# Deep profiling of mouse splenic architecture with CODEX multiplexed imaging

Yury Goltsev[1,3], Nikolay Samusik[1,3], Julia Kennedy-Darling[3],
Salil Bhate[3], Matthew Hale[3,4], Gustavo Vasquez[3], and Garry P. Nolan[2, 3]

**Department of Microbiology and Immunology,**
**Baxter Laboratory in Stem Cell Biology,**
**Stanford University School of Medicine**
**Stanford, California, USA.**

1 – Contributed equally

2 – Corresponding author

3 – Department of Microbiology and Immunology, Baxter Laboratory in Stem Cell Biology, Stanford University School of Medicine

4 – Current address: Smart Tube, Inc., 922 Industrial Ave., Palo Alto, CA 94303

## Abstract

A cytometric imaging approach, called CO-Detection by indEXing (CODEX), that enables high parameter multiplexing of antibody-tagged target epitopes is used here to create high parameter imaging datasets of normal mouse and lupus (MRL/lpr) spleens. In this procedure, antibody binding events are rendered iteratively using DNA barcodes, fluorescent dNTP analogs, and an in-situ polymerization-based indexing procedure. Fluorescent signals from multiple rounds of indexing are computationally combined into a multi-channel image stack and subjected to image segmentation and quantification. A segmentation and linear model algorithm was developed to accurately quantify membrane antigen levels on dissociated cells as well as tissue sections. Leveraging the spatially resolved nature of CODEX multiplexed single-cell imaging data, quantitative *de novo* characterization of lymphoid tissue architecture was enabled and overlaid onto previously described morphological features. We observed an unexpected, profound impact of the cellular neighborhood on the expression of protein receptors on immune cells. By comparing normal murine spleen to spleens from animals with systemic autoimmune disease (MRL/*lpr*), extensive and previously uncharacterized splenic cell interaction dynamics in the healthy versus diseased state was observed. The fidelity of multiplexed imaging data analysis demonstrated here will allow deep proteomic analysis and systematic characterization of complex tissue architecture in normal and clinically aberrant samples.

## Introduction

Tissue imaging has been at the foundation of basic research and clinical studies since the advent of the microscope. Dyes that recognize cellular constituents and the conventional use of fluorophore-conjugated antibodies to tag specific epitopes typically allow only limited multiplexing--largely based on the number of fluorophores that can be simultaneously imaged. Microscopes with associated optical systems capable of reading five fluorophores simultaneously are common in academic practice, with specialty instruments reaching eight parameters. Approaches have been developed to overcome such limitations[1-3], but these protocols have required multiple stain/strip/wash cycles of the antibodies that can be time consuming or lead to sample degradation over time. Multiplexed ion beam imaging[4] and imaging mass cytometry[5] methods that detect antibodies tagged with metal isotopes on platform-dependent mass spectrometry instruments have a multiplexing capability reported to 40 channels; however, instrument availability and user expertise has so far limited the adoption of these technologies.

The need for highly multiplexed imaging is especially critical to our understanding of tissue architecture at the subcellular, cellular, and tissue level. Indeed, initiatives have been recently proposed to detail a cellular scale map in the entire human body as well as model organisms (Human Cell Atlas, Chan-Zuckerberg Initiative). As such, among the many ‘omics systems of interest, a deep phenotyping, multiplexed histological analysis would provide linkage to existing developmental and clinical knowledge of tissue structure. The potential for multiplexed in situ imaging at cellular and subcellular scales can be measured against the understandings brought by decades of fluorescence-based flow cytometry to basic researchers and clinicians alike[6]. The advent of cytometry technologies based on mass spectroscopy has increased the depth at which single cell proteomics and RNA expression can be accomplished[7-14]. Such flow cytometric procedures have provided critical biological information at the single-cell level regarding ploidy, immunophenotype, frequency of cell subsets, and expression levels of proteins, as well as functional characterization in basic biology and biomedical arenas.

The approach described here (CODEX, for CO-Detection by indEXing) extends these deep phenotyping capabilities to most standard three-color fluorescence microscope platforms for imaging of solid tissues. Using polymerase driven incorporation of dye labeled nucleotides into the DNA tag of oligonucleotide-conjugated antibodies accurate high multiplexed single-cell quantification of membrane protein expression in densely packed lymphoid tissue images, (which was once deemed impossible [15] ) was achieved. Automatic delineation of cell types from multidimensional marker expression and positional data generated by CODEX enabled deep characterization of cellular niches and their dynamics during autoimmune disease both for major and rare cell types populating mouse spleen. A rich source of multivariate data is generated and provided for the community to further efforts in refining algorithms for tissue segmentation, sub tissue neighborhood analysis, and rare cell type detection (http://welikesharingdata.blob.core.windows.net/forshare/index.html).

## Results

### Single base primer extension enables multiplexed antigen staining

DNA provides an ideal substrate for the design and construction of molecular tags due to its combinatorial polymer nature. An indexable tagging system whereby specific tags are iteratively revealed *in situ* by a stepwise enumeration procedure was designed. Antibodies (or other affinity-based probes) are first labeled with oligonucleotide duplexes that encode uniquely designed sequences with 5′ overhangs (**Figure 1A**). Cells are stained with a mixture containing all tagged antibodies at once. During iterative cycles of visualization of labeling (rendering) the sequence of the 5′ overhang determines the index (the combination of a polymerization cycle and a fluorescent channel) at which a given DNA tag incorporates one of two fluorescently labeled dNTP species. Specifically, the antibody-matched overhangs (indexes) include a region to be filled by blank letters and a dedicated position for a dye labeled nucleotide at the end. The antibodies to be revealed first generally have shorter overhangs than the antibodies to be visualized (rendered) in later cycles.

In **Figure 1**, we show how the CODEX (CO-Detection by indEXing) method works by iteratively extending the 3′ of the overhang (with DNA polymerase) in the presence of one of two non-labeled INDEX nucleotides (dATP or dGTP) along with two fluorescently labeled “rendering” nucleotides (dU-SS-Cy5 and dC-SS-Cy3). A set of 3 pairs of indexed antibodies bound to a cell are schematized. The first indexing nucleotide (in this case G (dGTP)) is added (by primer extension) along with fluorescent nucleotides dUTP and dCTP. Note that all 6 antibodies have the unlabeled INDEX “G” at the first position (across from the C in the lower strand). Only antibodies 1 and 2 are competent to be extended with (dU-SS-Cy5 and dC-SS-Cy3) that is present in the mixture during the primer extension because Abs 3, 4, 5 and 6 have a T in lower strand immediately after the C in the lower strand. Cells are washed of free nucleotides and the slide is imaged. At this step, only cells which have absorbed the antibodies 1 and 2 will be visualized. Then, a clearing step is performed using TCEP—which cleaves the disulfide linkers to release the fluorophores—and the slide is washed. The slide is now ready for indexing cycle #2. To reveal the next index, observe the T nucleotide in lower strand of the antibodies 3 , 4, 5 and 6. At this second cycle, a non-fluorescent A (dATP) is used to fill index position 2-- which then allows for fluorescent nucleotides U and C to be incorporated onto Abs 3 and 4. The cells are reimaged with this cycle 2 set of antibodies. Again, the fluorophores are cleaved with TCEP and the cells are washed. One repeats these two indexing cycles where the next index position is filled by G, such that the Abs 5 and 6 are now competent to be filled with fluorescent dUTP and dCTP. The fourth indexing cycle would employ non-fluorescent A again, and so forth. Each cycle requires approximately 10 minutes. Imaging time on every cycle can vary from minutes to hours depending on the sample dimensions, resolution at which the image is taken, and the microscope specifications. Importantly, the system enables multiplexed tissue imaging analysis by means of a standard fluorescence microscope.

To test the premise of the system, isolated mouse spleen cells were incubated with a CD4 antibody conjugated to an indexing oligonucleotide duplex (as represented by Ab 1 in **Figure** 1A). In this trial experiment TCRβ-Alexa 488 was used as a counterstain. A single round of primer extension was done with a mix of unlabeled dGTP and dUTP-ss-Cy5. A cell population positive for both CD4 and TCRβ was observed by flow cytometry. Observation of this population was dependent on the addition of Klenow DNA polymerase to the reaction mixture (**Figure** 1B, C) proving the feasibility of rendering the antibody binding pattern by primer extension. Similarly, in tissue sections, CODEX tag-conjugated antibodies produced lineage-specific staining comparable to regular fluorescent antibodies (see staining patterns of B220-CODEX and B220-APC in mouse spleen, Figure 1D-F).

In a simulated multicellular mix created by combining 30 batches of mouse splenocytes barcoded with pan-leukocytic CD45 antibody labeled with a set of 30 distinct CODEX tags, the visualization of the CODEX 15-cycle staining pattern indicated the comparable quantitative rendering of every cellular batch per designated cycle, with low background, no signal carryover between cycles, and lack of signal deterioration across the 15 cycles (**Supplementary Figure** 1A-F). As per flow cytometric and other cell staining approaches, appropriate titration of reagents minimizes background binding events and increases signal to noise.

## Multidimensional staining of mouse hematopoietic cells

To validate the quantitative performance of CODEX, cells isolated from mouse spleens were co-analyzed by mass-cytometry (CyTOF) and CODEX using identical 24-antibody panels (**Supplementary Table** 1 available online[38]). Use of the same antibody clones and the same splenocyte preparation ensured the validity of comparisons. CyTOF analysis was performed on cell suspensions stained with metal-tagged antibodies as previously described. For CODEX analysis, isolated spleen cells were stained a panel of antibodies conjugated to indexing oligonucleotides. Samples were fixed to a coverslip (**Figure** 2A) and imaged over 12 cycles of CODEX protocol. Images were segmented using the *in situ* cytometry software toolkit developed for this study (see Materials and Methods and **Figure** 2A for exemplary segmentation of the cell spread), and the staining of individual cells across the indexing cycles was quantified. Segmentation data was converted into flow cytometry standard (FCS) format and analyzed using the conventional flow cytometry analysis software Cytobank. Biaxial scatterplot gating analysis revealed a consistent similarity in lineage-positive populations between CODEX and CyTOF data (**Figure** 2B).

In further experiments (see below) the scope of CODEX was expanded to analysis of tissue sections. A 3D segmentation algorithm (see Materials and Methods) was therefore created to combine information from the nuclear staining and a ubiquitous membrane marker (in this case CD45) to define single-cell boundaries in crowded images such as lymphoid tissues. For each segmented object (i.e., cell) a marker expression profile (**Supplementary Table 2** available online[38]), as well as the identities of the nearby neighbors were recorded (using Delaunay triangulation, **Supplementary Table 3** available online[38]). In contrast to cells in suspension or dissociated cells spread on glass (**Figure** 2A), cells in tissue sections are adjacent to each other-- therefore a large fraction of each cell's membrane is in direct contact with the membranes of neighboring cells (**Figure** 2C). Depending on how the quantitation of marker expression per cell is performed, this might lead to a contribution of fluorescence from neighboring cells to the cell of interest (**Figure** 2D).

To address this latter challenge, a novel linear algorithm for positional spillover compensation was created. This algorithm is based on the same principles used in fluorescent spillover compensation in traditional flow cytometry, except that our algorithm performs compensation between physically adjacent cells based on approximated pairwise cell-to-cell contact ratios (**Figure** 2D). Use of this compensation method resulted in a considerable reduction of spillover signal (**Figure** 2E).

### Multidimensional analysis of cellular neighborhoods in murine spleen

Despite the large number cells that traffic through the mouse spleen, this lymphoid organ maintains clear structural sub-compartments-- each with distinct cellular compositions and functions. The white pulp, wherein T and B lymphocytes are spatially segregated into distinct T cell-rich zones (peri-arteriolar lymphoid sheath, PALS) and B cell-rich zones (follicles), is circumscribed by the marginal zone. The red pulp contains cells of erythroid lineage and a variety of innate immune cells including granulocytes, macrophages, and dendritic cells (see schematics in Figure 3B). Much of this structure and function has been laboriously determined over many years — and has never been visualized in a multiplexed system comparable to high dimensional cytometry.

A 30-antibody panel was therefore designed to identify splenic-resident cell types (lymphocytes, macrophages, microvessels, conduit system, splenic stroma; **Figure** 3A, **Supplementary Table** 1 available online[38]) and applied to the cryo-sections of spleens from wild-type (3 spleens) and MRL/lpr mice (6 spleens) (**Supplementary Figure 7**). The staining patterns of 28 DNA-conjugated antibodies were acquired over 14 cycles of CODEX imaging and overlaid with 2 additional fluorescent antibodies, CD45-FITC and NKp46-Pacific Blue and a DRAQ5 nuclear stain (**Figure** 3A and low-resolution views in **Supplementary Movie** 1 available online[38]). Each tissue was imaged with a 40x oil immersion objective in a 7x9 tiled acquisition at 1386x1008 pixels per tile and 188 nm/pixel resolution and 11 z-planes per tile (axial resolution 900 nm). Images were subjected to deconvolution to remove out-of-focus light. After drift-compensation and stitching, we obtained a total of 9 images (one per tissue) with x=9702 y=9072 z=11 dimensions, each consisting of 31 channels (30 antibodies and 1 nuclear stain).

4 major classic splenic compartments: red pulp, B-cell follicle, PALS and marginal zone (MZ) (**Figure 3B**) could be easily discerned in CODEX imaging data (**Figure 3A)**. Next, the CODEX data was subjected to segmentation, quantification and compensation, as described above, yielding a total of 734101 30-dimensional single-cell protein marker expression profiles (**Figure** 3C, **Supplementary Table**

**2** available online[38]). The segmented CODEX data was subject to automated phenotype mapping algorithm X-shift that was previously developed and validated on CyTOF data [16] (**Figure 3C**). 58 phenotypic clusters inferred by X-shift clustering were manually annotated (**Figure** 3C, D and **Supplementary Figure 2** available online[38]) based on the 30-color marker expression profile and thorough visual inspection of the representative image samples (**Supplementary Figure 2.1-2.27** available online[38]). Some clusters were found to originate from imaging artifacts such as dust and tissue sectioning defects. That reduced the overall number of cell-like objects to 707466. Each cluster was assigned to one of 27 broadly defined single-cell phenotypic groups (cell types), which in some cases could be clearly matched to major immune cell types and in others were named according to expression of distinguishing surface markers (see cluster annotation and cell counts in **Supplementary Table 4** available online[38]).

Notably this analysis confirmed that even rare computationally derived cellular phenotypes closely matched cell types expected to be observed in normal spleen. For example, CD4$^{hi}$/CD3$^{-}$/MHCII$^{hi}$ cells were identified by X-shift clustering as a rare (1321 out of total 707466 cells) yet distinct cell type present in the spleen (**Supplementary Figure 4**A, B). The CD4$^{hi}$/CD3$^{-}$/MHCII$^{hi}$ cells were sorted out with FACS and subject to microarray expression profiling, which revealed that they were similar to lymphoid tissue inducer (LTi) cells of the ILC3 subtype of innate lymphoid cells [17] (**Supplementary Figure 4**, **Supplementary Figure 3**.**28** available online[38]). These cells are crucial for the creation of proper splenic microarchitecture during development. Ectopic introduction of LTis can induce formation of secondary and tertiary lymphoid structures [18] [19]. These cells occupied a distinct location at the border between T and B cells in the normal tissue (**Supplementary Figure 4**, **Supplementary Figure 3**.**28** available online[38]) and lost their distinct localization pattern in the MRL/lpr mice. CD11c$^{+}$ B cells (age associated B cells (ABCs), have been shown to be a key participant in the triggering of certain autoimmune responses [20,21]) are another example of unsupervised identification of a rare cell type in CODEX data, the splenic location of which has not been previously described in the literature. We observed them to tightly associate with conventional dendritic cells (cDC) and occupy a distinct peri-follicular space in the boundary between PALS and B-zone. Interestingly, these cells diminished in numbers and redistributed towards intra-follicular space in the MRL/lpr spleens (**Supplementary Figure 2.5, Supplementary Figure 3**.**2** available online[38])**.** While multiple additional interesting cell-cell associations were observed, it is obviously beyond the scope and the goals of this report to describe all of the apparently new observations pertaining to every cellular subset delineated in this study. While we have created a compendium of observed cell types and their associations, we will leave this as a resource to the community—and relevant experts—to mine this data for biological or clinical significance.

Compared to CyTOF data on splenocytes isolated from homogenized spleen, CODEX *in situ* analysis produced a similar distribution of cell counts for major cell types except that CODEX identified larger numbers of resident and stromal cell types such as erythroblasts and F4/80 macrophages than CyTOF did (**Figure 3E**). This result can be explained given that these cell types are tightly intertwined with the splenic stroma and extracellular matrix, and thus many of these cells can get discarded during cell suspension extraction procedures used during traditional flow cytometry approaches.

CODEX analysis provided a unique view at cell type composition of the splenic tissue on the macro as well as the micro levels from the perspective of cell-to-cell contacts. Pseudo-color diagrams of cell type distribution in splenic tissue provided a visual overview of distribution of the 27 cell types in the splenic tissue, providing patterns that were otherwise not obvious from single channel overlays. (**Figure 3F and 5A**). Various automated approaches were then developed to quantitatively describe the splenic architecture as defined by cell-to-cell contacts and composition of cellular neighborhoods. To provide a high-level view of the cell type interaction landscape, the total counts of contacts between every pair of cell types in the Delaunay neighborhood graph (see schematics in right panel of **Figure 4A** and **Supplementary Table 3** for the data, available online[38]) for each condition was determined. The specificity of cell-to-cell interaction was estimated from the “log odds ratio” metric (ratio of observed probability to expected probability of cell-to-cell contact occurring by chance) (**Supplementary Table 5** available online[38]**)**. When visualized as heatmaps, this metric revealed a significant non-random distribution of cells in the spleen. In the majority of cases cell types were either selectively associating or avoiding each other (red or blue on the heatmap) pointing to prevalence of specific cell-to-cell interactions in shaping the spleen architecture. The major splenic anatomic compartments were reflected in two large mutually exclusive clusters of positive associations, which appeared to correspond to cell types populating the red pulp and the white pulp, respectively (indicated with black rectangular outlines on **Figure 3G**). For example, a significant positive association was observed between F4/80$^+$ macrophages and erythroid cells, as these cell types are both found in the red pulp and are closely associated in so-called erythroblast islands [22,23]. Also, as expected, a mutual avoidance was observed between cells known to more exclusively inhabit only the red or the white pulp areas (**Figure 3G**). An avoidance of interaction was also observed between T and B cells, reflecting concentration of these cell types in B cell follicles and PALS, respectively (**Figure 3G**).

Unexpectedly, the highest degree of association was observed between the cells of same phenotypic class (**Figure 3G**, bright red diagonal), suggesting that homotypic adhesion constitutes a major force driving the heterogeneity of cellular distribution in immune organs. This observation held true both for the major constituents of white pulp, T and B cells, as well as for rare cell types such as NK cells. Interestingly, even though CD8 and CD4 T cells tended to mix in the PALS, their mutual distribution was nonrandom and consisted of intertwined threads of homotypic cells (**Supplementary Figure 5A**). Interestingly, as an aside, similar structures could be reproduced in vitro by incubating heterotypic mixtures of sorted splenic cell populations (**Supplementary Figure 5B, C**). These data suggest that homotypic cell association might be an important driver of the white pulp substructure and is worth investigation under other auspices.

The precision in situ cytometry analysis of CODEX data allowed enumeration of cellular contexts in a manner not possible previously. We specify here an indexed “niche” (i-niche) as a ring of cells (excluding the central, or here defined as "index" cell) in no specific circumferential order that are in direct contact with the index cell (**Figure 4A**). We distinguish i-niche from the more formal understanding of "niche", which is often used in stem cell literature and where numbers of cells in the niche and their placement within the niche is undefined. In our definition, we allow the central cell to be of any type and are counting the cell types present in the ring. This flexible definition allows for multi-cellular interactions around a central cell to define the biology of that cell (and vice versa). Computationally, the i-niche window slides from cell to cell, considering each set of adjoining cells—and therefore allows consideration of the constituencies of different central cell types that might populate a given i-niche. We understand that our current definition is arbitrary and could be extended to include other specific cell arrangements—including, though beyond the scope of the current work, a 3D sphere of cells contacting the index cell.

