

RESEARCH

Open Access

Integrating multi-platform genomic data using hierarchical Bayesian relevance vector machines

Sanvesh Srivastava¹, Wenyi Wang², Ganiraju Manyam², Carlos Ordonez³
and Veerabhadran Baladandayuthapani^{4*}

Abstract

Background: Recent advances in genome technologies and the subsequent collection of genomic information at various molecular resolutions hold promise to accelerate the discovery of new therapeutic targets. A critical step in achieving these goals is to develop efficient clinical prediction models that integrate these diverse sources of high-throughput data. This step is challenging due to the presence of high-dimensionality and complex interactions in the data. For predicting relevant clinical outcomes, we propose a flexible statistical machine learning approach that acknowledges and models the interaction between platform-specific measurements through nonlinear kernel machines and borrows information within and between platforms through a hierarchical Bayesian framework. Our model has parameters with direct interpretations in terms of the effects of platforms and data interactions within and across platforms. The parameter estimation algorithm in our model uses a computationally efficient variational Bayes approach that scales well to large high-throughput datasets.

Results: We apply our methods of integrating gene/mRNA expression and microRNA profiles for predicting patient survival times to The Cancer Genome Atlas (TCGA) based glioblastoma multiforme (GBM) dataset. In terms of prediction accuracy, we show that our non-linear and interaction-based integrative methods perform better than linear alternatives and non-integrative methods that do not account for interactions between the platforms. We also find several prognostic mRNAs and microRNAs that are related to tumor invasion and are known to drive tumor metastasis and severe inflammatory response in GBM. In addition, our analysis reveals several interesting mRNA and microRNA interactions that have known implications in the etiology of GBM.

Conclusions: Our approach gains its flexibility and power by modeling the non-linear interaction structures between and within the platforms. Our framework is a useful tool for biomedical researchers, since clinical prediction using multi-platform genomic information is an important step towards personalized treatment of many cancers. We have a freely available software at: <http://odin.mdacc.tmc.edu/~vbaladan>.

Keywords: Bayesian modeling; Multiple kernel learning; Genomics; High-dimensional data analysis; Prediction; Variational inference

*Correspondence: veera@mdanderson.org

⁴Department of Biostatistics, Division of Quantitative Sciences, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Unit 1411, Houston, Texas, USA

Full list of author information is available at the end of the article

1 Introduction

Recent advances in genome technologies such as microarrays and next-generation sequencing have enabled the measurement of genomic activity at a very detailed resolution (e.g., base pair, single-nucleotide polymorphisms) as well as across multiple molecular levels: the epigenome, transcriptome and proteome. The collection of genomic information at various resolutions holds promise to accelerate the amalgamation of discovery science and clinical medicine [1]. One of the overarching goals of such studies is to relate these genomic data to relevant (patient-specific) clinical outcomes, not only to find significant biomarkers of disease progression/evolution but also to use the biomarkers to develop prediction models for deployment in future therapeutic studies. Furthermore, genomic data are now available from multiple platforms and resolutions for the same individual, thus allowing a researcher to simultaneously query these multiple sources of data to achieve these goals. Such motivating data have been collected under the aegis of The Cancer Genome Atlas (TCGA) project, wherein data from multiple genomic platforms such as gene/mRNA expression, DNA copy number, methylation and microRNA expression profiles are available for multiple tumor types (see <http://cancergenome.nih.gov> for more details). In addition, the available clinical information, such as stage of disease and survival times, motivates the analytic frameworks that integrate patient-specific data.

One of the main challenges in modeling the statistical dependence between such high-throughput studies is that a large number of measurements (usually in thousands) is available for a relatively small number (usually in tens or hundreds) of patient samples; therefore, classical statistical approaches based on linear models and hierarchical clustering are prone to over-fitting [2,3]. In these situations, [3] recommends accounting for high-dimensionality by using approaches that borrow information across covariates to compensate for the limited information available across samples, which leads to better and more reliable inference. Several approaches have been developed to address these challenges in various contexts. Some examples include linear parametric models and hierarchical clustering for inferring the relation between phenotypes and genomic features [4], hierarchical Bayesian modeling approaches based on linear shrinkage estimators [5], linear canonical correlation analysis [6], intensity-based approaches for merging datasets [7], and regularized linear regression approaches [8].

Although these approaches are computationally efficient, interpretable, and simple, they make two unrealistic assumptions for practical data analysis. First, due to the parametric and linear assumptions, they might miss the underlying non-linear patterns in the data. Second, and more importantly, these non-linear patterns are further

amplified in the presence of complex interactions within and between the different platforms that must be modeled while integrating data from these platforms. In this paper, we present a statistical machine learning approach called the hierarchical relevance vector machines (H-RVM) to address these modeling and inferential challenges. Briefly, the framework presented here: (a) models the relation between a relevant clinical outcome (scalar) and high-dimensional covariates/features through a data-adaptive and flexible nonparametric approach, (b) borrows information within and between platforms through a hierarchical Bayesian framework, (c) acknowledges and models the interaction between platforms through nonlinear kernel-based functionals, (d) has parameters that have explicit interpretation as the effects of the platforms and their interactions on the outcome, and (e) uses a computationally efficient variational Bayes approach that can be readily scaled to large datasets.

Our methods are motivated by and applied to a TCGA based glioblastoma multiforme (GBM) dataset, for which we integrate gene (mRNA) and microRNA (miRNA) expression profiles to predict patient survival times^a. There is an increasing interest in identifying subtypes of GBM based on its gene expression data. The ultimate goal of subtyping GBM is to identify gene expression profiles that are prognostic or predictive of treatment outcomes. The known subtypes of GBM samples in TCGA include pro-neural, neural, classical, and mesenchymal; with the first two classes of which are suspected to differ from the other two in the cell of origin, which is a critical determinant of effective treatment regimens [9]. Differential expressions of miRNAs were recently found to be associated with many diseases, including cancers [10,11]. Previous studies have shown that combining multiple types of data, such as mRNA and miRNA expressions, could significantly improve the accuracy of detecting GBM subtypes, and thereby potentially predict the clinical outcomes [12]. However, methods are lacking to accurately model the effect of interactions between these data types directly on clinical outcomes. Here we show that our nonlinear and interaction-based integrative methods have better prediction accuracy than linear alternatives and non-integrative methods that do not account for the interactions between the platforms. We also find several prognostic mRNAs and microRNAs that are related to tumor invasion and that are known to drive tumor metastasis and severe inflammatory response in GBM. In addition, our analysis reveals several interesting mRNA-miRNA interactions that have known implications in the etiology of GBM. The paper is structured as follows. The basic construction of H-RVM is detailed in Section 2. The analysis of GBM data is presented in Section 3, and concluding remarks about the H-RVM framework are presented in Section 4.

2 Hierarchical Relevance Vector Machine model

For ease of exposition, we illustrate the model building process of H-RVM using data from two sources: gene/mRNA and miRNA expression measurements. The framework is easily extended to multiple platforms as discussed in Section 4. Suppose, we have data for N patients, and \mathbf{X} and \mathbf{Y} represent mean-centered and -standardized gene and miRNA expression matrices, with rows corresponding to patients and columns representing the G genes and M miRNAs, respectively^b. Centering and standardizing the gene and miRNA expression matrices remove any systematic mean or scaling effects caused by the use of different data sources, and make them compatible for model fitting. We denote the gene and miRNA expression for the i -th patient as row vectors $\mathbf{x}_i^T = (x_{i1}, \dots, x_{iG})$ and $\mathbf{y}_i^T = (y_{i1}, \dots, y_{iM})$. These covariates are high-dimensional, that is, both G and M are much larger than N ; for example, in the GBM data $G \approx 12000, M \approx 540, N \approx 250$. Based on these measurements, our aim is to predict a relevant clinical outcome, which in our case is the (log-transformed) survival time measured from time of diagnosis to death, denoted by the column vector $\mathbf{t} = (t_1, \dots, t_N)$ for the N patients.

