Induction of bronchus-associated lymphoid tissue is an early life adaptation for promoting human B cell immunity

Rei Matsumoto[1,12], Joshua Gray [2,12], Ksenia Rybkina[2], Hanna Oppenheimer[3,4], Lior Levy[3,4], Lilach M. Friedman [3,4], Muhammad Khamaisi[5], Wenzhao Meng[6], Aaron M. Rosenfeld [6], Rebecca S. Guyer[2], Marissa C. Bradley[6], David Chen [7], Mark A. Atkinson[8], Todd M. Brusko [8], Maigan Brusko[8], Thomas J. Connors [9], Eline T. Luning Prak [6], Uri Hershberg[5], Peter A. Sims[7,10], Tomer Hertz[3,4,11] & Donna L. Farber [1,2] ✉

Infants and young children are more susceptible to common respiratory pathogens than adults but can fare better against novel pathogens like severe acute respiratory syndrome coronavirus 2. The mechanisms by which infants and young children mount effective immune responses to respiratory pathogens are unknown. Through investigation of lungs and lung-associated lymph nodes from infant and pediatric organ donors aged 0–13 years, we show that bronchus-associated lymphoid tissue (BALT), containing B cell follicles, CD4+ T cells and functionally active germinal centers, develop during infancy. BALT structures are prevalent around lung airways during the first 3 years of life, and their numbers decline through childhood coincident with the accumulation of memory T cells. Single-cell profiling and repertoire analysis reveals that early life lung B cells undergo differentiation, somatic hypermutation and immunoglobulin class switching and exhibit a more activated profile than lymph node B cells. Moreover, B cells in the lung and lung-associated lymph nodes generate biased antibody responses to multiple respiratory pathogens compared to circulating antibodies, which are mostly specific for vaccine antigens in the early years of life. Together, our findings provide evidence for BALT as an early life adaptation for mobilizing localized immune protection to the diverse respiratory challenges during this formative life stage.

Infants and young children lack preexisting T and B cell immunity to common respiratory pathogens, which develops over repeated exposures to provide long-term immune protection. As a result, young children are particularly susceptible to ubiquitous respiratory pathogens and develop frequent respiratory illness during the early years of life. However, most children recover fully from a wide range of newly encountered respiratory infections and even novel pathogens, such as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), without serious complications[1–3] through mechanisms that are poorly understood. Infants generate adaptive immune responses to respiratory pathogens starting in the early months of life and require these de novo responses for survival, as babies with severe
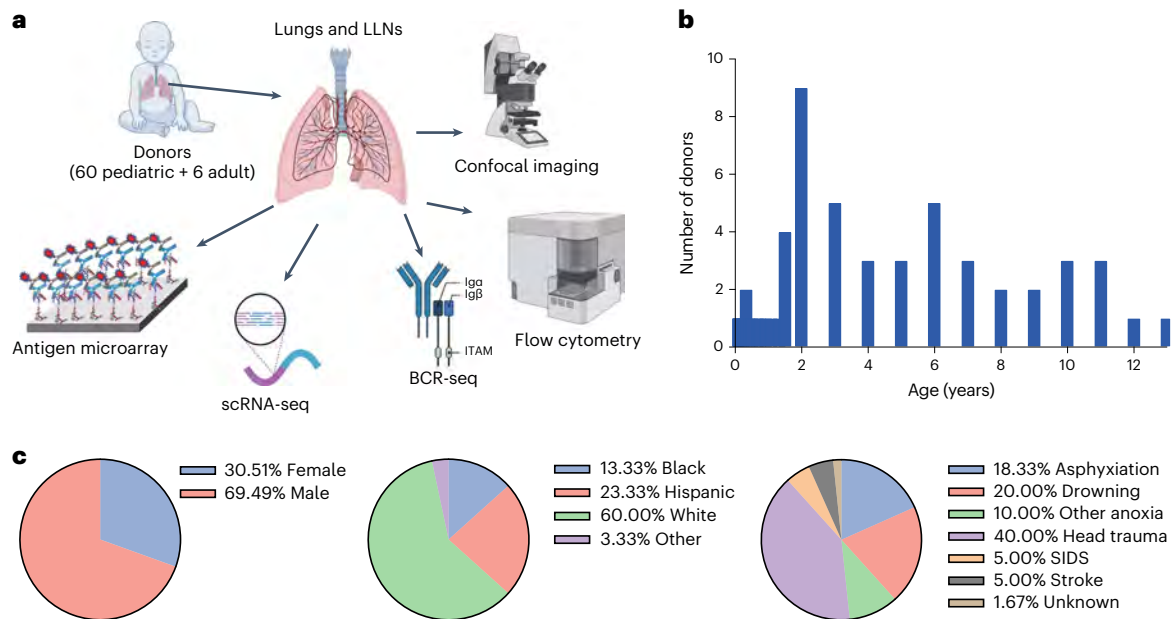
Fig. 1 | Pediatric organ donors and their characteristics. a, Study schematic. Lungs and LLNs were obtained from deceased pediatric organ donors (n = 60) aged 0–13 years. Samples were processed as tissue sections for analysis by immunofluorescence imaging and as single-cell suspensions for high-dimensional flow cytometry, BCR-seq, scRNA-seq and assays for antibody specificity using antigen array analysis. **b,c**, Donor demographics shown as age distribution, reported sex (female n = 18; male n = 41; unknown n = 1), ethnicity/race (African American, n = 8, blue; Hispanic, n = 14, red; white, n = 36, green; American Indian, n = 1, purple) or cause of death (asphyxiation, n = 11, blue; drowning, n = 12, red; other anoxia, n = 6, green; head trauma, n = 24, purple; sudden infant death syndrome (SIDS), n = 3, pink; stroke, n = 3, black; unknown, n = 1, brown).

combined immunodeficiencies (lacking T and B lymphocytes) or with impairments in helper T cell–B cell interactions develop life-threatening lung infections[4,5]. The distinct adaptations and mechanisms by which children develop protective responses in the respiratory tract over the crucial years of infancy and childhood have eluded investigation due to sampling issues and other practical and logistical challenges.

In mice, adaptive immune responses to respiratory pathogens are initiated in lung-associated lymph nodes (LLNs); antigen-specific T cells are primed and differentiate to effector cells, and antigen-specific B cells differentiate to antibody-secreting plasma cells or memory B cells through interactions with helper T cells in specialized structures called germinal centers (GCs)[6,7]. In humans, these tissue-priming events are challenging to sample and study; limited studies have shown that GC formation can be observed in LN aspirates in response to vaccines injected at a known time point[8–10]. However, little is known regarding the genesis of human adaptive immunity to natural exposures in tissues, particularly during the critical early years of life when newly encountered antigens predominate.

In this study, we analyzed the development of respiratory immunity in human lungs and LLNs across infancy and childhood through access to samples from organ donors aged 0–10 years old. We identify that bronchus-associated lymphoid tissue (BALT) develops in the lungs during the first year of life, reaching peak densities at 2–3 years of age and declining in density and follicular integrity starting at 4–5 years of age. Early life BALT supports GC formation, B cell differentiation, class switching and somatic hypermutation (SHM). High-dimensional profiling reveals distinct features of lung B cells that are more activated and GC-like than LN B cells, suggesting in situ activation. Importantly, during the time of peak BALT formation, lung and LLN B cells produce antibodies with biased specificities for respiratory pathogens distinct from the reactivity of circulating antibodies. Together, our results reveal compartmentalization of lung B cell immunity during the early years of life through establishment of ectopic lymphoid structures to augment lymphoid responses in the critical years of high pathogen load.

Results

We have previously investigated human tissue immunity in samples obtained from organ donors of all ages[11–13], enabling the analysis of immune responses at sites where they function and are maintained. For analysis of respiratory immune development from birth to child-hood, we obtained lungs and LLNs from 63 pediatric organ donors aged 0–13 years procured through the Human Atlas of Neonatal Development and Early Life Immunity (HANDEL-I) program, which maintains collaborations with organ procurement organizations throughout the United States. This pediatric cohort, composed of 60 donors who died of non-infectious causes and 3 donors who died with ongoing symptomatic respiratory infection, represented a racially and ethnically diverse population (Fig. 1a–c and Supplementary Table 1). In some cases, we included adult samples for reference (Supplementary Table 2). Tissues were analyzed by confocal imaging and processed to single-cell suspensions (Methods) for high-dimensional flow cytometry, single-cell transcriptional profiling, B cell receptor (BCR) sequencing (BCR-seq) and antigen specificity assays (Fig. 1a).

BALT induction and GC formation during early life

Lymphocyte accumulation and localization within the lungs during infancy and childhood were assessed by confocal imaging. In the early weeks of life, there were sparse lymphocytes in the lung (Fig. 2a). By 6 months of age, we observed lymphocyte clusters around the major airways consisting of large foci of CD20[+] B cells surrounded by CD4[+] T cells exhibiting memory (CD45RO[+]) phenotypes (Fig. 2a,b). These follicular structures, which were most prominent by 2–3 years of age (Fig. 2b), were situated between major airways and the adjacent pulmonary artery and resembled BALT. In humans, BALT is associated with lung injury or chronic inflammation in adults and is not present in healthy lungs[14] while in mice, BALT is induced by respiratory infection[15] and inhaled lipopolysaccharide (LPS)[16]. Lung imaging and quantitation revealed coordinate dynamics of BALT and lung B cells (mostly localized in BALT) over early life and childhood; there was a high proportion
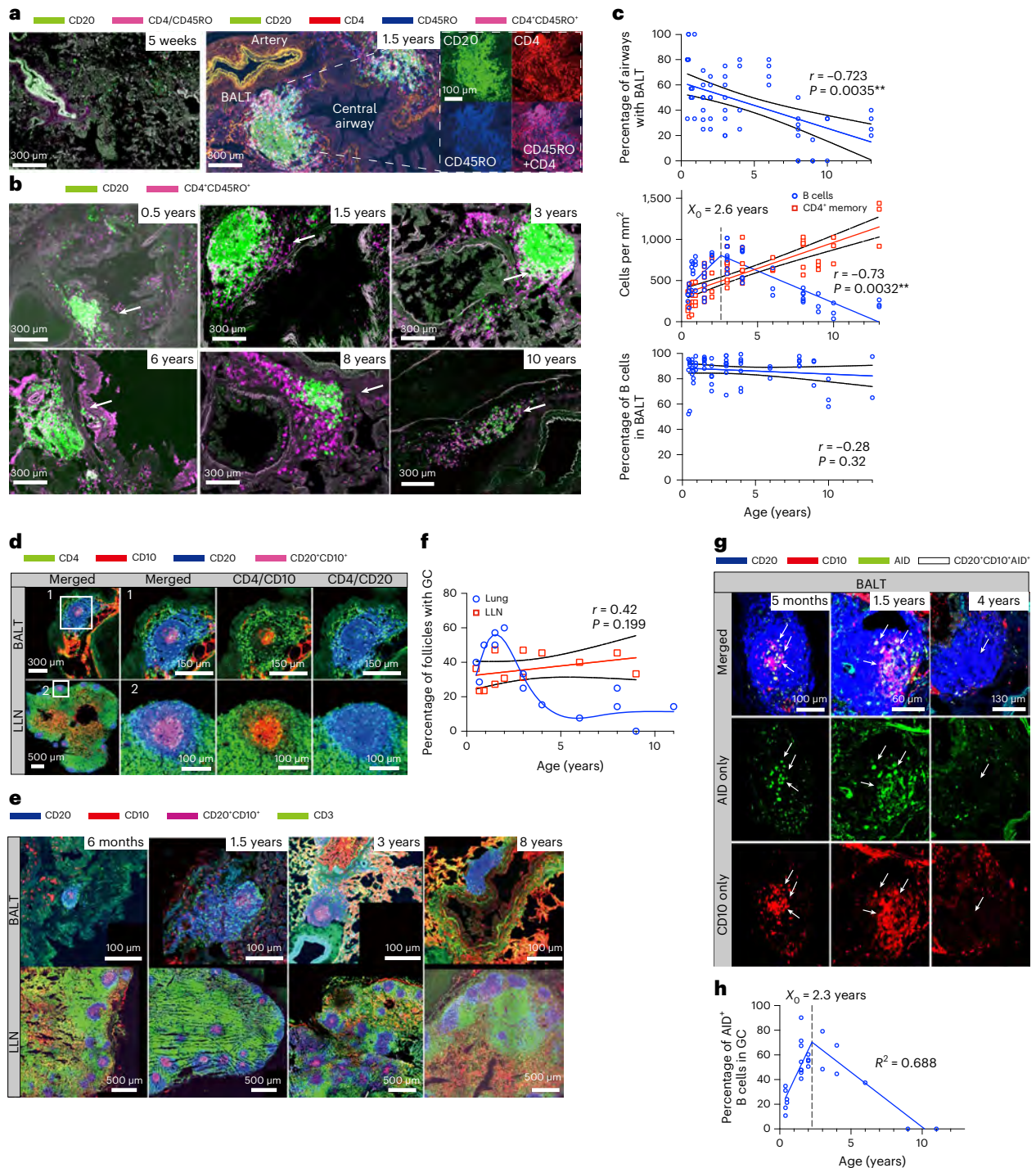
Fig. 2 | Formation and regression of BALT containing GCs during infancy and childhood. a, Confocal imaging of lungs from pediatric donors showing immunostaining with antibodies to CD20 (green), CD4 (red), CD45RO (blue) and CD4/CD45RO (purple, white arrow) in an infant donor (left) and 1.5-year-old donor showing BALT near airways (right). **b**, Lung imaging as in **a** in donors of indicated ages (white arrows indicate BALT). **c**, Quantitative analysis of BALT and lymphocyte content over age compiled from 21 donors. Top, percentage of central airways containing BALT from two to three sections per donor (95% confidence interval (95% CI) of −0.91 to −0.31, d.f. = 19). Middle, dynamics of B cell and memory CD4+ T cell content in lungs throughout childhood. Data are shown as cells per mm^2 from three to five sections per donor (memory T cells: 95% CI of 0.82 to 0.98, d.f. = 63; B cells: R^2 = 0.58, d.f. = 60). Bottom, proportion of B cells within BALT over age from three to five sections per donor (95% CI of −0.71 to 0.29, d.f. = 63). A linear regression analysis shows the line of best fit (colored line) and 95% confidence bands (black lines). For B cell dynamics, segmental

linear regression was best fitted; dotted lines show the intersection of two linear regressions (X_0). **d**,**e**, GC content in lungs and LLNs of a representative 1.5-year-old donor (**d**) and donors of indicated ages (**e**) shown by staining with anti-CD3 (green), anti-CD20 (blue) and anti-CD10 (red). GC B cells (CD10+CD20+) appear as pink. **f**, Dynamics of GC formation in lungs and LLNs over childhood shown as proportion of B cell follicles containing GCs calculated as mean percentage from three to five sections per donor (n = 12). A linear regression analysis shows the line of best fit (colored line) and 95% confidence bands (black lines) for GCs in LLNs (R^2 = 0.64; 95% CI of −0.24 to 0.81; d.f. = 10), while a spline was best fit for GC dynamics within BALT. **g**, AID expression (green) by GC B cells (CD20+CD10+AID+) from three representative donors of indicated ages. **h**, Frequency of AID+ B cells (two to three sections per donor) within GCs over age (n = 12). A segmental linear regression was best fit (R^2 = 0.688; d.f. = 20); r values were generated by Pearson's correlation analysis, and significance was determined by two-tailed test.

of central airways containing BALT and an increase in B cell numbers during the first 3 years of life and a significant decline in BALT and B cells after 4–5 years of age (Fig. 2c). Conversely, the number of memory CD4[+] T cells around airways increased progressively over childhood (Fig. 2c, middle). These results show a high level of organized BALT during the critical window of early life that subsequently diminishes over age concomitant with the accrual of lung memory T cells.

