes in different B cell states.

I) Clonal diversity scores (±SD) of B cell clones identified in scRNA-seq dataset.

J) Number of members per clonotype in B cell subsets from integrated repertoire analysis.

K) Co-occurrence of expanded scVDJ clones across B cell subsets.

L) Clonal relationships between scVDJ and sorted B cell subset repertoires.

M) Relative frequencies of B cell subsets separated by reason for tonsillectomy. Obstructive sleep apnoea (OSA; *n*=2), recurrent tonsillitis (RT; *n*=3), RT+OSA (*n*=2).

Figure 3. Mapping dynamics of human B cell activation and GC entry.

A) Marker gene expression for activated B cells in scRNA-seq.

B) Frequency of significant predicted ligand-receptor pair interactions between major B cell states, T cells and antigen-presenting cells (APCs).

C) Selected ligand–receptor interactions between B cells and APCs, CD4+ T cells and TfH cells.

D) Grid-based visualisation of tonsillar B cell RNA velocities. Arrow size conveys strength of directionality.

E) Relative frequencies of B cell types in a velocity-based pseudotime reconstruction of B cell activation and GC formation.

F) Heatmap depicting dynamic gene expression across velocity-based pseudotime reconstruction in E).

G) Anti-IgM-stimulation gene signature scores (±95% CI) across velocity-based pseudotime.

H) Expression of CSR genes through velocity-based pseudotime.

I) CSR gene signature score through velocity-based pseudotime.

J) Relative frequencies of cells with high CSR signature scores in different B cell states ($n = 7$). *p* denotes Student's T test.

K) Expression of genes implicated in CSR across B cell subsets.

L) Antibody subclass frequencies across velocity-based pseudotime.

M) Expression of germline transcripts (GLT) for IgM (I_μ), IgA (I_α) and IgG (I_γ) through velocity-based pseudotime. $I_{\alpha+\gamma}$ denotes sum of I_α and I_γ expression.

Figure 4. Resolving antibody class-dependent gene expression in GC B cells at single-cell resolution

A) Diffusion-based graph visualisation and pseudotemporal ordering of GC B cell scRNA-seq populations (6,475 cells). DC = diffusion component.

B) Single-cell gene expression heatmap of major GC B cell states ordered by pseudotime.

C) Expression of marker genes for prePB and FCRL3^{high} GC B cells.

D) Relative frequency of class-switched FCRL3^{high}, prePB cells and other GC B cells in scRNA/VDJ. (±SEM). p values denote result of Student's T test.

E) Percentage of unswitched members within expanded GC B cell clones. p values denote result of Wilcoxon Ranked Signed Sum test.

F) High and low affinity gene signature scores for GC B cells grouped by antibody SHM frequency (2,045 cells). Also shown are *CD27* expression, integrated clone size and predicted IgH functionality. p values denote result of Wilcoxon Ranked Signed Sum test.

G) Log enrichment of genes in class-switched vs unswitched and high vs low SHM GC B cells. Colour denotes statistical significance. c*or* denotes Pearson's Correlation coefficient.

H) Enrichment (z-scores) of genes differentially expressed by class-specific GC B cells with high SHM in gene ontologies and predicted targets of cytokine signalling from Ingenuity Pathway Analysis.

I) Mean expression of selected genes enriched in class-specific high SHM GC B cells.

Figure 5. Diverse MBC states and antibody class-dependent gene expression.

A) Clustering and UMAP visualisation of 21,595 MBC single-cell transcriptomes.

B) scVDJ-derived antibody isotypes of MBC single-cell transcriptomes with high quality VDJ sequences (*n*=15,531 cells).

C) Mean expression of top marker genes for MBC states.

D) Log fold change of genes significantly enriched in switched or unswitched MBCs with high SHM. CD21^{low} clusters were excluded due to low cell number.

E) Gene ontologies for genes significantly enriched in switched or unswitched MBCs with high SHM.

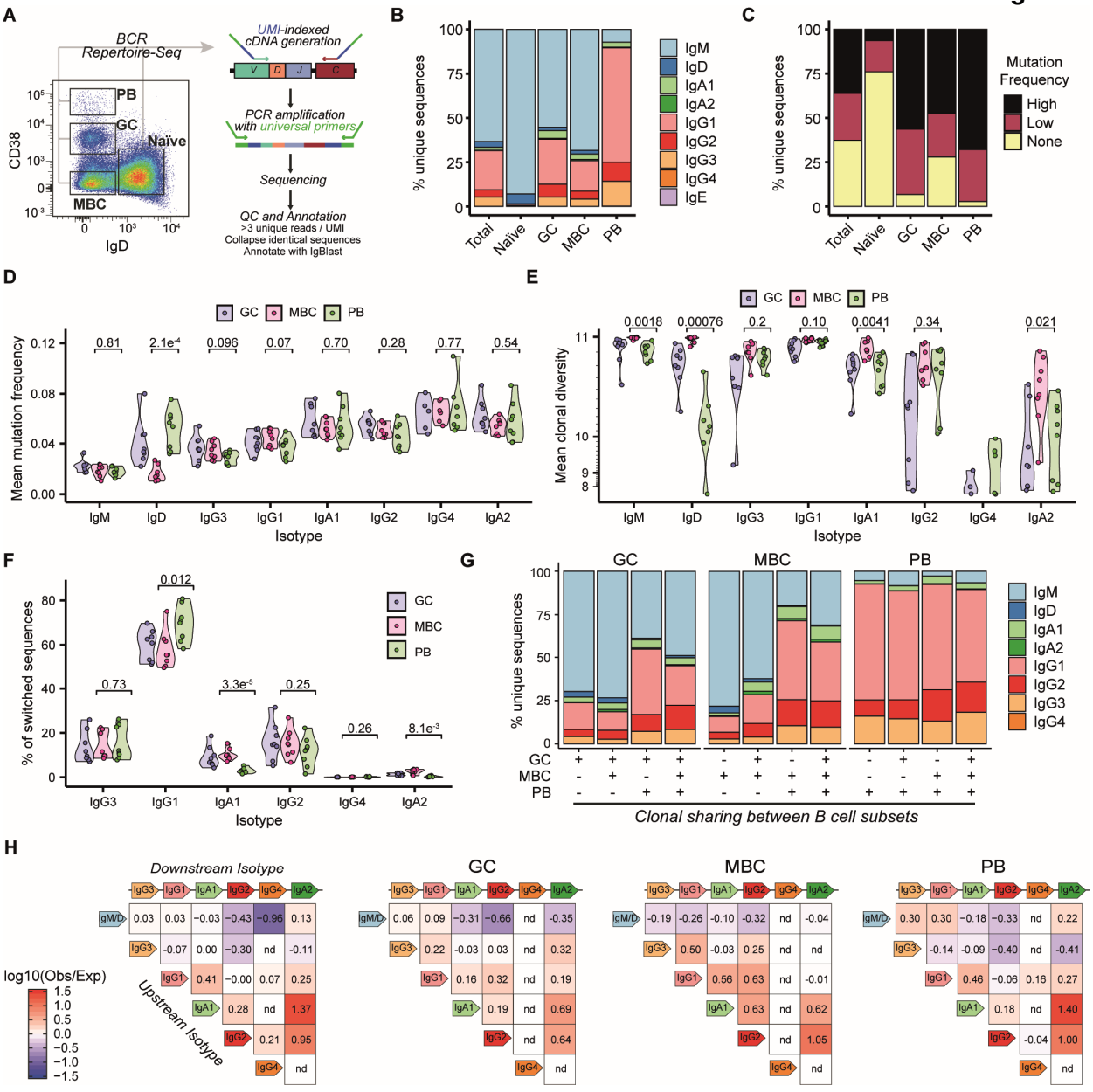
F) Log fold change of selected genes significantly enriched in switched or unswitched MBCs with high SHM.

G) Single-cell scores for GC entry and CSR signature genesets in switched and unswitched preGC MBC with high SHM. p values denote result of Wilcoxon Ranked Signed Sum test.

H) Single-cell gene expression of key CSR genes in switched and unswitched preGC MBC with high SHM. p values denote result of Wilcoxon Ranked Signed Sum test.