2.1 Basic construction

A basic (conceptual) model can be written in a high-dimensional regression setting as,

$$t_i = \alpha_0 + f_{\mathbf{x}}(\mathbf{x}_i, \boldsymbol{\alpha}_1) + f_{\mathbf{y}}(\mathbf{y}_i, \boldsymbol{\alpha}_2) + f_{(\mathbf{x} \otimes \mathbf{y})}(\mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\alpha}_3) + \epsilon_i, \quad (1)$$

where α_0 is the overall mean effect and ϵ_i is the random error; $f(\bullet)$'s, generally referred to as *basis functions*, are chosen to achieve a desired level of flexibility for modeling the effects of \mathbf{X} , \mathbf{Y} , and their functions on \mathbf{t} . Of these functions, $f_{(\mathbf{x} \otimes \mathbf{y})}$ models the *interactions* between \mathbf{X} and \mathbf{Y} , and the remaining basis functions, $f_{\mathbf{x}}$ and $f_{\mathbf{y}}$, respectively, model the main effects of \mathbf{X} and \mathbf{Y} for predicting \mathbf{t} . In most situations the regression coefficients, $\boldsymbol{\alpha} = (\alpha_0, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)$, linearly relate the covariate effects (i.e., values of the basis functions evaluated at the covariates) to the response. Linear regression is a special case of (1) when all the basis functions are linear, and the response for the i -th patient,

$$t_i = \mathbf{x}_i^T \boldsymbol{\alpha}_1 + \mathbf{y}_i^T \boldsymbol{\alpha}_2 + (\mathbf{x}_i \otimes \mathbf{y}_i)^T \boldsymbol{\alpha}_3 + \epsilon_i, \quad (2)$$

where $(\mathbf{x}_i \otimes \mathbf{y}_i) = (x_{i1}y_{i1}, \dots, x_{i1}y_{iM}, \dots, x_{iG}y_{i1}, \dots, x_{iG}y_{iM})$ models the first order interactions between genes and miRNAs and $\alpha_0 = 0$ because of the centered covariates. Further, due to the high-dimensional covariates \mathbf{x}_i 's and \mathbf{y}_i 's, a penalty is imposed on the regression coefficients $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)$ to avoid overfitting. The most popular of such penalties is the Lasso because it has many desirable properties for high-dimensional linear regression and variable selection [13,14]. Although (2) with a Lasso penalty is a popular choice for high-dimensional regression, the linearity of the basis functions imposes serious

restrictions on the flexibility of the model. For example, (2) does not model nonlinear covariate effects as well as second or higher order interactions between genes and miRNAs.

Through H-RVM, we propose a regression model as a special case of (1), using kernel-based functions to respectively model $f_{\mathbf{x}}, f_{\mathbf{y}}$, and $f_{(\mathbf{x} \otimes \mathbf{y})}$. The kernel functions incorporate nonlinear effects of possible interactions within and between high-dimensional gene and miRNA expression measurements. Further, H-RVM estimates the respective contributions of genes, miRNAs, and their interactions in predicting survival times, which is of primary importance in developing novel drug targets. H-RVM posits the following regression of \mathbf{t} on \mathbf{X} and \mathbf{Y} for the i -th patient:

$$t_i = \beta_1 f_{\mathbf{x}}\{\mathbf{x}_i, \boldsymbol{\alpha}_1\} + \beta_2 f_{\mathbf{y}}\{\mathbf{y}_i, \boldsymbol{\alpha}_2\} + \beta_3 f_{(\mathbf{x} \otimes \mathbf{y})}\{(\mathbf{x} \otimes \mathbf{y})_i, \boldsymbol{\alpha}_3\} + \epsilon_i, \quad (3)$$

where $(\mathbf{x} \otimes \mathbf{y})_i = (x_{i1}, \dots, x_{iG}, y_{i1}, \dots, y_{iM})$ is a vector of length $G + M$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ is such that its components lie on a probability simplex i.e. $\beta_m > 0$ for $m = 1, 2, 3$ and $\sum_{m=1}^3 \beta_m = 1$. H-RVM posits different kernels for the data sources and combines them through weights $\boldsymbol{\beta}$. The model parameters have the following interpretation:

- The kernel functions $f_{\mathbf{x}}(\bullet)$ and $f_{\mathbf{y}}(\bullet)$ model all possible interactions among genes and among miRNAs, respectively, and $f_{(\mathbf{x} \otimes \mathbf{y})}(\bullet)$ models all possible interactions between genes and miRNAs. The three kernels together account for the high-dimensionality and non-linearity of the covariate effects of \mathbf{X} and \mathbf{Y} by embedding them in the space of kernels.
- The m -th component of $\boldsymbol{\beta}$, β_m , denotes the influence of the m -th source on predicting the log survival time and has the following interpretation: if $\boldsymbol{\beta} = \{1, 0, 0\}$, then (3) corresponds to a functional regression model that predicts \mathbf{t} (log survival time) with only \mathbf{x} (gene expressions) as covariates. Conversely, if $\boldsymbol{\beta} = \{1/3, 1/3, 1/3\}$, then (3) corresponds to a regression model, with the platforms and their interactions contributing equally to the prediction of the survival time. In reality, we expect (and show) different contributions from each platform and estimate these weights from the data.

The task now remains to explicitly characterize the functions $f_{\mathbf{x}}(\bullet)$, $f_{\mathbf{y}}(\bullet)$ and $f_{(\mathbf{x} \otimes \mathbf{y})}(\bullet)$ using multiple kernels, as detailed below.

2.2 Multiple kernel learning

Kernel learning (KL) is an approach for nonparametric classification and regression that can be used for predicting \mathbf{t} based on \mathbf{X} and \mathbf{Y} [14]. First, for simplicity, assume

that we want to predict \mathbf{t} based on \mathbf{X} . KL posits the following relation between \mathbf{t} and \mathbf{X}

$$t_i = \alpha_0 + \sum_{j=1}^N \alpha_j \mathbf{K}(\mathbf{x}_j, \mathbf{x}_i | \sigma^2) + \epsilon_i \Rightarrow \mathbf{t} = \mathbf{K}^T \boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad (4)$$

where σ^2 is a kernel-specific “bandwidth” parameter and depends on the choice of kernel, $K(\bullet)$ (detailed later in the section) and $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_N)$ is the (white-noise) error. The primary parameter of interest is $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_N)^T$, and $\alpha_1, \dots, \alpha_N$ correspond to weights assigned to the features for N \mathbf{x}_j 's. Support Vector Machine (SVM) and Relevance Vector Machine (RVM) are canonical examples of KL [14]. We prefer RVM because of its probabilistic interpretation and other optimal properties compared to those of SVM [15]. There are cases, however, where one feature matrix may not fit the data well. Based on this observation, Multiple Kernel Learning (MKL) extends (4) and replaces \mathbf{K} by a weighted average of L feature matrices $\{\mathbf{K}_l\}_{l=1}^L$,

$$t_i = \alpha_0 + \sum_{j=1}^N \alpha_j \sum_{l=1}^L \beta_l \mathbf{K}_l(\mathbf{x}_j, \mathbf{x}_i | \sigma_l^2) + \epsilon_i \implies \quad (5)$$

$$\mathbf{t} = \beta_1 \mathbf{K}_1^T \boldsymbol{\alpha} + \dots + \beta_l \mathbf{K}_l^T \boldsymbol{\alpha} + \dots + \beta_L \mathbf{K}_L^T \boldsymbol{\alpha} + \boldsymbol{\epsilon}.$$

MKL improves the flexibility of KL by introducing L bandwidth parameters $\{\sigma_l^2\}_{l=1}^L$ and L weights for feature matrices $\boldsymbol{\beta} = (\beta_1, \dots, \beta_L)^T$. A variety of approaches exist to learn $\{\sigma_l^2\}_{l=1}^L$, $\boldsymbol{\beta}$, and $\boldsymbol{\alpha}$ for MKL (for details see [14,16,17]). Note that in all these works the data source (i.e., \mathbf{X}) remains the same for both KL and MKL. The H-RVM framework developed in this article extends KL to include multiple data sources and their interactions, and uses a learning algorithm similar to the MKL framework.

