picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

xxxxxx
Advance Access Publication Date: Day Month Year
Manuscript Category

Bioimage informatics

MIAMI: Mutual Information-based Analysis of Multiplex Imaging data

Souvik Seal [1,*] and Debashis Ghosh [1,*]

[1] Department of Biostatistics and Informatics, University of Colorado CU Anschutz Medical Campus, Aurora, Colorado

[*] To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Studying the interaction or co-expression of the proteins or markers in the tumor microenvironment (TME) of cancer subjects can be crucial in the assessment of risks, such as death or recurrence. In the conventional approach, the cells need to be declared positive or negative for a marker based on its intensity. For multiple markers, manual thresholds are required for each marker, which can become cumbersome. The performance of the subsequent analysis relies heavily on this step and thus suffers from subjectivity and lacks robustness.
Results: We present a new method where different marker intensities are viewed as dependent random variables, and the mutual information (MI) between them is considered to be a metric of co-expression. Estimation of the joint density, as required in the traditional form of MI, becomes increasingly challenging as the number of markers increases. We consider an alternative formulation of MI which is conceptually similar but has an efficient estimation technique for which we develop a new generalization. With the proposed method, we analyzed a lung cancer dataset finding the co-expression of the markers, HLA-DR and CK to be associated with survival. We also analyzed a triple negative breast cancer dataset finding the co-expression of the immuno-regulatory proteins, PD1, PD-L1, Lag3 and IDO, to be associated with disease recurrence. We demonstrated the robustness of our method through different simulation studies.
Availability: The associated *R* package can be found here, https://github.com/sealx017/MIAMI.
Contact: souvik.seal@cuanschutz.edu
Supplementary information: The Supplementary Material is attached.

1 Introduction

In recent years, multiplex tissue imaging (Bataille *et al.*, 2006) technologies like, imaging mass cytometry (IMC) (Ali *et al.*, 2020), multiplex immunohistochemistry (mIHC) (Tan *et al.*, 2020) and multiplexed ion beam imaging (MIBI) (Angelo *et al.*, 2014) have become increasingly popular for probing single-cell spatial biology. The technologies help in understanding the biological mechanisms underlying cellular and protein interactions in a wide array of scientific contexts. MIBI (Ionpath Inc.) (Angelo *et al.*, 2014) platform and the mIHC platforms such as Vectra 3.0 (Akoya Biosciences) (Huang *et al.*, 2013) and Vectra Polaris (Akoya Biosciences) (Pollan *et al.*, 2020) produce images of similar structure. In particular, each image is two-dimensional, collected at cell- and nucleus-level resolution and proteins in the sample have been labeled with antibodies that attach to cell membranes. We will refer to the antibodies as markers in the paper. Typically, mIHC images have 6-8 markers, whereas MIBI images can have 40 or more markers.

Many of the above markers are surface or phenotypic markers (Zola *et al.*, 2007; Shipkova and Wieland, 2012) which are primarily used for cell type identification. Additionally, there are functional markers (Ijsselsteijn *et al.*, 2019) such as HLA-DR (Saraiva *et al.*, 2018), PD1, PD-L1 (Alsaab *et al.*, 2017) and CD45RO (Lee *et al.*, 2008) that dictate or regulate important cell-functions. Both surface and functional markers are quantified as continuous valued marker intensities. However, in the traditional method, the interaction or co-expression effects of the markers are studied by binarizing them. For every marker, a threshold is chosen to indicate whether a cell (or, a pixel) in the tumor microenvironment (TME) (Binnewies *et al.*, 2018) is positive or negative for that marker. If a cell is positive for two markers, it implies that the markers have co-expressed or co-occurred in that cell. Next, for all the subjects considering every unique pair of markers, the proportion of cells positive for both the markers, the proportion of cells positive for only the first marker and the proportion of cells positive for only the second marker are computed. These proportions can then be used in hiearchical clustering (Murtagh and Legendre, 2014) to group the subjects. It can then be tested if the cluster labels correlate with clinical outcomes, such as disease recurrence or time to death (Koguchi *et al.*, 2015; Johnson *et al.*, 2021; Jackson *et al.*, 2020). The step of binarizing the marker expression profiles can be performed either using compatible commercially available software or

© The Author 2015. All rights reserved. For permissions, please e-mail: souvik.seal@cuanschutz.edu

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

manual assessment of the segmented images. For example, Johnson *et al.* (2018) used AQUAnalysis® software (McCabe *et al.*, 2005; Dolled-Filhart *et al.*, 2010) for binarizing the functional markers, PD1, PDL1, HLA-DR and IDO, and discovered that their co-expression predicted improved outcomes of anti-PD-1 therapies in metastatic melanoma. On the other hand, Patwa *et al.* (2021) used their own clinical expertise and careful evaluation of the segmented images to determine the binarizing thresholds and discovered the interaction between the markers, PD-1, PD-L1, IDO, and Lag3 to be associated with recurrence. It should be pointed out that such expression-thresholding or binarization is also popular in cell-cell communication or interaction analysis (Jin *et al.*, 2021) in the context of single-cell RNA sequencing (scRNA-seq) data. As an example, Armingol *et al.* (2021) defined a pair of cells to be interacting if the expressions of both ligand and receptor (Wong *et al.*, 1997) in those cells exceed certain chosen thresholds.

The manual threshold selection process for multiple markers can be extremely challenging and is subjective. Alternatively, commercially available softwares, such as AQUAnalysis® and inForm (Dolled-Filhart *et al.*, 2010; Kramer *et al.*, 2018), can be used for automated threshold selection in the context of a few particular data types. However, these softwares generally follow a "black box approach" and can be difficult to interpret. On top of these difficulties, many authors have criticized binarizing continuous random variables in general due to the resulting loss of power (Irwin and McClelland, 2003; Altman and Royston, 2006).

In this paper, we propose a threshold-free approach for studying marker co-expression. We treat the marker intensities as continuous random variables and use their marginal and joint probability density functions (PDF's) to construct a metric of co-expression based on mutual information (MI) (Cover and Thomas, 2006). Unlike the correlation coefficient, MI is capable of capturing non-linear patterns of dependence between the markers and is easily extendable for more than two markers. However, as the number of markers increases, computing the joint PDF becomes increasingly challenging which makes the computation of MI infeasible as well. Therefore, we use a slightly different formulation of MI known as Euclidean quadratic mutual information (EQMI) (Principe *et al.*, 2000) which has a similar interpretation but can be computed more efficiently. The computation algorithm is discussed in Principe (2010) with a simpler assumption that we further generalize. The vector of estimated values of the EQMI of all the subjects is tested for association with clinical outcomes. With the proposed method, we analyzed an mIHC lung cancer dataset (Seal *et al.*, 2022) finding that a higher co-expression of the markers, HLA-DR and CK was significantly associated with better five year overall survival. We analyzed a MIBI triple-negative breast cancer (TNBC) dataset (Keren *et al.*, 2018) studying the co-expression of two sets of functional markers, (a) HLA-DR, CD45RO, H3K27me3, H3K9ac and HLA-Class-1, and (b) PD1, PD-L1, Lag3 and IDO, which are also known as immuno-regulatory proteins (IRP's). We found the co-expression of the IRP's to be significantly associated with recurrence. We demonstrated the robustness of our method over the existing approaches through different simulation studies.

2 Materials and Methods

Suppose there are p markers and N subjects with the j-th subject having n_j cells. Let X_{kij} denote the expression of the k-th marker in the i-th cell of j-th subject for $k = 1, 2, \ldots, p$, $i = 1, 2, \ldots, n_j$, and $j = 1, 2, \ldots, N$. Note that we focus on cell-level data in this paper but the framework is readily usable on pixel-level data as well. Let $Y = (Y_1, Y_2, \ldots, Y_N)^T$ denote a subject-level outcome vector and C be an $N \times S$ matrix of subject-level covariates. Next, we discuss the existing and proposed methods and a brief summary of both is provided in Figure 1.

