

RESEARCH ARTICLE SUMMARY

IMMUNOLOGY

Spatial transcriptomics of B cell and T cell receptors reveals lymphocyte clonal dynamics

Camilla Engblom[†], Kim Thrane[‡], Qirong Lin[‡], Alma Andersson, Hosein Toosi, Xinsong Chen, Embla Steiner, Chang Lu, Giulia Mantovani, Michael Hagemann-Jensen, Sami Saarenpää, Mattias Jangard, Julio Saez-Rodriguez, Jakob Michaëlsson, Johan Hartman, Jens Lagergren, Jeff E. Mold^{*†}, Joakim Lundeberg[‡], Jonas Frisén[‡]

INTRODUCTION: B cells and T cells respond to infections and inflammation, regulate tissue homeostasis, and maintain immunological memory. The targeted reactivity of B and T cells is determined by their clonally heritable antigen receptors, encoded by immunoglobulin (IG) and T cell (TR) antigen receptor genes. Spatial analysis of lymphocyte clonality may link specific antigen receptors to their regulatory niche or antigen (tumor-associated, self, or foreign). This, in turn, may help identify and harness antigen-specific clones for therapy.

RATIONALE: Single-cell technologies permit the study of antigen receptors at a cellular level but lack spatial resolution. Current spatial transcriptomics methods locate gene expression in tissues, but do not retain full-length antigen receptor sequences. Thus, methodologies that comprehensively map B and T cell receptor sequences within tissues are needed. We therefore developed spatial transcriptomics for variable, diversity, and joining (VDJ) sequences (Spatial VDJ), which spatially resolves full-length IG and TR transcripts, tissue mor-

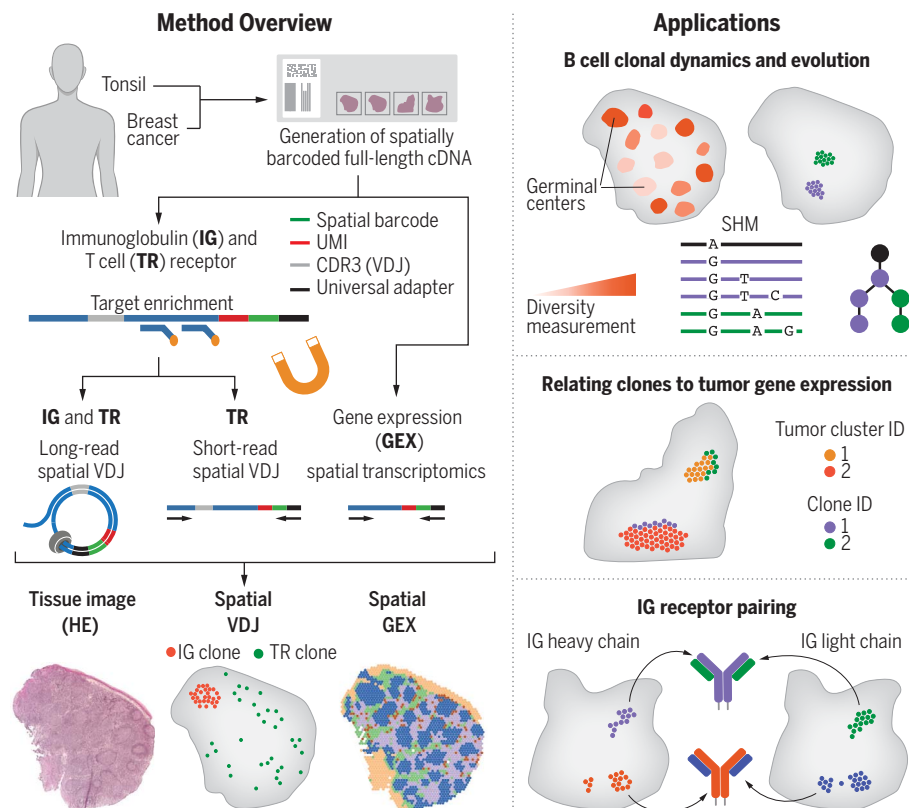
phology, and the whole transcriptome simultaneously in human tissue.

RESULTS: We developed two versions of Spatial VDJ that are compatible with whole-transcriptome 3' arrays for frozen human tissue sections: (i) long-read (LR) Spatial VDJ for full-length IG and TR antigen receptor transcripts and (ii) short-read (SR) Spatial VDJ for TR only. For both approaches, we enriched for antigen receptor transcripts using hybridization capture targeting the IG and TR constant regions.

Spatial VDJ captured thousands of B and T cell receptor sequences in human tonsil tissue sections. The antigen receptor distribution matched canonical B and T cell locations and spatial gene expression patterns. Amplified clonal sequences were also confirmed by orthogonal methods, and the LR and SR Spatial VDJ versions to capture TR clones yielded comparable results. We found spatial congruency between paired receptor chains and developed a computational framework that identified IG heavy and light chain receptor pairs validated by single-cell VDJ analysis in human breast cancer. Distinct IG receptors spatially segregated across different tumor-associated areas within the same tumor, opening up the possibility of using Spatial VDJ as a first screen for tumor-associated antibodies.

Spatial VDJ also uncovered B cell clonal family diversity measures within and between distinct B cell follicles (containing germinal centers), which are key sites for B cell clonal expansion and maturation. Spatial phylogenetic analysis of B cell clones identified simultaneous lineage tree branching and geographical spreading to distinct germinal centers.

CONCLUSION: This study describes a method that allows for full-length antigen receptor mapping within human tissue. Spatial VDJ reveals B and T cell clonal dynamics in both lymphoid and cancer tissue. This includes linking clones to tumor-associated gene expression programs and spatially reconstructing B cell clonal diversity and lineage trajectories across their anatomical niches. Ultimately, Spatial VDJ may be useful to discriminate B and T cell clonal dynamics in the context of infections, vaccination, and anticancer immune responses, with the goal of harnessing antigen-specific clones for engineered cell- and antibody-based therapeutics. ■



Spatial transcriptomics-based method to map B cell and T cell receptors. Targeted capture of antigen receptor sequences from spatial transcriptomics gene expression libraries resolves full-length receptor distribution alongside tissue anatomy and whole transcriptome in human tissue sections. Spatial VDJ uncovers B and T cell clonal dynamics, including linking clones to tumor-associated gene expression, pairing receptor chains, and reconstructing spatial B cell lineage trajectories.

The list of author affiliations is available in the full article online.

*Corresponding author. Email: jeff.mold@ki.se

†These authors contributed equally to this work.

‡These authors contributed equally to this work.

Cite this article as C. Engblom *et al.*, *Science* **382**, eadf8486 (2023). DOI: 10.1126/science.adf8486

READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.adf8486>

RESEARCH ARTICLE

IMMUNOLOGY

Spatial transcriptomics of B cell and T cell receptors reveals lymphocyte clonal dynamics

Camilla Engblom^{1†}, Kim Thrane^{2†}, Qirong Lin^{1†}, Alma Andersson², Hosein Toosi³, Xinsong Chen⁴, Embla Steiner¹, Chang Lu⁵, Giulia Mantovani¹, Michael Hagemann-Jensen¹, Sami Saarenpää², Mattias Jangard⁶, Julio Saez-Rodriguez⁵, Jakob Michaëlsson⁷, Johan Hartman^{4,8}, Jens Lagergren³, Jeff E. Mold^{1*†}, Joakim Lundeberg^{2†}, Jonas Frisén^{1†}

The spatial distribution of lymphocyte clones within tissues is critical to their development, selection, and expansion. We have developed spatial transcriptomics of variable, diversity, and joining (VDJ) sequences (Spatial VDJ), a method that maps B cell and T cell receptor sequences in human tissue sections. Spatial VDJ captures lymphocyte clones that match canonical B and T cell distributions and amplifies clonal sequences confirmed by orthogonal methods. We found spatial congruency between paired receptor chains, developed a computational framework to predict receptor pairs, and linked the expansion of distinct B cell clones to different tumor-associated gene expression programs. Spatial VDJ delineates B cell clonal diversity and lineage trajectories within their anatomical niche. Thus, Spatial VDJ captures lymphocyte spatial clonal architecture across tissues, providing a platform to harness clonal sequences for therapy.

B cells and T cells respond to infections and inflammation, regulate tissue homeostasis, and maintain immunological memory. The targeted reactivity of B and T cells is determined by their clonally heritable antigen receptors. Single-cell technologies permit the study of antigen receptors at a cellular level but lack spatial resolution. Current spatial transcriptomics (ST) methods do not retain full-length antigen receptor transcripts because library preparation steps fragment away the complementarity determining region 3 (CDR3) needed to define lymphocyte clonality (fig. S1A). Spatial analysis of lymphocyte clonality may link specific antigen receptors to tumor-associated, self, or foreign antigens, which may help identify and harness antigen-specific clones for therapy. Several techniques have been developed to spatially resolve antigen receptor sequences within human tissues. These include polymerase chain reaction (PCR)-based amplification of T cell receptors

(TCRs) from either ST (1) or Slide-seq libraries (2), laser-guided microdissection coupled with sequencing (3, 4), and capture through deep sequencing of rare longer immunoglobulin (IG) transcripts containing the CDR3 region sequence that remained post-fragmentation in ST libraries (5). However, methodologies that comprehensively map full-length B cell receptor (BCR) and TCR sequences and their lineage relationships within tissues in a high-throughput, user-friendly manner are lacking.

In this work, we developed spatial transcriptomics for variable, diversity, and joining (VDJ) sequences (Spatial VDJ), which maps full-length BCR and TCR transcripts in human tissue (Fig. 1A). Spatial VDJ is an extension of Vismium Spatial Gene Expression (Spatial GEX) for fresh frozen tissue, which relies on 3'-targeted barcoding of polyadenylated RNAs in a tissue section (fig. S1A) (6). Spatial GEX libraries from human tonsil tissue showed ample BCR and TCR constant gene expression (fig. S1, B to E). Still, they constituted a small fraction of the total library. Prior to the fragmentation steps, the CDR3 region sequence could be PCR amplified from spatially barcoded full-length Vismium cDNA (fig. S1F). Thus, sequencing the CDR3 region and the spatial barcode from Spatial GEX cDNA libraries should be a feasible approach to determine both clonality and the location of antigen receptor transcripts.

We developed two Spatial VDJ versions (Fig. 1A): (i) long-read (LR) Spatial VDJ, which generates spatially barcoded libraries of full-length IG and T cell (TR) antigen receptor transcripts, and (ii) short-read (SR) Spatial VDJ for TR sequences only, which uses a two-step seminested PCR-based approach with CDR3-

adjacent V primers (data S1). For both approaches, we used hybridization capture probes directed at the TR and IG constant regions to enrich for antigen receptor transcripts while preserving the spatial barcode, unique molecular identifier (UMI), and full-length receptor sequences (7, 8). Hybridization capture significantly increased the TR and IG UMI and spatial barcode counts compared with the non-enriched Spatial GEX library and preserved the IG and TR transcript distribution (fig. S2).

Spatial VDJ permits high-fidelity B and T cell clonal mapping

We applied LR and SR Spatial VDJ to the human tonsil owing to its abundance and spatial compartmentalization of B and T cell lineages (Fig. 1, A to E, and fig. S3). We identified 62,533 unique IG (*IGH*, *IGK*, and *IGL*) and TR (*TRB* and *TRA*) clonal sequences (fig. S4A), hereafter referred to as “clonotypes” or “clones.” We detected fewer TRA than TRB clones, likely because of *TRA*'s lower transcriptional abundance (fig. S1D) (7). The TR clone counts were slightly lower in SR than LR Spatial VDJ libraries (Fig. 1B), but the SR method recapitulated more than half of the clonal sequences, with increased sharing for expanded clones (fig. S4). Additionally, clonal abundance correlated well between LR and SR Spatial VDJ datasets (fig. S4E), and clones shared between both datasets showed greater expansion (fig. S4, E to G). We visualized the overall antigen receptor distribution (Fig. 1, C to E) and individual expanded clones (figs. S5 and S6). Notably, the same TR clonal sequence retrieved by either LR or SR Spatial VDJ exhibited overlapping spatial distribution, demonstrating that these approaches yield comparable results.

The antigen receptor distribution differed across the tonsil gene expression landscape (Fig. 1F and fig. S7). The unique IG clone count was the highest within tissue regions annotated as “isotype-switched cells,” “squamous epithelium,” and “B cell follicles/germinal centers (GCs)” (Fig. 1G). Isotype here refers to the *IGH* constant chain (*IGHA*, *IGHD*, *IGHG*, *IGHG*, and *IGHM*) that can be exchanged to confer diverse antibody effector functions while retaining antigen specificity (9). The *IGH* isotype fraction was consistent across sections (fig. S8A) and corresponded well to the Spatial GEX cluster groups (fig. S8B). The unique TR clone count was the highest in T cell zone-associated areas (Fig. 1G). Antigen receptor chain expression was also concordant between the Spatial VDJ and GEX datasets (fig. S8 and data S2). We next inferred the cell-type distribution by deconvoluting the Spatial GEX data using the *stereoscope* software package (10, 11) (Fig. 1H and fig. S9, A to C), which matched well-established B cell, plasma cell, and T cell-associated marker-gene expression

¹Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden. ²SciLifeLab, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden. ³SciLifeLab, Computational Science and Technology department, KTH Royal Institute of Technology, Stockholm, Sweden. ⁴Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden. ⁵Heidelberg University, Faculty of Medicine and Heidelberg University Hospital, Institute for Computational Biomedicine, Heidelberg, Germany. ⁶ENT Unit, Sophiahemmet University Research Laboratory and Sophiahemmet Hospital, Stockholm, Sweden. ⁷Center for Infectious Medicine, Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden. ⁸Department of Clinical Pathology and Cancer Diagnostics, Karolinska University Hospital, Stockholm, Sweden.

*Corresponding author. Email: jeff.mold@ki.se

†These authors contributed equally to this work.

‡These authors contributed equally to this work.

