Original Article

Pointwise mutual information quantifies intratumor heterogeneity in tissue sections labeled with multiple fluorescent biomarkers

Daniel M. Spagnolo[1,2], Rekha Gyanchandani[3], Yousef Al-Kofahi[4], Andrew M. Stern[2,5], Timothy R. Lezon[2,5], Albert Gough[2,5], Dan E. Meyer[4], Fiona Ginty[4], Brion Sarachan[6], Jeffrey Fine[7], Adrian V. Lee[3,8], D. Lansing Taylor[2,5,8], S. Chakra Chennubhotla[2]

[1]Program in Computational Biology, Joint Carnegie Mellon University-University of Pittsburgh, Departments of [2]Computational and Systems Biology and [3]Pharmacology and Chemical Biology, University of Pittsburgh, [5]Drug Discovery Institute, University of Pittsburgh, [7]Department of Pathology, School of Medicine, University of Pittsburgh, [8]University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania, [4]GE Global Research Center, Diagnostics, Imaging and Biomedical Technologies, [6]GE Global Research Center, Software Science and Analytics Organization, Niskayuna, NY, USA

E-mail: *Prof. S. Chakra Chennubhotla - chakracs@pitt.edu
*Corresponding author

Abstract

Background: Measures of spatial intratumor heterogeneity are potentially important diagnostic biomarkers for cancer progression, proliferation, and response to therapy. Spatial relationships among cells including cancer and stromal cells in the tumor microenvironment (TME) are key contributors to heterogeneity. **Methods:** We demonstrate how to quantify spatial heterogeneity from immunofluorescence pathology samples, using a set of 3 basic breast cancer biomarkers as a test case. We learn a set of dominant biomarker intensity patterns and map the spatial distribution of the biomarker patterns with a network. We then describe the pairwise association statistics for each pattern within the network using pointwise mutual information (PMI) and visually represent heterogeneity with a two-dimensional map. **Results:** We found a salient set of 8 biomarker patterns to describe cellular phenotypes from a tissue microarray cohort containing 4 different breast cancer subtypes. After computing PMI for each pair of biomarker patterns in each patient and tumor replicate, we visualize the interactions that contribute to the resulting association statistics. Then, we demonstrate the potential for using PMI as a diagnostic biomarker, by comparing PMI maps and heterogeneity scores from patients across the 4 different cancer subtypes. Estrogen receptor positive invasive lobular carcinoma patient, AL13-6, exhibited the highest heterogeneity score among those tested, while estrogen receptor negative invasive ductal carcinoma patient, AL13-14, exhibited the lowest heterogeneity score. **Conclusions:** This paper presents an approach for describing intratumor heterogeneity, in a quantitative fashion (via PMI), which departs from the purely qualitative approaches currently used in the clinic. PMI is generalizable to highly multiplexed/hyperplexed immunofluorescence images, as well as spatial data from complementary in situ methods including FISSEQ and CyTOF, sampling many different components within the TME. We hypothesize that PMI will uncover key spatial interactions in the TME that contribute to disease proliferation and progression.

Key words: Computational pathology, multiplexed immunofluorescence, pointwise mutual information, tumor heterogeneity, tumor microenvironment

INTRODUCTION

For many malignancies, molecular and cellular heterogeneity is a prominent feature among tumors from different patients, between different sites of neoplasia in a single patient and within a single tumor.[1] Intratumor heterogeneity involves phenotypically distinct cancer cell clonal subpopulations and other cell types that include local and bone marrow-derived stromal stem and progenitor cells, subclasses of immune inflammatory cells that are either tumor promoting or tumor-killing, cancer-associated fibroblasts, endothelial cells, and pericytes that comprise the tumor microenvironment (TME) or "tumor tissue system."[2-5] The TME can be viewed as an evolving ecosystem where cancer cells engage in heterotypic interactions with these other cell types and use available resources to proliferate and survive.[6,7] Consistent with this perspective, the spatial relationships among the cell types within the TME (i.e. spatial heterogeneity) appear to be one of the main drivers of disease progression and therapy resistance.[8-12] Thus, it is imperative to define the spatial heterogeneity within the TME to properly diagnose the specific disease subtype and identify the optimal course of therapy for individual patients.

To date, intratumor heterogeneity has been explored using three major approaches [Figure 1]. The first approach is to take core samples from specific regions of tumors to measure population averages. Heterogeneity is measured by analyzing multiple cores within the tumor. The specific analyses include whole exome sequencing,[13-16] epigenetics,[17] proteomics,[18,19] and metabolomics [Figure 1b].[19] The second approach involves "single cell analyses" using the above methods,[20,21] RNA-Seq,[22] imaging,[23] or flow cytometry[24] after separation of the cells from the tissue. The third approach uses the spatial resolution of light microscope imaging to maintain spatial context and is coupled with molecular-specific labels to measure biomarkers in the cells *in situ*.[3,25-27]
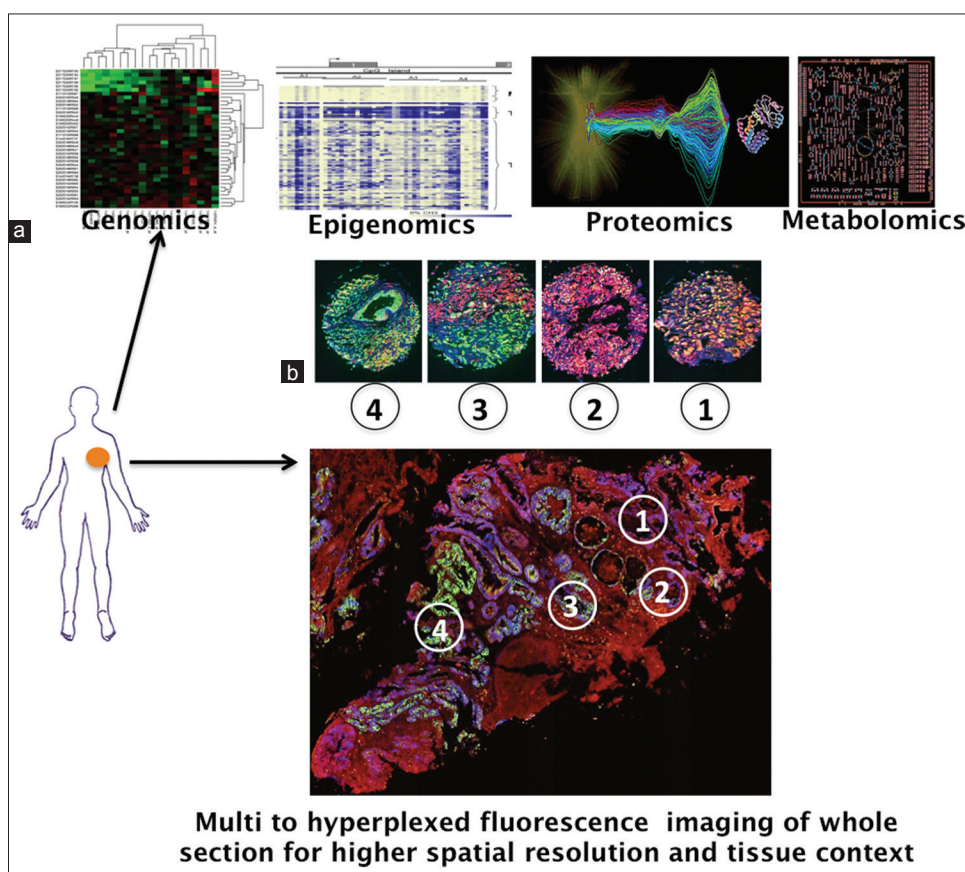
Figure 1: Quantifying spatial intratumor heterogeneity. (a) The current state-of-the-art uses genomics, epigenomics, proteomics, and/or metabolomics to study heterogeneity on either ground up tissue samples or single cells. These approaches do not account for the spatial organization of the tumor microenvironment. (b) We present a method using multiplexed immunofluorescence imaging, which incorporates spatial distribution of biomarkers, in addition to their intensities, to characterize spatial intratumor heterogeneity. As shown in regions 1–4, the spatial organization of biomarker signals varies across the sample. Our method can capture this variation in a whole slide sample and is applicable to single-protein, multiplexed (up to 7 biomarkers), and hyperplexed (>7 biomarkers) immunofluorescence images. Furthermore, our method may be applied beyond the realm of cellular constituents, where we can study spatial interactions between cells and noncellular components (e.g., secretory elements, extracellular matrix)

Spatial analyses using light microscope imaging facilitate analysis of large areas of tissue sections and/or multiple tumor microarray sections at the cellular and subcellular levels. Subcellular resolution, for example, permits the identification of the activation state of specific biomarkers (e.g., translocation of transcription factors into the nucleus).[28] In addition, recent developments in mass spectrometry imaging permit many cellular constituents to be measured across a tissue section but at a lower resolution than optical microscopy.[29]

Several light microscopy imaging platforms have been developed to characterize cellular biomarker expression levels within tumors including transmitted light and fluorescence.[30] Multivariate information based on fluorescence has been acquired from images of large area tissue sections and tissue microarrays (TMAs) based on DNA, RNA, and protein biomarkers, usually from 1 up to 7 fluorescently labeled biomarkers in the same sample (multiplexed fluorescence).[28,31] Multiple commercial platforms can now be used to acquire, process, segment and perform some basic analyses of biomarker signal levels in tissue samples (e.g., Genoptix, Carlsbad, CA, USA; Olympus, Center Valley, NJ, USA; Carl Zeiss, Inc., Thornwood, NY, USA; Hamamatsu Photonics, K.K., Hamamatsu City, Japan; Leica Biosystems, Inc., Buffalo Grove, IL, USA). Recently, platforms have been demonstrated permit up to 60 fluorescently labeled antibodies and a few DNA or RNA hybridization probes to be acquired in an iterative cycle of labeling, imaging, and quenching fluorescence.[32,33] It is now possible to "map" the location of specific cell types, states of cell activations, cell biomarker expression levels, and localizations, as well as extracellular constituents in tissue sections and TMAs.

A major challenge is to develop algorithms that can quantify key spatial relationships (interactions and lack thereof) within the TME based on panels of biomarkers. Initial efforts in measuring heterogeneity in tissue sections applied diversity metrics from ecological studies, such as Shannon entropy and Rao's quadratic entropy (QE).[9,34-37] However, these methods have not been adapted for multiplexed (up to 7 biomarkers) or hyperplexed (>7 biomarkers) immunofluorescence (IF) data.[2] Other methods that account for high-dimensional data may not have sophisticated cell phenotyping methods, allowing each biomarker to be only "on" or "off."[38] Furthermore, a few of these methods incorporate the spatial relationships between biomarker patterns in their heterogeneity scores.[9,36] Indeed, the spatial organization of the TME has been hypothesized to be an important diagnostic biomarker in addition to the expression levels of selected biomarkers from both cancer and noncancer cells.

