Supplementary Materials for

An Interactive Reference Framework for Modeling a Dynamic Immune System

Matthew H. Spitzer[1,2,3]*‡, Pier Federico Gherardini[1]*, Gabriela K. Fragiadakis[1], Nupur Bhattacharya[2], Robert T. Yuan[2,3], Andrew N. Hotson[1], Rachel Finck[1], Yaron Carmi[2], Eli R. Zunder[1], Wendy J. Fantl[1], Sean C. Bendall[2,3], Edgar G. Engleman[2,3]† and Garry P. Nolan[1,3]†‡

[1]Baxter Laboratory in Stem Cell Biology, Department of Microbiology and Immunology
[2]Department of Pathology
[3]Program in Immunology, Stanford University, Stanford, CA 94305, USA.

* Co-first authors.
† Co-senior authors.
‡To whom correspondence should be addressed:
gnolan@stanford.edu, matthew.spitzer@stanford.edu

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S20
Tables S1 to S4
Author Contributions

Materials and Methods

Mice

Wild-type male C57BL/6, 129S1/Sv and Balb/c mice were purchased from The Jackson Laboratory at 11 weeks of age. Animals were rested in our animal facility for 1 week and sacrificed at 12 weeks of age. All mice were housed in an American Association for the Accreditation of Laboratory Animal Care–accredited animal facility and maintained in specific pathogen-free conditions. Animal experiments were approved and conducted in accordance with Stanford University Asia Pacific Laboratory Accreditation Cooperation #13605.

Antibodies

A summary of all mass cytometry antibodies, reporter isotopes and concentrations used for analysis can be found in Methods Table 1. Primary conjugates of mass cytometry antibodies were prepared 100 µg at a time using the MaxPAR antibody conjugation kit (DVS Sciences, Toronto, Canada) according to the manufacturer's recommended protocol. Following labeling, antibodies were diluted in Candor PBS Antibody Stabilization solution (Candor Bioscience GmbH, Wangen, Germany) supplemented with 0.02% NaN3 to between 0.1 and 0.3 mg/mL and stored long-term at 4°C. Each antibody clone and lot was titrated to optimal staining concentrations using primary murine samples.

Cell Preparation

All tissue preparations were performed simultaneously from each individual mouse. After euthanasia by C02 inhalation, peripheral blood was collected via the posterior vena cava prior to perfusion of the animal and transferred into sodium heparin-coated vacuum tubes 1:1 dilution in RMPI 1640. Spleens, lymph nodes and thymi were homogenized in PBS at 4˚C. Bone marrow was flushed from femuri and resuspended in PBS at 4˚. Lungs and liver were finely minced and digested in RPMI 1640 with 2% FCS, 1 mg/ml collagenase IV, and 0.1 mg/ml DNase I. Liver suspensions were resuspended in ACK lysis buffer, washed in PBS with 2% FCS and 2 mM EDTA, and enriched by density gradient centrifugation using Optiprep (AxisShield, Oslo, Norway). Colons were fileted in RPMI with 10% FCS and cut into 0.5cm pieces before washing in HBSS with 0.015% DTT for 15 min at on a stir plate 37˚C. Colons were then washed in HBSS with 5% FCS and 25mM HEPES for 30 min at on a stir plate 37˚C and then digested in RPMI with 10% FCS, 0.167mg/ml Liberase TL (Roche Applied Science, Indianapolis, IN) and 0.25mg/ml DNase I for 30 min on a stir plate at 37˚C. Cells were then homogenized in RPMI with 10% FCS. Small intestines were fileted in RPMI with 10% FCS and cut into 0.5cm pieces before washing twice in HBSS with 10% FCS, 25mM EDTA and 2mM HEPES for 15 min on a stir plate at 37˚C. Small intestine were then washed in RPMI 1640 with 10% FCS for 10 min on a stir plate at 37˚C and then digested three times in RPMI with 10% FCS, 0.167mg/ml Liberase TL (Roche Applied Science, Indianapolis, IN) and 0.25mg/ml DNase I for 20 min on a stir plate at 37˚C. After each digestion, the re-suspended cells were quenched with RPMI with 10% FCS at 4˚ until the last digestion was completed. Cells were then homogenized in RPMI with 10% FCS. All tissues except peripheral blood were washed with PBS with 5mM EDTA and resuspended 1:1 with PBS with 5mM EDTA and 100μM Ciaplatin (Enzo Life Sciences, Farmingdale, NY) for 60s before quenching 1:1 with PBS with 0.5% BSA and 5mM EDTA to determine viability as previously described *(64)*. Cells were centrifuged at 500g for 5 min at 4˚C and resuspended in PBS with 0.5% BSA and 5mM EDTA at a density between 1-10*106 cells/ml. Suspensions and blood were fixed for 10 min at RT using 1:1.4 Proteomic Stabilizer according to the manufacturer's instructions (Smart Tube Inc., Palo Alto, CA) and frozen at -80˚C.

Mass-Tag Cellular Barcoding

Mass-tag cellular barcoding was preformed as previously described with the following modifications (Bodenmiller et al., 2012). For each tissue, 1.5*106 cells from each animal were barcoded with distinct combinations of stable Pd isotopes chelated by isothiocyanobenzyl-EDTA in 0.02% saponin in PBS *(65, 66)*. Cells were washed two times in PBS with 0.5% BSA and 0.02% NaN3 and pooled into a single FACS tube (BD Biosciences). After data collection, each condition was deconvoluted using a mass cytometry debarcoding algorithm *(65)*.

Mass Cytometry Staining and Measurement

Cells were resuspended in PBS with 0.5% BSA and 0.02% NaN3 and metal-labeled antibodies against CD16/32 were added at 20μg/ml for 5 min at RT on a shaker to block Fc receptors. Surface marker antibodies were then added, yielding 500 uL final reaction volumes and stained at room temperature for 30min at RT on a shaker. Following staining, cells were washed 2 more times with PBS with 0.5% BSA and 0.02% NaN3 then permeabilized with 4°C methanol for at 10 min at 4°C. Cells were then washed twice in PBS with 0.5% BSA and 0.02% NaN3 to remove remaining methanol, and then stained with intracellular antibodies in 500 μL for 30 min at RT on a shaker. Cells were washed twice in PBS with 0.5% BSA and 0.02% NaN3 and then stained with 1 mL of 1:4000 191/193Ir DNA intercalator (DVS Sciences, Richmond Hill, Ontario, Canada) diluted in PBS with 1.6% PFA overnight. Cells were then washed once with PBS with 0.5% BSA and 0.02% NaN3 and then two times with double-deionized (dd)H20. Care was taken to assure buffers preceding analysis were not contaminated with metals in the mass range above 100 Da. Mass cytometry samples were diluted in ddH2O containing bead standards (see below) to approximately 106 cells per mL and then analyzed on a CyTOFTM mass cytometer (DVS Sciences, Toronto, Canada) equilibrated with ddH2O.

Bead Standard Data Normalization

Just before analysis, the stained and intercalated cell pellet was resuspended in ddH2O containing the bead standard at a concentration ranging between 1 and 2*104 beads per ml as previously described *(35)*. The bead standards were prepared immediately before analysis, and the mixture of beads and cells were filtered through a 35-ml filter cap FACS tubes (BD Biosciences) before analysis. All mass cytometry files were normalized together using the mass cytometry data normalization algorithm freely available for download from www.cytobank.org/nolanlab.

Scaffold Map Generation

The datasets for each tissue were prepared by combining in a single file all the cell events from the different mice. Each sample was clustered independently in 200 clusters using the *clara* function in R. This approach was chosen due to its ability to handle the large numbers of events in mass cytometry data. While density peak-based clustering algorithms could be incorporated into this pipeline in the future, leading approaches *(67)* require an explicit computation of a distance matrix and the application of a heuristic for identifying density peaks that did not translate well to the nature of mass cytometry data (Figure S16). For the bone marrow, a graph was constructed by first connecting together the nodes representing the manually gated landmark populations and then connecting to them the nodes representing the cell clusters. Each node is associated with a vector containing the median marker values of the cells in the cluster (blue nodes) or gated populations (red nodes). Edge weights were defined as the cosine similarity between these vectors after comparing the results from the implementation of several distance metrics (Figures S17-S19)*(20)*. Edges of low weight were filtered out. We experimented with different threshold values for the weights and we found values of 0.8 for the initial subgraph of landmark nodes, and 0.7 for the complete graph to produce satisfying results. The graph was then laid out using an in-house R implementation of the ForceAtlas2 algorithm from the graph visualization software Gephi (*68*). To overlay the additional samples on the bone marrow map the position and identity of the landmark nodes was fixed and the clusters of each sample were connected to the landmark nodes as described above. Once again the graphs were laid out using ForceAtlas2 but this time only the blue nodes were allowed to move. To determine the reliance of the method on the high-dimensional nature of the data, we additionally generated Scaffold maps using only 30 or 20 protein parameters to (1) generate the entire map or (2) map unsupervised clusters onto the landmarks defined by all measured protein parameters. These results are displayed in Figure S20.

Distance Distributions for Statistical Tests

For the statistical comparison of populations across samples (i.e. Liver versus Bone Marrow Macrophages in Figure 2g), the cell clusters were classified according to their nearest Landmark population for each sample. Cells belonging to clusters classified into the population of interest were pooled for each sample, and cosine similarity was calculated from the Landmark position for each cell. A summary statistic was calculated for the distribution belonging to each sample (in the case of the above example, we chose the median). These values were the compared across the groups of interest i.e. Liver samples versus Bone Marrow samples) by a Wilcoxon rank sum test to test the hypothesis of interest.

Unsupervised Force-Directed Graph Generation

Cells were manually gated as Live CD45+ lineage- (Ter119, Ly6G, Siglec-F, CD19, B220, CD3, TCRβ, DX5, F4/80, CD11c, PDCA-1, FcεR1α) except for markers as noted. The gated cell populations for each tissue were clustered independently in 50 clusters using clara in R. The clusters for all the tissues were combined in a single graph with edge weights defined as the cosine similarity between the vectors of median marker values of each cluster. All the pairwise distances were calculated and for each node only the 10 edges of highest weight were retained. The graph was then laid out using the ForceAtlas2 algorithm in Gephi.

Fluorescence-Based Flow Cytometry

Lung single cell suspensions were prepared as previously described. Cells were incubated with 15µg/ml FcX Block (Biolegend) for 5 min at 4˚C to block Fc receptors and then stained with the following fluorochrome-conjugated antibodies and DAPI (Life Technologies) in 100µl reactions in PBS with 0.5% BSA and 5mM EDTA for 20 min at 4˚C: CD45.2 - Pacific Blue, SiglecF - AlexaFluor 647, CD11c - APC-Cy7, CD19 - FITC, CD49b - PerCP-Cy5.5, B220 - eFluor650, F4/80 - PE, CD8 - PE-Cy7. Cells were washed in PBS with 0.5% BSA and 5mM EDTA and centrifuged at 300g for 5 min at 4˚C two times and analyzed on an LSR II flow cytometer (BD Biosciences).

Supplementary Text

Additional Comments on Scaffold Maps

On Node Sizes

As a result of the clustering algorithm, the clusters do not contain an equal number of cells, which is the case for most clustering methods. In the paper, we have shown plots with clusters scaled in size to reflect the relative number of cells in each. However, since larger clusters can occasionally obscure smaller ones, an option in the interface exists to scale all equally such that each cluster can be fully viewed. Ultimately, static pictures of such complex datasets are a necessary compromise between comprehensiveness and interpretability. For this reason, in the interface, we make it easy to toggle between these different scaling options.

On the Inclusion of Landmark Nodes

The inclusion of landmark nodes is meant to facilitate and anchor the interpretability of the analysis when assessing global immune system organization. For instance, an unsupervised analysis of the dataset is presented **(Fig. S14)**. While the structure can be observed, it is not readily relatable to current knowledge. Nevertheless, these approaches are not mutually exclusive. We find utility in these unsupervised graphs when limiting our analysis to a single cell population **(Fig. 6)**. However, for global or otherwise broad analyses, the main advantages of incorporating landmarks are two-fold. Landmarks (1) provide points of reference for intuition during data analysis, facilitating the interpretation of the mapping results and (2) provide a logical way to incorporate new data files, making the reference framework extensible moving forward. Without landmarks, there would be no way to intuitively incorporate new data other than aligning cell clusters to their nearest neighbor in a prior analysis. Doing so would necessarily bias the analysis and prevent the discovery of new cell states by constraining the realm of possibilities to those observed in a prior experiment. By using landmarks, the data from independent samples can be oriented with respect to these reference points, but new data are not confined to the space occupied by the prior data.