We identified 100 of the major i-niches (by K-means clustering) according to the relative frequency of the identified cell types present in the ring of cells surrounding the index cell (Figure 4A) —where in our definition the index cell in the center can be ***any*** cell. For instance, in Figure 4B-i (top panel) and expandable view in **Supplementary Figure 6** the first column represents an i-niche which only has B cells surrounding the index cell. As a second example, the i-niches #10 and #53 (indicated by red arrows) contain variable numbers of B cells and marginal zone macrophages. The combinations of cells that appear in the ring structure are limited (Figure 4B-i). In fact, most of the i-niche structures do not contain more than 3 primary cell types—indicating that in most cell niches there is likely to be homotypic association of multiple cells of a given type.

Considering again the first column of niche composition heatmap (Figure 4B-i, i-niche #96), where the i-niche ring consists of only B cells, we mapped these i-niches back into the original tissue structure where they can be seen primarily in the follicular zone B cell region (**Figure 4C, left panel**). However, when the i-niche ring had B cells with at least one marginal macrophage neighbor (as identified by presence of CD169), these i-niches mapped back to the tissue—and as such would be conventionally identified as marginal zone B cells (**Figure 4C, right panel**). While this might seem a circular argument, it is important to remember that traditional surface marker analysis and cytometric gating do not permit unequivocal separation between marginal zone B cells and follicular B cells with the markers used (e.g. using B220, CD19, CD21/35, IgD, IgM etc). However, using this i-niche strategy we can identify and rapidly map all such B cells to their sub-tissue origin simply by considering the immediate neighbors of the index cell. In other words, the signature of tissue substructure is already evident by the most immediate neighbors in each cellular region.

Similar to the case of B cells, specific populations of T cells could be discerned based on the residence within i-niches. T and B cells are known to utilize the extracellular matrix (ECM) secreted by stromal cell network (conduit system) as cues for migration. Therefore, in the absence of any known markers, classification of T cells into those residing in the PALS versus red pulp versus residing in the red pulp and being in the contact with ERTR7 ECM is enabled by neighborhood context analysis. For instance, CODEX data enabled precise selection of the T-cells residing in ERTR7 enriched niches (in Figure 4B-iii see the column below the grey rectangle indicating a family of niches where index T cells contact ERTR7 stroma; as well see **Supplementary Figure 8.2 and 8.3** available online[38]) showing distribution of T cells in contact with ERTR7$^+$ stromal cells).

Taken together, while we see that surface marker expression alone was insufficient to associate many cell subsets within a given tissue subcompartment (e.g. CD4$^+$ T cells can be found both in the PALS and in the red pulp) -- i-niche designation does provide such mapping data (most of i-niches were enriched within a specific splenic subdivision Figure 4B-v). This begs interesting questions-- can new cell types, or functional subsets, be discerned by this approach? What is the frequency of a repeating i-niche structure that must be observed to suggest a function? And what would constitute proof that a given i-niche enumerates a new cell type or functionality?

One way to address these latter questions is to consider the phenotypes of the index cell in various i-niches. We observed that for several index cell types there was significant biasing of the surface marker expression depending on the i-niche in which the index cell resides. To systematically profile this effect, we subsetted the B and CD4 T index cells according to the i-niche order in **Figure 4B-i**, and depicted the average marker expression level of the index cell depending on the i-niche in **Figure 4B-ii and –iii**, respectively. For B cells, a compelling difference is observed in the CD21/35 and CD35 expression levels depending on the cell i-niche (neighborhood). In i-niches #10, 72, 53 and 33 (two red rectangles above **Figure 4B-ii**), CD21/35 and CD35 expression is high in these B cells when they are near marginal zone macrophages, follicular zone dendritic cells, other B cells. This is in stark contrast to the lower expression of CD21/35 and CD35 for the index B cell in every other i-niche. Notably, the literature supports higher levels of CD21/35 as one of the key markers of marginal zone B-cells [24].

Another example for B cells is the level of expression of B220 and CD19, which are low especially when B-cells sits in the family of i-niches dominated by presence of F4/80$^+$ macrophages (cyan rectangle at top of **Figure 4B-ii)**, B220 is a membrane associated protein tyrosine phosphatase known to be an essential regulator of BcR signaling. In view of similar co-stimulatory role of CD19 this observation points to attenuated signaling state of B cells populating the red pulp. However, when B cells are adjacent to cells of the ERTR7-positive stromal mesh - the B220 and CD19 levels on the index B cells are significantly higher (see **Figure 4B-ii** columns above the purple rectangle – for the composition of ERTR7 high i-niches and below in **Figure 4B-ii** where the CD19 and B220 expression is high). These observations suggest a potential link between signaling capacity of B cell membrane complexes and B cell proximity to splenic stroma.

For T cells, the story is analogous. For instance, CD27 and CD90 expression levels in the index CD4 T cells are highly variable across the various i-niches (consider the columns under the yellow and green bars in Figure 4B-iii). Interestingly, in association with strongly B cell rich i-niches (yellow bar) CD90 is diminished and CD27 shows varying levels of expression. Given that CD27 is a known T cell activation marker—associated with long term T cell memory amongst other functions[25], it might not be surprising to see varying levels of CD27 expression (columns beneath yellow bar, Figure 4B-iii). But why is CD90 expression high in CD4 cells when they are in CD4 or CD8 cell contexts (rows beneath the green bar)? Considering the elusive role of CD90--which can in some situation substitute for CD28 co-activation [26] – one can hypothesize that perhaps CD90 provides an alternative tonic signal that is required for the activity of these cells.

As another example consider the complex non-linear relationship between i-niche dependent levels of two other molecules, CD79b and B220, on B cells (**Figure 4D**). CD79b is a co-activator chain of the B cell receptor complex. CD79b is co-expressed with B220 as a large spread of the CD79b vs. B220 levels (shown on a scatter plot of isolated single cell splenocytes **Figure 4D**, top right panel). Such a distribution of expression is sometimes attributed to staining issues, measurement noise, or a simple lack of understanding of the underlying biology. However, as seen on Figure 4D, upper left panel, there is a non-random pattern of CD79b and B220 expression across i-niches, and , depending on the B220/CD79b levels, the i-niches map to specific regions in the splenic architecture (**Figure 4D**, lower panel). For instance, index B cells that were B220$^{int}$, CD79b$^{lo}$ (i-niche “59”) inhabited the boundary areas between the PALS and the follicles (**Figure 4D** image montage on the bottom). Index B cells that were B220$^{lo}$, CD79b$^{int}$ (i-niche “91”) were mostly found in the red pulp. And, B cells that were B220$^{int/hi}$, CD79b$^{hi}$ (i-niche “76”) were yet different again and were found at the boundary of the red pulp and the follicles. These observations suggest that the spread of the CD79b-B220 levels as well as of other marker levels on splenic B-cells could be, to a large degree, accounted for by the niche composition around those B-cells – and that the expression levels on these cells might be influenced by (or influences) the cells in their immediate surrounding.

We confirmed that these expression level observations are not quantification artifacts or signal spillover from neighboring cells. For instance, when a CD4 T cell was the index cell, such index cells exhibit a wide variability of CD4 expression across i-niches (expression levels in the CD4 row in **Figure 4B-iii**). Note that even those CD4+ index cells in B cell enriched niches (columns spanning the yellow bar in Figure 4B-iii) show little to no B220 spillover. And, when a B cell served as the index cell in CD4 rich environments (see the niches under the “CD4-rich” label in the Figure 4-i) the compensation algorithm effectively removed all CD4+ expression contributions into the index B cell (bottom row, Figure 4B-ii).

Most i-niches could be readily mapped into one of major anatomical compartments of the spleen (B cell follicle, PALS, marginal zone, or red pulp – per **Figure 3C**). In most cases, any given i-niche resided within a single anatomical compartment (although several i-niches were observed in more than one compartment), and every splenic compartment was populated by many i-niches (**Figure 4B-iv and -v**). The overall utility of the i-niche in determining any given surface marker expression value for an index cell was evaluated by constructing a linear regression model of marker expression using both the cell type identity and the i-niche constituency in a two-featured variable model (the other variable being the cell type identity). Notably, adding the i-niche information as a dependent variable significantly improved the fitness of the model for all markers (**Supplementary Table 6** available online[38]) with highest improvement F-values for CD90, B220, CD21/35, and ERTR7 and the lowest prediction rates for Ly6G, CD5, CD11b, CD5, and TCR$\beta$. Thus, the high variability in B220 expression levels is correlated to (driven by?) the i-niche in which the B cell resides. In other words, B220 expression levels are not independent of tissue locale, and are either driving the constituency of the i-niche partners, or are driven by them. As a counter point, the data also shows that i-niche does not reliably predict CD5 or TCR$\beta$ expression levels which is perhaps not surprising given that the levels of these surface receptors do not vary significantly across the identified niches **Figure 4B-iii**). Therefore—to the extent CD5 or TCR$\beta$ levels do differ across i-niches, the level of expression of these proteins is an autonomously determined state of the cell and is not greatly influenced by the i-niche in which it resides—a corollary is that this cell-autonomous level of CD5 or TCR$\beta$ does not drive the constituency of the i-niche. Of course, we cannot exclude that there might be other markers in T cells that are correlated to i-niche residence.

This result quantitatively demonstrates that for many markers a cell's i-niche (neighbors) determines a significant proportion of variance in marker expression. This analysis showed that many splenic cell types populate a wide variety of i-niches, suggestive of a multiplicity of functional state for any given immune cell type (**Supplementary Figure 6**). Further, tissue locale (i-niches) is a powerful indicator of potential differential function (to the extent tissue locale drives function) and these deterministic changes in surface marker protein expression are surrogate indicators of this locale or function.

## Changes in splenic composition associated with disease progression

It is long observed that inflammatory disease states change how cells traffic in tissues, especially immune organs. Dramatic examples of immune re-organization have been seen in many autoimmune diseases—wherein the tissue targets of autoimmune activity are often infiltrated by a variety of auto-reactive or inflammatory immune cells. One such example wherein the spleen is particularly affected is lupus erythematosus [27] [28]. In this disease, a variety of organs (from skin, to kidney, and other body organs) can be targeted in relapsing-remitting flares. We chose mice with MRL/lpr genotype as a model of autoimmune response because with age they are known to spontaneously develop symptoms closely resembling lupus [29].

A comparable region of spleen was visualized by CODEX for 3 normal BALBc spleens, and 6 spleens from MRL/lpr mice. Image segmentation revealed strong variation in cell counts between the norm and the disease (**Figure 5B**) for most (19 out of 27) of the cell types identified by X-shift clustering. Examples include a dramatic increase in CD71$^+$ erythroblasts (green cells on **Figure 5A** maps), a reduction in numbers of B cells and FDC, and increases in so-called B220$^+$ DN T cells (CD4/CD8 double-negative B220$^+$ T cells). Of the many observed changes, two features were used to broadly classify the

MRL/lpr spleens into early, intermediate, and late disease stages: (1) marginal zone disintegration associated with disease progression evident from a drop in marginal zone macrophage (MZM) counts (see black asterisk on **Figure 5B** and yellow arrow in **Supplementary Figure 7** pointing to the area where CD169 positive (red) rim of MZMs is expected to be observed) and (2) the emergence of atypical B220$^+$ DN T cells (see red asterisk on **Figure 5B**). Early stage disease was represented by three MZM-positive and B220$^+$ DN T cell – low spleens (**Supplementary Figure 7** and **Figure 5A** panels 4, 5, and 6). Two spleens represented the intermediate stage: an example of a spleen with MZM and B220$^+$ DN T cells (**Figure 5A** panel 8), and the other negative for both (**Figure 5A** panel 7). Late stage was represented by a single MZM-negative, B220$^+$ DN T cell – high spleen (**Figure 5A** panel 9).

We undertook a deep quantitative characterization of cell neighborhood and tissue architecture changes associated with the MRL/lpr phenotype. In accordance with the gross morphological similarity of red-white pulp distribution seen in cross-sections (**Figure 3A**), the superficial comparison of the cell-cell log odds ratio heatmaps revealed a general similarity of cell type interaction patterns between the normal and the MRL/lpr spleens (**Supplementary Figure 7**). This is exemplified by a consistent presence of larger cell-adjacency clusters corresponding to red and white pulp and the positive values on the diagonal indicating the persistence of homotypic cell-to-cell interaction across the datasets.

A deeper statistical analysis of changes in cell interactions revealed, in fact, many disease-associated changes in frequency of contacts between different cell types (see **Supplementary Table 5**). Among the changes we observed an increase in interaction between B cells and CD4$^-$/CD8$^+$ cDC in the early MRL/lpr spleen compared to normal, (**Figure 5C** left panel) suggesting an increase in B-cell activation. We also observed a higher interaction frequency of granulocytes with T cells, erythroblasts, and dendritic cells; a higher number of contacts between erythroblasts and various kinds of stromal cells, as well as B220$^+$ DN T cells (**Supplementary Table 5** available online[38]**, Supplementary Figure 8.15, 8.17, 8.20** available online[38]). In the intermediate and late stage MRL/lpr spleens, there was a significant increase in interaction of B220$^+$ DN T cells with CD4$^+$ T cells (**Figure 5C** right panel), CD8$^+$ T cells, erythroblasts, and a variety of other cell types compared with numbers of these interactions in the early MRL/lpr stage (**Supplementary Table 5** available online[38] and **Supplementary Figure 8.33, 8.37, 8.29-39** available online[38]). So, while there was no obvious gross rearrangement of the tissues, there were many homotypic and heterotypic cell-cell associations that are altered. A key question then becomes—can we identify the critical changes that are driving this disruption?

## Disease driven change in cell counts determines the frequency of specific cell–to–cell contacts

What could be the drivers of changes in frequency of pairwise cell-cell contacts? If the kinetics of cell contact, stickiness, and dissociations, follows a rate law— one possibility would be that modulation of specific cell–to-cell interaction potential—or “attraction” (for which the odds ratio score was used as an estimate across this study) is the main driver. In other words, it would be expected that when the affinity of such an interaction goes up, the fraction of interacting cells of a given cell pair would increase. At the same time, even in the absence of change in cell-to-cell affinity, the absolutely number of the cell-cell pairs (defined here as cell pair aggregates, or CPAs) and the number of interacting cell pairs should correlate with the frequencies of interacting cell types (analogous to concentrations in the rate law equation). Importantly, the latter scenario could be as biologically significant as the former. Finally, some of the cell-to-cell contacts may be observed due to low cellular motility of randomly meeting cells. Such interactions would not produce spatially defined sub-splenic CPAs and would have and odds ratio close to 1.

The perturbation introduced to normal splenic composition with MRL/lpr genotype allowed us to identify the mechanisms implicated in transition from norm to diseased like spleen. In short, we found that, for most cell-cell pairs observed, cell “stickiness” or mutual attraction, was not the primary determinant driving the change in absolute counts of interactions observed between the MRL/lpr and the norm. In **Figure 5D** we plot the change in counts of interactions of two cell types (e.g. A:B) between the MRL/lpr and the normal BALBc spleens. Each dot represents a pair of cell types. The value on the Y axis is the difference in the total number of observed interactions between BALBc and MRL/lpr. The X axis shows the difference between log odds ratios of interactions between the same conditions. There was no overall correlation observed ($R^2 = 0.058$). In fact of the 26 top scoring (FDR < 0.05 and change in absolute interaction counts > 150) cell type pairs of this cross comparison only 2 showed corresponding significant (FDR < 0.05) change in odds ratio score. Curiously these two interactions with a modest 1.5 times increase in interaction count and, concomitantly, a ~0.8 increase in log odds ratio score were the ones between the CD4 or CD8 T cells and ERTR7$^+$ stroma (see **Supplementary Table 5**, rows 6 and 7, and **Supplementary Figure 8** available online[38]), Visually they appeared as persistent co-clustering of T cells with ERTR7$^+$ stroma despite the overall drop of T cell numbers in the “early” MRL/lpr samples. Curiously, ERTR7 positive fibers of splenic stroma as well as ERTR7 protein itself were recently shown to be critically involved in T cell trafficking [30], suggesting that this increase in the spatial association could be reflective of the T cell activation.

For the rest 24 of the 26 changing interactions mentioned above at least one of the cells of the pair was scored as significantly (FDR<0.05) changing the frequency across scored conditions (**Supplementary Table 5** last column of the “EarlyMRL vs BALBc control” spreadsheet, available online[38]). We therefore conclude that—at least in the diseased state of early stage MRL/lpr—most of the change in counts of cell-cell interactions are driven simply by increases or decreases in cell type frequencies. In agreement with this, we observed a correlation ($R^2 = 0.288$) between the cell count changes and the interaction changes (**Figure 5E**) while there was no apparent correlation between cell count changes and the log odds ratio changes ($R^2 = 0.058$ – see **Figure 5D**). What suggests the changing frequency of any given cell type in the spleen (driven by unknown processes in MRL) accounts for most of changes of absolute counts of a given A:B pairing that is observed. Of course, there appear to be outliers for each case, where there are apparently odds-fold changes that appear to explain the MRL disease-driven changes in certain cell-cell pairings. While further work is required to determine which of these changes are instrumental to the MRL disease state, the dataset here provides a pipeline to applications in this and other disease areas for therapeutic targeting.

For additional evidence, $\chi^2$ statistics were used to compare the total magnitude of changes in pairwise cell type interaction matrices (total interaction count) versus changes in log-odds ratio matrices (propensity for non-random interaction). The $\chi^2$ deviation (sum of squares of z-score-normalized values) was computed for each disease matrix compared to the control. In every case, the $\chi^2$ values of cell interaction matrices were larger than of the respective log odds ratio matrices of the same biological sample (**Figure 5F**). This suggests that as the cell type frequencies change due to disease progression, the absolute numbers of interactions change dramatically whereas the frequency-normalized likelihoods of cell interactions change to a much smaller extent indicating a great degree of robustness of the ‘design principles’ of the splenic tissue and that many of the more dramatic disease-associated variations occur primarily through the shift in cell numbers.

Thus, in MRL there are significant changes in the pattern of cell-to-cell interactions and, accordingly, in the splenic architecture. Largely these changes are induced by modulation of cell type frequencies associated with disease progression. Later studies will be required to determine which of those changes are “null” for driving the disease state, and which reflect interactions that lead to further deterioration of immune control and splenic architecture.

## Reorganization of cells in disease-associated tissue substructures

We catalogued the cell-cell interaction “connectivity” in a circular correlation diagram. Rarely, if ever, there was any cell type found adjacent to only one other type of cell. The highest degree of connectivity was observed for the most abundant cell types such as B cells in normal spleen and erythroblasts (**Figure 6A**) in early MRL/lpr. This high connectivity in turn led to large effect on i-niches caused by changes in cell numbers associated with progression of disease from norm to autoimmunity. Most dramatic changes in cell frequencies were the increase in erythroblasts in the early MRL/lpr and the emergence of B220$^+$ DN T cells in late MRL/lpr – which were associated with appearance of novel i-niches relative to the normal spleen (spatial localization of B220$^+$DN T cell dominated i-niche 18, erythroblast driven i-niche 29 and and B-cells rich i-niche 96 is indicated on **Figure 6B** and their cell type composition is shown on heatmap on **Figure 6C**). A corollary to this is the question of whether the presence of these cells, and new i-niches dependent on these cells, somehow changed the observable biology of the cells they contact? We found an example of this behavior, where the proximity of CD4 T cells to B220$^+$ DN T leads to CD4 T cell activation in spleens of MRL/lpr mice: **Figure 6C** shows increased levels of CD27 expression in CD4 T cells present in i-niches dominated by B220$^+$ DN T cells (**Figure 6C** red circle).

