PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Graph of graphs analysis for multiplexed data with application to imaging mass cytometry

Ya-Wei Eileen Lin[1]*, Tal Shnitzer[1], Ronen Talmon[1], Franz Villarroel-Espindola[2], Shruti Desai[2], Kurt Schalper[2,3], Yuval Kluger[2,4,5]

1 Viterbi Faculty of Electrical Engineering, Technion - Israel Institute of Technology, Haifa, Israel, **2** Department of Pathology, School of Medicine, Yale University, New Haven, Connecticut, United States of America, **3** Department of Medicine, Yale School of Medicine and Yale Cancer Center, New Haven, Connecticut, United States of America, **4** Computational Biology and Bioinformatics Program, Yale University, New Haven, Connecticut, United States of America, **5** Program of Applied Mathematics, Yale University, New Haven, Connecticut, United States of America

* lin.ya-wei@campus.technion.ac.il

Citation: Lin Y-WE, Shnitzer T, Talmon R, Villarroel-Espindola F, Desai S, Schalper K, et al. (2021) Graph of graphs analysis for multiplexed data with application to imaging mass cytometry. PLoS Comput Biol 17(3): e1008741. https://doi.org/10.1371/journal.pcbi.1008741

Editor: Min Xu, Carnegie Mellon University, UNITED STATES

Received: August 10, 2020

Accepted: January 26, 2021

Published: March 29, 2021

Data Availability Statement: Data are available from the Yale University Human Investigation Committee (protocols #9505008219 and #1608018220); or local institutional protocols which approved the patient consent forms or, in some cases waiver of consent when retrospectively collected archive tissue was used in a de-identified manner; (contact via <kschalperlab@yale.edu>) for researchers who meet the criteria for access to confidential data.

Abstract

Imaging Mass Cytometry (IMC) combines laser ablation and mass spectrometry to quantitate metal-conjugated primary antibodies incubated in intact tumor tissue slides. This strategy allows spatially-resolved multiplexing of dozens of simultaneous protein targets with $1\mu m$ resolution. Each slide is a spatial assay consisting of high-dimensional multivariate observations (m-dimensional feature space) collected at different spatial positions and capturing data from a single biological sample or even representative spots from multiple samples when using tissue microarrays. Often, each of these spatial assays could be characterized by several regions of interest (ROIs). To extract meaningful information from the multi-dimensional observations recorded at different ROIs across different assays, we propose to analyze such datasets using a two-step graph-based approach. We first construct for each ROI a graph representing the interactions between the m covariates and compute an m dimensional vector characterizing the steady state distribution among features. We then use all these m-dimensional vectors to construct a graph between the ROIs from all assays. This second graph is subjected to a nonlinear dimension reduction analysis, retrieving the intrinsic geometric representation of the ROIs. Such a representation provides the foundation for efficient and accurate organization of the different ROIs that correlates with their phenotypes. Theoretically, we show that when the ROIs have a particular bi-modal distribution, the new representation gives rise to a better distinction between the two modalities compared to the maximum a posteriori (MAP) estimator. We applied our method to predict the sensitivity to PD-1 axis blockers treatment of lung cancer subjects based on IMC data, achieving 97.3% average accuracy on two IMC datasets. This serves as empirical evidence that the graph of graphs approach enables us to integrate multiple ROIs and the intra-relationships between the features at each ROI, giving rise to an informative representation that is strongly associated with the phenotypic state of the entire image.

Funding: The work of Y.-W E. L., T. S. and R. T. was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 802735-ERC-DIFFOP (https://erc.europa.eu). The work of F. V.-E., S. D. and K. S. was supported by the NIH grant R37CA245154 (https://www.nih.gov/), Yale SPORE in Lung Cancer P50CA196530 (https://www.yalecancercenter.org/research/excellence/lung/), Stand Up To Cancer – American Cancer Society Lung Cancer Dream Team Translational Research Grants SU2C-AACR-DT1715 and SU2C-AACR-DT22-17 (https://standuptocancer.org/research/research-portfolio/dream-teams/), Department of Defense-Lung Cancer Research Program Career Development Award W81XWH-16-1-0160 (https://cdmrp.army.mil/lcrp), Yale Cancer Center Support Grant P30CA016359 (https://www.yalecancercenter.org/research/people/ccsg/), sponsored research by Navigate Biopharma (https://www.navigatebp.com) and AstraZeneca (https://www.astrazeneca.com). The work of Y. K. was supported by NIH grant R01RGM131642, UM1DA051410 and P50CA121974 (https://www.nih.gov). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

We propose a two-step graph-based analyses for high-dimensional multiplexed datasets characterizing ROIs and their inter-relationships. The first step consists of extracting the steady state distribution of the random walk on the graph, which captures the mutual relations between the covariates of each ROI. The second step employs a nonlinear dimensionality reduction on the steady state distributions to construct a map that unravels the intrinsic geometric structure of the ROIs. We theoretically show that when the ROIs have a two-class structure, our method accentuates the distinction between the classes. Particularly, in a setting with Gaussian distribution it outperforms the MAP estimator, implying that the mutual relations between the covariates within the ROIs and spatial coordinates are well captured by the steady state distributions. We apply our method to imaging mass cytometry (IMC). Our analysis provides a representation that facilitates prediction of the sensitivity to PD-1 axis blockers treatment of lung cancer subjects. Particularly, our approach achieves state of the art results with average accuracy of 97.3% on two IMC datasets.

This is a *PLOS Computational Biology* Methods paper.

Introduction

Consider multi-feature observations collected at different spatial positions. Data structure of this type requires analysts to address two immediate natural questions. First is how to characterize the associations between the different features in each position. Second is how to organize the observations from different spatial positions into an informative representation.

We approach these two questions from the standpoint of manifold learning, which is a class of nonlinear dimensionality reduction techniques for high-dimensional data [1–4]. The common assumption in manifold learning is that the multi-feature observations lie on a hidden lower-dimensional manifold. Such an assumption facilitates the incorporation of geometric concepts such as metrics, geodesic distances, and embedding, into useful data analysis techniques. In order to learn a (continuous) manifold from discrete data samples, commonly-used manifold learning methods rely on the construction of a graph. Typically, the data samples form the graph nodes and the edges of the graph are determined according to some similarity notion that is usually application-specific.

In our work, we adhere to manifold learning techniques and propose a method consisting of two-step graph analysis. At the first stage, we build a graph for each spatial position, where the graph nodes are the multi-feature observations. The motivation to build such a graph rather than using the observations directly stems from an assumption that the information about the sample at each spatial position is better expressed by the mutual-relations between the features. Then, we define a random walk on this graph and build a characteristic vector of the respective spatial position by computing the steady state distribution (SSD) of the random walk. For analysis purposes, we define a new notion of heterogeneity, representing a statistical diversity of the multiple features, and show that the SSD characterizes each spatial position in

terms of this heterogeneity. In addition, using this notion of heterogeneity, when the density of the observations at the spatial positions is bi-modal, we show that these SSDs can lead to an accurate identification of the two modes, outperforming the maximum a-posteriori (MAP) estimator [5] in a statistical setting with Gaussian distributions.

At the second stage, we build a graph whose nodes are the new characteristic vectors (SSDs) of all the spatial positions. We apply diffusion maps [4] to this second graph and obtain a low dimensional representation of the spatial positions. The dimension of the computed representation is determined by a nonlinear variant of the Jackstraw algorithm [6].

Broadly, the proposed algorithm could be viewed as building a *graph of graphs*. From a manifold learning standpoint, this two-step procedure could be viewed as inferring a *manifold of manifolds*. Namely, at the first stage, we recover the local manifolds that underlie the multiple features at each spatial location, and then, at the second stage, we recover the global manifold between the spatial positions, formed by the collection of all local manifolds. This standpoint is related to a large body of recent work involving the discovery and analysis of multi-manifold structures, e.g., alternating diffusion [7–10], multi-view diffusion maps [11], joint Laplacian diagonalization [12], to name just a few. Therefore, the proposed method can be viewed as a follow up work along this line of research.

We apply our method to imaging mass cytometry (IMC) [13–15]. IMC is a new technique for multiplexed simultaneous imaging of proteins and protein modifications at subcellular resolution, ideally suited to uncover molecular and structural alterations of diseased tissues such as in cancer. IMC analysis can also be used to study the composition of non-diseases tissue samples such as histology studies or molecular profiles. The acquired intensities of the protein expression levels are viewed as markers, providing important biological information on the tissues of interest. This acquisition procedure gives rise to multi-feature observations at different spatial positions, where the multiple features are the markers and a selected subset of the spatial positions are ROIs within pathology slides.

Our experimental study focuses on one of the important tasks of IMC data analysis: associating the response status of a patient to a therapeutic intervention with a high-dimensional spatial IMC sample from the relevant patients' tissues. Here, we propose to recast this problem as a binary hypothesis testing problem. We assume that all the ROIs of each patient can be labeled by the patient's response or non-response status. The collection of all ROIs from the patients' cohort induces a bi-modal density of expression signatures. Then, given the protein expression levels within the ROIs of a certain tissue type, we ask whether the subject was responsive to treatment. We showcase the performance of the proposed method on two IMC cohorts consisting of samples taken from lung cancer subjects. We achieve a average 97.3% prediction accuracy of response to treatment (PD-1 axis blockers) in an unsupervised manner. This result outperforms competing methods, specifically, the results obtained by (i) diffusion maps (DM) [4] directly applied to the multi-feature observations, (ii) the heat kernel signature (HKS) [16], and (iii) the wave kernel signature (WKS) [17].

Results

Ethics statement

The study was approved by the Yale University Human Investigation Committee protocols #9505008219 and #1608018220; or local institutional protocols which approved the patient consent forms or, in some cases waiver of consent when retrospectively collected archive tissue was used in a de-identified manner.

Overview

We start by presenting the problem setting. Consider n data points $\{x_i\}_{i=1}^n$ from a hidden manifold $\mathcal{M}$ embedded in a high-dimensional Euclidean space $\mathbb{R}^k$. Assume we do not have direct access to these data points; instead, these data points are measured through m observation functions $f_j : \mathcal{M} \to \mathbb{R}^d$, where $j = 1, \ldots, m$ is the index of the observation function. Given n multi-feature observations $f_j(x_i)$ of the data point x_i for $i = 1, \ldots, n$, each consisting of m features $j = 1, \ldots, m$, our goal is to recover the data points x_i on the hidden manifold $\mathcal{M}$.

In the context of IMC, the data points represent the treatment outcome based on n spatial positions located at ROIs within pathology slides of tissues from several patients. At each spatial position $i = 1, \ldots, n$, the observations $f_j(x_i)$ for $j = 1, \ldots, m$ are the expression levels of m markers. Each observation $f_j(x_i) \in \mathbb{R}^d$ is a patch of d pixels of the expression level image of marker j at position i.

To simplify the presentation of our approach, we begin with an illustrative localization problem, which is simpler than the IMC problem. Suppose we have a surface $\mathcal{M}$ and objects located at x_i on $\mathcal{M}$. The locations x_i are hidden, but measured through m sensors, such that for each location x_i we have m multi-feature observations $\{f_j(x_i)\}_{j=1}^m$. That is $f_j(x_i)$ is a d-dimensional observation of sensor j when the object is at x_i. The goal is to recover the locations x_i on the surface $\mathcal{M}$ given $\{f_j(x_i)\}_{j=1}^m$.

Our approach consists of two stages. At the first stage, we construct a graph for each data point x_i in order to capture associations between its m multi-feature observations $\{f_j(x_i)\}_{j=1}^m$. Capturing such mutual-relationships is natural in the context of localization problems, e.g., the triangulation property [18] in which the relative locations of the sensors are exploited. In addition, these mutual-relations are typically more robust to noise in comparison with the nominal values of the multi-feature observations, $f_j(x_i)$, themselves. Concretely, consider the m observations $\{f_j(x_i)\}_{j=1}^m$ associated with the data point x_i. Each observation $f_j(x_i)$ forms a single node in the graph, hereby denoted as node j, giving rise to a graph with a total of m nodes. The graph we consider is the complete graph, where the weights of the edges are determined based on the Euclidean distance between the corresponding observations: the weight of the edge connecting nodes j and k is proportional to $\exp\{-\|f_j(x_i) - f_k(x_i)\|^2\}$. Then, we compute the SSD of a random walk defined on this graph at each location. SSD has a vector form that embodies the multi-feature inter-relationships of the data point x_i.

At the next step, we define a second graph based on the SSDs, characterizing the points $\{x_i\}_{i=1}^n$. Concretely, each data point x_i is represented by a node, and the pairwise distances between the SSDs form the adjacency matrix of the graph. Then, we apply a particular manifold learning technique, diffusion maps [4], to this graph. This application facilitates the recovery of the hidden manifold $\mathcal{M}$ in the sense that an embedding of the points x_i is learned, such that the distances between the embedded points respect a notion of an intrinsic distance (the diffusion distance [4]) on $\mathcal{M}$. The application of diffusion maps to the second graph in the context of the localization problem gives rise to an embedding that serves as an accurate representation of the hidden locations of the data point. In Localization toy problem, we demonstrate the proposed method on several simulations of localization toy problems.

We remark that the IMC problem and the localization problem share many aspects. For example, in both problems, the multi-feature observations are noisy and the mutual-relationship between them carry important information. Yet, there is a particular aspect that makes the IMC problem more challenging; while the points on the hidden manifold in a localization problem are homogeneous because they all represent location coordinates, the points in the

IMC problem could be significantly different due to the large variability in the tissue structure. Importantly, the proposed method accommodates the joint processing of such different points through their representation by the SSD.

Proposed method

The first step of the proposed method is to construct an undirected weighted graph $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i, \boldsymbol{W}_i)$ for each data point $x_i \in \mathcal{M} \subset \mathbb{R}^k$, where the vertex set is $\mathcal{V}_i = \{f_1(x_i), f_2(x_i), \ldots, f_m(x_i)\}$, the edge set is $\mathcal{E}_i \subseteq \mathcal{V}_i \times \mathcal{V}_i$, and the graph weights matrix $\boldsymbol{W}_i \in \mathbb{R}^{m\times m}$ is given by the Gaussian kernel

$$W_i(k,l) = \exp\left(-\frac{||f_k(x_i) - f_l(x_i)||_2^2}{2\epsilon}\right), \tag{1}$$

where $k, l \in \{1, \ldots, m\}$ and $\epsilon > 0$ is a scale parameter. Note that since $\boldsymbol{W}_i$ is symmetric, $\boldsymbol{W}_i$ is diagonalizable. That is, there is a set of real eigenvalues $\{\lambda_j\}_{j=1}^m$ with a corresponding orthonormal basis of eigenvectors $\{\boldsymbol{v}_j\}_{j=1}^m$ such that

$$W_i(k,l) = \sum_{j=1}^m \lambda_j \boldsymbol{v}_j(k)\boldsymbol{v}_j(l). \tag{2}$$

Next, we define a random walk on the graph $\mathcal{G}_i$. Let $\boldsymbol{P}_i \in \mathbb{R}^{m\times m}$ be a row stochastic matrix given by

$$\boldsymbol{P}_i = \boldsymbol{D}_i^{-1}\boldsymbol{W}_i, \tag{3}$$

where $\boldsymbol{D}_i$ is a diagonal matrix whose diagonal elements are given by $D_i(k,k) = \sum_{l=1}^m W_i(k,l)$. The value of $P_i(k, l)$ can be interpreted as a transition probability from a vertex $f_k(x_i)$ to a vertex $f_l(x_i)$ in one step of a random walk on the graph $\mathcal{G}_i$.

The transition probability matrix $\boldsymbol{P}_i$ is self-adjoint and compact, and therefore, the spectral decomposition of $\boldsymbol{P}_i^t$ for $t > 0$ is given by

$$P_i^t(k,l) = \sum_{j=1}^m \lambda_j^t \boldsymbol{\psi}_j(k)\boldsymbol{\phi}_j(l), \tag{4}$$

where $\{\boldsymbol{\psi}_j, \boldsymbol{\phi}_j\}_{j=1}^m$ are the right- and left-eigenvectors with the corresponding eigenvalues $\{\lambda_j\}_{j=1}^m$. By the construction of $\boldsymbol{P}_i$ from $\boldsymbol{W}_i$, the relations between their respective eigenvectors are given by

$$\boldsymbol{\psi}_j(k) = \frac{\boldsymbol{v}_j(k)}{\sqrt{D_i(k,k)}} \tag{5}$$

and

$$\boldsymbol{\phi}_j(k) = \boldsymbol{v}_j(k)\sqrt{D_i(k,k)}. \tag{6}$$

Interestingly, in the special case where $t = 1$, the probability of the node $f_k(x_i)$ to stay in place is given by

$$P_i(k,k) = \sum_{j=1}^{m} \lambda_j \boldsymbol{v}_j^2(k) = \frac{1}{D_i(k,k)}. \tag{7}$$

Note that $\boldsymbol{\phi}_1 \in \mathbb{R}^m$ is the left eigenvector of $\boldsymbol{P}_i$ corresponding to $\lambda_1 = 1$, satisfying:

$$\boldsymbol{P}_i^\top \boldsymbol{\phi}_1 = \boldsymbol{\phi}_1, \tag{8}$$

where $\boldsymbol{P}_i^\top$ is the transpose of the matrix $\boldsymbol{P}_i$. Consider an arbitrary distribution vector $\boldsymbol{\pi}_0 \in \mathbb{R}^m$ and observe the expansion:

$$\boldsymbol{\pi}_0^\top \boldsymbol{P}_i^t = \sum_{j=1}^{m} \lambda_j^t \langle \boldsymbol{\pi}_0, \boldsymbol{\psi}_j \rangle \boldsymbol{\phi}_j^\top, \tag{9}$$

where $\langle \boldsymbol{\pi}_0, \boldsymbol{\psi}_j \rangle = \boldsymbol{\pi}_0^\top \boldsymbol{\psi}_j$ is the standard Euclidean product. In the limit $t \to \infty$, $\lambda_j^t \to 0$ for $j > 1$ and therefore

$$\boldsymbol{\pi}_0^\top \boldsymbol{P}_i^t \longrightarrow \lambda_1^t \langle \boldsymbol{\pi}_0, \boldsymbol{\psi}_1 \rangle \boldsymbol{\phi}_1^\top = \boldsymbol{\phi}_1^\top \triangleq \boldsymbol{\pi}_i^\top, \tag{10}$$

where $\boldsymbol{\pi}_i \in \mathbb{R}^m$ since $\lambda_1 = 1$, $\boldsymbol{\psi}_1 = \mathbf{1}$, and $\sum_{j=1}^{m} \boldsymbol{\pi}_0(j) = 1$. Since the random walk defined by $\boldsymbol{P}_i$ is irreducible, finite, and aperiodic [19], the stationary distribution $\boldsymbol{\pi}_i$ is a *unique* stationary distribution. The convergence in Eq (10) and the uniqueness allow us to treat the stationary distribution in this case as the steady state distribution (SSD). Note that the SSD $\boldsymbol{\pi}_i \propto \boldsymbol{D}_i \mathbf{1}$, where $\mathbf{1} \in \mathbb{R}^m$ is an all-ones vector. In other words, $\boldsymbol{\pi}_i$ can be viewed as a normalized degrees vector of the graph $\mathcal{G}_i$.

We will use $\boldsymbol{\pi}_i$ as a new characteristic vector, or a signature, of x_i, and consequently, the induced pairwise distances $||\boldsymbol{\pi}_i - \boldsymbol{\pi}_{i'}||$, where $i, i' \in 1, \ldots, n$, will be used as the desired distances between the respective graphs $\mathcal{G}_i$ and $\mathcal{G}_{i'}$ for recovering $\mathcal{M}$. At first glance, using $\boldsymbol{\pi}_i$ may seem too simplistic. Instead, one could use the broad spectral information. Consequently, define the Diffusion Kernel Signature (DKS) by

$$x_i \mapsto [\mathrm{DKS}_t(f_1(x_i)), \mathrm{DKS}_t(f_2(x_i)), \ldots, \mathrm{DKS}_t(f_m(x_i))], \tag{11}$$

where

$$\mathrm{DKS}_t(f_k(x_i)) = \sum_{j=1}^{m} \lambda_j^t \boldsymbol{\psi}_j^2(k). \tag{12}$$

Since λ_j is in descending order and in [0, 1], the weight they assign to the eigenvectors in (12) becomes smaller as t increases. As a result, the DKS can be viewed as a low-pass filter, which controls the spectral bandwidth. In addition, the DKS can be recast in terms of the diffusion distance, a notion of distance induced by diffusion maps [4] that was shown useful in a broad range of applications, e.g., [20–22]. For more details see Diffusion maps. Specifically,

when $t = 1$, we can show that

$$\begin{aligned} \mathrm{DKS}_{t=1}(f_k(x_i)) &= \sum_{j=1}^{m} \lambda_j \boldsymbol{\psi}_j^2(k) \\ &= \sum_{j=1}^{m} \lambda_j \left(\frac{\boldsymbol{v}_j(k)}{\sqrt{D_i(k,k)}} \right)^2 \\ &= \frac{1}{D_i(k,k)} \sum_{j=1}^{m} \lambda_j \boldsymbol{v}_j^2(k) \\ &= \frac{1}{\boldsymbol{\pi}_i^2(k)}, \end{aligned} \tag{13}$$

indicating that the SSD is a special case of DKS.

We note that DKS has already appeared in previous work in the context of spectral distances in [23, 24], where it was shown that it describes the underlying geometry of $\mathcal{M}$. We show in the following that the seemingly simple SSD, despite the lack of broad spectral information as in the DKS, still carries substantial information.

Note that ϵ is a scale parameter of the Gaussian kernel, where it can be used to infer locality. If ϵ is set to a small value, then $\boldsymbol{\pi}_i$ captures local properties. Conversely, if ϵ is large, then $\boldsymbol{\pi}_i$ represents the global structure. As a result, a multiscale signature can be formed, consisting of multiple SSDs π_i computed with different values of ϵ.

The final stage of our method is building a low-dimensional representation of all the data points $\{x_i\}_{i=1}^n$. To this end, we apply diffusion maps to the corresponding characteristic vectors (signatures) $\{\boldsymbol{\pi}_i\}_{i=1}^n$ as follows. First, we build a global graph $\mathbf{G}^{(2)}$ whose nodes are $\boldsymbol{\pi}_i$ and edge weights are determined by a Gaussian kernel based on the l_1 *distance* between $\boldsymbol{\pi}_i$.

That is, the global graph weights matrix $\mathbf{W}^{(2)}$ is defined by

$$\mathrm{W}^{(2)}(k,l) = \exp\left(-\frac{||\boldsymbol{\pi}_k - \boldsymbol{\pi}_l||_1^2}{2\epsilon'} \right), \tag{14}$$

where $k, l \in [1, n]$, $||\cdot||_1$ is the l_1 distance and $\epsilon' > 0$ is a scale parameter. We remark that common practice is to use the l_2 distance in the Gaussian kernel. The reason we use the l_1 distance is described in Binary hypothesis testing, which indeed leads to better empirical performance reported in Imaging mass cytometry (IMC).

Second, we construct a random walk, denoting its transition probability matrix by $\mathbf{P}^{(2)}$. Third, we apply the eigendecomposition to $\mathbf{P}^{(2)}$. Fourth, we set the dimension of the new representation according to the variant of the Jackstraw method [6], as shown in Determining the dimension of data.

The entire method is summarized in Box 1 and a block diagram is illustrated in Fig 1.

Theoretical analysis

We propose a statistical model that allows for a tractable analysis, showing the advantages of the SSD signature. Consider a data point $x_i \in \mathcal{M}$ and denote the set of the observations by $S = \{f_j(x_i)\}_{j=1}^m$. Assume the j-th observation $f_j(x_i) \in \mathbb{R}^d$ is a realization of a d-dimensional random vector V_j^i following a multivariate normal distribution given by

$$V_j^i \sim \mathcal{N}(\mu_j^i \mathbf{1}_d, \sigma_j^i \boldsymbol{I}_d). \tag{15}$$

Box 1. A summary of the proposed method

Input: A set of multi-feature observations $\{f_j(x_i)\}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

Output: l-dimensional representation $\boldsymbol{\Psi}(x_i) \in \mathbb{R}^l$ for $i = 1, \ldots, n$.

1. For each x_i:
   a Construct a local graph $\mathcal{G}_i$ with vertex set
   $$\mathcal{V}_i = \{f_1(x_i), f_2(x_i), \ldots, f_m(x_i)\},$$
   edge set $\mathcal{E}_i \subseteq \mathcal{V}_i \times \mathcal{V}_i$, and edge weights matrix $\boldsymbol{W}_i \in \mathbb{R}^{m \times m}$ given in Eq (1).
   b Build a random walk on the local graph $\mathcal{G}_i$ with transition probability matrix $\boldsymbol{P}_i$ defined in (3).
   c Compute the SSD $\boldsymbol{\pi}_i \in \mathbb{R}^m$ of $\boldsymbol{P}_i$.
2. Construct a global graph $\mathbf{G}^{(2)}$ with vertex set $\{\boldsymbol{\pi}_i\}_{i=1}^n \in \mathbb{R}^{m \times n}$ and the graph weights matrix $\mathbf{P}^{(2)}$ given in Eq (14).
3. Build a random walk on $\mathbf{G}^{(2)}$ with transition probability matrix $\mathbf{P}^{(2)}$.
4. Apply eigenvalue decomposition to $\mathbf{P}^{(2)}$ and obtain the eigenvectors $\{\varphi_k\}_{k=1}^n$.
5. Determine the number of dimensions l as described in Determining the dimension of data.
6. Build the mapping: $x_i \mapsto (\varphi_1(x_i), \varphi_2(x_i), \ldots, \varphi_l(x_i))^\top \triangleq \boldsymbol{\Psi}(x_i)$ for $i = 1, \ldots, n$.

By collecting the m observations $\{f_j(x_i)\}_{j=1}^m$ and denoting the covariance matrix between the k and l observation functions by $\boldsymbol{\Sigma}_{k,l}^i$, we obtain an md-dimensional vector that can be viewed as a realization of the random vector $V^i = (V_1^i, \ldots, V_m^i)$ with the corresponding multivariate Gaussian distribution given by

$$\mathcal{N}(\boldsymbol{\mu}^i, \boldsymbol{\Sigma}^i), \tag{16}$$

where

$$\boldsymbol{\mu}^i = \begin{pmatrix} \left.\begin{matrix} \mu_1^i \\ \vdots \\ \mu_1^i \end{matrix}\right\} d \text{ elements} \\ \left.\begin{matrix} \mu_2^i \\ \vdots \\ \mu_2^i \end{matrix}\right\} d \text{ elements} \\ \vdots \\ \left.\begin{matrix} \mu_m^i \\ \vdots \\ \mu_m^i \end{matrix}\right\} d \text{ elements} \end{pmatrix} \in \mathbb{R}^{md}, \tag{17}$$

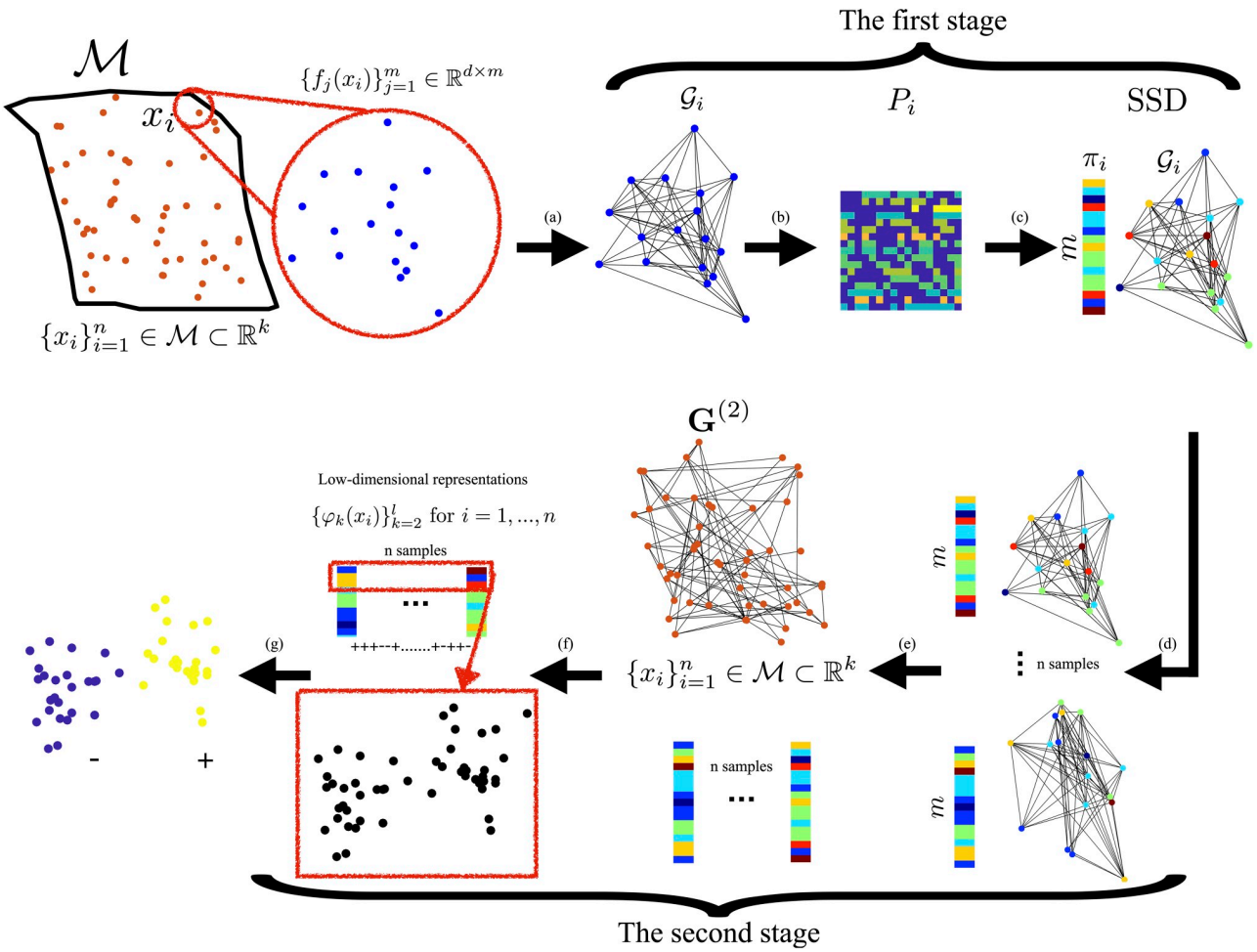

Fig 1. Illustrative diagram of the proposed method. (a) For each data point x_i, we build a local graph $\mathcal{G}_i$ based on its multi-feature observations $\{f_j(x_i)\}_{j=1}^m$. (b) We construct a random walk with transition probabilities matrix $\boldsymbol{P}_i$ on $\mathcal{G}_i$. (c) We extract the SSD signature $\boldsymbol{\pi}_i$ from $\boldsymbol{P}_i$. (e) We collect the SSDs of $\{x_i\}_{i=1}^n$ into an SSD representation matrix. (g) Subsequently, the matrix is subjected to a nonlinear dimensionality reduction using diffusion maps by the construction of the global graph $\mathbf{G}^{(2)}$ and the corresponding random walk with $\mathbf{P}^{(2)}$. (f) Via eigenvalue decomposition, we obtain a low-dimensional representation $\boldsymbol{\Psi}(x_i)$ for $i = 1, \ldots, n$, which is used in the subsequent tasks (g).

https://doi.org/10.1371/journal.pcbi.1008741.g001

and

$$\boldsymbol{\Sigma}^i = \begin{bmatrix} \sigma_1^i \boldsymbol{I}_d & \boldsymbol{\Sigma}_{1,2}^i & \dots & \dots & \dots & \dots & \boldsymbol{\Sigma}_{1,m}^i \\ \boldsymbol{\Sigma}_{2,1}^i & \sigma_2^i \boldsymbol{I}_d & \boldsymbol{\Sigma}_{2,3}^i & \dots & \dots & \dots & \boldsymbol{\Sigma}_{2,m}^i \\ \boldsymbol{\Sigma}_{3,1}^i & \boldsymbol{\Sigma}_{3,2}^i & \sigma_3^i \boldsymbol{I}_d & \boldsymbol{\Sigma}_{3,4}^i & \dots & \dots & \boldsymbol{\Sigma}_{3,m}^i \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{m,1}^i & \dots & \dots & \dots & \dots & \boldsymbol{\Sigma}_{m,m-1}^i & \sigma_m^i \boldsymbol{I}_d, \end{bmatrix} \in \mathbb{R}^{md\times md} \tag{18}$$

such that

$$\boldsymbol{\Sigma}_{\boldsymbol{k,l}}^{\boldsymbol{i}} = \sigma_{k,l}^i \boldsymbol{I}_d, \tag{19}$$

and $\sigma_{k,l}^i$ is the covariance between the k-th and l-th random vectors. In words, $\boldsymbol{\mu}^i$ in Eq (17) is a vector of md elements, consisting of the concatenation of m vectors. We index each of these

vector by $j = 1, \ldots, m$. All the entries of the j-th vector are equal and are set to the mean observation of the j-th sensor (marker) at the i-th point (μ_j^i). In Eq (18), $\boldsymbol{\Sigma}^i$ is a matrix of size of $md \times md$, which is the analogous concatenation of the covariance matrices of the observations. Specifically, the diagonal blocks are the diagonal matrices $\sigma_j^i \boldsymbol{I}_d$, whose diagonal elements are the variances of the j-th observation, and the off-diagonal blocks are the diagonal matrices $\sigma_{k,l}^i \boldsymbol{I}_d$, whose diagonal elements are the covariance between sensor (marker) k and sensor (marker) l.

Definition 1 (empirical mean) *Given a set* Γ *and some real function on the set* $\boldsymbol{q} \in \mathbb{R}^{|\Gamma|}$, *and a subset* $\Omega \subset \Gamma$, *the empirical mean of* $\boldsymbol{q}$ *in* Ω *is defined by*

$$\langle \boldsymbol{q} \rangle_\Omega = \frac{1}{|\Omega|} \sum_{j \in \Omega} \boldsymbol{q}(j). \tag{20}$$

Definition 2 (heterogeneity) *Define the heterogeneity of a data point* x_i *by*

$$\boldsymbol{h}_i = \boldsymbol{M}_i^{-1}(\boldsymbol{\Theta}^i - \langle \boldsymbol{\mu}^i \rangle_S \mathbf{1}) \in \mathbb{R}^m, \tag{21}$$

where $\langle \boldsymbol{\mu}^i \rangle_S$ *is given by Definition 1,* $\boldsymbol{M}_i$ *is an* $m \times m$ *diagonal matrix, whose diagonal elements are* μ_j^i, *and*

$$\boldsymbol{\Theta}^i = \begin{bmatrix} \mu_1^i \\ \mu_2^i \\ \vdots \\ \mu_m^i \end{bmatrix} \in \mathbb{R}^m. \tag{22}$$

The heterogeneity $\boldsymbol{h}_i \in \mathbb{R}^m$ captures the mutual-relationships between the expected values of the observations μ_j^i of a particular data point x_i; if $\boldsymbol{\Theta}(j)^i$ significantly deviates from μ^i, then $\boldsymbol{h}_i(j)$ is large. Conversely, if $\boldsymbol{\Theta}^i(j)$ is close to μ^i, then $\boldsymbol{h}_i(j)$ is close to zero.

Definition 3 (weighted heterogeneity) *Define* $\boldsymbol{g}_i$ *as a weighted heterogeneity, whose jth element is given by*

$$\boldsymbol{g}_i(j) = \boldsymbol{\Theta}^i(j) \boldsymbol{h}_i(j). \tag{23}$$

Under the considered statistical model, with the above definitions, the SSD $\boldsymbol{\pi}_i$ can be written explicitly.

Proposition 1 *The j-th element of* $\boldsymbol{\pi}_i$ *can be approximated by*

$$\boldsymbol{\pi}_i(j) = \frac{1}{2m} + \frac{\epsilon - (\boldsymbol{g}_i^2(j) + \sigma_j^i) - 2\langle \boldsymbol{\Sigma}_{j,\cdot}^i \rangle_S + \langle \boldsymbol{\Sigma}^i \rangle_S}{2\epsilon m - 2m(\langle \boldsymbol{g}_i^2 \rangle_S + \langle \boldsymbol{\Sigma}^i \rangle_S)}. \tag{24}$$

The derivation is based on the Taylor expansion of the Gaussian function in Eq (1), where ϵ is the scale of the function. The proof appears in S1 Appendix.

In order to give some intuition, we consider the following special cases, where the SSD assumes a simpler form.

Special case 1. Suppose that the random vectors of the observation functions are independent and identically distributed, i.e., $\boldsymbol{\Sigma}_{k,l}^i = \mathbf{0}\ \forall l \in \{1, \ldots, m\}$ and $k \neq l$. In this case, the k-th element of $\boldsymbol{\pi}_i$ is

$$\boldsymbol{\pi}_i(k) = \frac{1}{2m} + \frac{\epsilon - (\boldsymbol{g}_i^2(k) + \sigma_k^i)}{2\epsilon m - 2m(\langle \boldsymbol{g}_i^2 \rangle_S + \langle \sigma^i \rangle_S)}, \tag{25}$$

where

$$\boldsymbol{\sigma}^i = \begin{bmatrix} \sigma_1^i \\ \sigma_2^i \\ \vdots \\ \sigma_m^i \end{bmatrix}. \tag{26}$$

Note that a small value is assigned to $\boldsymbol{\pi}_i(k)$ if the weighted heterogeneity $\boldsymbol{g}_i(k)$ is large. In contrast, a large value is assigned to $\boldsymbol{\pi}_i(k)$ if $\boldsymbol{g}_i(k)$ is small. As a consequence, $\boldsymbol{\pi}_i(k)$ carries the heterogeneity information of the observations.

Special case 2. When the kernel scale $\epsilon \to \infty$, the information about the heterogeneity of each observation is lost, since the same weights are assigned to all the edges. As a consequence, $\boldsymbol{\pi}_i$ becomes just a constant vector, given by

$$\boldsymbol{\pi}_i \underset{\epsilon \to \infty}{\longrightarrow} \frac{1}{m}\mathbf{1}, \tag{27}$$

where $\mathbf{1} \in \mathbb{R}^m$ is an all-ones vector.

Binary hypothesis testing

Suppose that $\{x_i\}_{i=1}^n \in \mathcal{M}$ are realizations of a random variable X, which follows a bimodal distribution stemming from two hypotheses: $\mathcal{H}_1$ and $\mathcal{H}_2$; $\mathcal{H}_1$ has probability α and $\mathcal{H}_2$ has probability $(1 - \alpha)$, where $0 < \alpha < 1$. Denote the set of data points from hypothesis $\mathcal{H}_1$ by Ω_1 and the set of data points from hypothesis $\mathcal{H}_2$ by Ω_2. Recall that for each data point x_i, $f_j(x_i)$ are the realizations of the elements of the random vector V^i. Since V^i depends on the random variable X, assume that V^i also follows a bimodal distribution, which is induced by the bimodal distribution of X. Particularly, consider a Gaussian setting, where V^i is sampled from $\mathcal{N}(\boldsymbol{m}_1, \boldsymbol{\Sigma}_1)$ with probability α and from $\mathcal{N}(\boldsymbol{m}_2, \boldsymbol{\Sigma}_2)$ with probability of $(1 - \alpha)$. Similarly, assume that the random variable V_j^i follows a bimodal distribution: sampled from $\mathcal{N}(\mu_j^1 \mathbf{1}_d, \sigma_j^1 \boldsymbol{I}_d)$ with probability α and from $\mathcal{N}(\mu_j^2 \mathbf{1}_d, \sigma_j^2 \boldsymbol{I}_d)$ with probability $(1 - \alpha)$, where the respective probability density functions are $f(v|\mu_j^1, \sigma_j^1)$ and $f(v|\mu_j^2, \sigma_j^2)$.

A naïve approach for binary hypothesis testing would be to directly compare the densities of the two hypotheses for each observation separately. Particularly, based on the realizations from only one observation function f_j, the average probability of error attained in a Bayesian setting with the MAP estimator [5] is given by

$$P_{e,j} = \alpha(1 - TV(\mathcal{N}(\mu_j^1 \mathbf{1}_d, \sigma_j^1 \boldsymbol{I}_d), \mathcal{N}(\mu_j^2 \mathbf{1}_d, \sigma_j^2 \boldsymbol{I}_d))), \tag{28}$$

where $TV(\mathcal{N}(\mu_j^1 \mathbf{1}_d, \sigma_j^1 \boldsymbol{I}_d), \mathcal{N}(\mu_j^2 \mathbf{1}_d, \sigma_j^2 \boldsymbol{I}_d))$ denotes the total variation between the realizations of $V_j^i | x_i \in \Omega_1$ and $V_j^i | x_i \in \Omega_2$, defined by

$$TV(\mathcal{N}(\mu_j^1 \mathbf{1}_d, \sigma_j^1 \boldsymbol{I}_d), \mathcal{N}(\mu_j^2 \mathbf{1}_d, \sigma_j^2 \boldsymbol{I}_d)) = \frac{1}{2}\int_v \left| f(v|\mu_j^1, \sigma_j^1) - f(v|\mu_j^2, \sigma_j^2) \right| dv. \tag{29}$$

According to [25], consider $\frac{|\sigma_j^1-\sigma_j^2|}{\sigma(j)} \leq \frac{2}{3}$, where $\sigma(j) = \max\{\sigma_j^1, \sigma_j^2\}$. The total variation above between two Gaussian distributions is bounded by

$$TV(\mathcal{N}(\mu_j^1\mathbf{1}_d, \sigma_j^1\boldsymbol{I}_d), \mathcal{N}(\mu_j^2\mathbf{1}_d, \sigma_j^2\boldsymbol{I}_d)) \leq \frac{|\mu_j^1-\mu_j^2|}{2\sqrt{\sigma(j)}} + \frac{|\sigma_j^1-\sigma_j^2|}{2\sigma(j)}. \tag{30}$$

We seek another more discriminative approach for binary hypothesis testing. For this purpose, we propose a method based on the SSDs. Since the obtained SSDs represent probability distributions, the average probability of error is given by

$$P_e' = \alpha(1 - TV(\langle\boldsymbol{\pi}\rangle_{\Omega_1}, \langle\boldsymbol{\pi}\rangle_{\Omega_2})). \tag{31}$$

According to Proposition 1, the total variation between two SSDs associated with data points from two hypotheses can be explicitly expressed as the l_1 distance given by

$$\begin{aligned} TV(\langle\boldsymbol{\pi}\rangle_{\Omega_1}, \langle\boldsymbol{\pi}\rangle_{\Omega_2}) &= \sum_{k=1}^{m} |\langle\boldsymbol{\pi}(k)\rangle_{\Omega_1} - \langle\boldsymbol{\pi}(k)\rangle_{\Omega_2}| \\ &= \sum_{k=1}^{m} \left| \frac{\epsilon - \boldsymbol{g}_1^2(k) - \sigma_k^1 - 2\langle\boldsymbol{\Sigma}_{k,\cdot}\rangle_{\Omega_1} + \langle\boldsymbol{\Sigma}\rangle_{\Omega_1}}{2\epsilon m - 2m(\langle\boldsymbol{g}^2\rangle_{\Omega_1} + \langle\boldsymbol{\Sigma}\rangle_{\Omega_1})} - \frac{\epsilon - \boldsymbol{g}_2^2(k) - \sigma_k^2 - 2\langle\boldsymbol{\Sigma}_{k,\cdot}\rangle_{\Omega_2} + \langle\boldsymbol{\Sigma}\rangle_{\Omega_2}}{2\epsilon m - 2m(\langle\boldsymbol{g}^2\rangle_{\Omega_2} + \langle\boldsymbol{\Sigma}\rangle_{\Omega_2})} \right|, \end{aligned} \tag{32}$$

which consists of three main components: the variances σ_k^1 and σ_k^2, the weighted heterogeneities $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$, and the covariance $\boldsymbol{\Sigma}$.

The total variation of the measurements in Eq (29) and the total variation between the SSDs in Eq (32) can be used to distinguish between the two hypotheses. In the following, we specify the conditions, under which the total variation based on the SSDs in Eq (31) is larger, and hence, leading to smaller error compared to the standard MAP estimator using a single observation as specified in Eq (28).

Proposition 2 *Suppose $\mu_j^1 = \mu_j^2$ and $\sigma_j^1 = \sigma_j^2$, which imply by definition (or by Eq (30)) that the total variation between the distributions corresponding to the two hypotheses is zero, i.e.*,

$$TV(\mathcal{N}(\mu_j^1\mathbf{1}_d, \sigma_j^1\boldsymbol{I}_d), \mathcal{N}(\mu_j^2\mathbf{1}_d, \sigma_j^2\boldsymbol{I}_d)) = 0. \tag{33}$$

This means that not only the standard MAP estimator but also any estimator based directly on single channel observations cannot distinguish between the two hypotheses. Conversely, from Eq (32), the SSDs may carry a distinction capability, that is,

$$\begin{aligned} &TV(\langle\boldsymbol{\pi}\rangle_{\Omega_1}, \langle\boldsymbol{\pi}\rangle_{\Omega_2}) \\ = &\sum_{k=1}^{m} \left| \frac{\epsilon - \boldsymbol{g}_1^2(k) - \sigma_k^1 - 2\langle\boldsymbol{\Sigma}_{k,\cdot}\rangle_{\Omega_1} + \langle\boldsymbol{\Sigma}\rangle_{\Omega_1}}{2\epsilon m - 2m(\langle\boldsymbol{g}^2\rangle_{\Omega_1} + \langle\boldsymbol{\Sigma}\rangle_{\Omega_1})} - \frac{\epsilon - \boldsymbol{g}_2^2(k) - \sigma_k^2 - 2\langle\boldsymbol{\Sigma}_{k,\cdot}\rangle_{\Omega_2} + \langle\boldsymbol{\Sigma}\rangle_{\Omega_2}}{2\epsilon m - 2m(\langle\boldsymbol{g}^2\rangle_{\Omega_2} + \langle\boldsymbol{\Sigma}\rangle_{\Omega_2})} \right| \geq 0. \end{aligned} \tag{34}$$

Proposition 2 demonstrates that there are cases where a single observation cannot be used for distinguishing between the two hypotheses. However, in such cases, the SSDs may enable us to distinguish the hypotheses due to possible differences in either the heterogeneity or the covariances. Proposition 2 is further demonstrated in the context of the localization toy problem in Simulation 2.

To further expand the analysis, we make the following assumptions.

Assumption 1 *The empirical mean of the weighted heterogeneity is approximately the same under the two hypotheses*:

$$\langle \boldsymbol{g}^2 \rangle_{\Omega_1} - \langle \boldsymbol{g}^2 \rangle_{\Omega_2} \simeq 0. \tag{A.1}$$

Assumption 2 *The empirical mean of the covariance matrices is approximately the same under the two hypotheses*:

$$\langle \boldsymbol{\Sigma} \rangle_{\Omega_1} - \langle \boldsymbol{\Sigma} \rangle_{\Omega_2} \simeq 0. \tag{A.2}$$

Note that if Assumptions (A.1) and (A.2) hold, implying that $2\epsilon m - 2m(\langle \boldsymbol{g}^2 \rangle_{\Omega_1} + \langle \boldsymbol{\Sigma} \rangle_{\Omega_1}) \simeq 2\epsilon m - 2m(\langle \boldsymbol{g}^2 \rangle_{\Omega_2} + \langle \boldsymbol{\Sigma} \rangle_{\Omega_2})$, then the l_1 distance between the SSDs in Eq (32) can be recast as

$$TV(\langle \boldsymbol{\pi} \rangle_{\Omega_1}, \langle \boldsymbol{\pi} \rangle_{\Omega_2}) \simeq \sum_{k=1}^{m} \left| \frac{\boldsymbol{g}_1^2(k) - \boldsymbol{g}_2^2(k) + \sigma_k^1 - \sigma_k^2 + 2(\langle \boldsymbol{\Sigma}_{k,\cdot} \rangle_{\Omega_1} - \langle \boldsymbol{\Sigma}_{k,\cdot} \rangle_{\Omega_2})}{2\epsilon m - 2m(\langle \boldsymbol{g}^2 \rangle_{\Omega_1} + \langle \boldsymbol{\Sigma} \rangle_{\Omega_1})} \right|. \tag{35}$$

Proposition 3 *Suppose that Assumptions (A.1) and (A.2) hold. Suppose* $\mu_j^1 = \mu_j^2$, *which implies that the upper bound of the total variation at the j-th element in Eq (30) only depends on the variance*

$$TV(\mathcal{N}(\mu_j^1 \mathbf{1}_d, \sigma_j^1 \boldsymbol{I}_d), \mathcal{N}(\mu_j^2 \mathbf{1}_d, \sigma_j^2 \boldsymbol{I}_d)) \leq \frac{|\sigma_j^1 - \sigma_j^2|}{2\sigma(j)}. \tag{36}$$

In addition, suppose that (i) the covariance between the jth observation and the other observations under the two hypotheses is approximately equal $\langle \boldsymbol{\Sigma}_{k,\cdot} \rangle_{\Omega_1} \simeq \langle \boldsymbol{\Sigma}_{k,\cdot} \rangle_{\Omega_2}$, *(ii) the empirical mean of the difference of variance and weighted heterogeneity of the two hypotheses is greater than the difference of j-th variance, i.e.,* $\langle (\boldsymbol{\sigma}^1 - \boldsymbol{\sigma}^2)^\top (\boldsymbol{g}_1^2 - \boldsymbol{g}_2^2) \rangle \geq |\sigma_j^1 - \sigma_j^2|$, *and (iii) the weighted heterogeneity of* $\mathcal{H}_1$ *is sufficiently large such that* $\langle \boldsymbol{g}^2 \rangle_{\Omega_1} \geq \epsilon - \langle \boldsymbol{\Sigma} \rangle_{\Omega_1} - \sigma(j)$. *Then, the* l_1 *distance between the SSDs can be recast and bounded from below by*

$$\begin{aligned} TV(\langle \boldsymbol{\pi} \rangle_{\Omega_1}, \langle \boldsymbol{\pi} \rangle_{\Omega_2}) &\simeq \sum_{k=1}^{m} \left| \frac{\boldsymbol{g}_1^2(k) - \boldsymbol{g}_2^2(k) + \sigma_k^1 - \sigma_k^2}{2\epsilon m - 2m(\langle \boldsymbol{g}^2 \rangle_{\Omega_1} + \langle \boldsymbol{\Sigma} \rangle_{\Omega_1})} \right| \\ &\geq \frac{\langle (\boldsymbol{\sigma}^1 - \boldsymbol{\sigma}^2)^\top (\boldsymbol{g}_1^2 - \boldsymbol{g}_2^2) \rangle}{2\sigma(j)} \geq \frac{|\sigma_j^1 - \sigma_j^2|}{2\sigma(j)}. \end{aligned} \tag{37}$$

It follows that

$$TV(\mathcal{N}(\mu_j^1 \mathbf{1}_d, \sigma_j^1 \boldsymbol{I}_d), \mathcal{N}(\mu_j^2 \mathbf{1}_d, \sigma_j^2 \boldsymbol{I}_d)) \leq TV(\langle \boldsymbol{\pi} \rangle_{\Omega_1}, \langle \boldsymbol{\pi} \rangle_{\Omega_2}). \tag{38}$$

This proposition implies that when the assumptions hold, the probability of error based on SSD, which indirectly takes into account the mutual-relations between all observations, facilitates a better distinction of the two hypotheses compared to the standard MAP estimator computed from the best sensor. This property is further demonstrated in the localization toy problem in Simulation 3.

Proposition 4 *Suppose the conditions of Proposition 3 hold. In addition, suppose that the random vectors of the observations are independent and identically distributed, i.e.,*

$\sigma^1_{k,j} = \sigma^2_{k,j} = 0 \; \forall k \neq l$, *then*

$$TV(\mathcal{N}(\boldsymbol{m}_1, \boldsymbol{\Sigma}_1), \mathcal{N}(\boldsymbol{m}_2, \boldsymbol{\Sigma}_2)) \leq TV(\langle \boldsymbol{\pi} \rangle_{\Omega_1}, \langle \boldsymbol{\pi} \rangle_{\Omega_2}). \tag{39}$$

This proposition shows that the SSDs enable us a better distinction between the two hypotheses compared to the MAP estimator based on the distributions of the multi-feature observations. In other words, the heterogeneity comprising the SSD has a significant contribution to the ability to recover the information about the latent data points $\{x_i\}_{i=1}^n$ on the underlying manifold $\mathcal{M}$, thereby leading to accurate binary hypothesis testing.

Imaging mass cytometry (IMC)

IMC is a relatively new imaging method, which enables to examine tumors and tissues at subcellular resolutions, giving rise to images consisting of the intensities of multiple proteins [13–15]. This acquisition platform, combined with computational methods, has recently been the subject of many studies. Various image processing and analysis techniques for IMC datasets can be found in [26], where it is shown that single-cell segmentation can be accomplished successfully with supervised classifiers, leading to the characterization of cell co-occurrence and cell composition of different types of tissues and samples. In [27], an IMC dataset with 37 markers is used for cell segmentation and cell clustering based on random $125 \times 125 \mu m^2$ patches collected from breast cancer patients. This dataset is jointly analyzed with multi-platform genomics data, where it is shown that classifiers can be iteratively trained in a supervised manner to learn from the IMC pixels the corresponding cell expression levels. In [14], spanning-tree progression analysis combined with samples' type provided by pathologists is used for cell population and cell transition identification. In contrast to these methods, our approach focuses on extracting the mutual-relationships between markers at likely tumor cells regions at large, circumventing cell segmentation.

In this work, our goal is to identify the sensitivity of lung cancer subjects to treatment with PD-1 axis blockers, given their IMC multiplexed observations. More specifically, we aim at a binary prediction task: identifying whether the subjects responded or did not respond to the treatment. We analyze two IMC datasets consisting of baseline/treatment tumor samples from non-small cell lung cancer subjects profiled with 29 markers, representing phenotype and functional properties of both tumor and immune cells. The markers are denoted by LipoR, VIM, T-BET, CD47, Cytokeratin, CD45RO, PD-L1, GAPDH, B7-H3, LAG-3, TIM-3, FOXP3, CD4, B7-H4, CD68, PD1, CD20, CD8, CD25, VISTA, KI-67, B2M, CD3, IDO-1, PD-L2, GZB, Histone 3, DNA1 and DNA2. The resolution of the IMC images is 1 μm^2 per pixel. The first dataset, denoted Dataset 1, consists of 55 subjects (samples), and the second dataset, denoted Dataset 2, consists of 29 subjects (samples). These subjects received treatment with PD-1 axis blockers. Based on the clinical sensitivity to the treatment, the subjects are categorized as: durable clinical benefit (denoted with the label `responders`) and no durable benefit (denoted with the label `non-responders`).

Prior to our analysis, a standard pre-processing using z-score normalization is applied to each marker. We remark that the mean and the standard deviation are computed based only on pixels in which the marker has non-zero values. In addition, we apply a 3×3 median filter to every image in Dataset 2. We note that the median filter is not applied to Dataset 1, because the typical expression levels in this dataset are sparse, and in this situation, application of a filter may destroy the signal. This is demonstrated in S1 Fig, where we apply the median filter to the expression levels of few important markers and observe a degenerate result. The markers exhibiting such sparse expression levels in Dataset 1 are VIM, T-BET, CD45RO, PD-L1, GAPDH, B7-H3, LAG-3, FOXP3, B7-H4, PD1, CD20, CD8, CD25, VISTA, KI-67, CD3,

PD-L2 and GZB. Conversely, in Dataset 2, the typical expression levels are dense, and as a result, the median filter enhances the content of the images.

Our analysis does not consider the entire image, but rather focuses on ROIs located at the highest Cytokeratin expression levels, as Cytokeratin is expressed only in the tumor cells. As noted above, this circumvents cell segmentation that is typically required in other analyses techniques. At each ROI, we consider a "stack" of m image patches from all the m markers, where each patch consists of $d = b \times b$ pixels. We select N ROIs per sample by searching the patches with the maximal mean value of Cytokeratin. This is implemented by a 2D convolution applied to the Cytokeratin image with a constant kernel of size $b \times b$. We deliberately avoid patch overlap by assigning zero values to the pixels of each ROI, once it is selected.