To assess the functionality of B cell follicles in BALT, we investigated GC formation within BALT and the associated LLNs over age. Staining with CD10, a surface glycoprotein expressed by GC B cells[17], showed a GC-like structure in the center of multiple B cell follicles within BALT and LLNs (Fig. 2d). Notably, GCs were detected in BALT at 1–3 years of age and declined precipitously thereafter with minimal GCs in BALT after 5 years of age, while GC content remained stable within LLN follicles throughout childhood (Fig. 2e,f). We confirmed the presence of functional GCs in BALT and LLNs by staining for activation-induced cytidine deaminase (AID), an enzyme required for BCR SHM and class switch recombination in GCs[18,19]. AID was expressed by CD10[+] B cells in BALT of infants and young children, with a significant reduction in AID[+] B cells after 5 years of age (Fig. 2g,h). GC structures identified by CD10 (and AID) expression in BALT and LLNs also contained follicular dendritic cells (Extended Data Fig. 1a–c), another key feature of functionally active GCs[20,21]. Therefore, early life BALT exhibits multiple features of active GCs, while later in childhood, GCs are found predominantly in LLNs.

Site-specific B cell subset composition over childhood

To further investigate the dynamic changes in B cell immunity in the lung compared to LLNs during early life and childhood, we analyzed B cell subsets representing major stages in development and differentiation, including transitional, naive, memory and GC B cells, by flow cytometry (Fig. 3a and Extended Data Fig. 2). Overall, the lung had significantly lower frequencies of total B cells (CD19[+]) and naive B cells but contained greater proportions of transitional, GC and IgA[+] class-switched memory B cells than LLNs (Fig. 3b). In addition to compositional differences, certain subsets expressed site-specific phenotypes (Fig. 3c). Naive B cells expressed higher levels of the follicular marker CXCR5 (ref. 22) in LLNs than those in the lungs, while memory B cells in the lung expressed higher levels of CD95 and CD69 (denoting activation and tissue residency, respectively[23,24]) than those in LLNs (Fig. 3c). These results suggest a more activated, tissue-adapted state for lung B cells.

Over age (up to 10 years), B cell composition changed in each site, albeit with different dynamics. In the lung, the frequency of B cells decreased with age (consistent with imaging; Fig. 2c), while in the LLNs, B cell frequency increased progressively over childhood (Fig. 3d, left). The frequency of transitional B cells was higher in the lung than in the LLNs during infancy but declined in both sites with age, reaching negligible levels by 8 years of age, while naive B cell frequency remained stable in both sites (Fig. 3d, left). For subsets generated from antigen-driven differentiation, there was a marked decline in GC B cells in the lung between infants and older children, while GC B cell frequency remained stable in the LLNs throughout childhood (Fig. 3d, right), consistent with the imaging results. In addition, the frequency of IgG and IgA class-switched memory B cells increased significantly over childhood in both sites, although lungs contained higher frequencies of IgA[+] memory B cells than LLNs (Fig. 3d, right). These findings show that age-associated dynamics of B cell differentiation and memory formation are site specific.

Follicular helper T (T_FH) cells in the lung and LLNs over childhood

Another component critical to the induction of adaptive immunity and the function of GCs are T_FH cells, which exhibit a CD4[+]CD45RA[–]PD-1[+]CXCR5[+] phenotype and express the lineage-defining transcription factor BCL-6 (ref. 6). In the lung, tissue T_FH-like cells were recently

identified in mice following influenza infection; however, they were distinct from LN T_FH cells due to decreased expression of CXCR5 (refs. 25,26). To assess whether GC-containing follicles in pediatric BALT also contained T_FH cells, we used high-dimensional flow cytometry to identify features of T_FH cells in LLNs and lungs of infants and children.

In LLNs, we identified the two major populations of T_FH cells based on PD-1 and CXCR5 expression: PD-1[hi]CXCR5[+] cells (GC T_FH) and PD-1[int]CXCR5[+] cells (conventional T_FH), each representing ~15% of the CD4[+] memory T cell population in children (Fig. 4a). By contrast, lung CD4[+] T cells had a substantial population of PD-1[hi] cells but negligible CXCR5 expression (Fig. 4a), similar to lack of CXCR5 expression by lung B cells (Fig. 3). Because of the lack of CXCR5 on lung T cells, we examined expression of other T_FH markers by PD-1[hi] compared to PD-1[lo] memory CD4[+] T cells. The PD-1[hi]CXCR5[+] subset of cells in LLNs expressed significantly higher levels of BCL-6 than PD-1[lo] (non-T_FH) cells, PD-1[int]CXCR5[+] T_FH cells and PD-1[hi]CXCR5[–] cells (Fig. 4b, bottom). However, PD-1[hi] and PD-1[lo] memory CD4[+] T cells in the lung showed equivalent BCL-6 expression (Fig. 4b, top), prompting examination of additional markers to define a T_FH-like counterpart in human lungs. PD-1[hi]CXCR5[+] cells in LLNs expressed significantly greater proportions of ICOS and TIGIT known to promote T cell–B cell interactions within follicular structures[6,27] (Fig. 4c). Lung PD-1[hi] cells also expressed significantly increased proportions of these two markers compared to PD-1[lo] cells (Fig. 4c). Moreover, PD-1[hi]ICOS[+]TIGIT[+]CD4[+] T cells in the lung expressed higher levels of BCL-6 than PD-1[lo] and PD-1[hi]ICOS[–]TIGIT[–] cells in both the lung and LLNs (Fig. 4d). Importantly, the levels of BCL-6 expression in LLNs were equivalent between PD-1[hi]ICOS[+]TIGIT[+] cells and canonically defined GC T_FH (PD-1[hi]CXCR5[+]) cells, indicating that coexpression of ICOS and TIGIT is a surrogate marker for T_FH cells in peripheral, non-lymphoid sites such as the lungs (Fig. 4e).

Over childhood, the proportion of PD-1[hi]ICOS[+]TIGIT[+] cells in the lung significantly declined with age (Fig. 4f), coincident with the age-associated reduction in BALT structures and GC B cells with age. By contrast, the frequency of GC T_FH cells in LLNs remained stable over childhood (Fig. 4f), similar to the stable proportions of GC-containing follicles and frequencies of GC B cells in LLNs. Together, these data reveal the presence of T_FH-like cells in the lung during the early years of life coincident with the timing of functional BALT, suggesting their role in promoting adaptive immunity in the lungs.

Activated, GC-like transcriptional profiles of lung B cells

To assess the connection between B cell differentiation, class switching and BCR repertoire in the two sites, we performed single-cell transcriptome profiling by single-cell RNA sequencing (scRNA-seq) of B cells in the lungs and LLNs of three donors at the peak age of BALT formation (1–3 years). The scRNA-seq data revealed consistent results between donors and distinct features of B cells in the lung compared to LLNs (Fig. 5a). Each site contained transitional, naive, memory and GC B cells based on differential gene expression (Fig. 5a,b, Extended Data Fig. 4 and Supplementary Table 3) and consistent with flow cytometry results. Transitional and naive B cells both expressed *IGHD*, *IGHM*, *TCL1A* and *IGLL5*, while *CD38*, *MME* (*CD10*), *IL7R* and *SOX4* were specifically expressed by transitional subsets[28]. Memory B cells expressed the memory markers *CD27* and *TNFRSF13B* (*TACI*)[29] and were defined as class switched based on expression of *IGHG1* or *IGHA1*. GC B cells expressed lineage-defining genes *BCL6*, *AICDA* (encoding AID), *MME* and *MKI67* (Fig. 5b and Extended Data Fig. 4a). There were higher proportions of transitional B cells in the lung than in LLNs and more naive B cells in LLNs than in the lungs (Fig. 5c), consistent with flow cytometry data.

Differential gene expression analysis between lung and LLN B cells for both naive/transitional and memory subsets (Fig. 5d,e, Extended Data Fig. 4b and Supplementary Tables 4 and 5) revealed lung-specific genes associated with activation (*NR4A1*, *NR4A2*, *NR4A3* (ref. 30) and *CD83* (ref. 31)), tissue residency (*CD69*, *EZR* and *KLF6*)[32] and GC function (*BCL6* and *LY9* (ref. 33)). Fewer genes were specifically upregulated by
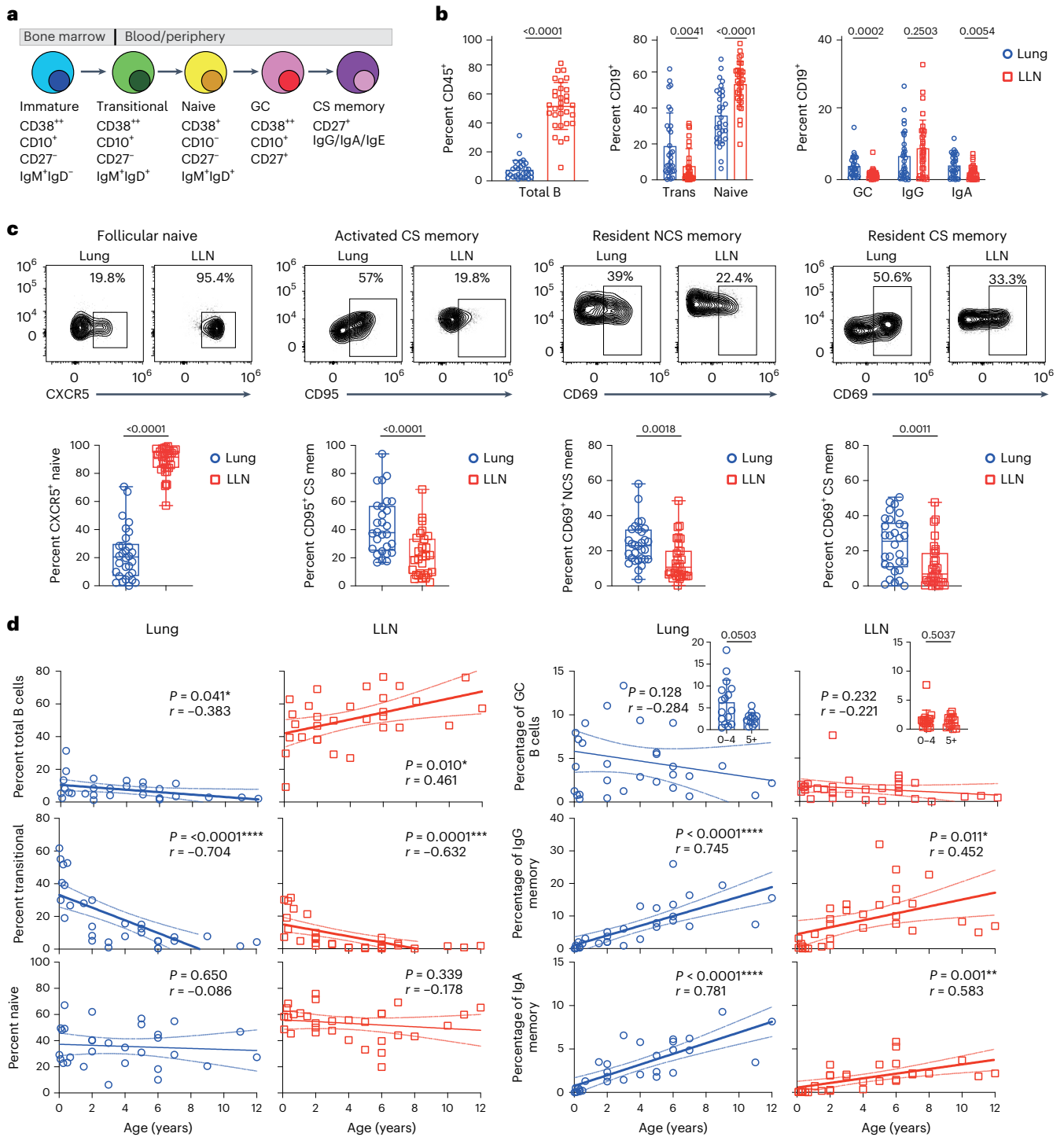
Fig. 3 | Distinct B cell subset composition and age-associated dynamics in lungs and LLNs. High-dimensional flow cytometry was performed on B cells isolated from lungs and LLNs from donors 0–13 years of age (*n* = 35). **a**, Schematic showing the major B cell subsets, their maturation state and distinguishing markers. **b**, Distinct proportion of B cells (among CD45⁺ cells) and composition of B cell subsets (among CD19⁺ cells) in lungs and LLNs compiled from 35 donors. Significance was determined by two-tailed *t*-test (significance cutoff of ≤0.05). Error bars indicate s.d. **c**, Differential expression of markers for follicular subsets (CXCR5), activation (CD95) and tissue residence (CD69) by indicated B cell subsets in lungs and LLNs shown in representative flow plots from individual donors (top) and box plots of compiled data from 35 donors (bottom). The midlines of box plots in **c** represent the median, the boxes represent the interquartile range, and the whiskers represent the minimum and maximum values. Significance was determined by two-tailed *t*-test (significance cutoff of

≤0.05). **d**, Distinct age-associated changes for B cell subsets in lungs (first and third columns, *n* = 29) and LLNs (second and fourth columns, *n* = 30) over the first 12 years of life. A linear regression analysis shows the line of best fit (solid line) and 95% confidence bands (dotted lines). The *r* and *P* values were generated by Pearson's correlation analysis; total B cells lung: 95% CI of −0.66 to −0.02, d.f. = 28; total B cells LLN: 95% CI of 0.12 to 0.70, d.f. = 29; transitional B cells lung: 95% CI of −0.85 to −0.46, d.f. = 28; transitional B cells LLN: 95% CI of −0.81 to −0.36, d.f. = 29; naive B cells lung: 95% CI of −0.43 to 0.28, d.f. = 28; naive B cells LLN: 95% CI of −0.50 to 0.19, d.f. = 29; GC B cells lung: 95% CI of −0.53 to 0.16, d.f. = 28; GC B cells LLN: 95% CI of −0.53 to 0.14, d.f. = 29; IgG memory B cells lung: 95% CI of 0.53 to 0.87, d.f. = 28; IgG memory B cells LLN: 95% CI of 0.12 to 0.69, d.f. = 29; IgA memory B cells lung: 95% CI of 0.59 to 0.89, d.f. = 28; IgA memory B cells LLN: 95% CI of 0.29 to 0.78, d.f. = 29; CS, class switched; Mem, memory; NCS, non-class switched.
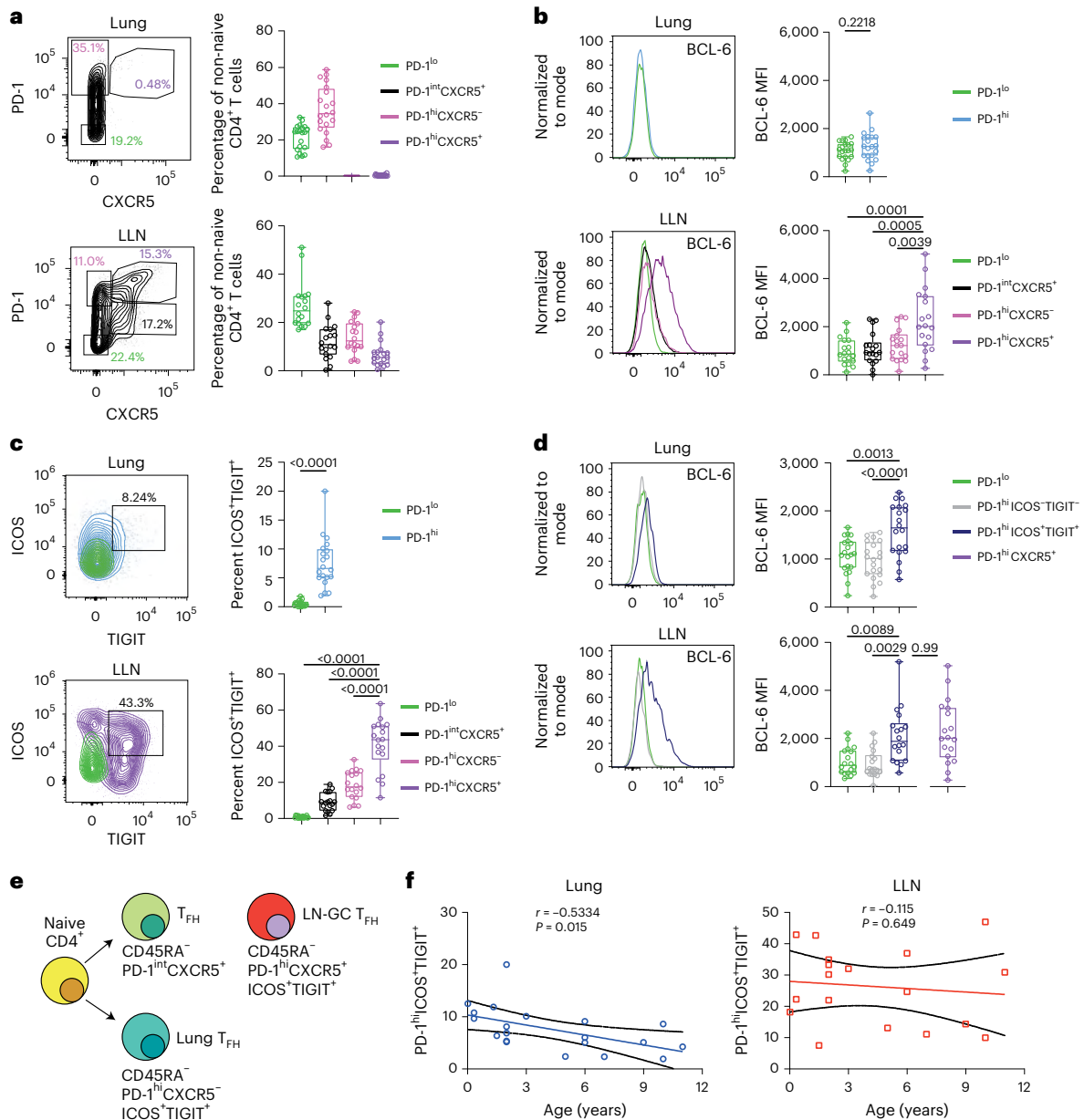
Fig. 4 | Analysis of helper T cell subsets in lungs and LLNs. T_{FH} cell content in lungs and LLNs. **a**, Representative flow cytometry plots from lung and LLN tissue from a 10-year-old old donor delineating subsets of non-naive (CD45RA⁻) CD4⁺ T cells based on PD-1 and CXCR5 expression (left). Box plots indicate the frequencies of the indicated subsets in the lung (top) and LLNs (bottom) compiled from 19 donors (PD-1^{lo}, green; PD-1^{int}CXCR5⁺, black; PD-1^{hi}CXCR5⁻, pink; PD-1^{hi}CXCR5⁺, purple). **b**, BCL-6 expression within CD4⁺ T cell subsets in the lung (top, stratified into PD-1^{lo} (green) versus PD-1^{hi} (light blue)) and LLNs (bottom, stratified based on PD-1 and CXCR5 expression as in **a**) shown in representative histograms (left) and box plots of BCL-6 mean fluorescence intensity (MFI; right). **c**, ICOS and TIGIT expression by subsets delineated in **b** from the lung (top) and LLNs (bottom) shown in representative flow cytometry plots (left) and box plots of the percentage of ICOS⁺TIGIT⁺ cells (right). **d**, BCL-6 expression by