Because the three data sources (gene expressions, miRNA expressions, and their interactions) can be used separately for predicting the log survival time, it is reasonable to combine their predictions to obtain more reliable estimates. To this end, H-RVM combines respective predictions obtained from different sources obtained using KL (4) through a weighted average, and chooses appropriate weights using MKL (5). Similar to (4), $\mathbf{K}_1^T \boldsymbol{\alpha}$, $\mathbf{K}_2^T \boldsymbol{\alpha}$, and $\mathbf{K}_3^T \boldsymbol{\alpha}$ are the predicted values of \mathbf{t} that correspond to genes, miRNAs, and their interactions, respectively. Using (5), we combine the predictions $\{\mathbf{K}_i^T \boldsymbol{\alpha}\}_{i=1}^3$ through the weight vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ to model \mathbf{t} as

$$\mathbf{t} = (\beta_1 \mathbf{K}_1^T + \beta_2 \mathbf{K}_2^T + \beta_3 \mathbf{K}_3^T) \boldsymbol{\alpha} + \boldsymbol{\epsilon} = \mathbf{K}_\beta^T \boldsymbol{\alpha} + \boldsymbol{\epsilon}. \quad (6)$$

We further constrain $\boldsymbol{\beta}$ such that its components lie on a probability simplex, i.e., $\sum_{m=1}^3 \beta_m = 1$. This constraint ensures that the joint (convolved) kernel, \mathbf{K}_β , is positive definite and that β_m denotes the influence of the m -th source in predicting the log survival time. Note that H-RVM is a special case of (3) with $f_x(\mathbf{x}_i, \boldsymbol{\alpha}) \equiv \mathbf{k}_{1,i}^T \boldsymbol{\alpha}$,

$f_y(\mathbf{y}_i, \boldsymbol{\alpha}) \equiv \mathbf{k}_{2,i}^T \boldsymbol{\alpha}$, and $f_{(\mathbf{x} \otimes \mathbf{y})}((\mathbf{x} \otimes \mathbf{y})_i, \boldsymbol{\alpha}) \equiv \mathbf{k}_{3,i}^T \boldsymbol{\alpha}$, where $\mathbf{k}_{m,i}$ is the i -th column of \mathbf{K}_m . Given $\{\mathbf{K}_i\}_{i=1}^3$, MKL can be used to learn $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Although similar to (5), (6) differs in two important ways. First, (6) obtains kernels using (4) for different data sources, namely gene expression, miRNA expression, and their interaction. Second, we allow for dependence between data sources via the interaction kernel (\mathbf{K}_3), but MKL does not; instead MKL uses a convex combination of the different kernels from the same data source to aid prediction.

The learning algorithm of H-RVM is independent of the choice of kernels, but in this work we use a Gaussian radial basis function (RBF) kernel (denoted by \mathbf{K}) [14]. The RBF kernel maps the m -th high-dimensional covariate to its feature space that is represented as feature matrix \mathbf{K}_m . The feature matrices \mathbf{K}_1 , \mathbf{K}_2 , and \mathbf{K}_3 correspond to genes, miRNAs, and interactions, and their (i, j) -th entries are as follows:

$$\begin{aligned} (\mathbf{K}_1)_{ij} &= e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_1^2}} = \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j | \sigma_1^2), \\ (\mathbf{K}_2)_{ij} &= e^{-\frac{\|\mathbf{y}_i - \mathbf{y}_j\|^2}{2\sigma_2^2}} = \mathbf{K}(\mathbf{y}_i, \mathbf{y}_j | \sigma_2^2), \\ (\mathbf{K}_3)_{ij} &= e^{-\frac{\|(\mathbf{x} \otimes \mathbf{y})_i - (\mathbf{x} \otimes \mathbf{y})_j\|^2}{2\sigma_3^2}} = \mathbf{K}((\mathbf{x} \otimes \mathbf{y})_i, (\mathbf{x} \otimes \mathbf{y})_j | \sigma_3^2), \end{aligned} \quad (7)$$

where σ_m^2 is the “bandwidth” parameter of the m -th kernel matrix and is chosen *a priori* through cross-validation (see [14] for details). The other choices of kernels include polynomial kernels and matern kernels [18]. To account for the overall mean (or intercept) in (1), an extra row of 1's is appended to the feature matrices in (7); therefore, $\{\mathbf{K}_i\}_{i=1}^3$ hereafter have dimensions $(N + 1) \times N$.

2.3 Generative Bayesian model for H-RVM

H-RVM reformulates (6) as a hierarchical Bayesian model for greater flexibility and better interpretation of its parameters. This reformulation serves two important purposes. First, H-RVM is interpreted as a hierarchical Bayesian extension of RVM [15], which is a special case of Bayesian KL. Second, instead of using MKL methods, H-RVM learns parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ from \mathbf{t} , \mathbf{X} , and \mathbf{Y} using the variational learning algorithm of hierarchical kernel learning (HKL) [14,16].

H-RVM posits the following generative model for the (noisy) log survival time measurements \mathbf{t} . Similar to MKL, $\mathbf{K}_\beta^T \boldsymbol{\alpha}$ represents the mean of \mathbf{t} . The error distribution is Gaussian with mean 0 and precision parameter γ (8). Further, we impose a Gamma prior on γ such that

$$\mathbf{t} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \mathbf{X}, \mathbf{Y} \sim \mathcal{N}(\mathbf{t} | \mathbf{K}_\beta^T \boldsymbol{\alpha}, \gamma^{-1} \mathbf{I}), \quad (8)$$

$$\gamma \sim \text{Gamma}(\gamma | c_\gamma, d_\gamma), \quad (9)$$

where $\mathcal{N}(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and $\text{Gamma}(\cdot|c_{\bullet}, d_{\bullet})$ represents a Gamma distribution with respective shape and rate parameters c_{\bullet} and d_{\bullet} .

Motivated by the “automatic relevance determination” idea of RVM, we impose independent Gaussian priors on the α_j 's with the same mean 0 and different precision parameters ϕ_j 's (10), where ϕ_j controls (*a priori*) predictive power of the j -th feature vector from the three data sources for the log survival time. A large ϕ_j indicates low predictive power. We also impose independent Gamma priors on the ϕ_j 's,

$$\boldsymbol{\alpha}|\boldsymbol{\phi} \sim \mathcal{N}(\boldsymbol{\alpha}|\mathbf{0}, \text{diag}(\boldsymbol{\phi}^{-1})), \quad (10)$$

$$\boldsymbol{\phi} \sim \prod_{j=0}^N \text{Gamma}(\phi_j|c_{\phi}, d_{\phi}), \quad (11)$$

where $\boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_n)$. This setting forces many α_j 's *a posteriori* to be near 0 with high precision. Most of the variance in \mathbf{t} is explained by a small number of feature vectors that correspond to nonzero α_j 's. These feature vectors are the “relevance vectors” of H-RVM that have the following three characteristics: they prevent over-fitting, represent a parsimonious description of the data, and correspond to feature vectors that are most predictive of the log survival time. An equivalent prior setting is found by marginalizing the ϕ_j 's from the joint distribution of $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ above, which imposes a multivariate Student's t prior on $\boldsymbol{\alpha}$ with mean $\mathbf{0}$.

Finally, we impose a Dirichlet prior on $\boldsymbol{\beta}$ to ensure that its components lie on a probability simplex:

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3) \sim \text{Dirichlet}(\boldsymbol{\beta}|a_1, a_2, a_3), \quad (12)$$

where the m -th component of $\boldsymbol{\beta}$, β_m , denotes the influence of m -th source in predicting the log survival time.