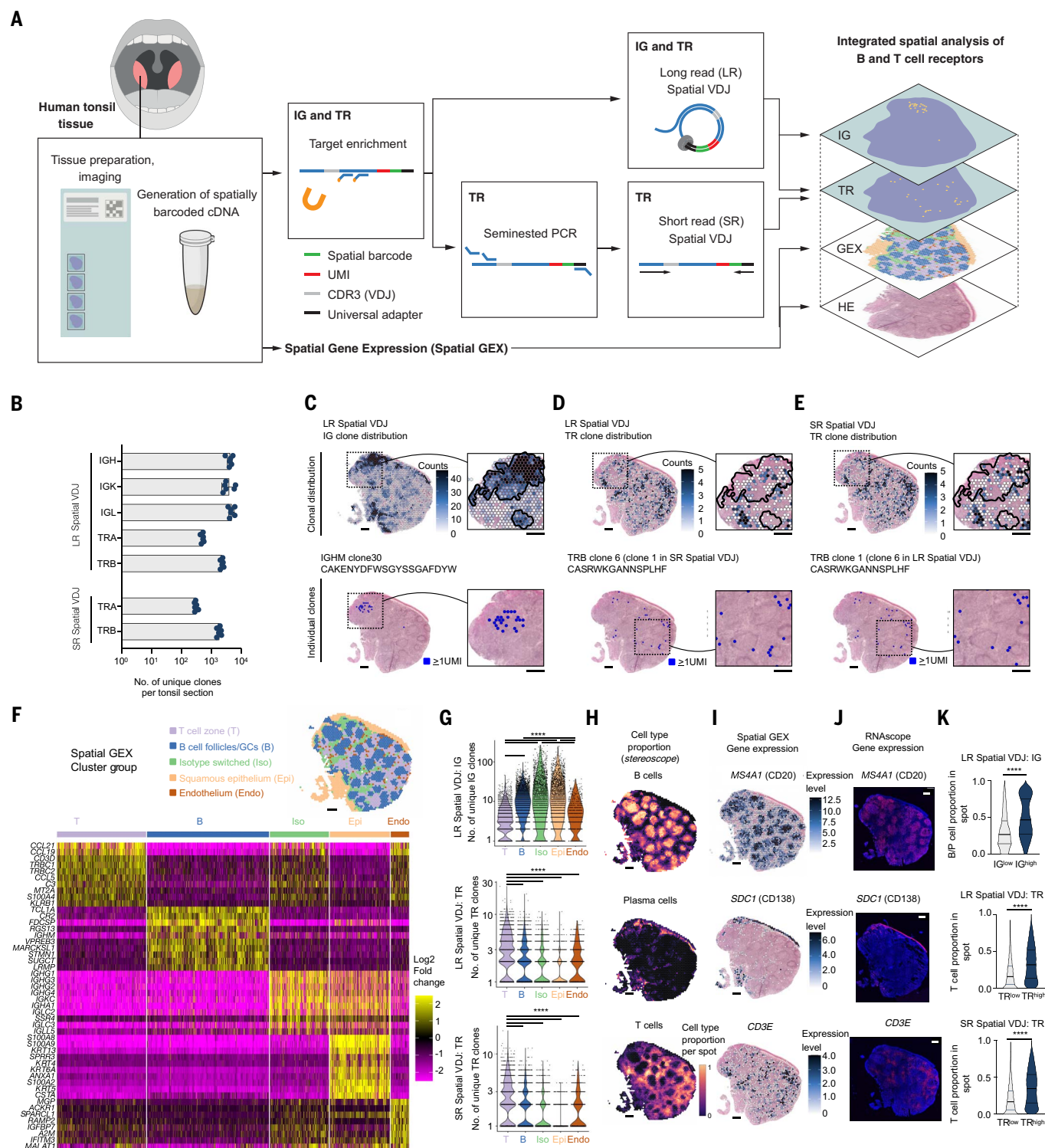


Fig. 1. Spatial VDJ maps antigen receptors in human tonsil—matching canonical B and T cell distributions. (A) Overview of long-read (LR) and short-read (SR) Spatial VDJ method in human tonsil tissue. (B) Unique *IGH*, *IGK*, *IGL*, *TRA*, and *TRB* clonotype sequences for each tonsil section ($n = 6$). (C to E) Unique IG (C) and TR (D) clonal receptor sequence distribution for LR Spatial VDJ and TR (E) clonal distribution for SR Spatial VDJ with a maximum cut-off above the 95th percentile. The black outline indicates regions with high IG clonal diversity. (F) Spatial GEX cluster group spatial distribution on tonsil tissue (top) and enriched genes (bottom). (G) Number of unique IG or TR clonotypes across Spatial GEX cluster groups (Kruskal–Wallis test, Dunn’s multiple comparison test;

**** $P < 0.0001$). (H) Visualization of major cell type distribution within the tonsil defined by single-cell deconvolution. (I) Spatial GEX of canonical B (top), plasma (middle), and T (bottom) cell markers (*MS4A1*, *SDC1*, and *CD3E*, respectively). Legends show UMI counts. (J) In situ hybridization of genes shown in (I) on proximal tissue sections to those analyzed by Spatial GEX or VDJ. (K) Cell type proportion for spots containing high or low clonotype counts (IG or TR) (Mann–Whitney U test; **** $P < 0.0001$). Scale bars, 500 μm . [(A) to (K)] Spatial VDJ was performed in two batches (two experiments). [(F), (G), and (K)] Data are pooled from six individual sections. [(C) to (E), (H), and (I)] Data are from one representative section (Tonsil section number 17). GC, germinal center.

(*MS4A1*, *SDCI*, and *CD3E*, respectively) shown by Spatial GEX (Fig. 1I) and in situ hybridization (Fig. 1J and fig. S9D). For spots that contained IG high (IG^{high}) versus low (IG^{low}) clonal content, the proportion of B cells and plasma cells was significantly higher, whereas TR^{high} versus TR^{low} spots had elevated T cell proportions (Fig. 1K and fig. S9E). The antigen-receptor distribution also matched known lymphoid tissue-associated chemokine gradients (fig. S10) (12). Thus, Spatial VDJ delineates diverse spatial distributions for B and T cell clones consistent with canonical secondary lymphoid tissue architecture and cell distribution.

The spatial segregation of individual clonotypes was highly reproducible and the most abundant clones were detected across sequential tonsil sections (Fig. 2, A to C, and figs. S5A, S6A, and S11A). However, TRA clones were rarely found across all six sections, likely because of lower TRA capture. Most clones appeared in one section, but expanded clones were detected across multiple or all sections, corresponding to higher UMI counts per clone (fig. S11, B and C). These results were expected, as many lymphocytes in secondary lymphoid organs are naïve or nonexpanded clones (17) and unlikely to be in multiple array spots or sections. To validate Spatial VDJ-defined clonal sequences, we analyzed nearby or adjacent sections from the same sample by a VDJ-targeted Smart-seq3 (SS3) protocol for bulk RNA (figs. S3A and S12). Bulk SS3 and Spatial VDJ showed comparable IGH isotype fractions (fig. S12, B and C). Of the IG clonotypes identified by LR Spatial VDJ, almost half were found by bulk SS3 VDJ (Fig. 2D and fig. S12, G and H). Bulk SS3 VDJ reproduced approximately a third of the TR clones captured by Spatial VDJ. The clones detected by multiple methods had higher UMI counts on average, and more expanded clones were more likely to be detected across methods (Fig. 2E and fig. S12, I to L). Additionally, the clone fraction of shared clonal sequences correlated well between datasets (Fig. 2F), in line with Slide-TCR-seq (2). Bulk SS3 VDJ captured substantially more TR clones, most of which were found in a single tissue section. Our combined analysis suggested that this was due to multiple factors, including better capture of rare mRNA transcripts by Bulk SS3 VDJ. Nonetheless, Spatial VDJ reproducibly detects expanded B and T cell clonal sequences.

Spatial VDJ captures B and T cell clones in human tumors

We next extended Spatial VDJ to human breast tumor tissue. Tumor infiltration by lymphocytes correlates with positive disease outcomes and treatment response across solid cancers, including breast cancer (13, 14). Mapping B and T cell clones within tumors may help identify and therapeutically harness individual or

groups of clones to boost tumor-associated immunity. We generated Spatial GEX and VDJ libraries from two untreated HER2-positive breast tumors (P1 and P2) (Fig. 3, figs. S13 to S19, and data S3). To better capture intrapatient heterogeneity, biopsies were divided into multiple regions, denoted as Region A (RegA) to RegE, each containing tumor, a tumor border, and adjacent nontumor tissue. We identified thousands of unique IG clonal sequences and hundreds of unique clonal TR sequences per patient (fig. S16) and section (Fig. 3C), with the highest number of distinct clones found outside the tumor area (Fig. 3D and fig. S17). The highest IG clonotype counts were in array spots enriched for plasmablasts and *APOE*-expressing macrophages (Fig. 3D and figs. S17C and S19), suggesting an association between tumor-infiltrating macrophages and plasma cells, as we observed previously (10). The spots with the highest TR receptor count were enriched for both myeloid and T cell-associated genes (fig. S17D) and cell subsets (fig. S19), which in P2 resembled mature tertiary lymphoid structures (TLS), but not in P1 (15). Some IGH clones were shared between the TLS-like and plasmablast-rich areas (fig. S17E), suggesting that at least some of the tumor-associated plasmablasts may derive from TLS-activated B cells.

Approximately half of the Spatial VDJ-defined TR receptor chains were confirmed by matched single-cell RNA sequencing (RNA-seq) VDJ (scVDJ) analysis of the same tumors (figs. S20 and S21A), similar to a spatial PCR-based TRB amplification protocol (1). For IG clones, we confirmed only 3.0 and 2.9% of the LR Spatial VDJ dataset by scVDJ. We detected approximately 10-fold more TR than IG clones by scVDJ, with the opposite ratio in Spatial VDJ (Fig. 3C and fig. S21B). Our analysis suggests that Spatial VDJ (compared with scVDJ) more easily captures plasma cell clones, likely owing to the abundant IG transcript expression per plasma cell (fig. S21C) (16) and sensitivity of plasma cells to single-cell dissociation protocols. Conversely, the reduced detection of TR clones by Spatial VDJ likely resulted from inefficient capture of rare transcript on the array (owing to low TR expression per cell and/or low cell number per clone) (fig. S21, C and D). These limitations may have also applied to rare B cell clones, because B cells express IG transcripts at lower levels compared with plasma cells. Supporting the notion that Spatial VDJ captures more expanded clones, the scVDJ TR clones that were shared with Spatial VDJ had more cells per clone (fig. S21E), and scVDJ clones detected in multiple tumor regions were more frequently captured by Spatial VDJ (fig. S20, F to I).

Expanded tumor-infiltrating B and T cell clones may indicate local antigen-specific interactions with tumor-regulatory potential (14, 17).

Tumors contain phenotypically diverse lymphocytes, but identifying which B or T cell clones and subsets interact directly with tumors is challenging. Spatial VDJ coupled with scVDJ on matched tumor biopsies revealed that expanded T cell clones were enriched in the tumor border relative to nontumor regions or within the tumor (Fig. 3E). These clones exhibited phenotypes ranging from CD4 or CD8 effectors to regulatory T cells (fig. S22). Thus, Spatial VDJ provides a platform to define putative tumor-reactive clonotypes according to spatial distribution and cell state, which is not possible when analyzing dissociated tissues. We found this particularly relevant for expanded B cell clones. In P2, we identified four distinct tumor clusters (figs. S15B and S18B). Different IG clones were distinctively enriched in the border to individual tumor clusters (Fig. 3F and fig. S23), two of which (IGLclone6 and IGLclone39) matched to class-switched B cells (IGHA1 and IGHG1, respectively). Most clones were enriched in limited parts of the tumor border, indicating localized antigen-binding responses (e.g., IG clonal colocalization with distinct tumor subclones expressing different antigens). All tumor clusters were enriched for cancer cells (figs. S18B and S19B) but exhibited differently enriched genes (fig. S23E). This intratumoral heterogeneity may result in differential tumor outgrowth and treatment responses within the same tumor [e.g., by distinct IG clones driving antibody-dependent tumor-cell killing in different tumor locations (14)]. Thus, the combination of unsupervised methodology with Spatial VDJ can identify potential tumor-reactive antigen receptors enriched in specific tumor regions.

Paired BCR or TCR sequences determine antigen binding. Spatial VDJ captures individual antigen receptor transcripts and cannot therefore define receptor pairs by itself. However, paired receptor chains should colocalize in the tissue. Using the matched single-cell VDJ dataset to verify paired receptor usage, we observed a high spatial correlation for most IG pairs across biopsied regions (Fig. 3, G and H, and fig. S24A). We lacked sufficient data points for TR pairs to evaluate their spatial distribution in a statistically robust manner. To predict antigen receptor pairing based on the Spatial VDJ data alone, we developed a computational framework called “repair.” We tested our model on data from P1 at different cut-offs of receptor chain abundance (minimal spot count per clone) and the pairing score (fig. S24, B and C, and data S4), which approximated the confidence of the pairing. We achieved the maximum accuracy (95.8%) and number of correct pairs ($n = 23$) when the IGH clonal chain had to be present in at least 10 spots and the pairing score was at least 0.35. At these settings, we obtained 157 de novo identified receptor pairs (i.e., receptor chains that were not both present in our scVDJ

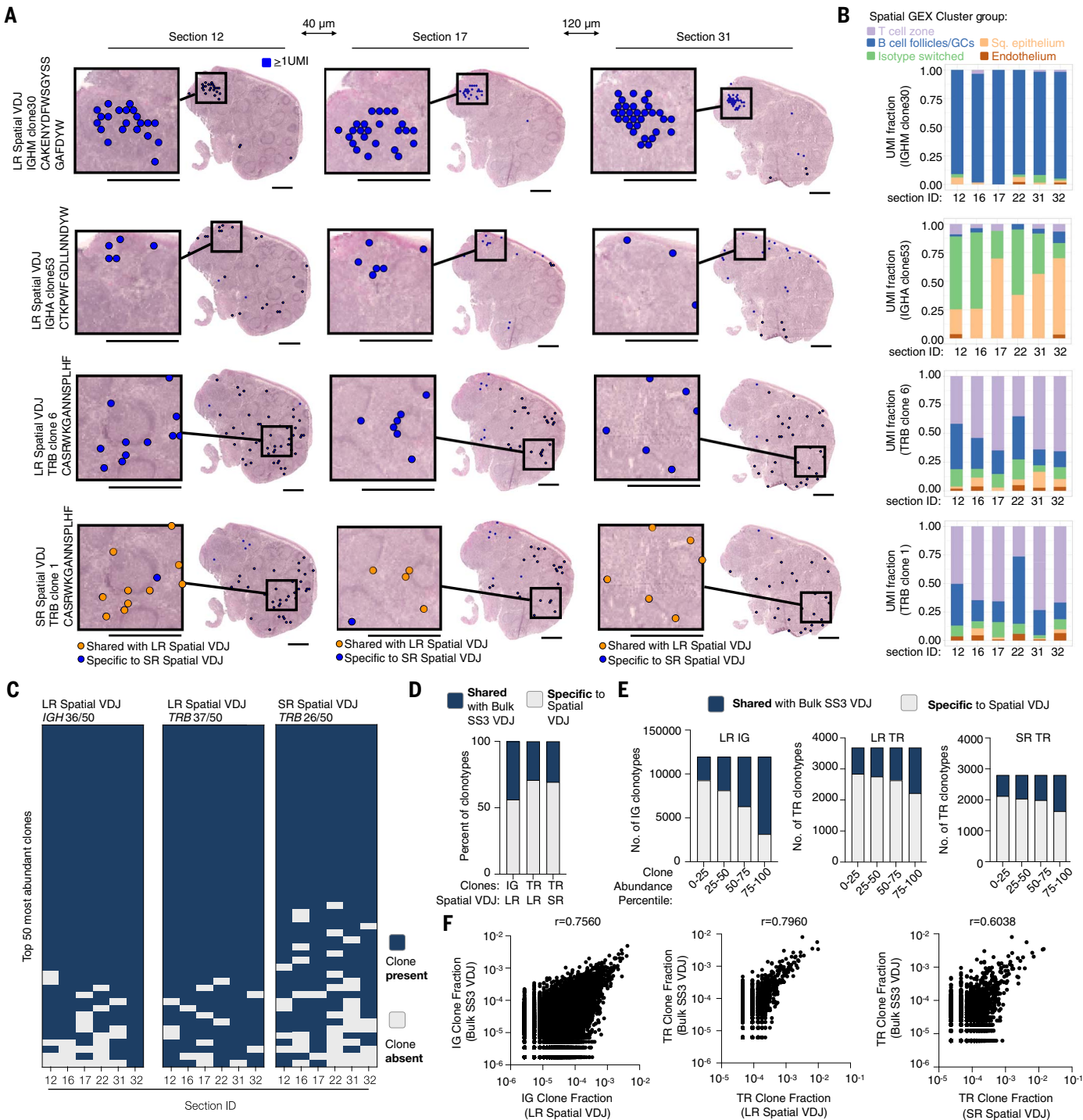


Fig. 2. Spatial VDJ reproducibly delineates B and T cell clones across tissue sections and amplifies validated clonal sequences. (A) Spatial distribution of individual clones across representative tonsil sections from the same tonsil sample. Distance between sections (orthogonal to the tissue plane) above the arrows. All sections ($n = 6$) are shown in figs. S5 (IG) and S6 (TR). Matched *TRB* clonotype sequence (identical ntCDR3; see materials and methods) captured by LR and SR Spatial VDJ is shown, and concordant spots are indicated in orange (bottom row). Clone ID numbering occurs independently within LR and SR Spatial VDJ datasets. Scale bars, 1 mm. (B) UMI distribution for each clone in (A) across Spatial GEX cluster groups and sections. (C) Presence or

absence of the top 50 *IGH*, *IGK*, or *TRB* clonotypes (based on UMI count and sorted on number of sections) across tonsil sections. (D) Percent shared and not shared clonotype sequences between Spatial VDJ and Bulk SS3 VDJ validation (see fig. S3A). (E) The number of shared versus not shared clonotype sequences with Bulk SS3 VDJ across increased clonal abundance in LR (left, middle) or SR (right) Spatial VDJ datasets. (F) Clone fraction of each shared *IG* (left) or *TR* (middle, right) clonotypes between the datasets listed. Pearson correlation coefficient (r) is denoted. [(A) to (F)] Spatial VDJ was performed in two batches (two experiments). Bulk SS3 VDJ was performed in a single experiment, with four technical replicate reactions per tissue section. [(D) to (F)] Data are pooled from six individual sections.