We have designed a method to quantify spatial intratumor heterogeneity [Figure 1]. The method can work with single biomarker, multiplexed, or hyperplexed IF data [Figure 2]. Other heterogeneity characterization methods, although insightful, may not incorporate spatial information or employ multiplexed methods, which implicate a larger number of biomarkers. One such method uses region of interest sampling to add spatial resolution, but this approach does not study spatial relationships between cell phenotypes, nor does it look at more than a single biomarker in its model of heterogeneity.[35] Another method does look at linear relationships among different biomarkers using multiplexed/hyperplexed IF data, but does not incorporate any spatial information, nor does it consider nonlinear associations.[39] Yet another looked at multiplexed phenotypic associations, in contrast to[39] which looked at biomarker associations, but also neglected spatial information.[40] Our method is holistic in its approach, using both the expression and spatial information of an entire tumor tissue section and/or spot in a TMA to characterize spatial associations. In addition, most other methods report intratumor heterogeneity as a single score, thus potentially mapping two spatially different organizations of the TMEs incorrectly to the same score. In comparison, we generate a two-dimensional (2D) heterogeneity map to explicitly elucidate spatial associations of both major and minor subpopulations. We hypothesize that the characterization of spatial intratumor heterogeneity will be an important diagnostic biomarker for cancer progression, proliferation, and response to therapy.

In this paper, the spatial intratumor heterogeneity measure we developed uses data processed on a TMA. We introduce these methods as a proof-of-concept, where we demonstrate the ability to quantify spatial heterogeneity using replicate cores of patient tumors and three breast cancer biomarkers (estrogen receptor [ER], human epidermal growth factor 2 [HER2], and progesterone receptor [PR]) combined with biomarkers for segmentation including the nucleus, plasma membrane, cytoplasm, and epithelial cells (see below). The impact of our method, using pointwise mutual information (PMI) to quantify spatial intratumor heterogeneity, will be extended in future studies to the analysis of whole-slide IF images, labeled with increasing numbers of cancer and stromal biomarkers.

METHODS

Tissue Microarray Preparation

A TMA of 99 spots (plus orientation cores) consisting of triplicate, 1 mm diameter cores from 24 invasive breast tumor tissues was constructed by the Tissue and Research Pathology Core Facility at the University of Pittsburgh Cancer Institute. Formalin-fixed paraffin-embedded (FFPE) tumor blocks were obtained from the University of Pittsburgh Health Sciences Tissue
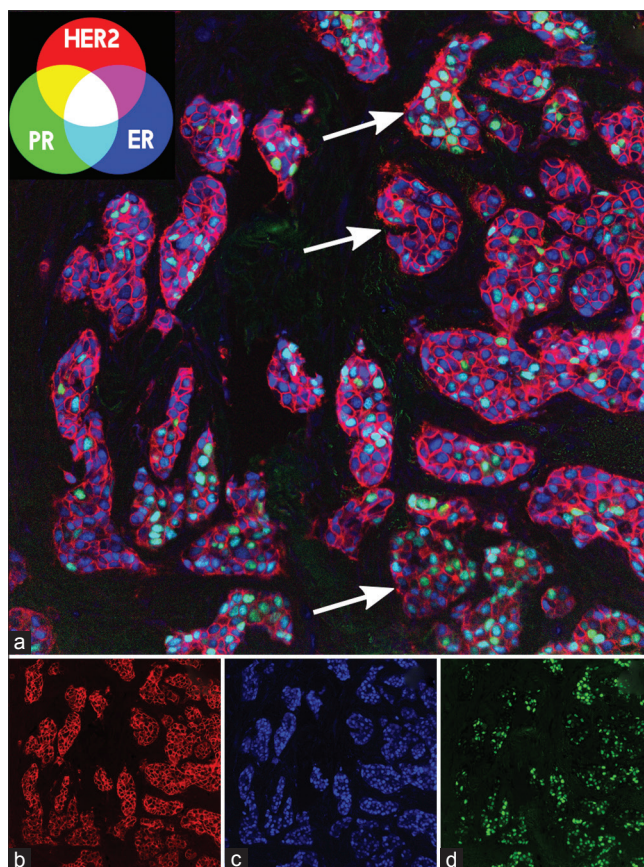
Figure 2: Multiplexed immunofluorescence image of a tissue microarray spot. (a) Pseudocolored multi-channel fluorescence image is shown for the estrogen receptor(+) invasive ductal carcinoma spot 55 of the tissue microarray [Table 1]. Human epidermal growth factor 2 is shown in red, estrogen receptor in blue, and progesterone receptor in green (see legend in the top left corner). Areas of progesterone receptor/estrogen receptor coexpression will appear in cyan, human epidermal growth factor 2/estrogen receptor co-expression in magenta, and progesterone receptor/human epidermal growth factor 2 coexpression in yellow. Arrows point to three different heterogeneous regions in the tumor sample, with varying populations of estrogen receptor(+), progesterone receptor(+), and estrogen receptor(+)/progesterone receptor(+) cells. The upper arrow indicates a tumor microdomain with higher than average estrogen receptor(+)/progesterone receptor(+) phenotyped cells. The middle arrow indicates a microdomain probably the best representative of the tumor sample en masse, containing mostly estrogen receptor(+) cells. In the third microdomain, indicated by the lower arrow, there is a higher than average population of progesterone receptor(+) cells. (b-d) Individual pseudocolored fluorescence images are shown for human epidermal growth factor 2, estrogen receptor, and progesterone receptor

Bank under the Institutional Review Board approved protocol (PRO13080285). Immunohistochemical expression levels for ER, PR, and HER2 were assessed according to the American Society of Clinical Oncology/College of American Pathologists guidelines. In total, three cases of ER(+) invasive ductal carcinoma (IDC), five cases of ER(+) invasive lobular carcinoma (ILC), eight cases of ER(−) IDC, and eight cases of HER2(+) IDC were each cored in triplicate for a total of 72 tumor tissue spots. The remaining 27 spots consisted of 1 mm cores from cell pellets made from MCF7, MCF10A, MDA-MB-231, and MDA-MB-468 breast cancer cell lines (Breast Cancer Panel, ATCC, Manassas, VA, USA), which were included as staining controls for the multiplexed IF study. Locations of all cores within the array were randomized. As an additional positive staining control, we also purchased a commercial breast cancer tissue array (BRC482 from Pantomics, Inc., Richmond, CA, USA).

Multiplexed Immunofluorescence Staining and Imaging

Multiplexed IF staining and imaging of the study and control microarray slides were conducted as described previously.[32] Two sequential 5 μm sections of the study array on positively charged glass microscope slides and one BRC482 control slide were taken through the process in parallel. The methods for slide preparation and iterative rounds of staining and imaging were described in detail by Gerdes *et al.*[32] and issued patents referenced therein. Briefly, slides were cleared of paraffin, subjected to a 2-step antigen retrieval protocol, and blocked with donkey serum and bovine serum albumin solution. A total of four rounds of staining, imaging, and dye deactivation were completed: Round 1: ribosomal protein S6 (Cy3 conjugate of rabbit anti-phospho-40S ribosomal protein S6 [Ser-240/244], #2215, cell signaling, used at 10 μg/mL), ER (Cy5 conjugate of mouse anti-ER-α, clone

1D5, #M7047, DAKO, used at 10 µg/mL); Round 2: PR (Cy3 conjugate, 13095, used at 5 µg/mL), HER2 (Cy5 conjugate of rabbit anti-HER2 clone D8F12.#4290, cell signaling, used at 5 µg/mL); Round 3: pan-cytokeratin (pan-CK) (Cy3 conjugate, 13010/13421, 2.5 µg/mL), Na$^+$/K$^+$-ATPase (plasma membrane) (Cy5 conjugate of rabbit anti-sodium-potassium-ATPase, clone EP1845Y, #2047-1, epitomics, used at 5 µg/mL); and Round 4: pan-cadherin (PCad) (Cy3 conjugate of rabbit anti-pan cadherin [RB-9036; Thermo Scientific], stained at 5 µg/mL), epidermal growth factor receptor (EGFR) (Cy5 conjugate of rabbit-anti-EGFR clone D38B1 [4267; cell signaling], used at 1 µg/mL). These eight primary antibodies were conjugated using NHS-ester dye chemistry as previously described, and each lot was revalidated on control tissue sections before used in this study. Background images were collected before the first round and following each dye deactivation step and subsequently used for subtraction of background autofluorescence, as described below. 4',6-diamidino-2-phenylindole (DAPI) staining of nuclei was collected for all rounds of imaging, and DAPI stain was recharged before each antibody staining round. Imaging was on Olympus I × 81 inverted fluorescence microscope with a 20 × 0.75 NA objective outfitted as described by Gerdes *et al.*[32]

Image Processing and Cell Quantification

The biomarker images, acquired in different rounds, were first aligned. Alignment of the different channels was achieved by registering each DAPI image from successive rounds of imaging to the DAPI image of the first round using rigid transformation (i.e., only translation and rotation).[41] After registration, autofluorescence, which is typical in FFPE tissue, was separated from the fluorophore signals. To do this, an autofluorescence removal process[42] was applied, in which an image of the unstained sample was acquired, normalized, and then subtracted from the corresponding normalized-stained image.

The subsequent step in the workflow was image segmentation, which consisted of several steps [Figure 1]. First, DAPI-stained nuclei were segmented using a wavelet-based segmentation algorithm, followed by applying shape-based watershed [Figure 1a and b].[43] Second, an epithelial segmentation algorithm was used to identify the epithelial cells in the image using a biomarker that is known to be specific to the epithelial cells or at least with known subcellular localization patterns in the epithelial cells (e.g., pan-CK). Third, the cell cytoplasm and membrane were segmented using an algorithm that detects tubular structures in the image based on computing Frangi vesselness [Figure 1c and d].[44] In parallel, a multi-level watershed algorithm was applied on the membrane segmentation marker to extract initial cell contours.[45] In the final step, initial cell segmentation

results were combined with the three individual compartment segmentations (i.e., nuclear, cytoplasm, and membrane) as well as the epithelial mask to generate final cell segmentation mask.

The last step of the workflow was image quantification. Given the segmentation masks and any number of biomarker images, a large number of measurements were computed. These measurements include different cell morphological features (e.g., cell size and shape) and several statistics (e.g., mean, variance, and kurtosis) of each biomarker at the image, cell, and subcellular levels. Although the subcellular measurements for biomarker intensity were available, for this study we used only the mean biomarker intensity at the cellular level. We experimented using the median biomarker intensity as well but observed qualitatively similar results.

Data Preparation and Automated Quality Control

After the quantification, two automated quality control (QC) steps were applied. The first QC step detected damaged or lost tissue from round-to-round imaging and artifacts such as light saturation or poor focus, by producing image masks that differentiate between good- and poor-quality regions within each image. The poor-quality regions were subsequently masked out and excluded from downstream analysis. Properly sectioned and imaged cells are expected to fall within a specific size range, and the majority of their nuclei should remain in the section with minimal fragmentation. Thus, the second QC step filtered out cells based on size and the number of enclosed nuclei fragments. For this work, we used minimum and maximum cell size thresholds of 100 and 3000 pixels (37 and 1110 µm), respectively. In addition, we rejected cells with zero or more than three nuclei fragments.[32,46,47]

Partitioning Cells into Two Subpopulations to Account for the Distinct Regimes in Biomarker Intensity Distributions

Figure 3 shows our strategy of quantifying heterogeneity, which is to describe the spatial organization of expression patterns in a tumor sample. To enable this description, we observe that the probability distributions of the mean cellular biomarker intensity values (ER, HER2, PR) can be partitioned into two subpopulations. We will refer the two populations as high- and low-intensity regimes, L1 and L2, respectively [Figure 4a].