The incorporation of landmark nodes does not, in any way, impact the clustering. The clustering is performed independently for each sample as the first step of the analysis with absolutely no input from the manually defined populations. The landmarks are later then used to determine distance from some standard. Using the bone marrow as the reference for this particular study was decided upon because the bone marrow is the only organ that contains most developing immune cells as well as mature immune cells. Therefore, it is well suited for defining the landmark populations for these analyses. As a comparison, we show maps for which the spleen is used as the reference **(Fig. S5)**. The organization is strikingly similar to that of the bone marrow maps, but no progenitor populations can be identified in this organ. Therefore, when viewing the bone marrow data, cell clusters containing these cells early in development are connected by longer lines and have no intuitive place to localize in the map. Along these lines, it is important to note that for studies limited to one organ (or more generally, one type of sample), a new reference can easily be defined using populations from that particular tissue of interest. As an example, the circadian rhythms in the lungs can alternatively be viewed using Scaffold maps with landmark populations derived from the lungs instead **(Fig. S7)**.

On Choosing a Clustering Algorithm

We chose to adopt the CLARA clustering algorithm for a few reasons. One is practicality for the data at hand: The algorithm can deal with very large datasets, which is something that neither hierarchical clustering nor more sophisticated methods requiring explicit computation of a distance matrix, such as that proposed recently by Rodriguez and Laio *(67)*, can do.

Two important issues in the evaluation of a clustering method include:

1) Defining the number of "true" clusters in the data
2) Assessing the quality of the resulting clusters

In terms of assessing the quality of the clusters themselves, this is a hotly debated topic in the literature *(29)*. Defining what constitutes a successful clustering is exceptionally difficult in general, but even more so for flow cytometry data, where there exists no general consensus on what defines a distinct cell population since function might change across a spectrum of values that themselves are composed by multiple protein marker changes. Unfortunately, the immune repertoire does not come in neatly defined bins, much as clustering algorithms have shown.

One of the reasons we initiated the development of Scaffold maps was precisely to sidestep these issues by providing investigators with an effective way to browse single-cell data. Arriving at the "correct" number of clusters is not required to meet this objective. Notably, the biological function of cells may not strictly partition by regions of peak density with the markers used for the clustering, and the measure of density will be dependent on the parameters measured. We therefore chose not to risk oversimplifying the analysis by establishing the expectation that any clustering algorithm will perfectly divide cells according to their biological behavior and function. Rather, we use clustering as a way to alleviate the inherent redundancy of the data and make the visual representation more manageable. In the past, when developing the algorithm on smaller datasets, we have also used hierarchical clustering without appreciable differences in the resulting maps.

In comparison, Rodriguez and Laio's clustering method based on density peaks is based on "a heuristic criterion for finding independent density peaks" *(67)*. Essentially this heuristic involves manually examining a plot of local density versus distance from all other points for all the observations in the dataset. Cluster centers should appear as isolated points of high density (i.e. high local density, large distance from other observations). We have tried to apply this heuristic to our data (albeit with some difficulties, as elaborated upon below) but did not get an answer as clear-cut as the ones presented in Rodriguez & Laio's paper. Depending on how the plot is interpreted, the number of clusters appears to be around 10-20 **(Fig. S16)**, which is unreasonably low given the fact that we can manually identify at least 20 populations from the data. Based on these results, it would seem that a straightforward application of the heuristics does not yield satisfactory results on high-dimensional flow cytometry data. It is absolutely possible that specific modifications of this method could be developed for the analysis of this kind of data, with potentially good results. As noted, the modular construction of the Scaffold maps algorithm would enable the incorporation of other clustering methods quite easily in the future.

Unfortunately, algorithms that require an explicit computation of the distance matrix could not handle the amount of data we have been able to cluster with CLARA—our average data file contains over 1-5*10^6 observations. It is possible that an adaptation could be eventually developed that allows the analysis of very large datasets with such methods, including Rodriguez & Laio's method. Future developments in density-peak clustering methods could easily be incorporated into the Scaffold maps pipeline moving forward.

On the Choice of Similarity Metric

A key question in developing the method was the selection of the similarity metric. We explored the use of cosine and angular similarities, as well as Euclidean, Manhattan and Mahalanobis distances, which were converted to similarities according to the formula similarity = (1 / (1 + distance)). We also considered measures based on a discretization of the data (e.g. Jaccard), but the complex shapes of the distributions observed in real data proved extremely difficult to discretize in a satisfactory and meaningful manner.

The best distance metric to use in a given application strongly depends on the specific characteristics of the data at hand. In this case, one logical way to investigate the effect of different distance metrics is to analyze how similar (or different) the twenty manually identified populations are under said metrics **(Fig. S17)**. Interestingly, cosine, angular, Euclidean and Manhattan similarities all behave very similarly in this analysis. If we define the population centroid as the median of the 39 marker values, and we then cluster the centroids using these different metrics, we obtain very similar groupings. For instance the different types of T cells form a separate group, as do the myeloid cells and the various progenitor populations. The Mahalanobis similarity gives slightly different results, with NKT and γδ T cells forming a separate group together with NK cells, and GMP clustering with monocytes and eosinophils.

The first step before running the layout algorithm is to filter the list of pairwise distances in order to retain only the highest-scoring edges. When using other metrics it was necessary to adopt a percentile-based filtering because the different similarity measures result in a different distribution of similarities **(Fig. S18)**. We retained the top 18% of similarity values, which results in a tractable but robust total number of edges for interpretation.

Having looked at the behavior of the distance metrics *per se*, it was then necessary to consider how these interact with the force-directed algorithm, which is responsible for translating the similarities in the final shape of the Scaffold map. In the ForceAtlas2 implementation the similarity between two connected nodes is transformed in attraction strength according to a parameter called "edge weight influence." This parameter acts as an exponent, so that the final attraction strength is calculated as F = (similarity) ^ (edge weight). Since all the similarity values are in the range [0, 1], higher edge weights have the effect of “compressing” the distribution of similarity scores. Using Euclidean, Manhattan and Mahalanobis metrics resulted in a much narrower distribution of similarities than did cosine similarity, while converting cosine similarity to angular similarity resulted in an intermediate situation **(Fig. S18)**. To account for this difference, we used an edge weight parameter of 1 for Euclidean, Manhattan and Mahalanobis metrics and 6 for angular similarity. It is important to note that this adjustment is only necessary because of the specific way the layout algorithm converts similarities to 2D distances—an inherently ad-hoc process intended to give visually pleasing results—and does not represent a fundamental difference in the behavior of the distance metrics.

Once these adjustments are made, Euclidean, Manhattan and Mahalanobis metrics result in similar maps, which are visually more compact compared to angular and cosine, which yield essentially identical results **(Fig. S19)**. While the results are similar we think that the less compact maps (implementing cosine or angular similarity) makes them easier to interpret. These maps have more “resolution,” probably a result of the fact that the cosine and angular distance metrics result in a broader distribution of similarity values **(Fig. S18)**.

We would also like to offer the following, more general considerations. First, this metric scales easily and efficiently into high-dimensional spaces. Because the metric is analogous to correlation, it offers an attractive compromise by considering the Boolean expression of any particular parameter (positive/negative) while also allowing its absolute magnitude of expression to factor into the similarity metric. Moreover, contrary to Euclidean distance, cosine similarity does not strongly favor dimensions with higher scales. This is a very attractive property because the expression of a specific marker can be crucial in defining a cell type even if its absolute expression value is lower compared to other parameters. For similar reasons, our laboratory chose to utilize cosine distance in a different, recent publication describing a new approach to define a linear differentiation path in single-cell data *(20)*. Cosine similarity is a more robust metric when mapping archival datasets, where measured values for the same marker can have vastly different numeric ranges. Cosine similarity is an ideal metric in this context because it is scale-independent, and the scale of the actual measured values can vary across different cytometers, fluorophores, antibody clones etc.

However, the system is meant to provide a community-organizable approach for bringing multiple datasets (and a modular approach for other metrics) to a framework for comparison. Any of the organizing or clustering algorithms that lay behind the resulting output are “modular” insofar as another researcher might decide that for their purposes a different landmark set, a different distance metric, or a different clustering approach would best represent the data.

On Metrics and the Utility of Multidimensional Distance

Enumeration of cell types, which is a utility of Scaffold maps, can be accomplished by manual identification through hand gating based on canonical phenotypic markers. Our method instead utilizes the similarities across *all measured dimensions* to compare any group of cells to well-established immune cell types (the so-called landmark nodes). This therefore leverages all of the information collected in order to make this determination rather than merely relying on the expression of the 1 or 2 canonical makers on which a manual analysis is based. For instance, the fidelity of some canonical lineage markers depends on the tissue of origin (as an example macrophages in the alveoli of the lung express the DC marker CD11c and the eosinophil marker Siglec-F) as well as the activation state. We therefore believe that utilizing multidimensional similarity provides distinct advantages. Moreover, while manual analysis of data requires expertise, is subjective and is laborious, the Scaffold maps algorithm increases the consistency, objectivity and throughput of data analysis while providing clusters that could be compared using appropriate metrics.

Most importantly, manual gating provides no information regarding the cellular expression of parameters that are not utilized for the identification of a particular cell type. In contrast, by plotting distance as similarity in Scaffold maps, this information is intrinsically reflected in the position of cell clusters with respect to the classical landmark cell populations. By enabling (1) the graph to be colored based on the expression of different parameters, (2) comparing the distribution of any parameters of interest for a cell cluster compared to its nearest landmark population(s) as histograms, and (3) the generation of unsupervised force-directed graphical landscapes for any population of interest, we provide several tools to discern why certain cells position themselves in their respective location in the Scaffold maps.

We also utilized distance in multidimensional space as a metric by which populations can be compared across samples for statistical inference. Multiple valid approaches to quantifying elements of graphical networks exist. Our idea with a reference map of this nature is to provide a framework and an obvious metric, but leave open the opportunities for other metrics to be applied. A straightforward metric that we have used is the distance of cells from a landmark as a means to compare differences in cell populations across samples. By building a distribution of distances for all cells closest to any given landmark, summary statistics of multidimensional distance (medians, percentiles, etc.) can be used to test whether the characteristics of any population are significantly different across sample types. For example, we used such an analysis to arrive at the conclusion that macrophages in the liver are significantly different from the macrophages in the bone marrow in their phenotypic characteristics **(Fig. 2G)**.

On Data Dimensionality and Effective Mapping

With respect to varying the number of proteins utilized for the mapping, two examples are included in the manuscript **(Fig. 5)**. The first of these shows the effect of including 15 proteins in the mapping of cells from the peripheral blood. The plot using all measured parameters is included as Panel A while the plot using only 15 is Panel C. Notably, the results are fairly similar, suggesting that there is some redundancy encoded in the biological system.

This is pushed further by using the information from only 8 proteins in the bottom of Fig. 5. We compare the organization of the bone marrow using all information from the experiment **(Fig. 5D)** to the organization resulting from only 8 proteins **(Fig. 5F)**. Comparing the results of these two graphs, it is apparent that the global organization is somewhat conserved. However, in Panel F, the graph using only 8 dimensions lacks resolution in both the myeloid region and the progenitor region of the graph. Because not enough information is included to distinguish between several related cell types, the nodes containing these cells collapse into center of these “zones,” unable to distinguish which specific population they belong to.

We also include the results of dropping the bone marrow analysis to 30 or 20 parameters used to map unsupervised clusters **(Fig. S20)**. These plots can be located on the left hand side of the figure, where the location of the landmark nodes is fixed from the original analysis, but the unsupervised clusters are mapped using fewer dimensions. The effect of using the lower-dimensional data to establish the localization of both the landmark and unsupervised nodes can be found in the right hand side of the figure. In these maps, reducing the information does influence the localization of the landmarks with respect to one another, especially in lower-dimensional space (≤20 parameters), emphasizing the advantage of utilizing higher-dimensional data to establish the initial map of the reference sample. That said, the difference in the map organization above this number of parameters is fairly minimal, strongly suggesting that the 39 dimensions we measure are sufficient to devise a robust organization. Notwithstanding, the maps still accurately represent the relationships between unsupervised cell clusters and landmark populations given the amount of information provided in the lower-dimensional space.