2.1 Traditional thresholding-based approach to study co-expression of the markers

For every marker k, we choose a cut-off t_k to define cell i of subject j as positive for that marker if $X_{kij} > t_k$. The choice of t_k can be guided by prior biological insight or by careful inspection of the marker intensity profile. For example, extreme quantiles like (e.g., the 90th or 95th percentiles) can serve as viable thresholds, as we see in the simulations. However, in most real datasets, it requires user-defined thresholds to obtain an appropriate threshold that leads to meaningful and interpretable conclusion. A cell can be positive for multiple markers. If a cell is positive for a pair of markers, (k_r, k_s), it would imply that these markers have co-expressed in that particular cell. For every subject, compute the proportion of such double-positive cells, denoted by $k_r^+ k_s^+$, the proportion of cells positive for only the first marker, denoted by $k_r^+ k_s^-$, and the proportion of cells positive for only the second marker, denoted by $k_r^- k_s^+$, for $k_r, k_s \in \{1, 2, \ldots, p\}$ and $k_r \neq k_s$. Next, the subjects are classified into two or more groups based on their vectors of proportions using hierarchical clustering. Note that if it is biologically relevant, instead of a pairwise analysis, one can also study multiple markers jointly e.g., with four markers, one can count the cells which are either $k_r^+ k_s^+ k_t^+ k_u^+$ or any of the possible $(2^4 - 1) = 15$ combinations and group the subjects based on these proportions.

Suppose that the subjects are grouped into M clusters. Let $\mathbf{Z} = (Z_1, \ldots, Z_N)^T$ be an $N \times M$ matrix of the cluster labels. Z_j is a vector corresponding to the subject j with $Z_{jm} = 1$ if the j-th subject belongs to group m and 0 otherwise. In most common practices, $M = 2$ is considered. When Y is a continuous outcome, a standard multiple linear regression model with $\mathbf{Z}$ as a predictor can be written as

$$Y = C\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \epsilon,$$

where $\boldsymbol{\beta}, \boldsymbol{\gamma}$ are fixed effects and ϵ is an $N \times 1$ error vector following multivariate normal distribution (MVN) with mean $\mathbf{0}$ and identity covariance matrix $\sigma^2 \mathbb{I}_N$. The null hypothesis, $H_0 : \boldsymbol{\gamma} = \mathbf{0}$, can be tested using the Wald test or likelihood ratio test (LRT) (Gourieroux *et al.*, 1982). Similarly, when Y is a categorical outcome, a logistic or multinomial logistic regression model (Kwak and Clayton-Matthews, 2002) can be considered.

Next, we consider the case of Y being a a right-censored failure time outcome. Let the outcome of the j-th individual be $Y_j = min(T_j, U_j)$, where T_j is the time to event and U_j is the censoring time. Let $\delta_j \equiv I(T_j \leq U_j)$ be the corresponding censoring indicator. Assuming that T_j and U_j are conditionally independent given the covariates for $j = 1, 2, \ldots, N$, the hazard function for the Cox proportional hazards (PH) model (Andersen and Gill, 1982) with fixed effects can be written as,

$$\lambda_j(t|C_j, Z_j) = \lambda_0(t) \exp(C_j^T \boldsymbol{\beta} + Z_j \boldsymbol{\gamma}), \quad j = 1, 2, \ldots, N \qquad (1)$$

where $\lambda_j(t|C_j, Z_j)$ is the hazard of the j-th subject at time t, given the vector of covariates C_j and the cluster label Z_j and $\lambda_0(t)$ is an unspecified baseline hazard at time t. To test the null hypothesis: $H_0 : \boldsymbol{\gamma} = \mathbf{0}$, an LRT (Therneau, 1997) can be considered.

2.2 Proposed Method: Mutual Information based analysis of marker co-expression

2.2.1 Theory of Mutual Information

Mutual information (MI) is an information theoretic measure of dependence between two or more random variables (r.v.'s). In contrast to the linear correlation coefficient, it captures dependences which do not manifest themselves in the covariance (Kraskov *et al.*, 2004). For two random variables, R_1 and R_2 with sample space $\mathcal{D}$, marginal probability

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

density functions (PDFs), f_1, f_2, and joint PDF f_{12}, the MI is defined as the following:

$$\text{MI}(R_1, R_2) = \int\int f_{12}(r_1, r_2) \log \frac{f_{12}(r_1, r_2)}{f_1(r_1) f_2(r_2)} dr_1 dr_2.$$

MI has been used as a tool for feature selection in many different contexts (Yang and Moody, 1999; Liu *et al.*, 2009; Hoque *et al.*, 2014; Song *et al.*, 2021). The measure can be easily generalized for more than two random variables. However, due to the curse of dimensionality (Langrené and Warin, 2019), it becomes extremely difficult to estimate the joint PDF and hence, the MI, as the number of random variables increases.

Xu (1998) and Principe *et al.* (2000) looked into alternative definitions of MI that would remain conceptually similar but can be computed efficiently. They discussed two measures, Euclidean quadratic mutual information (EQMI) and Cauchy-Schwarz quadratic mutual information (CSQMI), defined as below,

$$\begin{aligned}
&\text{EQMI}(R_1, R_2) \\
&= \int\int f_{12}(r_1, r_2)^2 dr_1 dr_2 - 2\int\int f_{12}(r_1, r_2) f_1(r_1) f_2(r_2) dr_1 dr_2 \\
&+ \int\int f_1(r_1)^2 f_2(r_2)^2 dr_1 dr_2, \\
&\text{CSQMI}(R_1, R_2) \\
&= \log \frac{(\int\int f_{12}(r_1, r_2)^2 dr_1 dr_2)(\int\int f_1(r_1)^2 f_2(r_2)^2 dr_1 dr_2)}{(\int\int f_{12}(r_1, r_2) f_1(r_1) f_2(r_2) dr_1 dr_2)}
\end{aligned} \tag{2}$$

It is trivial to verify that $\text{EQMI}(R_1, R_2) \geq 0$ with equality occurring if and only if R_1, R_2 are independent i.e., $f_{12}(r_1, r_2) = f_1(r_1) f_2(r_2)$ for any $r_1, r_2 \in \mathbb{R}$. Using the Cauchy-Schwarz inequality, we have $\text{CSQMI}(R_1, R_2) \geq 0$ with equality happening if and only if R_1, R_2 are independent. Xu (1998) argued that EQMI shares more properties, such as convexity with respect to the PDF's, of the traditional MI compared to CSQMI and thus, is better suited as a dependence measure. It should also be noticed that the form of EQMI is very similar to the distance covariance measure proposed by Székely *et al.* (2007). There is one more generalized measure of dependence, known as kernel canonical correlation analysis (KCCA) (Huang et al., 2006), which does not share forumlaic similarity with EQMI but can potentially serve as an alternative for detecting non-linear dependence patterns. In this paper, we focus on EQMI and propose its usage in the co-expression analysis of the markers in general multiplex imaging datasets used in the study of spatial biology of TME.

