Antibody repertoire and gene expression dynamics of diverse human B cell states during affinity maturation.

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
§ Lead contact
* To whom correspondence can be addressed: h.king@qmul.ac.uk, drhamishking@gmail.com and louisa.james@qmul.ac.uk

Abstract

In response to antigen challenge, B cells clonally expand, undergo selection and differentiate to produce mature B cell subsets and high affinity antibodies. However, the interplay between dynamic B cell states and their antibody-based selection is challenging to decipher in primary human tissue. We have applied an integrated analysis of bulk and single-cell antibody repertoires paired with single-cell transcriptomics of human B cells undergoing affinity maturation. We define unique gene expression and antibody repertoires of known and novel B cell states, including a pre-germinal centre state primed to undergo class switch recombination. We dissect antibody class-dependent gene expression of germinal centre and memory B cells to find that class switching prior to germinal centre entry dictates the capacity of B cells to undergo antibody-based selection and differentiate. Together, our analyses provide unprecedented resolution into the gene expression and selection dynamics that shape B cell-mediated immunity.

Introduction

Effective immunity depends on the ability of B cells to evolve a functional antibody repertoire in response to antigen challenge. Following 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) within secondary lymphoid tissues, such as the spleen, peripheral lymph nodes and tonsils (Cyster and Allen, 2019). These GCs are transient structures in which B cells undergo iterative cycles of clonal expansion and somatic hypermutation in the variable regions of their immunoglobulin heavy and light chain genes followed by affinity-based selection for clones with high antigen-specificity. This highly dynamic process occurs in spatially and transcriptionally 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 (Mesin et al., 2016). B cells can differentiate and exit the GC reaction either as antibody-secreting plasmablasts committed to the plasma cell lineage or memory B cells, which are long-lived quiescent cells capable of being reactivated upon secondary exposure to the antigen (Suan et al., 2017b). The effector functions of antibodies expressed by B cells are broadly determined by antibody class (IgM, IgD, IgG, IgA, IgE) and more precisely by isotype or subclass (IgG1-4, IgA1-2), specified by the constant domain genes in the immunoglobulin heavy chain (IgH) locus. Although all naïve B cells express IgM and IgD, during maturation they may undergo class switch recombination, which involves the deletional recombination of IgM and IgD constant domain genes and expression of a different downstream constant domain gene (IgG1-4, IgA1-2 or IgE) (Stavnezer and Schrader, 2014). The combined outcomes of B cell differentiation, antigen affinity maturation and class switch recombination ultimately shape the antibody repertoire and the B cell-mediated immune response more broadly.

During their maturation in germinal centres, B cells express their antibody immunoglobulin genes as part of a membrane-bound complex termed the B cell receptor (BCR). Antigen-binding and downstream signalling of the BCR is a primary determinant of GC B cell survival and even differentiation through differential expression of key transcription factors regulated by BCR-mediated signalling (Kwak et al., 2019, Shlomchik et al., 2019). Studies have shown that BCR activation thresholds and downstream signalling can differ as a result of isotype-specific differences in the extracellular, transmembrane and intracellular domains of immunoglobulin proteins forming the BCR (Martin and Goodnow, 2002, Engels et al., 2014, Xu et al., 2014b). Thus, as well as shaping the effector functions of the subsequent antibody repertoire, class switch recombination contributes towards B cell survival or fate specification within the GC reaction. However, resolving the combined contribution of somatic hypermutation, maturation and class switching in the polyclonal context of primary human lymphoid tissues remains an enormous challenge for the field (Mesin et al., 2016). Furthermore, while it has long been held that class switch recombination occurs exclusively within the GC, as this is where the highest AICDA expression is detected, other work has demonstrated that class switch recombination often occurs prior to formation of the GC response (Roco et al., 2019, Toellner et al., 1996, Pape et al., 2003). This has raised questions about our current understanding of the cellular states and dynamics during human B cell maturation *in vivo* and demands a systematic and unbiased approach to better define how the human antibody repertoire is shaped through somatic hypermutation, class switch recombination and differentiation into different B cell fates.

We have applied an integrated strategy of bulk and single-cell antibody repertoire analysis paired with single-cell transcriptomics of human B cells from a model secondary lymphoid tissue, tonsils. We compare and contrast the antibody repertoires of major B cell subsets to reveal unique class switch hierarchies of memory B cells and plasmablasts. We then discover and define novel transcriptional B cell states during the GC response using single-cell RNA-seq. In particular, we reveal the unique gene expression of a B cell state primed to undergo class switch recombination before entering the GC. By leveraging the single-cell resolution of our datasets, we deconvolve the contribution of somatic

hypermutation and antibody class to gene expression patterns linked with altered BCR signalling, B cell maturation and fate decisions within the GC. Finally, we define diverse memory B cell states within secondary lymphoid tissue and explore the impact of class switch recombination on their functional potential. Our analyses reveal a striking importance for class switch recombination in shaping B cell fate and maturation in the GC and memory B cell fate that reframes our understanding of antibody-based selection and B cell differentiation.

Results

Subset- and subclass-specific antibody class switch recombination landscapes.

To begin to untangle the antibody-based selection of different B cell fates, we characterised the antibody repertoires of four broadly defined B cell subsets from the human tonsil; naïve, germinal centre (GC) B cells, memory B cells (MBCs) and plasmablasts, in addition to total CD19+ B cells (Figure 1A). We applied a subclass-specific and quantitative unique molecular identifier-based repertoire sequencing protocol for IgH VDJ sequences (Horns et al., 2016). Naïve B cells were comprised of more than 95% unswitched and unmutated IgM and IgD sequences, while GC and MBC samples consisted of both switched and unswitched IgH sequences with elevated somatic hypermutation (Figure 1B-C). Plasmablasts were nearly all switched and highly mutated. Consistent with the low abundance of IgE^+ B cells in human tonsils, we detected only a single IgE sequence, preventing us from drawing any conclusions about IgE-expressing B cells.

Higher IgH somatic hypermutation frequencies within the GC are typically a reflection of higher affinity BCRs and are proposed to bias GC B cells towards the plasmablast cell fate rather than the MBC fate (Shinnakasu et al., 2016, Suan et al., 2017a). In keeping with this, plasmablast-derived antibody repertoires generally contained higher somatic hypermutation frequencies than those of MBCs at a bulk level (Figure 1C). However, by resolving for antibody subclass, we found that while MBC-derived IgM and IgD sequences were consistently less mutated than plasmablast-derived IgM and IgD, the somatic hypermutation levels for switched isotypes were broadly similar between different B cell subsets (Figure 1D). 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 (Figure 1E). This is not likely explained by differences in their somatic hypermutation frequencies (Figure 1D), but instead may reflect differential selection or differentiation processes linked with specific class switch recombination outcomes.

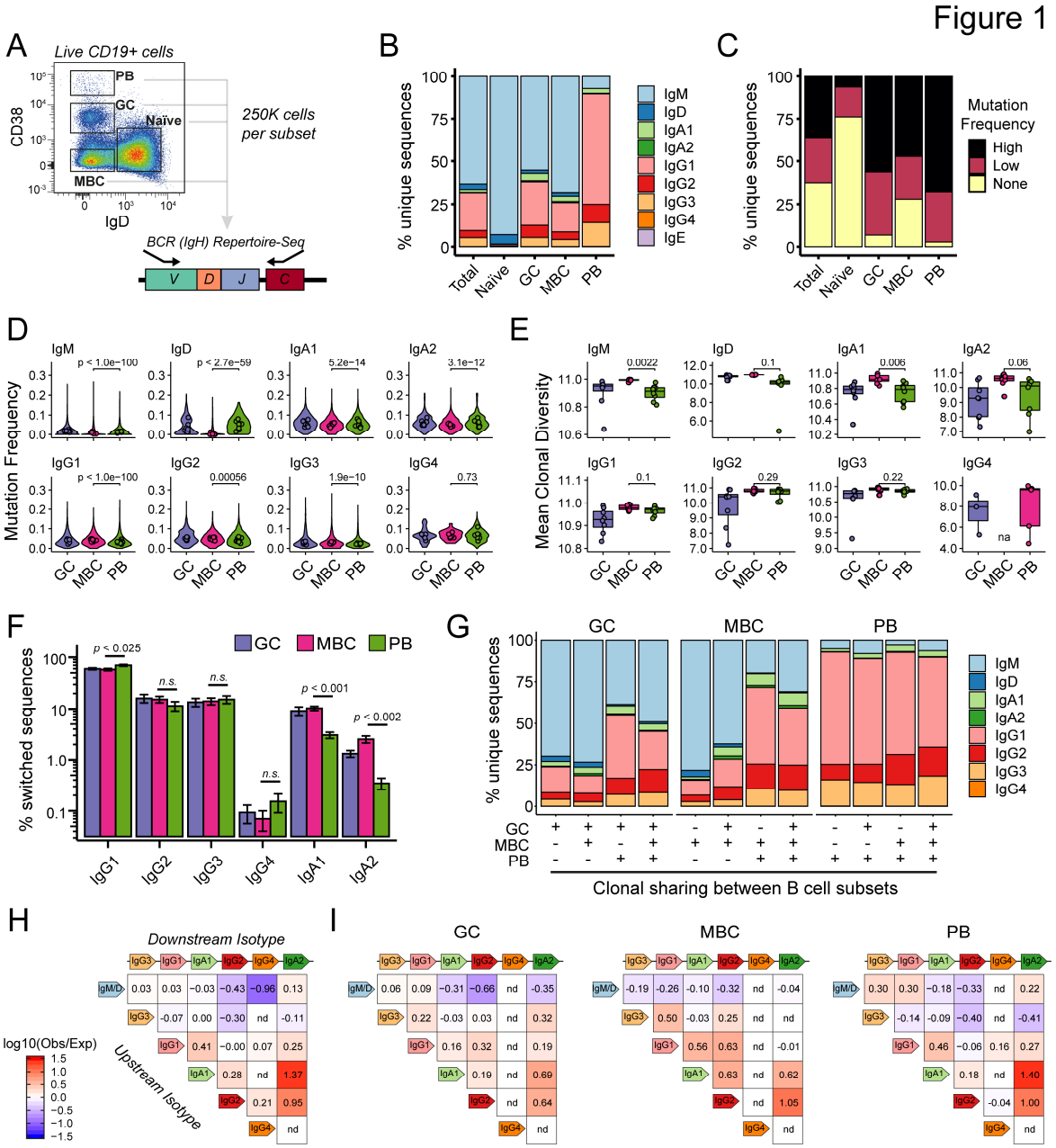

Figure 1. Subclass- and subset-specific features of human tonsillar B cell repertoires.

A) Cell sorting strategy to isolate naïve, germinal centre (GC), memory B cells (MBC) and plasmablasts (PB) from live CD19+ human tonsillar B cells for antibody repertoire analysis. Representative of *n*=8.

B) Mean antibody subclass frequencies within each B cell subset across donors.

C) Mean frequencies of antibody somatic hypermutation levels (None, Low, High) within each B cell subset across donors.

D) Somatic hypermutation frequencies for subclass-specific antibody sequences within each B cell subset. Violin plots show all unique sequences per subset, while points represent mean for each donor.

E) Mean clonal diversity scores per donor of subclass- and subset-specific B cell clones.

F) Mean switched antibody subclass frequencies within each B cell subset across donors.

G) Mean 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 top).

H) Observed/expected frequencies for isotype pairs detected in reconstructed phylogenies of clonally-related sequences. Antibody subclasses are ordered according to the IgH locus. nd denotes not detected.

I) Same as in H), except phylogenies are restricted to subset-specific sequences.

123 We next examined whether there were differences in antibody subclass frequencies or the manner in
124 which subclasses might have arisen in our broadly defined B cell populations. Intriguingly, we observed
125 that as well as an increased propensity to retain IgM expression (Figure 1B), MBCs were 3.3- or 7.3-
126 fold more likely than plasmablasts to express IgA1 or IgA2 respectively, while plasmablasts were
127 significantly more likely to express IgG1 (Figure 1F). These enrichments were linked with specific B cell
128 fates even for expanded clones spanning different B cell populations (Figure 1G). Finally, to explore
129 how these unique class switch patterns might have arisen, we reconstructed phylogenies for 28,845
130 expanded B cell clones and calculated the likelihood that specific class switch recombination events
131 were observed compared to that expected by chance (Figure 1H-I). Clonal lineages within the MBC
132 pool exhibited greater likelihoods for switching of isotype pairs located close to each other in linear
133 space along the IgH locus, compared to plasmablast clones which demonstrated a more eclectic pattern
134 of class-switch likelihoods (Figure 1I). Of note, both the antibody subclass frequencies and
135 reconstructed class switch hierarchies of MBCs closely resembled those of GC cells, consistent with
136 models that propose a stochastic exit of MBCs from the GC (Duffy et al., 2012, Good-Jacobson and
137 Shlomchik, 2010). Together, these analyses reveal that the antibody-based selection mechanisms of
138 two major mature B cell subsets (MBCs and plasmablasts) exhibit important differences related
139 primarily to their propensity to have undergone class switch recombination earlier in their maturation.

140

141 Single-cell atlas of tonsillar immune cells defines B cell states during affinity maturation.

142 To better understand the antibody-based selection mechanisms shaping the maturation of different B
143 cell subsets, we performed single-cell RNA-seq (scRNA-seq) paired with single-cell B cell repertoire
144 VDJ sequencing (scVDJ-seq) for unsorted tonsillar immune cells from the same samples used for our
145 bulk B cell repertoire analyses (Figure 2A-C). After stringent quality control, we retained the
146 transcriptomes of 32,607 immune cells (n=7; median of 3142 and mean of 4658 cells per donor) from
147 which we identified 30 transcriptionally distinct cell types (Figure 2D-F). Although our primary focus was
148 understanding B cell maturation, given the importance of other immune cell populations in this process
149 we also annotated varied T cell and non-lymphoid populations, including naïve and/or central memory
150 (NCM) $CD4^+$ and $CD8^+$ T cells, follicular helper (TfH) T cells, follicular regulatory T (Tfr) cells, regulatory
151 T cells (Treg), cytotoxic $CD8^+$ T cells, innate lymphoid cells (ILCs), natural killer (NK) cells, follicular
152 dendritic cell (FDC), plasmacytoid-derived dendritic cell (pDCs), classical dendritic cell (cDC1), and
153 several macrophage (MAC) clusters (Figures 2D-E and S1) that will act as a valuable resource for those
154 studying immune cell dynamics in human secondary lymphoid tissues.

