Tumor radiogenomics in gliomas with Bayesian layered variable selection

Shariq Mohammed [a,b,c,*], Sebastian Kurtek [d], Karthik Bharath [e], Arvind Rao [b,c,f], Veerabhadran Baladandayuthapani [b]

[a] *Department of Biostatistics, Boston University, 801 Massachusetts Ave, Boston, MA 02118, United States*
[b] *Department of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48103, United States*
[c] *Department of Computational Medicine and Bioinformatics, University of Michigan, 100 Washtenaw Avenue, Ann Arbor, MI 48109, United States*
[d] *Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH 43210, United States*
[e] *School of Mathematical Sciences, University Park, Nottingham, NG7 2RD, United Kingdom*
[f] *Department of Radiation Oncology, University of Michigan, 1500 E Medical Center Dr, Ann Arbor, MI 48109, United States*

ARTICLE INFO

ABSTRACT

We propose a statistical framework to analyze radiological magnetic resonance imaging (MRI) and genomic data to identify the underlying radiogenomic associations in lower grade gliomas (LGG). We devise a novel imaging phenotype by dividing the tumor region into concentric spherical layers that mimics the tumor evolution process. MRI data within each layer is represented by voxel–intensity-based probability density functions which capture the complete information about tumor heterogeneity. Under a Riemannian-geometric framework these densities are mapped to a vector of principal component scores which act as imaging phenotypes. Subsequently, we build Bayesian variable selection models for each layer with the imaging phenotypes as the response and the genomic markers as predictors. Our novel hierarchical prior formulation incorporates the interior-to-exterior structure of the layers, and the correlation between the genomic markers. We employ a computationally-efficient Expectation–Maximization-based strategy for estimation. Simulation studies demonstrate the superior performance of our approach compared to other approaches. With a focus on the cancer driver genes in LGG, we discuss some biologically relevant findings. Genes implicated with survival and oncogenesis are identified as being associated with the spherical layers, which could potentially serve as early-stage diagnostic markers for disease monitoring, prior to routine invasive approaches. We provide a R package that can be used to deploy our framework to identify radiogenomic associations.

1. Introduction

Lower-grade gliomas (LGG) are infiltrative brain neoplasms characterized as World Health Organization (WHO) grade II and III neoplasms. LGG is a uniformly fatal disease of young adults (mean age: 41 years), with survival times averaging approximately seven years (Claus et al., 2015). LGG patients usually have better survival than patients with high-grade (WHO grade IV) gliomas. However, some of the LGG tumors recur after treatment and progress into high-grade tumors, making it essential to understand the underlying etiology, and to improve the treatment management and monitoring for LGG patients.

Due to recent technological advances, large multi-modal datasets are being produced; these include imaging as well as multi-platform genomics data. In complex disease systems such as cancer, integrative analyses of such data can reveal important biological insights into specific disease mechanisms and its subsequent clinical translation (Wang et al., 2013; Morris and Baladandayuthapani, 2017). Genomic profiling technologies, such as microarrays, next-generation sequencing, methylation arrays and proteomic analyses, have facilitated thorough investigations at the molecular level. Such multi-platform genomic data resources have been used to develop models to better understand the molecular characterization across cancers and especially in gliomas (Ohgaki et al., 2004; Ohgaki and Kleihues, 2007). While genomic data provides information on the molecular characterization of the disease, radiological imaging, such as magnetic resonance imaging (MRI), computed tomography (CT) and positron emission tomography (PET), provide complementary information about the structural aspects of the disease. In this context, *radiomic* analysis involves mining and extraction of various types of quantitative imaging features from different modalities obtained through high-throughput radiology images (Bakas et al., 2017b). These image-derived features describe various characteristics such as morphology and texture, among others. Integrative

analyses of genomic and radiomic features that uncover underlying associations between them, commonly referred to as *radiogenomic* analyses, capture complementary characteristics of the tumor (Gevaert et al., 2014; Mazurowski et al., 2017). However, such integrative modeling approaches present multiple analytical and computational challenges including incorporating complex biological structure, and high-dimensionality of quantitative imaging/genomic markers, necessitating principled and biologically-informed dimension reduction and information extraction techniques.

Existing radiomic and radiogenomic studies typically consider voxel-level summary features (based on intensity histogram, geometry, and texture analyses) as representations of a tumor image, to assess the progression (or regression) of tumors using statistical or machine learning models (Lu et al., 2018; Elshafeey et al., 2019). The main drawbacks of using voxel-intensity summary statistics/features are: (i) they might not capture the entire information in the voxel intensity histogram; (ii) there is subjectivity in the choice of the number of features and the location where they are computed. Consequently, the radiomic/radiogenomic analyses using summary features are unable to detect potential small-scale and sensitive changes in the tumor (Just, 2014). To address these challenges, we propose a *de novo* strategy for quantifying the tumor heterogeneity using voxel-intensity based imaging phenotypes that mimic the tumor evolution process.

1.1. De novo quantification of imaging phenotypes

Tumors typically evolve from a single (or group of) cancerous cell(s) and grow multiplicatively via cross-talk with other nearby cells, thereby producing a central necrosis region (dead cells) (Swanson et al., 2003). Hence, as we move from the inner core of a tumor toward the exterior, different layers of the tumor possess distinct tissue characteristics. Therefore, it is important to construct imaging phenotypes which (a) mimic the underlying tumor evolution process, and (b) efficiently harness the tumor heterogeneity from the imaging scans, for downstream modeling and interpretations. To address (a), we divide the tumor region into one inner sphere followed by concentric spherical shells which potentially capture the tumor growth process. In Fig. 1, we show a visualization of an example tumor region divided into three spherical layers. Tumor spheroids (cell cultures with necrotic core and peripheral layer of proliferative cells) have been used as reliable models of in vitro solid tumors (Breslin and O'Driscoll, 2013; Costa et al., 2016). To address (b), we propose to work with the probability density functions (smoothed histograms) constructed using layer-wise tumor voxel intensities from three-dimensional (3D) MRI scans. These probability density functions (PDFs) are the data objects of interest and capture the entire information about the distribution of voxel intensities, no longer requiring any summaries of the distribution. Recent studies have successfully utilized density-based approaches in unsupervised clustering (Saha et al., 2016) and regression models for radiogenomic analyses (Mohammed et al., 2021; Yang et al., 2020).

Briefly, we propose a novel approach to identify the genomic markers that have significant associations with the PDF-based radiomic phenotypes. The imaging phenotype we consider arises from MRI, as it provides high-resolution images with a wide range of contrasts through multiple imaging sequences (see Section 2). The radiomic phenotypes are constructed by considering the aforementioned topological structure of the tumor and mimicking its evolution process. Our statistical framework identifies significant genomic associations on the different concentric spherical shells. These associations better inform the complex interplay between the molecular signatures and imaging characteristics in LGGs, and can provide effective non-invasive diagnostic options by monitoring the associated genomic markers prior to invasive alternatives such as biopsy. A schematic representation of our approach is presented in Fig. 2 (see Section 2 for details). This figure depicts an end-to-end workflow, which includes (i) constructing spherical layers from the tumor voxels in MRI scans and computing
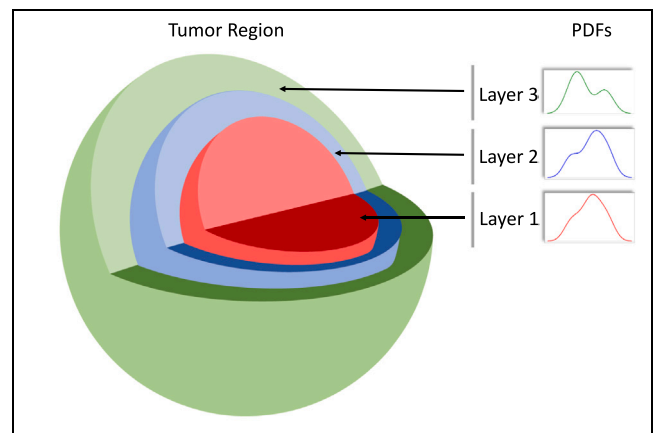

Fig. 1. A graphic to visualize an example tumor region divided into three spherical layers. The object on the left shows three spherical layers colored in shades of red, blue and green. For each of these layers, the curves on the right depict corresponding PDFs.

the corresponding imaging phenotypes, and (ii) building a statistical framework which includes the structural information from the data to identify radiogenomic associations, elements of which are summarized below.

Constructing one global PDF for each tumor disregards structural information about tissue characteristics of the tumor. This would prohibit us from analyzing any potential associations of genomic markers with local (regional) aspects of the tumor. We propose to evaluate associations between genomic signatures of the tumor through large (or subtle) changes in the overall shape of the voxel-intensity PDFs from biologically motivated tumor compartments/layers. We address this by constructing τ PDFs corresponding to τ tumor layers. Tumors for different subjects will have distinct characteristics of these layers, i.e., not all tumors might have structural information about all of the layers. However, tumor spheroids have been used as in vitro models of solid tumors (Breslin and O'Driscoll, 2013; Costa et al., 2016). Hence, we divide the tumor region into τ spherical shells to mimic the tumor growth process. The number of layers the tumor is divided into remains the same across subjects. We effectively have τ voxel-intensity PDFs as $x \mapsto (f_1(x), \ldots, f_\tau(x))$, a multivariate functional data object with each component as a PDF. The space of PDFs is a non-linear, infinite-dimensional manifold, which poses significant challenges in their analysis. We use the Riemannian-geometric framework based on the Fisher–Rao metric on the space of PDFs (Section 2) that (i) leads to a tractable representation of each PDF, and (ii) engenders dimension reduction of the functional objects, in a manner compatible with the requirements of downstream statistical analyses. This process yields a set of imaging phenotypes based on PDFs corresponding to each spherical layer in each MRI sequence. The natural dependence between the component functions is incorporated through a novel Bayesian prior formulation (Section 3).

1.2. Statistical framework

We build sequential regression models for each spherical layer starting from the inner-most sphere and moving outwards towards the edge of the tumor. The sets of imaging phenotypes from a spherical layer across multiple MRI sequences are then used as the multivariate response in the regression corresponding to that spherical layer, with the genomic markers as the corresponding predictors. This specific structured form of the imaging phenotypes as well as assessing their interaction with genomic covariates leads to the following challenges:
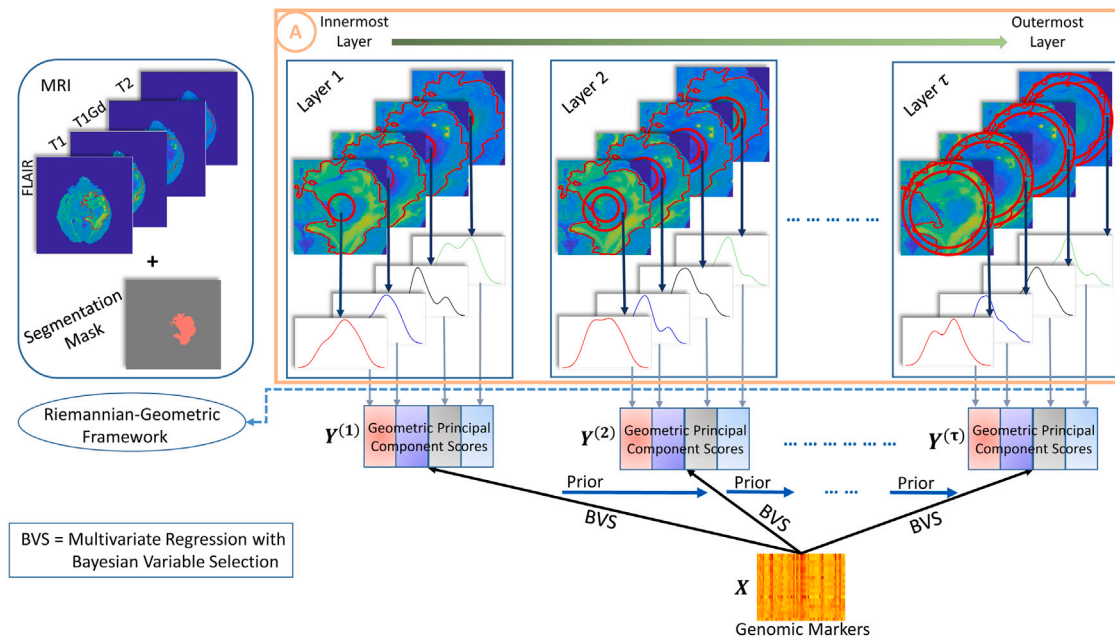
Fig. 2. Schematic representation of the proposed modeling approach. The tumor voxels are divided into spherical layers across all of the MRI sequences. Under a Riemannian-geometric framework, we construct principal component (PC) scores for each spherical layer across all four MRI sequences. With these PC scores as the multivariate response, and the genomic markers as predictors, we build a layer-wise Bayesian variable selection model. The estimation occurs sequentially, allowing for borrowing of information between layers through prior specification.