937

Figure 1



938

Figure 2

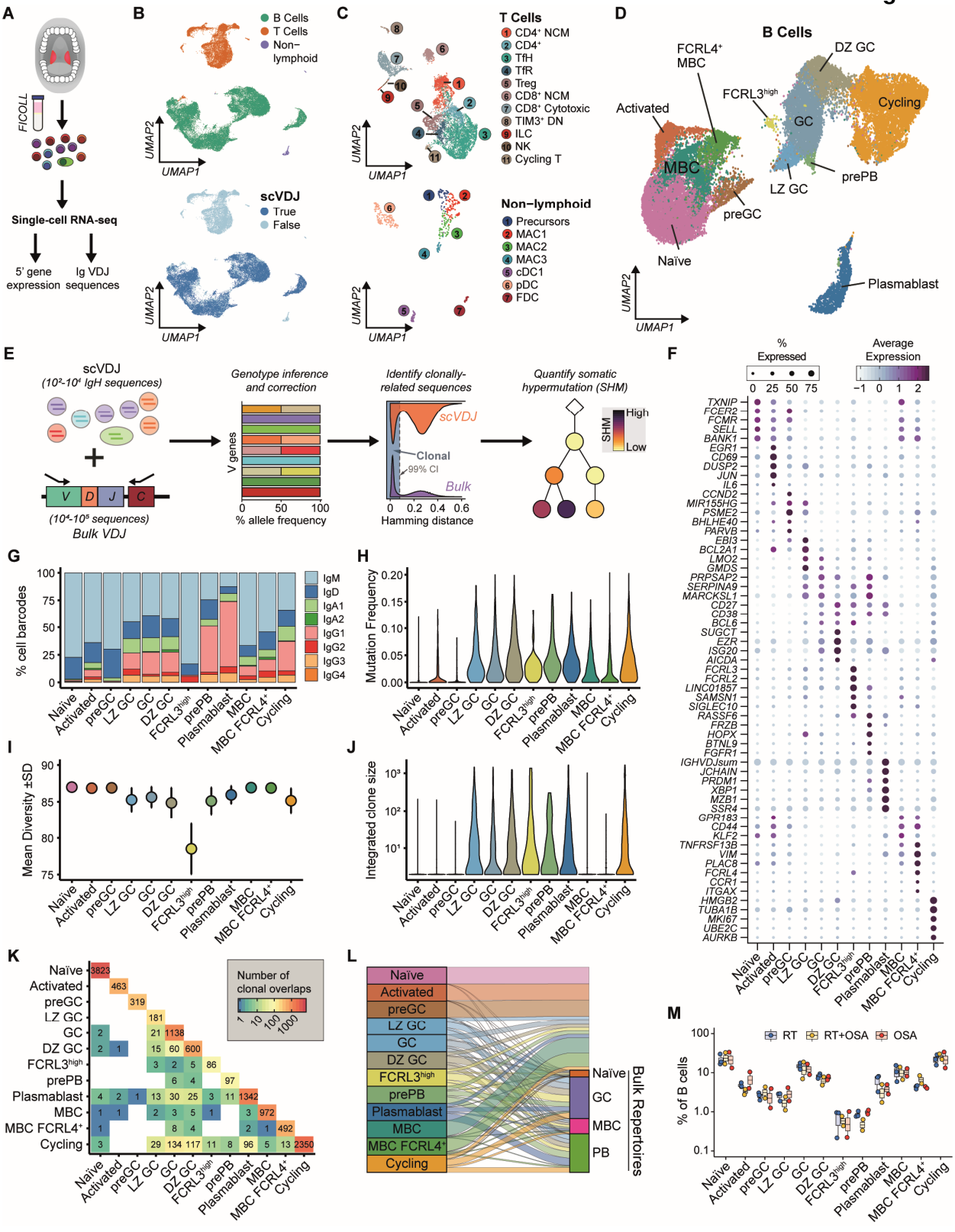

A
FICOLL
Single-cell RNA-seq
5' gene expression
Ig VDJ sequences
B
B Cells
T Cells
Non–lymphoid
scVDJ
True
False
UMAP1
UMAP2
C
T Cells
1 CD4+ NCM
2 CD4+
3 TfH
4 TfR
5 Treg
6 CD8+ NCM
7 CD8+ Cytotoxic
8 TIM3+ DN
9 ILC
10 NK
11 Cycling T
Non–lymphoid
1 Precursors
2 MAC1
3 MAC2
4 MAC3
5 cDC1
6 pDC
7 FDC
D
B Cells
FCRL4+ MBC
DZ GC
FCRL3high
Cycling
Activated
GC
MBC
LZ GC
prePB
preGC
Naïve
Plasmablast
E
scVDJ
(10²-10⁴ IgH sequences)
V D J C
(10⁴-10⁵ sequences)
Bulk VDJ
Genotype inference and correction
V genes
% allele frequency
Identify clonally-related sequences
scVDJ
Clonal
99% CI
Bulk
Hamming distance
Quantify somatic hypermutation (SHM)
SHM
High
Low
F
% Expressed
Average Expression
TXNIP
FCER2
FCMR
SELL
BANK1
EGR1
CD69
DUSP2
JUN
IL6
CCND2
MIR155HG
PSME2
BHLHE40
PARVB
EBI3
BCL2A1
LMO2
GMDS
PRPSAP2
SERPINA9
MARCKSL1
CD27
CD38
BCL6
SUGCT
EZR
ISG20
AICDA
FCRL3
FCRL2
LINC01857
SAMSN1
SIGLEC10
RASSF6
FRZB
HOPX
BTNL9
FGFR1
IGHVDJsum
JCHAIN
PRDM1
XBP1
MZB1
SSR4
GPR183
CD44
KLF2
TNFRSF13B
VIM
PLAC8
FCRL4
CCR1
ITGAX
HMGB2
TUBA1B
MKI67
UBE2C
AURKB
G
% cell barcodes
IgM
IgD
IgA1
IgA2
IgG1
IgG2
IgG3
IgG4
H
Mutation Frequency
I
Mean Diversity ±SD
J
Integrated clone size
K
Number of clonal overlaps
L
Naïve
Activated
preGC
LZ GC
GC
DZ GC
FCRL3high
prePB
Plasmablast
MBC
MBC FCRL4+
Cycling
Bulk Repertoires
M
RT
RT+OSA
OSA
% of B cells

Figure 3

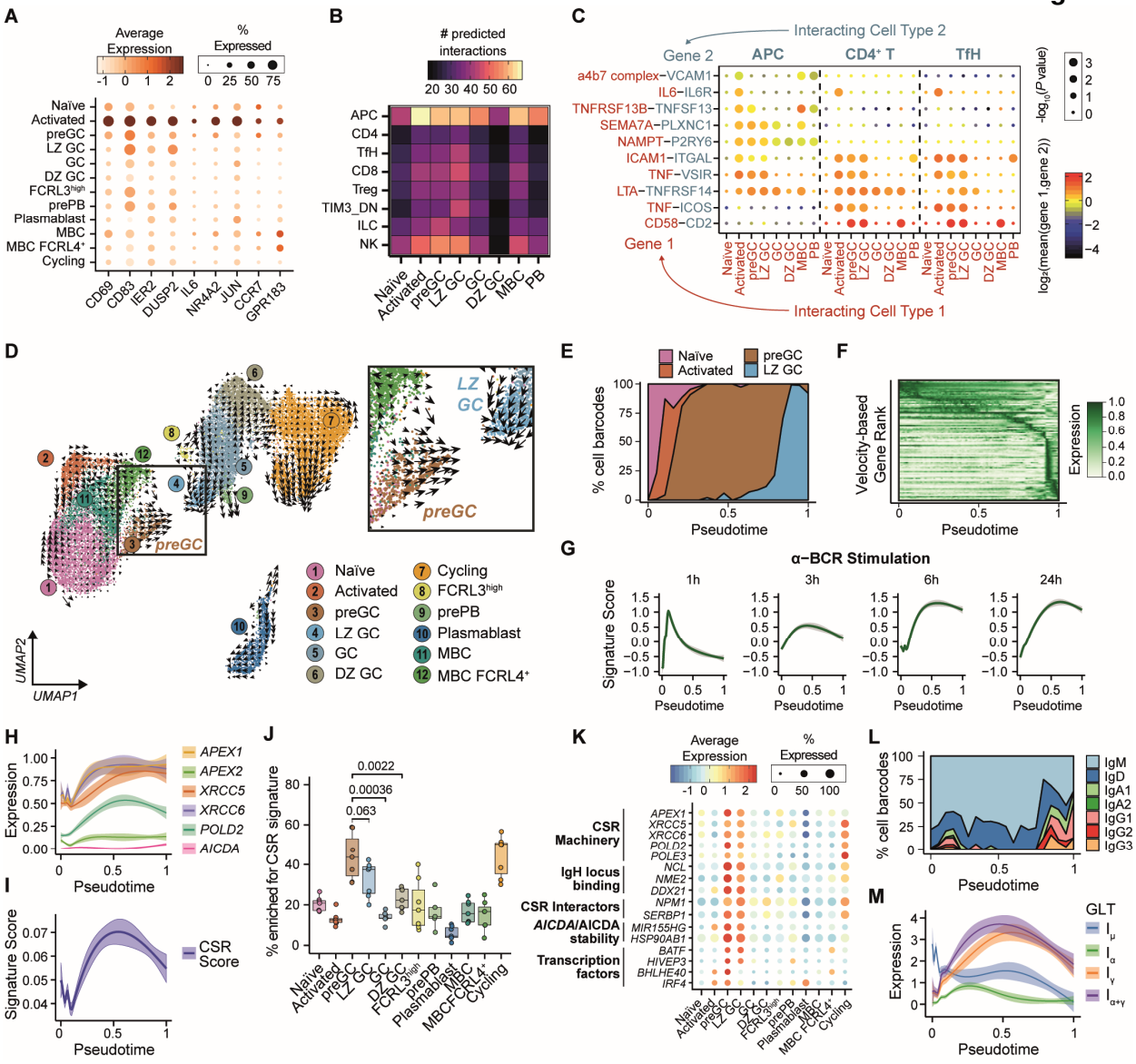

A
Average Expression
% Expressed
Naïve
Activated
preGC
LZ GC
GC
DZ GC
FCRL3high
prePB
Plasmablast
MBC
MBC FCRL4+
Cycling
CD69
CD83
IER2
DUSP2
IL6
NR4A2
JUN
CCR7
GPR183
B
predicted interactions
APC
CD4
TfH
CD8
Treg
TIM3_DN
ILC
NK
C
Interacting Cell Type 2
Gene 2
APC
CD4+ T
TfH
a4b7 complex–VCAM1
IL6–IL6R
TNFRSF13B–TNFSF13
SEMA7A–PLXNC1
NAMPT–P2RY6
ICAM1–ITGAL
TNF–VSIR
LTA–TNFRSF14
TNF–ICOS
CD58–CD2
Gene 1
Interacting Cell Type 1
-log10(P value)
log2(mean(gene 1,gene 2))
D
LZ GC
preGC
UMAP2
UMAP1
1 Naïve
2 Activated
3 preGC
4 LZ GC
5 GC
6 DZ GC
7 Cycling
8 FCRL3high
9 prePB
10 Plasmablast
11 MBC
12 MBC FCRL4+
E
Naïve
Activated
preGC
LZ GC
% cell barcodes
Pseudotime
F
Velocity-based Gene Rank
Expression
Pseudotime
G
α–BCR Stimulation
1h
3h
6h
24h
Signature Score
Pseudotime
H
Expression
APEX1
APEX2
XRCC5
XRCC6
POLD2
AICDA
Pseudotime
I
Signature Score
CSR Score
Pseudotime
J
% enriched for CSR signature
0.0022
0.00036
0.063
Naïve
Activated
preGC
LZ GC
GC
DZ GC
FCRL3high
prePB
Plasmablast
MBC
MBCFCRL4+
Cycling
K
Average Expression
% Expressed
CSR Machinery
IgH locus binding
CSR Interactors
AICDA/AICDA stability
Transcription factors
APEX1
XRCC5
XRCC6
POLD2
POLE3
NCL
NME2
DDX21
NPM1
SERBP1
MIR155HG
HSP90AB1
BATF
HIVEP3
BHLHE40
IRF4
L
% cell barcodes
IgM
IgD
IgA1
IgA2
IgG1
IgG2
IgG3
Pseudotime
M
Expression
GLT
Pseudotime

Figure 4

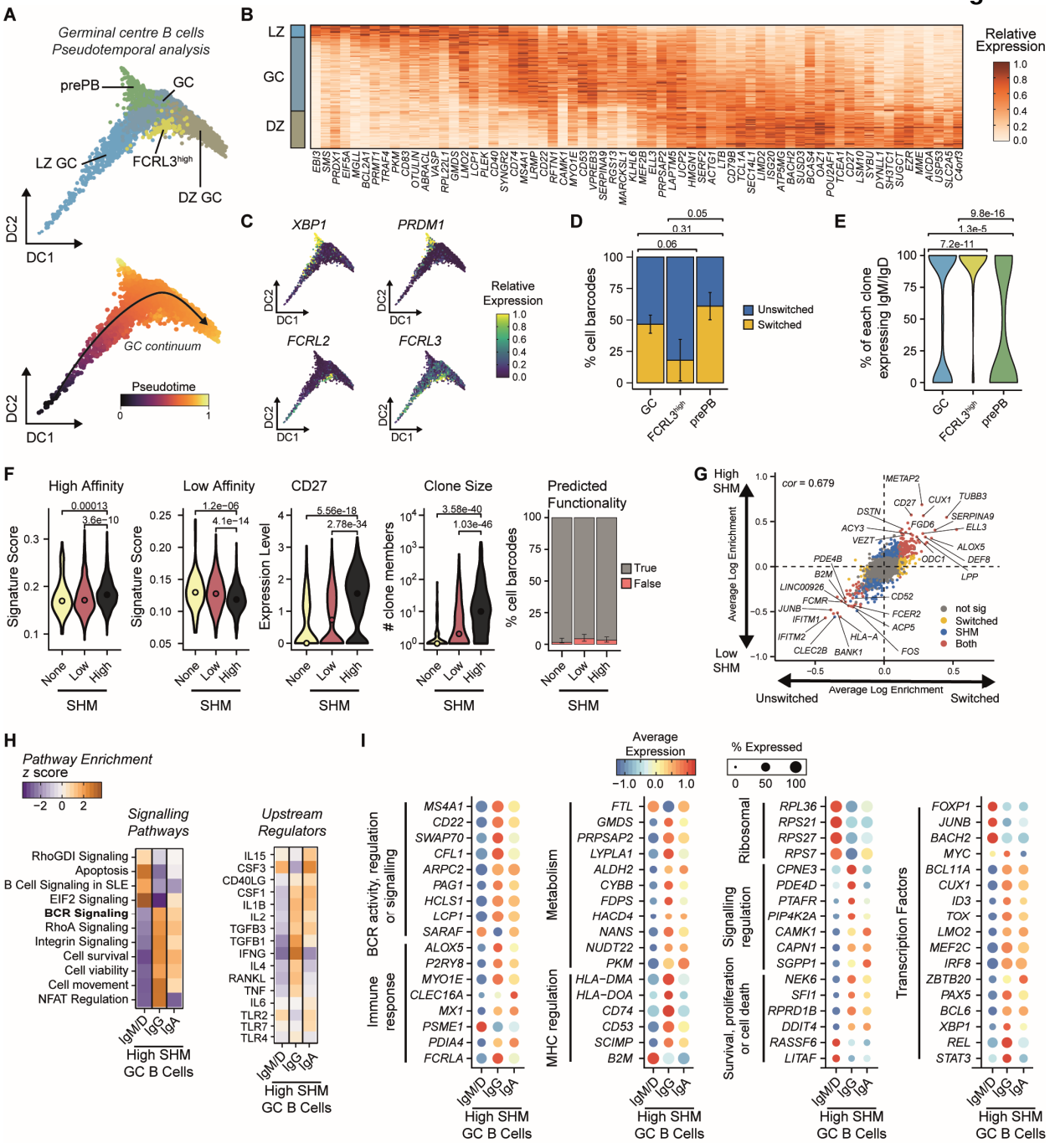
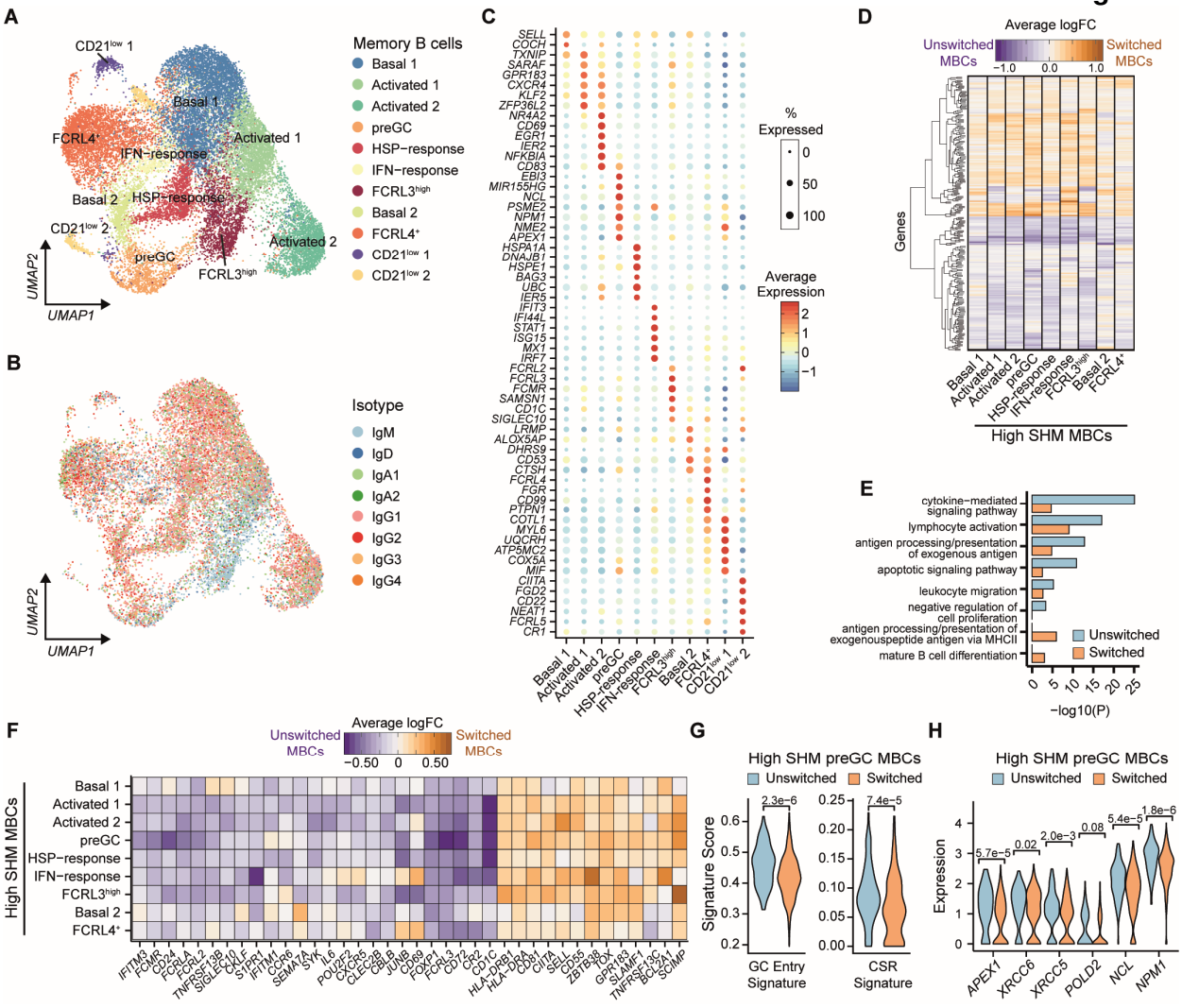