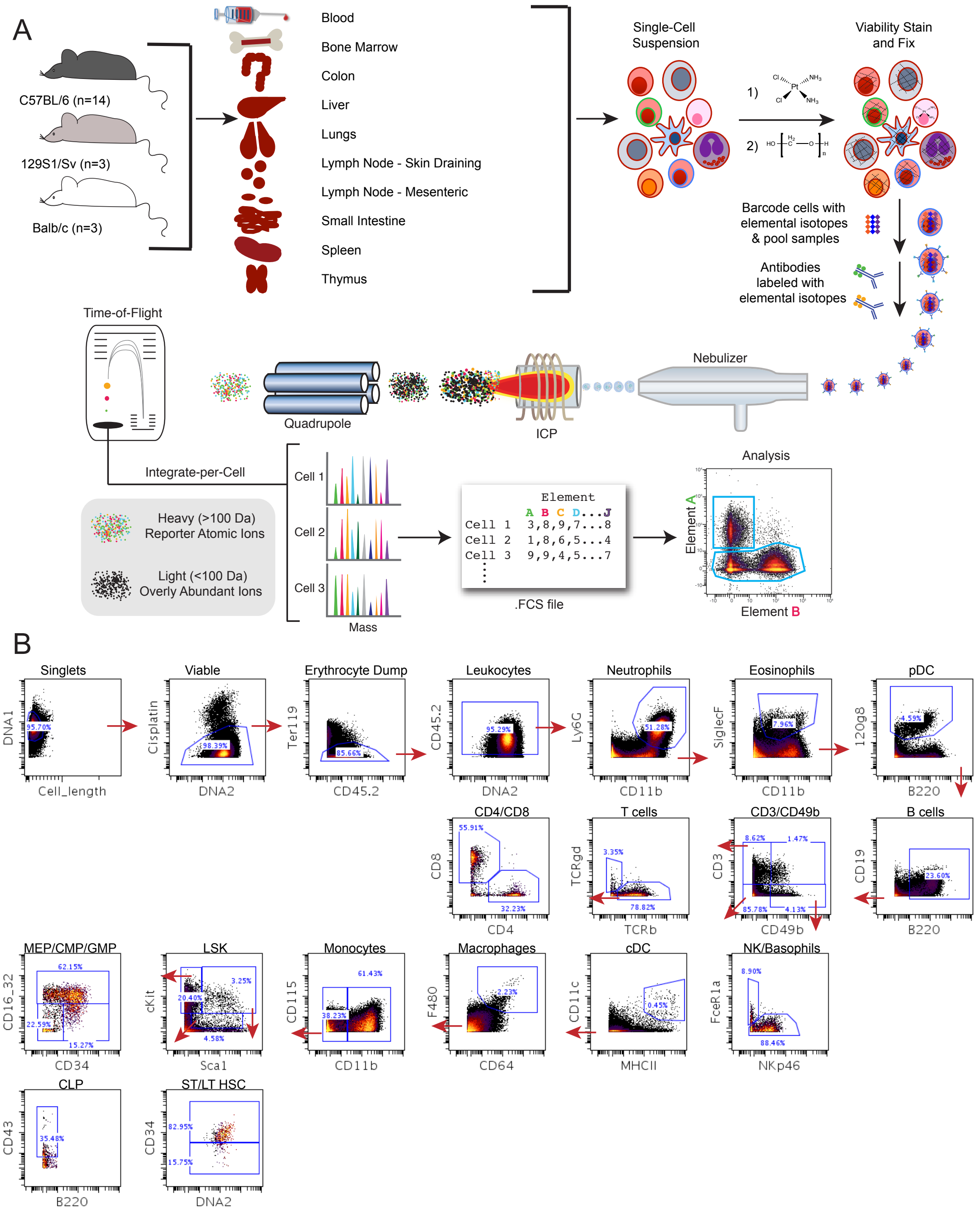


Fig. S1.

Mass cytometry experimental design and workflow.
(A) Model workflow for mass cytometry reference map experiment as described in the Materials and Methods.
(B) Typical gating strategy for murine bone marrow to define landmark populations for Scaffold maps.

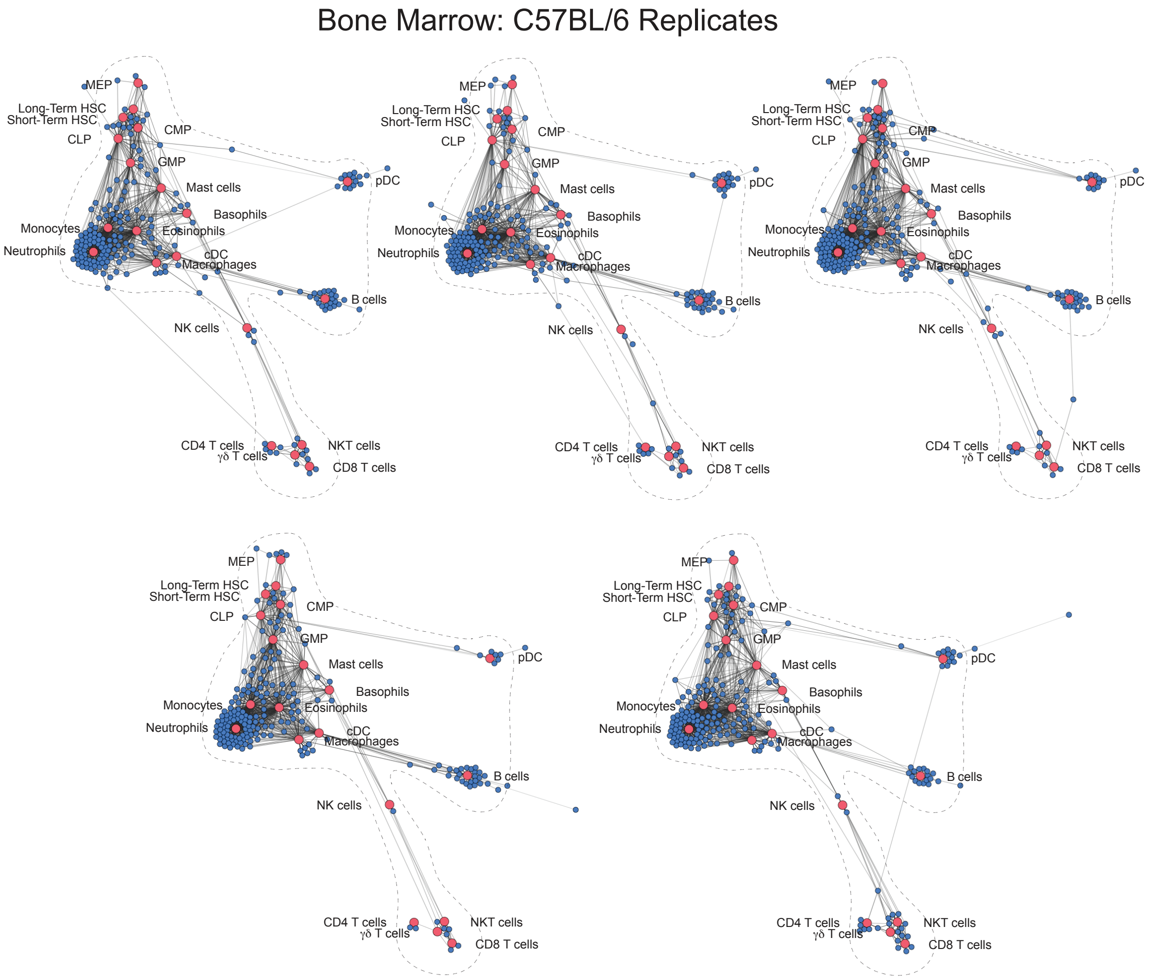


Fig. S2

Bone marrow Scaffold maps from individual C57BL/6 mice. Representative Scaffold maps are shown using clusters from bone marrow samples from individual C57BL/6 mice.

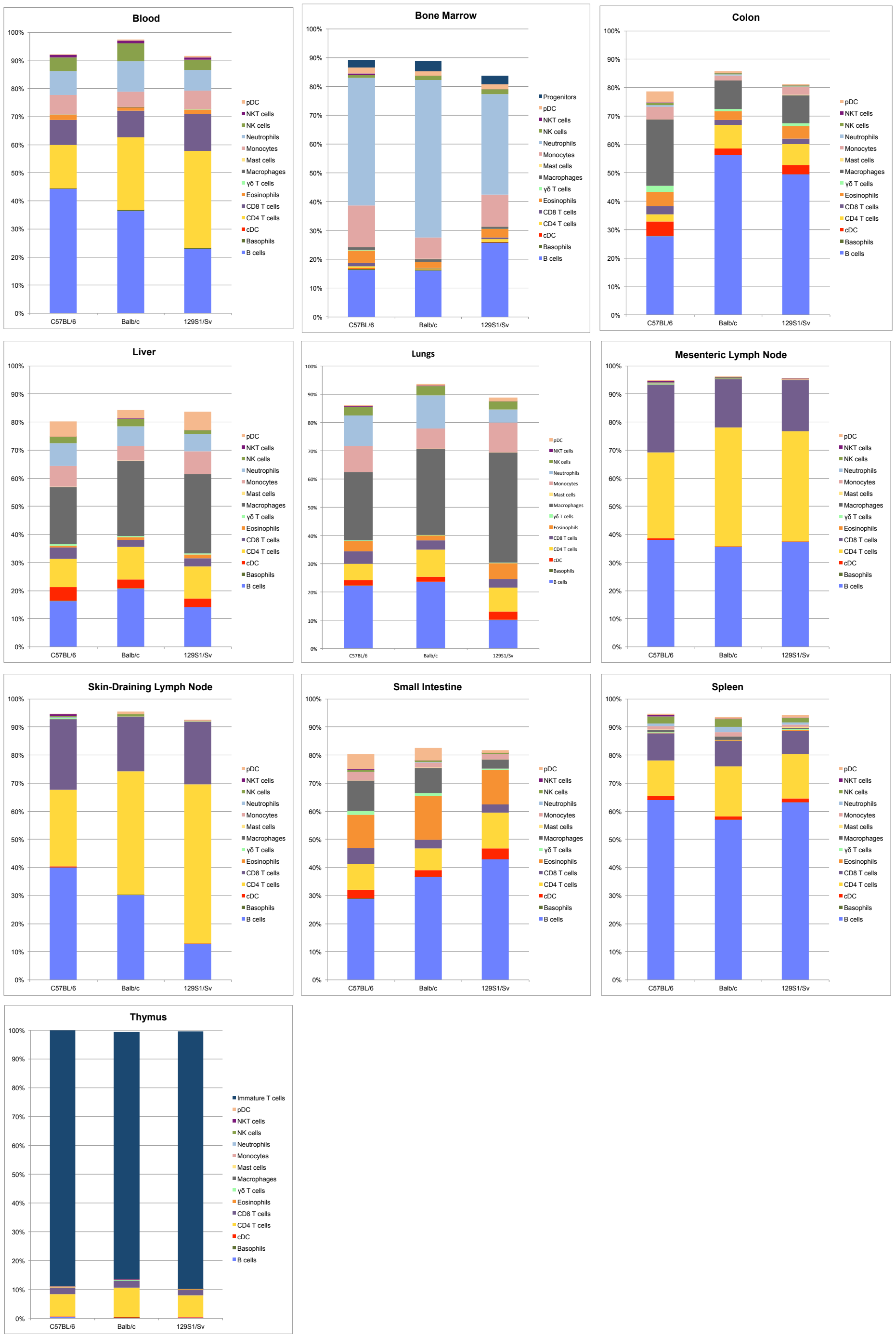


Fig. S3

Immune composition by tissue of origin. Median frequencies of each major cell population by tissue of origin for C57BL/6, 129S1/Sv and Balb/c mice as a percent of total leukocytes (Live Single Ter119-CD45+ cells).