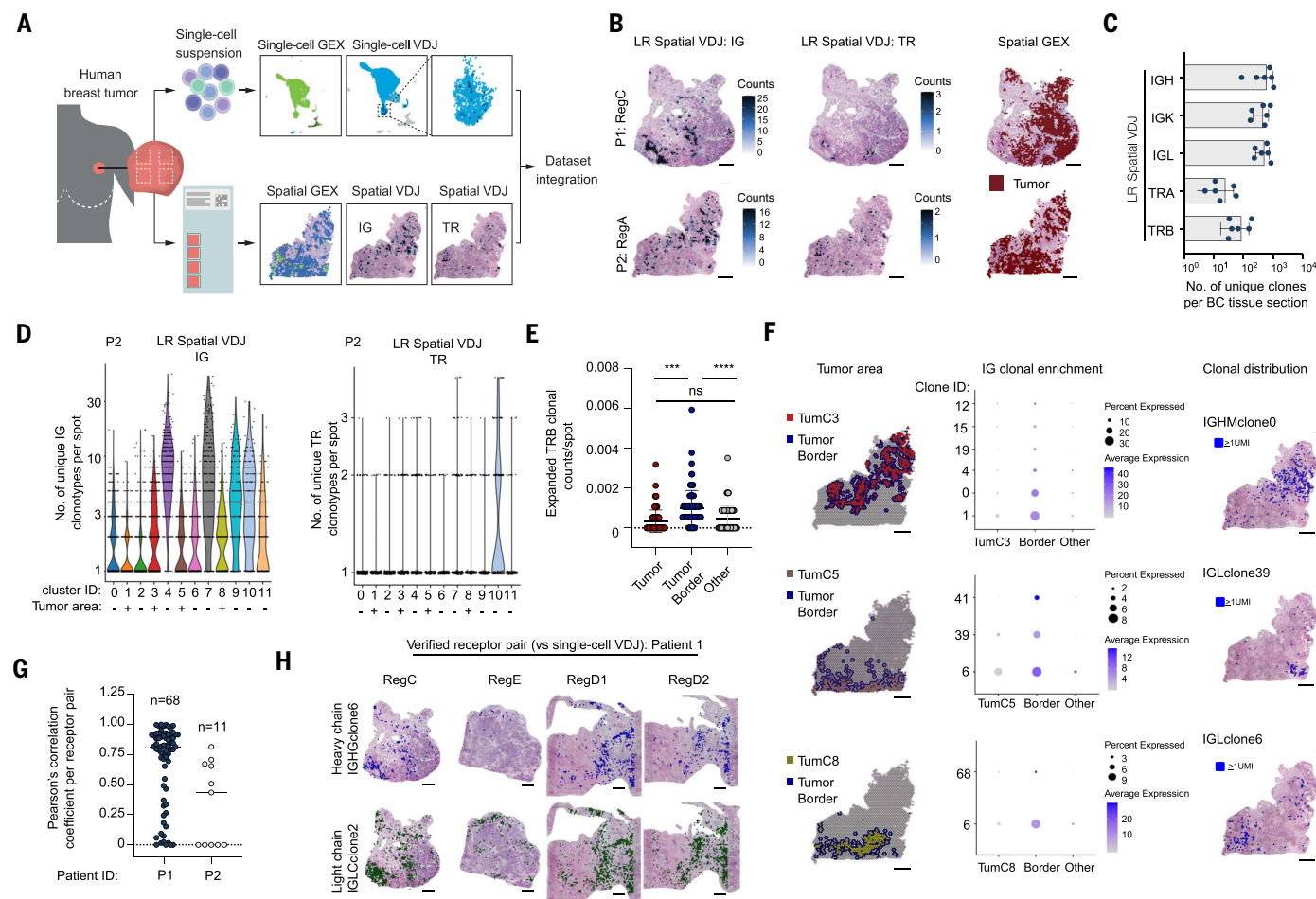


Fig. 3. Distinct B cell lineage clones spatially segregate along different breast tumor areas. (A) Outline of the breast cancer patient experiments. (B) Unique clonal distribution for LR Spatial VDJ (IG, left; TR, middle) and tumor area (right) for two biopsy regions from patients 1 (P1) and 2 (P2) with a maximum cut-off above the 95th percentile. (C) Unique clonotype count from LR Spatial VDJ libraries from breast cancer sections: P1 ($n = 3$ sampled regions, with one section for regions C and E, and two replicate sections for region D) and P2 ($n = 2$ sampled regions, one section prepared per region). (D) Number of unique IG (left) and TR (right) clonotypes per spot in different Spatial GEX clusters (+ denotes tumor cluster). (E) UMI count per spot for expanded TRB clones across spots belonging to the tumor, tumor border, and other areas. Statistical significance was calculated using the Friedman's test, then Dunn's

multiple comparison's test ($***P < 0.001$, $****P < 0.0001$). (F) (Left) Tumor 3, 5, and 8 clusters (TumC3, TumC5, and TumC8) and their surrounding borders visualized on P2 Region A. (Middle) Dot plots showing expression of significantly enriched IG clones per corresponding area. (Right) Spatial distribution of representative IG clonotypes. See fig. S23. (G) Pearson correlation coefficient (r) between the light versus heavy chain UMI count per spot for each IG receptor pair verified by scVDJ. Data are pooled from all tumor regions on a per patient basis. See fig. S24. (H) Spatial distribution of two IG light and heavy chain receptor pairs, verified by single-cell analysis. Scale bars, 1 mm. [(A) to (H)] Spatial VDJ was performed in two batches (two experiments). Single-cell processing was performed for each patient separately. [(D) to (F)] Data are from both P2 tumor regions. BC, breast cancer; Reg, region; Tum, tumor.

dataset) (fig. S24D and data S4). When “repair” was applied to P2, we found 25 predicted pairs with one out of two pairs present in the scVDJ dataset verified at these settings. Thus, “repair” with Spatial VDJ may potentially be used to predict de novo paired IG receptors in human tissues for further antibody studies.

Spatial VDJ uncovers B cell spatial evolution

B cells undergo affinity-dependent selection and clonal expansion in GCs. Multiple clones can simultaneously cycle through a single or several GCs, generating a dynamic clonal interplay (18). In the human tonsil, we annotated individual B cell follicles (Fig. 4A and fig. S7)

that contained GCs with compartments resembling the dark and light zone (fig. S25), which are enriched for somatic hypermutation and clonal expansion versus clonal selection, respectively (18). The highest IGH clonal family density (i.e., clone count per spot) was in the light zone (fig. S25, E and F), possibly reflecting ongoing clonal selection. IGH clonal family diversity measures, including species richness (*Chao1*), D50, largest clone fraction, the inverse Simpson index, and Shannon entropy, varied across follicles (Fig. 4B and fig. S26), indicating diverse interfollicular B cell clonal responses. The total IGH clonal family number per follicle positively correlated with follicle size

(fig. S26C). Most clonal families were present in a single follicle, but some appeared in multiple follicles, with 150 clonal families coincident in four or more follicles (Fig. 4C). Expanded follicular IGH clonal families both appeared primarily in one follicle (“restricted”) or spread more evenly across multiple follicles (“nonrestricted”) (Fig. 4, D and E). Thus, IGH clonal families may expand and presumably undergo affinity hypermutation in multiple GCs simultaneously. This is in line with laser microdissection studies in humans (3, 4) and photoactivation studies in mice (19). Nonrestricted IGH clonal families may be more mature as they had an increased fraction of downstream isotypes and a higher fraction

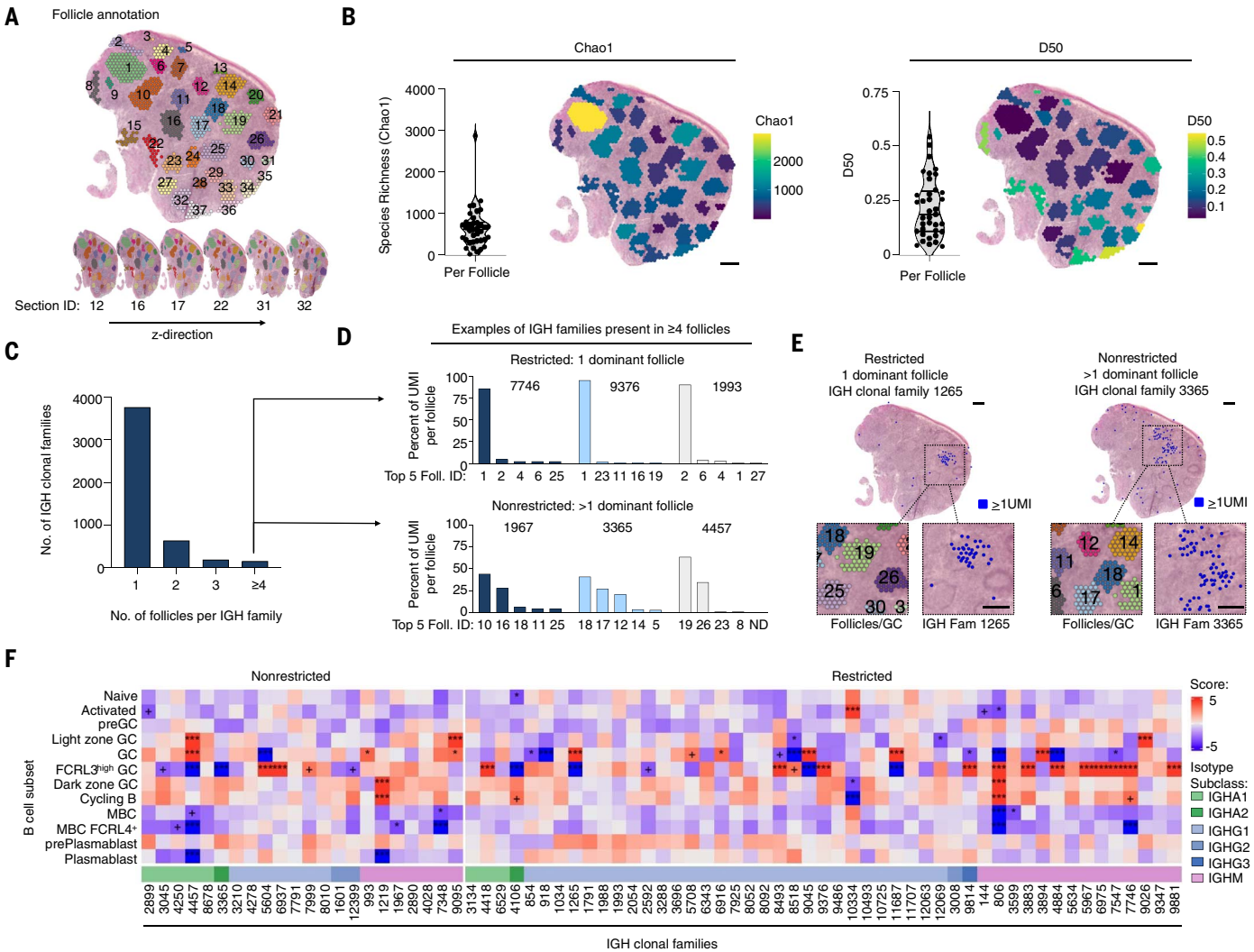


Fig. 4. Spatial VDJ delineates IG clonal diversity in B cell follicles. (A) B cell follicle annotation across human tonsil replicate sections. 36 of 37 follicles contained GC-like areas. (B) The IGH clonal family species richness Chao1 (left) and D50 (right) for each follicle and visualized on tonsil tissue. For one follicle, the D50 value was higher than 0.5 because 7 of 13 clones with UMI of 1 were needed to reach 50% of total UMIs. (C) Individual follicular IGH clonal family distribution across follicles. (D) The percentage UMI distribution across the top five follicles for six representative IGH clonal families present in ≥ 4 follicles. ND, not detected. (E) Spatial distribution for

a clonal family with (left) one or (right) multiple dominant follicles. (F) Association analysis between IGH clonal families and the B cell single-cell deconvolution data. P values were corrected with the Benjamini-Hochberg procedure and denoted by + ($P_{\text{adj}} < 0.1$; $*P_{\text{adj}} < 0.05$; $**P_{\text{adj}} < 0.01$; $***P_{\text{adj}} < 0.001$, with statistical significance considered at $*P_{\text{adj}} < 0.05$.) Scale bars, 500 μm . [(A) to (F)] Spatial VDJ was performed in two batches (two experiments) and data from six replicate sections of the same tonsil sample are pooled. [(B) and (E)] Pooled data are visualized on one representative section (tonsil section number 17).

of multiple isotypes within a single clonal family (fig. S27). Both restricted and nonrestricted clonal families were broadly associated with GC-related cell subset phenotypes (Fig. 4F). Restricted families were more often significantly associated with an Fc receptor-like 3 (FCRL3^{high}) GC cell phenotype than nonrestricted families (12 of 49 and 24.4% versus 2 of 22 and 9.1%, respectively), possibly reflecting different transcriptional states or colocalization with such cells. These multifollicular clonal families may derive from recently activated B cells that migrated from one follicle to another or from reactivated clonally related memory B cell clones that entered separate follicles. Thus, high B cell clonal diversity exists across GCs

with both restricted and multifollicular clonal expansion.

Antibodies further diversify their functions by class switch recombination (CSR), which is an intrachromosomal deletion recombination event that replaces the heavy chain constant region to confer diverse effector functions. In contrast to earlier work, recent studies have shown that B cells mainly undergo CSR prior to GC entry (18). However, where CSR occurs in human lymphoid tissues remains understudied. A recent putative CSR (pCSR) event may appear in our dataset as two (or more) colocalized transcripts from the same IGH clonal family with different isotypes but identical variable (V) gene sequences. We detected 810 IGH clonal

families containing multiple isotypes. Of these, 137 had identical V gene sequence alignments matched with multiple isotypes, and 17 occurred in the same spatial barcode (Fig. 5A and fig. S28). These 17 pCSR events all occurred extrafollicularly and were mainly switches from IGHG1 to IGHG2 (Fig. 5B); some of these pCSR events (e.g., clonal family I2298) occurred in an array spot with concurrent expression of genes encoding critical enzymes that mediate CSR (Fig. 5C). In sum, Spatial VDJ detected pCSR in human tissue that occurred extrafollicularly.