155 We next characterised tonsillar B cell states using both their unique gene expression and scVDJ-
156 derived antibody repertoire features. To improve the power and accuracy of our scVDJ analyses we
157 applied a novel strategy of integrating single-cell and bulk repertoires to benefit from the deeper
158 sampling of IgH sequences from bulk repertoires to enhance genotype correction, identification of

159 clonally-related sequences and quantitation of somatic hypermutation levels for scVDJ-derived
160 sequences (Figure 2G). We defined gene expression signatures for 12 distinct B cell types or states
161 (Figure 2H) and complemented marker gene-based annotation with antibody isotype frequencies
162 (Figure 2I), somatic hypermutation levels (Figure 2J), clonal diversity (Figure 2K-L) and relationships
163 with other B cell subsets (Figure 2M-N). We identified all major stages of B cell maturation, including
164 naïve, activated, GC (including both LZ and non-proliferating DZ cells), MBCs, tissue-resident FCRL4+
165 MBCs, plasmablasts and a cycling population consisting mostly of DZ GC cells (Figure 2H). Whilst
166 apoptotic cells normally comprise a sizeable proportion of GC B cells these were not retained in our
167 analysis as they failed to generate sufficiently high quality transcriptomic data. Our analysis of B cell
168 states also identified a "preGC" B cell state expressing unmutated IgM and IgD that transcriptionally
169 shared markers with both naïve and LZ GC populations, but had yet to acquire features consistent with
170 B cell maturation in the GC such as *CD27* and *CD38* expression, hypermutated antibody genes or
171 clonal expansion (see Figure 3 for further details). We also annotated a population of class switched
172 and hypermutated GC B cells that amongst other transcriptionally unique features (*FRZB*, *BTNL9*,
173 *FGFR1*) express low to intermediate levels of the plasmablast-specific transcription factors *PRDM1* and
174 *XBP1* (Figure 2H), suggesting that these cells may be a pre-plasmablast (prePB) state within the GC.
175 Finally, we discover a transcriptionally distinct and clonally-expanded IgM^{+} B cell population in the GC
176 with elevated expression of genes associated with inhibitory BCR signalling, such as *FCRL2*, *FCRL3*,
177 *SAMSN1*, and *SIGLEC10* that we have labelled as $FCRL2/3^{high}$ GC B cells. Some of these cells were
178 part of large expanded GC-derived clones that also contained MBCs or plasmablasts (Figure 2M-N),
179 indicating that this cell state arises as part of productive GC reactions and is unlikely to be derived from
180 a separate B cell lineage.

181 Crucially, all annotated B cell states were observed at reproducible frequencies across patients,
182 regardless of their history of tonsillitis (Figure 2O), although exactly how some of these transcriptional
183 states relate to other human B cell populations will require further work. We also note that we have
184 identified a greater diversity of B cell fates in our paediatric tonsil samples (typically <10 years old) than
185 previous single-cell studies of other human lymphoid tissues from adult donors (typically >40 years old)
186 (Madissoon et al., 2019, James et al., 2020), highlighting the importance of profiling immunologically
187 active tissues to understand B cell maturation. Together, our uniquely comprehensive transcriptomic
188 and repertoire analyses of a model secondary lymphoid organ have allowed us to create a detailed
189 overview of human B cell maturation that will allow us to interrogate gene expression and antibody
190 repertoire dynamics before, during and after maturation in the GC.

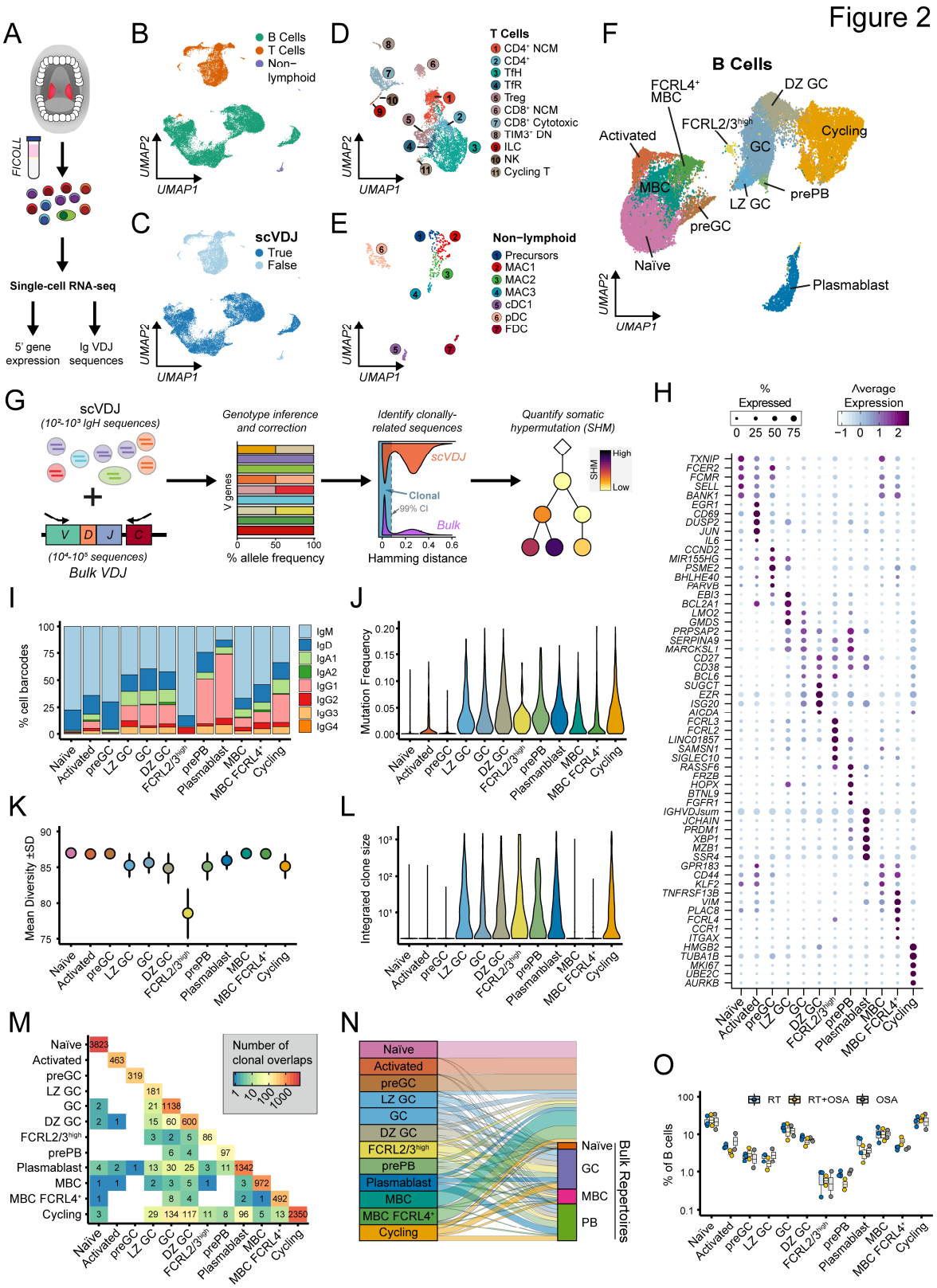

Figure 2. A single-cell atlas of human tonsillar immune cells to understand B cell maturation.

A) Schematic of tonsillar immune cell isolation followed by single-cell profiling of gene expression and antibody sequences.

B) UMAP projection of tonsillar immune scRNA-seq data (32,607 cells; 7 donors) annotated as B, T or non-lymphoid cells.

C) Same as in B), with cells annotated for whether a high quality immunoglobulin scVDJ sequence was assembled.

….continued on following page…

D) UMAP projection of T cell populations in the tonsillar immune scRNA-seq data, 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, $TIM3^+$ CD4/CD8 double-negative ($TIM3^+$ DN) and cycling T cells, in addition to innate lymphoid cells (ILC) and natural killer (NK) cells.
E) UMAP projection of non-lymphoid cell populations in the tonsillar immune scRNA-seq data, including monocyte/macrophage precursors (Precursors), macrophages (MAC1, MAC2, MAC3), conventional dendritic cell 1 (cDC1), plasmacytoid-derived dendritic cell (pDC) and follicular dendritic cell (FDC) subsets.
F) UMAP projection of B cell populations in the tonsillar immune scRNA-seq data, including naïve, activated, pre-germinal centre (preGC), light zone GC (LZ GC), GC, dark zone (DZ GC), cycling, $FCRL2/3^{high}$ GC, pre-plasmablasts (prePB), plasmablasts, MBC and $FCRL4^+$ MBC subsets.
G) Schematic of scVDJ and bulk repertoire integration, including genotype inference and correction, identification of clonally-related sequences and quantitation of somatic hypermutation.
H) Mean expression of key marker genes used to define B cell scRNA-seq clusters. Frequency of cells for which each gene is detected is denoted by size of the dots.
I) Relative scVDJ-derived antibody subclass frequencies in different B cell states.
J) Somatic hypermutation frequencies of scVDJ-derived antibody genes in different B cell states.
K) Clonal diversity scores of B cell clones identified in scRNA-seq dataset. Error bars denote ±SD.
L) Number of members per clonotype in different B cell states from integrated repertoire analysis.
M) Co-occurrence of expanded clones across B cell states. Numbers reflect a binary detection event.
N) Clonal relationships between scRNA-seq-defined B cell states and repertoires of sorted B cell subsets.
O) Relative frequencies of different B cell states separated by clinical indication for tonsillectomy. OSA; obstructive sleep apnoea (n=2), RT; recurrent tonsillitis (n=3), RT+OSA (n=2).

Dynamic gene expression during human B cell activation and GC formation.

Initially in a B cell response, B cells acquire antigen, either in soluble form or displayed on the surface of follicular antigen-presenting cells (APCs), which results in their activation and migration from the follicle to the T cell zone where they can form or participate in GC reactions. However, reconstructing the events during this important process in primary human tissues is extremely challenging. We therefore used our single-cell profiling of human B cell maturation to explore the dynamics of early B cell activation and GC formation. We first identified an activated B cell state with elevated expression of known activation marker genes (Activated; *CD69, CD83*, *JUN*) (Figures 2 and 3A) and which demonstrate the highest frequency of predicted cell-cell communication with APCs (FDCs, other dendritic cells, macrophages) (Figure 3B). This activated B cell state appeared capable of coordinating help from both APCs and T cells through IL6 signalling (Arkatkar et al., 2017, Ise et al., 2018) and/or ICAM1-ITGAL1 (LFA-1) interactions (Zaretsky et al., 2017) (Figure 3C). Many of these same predicted cell-cell interactions were also detected in preGC B cells (Figure 3C), suggesting that these might reflect a transitional cell state between antigen dependent-activation and GC entry or formation. Indeed, we found that preGC B cells exhibit a strong directionality towards the light zone of the GC using RNA velocity inference (Figure 3D). We next reconstructed a pseudotemporal trajectory of naïve, activated, preGC and LZ GC B cells, revealing a continuum of gene expression from early activation events to *bona fide* GC B cells and allowing us to identify dynamic expression of key signalling molecules and transcription factors (Figures 3E-F and S2A). Crucially, these gene expression changes through pseudotime were well correlated with an experimentally-derived time course of *in vitro*-stimulated B cells (Figure 3G; Shinohara et al., 2014), strongly supporting our pseudotemporal ordering of defined B cell states during B cell activation and GC entry. This roadmap of B cell activation may allow future improvement to *in vitro* B cell culture protocols to better model B cell activation dynamics *in vivo*.

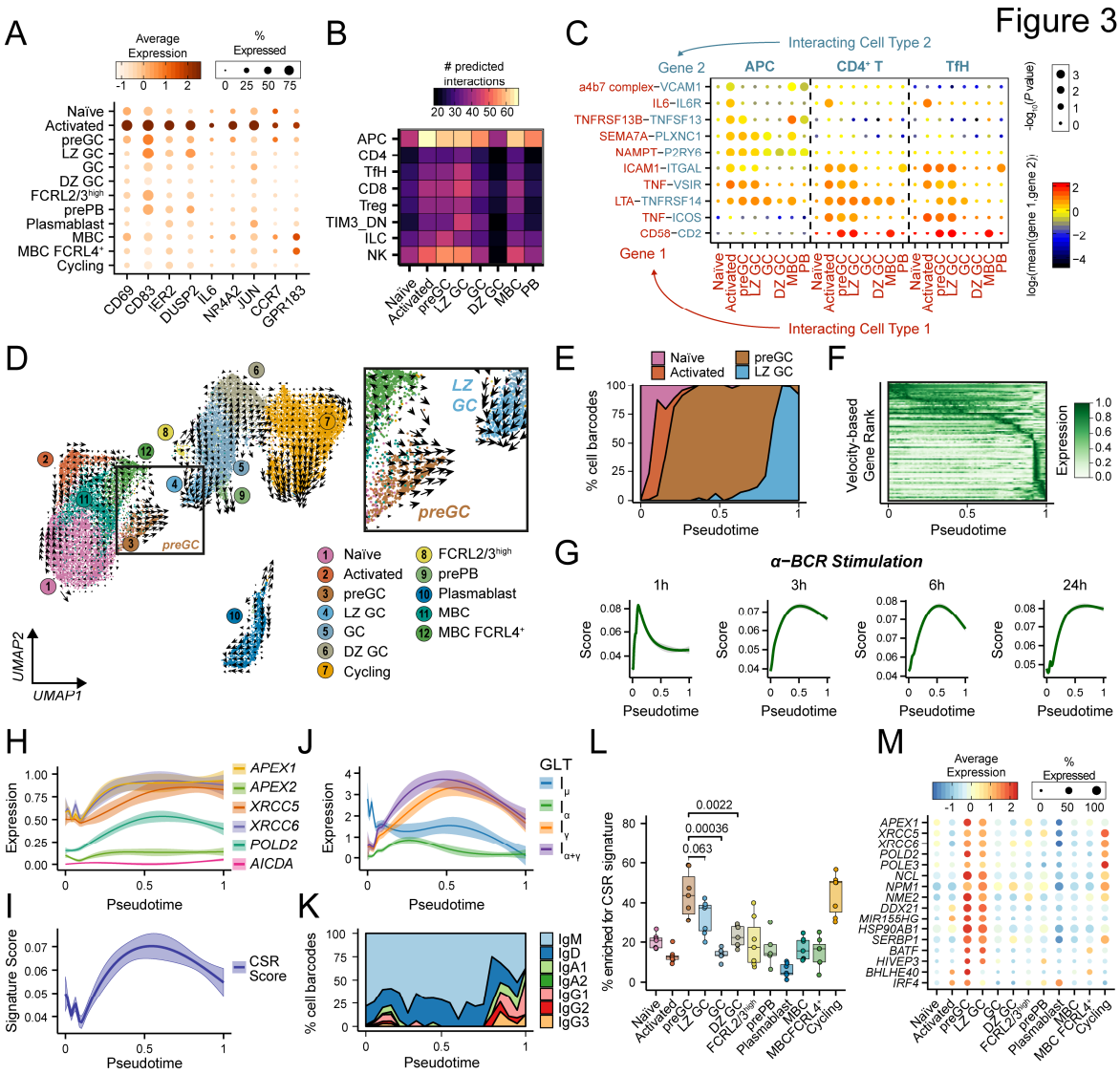

Figure 3. Reconstruction of *in vivo* B cell activation reveals class switch recombination in a preGC state.