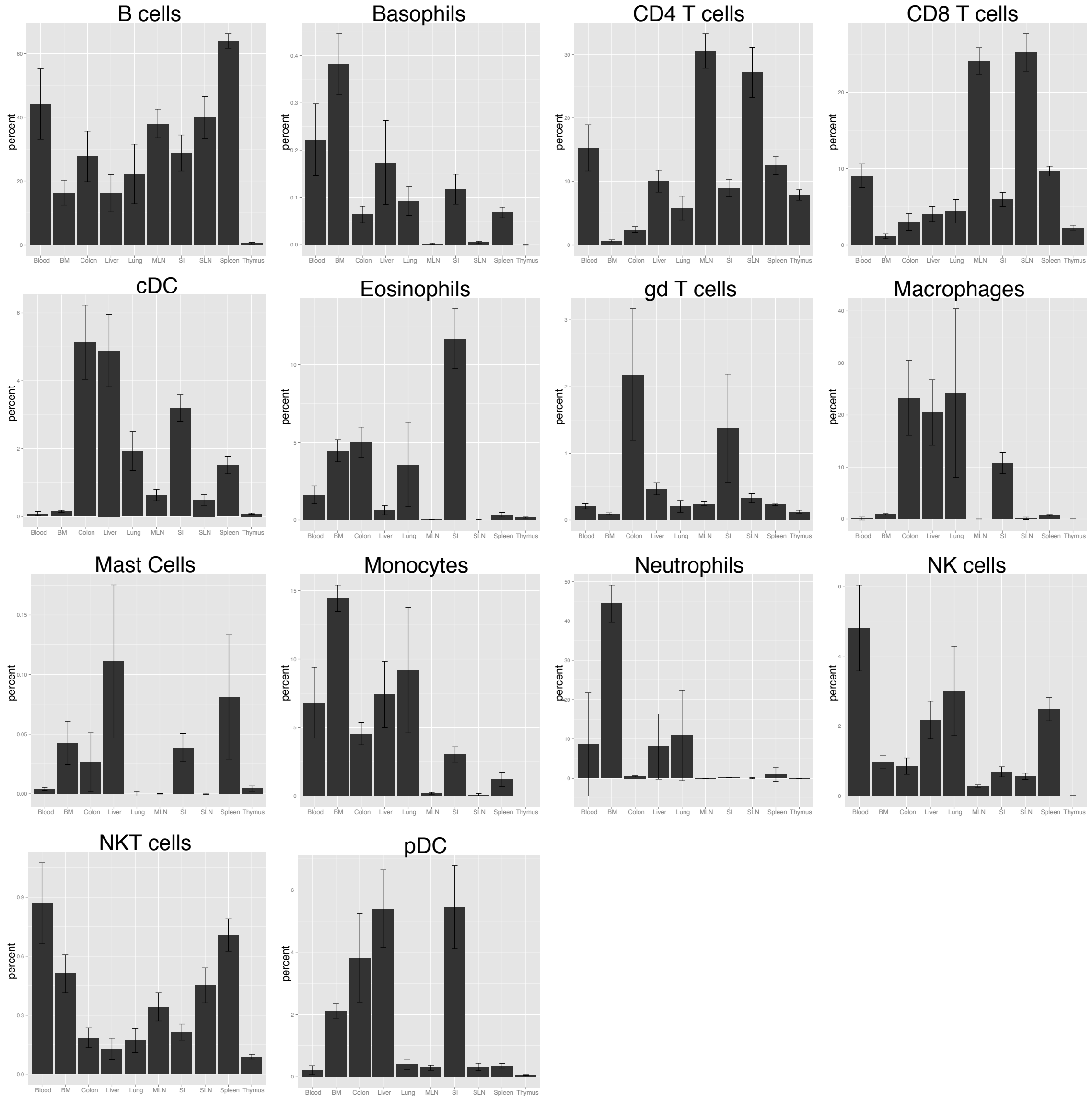


Fig. S4

Immune cell population frequencies by tissue of origin. Mean frequency and standard deviation of every major immune cell population across all tissues surveyed by mass cytometry in C57BL/6 mice (4 independent experiments, n=3-4 for each).

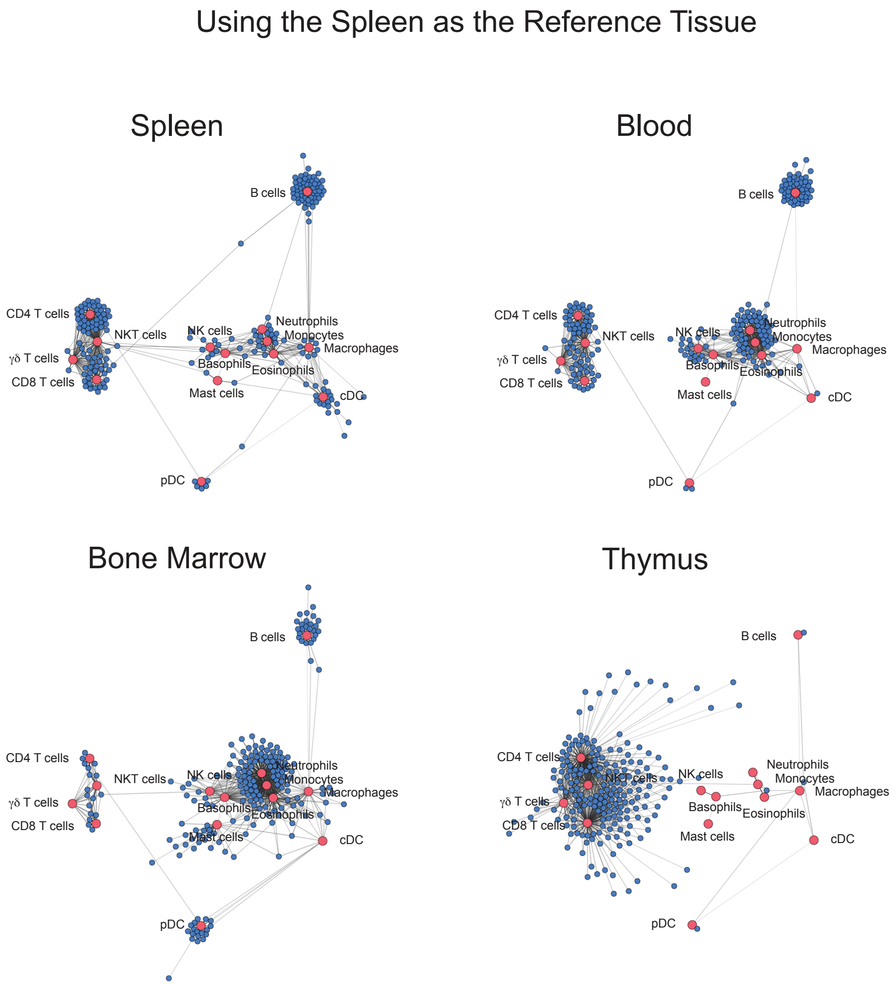


Fig. S5

Scaffold map analysis using populations from the spleen as the Landmark nodes. Maps shown are for C57BL/6 mice.

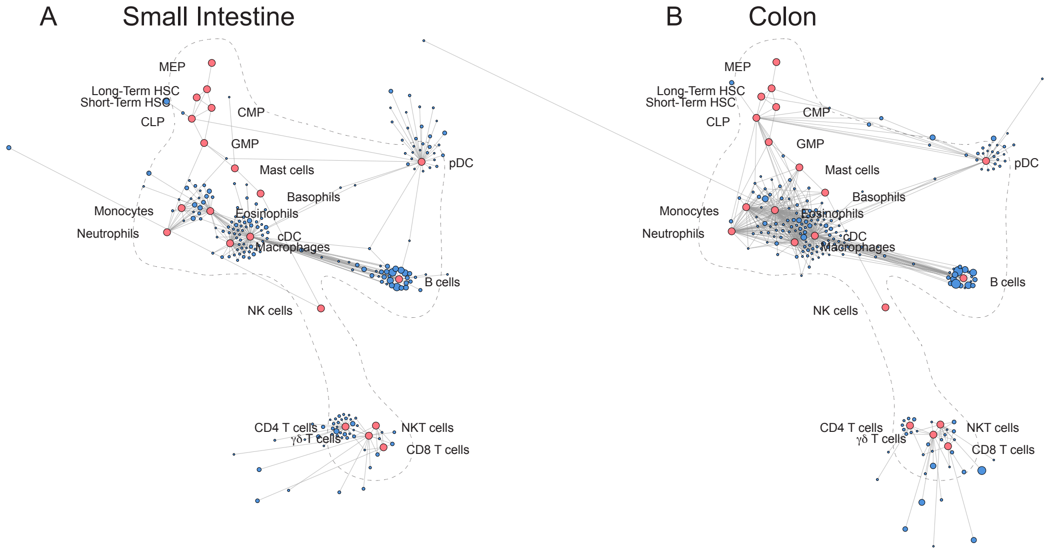


Fig. S6

Mapping immune organization in the gut. Scaffold maps for the gastrointestinal tract from C57BL/6 mice using bone marrow as the reference sample to define landmark nodes (red). (A) Small Intestine (B) Colon.

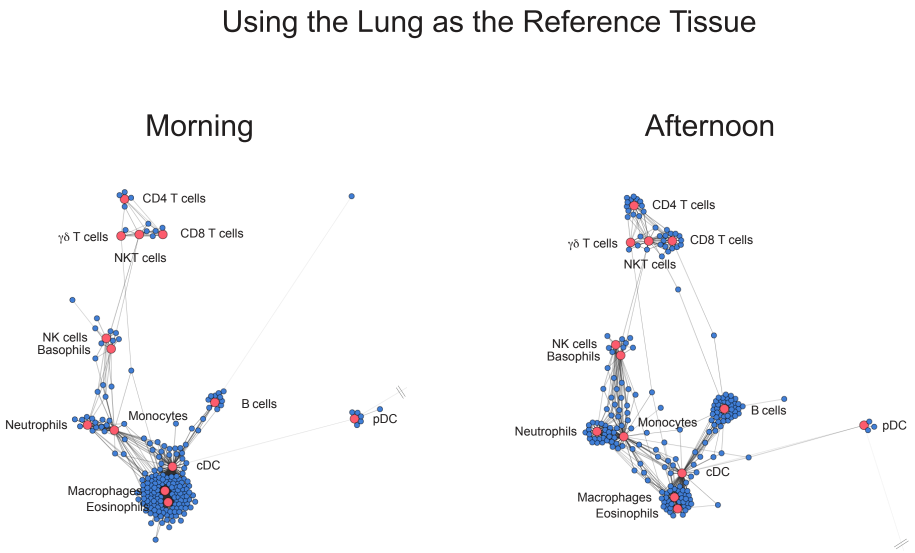


Fig. S7

Circadian remodeling of immune organization in the lungs viewed on Scaffold maps built using lung immune populations as the Landmark nodes.

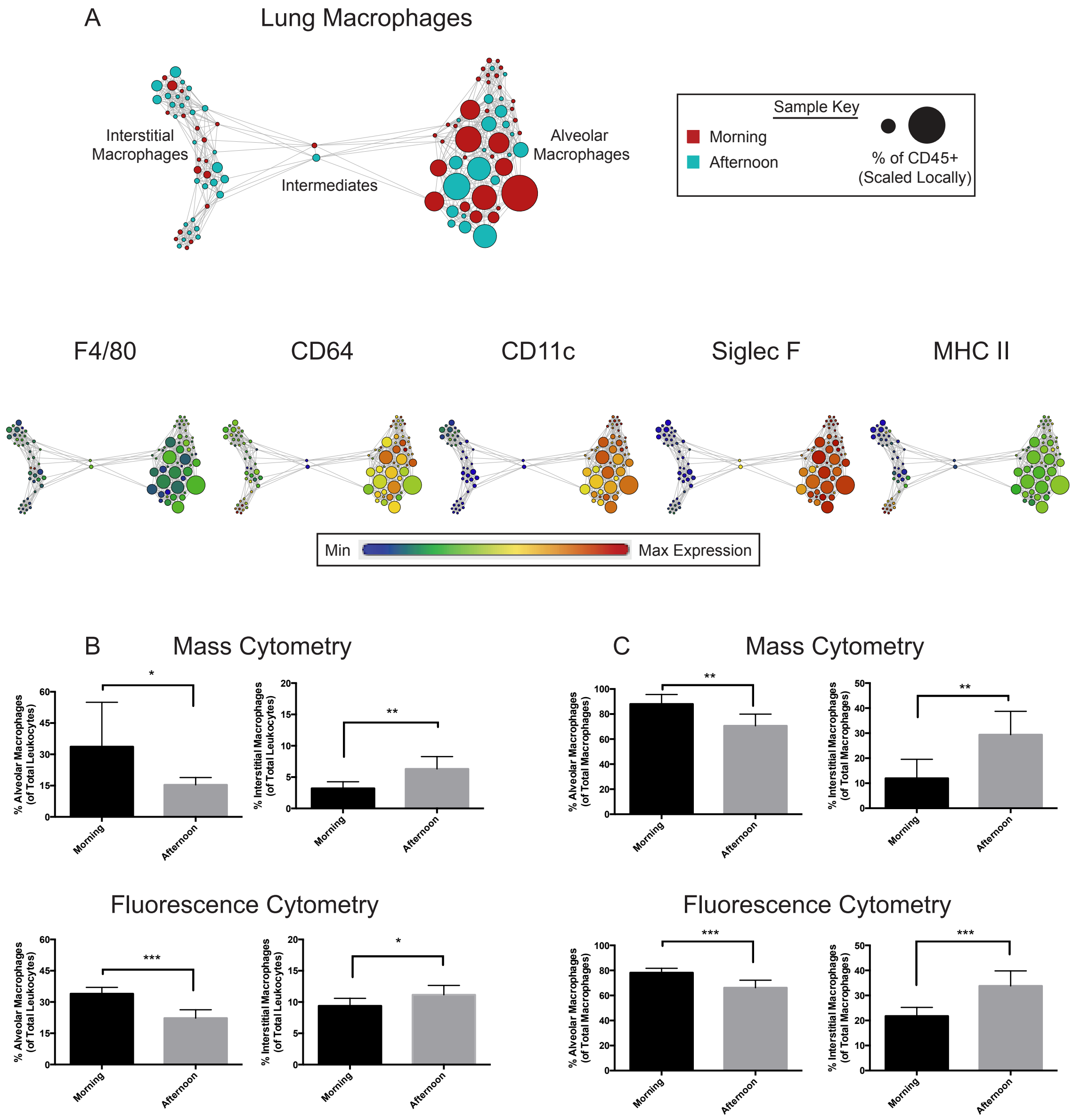


Fig. S8

Macrophage sub-types in the lung are differently remodeled according to circadian rhythms. (A) Unsupervised force-directed graph of macrophages from the lung colored by (top) time point of collection or (bottom) expression of the indicated protein. (B) Quantitation of alveolar and interstitial macrophages as a frequency of total leukocytes (CD45+) or (C) total macrophages (F4/80+CD64+) from the original mass cytometry dataset (n=7 morning and afternoon) and a follow-up fluorescence experiment (n=7 morning; n=8 afternoon). Bars represent mean ± SEM, and p-values result from two-sided t-test. * denotes $p < 0.05$; ** denotes $p < 0.01$; *** denotes $p < 0.001$.