Other cell types noticeably changed their characteristic distribution and their propensity to engage, or evade, specific cell-to-cell contacts (as estimated by odds ratio score) during disease progression. For example, stromal cells of $CD106^{+}CD16/32^{-}Ly6C^{hi}CD31^{+}$ phenotype were randomly distributed in the red pulp of normal spleens, but were found to aggregate in the areas proximal to the germinal centers of the MRL/lpr white pulp (**Supplementary Figure 3.16** available online[38]). This re-distribution correlated with erythroid proliferation and reduced odds ratio score for the interaction of $CD106^{+}CD16/32^{-}Ly6C^{hi}CD31^{+}$ and erythroblasts in lupus spleens (**Supplementary Table 5** available online[38]).

As noted, the analysis reveals that the development of the autoimmune disease in mice (as exemplified by MRL/lpr lupus) is associated with vast rearrangement of normal spleen architecture, which is likely to cause loss of cell-cell contexts normally hosting the cells crucial for proper splenic function, as well as the observed emergence of novel i-niches that are not found in the normal BALBc spleen. Additionally, certain i-niches were sequestered to specific anatomic compartments of the spleen, which allowed us to use such i-niches as reference points to quantitatively monitor high-order morphological changes. The i-niches that in normal spleen were localized to one distinct compartment (more than 90% of central cells reside within a particular splenic compartment) were used to evaluate the dynamics of splenic cells associated with progression of autoimmune disease (**Figure 7A, middle heatmap**). This analysis confirmed the dissipation of the marginal zone starting from early stages of MRL/lpr and revealed a progressive distortion of PALS. Curiously, depending on whether a i-niche was based on F4/80 macrophages or primarily contained erythroblasts, the red pulp appeared to reorganize in the diseased tissue (**Figure 7A, right heatmap**), pointing to the fact that more than one compartment-specific niche is required to reliably trace the fate of specific anatomic compartments. In many cases the definition of subsets/morphological units constituting the tissue is subjective, yet this study employed niches that were algorithmically defined. Therefore, using niches as markers of morphology can quantitatively monitor the changes of high-order anatomic architecture.

## Automatic definition of disease-specific tissue regions using convolutional neural networks

To automatically isolate the specific local combinations of expression patterns characteristic of the disease state, a fully convolutional neural network was trained to distinguish image patches from normal and MRL/lpr mice. The neural network operated by identifying, in each training image patch, the specific areas that corresponded to the disease state. To avoid the learning of trivial sample-specific staining variation, data were quantile normalized sample-wise and each marker was discretized to four levels. Since disease-specific hallmarks could potentially have multiple scales, the training data for our neural network was extracted at multiple levels of magnification. A simple regularized logistic regression model that considered only average marker expression and did not incorporate spatial information was unable to successfully distinguish patches normal and MRL/lpr spleens, whereas the trained neural network model consistently achieved a 90% precision of classification of image patches during cross-validation.

The neural network highlighted the regions in each multiparameter spleen image that corresponded to the disease state (**Figure 7B**), despite having seen no images from these spleens during training. To investigate the specific features learned by the neural network, the cell-type compositions of the regions identified as diseased versus those regions identified as normal were compared. There was significant enrichment of several cell types in these regions (**Figure 7C**). Although some cell types enriched in diseased regions, for example B220$^+$ DN T cells, were present only in the diseased tissue, the most highly enriched cell type (CD4$^+$/CD8$^-$ cDCs) were present in both the disease state and the healthy state.

To assess the specific contextual changes recognized by the neural network, the local neighborhoods of the CD4$^+$/CD8$^-$ cDCs that the neural network found to be enriched in MRL/lpr regions were analyzed. In these neighborhoods we observed a significant enrichment of other CD4$^+$/CD8$^-$ cDCs, as well as significant depletion of CD106$^+$/CD16/32$^+$/Ly6C$^-$/CD31$^-$ stromal cells (FDR $< 10^{-7}$). This suggests that the neural network had identified an altered context for CD4$^+$/CD8$^-$ cDCs (distant from stromal regions) as a key descriptor for the disease. Thus, the neural network approach described here enabled both automatic classification of samples according to disease state and an automatic identification of high-dimensional regions of interest and corresponding cellular niches.

## Primer dependent panels to extend the multiplexing capacity of CODEX

CODEX operates using an indexed polymerization step that enables precise incorporation of fluorophores into oligonucleotide-Ab conjugates at predetermined cycles. Although consistent performance of a model antigen (CD45) was observed across 15 cycles of CODEX (**Supplementary Figure 1A-F**), a gradual accumulation of polymerization errors during each cycle could potentially result in non-cognate rendering, and thus diminished and/or non-specific signals at later index cycles. In addition, the use of long single-stranded oligonucleotides that would enable indexing beyond 15 rounds might be problematic due to non-specific binding events to tissues under study.

For the polymerization event to initiate, a 3' hydroxyl is required. Thus, we reasoned that dedicated primers (each containing a distinct initiating sequence with a 3' hydroxyl) could be used to activate distinct subpanels of antibodies (**Supplementary Figure 9A**). This would allow design of antibody panels exceeding 30 markers into subpanels, each with a subpanel-specific activation sequence designed 5' to the indexing region. In this design, the antibody attachment linker is terminated with ddC, such that the extension is only possible after a hybridization of a hydroxyl-containing panel-specific activation primer.

The feasibility of such multipanel CODEX design and the robustness of CODEX protocol after many cycles and its independence of staining from the cycle number were tested in a model experiment. A 22-color panel of antibodies (11 cycles) conjugated to a terminated top oligonucleotide, was hybridized with bottom oligonucleotides of 1st ,2nd , and 3rd panels (**Supplementary Figure 9B)**. Thus, every antigen is detected thrice by the same antibody conjugated to oligonucleotides of 3 different panels. Each panel can only be rendered after annealing of a panel-specific activator oligonucleotide. The staining was rendered in 36 cycles (11 detection cycles + 1 blank no-antibody cycle per activator oligo) of CODEX with additional activator oligonucleotide hybridization step between each of the 3 panels. The signal for same antibody detected at different cycles (e.g., 1st, 13th, and 24th) was consistent across the three panels (**Supplementary Figure 9C**). This panel-activator design extends CODEX to a theoretically unlimited multiplexing capacity, bounded only by the speed and resolution of the imaging process itself and the time required for each imaging cycle.

## Discussion

Here the feasibility of polymerase-driven highly multiplexed visualization of antibody binding events to dissociated single cells as well as tissue sections was demonstrated. Critically, CODEX enables co-staining of all antigens simultaneously with the staining iteratively revealed by primer extension cycles wherein no diminution of epitope signal detection was observed. CODEX results were validated by comparison with CyTOF analysis demonstrating that CODEX data qualitatively matches the data generated by conventional flow cytometry while vastly exceeding it in dimensionality. A consistent lossless performance of CODEX for co-detection of up to 66 antigens was observed, and the primer-based extension of the system could enable, theoretically visualization of additional antigens per sample. The CODEX platform can be performed on any three-color fluorescence microscope enabling conversion of regular fluorescence microscope into a tool for multidimensional tissue rendering and cell cytometry. Given the low cost of converting a scope to this platform (a simple custom fluidics device for liquid handling in a customized stage is all that is required) this would enable studies in complex tissues where the availability of the complex instrumentation is limited by logistics and cost factors.

The unique set of algorithms described here successfully identified individual cells in the crowded environment of lymphoid tissue by relying both on the information from nuclear and the membrane staining. An accurate quantification of single-cell expression data was obtained directly from the images by creating a special algorithm for positional spill compensation. This allowed extraction of FACS-like data from tissue imaging and leveraged the automated phenotype mapping framework previously developed for CyTOF and multicolor FACS.

Other groups have reported successful multiplexed detection of up to 100 proteins in tissue sections by cyclic re-staining of a sample coupled to photo or chemical inactivation of fluorophores [1-3].

These approaches require time-intensive re-staining (days to weeks) of the sample for each round of antigen rendering. Moreover, since each round of staining and bleaching or fluorophore inactivation leads to epitope degradation, there is an upper limit on co-detected antigens [3] . CODEX completes a 30-antibody visualization in approximately 3.5 hours. Modifications to the technology that increase the measurements per cycle, reduce the cycling time, faster imaging methods such as light sheet microscopy, or an increased size of the imaging the field of view offer potential opportunities for increasing the depth and speed of the visualization process.

Performance of CODEX on tissue sections was validated in analysis of spleen sections of normal and lupus afflicted mice (MRL/lpr). Much like with conventional flow cytometry, CODEX discerned all major cell types commonly observed in mouse spleen. Moreover, application of a phenotype-mapping algorithm recently developed in our lab [16] and tailored to parsing the multidimensional single-cell data enabled detection of rare cells types (examples are CD4$^{hi}$ MHCII$^{hi}$ (Lti) cells, CD11c(+) B cells) and simultaneously placement in the tissue architecture. By mapping the cell type identity back onto the tissue and counting the cell interactions, the known tissue architecture of the normal spleen was recapitulated using CODEX. The analysis revealed that most splenic cell types were involved in homotypic interactions—which might underscore a novel driving principle of lymphoid tissue architecture. Further, the effect of the local niche on marker expression in multiple splenic cell types was evaluated demonstrating the significant impact of niche on expression and revealing unexpected correlations between levels of surface markers detected when cell types are measured across niches.

Using CODEX, the changes in tissue architecture that occurred in spleen in the wake of autoimmune disease were quantitatively probed. Among hallmarks of MRL/lpr progression were dissipation of marginal zone, disintegration of PALS, invasion of red pulp with erythroblasts and the infiltration of mixed-identity B220$^+$ DN T cells, which, interestingly, localize in a niche in between PALS and the B cell zone and in the marginal zone. A contact-dependent effect of B220$^+$ DN T cell on CD4 T cells reflected in increased levels of activation marker CD27 was observed. An account of statistically significant differences in frequency and strength of pairwise cell type contacts was created. From these observations and their quantitative analysis, we concluded that cellular interaction strength estimated from ratio of observed to expected probability of interaction and frequency of pairwise cell contacts do not correlate. Largely we found cell interaction frequency to be related to cell counts-- therefore while cell intrinsic adhesion properties and cell abundancy are both implicated in shaping the core splenic architecture, it is largely the change in cell numbers that is involved in reorganization of spleen during transition from norm to autoimmunity. This comprehensive, high-parameter description of changes that occur in the splenic architecture of MRL/lpr mice demonstrates the power of neighborhood analysis approaches. Previous analyses had relied on two-color immunohistochemical analyses of sequential sections to achieve multicolor single-cell resolution [31] [32] [28] [27], and as such would not correctly capture true cell-cell adjacencies.

An important principle was observed that is wholly unique to the quality and depth of the data presented herein—that being the context-dependent changes in expression of surface markers on cells. As clearly observed in experiments that drove Figures 4 & 6, cell populations that would otherwise be thought of as 'broadly'-expressing a given marker set (Figure 4D), in fact were composed of multiple cell phenotypes—said phenotypes being determined by the cells participating in their i-niche. In other words, what immunologists previously thought of as a single cell type could be subdivided into more subtle cell subsets that are defined by the neighborhood in which they reside. We leave open the question of whether the cells with different properties are attracted to a set of neighbor cells, or a given expression level of markers attracts the neighbors, or some dialectics thereof. What is clear, however, is that there are more subtle phenotypes in tissues that previously assumed, and that the kinds of deep cellular imaging phenotyping presented here is only the beginning of what is possible in the future as the technology develops.

Recent advances in genomics suggest that despite vastness of a genetic repertoire a limited number of cellular states have unique gene expression patterns. These countable patterns are reflected in existence of surface marker phenotypes recognizable as cell types. It is therefore reasonable to suggest that cell-to-cell interactions should be countable as well. The data collected in this study lay the foundation for a pan-cellular reference database defining cellular types not only by identities of proteins expressed but also by capacities for specific cell-to-cell interactions. Such a deep characterization was performed here for normal and diseased tissue from a perspective of cellular arrangements. CODEX thus opens new horizons for application of multidimensional cytometry in tissue diagnostics currently performed in a very low-dimensional space (1-3 markers). Further, disease classification was accurately performed by a neural network operating on multidimensional marker expression data, even after training on just one sample. We present here, for the research community, a large (~700,000 cells) dataset encompassing segmentation, quantification, and, most uniquely, spatial data from normal and disease-afflicted spleens (http://welikesharingdata.blob.core.windows.net/forshare/index.html and **Supplementary Tables 2 and 3** available online[38]) for further development of computational algorithms for tissue cytometry and digital pathology.

Automated identification and deep phenotypic profiling of cellular microenvironments in tissues is an unmet and important need. Results from high parameter mass cytometry have demonstrated the potential for biologic and clinical insights when such parameterization is matched with a focused computational capability [8,33,34]. Approaches such as CODEX should facilitate acceleration of multidimensional imaging of numerous tissues, could provide data necessary to infer biological and clinical correlates of tissue micro-architecture, and will be important for supporting biomedical inquiries into pathologies such as cancer immunotherapy and inflammatory disease states of tissues.

## Acknowledgements:

We thank A. Trejo and A. Jager for technical assistance. We are grateful to Peter Jackson for advice and access to the microscope used for this work. This work was further supported by grants to GPN: U19 AI057229, 1U19AI100627, Department of Defense (CDMRP), Northrop-Grumman Corporation, R01CA184968, 1R33CA183654-01, R33CA183692, 1R01GM10983601, 1R21CA183660, 1R01NS08953301, OPP1113682, 5UH2AR067676, 1R01CA19665701, R01HL120724. GPN is supported by the Rachford & Carlotta A. Harris Endowed Chair.

## Materials and Methods

### Animals

9 months old female MRL/lpr (chosen to represent lupus disease at a pronounced splenomegaly stage) and age/sex matched control BALBc mice purchased from Jackson Laboratory were used for the study. All animal studies were done in compliance with ethical regulations and procedures set in the Stanford Administrative Panel on Laboratory Animal Care Protocol 15986. In coherence with the primarily technical purpose of the study no animal cohort randomization or investigator blinding to group allocation was performed.

### Antibodies

CD27(LG.3A10), CD11c(N418), CD106(429), CD19(1D3), Ly6G(1A8), CD169(MOMA-1), CD16/32(2.4g2), CD3(17A2), CD90(Thy-1/G7), CD8a(53-6.7), Ly6c(HK1.4), F4/80(T45-2342), CD11b(m1/70), Ter119(ter119), TCR(h57-597), IgD(11-26c.2a), CD79b(HM79-12), CD5(53-7.3), CD31(mec13.3), CD71(C2F2), IgM(R6-60.2), CD4(rm4-5), ERTR7(ER-TR7), B220(RA3-6B2), CD35(8C12), MHCII(M5/114.15.2), CD44(im7), CD21/35(8D9), cd43(S7), cd8(53-6.7), CD45(30-F11)

### Oligonucleotide sequences

Single base extension during CODEX can be achieved by either a “missing base” approach (**Figure 1A**) or a “reversible terminator” method (see **Supplementary Figure 10**). In the case of the “missing base” approach, which was chosen for the experiments outlined in this paper, the top strand of the double-stranded oligonucleotide is covalently bound to the capture agent (antibody or RNA probe) and the bottom strand is annealed through hybridization to the top strand. All capture agents contain the same top strand and different bottom strands. The sequence of the bottom strands contains a common region that hybridizes to the top strand as well as a variable sequence region that serves as the indexing region. As shown in **Figure** 1 A, the overhanging 5’ end of the lower strand of the double-stranded oligonucleotide tag (which forms the overhang) is of the general formula 3’-YN$_1$/N$_2$-5’ followed by a short stretch of random composition on the 5’ end to increase the overall polymerase residence on the DNA duplex. N$_1$, N$_2$, are nucleotides (A and G in our case) allow for incorporation of labelled dNTPs (dU-ss-Cy5 and dC-ss-Cy3). Y is a nucleotide sequence of length n (n is 0, 1 or more) composed of alternating random-length stretches of “indexing” bases N$_3$ (dC) and N$_4$ (dT) that are complementary to unlabeled nucleotides. Stretches of random length (e.g. CCCTCCTTTCTT) rather than single base design (e.g. CTCTCTCTCT) were introduced to prevent misalignment of upper and lower strands of double-stranded oligonucleotide tags. All oligonucleotide sequences can be found in **Supplementary** Table 3.

## Antibody conjugation protocol

For full list of antibody clones and vendors see **Supplementary Table 1**. 100ug of each antibody was partially reduced by 30min incubation at room temperature with TCEP (final concentration 5mM) in PBS. Antibodies were purified by buffer exchange on BioGel P-30 spin-columns saturated with conjugation buffer (2mMTris, 150mM NaCl, 1mM EDTA, pH 7.2). Oligonucleotides bearing protected maleimide group were ordered from Trilink Inc. The maleimide group was de-protected/activated by Adler reaction (4h at 90C in toluene). After activation traces of toluene were removed from oligonucleotides by five washes in absolute ethanol. Activated oligonucleotides were resuspended in conjugation buffer and mixed with reduced antibodies at a molar ratio of 25:1. NaCl was added to conjugation reaction to final concentration of 1M. The conjugation reaction proceeded for 1h or longer. To remove unbound oligonucleotide with unquenched maleimide group the conjugated antibodies were filtered 4 times on molecular weight cutoff filters (Amicon 50 KDa). Final wash and storage were performed in phosphate buffer with 0.5M sodium chloride and 0.1% Tween-20. We found that about half of starting amount of the antibody was lost during conjugation with remaining half bearing with 2 to 12 oligos per 150kDa antibody molecular dimer.

To assemble DNA duplex tag 0.2ug of conjugated antibody was mixed with 100pmoles of bottom strand oligonucleotide in phosphate buffer with 0.6M Sodium Chloride and incubated for 30min at 40C.

## CODEX staining protocol

Mouse spleen and bone marrow cells were prepared according to standard procedure and fixed in 2% formaldehyde for 10min at room temperature. Following fixation, cells were spun and either stored frozen at -80 in PBS with 5% DMSO or permeabilized by incubation in ice-cold methanol for 10 min. and further stored at -80 °C in methanol.

Before staining stored cells were washed with SME (0.5%BSA in PBS, 5mMEDTA) once and blocked for 30min at room temperature in staining buffer (0.6M NaCl, 0.5%BSA, 50ug/ml rat IgG, 200ug/ml ssDNA, 5mMEDTA, 3nmoles per ml of blocking oligonucleotide TTTTccctctcctcttcctttCcTCT ddC in phosphate buffer pH 7.4).

In case cryo-sections were used – tissue sections were picked by warm coverslips and immediately placed into dry ice without allowing the section to dry. Coverslips with sections were dipped for 30 sec into acetone pre-chilled to dry-ice temperature, fully dried for 10 min at RT and transferred to SME with 1.6% formaldehyde for 20min. After that the fixed sections were washed twice in SME and further blocked in staining buffer for 30min. Suspension cells or frozen sections were stained in staining buffer for 2-3h at room temperature with a mixture of conjugated antibodies taken at 0.2ug of each antibody per 100ul of solution. After staining cells were washed twice with SME05 (SM supplemented with NaCl up to 0.65M final concentration). If methanol stored suspension cells were used – washed cells were allowed to adhere to poly-L-lysine coated coverslips and further fixed to coverslip surface by 20min incubation with 5mM BS3 cross linker in PBS. If DMSO stored suspension cells were used – following last wash, cells were spun and further fixed in ice cold methanol fro 10min. Following methanol fixation/permeabilization cells were washed once with SM05, allowed to adhere to coverslip surface and further fixed to coverslip surface by 20min incubation with 5mM BS3 cross linker in PBS. If frozen sections were used – following staining sections were washed twice by SM05, dipped into ice cold acetone for 30sec, dried at RT for 10min, and further postfixed by 20min incubation with 5mM BS3 cross linker in PBS. Following staining procedure and converting into planar form (in case of suspension cells) all kinds of samples were subjected to similar CODEX rendering protocol.