A
Germinal centre B cells
Pseudotemporal analysis
prePB
GC
LZ GC
FCRL3high
DZ GC
DC1
DC2
GC continuum
Pseudotime
B
LZ
GC
DZ
Relative Expression
C
XBP1
PRDM1
FCRL2
FCRL3
D
% cell barcodes
Unswitched
Switched
E
% of each clone expressing IgM/IgD
F
High Affinity
Low Affinity
CD27
Clone Size
Predicted Functionality
Signature Score
Expression Level
clone members
% cell barcodes
True
False
None
Low
High
SHM
G
High SHM
Low SHM
cor = 0.679
Average Log Enrichment
Unswitched
Switched
not sig
Switched
SHM
Both
H
Pathway Enrichment
z score
Signalling Pathways
RhoGDI Signaling
Apoptosis
B Cell Signaling in SLE
EIF2 Signaling
BCR Signaling
RhoA Signaling
Integrin Signaling
Cell survival
Cell viability
Cell movement
NFAT Regulation
Upstream Regulators
IL15
CSF3
CD40LG
CSF1
IL1B
IL2
TGFB3
TGFB1
IFNG
IL4
RANKL
TNF
IL6
TLR2
TLR7
TLR4
IgM/D
IgG
IgA
High SHM GC B Cells
I
Average Expression
% Expressed
BCR activity, regulation or signalling
Immune response
Metabolism
MHC regulation
Ribosomal
Signalling regulation
Survival, proliferation or cell death
Transcription Factors

Figure 5


A
CD21low 1
Basal 1
FCRL4+
Activated 1
IFN−response
Basal 2
HSP−response
CD21low 2
Activated 2
preGC
FCRL3high
UMAP2
UMAP1
Memory B cells
Basal 1
Activated 1
Activated 2
preGC
HSP−response
IFN−response
FCRL3high
Basal 2
FCRL4+
CD21low 1
CD21low 2
B
UMAP2
UMAP1
Isotype
IgM
IgD
IgA1
IgA2
IgG1
IgG2
IgG3
IgG4
C
SELL
COCH
TXNIP
SARAF
GPR183
CXCR4
KLF2
ZFP36L2
NR4A2
CD69
EGR1
IER2
NFKBIA
CD83
EBI3
MIR155HG
NCL
PSME2
NPM1
NME2
APEX1
HSPA1A
DNAJB1
HSPE1
BAG3
UBC
IER5
IFIT3
IFI44L
STAT1
ISG15
MX1
IRF7
FCRL2
FCRL3
FCMR
SAMSN1
CD1C
SIGLEC10
LRMP
ALOX5AP
DHRS9
CD53
CTSH
FCRL4
FGR
CD99
PTPN1
COTL1
MYL6
UQCRH
ATP5MC2
COX5A
MIF
CIITA
FGD2
CD22
NEAT1
FCRL5
CR1
Basal 1
Activated 1
Activated 2
preGC
HSP−response
IFN−response
FCRL3high
Basal 2
FCRL4+
CD21low 1
CD21low 2
% Expressed
0
50
100
Average Expression
2
1
0
−1
D
Average logFC
Unswitched MBCs
Switched MBCs
−1.0
0.0
1.0
Genes
Basal 1
Activated 1
Activated 2
preGC
HSP−response
IFN−response
FCRL3high
Basal 2
FCRL4+
High SHM MBCs
E
cytokine−mediated signaling pathway
lymphocyte activation
antigen processing/presentation of exogenous antigen
apoptotic signaling pathway
leukocyte migration
negative regulation of cell proliferation
antigen processing/presentation of exogenouspeptide antigen via MHCII
mature B cell differentiation
Unswitched
Switched
0
5
10
15
20
25
−log10(P)
F
Average logFC
Unswitched MBCs
Switched MBCs
−0.50
0
0.50
High SHM MBCs
Basal 1
Activated 1
Activated 2
preGC
HSP−response
IFN−response
FCRL3high
Basal 2
FCRL4+
IFITM3
FCMR
CD24
FCRL4
FCRL2
TNFRSF13B
SIGLEC10
CKLF
STYP1
IFITM1
CCR6
SEMA7A
SYK
IL6
POU2F2
CXCR5
CLEC2B
CBLB
JUNB
CD69
FOXP1
FCRL3
CD72
CR2
CD1C
HLA−DRB1
HLA−DRA
CD81
CIITA
SELL
CD55
ZBTB38
TOX
GPR183
SLAMF1
TNFRSF13C
BCL2A1
SCIMP
G
High SHM preGC MBCs
Unswitched
Switched
Signature Score
2.3e−6
7.4e−5
GC Entry Signature
CSR Signature
H
High SHM preGC MBCs
Unswitched
Switched
Expression
5.7e−5
0.02
2.0e−3
0.08
5.4e−5
1.8e−6
APEX1
XRCC6
XRCC5
POLD2
NCL
NPM1

Supplementary Methods

Bulk VDJ repertoire library protocol

Bulk B cell repertoire libraries of immunoglobulin heavy chains (IgH) were generated as described previously (*5*), with minor changes. 50 to 100 ng RNA from sorted B cell subsets were annealed to a pooled set of five isotype-specific IgH constant region primers containing unique molecular identifiers (UMIs) of either 10 or 12 nucleotides at 72°C for 5 min before being immediately placed on ice for 2 min. First-strand cDNA synthesis was performed using SuperScript IV reverse transcriptase (ThermoFisher Scientific) with recommended reagent concentrations and the following cycling conditions in a thermocycler: 105°C lid; 55°C 10 min; 80°C 10 min; 4°C hold. Second-strand cDNA synthesis was performed using Phusion® High-Fidelity DNA Polymerase (NEB) and six IgH variable region primers containing 10 or 12 nucleotide UMIs with the following cycling conditions: 105°C lid; 98°C 4 min; 52°C 1 min; 72°C 5 min; 4°C hold. Double-stranded cDNA was then purified using (0.6X) Ampure XP beads (Beckman Coulter) before amplification with Illumina adapter-containing primers (Nextera i7 indices) and NEBNext Ultra II Q5 Master Mix (NEB) as follows: 105°C lid; 98°C 30 seconds; (98°C 10 seconds, 72°C 50 seconds) × 22 to 28 cycles; 72°C 2 min; 4°C hold. Amplified libraries were purified using (0.6X) Ampure XP beads and quantified by Qubit™ dsDNA HS Assay Kit prior to multiplexing. Libraries were sequenced with a 5% PhiX spike-in using paired-end 301 bp reads on the Illumina MiSeq.

Quality control and sequence assembly of bulk B cell repertoires

Raw sequencing reads from bulk VDJ libraries were processed to generate UMI-collapsed consensus VDJ sequences using pRESTO (v0.5.10) (*59*). Paired-end sequencing reads with mean Phred quality scores less than 25 were removed, and remaining sequences were annotated and trimmed for PCR primer and UMI sequences. UMI barcodes were then filtered by length and the presence of ambiguous nucleotides, prior to UMI alignment using MUSCLE (v3.8.31) (*60*). To correct for sequencing or other errors, we generated consensus sequences from UMIs with at least 3 unique sequencing reads required, prior to assembly of paired-end UMI consensus sequences into a single VDJ contig and annotation of constant region isotype using MaskPrimers.py align to correct for primer misalignment.

Duplicate VDJ sequences within each subset were collapsed using CollapseSeq.py before VDJ gene assignment and functional annotation with AssignGenes.py (ChangeO v0.4.5) (*61*) and IgBLAST (v1.12.0) (*62*).

Identification of clonally-related sequences, genotype inference and calculation of IgH mutation frequencies.

Following initial quality control, all single-cell VDJ sequences were combined together with bulk BCR repertoire sequences from the same donor for subsequent processing. IgH sequences were annotated using AssignGenes.py and IgBLAST before isotype class assignment prior to correction of ambiguous V gene assignments using TIgGER (v0.3.1) (*61,63*). Clonally-related IgH sequences were identified using DefineClones.py (ChangeO) with a nearest neighbour distance threshold of 0.0818, as determined by the mean 99% confidence interval of all 8 donors with distToNearest (Shazam v0.1.11) (*61*). CreateGermlines.py was then used to infer germline sequences for each clonal family and observedMutations was used to calculate somatic hypermutation frequencies for each IgH sequence. Sequences with somatic hypermutation frequencies greater than 0.02 were annotated as “High” mutation levels, those between 0 and 0.02 as “Low” mutation levels and 0 as “None”. For bulk BCR repertoire analysis in Figure 1, single-cell VDJ sequences were excluded, providing ~1.5 million high-confidence and unique IgH sequences, with a median of 14 UMIs per sequence, a median of 28,918 unique sequences per donor per subset and approximately 96-99% of these sequences annotated as functional by IgBlast. To quantify antibody class switch hierarchies, lineage trees for expanded clonotypes were constructed via maximum parsimony using buildPhylipLineage (Alakazam v0.2.11) (*61*). The observed frequency of direct edges between unique sequences of different isotypes were counted and expressed relative to the frequency of such transitions expected by chance (calculated by iterative (k=100) random sampling of isotype frequencies) within each B cell subsets.