A threshold was defined using the knee observed in the probability distribution plots for each of the biomarkers (shown as vertical lines in Figure 4a). Since the data had already gone through several stages of QC (as described in the Methods), it was assumed that both the low-intensity and high-intensity populations were unique properties of the data, separate from nonspecific fluorescence. A given cell was assigned to
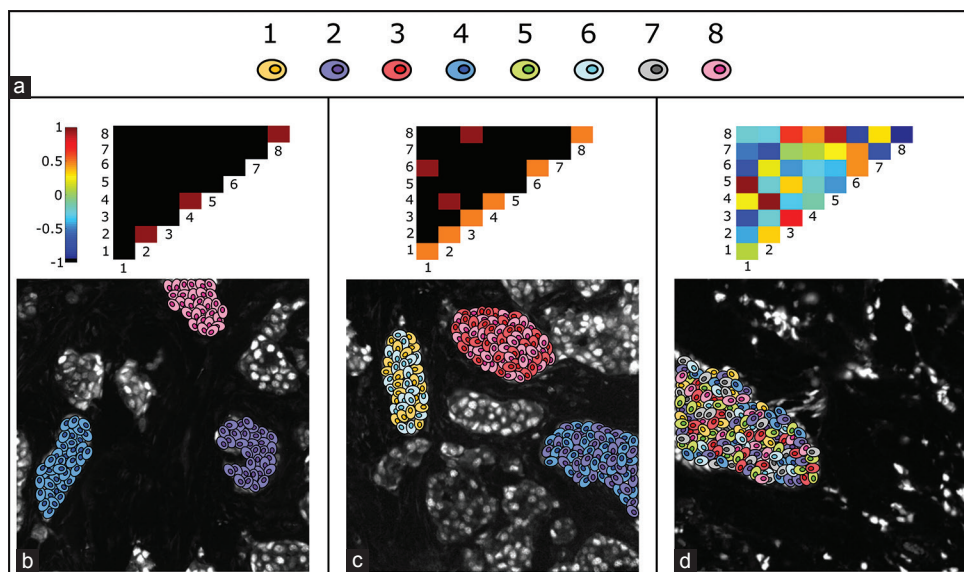
Figure 3: Canonical pointwise mutual information maps depicting various forms of spatial intratumor heterogeneity. (a) Cartoon representation of eight different cellular phenotypes based on high-dimensional biomarker intensity patterns acquired via pattern recognition algorithms [see Figure 4 for more details]. (b) A pointwise mutual information map with strong diagonal entries and weak off-diagonal entries describes a globally heterogeneous but locally homogeneous tumor. In this example, the pointwise mutual information map highlights locally homogeneous tumor microdomains containing cells of only one type each, phenotypes 2, 4, and 8 respectively. (c) On the contrary, a pointwise mutual information map with strong off-diagonal entries describes a tumor that is locally heterogeneous. In this example, locally heterogenous tumor microdomains exist as portrayed by the off-diagonal entries. One domain contains phenotypes 1 and 6, another contains phenotypes 2 and 4, and yet another containing phenotypes 3 and 8. (d) Pointwise mutual information maps can also portray anti-associations [e.g., if phenotype 1 never occurs spatially near phenotype 3, see Figures 5 and 6]. The ensemble of associations and anti-associations of varying intensities along or off the diagonal represents the true complexity of tumor images in a format that can be summarized and interrogated

the high-intensity subpopulation if any one of its mean biomarker values was greater than its respective threshold. A cell was assigned to the low-intensity subpopulation if and only if all of its mean biomarker values were below their respective thresholds. Biomarker patterns were then learned for each subpopulation of the data separately.

### Learning Dominant Biomarker Intensity Pattern Set from Each Cell Subpopulation

We aimed to discover a set of biomarker intensity patterns that adequately describe the high- and low-intensity subpopulations (L1 and L2) of IF data. While we could have imposed a predefined set of biomarker intensity patterns (for example, ER high/HER2 low/PR low and ER low/HER2 high/PR low), it was more compelling to learn the dominant patterns from the data, via machine learning (see Methods). The number of patterns needed to describe the data is likely to be greater than its dimensionality. This formulation is known as an *overcomplete* representation (see Supplementary Text for a glossary of machine learning and information theory terms). For example, if ER, HER2, and PR biomarker intensities were binarized into high and low signals, there would be eight potential combinations to describe the biomarkers intensities in three dimensions. In addition, we sought to develop methods that were applicable to

single biomarker expression data, multiplexed co-staining methods (<7 biomarkers), and emerging hyperplexed technologies (>7 biomarkers). To prescribe a predefined set of patterns on high-dimensional data, one would have to enumerate all potential combinations of biomarker intensities for that dataset: an exponentially complex endeavor. Machine-learning methods have the added benefit of scalability to these higher-dimensional problems.

A previous study suggests using mixtures of Gaussians to model high-dimensional biomarker distributions.[40] However, the formulation we describe below promotes interpretability via the mapping of cells to specific biomarker patterns. In addition, we are making no assumptions about the Gaussianity or distribution of cellular biomarker intensity profiles.

For our pattern recognition, we learned an overcomplete dictionary where the number of biomarker patterns used to represent the data is larger than its dimensionality (number of elements comprising each pattern). For example, we used eight biomarker patterns to describe IF data in three dimensions (ER, HER2, PR). Hereafter, we will refer to an overcomplete dictionary with m patterns as being m-overcomplete. Because the representation was overcomplete, we forced the cells to
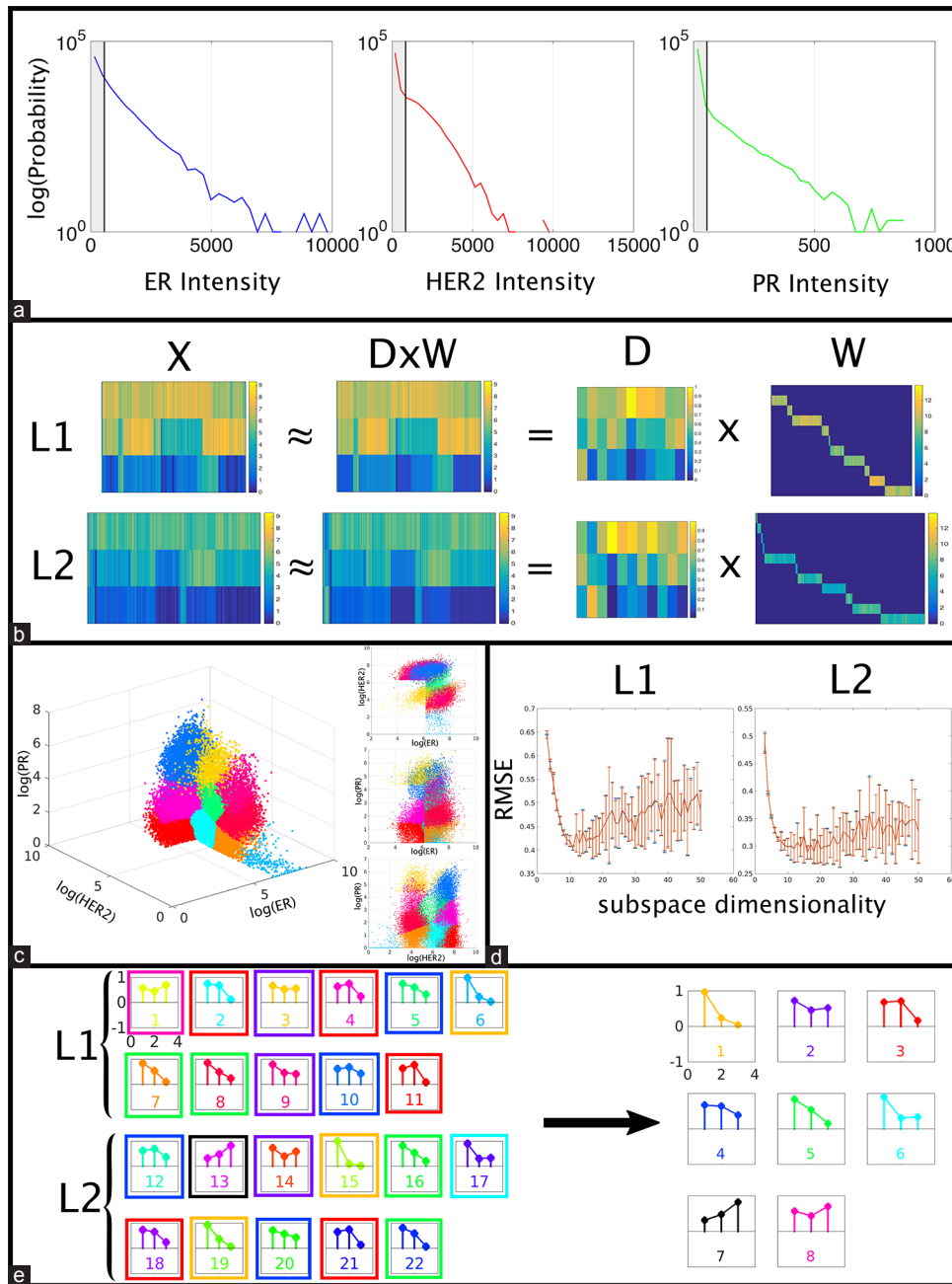
Figure 4: Learning a dictionary of cellular expression patterns. (a) Thresholds drawn as vertical lines for partitioning dataset into high-signal and low-signal subpopulations (L1 and L2, respectively). **(b)** Linear approximations of the L1 (high signal) and L2 (low signal) data matrices by the overcomplete dictionaries *D* and the sparse coding matrices *W*. Data matrix *D × W* is a reconstruction of the dataset, *X*. The rows of *X* and *D* correspond to estrogen receptor, progesterone receptor, and human epidermal growth factor 2 biomarker intensities, as labeled. The columns of *X* and *W* correspond to each individual cell. The columns of *D* correspond to the unique dictionary elements and the rows of *W* correspond to their weights. **(c)** Each cell is phenotyped to a single pattern in dictionary *D*. A three-dimensional representation of the L1 matrix is shown, where each cell is color coded by its phenotype. **(d)** Subspace selection of overcomplete dictionaries *D*, for L1 and L2, leads to a pattern size of 11 for each subpopulation. **(e)** Each pattern in the dictionary is shown as a colored stem plot and refers to (from left to right) the estrogen receptor, human epidermal growth factor 2, and progesterone receptor intensity levels. It is convenient to describe these intensities as high, medium, and low as we will do in the main text. For example, the cyan-colored pattern 2 in the L1 dictionary (left), which accounts for the cyan-colored cloud in panel c, may be described as estrogen receptor high, human epidermal growth factor 2 high, and progesterone receptor low. Next, using *k*-means clustering, we consolidate the L1 and L2 dictionaries into a final dictionary set of size 8. To denote the outcome of *k*-means clustering, we draw a colored box around each pattern in the L1 and L2 dictionaries, corresponding to the eight different consolidated clusters and show the mean patterns of the consolidated dictionary to the right

have a sparse coding, i.e., each cell is phenotyped with only one or few of the biomarker patterns. A sparse coding of the data improves interpretability for both cancer biologists and pathologists.

