RESTORE: **R**obust int**E**n**S**i**T**y n**OR**malization m**E**thod for Multiplexed Imaging

Young Hwan Chang*, Koei Chin, Guillaume Thibault,
Jennifer Eng, Lydia Grace, Joe W. Gray*

October 3, 2019

Abstract

Recent advances in multiplexed staining and imaging technologies promise to significantly improve the understanding of the functional states of individual cells and the interactions between the cells that comprise complex tissues. This often requires compilation of results from multiple samples. Quantitative integration of information between samples is complicated by variations in staining intensity and background fluorescence due to differences in fixation and staining between samples, and other technical artefacts that obscure biological variations. Failure to remove these unwanted artefacts will complicate downstream analysis and diminish the value of multiplexed imaging for clinical applications. Here, we describe a **R**obust int**E**n**S**i**T**y n**OR**malization m**E**thod (**RESTORE**) to compensate for unwanted variations in multiplexed imaging. To do this, first, we automatically identify negative control cells for each marker within the same tissue and then use their expression levels to infer background signal level. Second, the intensity profile is normalized by the inferred background level of the negative control cells to remove between-sample variation. We evaluated the performance with theoretical simulations and tested the method using two different datasets; one comprised of multiplex images of three adjacent sections cut from a tissue microarray (TMA) carrying 59 cores from separate breast cancers and stained on different days, and one comprised of a pair of longitudinal biopsies obtained from one patient where tissue biopsy, fixation, processing and multiplexed staining/imaging were done at different times. We demonstrate that **RESTORE** can remove unwanted variations effectively and shows robust performance. Our code and the demonstration datasets are freely available for download at `https://gitlab.com/Chang_Lab/cycif_int_norm`.

1 Introduction

Recent developments in multiplexed staining and imaging such as cyclic immunofluorescence (CyCIF) [1] [2], multiplexed IHC [3], CODEX [4] and other multiplexed imaging methods have greatly expanded the palette that researchers and pathologists can use to visualize and analyze tissue sections, allowing deep in situ assessment of the tumor microenvironment complexities [5–9]. Multiplexed analyses have the advantages of both highly multiplexed detection and retention of morphological context at the level of single cell and subcellular compartments. However, integration of information from multiple samples is challenging due to the lack of normalization procedures that can correct for technical variations in staining intensities between samples that result from variations in fixation, antibody concentration, etc so that biological differences between samples can be accurately assessed [5, 10, 11].

At present, the most common approach for the quantitative assessment of images of IHC- and IF-labelled material is an analysis technique commonly referred to as 'gating' or 'binary thresholding' based on single cell features. Essentially, a particular pixel intensity level (the threshold) is manually defined and then used to demarcate what is considered to be 'signal' (the immuno-labelled material of interest) and 'noise' (non-specific material attributable to the immuno-labelling process). This manual thresholding procedure can only provide genuinely valid results if one adjusts the threshold cut-point for each individual sample to deal with such intensity variations. This approach is often used during analysis of multiplex IHC data [3]. However, manual thresholding is subjective and cannot be scalable by its nature. Our recent analyses of images acquired during Cyclic IF analyses of tissue microarrays showed that tissue to tissue variation in autofluorescence and/or nonspecific immunofluorescent staining required manual setting of individual thresholds for each tissue and marker, creating a bottleneck and introducing bias. As an extreme example, we observed in some cases that the intensity value of negative cells in one tissue were higher than the values of positive cells in a separate sample. In this case, setting a global threshold was not possible. This extreme variation also precluded taking full advantage of the quantitative nature of the fluorescence images.

Recently, unsupervised clustering approaches [12–15] have been adopted for identification of different cell types from a continuous intensity distribution, instead of binarization. However,

*Corresponding author: {chanyo, grayjo}@ohsu.edu

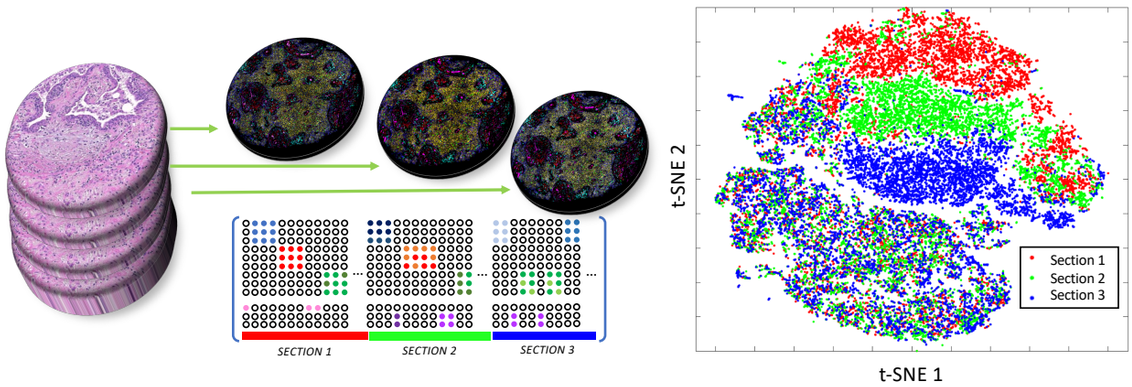

Figure 1: Multiplexed immunostaining intensity varies across samples; intensity profiles from three adjacent sections ($5\mu m$) show intensity variations. t-SNE embedding of individual single cell intensity features from three adjacent sections show uniformly distributed in immune cell types (bottom half) but clustered group in cancer cell type (top half) due to the intensity variation in cytokeratin (CK) markers.

the sample-to-sample intensity variations due to the technical artifacts throughout the procedure cause cells to cluster by samples or batches, instead of their cell types. For the above example we mentioned, applying unsupervised clustering without compensating intensity variation will yield a mixture of positive and negative cells from two samples due to batch effects. Therefore, the corrections of the unwanted intensity variations due to technical artifact and batch-effect within a group of samples are required as pre-processing steps for unsupervised clustering approaches to identify cell types. Quantile normalization in which intensity measurements that encompass values between 1st and the 99th percentile are aligned, is often used as a preprocessing step [1] to normalize intra-sample variations. However, if samples contain few or no positive cells for a certain marker, different cell populations, or different intensity distributions, quantile normalization may cause confounding variations by changing overall intensity profiles.

We now propose **RESTORE** for image quantification in a multiplexed imaging platform to address these issues. A key feature in **RESTORE** is the recognition that some cell types defined by reference markers in the multiplex can be safely assumed to have background levels of for other markers in the multiplex. For example, immune cells can be assumed to have background levels of cancer associated cytokeratins and vice versa. **RESTORE** uses these orthogonal staining patterns to set negative staining thresholds as a guide to further image intensity normalization. We then make the simplifying assumption that signals in the positive channels can be normalized by dividing by the inferred background level. This assumption is inaccurate for low signal levels but becomes increasingly accurate for high signals based on our simulation study (Section 4.5). We demonstrate the feasibility of the proposed approach for intensity normalization of multiplexed image data for robust analysis and comparison of tissue samples during analysis of a data set comprised of Cyclic IF analyses of three adjacent sections cut from a tissue microarray comprised of 59 cores from diverse breast cancers and stained for 37 proteins and phosphoproteins selected to identify diverse tumor and stromal cells and functional status (See Table S1).

2 Background and Motivation

The general problem is that the intensity features produced by multiplexed immunostaining have different intensity variations across markers, tissues, batches, etc. As an example, Figure 1 shows intensity variation from three adjacent sections (considered as *almost* technical replicates since each section is acquired by $5\mu m$ thickness difference) where these unwanted variations may be caused by technical issues such as batch effect, exposure time, protocols, etc., or tissue preparation. These sections were stained on separate days using Cyclic IF for 37 proteins and phosphoproteins identifying tumor, immune and stromal cells and functional states. Individual cells in the multiplex images were segmented using watershed segmentation followed by morphological operation and staining intensities were calculated for each segmented cell [2]. Figure 1 (right) shows t-SNE projection based on mean intensity profiles where each dot represents a single cell feature and red, green and blue color indicate section 1, 2, and 3 respectively. The top region of t-SNE shows cancer cell clusters from each section but due to intensity variation of cytokeratin (CK) markers, it shows batch effect (i.e., red, green and blue clusters). In contrast to cancer cell types, the bottom region of t-SNE represents immune cell types showing more a uniformly distributed pattern between samples

which we expect to see in the ideal setting, i.e., if there is no intensity variation across three adjacent sections.