2.2.2 Formulation of EQMI

For every subject j, we assume that the expression of marker k is a continuous random variable, denoted by X_{kj}, with sample space $\mathcal{D} = [0, 1]$. X_{kj} is observed in n_j cells as, $X_{k1j}, X_{k2j}, \ldots, X_{kn_jj}$. Suppose there are $p = 2$ markers and for every subject j, denote their joint PDF as $f_{12j}(x_1, x_2)$ and their marginal PDFs as $f_{1j}(x_1)$ and $f_{2j}(x_2)$ respectively. Following Equation (2), the EQMI between the markers 1 and 2 for subject j can be defined as,

$$\text{EQMI}(X_{1j}, X_{2j}) = V_J - 2V_C + V_M, \tag{3}$$

where $V_J = \int\int f_{12j}(x_1, x_2)^2 dx_1 dx_2$, $V_C = \int\int f_{12j}(x_1, x_2) f_{1j}(x_1) f_{2j}(x_2) dx_1 dx_2$, and $V_M = \int\int f_{1j}(x_1)^2 f_{2j}(x_2)^2 dx_1 dx_2$. $\text{EQMI}(X_{1j}, X_{2j})$ can be interpreted as a generalized measure of co-expression of the markers, capable of capturing non-linear dependences. A large value of $\text{EQMI}(X_{1j}, X_{2j})$ will imply that the markers $1, 2$ have significantly co-expressed in the TME of subject j. $\text{EQMI}(X_{1j}, X_{2j})$ is bounded below by 0 for every j but there is no common upper bound for different j's. Therefore, to compare $\text{EQMI}(X_{1j}, X_{2j})$'s across different subjects, we need to appropriately standardize the values so that they lie in the same scale. We define a new measure as, $\text{EQMI}^*(X_{1j}, X_{2j}) = (V_J + V_M)^{-1}(V_J - 2V_C + V_M)$. We observe that $\text{EQMI}^*(X_{1j}, X_{2j})$ lies between $[0, 1]$ because $V_C \geq 0$, and has a similar interpretation as $\text{EQMI}(X_{1j}, X_{2j})$.

It is intuitive to generalize the measure for any $p \geq 2$ markers. For subject j, letting $f_{12\ldots pj}(x_1, x_2, \ldots, x_p)$ to be the joint PDF and $f_{1j}(x_p), f_{2j}(x_p), \ldots, f_{pj}(x_p)$ to be the marginal PDFs, EQMI^* can be defined as,

$$\begin{aligned}
&\text{EQMI}^*(X_{1j}, X_{2j}, \ldots, X_{pj}) = \frac{V_J - 2V_C + V_M}{V_J + V_M} \\
&V_J = \int\int \ldots \int f_{12\ldots pj}(x_1, x_2, \ldots, x_p)^2 dx_1 dx_2 \ldots dx_p \\
&V_C = \int\int \ldots \int f_{12\ldots pj}(x_1, x_2, \ldots, x_p) f_{1j}(x_1) f_{2j}(x_2) \\
&\ldots f_{pj}(x_p) dx_1 dx_2 \ldots dx_p \\
&V_M = \int\int \ldots \int f_{1j}(x_1)^2 f_{2j}(x_2)^2 \ldots f_{pj}(x_p)^2 dx_1 dx_2 \ldots dx_p.
\end{aligned}$$

Notice that to estimate $\text{EQMI}^*(X_{1j}, X_{2j}, \ldots, X_{pj})$ when the true PDFs are unknown, a naive approach would be to estimate the PDFs first by using a kernel density estimation (KDE) approach (Silverman, 1981). Then, we use the estimated PDFs to compute the terms V_J, V_C and V_M via numerical integration (Davis and Rabinowitz, 2007). However, such an approach would be computationally infeasible for a large p and will defeat the purpose of considering EQMI instead of the standard form of MI.

$\text{EQMI}^*(X_{1j}, X_{2j}, \ldots, X_{pj})$ can be estimated efficiently as, $\widehat{\text{EQMI}}^*(X_{1j}, X_{2j}, \ldots, X_{pj}) = (\hat{V}_J + \hat{V}_M)^{-1}(\hat{V}_J - 2\hat{V}_C + \hat{V}_M)$, where

$$\begin{aligned}
&\hat{V}_J = \frac{1}{n_j^2} \sum_{i=1}^{n_j} \sum_{s=1}^{n_j} \prod_{k=1}^{p} \hat{V}_k(i, s) \\
&\hat{V}_C = \frac{1}{n_j} \sum_{i=1}^{n_j} \prod_{k=1}^{p} \hat{V}_k(i); \hat{V}_k(i) = \frac{1}{n_j} \sum_{s=1}^{n_j} \hat{V}_k(i, s), \\
&\hat{V}_M = \prod_{k=1}^{p} \hat{V}_k; \hat{V}_k = \frac{1}{n_j^2} \sum_{i=1}^{n_j} \sum_{s=1}^{n_j} \hat{V}_k(i, s),
\end{aligned} \tag{4}$$

and $\hat{V}_k(i, s) = G_{\sqrt{2}h_k}(X_{kij} - X_{ksj})$. G_h stands for a Gaussian kernel with bandwidth parameter h. This clever way of estimating the terms, $\hat{V}_J$, $\hat{V}_C$ and $\hat{V}_M$ is described in Principe (2010) with a simpler assumption that h_k's are equal for all k i.e., $h_k = h$ for $k = 1, 2, \ldots, p$. We provide the general derivation (i.e., $h_k \neq h_{k'}$, for $k \neq k'$) and other associated details in the Supplemetary material. For choosing the optimal values of h_k's, we generally consider the diagonal multivariate plug-in bandwidth selection procedure (Wand *et al.*, 1994; Chacón and Duong, 2010). However, when p is large ($p > 6$), to avoid computational deadlock we suggest using Silverman's rule of thumb (Silverman, 1981) for choosing h_k's individually.

2.2.3 Using Mutual Information in association analysis

Denote the vector of estimated values of EQMI^* as $\mathbf{E} \equiv (E_1, \ldots, E_N)^T$, where $E_j = \widehat{\text{EQMI}}^*(X_{1j}, X_{2j}, \ldots, X_{pj})$. Our goal is to test if $\mathbf{E}$ is associated with the clinical outcome Y. When Y is a continuous outcome, a standard multiple linear regression model with $\mathbf{E}$ as a predictor can be written as,

$$Y = C\boldsymbol{\beta} + \mathbf{E}\gamma + \epsilon,$$

where $\boldsymbol{\beta}, \gamma$ are fixed effects and ϵ is an $N \times 1$ error vector following multivariate normal distribution (MVN) with mean $\mathbf{0}$ and identity covariance matrix $\sigma^2 \mathbb{I}_N$. After estimating the parameters, the null hypothesis, $H_0 : \gamma = 0$, can be tested using the Wald test. Note that

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

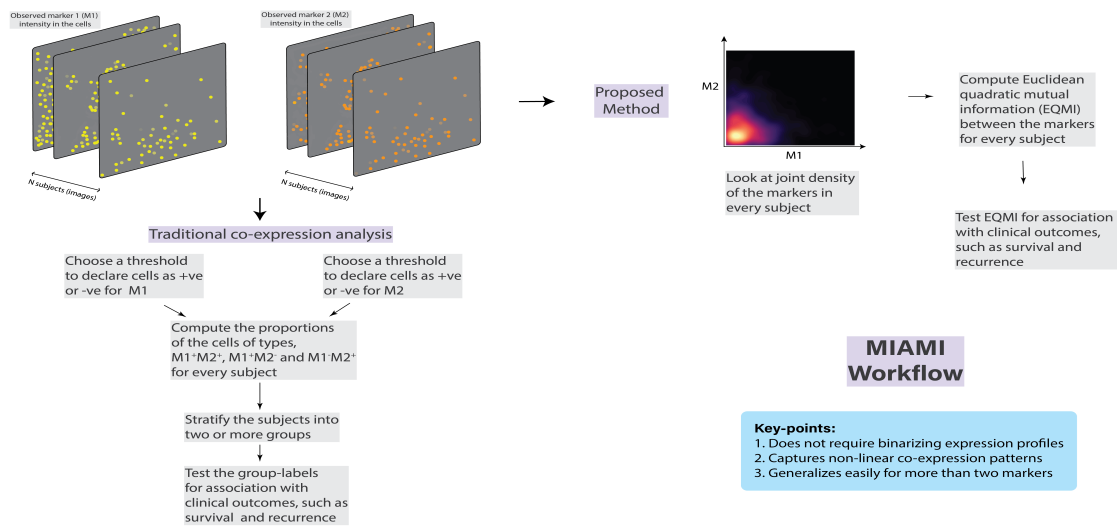

Fig. 1. Comparison of the workflow of the proposed method with the traditional method. We used segmented cell-level data in this paper but the method is applicable on a pixel-level data as well.

we can also add higher order terms of E_j, such as $E_j^2, E_j^3, \ldots$ as in a polynomial regression model (Ostertagová, 2012) to account for non-linear relationship between Y and E_j. Similarly, when Y is a categorical outcome, a logistic or multinomial logistic regression model (Kwak and Clayton-Matthews, 2002) can be considered.