A) Mean expression of key marker genes for activated B cells in scRNA-seq.
B) Frequency of significant predicted ligand-receptor pair interactions between major B cell states and different T cell populations and antigen-presenting cells (APCs) performed using CellPhoneDB.
C) Selected ligand–receptor interactions between B cell subsets and APCs, CD4+ T cells and TfH cells with CellPhoneDB.
D) Grid-based visualisation of tonsillar B cell RNA velocities. Arrow size conveys strength of predicted 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) Smoothed anti (α)-IgM-treatment gene signature scores (±95% CI) across velocity-based pseudotime.
H) Smoothed expression of class switch recombination genes differentially expressed through velocity-based pseudotime.
I) Smoothed class switch recombination gene signature score through velocity-based pseudotime.
J) Smoothed expression of I promoter germline transcripts for IgM (I_μ), IgA (I_α) and IgG (I_γ) through velocity-based pseudotime. $I_{\alpha+\gamma}$ denotes the sum of I_α and I_γ expression.
K) Relative antibody subclass frequencies across velocity-based pseudotime.
L) Relative frequencies of cells with high class switch recombination signature scores in different B cell states for each donor ($n = 7$). p values denote results of Student's T test.
M) Mean expression of genes implicated in class switch recombination for B cell subsets.

One surprising result from our reconstruction of human B cell activation and GC entry was the discovery that several genes associated with class switch recombination were most highly expressed prior to GC entry, with a specific enrichment of these genes in the preGC B cell population (Figure 3H-I). This

265 included *APEX1* expression, of which its translated product APE1 is required for class switch
266 recombination to occur in a dose-dependent manner (Masani et al., 2013, Xu et al., 2014a) and is
267 expressed almost exclusively by non-GC B cells (Figure S2B; Roco et al., 2019). We also found that
268 preGC B cells had the highest expression of IgH germline transcripts (GLTs) (Figure 3J), which
269 preceded switching from IgM/IgD to other isotypes (Figure 3K), in fitting with a recent study describing
270 GLT transcription prior to GC formation in mouse models (Roco et al., 2019). These findings contrast
271 with the prevailing view that class switch recombination occurs in the DZ of the GC, where AICDA
272 expression is highest. However, we detected little enrichment of our class switch recombination
273 signature in non-cycling DZ GC B cells (Figure 3L), and although there was an enrichment of class
274 switch recombination genes in cycling B cell populations, this likely reflects their involvement in cell
275 cycle-linked DNA recombination and repair. Our analysis also highlights many other genes previously
276 linked with class switch recombination (Figure 3M), including those capable of binding to switch region
277 sequences within the IgH locus (*NME2*, *NCL*, *DDX21*), interacting with the class switch recombination
278 machinery (*NPM1*, *SERBP1*) or regulating *AICDA*/AICDA transcript/protein stability (*mir155HG,*
279 *HSP90AB1*) (Borggrefe et al., 1998, Hanakahi et al., 1997, Mondal et al., 2016, Orthwein et al., 2010,
280 Shinozaki et al., 2006, McRae et al., 2017, Zheng et al., 2019). Notably, the microRNA gene *miR155HG*,
281 formerly called BIC (B-cell Integration Cluster), was up-regulated in preGC B cells and has been shown
282 to be essential for B cells to form GCs and undergo class switch recombination in mice (Thai et al.,
283 2007, Vigorito et al., 2007). Furthermore, the transcription factor genes *BATF*, *IRF4* and *BHLHE40* are
284 enriched in the preGC B cell state, of which BATF and IRF4 are known to regulate GC formation in a B
285 cell-intrinsic manner (Morman et al., 2018, Willis et al., 2014). Intriguingly, BHLHE40 is capable of
286 binding to the major regulatory regions α_1 RR and α_2 RR of the IgH locus (Figure S2C-D), implicating
287 this poorly understood transcription factor in the regulation of class switch recombination prior to GC
288 formation. Together, these analyses strongly support a model where class switch recombination occurs
289 primarily before formation of the GC response (Roco et al., 2019, Toellner et al., 1996, Pape et al.,
290 2003) and our detailed gene expression analyses define the cellular state involved.

291

292 Antibody-based selection of B cell fate in the germinal centre

293 Class switch recombination before entry into the GC has the potential to dramatically influence the
294 antibody-based selection of B cells within the subsequent GC reaction as a consequence of differential
295 signalling through the membrane-bound immunoglobulin BCR. We therefore turned to dissect the gene
296 expression dynamics linked with antibody-based selection and fate specification of B cells within the
297 GC reaction.

298 During affinity maturation and selection in the GC, B cells cycle between physically distinct light and
299 dark zones. While we clearly identified LZ and DZ B cell populations in our scRNA-seq dataset, we also
300 found that many GC B cells existed in a continuum between these two states (Figure 4A-B) similar to

301 previous reports (Milpied et al., 2018), with the exception of the FCRL2/3^{high} and prePB clusters which
302 existed as transcriptionally distinct states (Figures 2H and 4C). In addition to unique gene expression
303 patterns, these two sub-populations of GC B cells also exhibit unique patterns of class switching, with
304 prePB B cells in the GC more likely to express class-switched isotypes and FCRL2/3^{high} GC B cells
305 more likely to retain expression of IgM and IgD (Figures 2I and 4D). Furthermore, cells clonally related
306 to FCRL2/3^{high} GC B cells were almost exclusively IgM^{+} and had rarely undergone class-switching
307 (Figure 4E). These observations suggested that the outcome of class switch recombination may be
308 linked with specific gene expression programs of GC B cells.

309 To address this, we first needed to examine the contribution of antibody maturation state to GC B cell
310 gene expression programs, given lower average somatic hypermutation frequencies of IgM^{+} B cells
311 compared to B cells expressing switched isotypes (Figure 1D). We leveraged our paired single-cell VDJ
312 and transcriptomic datasets to stratify all non-cycling GC B cells (excluding prePB and FCRL2/3^{high}
313 populations) based on their IgH somatic hypermutation frequencies as a proxy for affinity (Figure 4F).
314 Strikingly, GC B cells with high or low somatic hypermutation were significantly enriched with gene sets
315 derived from experimentally determined high- or low-affinity antigen-binding B cells respectively
316 (Shinnakasu et al., 2016), higher expression of the B cell maturation marker CD27 and larger clone
317 sizes (Figure 4F), reflecting increased expansion and maturation based on BCR affinity. We found that
318 the direct comparison of GC B cells expressing different antibody classes was confounded by gene
319 expression linked with high and low somatic hypermutation frequencies (Figure 4G), consistent with
320 high and low affinity binding events differentially regulating GC B cells (Shinnakasu et al., 2016). To
321 overcome this, we examined GC B cells with matched somatic hypermutation levels expressing
322 different antibody classes (IgM/D, IgG or IgA) at single-cell resolution. This revealed differential
323 expression of genes involved directly in cell survival, BCR signalling, antigen presentation, immune
324 responses and metabolism, as well as other pathways more indirectly linked with BCR activity
325 (RhoA/RhoGDI, Integrin, NFAT, eIF2), between unswitched and switched GC B cells (Tybulewicz and
326 Henderson, 2009, Mielke et al., 2011, Arana et al., 2008, Scharenberg et al., 2007) (Figure 4H-J). These
327 differences may be linked with differential exposure to T cell-derived cytokines such as IL4, TGFB,
328 IFNG and CD40LG, or signalling through different toll-like receptors (TLR) (Figure 4I). Intriguingly,
329 several genes involved with GC confinement or regulating B cell niche homing were up-regulated in
330 IgG^{+} and IgA^{+} GC B cells, such as genes required for CXCL12-mediated migration to GCs (*LCP1* and
331 *MYO1E* (Dubovsky et al., 2013, Girón-Pérez et al., 2020)) and the GC confinement receptor *P2RY8*
332 (Muppidi et al., 2014).

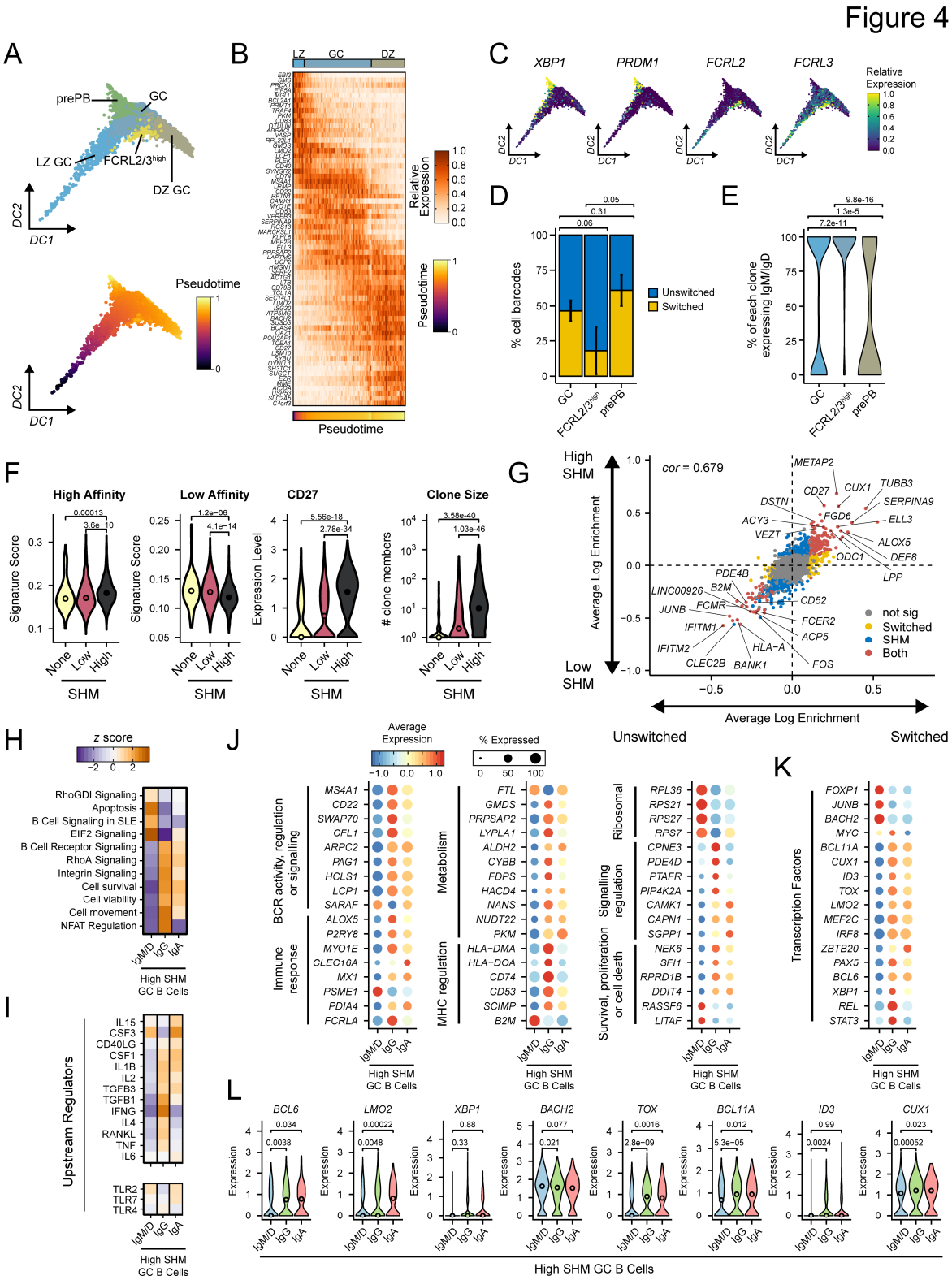

Figure 4. Influence of antibody class on B cell fate and function within the germinal centre.

A) Diffusion-based graph visualisation and pseudotemporal ordering of GC B cell scRNA-seq populations.

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

C) Expression of key marker genes for prePB and FCRL2/3^{high} GC B cells.

D) Relative frequency of GC B cells that have undergone class switch recombination (n = 6). "GC" is the combination of LZ GC, GC and DZ GC clusters in (A). Error bars denote SEM.

….continued on following page…

E) Percentage of members within each GC B cell clone that express IgM or IgD for clones that contain FCRL2/3^{high} or prePB cells. Clonal families include sequences from both scVDJ and bulk repertoires.
F) High and low affinity gene signature scores for GC B cells (LZ GC, GC and DZ GC clusters) grouped by antibody somatic hypermutation (SHM) frequency (n = 2045 cells). Also shown are *CD27* expression levels and integrated clone sizes.
G) Scatterplot comparing log enrichment of genes in class-switched vs unswitched (x axis) and high vs low affinity/SHM (y axis) GC B cells. Colour denotes statistical significance and c*or* denotes Pearson's Correlation coefficient.
H) Heatmap of significantly enriched gene ontologies for genes enriched in class-specific high affinity/SHM GC B cells.
I) Same as in (H), but for gene pathways downstream of cytokines (upper panel) and toll-like receptors (TLRs).
J) Mean expression of genes enriched in class-specific high affinity/SHM GC B cells grouped by predicted functions.
K) Mean expression of transcription factor genes enriched in class-specific high affinity/SHM GC B cells.
L) Single-cell gene expression of transcription factors in class-specific high affinity/SHM GC B cells.

Although most gene expression differences were comparable between IgG and IgA and we identified few significant or meaningful differences for subclass-specific B cells (Figure S3), one interesting example of class-specific gene expression was the enrichment of *CLEC16A* in IgA^{+} GC B cells given that this gene is associated with a selective IgA immunodeficiency (Ferreira et al., 2010). Finally, to try and understand the upstream regulation of these class-specific gene expression networks we examined the expression of transcription factors within class-specific GC B cells (Figure 4K-L). We found that IgM^{+} B cells express lower levels of transcription factors like *BCL6*, *XBP1* and *ID3* known to regulate the ability of B cells to remain in the GC or differentiate but higher levels of the transcription factor *BACH2* that represses plasma cell differentiation (Todd et al., 2009, Gloury et al., 2016, Huang et al., 2014, Shinnakasu et al., 2016). We also found differential expression of other transcription factors such as *LMO2*, *TOX*, *BCL11A* and *CUX1*, suggesting that they may play a role in the unique transcriptional wiring of switched and unswitched B cells within the GC. Our single-cell resolution of the GC B cell response has allowed us to uncouple antibody affinity and class and to dissect the differential contributions of these two critical arms of the B cell repertoire in shaping B cell fate and function in the GC. Importantly, these differential gene expression patterns suggest varying abilities of switched and unswitched B cells to survive and reside in the GC and establish that one of the dominant influences in shaping antibody-based selection in the GC is whether a B cell has undergone class switch recombination.

Diverse memory B cell states and activation dynamics in secondary lymphoid tissue.

Maturation state and antibody class appeared to both impact gene expression dynamics within the GC, presumably through membrane-specific isoforms of immunoglobulin as part of the BCR. Upon GC exit and differentiation, plasmablasts lose membrane-bound immunoglobulin expression and instead start to secrete large amounts of soluble antibody. In contrast MBCs retain BCR expression of which the antibody isotype may influence the phenotypic properties of different MBC subsets (Engels and Wienands, 2018).