The hierarchical Bayesian model (8) – (12) specifies a complete sampling model for the H-RVM framework. It can also be interpreted as a probabilistic approach for combining the predictions of log survival times from the three RVMs respectively corresponding to gene expressions, miRNA expressions, and their interactions. H-RVM introduces an additional hierarchy and combines the predictions of these three RVMs as a weighted average, with the weights generated from a Dirichlet distribution (12). The increased flexibility of H-RVM over RVM comes at the cost of analytic intractability of the posterior distributions of the H-RVM parameters. Estimation of the posterior distributions of the H-RVM's parameters can proceed via either simulation-based Markov chain Monte Carlo (MCMC) approaches or deterministic variational Bayes approaches. Given the complexity and size of high-throughput data in general and

GBM data in particular, MCMC approaches tend to be computationally expensive and slow. We employ variational Bayes methods from HKL [16] and obtain the analytically tractable variational posterior distribution, $q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\phi}, \gamma|\mathbf{t}, \mathbf{X}, \mathbf{Y}, c_{\phi}, d_{\phi}, c_{\gamma}, d_{\gamma}, a_1, a_2, a_3)$, that approximates analytically intractable true posterior distribution, $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\phi}, \gamma|\mathbf{t}, \mathbf{X}, \mathbf{Y}, c_{\phi}, d_{\phi}, c_{\gamma}, d_{\gamma}, a_1, a_2, a_3)$. This approximation achieves analytic tractability by assuming that $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\phi}$, and γ are independent under the variational posterior distribution. The analytic tractability leads to improved computational efficiency of the variational Bayes approach over sampling-based MCMC approaches. The derivations for variational posterior distributions are provided in Appendix A.

3 Data analysis

We apply the H-RVM approach to the GBM data as introduced in Section 1. GBM was one of the first cancers evaluated by the TCGA. GBM data have multiple molecular measurements on over 500 samples that include gene expression, copy number, methylation and microRNA expression. TCGA datasets are available at <http://tcga-data.nci.nih.gov/tcga/>. The dataset we analyze here includes information about the gene expressions (11972 probes), miRNA expressions (534 probes), and (uncensored) survival times for matched patient samples (248).

To remove the irrelevant noise variables before model fitting, we prescreened the gene and miRNA probes as follows. We performed univariate regression of the log survival times on the gene expression values, obtained p-values, and retained gene and miRNA probes that cross a liberal p-value threshold (≤ 0.05 here) – to balance the practical and statistical significance. This pre-selection identifies 1747 and 43 gene expression and miRNA probes, respectively, for downstream modeling. All our analyses and comparisons were based on these selected gene and miRNA probes.

We compare the predictions of H-RVM and three linear methods: penalized regressions (2) with the Lasso penalty [13] and with the following covariates: *i.* gene expressions (Gene-Lasso), *ii.* miRNA expressions (MiRNA-Lasso), and *iii.* both gene and miRNA expressions, and their first order interactions (Interaction-Lasso). We randomly split the GBM survival data into a training data and a test data with 223 (90%) and 25 (10%) patients, respectively. H-RVM, Gene-Lasso, MiRNA-Lasso, and Interaction Lasso are fit using the gene and miRNA expressions and log survival times in the training data. The variational inference algorithm is used for fitting H-RVM (see Appendix A). The R package `glmnet` is used for the three penalized linear regressions [19,20]. The log survival times of the test data are predicted for the four methods using the model fits on the training data. The mean squared

prediction errors (MSPEs) are respectively calculated for the four models as the average of the squared difference between the observed and predicted values for the test data. This process of randomly splitting the GBM survival data into training+test data and fitting the four models is repeated 50 times. The results are summarized below.

Figure 1 shows the prediction results for H-RVM, Gene-Lasso, MiRNA-Lasso, and Interaction-Lasso using the kernel density estimates (KDEs) of the MSPEs for the 50 random splits. The KDEs of the MSPEs for H-RVM is shifted to lower values than those for the three penalized linear regressions. The KDEs of the MSPEs for Gene-Lasso, MiRNA-Lasso, and Interaction-Lasso are close to each other, which implies that the MSPEs for these models are fairly similar. Two observations arise from these results. First, the results indicate that penalized linear regression with the Lasso penalty does not lead to improved performance after accounting for interactions among covariates, which has been well-established in literature [19]. Second, the prediction results of the penalized linear regressions do not improve after modeling the first order interactions among genes and miRNAs, thus indicating the presence of non-linear genomic effects and second or higher order interactions among them. For this case study, we see that H-RVM performs better than the penalized linear regression methods. This may be because of H-RVM accounts for the nonlinear

effects of genes and miRNAs and models the interactions within genes, within miRNAs, and between genes and miRNAs. Further, because we model the log survival time, the gain for survival time predictions is, in fact, *exponentially higher* for H-RVM compared to those for the other methods.

We compared the performance of the predictions of the log survival times from H-RVM and the observed survival times using Kaplan-Meier estimates of the survival curves. We used the R package *survival* to perform the log rank test and estimate the Kaplan-Meier survival curves [21]. Figure 2 compares the survival probability curves of the log survival times of patients predicted to be in the long and short survival groups by H-RVM. The patients are assigned to the long and short survival groups based on a median cut-off of the predicted log survival times obtained from H-RVM. The p-value of the log rank test that the two survival curves are the same is close to 0, indicating that the survival group predictions of H-RVM closely agree with the observed survival groups of the patients. In addition, Figure 3 compares the actual survival probability curves of the observed and predicted log survival times of patients in the test data with the minimum MSPE. The p-values and the survival probability curves indicate that the log survival time predictions of H-RVM agree closely with the observed log survival times, as well.

One of the additional gains of our modeling framework is the determination of which platform has a more profound influence on predicting the response, as captured by the weight parameter β . Figure 4 shows the estimates of the weights β for predicting the log survival time of the patients for gene expression, miRNA expressions, and their interactions obtained from H-RVM. The medians of the weights (25% and 75% quartiles) for the three data sources are 0.239 (0.113 and 0.360), 0.504 (0.408 and 0.583), and 0.201 (0.108 and 0.404), respectively. Interestingly, H-RVM shows that miRNAs have better predictive power than genes in predicting the log survival times of patients in the GBM data. The nonzero weight for interactions between gene and miRNA expressions further confirms the presence of nonlinear interactions.

To gain biological insights into our results, we performed a functional analysis of our model fitting results. We used Ingenuity Pathway Analysis software to perform functional analysis on selected significant genes used in fitting H-RVM. We used targetHub [22] to find the known and predicted interactions between significant genes and miRNAs. mirTarBase, a curated database of experimentally validated miRNA targets, was our choice as a source of known gene and miRNA interactions [23]. To identify the predicted gene and miRNA interactions, we used targetScan data [24] to filter out miRNA-gene interactions