PD-1 subsets differentially expressing ICOS and TIGIT in the lung (top) and LLNs (bottom) shown as representative histograms (left) and box plots of BCL-6 MFI (right). The box plot midlines in **a**–**d** represent the median, the boxes indicate the interquartile ranges, and the whiskers are the minimum and maximum values. Significance for **b**–**d** was determined by a two-tailed *t*-test (if two groups) or one-way analysis of variance and Tukey's multiple comparisons test (if three or more groups; significance cutoff of ≤0.05). **e**, Diagram showing distinct phenotypes of T_{FH}-like cells in LNs and lungs. **f**, Frequency of PD-1^{hi}ICOS⁺TIGIT⁺ cells in the lung and LLNs over age. Linear regression analysis showing the line of best fit (colored line) and 95% confidence bands (black lines); *r* and *P* values were generated by Pearson's correlation analysis; lung: 95% CI of −1.14 to −0.14, d.f. = 18; LLN: 95% CI of −2.1 to 1.4, d.f. = 16.

B cells in LLNs compared to in the lungs, such as the gene encoding complement receptor CR2 and the transcription factor ID3 with roles in GC formation[34] (Fig. 5e). The lung-specific profile was highly enriched for multiple genes associated with B cell activation not expressed by LLN B cells (Fig. 5f). Moreover, the subset composition and gene expression profiles of B cells in LLNs were strikingly similar to those in the mesenteric LN associated with intestines (Extended Data Fig. 5 and Supplementary Table 6), indicating common features for lymphoid

B cells and providing additional evidence for the distinct profile of lung-derived B cells.

We extracted BCR sequences from each donor and site, revealing a highly diverse clonal repertoire in the lung and LLNs (Fig. 5g). There were very few expanded clones overall, and clones representing two or more copies comprised <1% of the repertoire in the youngest (1.3 years old) donor and >1–2.5% in the 2- to 3-year-old donors (Fig. 5g). The majority of total BCR clones were derived from naive B cells in
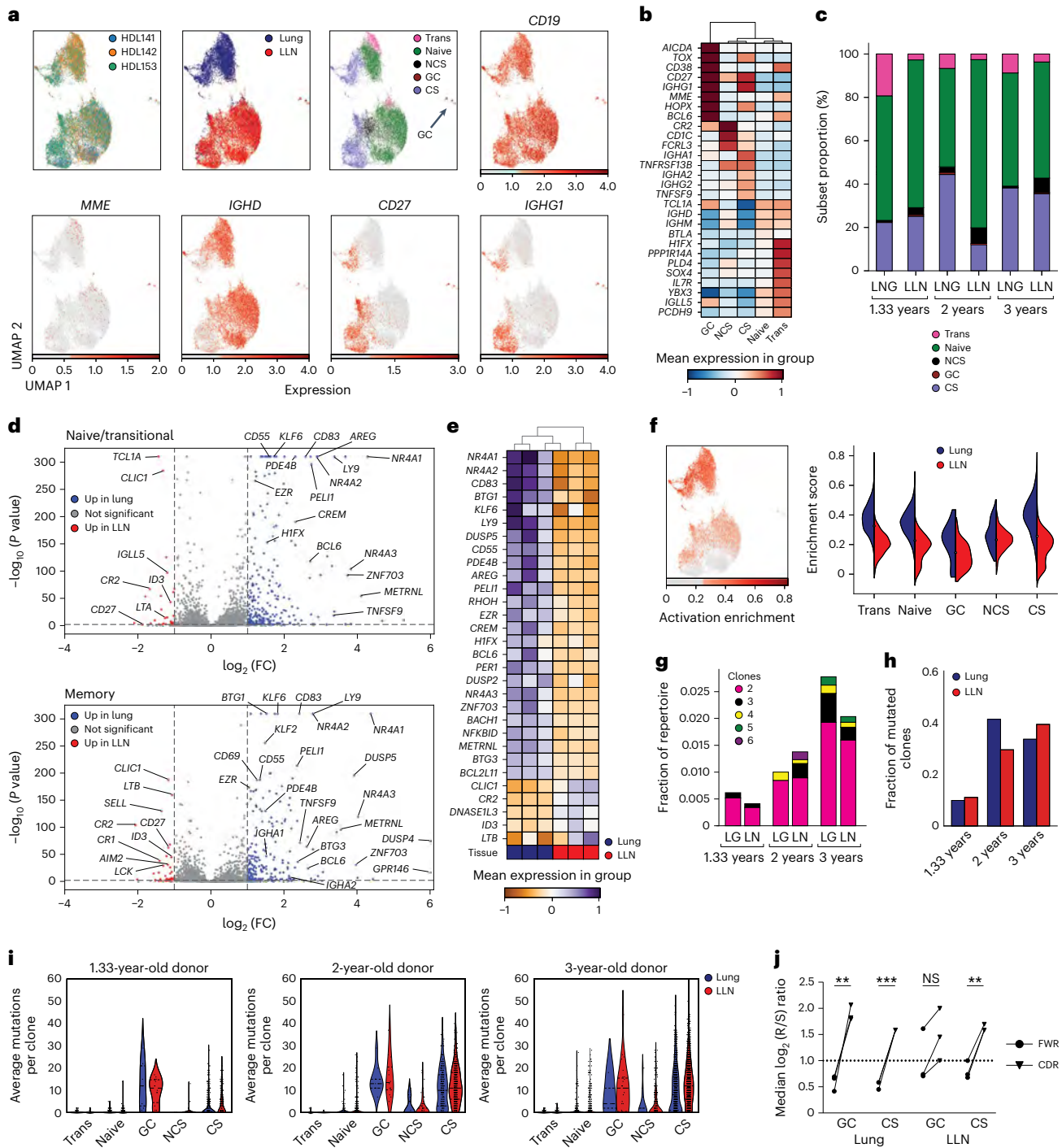
Fig. 5 | Transcriptional signatures and clonal dynamics for lung and LLN B cells in early life. Single-cell transcriptome and BCR-seq was conducted on 26,745 B cells from the lungs and LLNs of three donors aged 1.33, 2 and 3 years. **a**, Uniform manifold approximation and projection (UMAP) of all B cells colored by donor, tissue and subset; Trans, transitional; NCS, non-class switched memory; CS, class-switched memory. **b**, Heat map showing differentially expressed genes and their z-score expression within each subset across sites. **c**, Bar plot showing B cell subset proportions in each site of individual donors; LNG, lung. **d**, Volcano plots showing differentially expressed genes between the lung and LLNs when comparing combined naive and transitional B cells (top) and memory subsets (bottom). Red dots indicate genes upregulated in LLNs, and blue dots indicate those upregulated in the lung (minimum $\log_2$ (fold change) ($\log_2$ (FC)) > 1, maximum P value of <0.01). Significance was calculated using a two-sided Wilcoxon test with tie correction. **e**, Heat map showing the z-score expression of the consensus genes between B cell subsets in the lung and LLNs.

f, Left, UMAP of the enrichment scores of genes associated with B cell activation (Methods). Right, violin plot showing the distribution of enrichment scores of these genes within each subset for each organ. **g**, BCR analysis showing the fraction of clones with ≥2 cells per donor per tissue. **h**, Bar plot indicating the fraction of mutated clones (those with an average of ≥2% mutated positions in a sequence) per donor by site (lung, blue; LLN, red). **i**, Violin plots indicating the average number of mutations per clone for each subset and tissue (lung, blue; LLN, red) within each donor. Horizontal lines indicate interquartile range. **j**, Paired dot plot indicating the median distribution of the ratios of the average number of non-synonymous (R) and synonymous (S) mutations per clone within the FWR (circles) and CDR (triangles) of the BCR displayed by tissue and subset (GC and class-switched) in log scale. Statistical significance was determined by multiple t-tests using the Holm−Sidak method; **, lung GC adjusted $P = 0.001341$; ***, lung CS adjusted $P = 0.000081$; NS (not significant), LLN GC adjusted $P = 0.323825$; **, LLN CS adjusted $P = 0.003012$.
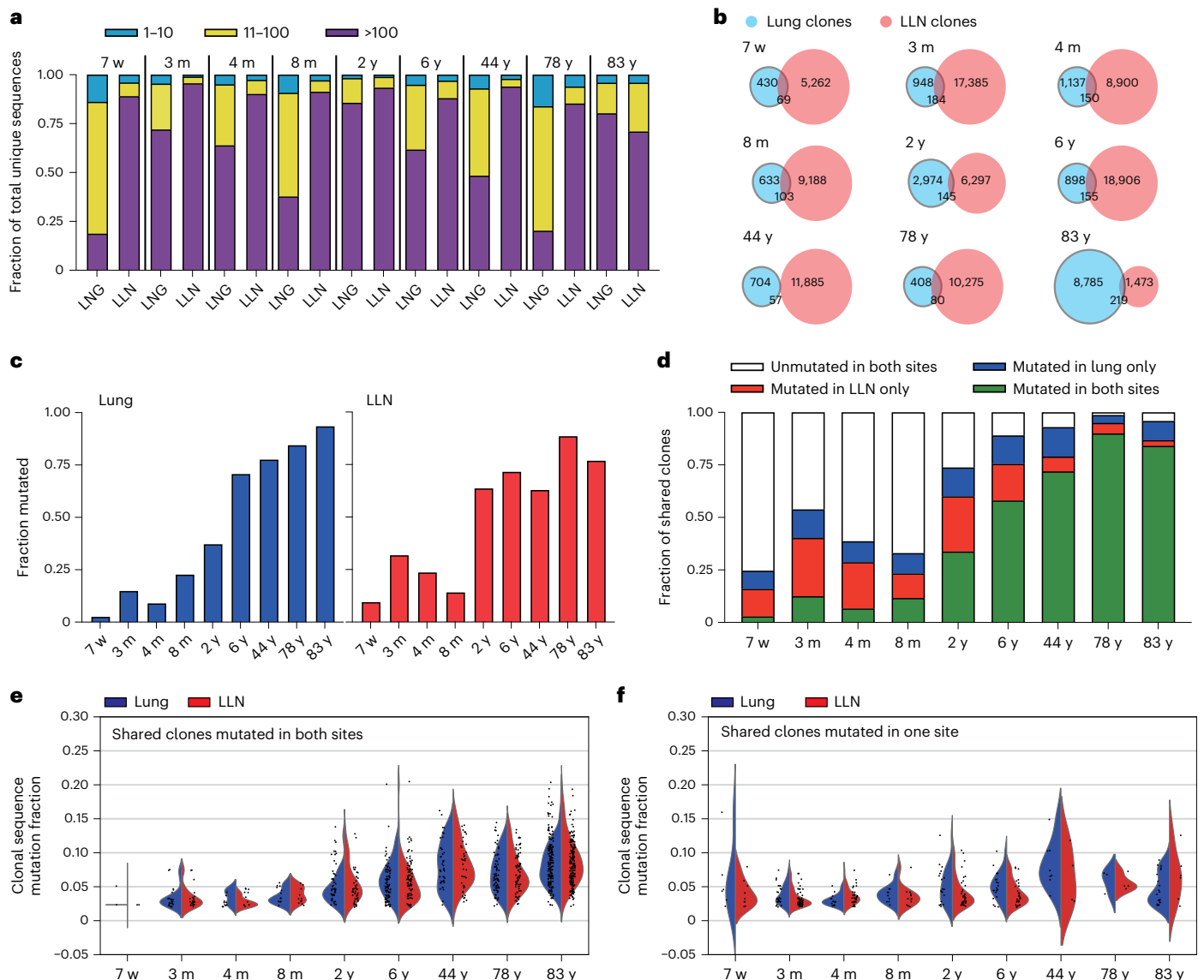
Fig. 6 | B cell clonal expansion and SHM in the lungs and LLNs during early life and childhood. BCR VH chains were sequenced from the lungs and LLNs of six pediatric donors (aged 7 weeks to 6 years) and three reference adult donors. **a**, Proportion of top *n* B cell clones (ranked by number of unique sequences) per total unique sequences from the lungs and LLNs. **b**, Clone numbers detected in the lungs (blue) and LLNs (red) shown in Venn diagrams as unique or shared between sites. **c**, Bar graph showing the fraction of clones found only in the lungs (blue; left) or LLNs (red; right) that are mutated (average per clone of fraction

of mutated positions in a sequence ≥0.02) in each donor. **d**, Fraction of shared clones that are unmutated in both sites (average per clone of fraction of mutated positions in sequence <0.02; white) or mutated (average per clone of fraction of mutated positions in a sequence ≥0.02) in both tissues (green), only in the LLNs (red) or only in the lung (blue). **e**,**f**, Range of mutation frequencies shown in violin plots and scatter plots of shared clones mutated in both sites (**e**) or in one site only (**f**); blue, distribution in the lungs; red, distribution in the LLNs; w, weeks; m, months; y, years.

the two younger donors, while a greater proportion were derived from class-switched memory cells in the 3-year-old donor (Extended Data Fig. 4). We were able to detect SHM, which develops following antigen exposure in GCs[35], in a subset of clones; the proportion of mutated clones was similar in both the lungs and LLNs across all donors (Fig. 5h). For both the lungs and LLNs, mutated clones were mostly associated with memory and GC B cells and were not found in naive and transitional subsets (Fig. 5i). Class-switched memory B cells in the youngest donor (1.33 years old) had fewer mutations across both sites than class-switched memory B cells in the older donors (2 and 3 years old). In addition, we analyzed BCR mutation patterns and found that the ratios of replacement to silent mutations were significantly greater within the complementarity-determining region (CDR) than in the framework (FWR), as displayed by the log₂ (non-synonymous/

synonymous) mutation ratio of class-switched B cells in both the lungs and LLNs (Fig. 5j), which indicates positive selection and affinity maturation[36]. Together, these results provide substantial evidence for functional B cell differentiation in the lung and LLNs and rapid maturation over the early years of life.