We used K-SVD, an iterative *k*-means derived method for designing overcomplete dictionaries with various sparsity constraints, to represent the IF data.[48] Each cell, $\vec{x}_i$ in the high- or low-intensity subpopulation was approximated by the following linear equation:

$$\vec{x}_i \approx w_i^1 \vec{d}_1 + w_i^2 \vec{d}_2 + \cdots + w_i^m \vec{d}_m \qquad \text{(Equation 1a)}$$

with the constraint

$$\text{Card } \vec{w}_i = 1 \qquad \text{(Equation 1b)}$$

where $\vec{w}_i = (w_i^1, \ldots, w_i^m)$ and is the pattern coefficient vector for the ith cell, $\vec{d}_1 \ldots \vec{d}_m$ represents the patterns in an m-overcomplete dictionary, w_i^j is the pattern coefficient for the jth pattern of the ith cell, and $\text{Card } \vec{w}_i$ is the cardinality (the number of nonzero elements) in pattern coefficient vector $\vec{w}_i$. By constraining the cardinality of the pattern coefficient vector to 1, we have explicitly set up the K-SVD algorithm to build a sparse representation such that each cell in a given dataset was phenotyped to only a single pattern in the dictionary [Figure 4b and c].

The biomarker pattern dictionary was learned using the logarithm of biomarker intensities for each cell in the TMA, for numerical stability. Since K-SVD is an iterative algorithm, we set it to run for 25 iterations well after it had reached a steady state result.

Determine Best Dictionary Size, *m*

A *priori*, it is not evident that how many elements should comprise the biomarker pattern dictionary. Representation error of the biomarker intensity data for each cell will tend to decrease as the number of patterns in the vocabulary increases. For example, if the number of patterns is equal to the number of cells in the TMA dataset, the representation error will be zero. However, the representation error of our dataset, given the linear approximations for different vocabulary sizes, decreased at smaller intervals as the vocabulary size approached the number of cells. Thus, we chose a vocabulary size where the representation error was low, but further increases to the vocabulary size had minimal returns. For each potential biomarker pattern dictionary size, *m*, the error of the linear representation was computed as:

$$e_m = \Sigma_i \left| \vec{x}_i - \left(w_1 \vec{d}_1 + w_2 \vec{d}_2 + \cdots + w_m \vec{d}_m \right) \right| \qquad \text{(Equation 2)}$$

To determine the best dictionary size, *m*, we performed a 10-fold cross-validation on the linear reconstruction for each IF data subpopulation. The data were split into ten equal but distinct groups, and the algorithm was trained on nine parts of the split and tested on the remaining one part. This procedure was repeated to cycle through each of the ten possible testing sets, with both the mean reconstruction error and the variance of the error reported at completion. An elbow criterion was used to choose the best subspace representation, where the allowance of another pattern in the dictionary was not found to decrease the error significantly [Figure 4d].

Construct Spatial Networks to Describe the Organization of Biomarker Patterns in a Tumor

To represent the spatial organization of the biomarker patterns in a tumor, a network was constructed for each spot in the TMA. The construction of spatial networks for tumor samples intrinsically couples cellular biomarker intensity data (in the nodes of the network) to spatial data (in the edges of the network). The assumptions in the network construction were that cells have the ability to communicate with nearby cells up to a certain limit, up to 250 μm as described by Francis and Palsson,[49] and that the ability for cells to communicate within that limit depends on cellular distance. Therefore, the probability distribution was computed for the distance between a cell in the TMA and its ten nearest neighbors. A hard limit was chosen based on the median value of this distribution times 1.5 (to estimate the standard deviation), where cells in the network were connected only within this limit. This limit was consistent with the 250 μm limit proposed by Francis and Palsson.[49] Then, the edges between cells in the network were weighted by the distance between the adjacent cells.

Using Pointwise Mutual Information to Quantify Spatial Biomarker Pattern Relationships

PMI was used to measure the association between each pair of biomarker patterns in the dictionary and thus different cell phenotypes, for a given sample of the data. This metric captures general statistical association, both linear and nonlinear, where previous studies[39] have used linear metrics such as Spearman's rho coefficient. PMI may be computed for an individual spot in the TMA, a single patient in the trial (using all of the spots sampled from that patient), a specific cancer cohort in the TMA, or the entire TMA. Once computed for each pair of biomarker patterns, a measure of all associations in the data is displayed in a PMI map. This map describes relationships between different cell phenotypes within the microenvironment, where differences may be compared from spot-to-spot and patient-to-patient.

Given a linear deconstruction of an IF dataset X, where each column of X is a cell x_k, into an overcomplete dictionary D, where each column of D is a distinct pattern *di*, and a sparse coding matrix W which assigns each cell to only a single biomarker intensity pattern, we

assign each cell to have a phenotype f_i, where i is the nonzero index in column w_k of W. A potential pitfall of the algorithm is that high- and low-signal intensity cells can be assigned to the same cell phenotype (more discussion of this in the Results section).

PMI between a pair of biomarker phenotypes (f_i, f_j) for a given network or network set S is defined as:

$$PMI_s \left(f_i, f_j\right) = \ \log \frac{P\left(f_{i_s}, f_{j_s}\right)}{P\left(f_{i_t}\right) P\left(f_{j_t}\right)} \qquad \text{(Equation 3)}$$

where $P\left(f_{i_s}\right)$ is the probability of phenotype f_i occurring in network set s, and $P\left(f_{i_t}\right)$ is the background probability distribution of phenotype f_i derived from the complete ensemble of networks. Note that the background distributions are based on the entire dataset, to compare individual networks to the distribution of the TMA as a whole. This construction is similar to the position-specific scoring matrices for either DNA or protein sequences, where the background distributions denote the probability of finding any particular nucleotide or amino acid over the dataset of sequences, for any given position.[50] A PMI map consists of the PMI score for every possible pair of patterns in the vocabulary for a given network set s. While we advocate the interpretation of the 2D PMI map for a thorough understanding of heterogeneity, we also derive a one-dimensional heterogeneity score value from the PMI map, for convenience of the reader interested in comparing with other one-dimensional scores in the literature. The information-deficient one-dimensional heterogeneity score is defined as:

$$HET_{PMI_s} = \sum_{i,j} \left| \log \frac{P\left(f_{i_s}, f_{j_s}\right)}{P\left(f_{i_t}\right) P\left(f_{j_t}\right)} \right| \qquad \text{(Equation 4)}$$

where higher scores denote a larger difference from the background distribution. The one-dimensional scores can incorrectly map two spatially different organizations of the TMEs, as seen by their PMI maps, to the same scale.

Visualize Spatial Networks for Specific Relationships

After computing the PMI map for a given tumor sample or patient and identifying significant interactions or interaction motifs, it is necessary to interrogate the cells which contributed to this significant association. A significant interaction would be considered when the PMI value is close to ±1. PMI values close to 1 signify that this particular spatial interaction of biomarker patterns occurs more frequently than observed in the background distribution. PMI values close to −1 signify that when one pattern is observed in the network, the other pattern is found to be observed less frequently than expected from the background distribution. PMI values

close to zero signify interactions that may adequately be described by the background distribution.

For a given interaction, a simple linear search through the network can extract the spatial connections in the network that contribute to the associations computed by PMI. These connections may then be superimposed onto the IF image to be examined in more detail by a pathologist [Figure 5].

Software

We used MATLAB (version R2015a, The MathWorks, Inc., Natick, MA, USA) to implement the analysis pipeline. We applied the segmentation algorithm developed by GE to output the cellular data into a comma separated value file containing the spatial location and the biomarker intensity for each cell in the TMA [Figure 2].[32,47] To partition the data into high- and low-intensity signals (L1 and L2, respectively [Figure 4a]), we applied a threshold value as determined by the elbow found in the probability distribution of the intensities of each biomarker channel. For biomarker pattern recognition via K-SVD, we used Ron Rubenstein's MATLAB implementation from http://www.cs.technion.ac.il/~ronrubin/Software/ksvdbo × 13.zip [Figure 4b and c]. This toolbox also requires the use of an orthogonal matching pursuit implementation in MATLAB from http://www.cs.technion.ac.il/~ronrubin/Software/ompbo × 10.zip. We selected the optimal subspace by performing multiple K-SVD trials at different subspace sizes and chose the ideal subspace dimension when the reconstruction error stopped decreasing [Figure 4d]. To consolidate the biomarker pattern sets learned from the L1 and L2 partitions, we used k-means in MATLAB, where the number of clusters k was chosen using the silhouette function in MATLAB, which provides a graphical representation of cluster membership confidence [Figure 4e].[51]

