Spatial Patterning Analysis of Cellular Ensembles (SPACE) enables statistically robust discovery of complex spatial organization at the cell and tissue level

Edward C. Schrom[1], Erin F. McCaffrey[2], Andrea J. Radtke[1,3], Emily Speranza[4], Leanne Arakkal[1], Nishant Thakur[1,3], Spencer Grant[5], Ronald N. Germain[1,3]
[1]Lymphocyte Biology Section, Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892-1892, USA.
[2]T-Lymphocyte Biology Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892-1892, USA.
[3]Center for Advanced Tissue Imaging, Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892-1892, USA.
[4]Florida Research and Innovation Center, Cleveland Clinic Lerner Research Institute, Port Saint Lucie, FL 34987, USA.
[5]Center for Alzheimer's and Related Dementias, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892-1892, USA.

ABSTRACT

Spatial patterns of cells and other entities drive both physiologic and pathologic processes within tissues. While many imaging and transcriptomic methods document tissue organization, discerning these patterns is challenging, especially when they involve multiple entities in complex arrangements. To address this challenge, we present Spatial Patterning Analysis of Cellular Ensembles (SPACE), an R package for analysis of high-plex tissue images generated using any collection modality. Unlike existing platforms, SPACE detects context-dependent associations, quantitative gradients and orientations, and other organizational complexities. Via a robust information theoretic framework, SPACE explores all possible ensembles – single entities, pairs, triplets, and so on – and ranks the strongest patterns of tissue organization. Using lymph node images for which ground truth has been defined, we validate SPACE and demonstrate its advantages. We then use SPACE to reanalyze a public dataset of human tuberculosis granulomas, verifying known patterns and discovering new patterns with possible insights into disease progression.

INTRODUCTION

Biological tissue function emerges from the local interactions of diverse cells and structures. Whether these interactions occur via direct contact or paracrine signaling, they are determined by the spatial arrangement of the involved entities. Thus discovering new spatial patterns can reveal new mechanisms of tissue (dys)function [Baccin et al 2020, Saviano et al 2020, Lewis et al 2021, Rendeiro et al 2021, Germain et al 2022].

Many methods exist to capture the spatial arrangements of deeply phenotyped cells and structures. Protein-targeted imaging methods, using fluorescence [Lin et al 2016, Radtke et al 2022], sequencing [Black et al 2021], or heavy metal isotopes [Giesen et al 2014, Keren et al 2019], capture 30-100 molecular parameters at submicron resolution. Transcript-targeted imaging methods capture hundreds of molecular parameters at similar resolution [Eng et al 2019, Xia et al 2019, Moffit et al 2022], or the entire transcriptome at lower spatial resolution [Stahl et al 2016, Rao et al 2021]. Multimodal techniques combine the advantages of several methods [Vandereyken et al 2023]. In each case, dozens to hundreds of cell types and structures can be identified, and the spatial patterns created by these many elements are too numerous for manual exploration.

To analyze high-plex spatial data, many computational platforms have been introduced [Schapiro et al 2017, Goltsev et al 2018, Stoltzfus et al 2020, Dries et al 2021, Bhate et al 2022, Bartolomeazzi et al 2022, Warchol et al 2023]. Existing platforms essentially perform two classes of analyses. The first examines pairwise associations based on the frequency of spatial co-occurrence [Schapiro et al 2017], spatial correlation [Stoltzfus et al 2020, Warchol et al 2023], or nearest distances

between individual cells [Bartolomeazzi et al 2022]. These analyses do not directly address patterns involving three or more cell types. Moreover, they only detect linear associations, i.e., associations that occur consistently throughout a tissue sample, but not nonlinear associations, i.e., context-specific associations that differ across subregions of a tissue sample. The second class of analysis uses machine learning (e.g., K-means) or topic modeling to define tissue microenvironments (MEs) of consistent cell type compositions [Goltsev et al 2018, Chen et al 2020, Schurch et al 2020, McCaffrey et al 2022, Nirmal et al 2022, Radtke et al 2023]. While this approach uses all cell types instead of just two, it often lacks statistical validation: the requested number of MEs will be found, whether or not their compositions are optimally distinguished from random assortment [Halkidi et al 2001, Baya & Granitto 2013]. Moreover, classification of discrete MEs ignores continuous gradients of cellular abundance. Finally, both classes of analyses rely on tabulated centroid locations of segmented cells, requiring customized approaches for irregularly shaped cells, acellular structures, or raw measurements of molecular expression.

To overcome these limitations of existing approaches, we present SPACE (Table 1). SPACE is a novel information-theoretic algorithm that systematically detects, statistically validates, and thoroughly characterizes spatial patterns of arbitrary complexity in imaged tissue samples. SPACE operates on raw molecular expression data, classified pixels, spatial maps of cellular segmentation, and/or centroid data simultaneously (Fig 1a-d), broadening the scope of entities that can be analyzed. For any combination of these entities, SPACE leverages information theory to quantify the deviation from random patterning in a metric we call "cis mutual information" (cisMI) (Fig 1e-g, Fig S1). Simple assumptions of parsimony allow SPACE to perform an exhaustive exploration and yet discern the top spatial patterns most likely to reveal interesting biology. Specific patterns are then characterized in detail, revealing not only co-assortment but also gradients, orientations, and other complex features that may be context-specific (Fig. 1h-j). SPACE is available as an open-source R package.

We demonstrate and validate SPACE on data from different imaging modalities, host species and tissues, and input data formats. We focus on the mouse lymph node, where ground-truth spatial patterning is known, as well as a recent study of human tuberculosis (TB) granulomas [McCaffrey et al 2022]. Although TB causes 1.6 million human deaths annually [WHO Global Tuberculosis Report 2022] and is widely studied, mechanisms of disease progression are not fully understood [Flynn & Chan 2022]. In addition to rediscovering known patterns in these data, SPACE provides a new level of exploratory power that translates the enormous complexity of high-plex spatial data into a small set of candidate spatial patterns most likely to reveal important underlying biological processes.

RESULTS

SPACE Detects Known Patterns and Gradients in the Mouse Lymph Node

To compare SPACE output against 'ground-truth' for a well-studied tissue, we used a 42-plex image of a mouse popliteal lymph node (LN) captured with the IBEX protocol [Radtke et al 2020] [Fig. S2a]. Using CellPose 2.0 [Pachitariu and Stringer 2022], we segmented individual cells [Fig. S2b]. To restrict our focus to well-studied cellular subsets, we chose 20 lineage markers to define canonical expression profiles [Table S1] and deployed a semi-supervised pipeline [Fig. S2c-h] to classify the cells into 19 types [Fig. 2a] and map them onto a segmentation mask [Fig. 2b]. We censused 3000 neighborhoods of radius 20 µm on this mask and measured cisMI for every ensemble of one, two, or three cell types. Of the possible 1159 ensembles, 498 (43.0%) have significant patterning, and these are ranked by cisMI (Fig. 2c). Individual cell types are also ranked by average cisMI (Fig 2d). Performing the same analysis on a table of cell centroids yields similar, albeit less powerful, results (Text S1, Fig S3).

The three most strongly patterned cell types on average are CD4+ T cells, CD8+ T cells, and naïve B cells (Fig. 2d). These populations define the largest microanatomical zones of the LN – the paracortical T zone and primary B follicles. These zones can be explained purely by self-clustering of each cell type, co-clustering of CD4+ and CD8+ T cells, and separation of naïve B cells from T cells (Fig. 2c, stars). Even

controlling for these patterns, the strongest ensemble includes all three populations (Fig. 2c, arrow), indicating that their patterning encodes more than just the obvious microanatomical zonation.