1. While handling high-resolution imaging data, the number of parameters to estimate is usually much higher than the number of subjects, leading to the curse of dimensionality ($p \gg n$).

2. Splitting the tumor voxels into concentric spherical layers induces a natural sequential ordering imitating tumor growth, which needs to be incorporated into the model formulation. Additionally, correlation between genomic markers needs to be accounted for as well.

3. High-dimensionality and sequential structure in imaging data, and correlation structure in genomic markers, create severe computational hurdles warranting scalable approaches.

In this paper, we build a novel statistical framework in a multiple-multivariate regression set-up, which addresses the above mentioned challenges incorporating the aforementioned dependence structures within and between the responses and covariates. Specifically, we leverage the hierarchical model specification to incorporate the sequential nature of the spherical layer-based imaging phenotypes as well as the correlation between the genomic markers. We propose a Bayesian variable selection approach that allows us to incorporate prior structural information in a natural and principled manner within a decision-theoretic framework, and deals with the issue of high-dimensionality. Specifically, we use continuous structured spike-and-slab priors (George and McCulloch, 1997; Ishwaran et al., 2005; Andersen et al., 2014) within the regression model for a specific layer to (i) blend prior information about the correlation between the genomic markers, and (ii) borrow information about the selected genomic markers from the estimation at one spherical layer and propagate the information through the prior of the model to the subsequent spherical layer(s). As opposed to existing approaches that incorporate apriori structural information (Li and Zhang, 2010; Vannucci and Stingo, 2010), our prior formulation incorporates structural information not only for the dependence between the predictors, but also for the dependence between the multivariate response and across the spherical layers.

Finally, the complex structure of the model warrants scalable estimation approaches. Toward this end, as an alternative to computationally expensive Markov chain Monte Carlo (MCMC) sampling techniques, we employ an Expectation–Maximization (EM)-based estimation strategy (Ročková and George, 2014), where we only search for the posterior mode to iteratively estimate the posterior selection probabilities. Our EM-based estimation is an generalization of Ročková and George (2014) to a multiple-multivariate regression setting, and provides a scalable alternative to sampling strategies; it also facilitates model selection. While developed in the context of this specific imaging-genomics case study, our sequential modeling framework is general, has independent methodological value and could also be used in other applications, which exhibit a natural ordering, e.g. multivariate data collected over time, space, or other axes.

Our work has innovations in two broad aspects. First, in our framework, we evaluate tumor heterogeneity by mimicking the tumor evolution process and constructing spherical layers from the tumor region. We then create layer-wise imaging phenotypes that comprehensively capture tumor heterogeneity by constructing PDFs that quantify the voxel-intensity distribution in each tumor layer. These two steps of evaluating the tumor as spherical layers and quantifying heterogeneity in these layers via PDFs are novel and have not been previously explored to the best of our knowledge. Second, the methodological novelty arises in developing a multivariate-multiple regression set-up that not only considers PDF-based imaging phenotypes as the multivariate response and genomic markers as predictors but also incorporates various types of structures/correlations from the imaging and genomic data into the modeling. These structures include passing information between the spherical layers, correlations between the genomic markers and correlations between the imaging phenotypes. Combining these structural attributes within a hierarchical Bayesian regression model with variable selection allows us to identify meaningful radiogenomic associations using a statistically rigorous framework.

The rest of this paper is organized as follows. In Section 2, we describe the radiological imaging data and the construction of the imaging phenotypes. We present our statistical modeling framework in Section 3, including model specification, details about the EM-based estimation (Section 4), and model selection (Section 4.2). In

Section 5, we include results from simulation studies under different scenarios to assess the performance of the proposed modeling approach. In Section 6, we provide details of the genomic data, present results of our radiogenomic analysis of LGGs, and shed light upon some important biological findings (Section 6). Finally, in Section 7, we conclude with a discussion and some directions for future work. The indices to sections, figures, tables, and equations in the supplementary material are preceded by 'S' (e.g. Section S1).

2. Data characteristics and imaging phenotypes

In this section, we describe the MRI data characteristics and present the explicit quantification of density-based imaging phenotypes. MRI scans are a rich source of imaging data as they provide a wide range of imaging contrasts. Consequently, different tissue characteristics in the brain are highlighted differently by the various MRI sequences. Primary MRI sequences include (i) native (T1), (ii) post-contrast T1-weighted (T1Gd), (iii) T2-weighted (T2), and (iv) T2 fluid attenuated inversion recovery (FLAIR). The utility of these imaging sequences in providing complementary information can be seen in Figure S1, where the tumor sub-regions appear with varying contrasts. MRI scans typically include data on all four imaging sequences and segmentation labels that generate a mask indicating the tumor and non-tumor regions in the MRI scan (top-left panel in Fig. 2). We specifically consider MRI scans of 65 LGG subjects obtained from The Cancer Imaging Archive (TCIA) (Clark et al., 2013). Further details about the data are provided in Section 6. Each MRI scan is structured as a 3D array where the third axis corresponds to the different axial slices of the image. Hence, we have four 3D arrays, each corresponding to one of the four imaging sequences, and an additional 3D array for the accompanying segmentation mask. All four MRI scans, and the segmentation mask, have a voxel-to-voxel correspondence which allows for clear specification of the tumor region. In Figure S1, we show an example of the axial slice from a brain MRI for a subject with LGG for each of the four imaging sequences. The region inside the red boundary overlaid on these images indicates the segmented tumor region. To integrate MRI data into our modeling, we first construct imaging phenotypes (or radiomic phenotypes) based on the voxel intensities of the MRI scans.

Construction of PDF-based imaging phenotypes. We divide the 3D tumor region into equal-volume concentric spherical shells with respect to a reference voxel (see Fig. 1). This reference voxel is chosen to be the mid-point for the axis-aligned minimal bounding box of the 3D tumor region. Consider MRI scans for n LGG subjects across all four imaging sequences and the corresponding tumor segmentation masks. Let N_i denote the number of voxels in the tumor region (i.e., tumor volume) for each subject i, and $\mathbf{c}_i$ denote array indices of the mid-point for the axis-aligned minimal bounding box of the tumor region. When dividing the tumor into $\tau \in \mathbb{N}$ concentric spherical shells, each spherical shell will have a volume of approximately N_i/τ. For each spherical shell, with $\mathbf{c}_i$ as the center, we make small increments to the radius to determine which voxels belong to that shell so that its volume is close to N_i/τ. When identifying voxels corresponding to shell t, we exclude all voxels corresponding to shells $1, \ldots, (t-1)$. While the innermost layer is a sphere, we refer to all of the τ layers as *spherical shells or layers* hereafter. Hence, for each subject, we have the tumor region divided into τ spherical shells across all four imaging sequences. This is represented in panel A of Fig. 2, with spherical layers from four MRI sequences and their corresponding PDFs. We outline this process in Algorithm S1 of Section S1, and describe construction of kernel density estimates $f_i^{(t,m)}$ corresponding to spherical shell $t \in \{1, \ldots, \tau\}$ from MRI sequence $m \in M = \{T1, T1Gd, T2, FLAIR\}$ for subject i. Via Algorithm S1, we obtain kernel density estimates $f_i^{(t,m)}$ based on the voxels from MRI sequence m in spherical shell $D_i^{(t)}$ for subject i. Using a Riemannian-geometric framework for analyzing PDFs (Srivastava et al., 2007), we implement a principal component analysis (PCA) at each layer $t = 1, \ldots, \tau$, which generates layer-specific PC scores (as shown
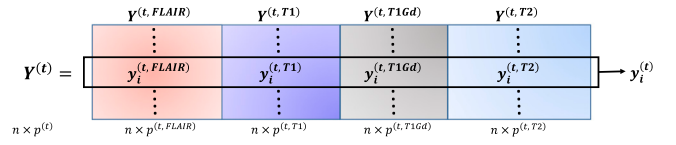

Fig. 3. Pictorial representation of $Y^{(t)}$ by augmenting PC scores across imaging sequences.

in Fig. 2). The geometric framework for PDFs is necessary in order to ensure that the PCA is compatible with the nonlinear structure of the space of the voxel-intensity PDFs, and has been successfully used in several works (Saha et al., 2016; Mohammed et al., 2021). For ease of exposition, we omit the mathematical details here and describe them in Section S1.2.

Layer-specific PCA of PDFs thus maps each PDF to a Euclidean vector of PC scores, which efficiently captures the variability in the voxel-intensity PDFs. The vector of PC scores is representative of tumor heterogeneity, and constitutes the imaging phenotype, which is the response in our Bayesian model. For each imaging sequence m and tumor layer t, we perform PCA using the sample of PDFs $f_1^{(t,m)}, \ldots, f_n^{(t,m)}$ to obtain the PC scores $Y^{(t,m)}$. The number of PCs included is decided such that the included PCs cumulatively explain 99% of the total variance. Note that unlike the case for traditional functional data on linear Hilbert spaces, the corresponding principal directions of variability for PDFs are not orthogonal on the space of PDFs—they are orthogonal in the linear tangent space at the mean. Thus, one needs to be careful when interpreting the PC scores as capturing magnitude associated with PDF eigenfunctions; details in Section S1 elucidate on this. Although we consider the PDF-based PC scores as response in our modeling framework, the methods developed in Section 3 are broadly applicable to multivariate Euclidean responses.

3. Models and methods

Following the above quantification, our data structure is of the following format. Let $\mathbf{y}_i^{(t,m)} = (y_{i1}^{(t,m)}, \ldots, y_{ip^{(t,m)}}^{(t,m)})$ represent a row vector of PC scores from imaging sequence $m \in M$ and layer $t \in T = \{1, \ldots, \tau\}$ for subject i. Here, $p^{(t,m)}$ represents the number of PCs included for layer t and imaging sequence m. Let $\mathbf{y}_i^{(t)} = (\mathbf{y}_i^{(t,T1)}, \ldots, \mathbf{y}_i^{(t,FLAIR)})$ represent the PC scores concatenated across all four imaging sequences and define $Y^{(t)} = (\mathbf{y}_1^{(t)^\top}, \ldots, \mathbf{y}_n^{(t)^\top})^\top \in \mathbb{R}^{n \times p^{(t)}}$ where $p^{(t)} = \sum_{m \in M} p^{(t,m)}$ is the total number of PCs included across all four imaging sequences. A pictorial representation of this augmentation is provided in Fig. 3.

Let $X \in \mathbb{R}^{n \times g}$ correspond to the predictors from g genomic markers for n subjects. For each $t \in T$, we model the PC scores $Y^{(t)}$ using the genomic markers X as

$$Y^{(t)} \sim N_{n,p^{(t)}}(XB^{(t)}, I_n \otimes \Delta^{(t)}), \text{ or equivalently,}$$

$$vec(Y^{(t)}) \sim N_{np^{(t)}}(vec(XB^{(t)}), I_n \otimes \Delta^{(t)}), \tag{1}$$

where $B^{(t)} = [B^{(t,T1)}, \ldots, B^{(t,FLAIR)}] \in \mathbb{R}^{g \times p^{(t)}}$ and $B^{(t,m)} = [\boldsymbol{\beta}_1^{(t,m)}; \ldots; \boldsymbol{\beta}_{p^{(t,m)}}^{(t,m)}] \in \mathbb{R}^{g \times p^{(t,m)}}$ for all $m \in M$. Here, $\boldsymbol{\beta}_j^{(t,m)} = (\beta_{1j}^{(t,m)}, \ldots, \beta_{gj}^{(t,m)})^\top$ for all $j = 1, \ldots, p^{(t,m)}$ and $m \in M$, and $vec(\cdot)$ is the row-wise vectorization of a matrix. The coefficient $\beta_{kj}^{(t,m)}$ corresponds to the association of genomic marker k and principal component j from the imaging sequence m in the layer t. Also, I_n is the identity matrix of dimension n, and $\otimes$ denotes the Kronecker product. Note that $\Delta^{(t)} \in \mathbb{R}^{p^{(t)} \times p^{(t)}}$ is a covariance matrix that captures the covariance between the PCs across imaging sequences, which naturally exists as these sequences are imaging the same region/tumor. In other words, although the PC scores corresponding to a specific imaging sequence are uncorrelated (by construction), cross-sequence correlation could still exist (e.g. between PC scores from T1 and PC scores from T2); this correlation is captured through $\Delta^{(t)}$. We can succinctly rewrite $vec(XB^{(t)}) = Z\boldsymbol{\beta}^{(t)}$, where $Z = X \otimes I_{p^{(t)}} \in$

$\mathbb{R}^{np^{(t)} \times gp^{(t)}}$ (Gupta and Nagar, 2018) and $\boldsymbol{\beta}^{(t)} = vec(B^{(t)}) \in \mathbb{R}^{gp^{(t)}}$. Let $\mathbf{y}^{(t)} = vec(Y^{(t)})$ and $\Theta^{(t)} = I_n \otimes \Delta^{(t)}$; then, Eq. (1) is equivalent to $\mathbf{y}^{(t)} \sim N_{np^{(t)}}(Z\boldsymbol{\beta}^{(t)}, \Theta^{(t)})$.