To construct cell networks, we computed the median distance, d_m, of the ten nearest neighbors for each cell and derived the standard deviation, s, from the median value, $s = 1.5d_m$ such that the connections between neighboring cells were Gaussian weighted. To keep the network sparse, cells separated by a distance >3 s were not connected. PMI was calculated by Equation 3, using 2D histograms to compute the joint probabilities. We observed that the distribution of PMI values was concentrated around 0, ranging between −10 and 10, with an additional population of extremely large negative values. After it was determined that the extremal values were the result of a lack of co-occurrences in combination with small smoothing coefficients used to calculate numerically stable logarithms, we suppressed these extremal values and focused only on the values around zero. PMI maps were normalized between −1 and 1 for ease of visualization, and the extrema were saturated to a normalized PMI value of −1. To visualize a specific interaction that contributes
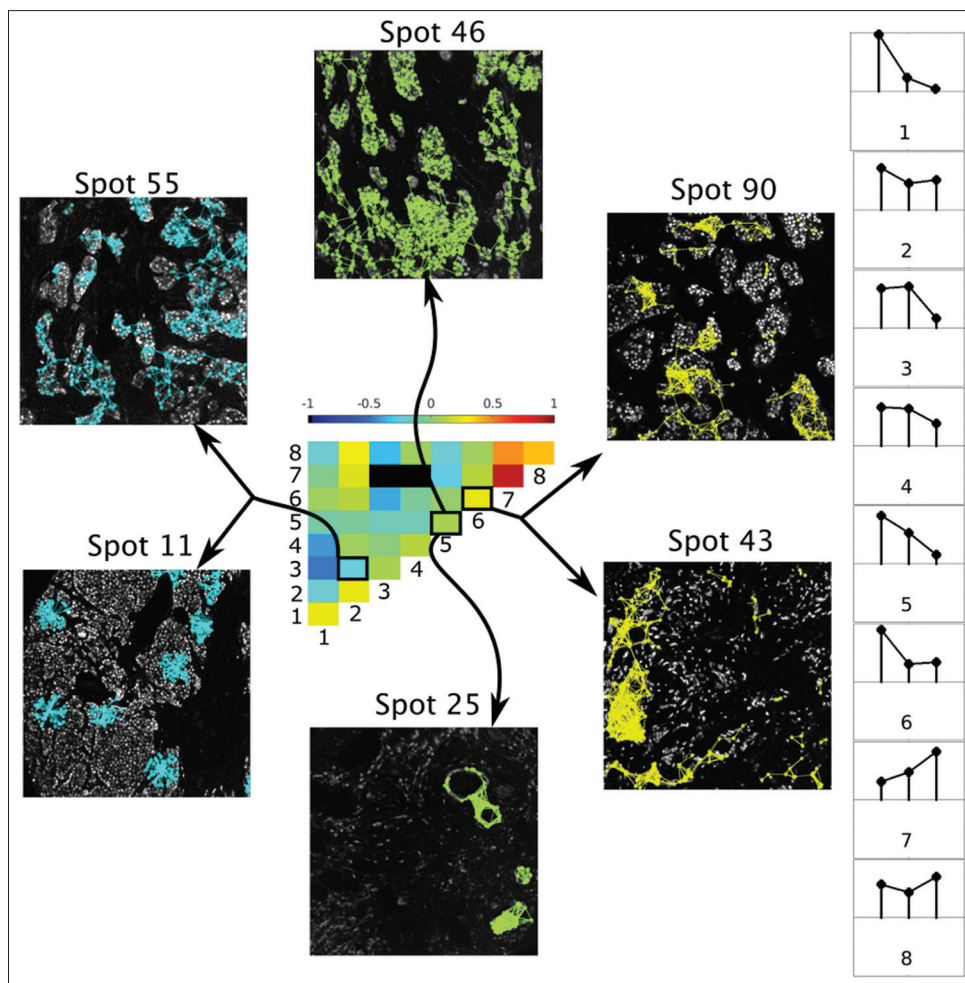
Figure 5: Visualizing networks of spatial interactions from pointwise mutual information maps. The pointwise mutual information map in the middle denotes the relative probability of finding two co-occurring phenotypes *i* and *j* in reference to a background distribution. In the colorbar above the pointwise mutual information map, red/blue indicates highest/lowest possible co-occurrence, and black indicates an absence of interactions. The stem plots to the right describe the eight phenotypes learned from the data, where each stem plot represents the relative estrogen receptor, human epidermal growth factor 2, and progesterone receptor intensities of the phenotype (left to right). The labels for each stem plot (1–8) correspond to the rows and columns of the pointwise mutual information map. This map allows us to probe any tumor sample for networks of spatial interactions that contribute to the pointwise mutual information calculation. We display representative networks of spatial interactions for three different pointwise mutual information map entries. The two networks shown in yellow are examples where phenotype 6 spatially co-occurs with itself more frequently than expected from the background distribution. The two networks shown in green indicate two spatial networks where phenotype 5 spatially co-occurs with itself as would be expected from a random phenotyping of cells, given phenotype background probabilities. The two networks shown in blue portray interactions between cells of phenotypes 2 and 3 spatially co-occurring, which happens less than is expected from the background distribution. In each of these cases, the nodes in these graphs are the spatially co-occurring cells of a specific phenotype, and edges are only drawn to cells in spatial proximity. Depending on individual tumor graph statistics, these spatial relationships may be localized or ubiquitous throughout the tumor

to the construction of the PMI map, a linear search may be executed on the tissue data to look for the specific pairwise interactions of cell phenotypes requested [Figure 5].

Dataset

Table 1 provides a summary statistics for the TMA with cohorts from (a) ER(+) IDC, (b) ER(+) ILC, (c) ER(−) IDC, and (d) HER2(+) IDC. Three biopsy cores were taken from each patient. For each tumor sample, we consider cells that passed QC (see the Methods section for more details) and divide them into two distinct sets, high and low, based on biomarker intensities.

RESULTS

Preprocessing of Multiplexed/Hyperplexed Immunofluorescence Image Data

A cellular segmentation algorithm was previously developed and applied to IF data taking advantage of the selectivity of segmentation biomarkers, DAPI, Na^+/K^+-ATPase, S6, and pan-cadherin.[32] The output of the segmentation algorithm includes cellular masks and subcellular masks for each cell at nuclear, cytoplasmic, and membrane resolution [Figures 1 and 2]. Using this segmentation, we extracted biomarker intensity at single

Table 1: Tissue microarray data with cohorts from (a) estrogen receptor(+) invasive ductal carcinoma, (b) estrogen receptor(+) invasive lobular carcinoma, (c) estrogen receptor(−) invasive ductal carcinoma, and (d) human epidermal growth factor 2(+) invasive ductal carcinoma. Spot number refers to the location on the TMA. The two numbers reported in the cell count column refer to the number of cells in the high-signal and low-signal populations, respectively, where the data were partitioned based on biomarker signal intensities (see Figure 3a for more details). (c) Spots 91 and 89 in ER(−) IDC are anomalous in that the ER staining was observed in the normal epithelium and not in the cancer cells. These spots were erroneously assigned the category of ER(−) IDC by the automated image processing algorithms. These errors could be avoided by having pathologists manually determine regions of interests. TMA: Tissue microarray, ER: Estrogen receptor IDC: Invasive ductal carcinoma

(a) ER(+) IDC		
Spot#	Patient ID	#Cells
000	AL13-1	23 (1181)
031	AL13-1	222 (382)
026	AL13-2	72 (376)
055	AL13-2	604 (641)
060	AL13-2	420 (624)
005	AL13-3	433 (589)
046	AL13-3	327 (1629)
086	AL13-3	164 (600)
(b) ER(+) ILC		
001	AL13-4	721 (249)
043	AL13-4	882 (192)
066	AL13-4	1065 (112)
011	AL13-5	2589 (166)
061	AL13-5	3339 (26)
080	AL13-5	2975 (52)
006	AL13-6	297 (621)
025	AL13-6	269 (458)
076	AL13-6	348 (246)
045	AL13-7	260 (192)
030	AL13-8	20 (125)
056	AL13-8	1062 (165)
096	AL13-8	479 (182)
(c) ER(-) IDC		
002	AL13-9	0 (1032)
039	AL13-9	43 (779)
079	AL13-9	1 (1084)
012	AL13-10	0 (1173)
052	AL13-10	0 (185)
091	AL13-10	0 (1779)
024	AL13-11	0 (849)
049	AL13-11	28 (831)
089	AL13-11	0 (895)

Contd...

Table 1: Contd...

(c) ER(-) IDC		
007	AL13-12	3 (381)
034	AL13-12	1 (482)
062	AL13-12	52 (917)
017	AL13-13	17 (84)
036	AL13-13	4 (1055)
072	AL13-13	44 (1219)
020	AL13-14	1 (1322)
044	AL13-14	0 (2296)
071	AL13-14	8 (1414)
032	AL13-15	12 (764)
057	AL13-15	65 (532)
029	AL13-16	118 (876)
067	AL13-16	4 (804)
095	AL13-16	4 (1771)
(d) HER2(+) IDC		
003	AL13-17	147 (53)
038	AL13-17	209 (36)
068	AL13-17	276 (39)
015	AL13-18	69 (39)
041	AL13-18	771 (184)
078	AL13-18	319 (678)
013	AL13-19	245 (215)
053	AL13-19	824 (123)
090	AL13-19	970 (178)
023	AL13-20	1044 (757)
063	AL13-20	799 (229)
088	AL13-20	669 (101)
008	AL13-21	39 (162)
033	AL13-21	194 (44)
065	AL13-21	97 (7)
018	AL13-22	677 (6)
048	AL13-22	890 (2)
073	AL13-22	521 (5)
021	AL13-23	86 (57)
058	AL13-23	439 (8)
083	AL13-23	1048 (28)
028	AL13-24	126 (1724)
070	AL13-24	64 (1440)
093	AL13-24	309 (2079)

cell resolution such that each cell was represented by its spatial coordinates and biomarker intensities, from which interaction networks were be built and biomarker intensity patterns were identified.

Figure 2 shows a pseudocolored IF image where ER signal is colored in blue, HER2 signal in red, and PR signal in green. The pseudocolored image makes colocalizations of biomarker signals within cells explicit and further helps assess heterogeneity. For example, the cytoplasm of a cell will appear either as cyan if ER (blue) colocalizes with PR (green) or as blue if ER (blue) localizes by itself. For spot 55 [Figure 2], we concluded that a significant

degree of heterogeneity is apparent. For demonstration, we highlighted three groups of cells shown by arrows in Figure 2. In the first group of cells (middle arrow), ER and HER2 were dominant in their respective cellular compartments and did not colocalize. In the second group of cells (top arrow), ER and PR colocalized to produce cyan, with HER2 signal localizing in the membrane. In the final group of cells (bottom arrow), PR and HER2 were dominant in their respective cellular compartments and did not colocalize, with a much smaller portion of ER(+) cells. This visualization example exemplifies a qualitative approach to describing heterogeneity, which was formalized and made quantitative with the PMI measure.

A Strategy for Quantifying Heterogeneity

Our strategy for quantifying heterogeneity has three components [Figure 3]. First, we learn a small set of dominant biomarker intensity patterns, for example, ER high/HER2 high/PR off; from the IF data based on biomarker intensity composition of each cell, we assign it to one of the dominant patterns. Figure 3a shows a cartoon representation of possible cell phenotypes. Second, we construct a spatial network to describe the organization of biomarker patterns in a tumor (see Methods). Finally, we quantitate heterogeneity in the form of a PMI map, where the entries measure how frequently a particular spatial interaction between two phenotypes (referenced by the row and column number) occurs in the dataset when compared to the interactions predicted by a random (or background) distribution over all phenotypes. In Figure 3b-d, PMI entries in red denote a strong spatial association between phenotypes while entries in black denote a lack of any colocalization. PMI entries colored green denote associations that are no better than a random distribution of cell phenotypes over the entire dataset. In addition, PMI maps can portray anti-associations denoted by blue (e.g., if phenotype 1 rarely occurs spatially near phenotype 3) as shown in Figure 3d.

A PMI map with strong diagonal entries and weak off-diagonal entries describes a globally heterogeneous but locally homogeneous tumor. To illustrate this, we show a canonical PMI map in Figure 3b where the associations in the diagonal entries for phenotypes 2, 4, and 8 are strong. This implies that these phenotypes are spatially associated with cells of the same phenotype as shown by the composition of the individual microdomains in the tumor sample in Figure 3b.

On the contrary, a PMI map with strong off-diagonal entries can describe a tumor that is locally heterogeneous. In the canonical PMI map shown in Figure 3c, the associations between the cellular phenotypes 1 and 6, 2 and 4, and 3 and 8 are spatially localized. In Figure 3d, we find associations between all phenotypes in the tumor

image, and hence, the PMI is thoroughly intermixed. The benefit of PMI maps over existing measures is that the maps evoke a spatial relationship between phenotypes. These provide not only a summary of cellular composition but also an approximation of the tumor topology. For the sake of brevity, we have not included more complicated PMI map examples, but all maps are built off of these simple interactions.