Single-cell VDJ analysis was performed broadly as described previously (*64*). Briefly, the number of quality filtered and annotated IgH, IgK or IgL were determined per unique cell barcode prior to integration with single-cell gene expression objects. If more than one contig per chain was identified,

metadata for that cell was ascribed as “Multi”. IgH diversity analyses were performed using the rarefyDiversity and testDiversity of Alakazam. To assess clonal relationships between cell types, co-occurrence of expanded clone members between cell types was reported as a binary event for each clone that contained a member within two different cell types in either single-cell or bulk repertoires.

Data quality control, processing and annotation of single-cell RNA-seq.

Gene expression count matrices from cellranger were used to calculate percentage mitochondrial expression per cell barcode prior to mitochondrial genes being removed from gene expression matrices. Similarly, the V, D and J gene counts for each immunoglobulin and T cell receptor were summed to calculate an overall expression before individual genes were removed from gene expression matrices. Counts of individual IgH constant region genes were also summed together (IgG1-4, and IgA1-A2) and removed. Modified gene-by-cell matrices were then used to create Seurat objects for each sample using Seurat (v3.0.3) (*65*), removing genes expressed in fewer than 3 cells. Cell barcodes with <1000 or >60000 UMIs and <500 or >7000 genes detected were removed, as were cell barcodes with >30% mitochondrial reads. Individual matrices were then log transformed, normalised by a factor of 10000 prior to predicting cell cycle phases using the CellCycleScoring command and identifying the 3000 most variable genes within each sample using the “vst” method. A preliminary integration of all unsorted immune cells or all sorted memory B cell datasets together was performed using FindIntegrationAnchors and IntegrateData (3000 genes) before regressing out cell cycle scores and mitochondrial gene expression, performing principle component analysis (PCA) and preliminary clustering and cell type annotation. One preliminary cluster was identified to be enriched with predicted doublets based on the results from DoubletFinder (v2.0.1) (*66*), and a small number of cell barcodes with co-expression of B/T/non-lymphoid markers were removed. Following the removal of poor quality cell barcodes based on these preliminary analyses, all normalised count matrices were integrated together using the unsorted immune samples as a reference with 4000 highly variable genes before scaling the integrated data and regressing cell cycle and mitochondrial gene expression, running PCA and identifying broad cell type lineages (B cell, T cell and non-lymphoid cells) using a broad resolution for clustering. These lineages were then separated for more detailed annotation by recomputing the

PCA (RunPCA), nearest neighbour graph (FindNeighbors) and unbiased clustering (FindClusters). Uniform Manifold Approximation and Projection (UMAP) was used to visualise integrated and lineage-specific datasets. B cells were annotated with scVDJ metadata from the integrated repertoire analysis detailed above and features such as isotype frequency, SHM levels and clonal properties were used to improve confidence of cell type annotation (such as between naïve and MBC clusters).

Prediction of cell-cell communication using CellPhoneDB.

To evaluate potential cell-cell communication, we used CellPhoneDB (v2.0.6) (*75*) to examine the expression of ligand-receptor pairs between different scRNA-seq clusters. Briefly, we exported raw gene count matrices from Seurat, converted gene IDs to Ensembl IDs using bioMart. We re-annotated all non-lymphoid cell type clusters as antigen-presenting cells (APCs), naïve and effector T cell groups by CD4 or CD8 expression, Treg and Tfr as "Treg" and rare GC subsets (prePB and FCRL3^{high}) as "GC" and exported cell type metadata for use with raw count data using the "statistical_analysis" command of CellPhoneDB with database v2.0.0. The number of unique significant ligand-receptor co-expression pairs (putative interactions; p value<0.05) between each cell type was then counted and visualised as a heatmap, while exemplar interacting pairs were visualised by calculating mean average expression level of gene 1 in cell type 1 and gene 2 in cell type 2 are indicated by colour and p values indicated by circle size.

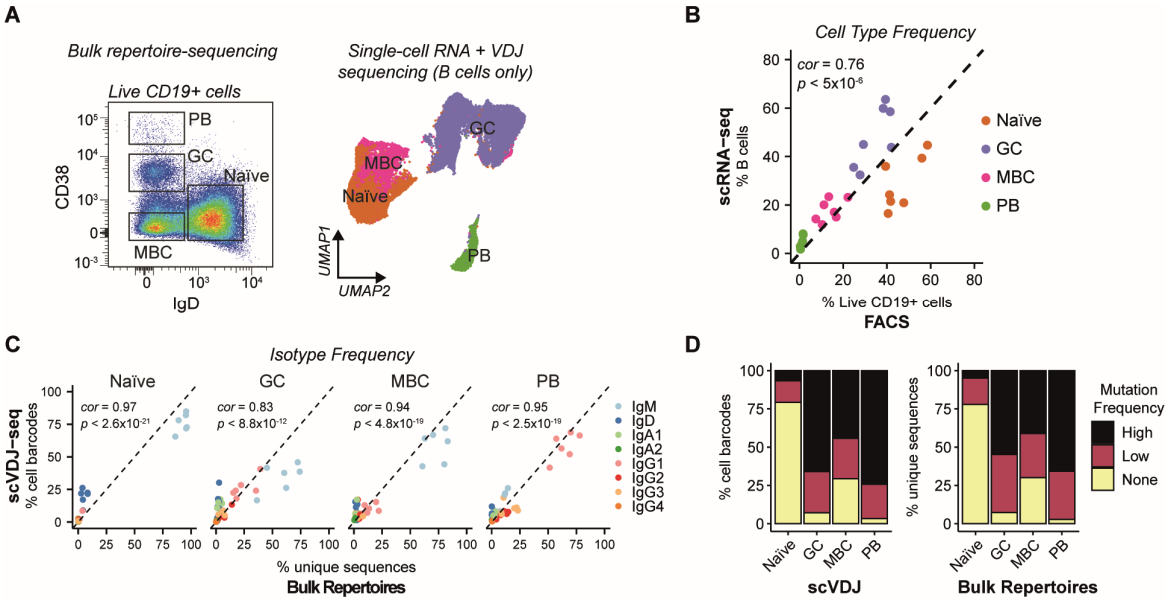


Figure S1. Comparison of FACS-based and scRNA-seq analysis of major B cell subsets.

A) Comparison of gates used for FACS-based sorting of B cell subsets for bulk repertoire sequencing (Fig1) with matched single-cell RNA-seq (scRNA) cell populations. See Fig2 and Methods for full details about scRNA-seq analyes.

B) Cell type frequencies derived from FACS or scRNA-seq of B cell populations shown in A) for each donor with both FACS and scRNA-seq data (*n*=7). *cor* denotes Pearson correlation coefficient and *p* denotes p value.

C) Antibody isotype frequencies derived from FACS or scVDJ-seq of B cell populations for each donor with both FACS and scVDJ-seq data (*n*=6). *cor* denotes Pearson correlation coefficient and *p* denotes p value.

D) Mutation frequency of IgH sequences derived from FACS or scVDJ-seq of B cell populations.

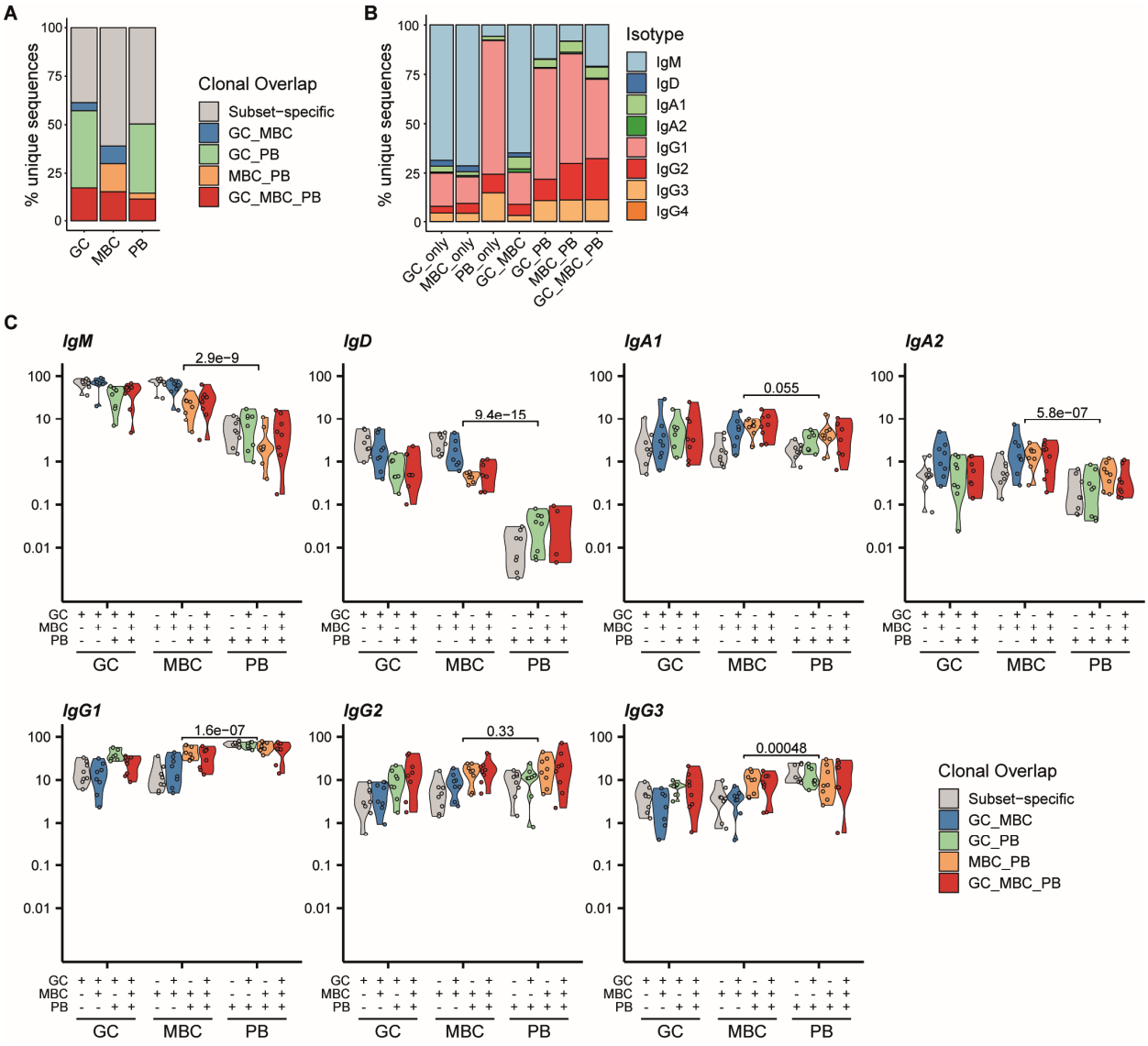


Figure S2. Clonal overlap of expanded B cell clones between GC, MBC and PB subsets.

A) Frequency of clonotype sharing for expanded B cell clones (>2 members) between major B cell subsets. Eg. GC_MBC denotes clone members were identified in both GC and MBC sorted repertoires.

B) Frequency of antibody isotypes for expanded B cell clones spanning multiple subsets.

C) Antibody subclass frequencies for expanded clones spanning different B cell populations (x axis). For each class of clone, subset-specific members are examined (see groups at bottom). Data shown are individual data points used to calculate mean frequencies in Fig1G. p values denote result of Student's T Test.