Structured variable selection prior. Our aim is to identify genomic markers associated with the PC scores, which translates to identifying the nonzero coefficients $B^{(t)}$ in the model in Eq. (1). The genomic markers could themselves be correlated or have an inherent dependence structure (e.g. similar biological function or signaling pathway). We incorporate this information within the variable selection framework by proposing a generalization of structured continuous spike-and-slab priors as follows (George and McCulloch, 1997; Andersen et al., 2014):

$$\beta_{kj}^{(t,m)} \Big| \zeta_k^{(t,m)}, v_{kj}^{2\ (t,m)} \overset{ind}{\sim} N\left(0, \left[(1 - \zeta_k^{(t,m)})v_0 + \zeta_k^{(t,m)} v_1\right] v_{kj}^{2\ (t,m)}\right),$$
$$\zeta_k^{(t,m)} \Big| \lambda_k^{(t,m)} \overset{ind}{\sim} Ber(\Phi(\lambda_k^{(t,m)})), \tag{2}$$
$$\lambda^{(t,m)} \overset{ind}{\sim} N(\boldsymbol{\mu}^{(t,m)}, \Lambda),$$

for all k, j, t, m, where $\eta_{kj}^{2\ (t,m)} := \left[(1 - \zeta_k^{(t,m)})v_0 + \zeta_k^{(t,m)} v_1\right] v_{kj}^{2\ (t,m)}$ is the hypervariance of $\beta_{kj}^{(t,m)}$ with $0 < v_0 \ll v_1$. The indicator $\zeta_k^{(t,m)}$ takes values 1 and 0 with probability $\Phi(\lambda_k^{(t,m)})$ and $(1 - \Phi(\lambda_k^{(t,m)}))$, respectively. Here $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. Note that any suitable injective function could be chosen instead of Φ. We use a Gamma prior on $v_{kj}^{-2(t,m)}$ with a_1 and a_2 as the shape and rate parameters, respectively. If $\zeta_k^{(t,m)} = 1$, the hypervariance for $\beta_{kj}^{(t,m)}$ is dictated by the prior on $v_{kj}^{-2(t,m)}$; when $\zeta_k^{(t,m)} = 0$, the hypervariance $v_0 v_{kj}^{2\ (t,m)}$ is small in magnitude allowing for shrinkage of the coefficient $\beta_{kj}^{(t,m)}$ toward zero. For this purpose, v_0 is chosen to be a much smaller positive number than v_1, encouraging the exclusion of much smaller nonzero effects. We use an inverse-Wishart (IW) prior on the covariance matrix $\Delta^{(t)}$ as $\Delta^{(t)} \overset{iid}{\sim} IW(\delta, \Psi)$, where δ is the degrees of freedom and Ψ is the scale matrix.

The indicator $\zeta_k^{(t,m)}$ corresponds to the genomic marker k across all PC scores j. That is, at each level t and imaging modality m, genomic marker k is assumed to be correlated with the entire voxel intensity PDF, and not just some of its features. This structure is built-in with the assumption that we are interested in associations of genomic markers with imaging sequences as a whole, and not just associations with specific PCs within an imaging sequence.

Borrowing strength using $\boldsymbol{\mu}^{(t,m)}$ and Λ. In essence, our model breaks down the components of a usual multiple-multivariate regression, into a sequence of such separate models for each layer $t \in T$. The dependence structure between the responses and hence, the corresponding regression coefficients is captured through the hyperparameters in our model specification. For $t = 1$, we choose a fairly vague prior by considering $\boldsymbol{\mu}^{(1,m)} = \mathbf{0}$. However, for the model at layer $t > 1$, we incorporate information from the estimation for models at the previous layers $1, \ldots, (t-1)$. We do so through a sequential posterior updating of the layer-specific parameters. Let $\hat{\lambda}^{(s,m)}$ be the posterior estimate of $\lambda^{(s,m)}$ for $s < t$; then, some potential choices for $\boldsymbol{\mu}^{(t,m)}$ include

1. $\boldsymbol{\mu}^{(t,m)} = \alpha \hat{\lambda}^{(t-1,m)} + (1 - \alpha)\mathbf{0}$, where $\alpha \in [0,1]$;
2. $\boldsymbol{\mu}^{(t,m)} = \sum_{s=1}^{t-1} \alpha_s \hat{\lambda}^{(t-s,m)} + \alpha_t \mathbf{0}$, such that $\sum_{s=1}^{t-1} \alpha_s = 1$;
3. $\boldsymbol{\mu}^{(t,m)} = \sum_{s=1}^{t-1} \alpha_s \hat{\lambda}^{(t-s,m)}$ for $\alpha_s \in \mathbb{R}$;
4. $\boldsymbol{\mu}^{(t,m)} = \alpha \hat{\lambda}_+^{(t-1,m)}$, where $\hat{\lambda}_{k,+}^{(t-1,m)} = \max(\hat{\lambda}_k^{(t-1,m)}, 0)$ for $k = 1, \ldots, g$ and $\alpha \in [0,1]$.

The first choice centers the prior on $\lambda^{(t,m)}$ based only on the estimates from the previous layer $(t-1)$. The second choice centers the prior in the convex hull created by estimates from all previous layers $1, \ldots, (t-1)$. The third choice induces an auto-regressive structure on the estimates from all previous layers $1, \ldots, (t-1)$. The fourth choice places higher prior probability of selection on those columns of X which were selected during the estimation for the previous layer $(t-1)$. That is, predictor k will have a higher chance of selection ($\zeta_k^{(t,m)} = 1$) in layer
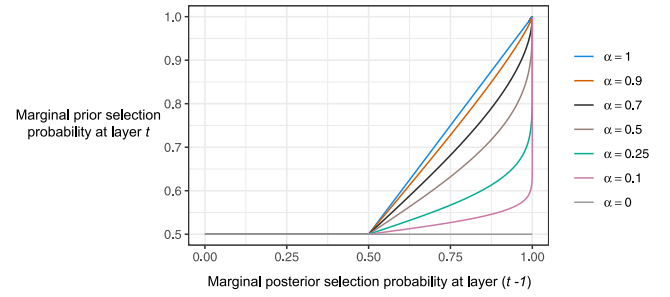

Fig. 4. Marginal prior selection probability $\Phi(\lambda_k^{(t,m)})$ (evaluated at $\lambda_k^{(t,m)} = \mu_k^{(t,m)} = \alpha \hat{\lambda}_{k,+}^{(t-1,m)}$) *versus* the marginal posterior selection probability $\Phi(\hat{\lambda}_k^{(t-1,m)})$ when $\Lambda_{kk} = 1$.

t for imaging sequence m when $\lambda_k^{(t,m)} > 0$. However, $\lambda_k^{(t,m)} > 0$ has higher probability of occurrence if $\mu_k^{(t,m)} > 0$. The construction of $\boldsymbol{\mu}^{(t,m)}$ in the fourth choice above ensures $\mu_k^{(t,m)} \geq 0$. Note that this choice does not penalize the selection of the predictor k in case it was not selected for the previous layer $(t-1)$. That is, the marginal prior probability of selection for that predictor at layer t will be 0.5. This induces a non-informative prior for those predictors that were not selected and lets the data dictate its posterior selection probability at layer t. In Fig. 4, for different choices of α, we plot the marginal prior selection probability $\Phi(\lambda_k^{(t,m)})$ (evaluated at $\lambda_k^{(t,m)} = \mu_k^{(t,m)} = \alpha \hat{\lambda}_{k,+}^{(t-1,m)}$) of gene k for imaging sequence m at layer t, versus the marginal posterior selection probability $\Phi(\hat{\lambda}_k^{(t-1,m)})$ at layer $(t-1)$ when $\Lambda_{kk} = 1$. For example, if $\hat{\lambda}_k^{(t-1,m)} = 1$, then predictor k has marginal posterior selection probability $\Phi(\hat{\lambda}_k^{(t-1,m)}) = 0.84$ for imaging sequence m at layer $(t-1)$. In this case, the marginal prior selection probability $\Phi(\lambda_k^{(t,m)})$ at layer t is $0.60, 0.69$ and 0.84 when $\alpha = 0.25, 0.5$ and 1, respectively. The marginal prior selection probability $\Phi(\lambda_k^{(t,m)}) = 0.5$ at layer t when $\hat{\lambda}_k^{(t-1,m)} \leq 0$.

We focus on the fourth case for the simulations and case study. Nevertheless, our estimation process proceeds in a sequential manner in t, where the estimation for model at layer t borrows information from the model for layer $(t-1)$, and possibly other previous layers. In the prior specification in Eq. (2), the parameter $\lambda_k^{(t,m)}$ serves as a structural parameter as its prior can be used to incorporate any dependence in the columns of X. The correlation structure between the genomic markers is incorporated into the prior for $\lambda^{(t,m)} = (\lambda_1^{(t,m)}, \ldots, \lambda_g^{(t,m)})^\top$ through its covariance specification Λ. If a certain genomic marker k is associated with the PC scores from imaging sequence m and layer t, then we assume that other positively correlated genomic markers might also have a similar association. This assumption is driven by the fact that voxels from the boundaries of the spherical layers are contiguous objects with potentially similar characteristics. This can be incorporated through a graph Laplacian or other choice of network structure-based covariance matrix Λ. If $\boldsymbol{\mu}^{(t,m)} = \mathbf{0}$ and $\Lambda = I_g$, then the joint prior on $\zeta_k^{(t,m)}$ and $\lambda_{(t,m)}$ reduces to $\zeta_k^{(t,m)} \overset{ind}{\sim} Ber(v_k^{(t,m)})$ with $v_k^{(t,m)} \overset{iid}{\sim} U(0,1)$.

Role of $\boldsymbol{\mu}^{(t,m)}$ and Λ. While $\boldsymbol{\mu}^{(t,m)}$ incorporates the layered dependence structure between responses for the selection of regression coefficients, Λ incorporates the correlation between the columns of X, and hence places higher prior probability of joint selection for any two correlated columns k and k'. On the other hand, $\boldsymbol{\mu}^{(t,m)}$ induces prior dependence for the same column across various layers t. In other words, $\mu_k^{(t,m)}$ centers the prior probability of selection for column k in layer t, based on its selection in previous layers. As a result, $\boldsymbol{\mu}^{(t,m)}$ and Λ contribute complementary information to the model and jointly contribute towards the selection of the regression coefficients.

4. Estimation and model selection

The full posterior distribution corresponding to the parameters of the model in Eq. (1) with the prior structure considered in Eq. (2) is

given in Section S2.1. Due to the multiple-multivariate regression setup, we have $g \times p^{(t)}$ coefficients (β_{kj}) and/or $|M| \times g$ posterior inclusion probabilities to sequentially estimate for each model at layer t. A full MCMC-based estimation can be employed, but the computation time would be additive with each additional spherical layer included, as the estimation is required to be sequential based on the model formulation. As a faster alternative to MCMC, we consider the EM-based strategy to variable selection as proposed by Ročková and George (2014). This EM-based estimation approach is a deterministic alternative to commonly used stochastic search algorithms, which exploits the prior formulation to rapidly find posterior modes. We emphasize that the EM-based estimation presented in this work is an adaptation of Ročková and George (2014) to the multiple-multivariate regression setting. We discuss sequential estimation using EM and describe the model selection procedure below.

4.1. Sequential EM updating

Our implementation of the EM-based estimation approach indirectly maximizes the parameter posterior given by $\pi(\beta_{kj}^{(t,m)}, \lambda^{(t,m)}, v_{kj}^{-2(t,m)}, \Delta^{(t)} | \mathbf{y}^{(t)})$, by proceeding iteratively over the complete-data log posterior $\log \pi(\beta_{kj}^{(t,m)}, \zeta_k^{(t,m)}, \lambda^{(t,m)}, v_{kj}^{-2(t,m)}, \Delta^{(t)} | \mathbf{y}^{(t)})$, where $\zeta_k^{(t,m)}$ are treated as missing data. At each iteration, the complete-data log posterior is replaced by its conditional expectation given the observed data and the current parameter estimates; this forms the E-Step. The expected complete-data log posterior is then maximized with respect to (w.r.t.) $\chi^{(t)} = (\beta_{kj}^{(t,m)}, \lambda^{(t,m)}, v_{kj}^{-2(t,m)}, \Delta^{(t)})$ for all m, j, k, which constitutes the M-Step. Iterating between the E-Step and M-Step generates a sequence of parameter estimates, which converge monotonically to the local maximum of $\pi(\chi^{(t)} | \mathbf{y}^{(t)})$, under appropriate regularity conditions.