We investigated this tripartite pattern using a covariation plot (Fig. 2e). In accordance with the general separation of the T zone and B follicles, the covariation plot shows separate regions where T cells vs. naïve B cells dominate (Fig. 2e, 15-65% along the x-axis vs. 65-100%). Additionally, within the T zone, CD4+ and CD8+ T cells form opposing gradients, where CD4+ T cells dominate closer to the main B follicle, whereas CD8+ T cells dominate farther away. Although the visual manifestation of this pattern depends on the slicing angle of the LN (Fig S4), it remains that CD4+ and CD8+ T cells co-occur at large but oppose one another specifically within the B-cell-devoid T zone. This nonlinear context-dependent pattern, which requires all three cellular partners to detect, was discovered recently [Baptista et al 2019] using manual histo-cytometry analyses [Gerner et al 2012], and SPACE recovers via automated exploration. Additionally, after mapping the B follicle and T zone (Fig. 2f), SPACE finds that Type 1 dendritic cells (cDC1s) peak in abundance closer to the center of the T zone, whereas cDC2s peak in abundance at the edge (Fig 2g), as originally discovered [Gerner et al 2012, Baptista et al 2019]. The combined organization of lymphocytes and DCs discovered by SPACE is clear from raw marker expression (Fig. 2h).

Finally, we chose a weaker ensemble: naïve B cells, light zone (LZ) germinal center (GC) B cells, and dark zone (DZ) GC B cells (Fig 2c, square). The covariation plot shows that these cell types are concentrated in three nearby zones, as expected. Additionally, DZ B cells are proximal to the bulk of the naïve B follicles which borders the T zone, whereas LZ B cells are distal (Fig. 2i). This is visually apparent (Fig. 2j). While the separation of the naïve B follicles, GC LZ, and GC DZ can be explained in terms of single or paired cell, positive cisMI for the tripartite pattern is statistical evidence that GC orientation within the B follicles, which requires all three cell types, is also significantly non-random. Neither GC orientation nor the T cell or cDC gradients are readily detectable with existing spatial analysis approaches (Text S2, Fig S5-S6).

SPACE Detects Known and New Patterns in Human TB

We next reanalyzed a publicly available data set of human TB granulomas. Using a 37-plex MIBI-TOF panel, two 500µm x 500µm fields of view (FoVs) were imaged from each of 15 human TB patients [McCaffrey et al 2022]. Of these patients, nine were undergoing diagnostic biopsies, three were undergoing therapeutic resections, and three were sampled postmortem. The original study uncovered consistent patterns of granuloma organization across clinical status. One of the strongest is the co-expression of immune-inhibiting molecules PD-L1 and IDO1 at the cellular level. After censusing 5 µm neighborhoods to capture single cells, we used SPACE to all pairs of the nine functional markers included in the original study. In many individual FoVs, SPACE finds that PD-L1 and IDO1 are the most strongly patterned pair of markers (Fig. 3a), due to co-expression (Fig. 3b). This is consistently observed across the full data set (Fig. 3c). Although PD-L1 and IDO1 consistently form the strongest spatial pattern (Fig. 3d, star), SPACE finds other marker pairs with average cisMI > 0 across the 30 FoVs (Fig. 3d, arrow), which were undetected in the original study. For example, H3K9Ac and IDO1 are negatively correlated in their expression, both in representative FoVs (Fig. 3e) and consistently across the data set (Fig. 3f). Because negative pairwise correlations are typically detectable by other approaches, this likely reflects SPACE's advantage of analyzing images of raw signal rather than tables of average cellular expression.

We also used SPACE to examine spatial patterns among the 20 segmented cell types defined in the original study. After censusing 50 µm neighborhoods to match the original study, we quantified cisMI for every ensemble of one, two, or three cell types in each FoV and then averaged across FoVs to find the strongest cellular patterns in the full data set (Fig. 3g). Several of the strongest patterns are formed by single cell types, e.g. CD11b/c+ CD206+ macrophages ("TP Macs"), which self-cluster more than expected by random chance (Fig. 3h, star). This agrees with the Spatial-LDA clustering analysis of

original study (Fig. 3i), which found that TP Macs were largely restricted to a single ME, termed "Myeloid Core 1" or "Mcore1."

We also studied the strongest tripartite pattern for which all three cell types were present in ≥50% of images, which includes TP Macs, B cells, and CD4+ T cells (Fig. 3g, arrow). In an illustrative FoV, these populations dominate in three separate areas of the granuloma (Fig. 3j). This agrees with the original Spatial-LDA analysis, which found that these three cell types were each restricted to a different ME (Fig. 3i). Gating on the covariation plot reveals MEs nearly identical to those discovered by Spatial-LDA (Fig. 3k). However, mere separation into distinct MEs only requires single and pairwise patterns: each cell type self-clusters, and each pair of cell types is mutually exclusive. The significant cisMI of the tripartite pattern indicates that other features of their spatial patterning are also important. Although B cells and CD4+ T cells dominate in separate MEs, they appear positively correlated specifically in non-Mcore1 niches where TP Macs are present in small numbers (Fig 3j, 10-20% on the x-axis). This context-dependent co-occurrence of B and CD4+ T cells is consistent across the full data set (Fig. 3l). This suggests that in TB granulomas, TP Macs at the edges of or outside Mcore1 can unite B and CD4+ T cells that would otherwise separate from one another. Thus, SPACE recovers patterns found by Spatial-LDA while also discovering context-specific gradients that are missed by clustering algorithms. Even patterns from the original study that integrate raw marker expression and cellular phenotypes/MEs are recovered and refined by SPACE (Text S3, Fig. S7).

SPACE Discovers Cellular Patterns Associated with TB Disease Severity

The range of clinical statuses – biopsy, resection, and post-mortem – likely represents progressing disease. While other factors such as ongoing treatment and co-morbidities are surely important, they are not available as metadata. Nonetheless, we searched for patterns associated with putative severity. Although cellular abundances do not predict clinical status, SPACE also measures two metrics of cellular diversity. Alpha diversity, which measures the number and evenness of cell types in each FoV (Fig. 4a), is higher for biopsy and resection samples compared to postmortem (Fig 4b). Beta diversity, which measures the distinctiveness ME cellular compositions (Fig. 4c), is higher in biopsy samples, compared to either resection or post-mortem (Fig. 4d). Together, this suggests that early in disease progression, the distinctions among MEs blur as cell types intermix. Later, one or a few cell types dominate the granuloma, though not the same types in each patient.

Additionally, we searched for ensembles whose strength of spatial patterning correlates with clinical status. After narrowing our focus to cell types that were present in ≥80% of images, cisMI significantly associates with clinical status for one ensemble: CD4+ T cells, CD14+ monocytes, and TP macrophages. The strength of non-random patterning among these three cell types decreases as disease progresses (Fig. 4e). For more granular detail, we decomposed the census for each FoV into eight fractions: neighborhoods with: i) none of the three cell types, ii) just CD4+ T cells, iii) just CD14+ monocytes, iv) just TP Macs, v) CD4+ T cells and CD14+ monocytes, vi) CD4+ T cells and TP Macs, vii) CD14+ monocytes and TP Macs, and viii) all three cell types. We compared these fractions to fractions from randomized censuses simulated in SPACE. Thus, for each FoV, the eight fractions can be expressed as a Z-scores relative to null expectations. Extreme Z-scores indicate deviations from random patterning, such that average absolute Z-score correlates strongly with cisMI (Fig. 4f). Feeding the Z-scores for each FoV into PCA reveals that PC1 captures a large fraction of the variation (42.6%) and distinguishes biopsy from resection and post-mortem samples (Fig. 4g). PC1 loadings show that as disease progresses, neighborhoods with just 0-1 of the three cell types become less enriched, whereas neighborhoods with 2+ of the cell types become more enriched (Fig. 4h, stars). This agrees with the increase in cellular

intermixing with more severe disease. However, one exception to this trend is the loss of CD4+ T cell colocalization with CD14+ monocytes (Fig. 4h, arrow).