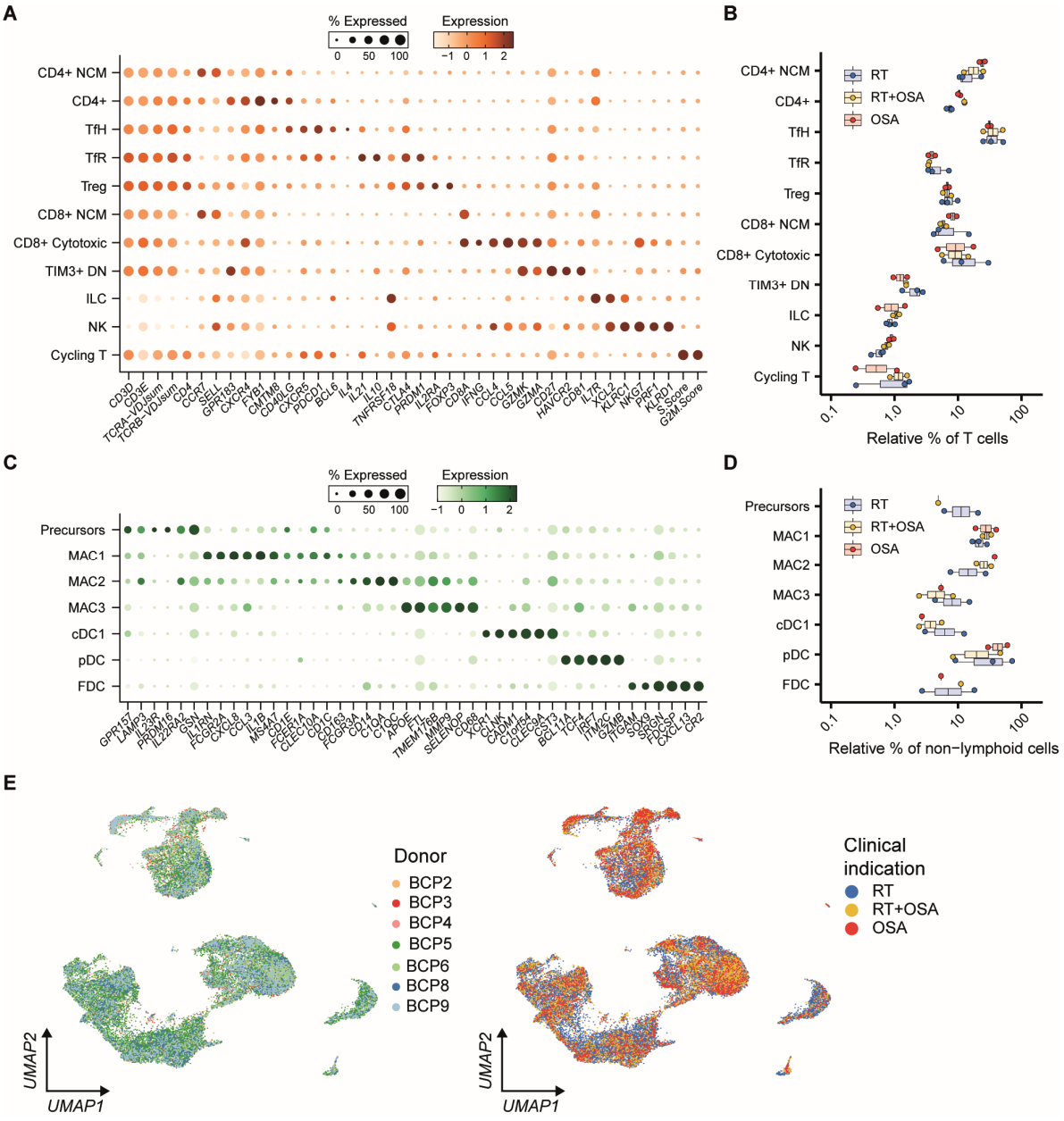


Figure S3. Analysis of T cell and non-lymphoid cell populations in human tonsils by scRNA-seq.

A) Mean expression of key marker genes used to define T cell scRNA-seq clusters, including CD4⁺ naïve or central memory (CD4⁺ NCM), CD4⁺, T follicular helper (TfH), T follicular regulatory (TfH), T regulatory (Treg), CD8⁺ naïve or central memory (CD8⁺ NCM), CD8⁺ cytotoxic (CD8⁺ CTL), TIM3⁺ CD4/CD8 double-negative (TIM3⁺ DN) and cycling T cells, and innate lymphoid cells (ILC) and natural killer (NK) cells. Size reflects frequency of cells in which each gene is detected.

B) Relative frequencies of different T cell subsets separated by clinical indication for tonsillectomy. OSA = obstructive sleep apnoea, RT = recurrent tonsillitis

C) Mean expression of key marker genes used to define non-lymphoid cell scRNA-seq clusters, including monocyte/macrophages precursor (Precursors), macrophage (MAC1, MAC2, MAC3), conventional dendritic cell 1 (cDC1), plasmacytoid-derived dendritic cell (pDC) and follicular dendritic cell (FDC) subsets.

D) Relative frequencies of different non-lymphoid cell subsets separated by clinical indication for tonsillectomy.

E) UMAP projection of tonsillar immune scRNA-seq data (32,607 cells; 7 donors) annotated by donor or clinical indication for tonsillectomy (OSA (n = 2), RT (n = 3), RT+OSA (n = 2)).

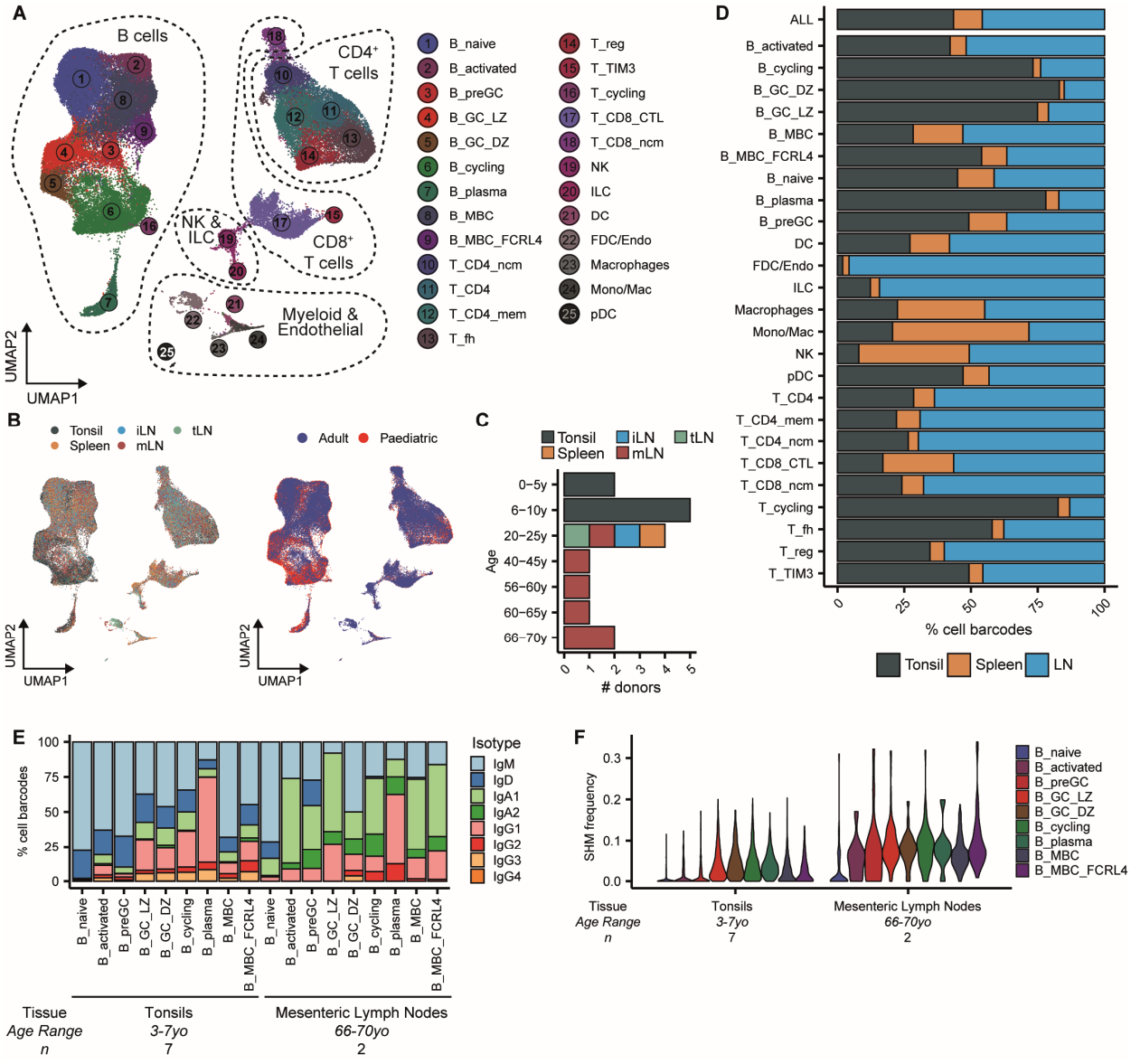


Figure S4. Integration of tonsillar scRNA-seq datasets with lymph node and spleen scRNA-seq.

A) UMAP visualisation of integrated secondary lymphoid organ scRNA-seq datasets from human tonsils (*n*=7), lymph nodes, (mesenteric (*n*=6); thoracic (*n*=1); inguinal (*n*=1)) and spleen (*n*=1) with cell type clusters annotated (74,607 cells).

B) UMAP of integrated secondary lymphoid organ scRNA-seq showing tissue and patient age at time of collection.

C) Age of patient donors by tissue.

D) Relative frequency of cells from different secondary lymphoid organs in annotated cell type clusters.

E) Relative frequency of scVDJ antibody isotype within B cell type clusters between tonsils (*n*=7) and mesenteric lymph nodes (*n*=2 for which scVDJ was available). Age range for samples from each organ is annotated.

F) SHM frequencies from scVDJ data for B cell type clusters between tonsils and mesenteric lymph nodes.

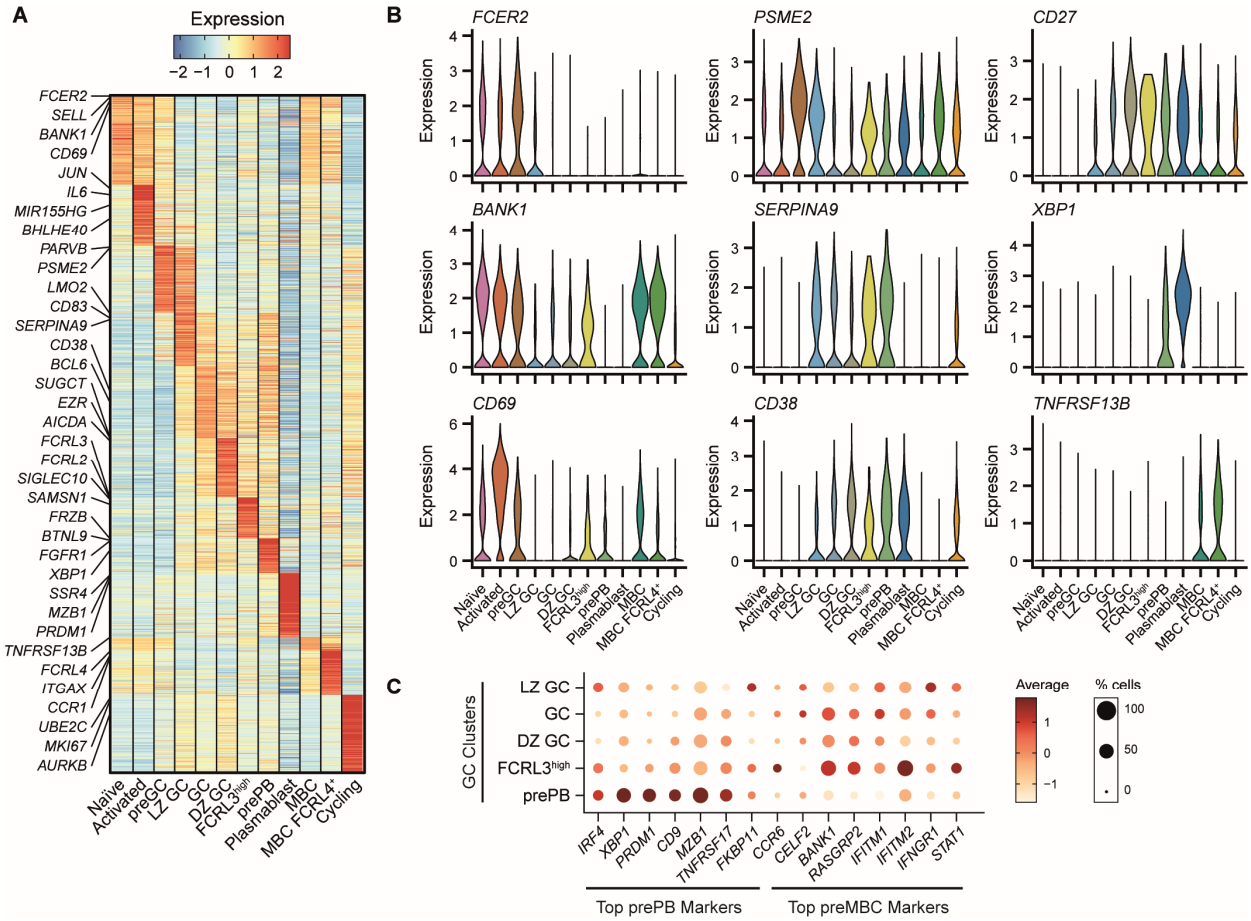


Figure S5. Marker gene analysis of human B cell subsets from tonsils by single-cell RNA-seq.