The expression for the complete-data log posterior is given in Section S2.1. For the uth iteration of the algorithm, conditional expectation of the complete-data log posterior is given by

$$Q\left(\beta_{kj}^{(t,m)}, \lambda^{(t,m)}, v_{kj}^{-2(t,m)}, \Delta^{(t)} | \beta_{kj}^{(t,m),(u)}, \lambda^{(t,m),(u)}, v_{kj}^{-2(t,m),(u)}, \Delta^{(t),(u)}\right)$$
$$:= E_{\zeta^{(t)} | \chi^{(t),(u)}}\left[\log\left(\pi(\beta_{kj}^{(t,m)}, \zeta_k^{(t,m)}, \lambda^{(t,m)}, v_{kj}^{-2(t,m)}, \Delta^{(t)} | \mathbf{y}^{(t)})\right)\right], \quad (3)$$

where $E_{\zeta^{(t)} | \chi^{(t),(u)}}$ is a conditional expectation with $\chi^{(t),(u)} = (\beta_{kj}^{(t,m),(u)}, \lambda^{(t,m),(u)}, v_{kj}^{-2(t,m),(u)}, \Delta^{(t),(u)})$ for all m, j, k and $\zeta^{(t)} = (\zeta_k^{(t,m)})_{m,k}$. For the $(u+1)$th iteration in the EM algorithm, we (i) find the expectation in Eq. (3) to obtain the Q-function, and (ii) maximize the Q-function w.r.t. $\chi^{(t)}$ to obtain new estimates $\chi^{(t),(u+1)}$.

4.1.1. The E-step

As the conditional expectation is computed w.r.t. $\zeta_k^{(t,m)}$, we note that the only two terms in the complete-data log posterior involving $\zeta_k^{(t,m)}$ are the terms arising from the prior for $\beta_j^{(t,m)}$ and the prior for $\zeta_k^{(t,m)}$. After some algebra (see Section S2.2), we see that these two terms are given by

$$- \sum_{k,j,m} \frac{1}{2} \log(v_{kj}^{2(t,m)}) + \frac{(\beta_{kj}^{(t,m)})^2}{2 v_{kj}^{2(t,m)}} E_{\zeta^{(t)} | \chi^{(t),(u)}}\left[\frac{1}{(1-\zeta_k^{(t,m)})v_0 + \zeta_k^{(t,m)}v_1}\right],$$

and

$$\sum_{k,m} \log(1 - \Phi(\lambda_k^{(t,m)})) + E_{\zeta^{(t)} | \chi^{(t),(u)}}[\zeta_k^{(t,m)}] \log \frac{\Phi(\lambda_k^{(t,m)})}{1 - \Phi(\lambda_k^{(t,m)})}.$$

To complete the E-Step, the two expectations we need to compute are given by

$$w_k^{(t,m)} := E_{\zeta^{(t)} | \chi^{(t),(u)}}[\zeta_k^{(t,m)}],$$

and

$$d_k^{(t,m)} := E_{\zeta^{(t)} | \chi^{(t),(u)}}\left[\frac{1}{(1-\zeta_k^{(t,m)})v_0 + \zeta_k^{(t,m)}v_1}\right].$$

Note then that $w_k^{(t,m)} = P(\zeta_k^{(t,m)} = 1 | \chi^{(t),(u)}) = \frac{a_k^{(t,m)}}{a_k^{(t,m)} + b_k^{(t,m)}}$, where $a_k^{(t,m)}$ and $b_k^{(t,m)}$ are defined in Equations (S1) and (S2), respectively. To compute $d_k^{(t,m)}$, we note than when $\zeta_k^{(t,m)} \in \{0,1\}$, we have $((1-\zeta_k^{(t,m)})v_0 + \zeta_k^{(t,m)}v_1)^{-1} = \frac{1-\zeta_k^{(t,m)}}{v_0} + \frac{\zeta_k^{(t,m)}}{v_1}$, almost surely. Hence, $d_k^{(t,m)} = \frac{1-w_k^{(t,m)}}{v_0} + \frac{w_k^{(t,m)}}{v_1}$.

4.1.2. The M-step

Once we compute $w_k^{(t,m)}$ and $d_k^{(t,m)}$ in the E-step, we focus on maximizing the expected log posterior, i.e., the Q-function, w.r.t. other parameters. The Q-function, which is a function of $\chi^{(t)} = (\beta_{kj}^{(t,m)}, \lambda^{(t,m)}, v_{kj}^{-2(t,m)}, \Delta^{(t)})$ for all m, j, k, can be split into three terms as

$$Q(\chi^{(t)} | \chi^{(t),(u)}) = Q_1(\beta_{kj}^{(t,m)}, v_{kj}^{-2(t,m)}, \Delta^{(t)} | \beta_{kj}^{(t,m),(u)}, v_{kj}^{-2(t,m),(u)}, \Delta^{(t),(u)})$$
$$+ Q_2(\lambda^{(t,m)} | \lambda^{(t,m),(u)}) + const,$$

where Q_1 only depends on $\beta_{kj}^{(t,m)}, v_{kj}^{-2(t,m)}$ and $\Delta^{(t)}$, and Q_2 depends on $\lambda^{(t,m)}$. The complete forms of Q_1 and Q_2 are given in Section S2.3. Completing the EM-based estimation procedure, we maximize Q w.r.t. $\chi^{(t)}$, i.e., we maximize Q_1 w.r.t. $\beta_{kj}^{(t,m)}, v_{kj}^{-2(t,m)}$ and $\Delta^{(t)}$, and Q_2 w.r.t. $\lambda^{(t,m)}$. The objective functions for all of the maximization problems are provided in Section S2.4.

Maximizing Q_1 w.r.t. $\beta_{kj}^{(t,m)}$. Assuming we have the parameters $v_{kj}^{2(t,m),(u)}, \Delta^{(t),(u)}$ at iteration u and the value $d_k^{(t,m),(u+1)}$ at iteration $(u+1)$, maximizing Equation (S3) is equivalent to finding the maximum a-posteriori (MAP) estimate of a Bayesian regression $\mathbf{y}^{(t)} \sim N_{np^{(t)}}(Z\beta^{(t)}, \Theta^{(t)})$ with the prior $\beta^{(t)} \sim N(\mathbf{0}, \Gamma^{(t)})$. Thus, the update (MAP estimate) for $\beta^{(t)}$ is given by

$$\beta^{(t),(u+1)} = \Sigma^{(t),(u)} Z^\top (\Theta^{(t),(u)})^{-1} \mathbf{y},$$

where $\Sigma^{(t),(u)} = \left(Z^\top (\Theta^{(t),(u)})^{-1} Z + (\Gamma^{(t),(u)})^{-1}\right)^{-1}$, $\Gamma^{(t)} = diag(v_{kj}^{(t,m)}/d_k^{(t,m)})$ and $\Gamma^{(t),(u)} = diag(v_{kj}^{(t,m),(u)}/d_k^{(t,m),(u+1)})$.

Maximizing Q_1 w.r.t. $v_{kj}^{-2(t,m)}$. Each term inside the product in Equation (S4) is the kernel of a Gamma distribution for $v_{kj}^{-2(t,m)}$. Assuming we have the parameters $\beta^{(t),(u+1)}, \Delta^{(t),(u)}$ at iteration $(u+1)$, the update for $v_{kj}^{-2(t,m)}$ (maximizer of Equation (S4)) for each m, k, j is given by

$$v_{kj}^{-2(t,m),(u+1)} = \left(a_1 - \frac{1}{2}\right) \Big/ \left(a_2 + \frac{(\beta_{kj}^{(t,m),(u+1)})^2 d_k^{(t,m),(u+1)}}{2}\right),$$

where $a_1 > \frac{1}{2}$.

Maximizing Q_1 w.r.t. $\Delta^{(t)}$. Equation (S5) is the kernel of a Wishart distribution for $(\Delta^{(t)})^{-1}$; thus, its maximizer is the mode of the corresponding Wishart distribution. Assuming we have the parameters $B^{(t),(u+1)}$ at iteration $(u+1)$, the update for $(\Delta^{(t)})^{-1}$ is given by

$$(\Delta^{(t),(u+1)})^{-1} = (n + \delta + p^{(t)} + 1)\left[\Psi + (Y^{(t)} - XB^{(t)})^\top (Y^{(t)} - XB^{(t)})\right]^{-1}.$$

Maximizing Q_2 w.r.t. $\lambda^{(t,m)}$. The posterior for $\lambda^{(t,m)}$ does not arise from a standard distribution. Thus, we resort to numerical methods, specifically gradient descent, to find the solution for the optimization in Equation (S6), to compute the update for $\lambda^{(t,m)}$ at iteration $(u+1)$. The first derivative of the objective function $F(\lambda^{(t,m)})$ is given by

$$\nabla F(\lambda^{(t,m)}) = (\lambda^{(t,m)} - \mu^{(t,m)})^\top \Lambda^{-1} - (s_1^{(t,m)}, \dots, s_g^{(t,m)}),$$

where

$$s_k^{(t,m)} = -\frac{(1 - w_k^{(t,m)})\phi(\lambda_k^{(t,m)})}{1 - \Phi(\lambda_k^{(t,m)})} + \frac{w_k^{(t,m)}\phi(\lambda_k^{(t,m)})}{\Phi(\lambda_k^{(t,m)})}.$$

Then, the update at the $(r+1)$th iteration of the gradient descent algorithm is given by

$$\lambda^{(t,m),(u+1,r+1)} = \lambda^{(t,m),(u+1,r)} - \kappa \nabla F(\lambda^{(t,m),(u+1,r)}),$$

where κ is the learning rate. This learning rate can be modified within each iteration to speed up convergence. The initial value $\lambda^{(t,m),(u+1,0)}$

at iteration $(u+1)$ of the EM algorithm can be set as the solution of Equation (S6) at iteration u. This solution for $\lambda^{(t,m)}$ is used to assign the prior mean for $\lambda^{(t+1,m)}$ as $\mu^{(t+1,m)} = \alpha\hat{\lambda}_{+}^{(t,m)}$.

One of the major drawbacks of the EM algorithm is its potential to get trapped in a local maximum. A possible solution for this is to run the algorithm with different starting values. In our estimation approach, we consider the deterministic annealing variant of the EM algorithm (Ueda and Nakano, 1998), referred to as the DAEM algorithm, which improves the chance of finding the global maximum (see Section S2.5 for details). All of the results presented in this paper are based on the DAEM algorithm. We denote the DAEM algorithm-based estimates for the parameters in our model by $\hat{\chi}^{(t)}$.

4.2. Model selection

Once we have the estimates of the parameters $\hat{\chi}^{(t)}$, we can compute the posterior inclusion probabilities for predictor k in layer t and imaging sequence m as $\hat{w}_{k}^{(t,m)} = P(\zeta_{k}^{(t,m)} = 1|\hat{\chi}^{(t)})$ using Equation (S8). In terms of variable selection, we set $\hat{\zeta}_{k}^{(t,m)} = 1$ when $P(\zeta_{k}^{(t,m)} = 1|\hat{\chi}^{(t)}) \rangle 0.5$, and set $\hat{\zeta}_{k}^{(t,m)} = 0$ otherwise. Note that v_0 and v_1 are fixed hyperparameters and their choice determines the solution set $\hat{\zeta}^{(t)}$. The speed of the EM algorithm makes it feasible to obtain solution sets $\hat{\zeta}^{(t)}$ for multiple choices of v_0 while fixing v_1. As v_0 is increased, we see that more coefficients with smaller magnitude get absorbed by the spike distribution, leading to sparser models. We consider a grid of values for v_0, fix v_1 to be significantly larger than v_0, and obtain solutions for each combination of v_0 and v_1. We denote the solution set corresponding to the choice of v_0 as $S(v_0) = \{(m,k,1),\ldots,(m,k,p^{(t,m)})|\zeta_{k}^{(t,m)} = 1 \ \forall \ m \in M, k \in \{1,\ldots,g\} \text{ and } t \in \{1,\ldots,\tau\}\}$.

We calculate the Bayesian Information Criterion (BIC) across all layers t, once the set of coefficients to be included in the models has been identified. This is expressed as: $BIC = \sum_{t=1}^{\tau} -2[K^{(t)}\log(n) + \log\pi(Y^{(t)}|\hat{\beta}_{kj}^{(t,m)}, \widehat{\Delta^{(t)}})]$, where $K^{(t)} = \#\{\hat{\beta}_{kj}^{(t,m)} \neq 0|m,k,j\}$ is the number of nonzero coefficients included in the model for layer t. The final model is based on the inclusion of coefficients, which minimize the BIC. We outline the overall step-by-step procedure from constructing the PC scores to final model selection in Algorithm 1.

5. Simulation study

In this section, we demonstrate the effectiveness of our approach through simulation studies under different scenarios. We do not directly simulate the MRI images; instead we simulate $Y^{(t)}$ corresponding to the PC scores for each layer t. To generate the matrix $X = [\mathbf{x}_1,\ldots,\mathbf{x}_n]^{\top} \in \mathbb{R}^{n\times g}$ of g genomic markers for n subjects, we simulate $\mathbf{x}_i$ independently from $N_g(\mathbf{0}, \Sigma^{(x)})$ for $i = 1,\ldots,n$, where $\Sigma^{(x)}$ is the covariance matrix which can be specified to induce different correlation structures between the columns of X. We broadly consider two cases for our simulation study: the response $Y^{(t)}$ is (i) generated from the model, and (ii) generated directly based on fixed coefficients $B^{(t)}$.

Case 1: $Y^{(t)}$ is generated from the model. In this case, we want to assess the performance of the proposed approach when the response $Y^{(t)}$ is generated directly from the model in Eq. (1), with the prior structure in Eq. (2). The step-by-step procedure to generate the data is described as an algorithm in Section S3. We construct subcases defined by different choices of Ψ, which incorporates the correlation structure between the columns of $Y^{(t)}$. For simplicity, we consider $\Psi = \sigma^2 I_p$, where $p = p^{(t)}$ for all t, indicating that the number of columns of $Y^{(t)}$ is the same for all t. Here, σ^2 acts as the noise parameter that controls the variance in the columns of $Y^{(t)}$. The dependence structure across the layers t and imaging sequences m is incorporated through the parameter $\mu^{(t,m)}$, as described in Section 3.

Algorithm 1 Outline of the statistical framework

1: Construct layer-wise kernel density estimates $f_i^{(t,m)}$ as described in Algorithm S1 for all subjects $i = 1,\ldots,n$, spherical layers $t = 1,\ldots,\tau$ and MRI sequence $m \in M$.

2: Consider $X \in \mathbb{R}^{n\times g}$ as the gene expression matrix for LGG-genes.

3: **for** each layer $t = 1,\ldots,\tau$ **do**

4:      **for** each MRI sequence $m \in M$ **do**

5:          Compute the PC scores $Y^{(t,m)} \in \mathbb{R}^{n\times p^{(t,m)}}$ using PCA in Algorithm S3.

6:          Define $Y^{(t)} = [Y^{(t,FLAIR)}; Y^{(t,T1)}; Y^{(t,T1Gd)}; Y^{(t,T2)}] \in \mathbb{R}^{n\times p^{(t)}}$.

7: **for** each $v_0 \in$ Set of values to consider **do**

8:      **for** each layer $t = 1,\ldots,\tau$ **do**

9:          Bayesian Modeling:

$$Y^{(t)} \sim N_{n,p^{(t)}}(XB^{(t)}, I_n \otimes \Delta^{(t)}),$$
$$\beta_{kj}^{(t,m)} \overset{ind}{\sim} N\left(0, \left[(1-\zeta_k^{(t,m)})v_0 + \zeta_k^{(t,m)}v_1\right]v_{kj}^{2\ (t,m)}\right),$$
$$\zeta_k^{(t,m)} \overset{ind}{\sim} Ber(\Phi(\lambda_k^{(t,m)})),$$
$$\lambda^{(t,m)} \overset{ind}{\sim} N(\mu^{(t,m)}, \Lambda),$$
$$v_{kj}^{-2\,(t,m)} \overset{iid}{\sim} Gamma(a_1, a_2),$$
$$\Delta^{(t)} \overset{iid}{\sim} IW(\delta, \Psi).$$

10:          Estimation (Section 4) with deterministic annealing:

a: E-Step: Compute the E-Step as described in Section 4.1.1.

b: M-Step: Sequentially maximize Q_1 w.r.t. $\beta_{kj}^{(t,m)}, v_{kj}^{-2\,(t,m)}, \Delta^{(t)}$ and Q_2 w.r.t. $\lambda^{(t,m)}$.

11: Model Selection: Choose the model corresponding to that v_0, which minimizes the BIC as described in Section 4.2.

We use $n = 100$ samples, $p = 12$ columns in each $Y^{(t)}$ with $p^{(t,m)} = 3$ for all $m \in M$ and $t \in T$, $g = 20$ genomic markers, and $\tau = 3$ layers. We consider two choices for the covariance in the columns of X: (i) $\Sigma^{(x)} = I_g$, indicating no correlation between the genomic markers, and (ii) $\Sigma^{(x)}$ as a (2×2)-block matrix, where $\Sigma_{11}^{(x)} = ((\Sigma_{11}^{(x)}))_{1\leq i,j\leq 10}$ with $(\Sigma_{11}^{(x)})_{ij} = 10$ if $i = j$ and 9 otherwise, $\Sigma_{12}^{(x)} = \Sigma_{21}^{(x)} = \mathbf{0}$ and $\Sigma_{22}^{(x)} = I_{10}$. The second choice of $\Sigma^{(x)}$ indicates two groups of genomic markers, with one group having high between-group correlation and the other group being uncorrelated. For the noise parameter, we consider $\sigma^2 = 1, 10, 20, 30$. For each combination of σ^2 and $\Sigma^{(x)}$, we replicate the simulation 30 times.

Table S1 presents the complete results of EM-based estimation and subsequent model selection (as described in Section 3) for the simulated data in Case 1. We include the true positive selection rate (TPR), the average absolute error for the inclusion probabilities $w_k^{(t,m)}$, i.e., $E_w = \frac{1}{g|M|\tau}\sum_{k,m,t}|\zeta_k^{(t,m)} - \hat{w}_k^{(t,m)}|$, and the average squared error for the estimates of β, i.e., $E_\beta = \frac{1}{gp|M|\tau}\sum_{k,j,m,t}(\beta_{kj}^{(t,m)} - \hat{\beta}_{kj}^{(t,m)})^2$. In Fig. 5, we plot the average TPR across 30 simulations for the four scenarios based on different choices of σ^2 for the case $\Sigma^{(x)} = I_g$, i.e., assuming no correlation between the genomic markers. As the magnitude of the noise σ^2 increases, the TPR decreases across all layers. In this case, we also evaluate the performance of estimation while ignoring the dependence between the layers, by setting $\mu^{(t,m)} = \mathbf{0}$ for all t; this result is shown as model (B) in Fig. 5. From the results for layers 2 and 3, we see that including dependence across layers within the prior of $\lambda^{(t)}$ improves the performance in terms of the TPR for all noise levels. Note that these two scenarios perform similarly in layer 1 as we have $\mu^{(1,m)} = \mathbf{0}$. The first panel in Table S1 shows these rates in more detail along with E_w and E_β. Since $\Sigma^{(x)} = I_g$ and we use Λ to denote the correlation matrix of X, i.e., $\Lambda = cor(X)$, we expect $\Lambda \approx I_g$.

Comparative analysis. In Fig. 6, we consider the case where $\Sigma^{(x)}$ has a block structure and evaluate performance of the model under
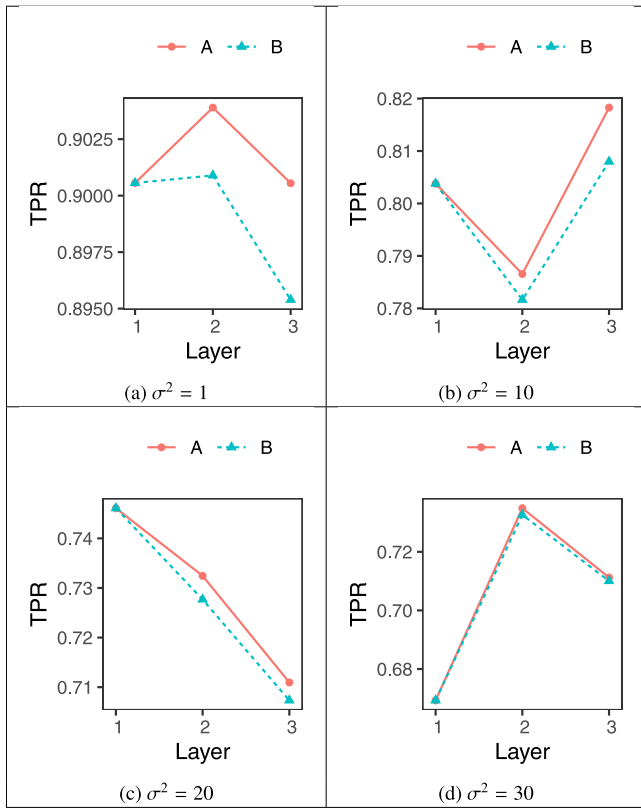
Fig. 5. Average TPR in each of the three layers for Case 1 when $\Sigma^{(x)} = I_g$. Model (A): $\boldsymbol{\mu}^{(t,m)} \neq \mathbf{0}; \Lambda = cor(X) \approx I_g$. Model (B): $\boldsymbol{\mu}^{(t,m)} = \mathbf{0}; \Lambda = cor(X) \approx I_g$.

Fig. 6. Average TPR in each of the three layers for Case 1 when $\Sigma^{(x)}$ is a (2×2)-block matrix. Model (A): $\boldsymbol{\mu}^{(t,m)} \neq \mathbf{0}; \Lambda = cor(X)$. Model (B): $\boldsymbol{\mu}^{(t,m)} = \mathbf{0}; \Lambda = cor(X)$. Model (C): $\boldsymbol{\mu}^{(t,m)} \neq \mathbf{0}; \Lambda = I_g$. Model (D): $\boldsymbol{\mu}^{(t,m)} = \mathbf{0}; \Lambda = I_g$.

four scenarios: (A) including both $\boldsymbol{\mu}^{(t,m)}$ and Λ, (B) ignoring only $\boldsymbol{\mu}^{(t,m)}$, (C) ignoring only Λ, and (D) ignoring both $\boldsymbol{\mu}^{(t,m)}$ and Λ. Ignoring $\boldsymbol{\mu}^{(t,m)}$ corresponds to choosing $\boldsymbol{\mu}^{(t,m)} = \mathbf{0}$ for all t; ignoring Λ implies choosing $\Lambda = I_g$. Scenario (A) refers to incorporating the structure as described in Section 3; Scenario (B) is similar in spirit to other variable selection approaches (Li and Zhang, 2010; Vannucci and Stingo, 2010) that incorporate dependence in the covariates through the indicator variable ζ (while ignoring the layer structure); Scenarios (C) and (D) correspond to vanilla Bayesian variable selection for each layer with uniform priors on selection probabilities for ζ (George and McCulloch, 1997). To establish comparison, we use the same data replications across all four scenarios. We see that incorporating the structure in X through the hyperparameter $\Lambda = cor(X)$, and borrowing information from previous layers for the mean $\boldsymbol{\mu}^{(t,m)}$, yields better estimation performance in terms of TPR, compared to having no structure by choosing $\boldsymbol{\mu}^{(t,m)} = \mathbf{0} \ \forall \ t$ or $\Lambda = I_g$, or both. This improvement in performance holds true for all choices of the noise parameter σ^2. Detailed results are shown in the second panel of Table S1. We have not included the false positive rates in the results as they were less than 0.1% on average in all scenarios. To the best of our knowledge, there are no other modeling frameworks that we could fairly compare with our proposed method that includes layer-wise modeling. However, LASSO-type regularization approaches have been used to identify imaging genetic associations in other applications (Du et al., 2016, 2020). Hence, we compared the estimation under a multivariate-multiple regression with LASSO penalty, which could be construed as a state-of-the-art approach, although it does not incorporate the layered nature of the responses. We observed high TPRs but also substantially higher false positive rates (for both Cases 1 and 2). Results presented in Table S1 report the averages obtained across 30 replications and the corresponding standard deviations. Details of the various hyperparameter choices in the data generation and model estimation processes are given in Section S3.
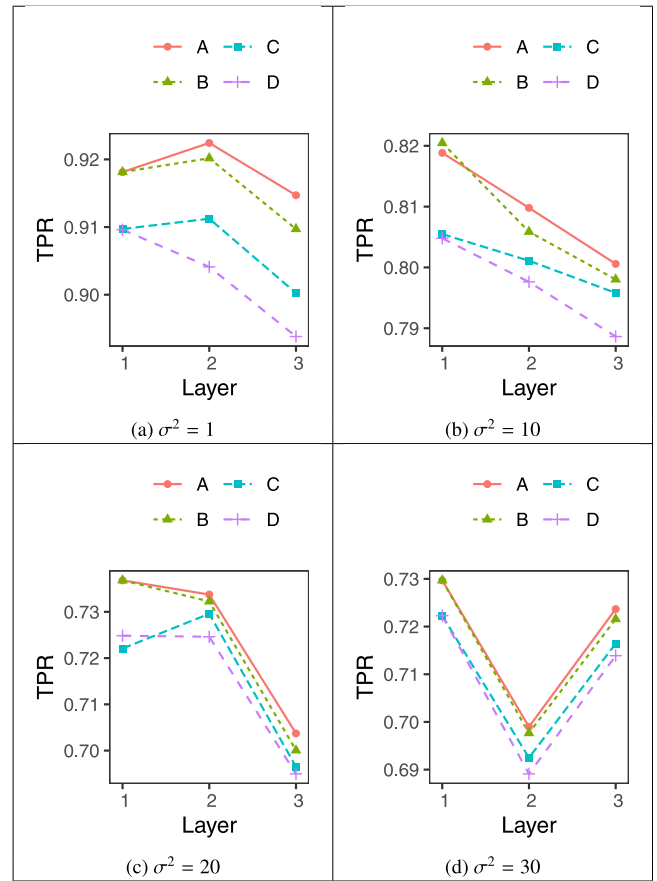
Case 2: $Y^{(t)}$ **is generated directly based on fixed** $B^{(t)}$. In this case, we evaluate the performance of our approach in identifying the associations if they are indeed sequentially dependent across layers. Similar to Case 1, we evaluate performance of the model under the four scenarios (A)-(D). Each of these scenarios determines the choices for $\boldsymbol{\mu}^{(t,m)}$ and Λ as described in Case 1.

We assume $p = p^{(t)} \ \forall \ t$ and consider fixed values of $B^{(t)}$ for all layers $t = 1, \ldots, \tau = 3$. We further assume that $p^{(t,m)} = 3$ for all m, t and $g = 20$. The choice of $B^{(t)}$ is considered so that the genomic markers which are associated with layer t are based on the associations in layers $1, \ldots, (t-1)$. These associations are built so that they do not necessarily need to hold across all imaging sequences. Specifically, we consider $B^{(1)} = \text{block-diag}(B^{(1)}_{15,9}, \mathbf{0})$, $B^{(1)} = \text{block-diag}(B^{(2)}_{10,6}, \mathbf{0})$, and $B^{(1)} = \text{block-diag}(B^{(3)}_{5,3}, \mathbf{0})$, where $B^{(t)}_{r,s}$ is an $(r \times s)$-matrix whose entries are filled by randomly sampling rs values from a double-exponential distribution centered at 0 with scale parameter θ. The value of θ acts as the effect size, and we consider four choices: $\theta = 0.7, 0.8, 0.9, 1$. We use $m = 1, \ldots, 4$ to index the four MRI sequences. Here $B^{(1)}$ indicates that the first 15 genomic markers are associated with all three PC scores corresponding to $m = 1, 2, 3$. As indicated by $B^{(2)}$, the first 10 of these 15 markers are further associated with all three PC scores for $m = 1, 2$. Similarly, from $B^{(3)}$ we see that the first five markers are associated with all three PC scores for $m = 1$. We construct X such that the first 15 columns are positively correlated with each other and uncorrelated with the next five (16-20) columns. The last five columns are also considered to be positively correlated with each other. That is, we consider $\Sigma^{(x)}$ as a (4×4)-block matrix as shown in Figure S6 of Section S3. To generate $Y^{(t)}$, we first simulate $\Delta^{(t)} \sim IW(\delta = p^{(t)}, \Psi = \sigma^2 I_g)$ with
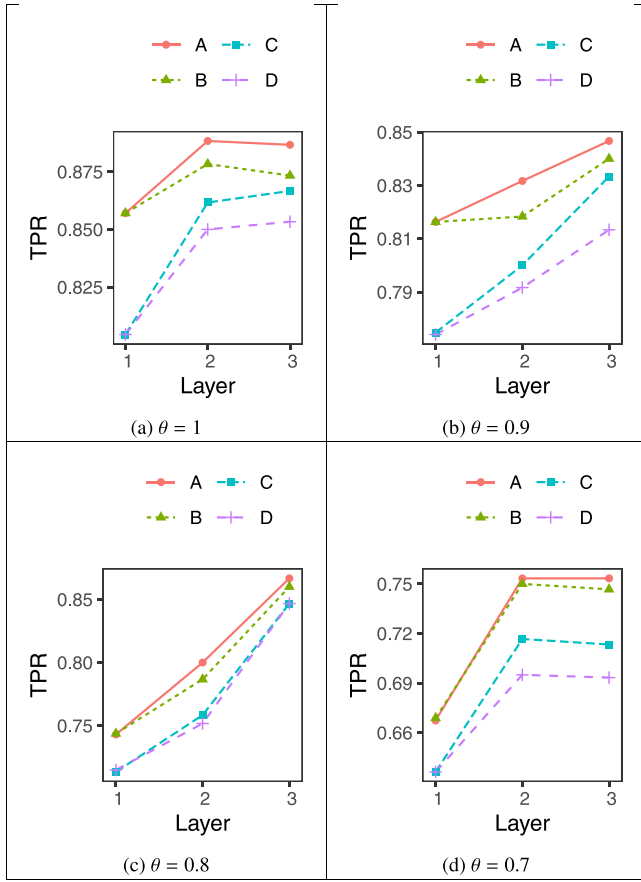
Fig. 7. Average TPR in each of the three layers for Case 2 when $\Sigma^{(x)}$ is a (4×4)-block matrix. Model (A): $\boldsymbol{\mu}^{(t,m)} \neq \boldsymbol{0}; \Lambda = cor(X)$. Model (B): $\boldsymbol{\mu}^{(t,m)} = \boldsymbol{0}; \Lambda = cor(X)$. Model (C): $\boldsymbol{\mu}^{(t,m)} \neq \boldsymbol{0}; \Lambda = I_g$. Model (D): $\boldsymbol{\mu}^{(t,m)} = \boldsymbol{0}; \Lambda = I_g$.

$\sigma^2 = 20$, which is one of the higher choices for the noise parameter in Case 1. We then generate $vec(Y^{(t)}) \sim N_{np^{(t)}}(vec(XB^{(t)}), I_n \otimes \Delta^{(t)})$ and appropriately reshape $vec(Y^{(t)})$ to $Y^{(t)} \in \mathbb{R}^{n \times p^{(t)}}$. This step-by-step procedure of generating the data under Case 2 is provided in Section S3.

We now discuss the results of the simulation study under Case 2. Similar to Case 1, we present the TPR, E_w and E_β for each of the $\tau = 3$ layers separately in Table S2. For each choice of the effect size θ, we present results corresponding to four scenarios, which are determined based on incorporating (or ignoring) the structure through $\boldsymbol{\mu}^{(t,m)}$ and Λ. Fig. 7 shows the average TPR under all four choices of θ. We note that including the structure through $\boldsymbol{\mu}^{(t,m)}$ and Λ provides higher values for TPR thereby improving performance, even for smaller effect sizes. From Table S2, we see that by including this structure we also obtain lower values for error metrics E_w and E_β. We note that including the dependence structure in the columns of X, through Λ, has a pronounced influence on the performance compared to ignoring it by considering $\Lambda = I_g$. Including the structure between the layers through $\boldsymbol{\mu}^{(t,m)}$ also has an incremental effect in the performance for all choices of θ. The results for layer 1 are similar for scenarios with $\boldsymbol{\mu}^{(t,m)} = \boldsymbol{0}$ and $\boldsymbol{\mu}^{(t,m)} \neq \boldsymbol{0}$ for fixed Λ as $\boldsymbol{\mu}^{(1,m)} = \boldsymbol{0}$. Table S2 reports the averages obtained across 30 replications, and are accompanied by the standard deviations; false positive rates are not reported as they were negligible across all scenarios. Details of the hyperparameter choices are provided in Section S3.

6. Radiogenomic analysis in lower grade gliomas

We leverage the radiology-based imaging data from TCIA and the genomic data from The Cancer Genome Atlas (TCGA; www.cancer.gov/tcga) in the context of LGG, to understand the associations between imaging phenotypes and genomic markers. In this section, we describe the data acquisition and pre-processing steps, and present results from our Bayesian modeling approach. A R package is provided with the supplementary material that includes relevant data and code.

Imaging data. The imaging data acquired from TCIA contain publicly available pre-operative multi-institutional MRI scans for the TCGA LGG cohort. Specifically, we obtained MRI scans and the corresponding tumor segmentation labels for 65 LGG subjects from Bakas et al. (2017a,b). The segmentation labels were constructed using an automated segmentation method called GLISTRboost (Bakas et al., 2015). One of the hurdles posed by MRI for statistical analysis is that the voxel intensity values are not comparable either between study visits for the same subject or across different subjects. Hence, it is required to pre-process these images to perform intensity value normalization. We employ a biologically motivated normalization technique using the R package *WhiteStripe* for intensity normalization (Shinohara et al., 2014). However, we discard imaging data for two subjects due to issues with segmentation masks and intensity value normalization, leaving us with a total of 63 LGG subjects with imaging data.

Imaging phenotypes. Gliomas contain various heterogeneous sub-regions broadly categorized into three groups: necrosis and non-enhancing core, edema, and enhancing core. These sub-regions reflect differences in tumor biology, have variable histologic and genomic phenotypes, and exhibit highly variable clinical prognosis (Bakas et al., 2017b). Typically, the area of dead cells in the center is referred to as necrosis or non-enhancing core, and the tissue swelling caused in the outer region due to accumulation of fluid is referred to as edema (https://mayfieldclinic.com/pe-braintumor.htm). Following suit, we divide the tumor region for each subject into $\tau = 3$ spherical shells, i.e., one inner sphere and two spherical shells. For each subject, we follow the procedure described in Algorithm S1 to construct the PDFs corresponding to the three spherical shells. The PDFs estimated in each layer are used to construct the PC scores $Y^{(t)}$, which act as the imaging phenotypes.

Genomic data. We obtained the genomic data from LinkedOmics (Vasaikar et al., 2017), a publicly available portal that includes multi-omics data from multiple cancer types in TCGA. The genomic data we acquire are normalized gene-level RNA sequencing data from the Illumina HiSeq system (high-throughput sequencing) with the expression values on the $\log_2$ scale. This data is obtained for all of the 63 matched LGG subjects across 20086 genes. We focus our analysis on *LGG-genes*, which are genes specifically reported as cancer drivers in LGG (Bailey et al., 2018). The cancer driver genes were cataloged as driver genes based on a PanCancer and PanSoftware analysis (comprising of all 33 TCGA projects and several computational tools). The set of cancer driver genes for LGG includes the 24 genes reported in Table S3, whose expression profile was retrieved from the expression of the full set of genes from LinkedOmics. We work with the gene expression data corresponding to these 24 LGG-genes as the genomic markers for the 63 subjects.

Prior elicitation. For our statistical analysis of the LGG data, we choose $\boldsymbol{\mu}^{(t,m)} = \alpha \hat{\lambda}_+^{(t-1,m)}$, which prioritizes selection of genes in the previous layer, and Λ as the correlation matrix of X capturing dependence structure between the genes. The choices of some of the other hyperparameters include $v_1 = 100, a_1 = 4, a_2 = 5, \alpha = 0.5, \Psi = I_g, \delta = p^{(t)}$. As v_1 is expected to be very large, we have compensated by placing an informative prior on $v_{kj}^{-2(t,m)}$ with more mass around the value 1 based on our choice of a_1 and a_2. The choice of δ and Ψ are made such that we have a non-informative prior around $\Delta^{(t)}$. Within the estimation for each model for layer t, we choose 10^{-5} as the threshold for convergence. The choice of α is such that we place the
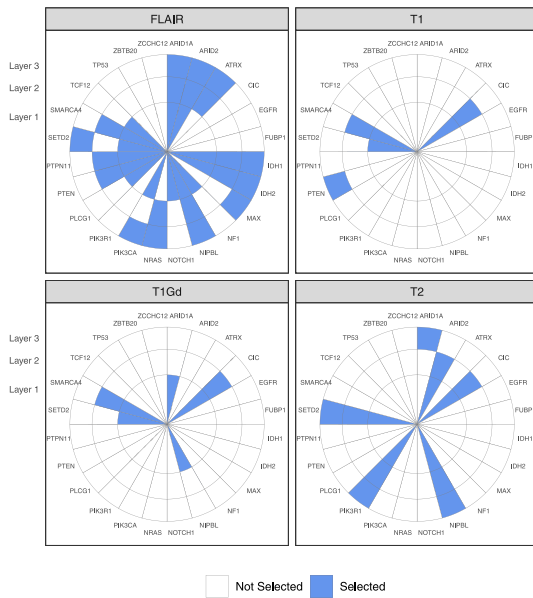
Fig. 8. Associations of layer-wise PC scores from the four imaging sequences with the gene expression. Each panel corresponds to an imaging sequence, and a block filled with color indicates association for that gene with one of the three specific layers ($\tau = 3$).

prior mean at the midpoint of zero and $\hat{\lambda}^{(t-1,m)}$ for the selected genes. Further details about hyperparameter choices are provided in Section S4.4. For model selection, we estimate the model on a grid of values for $v_0 \in \{0.001 + (i-1)0.001 | i = 1, \ldots, 40\}$ and compute the BIC based on each value of v_0 to choose the best model as the one with lowest BIC.

We implement the multiple-multivariate Bayesian regression model with the gene expression of the LGG-genes as predictors and the PC scores from the MRI images as responses. Fig. 8 presents the results of our analysis, where each panel corresponds to one of the four imaging sequences (FLAIR, T1, T1Gd and T2). The inner sphere from the tumor region is represented as the inner circle (layer 1). Similarly, the two spherical shells are represented as concentric annuli (layers 2 and 3). The circular region in the plot is divided into 24 sectors, with each representing one of the 24 LGG-genes under consideration. Consequently, each sector defined by a gene is divided into three concentric blocks corresponding to the three layers. If $\hat{\zeta}_k^{(t,m)} = 1$ for gene k in layer t of imaging sequence m, the corresponding block is colored in blue. Based on our estimation and model selection, if $\hat{\zeta}_k^{(t,m)} = 1$, the expression of gene k has significant association with the PC scores corresponding to layer t from the imaging sequence m. For example, in FLAIR, the block in layer 3 corresponding to the sector for gene ATRX is colored, indicating significant association of the expression of ATRX with PC scores from layer 3 in this imaging sequence.

Figure S13 shows plots corresponding to the Q-function maximization in the EM algorithm. In terms of computation time, our algorithm requires about 20 s for estimation and model selection across $\tau = 3$ layers with $g = 24$ genes and $p^{(1)} = 22, p^{(2)} = 18, p^{(3)} = 17$ PC scores. The analysis was executed on a computer with an Intel(R) Core(TM) i7-7700 CPU @ 3.60 GHz, 3601 MHz, 4 cores, 8 logical processors with 32 GB RAM; model selection for 40 values of v_0 was done in parallel across 4 processors. Under the same settings, generating 200 MCMC samples of the model parameters for each layer sequentially would require about 50 s. We demonstrate the utility of our EM-based estimation compared to a MCMC-based approach. Details of this comparison under different choices for the number of parameters are provided in Section S5.

Biological findings and implications. We now describe the biological significance of the radiogenomic associations identified by

our model. These results are presented separately for each imaging sequence in Fig. 8. Most of the radiogenomic associations appear in FLAIR, which presents most of the tumor regions in LGG with relative hypointensity (dark signal on MRI), except for a hyperintense (bright) rim (Weerakkody et al., 2020). The other three sequences usually present a hypointense signal compared to other brain tissues, with exceptions at certain times. We specifically focus on the following identified associations:

1. The genes IDH1 and IDH2 demonstrate significant associations with all three layers for the FLAIR sequence. Several recent studies (Yan et al., 2009; Leu et al., 2013, 2016) have highlighted the prognostic importance of IDH mutations in gliomas as well as their distinctive genetic and clinical characteristics. Mutations that affect IDH1 were identified in more than 70% of low grade gliomas and in glioblastomas that developed from lower-grade lesions; most of the other tumors often had mutations affecting the gene IDH2 (Yan et al., 2009). These mutations were also shown to be strong predictors for survival, specifically in the case of LGG (Leu et al., 2013). During the developmental stage of most LGG tumors, one of the earliest genetic events is the acquisition of IDH1 or IDH2 mutation (Leu et al., 2016).

2. The PI3K pathway (group of genes), which includes genes such as PTEN, PIK3CA and PIK3R1, is one of the most deregulated and druggable pathways in human cancer (Arafeh and Samuels, 2019). We see associations of the genes PTEN, PIK3CA and PIK3R1 with the spherical shells in both T2 and FLAIR. The PI3K pathway controls multiple cellular processes including metabolism, motility, proliferation, growth and survival (Janku et al., 2018). Somatic mutations in PIK3R1 act as oncogenic driver events in gliomas and are known to increase signaling through the PI3K pathway; they also promote tumorigenesis of primary normal human astrocytes in an orthotopic xenograft model (Quayle et al., 2012).

3. Another significant association across all of the three spherical shells in the FLAIR imaging sequence, and partially in T2, is with the genes ARID1 A and ARID2. Studies have shown that somatic mutations in the chromatin remodeling gene ARID1 A occur in several tumor types (Jones et al., 2012). ARID1 A is known to be a tumor suppressor and inhibits glioma cell proliferation via the PI3K pathway (Zeng et al., 2013). ARID2 is a gene associated with chromatin organization and has been predicted to be a glioma driver (Ceccarelli et al., 2016). Moreover, recent studies have identified ARID2 as a direct target and functional effector in other cancers such as oral cancer (Wu et al., 2020).

4. The genes PTPN11 and NIBPL also present significant associations across all three spherical shells in the FLAIR sequence. Additionally, NIPBL presents association across all spherical shells in T2. NIPBL is a somatically altered glioma gene that is known to be a crucial adherin subunit, and is essential for loading cohesins on chromatin Ceccarelli et al. (2016). PTPN11 is known to mediate gliomagenesis in both mice and humans (Liu et al., 2011).

5. NRAS is a member of the RAS oncogene family, which encodes small enzymes involved in cellular signal transduction (Fiore et al., 2016). In our analysis, NRAS shows significant association with the exterior layers of the tumor in FLAIR. RAS pathways, when activated, trigger downstream signaling pathways such as MAPKs and PI3K/AKT, which modulate cell growth and survival (Schubbert et al., 2007). NRAS was also identified as a gene which promoted oncogenesis in glioma stem cell (Gong et al., 2016).

6. We see significant associations of the gene CIC with the inner and middle layers in T1, T1Gd and T2. CIC is a transcriptional repressor that counteracts activation of genes downstream of receptor tyrosine kinase (RTK)/RAS/ERK signaling

pathways (Bunda et al., 2019). Loss of CIC potentiated the formation and reduced the latency in tumor development in an orthotopic mouse model of glioma (Yang et al., 2017). Tumorigenesis in high-grade gliomas is attributed to hyperactive RTK/RAS/ERK signaling, while CIC has a tumor suppressive function (Bunda et al., 2019). This indicates an inverse functional relationship between CIC and the genes in RTK/RAS/ERK pathways (e.g., NRAS). Interestingly, we see that associations with CIC are picked only when there are no significant associations with NRAS.

These findings indicate several associations between the LGG-genes, and related pathways, with imaging phenotypes from the spherical shells. It is promising that several of these genes have been found to play important roles in the genesis of glioma. It is not surprising that we see most of the associations in the FLAIR sequence, as it exhibits the best contrast between normal brain tissue and presumed infiltrating tumor margins (Grier and Batchelor, 2006). It is also the principal imaging sequence for assessment of LGG growth (Bynevelt et al., 2001). Furthermore, we provide a detailed analysis of consistency in identifying these radiogenomic associations by refitting the model multiple times while including randomly chosen genes as predictors in addition to the LGG-genes. Further details are provided in Section S4.5 of the supplementary material. Results from this analysis (see Figure S14) show that the associations we reported above are consistently being identified even when the LGG-genes are paired with other genes as predictors. These findings encourage a deeper corroboration, which could potentially indicate intricate characteristics of tumor growth.

In the analysis presented above, the tumor region was divided into three spherical shells to conform with the number of natural sub-regions: necrosis or non-enhancing core, edema and enhancing tumor. However, we also deployed our framework with other choices for the number of spherical shells including $\tau = 4, 5$, and 6. We plot these results in Figures S10–S12 in Section S4.2 of the supplementary material. We see that the results are reasonably robust, i.e., most of the genes associated with the layers from the imaging sequences are consistent (except for a few additions/deletions) across different choices of τ. This indicates that our approach is broadly able to identify the underlying radiogenomic associations between the radiomic phenotypes and gene expression.

7. Discussion and future work

In this paper, we propose a new statistical modeling framework to analyze radiological imaging and genomic data that is biologically-informed, mimics the tumor evolution process, and accounts for (structural) tumor heterogeneity. In the context of LGG, we identify molecular determinants of this tumor evolution process by borrowing information sequentially from concentric spherical shells; this process tries to emulate tumor growth since 3D spheroids are known to better mimic real tumors (Breslin and O'Driscoll, 2013). We build layer-wise sequential Bayesian regression models that consider (i) the PC scores constructed from the PDFs of the MRI-based tumor voxel intensities as the responses, which effectively capture tumor heterogeneity, and (ii) the transcriptomic profiling data in terms of the gene expression of LGG-genes as predictors. Information about the multiple imaging sequences, multiple PC scores from a given sequence, and the correlation between the genes was encoded into the prior structure of the regression model. The estimation is based on an computationally-efficient EM algorithm, and it identifies genes which have significant association with tumor heterogeneity in each spherical shell. We have proposed novel methodology in terms of multiple-multivariate regression to simultaneously incorporate complex dependence structure between (i) the genomic covariates, and (ii) the components of the multivariate response. Our model incorporates a sequential variable selection strategy to borrow information across layers of the tumor. Furthermore, this sequential strategy can be deployed to model similar structured data, e.g., any imaging data that mimics a growth process or data where the sequencing arises due to time.

An important aspect of our analysis is that we borrow information about the selection of genes starting from the interior sphere and moving in an outward direction. This approach captures the process of tumor growth which starts as a single cancerous cell and grows radially outward to form the entire tumor. The choice of spherical shapes for the different tumor layers is an uninformative one, due to the lack of prior knowledge about the tumor's growth process. If the tumor's growth dynamics are available, this new information can be used to build the layers. This growth process is incorporated through the parameter $\boldsymbol{\mu}^{(t,m)}$, necessitating sequential estimation. However, we are only interested in the posterior inclusion probabilities for each of the LGG-genes to identify the radiogenomic associations. Hence, instead of using MCMC sampling techniques, we perform estimation using an EM-algorithm, which iteratively searches for the posterior mode of the inclusion probabilities; this, in turn, reduces computational time of estimation. Information about the selection of genes across spherical shells is propagated, and many significant associations in the inner sphere carry their effect forward to subsequent spherical shells. Clearly, this is not always the case, which is reassuring, as these results are not dominated by our choice of prioritizing the previously selected genes through $\boldsymbol{\mu}^{(t,m)}$. Additionally, dependence structure between the genes is incorporated using Λ. We see that gene pairs, such as NIBPL and ARID2 or NF1 and ZBTB20, have high positive correlation (Figure S9), and are either jointly selected or jointly not selected in most layers and imaging sequences. Similarly, the genes ATRX and CIC have negative correlation, and their selection alternates, i.e., one of them is not selected while the other one is selected.

Our analysis identifies several associations between image-based tumor characteristics and expression profiles of LGG-genes. The genes associated with the imaging phenotypes in the inner spherical layers can act as potential biomarkers for early events. Monitoring these selected genes from the onset, or the time of diagnosis, could lead to a better understanding of the molecular underpinnings as well as provide effective diagnostic options prior to invasive approaches such as a biopsy. One of the limitations of our data is that the gene expression might not be derived from the spherical layers separately during the biopsy. This can be potentially addressed in the future using more modern technologies that facilitate analyses using spatial transcriptomics (Burgess, 2019)—an area which is still maturing. In the proposed model setup, the parameter space is high-dimensional due to the multivariate regression setting. This results in a high computational burden during estimation while incorporating an extremely large number of genomic markers within the sequential multivariate regression framework; this issue needs to be explored further using highly scalable or approximate estimation algorithms.

Future work. Our work tries to identify radiogenomic associations between radiomic phenotypes and genomic markers in LGGs by incorporating the biological structure of the tumor into the modeling framework. However, this could be further explored in several directions such as identifying the tumor regions/spherical shells based on known segmentation of the tumor sub-regions. This could additionally be informed by the shape of the tumor, which the PDFs do not capture, as they only assess the heterogeneity in the tumor voxel values. Also, our model explores linear relationships between the imaging-based PC scores and the gene expression, which could potentially be extended to study nonlinear associations (e.g. using semi-parametric additive models Scheipl et al., 2012; Ni et al., 2015). Our future work will include integrating these radiogenomic findings into predictive models for clinical outcomes such as overall survival, time to progression.

Software and supplementary material

We provide the R package *marbles* (MultivARiate Bayesian Layered variablE Selection) that includes relevant data as an online supplementary file. For the most recent version of the package, see www.github.com/shariq-mohammed/marbles. All of the details in the text which were referenced as supplementary material are provided as a separate online document. This document includes details of: (i) construction of layer-wise radiomic phenotypes, and computation of the principal component scores from a sample of probability density functions, (ii) full posterior of the model, and its logarithm, as well as some computational details related to the Expectation-Maximization (EM) algorithm used for estimation, (iii) results from simulation studies, (iv) gene expression, and some additional results for the The Cancer Genome Atlas lower grade glioma data, and (v) comparison of computational burden for the EM-based estimation and model selection versus a Markov Chain Monte Carlo sampling-based approach.

CRediT authorship contribution statement

Shariq Mohammed: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Sebastian Kurtek:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition. **Karthik Bharath:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition. **Arvind Rao:** Conceptualization, Methodology, Writing – review & editing, Resources, Supervision, Project administration, Funding acquisition. **Veerabhadran Baladandayuthapani:** Conceptualization, Methodology, Writing – review & editing, Resources, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We provide open-source software for implementation, in the form of an easy-to-use R package. This is publicly available via GitHub: https://github.com/shariq-mohammed/marbles

Acknowledgments

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.media.2023.102964.

References

Andersen, M.R., Winther, O., Hansen, L.K., 2014. Bayesian inference for structured spike and slab priors. In: Advances in Neural Information Processing Systems. pp. 1745–1753.