Building a Dictionary of Biomarker Expression Patterns

We segregated the data into two partitions based on the distribution of signal intensity for each biomarker, under the assumption that signal intensity indicates true biomarker expression [Figure 4a]. Notice that each of these log-occurrence distributions may be modeled by two or more linear equations. The notch where these two different models would meet is set to be the threshold for that particular biomarker channel and is drawn as vertical lines in the biomarker intensity distribution graphs. For any given cell, if one or more of its biomarker intensities is above threshold, then that cell is classified as Level 1 (L1). If all of the biomarker intensities for any given cell are below the thresholds in their corresponding biomarker channels, then that cell is classified as Level 2 (L2). These two partitions can be interpreted in terms of their signal-to-noise ratio, where L1 has a higher signal-to-noise ratio and L2 has a lower signal-to-noise ratio in comparison. Each partition of cells is used to learn its own set of biomarker patterns. This approach seems particularly judicious given that the distribution of pattern coefficients for L1 and L2 data has different Gaussianity in general [Figure 2]. As shown in Figure 4a, the studied biomarker intensities have long-tailed distributions, so we chose a log-intensity representation to derive a numerically stable pattern recognition algorithm.

For each partition of the data, L1 and L2, we arrived at a sparse signal representation [Figure 4b]. A given data matrix X, where the columns represent each cell in the dataset and the rows represent the log biomarker intensities of each cell (top to bottom, ER, HER2, PR, respectively), can be approximated by the product of matrices D and W. D represents a dictionary of potential biomarker intensity patterns learned from the ensemble of cells in the dataset X, where each column represents one of the patterns learned from the data and each row represents the respective biomarker intensities of each pattern. W is a sparse matrix, which phenotypes each cell in X to a specific pattern in D with a particular scaling coefficient. Thus, each cell (column in W) is represented by only one cell phenotype, which corresponds to the biomarker pattern (column in D) where the sparse code lies. The color spectrum for each matrix varies from blue (low intensity) to yellow (high intensity).

We also display matrix *DW* to portray the similarity between the actual data matrix and its reconstruction. By viewing matrices *X* and *DW*, which are column sorted by the dictionary element they have the most consensus with, we can observe that each of the biomarker patterns is present in the data. The benefit of this reconstruction of the data is the ability to represent a large array of cell-level data with a small number of interpretable biomarker patterns, describing highly clustered clouds inherent to the dataset as shown in Figure 4c. Each cell in the three-dimensional log biomarker intensity space is color coded by its phenotype [Figure 4c].

The reconstruction error of our linear representation of a given dataset *X* into dictionary *D* and dictionary coefficient matrix *W* highly depends on the dimensionality of *D*, i.e., the number of patterns that will be used to describe the dataset *X*. To choose the ideal dimensionality of *D*, we perform a 10-fold cross-validation of the data reconstruction [Figure 4d]. As is typical in these analyses, we note that as we increase the dimensionality, reconstruction error and the variance of the error decrease until a certain point where the error variance begins to increase with dimensionality. We found that a dictionary size of 11 patterns optimizes both reconstruction error and variance of the error, for both data partitions, L1 and L2.

Having learned a set of 11 patterns for each nonoverlapping partition of the data L1 and L2, we could merge the two dictionaries into a large single dictionary of biomarker intensity patterns that can describe the entire dataset. However, since these patterns were learned separately from partitions deriving from the same dataset, captured under the same experimental conditions, we noted that there were some redundancies between the dictionary learned from L1 data and the dictionary learned from L2 data. Thus, we used *k*-means clustering to consolidate the large 22-pattern dictionary (with 11 patterns from each partition) into a smaller final dictionary containing only the unique patterns discovered from our approach [Figure 4e]. In Figure 4e, the 11 patterns learned from L1 and the 11 patterns learned from L2 are shown to the left. Each biomarker pattern is represented as a stem plot of its ER, HER2, and PR intensity, respectively. For convenience, we will describe the intensity patterns in the stem plots as being high, medium, and low. For example, pattern 8 in the L1 dictionary (shown to the left) may be described as ER high, HER2 medium, and PR low.

The outcomes of *k*-means clustering, shown to the right, result in a final dictionary dimensionality of 8 biomarker intensity patterns. The final dimensionality was chosen based on the results of a silhouette criterion for clustering evaluation.[51] The boxes around each of the initial patterns to the left signify their cluster membership

and correspond to the color of the pattern in the final pattern set on the right. Note that one pattern was unique to partition L2, pattern 7 of the final pattern set, with low ER expression, intermediate HER2 expression, and high PR expression. This demonstrates the value of partitioning the data into two groups, L1 and L2, where patterns dominant in one partition, but not the other, may be elucidated.

Visualizing Networks of Spatial Interactions from Pointwise Mutual Information Maps

We generated PMI maps to summarize the relative probabilities of all pairwise spatial interactions within a given tumor sample. In our reconstruction of the cellular IF data, each cell was assigned a specific phenotype by its dominant pattern under sparse coding (see Methods). Each bin of the PMI plot represents the dependence of a cell phenotype upon other phenotypes or itself, relative to the background distribution of the individual biomarkers over the entire dataset. After identifying important spatial dependencies between phenotypes, we can reference the tumor spots and their respective interactions which contribute to the PMI score.

Figure 5 displays an example PMI map for the entire dataset [Table 1], excluding the tumor samples containing <100 cells. Each row and column are numbered from 1 to 8, representing one of the 8 potential cell phenotypes displayed to the right [learned via the approach described in Methods and Figure 4]. Each bin of the PMI map is colored from blue to red, according to its PMI score. Scores close to red signify that a particular spatial interaction between two phenotypes (referenced by the row and column number) occurs in the dataset more frequently than a random distribution of all the phenotypes in the data would account for. Scores close to blue signify that a particular spatial interaction between two phenotypes occurs in the dataset much less frequently than a random distribution would account for. If a bin is green, then it signifies that the background distribution of the phenotypes in the dataset adequately describes the spatial interaction between those two phenotypes. Bins colored in black signify that this potential spatial interaction is not found in this sample of the data.

In Figure 5, we show networks of spatial interactions that contribute to any one entry in the PMI map. The spatial dependencies shown in green for spots 46 and 25 signify similarity to the background distribution in the co-occurrence of phenotype 5 (ER high, HER2 medium, PR low) with itself. Note that these interactions can be either ubiquitous throughout the tumor sample (spot 46) or localized to specific tumor structures (spot 25). The spatial dependencies shown in light blue for spots 55 and 11 signify lower probability than background distribution in the co-occurrence of phenotype 2 (ER high, HER2 medium, PR medium) with phenotype 3 (ER high, HER2
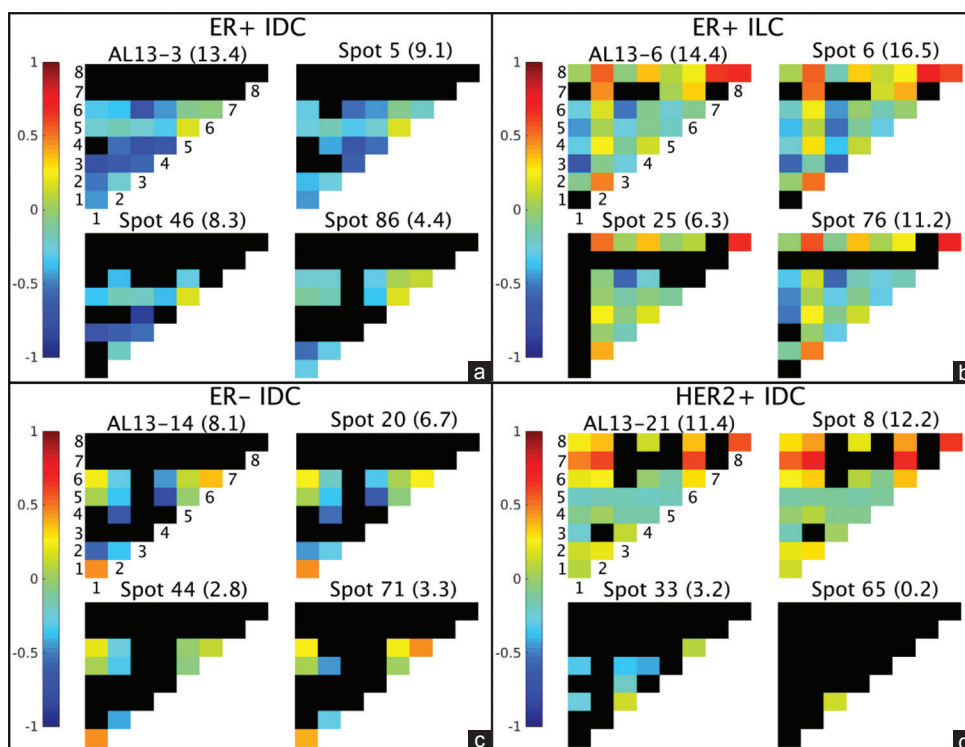
Figure 6: Pointwise mutual information maps as potential diagnostic biomarkers. Pointwise mutual information maps were constructed for individual cores using the background distributions of cell phenotypes in the entire dataset and were pooled together for patient-level pointwise mutual information (entire tumor) to better assess intratumor heterogeneity. A representative (a) estrogen receptor(+) invasive ductal carcinoma patient, (b) estrogen receptor(+) invasive lobular carcinoma patient, (c) estrogen receptor(−) invasive ductal carcinoma patient, and (d) human epidermal growth factor 2(+) invasive ductal carcinoma patient pointwise mutual information map was shown, as well as pointwise mutual information maps for the three cores taken from each patient. A heterogeneity score was assigned to each core/patient based on the entries in each pointwise mutual information map (see Methods for the relevant equation). Based on this heterogeneity score, patients AL13-3 estrogen receptor(+) invasive ductal carcinoma and AL13-6 estrogen receptor(+) invasive lobular carcinoma show more heterogeneity (difference from background distribution) than AL13-14 estrogen receptor(−) invasive ductal carcinoma and AL13-21 human epidermal growth factor 2(+) invasive ductal carcinoma. The degree to which the core-level pointwise mutual information maps change with respect to each other and the patient-level map can elucidate how much or little intratumor heterogeneity exists. For example, the core-level pointwise mutual information maps for patient AL13-14 are very similar, signifying that each core is a reasonable approximation for the patient-level analysis. As a contrary example, patient AL13-21 has highly differing core-level pointwise mutual information maps, signifying a high degree of intratumor heterogeneity in this patient. The summary heterogeneity score can provide a simple low-level understanding of heterogeneity between or within patient samples while the pointwise mutual information maps can provide a higher-level understanding, providing insight into the spatial relationships of different cell types which bring about the heterogeneity. We also propose visualization tools that can help elucidate these relationships in this higher-level understanding [Figure 5]

high, PR low). In comparison, spatial dependencies shown in yellow for spots 90 and 43 signify higher probability than background distribution in the co-occurrence of phenotype 6 (ER high, HER2 low, PR low) with itself. Note that these interactions are spatially localized. Observe that the PMI maps will change when we evaluate the relative probabilities over different subpopulations.

Pointwise Mutual Information Maps as Potential Diagnostic Biomarkers

In Figure 6, we show the construction of PMI maps for subpopulations of the dataset: patient AL13-3 ER(+) IDC cores (spots) [Figure 6a], patient AL13-6 ER(+) ILC cores [Figure 6b], patient AL13-14 ER(−) IDC cores [Figure 6c], and patient AL13-21 HER2(+) IDC cores [Figure 6d]. For each panel of this figure, we display the PMI maps for the three replicate cores of

a given case, then combine the spatial networks for all three cores to build a patient-level PMI map, and finally report HET_{PMI_s} scores for each PMI map.