### CODEX rendering protocol

Following fixation coverslips with cells or frozen sections were washed twice with buffer 405(10mM Tris 7.5, 650mM NaCl, 0.1% Triton x100). Cells or sections were treated with 50mM TCEP in buffer 405 for 4min, washed 3 times with buffer 405 and further blocked with 100mM iodoacetamide for 1h in buffer 405 adjusted to pH8 with NaOH. Following blocking if necessary cells were stained for 40min with phalloidin 2ug/ml in buffer 4(10mM Tris 7.5, 10mM MgCl2, 150mM NaCl, 0.1% Triton x100).

Staining was rendered by cycles. In odd cycles (1,3,5…) - cells were incubated for 2min in G-mix (150nM dG, 150nM dUssCy5, 150nM dCssCy3, 25ul NEB exo- Klenow, 3ug phalloidin-FITC per ml in buffer 4); washed 3 times with 405 (buffet 4 without MgCl and supplemented with NaCl up to final 0.65M); photographed; incubated 2min in 50mM TCEP in buffer 405; washed twice with 405; photographed; incubated for 1min in freshly made 100mM iodoacetamide in buffer 405; washed three times with buffer 4. In even cycles (2,4,6…) - cells were incubated for 2min in A- mix (150nM dATP, 150nM dUTPssCy5, 150nM dCTPssCy3, 25ul NEB exo- Klenow, 3ug phalloidin-FITC per ml in buffer 4); washed 3times with 405 (buffet 4 without MgCl and supplemented with NaCl up to final 0.65M); photographed; incubated 4min in 50mM TCEP in buffer 405; washed twice with 405; photographed; incubated for 1min in freshly made 100mM iodoacetamide in buffer 405; washed three times with buffer 4. Fluorescent nucleotides dU-ss-Cy5 and dU-ss-Cy3 were custom synthesized by Jena Bioscience.

Microfluidic setup custom manufactured by M.H. was used to automate CODEX solution exchange and image acquisition.

### Primer dependent panels

Rendering of antibodies with spacers followed the same procedure as the standard CODEX protocol with the exception of the following differences. Before proceeding to rendering next spacer dependent panel, the stained cells were incubated with a spacer oligonucleotide (1µM final concentration in buffer 405) at room temperature for 10 minutes. Cells were washed 4X with buffer 4 and rendering proceeded as usual. To initiate each additional spacer set, the spacer incubation step was repeated using corresponding spacer samples.

## Imaging

Images were collected using a Keyence BZ-X710 fluorescent microscope configured with 3 fluorescent channels (FITC, Cy3, Cy5) and equipped with Nikon PlanFluor 40x NA 1.3 oil immersion lens. Imaging and washes were iteratively performed automatically using a specially developed fluidics setup. Images were subject to deconvolution using Microvolution software (www.microvolution.com).

## Data analysis

For each imaging field analyzed by CODEX multidimensional staining multi-color z-stacks collected during individual cycles were aligned against reference channel (CD45) by 3D drift compensation[35]. If necessary individual fields covering large tiled areas were "stitched" using dedicated ImageJ plugin[36]. For 22 color experiment on dissociated cells attached to coverslip (**Figure** 2) images corresponding to the best focal plane of vertical image stacks collected at each acquisition step of CODEX were chosen for quantification.

Image stacks were subject to a purposefully developed image segmentation algorithm that creates 3D voxel regions around nuclei using a combination of low-pass FFT filtering and a watershed algorithm. Per-cell intensities were quantified by integrating the intensity of each channel within a given cell region and divided by the region size in voxels. The cell-to-cell signal spill coefficients were estimated based on the fraction of shared boundary between each pair of cell regions, resulting in a banded matrix (most cells don't have any shared boundaries). To compensate the cell-to-cell spill, the raw intensity vector was multiplied by the inverse spill matrix.

Compensated intensities of cells from normal and MRL/lpr spleen were pooled together for clustering, resulting in a dataset of 707466 cells (**Supplementary Table 2**). This dataset was subject to clustering using X-shift [16] and K was automatically set to 60.

To define for each cell the neighbors of the first (immediate) tier of proximity Delaunay graph was computed for the dataset (**Supplementary Table 3**). The odds ratio of co-occurrence of cell type A and cell type B was estimated as the observed frequency of co-occurrence (mean of the beta-distribution, with parameter alpha = number of edges connecting cell types A and B and parameter beta = total number of edges minus number of edges connecting A-B) divided by the theoretical frequency of co-occurrence (total frequency of edges incident to type A multiplied by the total frequency of edges incident to type B) see **Supplementary Table 5**. The odds ratios are represented in heatmaps on **Figure 3G** , with a range of values from less than 1 to more than 1 meaning that two cell types are, respectively, less or more likely to co-occur than expected by chance. The significance of the difference from zero was tested using binomial distribution (probability of getting an observed number of interactions between A and B (successes) amongst the total number of registered interactions (number of trials) given the theoretical probability of A-B interaction (probability of success)).

The significance of change of interaction frequencies or log-odds ratios were computed between BALB/c and Stage 1 (early) MRL using pairwise T-test. However, the same procedure could not be applied to testing BALB/c versus MRL/lpr Stages 2 and 3 because of high sample-specific variation in those more advanced disease stages. Therefore we scored computed the deviation of those Stage2/3 values from BALB/c using $\chi^2$ statistics because it does not require Stage 2/3 samples to have a common mean.

The P-values were subject to FDR correction using Benjamini–Hochberg procedure. Interactions that were considered significant for FDR q-value < 0.05 or > 0. (**Supplementary Table 5**).

In order to estimate the overall deviation of either interaction frequency matrices or log-odds ratio matrices, the matrices were subject to z-transformation based on the mean and the SDs of the BALB/c samples, and then $\chi$ statistics was computed as square root of the sum of squares of all elements of the z-score transformed matrices (**Figure 5 F**).

## Neural network training and data analysis[37]

Preprocessing:
Image stacks were maximum-intensity projected following deconvolution. Data was quantile normalized to 4 levels (0, 0.25, 0.5 and 0.75 quantiles). A baseline model was able to distinguish models without this discretization and normalization, suggesting strain-specific differences in antibody staining intensity.

Training and cross validation split:
Four spleen samples (two BALB/c and two MRL/lpr) were chosen as training samples. The remaining five spleens tissue samples (one BALB/c and four MRL/lpr) were used for testing the trained model. For cross-validation, different combinations of spleens were allocated to training and test sets. During training, 224x224 images were randomly extracted from the training tissue samples, at 1x, 0.5x and 2x zoom. At 1x zoom, there would be 6804 non-overlapping image patches in the training dataset. The trained models were tested on 4500 patches, at 1x zoom. Hyperparameters were manually tuned on 500 randomly selected images from the testing spleens. The Adam optimizer was used for training with an initial learning rate of 0.0001.

Baseline model:
A logistic regression model was trained by averaging marker intensities across the image. L2 regularization was used for weights.

Neural network architecture
A fully convolutional network architecture was used, with the following layers. To generate a prediction for an entire image patch, a global max-pooling layer was used.

1. Conv3 60
2. Conv3 120
3. Conv3 64
4. Batch Norm
5. Conv3, 64
6. Max pooling 2x2

7. Conv3, 128
8. Conv3, 128
9. Max pooling 2x2
10. Conv3, 256,
11. Conv3,256
12. Conv3,256
13. Max pooling 2x2
14. Conv3,512
15. Conv3,512
16. Conv3,512
17. Conv1,256
18. Conv1,64
19. Conv1,1
20. Global max pooling
21. Sigmoid

Weights for layers 5-16 were initialized from the VGG-16 pretrained model. The model was trained with cross-entropy loss.

Regularization:

L2 regularisation (0.1) was used for network weights. L1 regularization was applied to the feature map output after layer 19 to encourage sparse activations

Whole sample activations for test set:

Since the network was fully convolutional, it could be applied to images of any dimension. The network was applied to entire fields of view individually. The activation maps were obtained as the output after layer 21.

Aligning cell type information:

Each cells were assigned the MRL/lpr score of the corresponding pixel in the image.

Enrichment and neighbourhood analysis:

FDR controlled chi-squared tests of proportions were carried out to determine enrichment of specific cell types in the top 10% of cells by MRL/lpr score. For neighbourhood analysis of dendritic cells, the composition of the neighborhoods (cell centers within 30 pixels) of the top 300 cells (by MRL/lpr score) were compared to the composition of the neighborhoods of the bottom 300 cells. Only cells with positive neural network assigned MRL/lpr score, in MRL/lpr regions, were considered for this analysis.

## References

1. Schubert, W., *et al.* Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. *Nat Biotechnol* **24**, 1270-1278 (2006).
2. Schubert, W., Gieseler, A., Krusche, A., Serocka, P. & Hillert, R. Next-generation biomarkers based on 100-parameter functional super-resolution microscopy TIS. *New biotechnology* **29**, 599-610 (2012).
3. Gerdes, M.J., *et al.* Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc Natl Acad Sci U S A* **110**, 11982-11987 (2013).
4. Angelo, M., *et al.* Multiplexed ion beam imaging of human breast tumors. *Nat Med* **20**, 436-442 (2014).
5. Giesen, C., *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Methods* **11**, 417-422 (2014).
6. Chattopadhyay, P.K. & Roederer, M. Cytometry: today's technology and tomorrow's horizons. *Methods* **57**, 251-258 (2012).
7. Krishnaswamy, S., *et al.* Systems biology. Conditional density-based analysis of T cell signaling in single-cell data. *Science* **346**, 1250689 (2014).
8. Spitzer, M.H., *et al.* IMMUNOLOGY. An interactive reference framework for modeling a dynamic immune system. *Science* **349**, 1259425 (2015).
9. Bendall, S.C., *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687-696 (2011).
10. Bendall, S.C., *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157**, 714-725 (2014).
11. Levine, J.H., *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**, 184-197 (2015).
12. Zunder, E.R., Lujan, E., Goltsev, Y., Wernig, M. & Nolan, G.P. A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. *Cell Stem Cell* **16**, 323-337 (2015).
13. P, F.A., *et al.* Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nature Methods* (in press).
14. Andrew, H., *et al.* Coordinate actions of innate immune responses oppose those of the adaptive immune system during Salmonella infection of mice. *Science Signaling* (in press).
15. Gerner, M.Y., Kastenmuller, W., Ifrim, I., Kabat, J. & Germain, R.N. Histo-cytometry: a method for highly multiplex quantitative tissue imaging analysis applied to dendritic cell subset microanatomy in lymph nodes. *Immunity* **37**, 364-376 (2012).

16. Samusik, N., Good, Z., Spitzer, M.H., Davis, K.L. & Nolan, G.P. Automated mapping of phenotype space with single-cell data. *Nat Methods* **13**, 493-496 (2016).

17. Robinette, M.L., *et al.* Transcriptional programs define molecular characteristics of innate lymphoid cell classes and subsets. *Nat Immunol* **16**, 306-317 (2015).

18. Cupedo, T., Jansen, W., Kraal, G. & Mebius, R.E. Induction of secondary and tertiary lymphoid structures in the skin. *Immunity* **21**, 655-667 (2004).

19. Meier, D., *et al.* Ectopic lymphoid-organ development occurs through interleukin 7-mediated enhanced survival of lymphoid-tissue-inducer cells. *Immunity* **26**, 643-654 (2007).

20. Rubtsova, K., *et al.* B cells expressing the transcription factor T-bet drive lupus-like autoimmunity. *J Clin Invest* **127**, 1392-1404 (2017).

21. Rubtsov, A.V., *et al.* Toll-like receptor 7 (TLR7)-driven accumulation of a novel CD11c(+) B-cell population is important for the development of autoimmunity. *Blood* **118**, 1305-1315 (2011).

22. Chasis, J.A. & Mohandas, N. Erythroblastic islands: niches for erythropoiesis. *Blood* **112**, 470-478 (2008).

23. Socolovsky, M., *et al.* Negative autoregulation by FAS mediates robust fetal erythropoiesis. *PLoS Biol* **5**, e252 (2007).

24. Palm, A.K., Friedrich, H.C. & Kleinau, S. Nodal marginal zone B cells in mice: a novel subset with dormant self-reactivity. *Scientific reports* **6**, 27687 (2016).

25. Hendriks, J., *et al.* CD27 is required for generation and long-term maintenance of T cell immunity. *Nat Immunol* **1**, 433-440 (2000).

26. Haeryfar, S.M. & Hoskin, D.W. Thy-1: more than a mouse pan-T cell marker. *J Immunol* **173**, 3581-3588 (2004).

27. Lieberum, B. & Hartmann, K.U. Successive changes of the cellular composition in lymphoid organs of MRL-Mp/lpr-lpr mice during the development of lymphoproliferative disease as investigated in cryosections. *Clinical immunology and immunopathology* **46**, 421-431 (1988).

28. Jacobson, B.A., *et al.* Anatomy of autoantibody production: dominant localization of antibody-producing cells to T cell zones in Fas-deficient mice. *Immunity* **3**, 509-519 (1995).

29. Kanauchi, H., Furukawa, F. & Imamura, S. Characterization of cutaneous infiltrates in MRL/lpr mice monitored from onset to the full development of lupus erythematosus-like skin lesions. *The Journal of investigative dermatology* **96**, 478-483 (1991).

30. Burrell, B.E., *et al.* Lymph Node Stromal Fiber ER-TR7 Modulates CD4+ T Cell Lymph Node Trafficking and Transplant Tolerance. *Transplantation* **99**, 1119-1125 (2015).

31. Mandik-Nayak, L., *et al.* MRL-lpr/lpr mice exhibit a defect in maintaining developmental arrest and follicular exclusion of anti-double-stranded DNA B cells. *J Exp Med* **189**, 1799-1814 (1999).

32. Morse, H.C., 3rd*, et al.* Abnormalities induced by the mutant gene Ipr: expansion of a unique lymphocyte subset. *J Immunol* **129**, 2612-2615 (1982).

33. Gaudilliere, B.*, et al.* Clinical recovery from surgery correlates with single-cell immune signatures. *Science translational medicine* **6**, 255ra131 (2014).

34. Bruggner, R.V., Bodenmiller, B., Dill, D.L., Tibshirani, R.J. & Nolan, G.P. Automated identification of stratifying signatures in cellular subpopulations. *Proc Natl Acad Sci U S A* **111**, E2770-2777 (2014).

35. Parslow, A., Cardona, A. & Bryson-Richardson, R.J. Sample drift correction following 4D confocal time-lapse imaging. *Journal of visualized experiments : JoVE* (2014).

36. Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25**, 1463-1465 (2009).

37. Oquab, Maxime & Bottou, Leon & Laptev, Ivan & Sivic, Josef. (2015). Is object localization for free? - Weakly-supervised learning with convolutional neural networks. 685-694. 10.1109/CVPR.2015.7298668.

38. Supplementary online materials can be found at http://welikesharingdata.blob.core.windows.net/forshare/index.html

## Figure Legends

### Figure 1 – Sequential primer extension on samples stained with DNA barcoded antibodies enables unlimited level of multiplexing.

**(A)** CODEX diagram. Cells are first simultaneously pre-stained with a mixture of all DNA tagged antibodies. At each rendering cycle the cells are exposed to a nucleotide mix that contains one of two non-fluorescent index nucleotides and two fluorescent nucleotides. The index nucleotides fills in the first index position across ALL antibodies bound to the cells. However, the DNA tags are designed such that only the first two antibodies are capable of being labelled with one of the two fluorescent dNTPs – and only if the index nucleotide was previously incorporated. Those two antibodies are then co-imaged by standard fluorescence microscopy. The fluorophores, which were incorporated in the first index cycle are cleaved away, the cells are washed, and the slide is ready for the next cycle. At the end of the multicycle rendering protocol each pair of antibodies is visualized at a known, pre-defined cycle of the sequencing protocol, and the multiparameter image can be reconstructed.

**(B, C)** Mouse spleen cells were fixed and co-stained with conventional TcRβ Ax488 antibody and CD4 antibody conjugated to CODEX oligonucleotide duplex as in first round of (A). After staining cells were either incubated in extension buffer with dG and dUTP-Cy5 without (B) or with (C) Klenow exo- polymerase. Note that TcRβ−positive T cells in (B) and (C) are indicated by Ax-488 staining.

Dependent upon the addition of Klenow, TcRβ−positive CD4 positive T cells are seen as a Cy5 positive subset of TcRβ−positive T cells in (C).

**(D)** Spleen cryo-section stained with B cell specific B220-APC (red) and T cell specific TCR-FITC (green) show mutually exclusive staining pattern in the marginal area between B cell follicle and the white pulp.

**(E)** Spleen cryo-section stained with CODEX DNA tagged B220 (red) and CODEX DNA tagged TCR- (green) shows staining similar to the one observed with regular antibodies in (D).

**(F)** Spleen sections were co-stained with regular B220-FITC and two antibodies (ERTR7 and CD169) tagged with cycel1 CODEX DNA duplexes. Localization of marginal zone CD169 positive macrophages in the area between the ERTR7 positive splenic conduit of the white pulp and the B220 positive follicular B cells (D) as reported previously has been observed.

**Figure 2 – Accuracy of surface marker quantitation by CODEX.**

**(A)** Microscopic image of mouse splenocytes stained with a 24-color antibody panel, showing one cycle of CODEX antibody rendering. Cell contours show the outlines produced by the cell segmentation algorithm

**(B)** Comparison of single-cell expression data derived from dissociated mouse splenocytes on an identical 24-color panel using CODEX and CyTOF.

**(C)** Example segmentation in a mouse spleen section based on combining nuclear and membrane (CD45) channel.

**(D)** Graphical explanation of the algorithm for compensating the spillover between neighboring cells using a cell-by-cell compensation matrix.

**(E)** The effect of the compensation algorithm on the estimated single-cell intensity of CD8 and CD4 markers in mouse spleen.

**Figure 3 – CODEX analysis of mouse spleen cryosections co-stained for 28 antigens.**

**(A)** Three collated images on the left correspond to: the legend of antibody renderings per cycle; gross morphology photograph of MRL/lpr (left) and normal (right) spleen embedded in O.C.T. block prior to sectioning. Green color corresponds to antibodies rendered by extension with dUTP-Cy5, red – dCTP-Cy3 On the right collage of the CODEX multicycle data for normal spleen (BALBc-2) and early MRL/lpr spleen (MRL/lpr -4). All images are derived from a single scan with a 40x oil objective of an area covered by 63 tiled fields.

**(B)** Schematic diagram of major known splenic anatomical subdivisions drawn based on cell distribution in BALBc-1 replicate.

**(C)** An exemplary profile of Vortex cluster (B-cells) used for manual matching of clusters to known cell types.

**(D)** Minimal spanning tree (MSP) built for all clusters identified by Vortex analysis. On the left middle and right panels the MSPs are colored by expression levels of B220, TCR and CD71 accordingly to indicate location of B-cells T-cells and erythroblasts on the tree.

**(E)** Vienne diagram showing for several major cell types their fraction of total cells as identified by CODEX analysis of splenic tissue and CYTOF analysis of isolated BALBc splenocytes

**(F)** Post-segmentation derived diagram of identified objects (cells) colored according to cell types in BALBc-1 replicate. Full size diagrams are available for every tissue analyzed in this study are available online at https://welikesharingdata.blob.core.windows.net/forshare/index.html

**(G)** Average cell type to cell type interaction strength heatmap for BALBc samples. Color from blue (<0) to white (around 0) to red (>0) indicates log of odds ratio of interaction (ratio of observed frequency versus expected frequency of interaction). The rows and columns are in the same order (annotation on the right). Black outlines indicate two largely exclusive mega-clusters of cross-interacting cell types loosely matching the cell types populating the red and the white pulp.

**Figure 4 – Unbiased identification of i-niches in multidimensional CODEX data.**

**(A)** On the left – diagram explaining the terminology used for defining i-niche (a ring of first tier neighbors for central cell). On the right – Delaunay triangulation graph used for identification of first tier of neighbors for every cell.