Rapid accumulation of SHMs in the lung and LLNs throughout childhood

To further investigate BCR mutation and clonality with age at greater depth and in more individuals, we applied next-generation sequencing of the variable region of the *IGH* gene, as previously described[37], in the lungs and LLNs of six pediatric donors (7 weeks to 6 years of age) and three reference adult donors (Fig. 6 and Supplementary Table 7). In the lungs of pediatric donors, there were fewer clones overall than
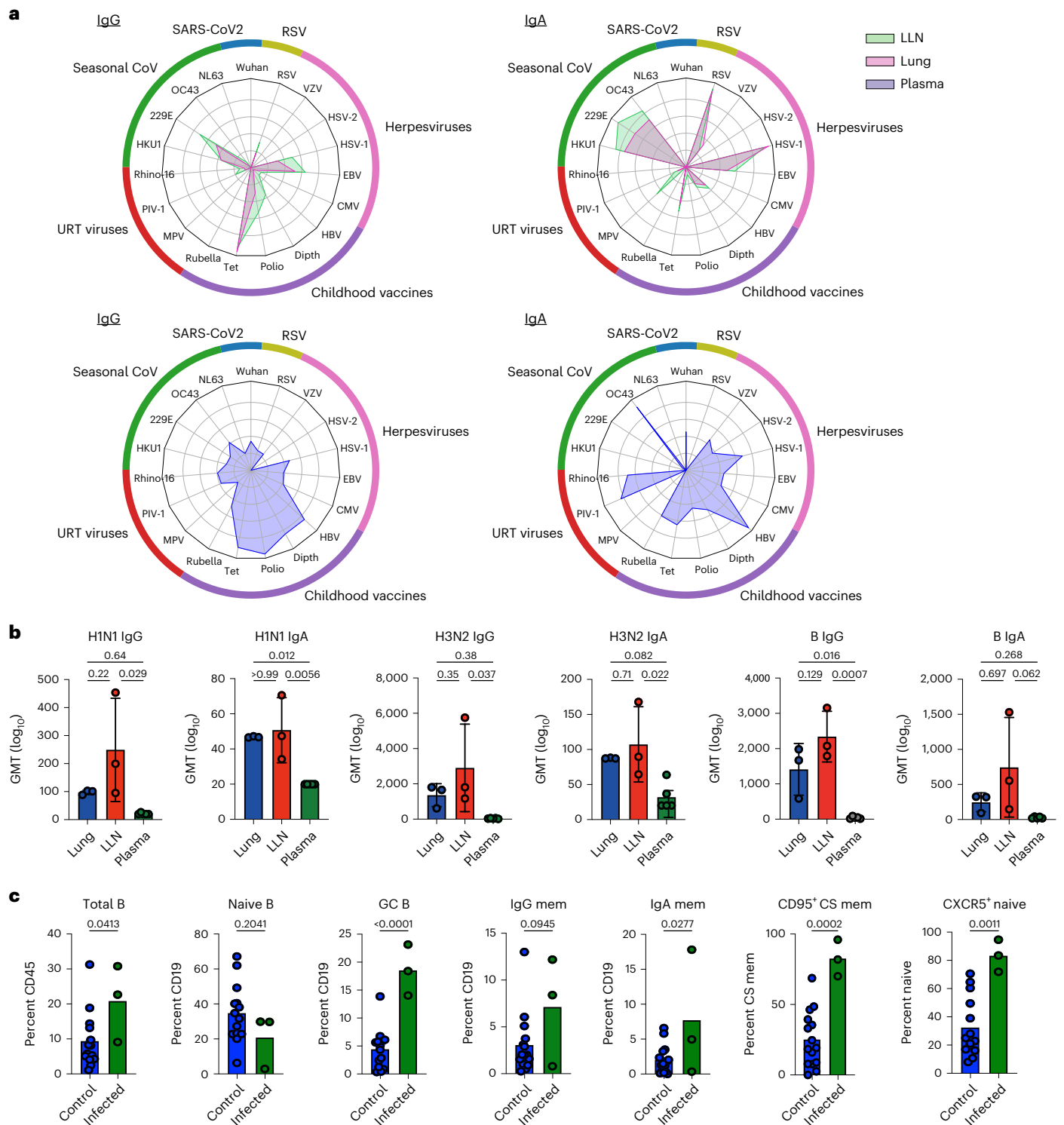
Fig. 7 | Distinct antigen specificities for tissue B cells compared to circulating antibodies. B cells were isolated from the lungs and LLNs of donors 1–3 years of age and were activated in vitro, and secreted antibodies were analyzed for specificity using antigen arrays (Methods) along with paired and age-matched plasma samples. **a**, Spider plots showing relative IgG (left) and IgA (right) reactivities for the different antigen types (top: LLN (green) and lung (pink); bottom: plasma (blue)) from pediatric donors (average of three donors for lungs/LLNs and average of six donors for plasma); VZV, varicella zoster virus; HSV, herpes simplex virus; EBV, Epstein–Barr virus; CMV, cytomegalovirus; HBV, hepatitis B; Dipth, diphtheria; Tet, tetanus; MPV, metapneumovirus; PIV-1,

parainfluenza virus type 1. **b**, Bar graph showing geometric mean titers (GMT) of each influenza strain for lungs (blue), LLNs (red) and plasma (green). Each circle represents an individual donor (lungs/LLNs: $n = 3$; plasma: $n = 6$). Significance was determined by one-way analysis of variance and Tukey's multiple comparisons test. Error bars indicate s.d. **c**, B cell activation in the lung during respiratory infection. Graphs show indicated B cell subset frequency in the lung from three donors (0–4 years old) with acute respiratory infection and age-matched healthy control donors ($n = 14$). P values were determined by unpaired t-tests (significance ≦0.05); Mem, memory; CS, class switched; URT, upper respiratory tract.

in LLNs, and lung clones had, on average, more clonal expansion than LLNs (Fig. 6a,b). Clonal overlap between the lung and LLNs was detected in pediatric and adult donors at similar frequencies (13.03%; Fig. 6b), suggesting that clonal expansion and/or migration was not age dependent. However, the extent of SHM differed with age; the proportion of mutated clones (SHM ≥ 2%) in the lung and LLNs was lowest in infant donors and substantially increased in tissues of older children and adults for both sites (Fig. 6c). Importantly, the percentage of mutated clones was low (<25%) for each site during infancy but increased thereafter, reaching adult SHM levels (75%) by 6 years in the lungs and 2 years in the LLNs (Fig. 6c). These results show rapid development of a somatically mutated B cell repertoire in the lungs and LLNs during the first 2 years of life.

To control for compositional differences in each site, we compared SHMs between expanded clones that were shared between the two sites (Fig. 6d–f). Our results show striking differences in SHM levels and distribution during infancy compared to older children and adults that indicate site-specific priming. First, shared clones in infants were largely unmutated, while the majority of the shared clones in the 2-year old, 6-year old and adult donors were mutated (Fig. 6d), indicating that expanded clones of infant B cells were not fully differentiated. Moreover, the shared clones that were mutated during infancy were mostly detected in single sites (that is, only in the lung or LLNs; Fig. 6d), suggesting site-specific differentiation in GCs. By contrast, the majority of shared clones in older children and adults were mutated in both sites and exhibited broad mutation ranges that increased with age (Fig. 6d,e). Violin plots of mutation frequencies in shared clones mutated in one site show broad ranges of mutations even in infants (Fig. 6f), further supporting in situ origin. Together, these findings provide evidence for compartmentalized B cell differentiation and somatic mutation in the lung and its associated LN during the early years.

Lung B cells respond to respiratory pathogens in early life

The compartmentalized differentiation of B cells in the lung and LLNs could derive from exposure to local antigens in the respiratory tract. Using array-based screening, we assessed the reactivity of antibodies generated from tissue B cells (Methods) compared to antibodies in plasma to multiple antigens, including childhood vaccines, common respiratory viruses, herpesviruses and influenza hemagglutinin (HA) proteins from different influenza strains (Supplementary Table 8). Our results revealed distinct specificity profiles for IgG and IgA antibodies produced in the lungs and LLNs compared to those in plasma that were observed in aggregate and in a consistent pattern from individual donors (Fig. 7a and Extended Data Fig. 6a). Lung- and LLN-derived antibodies exhibited comparable specificities, including IgG specific for seasonal coronaviruses, herpesviruses and some vaccine antigens (for example, tetanus, diphtheria and polio) and IgA specific for respiratory syncytial virus (RSV), seasonal coronaviruses and metapneumovirus (Fig. 7a and Extended Data Fig. 6a). By contrast, IgG and IgA from plasma were mostly specific to multiple childhood vaccines, herpesvirus strains and some upper respiratory pathogens (for example, rhinovirus) with relatively low or undetectable reactivity to coronaviruses and RSV compared to IgG and IgA from the lungs and LLNs (Fig. 7a). Similar results were found for influenza HA antigens; there was a higher level of reactivity to multiple influenza HA types in the lungs and LLNs than in plasma in pediatric donors (Fig. 7b), while adult donors had significantly increased specificity for influenza HA compared to children (Extended Data Fig. 6b) as a control. Together these results demonstrate a compartmentalized B cell response to respiratory pathogens during early life. Indeed, we found a significant increase in the proportion of GC B cells, activated memory B cells and follicular naive B cells in lungs from infants with respiratory infections compared to healthy control donors (Fig. 7c). These results indicate that localized B cells respond to respiratory pathogens in the lungs and LLNs during early life.

Discussion

Here, we present new evidence for BALT induction in humans as a feature of normal immune development that promotes respiratory immunity. We show that BALT develops in the lung during the early years of life and supports GC formation, B cell differentiation, SHM and, importantly, functional B cell responses. However, after 4–6 years and coincident with an increase in lung memory T cells, these BALT structures begin to regress, lose structural integrity and no longer support GC formation. The timing of maximal BALT formation in the early years further correlates with extensive B cell differentiation, SHM and direct in situ responses to respiratory viruses. Our results provide evidence that BALT induction is an early life adaptation for promoting localized immunity and B cell differentiation, which, together with the LLN response, can meet the diverse respiratory challenges during this vulnerable life stage.

Induction of BALT has been demonstrated in mouse infection models. Previous studies by Randall and colleagues demonstrated that BALT was induced following respiratory infection of mice lacking secondary lymphoid organs and supported GC formation, viral clearance and establishment of immune memory[15,38]. Induced BALT structures, referred to as iBALT, could be triggered by local inflammation in the lung (for example, inhaled LPS), particularly in neonatal mice[16]. In humans, BALT was previously identified as lymphoid aggregates in lung autopsy studies of infants who died of SIDS and other causes[39,40]; however, a role for BALT in the immune response was not possible to investigate in these samples. Here, we show multiple functional roles for BALT in supporting localized B cell differentiation and immune responses, indicating that early life human BALT plays an analogous functional role to iBALT in mice. The extensive inflammation that results from exposure to multiple diverse microorganisms during the early life period may likewise be a trigger for induction of BALT in the human infant.

Our results provide several lines of evidence for in situ priming and differentiation of B cells in the lung (in addition to the associated LN) in the early years of life. First, we identify functional GC containing GC B cells expressing AID, T_{FH} cells and follicular dendritic cells in both the lung and LLNs during early life. Second, we found that B cells in the lung exhibit a more activated profile, including increased expression of *BCL6* transcripts, than those in LLNs, consistent with their localization at the frontline of antigen exposure. Third, our repertoire analysis revealed somatic mutation of expanded B cell clones specific to either lung or LN, suggesting site-specific differentiation in GCs. Finally, we detected biased specificity of antibodies derived from early life lung and LLN-derived B cells to respiratory pathogens, while circulating antibodies were specific to antigens derived from vaccines and herpesvirus strains. These localized responses are more prevalent in early life. B cell clones in adult tissues contain expanded and somatically mutated clones dispersed across multiple sites[37], and antibody specificities to respiratory and systemic pathogens are readily detected in adult blood.

The steep decline in BALT and associated lung GC B cells and T_{FH} cells occurs after age 5 when LLNs become the primary site for GC formation and induction of adaptive immunity. In adults, BALT is rarely present and is instead associated with chronic diseases such as chronic obstructive pulmonary disease and asthma[41]. We and others observed a similar age-associated decline in gut-associated lymphoid tissue; isolated lymphoid follicles in the jejunum can be detected in the early years and decline thereafter, and Peyer's patches in the ileum are highest in children and decline, albeit more gradually, with age in adults[42,43]. There could be multiple mechanisms for the decline in functional BALT. The extent of lung inflammation may be reduced by the development of adaptive immune responses to ubiquitous pathogens in the early years. Different proinflammatory cytokines and chemokines have been shown to promote BALT formation in mice[44], and it is possible that certain inflammatory cytokines may be necessary for induction and/or maintenance of BALT in humans. Another possibility is that the

establishment of tissue-resident memory T cells in the lung may promote more rapid clearance of pathogens and stabilize the lung environment by an overall reduction in pathogen load. We show here that the space where BALT forms in the lung is occupied by memory T cells in older children, and we recently demonstrated that functionally mature tissue-resident memory T cells form in human lungs after age 4 (ref. 45), the age when BALT starts to regress. Additional feedback mechanisms to regulate B cell immunity, such as the presence of antibodies[46], that may accumulate in the lung could also contribute to decreased signals for BALT maintenance. Together, our results and mouse studies indicate that maintenance of ectopic lymphoid structures in mucosal sites such as the lung is an active process that requires ongoing signals.

We propose that induction of ectopic lymphoid structures in mucosal sites could be a generalized adaptation of the early life immune system for promoting efficacious T cell–B cell interaction at sites of high and diverse pathogen loads. This tissue-directed response to respiratory pathogens during the early years of life can account for the better outcome of young children to newly encountered respiratory pathogens than observed in adults. There are also major implications for promoting respiratory immunity through vaccines. The presence of BALT in young children suggests that intranasal administration of vaccines may be particularly effective in children when priming outside lymphoid organs can occur. Moreover, induction of BALT could also be targeted in adults. Certain experimental vaccines were shown to induce BALT in adult mice and correlated with protection to respiratory pathogens[47,48]. Promoting BALT in future vaccines could be a beneficial strategy when systemic administration does not prevent infection or illness.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-023-01557-3.

References

1. Mohr, E. & Siegrist, C.-A. Vaccination in early life: standing up to the challenges. *Curr. Opin. Immunol.* **41**, 1–8 (2016).
2. PrabhuDas, M. et al. Challenges in infant immunity: implications for responses to infection and vaccines. *Nat. Immunol.* **12**, 189–194 (2011).
3. Lu, X. et al. SARS-CoV-2 infection in children. *N. Engl. J. Med.* **382**, 1663–1665 (2020).
4. Buckley, R. H. The multiple causes of human SCID. *J. Clin. Invest.* **114**, 1409–1411 (2004).
5. Franca, T. T. et al. CD40 ligand deficiency: treatment strategies and novel therapeutic perspectives. *Expert Rev. Clin. Immunol.* **15**, 529–540 (2019).
6. Crotty, S. Follicular helper CD4 T cells (T_FH). *Annu. Rev. Immunol.* **29**, 621–663 (2011).
7. Victora, G. D. & Nussenzweig, M. C. Germinal centers. *Annu. Rev. Immunol.* **30**, 429–457 (2012).
8. Kim, W. et al. Germinal centre-driven maturation of B cell response to mRNA vaccination. *Nature* **604**, 141–145 (2022).
9. Turner, J. S. et al. Human germinal centres engage memory and naive B cells after influenza vaccination. *Nature* **586**, 127–132 (2020).
10. Turner, J. S. et al. SARS-CoV-2 mRNA vaccines induce persistent human germinal centre responses. *Nature* **596**, 109–113 (2021).
11. Thome, J. J. et al. Early-life compartmentalization of human T cell differentiation and regulatory function in mucosal and lymphoid tissues. *Nat. Med.* **22**, 72–77 (2016).
12. Farber, D. L. Tissues, not blood, are where immune cells act. *Nature* **593**, 506–509 (2021).
13. Kumar, B. V., Connors, T. J. & Farber, D. L. Human T cell development, localization, and function throughout life. *Immunity* **48**, 202–213 (2018).
14. Randall, T. D. Bronchus-associated lymphoid tissue (BALT) structure and function. *Adv. Immunol.* **107**, 187–241 (2010).
15. Moyron-Quiroz, J. E. et al. Role of inducible bronchus associated lymphoid tissue (iBALT) in respiratory immunity. *Nat. Med.* **10**, 927–934 (2004).
16. Rangel-Moreno, J. et al. The development of inducible bronchus-associated lymphoid tissue depends on IL-17. *Nat. Immunol.* **12**, 639–646 (2011).
17. Bilalovic, N. et al. Expression of BCL-6 and CD10 protein is associated with longer overall survival and time to treatment failure in follicular lymphoma. *Am. J. Clin. Pathol.* **121**, 34–42 (2004).
18. Muramatsu, M. et al. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
19. Muramatsu, M. et al. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* **274**, 18470–18476 (1999).
20. Cyster, J. G. et al. Follicular stromal cells and lymphocyte homing to follicles. *Immunol. Rev.* **176**, 181–193 (2000).
21. Heesters, B. A., Myers, R. C. & Carroll, M. C. Follicular dendritic cells: dynamic antigen libraries. *Nat. Rev. Immunol.* **14**, 495–504 (2014).
22. Junt, T. et al. CXCR5-dependent seeding of follicular niches by B and T_H cells augments antiviral B cell responses. *J. Immunol.* **175**, 7109–7116 (2005).
23. Weisel, N. M. et al. Comprehensive analyses of B-cell compartments across the human body reveal novel subsets and a gut-resident memory phenotype. *Blood* **136**, 2774–2785 (2020).
24. Glass, D. R. et al. An integrated multi-omic single-cell atlas of human B cell identity. *Immunity* **53**, 217–232 (2020).
25. Son, Y. M. et al. Tissue-resident CD4+ T helper cells assist the development of protective respiratory B and CD8+ T cell memory responses. *Sci. Immunol.* **6**, eabb6852 (2021).
26. Swarnalekha, N. et al. T resident helper cells promote humoral responses in the lung. *Sci. Immunol.* **6**, eabb6808 (2021).
27. Locci, M. et al. Human circulating PD-1+CXCR3−CXCR5+ memory T_FH cells are highly functional and correlate with broadly neutralizing HIV antibody responses. *Immunity* **39**, 758–769 (2013).
28. Stewart, A. et al. Single-cell transcriptomic analyses define distinct peripheral B cell subsets and discrete development pathways. *Front. Immunol.* **12**, 602539 (2021).
29. He, B. et al. The transmembrane activator TACI triggers immunoglobulin class switching by activating B cells through the adaptor MyD88. *Nat. Immunol.* **11**, 836–845 (2010).
30. Tan, C. et al. NR4A nuclear receptors restrain B cell responses to antigen when second signals are absent or limiting. *Nat. Immunol.* **21**, 1267–1279 (2020).
31. Krzyzak, L. et al. CD83 modulates B cell activation and germinal center responses. *J. Immunol.* **196**, 3581–3594 (2016).
32. Weisel, N. M. et al. Surface phenotypes of naive and memory B cells in mouse and human tissues. *Nat. Immunol.* **23**, 135–145 (2022).
33. Cannons, J. L. et al. Optimal germinal center responses require a multistage T cell:B cell adhesion process involving integrins, SLAM-associated protein, and CD84. *Immunity* **32**, 253–265 (2010).
34. Chen, S. et al. ID3 orchestrates germinal center B cell development. *Mol. Cell. Biol.* **36**, 2543–2552 (2016).

35. Davis, C. W. et al. Longitudinal analysis of the human B cell response to Ebola virus infection. *Cell* **177**, 1566–1582 (2019).

36. Yaari, G., Benichou, J. I., Vander Heiden, J. A., Kleinstein, S. H. & Louzoun, Y. The mutation patterns in B-cell immunoglobulin receptors reflect the influence of selection acting at multiple time-scales. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **370**, 20140242 (2015).

37. Meng, W. et al. An atlas of B-cell clonal distribution in the human body. *Nat. Biotechnol.* **35**, 879–884 (2017).

38. Moyron-Quiroz, J. E. et al. Persistence and responsiveness of immunologic memory in the absence of secondary lymphoid organs. *Immunity* **25**, 643–654 (2006).

39. Gould, S. J. & Isaacson, P. G. Bronchus-associated lymphoid tissue (BALT) in human fetal and infant lung. *J. Pathol.* **169**, 229–234 (1993).

40. Heier, I. et al. Characterisation of bronchus-associated lymphoid tissue and antigen-presenting cells in central airway mucosa of children. *Thorax* **66**, 151–156 (2011).

41. Pabst, R. & Tschernig, T. Bronchus-associated lymphoid tissue: an entry site for antigens for successful mucosal vaccinations? *Am. J. Respir. Cell Mol. Biol.* **43**, 137–141 (2010).

42. Morbe, U. M. et al. Human gut-associated lymphoid tissues (GALT); diversity, structure, and function. *Mucosal Immunol.* **14**, 793–802 (2021).

43. Senda, T. et al. Microanatomical dissection of human intestinal T-cell immunity reveals site-specific changes in gut-associated lymphoid tissues over life. *Mucosal Immunol.* **12**, 378–389 (2019).

44. Randall, T. D. & Mebius, R. E. The development and function of mucosal lymphoid tissues: a balancing act with micro-organisms. *Mucosal Immunol.* **7**, 455–466 (2014).

45. Connors, T. J. et al. Site-specific development and progressive maturation of human tissue resident memory T cells over infancy and childhood. *Immunity* https://doi.org/10.1016/j.immuni.2023.06.008 (2023)

46. Xu, H., Zhang, L. & Heyman, B. IgG-mediated immune suppression in mice is epitope specific except during high epitope density conditions. *Sci. Rep.* **8**, 15292 (2018).

47. Wiley, J. A. et al. Inducible bronchus-associated lymphoid tissue elicited by a protein cage nanoparticle enhances protection in mice against diverse respiratory viruses. *PLoS ONE* **4**, e7142 (2009).

48. Chiavolini, D. et al. Bronchus-associated lymphoid tissue (BALT) and survival in a vaccine mouse model of tularemia. *PLoS ONE* **5**, e11156 (2010).

[1]Department of Surgery, Columbia University Irving Medical Center, New York, NY, USA. [2]Department of Microbiology and Immunology, Columbia University Irving Medical Center, New York, NY, USA. [3]Department of Microbiology, Immunology and Genetics, Ben-Gurion University of the Negev, Be'er-Sheva, Israel. [4]The National Institute for Biotechnology in the Negev, Ben-Gurion University of Negev, Be'er-Sheva, Israel. [5]Department of Human Biology, University of Haifa, Haifa, Israel. [6]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [7]Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA. [8]Department of Pathology, Immunology and Laboratory Medicine, University of Florida Diabetes Institute, Gainesville, FL, USA. [9]Department of Pediatrics, Columbia University Irving Medical Center, New York, NY, USA. [10]Department of Biochemistry & Molecular Biophysics, Columbia University Irving Medical Center, New York, NY, USA. [11]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [12]These authors contributed equally: Rei Matsumoto, Joshua Gray. ✉e-mail: df2396@cumc.columbia.edu

Methods

Ethics statement

All research complies with ethical regulations outlined by the Columbia University Institutional Review Board (IRB).

Human samples

Organ donors. All human tissue samples were obtained from deceased organ donors at the time of organ procurement for clinical transplantation. The use of these organ donor samples is not human subjects research as confirmed by the Columbia University IRB, as tissues are obtained from deceased, not living, individuals. Tissues from pediatric organ donors were obtained through the HANDEL-I program at the University of Florida, which maintains collaborations with organ procurement organizations through the United States[45,49]. Samples from adult organ donors were obtained through an approved protocol and material transfer agreement with LiveOnNY, the organ procurement organization for the New York metropolitan area, as previously described[50,51]. Donors were free of malignancy, seronegative for hepatitis B and C and HIV and free of ongoing respiratory infections, except where indicated (Supplementary Table 1).

Pediatric donors. Serum samples for antigen microarray assays were collected from children of ages 1.25–3 years who were at the clinic for well child visits. We obtained informed consent from their parents under a protocol approved by the IRB of Columbia University Medical Center.

Sample processing

Organs were packaged in saline or University of Wisconsin solution, placed on ice and transported to the laboratory. Tissues were processed within 24–36 h of aortic cross-clamp (separation of systemic circulation from the outflow of the heart). Lungs and LNs were processed to single-cell suspensions using mechanical and enzymatic digestion, as previously described[52,53]. LNs were isolated by dissection from the tracheobronchial tree of the lung. Isolated LNs were placed in complete medium (IMDM (Gibco, 12440-053), 1% PSQ (100 U ml^{-1} penicillin, 100 µg ml^{-1} streptomycin and 2 mM L-glutamine; GeminiBio, c400-110) and 10% fetal bovine serum (FBS; GeminiBio, 100–106)), mechanically disrupted with scissors and double filtered through 100-µm strainers. Suspensions were washed with wash buffer (PBS (Corning, 20-030-CV), 2 mM EDTA (Corning, 46-034-CI) and 5% FBS) and resuspended in complete medium. Lung parenchyma were separated from the large airways (trachea) using scissors and placed into digestion medium (IMDM, collagenase D 1 mg ml^{-1} (Millipore Sigma, 11088882001) and DNase 0.1 mg ml^{-1} (Millipore Sigma, DN25-5G)) on a shaker at 37 °C for 30 min. Following incubation, lung samples were strained and washed with wash buffer. Samples were centrifuged with resultant cell pellets diluted for mononuclear cell isolation by density gradient centrifugation (Lymphocyte Separation Medium, Corning, 25-072-CI), followed by resuspension in complete medium. In some cases, blood was also obtained from organ donors, and mononuclear cells were isolated by density gradient centrifugation, as described above.

Immunofluorescence imaging

Lung samples obtained from both lobes including major airways and LLNs were cut into 0.5-cm sections and fixed with 4% paraformaldehyde, lysine and periodate buffer (PLP; 0.05 M phosphate buffer (Sigma), 0.1 M L-lysine (Sigma; pH 7.4), 2 mg ml^{-1} sodium periodate and 10 mg ml^{-1} paraformaldehyde (Fisher)). Samples were then placed on a shaker at 4 °C overnight, washed in PLP buffer and placed into 30% sucrose solution for 48 h at 4 °C, followed by embedding in optimal cutting temperature medium. Samples were then frozen and stored at −80 °C. Frozen samples used for imaging were sectioned using a Leica CM1950 cryostat to 20 µm thickness. Before antibody staining, Fc blocking was performed using anti-CD16/32 Fc block (BioLegend)

diluted in PBS containing 2% goat serum and 5% FBS for 1 h at 20 °C. Sections were then stained with DAPI and multicolor immunofluorescence panel (Supplementary Table 9) for 1 h at 20 °C protected from light. Images were acquired using a Nikon A1 confocal microscope. Imaged data were analyzed using Imaris software (Bitplane; Oxford Instruments) using surface creation and spot detection functions. Four to five sections per donor were imaged for the lungs with area and cell number calculated for three to four central airways per donor. For LLNs, two sections per donor were imaged, and the whole cross-section was analyzed.

Flow cytometry

Single-cell suspensions ($5 × 10^6$–$10 × 10^6$ cells) from lungs, LLNs and blood were stained using a 26-marker spectral flow cytometry panel (Supplementary Table 9). Cells were first washed with PBS, resuspended with LIVE/DEAD Fixable Blue Dead Cell stain (Thermo Fisher Scientific) and incubated at 20 °C in the dark for 20 min, followed by washing with fluorescence-activated cell sorting (FACS) buffer (PBS, 2% FBS and 0.1 mM EDTA) and subsequent incubation with TrueStain FcX for 10 min in the dark. Next, cells were washed with FACS buffer and resuspended in staining antibodies for 30 min at 20 °C, followed by two washes with FACS buffer. Cells were then fixed and permeabilized using a Foxp3 Transcription Factor Staining Buffer set (eBioscience), as per the manufacturer's protocols. Last, cells were incubated in the dark at 4 °C with intracellular antibodies and washed twice in permeabilization buffer. Data were collected on a five-laser Cytek Aurora flow cytometer (Cytek Bio) and analyzed using FlowJo version 10.8.1 (Treestar); statistical testing of compiled data was performed using Prism (GraphPad).

Immunoglobulin heavy chain gene (*IGH*) sequencing

DNA was extracted from cryopreserved mononuclear cells from the lungs or LLNs using a Gentra Puregene Cell kit (Qiagen). Immunoglobulin heavy chain family-specific PCRs were performed on genomic DNA samples using primers in *FR1* and J$_H$ as described previously[37]. Two biological replicates were run on all samples. Sequencing was performed in the Human Immunology Core Facility at the University of Pennsylvania using an Illumina 2 × 300-base pair (bp) paired-end kit (Illumina MiSeq Reagent kit v3, 600-cycle, Illumina MS-102-3003). Reads from an Illumina MiSeq were filtered, annotated and grouped into clones as described previously[54]. Briefly, pRESTO[55] was used to pair reads, remove sequences with overall low quality and mask individual low-quality base calls. IgBLAST[56] was then applied to the filtered sequences for alignment and gene annotation. Finally, the aligned sequences were imported into ImmuneDB[57] for clonal inference, lineage construction and further processing. Sequences sharing the same V$_H$ gene, J$_H$ gene and 85% amino acid similarity in CDR3 were clustered into clones. Metadata for the input DNA and clone summary can be found in Supplementary Table 7. All downstream analyses exclude non-productive clones and clones with fewer than two copies across all libraries from each donor.

Single-cell transcriptomic profiling by scRNA-seq

Single-cell suspensions of each tissue were prepared as described above. B cells were enriched by FACS using antibodies to CD45, CD33, CD66b, CD3 and CD19. Enrichment resulted in ~90% purity of live CD45$^+$CD66b$^-$CD33$^-$CD3$^-$CD19$^+$ cells from each tissue. Enriched B cell populations from each sample were resuspended in PBS, FBS and EDTA and stained with 3 µl of a unique TotalSeq-C DNA-barcoded hashtag (Supplementary Table 9) for 30 min on ice. Cells were washed, counted and combined in PBS and FBS, ensuring equal numbers of live cells were present from each sample.

CITE/single-cell BCR-seq (scBCR-seq) libraries were constructed using the Chromium Next GEM Single Cell 5′ Reagent kit v2 (10x Genomics) with both a 5′ Feature Barcode kit (10x Genomics) for antibody-derived tag (ADT) amplification and a Chromium Single Cell

Human BCR Amplification kit (10x Genomics) for targeted BCR amplification, according to the manufacturer's instructions. The ADT library pool and scRNA-seq library pool were each sequenced on one lane of an S4 flow cell on an Illumina NovaSeq 6000 (~2.5 billion 2 × 100-bp reads per lane). The scBCR-seq library pool was sequenced with a 150-cycle high-output kit on an Illumina NextSeq 550 (read 1: 26 bp; read 2: 114 bp; index read 1: 10 bp; index read 2: 10 bp).

Data processing and analysis

Quantification of scRNA-seq library pools was performed by pseudoalignment to the human transcriptome (Gencode v24 annotation derived from GRCh38) using kallisto v0.46.2 in 'BUS' mode[58,59]. ADT barcodes were extracted from demultiplexed fastq files using custom scripts (https://github.com/simslab/DropSeqPipeline8). Hashtag barcodes from cell-identifying barcodes containing >1,000 unique transcripts were demultiplexed using Hashsolo with default parameters[60]. scBCR-seq data were demultiplexed and aligned with VDJ cassette and variable sequencing identification using Cell Ranger VDJ v7.0.1 and the GRCh38/Ensembl v7.0.0 annotation from 10x Genomics. Finally, the demultiplexed scRNA-seq and scBCR-seq data were combined into a single h5ad-formatted anndata object using scirpy[61] for downstream analysis.

A count matrix was created of 49,734 individual events using the pandas package in Python (version 1.3.5). Individual events with less than 600 total counts, less than 400 genes and more than 20% mitochondrial content were excluded from further analysis as well as doublets identified using the Scrublet package (version 0.2.3). Highly variable genes were calculated using a minimum dispersion of 0.5 for each library, and gene expression was scaled without zero center and a max value of 10. The top 30 principal components were used to calculate the 10 nearest neighbors, which were used to generate a UMAP and run Leiden clustering with the scanpy package and default settings. Differential gene expression for the Leiden clusters was calculated using the Wilcoxon mode and tie correction in the rank_genes_groups function. Clusters identified without *CD19* or *MS4A1* expression were removed as non-B cell contaminants.

Highly variable genes were recalculated on the cleaned B cell dataset consisting of 43,552 B cells. The data were scaled, and the regress_out function was used to reduce the influence of cell cycle genes on the principal-component analysis. The principal-component analysis was harmonized by individual donor using the Python harmonypy package (version 0.0.5) to reduce donor-to-donor variability. Nearest neighbors and UMAP were then calculated as described above. Leiden clustering was performed on the data at a resolution of 2.0, providing 24 clusters. Differential expression analysis was conducted on the individual clusters to determine transcriptional profiles and was displayed in a dot plot using scanpy's rank_genes_groups_dotplot function. Clusters with similar gene expression profiles for key genes related to B cell subsets were grouped as one cluster and annotated as 'transitional', 'naive', 'non-class-switched memory', 'class-switched memory' or 'GC'.

Differential gene expression analysis

Differential expression analysis was conducted on the named B cell subsets using the rank_genes_groups function. Significant gene lists were filtered for a minimum $\log_2$ (fold change) of 1.0, and key genes were plotted using the rank_genes_groups_matrixplot function and a scaled z score. Proportions of each B cell subset were plotted as a stacked bar plot by each donor's respective tissue. To determine a tissue-specific signature of B cells, we created tissue-specific subsets by combining each event's tissue of origin with its assigned B cell subset. We then combined transitional and naive B cells as well as memory subsets and subsampled down to the tissue with the lowest cellular contribution (GC B cells were excluded as they contained less than 100 events). Differential expression was calculated for the combined transitional and

naive B cells and memory subsets between the organs. Volcano plots were created using the bioinfokit package v2.0.8. The genes that were significantly different in both comparisons (LLN naive/transitional versus lung naive/transitional and LLN memory versus lung memory) were then plotted using the rank_genes_groups_matrixplot function as described above.

B cell activation scores were calculated using the top 100 genes from Fowler et al.[62] that were the most differentially expressed when comparing B cells activated through the BCR and with LPS compared to resting B cells. These genes were then supplied to the score_genes function, projected as a function of expression on the UMAP and plotted using Seaborn's violinplot by each tissue's subset.

Single-cell B cell mutation, clonal annotation and analysis

To perform single-cell mutation and clonal diversity analysis, we annotated gene sequences based on heavy chains[63]. Using the Immcantation pipeline (https://immcantation.readthedocs.io/en/latest/index.html#), we aligned the heavy chains and light chains in each cell to the appropriate germline using igBlast based on IMGT germline references (https://changeo.readthedocs.io/en/stable/examples/igblast.html). Cells that expressed more than one heavy chain or had only light chains were removed from further analysis. For all datasets, only productive-rearranged BCR contigs with valid V and J gene annotations were used. We then divided all usable cells into groups with the same heavy chain V and J germline assignment and the same CDR3 length. These groups were then further separated into individual clones by hierarchical clustering (cutoff of 0.15). For each sequence, we calculated the number of synonymous and non-synonymous nucleotide mutations across the whole sequence and separately in CDR and FWR regions (including CDR3 and FWR4).

B cell culture and polyclonal activation

Single-cell suspensions (30×10^6–100×10^6 cells) from lungs and LLNs were resuspended in a warm stimulation medium (RPMI, 5% human antibody serum, 1% PSQ, 10 µg ml^{-1} LPS, 1 µg ml^{-1} R848, 1 µg ml^{-1} Pokeweed mitogen, 1:10,000 *Staphylococcus aureus* Cowean and 0.5 mM β-mercaptoethanol) in a six-well plate at a ratio of 30×10^6–50×10^6 cells per well. The plated cells were put in an incubator at 37 °C and 5% CO_2. On the fifth day, the cells were centrifuged at 400g for 5 min, and the supernatants were collected.

Antigen microarray assay

Viruses and recombinant proteins were spotted onto *N*-hydroxysuccinimide ester-derivatized hydrogel slides (H slides type B, Schott-Nexterion) using a Scienion Sx non-contact array spotter (Scienion). Spot volumes ranged between 300 and 360 pl. Antigens at each concentration were spotted in triplicate. Antigens were diluted in Triton (Merck, 9036-19-5) or Scispot D1 spotting buffer (Scienion) at multiple concentrations ranging from 65 to 16 µg ml^{-1}. Eight identical microarrays were spotted on each microarray slide. Antigens included (1) antigens of common childhood vaccines, (2) herpesvirus antigens (including herpes simplex virus 1 and 2, Epstein–Barr virus, cytomegalovirus and varicella zoster virus), (3) surface glycoproteins of common respiratory viral infections (including RSV, adenoviruses and parainfluenza), (4) influenza HA recombinant proteins spanning the antigenic evolution of H1N1, H3N2 and B subtypes and (5) antigens of the SARS-CoV-2 Wuhan strain and seasonal coronavirus strains such as HKU-1, NL63, 229E and OC43 (Supplementary Table 8).

Slides were hybridized with samples using a previously developed protocol[64]. Briefly, array slides were chemically blocked at 20 °C on a shaker. Following the blocking step, slides were washed twice in PBST (PBS and 0.05% Tween 20), twice in PBS and twice in deionized water. After washing, the slides were dried by centrifugation for 5 min and loaded on incubation trays (PepperChips, PepperPrint). Samples from sera, lungs and LLNs were diluted to 10 µg ml^{-1} for both IgG and

IgA in a hybridization buffer (1% bovine serum albumin and 0.025% Tween 20 in PBS). After 2 h of incubation at 20 °C on a shaker, the slides were washed twice using the washing buffer and twice with PBS, as mentioned above. Slides were then incubated for 45 min on a shaker at 20 °C with a fluorescently labeled polyclonal secondary antibody diluted in hybridization buffer (IgG: Alexa 544 Fluor 647 AffiniPure donkey anti-human IgG (H+L), 709-605-149, Jackson ImmunoResearch, 1:1,000 dilution; IgA: Alexa Fluor 647 AffiniPure goat anti-human serum IgA chain specific, 109-605-011, Jackson ImmunoResearch, 1:3,000 dilution). Slides were scanned on a three-laser GenePix 4400 scanner. Images were analyzed using GenePix Pro version 7 to obtain the MFI of each spot after subtracting the main local background fluorescence intensity ($0 \leq MFI \leq 65{,}000$).

Antigen microarray analysis

The microarray results were analyzed using an in-house pipeline developed in Python. Because antigens at each concentration were spotted in triplicate, the median MFI intensity of each triplicate was calculated. During each experiment, a negative-control array was incubated with the incubation buffer only. The background staining of each triplicate of spots in the negative-control microarray was subtracted from all other microarrays. To minimize background noise, antigens with a maximal value of <2,000 were excluded from the analysis. All responses below an MFI of 1,000 were considered background. For the influenza microarrays, we sorted the HA and neuraminidase proteins into groups according to their subtype. For respiratory and childhood vaccine antigens, we also divided the antigens into appropriate analysis groups. The magnitude of antibody response to a group of antigens was defined as the geometric mean of the MFI levels to all the proteins included in the group.

Statistical analysis

For comparisons between experimental groups, we performed Wilcoxon rank-sum tests using the scipy.stats package. In the spider plots, all samples were normalized together to transform the values to the range of 0–1, where each antigen was normalized individually across all samples.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

BCR-seq data that support the findings of this study have been deposited in the SRA with the accession code PRJNA847585. The single-cell transcriptome and CITE-seq data have been deposited in Gene Expression Omnibus with accession number GSE223646. Source data are provided with this paper.

Code availability

Custom computer code used in this study is available at https://github.com/simslab/DropSeqPipeline8 and https://github.com/arosenfeld/immunedb.

References

49. Poon, M. M. L. et al. Heterogeneity of human anti-viral immunity shaped by virus, tissue, age, and sex. *Cell Rep.* **37**, 110071 (2021).
50. Thome, J. J. et al. Spatial map of human T cell compartmentalization and maintenance over decades of life. *Cell* **159**, 814–828 (2014).
51. Carpenter, D. J. et al. Human immunology studies using organ donors: impact of clinical variations on immune parameters in tissues and circulation. *Am. J. Transpl.* **18**, 74–88 (2018).
52. Dogra, P. et al. Tissue determinants of human NK cell development, function, and residence. *Cell* **180**, 749–763 (2020).
53. Szabo, P. A. et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, 4706 (2019).
54. Goel, R. R. et al. Distinct antibody and memory B cell responses in SARS-CoV-2 naive and recovered individuals following mRNA vaccination. *Sci. Immunol.* **6**, eabi6950 (2021).
55. Vander Heiden, J. A. et al. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930–1932 (2014).
56. Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41**, W34–W40 (2013).
57. Rosenfeld, A. M., Meng, W., Luning Prak, E. T. & Hershberg, U. ImmuneDB: a system for the analysis and exploration of high-throughput adaptive immune receptor sequencing data. *Bioinformatics* **33**, 292–293 (2017).
58. Melsted, P. et al. Modular, efficient and constant-memory single-cell RNA-seq preprocessing. *Nat. Biotechnol.* **39**, 813–818 (2021).
59. Melsted, P., Ntranos, V. & Pachter, L. The barcode, UMI, set format and BUStools. *Bioinformatics* **35**, 4472–4473 (2019).
60. Bernstein, N. J. et al. Solo: doublet identification in single-cell RNA-seq via semi-supervised deep learning. *Cell Syst.* **11**, 95–101 (2020).
61. Sturm, G. et al. Scirpy: a Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. *Bioinformatics* **36**, 4817–4818 (2020).
62. Fowler, T. et al. Divergence of transcriptional landscape occurs early in B cell activation. *Epigenetics Chromatin* **8**, 1–14 (2015).
63. Zhou, J. Q. & Kleinstein, S. H. Cutting edge: Ig H chains are sufficient to determine most B cell clonal relationships. *J. Immunol.* **203**, 1687–1692 (2019).
64. Levy, S. et al. FLU-LISA: high throughput antibody profiling using antigen microarrays. *Immunol. Cell Biol.* **101**, 231–248 (2022).

Author contributions

R.M. designed and performed experiments, analyzed data and wrote the manuscript. J.G. designed and performed the flow cytometry studies and single-cell transcriptome profiling, analyzed data and wrote the manuscript. K.R. designed and performed the flow cytometry studies for the T_{FH} analysis. L.M.F., H.O., L.L. and T.H. performed antigen microarray assays, analyzed the data and helped write the manuscript. W.M., A.M.R. and E.T.L.P. performed the V_H chain sequencing, analyzed the data and helped write and edit the manuscript. M.K. and U.H. analyzed bulk and scBCR repertoires and helped write the manuscript. D.C. constructed libraries for single-cell transcriptome profiling. P.A.S. analyzed scRNA-seq data. T.M.B., M.A.A. and M.B. coordinated acquisition of human infant tissues. T.J.C., R.S.G., J.G., M.C.B. and K.R. coordinated acquisition and/or processed human infant tissue. D.L.F. planned experiments, analyzed data and wrote and edited the manuscript.

**Extended Data Fig. 1 | Germinal Centers in pediatric lungs and associated lymph nodes contain follicular dendritic cells.** Tissue sections from lung-associated lymph nodes (LLN) and lungs were stained with anti-CD3 (green), anti-CD20 (blue), anti-CD10 (red), and anti-follicular dendritic cell (FDC) (light blue) antibodies. **a**, Representative image of LLN from a 1.5 yr old donor showing FDC in a germinal center (GC) structure (scale bar: 150μm) (n = 3, with 3 sections imaged per donor). **b**, FDC in BALT shown in representative images from 4 organ donors paired by row. FDC, and CD10 channels are shown for each donor and merged column shows staining with all four fluorochromes (n = 5, with 3 sections imaged per donor). **c**, LLN from a 4 yr old donor showing FDC (white arrows) in GC structures (pink) (scale bar: 200μm). Heterogeneous appearance of FDC with GC B cells across different follicles suggests multiple stages of GC formation (n = 3, with 3 sections imaged per donor). FDC=follicular dendritic cell, yr= year, LLN=lung lymph node, GC=germinal center.
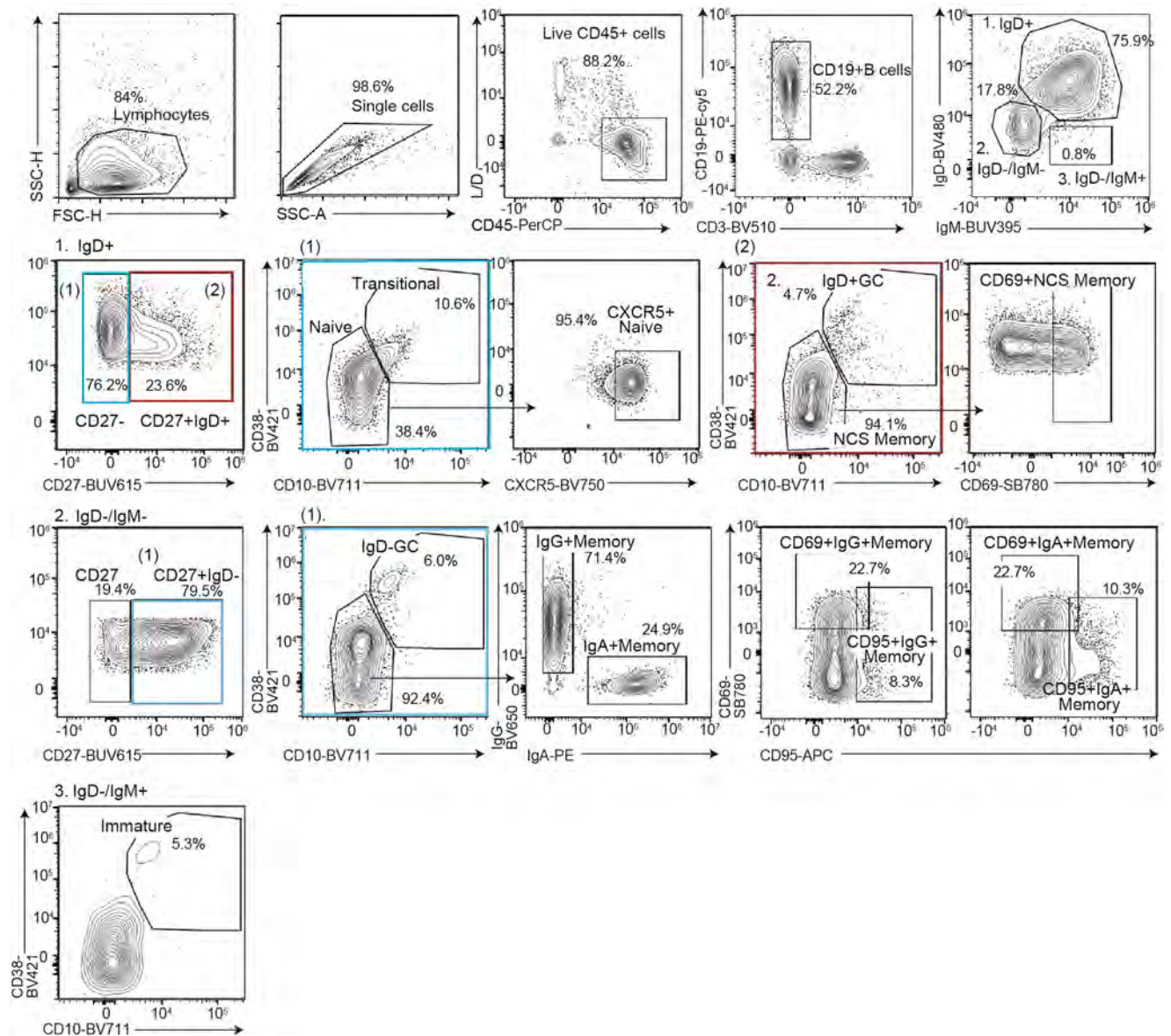
Extended Data Fig. 2 | Gating strategy for analysis of B cell subsets by flow cytometry. Lymphocytes were selected based on forward (FSC) vs side scatter (SSC) height (H) properties, then single cell gating was applied based on SSC-H vs area (-A), followed by selection of CD45+ live cells. B cells (CD19 + CD3-) were selected by CD19 expression and lack of CD3 expression and three subpopulations were identified based on IgD and IgM expression: 1. IgD+ cells (second row) were delineated into subsets based on CD27 expression with IgD+CD27- cells further divided based on CD38 and CD10 expression into transitional and naive B cells and IgD+CD27+ cells to IgD+ GC and non-class switched (NCS) memory B cells, which were analyzed for CD69 expression for tissue residency. 2. IgD-IgM- cells were also delineated based on CD27 expression to GC B cells based on CD38 and CD10 expression and CD27+IgD-IgM-CD10- B cells were further delineated into IgA+ memory and IgG+ class-switched (CS) memory B cells which were further analyzed for CD69 expression for tissue residency and CD95 for activation. 3. IgD-IgM+ B cells were delineated into immature subsets based on CD38 and CD10 expression.
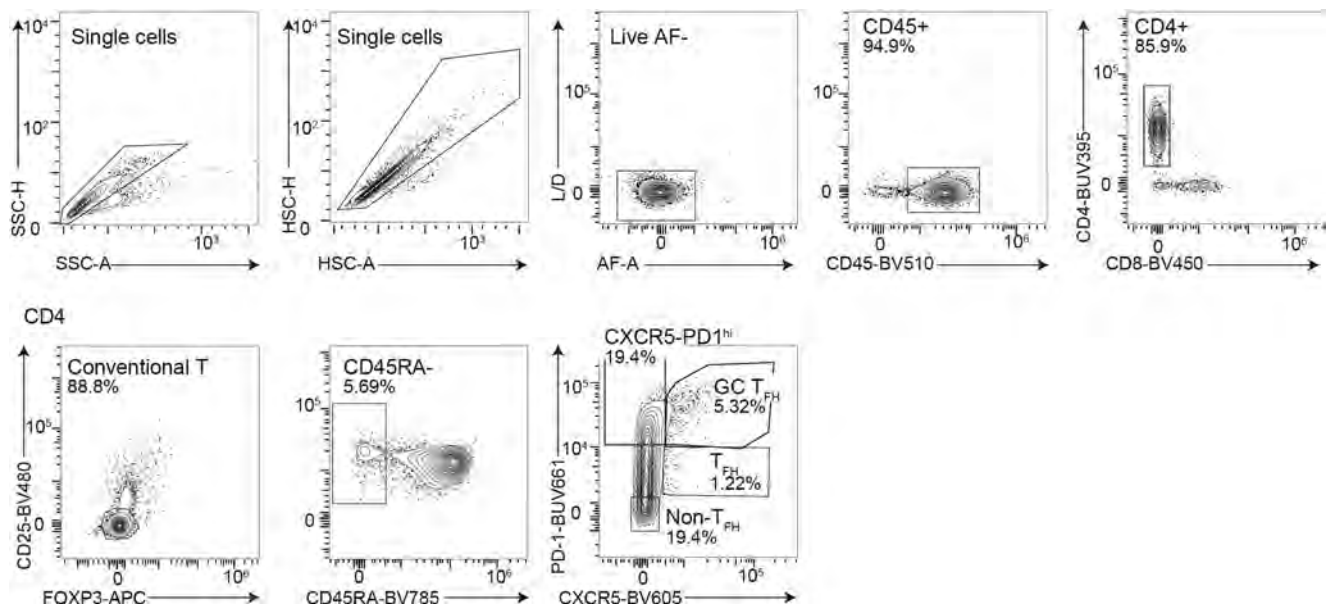
Extended Data Fig. 3 | Gating strategy for analysis of follicular helper T cells (T$_{FH}$) by flow cytometry. Single cells were selected based on side scatter (SSC), height (H) vs area (A) properties, followed by selection of CD45+ live cells. Conventional CD4+ T cells were gated based on lack of CD25 and FoxP3 expression and further gated on CD45RA$^-$ non-naïve cells. Tfh subsets were then identified based on expression of CXCR5 and PD-1. GC = germinal center.
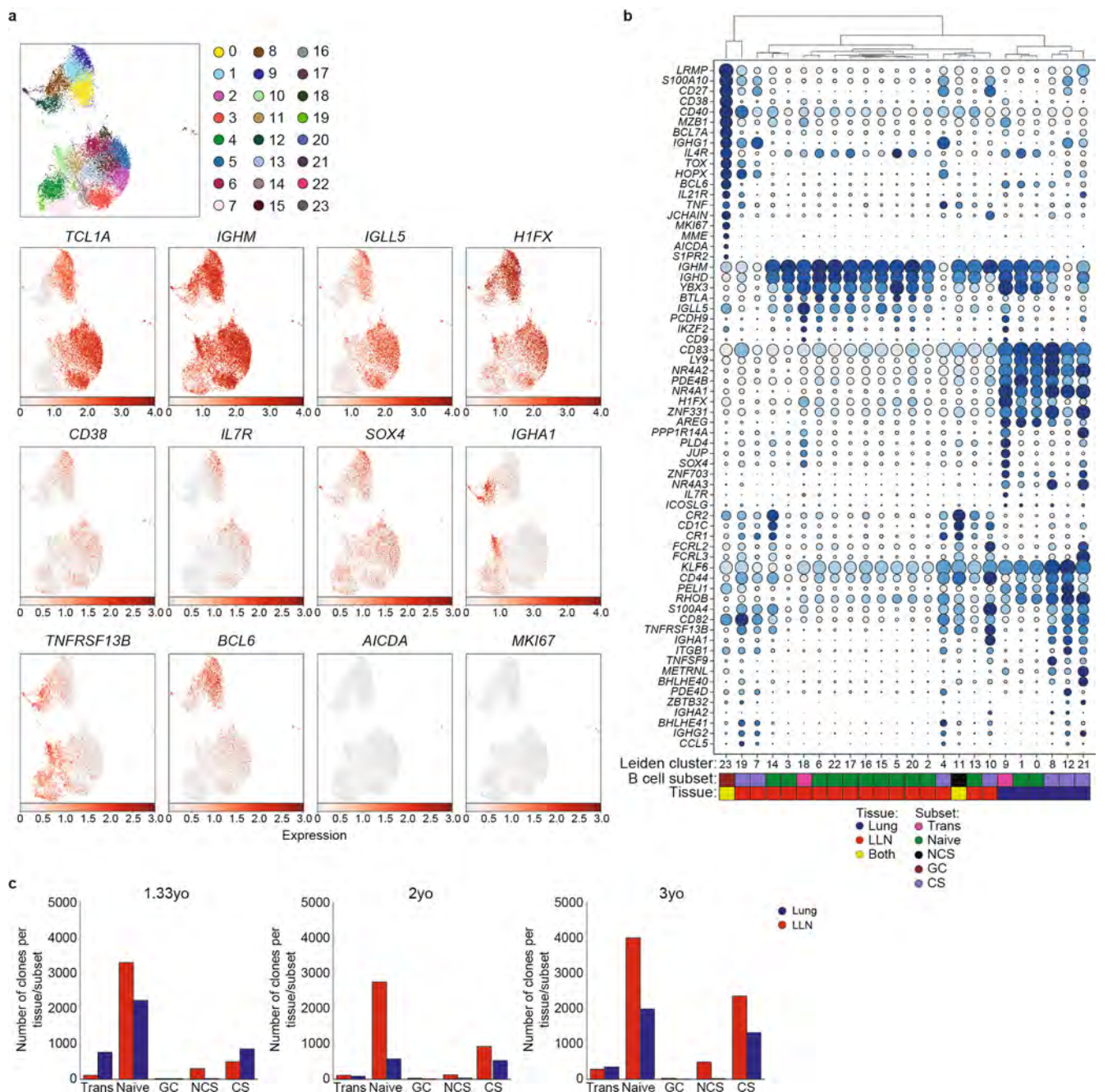
Extended Data Fig. 4 | Determination of B cell subsets in the LLN from Leiden clustering and BCR clone identification. a, UMAP projection of all B cells colored by original Leiden cluster and additional key markers delineating known B cell subsets (bottom rows). **b**, Dotplot shows top differentially expressed genes (DEGs) per Leiden cluster indicating how clusters were collapsed to delineate known subsets. Gene expression values were scaled to a log2 fold change. Dots are colored by average logFC expression and sized by the percentage of cells per cluster that expressed the particular gene. Gene lists were filtered based on a minimum logFC of >1, adjusted P < 0.01 and detected on at least 10% of cells within its cluster. Significant genes calculated using a two-sided Wilcoxon with tie correction. Indicated at the bottom of the dotplot is the Leiden clusters that make up the different B cell subsets based on gene expression and the tissue they are predominantly expressed in. **c**, Barplots indicating the number of clones for each individual donor in each site, stratified by subsets as identified in (b).
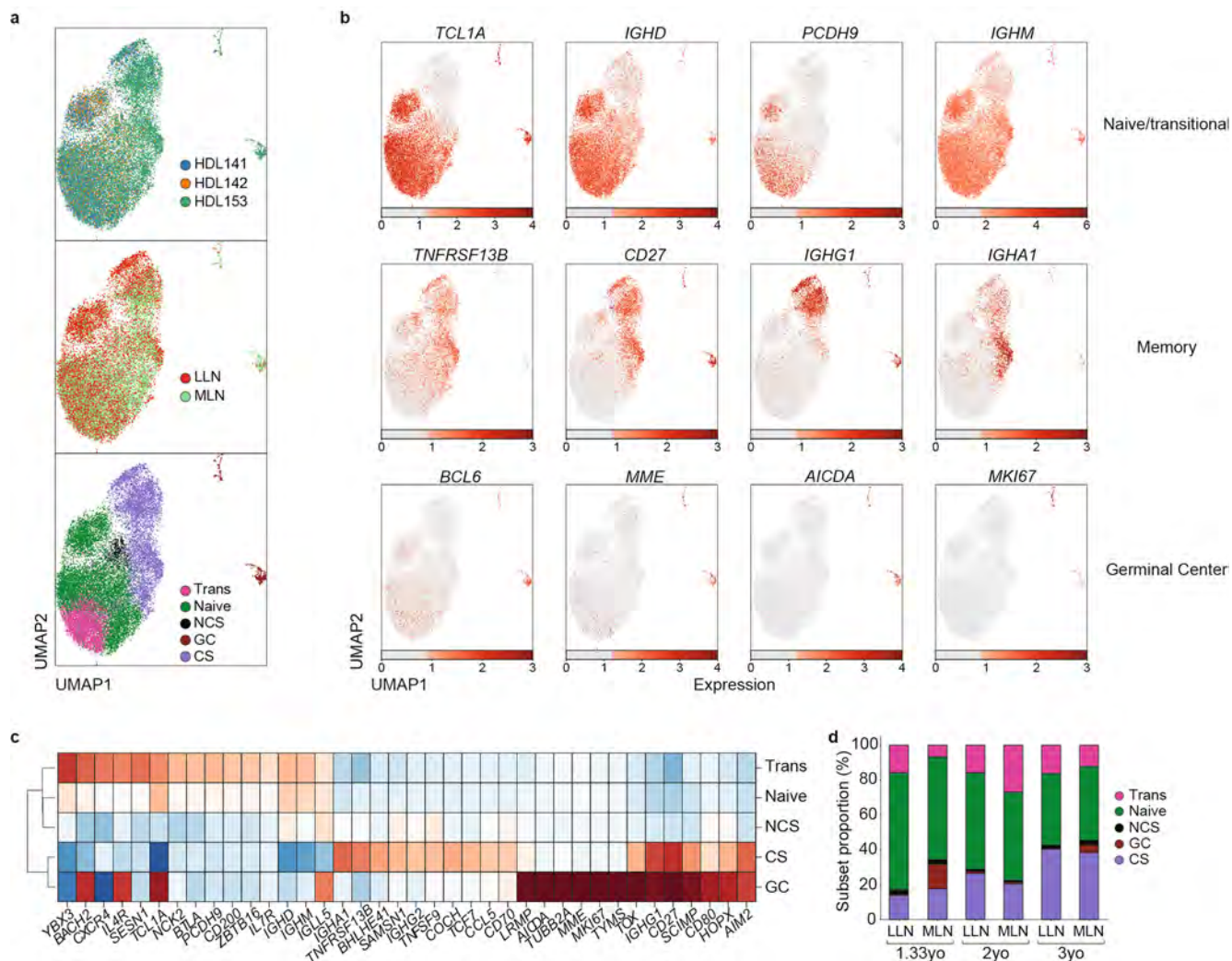
Extended Data Fig. 5 | B cells from lung- and gut-associated lymph nodes exhibit similar transcriptional profiles. Single cell RNA-sequencing data derived from 33,737 B cells from the lung-associated lymph node (LLN) and gut-associated mesenteric lymph nodes (MLN) of three donors aged 1–3 years was analyzed as in Fig. 5 and Extended Data Fig. 4. **a**, UMAP projection of all B cells colored by donor (top), tissue (middle) and subset (bottom), **b**, Expression of top lineage-defining genes for naïve/transitional (top row), memory (middle row) and germinal center (bottom row) B cells in the UMAP. **c**, Heat map showing z-score expression of the top differentially expressed genes associated with the major B cell subsets in each row: Transitional (Trans), Naïve, Non-class switched (NCS) memory, Class-switched (CS) memory, and germinal center (GC) B cells. **d**, Proportional composition of each B cell subset in the LLN and MLN for each donor displayed as a barplot.
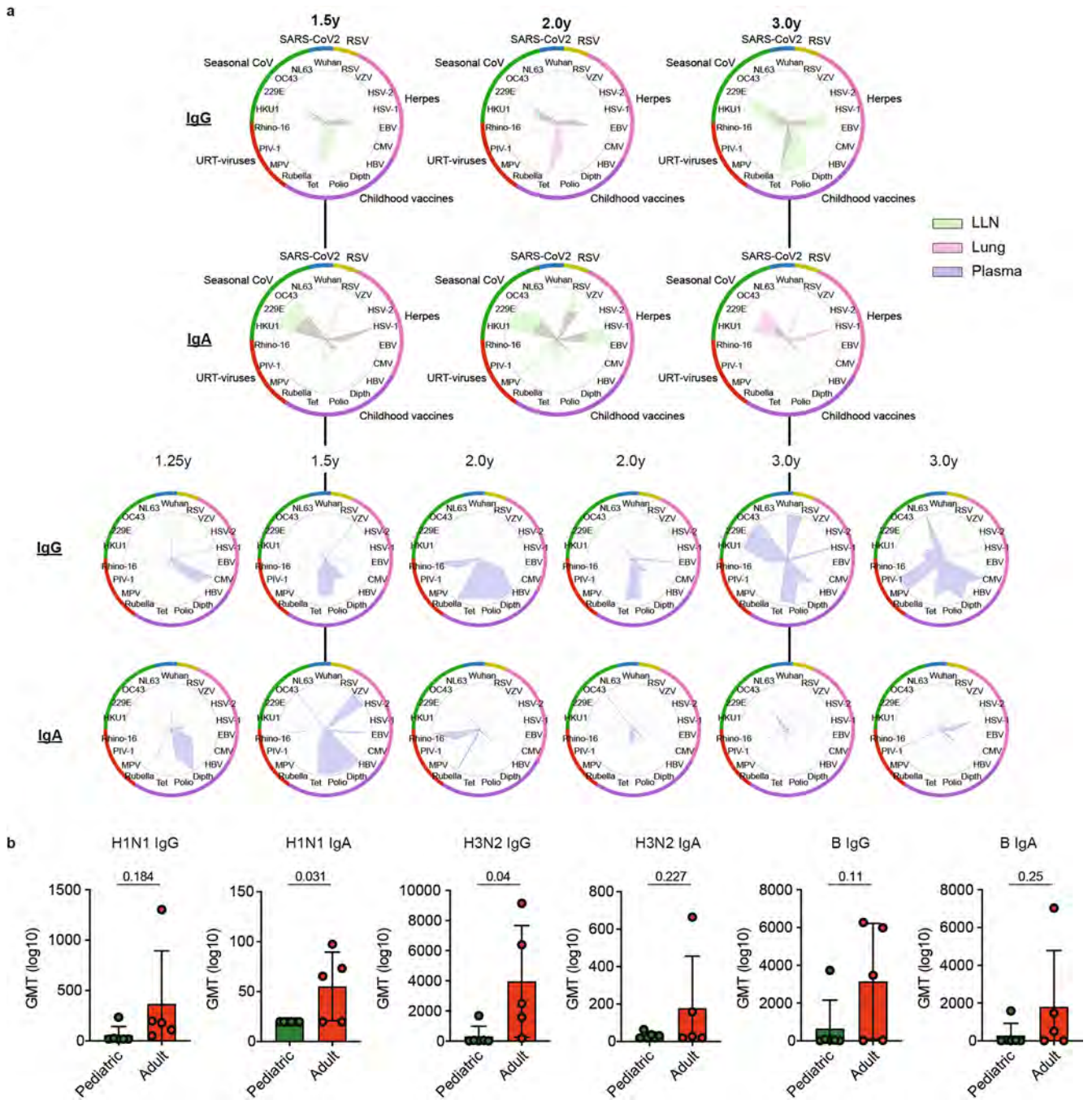
Extended Data Fig. 6 | Antigen specificity of antibodies derived from lungs, lung-associated lymph nodes, and plasma of individual donors. a, Spider plots showing relative reactivities of IgG and IgA for the different antigen types (green=LLN, pink=lung, violet=plasma) from each pediatric donor (lung/LLN: n = 6, plasma: n = 6 (3 samples from organ donors and 3 samples from living blood donors) as determined by the antigen array assay (see methods). Tissue and plasma samples derived from the same donors are linked with vertical lines. **b,** Bar graphs showing geometric mean titers (GMT) of each Influenza strain in serum of adult (red, n = 5) and pediatric (green, n = 6) donors. Significance was determined by unpaired t test and p value generated by two-tailed test (p ≦ 0.05). Error bars indicate SD.

nature portfolio

Corresponding author(s): Donna L. Farber

Last updated by author(s): Jun 6, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	5-laser Cytek Aurora flow cytometer (Cytek Bio) Nikon A1 Scanning Confocal Microscope Illumina NovaSeq 6000 three-laser GenePix 4400 scanner

Data analysis	Flowjo: Tree Star, v10.8.1, https://www.flowjo.com/
	 Imaris: Bitplane; Oxford Instruments, v9.8.2, https://imaris.oxinst.com/
	 Prism: GraphPad Software, v9.3.1 https://www.graphpad.com/scientific-software/prism/
	 ImmuneDB: v0.29.11 (available at github.com/arosenfeld/immunedb)
	 pRESTO: v0.7.0, https;//prestodb.io/
	 IgBLAST: v1.19.0, https://www.ncbi.nlm.nih.gov/igblast/
	 pandas: v1.3.5, https://pandas.pydata.org/
	 seaborn: v0.11.0, https://seaborn.pydata.org/
	 scipy: v1.7.1, https://scipy.org/
	 scrublet: v0.2.3, https://github.com/swolock/scrublet
	 bioinfokit: v2.0.8, https://pypi.org/project/bioinfokit/
	 kallisto: v0.46.2, https://pachterlab.github.io/kallisto/
	 Cell Ranger VDJ: v7.0.1, https://support.10xgenomics.com/single-cell-vdj/software/downloads/latest
	 harmonypy: v0.0.5, https://pypi.org/project/harmonypy/
	 Immcantation pipeline: https://immcantation.readthedocs.io/en/latest/index.html#

Hierarchical clustering: https://scoper.readthedocs.io/en/stable/topics/hierarchicalClones/
GenePix Pro: v7.0, https://support.moleculardevices.com/s/article/GenePix-Pro-Software-Manual-Download-Page

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

BCR sequencing data that support the findings of this study have been deposited in SRA with the accession code: PRJNA847585. The single cell transcriptome and CITE-seq data have been deposited in GEO accession number GSE223646. Other data that support the findings of this study are provided as source data.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	3 females. Please note that the human tissue samples analyzed here were obtained from deceased organ donors which are not human subjects as determined by Columbia University Institutional Review Board (See methods)
Population characteristics	The age of the blood donors was 1.25-3 years old. These donors were otherwise healthy, and approached for routine phlebotomy.
Recruitment	Children who were at the clinic for well-child visit were approached for enrollment in the study. Following parental consent for study participation, blood sampling was obtained in coordination with clinical blood work. Recruitment of donor serum samples was carried out by a clinical coordinator within the lab, therefore negates any potential self-selection bias. No self-selection bias present
Ethics oversight	Institutional Review Board of Columbia University Medical Center

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size is determined by the availability of donor tissues, which were acquired over the last 5 years. These are a precious resource, and were optimally utilized to extract as much information as possible from these limited samples. Younger donor samples were prioritized to capture the highly dynamic events occuring in the early years of life that we present in this study.
Data exclusions	No data were excluded from any of the analyses.
Replication	Figures show data from multiple individuals. Additionally, for imaging, 3-5 sections from lung and 2 sections from lymph nodes were analyzed for each individual.
Randomization	All donors samples obtained were analyzed. The experimental groups in this study were predicated on the tissue of origin of the analyzed cells, as well as age, therefore no randomization was possible.
Blinding	The experiments did not require blinded analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Alexa Fluor 488 Anti-human CD20 Antibody (clone: L26) : Invitrogen, Cat# 53-0202-82, LOT 2247482 Alexa Fluor 647 Anti-human CD10 Antibody (clone: SN5c) : Invitrogen, Cat# 51-0108-42, LOT 2351401 Brilliant Violet(BV) 421 Anti-human CD4 Antibody (clone: RPA-T4) :Biolegend, Cat# 300532, LOT B302124 Alexa Fluor 488 Anti-human CD8 Antibody (clone: HIT8a) :Biolegend, Cat# 300916, LOT B248835 PE Anti-human CD45RO Antibody (clone: UCHL1) :abcam, Cat# ab77217, LOT GR3201396-1 PerCP Anti-human CD45 Antibody (clone: HI30) :Biolegend, Cat# 304025, LOT B331250 BV510 Anti-human CD3 Antibody (clone: SK7) :Biolegend, Cat# 344827, LOT B310746 BUV737 Anti-human CD4 Antibody (clone: SK3) :BD Horizon, Cat# 612748, LOT 0163845 BUV805 Anti-human CD8Antibody (clone: SK1) :BD Horizon, Cat# 612889, LOT 0195677 PE-Cy5 Anti-human CD19 Antibody (clone: HIB19) :Biolegend, Cat# 302210, LOT B263542 BV480 Anti-human IgD Antibody (clone: IA6-2) :BD Horizon, Cat# 566138, LOT 0302854 BUV395 Anti-human IgM Antibody (clone: G20-127) :BD Horizon, Cat# 563903, LOT 0192229 BV650 Anti-human IgG Antibody (clone: G18-145) :BD Horizon, Cat# 740596, LOT 0352502 PE Anti-human IgA Antibody (clone: IS11-8E10) :Miltenyi Biotec, Cat# 130-113-476, LOT 5210509158 BUV615 Anti-human CD27 Antibody (clone: L128) :BD OptiBuild, Cat# 751463, LOT 0329942 BV421 Anti-human CD38 Antibody (clone: HIT2) :BD Horizon, Cat# 562444, LOT 0266689 BV711 Anti-human CD10 Antibody (clone: HI10a) :BD OptiBuild, Cat# 740770, LOT 1074158 BV750 Anti-human CXCR5 Antibody (clone: RF8B2) :BD OptiBuild, Cat# 747111, LOT 1061310 PE/Dazzle-594 Anti-human CD24 Antibody (clone: MI5) :Biolegend, Cat# 311134, LOT B314081 APC Anti-human CD95 Antibody (clone: DX2) :Biolegend, Cat# 305612, LOT B266759 SuperBright 780 Anti-human CD69 Antibody (clone: FN50) :Invitrogen, Cat# 78-0699-42, LOT 2057235 PE/Cyanine 7 Anti-human CD138 Antibody (clone: DL-101) :Biolegend, Cat# 352318, LOT B299531 Alexa Fluor 488 Anti-human CCR7 Antibody (clone: G043H7) :Biolegend, Cat# 353206, LOT B277021 BV605 Anti-human CD45RA Antibody (clone: HI100) :Biolegend, Cat# 304134, LOT B281355 Alexa Fluor 700 Anti-human Ki67 Antibody (clone: Ki-67) :Biolegend, Cat# 350530, LOT B240384 Alexa Fluor 647 Anti-human Bcl-2 Antibody (clone: 100) :Biolegend, Cat# 658706, LOT B316645 BUV395 Anti-human CD4 Antibody (clone: SK3) :BD Horizon, Cat# 563550 BUV496 Anti-human CD31 Antibody (clone: L133.1) :BD Horizon, Cat# 749833 BUV563 Anti-human CCR6 Antibody (clone: 11A9) :BD Horizon, Cat# 749362 BUV615 Anti-human CD161 Antibody (clone: DX12) :BD Horizon, Cat# 751151 BUV661 Anti-human PD-1 Antibody (clone: EH12.1) :BD Horizon, Cat# 750260 BUV737 Anti-human CXCR3 Antibody (clone: 1C6) :BD Horizon, Cat# 741866 BUV805 Anti-human CX3CR1 Antibody (clone: 2A9-1) :BD Horizon, Cat# 749353 BV421 Anti-human CTLA4 Antibody (clone: BNI3) :Biolegend, Cat# 562743 V450 Anti-human CD8 Antibody (clone: RPA-T8) :Biolegend, Cat# 561426 BV480 Anti-human CD25 Antibody (clone: M-A251) :Biolegend, Cat# 566102 BV510 Anti-human CD45 Antibody (clone: 2D1) :Biolegend, Cat# 368526 BV570 Anti-human CD127 Antibody (clone: A019D5) :Biolegend, Cat# 351308 BV605 Anti-human CXCR5 Antibody (clone: RF8B2) :BD Horizon, Cat# 740379 BV650 Anti-human OX40 Antibody (clone: ACT35) :Biolegend, Cat# 563658 BV711 Anti-human CD69 Antibody (clone: FN50) :Biolegend, Cat# 310944 BV750 Anti-human CCR4 Antibody (clone: 1G1) :BD Horizon, Cat# 746980 BV785 Anti-human CD45RA Antibody (clone: HI100) :Biolegend, Cat# 304140 AF488 Anti-human CCR7 Antibody (clone: G043H7) :Biolegend, Cat# 353206 Spark Blue 550 Anti-human CD3 Antibody (clone: SK7) :Biolegend, Cat# 344852 BB700 Anti-human CCR2 Antibody (clone: LS132.1D9) :BD Horizon, Cat# 747847 PerCP-e710 Anti-human TIGIT Antibody (clone: MBSA43) : Invitrogen, Cat# 50-245-943 PE Anti-human 4-1BB Antibody (clone: 4B4-1) :Biolegend, Cat# 309804 PE-Cy5 Anti-human CCR5 Antibody (clone: 2D7/CCR5) :Biolegend, Cat# 556889 Pe-Cy7 Anti-human CD40L Antibody (clone: 24-31) :Biolegend, Cat# 310832 PE-CF594 Anti-human BCL6 Antibody (clone: K112-91) :Biolegend, Cat# 562401 Zombie NIR Live/Dead stain :Biolegend, Cat# 423105 APC Anti-human ICOS Antibody (clone: C398.4A) :Biolegend, Cat# 313510 APC Cy7 Anti-human CD14 Antibody (clone: MΦP9) :Biolegend, 557831 AF647 Anti-human Blimp-1 Antibody (clone: 6D3) :Biolegend, 565002

R718 Anti-human FOXP3 Antibody (clone: 259D/C7) :Biolegend, 566936
APC Fire810 Anti-human CD27 Antibody (clone: QA17A18) :Biolegend, 393214
PerCP-eFluor 710 Anti-human CD103 Antibody (clone: Ber-ACT8) :Invitrogen, Cat# 50-113-0774, LOT 2088791
GTCAACTCTTTAGCG-TotalSeq-C0251_anti-human_Hashtag_1_Antibody (clone:LNH-94; 2M2): Biolegend, Cat#394661
TGATGGCCTATTGGG-TotalSeq-C0252_anti-human_Hashtag_2_Antibody (clone:LNH-94; 2M2): Biolegend, Cat#394663
TTCCGCCTCTCTTTG-TotalSeq-C0253_anti-human_Hashtag_3_Antibody (clone:LNH-94; 2M2): Biolegend, Cat#394665
AGTAAGTTCAGCGTA-TotalSeq-C0254_anti-human_Hashtag_4_Antibody (clone:LNH-94; 2M2): Biolegend, Cat#394667
AAGTATCGTTTCGCA-TotalSeq-C0255_anti-human_Hashtag_5_Antibody (clone:LNH-94; 2M2): Biolegend, Cat#394669
GGTTGCCAGATGTCA-TotalSeq-C0256_anti-human_Hashtag_6_Antibody (clone:LNH-94; 2M2): Biolegend, Cat#394671
TGTCTTTCCTGCCAG-TotalSeq-C0257_anti-human_Hashtag_7_Antibody (clone:LNH-94; 2M2): Biolegend, Cat#394673
CTCCTCTGCAATTAC-TotalSeq-C0258_anti-human_Hashtag_8_Antibody (clone:LNH-94; 2M2): Biolegend, Cat#394675
CAGTAGTCACGGTCA-TotalSeq-C0259_anti-human_Hashtag_9_Antibody (clone:LNH-94; 2M2): Biolegend, Cat#394677

Validation | All antibodies were validated for specificity by the manufacturer. Staining was validated by negative control samples.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation | Following procurement, organs were transported to the laboratory and maintained in cold media supplemented with 5% FBS, penicillin/streptomycin, and glutamine (PSQ). Lymph node (LN) tissues were dissected out from the lung, cleaned of fat and connective tissue. Lungs and LN were chopped into pieces, and incubated with IMDM (Fisher) media containing collagenase D (Sigma) and DNase (Sigma) for 30 minutes at 37°C. Collagenase digestion was inactivated by addition of 1mM EDTA (Fisher). To remove the additional connective tissue, cell suspensions were filtered through 100μM cell strainers and washed with DPBS (Fisher Scientific) media containing FBS (GeminiBio) and EDTA (Fisher). For purity of the cell suspension, cells were centrifuged through density gradient centrifugation using Ficoll-Paque (Fisher) and washed twice with PBS.
These cells were first washed with PBS and were re-suspended with LIVE/ DEAD Fixable Blue Dead Cell stain(Thermo Fisher Scientific) and incubated at room temperature (RT) in the dark for 20 minutes, followed by washing with FACS-buffer (PBS + 2%FBS + 0.1 mM EDTA) and subsequently incubated with TrueStain FcX for 10 minutes in the dark. Next, cells were washed with FACS-buffer and re-suspended in staining antibodies for 30 minutes in RT, followed by two washes with FACS-buffer, and cells were fixed and permeabilized using a Foxp3 Transcription Factor Staining Buffer Set (eBioscience) as per manufacturer's protocols. Lastly, cells were incubated in the dark at 4°C with intracellular antibodies and washed twice in permeabilization buffer. Data were collected on 5-laser Cytek Aurora flow cytometer (Cytek Bio).

Instrument | 5-laser Cytek Aurora flow cytometer (Cytek Bio)

Software | Flowjo: Tree Star, v10.8.1, https://www.flowjo.com/

Cell population abundance | All cell population frequencies were clearly indicated. No sorting experiments were performed.

Gating strategy | (B cell gating)
Lymphocytes were selected based on forward (FSC) vs side scatter (SSC) height (H) properties, then single cell gate was applied based on SSC and H vs area (A) properties, followed by selection of CD45+ live cells. B cells were selected by expression of CD19, subsets were identified based on expression of IgD vs IgM. IgD+ cells were then divided based on CD27 expression. Based on expression of CD38 and CD10, IgD+CD27- cells were further divided into transitional and naïve B cells, and IgD+CD27+ cells to IgD+GC and NCS memory B cells. IgD-IgM- cells were also delineated based on CD27 expression and IgD- GC B cells were gated based on CD38 and CD10 expression. CD27+IgD-IgM-CD10- B cells were further delineated into IgA memory and IgG memory B cells.
(TFH gating)
Single cells were selected based on forward (FSC) and side scatter (SSC), height (H) vs area (A) properties, followed by selection of CD45+ live cells. CD45RA- CD4 conventional T cells were selected by expression of CD4 and lack of CD25, Foxp3 and CD45RA expression. Subsets were identified based on expression of CXCR5 vs PD-1.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.