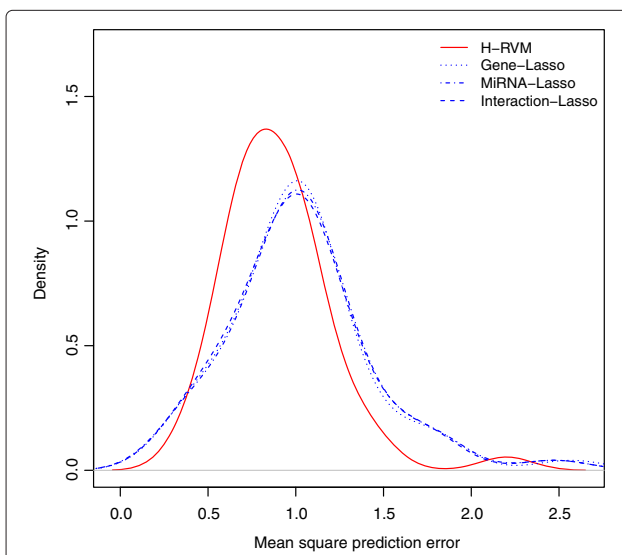
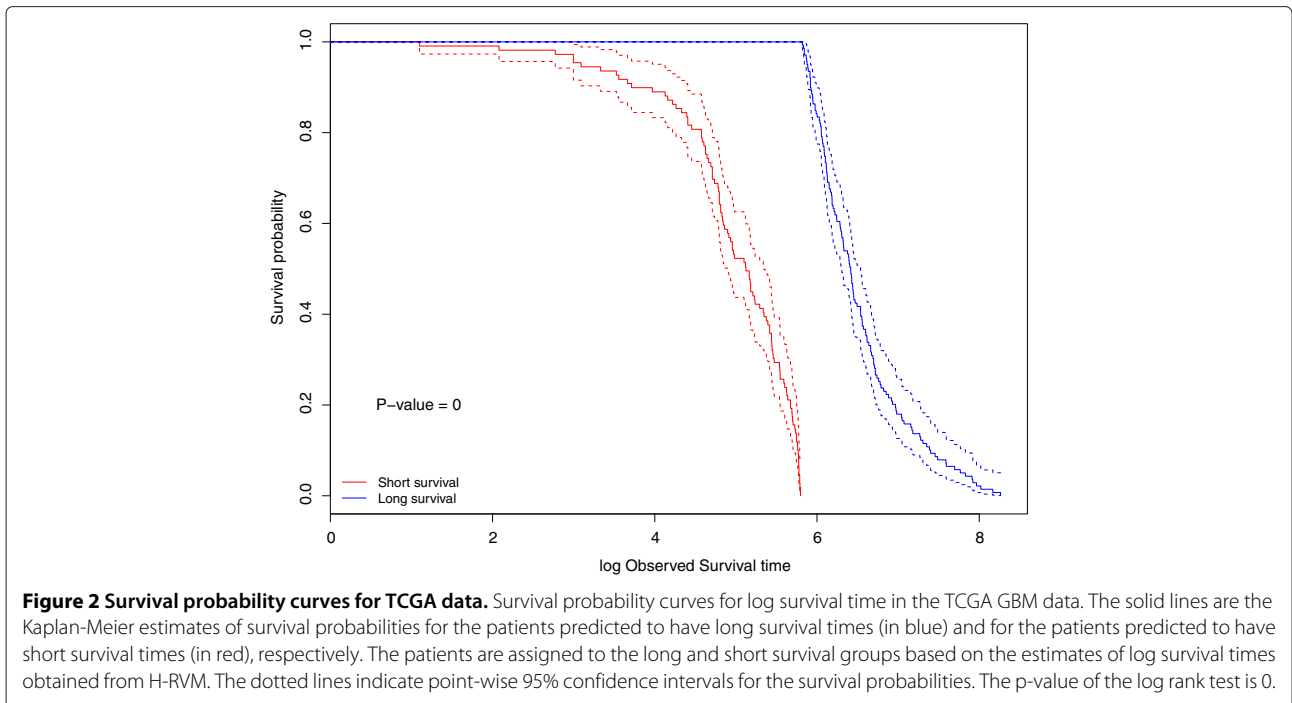


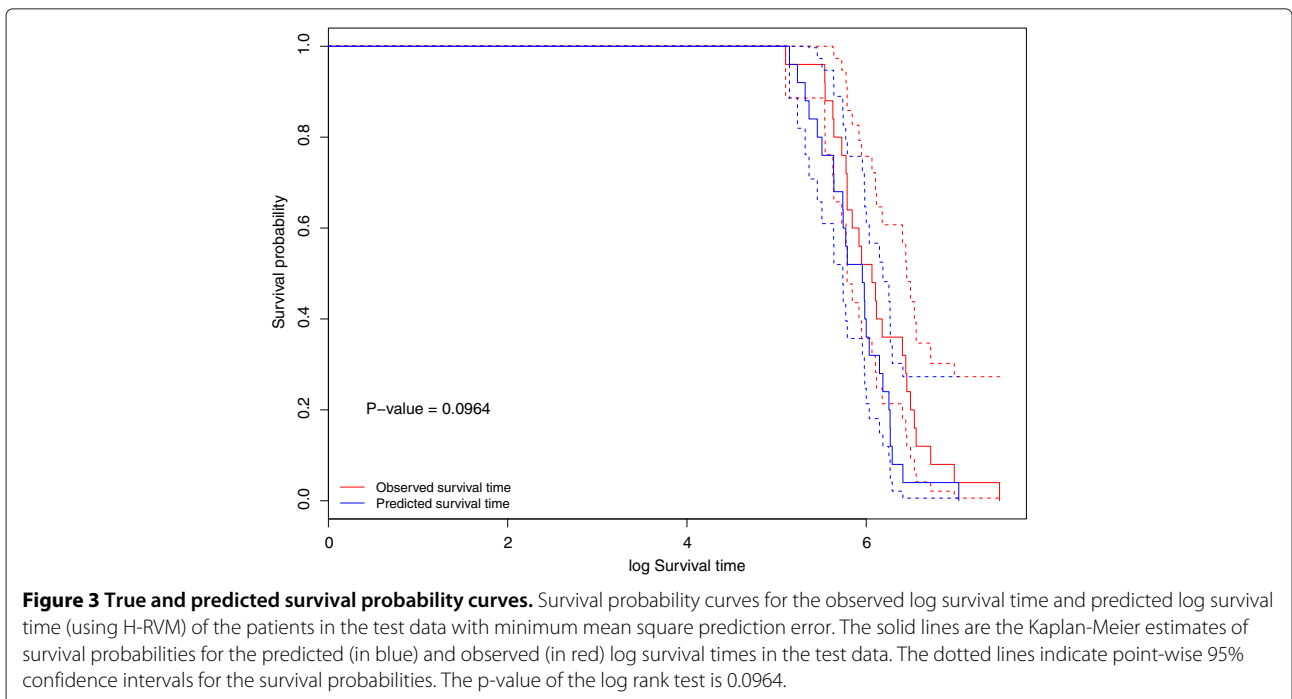
Figure 1 Kernel density estimates. Kernel density estimates (KDEs) of mean square prediction errors (MSPEs) for H-RVM, Gene-Lasso, MiRNA-Lasso, and Interaction-Lasso. The GBM survival data is randomly split 50 times into training and test data, all four models are fit on the training data, and MSPEs for log survival times are calculated for the test data using the model fit on training data. The x-axis represents the MSPEs and the y-axis represents the respective KDEs for H-RVM (in solid red), Gene-Lasso (in dotted blue), MiRNA-Lasso (in dotted and dashed blue), and Interaction-Lasso (in dashed blue).



in which the miRNA and gene effects on survival were concordant, since discordant behavior is expected in biological systems for a direct interaction between miRNA and its targets.

Pathway analyses indicates that the anti-survival genes (i.e., genes with negative effects on survival times) are enriched with signaling pathways related to tumor invasion (see Figure 5). HMGB1 and TWEAK signaling

pathways, which are known to drive tumor metastasis and severe inflammatory responses in GBM and other cancers, are associated with these genes [25-28]. Pro-survival genes are represented by PDGF, PTEN and other signaling pathways. It is well-established that the PDGF signaling pathway dominates the proneural subgroup, which correlates with a good survival time for patients with GBM [29]. The functional terms cellular movement



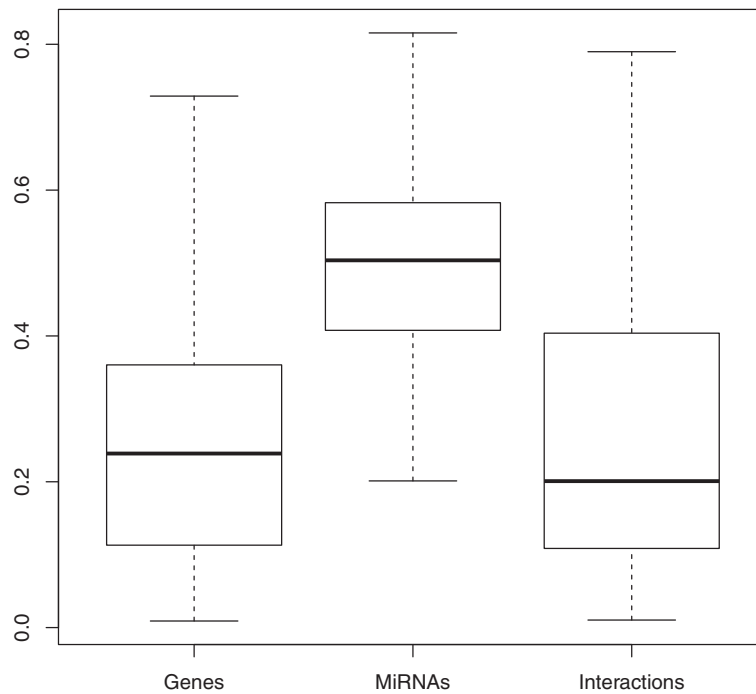
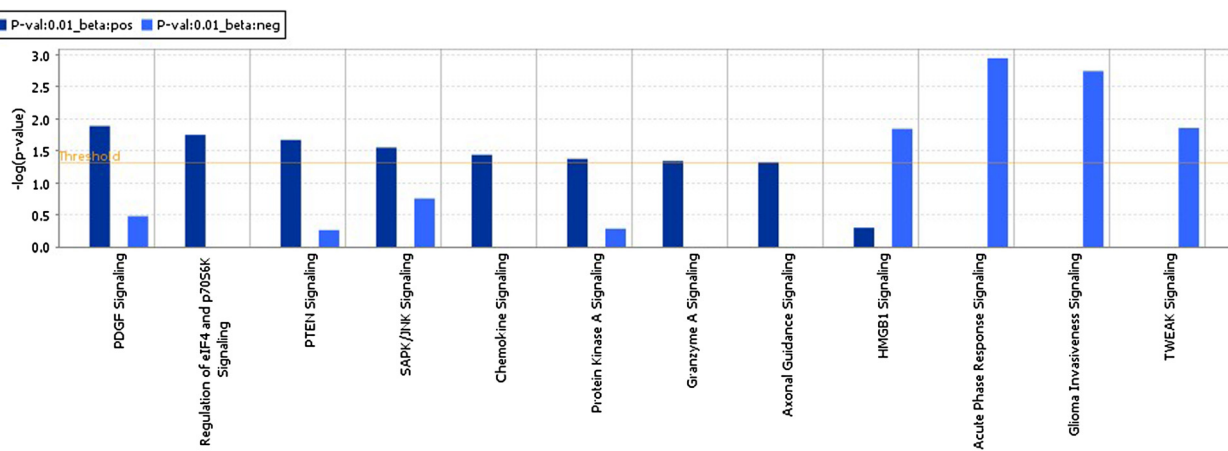


Figure 4 Boxplots of weights. Boxplots of weights $\beta = (\beta_1, \beta_2, \beta_3)$ of gene expressions, miRNA expressions, and their interactions in predicting log survival time. The GBM survival data is randomly split 50 times into training and test data, H-RVM is fit on the training data, and β is obtained from the fit on training data. The y-axis shows the distributions of respective weights for gene expressions, miRNA expressions, and their interactions in predicting log survival time of patients across 50 random splits of the GBM survival data.

and cell-to-cell signaling and interaction pathways are enriched for anti-survival genes, reinforcing their role in invasive GBM.

The target analysis of miRNA revealed 22 known interactions between 8 miRNAs and 20 genes, as shown in Table 1. Four of these eight miRNAs (hsa-miR-31, hsa-miR-146b, hsa-miR-221 and hsa-miR-222) were previously identified as anti-survival markers of GBM [30].

Mir-21 is an established marker of GBM and is known to target many tumor suppressor genes [31]. Mir-34a expression is higher in other GBM subtypes compared to that in the pro-survival proneural glioma subtype [32]. The anti-survival patterns of all these miRNAs indicate that these gene and miRNA interactions can be targeted for therapy of GBM subgroups with expected poor survival. We also identified 1006 predicted interactions



© 2000-2013 Ingenuity Systems, Inc. All rights reserved.

Figure 5 Comparison of the signaling pathways. Comparison of the signaling pathways associated with significant prognostic genes in Glioblastoma multiforme.

Table 1 List of known gene-microRNA interactions identified as significant in the H-RVM model using target analysis

Gene symbols	microRNA
FOXP3, YY1, KLF13, ETS1	hsa-mir-31
FOXO3, DDIT4	hsa-mir-221
ATAT1	hsa-mir-23a
FOXO3	hsa-mir-222
FGG, CPEB3, FGB, PIK3R1	hsa-mir-29a
PDGFB	hsa-mir-146b
PDCD4, TOPORS, BASP1, MARCKS, TP53BP2	hsa-mir-21
SIRT1, YY1, E2F3, CDC25C	hsa-mir-34a

(by TargetScan) between 31 miRNA and 484 genes that are significant (see Additional file 1).

4 Conclusions and future work

We have presented an integrative framework, H-RVM, that generalizes the multiple kernel learning framework for integrating high-dimensional data from multiple sources, incorporating within and between platform interactions to develop a prediction model for clinical outcomes. We applied H-RVM to a high-dimensional TCGA GBM data to predict patient survival times using two data sources: gene and miRNA expressions, and found that the predictive performance of H-RVM is better than those of linear methods that do not model nonlinear effects and interactions. We hypothesize that H-RVM gains flexibility and power by modeling the non-linear interaction structures between gene and miRNA expressions. H-RVM will be a useful tool for biomedical researchers, as clinical prediction using multi-platform genomic information is an important step towards identifying personalized treatments for many cancers. We have code for fitting H-RVM that is freely available at the corresponding author's website (see Additional file 2).

Although we have presented the application of H-RVM in the context of two platforms, the framework is general and can be extended and adapted to data from multiple platforms with different distributional assumptions. This will essentially entail a generalization of the H-RVM model by assuming additional terms for the different platforms. One key issue that warrants further investigation is an increase in the number of (multiplicative) between-platform interaction terms. We used the computationally efficient variational Bayes methods, which are extremely useful for handling large datasets from projects such as TCGA. In addition, [17] presents more scalable versions of HKL and MKL that can be adapted to our framework. Our future work will concentrate on extending the H-RVM framework using Bayesian spike and slab priors to select variables from the interacting covariates

before embedding the data in the space of kernels, as well as incorporating uncertainty estimations of the scale parameters – thus aiding the joint model building process.

Endnotes

^a We use gene and mRNA interchangeably to mean transcript-level expression.

^b We use bold lowercase and uppercase alphabets to denote column vectors and matrices, respectively.

A Appendix: Variational inference for H-RVM

Following the hierarchic kernel learning algorithm (HKL) of [16], we provide the derivation for the variational inference algorithm that estimates the variational posterior distributions for the parameters of H-RVM. Our interest lies in the posterior distributions of α and β that are obtained using the Bayesian model (8) – (12). Unlike RVM, the posterior distributions of α and β in H-RVM are analytically intractable. There are several techniques that can be used to obtain these posteriors distributions. We employ the variational Bayes methods from HKL [16] and obtain analytically tractable variational posterior distribution, $q(\alpha, \beta, \phi, \gamma | \mathbf{t}, \mathbf{X}, \mathbf{Y}, c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3)$, that approximates analytically intractable true posterior distribution, $p(\alpha, \beta, \phi, \gamma | \mathbf{t}, \mathbf{X}, \mathbf{Y}, c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3)$. The variational approach minimizes the Kullback-Liebler (KL) divergence between $q(\alpha, \beta, \phi, \gamma | \mathbf{t}, \mathbf{X}, \mathbf{Y}, c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3)$ and $p(\alpha, \beta, \phi, \gamma | \mathbf{t}, \mathbf{X}, \mathbf{Y}, c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3)$. This approximation achieves analytic tractability by assuming that α, β, ϕ , and γ are independent under $q(\alpha, \beta, \phi, \gamma | \mathbf{t}, \mathbf{X}, \mathbf{Y}, c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3)$. Therefore,

$$q(\alpha, \beta, \phi, \gamma | \mathbf{t}, \mathbf{X}, \mathbf{Y}, c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3) = q(\alpha)q(\beta)q(\phi)q(\gamma), \quad (13)$$

where we have suppressed the conditioning on the data and hyperparameters for the variational posteriors on the right. Notice that the factorization (13) alone guarantees the analytic tractability of $q(\alpha, \beta, \phi, \gamma | \mathbf{t}, \mathbf{X}, \mathbf{Y}, c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3)$, and we do not assume any distributional form for the q 's. Following [16] and [14], the variational posterior distributions are derived as

$$\log q(\alpha) \propto \mathbb{E}_{\beta, \phi, \gamma} [p(\alpha, \beta, \phi, \gamma, \mathbf{t}, \mathbf{X}, \mathbf{Y}, c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3)], \quad (14)$$

$$\log q(\beta) \propto \mathbb{E}_{\alpha, \phi, \gamma} [p(\alpha, \beta, \phi, \gamma, \mathbf{t}, \mathbf{X}, \mathbf{Y}, c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3)], \quad (15)$$

$$\log q(\phi) \propto \mathbb{E}_{\alpha, \beta, \gamma} [p(\alpha, \beta, \phi, \gamma, \mathbf{t}, \mathbf{X}, \mathbf{Y}, c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3)], \quad (16)$$

$$\log q(\gamma) \propto \mathbb{E}_{\alpha, \beta, \phi} [p(\alpha, \beta, \phi, \gamma, \mathbf{t}, \mathbf{X}, \mathbf{Y}, c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3)], \quad (17)$$

where all expectations are with respect to the variational posterior distributions. Hereafter, we will denote $\mathbb{E}_\bullet[f]$ as $\langle f \rangle_\bullet$ for notational simplicity.

Following [16] and using (14), the variational posterior distribution of α ,

$$\begin{aligned} q(\alpha) &= \mathcal{N}(\mu_\alpha, \Sigma_\alpha); \\ \Sigma_\alpha &= \left[\langle \gamma \rangle_\gamma \langle \mathbf{K}_\beta \mathbf{K}_\beta^T \rangle_\beta + \text{diag}(\langle \phi \rangle_\phi) \right]^{-1}, \\ \mu_\alpha &= \langle \gamma \rangle_\gamma \Sigma_\alpha \langle \mathbf{K}_\beta \rangle_\beta \mathbf{t}. \end{aligned} \quad (18)$$

Following [16] and using (16), the variational posterior distribution of ϕ ,

$$\begin{aligned} q(\phi) &= \prod_{j=0}^N q(\phi_j), \text{ where} \\ q(\phi_j) &= \text{Gamma} \left(\frac{1}{2} + c_\phi, \frac{1}{2} \langle \alpha_i^2 \rangle_\alpha + d_\phi \right), \text{ for } j = 1, \dots, N. \end{aligned} \quad (19)$$

Following [16] and using (17), the variational posterior distribution of γ ,

$$\begin{aligned} q(\gamma) &= \text{Gamma} \left(\frac{N}{2} + c_\gamma, \frac{1}{2} \langle \|\mathbf{e}\|^2 \rangle_{\alpha, \beta} + d_\gamma \right), \text{ where} \\ \|\mathbf{e}\|^2 &= \|\mathbf{t} - \mathbf{K}_\beta^T \alpha\|^2. \end{aligned} \quad (20)$$

All the expectations above are determined using $\langle \alpha \rangle_\alpha$, $\langle \alpha \alpha^T \rangle_\alpha$, $\langle \phi \rangle_\phi$, and $\langle \gamma \rangle_\gamma$, which are available from the distributional forms of $q(\alpha)$, $q(\phi)$, and $q(\gamma)$ in (18) – (20). Specifically,

$$\begin{aligned} \langle \alpha \rangle_\alpha &= \mu_\alpha, \quad \langle \alpha \alpha^T \rangle_\alpha = \Sigma_\alpha + \mu_\alpha \mu_\alpha^T, \\ \langle \phi_j \rangle_{\phi_j} &= \frac{\frac{1}{2} + c_\phi}{\frac{1}{2} (\langle \Sigma_\alpha \rangle_{jj} + \langle \mu_\alpha \mu_\alpha^T \rangle_{jj}) + d_\phi}, \\ \langle \gamma \rangle_\gamma &= \frac{\frac{N}{2} + c_\gamma}{\frac{1}{2} \langle \|\mathbf{e}\|^2 \rangle_{\alpha, \beta} + d_\gamma}, \\ \langle \|\mathbf{e}\|^2 \rangle_{\alpha, \beta} &= \sum_{n=1}^N t_n^2 - 2 \sum_{n=1}^N t_n \sum_{m=1}^3 \langle \beta_m \rangle_\beta \langle \alpha^T \rangle_\alpha \mathbf{k}_{mn} \\ &\quad + \sum_{m=1}^3 \sum_{l=1}^3 \langle \beta_m \beta_l \rangle_\beta \Omega_{ml}, \\ \Omega_{ml} &= \sum_{n=1}^N \mathbf{k}_{mn}^T \langle \alpha \alpha^T \rangle_\alpha \mathbf{k}_{ln}. \end{aligned}$$

Instead of $q(\beta)$, its non-normalized version $q^*(\beta)$ is available from [16] as

$$\begin{aligned} q^*(\beta) &= \prod_{m=1}^3 \beta_m^{a_m-1} \exp \left(-\frac{\langle \gamma \rangle_\gamma}{2} (\beta^T \Omega \beta - 2\beta^T \mathbf{b}) \right), \text{ where} \\ \Omega_{ml} &= \sum_{n=1}^N \mathbf{k}_{mn}^T \langle \alpha \alpha^T \rangle_\alpha \mathbf{k}_{ln}, \text{ for } m, l = 1, 2, 3 \text{ and} \\ b_m &= \langle \alpha^T \rangle_\alpha \mathbf{K}_m \mathbf{t}, \text{ for } m = 1, 2, 3. \end{aligned} \quad (21)$$

Following [16], calculate $\langle \beta \rangle_\beta$, $\langle \log \beta \rangle_\beta$, and $\langle \beta \beta^T \rangle_\beta$, as follows. Sample S β 's from Dirichlet(a_1, a_2, a_3) and estimate the expectations as $\langle f(\beta) \rangle_\beta \approx \sum_{s=1}^S f(\beta_s) w(\beta_s)$ where $f(\beta) \equiv \beta, \log \beta$, and $\beta \beta^T$, respectively, and

$$w(\beta_s) = \frac{\exp \left(-\frac{\langle \gamma \rangle_\gamma}{2} (\beta_s^T \Omega \beta_s - 2\beta_s^T \mathbf{b}) \right)}{\sum_{i=1}^S \exp \left(-\frac{\langle \gamma \rangle_\gamma}{2} (\beta_i^T \Omega \beta_i - 2\beta_i^T \mathbf{b}) \right)}.$$

The analytic tractability of $q(\alpha, \beta, \phi, \gamma | \mathbf{t}, \mathbf{X}, \mathbf{Y}, c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3)$ in variational inference guarantees that the marginal variational distribution (or likelihood) of the data $q(\mathbf{t}, \mathbf{X}, \mathbf{Y} | c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3)$ is also analytically tractable. Estimate hyperparameters $c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2$, and a_3 as

$$\arg \max_{c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3} \log q(\mathbf{t}, \mathbf{X}, \mathbf{Y} | c_\phi, d_\phi, c_\gamma, d_\gamma, a_1, a_2, a_3),$$

which is the type II maximum likelihood procedure as recommended in [16]. The kernel parameters $\{\sigma_i^2\}_{i=1}^3$ are learned respectively from three RVMs for each of the three sources using cross-validation as recommended by [15].

Additional files

Additional file 1: mRNA-miRNA-predicted-interactions.xlsx. Excel file containing all predicted mRNA and microRNA interactions flagged as significant in our analysis. The file is available at: http://odin.mdacc.tmc.edu/~vbaladan/Veera_Home_Page/Software_files/mRNA-miRNA-predicted-interactions.xlsx.

Additional file 2: hrvm-0.1.1.tar.gz. R package for fitting H-RVM available at: http://odin.mdacc.tmc.edu/~vbaladan/Veera_Home_Page/Software_files/hrvm_0.1.1.tar.gz.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

VB conceived the research. SS and VB worked out the detailed algorithms and derivations. SS implemented the software and conducted the data analysis. WW and VB oversaw the entire research. WW and GM provided the biological insights into the findings. CO provided insights into the predictive analysis. All authors contributed towards writing the manuscript.

Acknowledgements

VB's research is partially supported by NIH grant R01 CA160736; NSF grant IIS-915196 and the Cancer Center Support Grant (CCSG) (P30 CA016672). WW's work is in part funded by 5U24 CA143883-04 and P30 CA016672. CO is

supported by NSF grant IIS-914861. The content is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. National Cancer Institute, the National Institutes of Health, or the National Science Foundation. We also thank LeeAnn Chastain for editorial assistance with the manuscript.

Author details

¹Department of Statistics, Purdue University, 250 N. University Street, West Lafayette, IN 47907, USA. ²Department of Bioinformatics and Computational Biology, Division of Quantitative Sciences, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Unit 1411, Houston, Texas, USA. ³Department of Computer Science, University of Houston, 4800 Calhoun, Houston, Texas, USA. ⁴Department of Biostatistics, Division of Quantitative Sciences, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Unit 1411, Houston, Texas, USA.

Received: 7 March 2013 Accepted: 10 June 2013

Published: 28 June 2013

References

1. L Chin, JN Andersen, PA Futreal, Cancer genomics: from discovery science to personalized medicine. *Nature Med.* **17**(3), 297–303 (2011)
2. D Witten, R Tibshirani, A framework for feature selection in clustering. *J. Am. Stat. Assoc.* **105**(490), 713–726 (2010)
3. B Efron, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction.* (Cambridge University Press, New York, USA, 2010)
4. M Diehn, C Nardini, M Kuo, Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. *Proc. Natl. Acad. Sci.* **105**(13), 5213 (2008)
5. W Wang, V Baladandayuthapani, JS Morris, BM Broom, G Manyam, KA Do, iBAG: integrative Bayesian analysis of high-dimensional multiplatform genomics data. *Bioinformatics.* **29**(2), 149–159 (2013). <http://bioinformatics.oxfordjournals.org/content/29/2/149.abstract>
6. DM Witten, RJ Tibshirani, et al, Extensions of sparse canonical correlation analysis with applications to genomic data. *Stat. Appl. Genet. Mol. Biol.* **8**, 28 (2009)
7. AA Shabalina, H Tjelmeland, C Fan, CM Perou, AB Nobel, Merging two gene-expression studies via cross-platform normalization. *Bioinformatics.* **24**(9), 1154–1160 (2008)
8. S Ma, Y Zhang, J Huang, Y Huang, Q Lan, N Rothman, T Zheng, Integrative analysis of cancer prognosis data with multiple subtypes using regularized gradient descent. *Genet. Epidemiol.* **36**(8), 829–838 (2012). <http://dx.doi.org/10.1002/gepi.21669>, doi:10.1002/gepi.21669
9. RGx Verhaak, RG Hoadley, CM Perou, DN Hayes, Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell.* **17**, 98–110 (2010)
10. MV Iorio, M Ferracin, M Negrini, CM Croce, MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* **65**(16), 7065–7070 (2005)
11. P Fasanaro, S Greco, M Ivan, MC Capogrossi, F Martelli, microRNA: emerging therapeutic targets in acute ischemic diseases. *Pharmacol. Ther.* **125**, 92–104 (2010)
12. W Tang, J Duan, JG Zhang, YP Wang, Subtyping glioblastoma by combining miRNA and mRNA expression data using compressed sensing-based approach. *EURASIP J. Bioinformatics Syst. Biol.* **2013**, 2 (2013)
13. R Tibshirani, Regression shrinkage selection via the lasso. *J. R. Stat. Soc. Ser. B (Methodological).* **58**(1), 267–288 (1996)
14. CM Bishop, *Pattern Recognition and Machine Learning*, vol. 4. (Springer, New York, 2006)
15. M Tipping, Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **1**, 211–244 (2001)
16. M Girolami, S Rogers, in *Proceedings of the 22nd International Conference on Machine Learning (ICML-05)*. Hierarchic Bayesian models for kernel learning (ACM, New York, USA, Bonn, Germany, 2005), pp. 241–248. <http://doi.acm.org/10.1145/1102351.1102382>, doi:10.1145/1102351.1102382
17. M Gönen, in *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, ed. by J Langford, J Pineau. Bayesian efficient multiple kernel learning (Omnipress, Edinburgh, Scotland, 2012), pp. 1–8
18. J Shawe-Taylor, N Cristianini, *Kernel Methods for Pattern Analysis.* (Cambridge university press, New York, USA, 2004)
19. J Friedman, T Hastie, R Tibshirani, Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1 (2010)
20. R Development Core Team, *R: A Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, Vienna, 2013). <http://www.R-project.org/>
21. TM Therneau, *Modeling Survival Data: Extending the Cox Model.* (Springer-Verlag New York, Inc., New York, USA, 2000)
22. targetHub, targetHub (2013). http://app1.bioinformatics.mdanderson.org/tarhub/_design/basic/index.html
23. SD Hsu, FM Lin, WY Wu, C Liang, WC Huang, WL Chan, WT Tsai, GZ Chen, CJ Lee, CM Chiu, et al, miRTarBase: a database curates experimentally validated microRNA–target interactions. *Nucleic Acids Res.* **39**(suppl 1), D163–D169 (2011)
24. RC Friedman, KKH Farh, CB Burge, DP Bartel, Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009)
25. D Tang, R Kang, HJ Zeh, MT Lotze, High-mobility group box 1 and cancer. *Biochimica et Biophysica Acta (BBA)-Gene Regul. Mech.* **1799**, 131–140 (2010)
26. NL Tran, WS McDonough, BA Savitch, TF Sawyer, JA Winkles, ME Berens, The tumor necrosis factor-like weak inducer of apoptosis (TWEAK)-fibroblast growth factor-inducible 14 (Fn14) signaling system regulates glioma cell survival via NF- κ B pathway activation and BCL-XL/BCL-W expression. *J. Biol. Chem.* **280**(5), 3483–3492 (2005)
27. M Huang, S Narita, N Tsuchiya, Z Ma, K Numakura, T Obara, H Tsuruta, M Saito, T Inoue, Y Horikawa, et al, Overexpression of Fn14 promotes androgen-independent prostate cancer progression through MMP-9 and correlates with poor treatment outcome. *Carcinogenesis.* **32**(11), 1589–1596 (2011)
28. L Dai, L Gu, C Ding, L Qiu, W Di, TWEAK promotes ovarian cancer cell metastasis via NF- κ B pathway activation and VEGF expression. *Cancer Lett.* **283**(2), 159–167 (2009). <http://www.sciencedirect.com/science/article/pii/S0304383509002286>
29. JT Huse, E Holland, LM DeAngelis, Glioblastoma: molecular analysis and clinical implications. *Ann. Rev. Med.* **64**, 59–70 (2012). doi:10.1146/annurev-med-100711-143028
30. S Srinivasan, IRP Patric, K Somasundaram, A ten-microRNA expression signature predicts survival in glioblastoma. *PLoS One.* **6**(3), e17438 (2011)
31. T Papagiannakopoulos, A Shapiro, KS Kosik, MicroRNA-21 targets a network of key tumor-suppressive pathways in glioblastoma cells. *Cancer Res.* **68**(19), 8164–8172 (2008)
32. J Silber, A Jacobsen, T Ozawa, G Harinath, A Pedraza, C Sander, EC Holland, JT Huse, miR-34a repression in proneural malignant gliomas upregulates expression of its target PDGFRA and promotes tumorigenesis. *PLoS ONE.* **7**(3), e33844 (2012). <http://dx.doi.org/10.1371>

doi:10.1186/1687-4153-2013-9

Cite this article as: Srivastava et al.: Integrating multi-platform genomic data using hierarchical Bayesian relevance vector machines. *EURASIP Journal on Bioinformatics and Systems Biology* 2013 **2013**:9.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com