**(B-i)** Heatmap depicting frequency of cell types in 100 types of i-niches identified by K-means (K=100) clustering of all index cells in the dataset (each cell is an index cell for its i-niche) based on frequency of different cell types in the first tier of neighbors. **(B-ii and -iii)** Two heatmaps from top to bottom show average expression of selected surface markers measured in a central cell across 100 i-niches when central cell is B-cells or CD4 T-cell accordingly. Two orange rectangles over top heatmap indicates position of i-niches with high CD35 (containing FDCs and marginal zone macrophages). Cyan rectangle shows location of family of i-niches with high content of F4/80 macrophages and low B220 and CD19 in central B – cell. Purple rectangle indicates family of i-niches enriched with ERTR-7 positive stroma. Below top heatmap location of selected i-niches shown in (E) are indicated. Over bottom heatmap yellow rectangle indicated family of i-niches dominated with dominating presence of B-cells. Two green rectangles indicate family of niches with high levels of CD90 and CD27 in the index CD4 T cells. Grey rectangle indicates family of i-niches enriched with ERTR-7 positive stroma. **(B-iv)** Abundance of 100 i-niches in normal spleen (top bar graph) and **(B-v)** relative distribution of i-niches between splenic histological subdivisions (PALS, red pulp, marginal zone and B-zone) shown as a heatmap. To illustrate a variety of tissue distribution pattern by i-niches an overlay of selected i-niches over a schematic of normal spleen (BALBc-1) is shown.

**(C)** An example of marginal zone and follicular (B-zone) B cells defined by residence in distinct i-niches (e.g. marginal zone i-niche includes a marginal zone macrophage marked by letter H and green color). Positions of B-cells in each i-niche is marked with red circles over the schematic of BALBc spleen.

**(D)** Top right shows a biaxial plot of flow data for CD79b and B220 measured in isolated splenocytes. Top left shows levels of CD79b and B220 in central B-cells as measured across all 100 i-niches. To illustrate i-niche dependent variability of surface marker expression – images of central cells (marked with red cross) with levels of surface marker indicated in pseudocolor palette are shown for selected exemplary i-niches in the bottom panels.

**Figure 5 – Differential effect of disease over i-niche presence across dataset.**

- **(A)** Post-segmentation diagrams of all objects (cells) colored according to cell types (see color map in Figure 3F) for all normal and MRL/lpr tissue sections imaged in the study. . Full size diagrams are available for every tissue analyzed in this study are available online at https://welikesharingdata.blob.core.windows.net/forshare/index.html
- **(B)** Cell counts across dataset for manually annotated Vortex clusters (cell types) across progression from normal to afflicted spleen. Cell types were split into four types according to the dynamics of counts across dataset as represented by average cell count line graphs on the right
- **(C)** Two examples of change in cell-to-cell interaction frequency during disease progression – between the B cells and dendritic cells in normal and early MRL/lpr spleen and between B220+ DN T cells and CD4 T cells during progression from early MRL/lpr to intermediate.
- **(D)** Co-distribution of odds ratio log-fold [log(odds ratio in early MRL/lpr)- log(odds ratio in BALBc)] on X axis and change in counts of interactions for early MRL/lpr versus control (BALBc) comparisons (on Y axis).
- **(E)** Co-distribution of cumulative cell frequency change [celltype1 freq. change + celltype1 freq. change] on X axis and change in counts of interactions for early MRL/lpr versus control (BALBc) comparisons (on Y axis).
- **(F)** Bar graph showing Chi square values across conditions computed for odds ratio and direct interaction counts.

**Figure 6 – Differential effect of disease over i-niche presence across dataset.**

- **(A)** Cell interaction networks built for BALBc early MRL/lpr and late MRL/lpr based on number of contacts observed between two cell types (only connections with more then 150 interactions per sample are shown on the diagrams). Thickness of connection correlates with number of contacts size of the node indicates number of cells per condition.

**(B)** Evolution of i-niche abundance across dataset. Selected three i-niches (marked above heatmap in (C) depicting i-niche composition) differentially represented across dataset (changing between norm and disease) are shown. Yellow circles overlaid over blank rectangle corresponding to imaged area indicate location of i-niche.

**(C)** Top heatmap shows frequencies of B220$^+$ DN T cells, erythroblasts and B-cells in the i-niche rings. Line above top heatmap indicates the composition of i-niches 18, 29, and 96 described in (B). Middle heatmap indicates expression of selected markers when the i-niche central cell is an erythroblast – primarily to show that CD27 is not expressed on erythroblasts in the vicinity of B220$^+$ DN T cells. Bottom heatmap indicates expression of selected markers when the i-niche central cell is a CD4 T cell. Red oval outline pinpoints i-niches with elevated CD27. Note that these i-niches as indicated by top heatmap have B220$^+$ DN T cells as a prevailing component. Lower panels show examples of central cells in i-niches marked under the lower heatmap. i-niche 50 is an example of i-niche without B220$^+$ DN T cells. Central cell does not express high CD27. i-niches 42 and 44 have high frequency of B220$^+$ DN T cells and accordingly central cells express high CD27.

**Figure 7 – Differential effect of disease over i-niche presence across dataset.**

**(A)** Selected i-niches (green heatmap shows i-niche composition) were chosen based on high (>90%) presence per single histological subdivision (blue heatmap). Abundance of these i-niches (brown heatmap) was used to judge the preservation or decay of a histological splenic subdivision corresponding to selected i-niches.

**(B)** Red color over blue rectangle indicates neural network predicted regions of interest (MRL specific regions) in entire spleen images. From top left, clockwise: BALBc #2, MRL/lpr #5, MRL/lpr #7, MRL/lpr #8.

**(C)** Cell types enriched (FDR < 0.1) in neural network predicted regions in MRL spleens (in red in A).

## Supplementary Figure Legends

(supplementary figures and tables not included in the main text can be found at at

http://welikesharingdata.blob.core.windows.net/forshare/index.html)

**Supplementary Figure 1. Robustness of CODEX rendering.**

**(A)** Experimental scheme for mimicking the tissue with 30 distinct cell types **(B)** Montage of a fragment of imaging field of the 15 cycles of CODEX used to render the mix of 30 barcoded spleens – first cycle top left last cycle bottom right. (**C**). An example of image quantification approach used for parsing the CODEX data. Images corresponding to the best focal plane of vertical image stacks collected at each acquisition step of CODEX were chosen for quantification. Images were segmented by Cell Profiler, tertiary objects encompassing the cell membrane (left panel) and an external to cell region immediately bordering the cell membrane were chosen for further quantitation. To account for local background the value corresponding to difference between the mean intensity value inside "cell membrane" object and the mean intensity inside the external ring object was chosen as a representation of the intensity of the antibody signal. **(D)** Time-lapse profile of median intensity per cell membrane for individual cells marked by white arrows on (B). **(E)** Heatmap (cycles in columns, cells in rows) showing mean fluorescence per cell membrane for each cell per in each of the 15 CODEX cycles performed on cells of 30 barcoded spleens. Odd columns correspond to imaging after labeled base incorporation. Even columns correspond to imaging after inactivation of staining by TCEP. (**F**) Average intensity of CD45 antigen expression in 15 distinct Cy5-positive populations sequentially rendered by 15 CODEX cycles of the experiment. (Similar results were obtained for Cy3-positive populations – data not shown). Note the overall lack of dependence of the average signal intensity on the cycle number, indicating the robustness of the stained tissue to the chemistry of rendering by CODEX.

**Supplementary Figure 2** (available online[38])**. In situ imaging examples and surface marker expression profiles of cell types identified in normal and MRL spleens.**

First panel shows schematics of how data represented across the individual pages of this multipage figure. Every panel shows per cell type marker expression profile; high resolution montage of images acquired in CODEX cycles with cell of indicated type marked with yellow crosses and distribution of the cell type across normal and autoimmune spleens of the dataset.

**Supplementary Figure 3**. **Cross tissue and cross samples distribution of cell types identified in normal and MRL spleens.** (available online[38])

First panel shows schematics of how data represented across the individual pages of this multipage figure. Every panel shows low resolution montage of tiled images of each tissue sections imaged in this study and distribution of cells of indicated type across these images the cell are marked with white circle.

**Supplementary Figure 4. CD4(+)MHCII(+) cells correspond to ILC3 of Lti type**

**(A)** Distribution of CD4(+)MHCII(+) cells (marked with white circles) in BALBc #2 spleen stained with IgD (green) and CD90 (red) to indicate positions of B and T cells accordingly. **(B)** CD4 and MHCII expression in isolated mouse splenocytes gated negative for all CODEX panel markers and in addition

120g8 (lineage depletion with BD 558451 and dump channel for FITC conjugated or biotinilated antibodies corresponding to the antigens stained with CODEX panel were used for negative gating) except CD4, MHCII, CD45 and CD44. **(C)** CD4(+)MHCII(+) cells within the gate shown in (B) were sorted out and subjected to microarray analysis. CD4 T cells, CD8 T cells, bulk B cells and Conventional CD11c positive dendritic cells were co–sorted as a control. Expression of Lti signature genes (two individual signature sets as inferred in {Robinette, 2015 #1194} ) were analyzed in all sorted cells

**Supplementary Figure 5. Homotypic adhesion drives thread-like arrangement of CD8 cells.**

**(A)** Thread like arrangement of CD8 T cells (purple, annotated with V-letter) has been noticed in PALS of splenic samples across dataset. To examine potential mechanisms driving these structures CD8 Tells and B220 positive B cells were sorted individually from BALBc spleen **(B)** and later combined in flat bottom microwell plates and mixed at 37C in culture medium **C.** After of mixing cells were stained for B220 (green) and CD8a (red) and imaged. Thread like structures similar to what was observed in spleen were detected.

**Supplementary Figure 6. Index cell types distribution between niches**

**(A).** Heat map showing average frequencies of cell types (rows of heatmap) in the ring of index cell neighbors (see schematics on the left) for all niche clusters (0-99 in columns) **(B)** Heat map shows how different cell types (in rows) are distributed between niches (in columns).

**Supplementary Figure 7. Types of samples in MRL/lpr dataset.**

MRL/lpr dataset had 9 samples: 3 control wild type BALBc spleens (BALBc -1,-2,-3 and 6 MRL spleens MRL -4,-5,-6,-7,-8,-9). Based on disintegration of marginal zone as measured by frequency of marginal zone macrophages (MZMPhs) and accumulation of double negative T-cells expressing B220 B cell marker (B220 DN T cells) MRL spleens were grouped into early (MRL -4,-5,-6), intermediate (MRL -7,-8), and late types. Early stage was represented by 3 MZM positive DN T cell-low spleens. Int1 stage was represented by single MZM low DN T cell-low spleen. Int2 stage was represented by single MZM positive DN T cell-positive spleen. Late stage was represented by single MZM positive DN T cell-positive spleen. A single representative spleen is shown for each stage together with interaction matrix. Color represents odd ratios (observed frequency of interaction/ expected frequency of interaction).

**Supplementary Figure 8. Cross tissue and cross samples distribution of interacting cell pairs for selected types of cell-to-cell interactions.** (available online[38])

First panel shows schematics of how data represented across the individual pages of this multipage figure. Every next page shows schematics of dataset tissue types and low resolution tiled view of interacting cell pairs across imaged sections of normal and MRL LPR spleen. One cell type is marked with white circle and the other with cyan. Due to cell proximity in most cases cyan circles practically completely overlay white.

**Supplementary Figure 9. Expanding the multiplexing limit of CODEX by "panels and activators" design.**

**(A)** Diagram of "multipanel"/"activator oligo" CODEX approach. The list of antibodies can be divided in sets such that number of antibodies in each individual set does not exceed the capacity of the multiplexing protocol to render staining without significant signal loss (e.g.30). Each such set of antibodies will be conjugated to "terminated" (the last 3' base is dideoxy- or propyl- modified) upper strand oligonucleotide of the same sequence as in the original version of the "missing base" approach. The lower strand oligonucleotides will incorporate an additional set-specific region, which will serve as a landing spot for the dedicated primer oligo which is to be on-slide hybridized to the particular subset of the total plurality of the antibodies at the time when they are to be rendered. This approach prevents extension of reads beyond certain threshold and at the same time have an unlimited potential number of antibodies in the sample. **(B)** Schematics of experiment demonstrating the "activator" method and its robustness. Each antigen of a set of 22 surface markers is redundantly detected by three CODEX tag conjugates of the same antibody. The first conjugate is detected during panel 1 rendering, second – during panel 2 etc... Thus the signal for same antigen is detected at different cycles (e.g., 1st, 13th, and 24th) **(C)** Montage of a fragment of imaging field of the 36 cycles of CODEX used to render a mixture of 18 barcoded spleens (similar to design in Figure 2A). Cycles N,N+12 and N+24 all three of which render same pair of antigens are shown per tile for all 11 pairs of antigens (see annotation in the black rectangle of each tile)

**Supplementary Figure 10. The "reversible terminator method" variation of the CODEX.**

This implementation of the CODEX relies on reversible terminators, i.e., chain terminator nucleotides that can be de-protected after incorporation, thereby allowing further nucleotides to be added to that nucleotide. Similar to approach in Fig.1 the planar sample is co-stained simultaneously using a panel of capture agents, each labeled with one oligonucleotide duplex designed according to the strategy outlined on the diagram. The duplexes are designed in such a way that each antibody has the same upper strand sequence linked, covalently or through streptavidin, to an antibody through the 5' end. The lower strand changes from antibody to antibody. In this implementation, the general formula for the lower strand is 3'-dideoxydC -sequence-complimentary-to-upper-strand $G_nA/T/C$-5' . One type of lower strand base (nucleotide C in this example) is reserved for step-wise progression and its complementary pair on the upper strand is never used in labeled form. The other three bases (as opposed to “missing base” approach see Fig.1) are complementary to labeled nucleotides and can be used to identify three capture agents per cycle. Each cycle includes: (a) a labeling step in which the three capture agents are labeled and duplexes on the rest are extended with an unlabeled nucleotide one base at a time, (b) an imaging step and (c) a destaining/deprotection step. During destaining/deprotection step of the cycle to cycle transition the added fluorescent labels from the previous cycle are inactivated by any of the suitable methods, including cleavage of fluorophore off the nucleotide (if the labeled nucleotide is linked to the fluorophore through a cleavable linker); peroxide based bleaching; photobleaching; chemically-assisted photobleaching; After or simultaneously with inactivation of the fluorophores added in the previous reaction, the unlabeled “extension” nucleotide that has been added to the remainder of the capture agents is activated by cleavage of the protective group off its 3’ end. Cleavage of the protective group, in turn, allows that nucleotide to be extended in the next cycle.

## Supplementary Table Legends

(supplementary figures and tables not included in the main text can be found at at

http://welikesharingdata.blob.core.windows.net/forshare/index.html)

**Supplementary Table 1. List of CODEX antibodies and oligonucleotides.**

Excel file with four spreadsheets corresponding to multidimensional staining experiments performed in the study **(CODEX panel for cell spreads)** List of 24 antibodies (23 DNA conjugated + CD45 FITC for counterstain), upper and lower nucleotides used for CODEX staining of isolated splenocytes. **(CODEX panel for spleen tissue)** List of 30 antibodies (28 DNA conjugated + CD45 FITC and NKp46 PacBlue), upper and lower nucleotides used for comparative CODEX staining of normal BALBc and lupus afflicted MRL/lpr spleen sections. **(CYTOF panel for spleen cells)** List of 23 metal conjugated antibodies antibodies used in CyTOF analysis of isolated splenocytes. **(Activator driven CODEX panels)** List of 22 antibodies (22 DNA conjugated + CD45 FITC for counterstain), upper, lower and activator nucleotides used for activator driven CODEX staining of isolated splenocytes (see exp. Schematics in Supplementary Figure 9).

**Supplementary Table 2. Segmentation and quantification of surface marker expression in cells of normal and MRL/lpr spleen.**

Excel file with cell types annotation, expression profiles and coordinates of all segmented objected identified in spleen sections analyzed in this study

## Supplementary Table 3. Delaunay neighborhood graph.

Excel file with annotation and coordinates of all pairwise cell to cell contacts (interactions) mapped across samples imaged in this study

## Supplementary Table 4. X-shift cluster annotations and cell counts

Excel file with 58 clusters identified by X-shift analysis, their annotations and resulting across dataset counts for 27 imaging phenotypes identified in this study

## Supplementary Table 5. Dynamics of average cell type to cell type interaction frequency and strength across dataset.

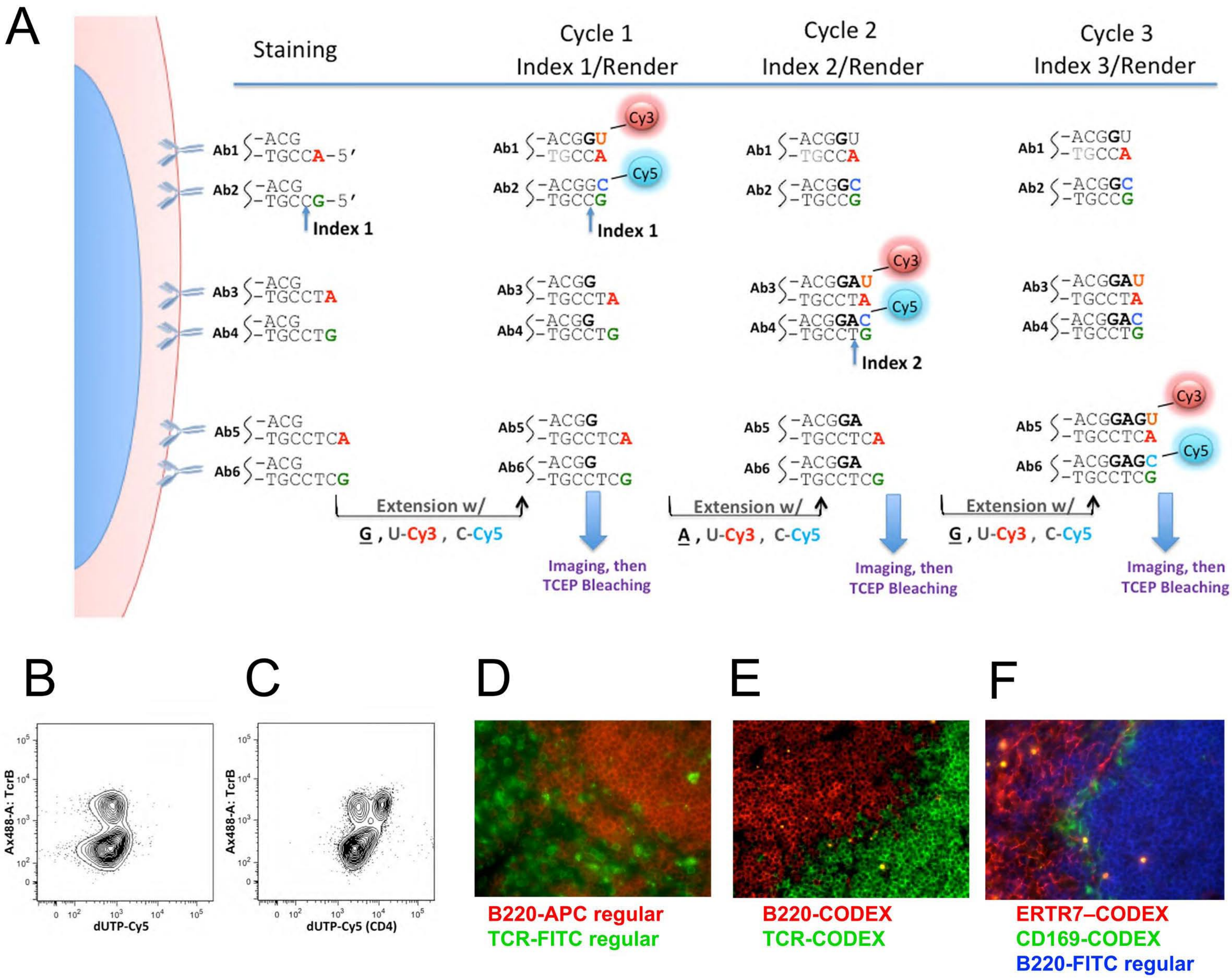
Excel table with three spread sheets. **Full data** contains odds ratios; direct counts of interactions as well as various differential metrics for comparisons off frequency and strength of cell type to cell type interactions between early MRL and control (BALBc) and intermediate-late MRL and early MRL. **Early vs control** shows top candidate cell type pairs selected based on the change in strength (odds ratios) or frequency of interactions between early MRL spleen and control spleens. **Late vs early** shows top candidate cell type pairs selected based on the change in strength (odds ratios) or frequency of interactions between combined intermediate and late MRL spleens and early MRL spleens.