To better understand how IGH clonal families undergoing somatic hypermutation spatially segregate, we generated phylogenetic trees for a subset of expanded IGH clonal families. First, we

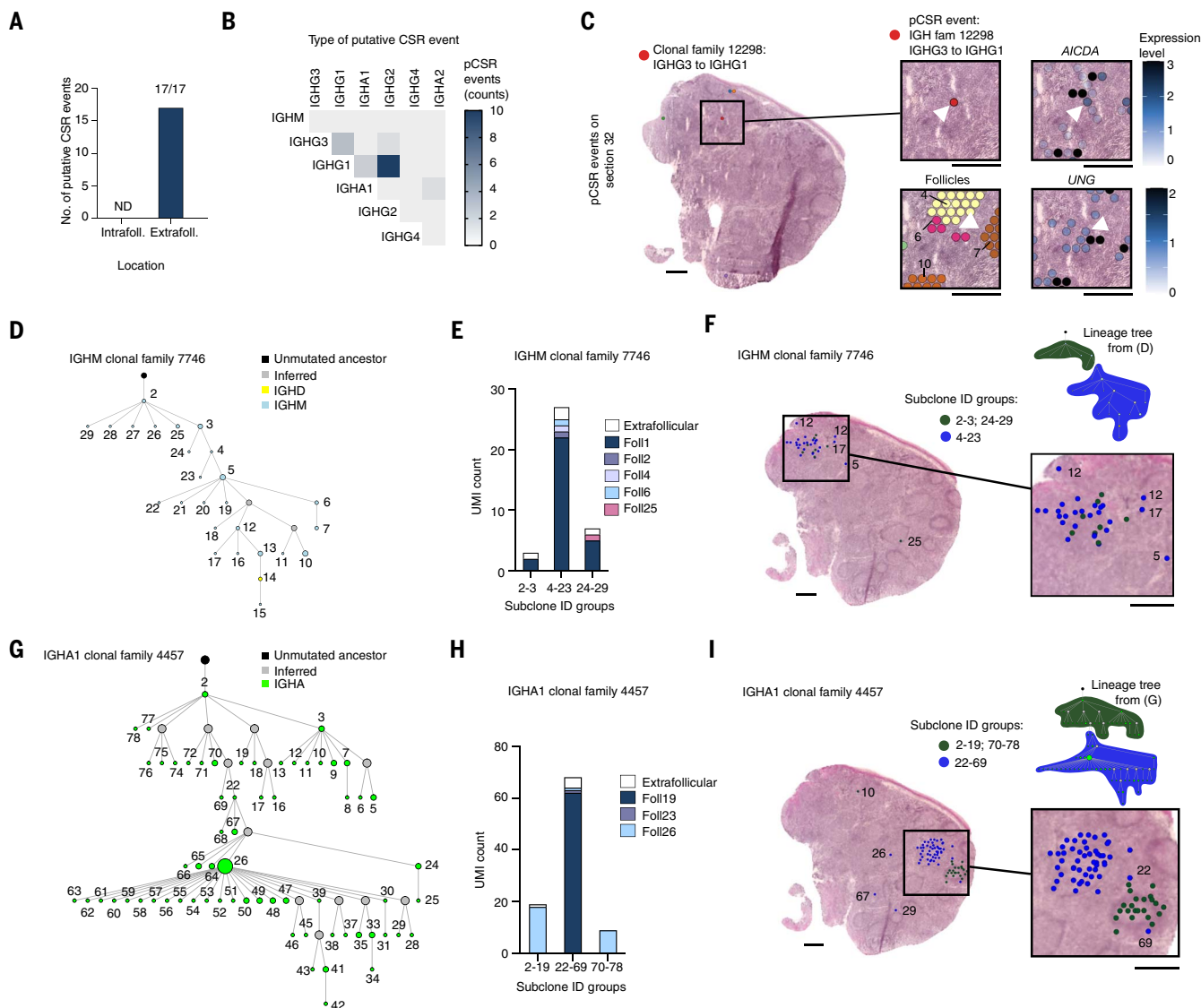


Fig. 5. Spatial VDJ uncovers IG spatial clonal evolution in human lymphoid tissue. (A) The number of pCSRs ($n = 17$) occurring within (+) or outside (–) follicle areas. (B) Types of pCSR events. (C) An individual pCSR event is visualized on section 32 (top left zoom-in), and its position is indicated relative to the follicles (bottom left) and the expression of CSR-associated enzymes *AICDA* (top right) and *UNG* (bottom right) with a maximum cut-off above the 95th percentile. The white arrows indicate the pCSR event in IGH clonal family 12298. (D) Lineage tree of IGHM clonal family 7746. Each node of the tree (“subclone”) represents a unique V sequence, and the size of the node is proportional to the number of distinct

reads in that node. All subclones are of the IGHM isotype (blue). Inferred clones (not present in the data) are gray and lack a subclone ID. (E) Intra- and extrafollicular UMI distribution for IGHM clonal family 7746 subclones. (F) Distribution of IGHM clonal family 7746 on tonsil tissue. (G) Lineage tree of IGHA1 clonal family 4457, as in (D). (H) Intra- and extrafollicular UMI distribution for IGHA1 clonal family 4457 subclones. (I) Spatial distribution of IGHA1 clonal family 4457. [(A) to (I)] Spatial VDJ was performed in two batches (two experiments). [(A), (B), and (D) to (I)] Data from six replicate sections are pooled. (C) Data are from section 32. [(F) and (I)] Pooled data are overlaid on section number 17 H&E image. Scale bars, 500 μm .

mapped mutational events occurring primarily in one follicle [e.g., clonal family 7746 (Fig. 5, D to F) and the most prominent clonal family, 806 (fig. S28, D to F)]. Given their considerable clonal size, exclusive and concurrent *IGHM* and *IGHD* expression, minor spreading of subclones across follicles, and almost exclusive follicular location, these clones were likely recently activated naïve B cells undergoing somatic hypermutation. By contrast, clonal family 4457 was an expanded, seemingly actively mutating family spread primarily across two follicles (Fig. 5, G to I). A branch of subclones (IDs 22 to

69) was almost exclusively found in follicle 19, whereas the remaining subclones (IDs 2 to 19 and 70 to 78) were only found in follicle 26. These findings suggest that a subclone 22 (or a closely related, undetected subclone), likely a GC B cell, migrated from follicle 19 to seed follicle 26. Thus, Spatial VDJ enables high-resolution demarcation of clonal composition and evolution across human GCs.

Discussion

Spatial VDJ maps full-length B and T cell receptor sequences in human tissue. The

performance of Spatial VDJ depends on several factors, including the quality of the starting material, lymphocyte abundance, tissue compatibility with the Visium Spatial GEX fresh frozen protocol, antigen receptor expression level, clonal cell count, and clonal tissue distribution. For example, Spatial VDJ optimally captures expanded clones and plasma cell clones owing to high cell count per clone and high receptor expression per cell, respectively. By contrast, Spatial VDJ’s ability to capture non-expanded clones or clones for which the antigen receptor is expressed at low levels is more limited owing

to transcript dropout and partial coverage of the tissue by array spots. Nevertheless, our TCR results were comparable to other spatial retrieval methods (1, 2). Spatial VDJ cannot (yet) fully resolve a clone's cell state(s) or receptor chain pairing as the capture array lacks single-cell resolution. Computational modeling developed in this study permitted receptor pairing for IG but not TR chains because of data sparseness for the latter. Improved analytical tools, increased resolution of the capture array, inclusion of $\gamma\delta$ T cell receptor gene probes, and protein co-capture could further extend Spatial VDJ's performance.

Several key elements distinguish Spatial VDJ from other spatial antigen receptor methods (1, 2). First, LR Spatial VDJ simultaneously maps TR and IG clones. Second, LR Spatial VDJ captures full receptor sequences, which for B cells permits spatial lineage tracing (including somatic hypermutation and class switching). Deep sequencing of Visium libraries permits some IG light chain clonal recovery, but not full transcript recovery, owing to fragmentation steps (5). Laser-guided microdissection coupled with BCR genomic repertoire sequencing (4) avoids IG expression bias but is more arduous, lower throughput, and at lower resolution. Third, Spatial VDJ is an extension of Visium for fresh frozen tissue, which is commercially available, frequently used, requires no specialized equipment, and allows data projection onto matched histologically stained (or fluorescent) images. Thus, Spatial VDJ should be straightforward to implement. Compared with Slide-seq, Visium has a larger capture area but lower spatial resolution (2). Lastly, Spatial VDJ offers a long-read and short-read version for spatial TCR analysis. Both methods reproducibly map the most abundant clones, but LR Spatial VDJ captured slightly higher clonal numbers and required fewer PCR cycles, potentially reducing PCR artefacts. However, SR Spatial VDJ may be more accessible, as short-read sequencing services are more broadly available and higher throughput.

We predict that Spatial VDJ will be useful across research fields to study urgent clinically relevant phenomena, particularly to discriminate B cell clonal dynamics during infections and vaccination. For example, robust and persistent GC responses have been linked to productive memory B cells and antibodies after severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccination, whereas a failure to generate such responses may relate to severe SARS-CoV-2 infections (20). The use of Spatial VDJ to resolve spatial dynamics of GC B cell clonal evolution across many individuals may help clarify what triggers protective versus insufficient B cell responses [e.g., by linking IG and TR sequence distributions to antigen receptor reactivities and affinities, IG and TR germline variations (21), protein expression, and clinical outcomes].

Likewise, extending Spatial VDJ to model organisms and/or settings in which the antigen is known and can be tracked may be highly useful to study such questions. These analyses are highly applicable to cancer (2, 5, 22) given the critical contribution of lymphocytes [and tertiary lymphoid structures (15, 23)] to antitumor immunity and the simultaneous challenge to identify functionally relevant B and T cell clones (14, 24). Spatial VDJ may offer a distinct vantage point to enrich for tumor-regulatory clones. Ultimately, we envision that Spatial VDJ may serve to link B and T cell clones to their spatial niches to identify candidate antigen-specific clones for engineered cell- or antibody-based therapeutics.

Materials and Methods

Tonsil samples

The tonsil samples were obtained from anonymous adult patients undergoing tonsillectomy due to obstructive sleep apnea syndrome. Permission to collect these unidentified tissues was obtained from the Ethical Review Authority in Stockholm, Sweden (2006/646-31/4, amendments 2015/1083-2, and 2017/1659-32). All patients gave informed consent prior to collecting tissue. From each tonsil a smaller piece ($\approx 0.5 \text{ cm}^3$) was cut out, immediately embedded in OCT, placed on dry ice, and stored at -80°C .

Breast tumor samples

Breast cancer samples were obtained by Dr. J. Hartman from the Department of Clinical Pathology and Cancer Diagnostics at Karolinska University Hospital, Stockholm, Sweden. Experimental procedures and protocols of the study were previously approved by the regional ethics review board (Etikprövningsnämnden) in Stockholm (2016/957-31, amendment 2017/742-32, 2021-00795, and 2022-05245-02). Breast tumor samples were collected from two patients with untreated invasive ductal carcinomas during surgery (fig. S13A). Histological evaluations of patient tumors were performed by pathologists for diagnostic purposes: tumor characteristics, including hormone receptor, HER2, and Ki67 expression (data S3). Both tumors were HER2 positive. For each tumor biopsy, different regions ($n = 4$ to 5) were selected by the pathologists depending on the size of the tumor (fig. S13A). From each region, tissue was isolated for immediate embedding in OCT and for gene expression analysis with spatial transcriptomics and Spatial VDJ analysis. We prepared Spatial GEX and VDJ libraries for three regions for Patient 1 (P1) (figs. S13 and 14), and two regions for Patient 2 (P2) (figs. S13 and S15). The remaining material for each tumor region was used for single-cell RNA and VDJ sequencing analysis. Samples for spatial transcriptomics were immediately frozen and stored at -80°C until further analysis. The tissues

obtained for single-cell analysis were processed directly.

Preparation and sequencing of Spatial GEX

Sections of fresh-frozen breast tumor and tonsil tissue were cut at $10\text{-}\mu\text{m}$ thickness and mounted onto slides from the Visium Spatial Gene Expression Slide & Reagent kit (10x Genomics). For Tonsil #2, six sections were analyzed, covering $200 \mu\text{m}$ in the z -axis (orthogonal to the tissue plane) of the sample (fig. S3A). Sequencing libraries were prepared following the manufacturer's protocol (Document number CG000239 Rev A, 10x Genomics). Prior to imaging, coverslips were mounted on the slides according to the protocol's optional step "Coverslip Application & Removal." Tissue images were taken at 20X magnification using the Metafer Slide Scanning platform (Microscope stand: AxioImager.Z2 with ScopeLED Illumination, Zeiss; Camera: CoolCube 4m, MetaSystems; Objective: Plan-Apochromat 20X/0.80 M27, $a = 0.55 \text{ mm}$, Zeiss, Software: Metafer5 version 3.14.192). Raw images were stitched with VSlide software (version 1.1.128; MetaSystems). Adaptations of the protocol were made in that the Hematoxylin staining time was reduced to 4 min and tissue permeabilization was performed for 12 min. Final libraries were sequenced on NextSeq2000 (Illumina) or NovaSeq6000 (Illumina).

Target enrichment of spatial cDNA with hybridization capture (LR and SR Spatial VDJ)

For target enrichment we used IDT xGen Hybridization and Wash Kit (#1805584) with one Discovery pool (IDT) each for BCR and TCR transcripts (IG and TCR pool) (data S1). The BCR consists of two heavy chains (encoded by the *IGH* gene loci) and two light chains (*IGK* or *IGL*) (25). The TCR is either composed of an α (*TRA*) and β chain (*TRB*) or a γ (*TRG*) and δ (*TRD*) chain (26), with T cells expressing TCRs encoded by *TRA* and *TRB* ($\alpha\beta$ T cells) being more abundant. We focused our spatial analysis in this study on $\alpha\beta$ T cells; (i.e., probes targeting *TRD* and *TRG* were not included). Custom blocking oligos (IDT) (data S1) were designed to hybridize to universal adaptor sequences of the cDNA library, to prevent off-target fragments from binding to BCR or TCR transcripts and contaminating the enriched library. To ensure sufficient product for the hybridization capture protocol, we cocaptured the TR and IG transcripts in the same reaction. Regardless of whether the user selects LR (both IG and TR analysis) versus SR Spatial VDJ (TR only), we would recommend always including both IG and TR capture probes for the hybridization capture step, since the low TR transcript count in the cDNA libraries is likely a limiting step to a successful target enrichment reaction. We did not pursue a short-read option for IG receptor sequences since we wanted to capture the full-length transcripts to map somatic hypermutation

and class switching events. Additionally, owing to its very long constant region, a similar semi-nested approach for IG receptor sequences would still likely result in an inability for the IG heavy chain transcripts to cluster on the sequencer instrument.

The input cDNA libraries were amplified prior to target enrichment to ensure sufficient material. We took 10 µl of Visium cDNA per library (since this is typically considered representative of the entire spatial GEX library) and split into subsamples of 10 ng per PCR reaction (number of subsamples varied depending on the cDNA concentration in each library). PCR reactions were performed using 25 µl of KAPA HiFi HotStart ReadyMix (2X) (KK2602, Roche), 7.5 µl of cDNA primers (10x Genomics), and 17.5 µl of sample in Milli-Q water. The following settings were used for the PCRs:

1. 98°C: 3 min
2. 98°C: 15 s
3. 63°C: 30 s
4. 72°C: 2 min
5. Repeat steps 2 to 5 six times for a total of five cycles
6. 72°C: 1 min

The subsamples were pooled again before bead wash with 0.6X SPRIselect (Beckman-Coulter) and eluted in 40 µl Elution buffer (#19086, Qiagen) each.