Next, we consider the case of Y being a survival or recurrence outcome. Using the same definitions and conditional independence assumptions of T_j, U_j and covariates as in Section 2.1, the hazard function for the Cox proportional hazards (PH) model can be written as,

$$\lambda_j(t|C_j, E_j) = \lambda_0(t)\exp(C_j^T\beta + E_j\gamma), \quad j = 1, 2, \ldots, N \qquad (5)$$

where $\lambda_j(t|C_j, E_j)$ is the hazard of the j-th subject at time t, given the vector of covariates C_j and the predictor E_j and $\lambda_0(t)$ is an unspecified baseline hazard at time t. To test the null hypothesis: $H_0 : \gamma = 0$, a likelihood ratio test (LRT) can be considered.

In a two-marker scenario (i.e., $p = 2$), one can also treat the absolute value of the Pearson correlation as a measure of marker co-expression and use it as E_j in the earlier equations for testing association with the outcome. However, as we later demonstrate, using correlation instead of EQMI can be sub-optimal in many cases. It is also difficult to generalize to more than two markers.

3 Real Data Analysis

We applied our method on two real datasets, an mIHC lung cancer dataset (Seal *et al.*, 2021) from Vectra 3.0 platform and a MIBI triple-negative breast cancer dataset (Keren *et al.*, 2018). We also applied the traditional thresholding-based method on both the datasets. Since it was hard to decide the optimal thresholds for binarizing the markers, we ran the method for varying values of the thresholds. For every marker, concatenating the intensity data of all the subjects, we computed the median, 95% and 99% quantiles. Next, three different thresholding-based methods using these quantiles were considered, respectively referred to as Median-Thresholding, Threshold 1 and Threshold 2. Note that Threshold 1 and 2 both captured the difference in the tails of the distributions, whereas Median-Thresholding captured the difference in the centers. For the first dataset, we also performed the correlation-based association analysis, referred to as Corr.

3.1 Application to mIHC Lung Cancer data

In the lung cancer dataset, there were 153 subjects each with 3-5 images (in total, 761 images). Two subject-level covariates, age and sex were available. For every subject, the images were non-overlapping and from the same tumor microenvironment (TME) region. After segmenting the images, the subjects had varying number of identified cells (from 3,755 to 16,949). We worked directly with the cell-level data as described next. The cells come from two different tissue regions: tumor and stroma. They were classified into either of the six different cell types: CD14+, CD19+, CD4+, CD8+, CK+ and Other, based on the expression of the phenotypic markers, CD19, CD3, CK, CD8 and CD14. These markers (and the cell-types) usually have clinical meaning e.g, CK is a type of a tumor cell marker and CD4 and CD8 are T-cell or cytotoxic T-cell markers. Apart from these, a functional marker HLA-DR (also known as MHCII), was measured in each of the cells. Johnson *et al.* (2021) classified the subjects into two groups, (a) MHCII: High and (b) MHCII: Low based on the proportion of CK^+ tumor cells that are also positive for HLA-DR (i.e., CK^+HLA-DR^+ cells). They discovered that group (a) had significantly higher five-year overall rate of survival (reported p-value of 0.046).

Note that having a large number of CK+HLADR+ tumor cells implies that these two markers had co-expressed in a lot of the tumor cells. In light of that, we studied if the degree of co-expression of these markers in the tumor cells, as quantified by our method, was associated with the survival. Considering the tumor cells of every subject j, we first estimated EQMI*, $E_j = \widehat{\text{EQMI}}^*(X_{1j}, X_{2j})$ between the two markers, HLA-DR and CK. Next, we tested the association of E_j with five-year overall survival using the Cox-PH model from Equation (5). The coefficient γ was -7.26 with the p-value of the LRT being 0.0286. Thus, subjects with high co-expression of the markers in the tumor cells were more likely to survive. The result was thus consistent with Johnson *et al.* (2021)'s finding. The estimated coefficient, hazard ratio (HR) and p-value of all the methods are listed in Table 1 and further details such as confidence interval of the HR, are provided in the Supplementary Material. The correlation-based analysis (Corr) yielded a negative coefficient estimate but was not statistically significant (at level 0.05). Out of the thresholding-based methods, only Threshold 2 had a significant p-value. It demonstrated that the traditional thresholding-based method could vary heavily based on the choice of the

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

thresholds leading to utterly different conclusions. On the other hand, inference using our method would be robust.

Table 1. Estimated coefficient, hazard ratio (HR) and LRT p-value for testing association with five-year overall survival using different methods in the mIHC lung cancer dataset.

Method	Coefficient (HR)	LRT p-value
EQMI*	-7.26 (0.0007)	0.0286
Corr	-1.38 (0.2512)	0.1838
Median-Thresholding	-0.48 (0.6189)	0.0495
Threshold 1	-0.58 (0.5585)	0.0575
Threshold 2	-1.03 (0.3555)	0.0098

We should point out that CK was a phenotypic marker, and studying its co-expression with a functional marker HLA-DR might not have much clinical relevance. However, the goal was to demonstrate that our method obviates the need to binarize the continuous-valued marker expression profiles and is potentially applicable even when one of the markers is a phenotypic marker.

3.2 Application to TNBC MIBI data

The triple-negative breast cancer (TNBC) MIBI dataset (Keren *et al.*, 2018) had 38 subjects, each with one image. There were 201,656 cells in total with each of the images having varying numbers of cells (between 1217 and 8212). There were 49 markers in total, the majority of which were lineage or phenotypic markers, used primarily for cell-type identification. We were interested in two sets of functional markers, (a) HLA-DR, CD45RO, H3K27me3, H3K9ac and HLA-Class-1, and (b) PD1, PD-L1, Lag3 and IDO, also known as immuno-regulatory proteins (IRP's). The reason for concentrating on these two sets of markers was the findings of Patwa *et al.* (2021). Employing the thresholding-based method with a set of very carefully chosen thresholds, they concluded that the pair-wise co-expression of the functional markers from the sets (a) and (b) were negatively associated with two clinical outcomes, namely disease recurrence and time to death (survival). For set (a), they did not find any statistical significance for either of the clinical outcomes (at level 0.05), whereas, for set (b), they were able to find statistical significance in the association test with recurrence (reported p-value of 0.0058).

For the co-expression analysis with the five markers from the set (a), we looked into all possible ($2^5 - 5 - 1 = 26$) two-way and higher-order combinations of the markers. There were four different types of combination, namely pair (10), triplet (10), quadruplet (5) and quintuple (1). We computed EQMI* for each combination of the markers and tested for association with recurrence and survival. In Table 2, we list the results for five marker combinations for which the lowest p-values were observed. We noticed that at a level of 0.05, several marker-combinations were found to be associated with recurrence. All the estimated coefficients were negative, implying that higher co-expression of the markers decreased the chance of disease recurrence. However, once we adjusted the p-values for multiple testing correction using Bonferroni's method (Bonferroni, 1936), only the combinations "HLA-DR, CD45RO" and "HLA-DR, CD45RO, H3K9ac" remained significant. Note that, we compared the p-values of every type of combination separately, meaning that for marker pairs and triplets, we compared the p-values at level 0.05/10 since the numbers of pairs and triplets were both 10, and for quadruplets, at level 0.05/5 since the number of quadruplets was 5. The marker-combinations were not independent and thus, a Bonferroni correction probably has been overly conservative in this case. For the survival outcome, we did not detect any statistical significance. But, the negative coefficient estimates hinted at a possible association of better rate of survival with higher co-expression.