A) Average gene expression from scRNA-seq for key marker genes in each of the 12 B cell subsets identified. The top 200 genes per cluster with p_val_adj < 0.05 and average log fold change > 0.3 are shown, with key markers of labelled.

B) Violin plots of important B cell markers from scRNA-seq in the 12 B cell subsets.

C) Expression in tonsillar GC B cell clusters of genes identified as prePB or preMBC markers in Holmes *et al.* 2020 (16).

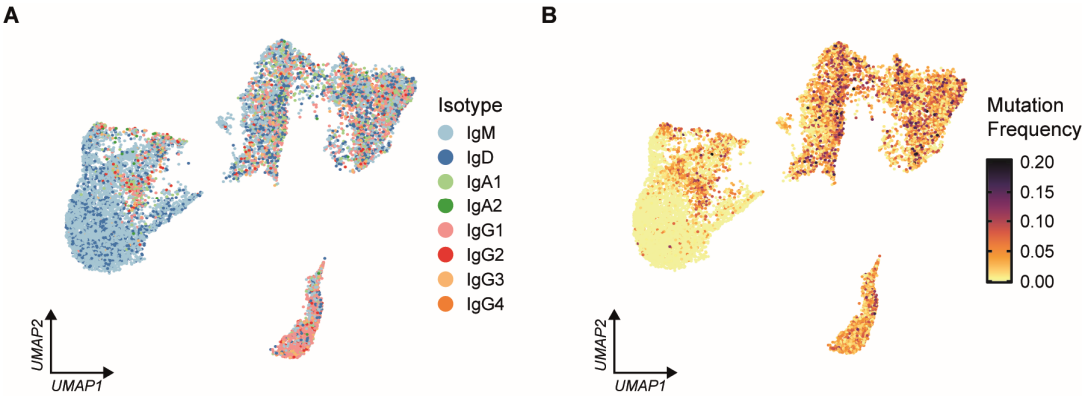


Figure S6. UMAP visualisation of B cell scVDJ antibody isotype and SHM frequency.

A) UMAP of tonsillar B cells (25,728 cells) annotated for scVDJ antibody isotype.

B) Same as in A), except for scVDJ SHM frequencies.

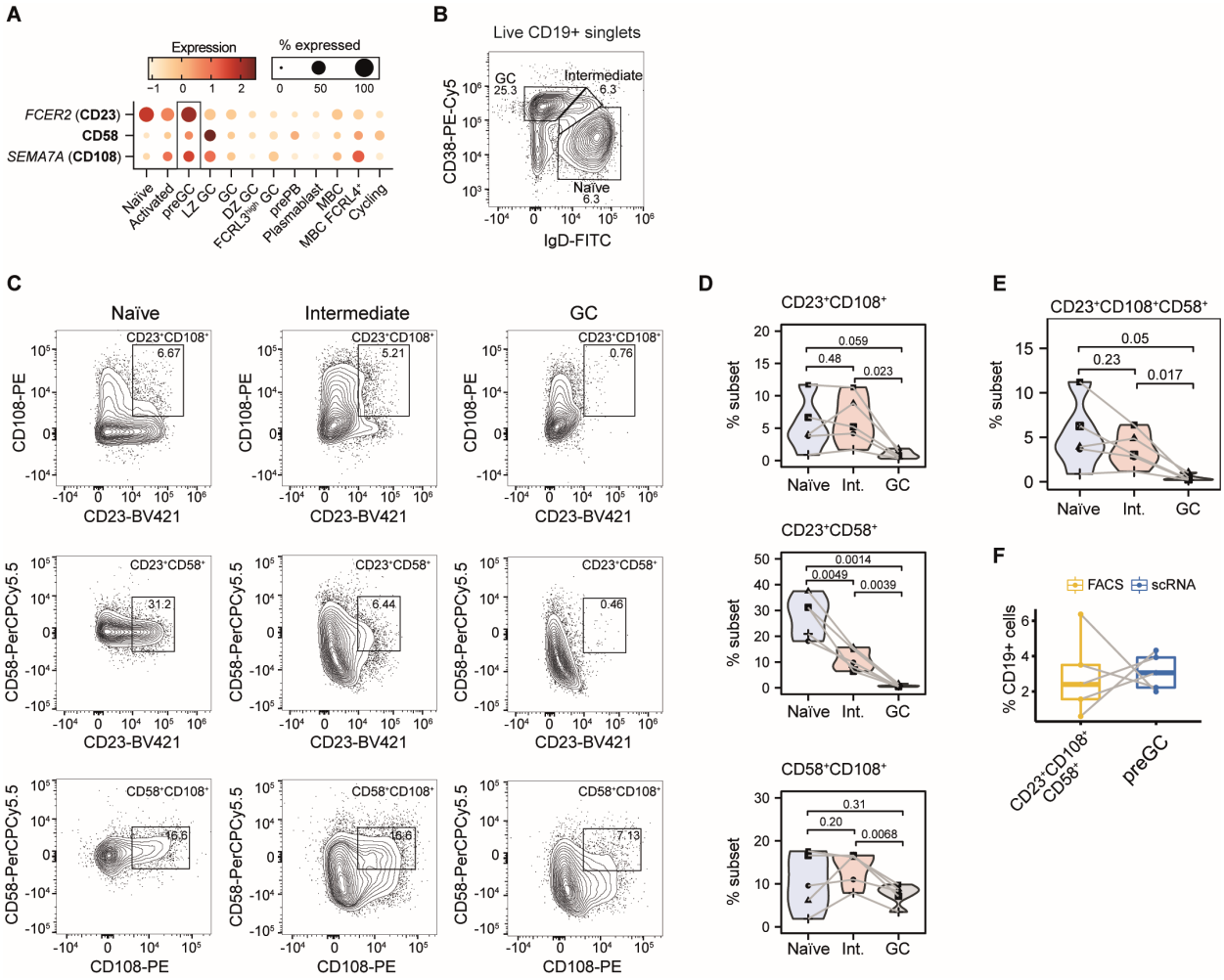


Figure S7. Flow cytometry analysis of CD23, CD58 and CD108 expression in naïve and GC B cells.

A) Expression of *FCER2* (CD23), *CD58*, and *SEMA7A* (CD108) in scRNA-seq B cell clusters.

B) Flow cytometry gating strategy for analysis of CD23, CD58 and CD108 surface marker expression.

C) Representative flow cytometry plots for CD23, CD58 and CD108 surface marker comparisons in naïve, intermediate and GC gates (see B).

D) Relative frequency of $CD23^+CD108^+$, $CD23^+CD58^+$ and $CD58^+$ $CD108^+$ B cells in naïve, intermediate and GC populations (n=5). Lines connect values from same donor. p values denote result of paired T tests.

E) Relative frequency of $CD23^+CD108^+CD58^+$ in naïve, intermediate and GC populations (n=5). p values denote result of paired T tests.

F) Frequencies of $CD23^+CD108^+CD58^+$ B cells in the naïve and intermediate gates relative to total live CD19+ B cells, compared with frequencies of preGC B cells relative to total B cells in matched scRNA-seq libraries (n=5).

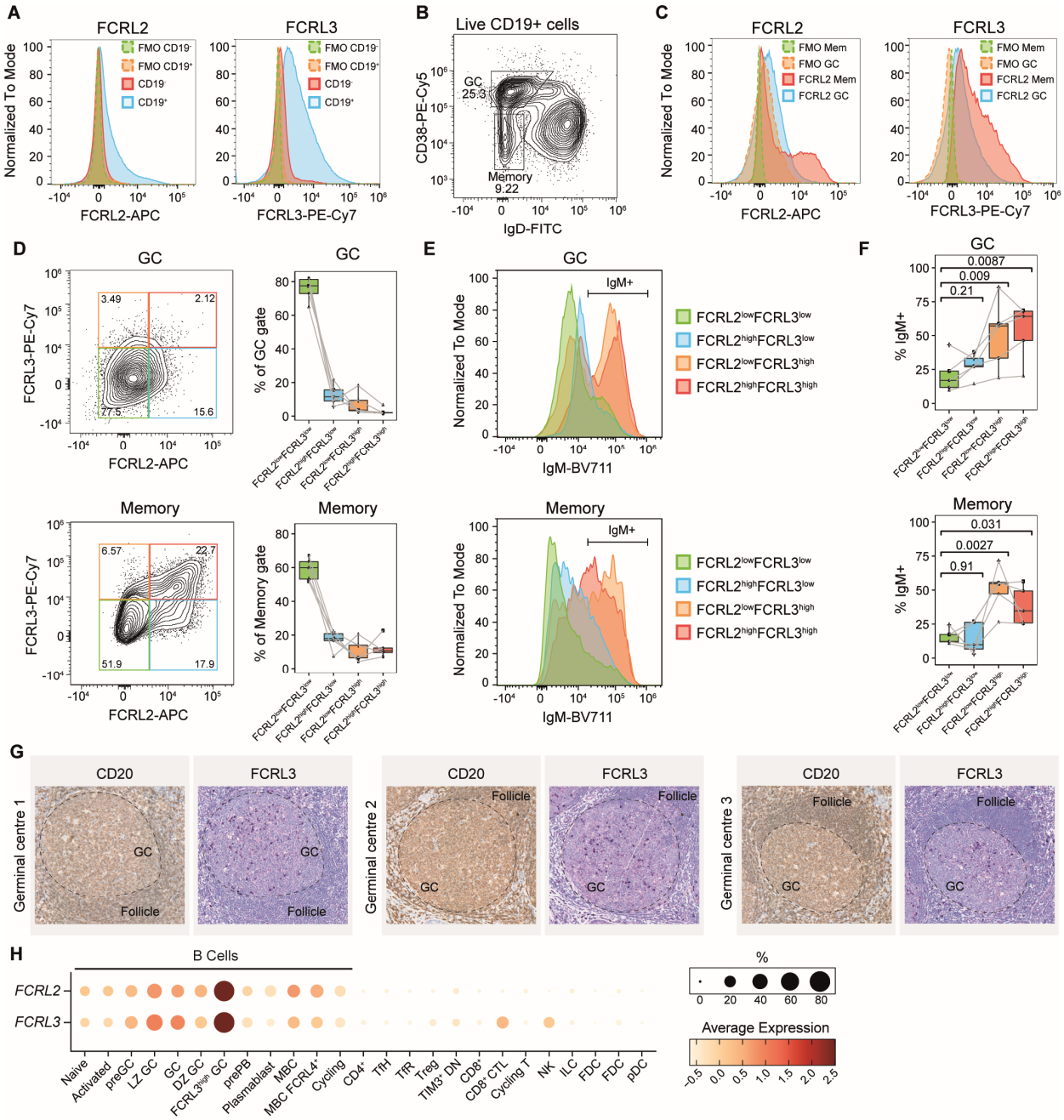


Figure S8. Flow cytometry analysis of FCRL3⁺ B cells.

A) FCRL2 and FCRL3 flow cytometry staining, with fluorescence minus one (FMO) controls, for live tonsillar B cells ($CD19^+$) and non-B cells ($CD19^-$).

B) Flow cytometry gating strategy for analysis of FCRL2 and FCRL3 surface marker expression in GC and memory B cells.

C) FCRL2 and FCRL3 flow cytometry staining, with FMO controls, for GC and memory B cells.

D) Representative flow cytometry plots for FCRL2 and FCRL3 surface marker expression in GC (upper) and memory B cells (lower). Relative frequencies of FCRL2 and FCRL3 populations within parent GC and memory subsets are shown (n=5).

E) Representative histogram of IgM expression by FCRL2 and FCRL3 populations in GC (upper) and memory B cells (lower).