## Supplementary Table 6.

## Linear regression model for marker expression level based on niche and cell type shows importance of niche.

The overall role of the niche in defining marker expression was evaluated by constructing a linear regression model of marker expression with cell type identity and niche as two feature variables. This Excel file shows F and P values for the contribution of niche to the model . The F value is the ratio of the mean regression sum of squares for the model including just cell type to the full model including both niche and the cell type. Its value ranges zero to an arbitrarily large number. A larger F value suggests that the niche has a larger contribution in explaining the variance observed in the expression levels of each marker. The value of $\Pr(>F)$ is the p-value against the null hypothesis that including the niche in the model does not improve the fit.

**FIGURE 1**



B220-APC regular
TCR-FITC regular

B220-CODEX
TCR-CODEX

ERTR7–CODEX
CD169-CODEX
B220-FITC regular

**FIGURE 2**

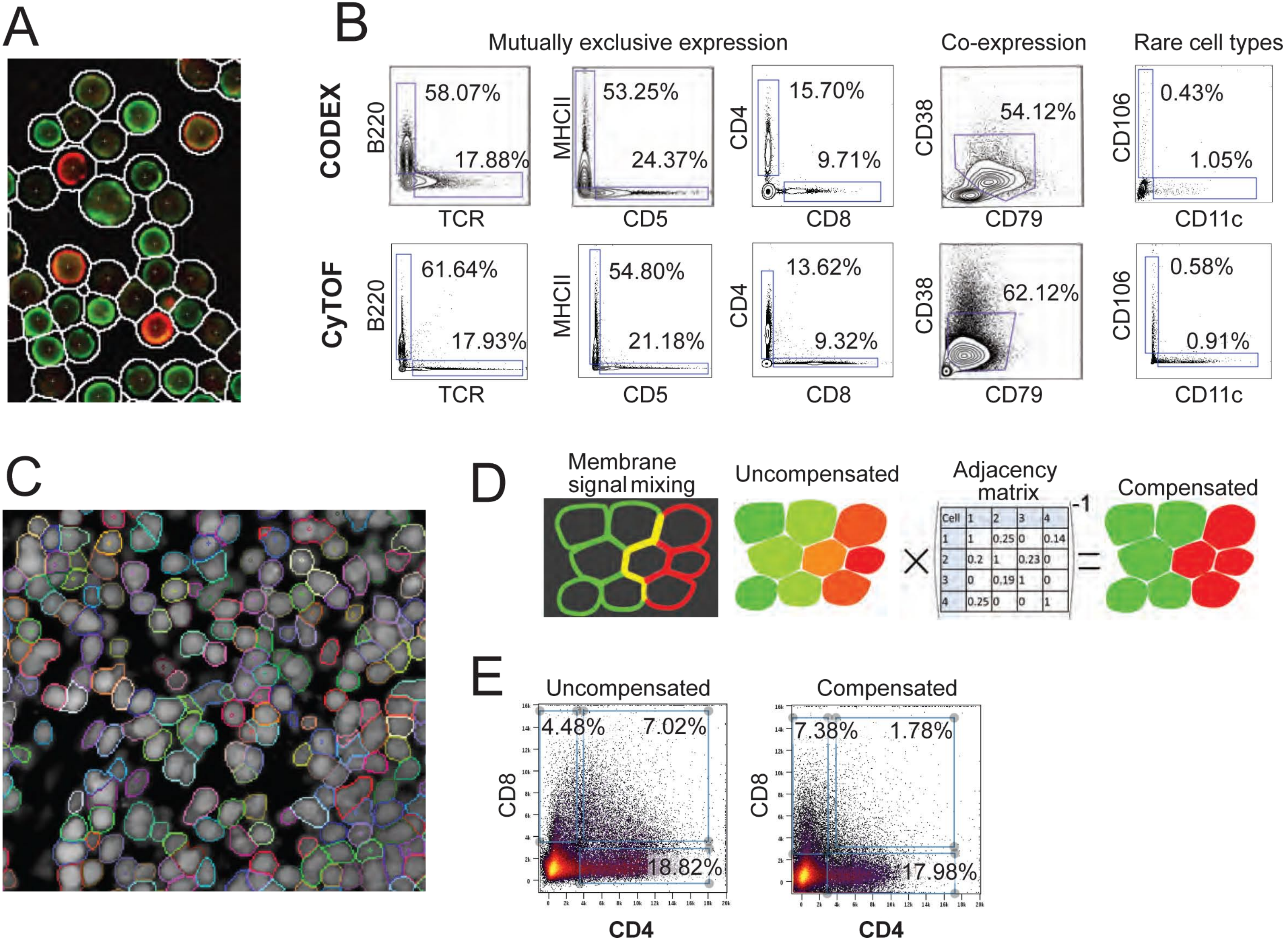
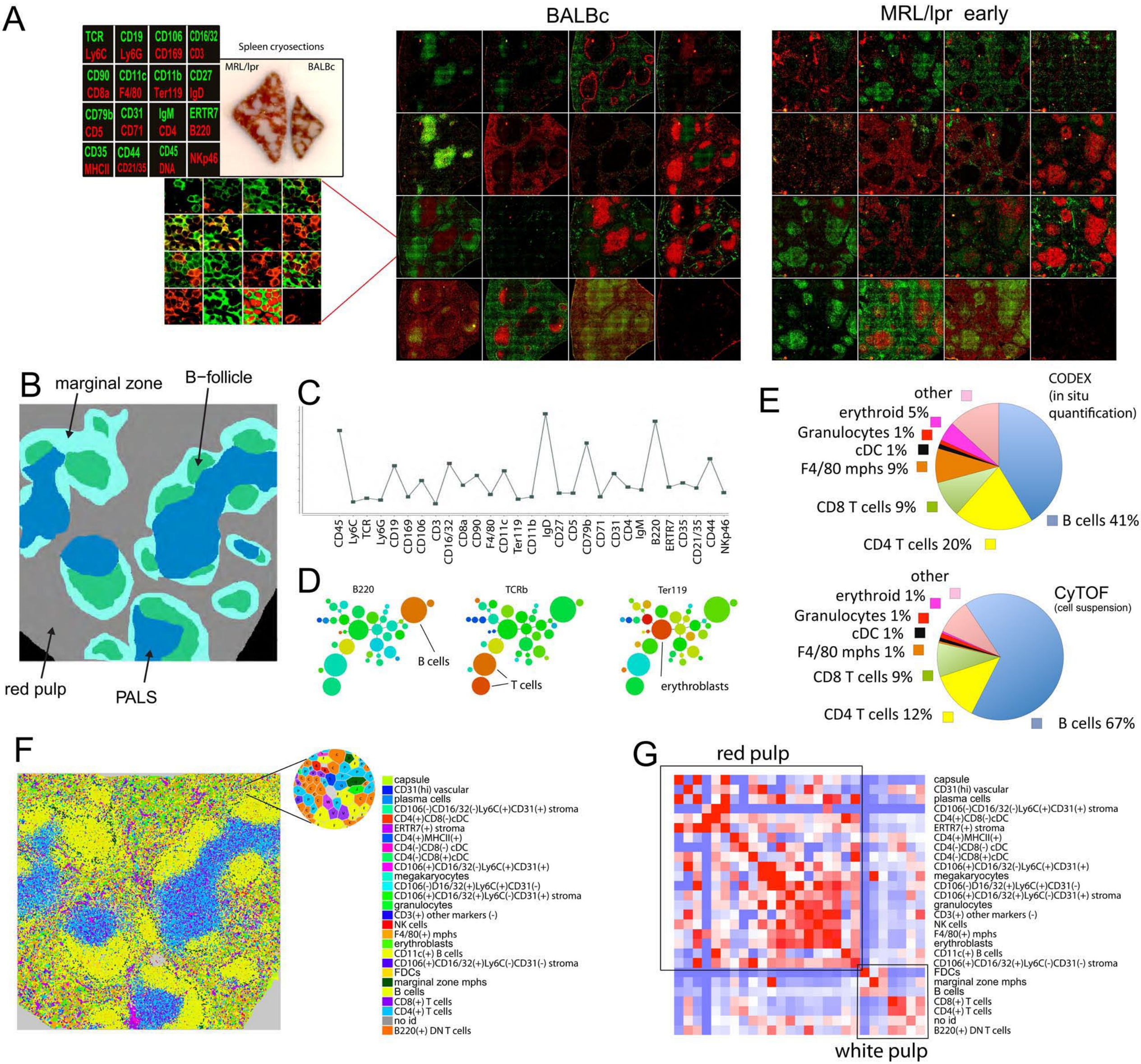

A
B
Mutually exclusive expression
Co-expression
Rare cell types
CODEX
CyTOF
B220
TCR
58.07%
17.88%
MHCII
CD5
53.25%
24.37%
CD4
CD8
15.70%
9.71%
CD38
CD79
54.12%
CD106
CD11c
0.43%
1.05%
61.64%
17.93%
54.80%
21.18%
13.62%
9.32%
62.12%
0.58%
0.91%
C
D
Membrane signal mixing
Uncompensated
Adjacency matrix
-1
Compensated
Cell 1 2 3 4
1 1 0.25 0 0.14
2 0.2 1 0.23 0
3 0 0.19 1 0
4 0.25 0 0 1
E
Uncompensated
Compensated
CD8
CD4
4.48%
7.02%
18.82%
7.38%
1.78%
17.98%
16k
14k
12k
10k
8k
6k
4k
2k
0
0
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k

FIGURE 3


A
TCR
Ly6C
CD19
Ly6G
CD106
CD169
CD16/32
CD3
CD90
CD8a
CD11c
F4/80
CD11b
Ter119
CD27
IgD
CD79b
CD5
CD31
CD71
IgM
CD4
ERTR7
B220
CD35
MHCII
CD44
CD21/35
CD45
DNA
NKp46
Spleen cryosections
MRL/lpr
BALBc
BALBc
MRL/lpr early
B
marginal zone
B−follicle
red pulp
PALS
C
CD45
Ly6C
TCR
Ly6G
CD19
CD169
CD106
CD3
CD16/32
CD8a
CD90
F4/80
CD11c
Ter119
CD11b
IgD
CD27
CD5
CD79b
CD71
CD31
CD4
IgM
B220
ERTR7
CD35
CD21/35
CD44
NKp46
D
B220
B cells
TCRb
T cells
Ter119
erythroblasts
E
other
erythroid 5%
Granulocytes 1%
cDC 1%
F4/80 mphs 9%
CD8 T cells 9%
CD4 T cells 20%
B cells 41%
CODEX
(in situ
quantification)
other
erythroid 1%
Granulocytes 1%
cDC 1%
F4/80 mphs 1%
CD8 T cells 9%
CD4 T cells 12%
B cells 67%
CyTOF
(cell suspension)
F
capsule
CD31(hi) vascular
plasma cells
CD106(-)CD16/32(-)Ly6C(+)CD31(+) stroma
CD4(+)CD8(-)cDC
ERTR7(+) stroma
CD4(+)MHCII(+)
CD4(-)CD8(-) cDC
CD4(-)CD8(+)cDC
CD106(+)CD16/32(-)Ly6C(+)CD31(+)
megakaryocytes
CD106(-)D16/32(+)Ly6C(+)CD31(-)
CD106(+)CD16/32(+)Ly6C(-)CD31(+) stroma
granulocytes
CD3(+) other markers (-)
NK cells
F4/80(+) mphs
erythroblasts
CD11c(+) B cells
CD106(+)CD16/32(+)Ly6C(-)CD31(-) stroma
FDCs
marginal zone mphs
B cells
CD8(+) T cells
CD4(+) T cells
no id
B220(+) DN T cells
G
red pulp
white pulp
capsule
CD31(hi) vascular
plasma cells
CD106(-)CD16/32(-)Ly6C(+)CD31(+) stroma
CD4(+)CD8(-)cDC
ERTR7(+) stroma
CD4(+)MHCII(+)
CD4(-)CD8(-) cDC
CD4(-)CD8(+)cDC
CD106(+)CD16/32(-)Ly6C(+)CD31(+)
megakaryocytes
CD106(-)D16/32(+)Ly6C(+)CD31(-)
CD106(+)CD16/32(+)Ly6C(-)CD31(+) stroma
granulocytes
CD3(+) other markers (-)
NK cells
F4/80(+) mphs
erythroblasts
CD11c(+) B cells
CD106(+)CD16/32(+)Ly6C(-)CD31(-) stroma
FDCs
marginal zone mphs
B cells
CD8(+) T cells
CD4(+) T cells
no id
B220(+) DN T cells

# FIGURE 4

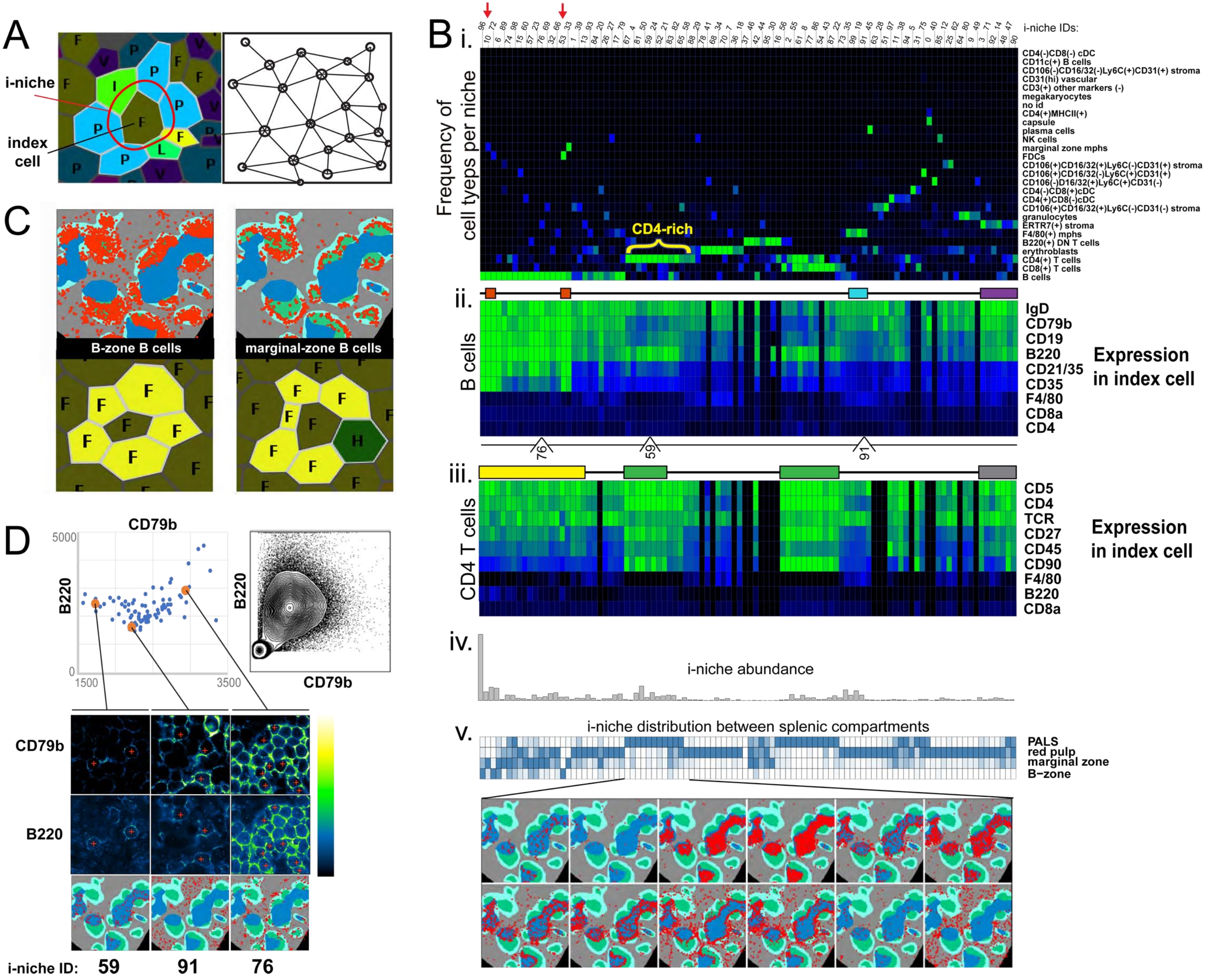
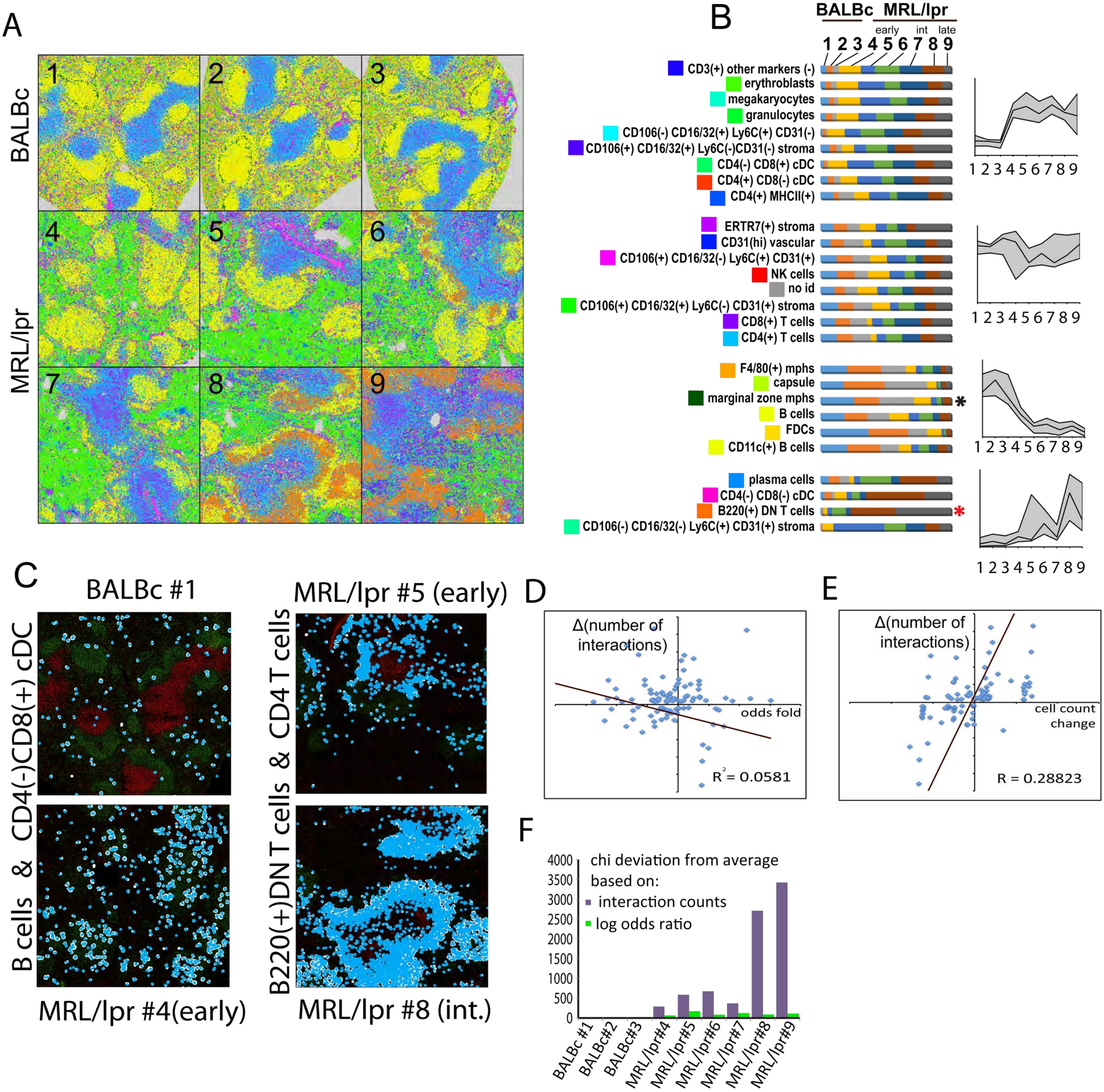

A
i-niche
index cell
F
P
V
I
L
B i.
i-niche IDs:
96 10 72 6 89 74 98 15 60 57 23 76 69 32 66 53 33 1 39 13 93 84 20 26 27 17 79 67 4 81 50 59 24 52 21 83 82 65 58 88 29 78 41 68 34 70 7 36 18 37 46 42 44 95 30 16 56 2 55 61 8 77 86 54 43 87 22 73 35 99 19 91 45 63 28 51 97 11 38 94 5 31 75 0 40 85 12 25 62 64 80 9 49 3 71 92 14 48 47 90
Frequency of cell tyeps per niche
CD4(-)CD8(-) cDC
CD11c(+) B cells
CD106(-)CD16/32(-)Ly6C(+)CD31(+) stroma
CD31(hi) vascular
CD3(+) other markers (-)
megakaryocytes
no id
CD4(+)MHCII(+)
capsule
plasma cells
NK cells
marginal zone mphs
FDCs
CD106(+)CD16/32(+)Ly6C(-)CD31(+) stroma
CD106(+)CD16/32(-)Ly6C(+)CD31(+)
CD106(-)D16/32(+)Ly6C(+)CD31(-)
CD4(-)CD8(+)cDC
CD4(+)CD8(-)cDC
CD106(+)CD16/32(+)Ly6C(-)CD31(-) stroma
granulocytes
ERTR7(+) stroma
F4/80(+) mphs
B220(+) DN T cells
erythroblasts
CD4(+) T cells
CD8(+) T cells
B cells
CD4-rich
ii.
B cells
IgD
CD79b
CD19
B220
CD21/35
CD35
F4/80
CD8a
CD4
Expression in index cell
76
59
91
iii.
CD4 T cells
CD5
CD4
TCR
CD27
CD45
CD90
F4/80
B220
CD8a
Expression in index cell
iv.
i-niche abundance
v.
i-niche distribution between splenic compartments
PALS
red pulp
marginal zone
B−zone
C
B-zone B cells
marginal-zone B cells
F
H
D
CD79b
B220
5000
0
1500
3500
B220
CD79b
CD79b
B220
i-niche ID:
59
91
76

FIGURE 5


A
BALBc
MRL/lpr
1
2
3
4
5
6
7
8
9
B
BALBc
MRL/lpr
early
int
late
1 2 3 4 5 6 7 8 9
CD3(+) other markers (-)
erythroblasts
megakaryocytes
granulocytes
CD106(-) CD16/32(+) Ly6C(+) CD31(-)
CD106(+) CD16/32(+) Ly6C(-)CD31(-) stroma
CD4(-) CD8(+) cDC
CD4(+) CD8(-) cDC
CD4(+) MHCII(+)
ERTR7(+) stroma
CD31(hi) vascular
CD106(+) CD16/32(-) Ly6C(+) CD31(+)
NK cells
no id
CD106(+) CD16/32(+) Ly6C(-) CD31(+) stroma
CD8(+) T cells
CD4(+) T cells
F4/80(+) mphs
capsule
marginal zone mphs
B cells
FDCs
CD11c(+) B cells
plasma cells
CD4(-) CD8(-) cDC
B220(+) DN T cells
CD106(-) CD16/32(-) Ly6C(+) CD31(+) stroma
*
*
1 2 3 4 5 6 7 8 9
1 2 3 4 5 6 7 8 9
1 2 3 4 5 6 7 8 9
1 2 3 4 5 6 7 8 9
C
BALBc #1
MRL/lpr #5 (early)
B cells & CD4(-)CD8(+) cDC
B220(+)DN T cells & CD4 T cells
MRL/lpr #4(early)
MRL/lpr #8 (int.)
D
Δ(number of interactions)
odds fold
$R^2 = 0.0581$
E
Δ(number of interactions)
cell count change
R = 0.28823
F
chi deviation from average based on:
interaction counts
log odds ratio
4000
3500
3000
2500
2000
1500
1000
500
0
BALBc #1
BALBc#2
BALBc#3
MRL/lpr#4
MRL/lpr#5
MRL/lpr#6
MRL/lpr#7
MRL/lpr#8
MRL/lpr#9

# FIGURE 6

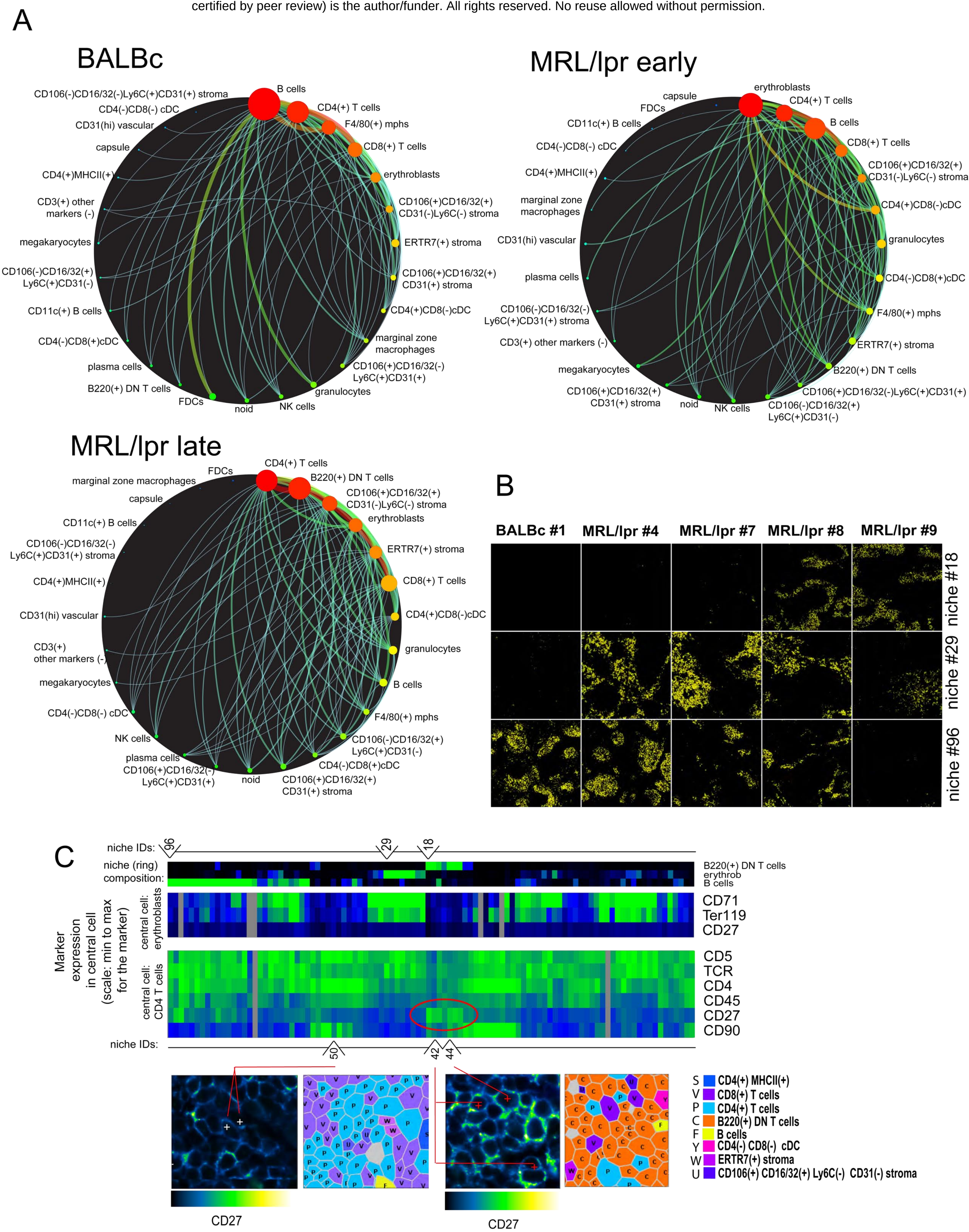
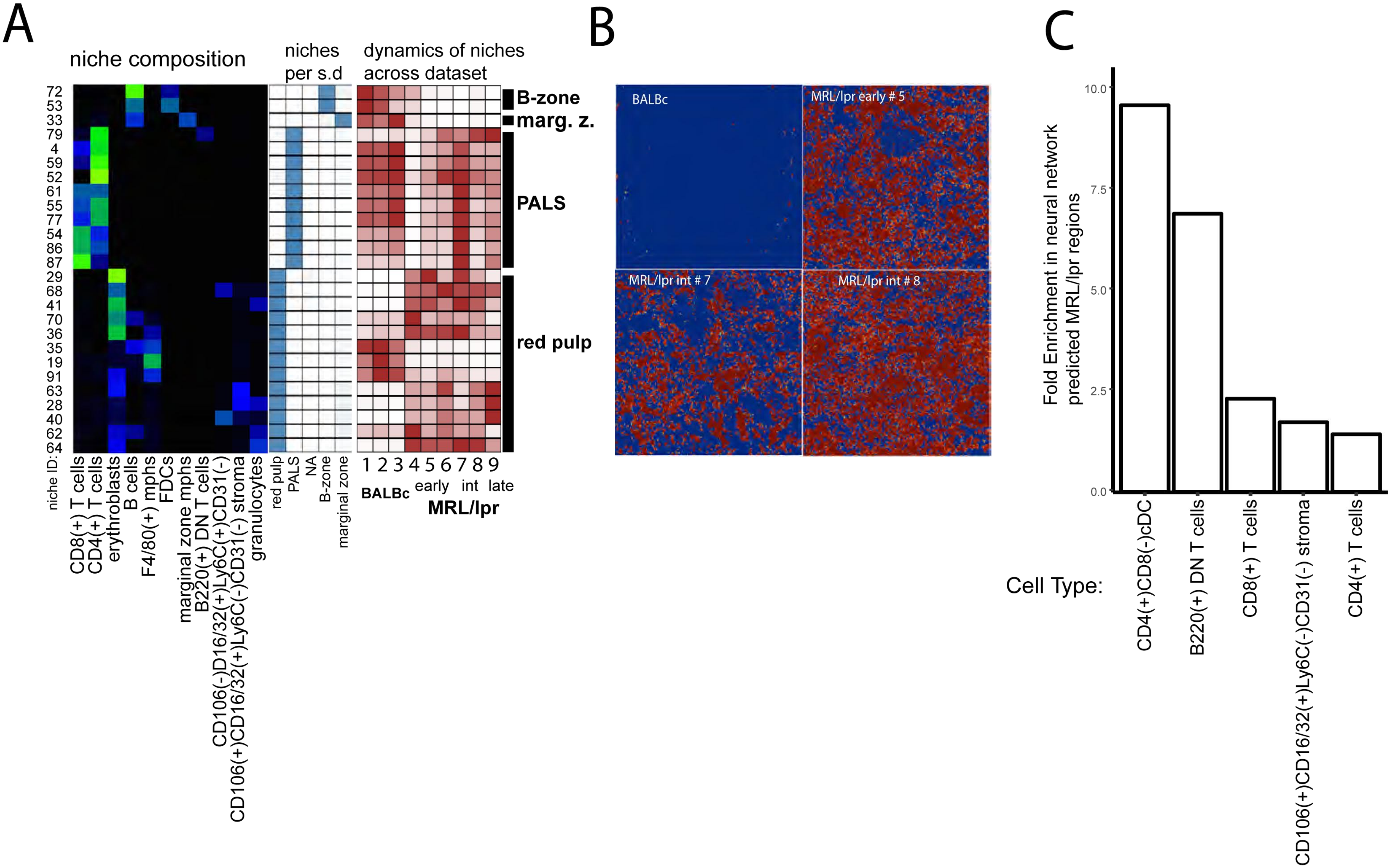
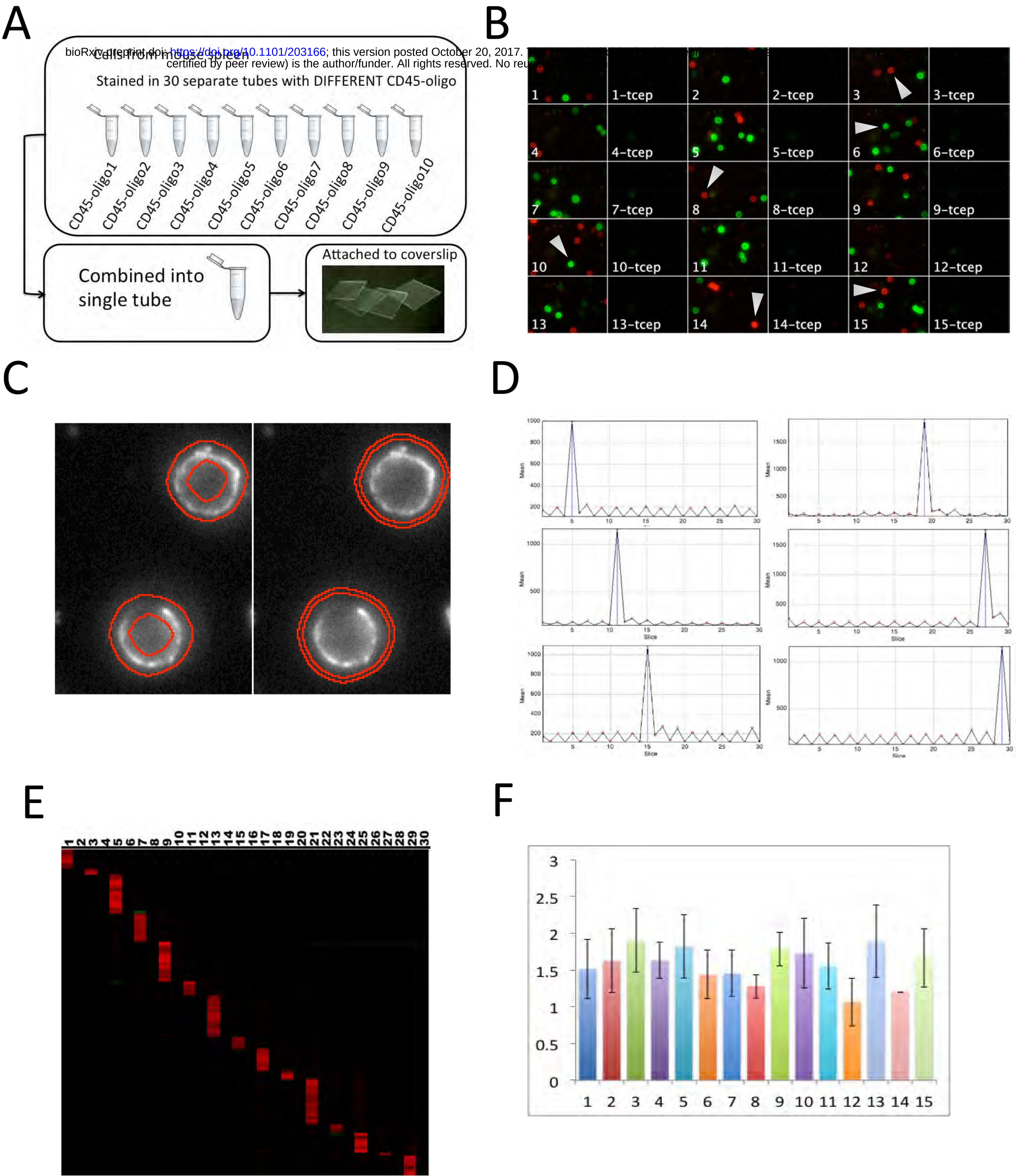

A
BALBc
B cells
CD106(-)CD16/32(-)Ly6C(+)CD31(+) stroma
CD4(-)CD8(-) cDC
CD4(+) T cells
CD31(hi) vascular
F4/80(+) mphs
capsule
CD8(+) T cells
CD4(+)MHCII(+)
erythroblasts
CD3(+) other markers (-)
CD106(+)CD16/32(+) CD31(-)Ly6C(-) stroma
megakaryocytes
ERTR7(+) stroma
CD106(-)CD16/32(+) Ly6C(+)CD31(-)
CD106(+)CD16/32(+) CD31(+) stroma
CD11c(+) B cells
CD4(+)CD8(-)cDC
CD4(-)CD8(+)cDC
marginal zone macrophages
plasma cells
CD106(+)CD16/32(-) Ly6C(+)CD31(+)
B220(+) DN T cells
granulocytes
FDCs
noid
NK cells
MRL/lpr early
erythroblasts
capsule
CD4(+) T cells
FDCs
CD11c(+) B cells
B cells
CD4(-)CD8(-) cDC
CD8(+) T cells
CD4(+)MHCII(+)
CD106(+)CD16/32(+) CD31(-)Ly6C(-) stroma
marginal zone macrophages
CD4(+)CD8(-)cDC
CD31(hi) vascular
granulocytes
plasma cells
CD4(-)CD8(+)cDC
CD106(-)CD16/32(-) Ly6C(+)CD31(+) stroma
F4/80(+) mphs
CD3(+) other markers (-)
ERTR7(+) stroma
megakaryocytes
B220(+) DN T cells
CD106(+)CD16/32(+) CD31(+) stroma
CD106(+)CD16/32(-)Ly6C(+)CD31(+)
noid
NK cells
CD106(-)CD16/32(+) Ly6C(+)CD31(-)
MRL/lpr late
CD4(+) T cells
FDCs
marginal zone macrophages
B220(+) DN T cells
capsule
CD106(+)CD16/32(+) CD31(-)Ly6C(-) stroma
CD11c(+) B cells
erythroblasts
CD106(-)CD16/32(-) Ly6C(+)CD31(+) stroma
ERTR7(+) stroma
CD4(+)MHCII(+)
CD8(+) T cells
CD31(hi) vascular
CD4(+)CD8(-)cDC
CD3(+) other markers (-)
granulocytes
megakaryocytes
B cells
CD4(-)CD8(-) cDC
F4/80(+) mphs
NK cells
CD106(-)CD16/32(+) Ly6C(+)CD31(-)
plasma cells
CD4(-)CD8(+)cDC
CD106(+)CD16/32(-) Ly6C(+)CD31(+)
noid
CD106(+)CD16/32(+) CD31(+) stroma
B
BALBc #1
MRL/lpr #4
MRL/lpr #7
MRL/lpr #8
MRL/lpr #9
niche #18
niche #29
niche #96
C
niche IDs:
96
29
18
niche (ring) composition:
B220(+) DN T cells
erythrob
B cells
Marker expression in central cell (scale: min to max for the marker)
central cell: erythroblasts
CD71
Ter119
CD27
central cell: CD4 T cells
CD5
TCR
CD4
CD45
CD27
CD90
niche IDs:
50
42
44
CD27
CD27
S CD4(+) MHCII(+)
V CD8(+) T cells
P CD4(+) T cells
C B220(+) DN T cells
F B cells
Y CD4(-) CD8(-) cDC
W ERTR7(+) stroma
U CD106(+) CD16/32(+) Ly6C(-) CD31(-) stroma

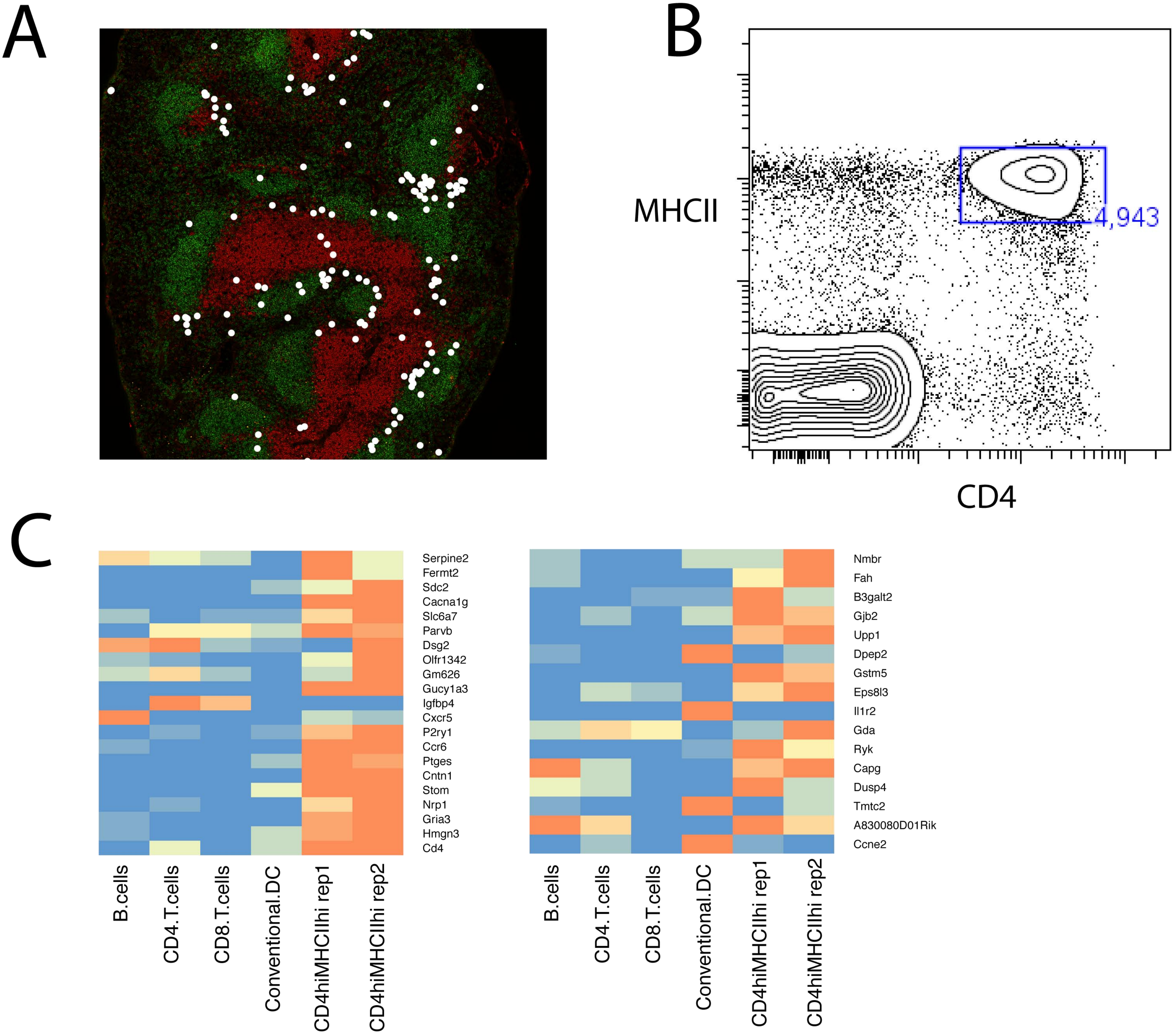
FIGURE 7


A
niche composition
niches per s.d
dynamics of niches across dataset
niche ID:
72
53
33
79
4
59
52
61
55
77
54
86
87
29
68
41
70
36
35
19
91
63
28
40
62
64
CD8(+) T cells
CD4(+) T cells
erythroblasts
B cells
F4/80(+) mphs
FDCs
marginal zone mphs
B220(+) DN T cells
CD106(-)D16/32(+)Ly6C(+)CD31(-)
CD106(+)CD16/32(+)Ly6C(-)CD31(-) stroma
granulocytes
red pulp
PALS
NA
B-zone
marginal zone
1 2 3 4 5 6 7 8 9
BALBc
early
int
late
MRL/lpr
B-zone
marg. z.
PALS
red pulp
B
BALBc
MRL/lpr early # 5
MRL/lpr int # 7
MRL/lpr int # 8
C
Fold Enrichment in neural network predicted MRL/lpr regions
10.0
7.5
5.0
2.5
0.0
Cell Type:
CD4(+)CD8(-)cDC
B220(+) DN T cells
CD8(+) T cells
CD106(+)CD16/32(+)Ly6C(-)CD31(-) stroma
CD4(+) T cells

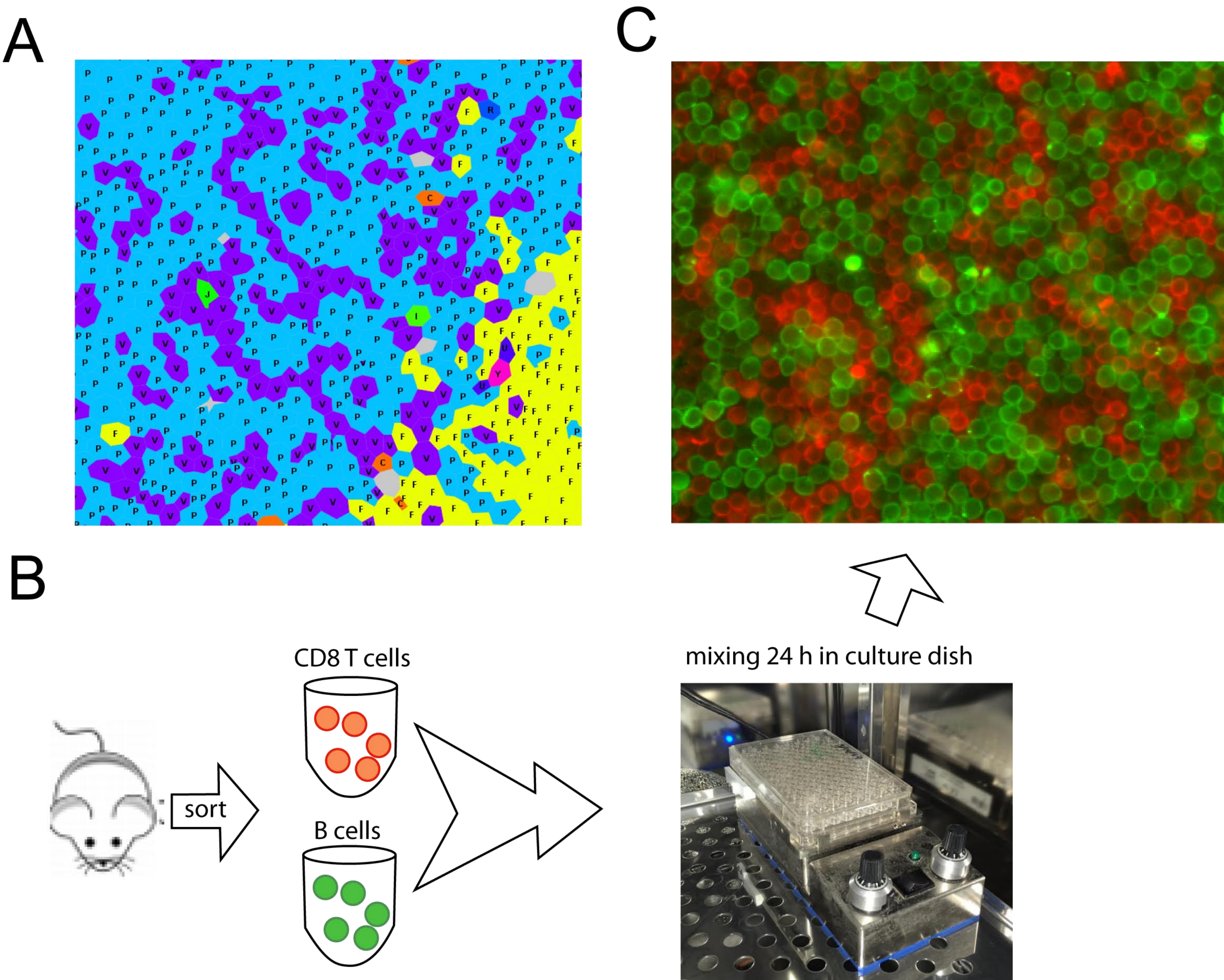
# Supplementary Figure 1



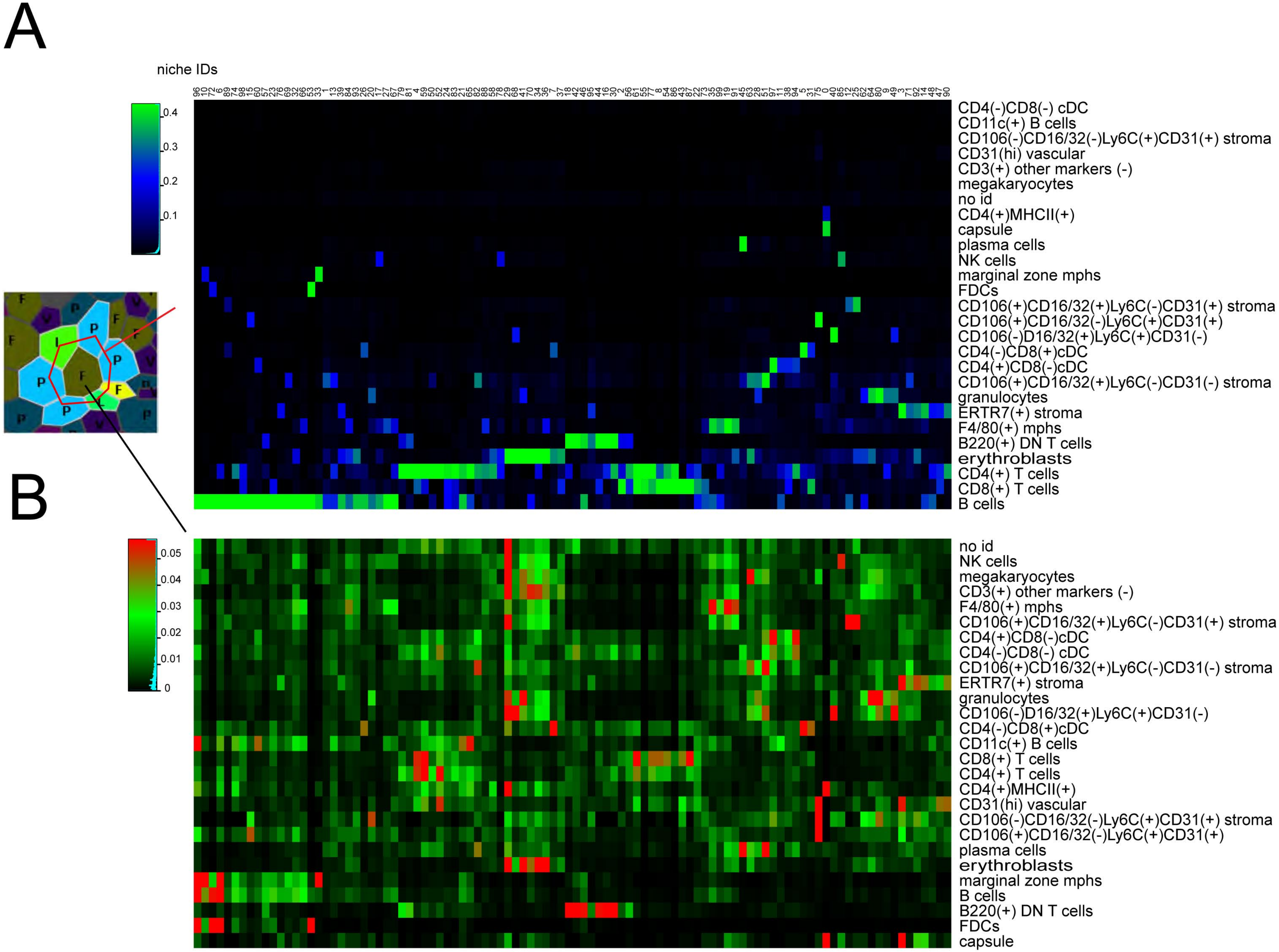
# Supplementary Figure 4



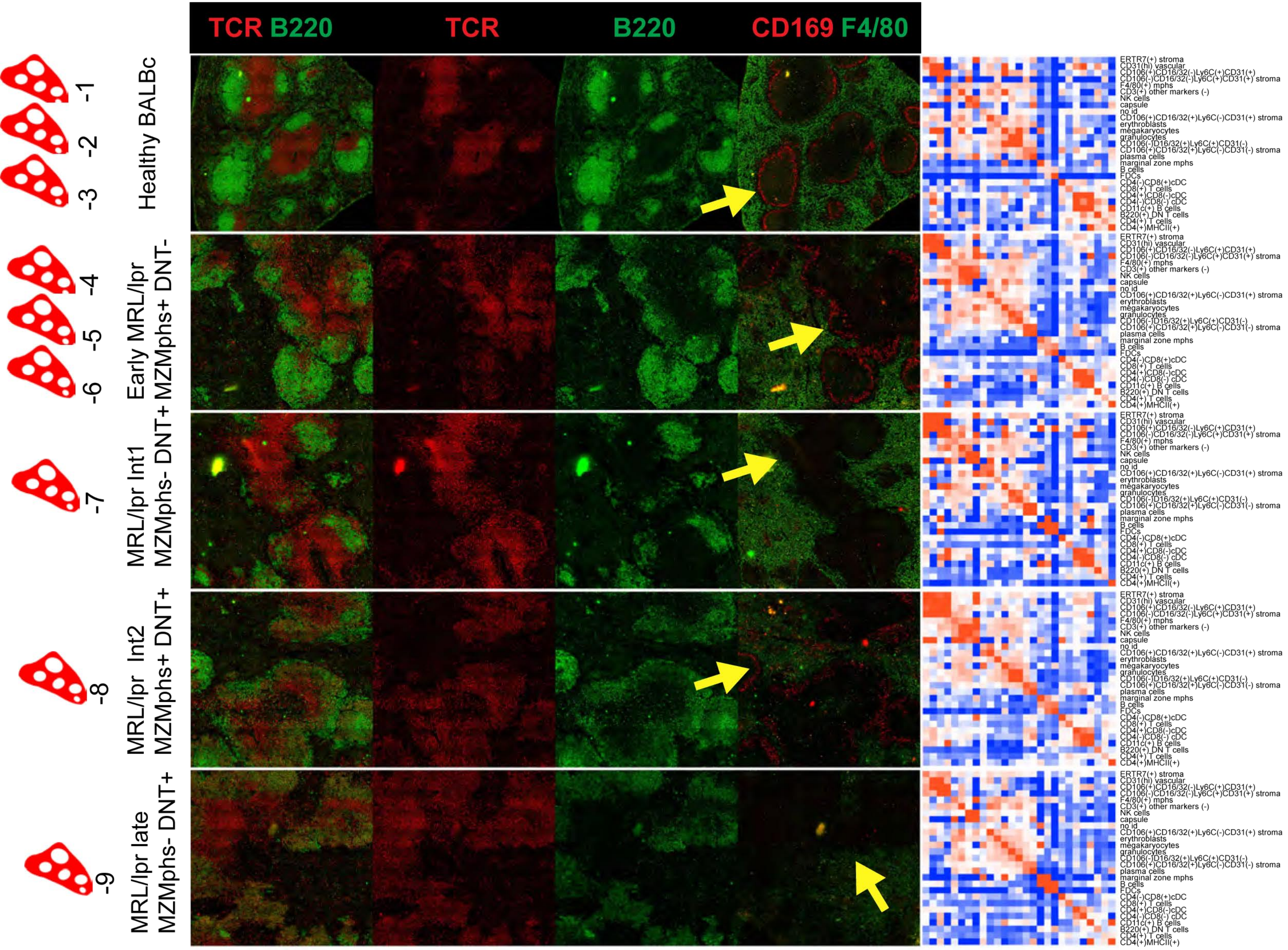
# Supplementary Figure 5



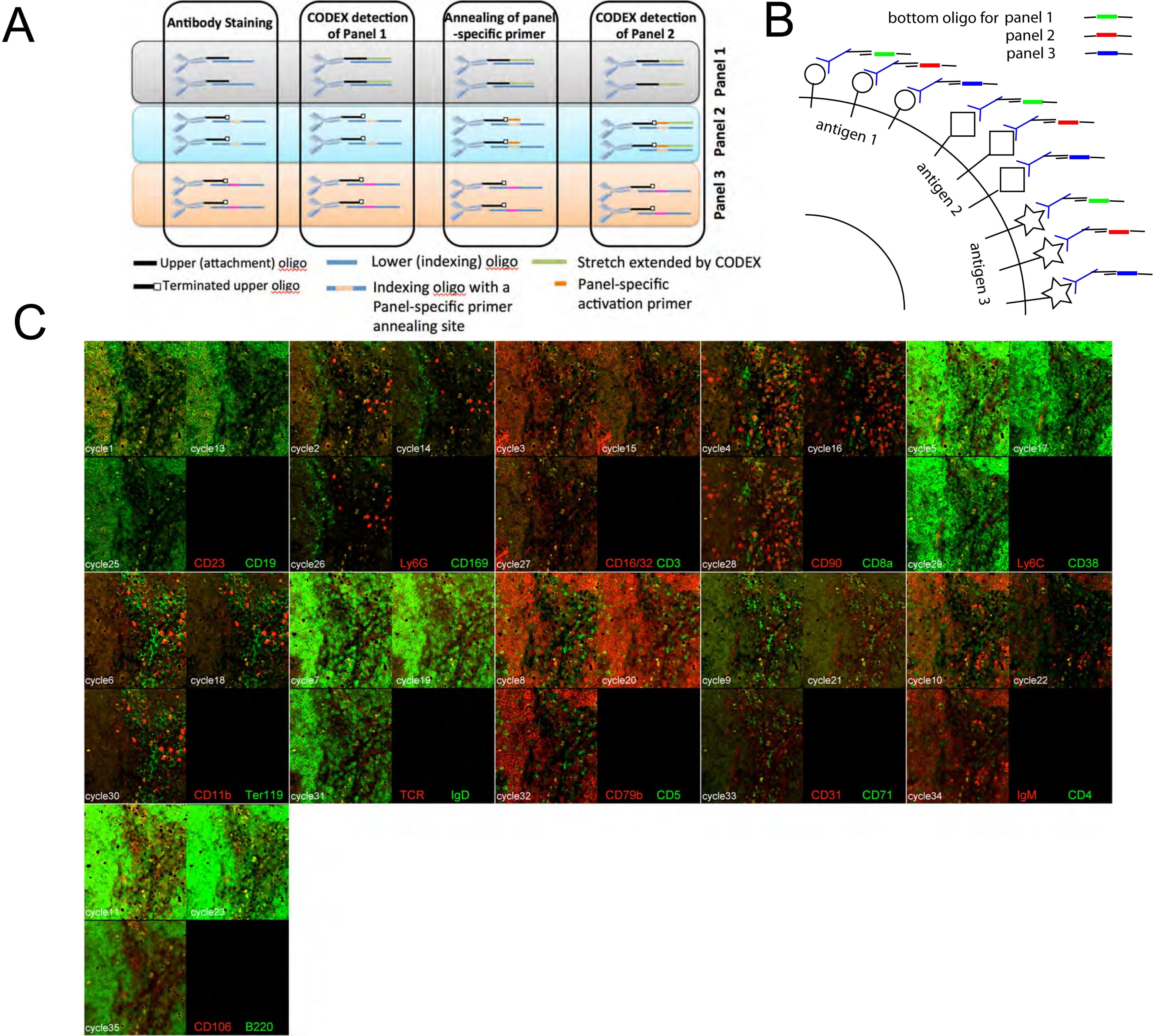
Supplementary Figure 6



# Supplementary Figure 7



# Supplementary Figure 9



## Supplementary Figure 10