Arafeh, R., Samuels, Y., 2019. PIK3CA in cancer: The past 30 years. In: Seminars in Cancer Biology. Elsevier.

Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al., 2018. Comprehensive characterization of cancer driver genes and mutations. Cell 173 (2), 371–385.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C., 2017a. Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. http://dx.doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C., 2017b. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Nat. Sci. Data 4, 170117.

Bakas, S., Zeng, K., Sotiras, A., Rathore, S., Akbari, H., Gaonkar, B., Rozycki, M., Pati, S., Davatzikos, C., 2015. GLISTRboost: combining multimodal MRI segmentation, registration, and biophysical tumor growth modeling with gradient boosting machines for glioma segmentation. In: International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer, pp. 144–155.

Breslin, S., O'Driscoll, L., 2013. Three-dimensional cell culture: the missing link in drug discovery. Drug Discov. Today 18 (5–6), 240–249.

Bunda, S., Heir, P., Metcalf, J., Li, A.S.C., Agnihotri, S., Pusch, S., Yasin, M., Li, M., Burrell, K., Mansouri, S., et al., 2019. CIC protein instability contributes to tumorigenesis in glioblastoma. Nature Commun. 10 (1), 1–17.

Burgess, D.J., 2019. Spatial transcriptomics coming of age. Nature Rev. Genet. 20 (6), 317.

Bynevelt, M., Britton, J., Seymour, H., MacSweeney, E., Thomas, N., Sandhu, K., 2001. FLAIR imaging in the follow-up of low-grade gliomas: time to dispense with the dual-echo? Neuroradiology 43 (2), 129–133.

Ceccarelli, M., Barthel, F.P., Malta, T.M., Sabedot, T.S., Salama, S.R., Murray, B.A., Morozova, O., Newton, Y., Radenbaugh, A., Pagnotta, S.M., et al., 2016. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. Cell 164 (3), 550–563.

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al., 2013. The cancer imaging archive (TCIA): maintaining and operating a public information repository. J. Digit. Imaging 26 (6), 1045–1057.

Claus, E.B., Walsh, K.M., Wiencke, J.K., Molinaro, A.M., Wiemels, J.L., Schildkraut, J.M., Bondy, M.L., Berger, M., Jenkins, R., Wrensch, M., 2015. Survival and low-grade glioma: the emergence of genetic information. Neurosurg. Focus 38 (1), E6.

Costa, E.C., Moreira, A.F., de Melo-Diogo, D., Gaspar, V.M., Carvalho, M.P., Correia, I.J., 2016. 3D tumor spheroids: an overview on the tools and techniques used for their analysis. Biotechnol. Adv. 34 (8), 1427–1441.

Du, L., Huang, H., Yan, J., Kim, S., Risacher, S.L., Inlow, M., Moore, J.H., Saykin, A.J., Shen, L., Alzheimer's Disease Neuroimaging Initiative, 2016. Structured sparse canonical correlation analysis for brain imaging genetics: an improved GraphNet method. Bioinformatics 32 (10), 1544–1551.

Du, L., Liu, K., Yao, X., Risacher, S.L., Han, J., Saykin, A.J., Guo, L., Shen, L., 2020. Detecting genetic associations with brain imaging phenotypes in Alzheimer's disease via a novel structured SCCA approach. Med. Image Anal. 61, 101656.

Elshafeey, N., Kotrotsou, A., Hassan, A., Elshafei, N., Hassan, I., Ahmed, S., Abrol, S., Agarwal, A., El Salek, K., Bergamaschi, S., et al., 2019. Multicenter study demonstrates radiomic features derived from magnetic resonance perfusion images identify pseudoprogression in glioblastoma. Nature Commun. 10 (1), 1–9.

Fiore, D., Donnarumma, E., Roscigno, G., Iaboni, M., Russo, V., Affinito, A., Adamo, A., De Martino, F., Quintavalle, C., Romano, G., et al., 2016. miR-340 predicts glioblastoma survival and modulates key cancer hallmarks through down-regulation of NRAS. Oncotarget 7 (15), 19531.

George, E.I., McCulloch, R.E., 1997. Approaches for Bayesian variable selection. Statist. Sinica 339–373.

Gevaert, O., Mitchell, L.A., Achrol, A.S., Xu, J., Echegaray, S., Steinberg, G.K., Cheshier, S.H., Napel, S., Zaharchuk, G., Plevritis, S.K., 2014. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. Radiology 273 (1), 168–174.

Gong, W., Zheng, J., Liu, X., Ma, J., Liu, Y., Xue, Y., 2016. Knockdown of NEAT1 restrained the malignant progression of glioma stem cells by activating microRNA let-7e. Oncotarget 7 (38), 62208–62223.

Grier, J.T., Batchelor, T., 2006. Low-grade gliomas in adults. Oncologist 11 (6), 681–693.

Gupta, A.K., Nagar, D.K., 2018. Matrix Variate Distributions. Chapman and Hall/CRC.

Ishwaran, H., Rao, J.S., et al., 2005. Spike and slab variable selection: frequentist and Bayesian strategies. Ann. Statist. 33 (2), 730–773.

Janku, F., Yap, T.A., Meric-Bernstam, F., 2018. Targeting the PI3K pathway in cancer: are we making headway? Nat. Rev. Clin. Oncol. 15 (5), 273–291.

Jones, S., Li, M., Parsons, D.W., Zhang, X., Wesseling, J., Kristel, P., Schmidt, M.K., Markowitz, S., Yan, H., Bigner, D., et al., 2012. Somatic mutations in the chromatin remodeling gene ARID1A occur in several tumor types. Hum. Mutat. 33 (1), 100–103.

Just, N., 2014. Improving tumour heterogeneity MRI assessment with histograms. Br. J. Cancer 111 (12), 2205–2213.

Leu, S., von Felten, S., Frank, S., Boulay, J.-L., Mariani, L., 2016. IDH mutation is associated with higher risk of malignant transformation in low-grade glioma. J. Neuro-Oncol. 127 (2), 363–372.

Leu, S., von Felten, S., Frank, S., Vassella, E., Vajtai, I., Taylor, E., Schulz, M., Hutter, G., Hench, J., Schucht, P., et al., 2013. IDH/MGMT-driven molecular classification of low-grade glioma is a strong predictor for long-term survival. Neuro-Oncol. 15 (4), 469–479.

Li, F., Zhang, N.R., 2010. Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. J. Amer. Statist. Assoc. 105 (491), 1202–1214.

Liu, K.-W., Feng, H., Bachoo, R., Kazlauskas, A., Smith, E.M., Symes, K., Hamilton, R.L., Nagane, M., Nishikawa, R., Hu, B., et al., 2011. SHP-2/PTPN11 mediates glioma-genesis driven by PDGFRA and INK4A/ARF aberrations in mice and humans. J. Clin. Invest. 121 (3), 905–917.

Lu, C.-F., Hsu, F.-T., Hsieh, K.L.-C., Kao, Y.-C.J., Cheng, S.-J., Hsu, J.B.-K., Tsai, P.-H., Chen, R.-J., Huang, C.-C., Yen, Y., et al., 2018. Machine learning–based radiomics for molecular subtyping of gliomas. Clin. Cancer Res. 24 (18), 4429–4436.

Mazurowski, M.A., Clark, K., Czarnek, N.M., Shamsesfandabadi, P., Peters, K.B., Saha, A., 2017. Radiogenomics of lower-grade glioma: algorithmically-assessed tumor shape is associated with tumor genomic subtypes and patient outcomes in a multi-institutional study with the cancer genome atlas data. J. Neuro-Oncol. 133 (1), 27–35.

Mohammed, S., Bharath, K., Kurtek, S., Rao, A., Baladandayuthapani, V., 2021. RADIOHEAD: Radiogenomic analysis incorporating tumor heterogeneity in imaging through densities. Ann. Appl. Statist. 15 (4), 1808–1830.

Morris, J.S., Baladandayuthapani, V., 2017. Statistical contributions to bioinformatics: Design, modelling, structure learning and integration. Statist. Model. 17 (4–5), 245–289.

Ni, Y., Stingo, F.C., Baladandayuthapani, V., 2015. Bayesian nonlinear model selection for gene regulatory networks. Biometrics 71 (3), 585–595.

Ohgaki, H., Dessen, P., Jourde, B., Horstmann, S., Nishikawa, T., Di Patre, P.-L., Burkhard, C., Schüler, D., Probst-Hensch, N.M., Maiorka, P.C., et al., 2004. Genetic pathways to glioblastoma: a population-based study. Cancer Res. 64 (19), 6892–6899.

Ohgaki, H., Kleihues, P., 2007. Genetic pathways to primary and secondary glioblastoma. Amer. J. Pathol. 170 (5), 1445–1453.

Quayle, S.N., Lee, J.Y., Cheung, L.W.T., Ding, L., Wiedemeyer, R., Dewan, R.W., Huang-Hobbs, E., Zhuang, L., Wilson, R.K., Ligon, K.L., et al., 2012. Somatic mutations of PIK3R1 promote gliomagenesis. PLoS One 7 (11).

Ročková, V., George, E.I., 2014. EMVS: The EM approach to Bayesian variable selection. J. Amer. Statist. Assoc. 109 (506), 828–846.

Saha, A., Banerjee, S., Kurtek, S., Narang, S., Lee, J., Rao, G., Martinez, J., Bharath, K., Rao, A.U., Baladandayuthapani, V., 2016. DEMARCATE: Density-based magnetic resonance image clustering for assessing tumor heterogeneity in cancer. NeuroImage: Clin. 12, 132–143.

Scheipl, F., Fahrmeir, L., Kneib, T., 2012. Spike-and-slab priors for function selection in structured additive regression models. J. Amer. Statist. Assoc. 107 (500), 1518–1532.

Schubbert, S., Shannon, K., Bollag, G., 2007. Hyperactive ras in developmental disorders and cancer. Nat. Rev. Cancer 7 (4), 295–308.

Shinohara, R.T., Sweeney, E.M., Goldsmith, J., Shiee, N., Mateen, F.J., Calabresi, P.A., Jarso, S., Pham, D.L., Reich, D.S., Crainiceanu, C.M., 2014. Statistical normalization techniques for magnetic resonance imaging. NeuroImage: Clin. 6, 9–19.

Srivastava, A., Jermyn, I.H., Joshi, S.H., 2007. Riemannian analysis of probability density functions with applications in vision. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8.

Swanson, K.R., Bridge, C., Murray, J., Alvord, Jr., E.C., 2003. Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion. J. Neurol. Sci. 216 (1), 1–10.

Ueda, N., Nakano, R., 1998. Deterministic annealing EM algorithm. Neural Netw. 11 (2), 271–282.

Vannucci, M., Stingo, F.C., 2010. Bayesian models for variable selection that incorporate biological information. Bayesian Statist. 9, 1–20.

Vasaikar, S.V., Straub, P., Wang, J., Zhang, B., 2017. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. Nucl. Aids Res. 46 (D1), D956–D963.

Wang, W., Baladandayuthapani, V., Morris, J.S., Broom, B.M., Manyam, G., Do, K.-A., 2013. IBAG: integrative Bayesian analysis of high-dimensional multiplatform genomics data. Bioinformatics 29 (2), 149–159.

Weerakkody, Y., Gaillard, F., et al., 2020. Anaplastic astrocytoma: radiology reference article. www.radiopaedia.org/articles/anaplastic-astrocytoma. Accessed: October 6, 2020.

Wu, M., Duan, Q., Liu, X., Zhang, P., Fu, Y., Zhang, Z., Liu, L., Cheng, J., Jiang, H., 2020. Mir-155-5p promotes oral cancer progression by targeting chromatin remodeling gene ARID2. Biomed. Pharmacotherapy 122, 109696.

Yan, H., Parsons, D.W., Jin, G., McLendon, R., Rasheed, B.A., Yuan, W., Kos, I., Batinic-Haberle, I., Jones, S., Riggins, G.J., et al., 2009. IDH1 and IDH2 mutations in gliomas. N. Engl. J. Med. 360 (8), 765–773.

Yang, H., Baladandayuthapani, V., Rao, A.U., Morris, J.S., 2020. Quantile function on scalar regression analysis for distributional data. J. Amer. Statist. Assoc. 115 (529), 90–106.

Yang, R., Chen, L.H., Hansen, L.J., Carpenter, A.B., Moure, C.J., Liu, H., Pirozzi, C.J., Diplas, B.H., Waitkus, M.S., Greer, P.K., et al., 2017. CIC loss promotes gliomagenesis via aberrant neural stem cell proliferation and differentiation. Cancer Res. 77 (22), 6097–6108.

Zeng, Y., Liu, Z., Yang, J., Liu, Y., Huo, L., Li, Z., Lan, S., Wu, J., Chen, X., Yang, K., et al., 2013. ARID1A is a tumour suppressor and inhibits glioma cell proliferation via the PI3K pathway. Head Neck Oncol. 5 (1), 6.