For AL13-3 ER(+) IDC [Figure 6a], we observe no interactions involving phenotype 7 (ER low, HER2 medium, and PR high) or phenotype 8 (ER medium, HER2 medium, PR high), denoting a lack of ER(−)/PR(+) cells for this patient. There are also a few interactions that have negative PMI scores, denoting a lower likelihood for co-occurrence than predicted by the background distribution (e.g., phenotype 3 [ER high, HER2 high, PR low] with phenotype 4 [ER high, HER2 high, PR medium]). One particular spatial interaction with a slightly higher likelihood for co-occurrence compared to the background distribution is that between phenotype 5 (ER high, HER2 medium, PR low) with itself. The observation that the replicate core PMI

maps are similar to the combined patient-level PMI map suggests that the cores are reasonably accurate representations of the entire tumor. However, larger PMI values in spot 5 and 46, compared to spot 86, lead to higher heterogeneity scores.

The PMI map for AL13-3 ER(+) ILC [Figure 6b] presents a high level of deviation from the background distribution as noted by its array of colors. Phenotype 8 (ER medium, HER2 medium, PR high), for example, presents a dynamic behavior, interacting with itself heavily, in addition to phenotype 2 (ER high, HER2 medium, PR medium) and phenotype 7 (ER low, HER2 medium, and PR high), above background distribution. This example suggests that PR(+) phenotypes co-occur spatially. There exists a certain degree of core-level heterogeneity within the tumor. Notably, spots 25 and 76 contain no cells of phenotype 7 (ER low, HER2 medium, and PR high) while spot 6 exhibits phenotype 7 under specific interactions with phenotype 2 (ER high, HER2 medium, PR medium), phenotype 5 (ER high, HER2 medium, PR low), phenotype 6 (ER high, HER2 low, PR low), and phenotype 8 (ER medium, HER2 medium, PR high). Note that, for core 25, the interaction of phenotype 5 (ER high, HER2 medium, PR low) with itself changed from being slightly positive [when the entire dataset was considered, Figure 5] to slightly negative in the core-level PMI map. Lack of interactions lessens the heterogeneity scores for spots 25 and 76 compared to spot 6. The heterogeneity score for spot 6 is heightened due to stronger anti-associations (for example, between phenotypes 3 and 5 and phenotype 3 with itself), which are then smoothed out by the addition of the two remaining spots for the AL13-6 patient level PMI map.

AL13-14 ER(−) IDC [Figure 6c] contains fewer PMI interactions than the previous two examples, with 23 of the potential 36 interactions not occurring at all in the patient (shown as black in the PMI map). Phenotype 1 (ER high, HER2 low, PR nearly off) cells, however, co-occur with other phenotype 1 cells more frequently than is described in the background distribution. It may seem counter-intuitive that ER high phenotype 1 occurs in the ER(−) IDC data. However, ER(−) data may very well be captured by an ER high phenotype because a given phenotype decision is made based on the angle between the pattern vectors and the data point in 3D space, and not by the projection distance of the data point onto the pattern vectors [Figure 3]. The observation that the replicate core PMI maps are similar to the combined patient-level PMI map suggests that the cores are reasonably accurate representations of the entire tumor.

For AL13-21 HER2(+) IDC [Figure 6d], phenotype 7 (ER low, HER2 medium, and PR high) co-occurs at greater than background probability with phenotype 1 (ER high, HER2 low, PR nearly off), phenotype 2 (ER high, HER2 medium, PR medium), and phenotype 6 (ER high, HER2 low, PR low). In addition, phenotype 8 (ER medium, HER2 medium, PR high) has a highly dependent interaction with itself. An interesting feature of this patient is that the tumor cores taken from this patient are highly heterogeneous. Spot 8 is very dynamic while spot 65 is completely homogenous, containing only phenotype 3 (ER high, HER2 high, PR low) interacting with itself.

Comparing heterogeneity scores across patients, we observe that ER(+) ILC patient AL13-6 has the highest degree of heterogeneity, followed by ER(+) IDC patient AL13-3, HER2(+) IDC patient AL13-21, and finally ER(−) IDC patient AL13-14. Patient AL13-6 has the largest diversity of interactions (few black bins) and contains many strong positive and negative PMI values. Patient AL13-3 has less phenotype co-occurrence diversity but contains mostly strong negative association scores. Following this, AL13-21 contains many interactions with low PMI values. Finally, AL13-14 has the lowest heterogeneity score, containing very few co-occurrences with strong PMI values. Clearly, a breadth of information pertaining to the dependencies between interactions of various cell types and various levels of local and global heterogeneity can be gleaned from core-level PMI maps and comparisons of these maps to their patient-level PMI.

DISCUSSION

With the ability to capture a growing number of biomarkers, IF and mass spectrometry imaging techniques will play a major role in the quantification of spatial intratumor heterogeneity. These emerging hyperplexed imaging technologies will increase the need for scalable algorithms. The output of these algorithms must remain interpretable and actionable for decision-making purposes in the diagnostic realm. Our method is flexible regardless of the number of biomarkers imaged and can incorporate spatial and expression data together to quantify spatial intratumor heterogeneity. The end product includes PMI maps, which represent spatial relationships between cell phenotypes and other constituents of the TME. In addition, we provide a score based on PMI values to summarize intra- and inter-tumor heterogeneity. Finally, the PMI maps permit the end user to visualize its individual components in the form of spatially interacting cellular networks.

PMI maps have the potential to be an innovative tool for the modern computational pathologist. For example, if a tumor sample has a PMI map with strong diagonal entries and weak off-diagonal entries, this describes a

tumor with several highly localized self-interactions of specific cell phenotypes, thus signifying a tumor sample exhibiting local homogeneity but global heterogeneity. As a contrary example, a tumor sample with a PMI map having strong off-diagonal entries describes a tumor with many localized interactions between different cell phenotypes, signifying a tumor sample exhibiting strong local heterogeneity. Another possibility is the presence of only a single dominant entry in the PMI map, which signifies either global homogeneity if along the diagonal or global heterogeneity with two mixing populations if off-diagonal. By looking at a patient's PMI map, and comparing it to the individual PMI maps for each tumor sample as well as cohort-summarized (e.g., for all ER[+] IDC patients) PMI maps, unique interactions between biomarkers in the sample can be identified and interrogated. HET_{PMI_S} scores (Equation 4) can be used to compare tumor samples quickly. Using these measures, a clinician can arrive at highly specific conclusions regarding the degree of heterogeneity within a single tumor sample or between different samples of the same tumor. The level of information surmised from PMI includes spatial

elements that are not achievable by other methods that quantify intratumor heterogeneity, including QE.[35]

Comparing Pointwise Mutual Information to Quadratic Entropy and Clinical Standards

QE measures heterogeneity via species diversity (relative abundance) without explicitly quantifying spatial interactions among species. To compare QE with PMI maps, we binned ER signal intensities to define different "species" of cells in the TME. Following work in ecological diversity,[34] Potts *et al.*[35] introduced a distance matrix between these species to penalize spatial interactions between disparate species (more heterogeneity) but reward spatial interactions between similar species (less heterogeneity). High QE values denote more heterogeneity, and low QE values denote less heterogeneity. We computed HET_{slide} as the QE of the entire tumor sample and HET_{ROI} as the QE of various regions of interests in the tumor sample [Figure 7a-c]. We observed for our data that HET_{slide} does not correlate consistently with any of the statistics for HET_{ROI} [Figure 7d and e]. This is similar to our
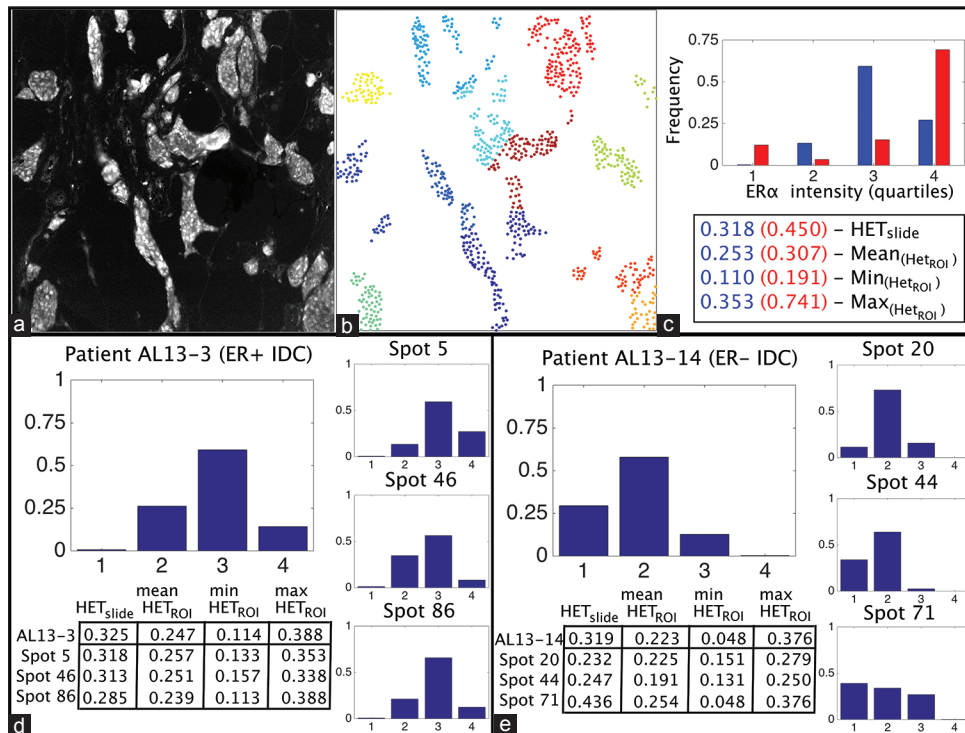
Figure 7: Quadratic entropy implementation with patient and core-level results. (a) The estrogen receptor immunofluorescence image is shown for spot 5. Note different levels of estrogen receptor signal intensity throughout the tumor sample. (b) Regions of interest are selected via *k*-means clustering on the XY coordinates of each cell (although one may also select regions of interests manually). (c) The frequencies of each estrogen receptor-intensity species (where 1 is lowest intensity and 4 is highest) are shown for spot 5 by the blue bars. The species frequencies for spot 76 (immunofluorescence and regions of interests not shown) are shown in red to provide another example of potential estrogen receptor-intensity distributions. Quadratic entropy is reported for the entire tumor sample (HET_{slide}), as well as mean, minimum, and maximum quadratic entropy of each regions of interest (HET_{ROI}). (d and e) Quadratic entropy was calculated for patients AL13-3 and AL13-14, from cohorts estrogen receptor(+) invasive ductal carcinoma and estrogen receptor(−) invasive ductal carcinoma, respectively. HET_{slide} is computed for each individual replicate core of each patient and then computed for the entire patient (consolidating the cells from each core). HET_{ROI} is computed for each region of interest in the cores and the mean, minimum, and maximum HET_{ROI} scores are reported for each core and patient

observations in Figure 6 where patient-level PMI maps do not necessarily correlate to their replicate core PMI maps, highlighting a degree of spatial heterogeneity between cores. In addition, HET_{slide} can be similar for different tumor samples even though the samples have radically different frequencies of ER expressing species [Figure 7, panel d vs. panel e].