F) Relative frequencies of IgM+ populations within different FCRL2/FCRL3 populations are shown (n=5). Lines connect values from same donor. p values denote result of paired T tests.

G) Immunohistochemistry of CD20 and FCRL3 in paediatric human tonsils.

H) Mean expression of FCRL2 and FCRL3 expression across all tonsillar immune cell subsets by scRNA-seq.

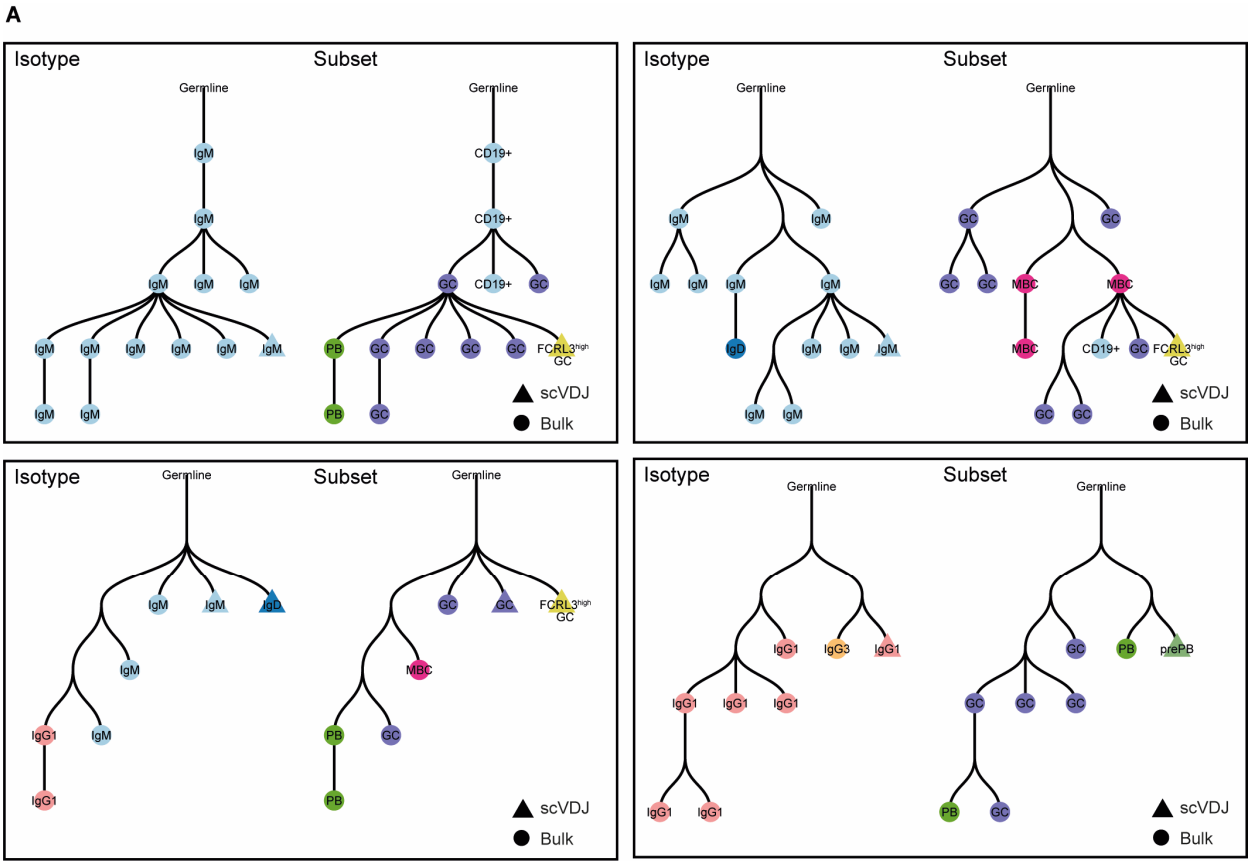


Figure S9. Example lineage trees of expanded FCRL3^{high} GC B cell clones.

A) Reconstructed lineage tree phylogenies of expanded B cell clones containing sequences from bulk B cell repertoires (circles) or scVDJ (triangles). Examples from three expanded FCRL3^{high} GC B cells and a prePB lineage are shown. Antibody isotype and sorted B cell subset (bulk) or annotated scRNA-seq cluster (scVDJ) are annotated for each tree.

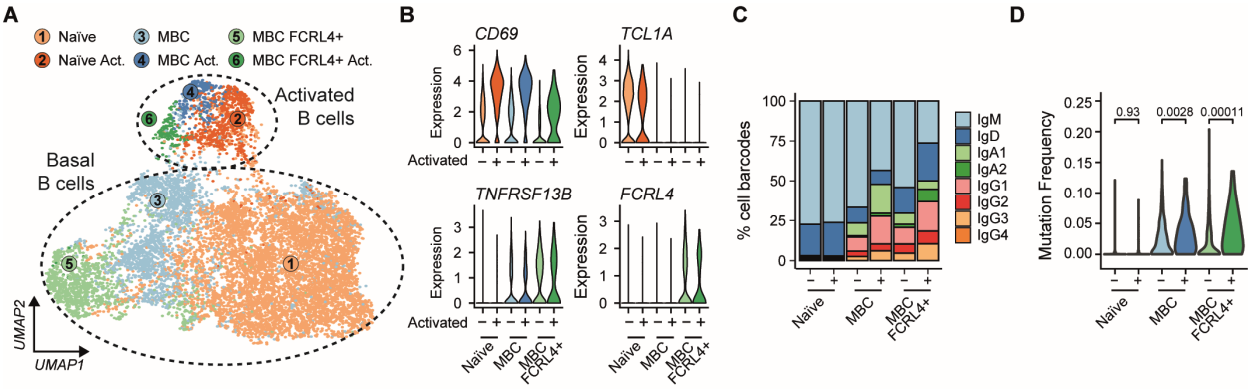


Figure S10. Sub-clustering of naïve, memory and activated B cell scRNA-seq clusters.

A) UMAP visualisation of naïve, memory and activated B cell subsets from Figure 2 that were subjected to reclustering (10,772 cells). The Activated B cell cluster from Figure 2 is comprised of naïve, MBC and MBC FCRL4+ populations.

B) Expression of key markers for naïve, activated and MBC populations confirming identity of subclusters.

C) Relative frequency of scVDJ antibody isotype within subclusters, comparing basal and activated states of different B cell populations.

D) Somatic hypermutation frequencies from scVDJ data of basal and activated states of different B cell populations. *p* values denote results from Wilcoxon Ranked Signed Sum test.

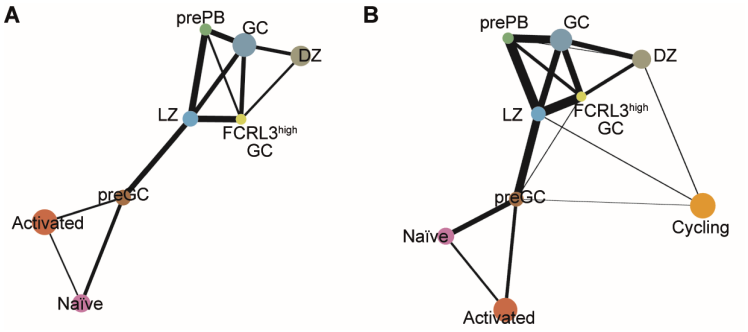


Figure S11. PAGA-based trajectory analysis of B cell scRNA-seq clusters.

A) Connectivity of naïve, activated, preGC and GC B cell scRNA-seq clusters using partition-based graph abstraction (PAGA) analysis. Line width denotes strength of connectivity.

B) Same as in A), but with cycling B cells also included.

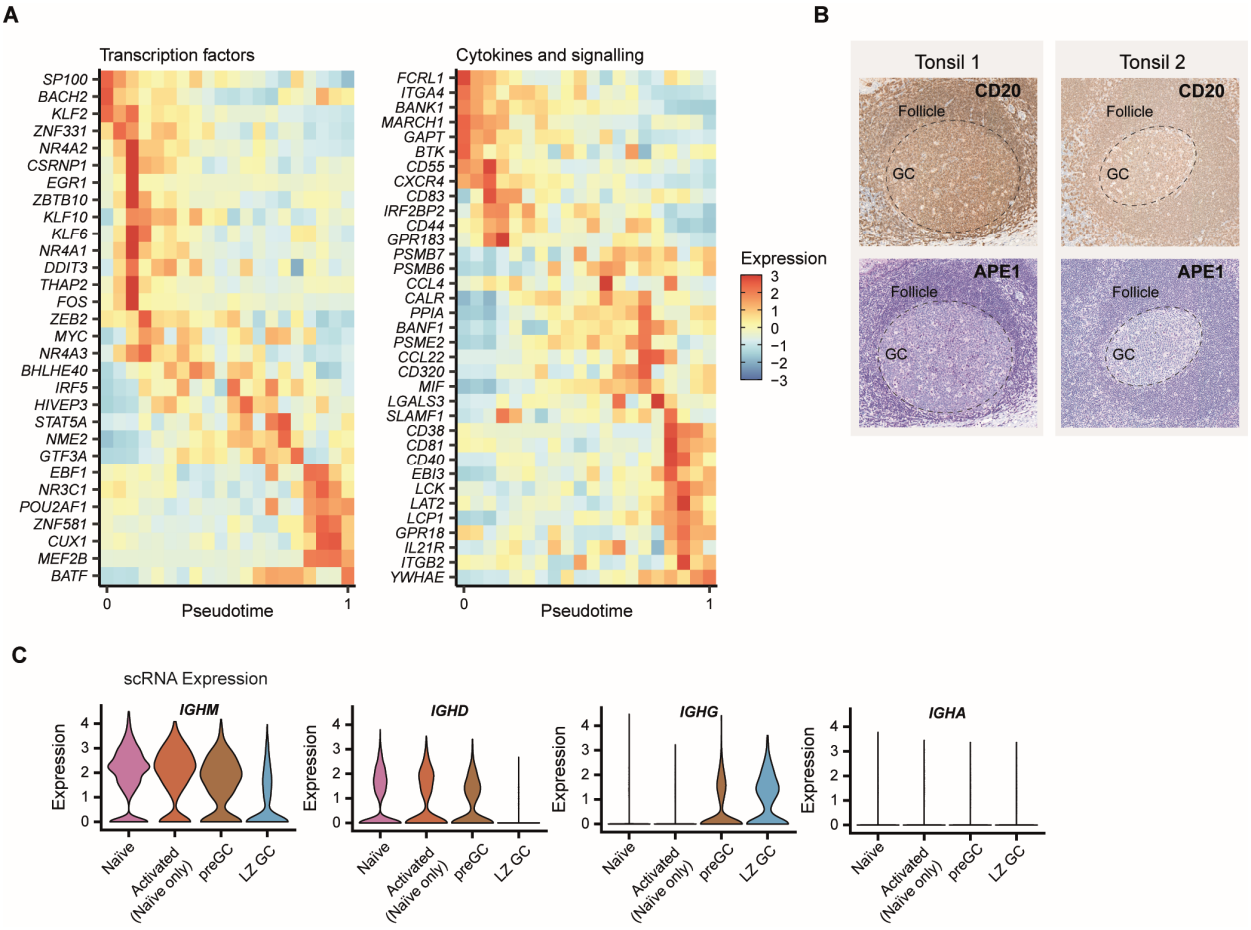


Figure S12. Dynamic gene expression during B cell activation and GC entry/formation.

A) Gene expression for selected genes encoding transcription factors or cytokines and signalling molecules that are significantly differentially expressed through velocity-based pseudotime of B cell activation and GC entry.

B) Immunohistochemistry of CD20 (B cell marker) and APE1 (*APEX1*) in two paediatric human tonsils reveals depleted expression of APE1 in germinal centres (GCs) compared to the follicular zone.

C) scRNA-based quantitation of *IGHM*, *IGHD*, *IGHG* and *IGHA* expression. *IGHG* and *IGHA* values are the sum of subclass counts. Activated cluster contains only naïve activated cells (see FigS10 for more information).