We followed the protocol “xGen hybridization capture of DNA libraries,” version 4 (Integrated DNA Technologies), with the optional Appendix A “AMPure XP Bead DNA concentration protocol” followed by “Plate protocol” from step 13 (in tube strips), with the hybridization reaction carried out overnight (16 hours). The IG and TCR discovery pools were mixed at a final ratio of 1:3 or 1:12 (IG:TR) in the hybridization reaction which yielded comparable results. We opted to use 1:3 for subsequent experiments. The 1:3 ratio was achieved by triplicating the TR probe sequences versus the IG in the IDT order and adding equivalent volumes of the produced pools in the hybridization reaction (i.e., two separate pools were ordered, with each TR bait probe sequence listed three times and each IG listed once). For one (out of two) Spatial VDJ experiment, the enriched and purified libraries were amplified twice with an AMPure bead wash after each PCR, using 25 µl of KAPA HiFi HotStart ReadyMix (2X) (KK2602, Roche), 7.5 µl of cDNA primers (10x Genomics), and 17.5 µl of the sample in Milli-Q water. The following settings were used for the PCRs:

7. 98°C: 3 min
8. 98°C: 15 s
9. 63°C: 30 s
10. 72°C: 2 min
11. Repeat steps 2 to 5 six times for a total of seven cycles (PCR 1) and four times for a total of five cycles (PCR 2)
12. 72°C: 1 min

For the second Spatial VDJ experiment, the enriched libraries were amplified in one PCR reaction as described above, but with 14 cycles.

Long-read library preparation and sequencing (LR Spatial VDJ for IG and TR)

The product from the hybridization capture was used as input into the SMRTbell library preparation (PacBio). We concentrated the DNA by AMPure Bead Purification (0.8X), eluting in 6 µl of EB buffer, using 1 µl for Qubit measurements. We used at least 1 µg of input for each library and multiplexed eight samples in one sequencing run. PacBio Barcoded Overhand Adapters were used for multiplexing and followed the manufacturer's instructions for the library preparations. This protocol indexed the enriched libraries by ligation to avoid unnecessary PCR cycles, which can introduce errors and chimeric reads that are particularly problematic for long-read sequencing. The pooled library had more than the minimum amount required for sequencing. A SMRT Enzyme clean-up kit was used to remove linear and single-stranded DNA. The final libraries were sequenced on a Sequel II at the National Genomics Infrastructure (NGI)/Uppsala Genome Center.

Short-read library preparation and sequencing (SR Spatial VDJ for TR only)

After hybridization capture and post-capture PCR amplification (14 cycles), further *TRA* and *TRB* enrichment was performed by semi-nested PCR reactions with the following primers: V primers targeting either the *TRAV* or *TRBV* genes, 5' of the CDR3 region (i.e., “Outer” *TRAV* or *TRBV* primers) and a primer (“partRead1”) targeting the universal partial read 1 sequence present on all poly-dT captured transcripts in Visium Gene Expression cDNA libraries (see data S1 for primer sequences). PartRead1 is also compatible with TruSeq indexes to allow sample multiplexing for sequencing. The Outer V primer PCR input was 2.5–5 ng of hybridization captured cDNA from tonsil Visium Gene Expression libraries and the reaction was run with KAPA HiFi HotStart ReadyMix (2X) (KK2602, Roche). Four replicate reactions were prepared per sample and pooled at the final indexing step. All primers were diluted 40X for a final concentration of 2.5 µM from 100 µM stocks (Integrated DNA Technologies). The PCR was run for 12 cycles for *TRB* and 14 cycles for *TRA*, under the following conditions:

1. 98°C: 5 min
2. 98°C: 20 s
3. 65°C: 30 s
4. 72°C: 1 min 30 s
5. Repeat steps 2–5 11 or 13 times for a total of 12 or 14 cycles
6. 72°C: 7 min

Quantitative real-time PCR was performed to determine the appropriate number of cycles (to limit exponential amplification). The Outer

V primer PCR product was purified using a magnetic bead clean-up with AMPure beads (0.6X), followed by two 80% EtOH washes. The Outer V primer PCR product was eluted in EB buffer after incubation at 15 min at 37°C. The cleaned PCR product was quantified using Qubit and Bioanalyzer (Agilent). Five ng of each PCR product was used as input to the subsequent Inner V primer PCR.

The Inner V primer PCR was performed with the following primers: V primers targeting either the *TRAV* or *TRBV* genes, close/adjacent to the CDR3 region (i.e., “Inner” V primers) and the same universal partRead1 primer as described for the Outer V primer PCR (see data S1 for primer sequences). These Inner V primers have a handle compatible with TruSeq indexing (data S1). The primer concentrations and reagents were as described for the Outer V primer PCR. Quantitative real-time PCR was performed again to determine the appropriate number of cycles (to avoid exponential amplification). The following conditions were used for the PCR reaction:

1. 98°C: 5 min
2. 98°C: 20 s
3. 72°C: 30 s
4. 72°C: 1 min 30 s
5. Repeat steps 2–5 seven times for a total of eight cycles
6. 72°C: 7 min

The PCR product was cleaned using AMPure magnetic bead-clean up and ethanol washes as described above. The final eluted PCR products were quantified using Qubit and Bioanalyzer (Agilent). For the indexing step, 10 ng of input material was used from each replicate. All replicates for each individual sample were pooled at this stage and received the same index. The samples were PCR indexed using TruSeq Indexes, according to the conditions listed below and as described previously (6):

1. 98°C: 3 min
2. 98°C: 20 s
3. 60°C: 30 s
4. 72°C: 1 min 30 s
5. Repeat steps 2–5 seven times for a total of eight cycles
6. 72°C: 5 min

The PCR product was cleaned using AMPure magnetic bead-clean up and ethanol washes as described above. The final eluted PCR products were quantified using Qubit and Bioanalyzer (Agilent) and pooled for sequencing. The samples were sequenced on a Novaseq6000 SP flowcell using a short read 1 and a longer read 2 to capture the entire CDR3 region and part of the constant region from the 5' end (read 1, 35 nt; index read, 6 nt; read 2, 259 nt).

RNAscope

Fresh frozen tissue sections were fixed for 1 hour in pre-chilled 4% PFA. The procedure was then performed according to the RNAscope® Multiplex

Fluorescent Reagent Kit v2 Assay User Manual (Document Number 323100-USM). For tonsil section #25 (fig. S3A), probes targeting *SDC1* (Cat#: 416961-C2; lot#: 22076C) and *MS4A1* (Cat#: 426771-C3; lot#: 22076C) were used. For tonsil section #26 (fig. S3A), probes targeting *CD3E* (Cat#: 553971-C2; lot#: 22076D) and *MS4A1* were used. Control tonsil sections were incubated with either the RNAscope® 3-plex Positive Control Probe_Hs (320861; lot# 21258A) or the RNAscope® 3-plex Negative Control Probe_Hs (320871; lot# 21285A).

The TSA Vivid Fluorophore kit 570 (7526/1 KIT; batch 1A) was assigned to channel 2 with a dilution of 1:1500 for tonsil section #25 and a dilution of 1:750 for tonsil section #26. The TSA Vivid Fluorophore kit 650 (7527/1 KIT; batch 1A) was assigned to channel 3 with a dilution of 1:750 for both tonsil sections. For the control tonsil sections the TSA Vivid Fluorophores were assigned to the channels according to the RNAscope® Multiplex Fluorescent Reagent Kit v2 Assay User Manual. The following reagents were also used:

RNAscope® multiplex fluorescent detection reagents v2 (Cat#: 323110; lot# 2014906)

RNAscope® hydrogen peroxide (Cat#322335; lot#2015184)

RNAscope® protease IV (Cat#322336; lot#: 2015288)

Images were acquired with Zeiss LSM700 confocal microscope and the Zen2012 software with 10X (Plan-Apochromat 10x/0.45) objective. The Diode555, Diode405-5 Diode488 and Diode639 excitation lasers were directed through a photomultiplier tube (LSM, Carl Zeiss). Tile scan stitching was performed with Zen 2012 software and image processing was performed with Image J/Fiji software (version 1.53 for Mac; Java 1.8.0_172). Images were processed by using Fiji functions of splitting channels, contrast enhancement, merging channels, rotating images, and adding scale bar.

Bulk SS3-based VDJ library preparation (Bulk SS3 VDJ)

We modified the recently described Smart-seq3 (SS3) method (27) to accommodate the use of low concentrations (10 to 100 ng) of purified RNA as the input (as opposed to single cells) and the use of many replicates to reduce amplification biases. This method will amplify UMI-barcode IG and TR cDNA using primers directed to constant regions and a universal sequence 5' of the UMI region, similar to the 10x Genomics VDJ 5' protocol. We first lysed each selected fresh-frozen tonsil tissue sections and isolated total RNA. We then performed reverse transcription and amplification following the Smart-seq3 protocol to generate cDNA libraries for each section. After bead purification, those libraries were split to amplify IG or TR amplicons by target-specific primers (forward: universal sequence introduced by a

template-switching step; reverse: constant region-specific primers for TR or IG), followed up by bead purification. A second PCR (with constant primers binding closer to the CDR3) with subsequent bead purification was then performed for both IG and TR amplicons to shorten the 3' end of each amplicon for Illumina sequencing of the internal VDJ recombinant sequence (CDR3). Individual samples were indexed and pooled for sequencing by the National Genomics Infrastructure, SciLifeLab (Solna). Detailed descriptions for this method can be found in the section labeled "Bulk SS3-based VDJ library preparation (Bulk SS3 VDJ)" in the Supplementary Materials file.

Cell processing for single-cell RNA sequencing

Single-cell suspensions from separate breast tumor regions (P1, $n = 5$; P2, $n = 4$) were prepared by enzymatic tissue dissociation using the human Tumor Dissociation Kit (Miltenyi Biotec, 130-095-929) and gentleMACS dissociator (Miltenyi Biotec). Cell suspensions were stained with the Zombie Aqua Fixable viability dye (Biolegend, 423101) at room temperature for 20 min, then washed with phosphate buffered saline (PBS). The cells were incubated with Human TruStain Fc block (Biolegend, 422302) for 10 min to limit nonspecific antibody binding, then stained for 20 min with anti-CD45 (1:40, Biolegend, 304021), anti-EPCAM (1:40, Biolegend, 324206) and cell hashing TotalSeq-C antibodies (Biolegend; 394661, 394663, 394665, 394667, 394669) in fluorescence-activated cell sorting (FACS) buffer (PBS + 0.5% bovine serum albumin). Each tumor region was labeled with a unique TotalSeq-C antibody to enable pooling all tumor regions into one sample for single-cell gene expression and VDJ analysis. The cells were subsequently washed and resuspended in FACS buffer. FACS was performed using an influx flow cytometer (BD Biosciences) to sort live EPCAM⁺CD45⁺ single cells for 10x Genomics Chromium Single Cell gene expression analysis. At least 2.5×10^4 to 3×10^4 cells were sorted per tumor region and all tumor regions for each patient were sorted into the same tube. Single-stain controls (cells and beads) and fluorescence minus one (FMO) controls—containing all the fluorochromes in the panel except the one being measured—were used to set voltages and to define the proper gating strategy.

10x Genomics Chromium Single-Cell library preparation and sequencing

Single-cell gene expression (GEX), VDJ clonotype, and feature barcoding libraries were generated from EPCAM⁺CD45⁺ cells using the 10x Genomics Chromium Single Cell 5' assay, following the manufacturer's instructions. Libraries were profiled and quantified using a Bioanalyzer High Sensitivity DNA kit (Agilent Technologies) and Qubit High sensitivity kit (ThermoFisher Scientific). Final single-cell gene expression

libraries were sequenced (30,000 reads per cell for GEX; 4000 to 6000 reads per cell for VDJ clonotype, and 4000 to 5000 reads per cell for feature barcode libraries) on a NovaSeq 6000 SP flowcell (Illumina, 150-8-8-150 read set-up) by the National Genomics Infrastructure, SciLifeLab (Solna).

Data processing of Spatial GEX libraries

Following demultiplexing of bcl files, read 2 fastq files were trimmed using Cutadapt (28) to remove full-length or truncated template switch oligo (TSO) sequences from the 5' end (beginning of read 2) and polyA homopolymers from the 3' end (end of read 2). The TSO sequence (AAGCAGTGGTATCAACGCAGAGTACATGGG) was used as a non-internal 5' adapter with a minimum overlap of 5, meaning that partial matches (up to five base pairs) or intact TSO sequences were removed from the 5' end. The error tolerance was set to 0.1 for the TSO trimming to allow for a maximum of 3 errors. For the 3' end homopolymer trimming, a sequence of 10 As was used as a regular 3' adapter to remove potential polyA tail products regardless of its position in the read, also with a minimum overlap of five base pairs. The trimmed data were processed with the Space Ranger pipeline (10x Genomics), version 1.2.1 (tonsil) and version 1.0.0 (BC) and mapped to the GRCh38 v93 genome assembly.

Long-read sequencing analysis (LR Spatial VDJ)

The input to the analysis was demultiplexed consensus reads from the long-read Pacbio sequencing. The analysis was performed using Python programming language, see the Zenodo repository (29): scripts/functions/long-read-processing-STAR-paper-main.zip. Scripts are also available on GitHub (30). First, the fastq files were parsed into a dataframe with readID, sequence, and quality columns. Second, we searched for the Truseq adapter and the TSO sequence to anchor the ends of each read and discarded the reads that lacked these sequences. Third, we searched for the pattern "CGACGCTCTTCCGATCT" which is part of the Truseq adapter starting in the first seven bases of either the read or its reverse complement. If any of the positions matched the sequence with hamming distance of one or less, we marked it. We did the same for the TSO sequence ("TCTGCGTTGATACCACT"). We reverse complemented the reads as needed so that all the reads have the Truseq adapter at the beginning and the TSO at the end. We identified the spatial barcode and the UMI by obtaining the first 16 bases following the Truseq adapter for the spatial barcode and the 12 bases following that as the UMI. We made sure that the following four bases were all T's and filtered out the reads that had any other bases in that interval. The most common UMI was a polydT, which was considered an artifact. Those reads were consequently

removed. We also defined the end of polyT region as the first matching position for the pattern “[^T]{0,2}[^T]{0,2}[^T]” after 28 bases after the end of the Truseq adapter location.

LR Spatial VDJ clonotype analysis

To run MiXCR (version 3.0.3), we trimmed the polyT and TSO and wrote the reads to a new fastq file. We used MiXCR (37) with this command:

```
mixcr analyze shotgun -s hsa-align -OsaveOriginalReads=true-starting-material rna <TrimmedFastq> <SampleName>
```

We then ran another MiXCR command to report alignments for each read:

```
mixcr exportAlignments -f -cloneIdWithMappingType -cloneId -readIds -descsR1 <SampleName>.cna <ReportFile>
```

We then used the resulting tabular file to assign the reads to the clonotypes in MiXCR output. To use UMIs, we filtered out reads that did not map to any clone (`cloneId == -1`), then grouped the reads table by the Visium barcode and UMI and counted how many reads they have and how many clones were associated with each UMI. We filtered out UMIs that had been assigned to more than one clonotype, as they were likely due to PCR or sequencing errors. A count matrix (clone ID × spatial barcode) was generated and merged to the clone list exported by MiXCR. After removing spots outside the tissue section (as is standard for the Space Ranger pipeline), we imported the gene expression cluster and follicle annotations in R (version 4.2.0), then calculated UMI, spatial barcode, and overlaps among different clusters, follicles, and sections. Script for clone analysis can be found in the Zenodo repository (29): tonsil: scripts/tonsil/1_LR-SpatialVDJ/tonsil LR clonotype analysis.Rmd; breast cancer: scripts/breast_cancer/1_LR-SpatialVDJ/BC LR clonotype analysis.Rmd.

SR Spatial VDJ clonotype analysis

To preprocess short-read sequences, we modified the example workflow of “UMI Barcoded Illumina MiSeq 325+275 paired-end 5’ RACE BCR mRNA” from pRESTO (version 0.6.2, Immcantation group) (32). Reads with less than 20 Phred quality scores were regarded as low-quality sequences and were removed. Then, we removed reads that lacked a valid spatial barcode by comparing the reads to the Visium barcode list. We only kept the reads that included the “inner V” primer sequences used in the second step of the *TRA* and *TRB* seminested PCR enrichment steps. Input fastq files that included more than 13 million were downsampled to 13 million reads by *SplitSeq.py* samplepair, owing to the platform limitation (downsample rate: 0.76–0.94). To have a more accurate UMI collapsing result, we regarded both the UMI and spatial barcode sequence as the “UMI” region in the pipeline. We selected UMIs with at least three reads per group in tonsil for repertoire analysis.

We merged all reads from different tonsil sections and used MiXCR (version 3.0.13) with the following commands:

```
mixcr analyze amplicon-library repseqio.v1.5-species hsa-starting-material rna-5-end no-v-primers-3-end c-primers-adapters adapters-present-align “-OsaveOriginalReads=true”-assemble “-write-alignments” <R1Fastq> <R2Fastq> <SampleName>
```

To report the alignments for each read, we further ran the MiXCR command:

```
mixcr exportAlignments -cloneId -descsR2 -targetSequences -nFeature CDR3 -jGene <SampleName.clns> <SampleName.tsv>
```

We then used customized R scripts to create the clonal UMI count matrix for each tonsil section (R version 4.2.0). The exported UMI collapsed reads that lacked a valid CDR3 call from MiXCR output, did not map to any clone (`cloneId == -1`), or had fewer read counts than three reads per UMI were removed. We observed some cases of index hopping (i.e., where individual UMI collapsed reads from different samples contained identical nucleotide CDR3 sequence, J gene, spatial barcode, and UMI sequence). The clone that contained the most read count in its UMI group was considered the true UMI collapsed reads and the rest discarded. At these settings, we still observed a small number (1 to 4 per section) of TR clones with an unexpectedly high spatial barcode number per clone (that was not reproduced in the LR-Spatial VDJ data), suggestive of PCR overamplification and artifacts in spatial barcode region. We therefore removed those clones from further analysis. Section overlap was calculated the same way as we did for LR reads. Scripts can be found in the Zenodo repository (29): scripts/tonsil/2_SR-SpatialVDJ/tonsil SR-Spatial VDJ analysis.Rmd and scripts/tonsil/2_SR-SpatialVDJ/tonsil seminested TR on-off analysis.Rmd.

Visualization and analysis of spatial data

Gene expression data of tonsil #1

Two sets of tissue images, spatial coordinates, and count matrices generated through the Space Ranger (10x Genomics) pipeline were imported into R (version 4.0.5, R Core Team, 2017) using the *InputFromTable()* function of the STUtility package (33). For fig. S1C, when applicable, features were collapsed based on antigen receptor chain identity and visualized using the *FeatureOverlay()* function. Tonsil #1 Space Ranger output data can be found in the Zenodo repository (29): data/tonsil/5_spatial_GEX/spaceranger_output/. R script for reproducing figures in fig. S1: scripts/tonsil/5_spatial_GEX/tonsil1_GEX_v2.Rmd.

Spatial GEX and LR and SR Spatial VDJ tonsil and tumor data

Space Ranger output data were imported into R (version 4.0.5) as described in previous para-

graphs. One Seurat object (34) per sample type or patient (tonsil, P1, and P2) was created using the STUtility package. For each dataset, genes were filtered for a minimum of 100 counts across the entire dataset and a minimum of five spots. Spots were also filtered based on the number of counts. Spots with at least 300 counts (P1) or 500 (tonsil and P2) were kept. To add clonotype data to the Seurat objects described above, the clonotype count matrices were modified to contain the same barcodes (bcs) as the gene expression count matrices. Accordingly, bcs absent in the gene expression count matrix (after filtering described above) were removed and only bcs present in the gene expression count matrix were added to the clonotype count matrix and filled with zeros for each clone. Finally, clonotype names were edited to include the first four letters from the “C hit” column of the MiXCR output, followed by “clone” and the number assigned by MiXCR. Each modified clonotype count matrix was then loaded as a new assay into its respective Seurat object (tonsil, P1, and P2), whereafter genes and clonotypes were visualized on the tissue images using built-in functions of the STUtility package. Scripts for generating all figure panels are listed as in the Zenodo repository (29): paths in data S5, and Seurat objects with descriptions can be found in:

Tonsil #2: data/tonsil/0_integrated/tonsil_se_PartIV.rds

Breast Cancer Patient 1: data/breast_cancer/0_integrated/P2/BC_P1_PartIII.rds

Breast Cancer Patient 2: data/breast_cancer/0_integrated/P2/BC_P2_PartIII.rds

Normalization and clustering of GEX data

The gene expression data of tonsil, P1, and P2, respectively, were normalized across spots with Seurat *SCTransform()* function in R (version 4.0.5). Dimensionality reduction was performed with nonnegative matrix factorization by using STUtility *RunNMF()* function with 24 factors. These factors were subsequently used for clustering using Seurat *FindNeighbors()* function followed by Seurat *FindClusters()* at resolution 0.8.

For Tonsil #2, this strategy rendered 13 clusters, which were annotated and grouped based on shared gene expression and single cell deconvolution results from *stereoscope* (method described below). The cluster groups were consistent with known anatomical structures observed in H&E stains of each section. For follicle annotation on tonsil sections, each follicle or GC had to contain at least two spots belonging to at least one of the four clusters annotated as B cell follicle or GC cluster group (c1, 4, 5, or 7) and be found in at least two independent sections.

For P1 and P2 breast tumor samples, the strategy described above generated 18 and 12 clusters, respectively. Tumor areas were annotated based on gene expression cluster and single-cell deconvolution results per Spatial GEX cluster

from *stereoscope* (method described below) (35). All tumor clusters had enriched tumor-associated gene expression and cancer epithelial cell signatures. Cluster 13 (P1) exhibited a heterogeneous epithelial signal, with cancer and normal epithelial cell enrichment varying across sample regions. Due to the presence of cancer epithelial signature in at least one sampled region, it was denoted as “Tumor,” but likely contained normal breast epithelial tissue as well. For P2, we identified four distinct tumor clusters, three of which were detected primarily within Region A (Tumor C3, C5, and C8). Tumor C3, C5, and C8 were all enriched for cancer cells, but exhibited differently enriched genes. Tumor C3 spots had elevated expression of genes involved in hormone-receptor signaling, including *TFF1* and *TFF3* (36), and expressed by breast glandular cells (e.g., *SCGB2A2*) (37–39). Conversely, Tumor C5 and C8 shared many enriched genes, including tumor-associated *S100A7-9*, *CD44*, and *PABPC1*, which are involved in cell migration, proliferation, and cell-to-cell interactions (40–43), as well as genes expressed by macrophages [e.g., *LYZ* and *APOE* (44)], including those infiltrating tumors. Based on the breast cancer spatial cluster annotation and border detection with STUtility RegionNeighbors() function, we adapted Seurat FindMarkers() function for calculating differentially expressed LR Spatial VDJ IG clones across regions. Specifically, we used the Poisson test, with a minimum of 0.2-fold difference and 0.01 as the minimum percentage of spots in an area expressing a clonotype. The latter parameter is drastically reduced from the default of 0.1, which is optimized for scRNA-seq analysis. We reasoned that spatial VDJ data should require a lower value to reveal clonotypes that are region-specific but only expressed in a minority of spots in that region. This workflow is provided as an R markdown file in the Zenodo repository (29): scripts/breast_cancer/0_integrated/IG_enrichment_BC_v1.Rmd. For script paths to all figure panels, see data S5.

We applied a previously developed script to analyze enrichment or depletion of the single-cell deconvolution analysis across Spatial GEX clusters (version used in this paper is in the Zenodo repository (29): scripts/functions/enriched-region-add-legend.py. Scripts are also available in the GitHub repository (45). In general, we calculated the average proportion of each cell type within the region as “true average”. Then, we randomly shuffled the spatial barcodes of deconvolution results to assign a different region and calculated the “permuted average” of them. This shuffling and calculation were conducted 10000 times, and the mean value of difference divided by standard deviation was visualized as enrichment/depletion score. Script for generating input datasheet can be found in the Zenodo repository (29): scripts/tonsil/7_enrichment/tonsil enrichment analysis.Rmd;

scripts/breast_cancer/8_enrichment/BC enrichment analysis.Rmd; scripts/breast_cancer/8_enrichment/bc enrichment depletion per clone.Rmd.

Single-cell deconvolution (*stereoscope*) analysis

We used the software *stereoscope* (v.0.3) to decompose the Spatial GEX profiles into contributions from specific cell types, also known as single-cell mapping (35). *Stereoscope* uses a probabilistic framework that models both single-cell and spatial transcriptomics data with a negative binomial distribution. In short, *stereoscope* first learns cell type and gene specific parameters from the single-cell data (where no mixing of cells occurs), to then use these parameters in a guided decomposition of the mixed gene expression profiles in the spatial transcriptomics data.

The input to *stereoscope* is the raw (unnormalized) single-cell UMI count data, cell type annotations, and raw (unnormalized) spatial transcriptomics UMI count data. The output from *stereoscope* is a [spot] × [cell type] matrix, where each element gives the proportion of cells at a given spot that belongs to a specific cell type. For more details we refer to the original *stereoscope* publication. In the supplementary materials, we outline the two analyses in more detail.

Single-cell gene expression and VDJ data processing

Sequencing outputs were processed by Cell Ranger (version 5.0, 10x Genomics) (46). Gene-barcode count matrices were analyzed with the Seurat package (version 4.0, Satija Lab) in R (version 4.0.1). Two steps of filtering were introduced here. First, raw gene expression matrices were subset by the barcode list in VDJ output, including B and T cell subsets. Based on the UMI count, gene count, and mitochondrial percentage of raw gene expression matrices and their subsets, each threshold was selected to keep the maximum count of high-quality cells and avoid losing B and T cells which have VDJ sequencing outputs. Second, feature barcodes are used for doublet detection and removal by the “Demultiplexing with hashtag oligos (HTOs)” workflow from Seurat. Specifically, doublets in each sample were detected by HTODemux() function and removed for the downstream analysis. Cutoffs for UMI and gene count, mitochondrial percentage and HTO threshold (positive.quantile) can be found in data folder deposited in the Zenodo repository (29): data/breast_cancer/3_single-cell_GEX/preprocessing_output. All samples were integrated and scaled into one count matrix by Seurat. Dimension reduction, UMAP generation, and clustering, were performed on the merged dataset by Seurat. The merged dataset was clustered by a gradient of the resolution, from 0.2 to 2. We chose 0.8 as the final resolution by comparing the top-listed differentially

expressed genes in each cluster. Cell types were annotated by differentially expressed genes and their marker genes expression level. Our scRNAseq dataset also lacked neutrophils and eosinophils, which is a common occurrence in single cell RNAseq analysis and may be due to cell loss during tissue digestion (of note, neutrophils typically have a lower per cell UMI count, but we did not observe neutrophils even when lowering the UMI threshold during analysis). To further account for doublets after the first round of annotation, we removed T/NK cells that had BCRs, B cells that had TCRs, and cells from nonlymphocyte clusters that had either TCRs or BCRs in the single-cell dataset. We also removed potential doublets from B and T cell VDJ output by excluding cells with more than two pairs of chain. After quality filtering, only cells that met either one of these criteria were kept: only one *TRB* chain, one *TRA* and one *TRB* chain, or one *TRB* chain and two *TRA* chains. Then, we reanalyzed the single-cell dataset and created new UMAP coordinates, differential gene expression list, and cluster/subcluster annotation. We used the following parameters for generating breast cancer single-cell UMAP and cluster: runUMAP: dims = 1:15, n.neighbors = 25, min.dist = 0.2; findNeighbors: dims = 1:30; FindClusters: resolution = 0.8. For T cell subclustering, we used following parameters: runUMAP: dims = 1:10, n.neighbors = 30, min.dist = 0.2; FindNeighbors: dims = 1:10; FindClusters: resolution = 0.3. Scripts can be found in the Zenodo repository (29): scripts/breast_cancer/4_single-cell_GEX/BC single-cell new cluster.Rmd; scripts/breast_cancer/4_single-cell_GEX/single-cell clustering without integration.Rmd; scripts/breast_cancer/4_single-cell_GEX/single-cell GEX versus VDJ coverage.Rmd.

For additional VDJ clonotype quality filtering and to calculate the clonal overlap with the Spatial VDJ output (defined by MiXCR), cells that belonged to different clones from Cell Ranger VDJ output but with identical nucleotide CDR3 sequence and J gene in both receptors (i.e., *TRA/TRB* in T cells, IG light chain/IG heavy chain in B cells, etc.) would be merged and assigned with a renamed clone ID by a customized R script, all cells within the new clone were assigned the minimum clone ID within the group. The T cells that had a single *TRB* sequence, which Cell Ranger had identified as separate clones, were merged to other clones expressing the matched *TRB* nucleotide CDR3 and J gene. We kept unmatched single chain *TRB* clones in the clonal list. In addition, we removed one potential doublet with cell barcode “TGCTACCGTTCAGGCC-4” by comparing its receptor chains to other cells in the same clone. All the dimension reduction and annotation results, along with the VDJ output files were imported into Loupe Browser (version 5.0, 10x Genomics) and Loupe VDJ Browser (version

4.0, 10x Genomics) for interactive analysis. Detailed code for VDJ clonotype quality filtering can be found in the Zenodo repository (29): scripts/breast_cancer/5_single-cell_VDJ/T_B cell VDJ analysis.Rmd.

Clonotype overlap analysis

To compare clonal lists from the MiXCR analysis result, including LR Spatial, SR Spatial, and Bulk SS3 VDJ, we used customized R script (R version 4.2.0) (Zenodo repository (29): scripts/tonsil/4_VDJ_overlap/tonsil overlap analysis.Rmd; scripts/tonsil/4_VDJ_overlap/tonsil LR-SR-bulk clonal overlap analysis.Rmd) and marked clones with identical nucleotide CDR3 sequence and J gene as the overlapped clones, which was consistent with the default components that MiXCR used for collapsing reads into clone. If multiple clones had identical nucleotide CDR3 sequence and J gene, we grouped them together as one overlapped case. To compare clonal lists from Cell Ranger VDJ (single-cell) and MiXCR (LR- and SR-Spatial VDJ) outputs, we analyzed the single-cell clone list with MiXCR output by matched nucleotide CDR3 sequence and J gene. We then calculated the percentage of LR Spatial VDJ clones that overlapped with other analysis methods. Clone fraction was calculated based on dividing each clonotype's UMI count with the total TR (*TRA* and *TRB*) or IG (*IGH*, *IGK*, and *IGL*) UMI count from either LR Spatial VDJ, SR Spatial VDJ, or bulk datasets. Detailed script can be found in the Zenodo repository (29): scripts/breast_cancer/6_VDJ_overlap/breast cancer LR single-cell VDJ overlap.Rmd; scripts/breast_cancer/6_VDJ_overlap/breast cancer single-cell VDJ clone overlap.Rmd.

Clonal evolution analysis

Preprocessing steps

For this analysis, we collapsed reads within the same UMI group to generate a consensus read by customized script in R (version 4.2.0). Individual reads from the same sample with identical UMI and spatial barcode were grouped by their full-length sequence. One or more read groups with the most abundant read count were selected. Then, we calculated the average sequencing quality score for each read group. First, we kept the read groups with the highest quality score. Second, if multiple groups were present, we selected the group with the longest sequence. Third, for the remaining read groups with identical length, average quality score, and read count, we masked the different bases between the groups as N and generated the final sequence for that UMI group. We then extracted the section information from the R1 header (descrsR1 column) from the reads alignment output, grouped reads by sections or replicates, and converted them to a count matrix (clone ID = row name; section number = column ID). Postfiltered LR Spatial VDJ reads were split into different segments including V, D, J, etc.,

by using anchor points ("refpoints" column) per sequence from the MiXCR alignment output. We referred to the anchor point definition and the script named "mixcr2imgt.py" from the Immcantation group to split LR-Spatial VDJ reads (32, 47). Then, we loaded the read list into R and converted the column name to the Adaptive Immune Receptor Repertoire (AIRR) compatible format (48). We restricted our analyses to the heavy chain (*IGH*), as it should provide the richest clonal structure (49). *IGH* clones defined by MiXCR have unique CDR3 sequences, which separates closely related *IGH* clones that have undergone somatic hypermutation. To group the MiXCR-defined clones together to define "*IGH* clonal families", we adapted the instruction named "Clustering sequences into clonal groups" from the Immcantation group. We calculated the nearest neighbor distances (minimum normalized hamming distance) between the nucleotide junction sequence and selected 0.1 as the threshold according to their distribution for tonsil LR *IGH* clonal families. The CDR3 sequence contains the conserved cysteine and tryptophan/phenylalanine residues on its 5' and 3' side according to the AIRR standards documentation. However, the MiXCR results already included those two residues in the CDR3 sequence. Therefore, we simply renamed the nucleotide MiXCR CDR3 sequence to "junction". Finally, we grouped the *IGH* clones into *IGH* clonal families by "DefineClones.py" script from SHazaM (version 1.1.1, Immcantation group) (47). We extracted the *IGH* reads, removed the ones without a valid C region annotation, and excluded one read that had an IGLC alignment. The final read list was used for class-switching and lineage analysis. Script for this analysis can be found in the Zenodo repository (29): scripts/tonsil/1_LR-SpatialVDJ/tonsil LR clonal family analysis.Rmd and script for previously described steps can be found in scripts/tonsil/1_LR-SpatialVDJ/tonsil LR read preprocessing.Rmd.

For the breast cancer samples, we defined the *IGH* clonal families only as the input data for receptor pairing (see section below on 'repair'). Here, we performed the same analysis as for the tonsil (described above) with the following thresholds (Patient 1: 0.1; Patient 2: 0.08).

IGH clonal family analysis

We imported the tonsil LR Spatial VDJ *IGH* clonal family read list and the STUtility object into the R (version 4.2.0), extracted gene expression cluster and follicle annotation and merged them into the read list. For follicle analysis (detailed script can be found in the Zenodo repository (29): scripts/tonsil/1_LR-SpatialVDJ/tonsil LR clonal family analysis.Rmd), we determined the follicle count per clonal family by counting distinct follicle ID per clone. Then, we converted all follicles as "in" to define intrafollicular *IGH* clonal fami-

lies. We grouped reads by follicle to calculate follicle stats including UMI count, spatial barcode count, clonal family count and their ratio. To describe the expansion level within each follicle, we ranked the follicles for each clonal family based on the max UMI percentage and extracted the value per follicle. We used vjtools to calculate *IGH* clonal family diversity measures (50, 51), including Chao1 estimate (52), Inverse Simpson index, and Shannon Entropy (script for this process can be found in the Zenodo repository (29): scripts/tonsil/1_LR-SpatialVDJ/vjtools_stats.sh). The D50 diversity index (defined as the minimum number of UMI needed to reach 50% of the total clone sequences) was calculated manually (53, 54). To compare *IGH* clonal families with different spatial distributions, we divided *IGH* clonal families into "restricted" versus "nonrestricted" using the following thresholds:

- Restricted *IGH* clonal families: more than 75% of its UMI count belonged to a single follicle.
- Nonrestricted *IGH* clonal families: less than or equal to 75% of its UMI count belonged to a single follicle and at least 25% of its total UMI belonging to its second largest follicle.

We only considered expanded *IGH* clonal families for this analysis (defined as having at least 20 UMI in total and being present in at least 10 spatial barcodes across all replicate tonsil sections).

To determine the *IGH* clonal family isotype fraction, we selected the dominant isotype (with the highest UMI) per clone family, since some of them had multiple isotypes per family. The number of *IGH* clonal families with or without multiple isotypes were also calculated. Script for this analysis can be found in the Zenodo repository (29): scripts/tonsil/1_LR-SpatialVDJ/tonsil follicle distribution.Rmd.

Class-switch recombination analysis

To analyze class-switching events, we used R (version 4.2.0) and removed *IGHGP* and *IGHD* reads from the tonsil LR Spatial VDJ *IGH* clonal family read list because *IGHGP* is a pseudogene and *IGHD* is regarded as an alternative splicing result from *IGHM*. B cells that exclusively express *IGHD* as a result of a "true" class-switching event are most often rare, but can occur in mucosal tissues, including the tonsil (11, 55, 56). However, we detected very few clones with *IGHD* as the dominant isotype. Therefore, we concluded that this individual had few, if any, true *IGHD* class-switched clones, and these were not investigated further for the purpose of class-switching. We also removed *IGH* clonal families with less than two reads since they would not contain class-switching events. We grouped reads within each clonal family by follicle, cluster, or spatial barcode, and extracted class-switched reads within the samples. Putative class switch recombination events (pCSR) were defined as two sequences with identical V

gene sequences with two different IGH constant chains/isotypes colocalizing in the same spatial barcode. For reads with identical V sequences within an IGH clonal family, we allowed one hamming distance in the nucleotide CDR3 sequence within an IGH clonal family (1 of 17 pCSR events). The script for this analysis can be found in the Zenodo repository (29): scripts/tonsil/8_IGH_clonal_evolution/tonsil_LR_IGH_class-switching_analysis.Rmd.

Lineage relationship analysis

To create lineage trees with SHazaM (Immucantation group) (47) in R (version 4.2.0), we imported the tonsil LR Spatial VDJ IGH clonal family read list and removed the reads that lacked a valid CDR3 sequence or contained indel(s) in the V sequence compared with the reference. We recovered the reference sequence based on the mutation patterns from the “bestVAlignment” column and inserted IMGT gaps. Then, we removed the reads that did not have a valid junction length (not divisible by 3, which potentially represented a frame-shift mutation). To create the lineage tree and assign unique subclone IDs to each node, we followed the vignettes of “Lineage reconstruction” in Alakazam (version 1.2.0, Immucantation group) (47) with phylip (version 3.697, J. Felsenstein) for creating lineage trees (57). Each node of the tree (hereafter referred to as “subclone”) represented a distinct V sequence versus the others and the size of the node is proportional to the number of UMIs belonging to that node. We traversed the lineage tree through depth-first search to assign adjacent ID to closely related subclones. We then plotted lineage trees and exported count matrices for visualizing nodes within the tissue section. We also matched the spatial barcodes to their follicle and cluster identity. Finally, we generated all possible trees and created the clone family subclones versus spatial barcode count matrix for statistical analysis and spatial visualization. The script for this analysis can be found in the Zenodo repository (29): scripts/tonsil/8_IGH_clonal_evolution/tonsil_LR_IGH_lineage_analysis.Rmd.

Visualizations of IGH clonal families and subclones on tissue

Count matrices of IGH clonal families and subclones were imported into R (version 4.0.5) and added to the tonsil object as new assays. To generate aligned x and y coordinates of all six tonsil sections, and enable plotting of all data jointly, we used the functions MaskImages() and AlignImages() of the STUtility package. Next, all nonzero observations of an IGH clonal family or subclone were plotted and colored by tissue section on a transparent background. The same IGH clonal family or subclone was then visualized on one of the sections using the STUtility FeatureOverlay() function. The

joint data image was superimposed on the tissue image and aligned by matching shared spots in the two image types using Adobe Illustrator. Details on how these plots were generated are provided in the Zenodo repository (29): scripts/tonsil/0_integrated/IGHfam_subclones_v2.Rmd. For script paths to all figure panels, see data S5

Association between single-cell deconvolution data and IGH clonal families in tonsil tissues

Given that $A \in R^{N \times c}$ is the cell type proportion matrix, and $B \in R^{N \times d}$ is the IGH clonal family matrix, where N is the total number of spots, c is the number of cell types and d is the number of selected IGH families. In this study, we removed the spots that were expressed as 0 on all selected IGH families. Thus, $N = 2047$, $d = 71$, and $c = 30$. The associations between cell subsets and 30 IGH families in tonsil tissue was defined as follows

$$Score = B^T A$$

The final normalized score was defined as the z -score of the quartiles of the association score on the null model (generated by 5000-times sample-wise permutation). P values were corrected using the Benjamini-Hochberg procedure. The script for this analysis was generated in R (version 4.2.2) and can be found in the Zenodo repository (29): scripts/tonsil/1_LR_SpatialVDJ/IGHfam_Celltype_Association.Rmd.

Pairing IG receptor chains with repair Repair analysis workflow

For pairing receptor chains, we filtered the LR Spatial VDJ count matrix to include all IGL and IGK clonotypes, but to simplify the pairing, only one IGH clonotype per IGH clonal family (the one with highest counts), since multiple IGH chains could pair with a single light chain due to somatic hypermutation. The resulting count matrix was used as input in repair (see below for details, deposited in the Zenodo repository (29): scripts/functions/star-repair-master.zip. Scripts are also deposited on GitHub (58), which is a method that we developed for pairing of receptor chains. Different values of spot and pairing score cut-offs were used to find the threshold that would generate the highest number of correct pairs and minimum number of incorrect pairs. A predicted pair was assigned as correct only if the two chains were a verified pair in the ground truth scVDJ dataset (which contained only pairs where both chains overlapped in both the spatial and single cell VDJ datasets), and incorrect if any of the chains were paired with another chain in the ground truth dataset.

The following command was used to run repair:

```
repair analyze -i [path_to_input] -o [path_to_output] -ap “^(IGH)” -bp “^(IGL)^(IGK)” -mxo [cut_off_value] -x 1 -b 100 -nb 1
```

For each sample (P1 and P2), repair was executed with different values (1, 5, 10, 15, and 20) for -mxo (minimum number of spots). Further details as well as input files and results are provided in the Zenodo repository (29): data/breast_cancer/10_repair/repair_run.

The repair method

As discussed above, repair was developed for unsupervised pairing of the receptor chains. In this method, we consider the task of pairing the different chains as an optimization problem. Below we elaborate on the details of this problem:

If $A \in R^{N \times c_a}$ is the normalized expression matrix for the chains of type a , and $B \in R^{N \times c_b}$ is the equivalent matrix for chains of type b , then we want to find a matrix $M \in R^{c_a \times c_b}$ and scalar $g \in R_+$ such that $A = gBM^T$. Here, N represents the number of observations, and the number of a type chains and b type chains are given by c_a and c_b respectively. M can be considered as a “mapping matrix” between chains of type b to chains of type a . In layman’s terms, element (i, j) in M tells us how much of chain b_j contributes to chain a_i ’s expression, once we’ve accounted for the difference in abundance between the different chain types (taken care of by g). The scaling factor is taken as a scalar, shared across all chains, an assumption that equates to assuming that the difference in abundance between chains of type a and b is not influenced by the chains’ identities.

During optimization we do not immediately learn the parameters M and g but rather the matrix Q and scalar p which relates to the former accordingly

$$M_{ij} = \frac{\text{pow}(c, Q_{ij})}{\sum_k \text{pow}(c, Q_{ik})} \text{ and } g = \exp(p)$$

Where $\text{pow}(c, x) = c^x$. Hence, M is the row-wise generalized softmax (if $c = e$, then this equals the standard softmax) of Q , meaning that $M_{ij} \in (0, 1)$. There are two benefits to using the (generalized) softmax function: (i) it allows us to interpret the element M_{ij} as the probability that chain b_j is paired with chain a_i , and (ii) the larger the value of the base c , the more focused the probability mass will be on a single value. The second feature is considered as a benefit since it favors sparse mapping from chains of type b to chains of type a , i.e., one b chain dominates the explanation of one a chain’s expression; notably, the same b chain can map to multiple a chains.

Thus, the final optimization problem becomes: $\min_{g, M} \|A - gBM^T\|^2$, with g and M defined as above. To solve this, we use stochastic optimization (Adams optimizer), implemented using the *jax* python package (v.0.3.1).

Once an optimal solution to M is found, as a pruning step, we formulate an unbalanced linear assignment problem (LAP) where we

negate this M and use the result as our cost matrix. This generates a bijective (one-to-one) pairing between a subset of all chain types of a and b (if $C_a > C_b$, each b chain will be paired to an a chain, but not every a chain will be paired with a b chain, and vice versa for $C_a < C_b$. If $C_a = C_b$, then each a chain and b chain will be assigned a partner). See the Zenodo repository (29): scripts/functions/star-repair-master.zip. Scripts are also available on GitHub (58).

Statistical analysis and figure presentation

Data was processed using R, Python, and Microsoft Excel for Microsoft 365 (version 16.76). For R and Python versions, see details under the correct subheading for each analysis. Statistical analyses were performed and figures were prepared using R, GraphPad Prism (version 9.4.01), and Keynote (version 11.2). The experimental outlines in Fig. 1A, Fig. 3A, and fig. S3A were prepared by graphic designer M. Karlén. Other illustrations were prepared in Keynote or Adobe Illustrator by the authors.

REFERENCES AND NOTES

- W. H. Hudson, L. J. Sudmeier, Localization of T cell clonotypes using the Vismium spatial transcriptomics platform. *STAR Protoc.* **3**, 101391 (2022). doi: [10.1016/j.xpro.2022.101391](https://doi.org/10.1016/j.xpro.2022.101391); PMID: 35707680
- S. Liu et al., Spatial maps of T cell receptors and transcriptomes reveal distinct immune niches and interactions in the adaptive immune response. *Immunity* **55**, 1940–1952.e5 (2022). doi: [10.1016/j.immuni.2022.09.002](https://doi.org/10.1016/j.immuni.2022.09.002); PMID: 36223726
- R. J. Bende et al., Germinal centers in human lymph nodes contain reactivated memory B cells. *J. Exp. Med.* **204**, 2655–2665 (2007). doi: [10.1084/jem.20071006](https://doi.org/10.1084/jem.20071006); PMID: 17938234
- A. Pellissier et al., Convergent Evolution and B-Cell Recirculation in Germinal Centers in a Human Lymph Node. *bioRxiv* 2022.11.09.463832 [Preprint] (2022); <https://doi.org/10.1101/2022.11.09.463832>
- M. Meylan et al., Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer. *Immunity* **55**, 527–541.e5 (2022). doi: [10.1016/j.immuni.2022.02.001](https://doi.org/10.1016/j.immuni.2022.02.001); PMID: 35231421
- P. L. Ståhl et al., Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016). doi: [10.1126/science.aaf2403](https://doi.org/10.1126/science.aaf2403); PMID: 27365449
- A. A. Tu et al., TCR sequencing paired with massively parallel 3' RNA-seq reveals clonotypic T cell signatures. *Nat. Immunol.* **20**, 1692–1699 (2019). doi: [10.1038/s41590-019-0544-5](https://doi.org/10.1038/s41590-019-0544-5); PMID: 31745340
- M. Singh et al., High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat. Commun.* **10**, 3120 (2019). doi: [10.1038/s41467-019-11049-4](https://doi.org/10.1038/s41467-019-11049-4); PMID: 31311926
- J. Stavnezer, J. E. J. Guikema, C. E. Schrader, Mechanism and regulation of class switch recombination. *Annu. Rev. Immunol.* **26**, 261–292 (2008). doi: [10.1146/annurev.immunol.26.021607.090248](https://doi.org/10.1146/annurev.immunol.26.021607.090248); PMID: 18370922
- A. Andersson et al., Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nat. Commun.* **12**, 6012 (2021). doi: [10.1038/s41467-021-26271-2](https://doi.org/10.1038/s41467-021-26271-2); PMID: 34650042
- H. W. King et al., Single-cell analysis of human B cell maturation predicts how antibody class switching shapes selection dynamics. *Sci. Immunol.* **6**, eabe6291 (2021). doi: [10.1126/sciimmunol.abe6291](https://doi.org/10.1126/sciimmunol.abe6291); PMID: 33579751
- J. G. Cyster, Chemokines, sphingosine-1-phosphate, and cell migration in secondary lymphoid organs. *Annu. Rev. Immunol.* **23**, 127–159 (2005). doi: [10.1146/annurev.immunol.23.021704.115628](https://doi.org/10.1146/annurev.immunol.23.021704.115628); PMID: 15771568
- W. H. Fridman, F. Pagès, C. Sautès-Fridman, J. Galon, The immune contexture in human tumours: Impact on clinical outcome. *Nat. Rev. Cancer* **12**, 298–306 (2012). doi: [10.1038/nrc3245](https://doi.org/10.1038/nrc3245); PMID: 22419253
- C. M. Laumont, A. C. Banville, M. Gilardi, D. P. Hollern, B. H. Nelson, Tumour-infiltrating B cells: Immunological mechanisms, clinical impact and therapeutic opportunities. *Nat. Rev. Cancer* **22**, 414–430 (2022). doi: [10.1038/s41568-022-00466-1](https://doi.org/10.1038/s41568-022-00466-1); PMID: 35393541
- C. Sautès-Fridman, F. Petitprez, J. Calderaro, W. H. Fridman, Tertiary lymphoid structures in the era of cancer immunotherapy. *Nat. Rev. Cancer* **19**, 307–325 (2019). doi: [10.1038/s41568-019-0144-6](https://doi.org/10.1038/s41568-019-0144-6); PMID: 31092904
- G. E. Phad et al., Clonal structure, stability and dynamics of human memory B cells and circulating plasmablasts. *Nat. Immunol.* **23**, 1076–1085 (2022). doi: [10.1038/s41590-022-01230-1](https://doi.org/10.1038/s41590-022-01230-1); PMID: 35761085
- T. N. Schumacher, R. D. Schreiber, Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015). doi: [10.1126/science.aaa4971](https://doi.org/10.1126/science.aaa4971); PMID: 25838375
- G. D. Victora, M. C. Nussenzweig, Germinal Centers. *Annu. Rev. Immunol.* **40**, 413–442 (2022). doi: [10.1146/annurev-immunol-120419-022408](https://doi.org/10.1146/annurev-immunol-120419-022408); PMID: 35113731
- L. Mesin et al., Restricted Clonality and Limited Germinal Center Reentry Characterize Memory B Cell Reactivation by Boosting. *Cell* **180**, 92–106.e11 (2020). doi: [10.1016/j.cell.2019.11.032](https://doi.org/10.1016/j.cell.2019.11.032); PMID: 31866068
- J. L. Laidlaw, A. H. Ellebedy, The germinal centre B cell response to SARS-CoV-2. *Nat. Rev. Immunol.* **22**, 7–18 (2022). doi: [10.1038/s41577-021-00657-1](https://doi.org/10.1038/s41577-021-00657-1); PMID: 34873279
- K. Peng et al., Diversity in immunogenomics: The value and the challenge. *Nat. Methods* **18**, 588–591 (2021). doi: [10.1038/s41592-021-01169-5](https://doi.org/10.1038/s41592-021-01169-5); PMID: 34002093
- L. J. Sudmeier et al., Distinct phenotypic states and spatial distribution of CD8⁺ T cell clonotypes in human brain metastases. *Cell Rep. Med.* **3**, 100620 (2022). doi: [10.1016/j.xcrm.2022.100620](https://doi.org/10.1016/j.xcrm.2022.100620); PMID: 35584630
- T. N. Schumacher, D. S. Thommen, Tertiary lymphoid structures in cancer. *Science* **375**, eabf9419 (2022). doi: [10.1126/science.abf9419](https://doi.org/10.1126/science.abf9419); PMID: 34990248
- V. Leko, S. A. Rosenberg, Identifying and Targeting Human Tumor Antigens for T Cell-Based Immunotherapy of Solid Tumors. *Cancer Cell* **38**, 454–472 (2020). doi: [10.1016/j.jccell.2020.07.013](https://doi.org/10.1016/j.jccell.2020.07.013); PMID: 32822573
- D. G. Schatz, Y. Ji, Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* **11**, 251–263 (2011). doi: [10.1038/nri2941](https://doi.org/10.1038/nri2941); PMID: 21394103
- M. M. Davis, P. J. Bjorkman, T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395–402 (1988). doi: [10.1038/334395a0](https://doi.org/10.1038/334395a0); PMID: 3043226
- M. Hagemann-Jensen et al., Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.* **38**, 708–714 (2020). doi: [10.1038/s41587-020-0497-0](https://doi.org/10.1038/s41587-020-0497-0); PMID: 32518404
- M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011). doi: [10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200)
- C. Engblom et al., Spatial transcriptomics of B and T cell receptors uncovers lymphocyte clonal dynamics, version 2.0, Zenodo (2022); <https://zenodo.org/record/7961605>
- H. Toosi, long-read-processing-STAR-paper, Github (2022); <https://github.com/HoseinT/long-read-processing-STAR-paper>
- D. A. Bolotin et al., MiXCR: Software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015). doi: [10.1038/nmeth.3364](https://doi.org/10.1038/nmeth.3364); PMID: 25924071
- J. A. Vander Heiden et al., pRESTO: A toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930–1932 (2014). doi: [10.1093/bioinformatics/btu138](https://doi.org/10.1093/bioinformatics/btu138); PMID: 24618469
- J. Bergenstråhle, L. Larsson, J. Lundberg, Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics* **21**, 482 (2020). doi: [10.1186/s12864-020-06832-3](https://doi.org/10.1186/s12864-020-06832-3); PMID: 32664861
- T. Stuart et al., Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019). doi: [10.1016/j.cell.2019.05.031](https://doi.org/10.1016/j.cell.2019.05.031); PMID: 31178118
- A. Andersson et al., Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun. Biol.* **3**, 565 (2020). doi: [10.1038/s42003-020-01247-y](https://doi.org/10.1038/s42003-020-01247-y); PMID: 33037292
- J. K. Perry, N. Kannan, P. M. Grandison, M. D. Mitchell, P. E. Lobie, Are trefoil factors oncogenic? *Trends Endocrinol. Metab.* **19**, 74–81 (2008). doi: [10.1016/j.tem.2007.10.003](https://doi.org/10.1016/j.tem.2007.10.003); PMID: 18054496
- M. Karlsson et al., A single-cell type transcriptomics map of human tissues. *Sci. Adv.* **7**, eabh2169 (2021). doi: [10.1126/sciadv.abh2169](https://doi.org/10.1126/sciadv.abh2169); PMID: 34321199
- Single cell type - SCGBID2, The Human Protein Atlas; <https://www.proteinatlas.org/ENSG00000124935-SCGBID2/single-cell-type>
- Single cell type - SCGB2A2, The Human Protein Atlas; <https://www.proteinatlas.org/ENSG00000110484-SCGB2A2/single-cell-type>
- M. Uhlen et al., Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015). doi: [10.1126/science.1260419](https://doi.org/10.1126/science.1260419); PMID: 25613900
- S100A8 protein expression summary, The Human Protein Atlas; <https://www.proteinatlas.org/ENSG00000143546-S100A8>
- CD44 protein expression summary, The Human Protein Atlas; <https://www.proteinatlas.org/ENSG00000026508-CD44>
- PABPC1 protein expression summary, The Human Protein Atlas; <https://www.proteinatlas.org/ENSG00000070756-PABPC1>
- R.-Y. Ma, A. Black, B.-Z. Qian, Macrophage diversity in cancer revisited in the era of single-cell omics. *Trends Immunol.* **43**, 546–563 (2022). doi: [10.1016/j.it.2022.04.008](https://doi.org/10.1016/j.it.2022.04.008); PMID: 35690521
- A. Andersson, Cell type enrichment within spatial regions, Github (2021); <https://github.com/almaan/her2st/blob/master/scripts/enriched-region.py>
- G. X. Y. Zheng et al., Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017). doi: [10.1038/ncomms14049](https://doi.org/10.1038/ncomms14049); PMID: 28091601
- N. T. Gupta et al., Change-O: A toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* **31**, 3356–3358 (2015). doi: [10.1093/bioinformatics/btv359](https://doi.org/10.1093/bioinformatics/btv359); PMID: 26069265
- J. A. Vander Heiden et al., AIRR Community Standardized Representations for Annotated Immune Repertoires. *Front. Immunol.* **9**, 2206 (2018). doi: [10.3389/fimmu.2018.02206](https://doi.org/10.3389/fimmu.2018.02206); PMID: 30323809
- J. Q. Zhou, S. H. Kleinstein, Cutting Edge: Ig H Chains Are Sufficient to Determine Most B Cell Clonal Relationships. *J. Immunol.* **203**, 1687–1692 (2019). doi: [10.4049/jimmunol.1900666](https://doi.org/10.4049/jimmunol.1900666); PMID: 31484734
- M. Shugay et al., VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput. Biol.* **11**, e1004503 (2015). doi: [10.1371/journal.pcbi.1004503](https://doi.org/10.1371/journal.pcbi.1004503); PMID: 26606115
- M. O. Hill, Diversity and evenness: A unifying notation and its consequences. *Ecology* **54**, 427–432 (1973). doi: [10.2307/1934352](https://doi.org/10.2307/1934352)
- A. Chao, C.-H. Chiu, Species richness: Estimation and comparison in Wiley StatsRef: Statistics Reference Online (Wiley, 2016), pp. 1–26; <https://doi.org/10.1002/9781118451152.stat03432.pub2>
- J. Han, Immunodiversity assessment method and its use, World Patent (2012); <https://patentimages.storage.googleapis.com/15/3d/a3/fc66068d388212/WO2012097374A1.pdf>
- Y. Yang et al., Distinct mechanisms define murine B cell lineage immunoglobulin heavy chain (IgH) repertoires. *eLife* **4**, e09083 (2015). doi: [10.7554/eLife.09083](https://doi.org/10.7554/eLife.09083); PMID: 26422511
- C. Gutzeit, K. Chen, A. Cerutti, The enigmatic function of IgD: Some answers at last. *Eur. J. Immunol.* **48**, 1101–1113 (2018). doi: [10.1002/eji.201646547](https://doi.org/10.1002/eji.201646547); PMID: 29733429
- B. S. Steiniger, L. Raimer, A. Ecke, B. A. Stuck, Y. Cetin, Plasma cells, plasmablasts, and AID⁺/CD30⁺ B lymphoblasts inside and outside germinal centers: Details of the basal light zone and the outer zone in human palatine tonsils. *Histochem. Cell Biol.* **154**, 55–75 (2020). doi: [10.1007/s00418-020-01861-1](https://doi.org/10.1007/s00418-020-01861-1); PMID: 32172287
- J. Felsenstein, PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166 (1989).
- A. Andersson, star-repair, Github (2022); <https://github.com/almaan/star-repair>

ACKNOWLEDGMENTS

We acknowledge M. Stoeckius, C. Gallant, and K. Pfeiffer at 10x Genomics, Inc., for resources and technical support; M. Karlén's graphical design; H. Lönnqvist, A. Mollbrink, S. Giatrellis, L. Larsson, J. Taneyvski, L. Franzén, and L. Bergenstråhle for expert assistance; and Frisén, Lundberg, and Sandberg lab members and G. Al-Eryani for critical feedback. **Funding:** This work was supported by the Swedish Research Council (201806217_VR) (J.La., J.F., and J.H.), the Swedish Cancer Society (J.F. and J.Lu.), the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Actions grant agreement no. 844712 (C.E.), and German state funds approved by the State Parliament of Baden-Württemberg for the Innovation Campus Health + Life Science Alliance Heidelberg Mannheim (C.L.). The authors acknowledge support from the National Genomics Infrastructure in Stockholm and Uppsala funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for

Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. **Author contributions:** Conceptualization: J.Mo., C.E., K.T., and J.F.; Methodology: C.E., K.T., Q.L., J.Mo., A.A., H.T., M.H.J., J.Lu., and J.F.; Data curation: K.T., Q.L., and C.E.; Formal analysis: C.E., K.T., Q.L., A.A., H.T., C.L., and S.S.; Funding acquisition: J.Mo., C.E., J.La., J.H., J.Lu., and J.F.; Investigation: C.E., K.T., Q.L., J.Mo., X.C., E.S., C.L., G.M., and M.H.J.; Project administration: C.E., K.T., and J.Mo.; Resources: M.J., J.Mi., J.H., J.La., J.Mo., J.Lu., and J.F.; Software: K.T., Q.L., A.A., C.L., and H.T.; Supervision: J.Mo., J.H., J.S.R., J.La., J.Lu., and J.F.; Validation: C.E., K.T., Q.L., and J.Mo.; Visualization: K.T., Q.L., and C.E.; Writing – original draft: C.E., K.T., Q.L., and J.Mo.; Writing – review and editing: all authors. **Competing interests:** C.E., K.T., J.Mo., J.F., J.Lu., and Q.L. are inventors on patent (US Patent no. US11692218B2) encompassing this work. C.E., K.T., Q.L., A.A.,

H.T., S.S., J.Mo., J.Lu., and J.F. are scientific consultants for 10x Genomics Inc, which holds IP rights to this spatial technology. J.F. and J.L. have stock options in 10x Genomics, Inc, and J.Mo. has stock in Pacific Biosciences. J.S.R. reports funding from GSK, Pfizer, and Sanofi and fees from Travele and Astex Therapeutics. The remaining authors declare no competing interests. **Data and materials availability:** All data, code, and images needed to reproduce the figures in the manuscript are deposited on Zenodo (29). References to data and code per figure panel are annotated in table S5. The sequencing data have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession numbers EGAD00001011061 (breast cancer) and EGAD00001011062 (tonsil). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.science.org/about/>

[science-licenses-journal-article-reuse](#). This research was funded in whole or in part by Swedish Research Council (201806217_VR) and the European Union's Horizon 2020 research and innovation program (844712). The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf8486

Materials and Methods

Figs. S1 to S28

References (59–66)

MDAR Reproducibility Checklist

Data S1 to S5

Submitted 22 November 2022; accepted 23 October 2023
10.1126/science.adf8486