Table 2. Estimated coefficient and LRT p-value for testing association with recurrence and survival for five combinations of the markers from sets (a) and (b) with the lowest p-values, obtained by the proposed method in the MIBI dataset.

Clinical Outcome	Marker Combination	Coefficient	LRT p-value
Recurrence	HLA-DR, CD45RO	-32.97	0.0022
	HLA-DR, CD45RO, H3K9ac	-12.48	0.0051
	HLA-DR, CD45RO, H3K27me3	-6.73	0.0245
	CD45RO, H3K9ac	-41.53	0.0329
	HLA-DR, CD45RO, HLA-Class-1	-7.20	0.0421
Survival	HLA-Class-1, H3K27me3	-12.75	0.1010
	HLA-DR, H3K9ac	-8.27	0.1542
	HLA-DR, CD45RO	-8.35	0.1583
	HLA-DR, HLA-Class-1	-29.69	0.1630
	HLA-DR, CD45RO, H3K9ac	-0.74	0.3158
Recurrence	PD1, PD-L1	-9.61e+03	0.0046
	PD1, PD-L1, IDO	-5.54e+02	0.0065
	PD1, PD-L1, Lag3, IDO	-4.37e+02	0.0069
	PD-L1, Lag3, IDO	-8.94e+02	0.0084
	PD1, PD-L1, Lag3	-1.91e+02	0.0090
Survival	PD-L1, Lag3	-1.20e+02	0.0103
	Lag3, IDO	-1.14e+02	0.0302
	PD-L1, Lag3, IDO	-5.67e+02	0.0449
	PD-L1, IDO	-6.85e+02	0.0490
	PD1, PD-L1, Lag3, IDO	-2.15e+02	0.0586

For the co-expression analysis with the four IRP's from the set (b), we looked into all possible ($2^4 - 4 - 1 = 11$) two-way and higher order combinations. There were four different types of combination, namely pair (6), triplet (4) and quadruplet (1). In Table 2, we list the results for five marker combinations for which the lowest p-values were observed. At a level of 0.05, many of the marker-combinations were found to be associated with both recurrence and survival. Again, all the estimated coefficients were negative, implying that higher co-expression of the markers decreased the chance of disease recurrence. Upon correcting the p-values for multiple testing using Bonferroni's method, all of the marker combinations listed in Table 2 remained significant for recurrence while for survival, none of them remained significant. We compared the p-values of every type of combination separately, meaning that for marker pairs we compared the p-values at level 0.05/6 since the number of pairs was 6, for triplets we compared the p-values at level 0.05/4 since the number of triplets was 4, and for quadruplet, at level 0.05 since there was just a single quadruplet. Thus, we also arrived at a similar conclusion as (Patwa *et al.*, 2021) that the inter-play or co-expression of the IRP's were significantly associated with recurrence and possibly also with survival. One added novelty of our method was that one could easily pinpoint which of the combinations of the IRP's had the most impact.

It should be kept in mind that the sample-size for this dataset was quite small with only 16 events for recurrence and 15 for survival. It might have affected the overall inference which was based on asymptotic distributional properties of the test statistics. For the same reason, we mainly focused on the sign of the coefficent estimates but not their CI's in this particular case. We also applied the simple thresholding-based methods, Median-Thresholding, Threshold 1 and Threshold 2 as described earlier using both the sets of markers. We found only a single statistically significant result which was for Median-Thresholding using set (b) in the association

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

test with recurrence. Associated tables are provided in the Supplementary Material.

4 Simulation

Next, we compared the performance of the EQMI*-based association analysis with the correlation-based association analysis and the thresholding-based methods in different simulation setups. We assumed that there were two groups of subjects, one with subjects having high marker co-expression and the other with subjects having low or almost zero marker co-expression. In Section 4.1, we considered two markers sharing linear and non-linear patterns of co-expression relationship. In Section 4.2, we considered three markers sharing varying degree of linear co-expression relationship.

4.1 Simulation with two markers

4.1.1 Simulation using Gaussian copula

We replicated the characteristics of the lung cancer dataset in this simulation. The mean marginal distribution of the markers HLA-DR and CK across the subjects could be approximated by Beta distribution respectively with parameters, $(1.5, 170)$ and $(1.6, 35)$ (refer to the Supplementary Material). We used a Gaussian copula (Masarotto and Varin, 2012) to simulate correlated intensity data for two markers which had the above marginal Beta distributions. The simulation strategy was as follows,

1. A random number I_j between 0 and 1 was chosen with probability 0.5 each, respectively standing for group (1), whose subjects had high co-expression of the markers and group (2), whose subjects had mild to none co-expression of the markers. It assigned j-th subject to either of the two groups.
2. The intensity vector of two markers, $(X_{1ij}, X_{2ij})^T$ for every individual j was simulated as follows,
   a. If $I_j = 0$, simulate a correlation parameter ρ_j from Unif$(0.75, 0.9)$, or else simulate ρ_j from Unif$(0, 0.15)$.
   b. Consider a correlation matrix, $R_j = \begin{bmatrix} 1 & \rho_j \\ \rho_j & 1 \end{bmatrix}$ and simulate $(u_{1ij}, u_{2ij})^T \sim N_2(0, R_j)$ for $i = 1, \ldots, n_j$.
   c. Compute $(v_{1ij}, v_{2ij})^T = (\Phi(u_{1ij}), \Phi(u_{2ij}))^T$, where $\Phi()$ denotes the cumulative distribution function (CDF) of the standard normal distribution.
   d. Perform inverse transformation as, $(X_{1ij}, X_{2ij})^T = (F_1^{-1}(v_{1ij}), F_2^{-1}(v_{2ij}))^T$, where F_1 and F_2 are Beta distributions with parameters $(1.5, 170)$ and $(1.6, 35)$. Refer to the Supplementary Material for plots of the true joint densities of the markers for two groups of subjects.
3. The clinical outcome of j-th subject is simulated as, $Y_j = I_j\beta + \epsilon_j$, where $\epsilon_j \sim N(0, \sigma^2)$. This step is repeated 100 times to generate 100 different datasets having different Y vectors but the same intensity vectors, X_1 and X_2. All the methods are applied on these 100 datasets and empirical power is computed.

Steps $1 - 3$ were repeated 20 times and the mean empirical power of different methods were displayed for varying values of the number of cells (n_{cells}) and the number of individuals (N) in Figure 1. The EQMI*-based association analysis (EQMI) and the correlation-based association analysis (Corr) achieved comparable performance in all the cases. This particular simulation strategy inherently assumed that the dependence between the markers was linear. Thus, the estimated values of the EQMI* and the correlation shared an almost one-to-one relationship making the association analysis using either of them equivalent. Median-Tresholding and Threshold 1 showed similar performance, whereas Threshold 2 had consistently lower power. It showed the importance of choosing a proper threshold in the traditional thresholding-based method for achieving a reasonable performance. All the methods expectedly had the least power when N was the smallest, whereas the value of n_{cells} did not have any major impact. It implied if the co-expression pattern was well captured even through a smaller number of cells, most of the methods would perform well.

4.1.2 Simulation with squared marker co-expression relationship

The last simulation strategy essentially assumed a linear pattern of co-expression between the markers. Next, we simulated a non-linear pattern of co-expression between the markers where we would expect the correlation-based association analysis (Corr) to perform worse since it could only capture a linear dependence pattern. Steps 1 and 3 of the last simulation strategy were kept the same here and the marker-intensity simulation step i.e., step 2 was changed as follows,

2a. If $I_j = 0$, simulate X_{1ij} from Unif$(0, 0.1)$, and e_{ij} from Unif$(0, 0.0005)$. Construct X_{2ij} as,

$$X_{2ij} = (X_{1ij} - 0.05)^2 + e_{ij}.$$

2b. If $I_j = 1$, independently simulate x_{1ij}, x_{2ij} both from Unif$(0, 0.1)$.