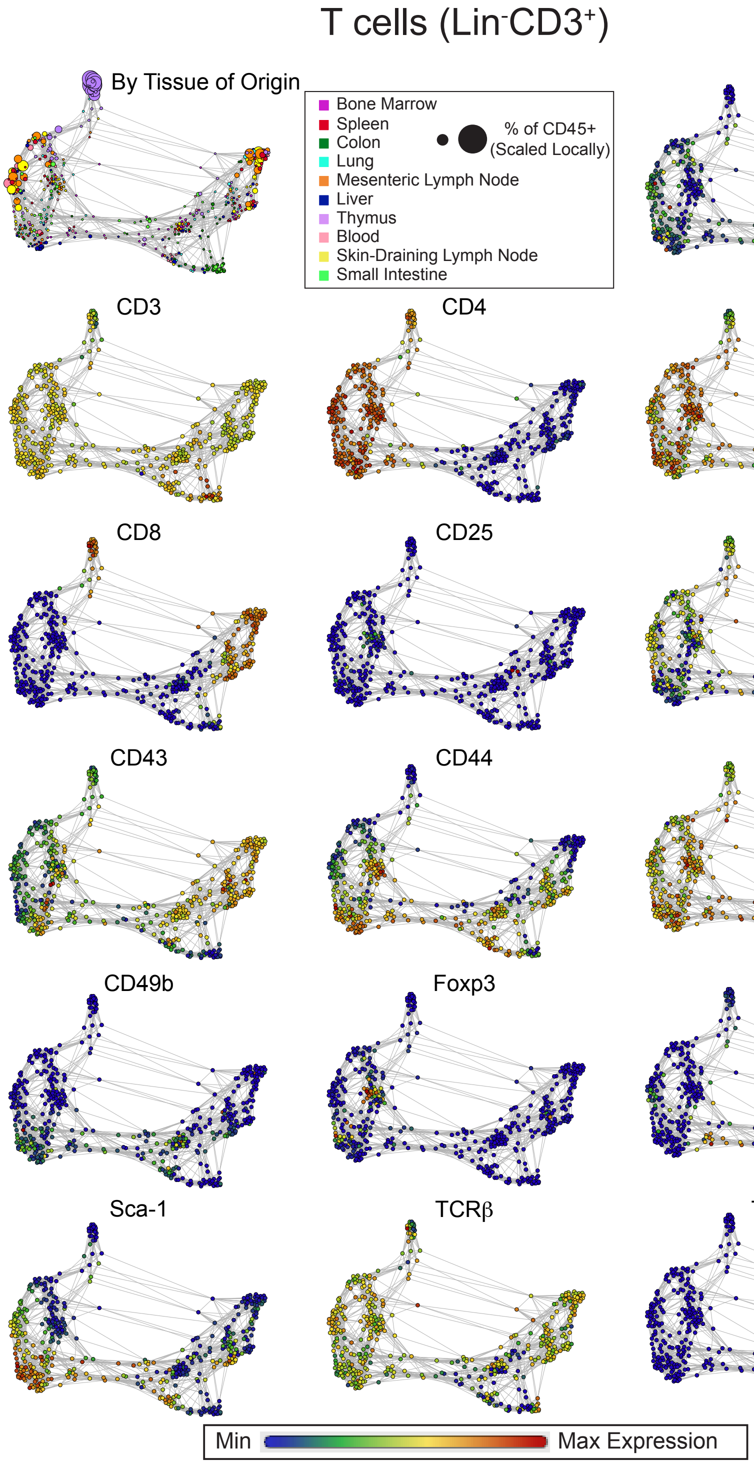


Fig. S9

Systemic T cell landscape colored by tissue of origin or by the dynamic range of the phenotypic marker as noted.

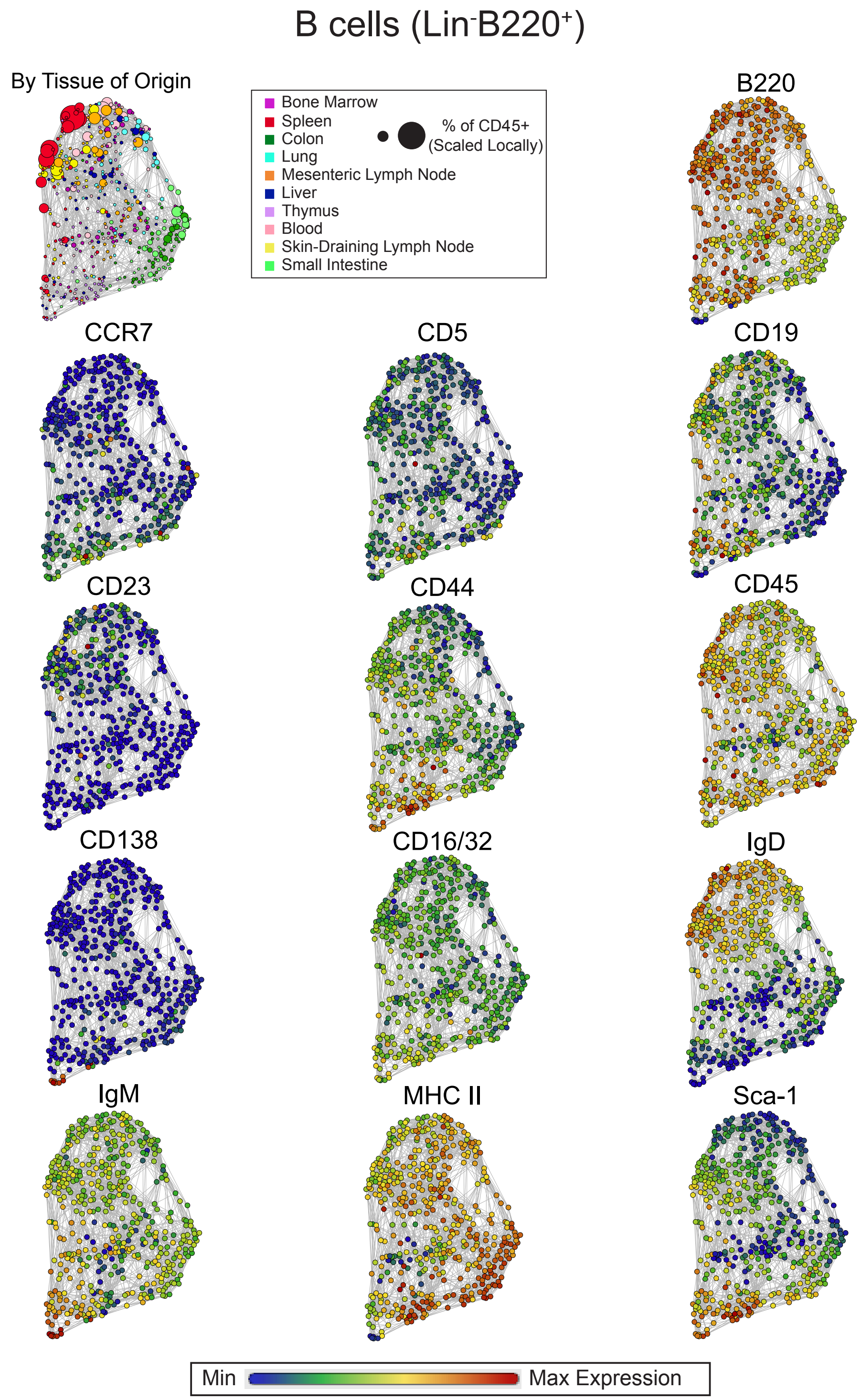


Fig. S10

Systemic B cell landscape colored by tissue of origin or by the dynamic range of the phenotypic marker as noted.

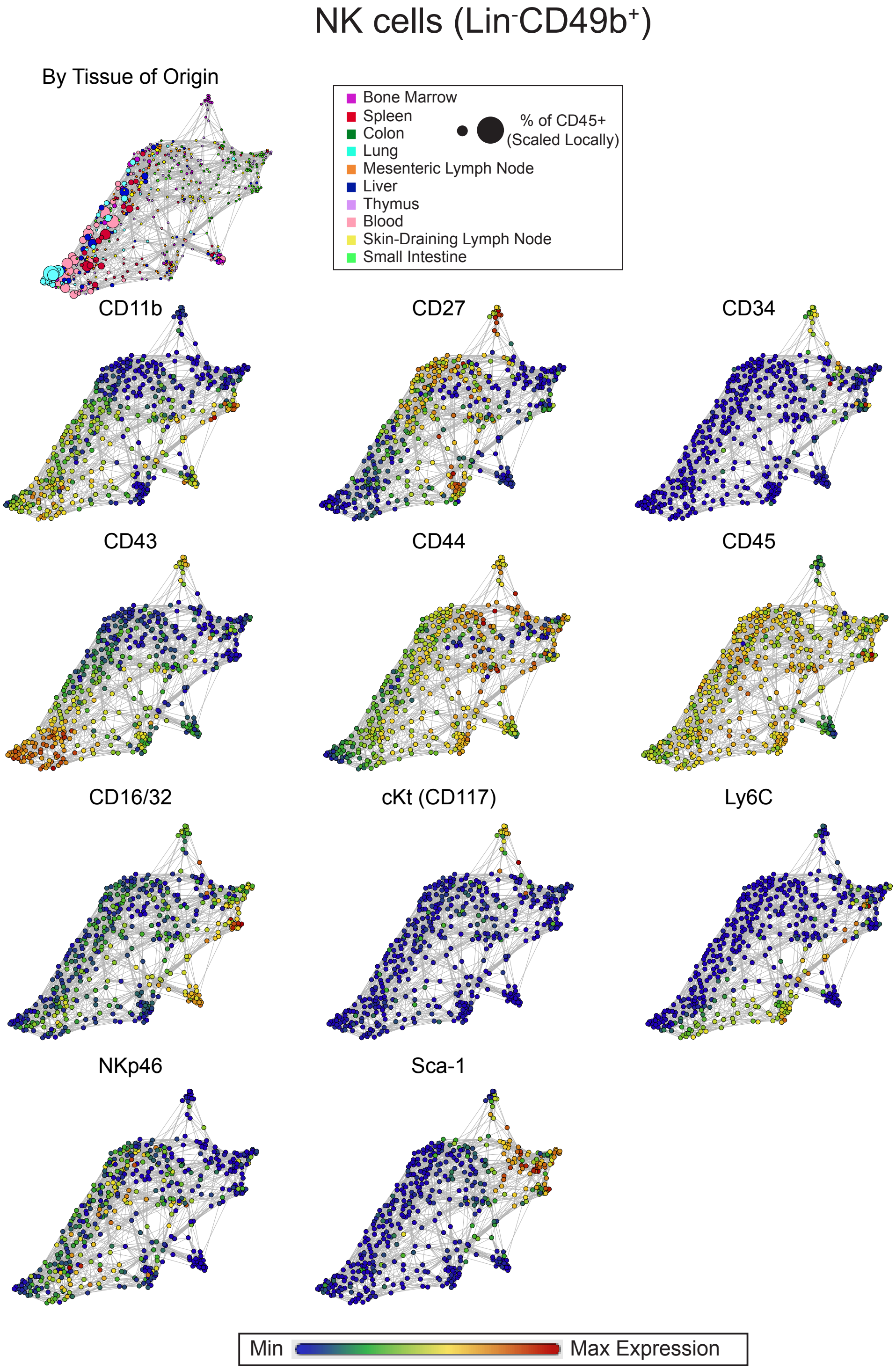


Fig. S11

Systemic NK cell landscape colored by tissue of origin or by the dynamic range of the phenotypic marker as noted.