We therefore sought to determine whether antibody class expression by MBCs might be linked with different functional abilities and to better define the heterogeneity within the MBC pool in secondary lymphoid tissue. A significant proportion of memory B cells are unswitched (Figure 1B), so to examine

382 potential differential gene expression across class switched memory B cells subsets we generated
383 paired single-cell transcriptomics and VDJ repertoires for IgD-depleted or IgM/IgD-depleted MBCs from
384 the same tonsillar immune cell preparations analysed previously. Dataset integration and quality control
385 provided 21,595 high-quality MBC single-cell transcriptomes that we annotated with 11 clusters
386 reflecting different MBC subsets and states (Figure 5A-C), all of which lacked marker gene expression
387 for naïve or GC B cells (Figure S4A). In addition to tissue-resident FCRL4^{+} MBCs previously identified
388 (Figure 2; Ehrhardt et al., 2005), we annotated two rare *CR2*/CD21^{low} MBC subsets resembling
389 populations described elsewhere (Lau et al., 2017, Thorarinsdottir et al., 2016) (Figures 5A,C and S4B).
390 However, the majority of MBC diversity within human tonsils appeared to reflect differences in cellular
391 state or signalling activity rather than distinct cell types, such as different activation states (Activated 1
392 and Activated 2), heat shock protein (HSP)-related gene activity (HSP-response), and an IFN-
393 responsive MBC state (IFN-response) (Figures 5C and S4C). We also identified a preGC MBC
394 population with similar gene expression to naïve preGC cells and an enrichment for class switch
395 recombination genes (Figures 5C and S4D-E), as well as an FCRL2/3^{high} MBC state similar to the
396 FCRL2/3^{high} GC population (Figures 5C and S4D), suggesting that these may be widely shared
397 functional states spanning multiple B cell fates.

398 We next considered whether class-switched and unswitched MBCs exhibit different gene expression
399 networks that might reflect unique functional abilities (Dogan et al., 2009). We found little evidence that
400 antibody class contributed towards the likelihood of an MBC to exist in a given state, with the exception
401 of FCRL2/3^{high} MBCs which, similar to FCRL2/3^{high} GC B cells, were enriched for IgM^{+} cells (Figures 5B
402 and S4F). Intriguingly, key marker genes associated with this state were broadly up-regulated in IgM^{+}
403 cells across all MBC clusters (Figure S4G), suggesting a close relationship between the expression of
404 these genes and IgM expression. We therefore compared gene expression of switched and unswitched
405 MBCs of equivalent somatic hypermutation levels as a proxy for affinity (similar to our analysis of GC B
406 cells in Figure 4), as MBCs expressing switched isotypes tended to have higher somatic hypermutation
407 frequencies than unswitched MBCs (Figure S4H). We found widespread differences between
408 unswitched and switched MBCs that were independent of MBC subset or state (Figure 5D), indicating
409 a specific transcriptional wiring of switched and unswitched MBCs that might reflect altered abilities to
410 signal to other immune cell populations, proliferate, survive, differentiate, migrate and respond to
411 challenges (Figure 5E). In particular, we noted elevated expression in unswitched MBCs of genes
412 known to regulate B cell migration within lymphoid tissues and genes with the potential to signal to and
413 activate other immune cell types, including *IL6*, *CLEC2B*, *CR2* (CD21), *CKLF*, *S1PR1*, *CCR6*, and
414 *SEMA7A* (Figure 5F; Arkatkar et al., 2017, Suzuki et al., 2008, Elgueta et al., 2015, Cinamon et al.,
415 2004, Han et al., 2001, Welte et al., 2006). This could all contribute to the increased capacity of IgM^{+}
416 MBCs to re-initiate GC reactions as part of a recall memory response (Dogan et al., 2009, Lutz et al.,
417 2015, Seifert et al., 2015).

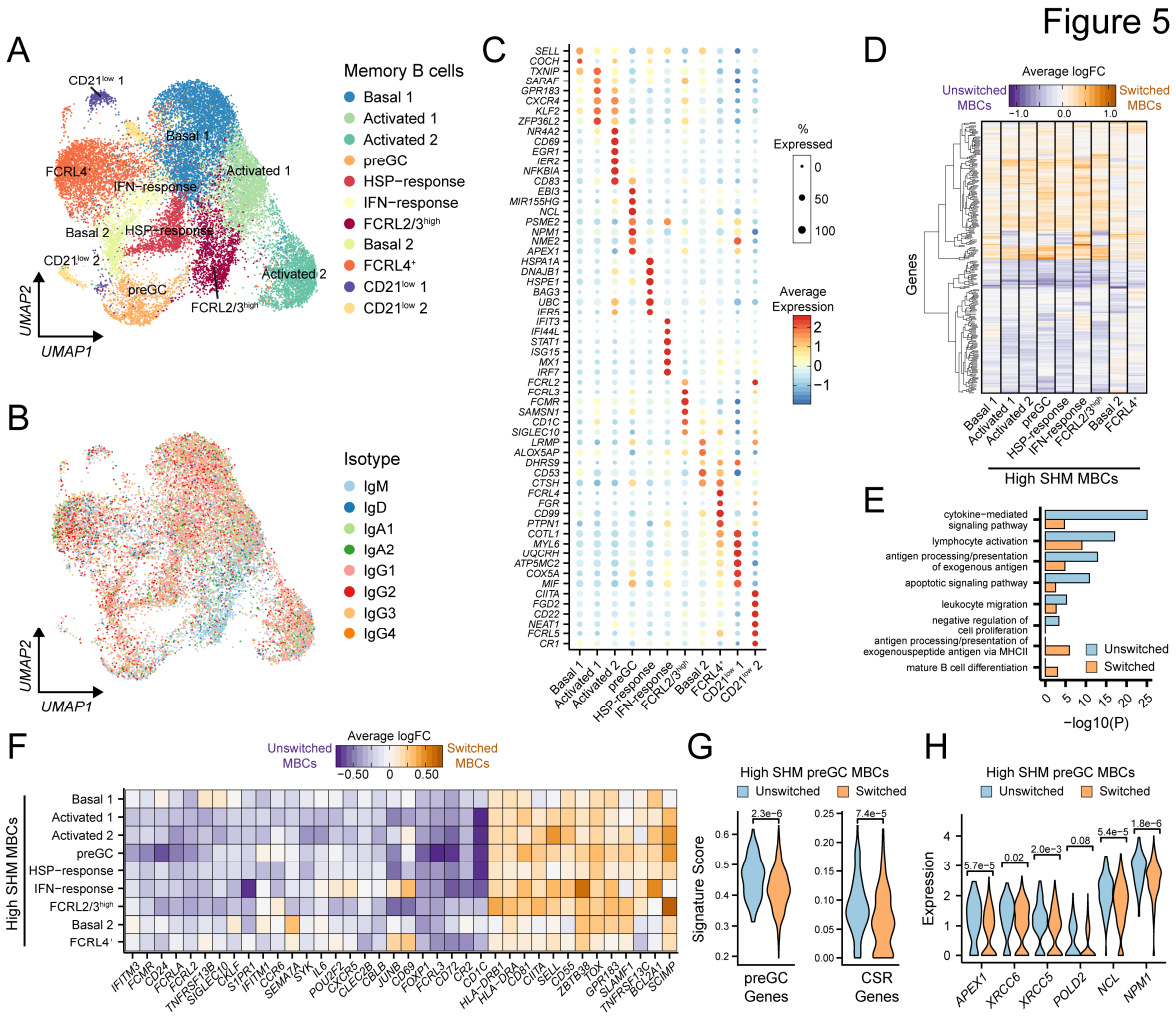

Figure 5. Diverse memory B cell states and antibody class-dependent gene expression.

A) Clustering and UMAP visualisation of 21,595 memory B cell (MBC) single-cell transcriptomes. Identified cell populations include multiple basal and activated states, MBCs enriched for heat shock protein (HSP)-response and interferon (IFN)-response genes, preGC MBCs, FCRL2/3^{high} MBCs, tissue-resident FCRL4^{+} MBCs and two CD21^{low} populations.

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

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

D) Average log fold change (logFC) of genes significantly enriched in switched or unswitched MBC states with similarly high affinity (based on SHM frequency). CD21^{low} clusters had too few cells and were excluded.

E) Gene ontologies for genes significantly enriched in switched or unswitched MBCs with high affinity/SHM, in any cell state.

F) Average logFC of selected immunologically-relevant genes significantly enriched in switched or unswitched MBC populations with high affinity/SHM.

G) Single-cell scores for preGC and class switch recombination (CSR) signature gene sets in switched and unswitched preGC MBC with high affinity/SHM.

H) Single-cell gene expression of key class switched recombination genes in switched and unswitched preGC MBC with high affinity/SHM.

However, IgM^{+} MBCs were also enriched for many genes proposed to regulate or inhibit B cell activation, such as *FCRLA*, *FCRL2*, *FCRL3*, *CBLB*, *CD72* and *SIGLEC10* (Wu and Bondada, 2009, Sohn et al., 2003, Meyer et al., 2018, Kochi et al., 2009, Shabani et al., 2014), which may reflect a fine regulatory balance controlling their activation threshold. Perhaps underpinning this, unswitched MBCs also expressed higher levels of the transcription factor genes *POU2F2* (OCT2) and *FOXP1* than class-switched MBCs (Figure 5F), which coordinate the capacity of B cells to respond normally to antigen

441 receptor signals and directly repress key regulators of plasma cell differentiation respectively (van
442 Keimpema et al., 2015, Corcoran et al., 2014). This is consistent with switched IgG$^+$ MBCs being more
443 likely to differentiate into plasma cells, while unswitched IgM$^+$ MBCs are more likely to re-enter or form
444 secondary GC responses to gain higher affinity (Dogan et al., 2009, Lutz et al., 2015, Seifert et al.,
445 2015). Indeed, we found that unswitched preGC MBCs exhibited significantly higher expression of many
446 genes linked with the preGC state, including those associated with class switch recombination of their
447 antibody genes (Figure 5G-H). This indicates that unswitched MBCs are more primed to undergo class
448 switching for the first time as they re-enter the GC reaction, which will have an important impact on their
449 subsequent selection dynamics in the GC.

450

451 Discussion

452 Antibody responses are the foundation of effective immune memory and our ability to manipulate them
453 through vaccination has contributed significantly to the success of modern medicine. While it is known
454 that high affinity class-switched antibodies are generated during GC reactions, the complexity and
455 dynamic nature of this response has presented a significant challenge for those seeking to understand
456 how human B cell-mediated immunity is derived. By combining bulk antibody repertoire analysis with
457 single-cell transcriptomics we have generated a detailed resource of the human GC response in a
458 model secondary lymphoid tissue. This allowed us to define gene expression signatures of known and
459 novel B cell states, most notably a population primed to undergo class switch recombination before
460 entering the GC reaction. Whether a B cell undergoes class switch recombination at this stage then
461 influences their capacity to undergo antibody-based selection within the GC and secondary activation
462 as MBCs.

463 Although first histologically observed over a century ago, many questions remain about how B cells
464 enter, experience and exit the GC reaction (Mesin et al., 2016, Shlomchik et al., 2019, Cyster and Allen,
465 2019). The practical challenges of sequentially sampling patient lymphoid tissues during an immune
466 response make this especially true for understanding the most dynamic aspects of human B cell
467 maturation. In particular, understanding the early events that facilitate GC entry by human B cells could
468 provide new targets for adjuvants during vaccination or other immunotherapies. We have used
469 pseudotemporal ordering to map the gene expression dynamics of both the early stages of B cell
470 activation that correspond to antigen-dependent signalling through the BCR and the subsequent
471 transition to a transcriptionally distinct preGC state, the latter of which is presumably under the
472 regulation of cognate antigen-specific T helper cells. Our discovery that this preGC state is primed to
473 undergo class switch recombination supports mounting evidence that class switching occurs before the
474 classical GC response (Toellner et al., 1996, Roco et al., 2019, Pape et al., 2003) and has profound
475 implications for understanding antibody-based selection dynamics in the GC.

476 Within the GC, B cell survival and selection is dependent on antigen binding to the BCR and its
477 downstream signalling pathways (Kwak et al., 2019). By using single-cell transcriptomics paired with
478 BCR sequence analysis we have been able to uncouple antibody class, affinity and B cell phenotype
479 at single-cell resolution. We show that whether a B cell has undergone class switch recombination
480 during GC entry is a major determinant of that B cell's capacity to expand, acquire high antigen affinity
481 and differentiate into plasmablasts or MBCs. IgG$^+$ and IgA$^+$ GC B cells have gene expression patterns
482 consistent with increased BCR signalling and a greater capacity to remain within the GC, acquire T cell
483 help and undergo somatic hypermutation to increase their affinity than GC B cells that have retained
484 IgM expression. If class switch recombination does indeed occur predominantly prior to GC entry, as
485 we and others suggest, these data support a model whereby the ability of a B cell to acquire high affinity
486 is primarily dictated by the outcome of a specific class switch recombination "checkpoint" at the preGC
487 stage. This would explain our observation that switched MBCs have comparable somatic hypermutation
488 frequencies to switched plasmablasts, in contrast to the prevailing paradigm that higher affinity GC B
489 cells preferentially differentiate towards the plasmablast fate whereas lower affinity clones seed the
490 memory compartment (Suan et al., 2017b, Phan et al., 2006, Shinnakasu et al., 2016). The differences
491 in affinity between these populations may instead be explained by the likelihood of whether they
492 retained IgM expression prior to entering affinity maturation in the GC. This class switch recombination
493 "checkpoint" is also relevant during the secondary activation of MBCs, where IgM$^+$ MBCs appear more
494 likely to undergo class switching compared to IgG$^+$ or IgA$^+$ MBCs, which may be important to provide a
495 higher affinity secondary response through more prolonged GC maturation as well as more diverse
496 effector functions from class-switched antibodies. Finally, our discovery of a conserved "unswitched"
497 signature of elevated FCRL2 and FCRL3 expression (amongst other genes) across different B cell
498 states, raises interesting questions about how these genes might regulate B cell function and immune
499 responses more broadly.

500 Although the direct mechanisms shaping these gene expression differences between class-specific B
501 cells remain to be elucidated, variations in the immunoglobulin tail tyrosine domain, linker flexibility or
502 glycosylation sites between IgM and other antibody isotypes may all contribute to differential BCR
503 signalling that could shape the behaviour of class-specific B cells (Martin and Goodnow, 2002, Engels
504 et al., 2014, Xu et al., 2014b). Of note, we did not identify many major or meaningful differences in gene
505 expression between IgG$^+$ and IgA$^+$ B cells, or between subclass-specific B cells, which may reflect the
506 need to increase the power of future studies to identify potentially subtle gene expression differences
507 between the less abundant isotypes. Unfortunately, we did not identify any IgE$^+$ GC B cells, which are
508 very rare in human tonsils, likely as they have been described to rapidly exit GCs due to IgE-specific
509 BCR signalling (Haniuda et al., 2016). Given their relative availability, human tonsils are a useful tissue
510 with which to examine the GC response. It will be of interest for future studies to compare class-specific
511 gene expression differences in other tissues and contexts to determine the contribution of the local
512 cellular environment on B cell maturation (James et al., 2020, Smillie et al., 2019). Similarly, recent

513 technical advances allowing the simultaneous readout of antigen specificity, antibody sequences and
514 gene expression (Setliff et al., 2019) make it possible to examine how different types of antigens may
515 be linked with different B cell phenotypes. Finally, our profiling of human MBCs in a secondary lymphoid
516 tissue revealed diverse states reflecting different activation, signalling and functional potential. Given
517 an emerging appreciation for heterogeneity within both human and mouse MBC populations (Good-
518 Jacobson and Shlomchik, 2010), our single-cell characterisation of different MBC states will act as a
519 valuable resource to interrogate the potential relevance for such diverse populations in mediating
520 humoral immunity.