Using the present work notation, the intrinsic representation of the information embodied at each ROI is denoted by x_i, which is assumed to be a data point from some hidden manifold $\mathcal{M}$. Our working hypothesis is that the distribution of $\{x_i\}_{i=1}^{n}$ on the hidden manifold $\mathcal{M}$ is bimodal, which is induced by the sensitivity of the subjects to the treatment; data points x_i at ROIs within tissues of `responders` are located in one region of the manifold, and data points x_i at ROIs within tissues from `non-responders` are located in another region of the manifold. Given a data point x_i, our two hypotheses, $\mathcal{H}_{\mathrm{r}}$ and $\mathcal{H}_{\mathrm{n}}$, are whether x_i is a realization from the distribution of `responders` or `non-responders`, respectively. Here, $n = N \times P$ is the total number of ROIs across all the samples, where P is the number of samples and N is the number of ROIs we consider in each sample. Next, recall that the intrinsic representation x_i is hidden. Instead, the accessible observations are the expression levels of the markers at the ROIs, which are represented mathematically by the observation functions $f_j(x_i) \in \mathbb{R}^d$ for $j = 1, \ldots, m$, where $m = 29$ is the number of markers. The domain of the observation functions is the hidden intrinsic manifold $\mathcal{M}$ and the range is of dimension $d = b \times b$, which is the size of a patch of an image of one marker. Namely, the values assigned by $f_j(x_i)$ to x_i correspond to the expression level of marker j at the ROI associated with x_i. The set of patches of all the markers at a specific ROI x_i is a set of multi-feature observations $\{f_j(x_i)\}_{j=1}^{29}$. In Fig 2, we illustrate the IMC multiplexed observations from a single subject and the set up under consideration.

The identification of the subject's response to treatment from the IMC data is based on the application of the proposed method described in Box 1. This algorithm fits well the problem at hand due to the following main reasons. First, directly comparing observations from the different markers, $f_j(x_i)$, is inapplicable since each ROI comprises different cells and different tissue structures. Our method circumvents this problem by computing an intrinsic signature of the observations. Second, due to the different dynamics of the nominal values of the observations from the different markers, a naïve concatenation of the multi-feature observations is inadequate (in contrast to the localization example).

Before applying the proposed method, we test empirically that Assumptions (A.1) and (A.2) hold. Note that for this test only, the true response status is used. To test Assumption (A.1), we compute

$$\frac{||\langle \boldsymbol{g}^2 \rangle_{\Omega_r} - \langle \boldsymbol{g}^2 \rangle_{\Omega_n}||_2}{\sqrt{||\langle \boldsymbol{g}^2 \rangle_{\Omega_r}||_2 \cdot ||\langle \boldsymbol{g}^2 \rangle_{\Omega_n}||_2}}$$

and to test Assumption (A.2) we compute

$$\frac{||\langle \boldsymbol{\Sigma} \rangle_{\Omega_r} - \langle \boldsymbol{\Sigma} \rangle_{\Omega_n}||_F}{\sqrt{||\langle \boldsymbol{\Sigma} \rangle_{\Omega_r}||_F \cdot ||\langle \boldsymbol{\Sigma} \rangle_{\Omega_n}||_F}},$$

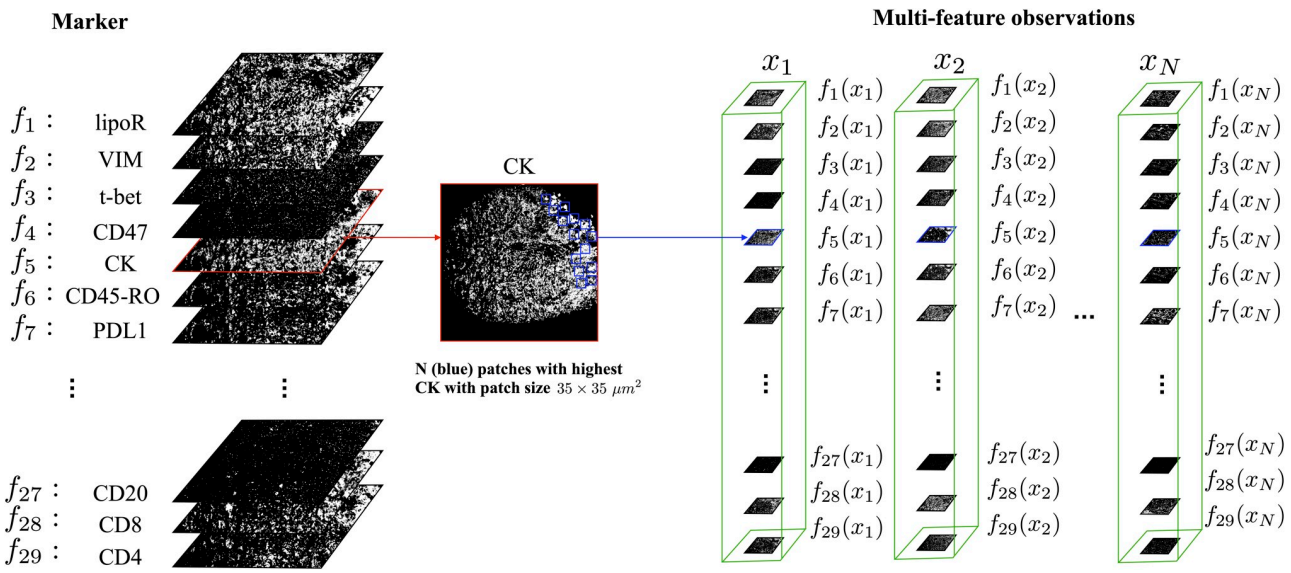

Fig 2. An illustration of the IMC multi-feature observations from a single subject. In the IMC data of a single subject, we focus first on the Cytokeratin marker, which is used as an indicator of tumor cells. We select N ROIs located at the highest Cytokeratin expression levels. These ROIs are patches of size $b \times b\ \mu m^2$, where here b set to 35. We assume that these ROIs have intrinsic hidden states represented by x_i. The expression of all 29 markers at these ROIs are viewed as the multi-feature observations $\{f_j(x_i)\}_{j=1}^{29}$ for $i = 1, \ldots, N$.

https://doi.org/10.1371/journal.pcbi.1008741.g002

where Ω_r and Ω_n denote the sets of `responders` and `non-responders`, respectively. The respective values we obtain for Dataset 1 are 0.12 and 0.08 and for Dataset 2 the values for `responders` and `non-responders` are 0.03 and 0.04. This implies that the conditions in the two assumptions are approximately satisfied.

We analyze the two datasets separately, since the datasets were collected around a year apart, and internal acquisition system parameters were modified during that time. We apply the proposed method presented in Box 1 to the observations, $\{f_j(x_i)\}_{j=1}^{29}$ for $i = 1, \ldots, n$, resulting in a low-dimensional representation of the ROIs. Then, we apply to the low-dimensional representation an RBF SVM classifier with a leave-one-subject-out (LOSO) cross-validation [5] in order to predict the response to treatment. In order to assess the prediction performance for each subject, we compute the average of the prediction results of all the ROIs of that subject. We note that the prediction is based on features computed per patch (rather than per subject). Therefore, in practice, the number of samples used for the cross-validation is $55 \times N$ for Dataset 1 and $29 \times N$ for Dataset 2. Importantly, at each cross-validation iteration, all N patches of a subject were removed from the training set, and were only used for testing the classifier.

We compare the SSD-based representation obtained by the propose method to other representations obtained by three competing algorithms. The first is a direct application of diffusion maps to the sets of multi-feature observations $\{f_j(x_i)\}_{j=1}^{29}$ for $i = 1, \ldots, n$. The second and the third are based on HKS [16] and WKS [17], respectively, replacing the SSD as the features of each ROI at the first stage of the proposed method in Box 1. The dimensions of the representation obtained by the proposed method and by the three competing algorithms are determined by a variant of the Jackstraw method, described in Determining the dimension of data.

Fig 3 presents the predictions of treatment response by the SSD, HKS, WKS based algorithms as well as DM. We show the 3D t-SNE visualization [28] of the patch representation obtained as an output of the different algorithms and the confusion matrix of the prediction obtained by an RBF SVM classifier applied to the patch representation. To complement the

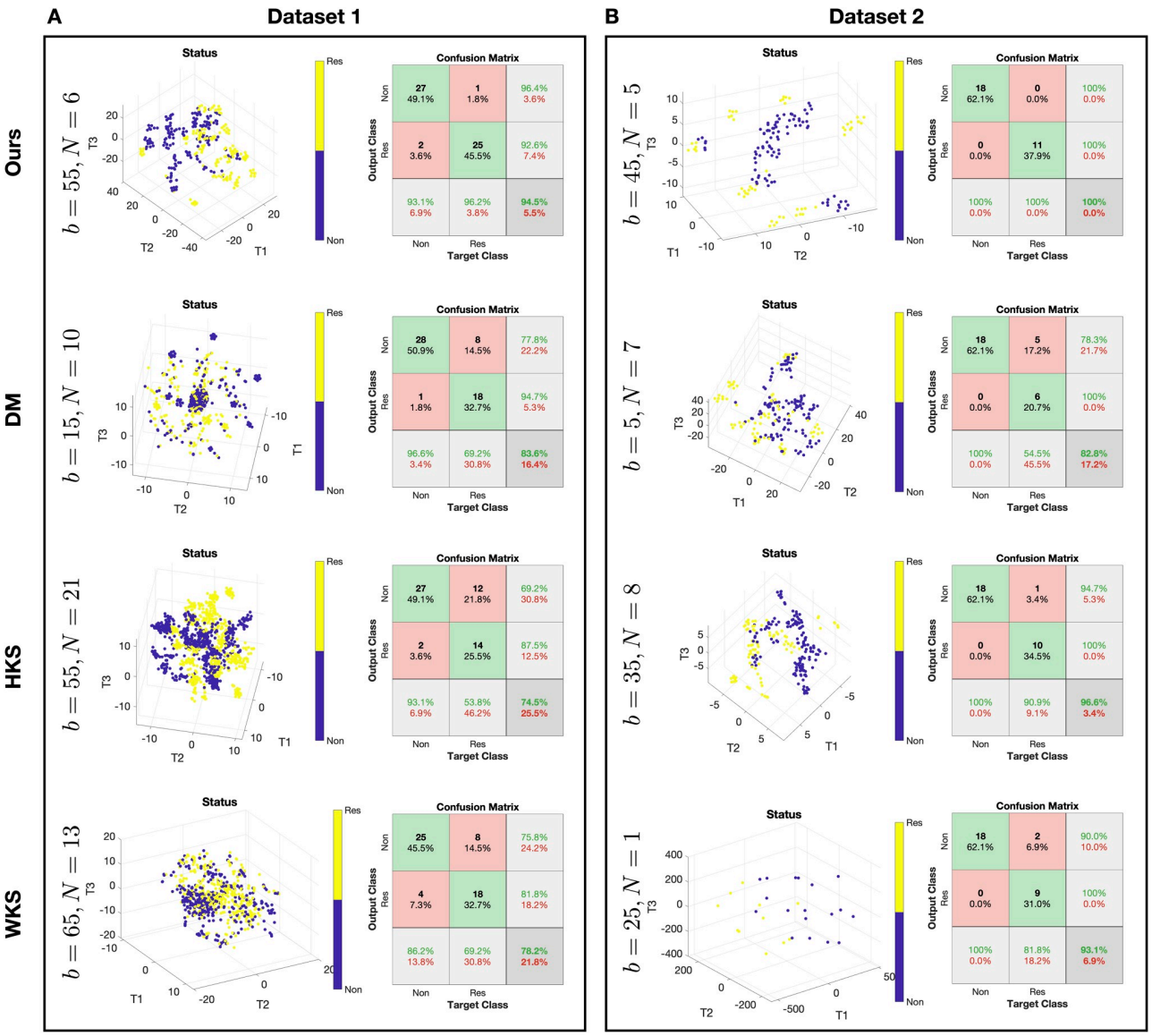

Fig 3. Treatment response predictions. (A) Prediction for Dataset 1. (B) Prediction for Dataset 2. In each panel, at each row, we plot the best algorithm configuration, the 3D t-SNE visualization of the patches representation colored by the response status, and the confusion matrix of the response prediction obtained by a leave-one-subject-out cross-validation using an RBF SVM classifier. The rows present the results of the different methods.

https://doi.org/10.1371/journal.pcbi.1008741.g003

results, in Table 1, we present the area under the ROC curve (AUC) of the treatment response predictions. We note that the AUC obtained for Dataset 2 without the median filter pre-processing is 0.931. For each algorithm, the presented prediction results are based on the patch size and number of patches configuration that yielded the best empirical performance using

Table 1. ROC AUC for predictions of treatment sensitivity.

	Proposed Method	DM	HKS	WKS
Dataset 1	0.9455	0.8364	0.7455	0.7818
Dataset 2	1	0.8276	0.9655	0.9310

https://doi.org/10.1371/journal.pcbi.1008741.t001

cross-validation as described above. The best configuration (patch size $b \times b$ and number of patches per subject N) of each algorithm is presented in Fig 3 on the left.

In the t-SNE plots, we observe that the unsupervised separation of the patches from `responders` and `non-responders` is most pronounced in the low-dimensional representation obtained by the proposed method. This distinct visual separation, which was obtained by the proposed algorithm without access to any response outcome information. The color in the figures indicates the response status. Note that the embedding is obtained by the unsupervised algorithm and the color labels are overlaid to demonstrate the degree of separability between patches from `responders` and `non-responders`, which implies that this patch representation is informative and useful for the subsequent response prediction. This result, which is obtained in an unsupervised manner, distinguishes the current work from the computational methods for IMC data described above that rely on supervised analysis. Indeed, we observe that the prediction accuracy obtained based on the proposed method is superior compared to the other three competing methods. We note that Dataset 1 consists of 26 `responders` and 29 `non-responders`, so that the chance level of accurate prediction of a subject with durable clinical benefit is 47.27%, and Dataset 2 consists of 11 `responders` and 18 `non-responders`, thus, the chance level of accurate prediction is 37.93%.

The derivations in Theoretical analysis imply that the capability to distinguish between the two hypotheses (sensitivity or insensitivity to treatment) highly depends on the mutual relationships between the markers, which we explicitly define and term heterogeneity (Definition 2). Since our empirical study demonstrates that our approach facilitates a distinct separation between ROIs of subjects according to their treatment response, by the theoretical analysis, we conclude that the heterogeneity between the markers is where the information about the sensitivity to treatment lies.

We examine the sensitivity of the tested algorithms to the choice of the hyperparameters: the number of ROIs per subject N and the size of the patch $b \times b$. In S2 Fig, we plot heatmaps of the treatment prediction accuracy obtained based on different choices of hyperparameters. We observe that within a relatively wide range of parameter values, the performance of the proposed method in Box 1 is high and insensitive to the particular choice of parameters. In addition, we note that the range of patch size where high performance is attained is centered at size 45×45 μm^2. As the size of tumor cells vary within a range of 10–30 μm^2 (a mean of 20 μm^2), and the size of a lymphocyte ranges from 8–12 μm^2 (a mean of 10 μm^2), it implies that high performance is attained when patches are likely including more than one cell and cell type. Conversely, we observe that the competing methods do not show the same degree of robustness to the choice of hyperparameters, and good performance is attained only for very particular (isolated) parameter values.

To further exploit the robustness of the proposed method, we implemented an ensemble of the classifiers based on different values of hyperparameters. The implementation of the combination is based on [29]. The performance of the combined classifiers is presented in S3 Fig. We observe that the prediction accuracy of this ensemble is comparable to the prediction accuracy obtained based on the classifier with the best parameter configuration, demonstrating that the particular choice of hyperparameters can be circumvented.

Our premise is that the resulting high prediction accuracy is attributed mainly to the unsupervised informative representation obtained by our approach, rather than the classifier type. To support this claim, we repeat the analysis by replacing the RBF SVM classifier with Random Forest [5]. The results are shown in S4 Fig and demonstrate comparable performance and similar trends.

To show that Stage 2 of the proposed method in Box 1 is essential, we plot in S5 Fig heatmaps of the multi-feature observations and the SSD features resulting from Stage 1 of the

algorithm. The heatmaps of the multi-feature observations demonstrate that there is no obvious difference between `responders` and `non-responders`, implying that the response status prediction is a non-trivial task. In addition, inspection of the heatmaps with the SSD features does not reveal apparent distinction between `responders` and `non-responders`. Recalling that after Stage 2, as presented in the t-SNE plots in Fig 3, this distinction becomes evident, demonstrating the contribution of Stage 2.

Discussion

We presented a two-step graph analysis approach. The first step is applied to the multi-feature observations of the data points, where their mutual-relationships are extracted. This step is implemented by computing the SSD of a random walk defined on the graph whose nodes are the observations. The resulting SSD can be viewed as a signature or a characteristic vector of the data point and is analogous to traditional signatures from other domains, such as the heat kernel signature (HKS) [16] or the wave kernel signature (WKS) [17] in the field of shape analysis, and PageRank [30] in web page ranking (see Related work). The second step is applied to the signatures obtained at the first step, for the purpose of constructing an intrinsic low-dimensional representation of all the data points.

Previous attempts to analyze such mulitplexed datasets involve various approaches, including direct comparisons of the marker expressions, the cell morphology, and interactions in cell neighborhoods, to name but a few [26]. Our method introduces a new approach, building a new representation of the multiplexed data in two steps. Since each of the steps involves a construction of a graph, the entire procedure can be viewed as building a graph of graphs.

While the algorithm is described in a general setting of multiplexed data fusion, our theoretical analysis is focused on binary hypothesis testing. In comparison with a traditional statistical estimation approach, we show that our method exhibits advantages, implying that the mutual-relationships between the multi-feature observations are well captured. In the context of IMC, this could minimize the effect of deviation in individual marker scores and cell/tissue heterogenenity.

We apply the proposed method to two IMC datasets and show that solely from the imaging data, we can distinguish between two different sensitivity levels to treatment. Since our approach does not rely on rigid prior knowledge or access to labels, it has the potential of identifying biological relevance of novel parameters or marker patterns in treatment responses by analyzing dominant factors contributing more to the model stratification. Importantly, we remark that in contrast to common practice, the proposed approach does not require cell-segmentation as a precursor.

In addition to the demonstrated advantage of the proposed method over the competing methods in terms of superior prediction accuracy of treatment response, the proposed method is also more computationally efficient. In S6 Fig, we present the run time of the two stages of the proposed method in Box 1 and compare them to the run time of the three competing methods. We note that the HKS and WKS replace the proposed SSD in Stage 1, but, their respective Stage 2 are similar. In Stage 1, we observe that the proposed method in Box 1 is faster than the competing methods. The main difference between the algorithms in Stage 1 is due to the fact that the SSD is proportional to the degree of the local graph, and as a result, it can be computed efficiently from the degree vector. Conversely, both HKS and WKS require eigenvalue decomposition, which is computationally more demanding. In addition, the direct application of DM just involves Stage 2 of the proposed algorithm. Seemingly, the run time of the algorithms in Stage 2 should have been the same, yet, we observe that DM is slower. This difference is attributed to the significantly different size of the feature space. In DM, the multi-

feature observations are simply concatenated, giving rise to feature space (input of Stage 2) of size $(b \times b \times m) \times (P \times N)$, where $b \times b$ is the size of patch, m is the number of biomarkers, and P is the number of subjects. Conversely, in Stage 2 of Algorithm 1, HKS, and WKS, the feature space (input of Stage 2) is only of size $m \times (P \times N)$.

It is conceivable that the most important hyperparameter of our method is the scale parameter. In S7 Fig, we present a toy example demonstrating that different values of the scale parameter ϵ lead to multiscale signatures capturing local and global features. We demonstrate that different scales facilitate the extraction of different features of the data. In future work, we plan to further explore the role of the scale and to devise multiscale signatures. Another possible direction for future research relies on the fact that our method is general and can be extended to other multiplexed datasets. For example, hyper spectral imaging, sensor networks, spatial multiplexed proteomics, and spatial transcriptomics assays is a representative subset of distinct technologies from diverse domains of science and engineering that share common data structures. The data in all these modalities consist of high-dimensional multivariate observations (m-dimensional feature space) collected at different spatial positions, and therefore, can be analyzed using similar computational methodologies. Furthermore, in many studies practitioners collect datasets consisting of multiple spatial assays of this type, each capturing such data from a single biological sample, patient, or hyper spectral image, etc. Each of these spatial assays could be characterized by several regions of interest (ROIs), giving rise to a setting similar to the IMC problem considered here. Specifically, we plan to examine applications of the proposed graph of graphs analysis to spatial transcriptomics such as Slide-seq [31], High-Density Spatial Transcriptomics [32, 33], MIBI-TOF [34], and DBiT-seq [35].

Materials and methods

Diffusion maps

Manifold learning is a class of nonlinear techniques that embeds high dimensional data points into a low dimensional space, relying on the assumption that the high-dimensional data lie on a low dimensional manifold $\mathcal{M}$ [2–4]. In order to "learn" the manifold from a discrete set of data points, a graph is typically defined, where the graph nodes are the data points and the edges are determined according to some similarity notion. Since the manifold information is entirely captured by its Laplacian, the discrete counterpart, the graph Laplacian is used to build a low-dimensional embedding that respects the manifold in some sense [36]. To this end, common practice is to compute and exploit the spectral decomposition of the graph Laplacian. Diffusion maps is one of these methods, which constructs a random walk on the graph and represents the data points in a low-dimensional space preserving the neighborhood information [4].

Consider a set of data points $\{y_i\}_{i=1}^{b}$, where $y_i \in \mathbb{R}^d$ for $i = 1, \ldots, b$. An undirected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{W})$ is constructed from the data points, where the vertex set is $\mathcal{V} = (y_1, y_2, \ldots, y_b)$ and the weights of the edges connecting two vertices are determined by a measure of similarity between any two data points, e.g., by

$$W(i,j) = \exp\left(-\frac{||y_i - y_j||_2^2}{2\epsilon}\right), \tag{40}$$

where $i, j \in \{1, \ldots, b\}$ and $\epsilon > 0$ is a scale parameter. Common practice is to set ϵ as the median of the distances between the graph nodes. Note that ϵ implicitly induces a notion of locality: it can be viewed as the (squared) radius of the neighborhood around each node, so that only nodes within this radius are considered as neighbors in the graph.

Next, a random walk $\boldsymbol{P}$ on the data points is constructed by normalizing the weight matrix $\boldsymbol{W}$

$$P(i,j) = \frac{W(i,j)}{d(i)}, \tag{41}$$

where $d(i) = \sum_{j=1}^{n} W(i,j)$. $\boldsymbol{P}$ is the transition matrix of a Markov chain defined on the data points $\{y_i\}_{i=1}^{b}$ (graph vertexes), where the entry $p(i,j)$ describes the probability of a random walk transitioning from the node y_i to the node y_j in a single step. Raising the transition matrix $\boldsymbol{P}$ to a power t can be viewed as applying the Markov chain to the data points t times.

Since $\boldsymbol{P}$ is similar to a symmetric and positive-define matrix, $\boldsymbol{P}$ has a biorthogonal right- and left-eigenvectors $\{\boldsymbol{\varphi}_i, \boldsymbol{\upsilon}_i\}_{i=1}^{b}$ with the eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_b \geq 0$. Consequently, the spectral decomposition of $\boldsymbol{P}^t$ is given by

$$P_t(i,j) = \sum_{k=1}^{b} \lambda_k^t \boldsymbol{\varphi}_k(i) \boldsymbol{\upsilon}_k(j). \tag{42}$$

The diffusion distance $D_t^2(i,j)$ between two data points y_i and y_j in the data set is defined by

$$D_t^2(i,j) = \sum_{k=1}^{b} \frac{(P_t(i,k) - P_t(j,k))^2}{\boldsymbol{\upsilon}_1(k)}, \tag{43}$$

which measures the similarity of two points based on the evolution of their probability distributions, and depends on all possible paths of length t in the graph between any two points. Namely, if two points are connected by a large number of paths, then the diffusion distance between them will be small. Conversely, if there are only few paths connecting two points, then the diffusion distance between them will be large.

The diffusion maps is defined by [4]

$$\Phi_t : y_i \mapsto (\lambda_2^t \varphi_2(y_i), \lambda_3^t \varphi_3(y_i), ..., \lambda_l^t \varphi_l(y_i))^\top. \tag{44}$$

We remark that in many cases, due to the typical fast decay of the eigenvalues of $\boldsymbol{P}^t$, l can be set to be smaller than d, thereby achieving dimension reduction. In addition, $\boldsymbol{\varphi}_1$ is a constant vector and therefore is not used in the mapping.

It can be shown that the diffusion distance can be approximated by the eigenvalues and eigenvectors by [4]

$$D_t^2(i,j) = \sum_{k=1}^{b} \lambda_k^{2t} (\varphi_k(i) - \varphi_k(j))^2 \approx ||\Phi_t(i) - \Phi_t(j)||^2, \tag{45}$$

where equality is reached for $l = b$. Namely, the diffusion distance can be approximated by the Euclidean distance between the diffusion maps of the data points.

Determining the dimension of data

A common problem in diffusion maps setting is how to choose the dimension l. The authors in [6] proposed *Jackstraw* to identify the number of principal components (PCs) in the context of principal component analysis (PCA) [37]. We present here a variant of *Jackstraw*, adapting it to diffusion maps.

Given a random walk $\boldsymbol{P}$ constructed from a set of b data points, the associated eigenvalues are $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_b$ with corresponding right-eigenvectors $\{\varphi_i\}_{i=1}^{b}$. Collect the

eigenvalues into a vector, denoted by $\boldsymbol{\lambda}$, where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_b)^\top$. Let $\boldsymbol{P}_k^*$ consist of the random permutation of $\boldsymbol{P}$. Apply eigenvalue decomposition to $\boldsymbol{P}_k^*$ and obtain the corresponding eigenvalues vector $\boldsymbol{\lambda}_k^*$. Repeat this shuffling procedure s times and obtain a set of vectors $\{\boldsymbol{\lambda}_k^*\}_{k=1}^s$. The dimension of the representations is determined by

$$l = \arg\min_{x=\{1,\ldots,b\}} (\boldsymbol{\lambda}(x) \geq \max\{|\boldsymbol{\lambda}_k^*(x)|\}), \quad \forall k = \{1, \ldots, s\}. \tag{46}$$

Note that the absolute values of the eigenvalues of $\boldsymbol{P}_k^*$ are considered because $\boldsymbol{P}_k^*$ is not necessarily symmetric and therefore its eigenvalues $\boldsymbol{\lambda}_k^*$ are not guaranteed to be real.

Related work

Heat kernel signature. There are several shape analysis signatures obtained by spectral methods with different geometric properties such as isometry and deformation invariance [16, 17, 38, 39], related to the proposed method. One of the notable shape signatures is based on the heat diffusion on a shape, called Heat Kernel Signature (HKS) [16]. Broadly, the HKS is obtained by the eigenvalue decomposition of the heat kernel defined on the shape. In the context of our problem, since the heat kernel and the Laplace-Beltrami operator Δ_i share the same eigenbasis, and since the discrete graph Laplacian converges (point-wise) to the Laplace-Beltrami [40]

$$\frac{1}{\epsilon}\boldsymbol{L}_i = \frac{\boldsymbol{I} - \boldsymbol{P}_i}{\epsilon} \xrightarrow[\epsilon \to 0]{N \to \infty} \Delta_i, \tag{47}$$

then, a discrete counterpart of HKS is given by

$$x_i \mapsto [\mathrm{HKS}_t(f_1(x_i)), \mathrm{HKS}_t(f_2(x_i)), \ldots, \mathrm{HKS}_t(f_m(x_i))], \tag{48}$$

where

$$\mathrm{HKS}_t(f_j(x_i)) = \sum_{k=1}^{m} \exp(-(1 - \lambda_k)t)\boldsymbol{\psi}_k^2(j), \tag{49}$$

λ_k and $\boldsymbol{\psi}_k$ are the k-th eigenvalue and k-th eigenvector of the random walk, respectively, and t is the number of random walk steps on the graph. For more details on HKS, we refer the readers to [16].

Similarly to the DKS in Eq (12), the HKS can also be viewed as a low-pass filter. Observe that, for small t, the HKS approximates the DKS by Taylor expansion; for other t values, the weights assigned by the HKS decay faster than the weights of DKS, and therefore, the DKS gives more attention to finer structures.

Wave kernel signature. Another related shape signature is built by the wave function to the Schrödinger equation describing the quantum mechanical particles, called Wave Kernel Signature (WKS) [17]. Similarly to the HKS, the WKS is given by

$$x_i \mapsto [\mathrm{WKS}_t(f_1(x_i)), \mathrm{WKS}_t(f_2(x_i)), \ldots, \mathrm{WKS}_t(f_m(x_i))], \tag{50}$$

where

$$\mathrm{WKS}_t(f_j(x_i)) = \sum_{k=1}^{m} C_t \exp\left(-\frac{(\log t - \log(1 - \lambda_k))^2}{2\sigma^2}\right)\boldsymbol{\psi}_k^2(j), \tag{51}$$

with

$$C_t = \left(\sum_{k=1}^{m} \exp\left(-\frac{(\log t - \log(1-\lambda_k))^2}{2\sigma^2} \right) \right)^{-1}, \tag{52}$$

λ_k and $\boldsymbol{\psi}_k$ are the k-th eigenvalue and k-th eigenvector of the random walk, and t is number of random walk steps on the graph. While the DKS and HKS are viewed as lowpass filters, we note that the WKS can be viewed as a band-pass filter. For more details, we refer the readers to [17].

Nodes ranking. In the context of web page ranking, there are several traditional algorithms based on the spectral analysis of directed graphs. Among them, the celebrated PageRank score is based on the stationary distribution of a random walk representing the popularity of linked web pages [30]. Hyper induced topic search (HITS) is another related algorithm, identifying the influential nodes using a random walk on a graph. There, the graph nodes are the web pages, which are divided into two groups: authorities and hubs [41]. Both algorithms address the problem of web page ranking, where PageRank depends on the incoming links whereas HITS focuses on the outgoing links.

Localization toy problem

To illustrate the challenge in the problem setting and the generality of the proposed solution, we present three simulations of different localization problems. The Matlab code is available in S1 Matlab Code.

Simulation 1. Consider 800 objects on a 2-sphere in $\mathbb{R}^3$ that can be located at four different regions. Each region consists of 200 objects. The positions of the objects are measured by 5 sensors, giving rise to the following set of observations $\{f_j(x_i)\}_{j=1}^{5} \in \mathbb{R}^{100}$, where j is the index of the sensor and i is the index of the object (position). Each sensor measures the position in $d = 100$ coordinates in the following way

$$\mathbb{R}^{100} \ni f_j(x_i) \sim \mathcal{N}(\mathbf{0}, \sigma_i^2 \boldsymbol{I}_{100}), \tag{53}$$

where the standard deviation of the measurement depends on the distance between the position of the sensor and the position of the object $\sigma_i = 20 \exp(- \| x_i - s_j \|_2^2)$, $\boldsymbol{I}_{100}$ is the identity matrix of size 100×100, $\| \cdot \|_2$ denotes the Euclidean norm and s_j denotes the 3D position of sensor $j \in \{1, \ldots, 5\}$. In other words, each object position is captured by $d = 100$ realizations of a Gaussian random variable with variance that is proportional to the distance between the sensor position and the object position. Note that the positions are captured by the sensor through the variance, therefore, they are difficult to infer directly by the multi-feature observations.

In Fig 4A, we present our setting consisting of objects and sensors located on and near a sphere, respectively. The objects positions are marked by dots, the different regions are marked by different colors (red, black, blue and yellow), and the sensors positions are marked by green stars. At the bottom, we present the (high-dimensional) sensor observations. Each block consists of 100×200 scalar observations corresponding to the observations of a single sensor from each region, where $d = 100$ is the dimension of each observation and 200 is the number of the positions per region. Visually, it is evident that distinguishing between the different regions merely based on these observations is non trivial.

For illustration purposes, we view the problem as a classification problem, where given the high-dimensional multi-feature observations, the task is to identify in which region the object at x_i resides.

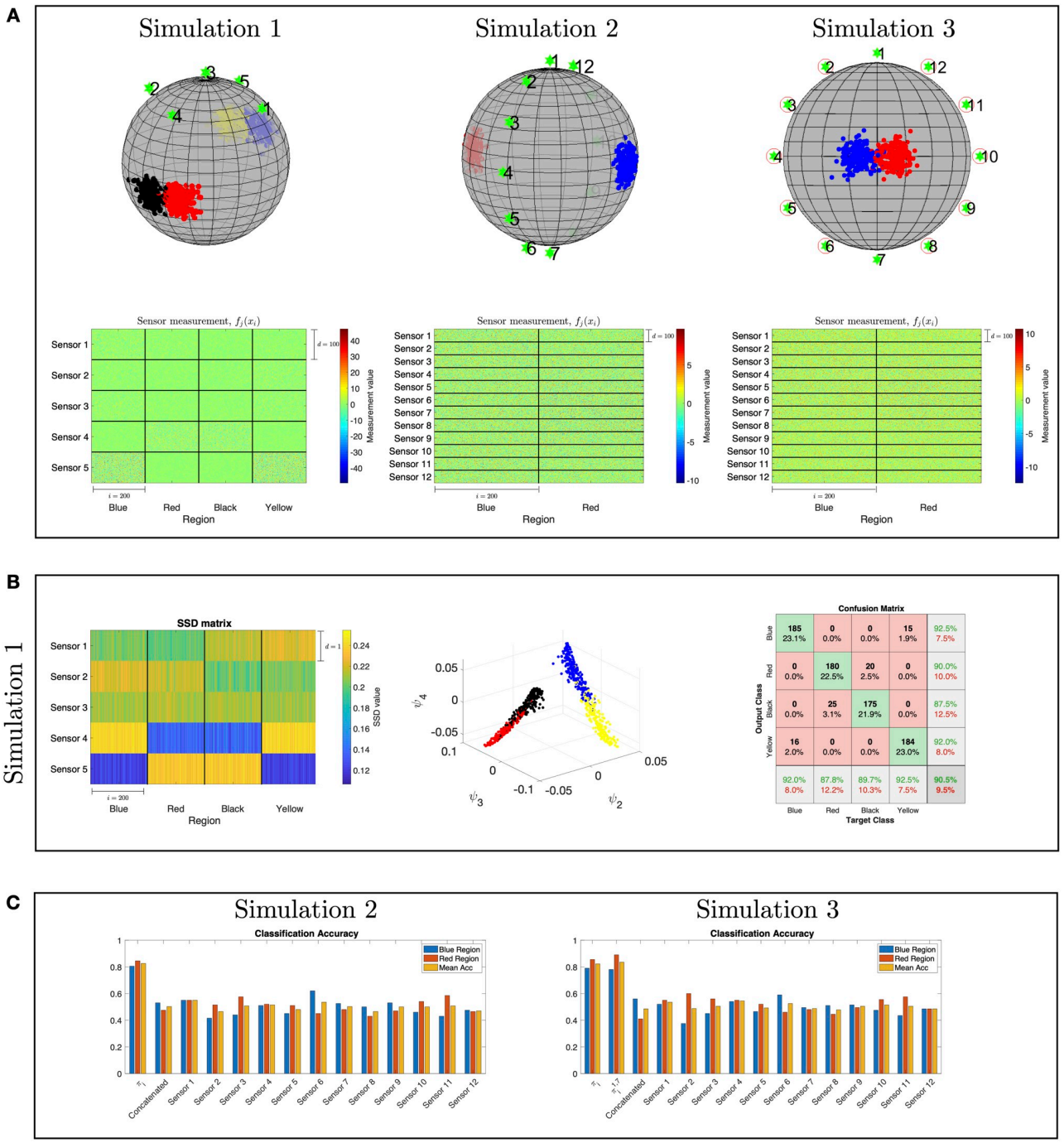

Fig 4. Illustration of localization toy problems. (A) The sensors and objects locations on a sphere are marked by green stars and dots, respectively. The multi-sensor observations correspond to Simulation 1 (left), Simulation 2 (middle), and Simulation 3 (right). (B) Results of the application of our approach to Simulation 1: the SSDs obtained by the proposed method (left), the diffusion maps embedding (middle), and the localization confusion matrix obtained by a 10-fold cross-validation with an RBF SVM classifier (right). (C) A comparison between the localization accuracy obtained by the proposed method based on SSDs and the localization accuracy obtained based on the output of each sensor as well as the concatenation of the output from all the sensors. The localization accuracy is the number of correctly identified positions divided by the true number of total positions in each region.

https://doi.org/10.1371/journal.pcbi.1008741.g004

Table 2. Localization accuracy from measurements of Simulation 1.

Region	Concatenated sensors	Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5
Blue	43%	46%	18%	44.5%	58%	40.5%
Red	54%	15.5%	52.5%	45.5%	48.5%	55%
Black	46.5%	49.5%	29%	40%	37.5%	42%
Yellow	38.5%	33.5%	52%	30.5%	42.5%	50.5%

The performances obtained by the 10-fold cross-validations are presented. The percentages indicate the number of correctly classified positions divided by the true number of total positions in each class.

https://doi.org/10.1371/journal.pcbi.1008741.t002

The observations from each sensor consisting of 800 object positions from four regions were processed in two stages: first, a 3D embedding is constructed by applying diffusion maps to the high-dimensional observations, and second, an RBF SVM is applied to the obtained embedding in order to classify the region. To evaluate the classifiers, we perform a 10-fold cross-validation. A similar two-step procedure is applied to the concatenation of the observations from all the sensors.

Table 2 presents the resulting classification accuracy, i.e., the number of correctly classified positions divided by the total number of positions in each region. The presented results are obtained using a 10-fold cross-validation. We observe that none of the sensors enables an accurate classification. Moreover, we show in Table 2 that a naïve concatenation of the observations from all the sensors do not yield a good classification either.

Seemingly, in order to mitigate the problem, we could simply represent the objects positions by a vector of the variance of the observations. However, it would require prior knowledge about the sensing model, whereas our approach is model-free.

In Fig 4B, we present the classification results obtained by the proposed method. In the diffusion maps embedding constructed from the SSDs, we observe a clear separation between the four regions. Finally, in the confusion matrix of the classification, we observe that the proposed method leads to significantly better classification results compared to the results in Table 2. This demonstrates the importance of taking into account the mutual-relationships between the sensor observations, rather than processing the nominal values of the observations directly, which in this case, give rise to correct identification of the four regions.

Simulation 2. The main purpose of this simulation is to demonstrate Proposition 2 from Binary hypothesis testing.

Consider $n = 400$ positions, $\{x_i\}_{i=1}^{400}$, such that $x_i \in S^2 \subset \mathbb{R}^3$, which are sampled from two different regions on the sphere. Suppose that the positions of the objects follow a bimodal distribution as depicted in the middle-top figure in Fig 4A: an object is located in the blue region with probability $\frac{1}{2}$ and in the red region with probability $\frac{1}{2}$. The two regions represent the two hypotheses, $\mathcal{H}_1$ and $\mathcal{H}_2$, where each region consists of 200 positions. Note the symmetry in this setting, that is, the distances from the blue and red regions to the sensors are approximately the same.

Here, we have 12 sensor observations $\{f_j(x_i)\}_{j=1}^{12}$, measuring the positions of the objects, located at s_j. The multi-feature observations are random samples from

$$\{f_j(x_i)\}_{j=1}^{12} \sim \frac{1}{2}\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_1) + \frac{1}{2}\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_2), \tag{54}$$

where $f_j(x_i) \in \mathbb{R}^{100}$, the standard deviation is $\sigma_j^i = 20 \exp(- \| x_i - s_j \|_2^2)$, and the covariance $\Sigma_{k,l} \sim |\mathcal{N}(0, 0.5)|$ for $\lfloor \frac{k}{100} \rfloor, \lfloor \frac{l}{100} \rfloor \in [1, \ldots, 12]$.

In this simulation, a direct computation can show that Assumptions (A.1) and (A.2) hold. Specifically, we note that

$$\frac{||\langle \boldsymbol{g}^2\rangle_{\Omega_r} - \langle \boldsymbol{g}^2\rangle_{\Omega_b}||_2}{\sqrt{||\langle \boldsymbol{g}^2\rangle_{\Omega_r}||_2 \cdot ||\langle \boldsymbol{g}^2\rangle_{\Omega_b}||_2}} = 0.03$$

and

$$\frac{||\langle \boldsymbol{\Sigma}\rangle_{\Omega_r} - \langle \boldsymbol{\Sigma}\rangle_{\Omega_b}||_F}{\sqrt{||\langle \boldsymbol{\Sigma}\rangle_{\Omega_r}||_F \cdot ||\langle \boldsymbol{\Sigma}\rangle_{\Omega_b}||_F}} = 0.02,$$

where $\boldsymbol{\Omega}_r$ and $\boldsymbol{\Omega}_b$ denote the sets of red or blue object locations, respectively. In addition, the conditions of Special Case 1 are satisfied, and thus, the total variation of these sensor observations is zero. In other words, using a single sensor is insufficient to distinguish between the two regions. Conversely, we show that since the SSDs take into account the covariance information between the sensors, they allow us to make this distinction.

In Fig 4C, we present a comparison between the classification results obtained by our approach using SSDs and the classification results obtained by using the output of each sensor as well as the concatenation of the output from all the sensors. The results are evaluated with a 10-fold cross-validation. We observe that the classification obtained by proposed algorithm is significantly better than the classification obtained using the "raw" sensor outputs.

Simulation 3. This simulation demonstrates Proposition 3 from Binary hypothesis testing. Consider $n = 400$ positions, $\{x_i\}_{i=1}^{400}$, such that $x_i \in S^2 \subset \mathbb{R}^3$, which are sampled from two different regions on the sphere, as depicted in the right-top figure in Fig 4A. The rest of the setting remains as described in Simulation 2.

Note that here, only two sensor observations satisfy the conditions of Special Case 1, namely, $\sigma_1^1 = \sigma_1^2$ and $\sigma_7^1 = \sigma_7^2$, and thus, the total variation of only these two sensor observations is zeros.

For each data point, we compute the difference between the left-hand side and the right-hand side of the inequality in Proposition 2 per sensor. Then, in the right-top figure in Fig 4A, we circle the sensors attaining the largest difference. As expected, we observe that the circled sensors are positioned in non-symmetric orientations with respect to the locations of the two regions.

Similarly to the previous simulation, we compare the classification results obtained by our proposed method to the results attained by applying the classification directly to the observations from each sensor separately and to the concatenation of the observations from all the sensors. In addition, we also compute the results of Algorithm 1 applied to all the sensors except Sensor 1 and Sensor 7, which are the least contributing according to the inequality in Proposition 2.

The classification results presented in Fig 4C imply that by removing the least contributing sensors, namely Sensor 1 and Sensor 7, the recomputed SSDs, denoted by $\boldsymbol{\pi}_i^{1,7}$, lead to slightly improved classification accuracy.

Supporting information

S1 Appendix. Detailed derivation of Proposition 1.
(PDF)

S1 Fig. Examples of expression levels of CD4, LAG3, B7H4 and CD20 before and after a median filter. (A): images from Dataset 1 with no filter (top) and after application of median filter (bottom). (B): same as (A) but for Dataset 2.
(TIFF)

S2 Fig. Heatmaps showing the accuracy of the prediction of treatment response obtained by RBF SVM classifiers based on different choices of hyperparameters. The prediction is based on A: the proposed method in Box 1, B: DM, C: HKS and D: WKS. At each panel, the prediction results for Dataset 1 and Dataset 2 are presented.
(TIFF)

S3 Fig. Treatment response prediction based on an ensemble of classifiers. The confusion matrices obtained by combining the RBF SVM classifiers based on different choices of hyperparameters. The combined parameter values are presented at the top. A: Dataset 1. B: Dataset 2.
(TIFF)

S4 Fig. Heatmaps showing the accuracy of the prediction of treatment response obtained by Random Forest classifiers based on different choices of hyperparameters. Same as S2 Fig, but the results are obtained by random forest (RF) classifiers.
(TIFF)

S5 Fig. Heatmaps of the IMC data and the corresponding SSD features. A: Dataset 1. B: Dataset 2. Each heatmap is divided into 2 vertical blocks representing the data collection from `non-responders` and `responders`. Each column in the heatmaps on the left consists of the multi-feature observations at one ROI. The column is composed of observations of $m = 29$ markers, where each marker observation is represented by a vector of size of $b \times b$, which is a column stack representation of the corresponding image patch. Each column in the heatmaps on the right consists of the SSD features of size $m = 29$ at one ROI. The hyperparameters used for extracting the SSD features are presented on the left.
(TIFF)

S6 Fig. Run time analysis. The run time (in seconds) of the proposed method in Box 1 and the three competing methods applied to three different choices of the number of ROIs (patches) N in Dataset 1 and Dataset 2. The run time is computed separately for the two stages of the algorithms. It is based on a Matlab implementation running on a single core 2.2GHz i7 CPU on a Macbook Pro from mid 2015 with 16GB 1600 MHz DDR3 RAM. A: the run time of Stage 1 for Dataset 1. B: the run time of Stage 2 for Dataset 1. C: the run time of Stage 1 for Dataset 2. D: the run time of Stage 2 for Dataset 2.
(TIFF)

S7 Fig. Multi-SSD. We illustrate this multiscale property using a 3D shape from Princeton ModelNet40 database [42]. Suppose the points on the shape are the graph nodes, and compute the SSD of the graph with different values of ϵ, where the ϵ are chosen in logarithmic spacing between $[10^{-3}, 10^{1.5}]$. The color represents the values of SSD with different scales ϵ computed based on data points from a 3D shape of flowers in a vase. The red color represents high values of SSD, and, the blue color represents the low value of SSD. We observe that $\boldsymbol{\pi}_i$ highlights junctions or hubs (in red), as in [43], both at local and global scales, depending on ϵ. We also observe that when ϵ is small, the SSD highlights the neck of each flower. When gradually increasing the value of ϵ, we observe that the SSD transitions toward the center of the shape, representing the global hub of the shape.
(EPS)

S1 Matlab Code. Localization toy examples code. The folder consists of a text file (readme.txt) and three Matlab scripts demonstrating the three simulations in Localization toy problem.
(ZIP)

Author Contributions

Conceptualization: Ya-Wei Eileen Lin, Tal Shnitzer, Ronen Talmon, Kurt Schalper, Yuval Kluger.

Data curation: Franz Villarroel-Espindola, Shruti Desai, Kurt Schalper.

Formal analysis: Ya-Wei Eileen Lin, Tal Shnitzer, Ronen Talmon, Yuval Kluger.

Funding acquisition: Ronen Talmon, Kurt Schalper, Yuval Kluger.

Investigation: Ya-Wei Eileen Lin, Tal Shnitzer, Ronen Talmon, Kurt Schalper, Yuval Kluger.

Methodology: Ya-Wei Eileen Lin, Tal Shnitzer, Ronen Talmon, Kurt Schalper, Yuval Kluger.

Project administration: Ronen Talmon, Yuval Kluger.

Resources: Franz Villarroel-Espindola, Shruti Desai.

Software: Ya-Wei Eileen Lin, Tal Shnitzer.

Supervision: Ronen Talmon, Kurt Schalper, Yuval Kluger.

Validation: Ya-Wei Eileen Lin, Tal Shnitzer, Ronen Talmon, Kurt Schalper, Yuval Kluger.

Visualization: Ya-Wei Eileen Lin, Tal Shnitzer, Ronen Talmon, Yuval Kluger.

Writing – original draft: Ya-Wei Eileen Lin, Tal Shnitzer, Ronen Talmon, Yuval Kluger.

Writing – review & editing: Ya-Wei Eileen Lin, Tal Shnitzer, Ronen Talmon, Kurt Schalper, Yuval Kluger.

References

1. Tenenbaum JB, De Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. science. 2000; 290(5500):2319–2323. https://doi.org/10.1126/science.290.5500.2319 PMID: 11125149
2. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. science. 2000; 290(5500):2323–2326. https://doi.org/10.1126/science.290.5500.2323 PMID: 11125150
3. Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation. 2003; 15(6):1373–1396. https://doi.org/10.1162/089976603321780317
4. Coifman R, Lafon S. Diffusion Maps. Appl Comput Harmon Anal. 2006; 21:5–30. https://doi.org/10.1016/j.acha.2006.04.006 PMID: 17063683
5. Murphy KP. Machine learning: a probabilistic perspective. MIT press; 2012.
6. Chung NC, Storey JD. Statistical significance of variables driving systematic variation in high-dimensional data. Bioinformatics. 2014; 31(4):545–554. https://doi.org/10.1093/bioinformatics/btu674 PMID: 25336500
7. Lederman RR, Talmon R. Learning the geometry of common latent variables using alternating-diffusion. Applied and Computational Harmonic Analysis. 2015;.
8. Talmon R, Wu HT. Latent common manifold learning with alternating diffusion: analysis and applications. Applied and Computational Harmonic Analysis. 2019; 47(3):848–892. https://doi.org/10.1016/j.acha.2017.12.006
9. Shnitzer T, Ben-Chen M, Guibas L, Talmon R, Wu HT. Recovering hidden components in multimodal data with composite diffusion operators. SIAM Journal on Mathematics of Data Science. 2019; 1(3):588–616. https://doi.org/10.1137/18M1218157
10. Katz O, Talmon R, Lo YL, Wu HT. Alternating diffusion maps for multimodal data fusion. Information Fusion. 2019; 45:346–360. https://doi.org/10.1016/j.inffus.2018.01.007
11. Lindenbaum O, Yeredor A, Salhov M, Averbuch A. Multi-view diffusion maps. Information Fusion. 2020; 55:127–149. https://doi.org/10.1016/j.inffus.2019.08.005
12. Eynard D, Kovnatsky A, Bronstein MM, Glashoff K, Bronstein AM. Multimodal manifold analysis by simultaneous diagonalization of laplacians. IEEE transactions on pattern analysis and machine intelligence. 2015; 37(12):2505–2517. https://doi.org/10.1109/TPAMI.2015.2408348 PMID: 26539854

13. Bodenmiller B, Zunder ER, Finck R, Chen TJ, Savig ES, Bruggner RV, et al. Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. Nature biotechnology. 2012; 30(9):858. https://doi.org/10.1038/nbt.2317 PMID: 22902532

14. 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. Nature methods. 2014; 11(4):417. https://doi.org/10.1038/nmeth.2869 PMID: 24584193

15. Chang Q, Ornatsky OI, Siddiqui I, Loboda A, Baranov VI, Hedley DW. Imaging mass cytometry. Cytometry part A. 2017; 91(2):160–169. https://doi.org/10.1002/cyto.a.23053

16. Sun J, Ovsjanikov M, Guibas L. A concise and provably informative multi-scale signature based on heat diffusion. In: Computer graphics forum. vol. 28. Wiley Online Library; 2009. p. 1383–1392.

17. Aubry M, Schlickewei U, Cremers D. The wave kernel signature: A quantum mechanical approach to shape analysis. In: 2011 IEEE international conference on computer vision workshops (ICCV workshops). IEEE; 2011. p. 1626–1633.

18. Krause H. Localization theory for triangulated categories. arXiv preprint arXiv:08061324. 2008;.

19. Lovász L, et al. Random walks on graphs: A survey. Combinatorics, Paul erdos is eighty. 1993; 2(1):1–46.

20. Talmon R, Cohen I, Gannot S. Single-channel transient interference suppression with diffusion maps. IEEE transactions on audio, speech, and language processing. 2012; 21(1):132–144. https://doi.org/10.1109/TASL.2012.2215593

21. Dov D, Talmon R, Cohen I. Audio-visual voice activity detection using diffusion maps. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2015; 23(4):732–745. https://doi.org/10.1109/TASLP.2015.2405481

22. Mishne G, Cohen I. Multiscale anomaly detection using diffusion maps. IEEE Journal of selected topics in signal processing. 2012; 7(1):111–123. https://doi.org/10.1109/JSTSP.2012.2232279

23. Bronstein MM, Bronstein AM. Shape recognition with spectral distances. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010; 33(5):1065–1071. https://doi.org/10.1109/TPAMI.2010.210

24. Cheng X, Mishne G. Spectral Embedding Norm: Looking Deep into the Spectrum of the Graph Laplacian. arXiv preprint arXiv:181010695. 2018;.

25. Devroye L, Mehrabian A, Reddad T. The total variation distance between high-dimensional Gaussians. arXiv preprint arXiv:181008693. 2018;.

26. Baharlou H, Canete NP, Cunningham AL, Harman AN, Patrick E. Mass Cytometry Imaging for the Study of Human Diseases—Applications and Data Analysis Strategies. Frontiers in Immunology. 2019; 10. https://doi.org/10.3389/fimmu.2019.02657 PMID: 31798587

27. Ali HR, Jackson HW, Zanotelli VR, Danenberg E, Fischer JR, Bardwell H, et al. Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer. Nature Cancer. 2020; 1(2):163–175. https://doi.org/10.1038/s43018-020-0026-6

28. Maaten Lvd, Hinton G. Visualizing data using t-SNE. Journal of machine learning research. 2008; 9(Nov):2579–2605.

29. Tax DM, Van Breukelen M, Duin RP, Kittler J. Combining multiple classifiers by averaging or by multiplying? Pattern recognition. 2000; 33(9):1475–1485. https://doi.org/10.1016/S0031-3203(99)00138-7

30. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: Bringing order to the web. Stanford InfoLab; 1999.

31. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science. 2019; 363 (6434):1463–1467. https://doi.org/10.1126/science.aaw1219 PMID: 30923225

32. Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, Schapiro D, et al. High-definition spatial transcriptomics for in situ tissue profiling. Nature methods. 2019; 16(10):987–990. https://doi.org/10.1038/s41592-019-0548-y PMID: 31501547

33. Vickovic S, Eraslan G, Klughammer J, Stenbeck L, Salmen F, Aijo T, et al. High-density spatial transcriptomics arrays for in situ tissue profiling. bioRxiv. 2019; p. 563338.

34. Keren L, Bosse M, Thompson S, Risom T, Vijayaragavan K, McCaffrey E, et al. MIBI-TOF: A multiplexed imaging platform relates cellular phenotypes and tissue structure. Science Advances. 2019; 5(10):eaax5851. https://doi.org/10.1126/sciadv.aax5851 PMID: 31633026

35. Liu Y, Yang M, Deng Y, Su G, Guo C, Zhang D, et al. High-Spatial-Resolution Multi-Omics Atlas Sequencing of Mouse Embryos via Deterministic Barcoding in Tissue. Available at SSRN 3466428. 2019;.

36. Bérard P, Besson G, Gallot S. Embedding Riemannian manifolds by their heat kernel. Geometric & Functional Analysis GAFA. 1994; 4(4):373–398. https://doi.org/10.1007/BF01896401

37. Dunteman GH. Principal components analysis. 69. Sage; 1989.

38. Raviv D, Bronstein MM, Bronstein AM, Kimmel R. Volumetric heat kernel signatures. In: Proceedings of the ACM workshop on 3D object retrieval. ACM; 2010. p. 39–44.

39. Rustamov RM. Laplace-Beltrami eigenfunctions for deformation invariant shape representation. In: Proceedings of the fifth Eurographics symposium on Geometry processing. Eurographics Association; 2007. p. 225–233.

40. Minakshisundaram S, Pleijel Å. Some properties of the eigenfunctions of the Laplace-operator on Riemannian manifolds. Canadian Journal of Mathematics. 1949; 1(3):242–256. https://doi.org/10.4153/CJM-1949-021-5

41. Kleinberg JM. Authoritative sources in a hyperlinked environment. Journal of the ACM (JACM). 1999; 46(5):604–632. https://doi.org/10.1145/324133.324140

42. Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X, et al. 3d shapenets: A deep representation for volumetric shapes. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 1912–1920.

43. Ma G, Lu CT, He L, Philip SY, Ragin AB. Multi-view graph embedding with hub detection for brain network analysis. In: 2017 IEEE International Conference on Data Mining (ICDM). IEEE; 2017. p. 967–972.