One feature of the QE model is the use of a distance matrix between species for characterizing heterogeneity. Another feature of the model is the encoding of spatially heterogeneous subregions in a tumor sample by the computation of both HET_{slide} for the entire tumor sample and HET_{ROI} for the subregions. Finally, the implementation of the QE model is simple and elegant, where species can be defined using only the signal intensity of single biomarkers. However, QE does not quantify the probability of spatial interactions between two cell phenotypes as done in PMI with its spatial network-based approaches. Currently, the QE approach works with one channel at a time while PMI can elucidate spatial relationships simultaneously in the space of all biomarkers. It may also be possible to develop a multichannel QE algorithm to quantify heterogeneity. In addition, QE and PMI are complementary in that a distance matrix can be incorporated into future implementations of PMI. Finally, we intend to incorporate the scaling factor associated with the assignment of cell phenotypes in the construction of PMI maps. It is important to emphasize that PMI maps should be interpreted in their entirety as 2D maps and not as one-dimensional information-deficient heterogeneity scores.

CONCLUSION

This paper presents an approach for describing intratumor heterogeneity, in a quantitative fashion, which departs from the purely qualitative approaches currently used in the clinic. With access to larger data samples and clinical outcome data, we will be able to correlate spatial relationships with disease progression and response to therapy. By increasing the number of biomarkers imaged, we can select for cells in different states of activation, as well as noncellular constituents (e.g., secretory elements, extracellular matrix), and quantify relationships between previously unstudied determinants in the TME. In combination with genomic, proteomic, and transcriptomic data, our PMI-based method for spatial intratumor heterogeneity using high-dimensional imaging modalities may be used as part of a multimodal approach to study the mechanisms of cancer.

Acknowledgments

Conflicts of Interest
There are no conflicts of interest.

REFERENCES

1. Alizadeh AA, Aranda V, Bardelli A, Blanpain C, Bock C, Borowski C, *et al.* Toward understanding and exploiting tumor heterogeneity. Nat Med 2015;21:846-53.
2. Gough AH, Chen N, Shun TY, Lezon TR, Boltz RC, Reese CE, *et al.* Identifying and quantifying heterogeneity in high content analysis: Application of heterogeneity indices to drug discovery. PLoS One 2014;9:e102678.
3. Critchley-Thorne RJ, Miller SM, Taylor DL, Lingle WL. Applications of cellular systems biology in breast cancer patient stratification and diagnostics. Comb Chem High Throughput Screen 2009;12:860-9.
4. Balkwill FR, Capasso M, Hagemann T. The tumor microenvironment at a glance. J Cell Sci 2012;125(Pt 23):5591-6.
5. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell 2011;144:646-74.
6. Tabassum DP, Polyak K. Tumorigenesis: It takes a village. Nat Rev Cancer 2015;15:473-83.
7. Marusyk A, Tabassum DP, Altrock PM, Almendro V, Michor F, Polyak K. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. Nature 2014;514:54-8.
8. Waclaw B, Bozic I, Pittman ME, Hruban RH, Vogelstein B, Nowak MA. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. Nature 2015;525:261-4.
9. Janiszewska M, Liu L, Almendro V, Kuang Y, Paweletz C, Sakr RA, *et al. In situ* single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer. Nat Genet 2015;47:1212-9.
10. Almendro V, Cheng YK, Randles A, Itzkovitz S, Marusyk A, Ametller E, *et al.* Inference of tumor evolution during chemotherapy by computational modeling and *in situ* analysis of genetic and phenotypic cellular diversity. Cell Rep 2014;6:514-27.
11. Shirinifard A, Thiagarajan S, Vogel P, Sablauer A. Detection of phenotypic alterations using high-content analysis of whole-slide images. J Histochem Cytochem 2016;64:301-10.
12. Gyanchandani R, Lin Y, Lin HM, Cooper KL, Normolle DP, Brufsky AM, *et al.* Intra-tumor heterogeneity affects gene expression profile test prognostic risk stratification in early breast cancer. Clin Cancer Res 2016. pii: Clincanres. 2889.2015. [Epub ahead of print].
13. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012;366:883-92.
14. Kumar A, Boyle EA, Tokita M, Mikheev AM, Sanger MC, Girard E, *et al.* Deep sequencing of multiple regions of glial tumors reveals spatial heterogeneity for mutations in clinically relevant genes. Genome Biol 2014;15:530.
15. Govindan R. Cancer. Attack of the clones. Science 2014;346:169-70.
16. Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A, *et al.* Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. J Pathol 2013;231:21-34.
17. Rivenbark AG, O'Connor SM, Coleman WB. Molecular and cellular heterogeneity in breast cancer: Challenges for personalized medicine. Am J Pathol 2013;183:1113-24.
18. Sugihara Y, Taniguchi H, Kushima R, Tsuda H, Kubota D, Ichikawa H, *et al.* Laser microdissection and two-dimensional difference gel electrophoresis reveal proteomic intra-tumor heterogeneity in colorectal cancer. J Proteomics 2013;78:134-47.

19. Balluff B, Frese CK, Maier SK, Schöne C, Kuster B, Schmitt M, *et al. De novo* discovery of phenotypic intratumour heterogeneity using imaging mass spectrometry. J Pathol 2015;235:3-13.

20. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, *et al.* Tumour evolution inferred by single-cell sequencing. Nature 2011;472:90-4.

21. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature 2014;512:155-60.

22. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 2014;344:1396-401.

23. Cohen AA, Geva-Zatorsky N, Eden E, Frenkel-Morgenstern M, Issaeva I, Sigal A, *et al.* Dynamic proteomics of individual cancer cells in response to a drug. Science 2008;322:1511-6.

24. Irish JM, Hovland R, Krutzik PO, Perez OD, Bruserud Ø, Gjertsen BT, *et al.* Single cell profiling of potentiated phospho-protein networks in cancer cells. Cell 2004;118:217-28.

25. Camp RL, Chung GG, Rimm DL. Automated subcellular localization and quantification of protein expression in tissue microarrays. Nat Med 2002;8:1323-7.

26. Chung GG, Zerkowski MP, Ghosh S, Camp RL, Rimm DL. Quantitative analysis of estrogen receptor heterogeneity in breast cancer. Lab Invest 2007;87:662-9.

27. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, *et al.* Fluorescent *in situ* sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. Nat Protoc 2015;10:442-58.

28. Gough AH, Lezon TR, Faeder JR, Chennubhotla SC, Murphy RF, Critchley-Thorne RJ, *et al.* High-content analysis with cellular and tissue systems biology: A bridge between cancer cell biology and tissue-based diagnostics. In: Mendelsohn J, Howley PM, Israel MA, Gray JW, Thompson CB, editors. The Molecular Basis of Cancer. Philadelphia, PA: Saunders/Elsevier; 2015. p. 369-92.

29. Giesen C, Wang HA, Schapiro D, Zivanovic N, Jacobs A, Hattendorf B, *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat Methods 2014;11:417-22.

30. Nederlof M, Watanabe S, Burnip B, Taylor DL, Critchley-Thorne R. High-throughput profiling of tissue and tissue model microarrays: Combined transmitted light and 3-color fluorescence digital pathology. J Pathol Inform 2011;2:50.

31. Stack EC, Wang C, Roman KA, Hoyt CC. Multiplexed immunohistochemistry, imaging, and quantitation: A review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis. Methods 2014;70:46-58.

32. Gerdes MJ, Sevinsky CJ, Sood A, Adak S, Bello MO, Bordwell A, *et al.* Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. Proc Natl Acad Sci U S A 2013;110:11982-7.

33. Schubert W, Bonnekoh B, Pommer AJ, Philipsen L, Böckelmann R, Malykh Y, *et al.* Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. Nat Biotechnol 2006;24:1270-8.

34. Rao CR. Diversity and dissimilarity coefficients – A unified approach. Theor Popul Biol 1982;21:24-43.

35. Potts SJ, Krueger JS, Landis ND, Eberhard DA, Young GD, Schmechel SC, *et al.* Evaluating tumor heterogeneity in immunohistochemistry-stained breast cancer tissue. Lab Invest 2012;92:1342-57.

36. Almendro V, Kim HJ, Cheng YK, Gönen M, Itzkovitz S, Argani P, *et al.* Genetic and phenotypic diversity in breast tumor metastases. Cancer Res 2014;74:1338-48.

37. Park SY, Gönen M, Kim HJ, Michor F, Polyak K. Cellular and genetic diversity in the progression of *in situ* human breast carcinomas to an invasive phenotype. J Clin Invest 2010;120:636-44.

38. Rose CJ, Naidoo K, Clay V, Linton K, Radford JA, Byers RJ. A statistical framework for analyzing hypothesized interactions between cells imaged using multispectral microscopy and multiple immunohistochemical markers. J Pathol Inform 2013;4:4.

39. Clarke GM, Zubovits JT, Shaikh KA, Wang D, Dinn SR, Corwin AD, *et al.* A novel, automated technology for multiplex biomarker imaging and application to breast cancer. Histopathology 2014;64:242-55.

40. Steininger RJ 3rd, Rajaram S, Girard L, Minna JD, Wu LF, Altschuler SJ. On comparing heterogeneity across biomarkers. Cytometry A 2015;87:558-67.

41. Bello M, Can A, Tao X. Accurate Registration and Failure Detection in Tissue Micro Array Images. In Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on 2008. IEEE; 2008.

42. Woolfe F, Gerdes M, Bello M, Tao X, Can A. Autofluorescence removal by non-negative matrix factorization. IEEE Trans Image Process 2011;20:1085-93.

43. Padfield D, Rittscher J, Roysam B. Spatio-Temporal Cell Segmentation and Tracking for Automated Screening. In Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on 2008. IEEE; 2008.

44. Frangi AF, Niessen WJ, Vincken KL, Viergever MA. Multiscale vessel enhancement filtering. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI'98. Berlin Heidelberg: Springer; 1998. p. 130-7.

45. Santamaria-Pang A, Huang Y, Rittscher J. Cell Segmentation and Classification via Unsupervised Shape Ranking. In Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on 2013. IEEE; 2013.

46. Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved automatic detection and segmentation of cell nuclei in histopathology images. IEEE Trans Biomed Eng 2010;57:841-52.

47. Al-Kofahi Y, Lassoued W, Grama K, Nath SK, Zhu J, Oueslati R, *et al.* Cell-based quantification of molecular biomarkers in histopathology specimens. Histopathology 2011;59:40-54.

48. Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 2006;54:4311-22.

49. Francis K, Palsson BO. Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. Proc Natl Acad Sci U S A 1997;94:12258-62.

50. Schneider TD, Stormo GD, Gold L, Ehrenfeucht A. Information content of binding sites on nucleotide sequences. J Mol Biol 1986;188:415-31.

51. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53-65.