521 Together, our integrated analyses of antibody repertories and gene expression of human B cell states
522 during affinity maturation highlight how the outcome of class switch recombination is a major
523 determinant of B cell fate and function. More broadly, our detailed annotation of diverse B cell states
524 provides a new and uniquely detailed framework through which to view B cell-mediated immune
525 responses in the context of both health and B cell-related pathologies such as allergy, multiple sclerosis,
526 rheumatoid arthritis and lymphoma.

527

528 # Methods

529 ## Human ethics, tissue collection and sorting of B cell subsets

530 Routine tonsillectomy patients at the Royal London Hospital aged between 3 and 14 were consented
531 for tissue collection with approval from North West - Greater Manchester East Research Ethics
532 Committee under REC reference 17/NW/0664. Following removal of dead tissue and clotted blood,
533 each palatine tonsil was bisected and processed separately as follows, with all repertoire and single-
534 cell analyses performed on a single bisected sample. Tonsillar tissue was dissected manually into
535 approximately 2-3 mm pieces prior to homogenisation with the gentleMACS™ Dissociator using C tubes
536 and two rounds of the Multi_C_01_01 setting in 8 mL RPMI + 10% fetal calf serum (FCS). Dissociated
537 cells were then passed through a 70 µM filter prior to isolation of mononuclear lymphocytes using Ficoll-
538 Paque™ gradients. Isolated mononuclear cells were then washed in RPMI + 10% FCS before cell
539 counting and viability determination using Trypan Blue staining. Cells to be used for 10X single-cell
540 transcriptomics were processed immediately, while cells for bulk repertoire sequencing were either
541 stored overnight at 4°C or cryopreserved in FCS with 10% DMSO at -70°C.

542 For bulk B cell repertoires, we labelled 1-1.3×10^8 tonsillar lymphocytes per donor for fluorescence-
543 activated cell sorting (FACS). Briefly, cells were washed and stained with Zombie NIR™ Fixable
544 Viability Kit (BioLegend) to label dead cells, followed by washing with FACS buffer (PBS + 0.5 % BSA
545 + 2 mM EDTA) and incubation with human FcR Blocking Reagent (Miltentyi Biotec). Cells were then
546 stained with CD19-APC (clone HIB19; BioLegend), CD38-PE-Cy7 (clone HB-7; BioLegend), CD27-
547 Pacific Blue™ (clone O323; BioLegend), IgD-PerCP-Cy5.5 (clone IA6-2; BioLegend), and IgM-FITC

548 (clone MHM-88; BioLegend). For bulk B cell repertoires, two aliquots of 250,000 cells for the following
549 populations were sorted using a BD FACSAria™ IIIu: total B cells (live $CD19^+$), naïve B cells (live $CD19^+$
550 IgD^+ $CD38^-$), germinal centre B cells (live $CD19^+$ IgD^- $CD38^+$), memory B cells (live $CD19^+$ IgD^- $CD38^-$),
551 and plasmablasts (live $CD19^+$ IgD^- $CD38^{++}$). Gates were set using fluorescence minus one (FMO)
552 controls. Sorted B cell samples were processed immediately for RNA extraction. For single-cell RNA-
553 seq, 50,000-200,000 class-switched memory B cells (live $CD19^+$ IgD^- $CD38^-$ IgM^+ (n=2) or IgM^- (n=4))
554 were sorted.

555

556 ## Bulk VDJ repertoire library preparation and sequencing

557 RNA was isolated from sorted B cell aliquots using the RNAqueous™-Micro Total RNA Isolation Kit
558 (ThermoScientific) supplemented with β-mercaptoethanol according to manufacturer's protocol. RNA
559 was stored long-term at -80°C or processed immediately to generate bulk repertoire sequencing
560 libraries of immunoglobulin heavy chains (IgH) as previously described (Horns et al., 2016), with minor
561 changes. Briefly, 50 to 100 ng RNA from sorted B cell subsets were annealed to a pooled set of five
562 isotype-specific IgH constant region primers containing unique molecular identifiers (UMIs) of either 10
563 or 12 nucleotides at 72°C for 5 minutes before being immediately placed on ice for 2 minutes. We then
564 performed first-strand cDNA synthesis using SuperScript IV reverse transcriptase (ThermoFisher
565 Scientific) with recommended reagent concentrations and the following cycling conditions in a
566 thermocycler: 105°C lid; 55°C 10 minutes; 80°C 10 minutes; 4°C hold. Second-strand cDNA synthesis
567 was performed using Phusion® High-Fidelity DNA Polymerase (NEB) and six IgH variable region
568 primers containing 10 or 12 nucleotide UMIs with the following cycling conditions: 105°C lid; 98°C 4
569 minutes; 52°C 1 minutes; 72°C 5 minutes; 4°C hold. Double-stranded cDNA was then purified using
570 (0.6X) Ampure XP beads (Beckman Coulter) before amplification with Illumina adapter-containing
571 primers (Nextera i7 indices) and NEBNext Ultra II Q5 Master Mix (NEB) as follows: 105°C lid; 98°C 30
572 seconds; (98°C 10 seconds, 72°C 50 seconds) × 22 to 28 cycles; 72°C 2 minutes; 4°C hold. Amplified
573 libraries were purified using (0.6X) Ampure XP beads and quantified by Qubit™ dsDNA HS Assay Kit
574 prior to multiplexing. Libraries were sequenced with PhiX spike-in using paired-end 301 bp reads on
575 the Illumina MiSeq platform.

576

577 ## 10x Genomics Chromium single-cell transcriptomics and VDJ library preparation,
578 sequencing and raw data processing

579 Total tonsillar immune cells (n=7) or FACS-enriched memory B cells (n=6) were loaded according to
580 the manufacturer's protocol for either for the Chromium single-cell 3' kit (v2; n = 1) or 5' gene expression
581 (v1; n = 6) to attain between 2000-5000 cells per well. Library preparation for both gene expression and
582 VDJ (BCR) was performed according to the manufacturer's protocol prior to sequencing on the Illumina

583 NextSeq 500 platform with 26/8/134 bp or 155/8/155 bp read configurations respectively. BaseCall files
584 were subsequently used to generate library-specific FASTQ files with cellranger mkfastq (v3.0.0) prior
585 to running cellranger count (v3.0.0) with the GRCh38 (release 92) reference to produce cell barcode-
586 gene expression matrices using default settings. For single-cell VDJ datasets, cellranger vdj (v3.0.0)
587 was run using the refdata-cellranger-vdj-GRCh38-alts-ensembl-2.0.0 reference from 10x Genomics
588 using default settings. Poor quality contigs that either did not map to immunoglobulin chains or were
589 designated incomplete by cellranger were discarded. We further filtered IgH contigs as to whether they
590 had sufficient coverage of constant regions to ensure accurate isotype assignment between closely
591 related subclasses using MaskPrimers.py (pRESTO v0.5.10; Vander Heiden et al., 2014).

592

593 Quality control and sequence assembly of bulk B cell repertoires

594 Raw sequencing data from bulk VDJ libraries were processed to generate UMI-collapsed consensus
595 VDJ sequences using pRESTO (v0.5.10; Vander Heiden et al., 2014). Paired-end sequencing reads
596 with mean Phred quality scores less than 25 were removed, and remaining sequences were annotated
597 and trimmed for PCR primer and UMI sequences. UMI barcodes were then filtered by length and the
598 presence of ambiguous nucleotides, prior to UMI alignment using MUSCLE (v3.8.31; Edgar, 2004).
599 UMI consensus sequences were then generated, with a minimum of three reads per UMI required, prior
600 to assembly of paired-end UMI consensus sequences into a single VDJ contig and annotation of
601 constant region isotype using MaskPrimers.py align (v0.5.10; Vander Heiden et al., 2014). Duplicate
602 VDJ sequences within each subset were then collapsed using CollapseSeq.py (v0.5.10; Vander Heiden
603 et al., 2014) before VDJ gene assignment and functional annotation using AssignGenes.py (ChangeO
604 v0.4.5; Gupta et al., 2015) and IgBLAST (v1.12.0; Ye et al., 2013).

605

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

608 Following initial quality control, all single-cell VDJ sequences were combined together with bulk BCR
609 repertoire sequences from the same donor for subsequent processing. IgH sequences were annotated
610 using IgBLAST (v1.12.0; Ye et al., 2013) and assigned isotype classes using AssignGenes.py prior to
611 correction of ambiguous V gene assignments using TIgGER (v0.3.1; Gupta et al., 2015, Gadala-Maria
612 et al., 2015). Clonally-related IgH sequences were identified using DefineClones.py (ChangeO v0.4.5;
613 Gupta et al., 2015) with a nearest neighbour distance threshold of 0.0818, as determined by the mean
614 99% confidence interval of all 8 donors with distToNearest (Shazam v0.1.11; Gupta et al., 2015).
615 CreateGermlines.py (ChangeO v0.4.5) was then used to infer germline sequences for each clonal
616 family and observedMutations (Shazam v0.1.11) was used to calculate somatic hypermutation
617 frequencies for each IgH sequence. Sequences with mutation frequencies greater than 0.02 were

618 annotated as “High” mutation levels, those between 0 and 0.02 as “Low” mutation levels and 0 as
619 “None”. For bulk BCR repertoire analysis in Figure 1, single-cell VDJ sequences were excluded and
620 analysed, providing ~1.5 million high-confidence and unique IgH sequences, with a median of 14 UMIs
621 per sequence, a median of 28,918 unique sequences per donor per subset and approximately 96-99%
622 of these sequences appeared to be functional.

623 Single-cell VDJ analysis was performed broadly as described previously (James et al., 2020). Briefly,
624 the number of quality filtered and annotated IgH, IgK or IgL were determined per unique cell barcode
625 prior to integration with single-cell gene expression objects. If more than one contig per chain was
626 identified, metadata for that cell was ascribed as “Multi”. IgH diversity analyses were performed using
627 the rarefyDiversity and testDiversity of Alakazam (v0.2.11; Gupta et al., 2015). To assess clonal
628 relationships between cell types, co-occurrence of expanded clone members between cell types was
629 reported as a binary event for each clone that contained a member within two different cell types in
630 either single-cell or bulk repertoires. For comparisons of somatic hypermutation and isotype frequencies
631 between subsets we used Wilcoxon Rank Sum Signed test while Student’s t test was used to compare
632 mean clonal diversity scores.

633

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

635 Gene expression count matrices from cellranger were used to calculate percentage mitochondrial
636 expression per cell barcode prior to mitochondrial genes being removed from gene expression matrices.
637 Similarly, the V, D and J gene counts for each immunoglobulin and T cell receptor were summed to
638 calculate an overall expression before individual genes were removed from gene expression matrices.
639 Counts of individual IgH constant region genes were also summed together (IgG1-4, and IgA1-A2) and
640 removed from gene expression matrices. Modified gene-by-cell matrices were then used to create
641 Seurat objects for each sample using Seurat (v3.0.3; Butler et al., 2018, Stuart et al., 2019), removing
642 genes expressed in fewer than 3 cells. Cell barcodes with >1000 and <60000 UMIs and >500 and
643 <7000 genes detected were removed, as were cell barcodes with >30% mitochondrial reads. Individual
644 samples were then log transformed, normalised by a factor of 10000 prior to predicting cell cycle phases
645 using the CellCycleScoring command and then identifying the 3000 most variable genes within each
646 sample using the “vst” method. We then performed a preliminary integration of all unsorted immune
647 cells or all sorted memory B cell datasets together using FindIntegrationAnchors and IntegrateData
648 (3000 genes) before regressing out cell cycle scores and mitochondrial gene expression, performing
649 principle component analysis (PCA) and preliminary clustering and cell type annotation. One cluster
650 was identified to be enriched with predicted doublets based on the results from DoubletFinder (v2.0.1;
651 McGinnis et al., 2019) and scrublet (v0.2; Wolock et al., 2019), and a small number of cell barcodes
652 with co-expression of B/T/non-lymphoid markers were manually removed by filtering on UMAP
653 coordinates. Following the removal of poor quality cell barcodes from gene expression matrices based

654 on these preliminary analyses of the unsorted immune cell and sorted memory B cell libraries, we then
655 integrated all normalised count matrices together using the unsorted immune cell count matrices as a
656 reference with 4000 highly variable genes before scaling the integrated data and regressing cell cycle
657 and mitochondrial gene expression, running PCA and identifying broad cell type lineages (B cell, T cell
658 and non-lymphoid cells) using a broad resolution for clustering. These lineages were then separated
659 for more detailed cell state annotation by recomputing the PCA (RunPCA), nearest neighbour graph
660 (FindNeighbors) and unbiased clustering (FindClusters). Uniform Manifold Approximation and
661 Projection (UMAP) was then used to visualise both integrated and lineage-specific datasets. B cells
662 were annotated with scVDJ metadata from the integrated repertoire analysis detailed above.

663

664 Differential gene expression and signature enrichment analysis.

665 Gene expression markers for different clusters of unsorted B cells, T cells, non-lymphoid cells and
666 sorted MBCs were identified using FindAllMarkers from Seurat with default settings, including Wilcoxon
667 test and Bonferroni *p* value correction (v3.0.3; Butler et al., 2018, Stuart et al., 2019). Differential gene
668 expression for antibody class-specific or somatic hypermutation frequency for GC B cells or class-
669 switched MBCs was performed using FindAllMarkers with Benjamini-Hochberg false discovery rate
670 (FDR) correction. Genes were deemed significantly different if FDR < 0.05, average log fold change >
671 0.1 and the gene was detected in >20% of cells in that group. Gene ontology analyses for high
672 affinity/SHM class-specific GC B cells were performed with Ingenuity Pathway Analysis (Qiagen)
673 software using avg_logFC values of all genes significantly enriched in at least one class. For IFN-
674 response MBC cluster gene ontology enrichment was performed with all significant gene markers for
675 this cluster in Metascape (Zhou et al., 2019) using default settings, as were all genes significantly
676 enriched or depleted in switched or unswitched in at least one MBC cluster.