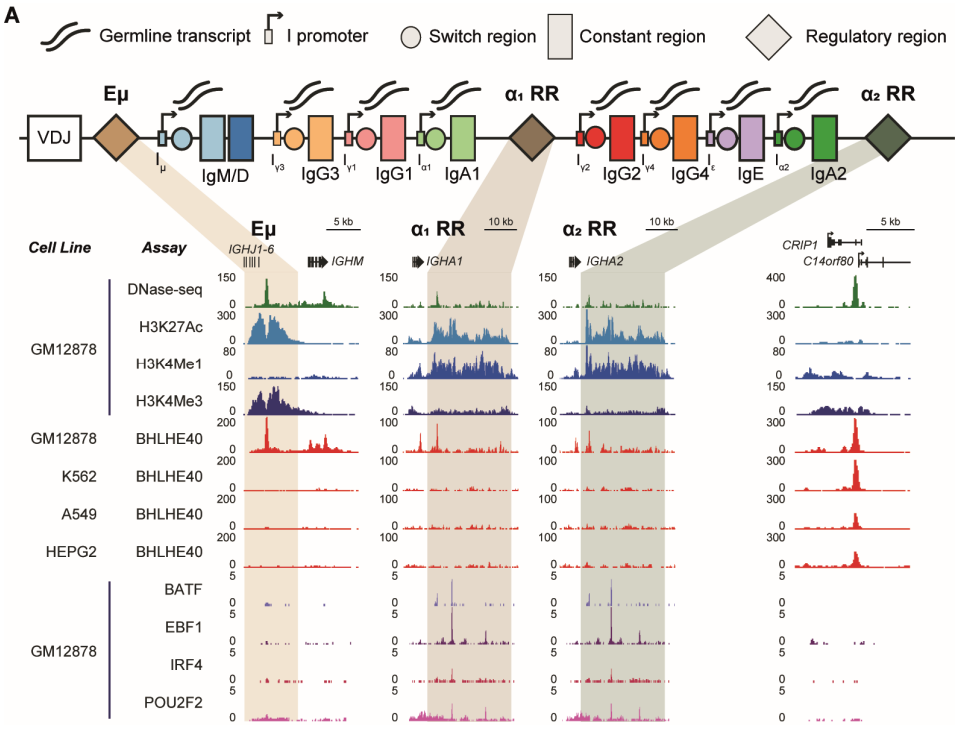


Figure S13. Transcription factor binding at the immunoglobulin heavy chain locus.

A) Schematic of the human immunoglobulin heavy chain (IgH) locus, with intergenic (I) promoters, switch regions, germline transcripts and regulatory regions (Eμ, α_1 RR, α_2 RR). Open chromatin (DNase-seq) and ChIP-seq from ENCODE at Eμ, α_1 RR, α_2 RR and a control neighbouring locus (*CRIP1 / C14orf80*) for EBV-transformed B lymphocyte cell line GM12878 and control non-B lymphocyte cell lines (K562, A549, HEPG2) are shown.

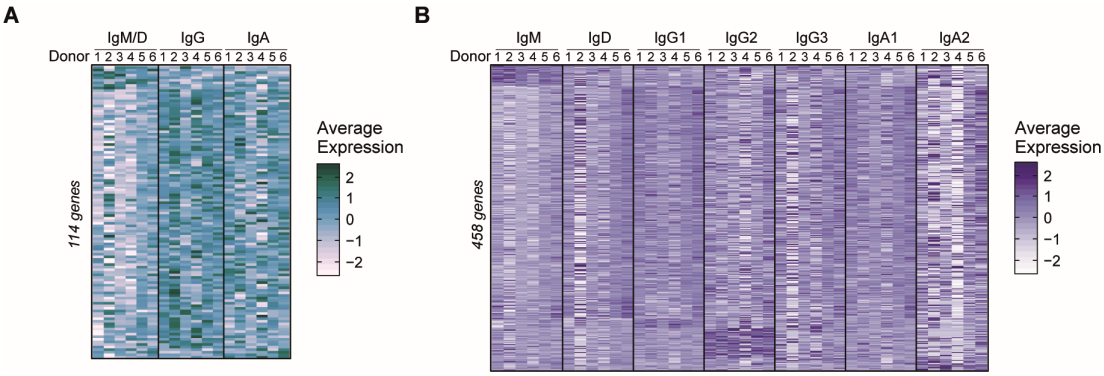


Figure S14. Class- and subclass-specific gene expression analyses of high SHM GC B cells.

A) Pseudobulk heatmaps of average expression per donor of differentially expressed genes between class-specific GC B cells with similar affinity (based on SHM frequency).

B) Same as in A), but for subclass-specific gene expression analyses.

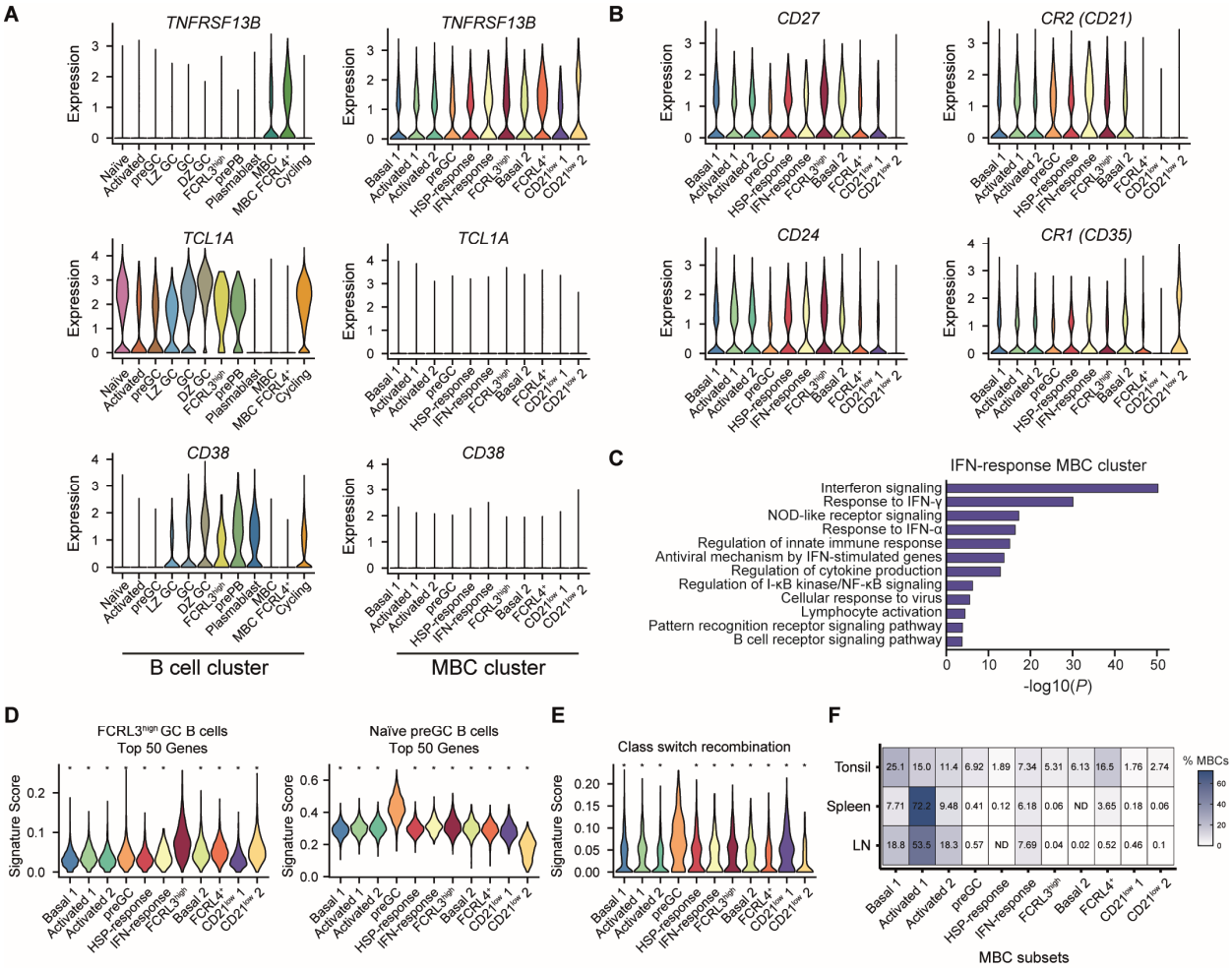


Figure S15. Characterisation of MBC states identified by scRNA-seq.

A) Single-cell expression of memory B cell (*TNFRSF1B*), naïve/undifferentiated (*TCL1A*) and germinal centre (*CD38*) markers across all B cell subsets (left) and sorted memory B cell subsets (right).

B) Single-cell expression of key marker genes differentially expressed by CD21^{low} MBC populations.

C) Top gene ontologies for significantly enriched genes in the IFN-response MBC cluster.

D) Single-cell AUCell-derived scores for top 50 marker genes of the naïve preGC B cells and FCRL3^{high} GC B cells in MBC subsets. * denotes *p* value < 0.001 from Wilcoxon Ranked Signed Sum test.

E) Single-cell AUCell-derived scores for CSR gene set in MBC subsets. * denotes *p* value < 0.001 from Wilcoxon Ranked Signed Sum test.

F) Relative frequencies of tonsillar MBC scRNA-seq clusters in lymph node and spleen MBCs.

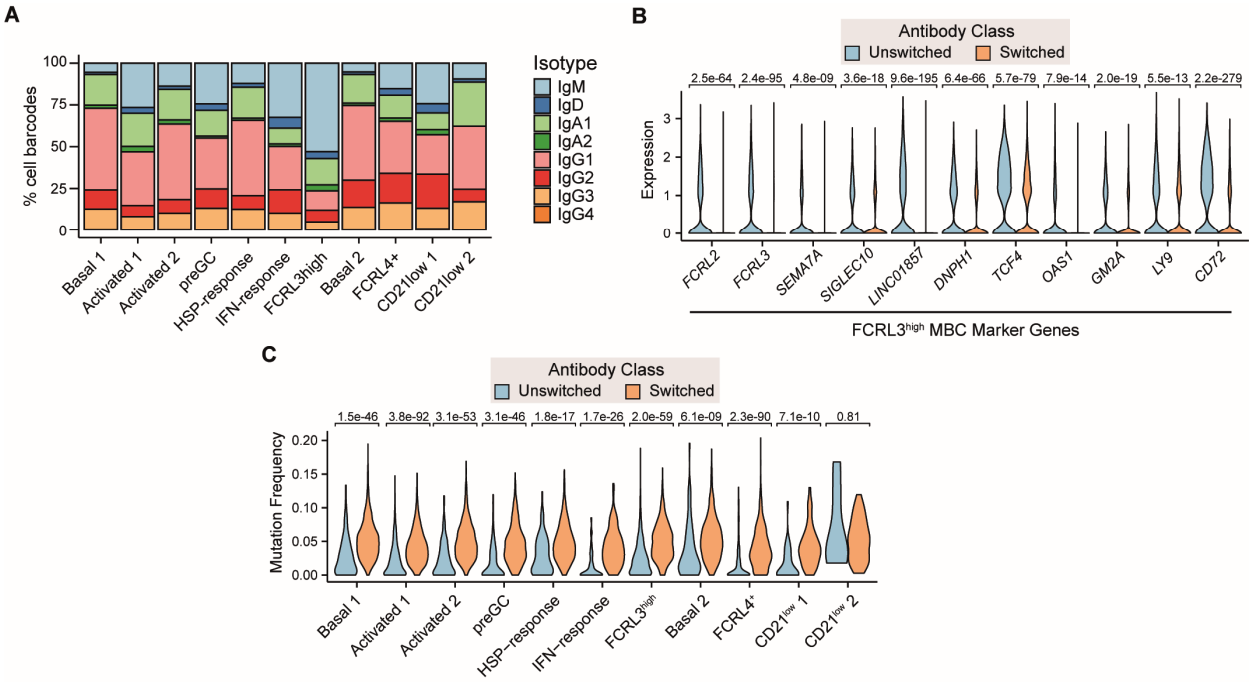


Figure S16. Antibody and gene expression features of switched and unswitched MBCs.

A) Relative frequencies of scVDJ-derived antibody subclass expression within different MBC scRNA-seq populations.

B) Single-cell expression of key marker genes of the FCRL3^{high} B cell states between switched and unswitched MBCs. *p* values denote results from Wilcoxon Ranked Signed Sum test.

C) SHM frequencies of scVDJ-derived antibody genes between switched and unswitched B cells in different MBC subsets. *p* values denote results from Wilcoxon Ranked Signed Sum test.