In IHC, the antibody validation reference is performed with control tissues known to contain the antigen of interest detected by an identical staining method. For example, "sausage technique" has been used where the entire sample has been proposed as a reference control standard [16, 17]. It is recommended that validation studies be carried out on multi-tissue control blocks containing both known-positive and known-negative normal and tumor tissues [18]. If we consider only a few markers in typical IHC or IF staining, it is feasible to include control tissue samples in staining/imaging process for individual markers and address antibody validation or staining variation by determining an appropriate transformation or normalization for each marker. Unfortunately, it is not practical to add control tissue samples for reference of individual Cyclic IF channels since there exist tens of markers (> 40) that may require more control tissue samples than the test sample to cover various cell types and functional states. In addition, even though we could normalize features based on reference samples, there still exist other uncontrollable factors causing intensity variations such as tissue fixation, processing in a different way than the test tissue, and antibody lot-to-lot variations especially for large cohort study.

In parallel, computational approaches have been proposed to adjust for unwanted variation and can be divided into two broad categories: 1) global adjustment and 2) an application specific method [19]. As an example, quantile normalization (QN), in which intensity measurements that encompass values between 1st and the 99th percentile are aligned, generally regarded as a self-contained step that plays no role in the downstream analysis of the data, belongs to the first category. In the second category, we find methods that incorporate adjustment into the main analysis of interest. For instance, the batch effects can be handled by explicitly adding "batch terms" to a linear model. Other linear model-based methods such as factor analysis attempt to infer the unwanted variation from the data and then adjust for it. However, these models require technical replicates or *a priori* information to identify batch effects, which is not feasible in our case, especially because tissue biopsies are precious and are often difficult to obtain in sufficient quantities. In addition, for needle biopsy samples, it is difficult to consider individual region of interests (ROIs) as technical replicates due to the heterogeneity.

3 Result and Discussion

One could consider two approaches to evaluate the proposed approach: 1) use cell classification based on manual gating as a ground truth, or 2) use technical replicates as a ground truth. The former approach needs manual gating for cell type classification, but it might be subjective and time consuming to classify various cell types from many TMAs (177 total, 59×3 adjacent sections as shown in Figure 2 (left)). For the latter one, we can use 59 TMAs that have three adjacent tissue sections. Since each TMA sample has three adjacent tissue samples, we could consider them as *almost* technical replicates where we expect to see a similar cell type component in their population. For instance, if there is no intensity variation across these three adjacent tissue samples, no matter where we draw threshold line or gate, the population of those cell types should be similar across three adjacent tissue samples. Thus, we measure the cell component as a metric and evaluate how the proposed approach compensates intensity variation. We also compare the correlation coefficient of cell composition between three adjacent sections with and without using the proposed normalization technique. Finally, we illustrate a clinical use case by applying the proposed approach to the study with longitudinal biopsies.

3.1 Application with TMAs analysis using 3 adjacent tissue sections

We compare correlation coefficient of individual cell population of three adjacent TMAs across 59 samples where r_{ij} represents correlation coefficient of cell population between the i-th and the j-th section as shown in Figure 2 (left). We define cell population by two approaches: 1) counting positive cell count for individual markers by inferring background signal and 2) using an unsupervised clustering approach by changing the number of clusters.

3.1.1 Cell population comparison by inferring background signal

First, we count positive cell population across all the Cyclic IF markers by inferring background signal. In order to illustrate intensity variations across three adjacent sections, we infer background signal in two different approaches. For a local approach, we infer background signal for an individual adjacent section, and for a global approach, we combine intensity features across three adjacent sections and infer background signal. If there is no intensity variation, we expect the local and global approaches to show similar results. With a local approach, we observe high correlation of cell composition across three adjacent TMAs as shown in Figure 2 (middle). Moreover, with

a global approach, the correlation of cell composition across three adjacent sections is poor, due to the intensity variation across samples as shown in Figure 2 (right). We note that in Figure 2 (middle, the proposed result), we found that three cores in TMAs show poor correlation coefficient due to the technical artifacts such as segmentation or staining issues.

Second, we compare the correlation of cell populations across individual markers ($n = 18$) as shown in Figure 3. Similar to the previous result, the local approach shows better correlation which confirms that the proposed method provides a robust cell classification result by inferring background signal from the negative control group and compensating intensity variation.

Lastly, we use the coefficient of variation ($c_v = \sigma/\mu$, known as relative standard deviation (RSD)) of positive cell count based on the inferred threshold of background signal where μ and σ represent mean and variation of positive cell counts across three adjacent TMAs respectively. The coefficient of variation is a standardized measure of dispersion of a frequency distribution, which is commonly used in analytical chemistry to express the precision and repeatability of an assay. We use coefficient of variation to address the positive cell number across individual markers. Figure 4 shows the extent of variability of positive cell count in relation to the mean of population across individual marker. Since there exist intensity variation across IF intensity features, the coefficient of variation of the global approach shows large variation compared to the coefficient of variation of the local approach. For CD45 marker, although the global approach shows slightly better performance, we note that a scale of the value is quite small compared to other c_v for different markers. Similarly, for Ki67 marker, we do not see much difference between the local and global approach.

3.1.2 Cell population comparison with and without the proposed intensity normalization by using unsupervised clustering approach

We apply an unsupervised clustering algorithm (k-means) with and without the proposed intensity normalization method on the three adjacent sections to define cell types. If the proposed method reduces intensity variation properly, we expect to see the identified cellular population by unsupervised clustering to be similar across the three adjacent sections. We run k-means clustering with $N = 5, 10, 15$ and 20 and calculate correlation coefficient of cell populations as shown in Figure 5. The top rows illustrate correlation of cell types across three adjacent sections without intensity normalization and bottom rows show correlation of cell types with the intensity normalization approach.

Since the proposed approach reduces intensity variation, the correlation coefficient based on clustered group component shows high correlation. On the other hand, due to the intensity variation, unsupervised clustering often identifies cell types by each section (i.e., one cluster originates from section 1 and the other from the other section). Thus, without normalization, unsupervised clustering identifies different cell populations, and the correlation of cell population is lower.

As the number of clusters (N) increases, the correlation decreases slightly although the proposed method still improves correlation compared to no normalization. Even though we expect to see similar populations across three adjacent sections, there are still small variations in cell populations which may separate as the number of clusters increases as shown in Figure 5 (more details shown in Figure S1) and cause lower correlation of cell composition. As an example, Figure S1 (top left) shows cell population distribution when we cluster cells into five clusters ($N = 5$). Without normalization, clustered ID 1 population is mainly from batch 1 (or section 1), clustered ID 2 population is mainly from batch 2 (or section 2), and clustered ID 4 is mainly from batch 3 (or section 3) due to the intensity variation. On the other hands, with the proposed approach, each

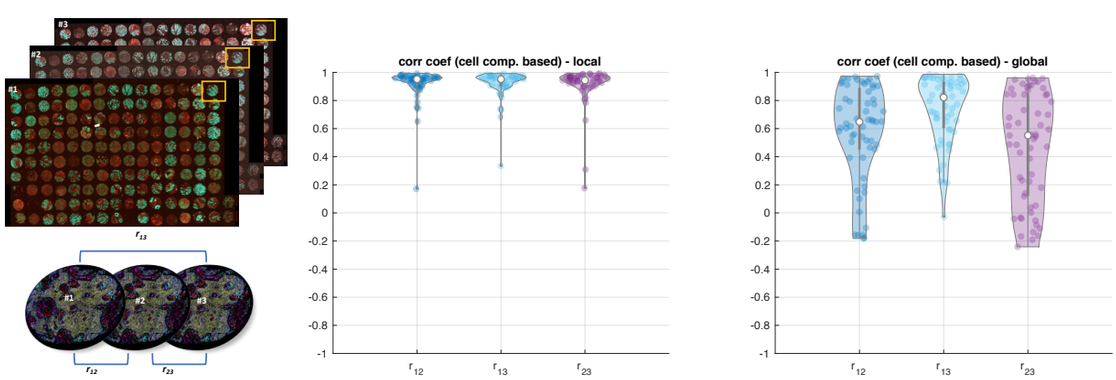

Figure 2: Example of three adjacent TMA sections (left, $n = 59$ with three adjacent sections) and correlation co- efficient based on group component using unsupervised clustering across these samples; middle and right figures represent correlation coefficient with and without the proposed approach respectively.

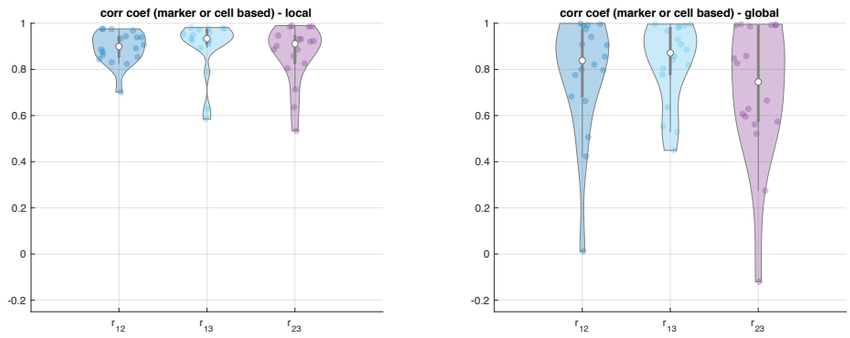

Figure 3: correlation coefficient based on cell component across individual markers ($n = 18$).

clustered ID shows similar distribution across three adjacent sections.