677 Enrichment of gene set signatures for single cells was calculated using AUCell (v1.5.5; Aibar et al.,
678 2017). For class switch recombination, a manually curated shortlist of genes was determined for genes
679 linked with CSR that were reliably detected in our sparse scRNA-seq datasets (Stavnezer and
680 Schrader, 2014). Gene signatures of high and low affinity mouse LZ GC B cells (Shinnakasu et al.,
681 2016) were obtained by quantifying RNA-seq transcript counts against GRCm38 transcriptome build
682 using Salmon (v1.0.0; Patro et al., 2017), collapsing protein-coding transcripts into a single gene count
683 using tximport (v1.10.1; Soneson et al., 2016), identifying significant gene expression differences
684 between the two groups using DESeq2 (v1.22.2; Love et al., 2014) with a threshold of log fold change
685 > 1.5 and *padj* < 0.05 and converting mouse gene IDs to human using bioMart (v2.38.0; Durinck et al.,
686 2005, Durinck et al., 2009). For IgM stimulation gene sets, Geo2R (Barrett et al., 2012) was used to
687 analyse microarray data from a timecourse of wild type splenic mouse B cells stimulated with 10 µg/ml
688 of anti-IgM (Shinohara et al., 2014) and identify genes enriched following α-IgM treatment compared to
689 control untreated cells with FDR < 0.05 and a fold change > 2. To calculate preGC and FCRL2/3^{high}

signature scores in MBC subsets, the top 50 most significantly enriched gene markers per cluster in the unsorted B cell subset analysis were used with AUCell. Unless indicated otherwise, Wilcoxon Ranked Signed Sum test was used to test for significant differences.

Prediction of cell-cell communication using CellPhoneDB.

To evaluate potential cell-cell communication, we used CellPhoneDB (v2.0.6; Vento-Tormo et al., 2018, Efremova et al., 2020) 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 (v2.38.0; Durinck et al., 2005). 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 FCRL2/3^{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.

RNA velocity and pseudotemporal ordering.

To calculate single-cell velocities we first quantified spliced and unspliced transcripts for the filtered barcodes output from cellranger using velocyto (v0.17.10; La Manno et al., 2018). Loom files were then combined using loompy (v2.0.17) before reformatting cell barcode names to be compatible with Seurat objects and merging with a Scanpy (v1.4; Wolf et al., 2018) object containing the raw gene expression matrix of high quality annotated single B cell transcriptomes (see above) using scVelo (v0.1.23; Bergen et al., 2019). scVelo was then used to pre-process, filter and normalise velocyto-derived counts with default settings prior to computation of the first- and second-order moments (scv.pp.moments) and subsequent velocity estimation using a dynamical model (scv.tl.recover_dynamics and scv.tl.velocity). Velocities were then projected and visualised onto UMAP embeddings at a grid level using an inverted transition matrix obtained from scv.tl.transition_matrix prior to scv.tl.velocity_embedding, basis='umap'.

For pseudotemporal ordering of the B cell activation and GC entry trajectory, a partition-based graph abstraction (PAGA) was performed for Naïve, Activated, preGC and LZ GC B cell clusters (Scanpy; tl.paga) before computing connectivity of single cells using a diffusion map (Scanpy; tl.diffmap). Velocity-based pseudotime reconstruction was performed using default settings for scVelo commands tl.recover_latent_time and tl.velocity_pseudotime, although the pseudotemporal order was reversed to place naïve cluster at pseudotime = 0. Dynamic gene expression changes were examined by using tl.rank_velocity_genes (scVelo) to sub-cluster original cell type annotations (resolution = 1) based on

725 RNA velocity and the top 200 genes per sub-cluster were reported before filtering out ribosomal genes
726 and collapsing to unique genes. Genes were then clustered through pseudotime for heatmap
727 visualisation with smoothed expression scores in scVelo. Quantitation of individual gene expression or
728 AUCell-derived signature scores of single-cells across pseudotime was performed using smoothed
729 normalised counts with geom_smooth() including 95% confidence intervals. For pseudotemporal
730 analysis of GC B cell subsets, a similar approach was taken for the LZ GC, GC, DZ GC, FCRL2/3^{high}
731 GC and prePB clusters, except that diffusion-based pseudotime was calculated with Scanpy (tl.dpt)
732 independent of RNA velocity measurements. Visualisation of a custom list of top GC subset marker
733 genes for LZ GC, GC and DZ GC clusters was performed using pl.paga.path heatmap with Scanpy.

734

735 Immunohistochemistry

736 Tonsil biopsies were cut from formalin fixed paraffin-embedded blocks then deparaffinized in xylene
737 and rehydrated through a series of ethanols to water. Endogenous peroxidase was blocked with 3%
738 hydrogen peroxide. Heat-mediated antigen retrieval was performed using a commercial citrate-based
739 unmasking buffer (Vector Labs) at 120°C using a pressure cooker. Sections (3µm thick) were then
740 incubated for 40 minutes at RT with 1:1000 dilution of anti-APE1 (HPA002564; Sigma). Detection of
741 primary antibody was performed using the Super-sensitive–Polymer HRP system (Biogenex) and
742 staining visualized using purple chromogen VIP (Vector Labs) and hematoxylin as a nuclear
743 counterstain. Slides were then scanned (Pannoramic 250 Flash) before being left to soak in xylene to
744 de-coverslip. Once the coverslips were removed, slides were rehydrated through ethanol to water. De-
745 staining and stripping of the primary antibodies and the heat-labile VIP chromogen was achieved using
746 a subsequent round of heat-mediated antigen retrieval as per the first round of staining. The second
747 primary antibody CD20 (M0755; Dako) was incubated for 40 minutes at RT at a dilution of 1:500,
748 followed by detection, visualization and scanning as before. Negative controls were performed by
749 treating sequential sections as above but without the second primary antibody (CD20) to confirm
750 complete antibody and signal stripping. Images were prepared using CaseViewer (3DHistTech).

751

752 Quantitation of IgH germline transcripts.

753 To quantify expression of IgH germline transcripts (GLT), all mapped reads to the IgH locus
754 (chr14:105540180-105879151) were extracted from cellranger-derived bam files. DropEst (v0.8.6;
755 Petukhov et al., 2018) was then used to count reads against a custom GTF containing coordinates for
756 I promoter GLT sequences (Sideras et al., 1989, Fujieda et al., 1996), annotation of membrane-specific
757 IgH exons and IgH switch regions identified by enrichment of the WRCY motif using HOMER2 (v4.9.1;
758 Heinz et al., 2010). Counts were then read into Seurat without filtering for log10-normalisation and
759 scaling. Due to mapping ambiguities between subclass-specific regions, subclass-specific counts were
760 summed together.

Accession codes and data availability

Raw sequencing and processed data files for single-cell RNA sequencing, single-cell VDJ sequencing, and bulk B cell repertoires are available at ArrayExpress (accession numbers: E-MTAB-8999, E-MTAB-9003 and E-MTAB-9005). RNA-seq and microarray data for high affinity and α-IgM stimulation gene expression signatures were obtained from GSE73729 and GSE41176 respectively. DNase-seq and ChIP-seq datasets from ENCODE were visualised using UCSC Genome Browser (Kent et al., 2002, Rosenbloom et al., 2012).

Acknowledgements

This research was supported by funding from the Wellcome Trust to L. K. J. (208961/Z/17/Z), S. A. T. (206194) and J. C. R. (110020/Z/15/Z). H. W. K. was funded by a Sir Henry Wellcome PostDoctoral Fellowship (213555/Z/18/Z). We would like to thank the Bart's and the London Genome Centre at QMUL for library sequencing support. We also thank members of the James and Teichmann labs for their help and support, especially Mirjana Efremova for her help with CellPhoneDB. Finally, we are grateful to Neil McCarthy, Kylie James, Kerstin Meyer, Lou Herman and Jo Spencer for their help reviewing the manuscript.

Author contributions

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

Conflict of Interest Statement

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.

References

Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J.*, et al.* 2017. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods,* 14**,** 1083-1086.

Arana, E., Harwood, N. E. & Batista, F. D. 2008. Regulation of integrin activation through the B-cell receptor. *J Cell Sci,* 121**,** 2279-2286.

Arkatkar, T., Du, S. W., Jacobs, H. M., Dam, E. M., Hou, B., Buckner, J. H., Rawlings, D. J. & Jackson, S. W. 2017. B cell–derived IL-6 initiates spontaneous germinal center formation during systemic autoimmunity. *J Exp Med,* 214**,** 3207-3217.

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M.*, et al.* 2012. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res,* 41**,** D991-D995.

Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. 2019. Generalizing RNA velocity to transient cell states through dynamical modeling. *bioRxiv***,** 820936; doi:10.1101/820936.

Borggrefe, T., Wabl, M., Akhmedov, A. T. & Jessberger, R. 1998. A B-cell-specific DNA Recombination Complex. *J Biol Chem,* 273**,** 17025-17035.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol,* 36**,** 411-420.

Cinamon, G., Matloubian, M., Lesneski, M. J., Xu, Y., Low, C., Lu, T., Proia, R. L. & Cyster, J. G. 2004. Sphingosine 1-phosphate receptor 1 promotes B cell localization in the splenic marginal zone. *Nat Immunol,* 5**,** 713-20.

Corcoran, L., Emslie, D., Kratina, T., Shi, W., Hirsch, S., Taubenheim, N. & Chevrier, S. 2014. Oct2 and Obf1 as Facilitators of B:T Cell Collaboration during a Humoral Immune Response. *Front Immunol,* 5**,** 108.

Cyster, J. G. & Allen, C. D. C. 2019. B Cell Responses: Cell Interaction Dynamics and Decisions. *Cell,* 177**,** 524-540.

Dogan, I., Bertocci, B., Vilmont, V., Delbos, F., Megret, J., Storck, S., Reynaud, C. A. & Weill, J. C. 2009. Multiple layers of B cell memory with different effector functions. *Nat Immunol,* 10**,** 1292-9.

Dubovsky, J. A., Chappell, D. L., Harrington, B. K., Agrawal, K., Andritsos, L. A., Flynn, J. M., Jones, J. A., Paulaitis, M. E., Bolon, B., Johnson, A. J.*, et al.* 2013. Lymphocyte cytosolic protein 1 is a chronic lymphocytic leukemia membrane-associated antigen critical to niche homing. *Blood,* 122**,** 3308-3316.

Duffy, K. R., Wellard, C. J., Markham, J. F., Zhou, J. H., Holmberg, R., Hawkins, E. D., Hasbold, J., Dowling, M. R. & Hodgkin, P. D. 2012. Activation-induced B cell fates are selected by intracellular stochastic competition. *Science,* 335**,** 338-41.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A. & Huber, W. 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics,* 21**,** 3439-40.

Durinck, S., Spellman, P. T., Birney, E. & Huber, W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc,* 4**,** 1184-91.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32**,** 1792-7.

Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. 2020. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc,* 15**,** 1484–1506.

Ehrhardt, G. R., Hsu, J. T., Gartland, L., Leu, C. M., Zhang, S., Davis, R. S. & Cooper, M. D. 2005. Expression of the immunoregulatory molecule FcRH4 defines a distinctive tissue-based population of memory B cells. *J Exp Med,* 202**,** 783-91.

Elgueta, R., Marks, E., Nowak, E., Menezes, S., Benson, M., Raman, V. S., Ortiz, C., O'connell, S., Hess, H., Lord, G. M.*, et al.* 2015. CCR6-dependent positioning of memory B cells is essential for their ability to mount a recall response to antigen. *J Immunol,* 194**,** 505-513.

Engels, N., Konig, L. M., Schulze, W., Radtke, D., Vanshylla, K., Lutz, J., Winkler, T. H., Nitschke, L. & Wienands, J. 2014. The immunoglobulin tail tyrosine motif upgrades memory-type BCRs by incorporating a Grb2-Btk signalling module. *Nat Commun,* 5**,** 5456.

Engels, N. & Wienands, J. 2018. Memory control by the B cell antigen receptor. *Immunol Rev,* 283**,** 150-160.

Ferreira, R. C., Pan-Hammarström, Q., Graham, R. R., Gateva, V., Fontán, G., Lee, A. T., Ortmann, W., Urcelay, E., Fernández-Arquero, M., Núñez, C.*, et al.* 2010. Association of IFIH1 and other autoimmunity risk alleles with selective IgA deficiency. *Nat Genet,* 42**,** 777-780.

Fujieda, S., Lin, Y. Q., Saxon, A. & Zhang, K. 1996. Multiple types of chimeric germ-line Ig heavy chain transcripts in human B cells: evidence for trans-splicing of human Ig RNA. *J Immunol,* 157**,** 3450-3459.

Gadala-Maria, D., Yaari, G., Uduman, M. & Kleinstein, S. H. 2015. Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. *PNAS,* 112**,** E862-70.

Girón-Pérez, D. A., Vadillo, E., Schnoor, M. & Santos-Argumedo, L. 2020. Myo1e modulates the recruitment of activated B cells to inguinal lymph nodes. *J Cell Sci,* 133**,** jcs235275.

Gloury, R., Zotos, D., Zuidscherwoude, M., Masson, F., Liao, Y., Hasbold, J., Corcoran, L. M., Hodgkin, P. D., Belz, G. T., Shi, W.*, et al.* 2016. Dynamic changes in Id3 and E-protein activity orchestrate germinal center and plasma cell development. *J Exp Med,* 213**,** 1095-1111.

Good-Jacobson, K. L. & Shlomchik, M. J. 2010. 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.

Gupta, N. T., Vander Heiden, J. A., Uduman, M., Gadala-Maria, D., Yaari, G. & Kleinstein, S. H. 2015. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics,* 31**,** 3356-3358.

Han, W., Lou, Y., Tang, J., Zhang, Y., Chen, Y., Li, Y., Gu, W., Huang, J., Gui, L., Tang, Y.*, et al.* 2001. Molecular cloning and characterization of chemokine-like factor 1 (CKLF1), a novel human cytokine with unique structure and potential chemotactic activity. *Biochem J,* 357**,** 127-35.

Hanakahi, L. A., Dempsey, L. A., Li, M.-J. & Maizels, N. 1997. Nucleolin is one component of the B cell-specific transcription factor and switch region binding protein, LR1. *PNAS,* 94**,** 3605-3610.

Haniuda, K., Fukao, S., Kodama, T., Hasegawa, H. & Kitamura, D. 2016. Autonomous membrane IgE signaling prevents IgE-memory formation. *Nat Immunol,* 17**,** 1109-17.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. & Glass, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell,* 38**,** 576-89.

863 Horns, F., Vollmers, C., Croote, D., Mackey, S. F., Swan, G. E., Dekker, C. L., Davis, M. M. & Quake, S. R. 2016. Lineage
864 tracing of human B cells reveals the in vivo landscape of human antibody class switching. *eLife,* 5**,** e16578.
865 Huang, C., Geng, H., Boss, I., Wang, L. & Melnick, A. 2014. Cooperative transcriptional repression by BCL6 and BACH2 in
866 germinal center B-cell differentiation. *Blood,* 123**,** 1012-20.
867 Ise, W., Fujii, K., Shiroguchi, K., Ito, A., Kometani, K., Takeda, K., Kawakami, E., Yamashita, K., Suzuki, K., Okada, T.*, et al.*
868 2018. T Follicular Helper Cell-Germinal Center B Cell Interaction Strength Regulates Entry into Plasma Cell or Recycling
869 Germinal Center Cell Fate. *Immunity,* 48**,** 702-715.e4.
870 James, K. R., Gomes, T., Elmentaite, R., Kumar, N., Gulliver, E. L., King, H. W., Stares, M. D., Bareham, B. R., Ferdinand, J.
871 R., Petrova, V. N.*, et al.* 2020. Distinct microbial and immune niches of the human colon. *Nat Immunol,* 21**,** 343-353.
872 Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., Haussler & David 2002. The Human
873 Genome Browser at UCSC. *Genome Res,* 12**,** 996-1006.
874 Kochi, Y., Myouzen, K., Yamada, R., Suzuki, A., Kurosaki, T., Nakamura, Y. & Yamamoto, K. 2009. FCRL3, an autoimmune
875 susceptibility gene, has inhibitory potential on B-cell receptor-mediated signaling. *J Immunol,* 183**,** 5502-10.
876 Kwak, K., Akkaya, M. & Pierce, S. K. 2019. B cell signaling in context. *Nat Immunol,* 20**,** 963-969.
877 La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M. E., Lönnerberg,
878 P., Furlan, A.*, et al.* 2018. RNA velocity of single cells. *Nature,* 560**,** 494-498.
879 Lau, D., Lan, L. Y.-L., Andrews, S. F., Henry, C., Rojas, K. T., Neu, K. E., Huang, M., Huang, Y., Dekosky, B., Palm, A.-K. E.*,*
880 *et al.* 2017. Low CD21 expression defines a population of recent germinal center graduates primed for plasma cell
881 differentiation. *Sci Immunol,* 2**,** eaai8153.
882 Love, M. I., Huber, W. & Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.
883 *Genome Biol,* 15**,** 550.
884 Lutz, J., Dittmann, K., Bosl, M. R., Winkler, T. H., Wienands, J. & Engels, N. 2015. Reactivation of IgG-switched memory B
885 cells by BCR-intrinsic signal amplification promotes IgG antibody production. *Nat Commun,* 6**,** 8575.
886 Madissoon, E., Wilbrey-Clark, A., Miragaia, R. J., Saeb-Parsy, K., Mahbubani, K. T., Georgakopoulos, N., Harding, P.,
887 Polanski, K., Huang, N., Nowicki-Osuch, K.*, et al.* 2019. scRNA-seq assessment of the human lung, spleen, and
888 esophagus tissue stability after cold preservation. *Genome Biol,* 21**,** 1.
889 Martin, S. W. & Goodnow, C. C. 2002. Burst-enhancing role of the IgG membrane tail as a molecular determinant of memory.
890 *Nat Immunol,* 3**,** 182-8.
891 Masani, S., Han, L. & Yu, K. 2013. Apurinic/Apyrimidinic Endonuclease 1 Is the Essential Nuclease during Immunoglobulin
892 Class Switch Recombination. *Mol Cell Biol*, 33**,** 1468-1473.
893 Mcginnis, C. S., Murrow, L. M. & Gartner, Z. J. 2019. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data
894 Using Artificial Nearest Neighbors. *Cell Systems,* 8**,** 329-337.e4.
895 Mcrae, E. K. S., Booy, E. P., Moya-Torres, A., Ezzati, P., Stetefeld, J. & Mckenna, S. A. 2017. Human DDX21 binds and
896 unwinds RNA guanine quadruplexes. *Nucleic Acids Res,* 45**,** 6656-6668.
897 Mesin, L., Ersching, J. & Victora, G. D. 2016. Germinal Center B Cell Dynamics. *Immunity,* 45**,** 471-482.
898 Meyer, S. J., Linder, A. T., Brandl, C. & Nitschke, L. 2018. B Cell Siglecs-News on Signaling and Its Interplay With Ligand
899 Binding. *Front Immunol,* 9**,** 2820-2820.
900 Mielke, N., Schwarzer, R., Calkhoven, C. F., Kaufman, R. J., Dörken, B., Leutz, A. & Jundt, F. 2011. Eukaryotic initiation factor
901 2alpha phosphorylation is required for B-cell maturation and function in mice. *Haematologica,* 96**,** 1261-1268.
902 Milpied, P., Cervera-Marzal, I., Mollichella, M.-L., Tesson, B., Brisou, G., Traverse-Glehen, A., Salles, G., Spinelli, L. & Nadel,
903 B. 2018. Human germinal center transcriptional programs are de-synchronized in B cell lymphoma. *Nat Immunol,* 19**,**
904 1013-1024.
905 Mondal, S., Begum, N. A., Hu, W. & Honjo, T. 2016. Functional requirements of AID's higher order structures and their
906 interaction with RNA-binding proteins. *PNAS,* 113**,** E1545-E1554.
907 Morman, R. E., Schweickert, P. G., Konieczny, S. F. & Taparowsky, E. J. 2018. BATF regulates the expression of Nfil3,
908 Wnt10a and miR155hg for efficient induction of antibody class switch recombination in mice. *Eur J Immunol,* 48**,** 1492-
909 1505.
910 Muppidi, J. R., Schmitz, R., Green, J. A., Xiao, W., Larsen, A. B., Braun, S. E., An, J., Xu, Y., Rosenwald, A., Ott, G.*, et al.*
911 2014. Loss of signalling via Galpha13 in germinal centre B-cell-derived lymphoma. *Nature,* 516**,** 254-8.
912 Orthwein, A., Patenaude, A.-M., Affar, E. B., Lamarre, A., Young, J. C. & Di Noia, J. M. 2010. Regulation of activation-induced
913 deaminase stability and antibody gene diversification by Hsp90. *J Exp Med,* 207**,** 2751-2765.
914 Pape, K. A., Kouskoff , V., Nemazee , D., Tang , H. L., Cyster , J. G., Tze , L. E., Hippen , K. L., Behrens , T. W. & Jenkins ,
915 M. K. 2003. Visualization of the Genesis and Fate of Isotype-switched B Cells during a Primary Immune Response. *J*
916 *Exp Med,* 197**,** 1677-1687.
917 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. 2017. Salmon provides fast and bias-aware quantification of
918 transcript expression. *Nat Methods,* 14**,** 417-419.
919 Petukhov, V., Guo, J., Baryawno, N., Severe, N., Scadden, D. T., Samsonova, M. G. & Kharchenko, P. V. 2018. dropEst:
920 pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. *Genome Biol,* 19**,**
921 78.
922 Phan, T. G., Paus, D., Chan, T. D., Turner, M. L., Nutt, S. L., Basten, A. & Brink, R. 2006. High affinity germinal center B cells
923 are actively selected into the plasma cell compartment. *J Exp Med,* 203**,** 2419-24.
924 Roco, J. A., Mesin, L., Binder, S. C., Nefzger, C., Gonzalez-Figueroa, P., Canete, P. F., Ellyard, J., Shen, Q., Robert, P. A.,
925 Cappello, J.*, et al.* 2019. Class-Switch Recombination Occurs Infrequently in Germinal Centers. *Immunity,* 51**,** 337-
926 350.e7.
927 Rosenbloom, K. R., Sloan, C. A., Malladi, V. S., Dreszer, T. R., Learned, K., Kirkup, V. M., Wong, M. C., Maddren, M., Fang,
928 R., Heitner, S. G.*, et al.* 2012. ENCODE Data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res,* 41**,**
929 D56-D63.

930 Scharenberg, A. M., Humphries, L. A. & Rawlings, D. J. 2007. Calcium signalling and cell-fate choice in B cells. *Nat Rev*
931 *Immunol,* 7**,** 778-789.
932 Seifert, M., Przekopowitz, M., Taudien, S., Lollies, A., Ronge, V., Drees, B., Lindemann, M., Hillen, U., Engler, H., Singer, B.
933 B.*, et al.* 2015. Functional capacities of human IgM memory B cells in early inflammatory responses and secondary
934 germinal center reactions. *PNAS,* 112**,** E546-55.
935 Setliff, I., Shiakolas, A. R., Pilewski, K. A., Murji, A. A., Mapengo, R. E., Janowska, K., Richardson, S., Oosthuysen, C., Raju,
936 N., Ronsard, L.*, et al.* 2019. High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity. *Cell,* 179**,**
937 1636-1646.e15.
938 Shabani, M., Bayat, A. A., Jeddi-Tehrani, M., Rabbani, H., Hojjat-Farsangi, M., Ulivieri, C., Amirghofran, Z., Baldari, C. T. &
939 Shokri, F. 2014. Ligation of human Fc receptor like-2 by monoclonal antibodies down-regulates B-cell receptor-mediated
940 signalling. *Immunology,* 143**,** 341-53.
941 Shinnakasu, R., Inoue, T., Kometani, K., Moriyama, S., Adachi, Y., Nakayama, M., Takahashi, Y., Fukuyama, H., Okada, T.
942 & Kurosaki, T. 2016. Regulated selection of germinal-center cells into the memory B cell compartment. *Nat Immunol,* 17**,**
943 861-869.
944 Shinohara, H., Behar, M., Inoue, K., Hiroshima, M., Yasuda, T., Nagashima, T., Kimura, S., Sanjo, H., Maeda, S., Yumoto, N.*,*
945 *et al.* 2014. Positive Feedback Within a Kinase Signaling Complex Functions as a Switch Mechanism for NF-κB
946 Activation. *Science,* 344**,** 760-764.
947 Shinozaki, F., Minami, M., Chiba, T., Suzuki, M., Yoshimatsu, K., Ichikawa, Y., Terasawa, K., Emori, Y., Matsumoto, K.,
948 Kurosaki, T.*, et al.* 2006. Depletion of hsp90beta induces multiple defects in B cell receptor signaling. *J Biol Chem,* 281**,**
949 16361-9.
950 Shlomchik, M. J., Luo, W. & Weisel, F. 2019. Linking signaling and selection in the germinal center. *Immunol Rev,* 288**,** 49-63.
951 Sideras, P., Mizuta, T. R., Kanamori, H., Suzuki, N., Okamoto, M., Kuze, K., Ohno, H., Doi, S., Fukuhara, S., Hassan, M. S.*,*
952 *et al.* 1989. Production of sterile transcripts of C gamma genes in an IgM-producing human neoplastic B cell line that
953 switches to IgG-producing cells. *Int Immunol,* 1**,** 631-42.
954 Smillie, C. S., Biton, M., Ordovas-Montanes, J., Sullivan, K. M., Burgin, G., Graham, D. B., Herbst, R. H., Rogel, N., Slyper,
955 M., Waldman, J.*, et al.* 2019. Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell,* 178**,**
956 714-730.e22.
957 Sohn, H. W., Gu, H. & Pierce, S. K. 2003. Cbl-b negatively regulates B cell antigen receptor signaling in mature B cells through
958 ubiquitination of the tyrosine kinase Syk. *J Exp Med,* 197**,** 1511-1524.
959 Soneson, C., Love, M. & Robinson, M. 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level
960 inferences [version 2; peer review: 2 approved]. *F1000Research,* 4.
961 Stavnezer, J. & Schrader, C. E. 2014. IgH Chain Class Switch Recombination: Mechanism and Regulation. *J Immunol,* 193**,**
962 5370-5378.
963 Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P. &
964 Satija, R. 2019. Comprehensive Integration of Single-Cell Data. *Cell,* 177**,** 1888-1902.e21.
965 Suan, D., Kräutler, N. J., Maag, J. L. V., Butt, D., Bourne, K., Hermes, J. R., Avery, D. T., Young, C., Statham, A., Elliott, M.*,*
966 *et al.* 2017a. CCR6 Defines Memory B Cell Precursors in Mouse and Human Germinal Centers, Revealing Light-Zone
967 Location and Predominant Low Antigen Affinity. *Immunity,* 47**,** 1142-1153.e4.
968 Suan, D., Sundling, C. & Brink, R. 2017b. Plasma cell and memory B cell differentiation from the germinal center. *Curr Opin*
969 *Immunol,* 45**,** 97-102.
970 Suzuki, K., Kumanogoh, A. & Kikutani, H. 2008. Semaphorins and their receptors in immune cell interactions. *Nat Immunol,*
971 9**,** 17-23.
972 Thai, T.-H., Calado, D. P., Casola, S., Ansel, K. M., Xiao, C., Xue, Y., Murphy, A., Frendewey, D., Valenzuela, D., Kutok, J.
973 L.*, et al.* 2007. Regulation of the Germinal Center Response by MicroRNA-155. *Science,* 316**,** 604-608.
974 Thorarinsdottir, K., Camponeschi, A., Cavallini, N., Grimsholm, O., Jacobsson, L., Gjertsson, I. & Mårtensson, I. L. 2016.
975 CD21(-/low) B cells in human blood are memory cells. *Clin Exp Immunol,* 185**,** 252-262.
976 Todd, D. J., Mcheyzer-Williams, L. J., Kowal, C., Lee, A. H., Volpe, B. T., Diamond, B., Mcheyzer-Williams, M. G. & Glimcher,
977 L. H. 2009. XBP1 governs late events in plasma cell differentiation and is not required for antigen-specific memory B cell
978 development. *J Exp Med,* 206**,** 2151-9.
979 Toellner, K. M., Gulbranson-Judge, A., Taylor, D. R., Sze, D. M. & Maclennan, I. C. 1996. Immunoglobulin switch transcript
980 production in vivo related to the site and time of antigen-specific B cell activation. *J Exp Med,* 183**,** 2303-12.
981 Tybulewicz, V. L. J. & Henderson, R. B. 2009. Rho family GTPases and their regulators in lymphocytes. *Nat Rev Immunol,* 9**,**
982 630-644.
983 Van Keimpema, M., Grüneberg, L. J., Mokry, M., Van Boxtel, R., Van Zelm, M. C., Coffer, P., Pals, S. T. & Spaargaren, M.
984 2015. The forkhead transcription factor FOXP1 represses human plasma cell differentiation. *Blood,* 126**,** 2098-2109.
985 Vander Heiden, J. A., Yaari, G., Uduman, M., Stern, J. N. H., O'connor, K. C., Hafler, D. A., Vigneault, F. & Kleinstein, S. H.
986 2014. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires.
987 *Bioinformatics,* 30**,** 1930-1932.
988 Vento-Tormo, R., Efremova, M., Botting, R. A., Turco, M. Y., Vento-Tormo, M., Meyer, K. B., Park, J.-E., Stephenson, E.,
989 Polański, K., Goncalves, A.*, et al.* 2018. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature,*
990 563**,** 347-353.
991 Vigorito, E., Perks, K. L., Abreu-Goodger, C., Bunting, S., Xiang, Z., Kohlhaas, S., Das, P. P., Miska, E. A., Rodriguez, A.,
992 Bradley, A.*, et al.* 2007. microRNA-155 Regulates the Generation of Immunoglobulin Class-Switched Plasma Cells.
993 *Immunity,* 27**,** 847-859.
994 Welte, S., Kuttruff, S., Waldhauer, I. & Steinle, A. 2006. Mutual activation of natural killer cells and monocytes mediated by
995 NKp80-AICL interaction. *Nat Immunol,* 7**,** 1334-1342.

996 Willis, S. N., Good-Jacobson, K. L., Curtis, J., Light, A., Tellier, J., Shi, W., Smyth, G. K., Tarlinton, D. M., Belz, G. T., Corcoran,
997 L. M.*, et al.* 2014. Transcription factor IRF4 regulates germinal center cell formation through a B cell-intrinsic mechanism.
998 *J Immunol,* 192**,** 3200-6.
999 Wolf, F. A., Angerer, P. & Theis, F. J. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol,*
1000 19**,** 15.
1001 Wolock, S. L., Lopez, R. & Klein, A. M. 2019. Scrublet: Computational Identification of Cell Doublets in Single-Cell
1002 Transcriptomic Data. *Cell Systems,* 8**,** 281-291.e9.
1003 Wu, H. J. & Bondada, S. 2009. CD72, a coreceptor with both positive and negative effects on B lymphocyte development and
1004 function. *J Clin Immunol,* 29**,** 12-21.
1005 Xu, J., Husain, A., Hu, W., Honjo, T. & Kobayashi, M. 2014a. APE1 is dispensable for S-region cleavage but required for its
1006 repair in class switch recombination. *PNAS,* 111**,** 17242-17247.
1007 Xu, Y., Xu, L., Zhao, M., Xu, C., Fan, Y., Pierce, S. K. & Liu, W. 2014b. No receptor stands alone: IgG B-cell receptor intrinsic
1008 and extrinsic mechanisms contribute to antibody memory. *Cell Research,* 24**,** 651-664.
1009 Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. 2013. IgBLAST: an immunoglobulin variable domain sequence analysis tool.
1010 *Nucleic Acids Res,* 41**,** W34-W40.
1011 Zaretsky, I., Atrakchi, O., Mazor, R. D., Stoler-Barak, L., Biram, A., Feigelson, S. W., Gitlin, A. D., Engelhardt, B. & Shulman,
1012 Z. 2017. ICAMs support B cell interactions with T follicular helper cells and promote clonal selection. *J Exp Med,* 214**,**
1013 3435-3448.
1014 Zheng, S., Kusnadi, A., Choi, J. E., Vuong, B. Q., Rhodes, D. & Chaudhuri, J. 2019. NME proteins regulate class switch
1015 recombination. *FEBS Lett,* 593**,** 80-87.
1016 Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C. & Chanda, S. K. 2019. Metascape
1017 provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun,* 10**,** 1523.

1018

Supplementary Figures

Figure S1

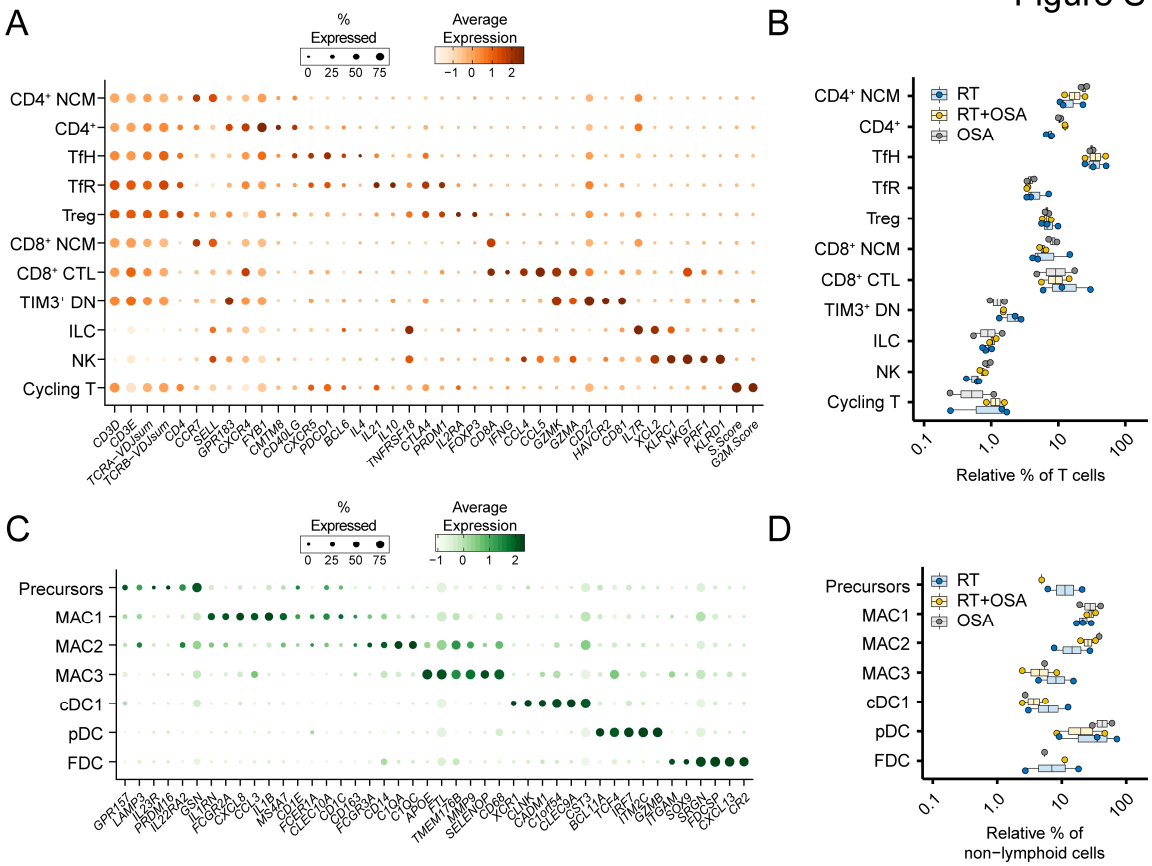

Figure S1. Annotation of non-B cell populations in the human tonsils.

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, in addition to innate lymphoid cells (ILC) and natural killer (NK) cells. Frequency of cells for which each gene is detected is denoted by size of the dots.

B) Relative frequencies of different T cell subsets separated by clinical indication for tonsillectomy. OSA; obstructive sleep apnoea ($n = 2$), RT; recurrent tonsillitis ($n = 3$), RT+OSA ($n = 2$).

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. Frequency of cells for which each gene is detected is denoted by size of the dots.

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

Figure S2

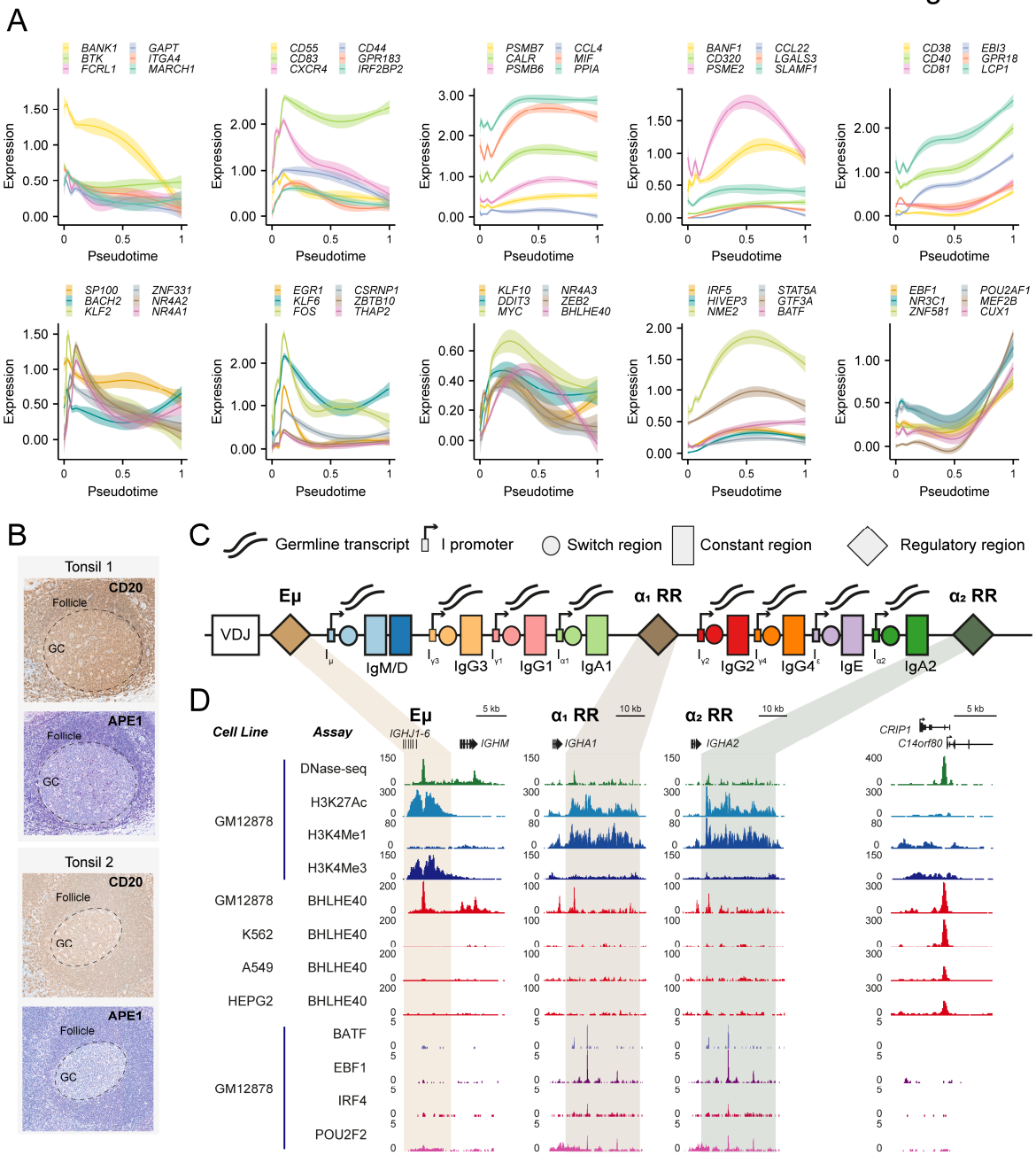

Figure S2. Dynamic gene expression during class switch recombination and transcription factor binding at the immunoglobulin locus.

A) Smoothed gene expression for example cell surface receptor or cytokine (top row) and transcription factor (bottom row) genes that are differentially regulated 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) 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).

D) Open chromatin (DNase-seq) and ChIP-seq from ENCODE consortium 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).

Figure S3

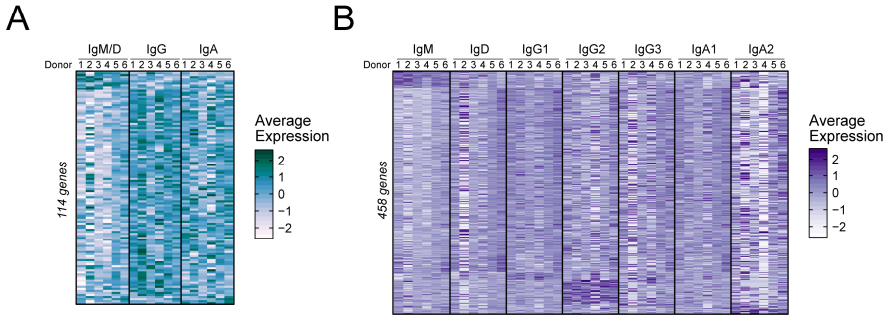

Figure S3. Class- and subclass-specific gene expression analyses.

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 S4

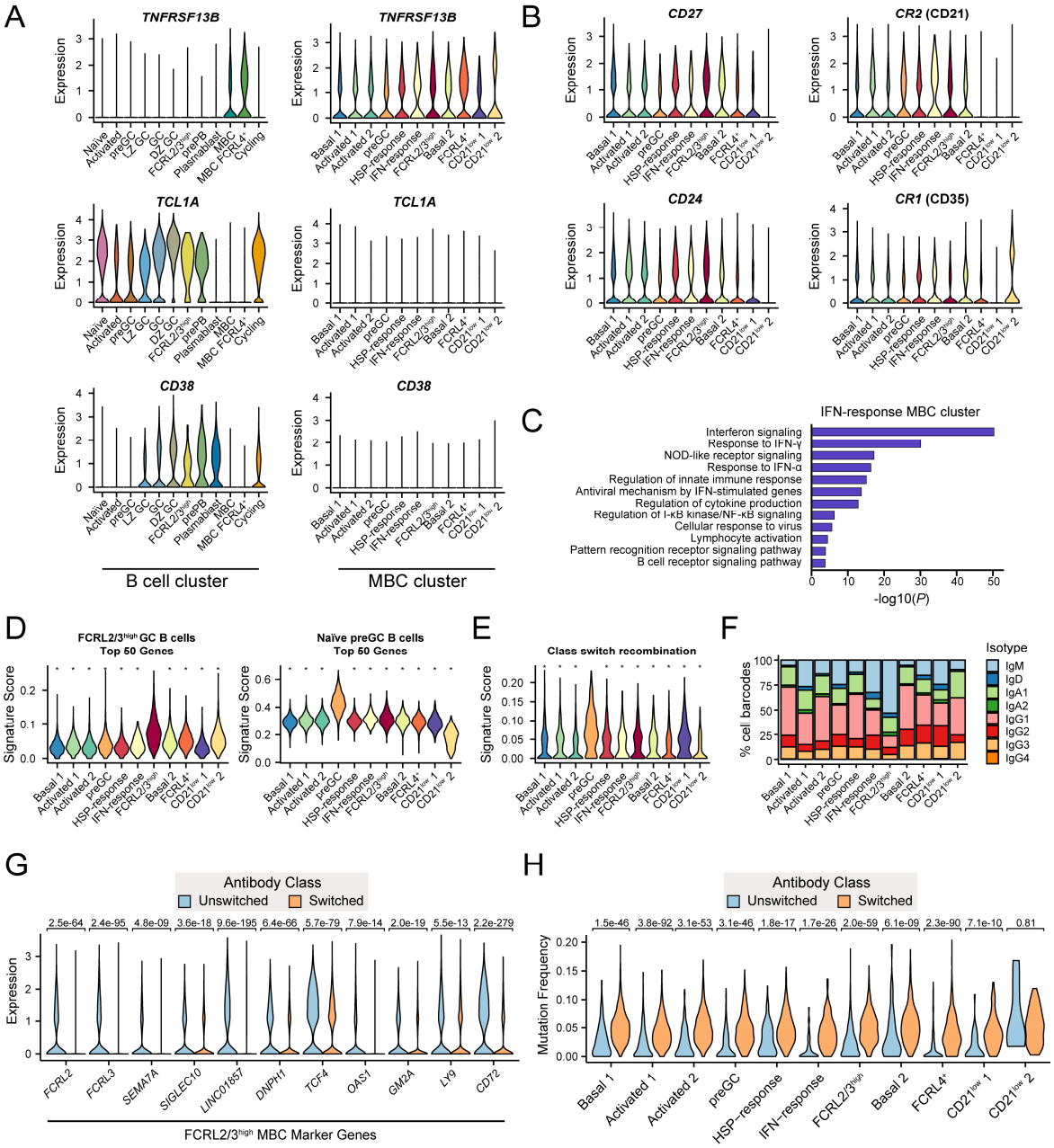

Figure S4. Characterisation of memory B cell states identified by scRNA-seq.

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 FCRL2/3^{high} GC B cells in MBC subsets. * denotes p value < 0.001.

E) Single-cell AUCell-derived scores for class switch recombination gene set in MBC subsets. * denotes p value < 0.001.

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

G) Single-cell expression of key marker genes of the FCRL2/3^{high} B cell states between switched and unswitched MBCs.

H) Somatic hypermutation frequencies of scVDJ-derived antibody genes between switched and unswitched B cells within different MBC populations.