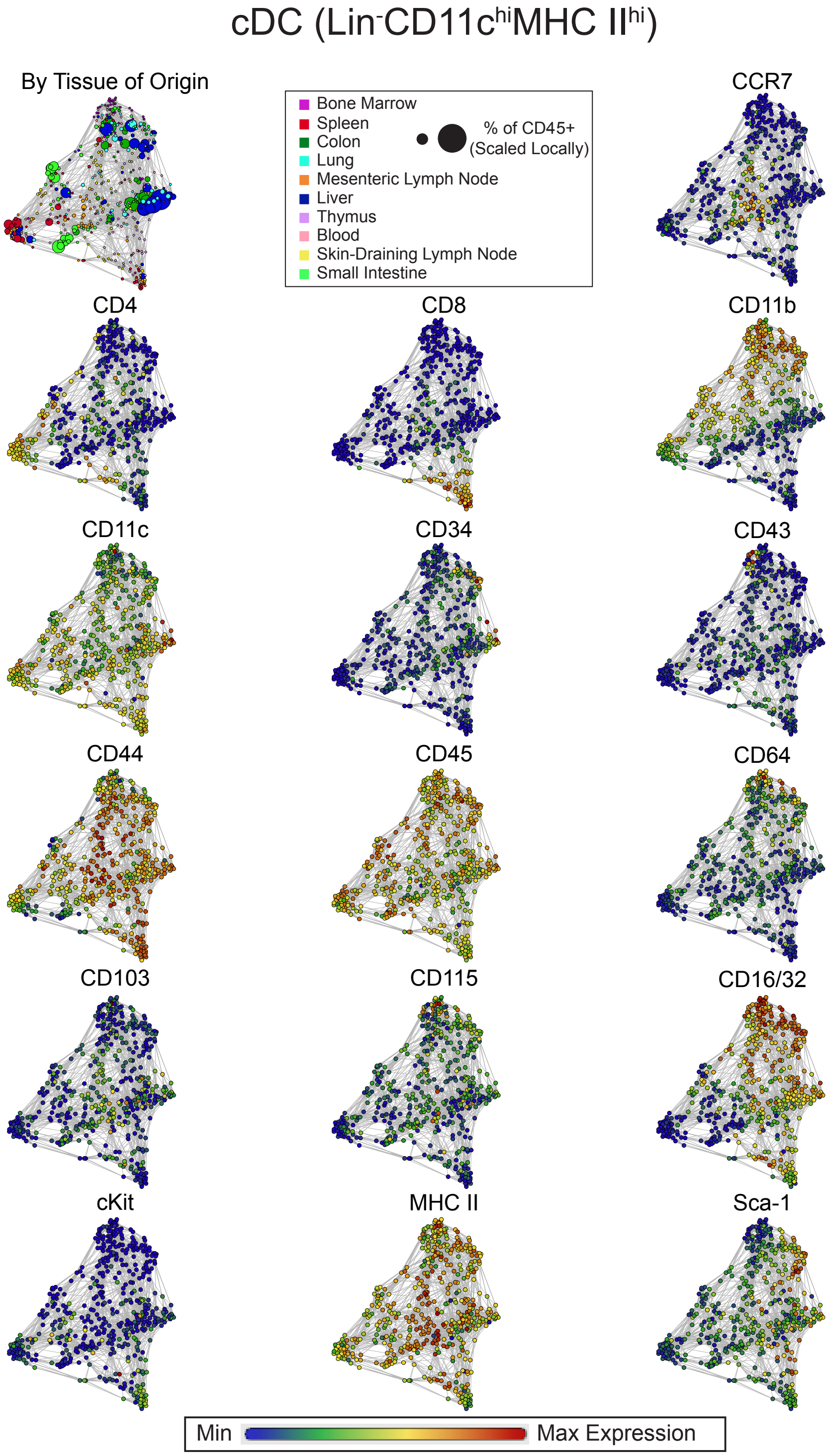


Fig. S12

Systemic conventional dendritic cell landscape colored by tissue of origin or by the dynamic range of the phenotypic marker as noted.

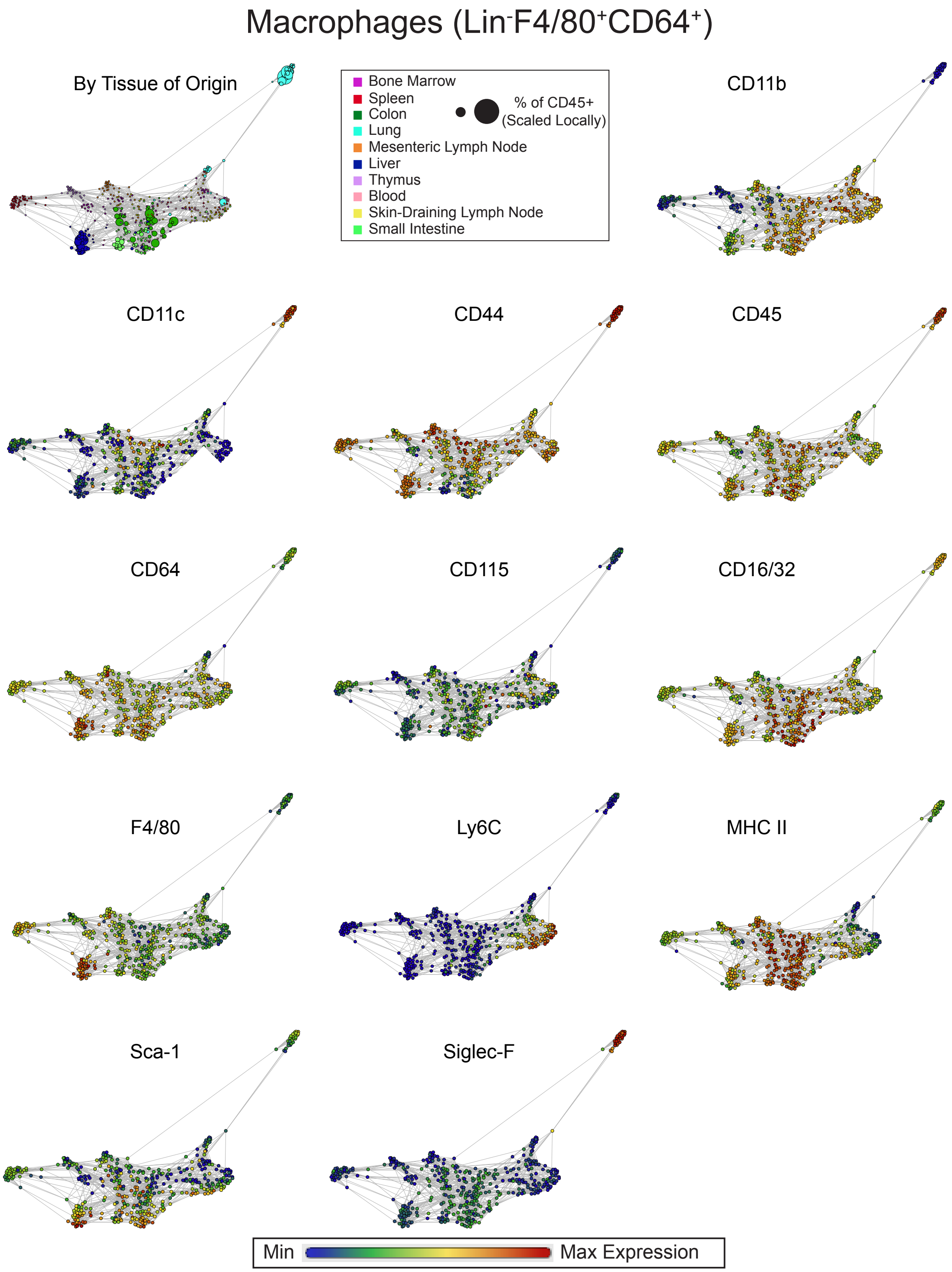


Fig. S13

Systemic macrophage landscape colored by tissue of origin or by the dynamic range of the phenotypic marker as noted.

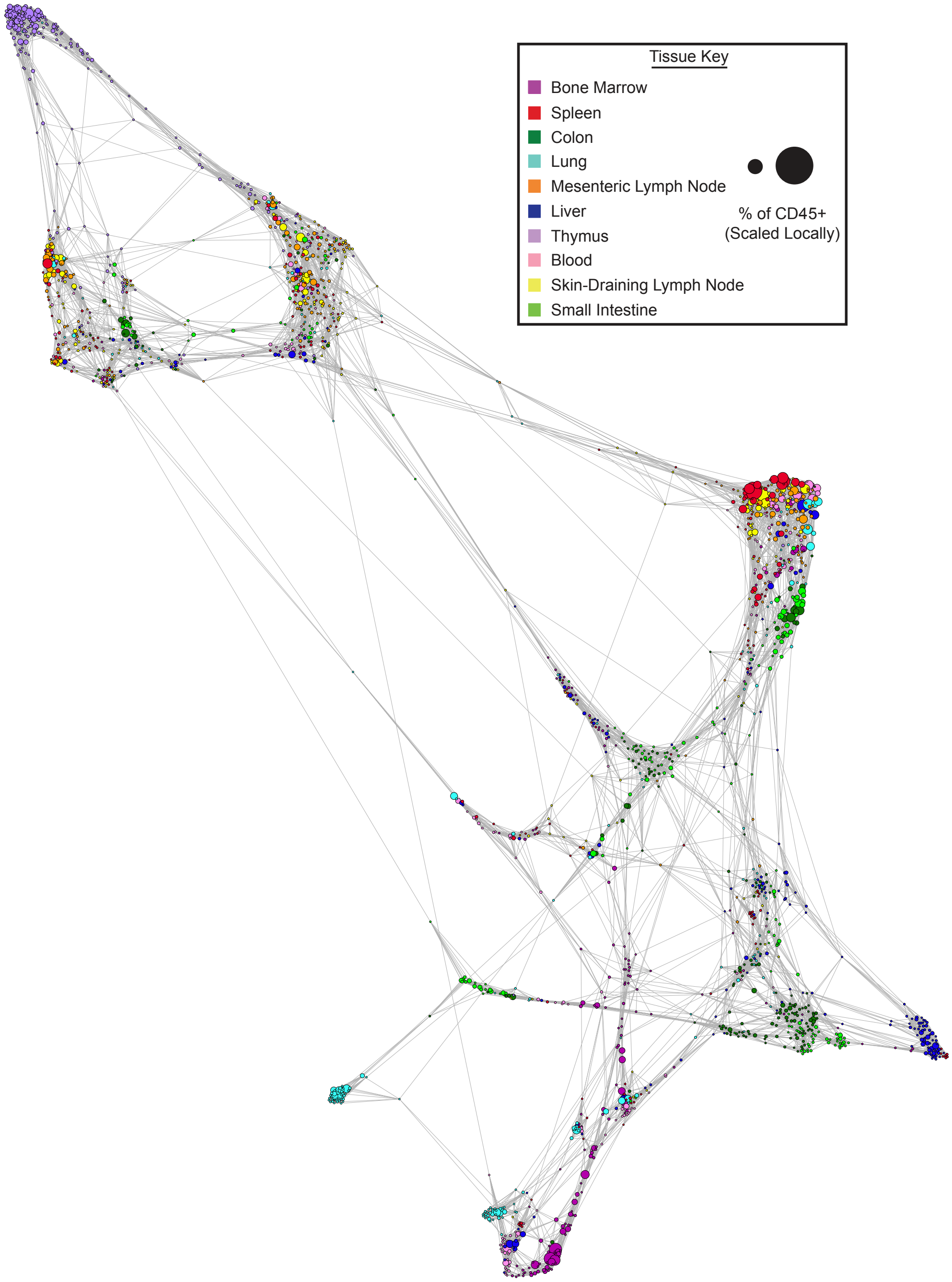
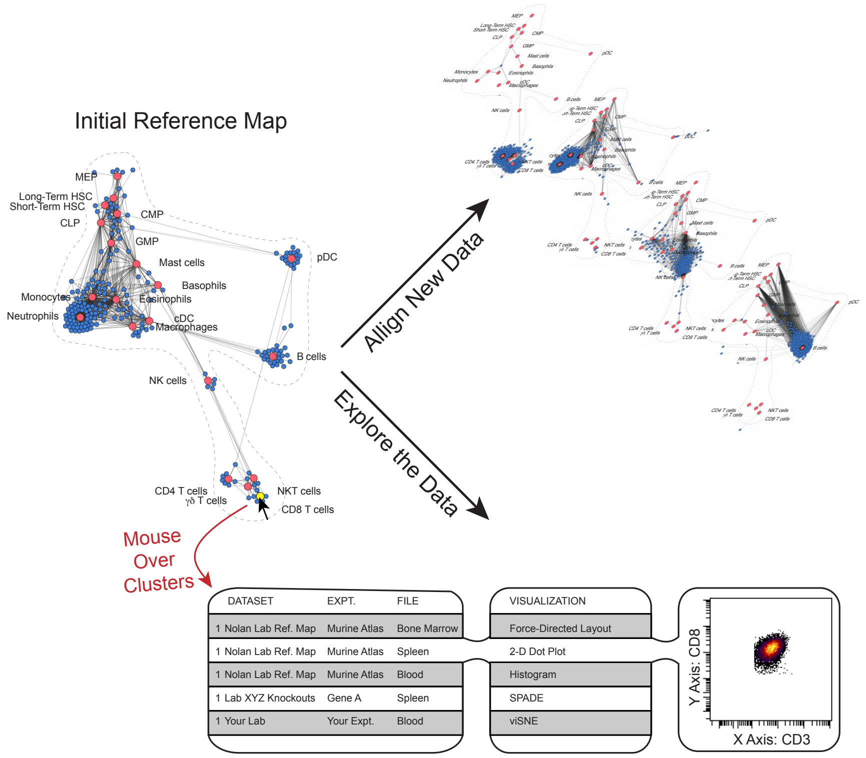


Fig. S14

Unsupervised force-directed graph including all cell clusters from all organs from C57BL/6 mice.