From Figure 1, we noticed that the EQMI*-based association analysis performed the best in all the cases. Threshold 1 and 2 achieved comparable performance for large value of n_{cells}, whereas Median-Thresholding mostly yielded poor performance. Note that for both the groups of subjects, the correlation between the markers were close to zero as the dependence pattern was non-linear, squared to be specific. Expectedly, the correlation-based association analysis (Corr) had almost no power in every case. The simulation strategy demonstrated why using a generalized measure of co-expression such as EQMI* would be more optimal in many cases.

4.1.3 Simulation with circular marker co-expression relationship

In the last simulation setup, the thresholding-based methods performed well despite the marker co-expression pattern being non-linear. Next, we looked into a slightly more complicated co-expression relationship for which the thresholding-based approach would suffer. Steps 1 and 3 of the last two simulation strategies were kept the same here and the step 2 was changed as follows,

2a. If $I_j = 0$, simulate X_{1ij} from Unif$(0, 0.1)$, e_{ij} from Unif$(0, 0.0005)$ and a random number s_{ij} between -1 and 1. Construct X_{2ij} as,

$$X_{2ij} = 0.05 + s_{ij}\sqrt{0.05^2 - (X_{1ij} - 0.05)^2} + e_{ij}.$$

2b. If $I_j = 1$, independently simulate X_{1ij}, X_{2ij} both from Unif$(0, 0.1)$.

From Figure 1, we noticed that the EQMI*-based association analysis performed the best in all the cases. The correlation-based association analysis (Corr) expectedly performed the worst. Threshold 1, unlike the last simulation, performed significantly worse. In this simulation setup, the subjects of one group had a circular pattern of marker co-expression and the others had almost zero co-expression. Recall that in a two-marker scenario, the thresholding-based methods depended on computing the proportions of the cells positive for both the markers and of the cells positive for only one of the markers (Section 2.1). The difference between these proportions

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

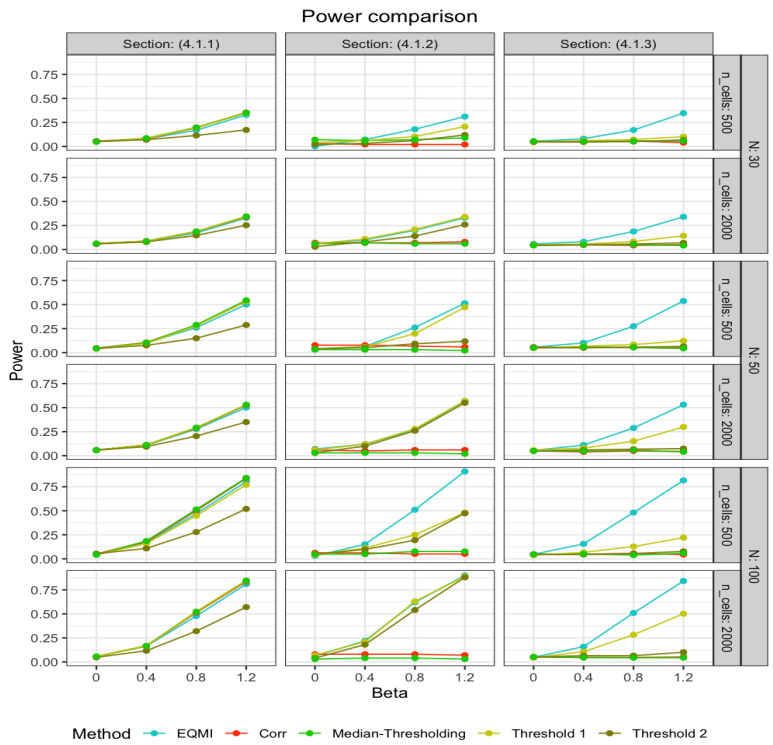

Fig. 2. The figure displays the power of different methods under different simulation scenarios from Section 4.1 with two markers for varying numbers of subjects (N) and cells (n_{cells}). On the x-axis, the fixed effect size β was varied from low to high.

across the two groups of subjects became negligible under this setup which made it difficult distinguishing between them. This explained the overall poor performance of all the thresholding-based methods.

4.2 Simulation with three markers

Next, we considered three markers and simulated varied degree of linear dependence between them using Gaussian copula. We only performed the EQMI*-based association analysis and the thresholding-based methods in this case. There were again two groups of subjects respectively with high and low co-expression. The simulation strategy was as follows,

1. A random number I_j between 0 and 1 was chosen with probability 0.5 each, respectively standing for groups (1) and (2).
2. The intensity vector of three markers, $(X_{1ij}, X_{2ij}, X_{3ij})^T$ for every individual j was simulated as follows,

   (i) Consider a correlation matrix, $R_j = \begin{bmatrix} 1 & \rho_{12j} & \rho_{13j} \\ \rho_{12j} & 1 & \rho_{23j} \\ \rho_{13j} & \rho_{23j} & 1 \end{bmatrix}$. For the subjects in group (1), the off-diagonal elements of the correlation matrix would have high values, whereas they would have low values for the subjects in group (2). We considered three different cases with varying differences between the correlation matrices of the two groups.

   Case (a) : If $I_j = 0$, the correlation parameters ρ_{12j}, ρ_{23j} and ρ_{13j} were independently simulated from Unif$(0.4, 0.6)$. Otherwise, they were independently simulated from Unif$(0.2, 0.4)$.

   Case (b) : Regardless of the value of I_j, ρ_{13j} was kept to be 0. If $I_j = 0$, ρ_{12j} and ρ_{23j} were independently simulated from Unif$(0.4, 0.6)$. Otherwise, these two were independently simulated from Unif$(0.2, 0.4)$.

   Case (c) : Regardless of the value of I_j, ρ_{13j} and ρ_{23j} were kept to be 0. If $I_j = 0$, ρ_{12j} was simulated from Unif$(0.4, 0.6)$. Otherwise, it was simulated from Unif$(0.2, 0.4)$.

   (ii) Simulate $(u_{1ij}, u_{2ij}, u_{3ij})^T \sim N_3(0, R_j)$ and compute $(v_{1ij}, v_{2ij}, v_{3ij})^T = (\Phi(u_{1ij}), \Phi(u_{2ij}), \Phi(u_{3ij}))^T$.

   (iv) Perform inverse transformation as, $(X_{1ij}, X_{2ij}, X_{3ij})^T = (F_1^{-1}(v_{1ij}), F_2^{-1}(v_{2ij}), F_3^{-1}(v_{3ij}))^T$, where F_1, F_2 and F_3 are the Beta distributions respectively with parameters $(1.5, 170)$, $(1.6, 35)$ and $(1.6, 35)$.

3. The clinical outcome of j-th subject was simulated as, $Y_j = I_j\beta + \epsilon_j$, where $\epsilon_j \sim N(0, \sigma^2)$. This step was repeated 100 times to generate 100 different datasets having different Y vectors but the same intensity data, X_1, X_2 and X_3. All the methods were applied on these 100 datasets and empirical power was computed.

Steps $1-3$ were repeated 20 times and in Figure 2 the mean empirical power of the methods were displayed. The power of the methods were quite low when N was small. The EQMI*-based method outperformed both the thresholding-based methods, Threshold 1 and 2 in every case. Threshold 2 had little to no power in most of the cases. Note that, the cases (a), (b), and (c) differed in how different the marker co-expression pattern of the two groups were. The difference between the marker co-expression pattern of the two groups was the largest in case (a) since all the three correlation parameters, ρ_{12j}, ρ_{23j} and ρ_{13j} were different across the groups. The difference was the smallest in case (c) as two of the three correlation parameters, ρ_{23j} and ρ_{13j} were kept to be 0 in both the groups. Quite expectedly, the power of the methods decreased going from case (a) to case (c), as the difference between the groups of subjects reduced. The decrease was more prominent with Threshold 1 suggesting the method's lack of robustness.