To investigate functional interactions between CD4+ T cells and CD14+ monocytes in TB, we used 5 µm neighborhoods censused by SPACE to examine molecular expression. CD14+ monocytes expressed the inhibitory ligand PD-L1 at significantly higher levels in biopsy samples compared to resection or post-mortem (Fig. 4i). Although PD-L1 expression can be driven by Type 1 inflammatory signaling, little if any IFNγ was detected in the 30 FoVs (Fig. 4i). PD-L1 on CD14+ monocytes may suppress the activation of nearby T cells, preventing a robust inflammatory environment from forming. Using 20 µm neighborhoods, we modeled PD-1 expression probability on CD4+ T cells (a readout of antigen-specific activation) as a function of the frequency of nearby CD14+ monocytes. Nearby CD14+ monocytes strongly diminish CD4+ T cell activation in biopsy samples, weakly diminish it in resection samples, and exert no detectable effect in post-mortem samples (Fig. 4j). Thus, as disease progresses, CD14+ monocytes interact with CD4+ T cells less frequently, and these interactions become less suppressive when they do occur. Notably, these changes happen in a largely non-inflammatory signaling environment in which most other cell types are intermixing more rather than less. Therefore, it is possible that suppressive CD14+ monocyte interactions with CD4+ T cells help maintain the granuloma structure which contains pathogenic bacteria. If this structure decays without a corresponding increase in Type 1 inflammation, disease worsens. While such conjectures cannot be proven using observational data, they demonstrate that SPACE rapidly narrows a plethora of possible spatial patterns to a small set of statistically verified correlates of clinical status, streamlining the process of hypothesis generation.

DISCUSSION

High-content spatial representations of biological tissues, including the images of human TB granulomas studied here, contain enormous amounts of information on tissue architecture and cellular arrangement. Existing analytical techniques – pairwise association metrics and MEs defined by clustering algorithms – only capture part of the insights available in such dense datasets. SPACE is, to the best of our knowledge, the first analysis platform to exhaustively quantify all possible spatial patterns. These patterns encompass any number of cell types or other entities and any degree of complexity, including context-dependent associations and quantitative gradients and orientations. Combined with the parsimonious assumption that higher-order patterns are built from lower-order patterns, SPACE ranks the exhaustively searched patterns, narrowing the enormous amount of information in the original data to a short list of candidate patterns most likely to reveal drivers of tissue organization.

As high-content spatial data on biological tissues becomes more sophisticated and prevalent, new innovations will be added to the SPACE platform. For example, SPACE currently operates on 2D image data, but 3D volumetric imaging promises to yield even more insights into complex tissue architecture [Li et al 2019]. While the information theoretic calculations behind cisMI are equally applicable, the program must be generalized to census 3D images. Similarly, while there is no theoretical limit to the number of ensembles for which cisMI can be calculated, there are practical limitations. For example, for a spatial transcriptomics data set with 1000 unique transcripts, exploring all ensembles including four mRNA species implies $>4\times10^{10}$ unique combinations, which would be computationally infeasible. Finally, SPACE analyses one image at a time, quantifying the difference between its true patterning vs. randomized expectations. Comparison across multiple images requires customized secondary analyses. As high-content spatial data becomes easier to collect in parallel for numerous samples, it will be more useful to directly quantify the spatial patterns that best distinguish sample groups from one another. To accomplish this in a single step, we are developing another information-theoretic metric, called transMI.

Even in its current form, the applications of SPACE are broad. Although we developed SPACE with high-content tissue images in mind, SPACE can census and analyze any properly formatted image-based data, from super-resolution microscopy of single cells to remote sensing of earth's surface. The information-theoretic calculation of cisMI is agnostic to the source of the data, whether spatial or not. So long as randomized null expectations can be credibly simulated, SPACE essentially performs a nonparametric test for divergences between two distributions. Whereas common nonparametric tests (e.g., Wilcoxon Rank Sum, Kruskal-Wallis) only compare the first moment of 1-dimensional distributions, SPACE compares all moments of any-dimensional distributions, even when the underlying data is compositional. Thus, SPACE could prove to be an enormously useful statistical innovation. SPACE is available as a free, open-source R package that can be widely adopted and modified to support advanced analytical capabilities in a variety of scientific fields.

METHODS

Principles of SPACE

SPACE operates on a variety of image-based input data types, including images of quantitative molecular expression, segmented cells classified into phenotypes, and pixels classified into cellular phenotypes and/or acellular structures (Table 1, Fig 1a). SPACE can also accept tables of cellular XY centroids reporting the phenotype and/or average marker expression for each cell. Importantly, multiple inputs can be analyzed simultaneously, to integrate data at the molecular, cellular, and structural levels. Biological components at any of these levels are simply called “variables.” Across each layer of input data, thousands of randomly located circular neighborhoods are delimited (Fig 1b). While such sampled neighborhoods may overlap, no two neighborhoods are identically placed. The amount of each variable in each neighborhood is reported in a table called a “census” (Fig 1c-d). A census may include a variety of neighborhood sizes, for downstream analysis of patterning at different scales (e.g., adjacent cell contacts at 5-10µm vs. microanatomical zonation at 50-100µm). Inside each neighborhood, the contiguous patches of each variable are also recorded, respecting any truncation by the neighborhood’s boundary. By reassorting these patches into new hypothetical neighborhoods, randomized censuses are simulated (Fig. 1c). Randomized censuses represent null expectations in the absence of spatial patterning (Fig. 1d), while controlling for factors such cell size, morphology, and tissue geometry. At least 100 randomized censuses are simulated for downstream bootstrapping.

From the observed and randomized censuses, the degree of non-random patterning for any ensemble of variables can be quantified. The observed census gives the true distributions of heterotypic co-occurrence and homotypic occurrence for the included variables, while a randomized census gives the null expectations for these same distributions (Fig 1e). Although this is most easily depicted for an ensemble of two variables, it generalizes to ensembles of any size. The difference between these distributions is quantified using an analogue of mutual information called “cisMI,” measured in bits, which measures the degree of non-random patterning (Fig. 1f). Calculating cisMI requires several steps, to control for several important factors. First, the observed and two randomized censuses provide distributions of co-occurrence for the variables in the ensemble of interest, along with distributions for all subsets of these variables (Fig. S1a). These distributions are discretized with a user-specified number of bins per variable (Fig. S1b). More bins can resolve more intricate spatial patterns but also require more neighborhoods to be censused. The empirical bin probabilities are smoothed by calculating the Chao-Jost coverage estimator C [Chao & Jost 2012], normalizing the empirical distribution to C, and distributing the remaining 1-C probability mass uniformly across all possible but unobserved bins (Fig. S1c). This excludes bins representing impossible combinations of variables; for example, two cell types cannot sum to >100% of the area in a neighborhood. Then the Kullback-Leibler divergence is calculated between the observed and one randomized distribution, normalizing by the total number of possible bins to control for the compositional constraint on distribution shape. Per-bin Kullback-Leibler

divergence between two randomized distributions is subtracted, to control for variation among randomized distributions that arises during stochastic simulation (Fig S1d). This yields an intermediate measure of cisMI. Because non-random patterning of single variables contributes to apparent non-random patterning of 2-variable ensembles (and non-random patterning of 2-variable ensembles contributes to apparent non-random patterning of 3-variable ensembles, and so on), the intermediate measure for all lower-order subsets is subtracted from higher-order ensembles, giving a final measure of cisMI (Fig. S1e). This regularizes cisMI for higher-order ensembles, invoking a parsimonious assumption that spatial patterning is driven by the simplest (lowest-order) interactions. Remaining cisMI for higher-order ensembles reflects patterning that cannot be explained purely by lower-order terms. CisMI is measured repeatedly using many simulations of randomized distributions, to estimate uncertainty via bootstrapping (Fig S1f). If ≥95% of cisMI estimates exceed 0, then the ensemble displays a significantly non-random spatial pattern. A single ensemble can be queried for hypothesis testing, or these calculations can be looped over many ensembles to explore many possible spatial patterns, where ensembles are ranked from strongest to weakest patterning (highest to lowest cisMI, Fig 1g).

Importantly, a high cisMI score indicates strong non-random patterning, but it does not indicate the type of pattern, whether co-occurrence, mutual exclusion, or non-linear. Thus, an ensemble of interest with significant cisMI can be selected for further exploration. From the observed census, the distribution of heterotypic co-occurrence is generated in V-dimensional space (where V is the number of variables in the ensemble), and a self-organizing map places a curve through the areas of highest data density (Fig. 1h). The curve is then projected into lower dimension by plotting the value of each variable along the curve separately. The resulting "covariation plot" shows how each variable in the ensemble covaries as the tissue is traversed along a hidden but illustrative (i.e., latent) path (Fig 1i). This reveals context-dependent associations, quantitative gradients, and other features of the ensemble's spatial patterning. To visualize this in the original tissue context, manual gates can be drawn on the covariation plot, splitting the image into regions with characteristic amounts or gradients of each of the involved variables (Fig 1i). These regions are mapped back onto the original image with false coloring for easy visualization (Fig 1j). Because mapped regions are identical in format to segmentation masks and pixel-classified images, they can be recycled as inputs for another round of SPACE analysis, to find hierarchical spatial relationships among the regions themselves (Fig 1j).

Implementation and Availability of SPACE

SPACE is written in the R computing language and is freely available as an open-source R package on Github [link page when ready], where code documentation and a tutorial are also available. All other analyses in this paper are performed in R as described in the text and figure legends. Data and code required for all analyses in this paper are freely available on Github [link page when ready].

Acknowledgments
This research was supported in part by the Intramural Research Program of the NIH.

Funding
This work was supported by the Intramural Research Program of the NIH and by a grant from the Chan-Zuckerberg Initiative (CZI)

Author Contributions
ECS conceived of the work, developed the mathematical approach, wrote the computer code, analyzed the primary data, interpreted the data analysis, and wrote the manuscript.

EFM contributed primary data, interpreted the data analysis, and edited the manuscript.
AJR contributed primary data, interpreted the data analysis, and edited the manuscript.
ES interpreted the data analysis and edited the manuscript.
LA interpreted the data analysis and edited the manuscript.
NT analyzed primary data.
SG analyzed primary data.
RNG conceived of the work, interpreted the data analysis, and edited the manuscript.

Human Data Approval
All data from human samples were previously published in [McCaffrey et al. 2022], for which they were collected under institutional review board protocol 46586 ('Generation of an Immune Atlas of Human Tuberculosis Granulomas with Multiplexed Ion Beam Imaging'). Patient consent was not required, because the clinical specimens were from archives, with no participation of human subjects.

Data and Code Availability
All data used in this study are publicly available from their original studies: mouse data are available in [Radtke et al 2020] and human data are available in [McCaffrey et al 2022]. The SPACE R package and documentation is currently available by request only, but it will be released freely and publicly on Github after peer review and publication, along with the complete R code used for all analyses in this study.

FIGURES AND TABLES

Table 1. Comparison of SPACE with existing spatial analysis platforms. "Input Data" refers to the actual data on which statistical calculations are performed. "Pattern Detection" refers to patterns that are systematically explored and reported.

Program Details		**Input Data**					**Pattern Detection**				
Name	**Platform**	**Full Images**			**Centroid Tables**		**Plexity**		**Complexity**		**Microenvironments**
		Expression	**Objects**	**Pixels**	**Expression**	**Objects**	**1-2**	**3+**	**Linear**	**Nonlinear**	**Categorical**
SPACE	R	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
CODEX Toolkit	GUI	no	no	no	yes	yes	yes	no	yes	no	yes
CytoMAP	Matlab	no	no	no	yes	yes	yes	no	yes	no	yes
Giotto	R	no	no	no	yes	yes	yes	no	yes	no	yes
HistoCAT	GUI	no	no	no	yes	yes	yes	no	yes	no	no
ImaCytE	GUI	no	no	no	no	yes	yes	no	yes	no	no
SIMPLI	Nextflow	no	no	no	yes	yes	yes	no	yes	no	no
SPIAT	R	no	no	no	yes	yes	yes	no	yes	no	yes
Visinity	GUI	no	no	no	no	yes	yes	no	yes	no	yes

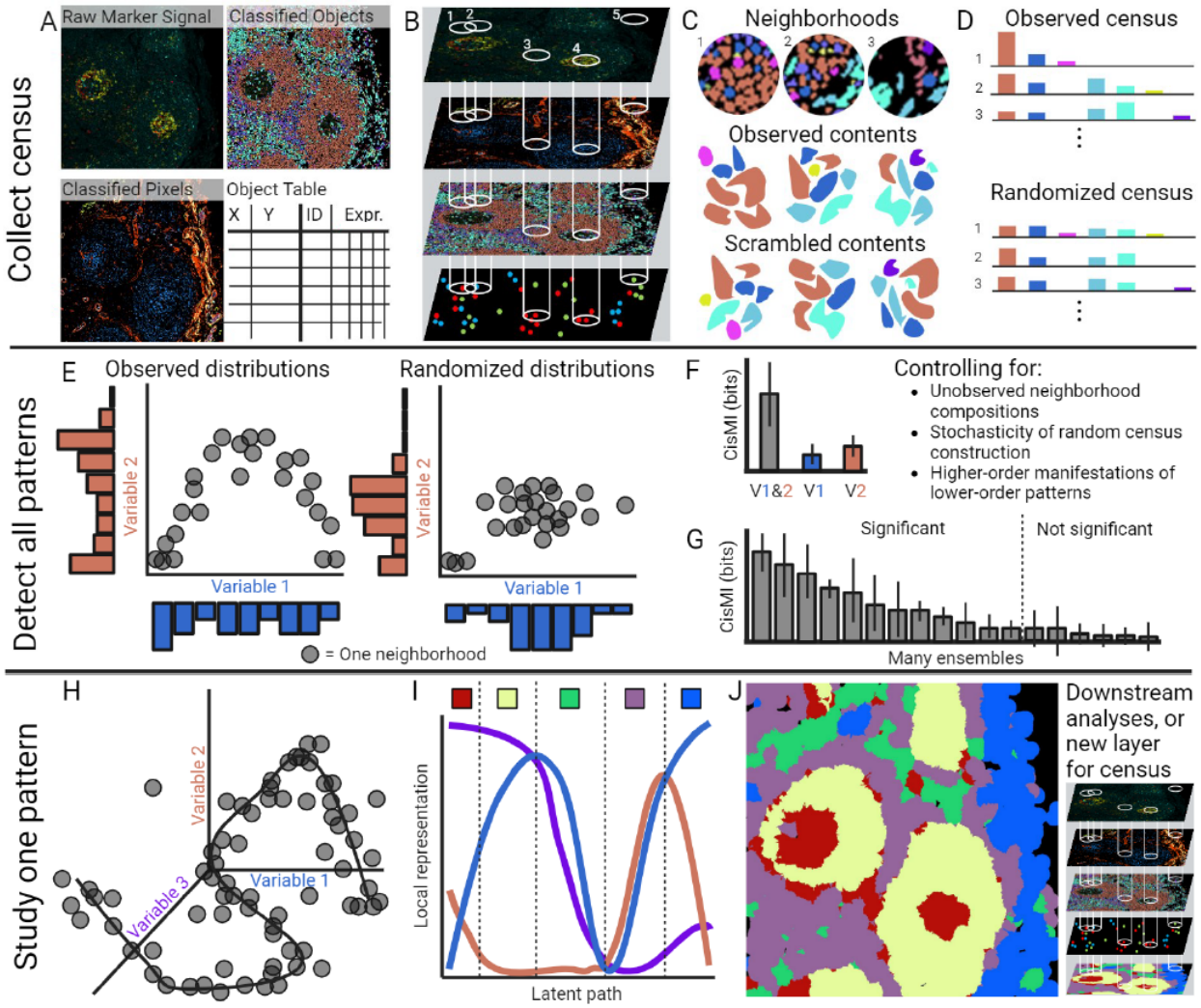

Figure 1. SPACE workflow. A) SPACE accepts images of molecular expression, images of segmented and classified cells, images of classified pixels, and centroid tables of cellular phenotypes and/or mean molecular expression. B) Neighborhoods are drawn through each layer of provided data. C) Neighborhood contents are tabulated, as are contiguous patches of each variable, which are scrambled to form randomized neighborhoods. D) True neighborhood contents form the observed census, while scrambled neighborhood contents form randomized censuses. E) For a given ensemble, observed and randomized distributions of co-occurrence are generated. F) The KL divergence between the observed vs. randomized version of each distribution is calculated. G) KL divergences are adjusted for several factors to yield cisMI, which is then used to rank the ensembles with the strongest spatial patterning. H) For a given ensemble, a self-organizing map learns the shape of non-random patterning as a curve in high dimension. I) The high dimensional curve is projected onto the covariation plot revealing specific features of the ensemble's patterning. J) Manually-defined regions are mapped back onto the tissue for further rounds of SPACE analysis. Created with Biorender.com.

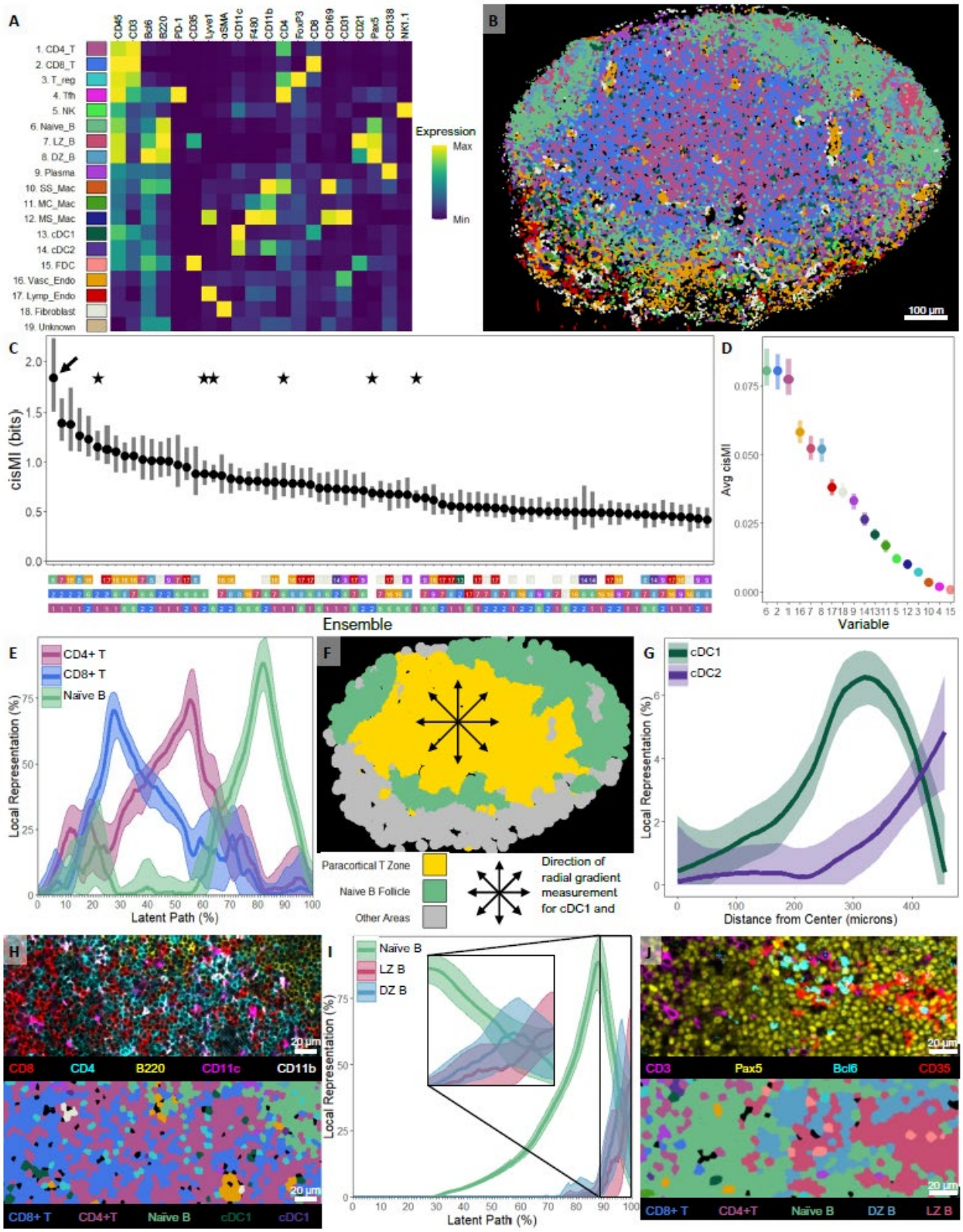

Figure 2. Validation of SPACE on a mouse LN. A) Marker expression profiles of 19 segmented cell types. B) Segmentation image used in SPACE analysis. C) SPACE quantification of cisMI for the 75 most significantly non-randomly patterned ensembles. D) Average cisMI for each cell type across all significant ensembles. E) Covariation plot for CD4+ T cells, CD8+ T cells, and naïve B cells. F) Zone image obtained from partitioning the covariation plot into T- vs. B-cell dominated areas. G) Gradients of cDC1 and

cDC2 abundance from the interior to the exterior of the mapped T zone. H) Visual example of the CD4+ T vs. CD8+ T and cDC2 vs. cDC1 gradients in the T cell zone, with respect to the B follicle. I) Covariation plot for naïve B, LZ B, and DZ B cells. J) Visual example of the orientation of GCs with the DZ facing the interior of the naïve B follicle and the LZ facing the exterior.

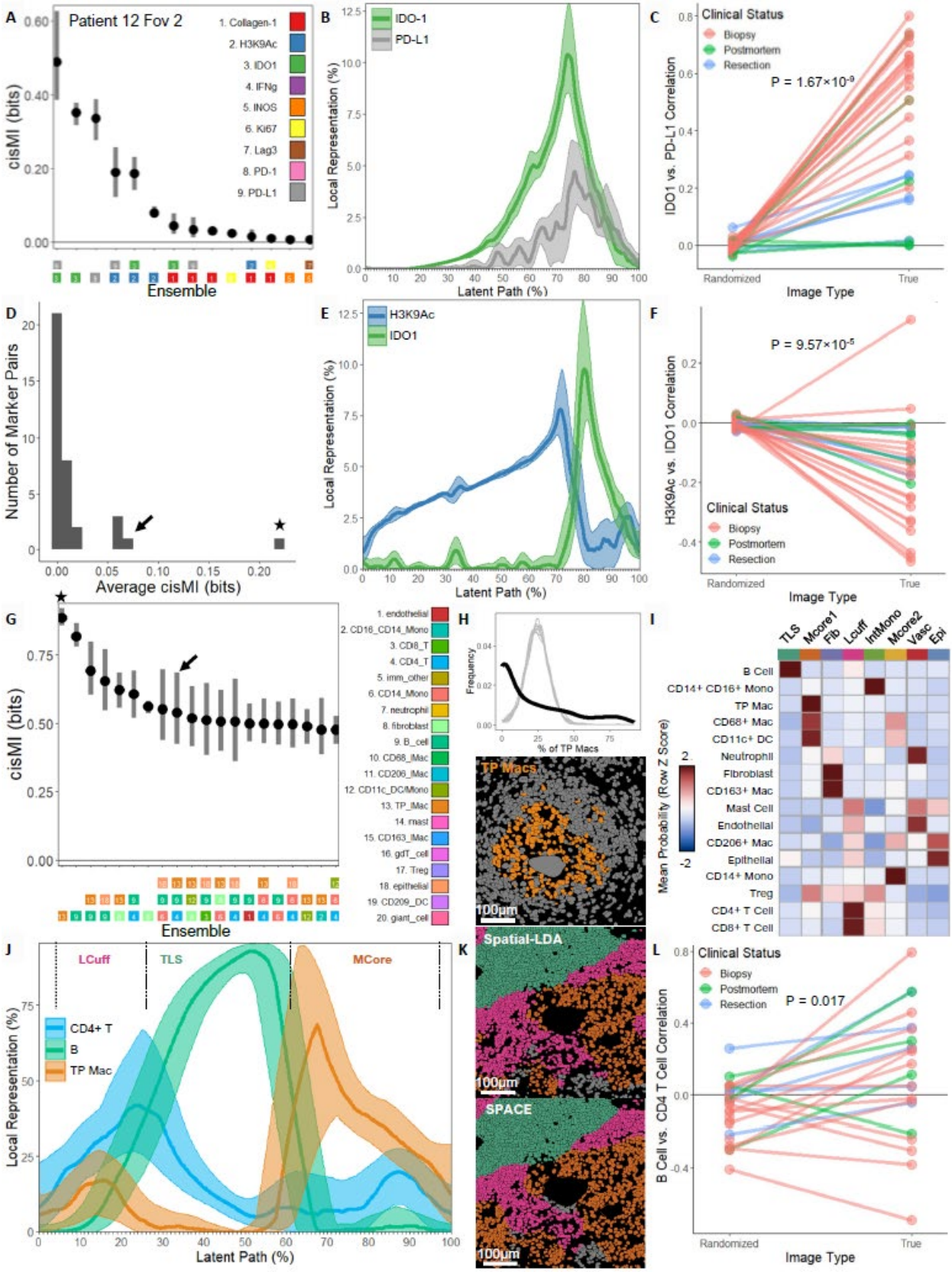

Figure 3. SPACE re-analysis of human TB granulomas. A) SPACE quantification of cisMI for all significant ensembles of functional markers in Patient 12 FoV 2. B) Covariation plot for PD-L1 and IDO1 in Patient 12 FoV 2. C) Correlation of PD-L1 and IDO1

expression in the true vs. randomized images across the full data set (paired T test). D) Average cisMI across all FoVs for each pair of functional markers, including PD-L1 and IDO1 (highest average cisMI) as well as H3K9Ac and IDO1 (second-highest average cisMI). E) Covariation plot of H3K9Ac and IDO1 in Patient 12 FoV 1. F) Correlation of H3K9Ac and IDO1 in the true vs. randomized images across the full data set (paired T test). G) The 25 strongest patterns among segmented cell types in terms of cisMI averaged across the full data set. H) Observed (black) vs. randomized (gray) distribution of TP macrophage abundance in 50µm neighborhoods in Patient 10 FoV 1. I) Compositions of the MEs discovered by Spatial-LDA in the original study, defined by the prevalence of 16 different segmented cell types. J) Covariation plot for TP Macs, B cells, and CD4+ T cells in Patient 8 FoV1, with the gates used to define the Lcuff, TLS, and Mcore1 MEs. K) Lcuff, TLS, and Mcore1 MEs for Patient 8 FoV 1 as drawn by Spatial-LDA vs. SPACE. L) Correlation of B cell and CD4+ T cell abundance specifically in 50µm neighborhoods where TP Macs occur outside the Mcore1 ME in the true vs. randomized images across the full data set (paired T test).

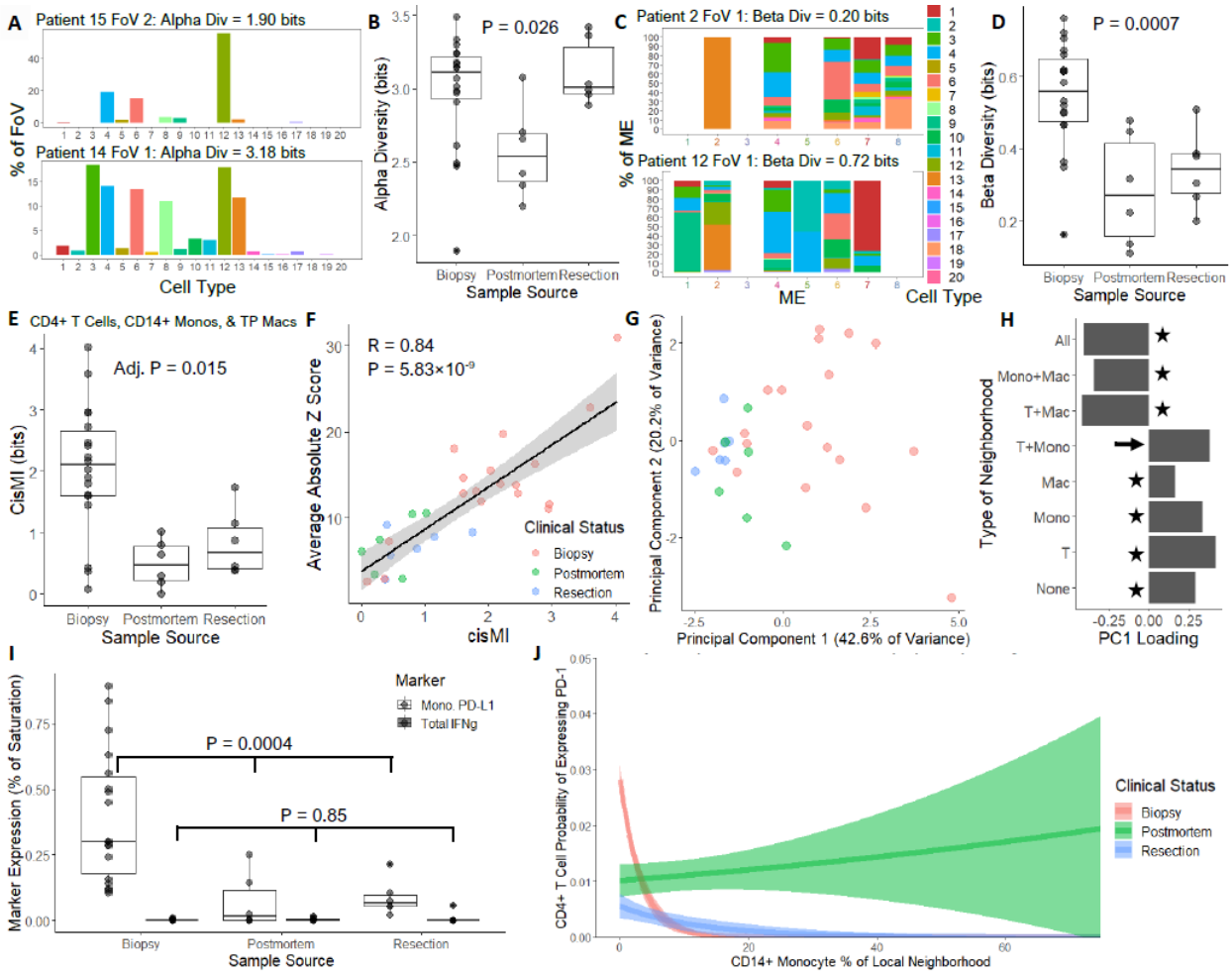

Figure 4. Novel correlates of TB disease progression discovered with SPACE. A) Example FoVs with low and high cellular alpha diversity. B) Comparison of cellular alpha diversity across clinical statuses (ANOVA). C) Example FoVs with low and high cellular beta diversity across MEs. D) Comparison of cellular beta diversity of MEs across clinical statuses (ANOVA). E) CisMI for the tripartite ensemble of CD4+ T cells, CD14+ monocytes, and TP macrophages compared across clinical statuses (ANOVA). F) Correlation of average absolute Z scores for the prevalence of each neighborhood category with cisMI. G) PCA of FoVs using the Z scores for each neighborhood category. H) Loadings for Principal Component 1. I) Expression of PD-L1 specifically on CD14+ monocytes and expression of IFNγ at large compared across clinical statuses (ANOVA). J) Logistic regression model of PD-1 expression probability on CD4+ T cells as a function of the local CD14+ monocyte abundance across clinical statuses ($P < 0.01$ for all main and interaction terms in the model).

REFERENCES

1. Baccin C, Al-Sabah J, Velten L, Helbling PM, Grünschläger F, Hernández-Malmierca P, Nombela-Arrieta C, Steinmetz LM, Trumpp A, Haas S. Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. Nat Cell Biol. 2020 Jan;22(1):38-48. doi: 10.1038/s41556-019-0439-6. Epub 2019 Dec 23. PMID: 31871321; PMCID: PMC7610809.
2. Saviano A, Henderson NC, Baumert TF. Single-cell genomics and spatial transcriptomics: Discovery of novel cell states and cellular interactions in liver physiology and disease biology. J Hepatol. 2020 Nov;73(5):1219-1230. doi: 10.1016/j.jhep.2020.06.004. Epub 2020 Jun 10. PMID: 32534107; PMCID: PMC7116221.
3. Lewis SM, Asselin-Labat ML, Nguyen Q, Berthelet J, Tan X, Wimmer VC, Merino D, Rogers KL, Naik SH. Spatial omics and multiplexed imaging to explore cancer biology. Nat Methods. 2021 Sep;18(9):997-1012. doi: 10.1038/s41592-021-01203-6. Epub 2021 Aug 2. PMID: 34341583.
4. Rendeiro AF, Ravichandran H, Bram Y, Chandar V, Kim J, Meydan C, Park J, Foox J, Hether T, Warren S, Kim Y, Reeves J, Salvatore S, Mason CE, Swanson EC, Borczuk AC, Elemento O, Schwartz RE. The spatial landscape of lung pathology during COVID-19 progression. Nature. 2021 May;593(7860):564-569. doi: 10.1038/s41586-021-03475-6. Epub 2021 Mar 29. PMID: 33780969; PMCID: PMC8204801.
5. Germain RN, Radtke AJ, Thakur N, Schrom EC, Hor JL, Ichise H, Arroyo-Mejias AJ, Chu CJ, Grant S. Understanding immunity in a tissue-centric context: Combining novel imaging methods and mathematics to extract new insights into function and dysfunction. Immunol Rev. 2022 Mar;306(1):8-24. doi: 10.1111/imr.13052. Epub 2021 Dec 16. PMID: 34918351.
6. Lin JR, Fallahi-Sichani M, Chen JY, Sorger PK. Cyclic Immunofluorescence (CycIF), A Highly Multiplexed Method for Single-cell Imaging. Curr Protoc Chem Biol. 2016 Dec 7;8(4):251-264. doi: 10.1002/cpch.14. PMID: 27925668; PMCID: PMC5233430.
7. Radtke AJ, Chu CJ, Yaniv Z, Yao L, Marr J, Beuschel RT, Ichise H, Gola A, Kabat J, Lowekamp B, Speranza E, Croteau J, Thakur N, Jonigk D, Davis JL, Hernandez JM, Germain RN. IBEX: an iterative immunolabeling and chemical bleaching method for high-content imaging of diverse tissues. Nat Protoc. 2022 Feb;17(2):378-401. doi: 10.1038/s41596-021-00644-9. Epub 2022 Jan 12. PMID: 35022622.
8. Black S, Phillips D, Hickey JW, Kennedy-Darling J, Venkataraaman VG, Samusik N, Goltsev Y, Schürch CM, Nolan GP. CODEX multiplexed tissue imaging with DNA-conjugated antibodies. Nat Protoc. 2021 Aug;16(8):3802-3835. doi: 10.1038/s41596-021-00556-8. Epub 2021 Jul 2. PMID: 34215862; PMCID: PMC8647621.
9. Giesen C, Wang HA, Schapiro D, Zivanovic N, Jacobs A, Hattendorf B, Schüffler PJ, Grolimund D, Buhmann JM, Brandt S, Varga Z, Wild PJ, Günther D, Bodenmiller B. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat Methods. 2014 Apr;11(4):417-22. doi: 10.1038/nmeth.2869. Epub 2014 Mar 2. PMID: 24584193.
10. Keren L, Bosse M, Thompson S, Risom T, Vijayaragavan K, McCaffrey E, Marquez D, Angoshtari R, Greenwald NF, Fienberg H, Wang J, Kambham N, Kirkwood D, Nolan G, Montine TJ, Galli SJ, West R, Bendall SC, Angelo M. MIBI-TOF: A multiplexed imaging platform relates cellular phenotypes and tissue structure. Sci Adv. 2019 Oct 9;5(10):eaax5851. doi: 10.1126/sciadv.aax5851. PMID: 31633026; PMCID: PMC6785247.
11. Eng CL, Lawson M, Zhu Q, Dries R, Koulena N, Takei Y, Yun J, Cronin C, Karp C, Yuan GC, Cai L. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. Nature. 2019 Apr;568(7751):235-239. doi: 10.1038/s41586-019-1049-y. Epub 2019 Mar 25. PMID: 30911168; PMCID: PMC6544023.

12. Xia C, Fan J, Emanuel G, Hao J, Zhuang X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proc Natl Acad Sci U S A. 2019 Sep 24;116(39):19490-19499. doi: 10.1073/pnas.1912459116. Epub 2019 Sep 9. PMID: 31501331; PMCID: PMC6765259.
13. Moffitt JR, Lundberg E, Heyn H. The emerging landscape of spatial profiling technologies. Nat Rev Genet. 2022 Dec;23(12):741-759. doi: 10.1038/s41576-022-00515-3. Epub 2022 Jul 20. PMID: 35859028.
14. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, Mollbrink A, Linnarsson S, Codeluppi S, Borg Å, Pontén F, Costea PI, Sahlén P, Mulder J, Bergmann O, Lundeberg J, Frisén J. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science. 2016 Jul 1;353(6294):78-82. doi: 10.1126/science.aaf2403. PMID: 27365449.
15. Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature. 2021 Aug;596(7871):211-220. doi: 10.1038/s41586-021-03634-9. Epub 2021 Aug 11. PMID: 34381231; PMCID: PMC8475179.
16. Vandereyken K, Sifrim A, Thienpont B, Voet T. Methods and applications for single-cell and spatial multi-omics. Nat Rev Genet. 2023 Aug;24(8):494-515. doi: 10.1038/s41576-023-00580-2. Epub 2023 Mar 2. PMID: 36864178; PMCID: PMC9979144.
17. Schapiro D, Jackson HW, Raghuraman S, Fischer JR, Zanotelli VRT, Schulz D, Giesen C, Catena R, Varga Z, Bodenmiller B. histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. Nat Methods. 2017 Sep;14(9):873-876. doi: 10.1038/nmeth.4391. Epub 2017 Aug 7. PMID: 28783155; PMCID: PMC5617107.
18. Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G, Black S, Nolan GP. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. Cell. 2018 Aug 9;174(4):968-981.e15. doi: 10.1016/j.cell.2018.07.010. Epub 2018 Aug 2. PMID: 30078711; PMCID: PMC6086938.
19. Stoltzfus CR, Filipek J, Gern BH, Olin BE, Leal JM, Wu Y, Lyons-Cohen MR, Huang JY, Paz-Stoltzfus CL, Plumlee CR, Pöschinger T, Urdahl KB, Perro M, Gerner MY. CytoMAP: A Spatial Analysis Toolbox Reveals Features of Myeloid Cell Organization in Lymphoid Tissues. Cell Rep. 2020 Apr 21;31(3):107523. doi: 10.1016/j.celrep.2020.107523. PMID: 32320656; PMCID: PMC7233132.
20. Dries R, Zhu Q, Dong R, Eng CL, Li H, Liu K, Fu Y, Zhao T, Sarkar A, Bao F, George RE, Pierson N, Cai L, Yuan GC. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. Genome Biol. 2021 Mar 8;22(1):78. doi: 10.1186/s13059-021-02286-2. PMID: 33685491; PMCID: PMC7938609.
21. Bhate SS, Barlow GL, Schürch CM, Nolan GP. Tissue schematics map the specialization of immune tissue motifs and their appropriation by tumors. Cell Syst. 2022 Feb 16;13(2):109-130.e6. doi: 10.1016/j.cels.2021.09.012. Epub 2021 Oct 14. PMID: 34653369.
22. Bortolomeazzi M, Montorsi L, Temelkovski D, Keddar MR, Acha-Sagredo A, Pitcher MJ, Basso G, Laghi L, Rodriguez-Justo M, Spencer J, Ciccarelli FD. A SIMPLI (Single-cell Identification from MultiPLexed Images) approach for spatially-resolved tissue phenotyping at single-cell resolution. Nat Commun. 2022 Feb 9;13(1):781. doi: 10.1038/s41467-022-28470-x. PMID: 35140207; PMCID: PMC8828885.
23. Warchol S, Krueger R, Nirmal AJ, Gaglia G, Jessup J, Ritch CC, Hoffer J, Muhlich J, Burger ML, Jacks T, Santagata S, Sorger PK, Pfister H. Visinity: Visual Spatial Neighborhood Analysis for Multiplexed Tissue Imaging Data. IEEE Trans Vis Comput Graph. 2023 Jan;29(1):106-116. doi: 10.1109/TVCG.2022.3209378. Epub 2022 Dec 16. PMID: 36170403; PMCID: PMC10043053.

24. Chen Z, Soifer I, Hilton H, Keren L, Jojic V. Modeling Multiplexed Images with Spatial-LDA Reveals Novel Tissue Microenvironments. J Comput Biol. 2020 Aug;27(8):1204-1218. doi: 10.1089/cmb.2019.0340. Epub 2020 Apr 3. PMID: 32243203; PMCID: PMC7415889.
25. Schürch CM, Bhate SS, Barlow GL, Phillips DJ, Noti L, Zlobec I, Chu P, Black S, Demeter J, McIlwain DR, Kinoshita S, Samusik N, Goltsev Y, Nolan GP. Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. Cell. 2020 Sep 3;182(5):1341-1359.e19. doi: 10.1016/j.cell.2020.07.005. Epub 2020 Aug 6. Erratum in: Cell. 2020 Oct 29;183(3):838. PMID: 32763154; PMCID: PMC7479520.
26. McCaffrey EF, Donato M, Keren L, Chen Z, Delmastro A, Fitzpatrick MB, Gupta S, Greenwald NF, Baranski A, Graf W, Kumar R, Bosse M, Fullaway CC, Ramdial PK, Forgó E, Jojic V, Van Valen D, Mehra S, Khader SA, Bendall SC, van de Rijn M, Kalman D, Kaushal D, Hunter RL, Banaei N, Steyn AJC, Khatri P, Angelo M. The immunoregulatory landscape of human tuberculosis granulomas. Nat Immunol. 2022 Feb;23(2):318-329. doi: 10.1038/s41590-021-01121-x. Epub 2022 Jan 20. Erratum in: Nat Immunol. 2022 May;23(5):814. PMID: 35058616; PMCID: PMC8810384.
27. Nirmal AJ, Maliga Z, Vallius T, Quattrochi B, Chen AA, Jacobson CA, Pelletier RJ, Yapp C, Arias-Camison R, Chen YA, Lian CG, Murphy GF, Santagata S, Sorger PK. The Spatial Landscape of Progression and Immunoediting in Primary Melanoma at Single-Cell Resolution. Cancer Discov. 2022 Jun 2;12(6):1518-1541. doi: 10.1158/2159-8290.CD-21-1357. PMID: 35404441; PMCID: PMC9167783.
28. Radtke AJ, Postovalova E, Varlamova A, Bagaev A, Sorokina M, Kudryashova O, Meerson M, Polyakova M, Galkin I, Svekolkin V, Isaev S, Perelman G, Lozinsky Y, Yaniv Z, Lowekamp BC, Speranza E, Yao L, Pittaluga S, Shaffer III AL, Jonigk D, Phelan JD, Davies-Hill T, Huang DW, Ovcharov P, Nomie M, Nuzhdina E, Kotlov N, Ataullakhanov R, Fowler N, Kelly M, Muppidi J, Davis J, Hernandez JM, Wilson WH, Jaffe ES, Staudt LM, Roschewski M, Germain RN. A multi-scale, multiomic atlas of human normal and follicular lymphoma lymph nodes. bioRxiv. 2023. doi: 10.1101/2022.06.03.494716
29. Halkidi M, Batistakis Y, Vazirgiannis M. On clustering validation techniques. J Intell Inf Syst. 2001 Dec;17:107-45. doi: 10.1023/A:1012801612483.
30. Bayá AE, Granitto PM. How many clusters: a validation index for arbitrary-shaped clusters. IEEE/ACM Trans Comput Biol Bioinform. 2013 Mar-Apr;10(2):401-14. doi: 10.1109/TCBB.2013.32. PMID: 23929864.
31. World Health Organization. Global Tuberculosis Report. 2022. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022
32. Flynn JL, Chan J. Immune cell interactions in tuberculosis. Cell. 2022 Dec 8;185(25):4682-4702. doi: 10.1016/j.cell.2022.10.025. PMID: 36493751.
33. Pachitariu M, Stringer C. Cellpose 2.0: how to train your own model. Nat Methods. 2022 Dec;19(12):1634-1641. doi: 10.1038/s41592-022-01663-4. Epub 2022 Nov 7. PMID: 36344832; PMCID: PMC9718665.
34. Radtke AJ, Kandov E, Lowekamp B, Speranza E, Chu CJ, Gola A, Thakur N, Shih R, Yao L, Yaniv ZR, Beuschel RT, Kabat J, Croteau J, Davis J, Hernandez JM, Germain RN. IBEX: A versatile multiplex optical imaging approach for deep phenotyping and spatial analysis of cells in complex tissues. Proc Natl Acad Sci U S A. 2020 Dec 29;117(52):33455-33465. doi: 10.1073/pnas.2018488117. Epub 2020 Dec 21. PMID: 33376221; PMCID: PMC7776876.
35. Baptista AP, Gola A, Huang Y, Milanez-Almeida P, Torabi-Parizi P, Urban JF Jr, Shapiro VS, Gerner MY, Germain RN. The Chemoattractant Receptor Ebi2 Drives Intranodal Naive CD4+ T Cell Peripheralization to Promote Effective Adaptive Immunity. Immunity. 2019 May 21;50(5):1188-1201.e6. doi: 10.1016/j.immuni.2019.04.001. Epub 2019 Apr 30. PMID: 31053504.

36. Gerner MY, Kastenmuller W, Ifrim I, Kabat J, Germain RN. Histo-cytometry: a method for highly multiplex quantitative tissue imaging analysis applied to dendritic cell subset microanatomy in lymph nodes. Immunity. 2012 Aug 24;37(2):364-76. doi: 10.1016/j.immuni.2012.07.011. Epub 2012 Aug 2. PMID: 22863836; PMCID: PMC3514885.
37. Li W, Germain RN, Gerner MY. High-dimensional cell-level analysis of tissues with Ce3D multiplex volume imaging. Nat Protoc. 2019 Jun;14(6):1708-1733. doi: 10.1038/s41596-019-0156-4. Epub 2019 Apr 26. PMID: 31028373; PMCID: PMC8690297.
38. Chao A, Jost L. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. Ecology. 2012 Dec;93(12):2533-47. doi: 10.1890/11-1952.1. PMID: 23431585.