3.2 Application with longitudinal biopsies sample study

We validate the proposed approach with longitudinal biopsy sample study, the Serial Measurements of Molecular and Architectural Responses to Therapy (SMMART) trials [20] where tissue biopsy, fixation, processing and multiplexed imaging are done at different times. Since we need to identify cellular composition changes for comparative study from longitudinal biopsies (before/after drug treatment), it is critical to remove unwanted variation and integrate two datasets together for unsupervised clustering analysis. Because of the previous aforementioned aim, and because we often observe different cell population in paired biopsies, we cannot use the quantile normalization approach.

Figure 6 (left) shows a heat map of intensity profiles. Without normalization, if we apply any unsupervised clustering, it will cluster cells based on their batches (i.e., Bx1 and Bx2) as we can see huge variations. By applying the proposed approaches, intensity profiles look similar to each other across all the markers. Figure 6 (right) shows intensity distribution changes before/after normalization. As we can see, the proposed approach aligns intensity profiles by removing any unwanted variation and matching the intensity level of negative control. Note that if we use quantile normalization instead of using the proposed approach, we may generate an obscure shift of intensity distribution depending on their cell populations as shown in CK17 and HER2 and we expect that this causes artifact. More detailed information such as a comparison between before/after normalization for all other markers is shown in Figure S2.

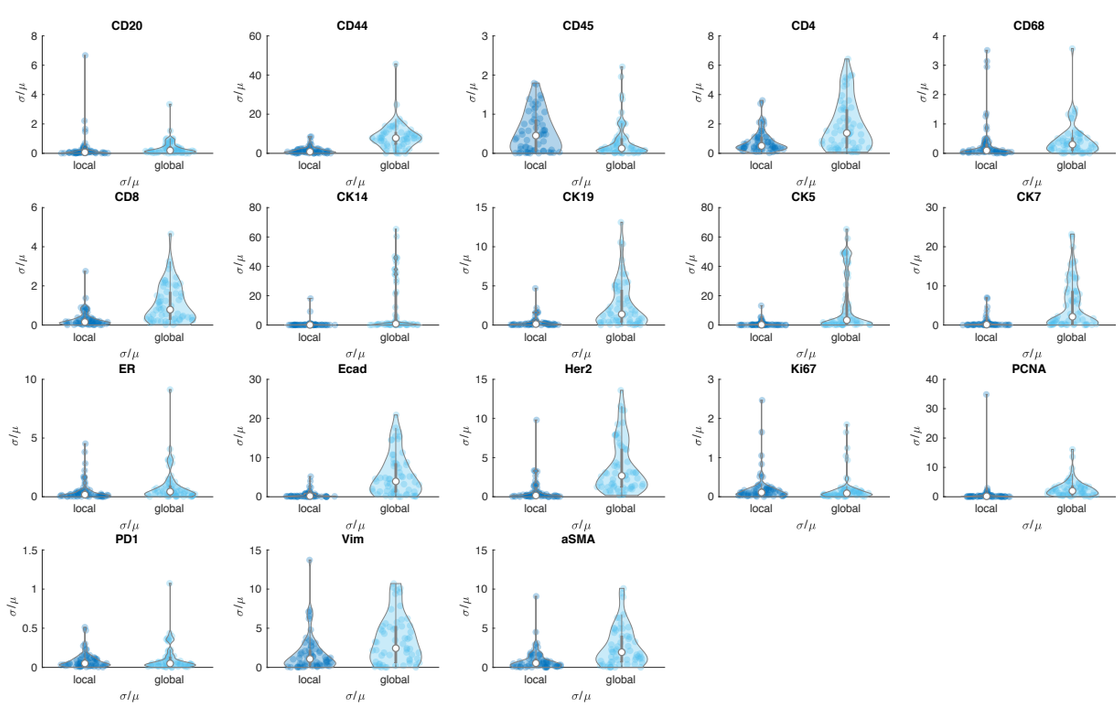

Figure 4: Coefficient of variation ($c_v = \frac{\sigma}{\mu}$) of positive cell counts based on local (left) and global (right) threshold values.

4 Method

We now propose **RESTORE** as a practical strategy for correction for batch effects during staining that does not require adding control tissues or using technical replicates in staining and processing. Our approach is composed of two parts: 1) Definition of mutually exclusive marker pairs or cell types that are known to be positive for a given markers and negative for others based on biological literature or data driven based, and 2) inference of background levels for specific markers in cells that are defined to be negative for those markers based on positive identifying markers. The positive/negative marker sets used in this study are defined in Table S1.

For a given reference marker, we can use target marker (shown in Table S1) which shows mutually exclusiveness or positive-negative associations to define the background levels of negative controls. For a given reference marker, since we expect the background levels of negative controls should be aligned within the same ranges across tissue samples, we can normalize a reference marker expression by the inferred background levels. Also, since we have thousands or millions of cells in tissue and various markers to characterize different cell types, it is not difficult to find a negative control for individual markers from the same tissue sample.

4.1 Concepts for intensity normalization

Here, we introduce the fundamental concepts for intensity normalization. For a given reference marker, expression levels of the negative control are known *a priori* to be truly unassociated with the factor of interest. On the other hand, positive control markers are those expression levels that are known *a priori* to be truly associated with the factor of interest. For example, if the factor of interest is finding immune cell, CD45 would be a positive control, and a negative control would be any cytokeratin (CK-) markers. Since the expression of the negative control group is known to be unassociated with the factor of interest, there is no danger in picking up any of the relevant biology and thus we could use them to remove unwanted variations. Even though individual tissue sample may have different background levels, we identify negative cells for each tissue sample by using mutually exclusive marker pairs in Table S1 and then make their expression level below the level of any positive cells by normalized with the inferred background levels.

For our cyclic multiplexed IF (cmIF) imaging [2], we have tens of markers as shown in Table S1. Thus, for a given i-th marker as a reference, there exists at least one marker, the j-th marker (target), showing a mutually exclusive expression pattern. The positive cell of the i-th marker cannot express the positive signal of j-th marker as shown in Figure 7 (left). This is a reasonable assumption; since current cmIF panel includes immune/cancer markers (biologically mutually exclusive). Therefore, there should exist at least one mutually exclusive marker pair. Herein, we do not consider if there exists no positive cell in the reference marker since this marker is not useful

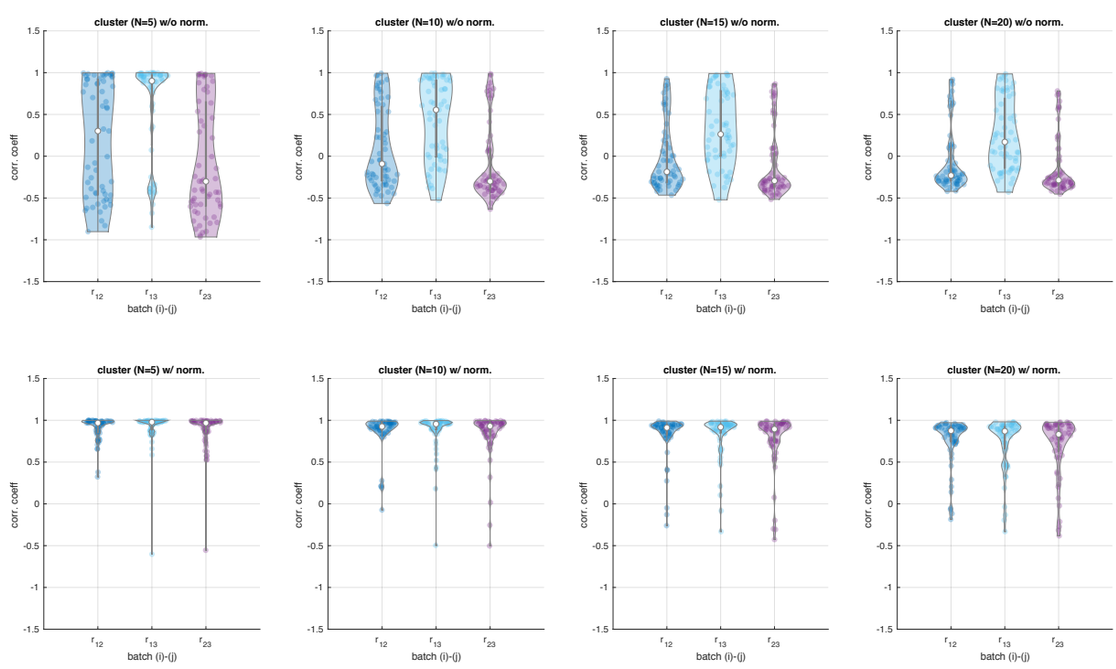

Figure 5: Correlation coefficient based on group component using unsupervised clustering across three adjacent TMA samples ($n = 59$) where top figure shows the result without the proposed approach and bottom figure shows the result with the proposed approach.

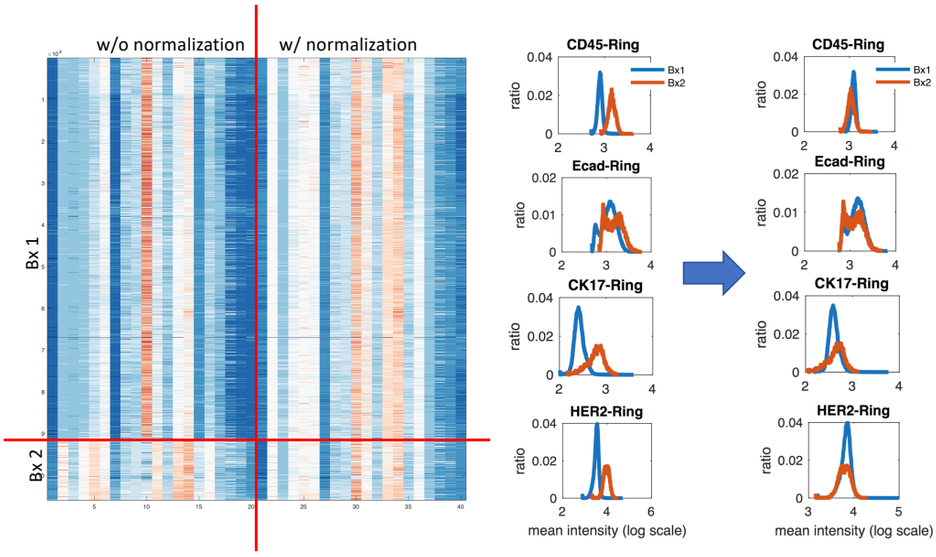

Figure 6: Cyclic multiplexed IF Intensity distribution across Bx 1 and Bx 2: (left) heat map shows intensity features without and with the proposed normalization and (right) intensity distribution of selected markers (CK45, Ecad, CK17, HER2) (all other markers' distribution is shown in Figure S2.

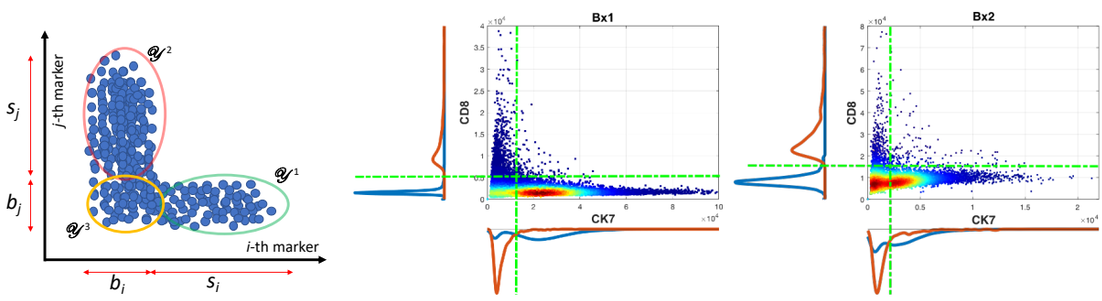

Figure 7: Conceptual illustration (left) and example of scatter plot using longitudinal biopsy samples (Bx1: middle and Bx2: right): mean intensity distribution of two markers shows different baseline and signal level of each marker (CK7 and CD8) in Bx1/2 where green lines illustrate example of inferred intensity of baseline for each marker, red and blue line shows corresponding cell population density.

for further analysis by definition (i.e., no positive staining in the tissue sample).

4.2 Identifying mutually exclusive marker pairs

In order to infer background signal, for a given i-th marker, we need to find most mutually exclusive marker pairs to identify negative control group. This procedure can be done by using either 1) biologically known mutually exclusive marker information (i.e., a cancer vs. immune marker), or 2) a data-driven approach by identifying mutually exclusive information from the i-th marker and all other maker pairs. For the latter case, we use a Singular Value Decomposition (SVD) to measure mutually exclusiveness.

Define $D = \begin{bmatrix} x_{i1} & x_{j1} \\ \cdots & \cdots \\ x_{in} & x_{jn} \end{bmatrix} \in \mathbb{R}^{n\times 2}$ where i, j represents the i-th and j-th markers respectively and $x_{\{\cdot\}}$ represent mean intensity of individual cell. By using SVD, we can factorize $D = U\Sigma V^*$ where $U \in \mathbb{R}^{n\times n}$, $\Sigma \in \mathbb{R}^{n\times 2}$ and $V^* \in \mathbb{R}^{2\times 2}$. U and V^* can be viewed as rotation matrix and the diagonal entries σ_i of Σ are known as the singular values of D which can be regarded as a scaling matrix. Since the singular values can be interpreted as the semi-axis of an ellipse in 2D, we measure mutually exclusive of two markers expression as a ratio (r $= \frac{\sigma_2}{\sigma_1}$). If two markers are highly correlated with each other, we will get an elongated ellipse (i.e., r is close to zero). On the other hand, if they are mutually exclusive, you will get a more circular shape (i.e., close to 1).

For each marker, we rank r across all pairs and choose top 5 mutually exclusive marker pairs. Supplementary Table S1 shows a comparison between the data-driven approach (left) and biologically known mutually exclusive marker pairs (right). We found that the data-driven approach

identifies similar and consistent marker pairs to biologically known mutually exclusive pairs. The red color in the data-driven approach indicates the matched marker pairs presented in biological knowledge-based approach. Similarly the green color in the biological knowledge-based approach indicates the matched marker pairs presented in the data-driven approach.

Mutually exclusive marker identification works reasonably well in many cases and confirms that biological known mutually exclusive marker pairs show mutually exclusiveness in our cmIF imaging data. From data-driven approach, in general, immune markers show most mutually exclusiveness with cytokeratin markers. Interestingly, we also observe more specific mutually exclusiveness patterns, for instance, the top mutually exclusive pairs with CD20 is B-cell marker (i.e., lymphoid lineage immune cell with $CD45^+$) are $CD68^+$, a macrophage (myeloid lineage) marker, and $CD31^+$, an endothelial cell marker and it is consistent with biologically known mutually exclusive cell types.

We do not see CK5 in the data-driven approach. This is because staining of CK5 marker is bad in our dataset. We also observe that a few cells often show stain in both cancer and immune markers and thus, we need to carefully check marker staining quality for this analysis. Herein, we simply use (fixed) biologically-driven mutually exclusive marker pairs to infer background signal, instead of identifying the most mutually exclusive pairs for individual cores from the tissue microarrays (TMAs) dataset and a longitudinal study.

4.3 Identifying cell types via non-negative matrix factorization (NNM) or sparse subspace clustering (SSC)

For a given mutually exclusive marker pair, define mean intensity profiles of two markers as $\mathcal{Y} = \begin{bmatrix} \mathcal{Y}^1 \\ \mathcal{Y}^2 \\ \mathcal{Y}^3 \end{bmatrix} \in \mathbb{R}^{N\times 2}$ where $N(=\sum_{i=1}^{3} n_i)$ represents the total number of cells, $\mathcal{Y}^1 \in \mathbb{R}^{n_1\times 2}$ represents 2-dimensional (i-th and j-th markers) mean intensity profile of a set of the i-th marker positive cells, $\mathcal{Y}^2 \in \mathbb{R}^{n_2\times 2}$ represents intensity profile of a set of the j-th marker positive cells and $\mathcal{Y}^3 \in \mathbb{R}^{n_3\times 2}$ represents intensity profile of a set of negative cells of both the i-th and the j-th marker where n_i represents the number of cells belong to $\mathcal{Y}^i$.

The l-th row of $\mathcal{Y}$ can be denoted by $y'_l = \begin{bmatrix} y_{l1} & y_{l2} \end{bmatrix} = \begin{bmatrix} b^l_i + s^l_i & b^l_j + s^l_j \end{bmatrix} \in \mathbb{R}^{1\times 2}$ where b^l_i and b^l_j represents baseline (autofluorescence level) of the i-, j-th marker respectively and s^l_i and s^l_j represents signal level of the i-th, j-th marker respectively. By definition (and expectation of high signal-to-noise ratio (SNR) of immunofluorescence intensity profile), we assume $s^l_i \gg b^l_i > 0$ and $s^l_j \gg b^l_j > 0$:

Lemma 1. *Consider mutually exclusive set (i, j)- markers, i.e., $\{\mathcal{Y}^1, \mathcal{Y}^2, \mathcal{Y}^3\}$ where $y'_p \in \mathcal{Y}^p$, p represents group index, i.e., $p = \{1, 2, 3\}$. Then y_p cannot be represented by linear combination of $\alpha y_q + \beta y_r$ with constraint ($\alpha > 0$ and $\beta > 0$) where $p \neq q \neq r$ and (p, q, r) represents any permutations of $(1, 2, 3)$.*

Proof. (suppose not) i.e., $y_p = \alpha y_q + \beta y_r$ where we simply consider $p = 1, q = 2$ and $r = 3$:

$$\begin{bmatrix} b^p_i + s^p_i \\ b^p_j \end{bmatrix} = \alpha \begin{bmatrix} b^q_i \\ b^q_j + s^q_j \end{bmatrix} + \beta \begin{bmatrix} b^r_i \\ b^r_j \end{bmatrix} \tag{1}$$

Then, reformulating this

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = -\frac{1}{s^q_j b^r_i + (b^r_i b^q_j - b^q_i b^r_j)} \begin{bmatrix} b^r_j & -b^r_i \\ -(b^q_j + s^q_j) & b^q_i \end{bmatrix} \begin{bmatrix} b^p_i + s^p_i \\ b^p_j \end{bmatrix} \tag{2}$$

Recall $s^l_i \gg b^l_i$ or simply assume $b^r_j \approx b^p_j \approx b^q_j$, $b^r_i \approx b^p_i \approx b^q_i$:

$$\begin{aligned} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} &= -\frac{1}{s^q_j b^r_i + (b^r_i b^q_j - b^q_i b^r_j)} \begin{bmatrix} (b^r_j b^p_i - b^r_i b^p_j) + b^r_j s^p_i \\ -(b^q_j s^p_i + s^q_j b^p_i + s^q_j s^p_i) + (b^q_i b^p_j - b^p_i b^q_j) \end{bmatrix} & (3) \\ &\approx -\frac{1}{s^q_j b^r_i} \begin{bmatrix} b^r_j s^p_i \\ -(b^q_j s^p_i + s^q_j b^p_i + s^q_j s^p_i) \end{bmatrix} \implies \begin{bmatrix} (-) \\ (+) \end{bmatrix} & (4) \end{aligned}$$

(by contradiction). □

By Lemma 1, simple non-negative matrix factorization (NNMF) or sparse subspace clustering (SSC) or other method (e.g. gaussian mixture model) with this constraint should be able to identify the group of cell types for a given mutually exclusive pattern.

4.3.1 Non-negative Matrix Factorization (NNMF)

$$\min \|\mathcal{Y} - WH\|_F \text{ subject to } W \geq 0, H \geq 0 \tag{5}$$

if we consider 2-dimensional feature set (i.e., mean intensity profile), there exists a trivial solution, i.e., $W = \mathcal{Y}$ and $H = I$ so we need to consider additional constraint such as $H \neq I$. If we include more than 2 dimensional features by adding more features including total intensity, other cellular mean intensity (nuclear-, cytoplasm-, or cellular), we can simply use (5).

4.3.2 Sparse Subspace Clustering (SSC)

We can use sparse subspace clustering with additional constraint by considering that each data point in a union of subspaces can be efficiently reconstructed by a combination of other points in the dataset. More precisely, each data point for data point $y_m = \mathcal{Y}' c_m$ where $c_{mm} = 0$ and $c_{ij} \geq 0$ (additional constraint). Then,

$$\min \|c_i\|_q \text{ subject to } y_m = \mathcal{Y}' c_m, c_{mm} = 0, c_{ij} \geq 0. \tag{6}$$

$$\min \|\mathbf{C}\|_1 \text{ subject to } \mathcal{Y}' = \mathcal{Y}'\mathbf{C}, diag(\mathbf{C}) = 0, c_{ij} \geq 0. \tag{7}$$

In practice, since the intensity values of y_m are all positive and l_1 optimization penalizes a sparsity of coefficient $\mathbf{C}$, it does not choose negative coefficient of c_{ij} even without having additional constraint (i.e.., $c_{ij} \geq 0$) in the optimization problem in (6) or (7). In this paper, we simply use the optimization problem in (7) to identify two groups, i.e., positive cell and negative cell group to infer background signal.

4.4 Inferring autofluorescence level with mutually exclusive marker pairs

Here, we simply use SSC to divide $\mathcal{Y}$ into two groups, either $[\mathcal{Y}^1 \oplus \mathcal{Y}^3$ and $\mathcal{Y}^2]$ or $[\mathcal{Y}^1$ and $\mathcal{Y}^2 \oplus \mathcal{Y}^3]$, by selecting the number of clusters is equal to 2 (note that one could cluster more than 2 and assign each cluster into the group of interest for inferring background or autofluorescence signal). Since we are interested in inferring baseline signal of the i-th marker, we measure the i-th marker expression from the clustered group and determine the group for measuring baseline signal:

$$\begin{bmatrix} \mathcal{Y}^1 \\ \mathcal{Y}^2 \\ \mathcal{Y}^3 \end{bmatrix} = \begin{bmatrix} Y_1^1 & Y_2^1 \\ Y_1^2 & Y_2^2 \\ Y_1^3 & Y_2^3 \end{bmatrix} = \begin{bmatrix} B_i^1 + S_i^1 & B_j^1 + S_j^1 \\ B_i^2 + S_i^2 & B_j^2 + S_j^2 \\ B_i^3 + S_i^3 & B_j^3 + S_j^3 \end{bmatrix} = \begin{bmatrix} B_i^1 & B_j^1 \\ B_i^2 & B_j^2 \\ B_i^3 & B_j^3 \end{bmatrix} + \begin{bmatrix} S_i^1 & 0 \\ 0 & S_j^2 \\ 0 & 0 \end{bmatrix} \tag{8}$$

For a given clustered group, either $[\mathcal{Y}^1 \oplus \mathcal{Y}^3$ and $\mathcal{Y}^2]$ or $[\mathcal{Y}^1$ and $\mathcal{Y}^2 \oplus \mathcal{Y}^3]$, we can infer baseline signal of the i-th marker from either B_i^2 (from $\mathcal{Y}^2$) or (B_i^2 or B_i^3) (from $\mathcal{Y}^2 \oplus \mathcal{Y}^3$). Note that we can measure the autofluorescence level or baseline signal from each clustered group Y_p^q and identify background signal and signal level respectively as follows:

$$\begin{aligned}
\mathbb{E}[B_i^2] &\approx \mathbb{E}[Y_1^2] \approx \mathbb{E}[B_i^3] \approx \mathbb{E}[Y_1^3] && (9)\\
\mathbb{E}[B_j^2] &\approx \mathbb{E}[Y_2^1] \approx \mathbb{E}[B_j^3] \approx \mathbb{E}[Y_2^3] && (10)\\
\mathbb{E}[S_i^1] &\approx \mathbb{E}[Y_1^1] - \mathbb{E}[Y_1^3] \approx \mathbb{E}[Y_1^1] - \mathbb{E}[Y_1^2] && (11)\\
\mathbb{E}[S_j^2] &\approx \mathbb{E}[Y_2^2] - \mathbb{E}[Y_2^1] \approx \mathbb{E}[Y_2^2] - \mathbb{E}[Y_2^3] && (12)
\end{aligned}$$

4.5 Intensity normalization by inferring autofluorescence signal

For a given reference marker (i.e., i-th marker), intensity normalization step is straightforward by the fact that the autofluorescence level of negative controls (the i-th marker expression of B_i^1, B_i^2 or B_i^3) should be in the same ranges across batches or samples. As we infer background signal based on the negative control, we can scale intensity values by the inferred background signal level of the negative control for individual sample respectively to align intensity distribution. One could use maximum intensity value from the negative control instead of using mean intensity value. Thus, for a given reference marker, all the background level or baseline level of the negative control is below one, i.e., in the same ranges across samples.

As an example, consider two sample tissues p and q which have different gains G_p and G_q. This gain reflects any possible source of intensity variation such as tissue fixation, exposure time, batch effect, etc. Due to the different gain, intensity value of p and q could be different. For a given i-th marker, mean intensity measurement of single cell can be defined as follows:

$$\begin{aligned}
y_i^p &= G_i^p(b_i^p + s_i^p) \triangleq B_i^p + S_i^p \\
y_i^q &= G_i^q(b_i^q + s_i^q) \triangleq B_i^q + S_i^q
\end{aligned}$$

where b_i^p, b_i^q represent baseline or autofluorescence signal, s_i^p, s_j^q represent signal, and y_i^p and y_j^q represent measurement signal, i.e., single-cell mean intensity from sample p and q respectively. G_i^p and G_i^q represent gain value of the i-th marker for sample p and q respectively.

We infer $\bar{B}_i^p$ and $\bar{B}_i^q$ using mutually exclusive maker pairs and herein, we assume that we choose $\bar{B}_i^p (= G_i^p \bar{b}_i^p)$ and $\bar{B}_i^q = (G_i^q \bar{b}_i^q)$ from the maximum values from the negative controls (i.e., $\bar{b}_i^p = \max(b_{i,k}^p)$ and $\bar{b}_i^q = \max(b_{i,l}^q)$ where k and l represents cell index for sample p and q respectively). Then, we normalize intensity profiles (i.e., y_i^p and y_i^q) based on these values as follows:

$$\begin{aligned}
\bar{y}_i^p &\triangleq \frac{y_i^p}{\bar{B}_i^p} = \frac{G_i^p(b_i^p + s_i^p)}{\bar{B}_i^p} = \frac{G_i^p(b_i^p + s_i^p)}{G_i^p \bar{b}_i^p} = \frac{b_i^p}{\bar{b}_i^p} + \frac{s_i^p}{\bar{b}_i^p} \\
\bar{y}_i^q &\triangleq \frac{y_i^q}{\bar{B}_i^q} = \frac{G_i^q(b_i^q + s_i^q)}{\bar{B}_i^q} = \frac{G_i^q(b_i^q + s_i^q)}{G_i^q \bar{b}_i^q} = \frac{b_i^q}{\bar{b}_i^q} + \frac{s_i^q}{\bar{b}_i^q}
\end{aligned}$$

Therefore,

$$\bar{y}_i^p = \begin{cases} \frac{b_i^p}{\bar{b}_i^p} (\leq 1), & \text{if negative cell, i.e., } s_i^p = 0 \\ 1 + \frac{s_i^p}{\bar{b}_i^p}, & \text{if positive cell, i.e., } s_i^p > 0 \end{cases}, \bar{y}_i^q = \begin{cases} \frac{b_i^q}{\bar{b}_i^q} (\leq 1), & \text{if negative cell, i.e., } s_i^q = 0 \\ 1 + \frac{s_i^q}{\bar{b}_i^q}, & \text{if positive cell, i.e., } s_i^q = 0 \end{cases}$$

Note that $\frac{b_i^p}{\bar{b}_i^p} \leq 1$, $\frac{b_i^q}{\bar{b}_i^q} \leq 1$ and $\frac{s_i^p}{\bar{b}_i^p} \gg 1$, $\frac{s_i^q}{\bar{b}_i^q} \gg 1$ by definition (high signal to background ratio). So, all the negative cell from the sample p and q will be less than or equal to 1 and positive signal will be above 1 by normalization. Also, note that normalized measurement $\bar{y}_i^p$ and $\bar{y}_i^q$ are not depend on the gain term G_i^p and G_i^q anymore and if we assume $\bar{b}_i^q \approx \bar{b}_j^q$, normalized signal $\bar{y}_i^p$ and $\bar{y}_j^p$ are representing the true signal, i.e., s_i and s_j with the same scaling factor ($\bar{b}_i^q \approx \bar{b}_i^q$).

It is important to see whether normalization conserves the true signal ratio ($\frac{s_i^q}{s_i^p}$) for positive cell. Since normalized intensity for negative cell will be less than or equal to 1, we do not consider here. Assuming $\bar{b}_i^q \approx \bar{b}_i^q = \bar{b}_i$, we consider positive cell:

$$\frac{\bar{y}_i^q}{\bar{y}_i^p} = \frac{1 + \frac{s_i^q}{\bar{b}_i}}{1 + \frac{s_i^p}{\bar{b}_i}} = \frac{\bar{b}_i + s_i^q}{\bar{b}_i + s_i^p} \tag{13}$$

If signal level is high enough (high signal-to-baseline ratio (SBR)), i.e., $s_i^p \gg \bar{b}_i$ and $s_i^q \gg \bar{b}_i$,

$$\frac{\bar{y}_i^q}{\bar{y}_i^p} \approx \frac{s_i^q}{s_i^p} \tag{14}$$

Thus, for high SBR region, we could preserve the true signal ratio well but we may have distorted signal ratio when signal level is close to near baseline or background level. On the other hand, without normalization, we assume $\bar{b}_i^p \approx \bar{b}_i^q = \bar{b}_i$ and consider positive cells as below:

$$\frac{y_i^q}{y_i^p} = \frac{G_i^p(\bar{b}_i + s_i^p)}{G_i^q(\bar{b}_i + s_i^q)} \approx \frac{G_i^p}{G_i^q} \frac{s_i^p}{s_i^q} \tag{15}$$

For the last approximated term, we assume that signal level is high enough. We can see that the true signal ratio is scaled by the gain ratio if we do not normalize them.

In order to show this, we simulate data and compare the result with and without normalization, i.e., $\frac{y_i^p}{y_i^q}$ and $\frac{\bar{y}_i^p}{\bar{y}_i^q}$. Simulated result and heat map version of the result are shown in Figure S3 and S4 respectively. As we described above, as SBR increases, the normalized measurement is close to the true signal ratio. In Figure S4, heat map shows clearly the region of distortion in low SBR region with high true signal ratio. On the other hand, without normalization, the measurement is converged to gain ratio (i.e., G_2/G_1) where we have $G_2/G_1 = 0.5$ in this simulation.

Data accessibility and code sharing

For research reproducibility, our data (`download`) and code (`https://gitlab.com/Chang_Lab/cycif_int_norm`) will be available.

Acknowledgement

We thank Lauren Kronebush for helping scientific editing. The resources of the Exacloud high performance computing environment developed jointly by OHSU and Intel and the technical support of the OHSU Advanced Computing Center are gratefully acknowledged. This work was supported in part by the National Cancer Institute (U54CA209988, U2CCA233280) and SBIR (R44CA224994). YHC acknowledges the OHSU Center for Spatial Systems Biomedicine, Brenden Colson Center for Pancreatic Care and Biomedical Innovation Program Award from the Oregon Clinical & Translational Research Institute.

Author Contribution

YHC and JWG conceived and designed the study. KC, JE and LC performed the experiments and GT performed segmentation and feature extraction. YHC developed algorithms, wrote the code, analyzed the data and interpreted the results. YHC drafted the manuscript and KC, GT, JE and JWG reviewed the manuscript. JWG supervised and oversaw the study. All authors critically revised the manuscript and provided intellectual content.

References

[1] Lin, J.R., et al., Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. Elife, 2018. 7.

[2] Jennifer Eng, G.T., Shiuh-Wen Luoh, Joe W. Gray, Young Hwan Chang, and Koei Chin, Cyclic Multiplexed-Immunofluorescence (cmIF), a Highly Multiplexed Method for Single-Cell Analysis, in Biomarkers for Immunotherapy of Cancer, A.C. Magdalena Thurin, Francesco Marincola, Editor. 2019, Humana Press.

[3] Tsujikawa, T., et al., Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. Cell Rep, 2017. 19(1): p. 203-217.

[4] Goltsev, Y., et al., Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. Cell, 2018. 174(4): p. 968-981 e15.

[5] Lin, J.R., M. Fallahi-Sichani, and P.K. Sorger, Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. Nat Commun, 2015. 6: p. 8390.

[6] Crosetto, N., M. Bienko, and A. van Oudenaarden, Spatially resolved transcriptomics and beyond. Nat Rev Genet, 2015. 16(1): p. 57-66.

[7] Thiery-Lavenant, G., A.I. Zavalin, and R.M. Caprioli, Targeted multiplex imaging mass spectrometry in transmission geometry for subcellular spatial resolution. J Am Soc Mass Spectrom, 2013. 24(4): p. 609-14.

[8] Giesen, C., et al., Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat Methods, 2014. 11(4): p. 417-22.

[9] Grabe, N., W. Roth, and S. Foersch, [Digital pathology in immuno-oncology-current opportunities and challenges : Overview of the analysis of immune cell infiltrates using whole slide imaging]. Pathologe, 2018. 39(6): p. 539-545.

[10] Qiu, P., et al., Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat Biotechnol, 2011. 29(10): p. 886-91.

[11] Amir el, A.D., et al., viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat Biotechnol, 2013. 31(6): p. 545-52.

[12] DiGiuseppe, J.A., et al., PhenoGraph and viSNE facilitate the identification of abnormal T-cell populations in routine clinical flow cytometric data. Cytometry B Clin Cytom, 2018. 94(5): p. 588-601.

[13] Samusik, N., et al., Automated mapping of phenotype space with single-cell data. Nat Methods, 2016. 13(6): p. 493-6.

[14] Brummelman, J., et al., Development, application and computational analysis of high-dimensional fluorescent antibody panels for single-cell flow cytometry. Nat Protoc, 2019. 14(7): p. 1946-1969.

[15] Young Hwan, C., et al., Multiplexed immunohistochemistry image analysis using sparse coding. Conf Proc IEEE Eng Med Biol Soc, 2017. 2017: p. 4046-4049.

[16] Dixon, A.R., et al., Recent developments in multiplexing techniques for immunohistochemistry. Expert Rev Mol Diagn, 2015. 15(9): p. 1171-86.

[17] Baker, G.M., et al., Development and validation of an immunohistochemistry assay to assess glucocorticoid receptor expression for clinical trials of mifepristone in breast cancer. Cancer Manag Res, 2015. 7: p. 361-8.

[18] Ramos-Vara, J.A., et al., Suggested guidelines for immunohistochemical techniques in veterinary diagnostic laboratories. J Vet Diagn Invest, 2008. 20(4): p. 393-413.

[19] Gagnon-Bartsch, J.A. and T.P. Speed, Using control genes to correct for unwanted variation in microarray data. Biostatistics, 2012. 13(3): p. 539-52.

[20] Mitri, Z.I., et al., Implementing a comprehensive translational oncology platform: from molecular testing to actionability. J Transl Med, 2018. 16(1): p. 358.

Supplementary Information

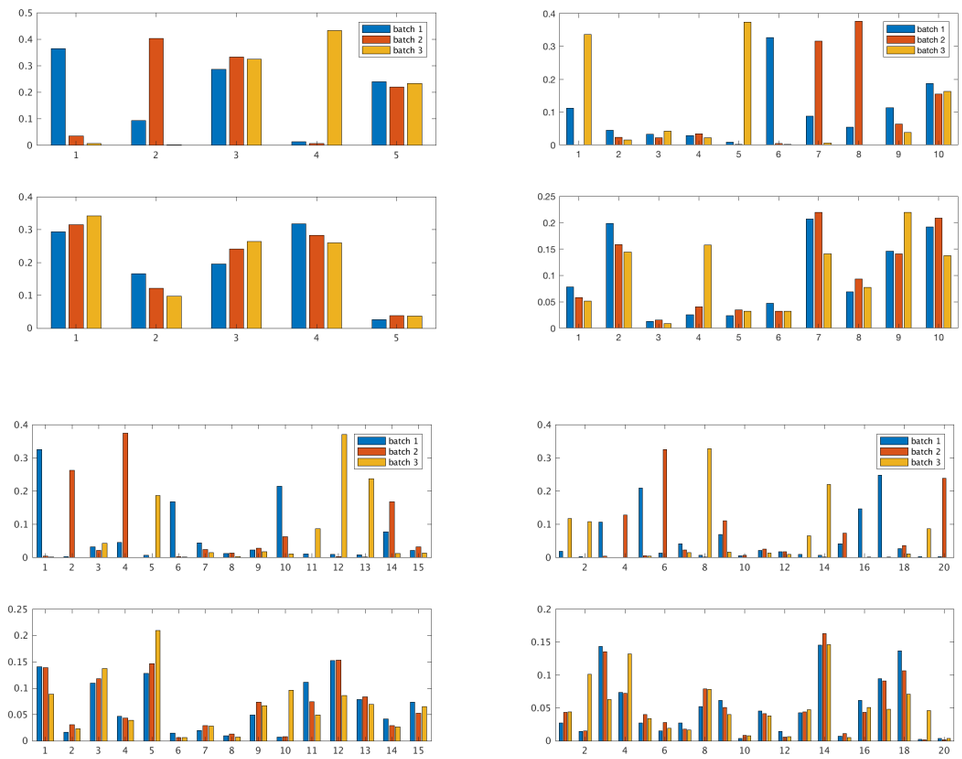

Figure S1: Example of cell composition comparison based on unsupervised clustering result between without normalization and with normalization with varying the number of cluster ($N = 5, 10, 15, 20$). For each figure, top row shows the result without normalization and bottom row shows the result with normalization where x-axis represents clustered group ID and y-axis represents their cell population. With normalization approaches, each cluster group shows uniform distribution from each batch.

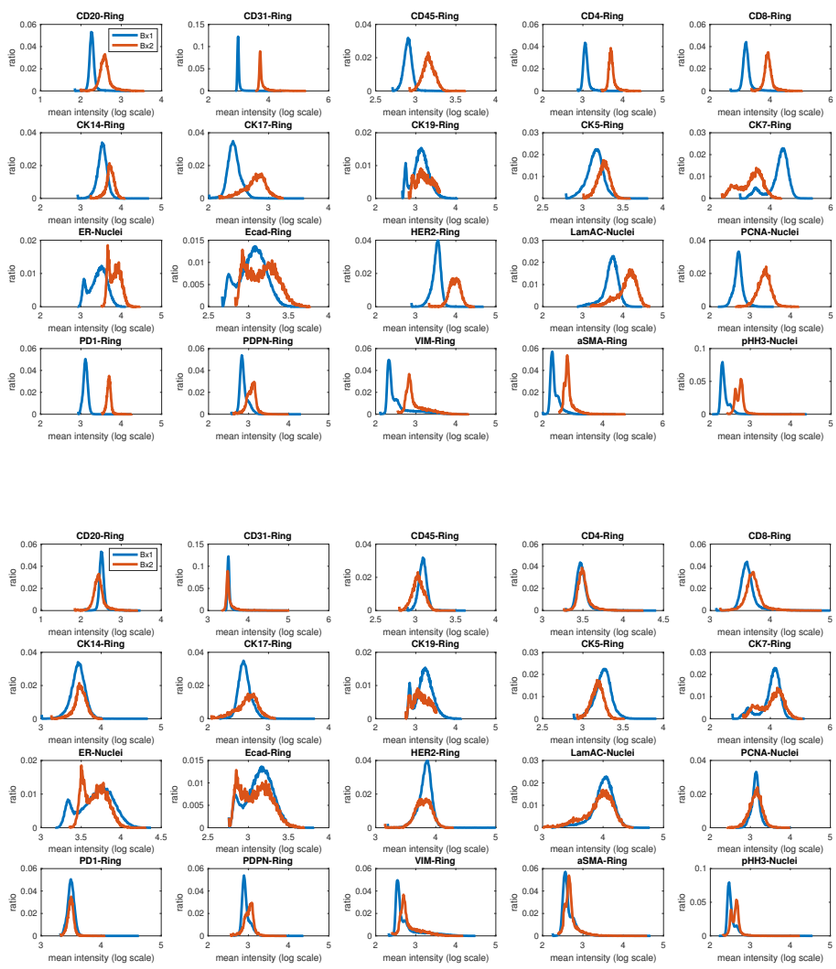

Figure S2: cmIF Intensity distribution across Bx 1 and Bx 2 where top figure shows before normalization and bottom figure shows after normalization based on inferred background signal.

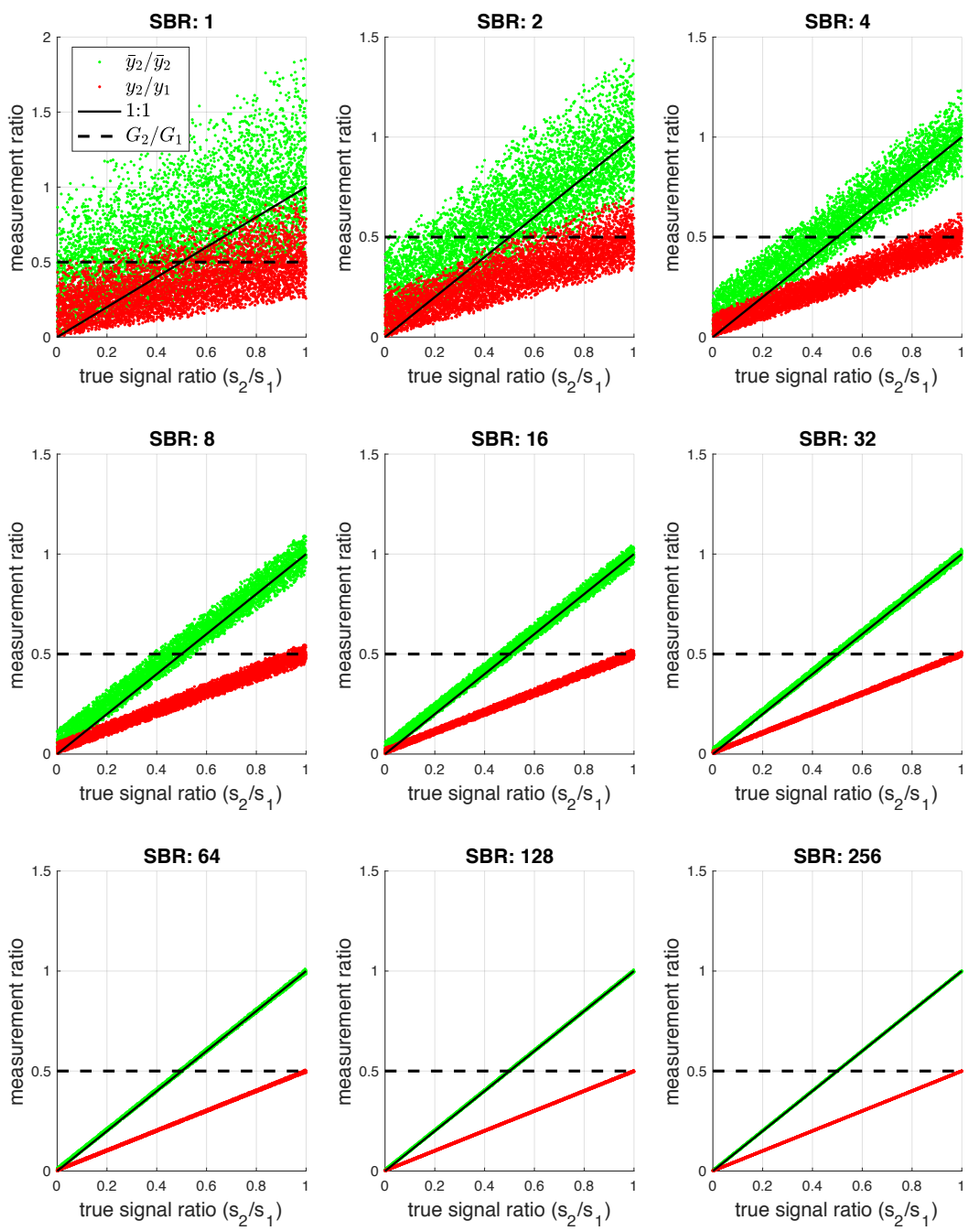

Figure S3: Comparison between true signal ratio and measurement signal ratio with and without normalization by changing signal to background ratio (SBR): as SBR increases, the normalized signal ratio is close to the true signal ratio. On the other hand, without normalization, measurement signal ratio is proportional to gain G_2/G_1 (0.5, in this simulation setting).

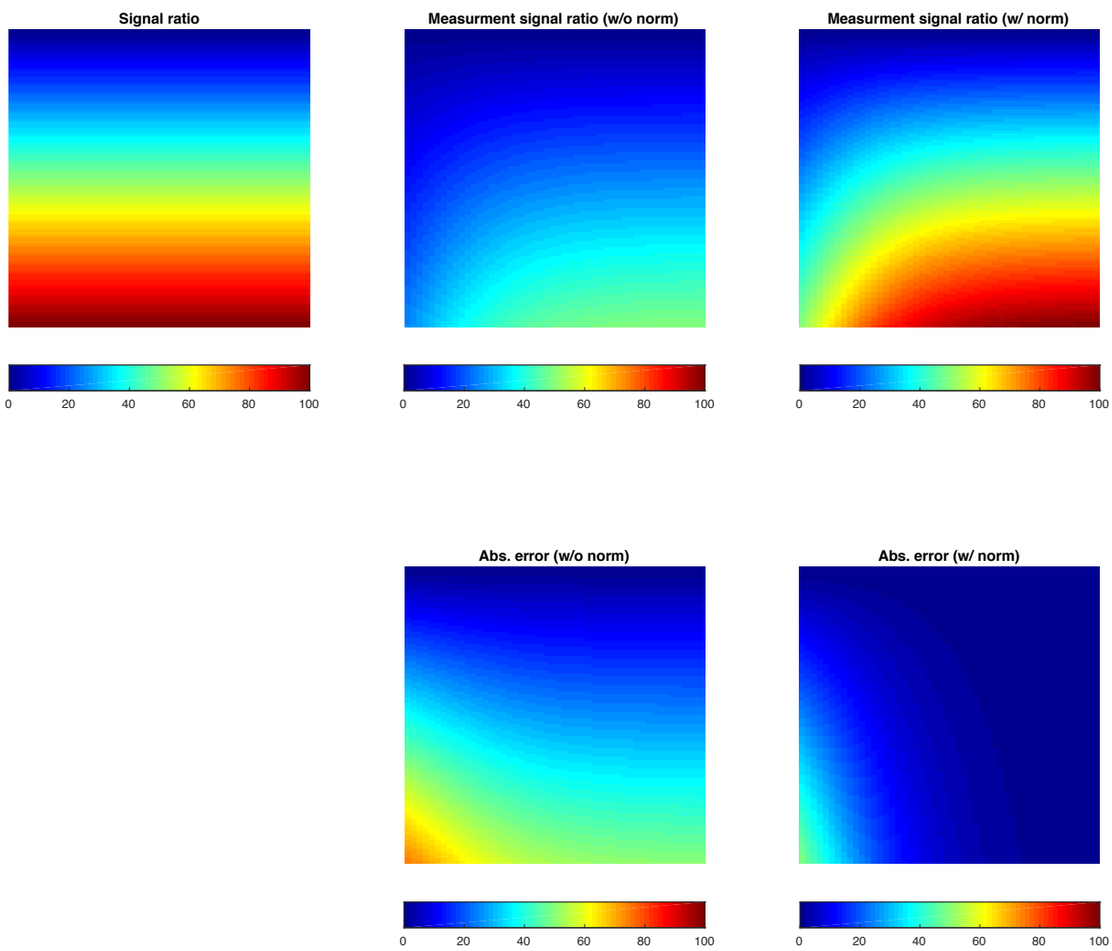

Figure S4: A heat map representation: (top left) true signal ratio, (top middle) measurement signal ratio [w/o norm], (top right) measure signal ratio [w/ norm], (bottom middle) error [w/o norm], (bottom right) error [w/ norm] where color represents signal ratio (low in blue and high in red). From left to right across columns, signal to background ratio (SBR) increases and thus, measurement signal ratio with normalization is close to the true signal ratio.

	Data-driven approach					Knowledge-based approach				
Markers	**1st**	**2nd**	**3rd**	**4th**	**5th**	**1st**	**2nd**	**3rd**	**4th**	**5th**
AR	CD31	HER2	CK7	CD45	LamB2	-	-	-	-	-
aSMA	CK19	HER2	CD68	PgR	H3K27	CK19	CK7	CD68		
CD20	CD31	HER2	CK7	CD68	PgR	CK19	CK7	CK14	CK5	
CD3	CK7	HER2	CD31	CK19	CD68	CK19	CK7	CK14	CK5	
CD4	HER2	CK19	CK7	CD31	H3K27	CK19	CK7	CK14	CK5	
CD44	CK19	CD68	Ecad	PgR	HER2	-	-	-	-	-
CD45	CK7	HER2	CD31	CK19	LamB2	CK19	CK7	CK14	CK5	
CD68	CK19	PgR	CK14	Ecad	H3K27	CK19	CK7	CD31		
CD8	CK7	HER2	CD31	CK19	LamB2	CK19	CK7	CK14	CK5	
CK14	CD68	Ki67	ColI	cPARP	GrNZB	CD31	CD68	Vim		
CK17	CD68	CK14	PgR	CK19	CD44	CD31	CD68	Vim		
CK19	CD68	CD44	Vim	CD31	CD4	CD68	CD4	CD31		
CK5	CD68	PgR	CK19	CD44	Ecad	CD31	CD68	Vim		
CK7	CD31	Vim	CD4	CD45	CD3	CD68	CD4	CD31		
CK8	CD31	CD68	Vim	CD4	HER2	CD68	CD4	CD31		
ColI	CK14	CD68	CK19	PgR	CK5	CK19	CK7	CK14	CK5	
ColIV	CD68	PgR	CK19	HER2	CD31	CK19	CK7	CK14	CK5	CD68
cPARP	CK14	CD68	GRNZB	PgR	Ki67	-	-	-	-	-
Ecad	CD68	CD44	CK14	Vim	PgR	CD68	CD4	CD31		
ER	CD31	HER2	CK7	CD68	Vim	CD68	CD4	CD31		
FoxP3	CD68	HER2	CK19	PgR	CD31	CK19	CK7	CK5		
GRNZB	CK14	Ki67	cPARP	ColI	CD68	CK19	CK7	CK5	CD31	
H3K27	CD68	HER2	CD44	CD31	PgR	-	-	-	-	-
H3K4	CD31	CD68	HER2	CK19	PgR	-	-	-	-	-
LamB2	HER2	CD31	CK7	CK19	CD68	-	-	-	-	-
LaminAC	CD68	PgR	CK19	CD44	CK14	-	-	-	-	-
PCNA	CK7	CD45	CD31	HER2	CD3	-	-	-	-	-
PD1	CD68	HER2	CD31	PgR	CK19	CK19	CK7	CK14	CK5	CD31
PDPNP	CD68	CK19	HER2	PgR	CK14	CK19	CK7	CK14	CK5	CD68
pERK	CD68	HER2	CD31	PgR	CK14	-	-	-	-	-
PgR	CD68	CK14	HER2	CD44	CK19	CD68	CD4	CD31		
pHH3	GRNZB	CK14	CD68	ColI	Ki67	-	-	-	-	-
pRB	CK14	CD68	PgR	CK19	GRNZB	-	-	-	-	-
pS6	CD68	CK14	PgR	CK19	CD44	-	-	-	-	-
Vim	HER2	CK19	CK7	Ecad	H3K27	CK19	CK7	CD68		
HER2	CD31	Vim	CD4	PgR	CD44	CD68	CD4	CD31		
Ki67	CK14	CD68	GRNZB	CD44	PgR	-	-	-	-	-

Table S1: A list of mutually exclusive marker pairs based on (left) data-driven and (right) biological known information. Red color on the left panel indicates that matched pairs exist on the right panel. Green color on the right panel indicates that matched pairs exist on the left panel.