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

5 Discussion

In multiplex imaging data, studying the interaction or co-expression of multiple functional markers in the cells of the tumor microenvironment (TME) can be crucial for subject-specific assessment of risks. The traditional approach requires a complex step of binarizing the continuous valued marker expression profiles which is prone to subjectivity and can be sub-optimal in many scenarios. The complexity gets exacerbated as the number of markers increases. In this paper, we propose a method for studying co-expression or co-expression of multiple markers based on the theory of mutual information (MI). We treat the subject-specific intensity or expression of every marker as a continuous random variable. We determine how much the markers have co-expressed in the TME of a particular subject by computing a measure known as Euclidean quadratic mutual information (EQMI), comparing the estimated marginal and joint probability density functions (PDFs) of the markers. The formula of EQMI has a similar interpretation as the standard formula of MI but allows a more efficient computation. We adopt and generalize an existing algorithm for computing EQMI that does not require explicitly estimating the joint PDF of the markers, a step which becomes increasingly intractable as the number of markers increases. Next, the subject-level EQMI values are tested for association with the clinical outcomes. The proposed method is free from the subjectivity bias of the traditional thresholding-based method and is readily applicable with any number of markers.

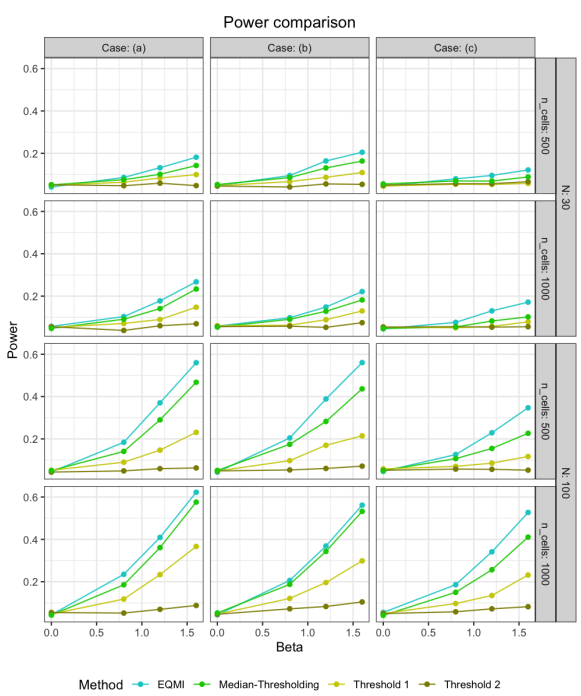

Fig. 3. The figure displays the power of different methods under different cases from Section 4.2 with three markers for varying numbers of subjects (N) and cells (n_{cells}). On the x-axis, the fixed effect size β was varied from low to high.

We applied the proposed method to two real datasets, one mIHC lung cancer dataset and one MIBI triple negative breast cancer dataset. In the former, we found high co-expression of the markers, HLA-DR and CK to be associated with the five-year overall survival of the subjects. In the latter, we found high co-expression of the immuno-regulatory proteins, PD1, PD-L1, IDO and Lag3 (IRP's) to be associated with disease recurrence. We evaluated the performance of our method through several simulation scenarios with two and three markers. In the scenarios with two markers, we showed that all the methods perform well and close to each other if the pattern of dependence (co-expression) between the markers is linear. However, with a more complex non-linear dependence pattern, only the proposed method could achieve respectable power. In the scenarios with three markers, we found that the proposed method performed consistently better than the thresholding-based method and showed superior robustness.

As we have shown in the simulation studies, EQMI can capture both linear and non-linear patterns of co-expression between the markers very well. However, the measure is not well suited for capturing the differences between the patterns. For example, it may happen that one subject has a linear pattern of co-expression, whereas some other subject has a non-linear pattern. The EQMI for both the subjects can be very similar, making it hard to distinguish between them. As a part of our future direction, we would like to improve the method by detecting and incorporating the type of the co-expression pattern. With more than two markers, we studied the co-expression patterns of all possible combinations of the markers and declared significance based on p-values corrected by Bonferroni's method. However, in future, we would like to explore the causal direction between the markers which can then be used to determine a smaller and optimal set of markers and would obviate the need of exploring all possible marker-combinations. In this paper, we have not used any information on the spatial locations of the TME cells. As a future direction, we would like to study the MI between the spatial information and the marker expression profiles with a goal to detect spatially variable markers and their spatial patterns.

Our method is available as an *R* package named MIAMI at this link, https://github.com/sealx017/MIAMI. The package is readily applicable to any multiplex imaging dataset which has cell-level intensity data on two or more markers. In future, we would like to further augment the package's capability by incorporating a pixel-level analysis as well.

Acknowledgments

S. Seal acknowledges funding from the Grohne-Stepp Endowment from the University of Colorado Cancer Center, NCI R01 CA129102 and NSF DMS 1914937. D. Ghosh acknowledges funding from NCI R01 CA129102 and NSF DMS 1914937.

Funding information

This work was supported by the National Cancer Institute [R01 CA129102]; and the National Science Foundation [DMS 1914937].

References

Ali, H. R., Jackson, H. W., Zanotelli, V. R., Danenberg, E., Fischer, J. R., Bardwell, H., Provenzano, E., Rueda, O. M., Chin, S.-F., Aparicio, S., *et al.* (2020). Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer. *Nature Cancer*, **1**(2), 163–175.

Alsaab, H. O., Sau, S., Alzhrani, R., Tatiparti, K., Bhise, K., Kashaw, S. K., and Iyer, A. K. (2017). Pd-1 and pd-l1 checkpoint signaling inhibition for cancer immunotherapy: mechanism, combinations, and clinical outcome. *Frontiers in pharmacology*, **8**, 561.

Altman, D. G. and Royston, P. (2006). The cost of dichotomising continuous variables. *Bmj*, **332**(7549), 1080.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, pages 1100–1120.

Angelo, M., Bendall, S. C., Finck, R., Hale, M. B., Hitzman, C., Borowsky, A. D., Levenson, R. M., Lowe, J. B., Liu, S. D., Zhao, S., *et al.* (2014). Multiplexed ion beam imaging of human breast tumors. *Nature Medicine*, **20**(4), 436–442.

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

Armingol, E., Officer, A., Harismendy, O., and Lewis, N. E. (2021). Deciphering cell–cell interactions and communication from gene expression. *Nature Reviews Genetics*, **22**(2), 71–88.

Bataille, F., Troppmann, S., Klebl, F., Rogler, G., Stoelcker, B., Hofstadter, F., Bosserhoff, A.-K., and Rümmele, P. (2006). Multiparameter immunofluorescence on paraffin-embedded tissue sections. *Applied Immunohistochemistry & Molecular Morphology*, **14**(2), 225–228.

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., Coussens, L. M., Gabrilovich, D. I., Ostrand-Rosenberg, S., Hedrick, C. C., *et al.* (2018). Understanding the tumor immune microenvironment (time) for effective therapy. *Nature Medicine*, **24**(5), 541–550.

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, **8**, 3–62.

Chacón, J. E. and Duong, T. (2010). Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *Test*, **19**(2), 375–398.

Cover, T. M. and Thomas, J. A. (2006). Elements of information theory 2nd edition (wiley series in telecommunications and signal processing).

Davis, P. J. and Rabinowitz, P. (2007). *Methods of numerical integration*. Courier Corporation.

Dolled-Filhart, M., Gustavson, M., Camp, R. L., Rimm, D. L., Tonkinson, J. L., and Christiansen, J. (2010). Automated analysis of tissue microarrays. In *Tissue Microarrays*, pages 151–162. Springer.