A model for an extensible, dynamic and community-collated immune reference map

Initial Reference Map

DATASET	EXPT.	FILE
1 Nolan Lab Ref. Map	Murine Atlas	Bone Marrow
1 Nolan Lab Ref. Map	Murine Atlas	Spleen
1 Nolan Lab Ref. Map	Murine Atlas	Blood
1 Lab XYZ Knockouts	Gene A	Spleen
1 Your Lab	Your Expt.	Blood

VISUALIZATION
Force-Directed Layout
2-D Dot Plot
Histogram
SPADE
viSNE

Fig. S15

A model for an extensible, dynamic and community-collated immune reference map. Using the approaches described here, it will become possible to assemble an analytical framework in which users can dynamically interrogate single-cell immune data. At the most basic level, this will enable mining of the datasets described here to investigate the characteristics of immunity across the body and between mouse strains by dynamically "mousing over" cell clusters to access their underlying metadata. Using Scaffold mapping to align data from independent experiments in a data-driven fashion, this resource could transform into a repository collated by the community, unifying investigations of immune cell characteristics and behaviors into a single, interactive resource.

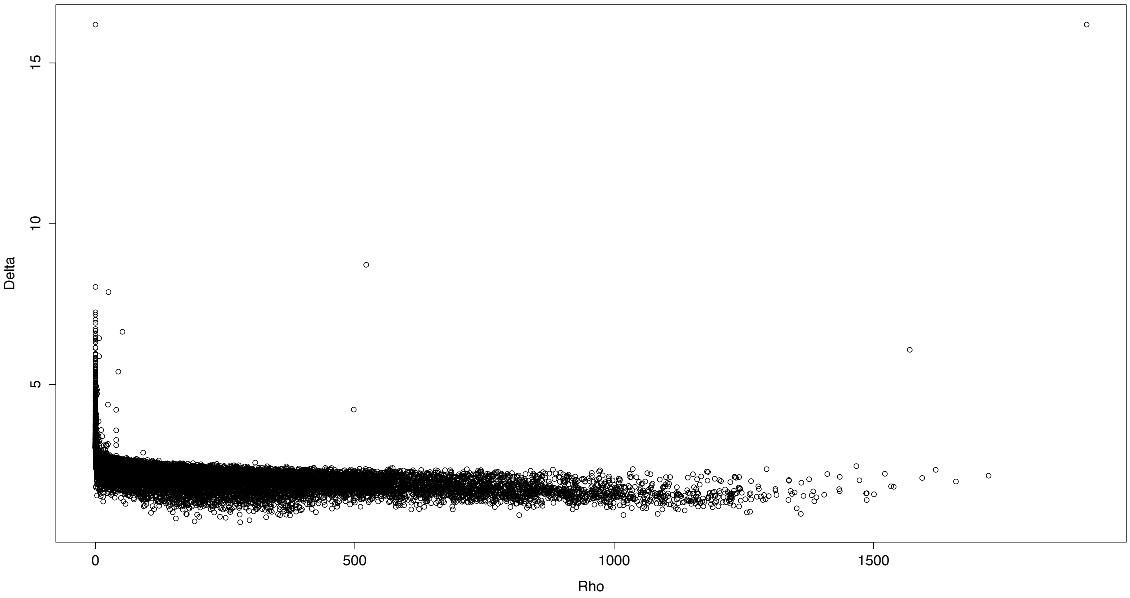


Fig. S16

Effect of using Rodriquez and Laio's clustering method (Science, 2014) on the bone marrow from a representative mouse. Note that computation of the distance matrix for all mice was extraordinarily time consuming. Additionally, their heuristic does not provide a useful threshold for cluster identification in these complex data.

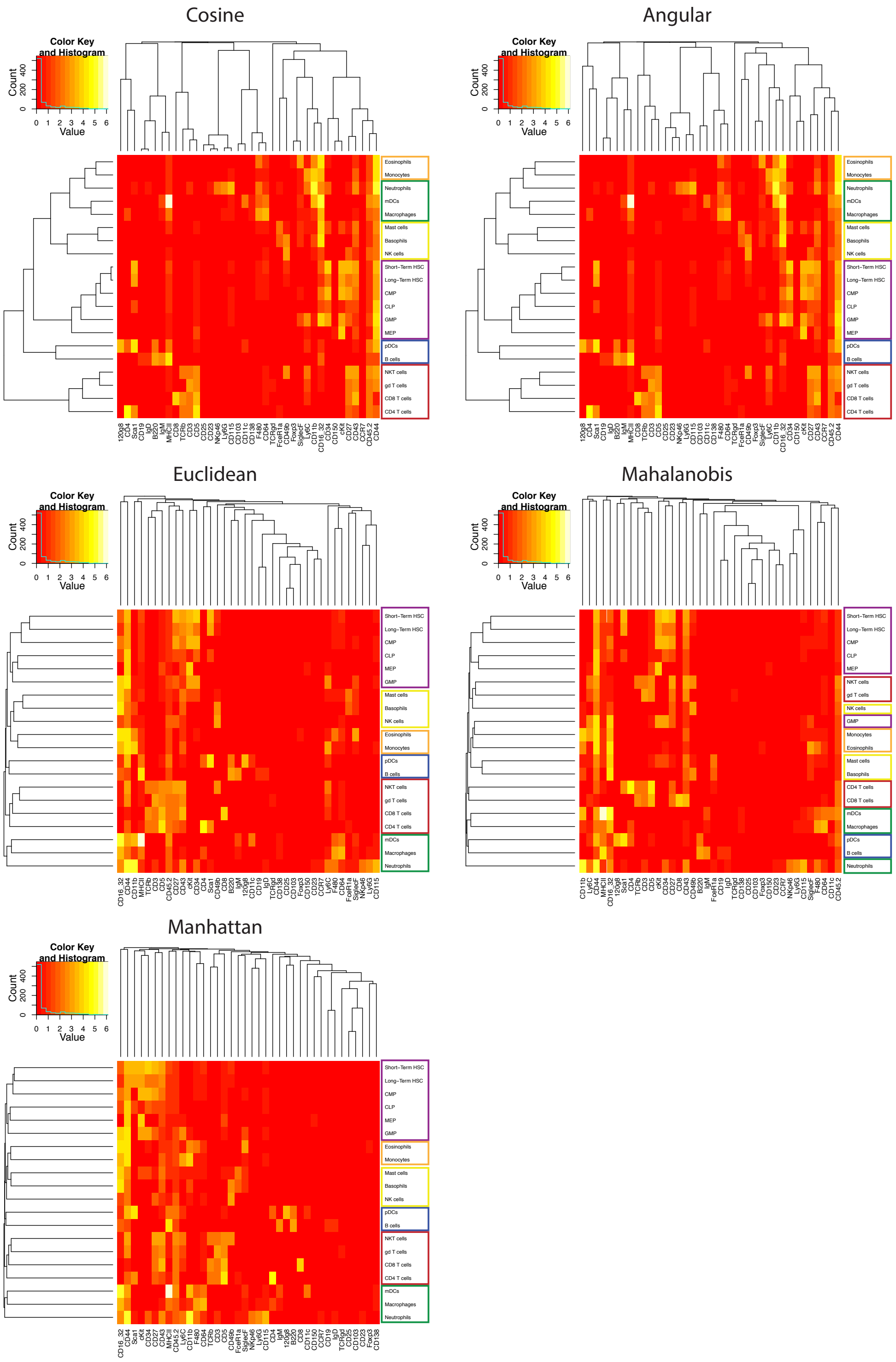


Fig. S17

Clustering of landmark cell populations for each of several distance metrics. The results are very similar, with Mahalanobis distance resulting in some minor differences.

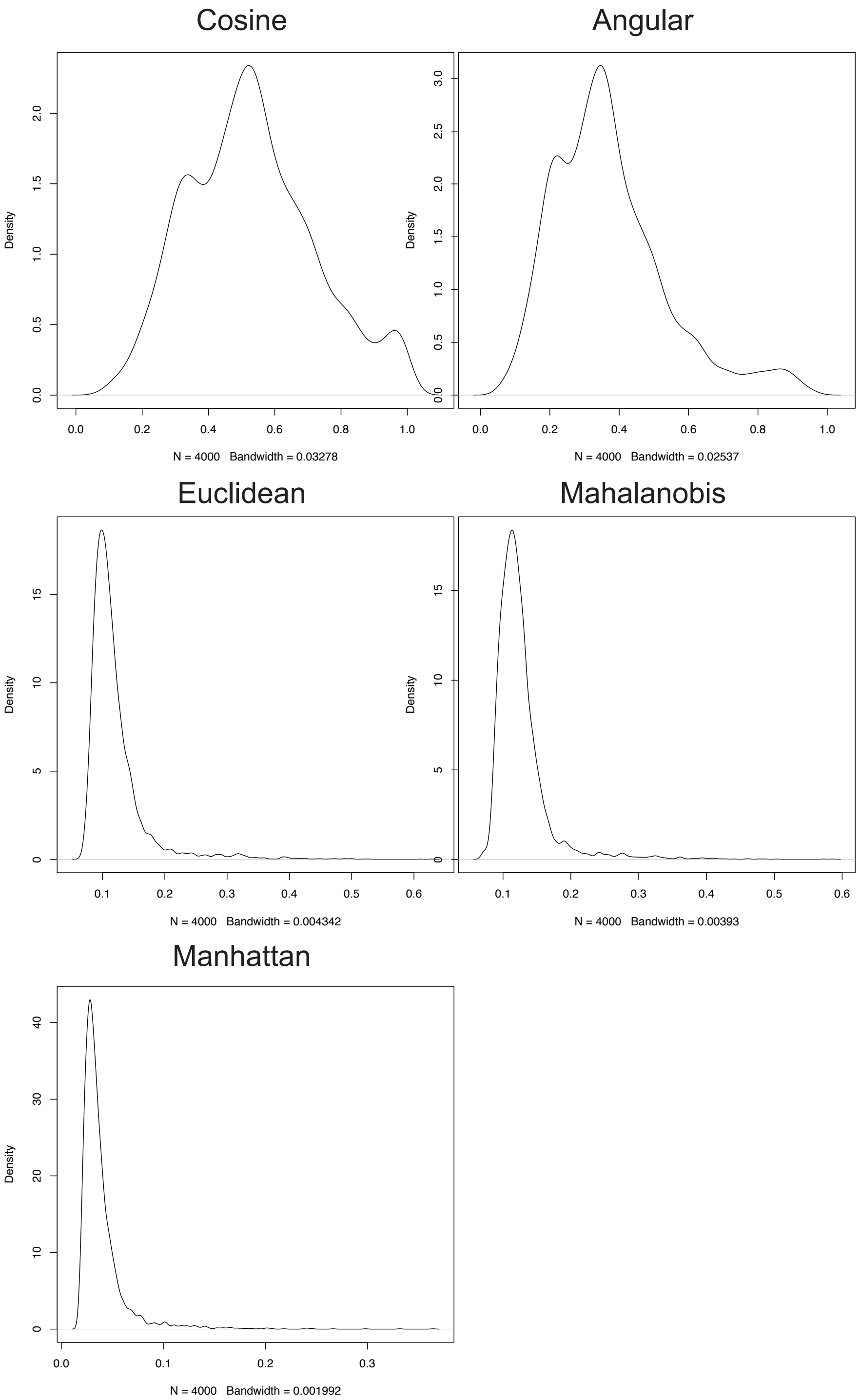
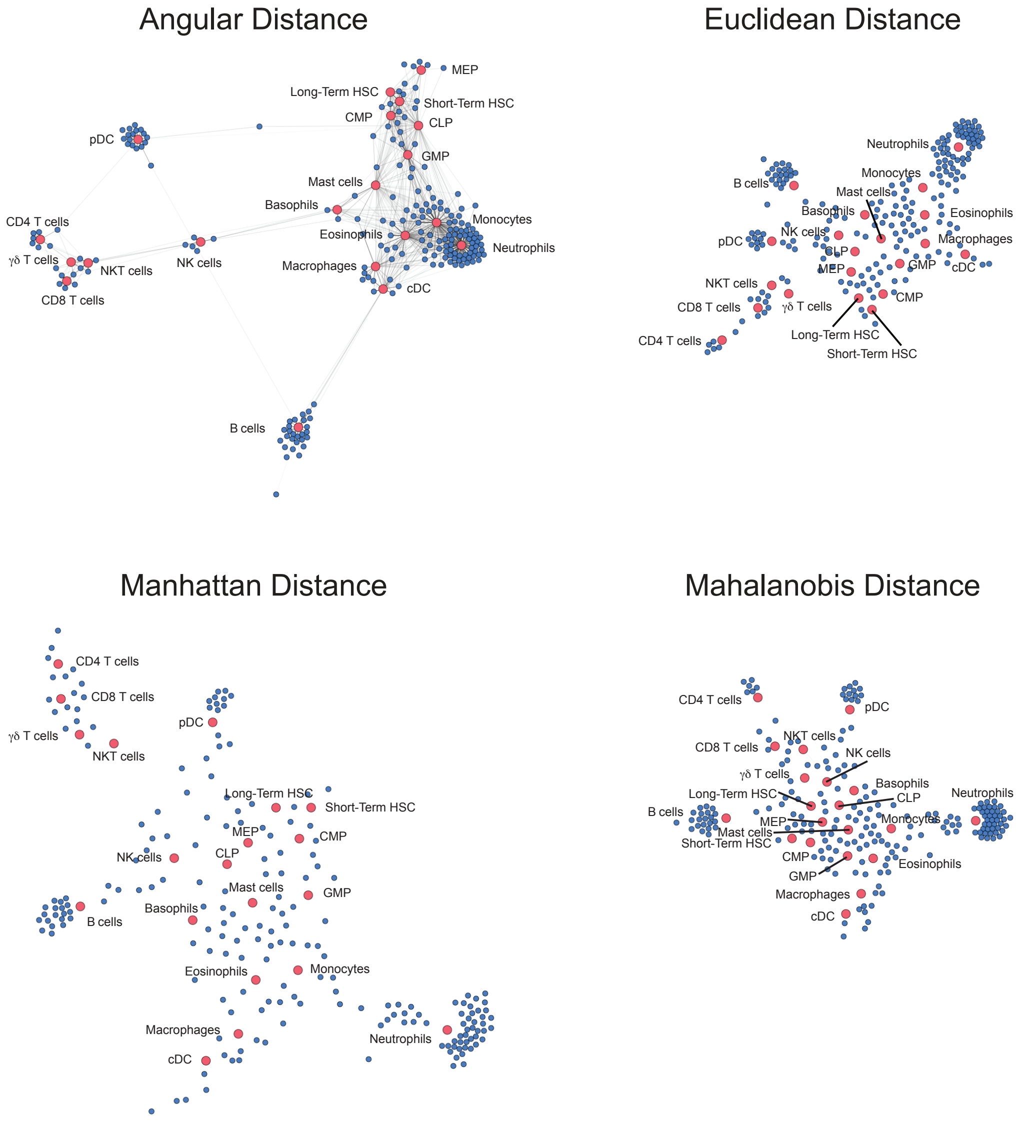


Fig. S18

The distribution of similarity measures for all clusters and landmarks for each potential choice of distance metric. Note the compressed range for Euclidean, Mahalanobis and Manhattan distances as compared to Cosine or Angular.



Angular Distance

Euclidean Distance

Manhattan Distance

Mahalanobis Distance

Fig. S19

Comparison of Scaffold map analysis using different distance metrics. Maps shown represent the bone marrow of C57BL/6 mice.

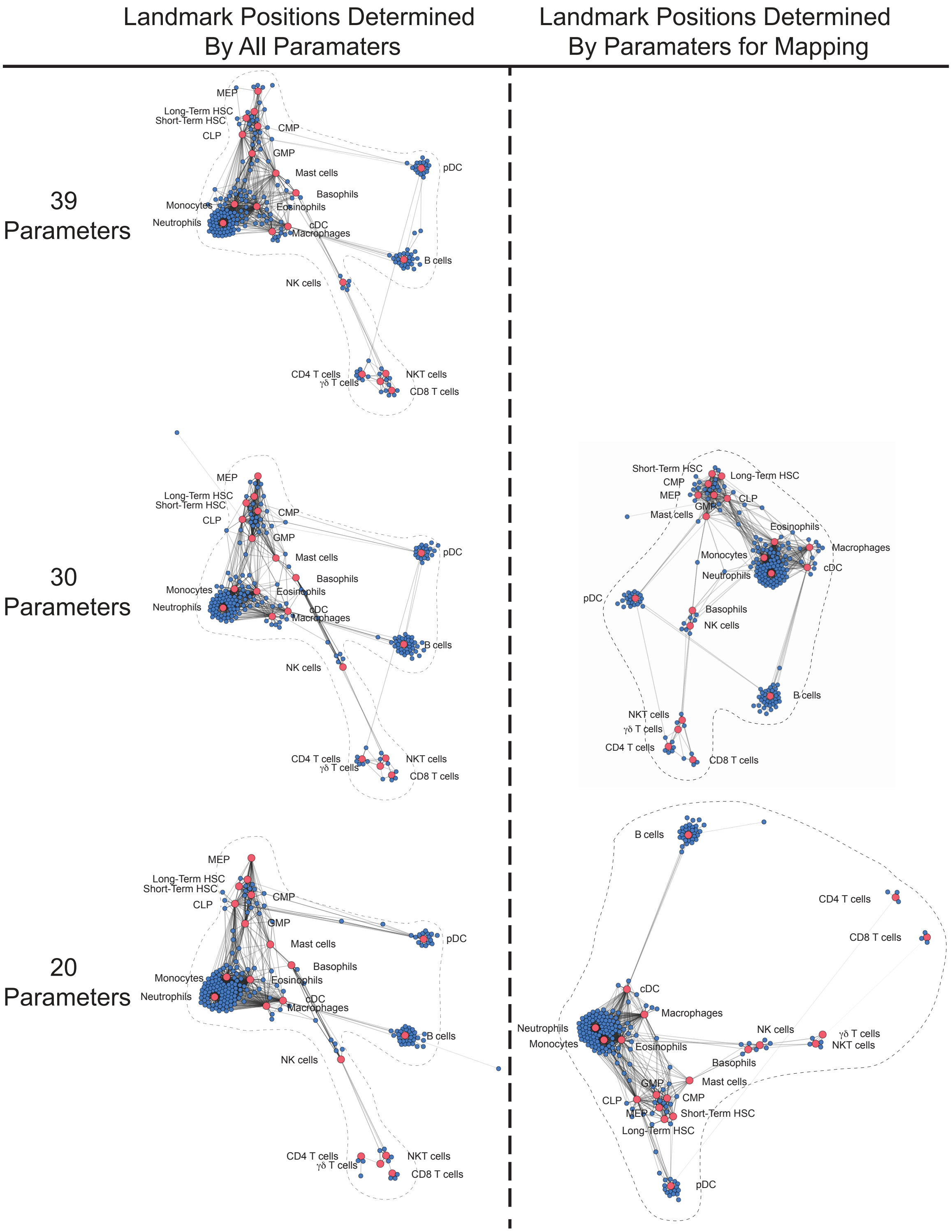


Fig. S20

Scaffold map analysis when fewer parameters are used either (1-Left) to map the unsupervised clusters onto the original Landmarks defined using all parameters or (2-Right) using fewer parameters to define both the Landmarks and the position of the unsupervised clusters.

Table S1.
Antibodies utilized for mass cytometry experiment. All reagents except those supplied by DVS Sciences were conjugated, validated and titrated in house.

Antigen	**Clone**	**Supplier**	**Elemental**	**Final Conc.**
Ter119	TER119	Biolegend	In113	4
CD45.2	104	Biolegend	In115	4
Ly6G	1A8	Biolegend	La139	2
IgD	11-26c.2a	BD	Pr141	1
CD11c	N418	DVS	Nd142	1 ul
F4/80	BM8	Biolegend	Nd143	8
CD3	17A2	BD	Nd144	4
NKp46	29A1.4	Biolegend	Nd145	8
CD23	B3B4	BD	Nd146	2
CD34	RAM34	BD	Sm147	16
CD115	AFS98	Biolegend	Nd148	16
CD19	1D3	BD	Sm149	2
PDCA-1	120g8	Imgenex	Nd150	2
CD8α	53-6.7	Biolegend	Eu151	4
Ly6C	HK1.4	Novus	Sm152	0.5
CD4	RM4-5	Biolegend	Eu153	2
CD11b	M1/70	Biolegend	Sm154	2
CD27	LG.3A10	Biolegend	Gd155	1
CD16/32	2.4G2	BD	Gd156	8
Siglec-F	E50-2440	BD	Gd157	4
Foxp3	NRRF-30	eBioscienc	Gd158	8
B220	RA3-6B2	BD	Tb159	2
CD5	53-7.3	DVS	Gd160	1 ul
FcεR1α	MAR-1	Biolegend	Dy161	8
TCRγδ	GL3	Biolegend	Dy162	2
CCR7	4B12	Biolegend	Dy163	8
Sca1	D7	DVS	Dy164	1 ul
CD49b	HMα2	Biolegend	Ho165	2
cKit	2B8	DVS	Er166	1 ul
CD150	TC15-	DVS	Er167	1 ul
CD25	3C7	BD	Er168	2
TCRb	H57-597	DVS	Tm169	2 ul
CD43	S7	BD	Er170	2
CD64	X54-5/7.1	Biolegend	Yb171	10
CD138	281-2	Biolegend	Yb172	1
CD103	2E7	Biolegend	Yb173	4
IgM	RMM-1	Biolegend	Yb174	8
CD44	IM7	BD	Lu175	0.25
MHC II	M5/114.15.2	Biolegend	Yb176	0.25

Table S2.

Frequencies of major immune cell populations by tissue of origin and inbred mouse strain as determined by manual gating. Values are mean ± SD.

Please see attached Excel file.

Table S3.

Deviant cellular phenotypes identified in the immune system reference map.

Major Distinguishing Markers	**Minor Distinguishing Markers**	**Tissue Specificity (if any)**	**Described as / Similar to**
CD3, TCRβ, TCRγδ, CD4, CD8	CD5, CD43, CD27	Small Intestine, Colon, Lungs	MAIT cells
B220, CD19, CD44, Sca-1	IgM, IgD, CD23	Thymus	B cell subset?
CD49b, CD11b, CD27, CD34, cKit	CD16/32, NKp46	Bone Marrow	NK cell subset?
CD43, CD11b, CD16/32 Ly6C, Ly6G	CD64, CD44, MHC II	Lung	Myeloid subset?
PDCA-1, B220, CD11c	MHC II, CD16/32	Small Intestine, Colon	pDC?
CD3, TCRβ, Foxp3, CD4, CD25	CD44, CD27	Mixed	$CD25^-$ $Foxp3^+$ Treg
CD16/32, F4/80, Siglec-F, CD44, PDCA-1, MHC II	Sca-1, Ly6C	Liver	Myeloid subset?
CD11c, CD16/32, CD44, MHC II	c-Kit, CD11b	Liver	Myeloid subset?

Table S4.

Readily observable stratifications of immune cell populations by expression of phenotypic markers in population-specific landscapes generated by force-directed graphs. Accompanying visualizations can be found in Figures S5-S9.

Major Population	Major Stratifying Markers	Minor Stratifying Markers	Other Relevant Markers Not Included
T cells	CD3, TCRβ, TCRγδ, CD4, CD8	CCR7, CD5, CD25, CD27, CD43, CD44, CD49b, Foxp3, Ly6C, Sca-1	CD62L, CD90, T-bet, GATA-3, RORγt, Bcl6, NK1.1, CD122, CD24
B cells	B220, CD19, IgM, IgD, MHC II, CD138	CCR7, CD5, CD23, CD44, CD16/32, Sca-1, CD43	CD93, CD127, CD24, CD21, IgG, IgE, IgA
NK cells	CD49b, CD11b, CD27	CD34, CD43, CD44, CD16/32, cKit, Ly6C, NKp46, Sca-1	NK1.1, Ly49H, Ly49C/I, NKG2D, NKG2A, KLRG1
cDC	CD11c, MHC II, CD11b, CD4, CD8,	CCR7, CD43, CD44, CD103, CD16/32, cKit, Sca-1	DNGR-1, CD205, CD207, Dectin-1, SIRPα
Macrophages	CD11b, CD11c, CD64, F4/80, Siglec-F	CD44, CD115, CD16/32, Ly6C, MHC II, Sca-1	CD169, MerTK
Monocytes	CD11b, CD115, Ly6C, MHC II	CCR7, CD4, CD11c, CD43, CD44, CD16/32, Sca-1	

Author Contributions:

MHS planned and performed the experiments, developed the algorithm, analyzed the data and wrote the paper. PFG developed the algorithm, analyzed the data and revised the paper. GKF analyzed the data and revised the paper. NB, RTY and YC assisted with tissue harvest, fluorescence flow cytometry and microscopy. ANH helped design the experiments. RF analyzed the data. ERZ adapted the force-directed layout algorithm and developed the mass-tag cell barcoding protocol. WJF and SCB helped design the experiments and revised the paper. EGE designed the experiments, revised the paper and supervised the study. GPN designed the experiments, wrote the paper and supervised the study.