Gourieroux, C., Holly, A., and Monfort, A. (1982). Likelihood ratio test, wald test, and kuhn-tucker test in linear models with inequality constraints on the regression parameters. *Econometrica: journal of the Econometric Society*, pages 63–80.

Hoque, N., Bhattacharyya, D. K., and Kalita, J. K. (2014). Mifs-nd: A mutual information-based feature selection method. *Expert Systems with Applications*, **41**(14), 6371–6385.

Huang, W., Hennrick, K., and Drew, S. (2013). A colorful future of quantitative pathology: validation of vectra technology using chromogenic multiplexed immunohistochemistry and prostate tissue microarrays. *Human Pathology*, **44**(1), 29–38.

Ijsselsteijn, M. E., van der Breggen, R., Farina Sarasqueta, A., Koning, F., and de Miranda, N. F. (2019). A 40-marker panel for high dimensional characterization of cancer immune microenvironments by imaging mass cytometry. *Frontiers in immunology*, page 2534.

Irwin, J. R. and McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, **40**(3), 366–371.

Jackson, H. W., Fischer, J. R., *et al.* (2020). The single-cell pathology landscape of breast cancer. *Nature*, **578**(7796), 615–620.

Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M. V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using cellchat. *Nature Communications*, **12**(1), 1–20.

Johnson, A. M., Boland, *et al.* (2021). Cancer cell-specific mhcii expression as a determinant of the immune infiltrate organization and function in the non-small cell lung cancer tumor microenvironment. *Journal of Thoracic Oncology*.

Johnson, D. B., Bordeaux, J., Kim, J. Y., Vaupel, C., Rimm, D. L., Ho, T. H., Joseph, R. W., Daud, A. I., Conry, R. M., Gaughan, E. M., *et al.* (2018). Quantitative spatial profiling of pd-1/pd-l1 interaction and hla-dr/ido-1 predicts improved outcomes of anti–pd-1 therapies in metastatic melanoma. *Clinical Cancer Research*, **24**(21), 5250–5260.

Keren, L., Bosse, M., Marquez, D., Angoshtari, R., Jain, S., Varma, S., Yang, S.-R., Kurian, A., Van Valen, D., West, R., *et al.* (2018). A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell*, **174**(6), 1373–1387.

Koguchi, Y., Hoen, H. M., Bambina, S. A., Rynning, M. D., Fuerstenberg, R. K., Curti, B. D., Urba, W. J., Milburn, C., Bahjat, F. R., Korman, A. J., *et al.* (2015). Serum immunoregulatory proteins as predictors of overall survival of metastatic melanoma patients treated with ipilimumab. *Cancer Research*, **75**(23), 5084–5092.

Kramer, A. S., Latham, B., Diepeveen, L. A., Mou, L., Laurent, G. J., Elsegood, C., Ochoa-Callejero, L., and Yeoh, G. C. (2018). Inform software: a semi-automated research tool to identify presumptive human hepatic progenitor cells, and other histological features of pathological significance. *Scientific reports*, **8**(1), 1–10.

Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Physical Review E*, **69**(6), 066138.

Kwak, C. and Clayton-Matthews, A. (2002). Multinomial logistic regression. *Nursing research*, **51**(6), 404–410.

Langrené, N. and Warin, X. (2019). Fast and stable multivariate kernel density estimation by fast sum updating. *Journal of Computational and Graphical Statistics*, **28**(3), 596–608.

Lee, H., Chae, S., Lee, Y., Kim, M., Lee, H., Lee, B., and Kim, W. (2008). Prognostic implications of type and density of tumour-infiltrating lymphocytes in gastric cancer. *British Journal of Cancer*, **99**(10), 1704–1711.

Liu, H., Sun, J., Liu, L., and Zhang, H. (2009). Feature selection with dynamic mutual information. *Pattern Recognition*, **42**(7), 1330–1339.

Masarotto, G. and Varin, C. (2012). Gaussian copula marginal regression. *Electronic Journal of Statistics*, **6**, 1517–1549.

McCabe, A., Dolled-Filhart, M., Camp, R. L., and Rimm, D. L. (2005). Automated quantitative analysis (aqua) of in situ protein expression, antibody concentration, and prognosis. *Journal of the National Cancer Institute*, **97**(24), 1808–1815.

Murtagh, F. and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *Journal of Classification*, **31**(3), 274–295.

Ostertagová, E. (2012). Modelling using polynomial regression. *Procedia Engineering*, **48**, 500–506.

Patwa, A., Yamashita, R., Long, J., Risom, T., Angelo, M., Keren, L., and Rubin, D. L. (2021). Multiplexed imaging analysis of the tumor-immune microenvironment reveals predictors of outcome in triple-negative breast cancer. *Communications Biology*, **4**(1), 1–14.

Pollan, S., Hanifi, A., Nagy, M., Stavrou, N., Parnell, E., Gozo, M., Attanasio, N., William, J., and Au, Q. (2020). Profiling exhausted t cells using vectra® polaris™ multiplex immunofluorescence assay in hnscc.

Principe, J. C. (2010). *Information theoretic learning: Renyi's entropy and kernel perspectives*. Springer Science & Business Media.

Principe, J. C., Xu, D., Zhao, Q., and Fisher, J. W. (2000). Learning from examples with information theoretic criteria. *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, **26**(1), 61–77.

Saraiva, D. P., Jacinto, A., Borralho, P., Braga, S., and Cabral, M. G. (2018). Hla-dr in cytotoxic t lymphocytes predicts breast cancer patients' response to neoadjuvant chemotherapy. *Frontiers in Immunology*, **9**, 2605.

Seal, S., Wrobel, J., Johnson, A. M., Nemenoff, R. A., Schenk, E. L., Bitler, B. G., Jordan, K. R., and Ghosh, D. (2021). On clustering for cell phenotyping in multiplex immunohistochemistry (mihc) and multiplexed ion beam imaging (mibi) data.

Seal, S., Vu, T., Ghosh, T., Wrobel, J., and Ghosh, D. (2022). Denvar: Density-based variation analysis of multiplex imaging data. *Bioinformatics Advances*, **2**(1), vbac039.

Shipkova, M. and Wieland, E. (2012). Surface markers of lymphocyte activation and markers of cell proliferation. *Clinica Chimica Acta*, **413**(17-18), 1338–1349.

Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society: Series B (Methodological)*, **43**(1), 97–99.

Song, X.-f., Zhang, Y., Gong, D.-w., and Sun, X.-y. (2021). Feature selection using bare-bones particle swarm optimization with mutual information. *Pattern Recognition*, **112**, 107804.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, **35**(6), 2769–2794.

Tan, W. C. C., Nerurkar, S. N., Cai, H. Y., Ng, H. H. M., Wu, D., Wee, Y. T. F., Lim, J. C. T., Yeong, J., and Lim, T. K. H. (2020). Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. *Cancer Communications*, **40**(4), 135–153.

Therneau, T. M. (1997). Extending the cox model. In *Proceedings of the First Seattle Symposium in Biostatistics*, pages 51–84. Springer.

Wand, M. P., Jones, M. C., *et al.* (1994). Multivariate plug-in bandwidth selection. *Computational Statistics*, **9**(2), 97–116.

Wong, J. Y., Kuhl, T. L., Israelachvili, J. N., Mullah, N., and Zalipsky, S. (1997). Direct measurement of a tethered ligand-receptor interaction potential. *Science*, **275**(5301), 820–822.

Xu, D. (1998). *Energy, entropy and information potential for neural computation*. University of Florida.

Yang, H. and Moody, J. (1999). Feature selection based on joint mutual information. In *Proceedings of International ICSC Symposium on Advances in Intelligent Data Analysis*, volume 23. Citeseer.

Zola, H., Swart, B., Nicholson, I., and Voss, E. (2007). *Leukocyte and stromal cell molecules: the CD markers*. John Wiley & Sons.

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture