

METHODOLOGY ARTICLE

Open Access



# Data integration by multi-tuning parameter elastic net regression

Jie Liu<sup>1\*</sup> , Gangning Liang<sup>2</sup>, Kimberly D Siegmund<sup>1</sup> and Juan Pablo Lewinger<sup>1</sup>

## Abstract

**Background:** To integrate molecular features from multiple high-throughput platforms in prediction, a regression model that penalizes features from all platforms equally is commonly used. However, data from different platforms are likely to differ in effect sizes, the proportion of predictive features, and correlations structures. Subtle but important features may be missed by shrinking all features equally.

**Results:** We propose an Elastic net (EN) model with *separate* tuning parameter penalties for each platform that is fit using standard software. In a comprehensive simulation study, we evaluated the performance of EN logistic regression with multiple tuning penalties. We found that when the number of informative features differs among the platforms, and when there is no notable correlation between the features from different platforms, the multi-tuning parameter EN yields more predictive models. Moreover, the multi-tuning parameter EN is robust, in the sense that there is no loss of predictivity relative to a single tuning parameter EN when features across all platforms have similar effects. We also investigated the performance of multi-tuning parameter EN using real cancer datasets.

**Conclusion:** The proposed multi-tuning parameter EN model, fit using standard penalized regression software, can achieve better prediction in sample classification when integrating multiple genomic platforms, compared to the traditional method where a single penalty parameter is used for all features in different platforms.

**Keywords:** Data integration, Classification, Elastic net

## Background

As multi-platform profiling of tissues enabled by advances in high-throughput ‘omic’ technologies becomes routine, efficient statistical methods to integrate multi-omic data is becoming increasingly important. Multi-omic profiling has been used to successfully investigate prognostic biomarkers and identify aberrant pathways in cancer [1, 2], enhance clustering and subclassification [3, 4], and improve prediction of cancer prognosis and therapeutic response [5–7]. Multi-omic data integration can be challenging for several reasons. First, different data types will typically have different scales of measurement. To meaningfully integrate data sets with diverse scales, proper standardization and/or data transformation is required. A second challenge is the increasingly high-dimensionality of multi-platform ‘omic’ data. This could be addressed by feature pre-screening methods such as sure independence

screening [8], or by dimensional reduction techniques such as principal component analysis [9] or partial least squares [10]. The third challenge, which to the best of our knowledge has not been yet fully addressed, is the potentially different contributions of individual data types in the final prediction models.

Regularized regression with a sparsity inducing penalty (e.g. LASSO [11], Elastic Net [12], SCAD [13]) is a common approach to feature selection for building predictive models based on high-dimensional data, and can be effectively used for a joint analysis of multi-omic profiles measured on the same samples. For example, Taskesen et al. used the standard LASSO for classification of samples measured on methylation and expression platforms [14] by including all the features from both platforms. In regularized regression, a tuning parameter controls the degree of shrinkage applied to the regression coefficients, and penalties that induce sparsity shrink many coefficients to exactly zero, performing in effect model selection. However, typical regularized regression approaches, would apply the same degree of shrinkage to

\* Correspondence: liu485@usc.edu

<sup>1</sup>Department of Preventive Medicine, USC Keck School of Medicine, 2001 N Soto Street, Los Angeles, CA 90089, USA

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



all features regardless of their omic type, which can be suboptimal if the number and effect sizes of predictive features differ between data types. If for example, there were fewer predictive gene expression features than DNA methylation features, but the predictive expression features had larger effects than the predictive methylation features, forcing a common degree of shrinkage could result in all methylation feature coefficients shrunk to zero, and a final model containing only expression features. Independently predictive methylation features with subtler effects would be missed. Approaches that account for these differences may offer improved predictive performance.

A natural question of interest is whether allowing for differential shrinkage across features from different omic types by using a separate tuning parameter for each type, can yield better predictive models. Taskesen et al. addressed this to some extent in a stage-wise analysis, performing separate analysis of omic platforms prior to creating a single classifier from the posterior probabilities of each model fit [14]. In this approach, individual features from the separate platforms were not combined in their final classifier but only used to pick the more predictive data type. Even if the selected features from multiple platforms were combined, such a two-stage approach would likely result in a different set of selected features having different prediction potential. It is this joint analysis using features from multiple platforms that is our goal. Note that allowing differential shrinkage across different platforms is a distinct issue from how to deal with subgroups of features, such as co-regulated genes within a cluster structure, that one may have reason to believe a-priori are either all predictive or not as a group, and which has been addressed by methods such as the group LASSO [15–17] that encourage the selection of either all or none of the features in each subgroup.

In this paper, we investigate the performance of a multi-tuning parameter elastic net regression (MTP EN) with separate tuning parameters for each omic type. Through simulations with a range of scenarios differing in number of predictive features, effect sizes, and correlation structures between omic types, we show that MTP EN can yield models with better prediction performance. We apply the MTP EN to publically available prostate cancer and acute myeloid leukemia data sets from the Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO).

## Methods

### Setup and notation

We assume we have training data consisting of high-dimensional features from multiple omic types (e.g. gene expression, DNA methylation, somatic mutations, etc.), measured on  $n$  samples, and an outcome of interest (e.g. cancer vs. normal tissue). The goal is to use the training

data to build a model that can be used to predict the outcome for new samples based on their corresponding multi-omic profiles. This is a standard supervised learning setting with the particularity that the features are from multiple omic types.

For simplicity, we focus our presentation on a binary outcome and omic features of two types, but the multi-tuning parameter regression framework can be applied to other outcomes (e.g. multinomial, continuous, time to event, etc.) and more than two data types. For sample  $i$ ,  $i = 1, \dots, n$ , we denote the binary outcome by  $y_i \in \{0, 1\}$  and its corresponding omic profile as the partitioned vector  $\mathbf{x}_i = (\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)})$ , where  $\mathbf{x}_i^{(1)}$  and  $\mathbf{x}_i^{(2)}$  represent the features from the first and second omic type, respectively. The number of features of each type is usually much larger than the sample size  $n$ , and is in the hundreds or thousands in typical studies. For example, gene expression is measured for thousands of genes/features, DNA methylation for hundreds of thousands of CpGs, and genotypes for millions of SNPs. We let  $p_1$  and  $p_2$  be the number of features of the first and second omic type respectively, and  $p = p_1 + p_2$  be the total number of features. In typical applications, there is also a low-dimensional vector of personal and/or clinical features (e.g. age, gender, etc.), that may also be predictive and included in the feature set, but for simplicity we omit here any additional non-omic features.

We denote by  $\mathbf{y} = (y_1, \dots, y_n)$  the vector of outcomes and by  $\mathbf{X}^{(1)}$  and  $\mathbf{X}^{(2)}$  the design matrices whose rows are the  $i^{\text{th}}$  sample vector  $\mathbf{x}_i^{(1)}$  and  $\mathbf{x}_i^{(2)}$  defined above. We denote by  $\mathbf{X} = [\mathbf{X}^{(1)} | \mathbf{X}^{(2)}]$  the matrix containing the full set of features obtained by row-wise concatenation of  $\mathbf{X}^{(1)}$  and  $\mathbf{X}^{(2)}$ .

In addition to constructing a model that generalizes well, i.e. accurately predicts  $\mathbf{y}$  based on  $\mathbf{X}$  in new subjects, a parsimonious prediction model based on a small subset of the full feature set is generally preferred. A model with fewer features is more interpretable and more easily translated into a custom assay deployable in a clinical setting.

Regularized regression with a sparsity inducing penalty is an effective way to simultaneously perform feature selection and parameter estimation to build a predictive model with a small subset of features. The simplest and most commonly used sparsity inducing penalty is the  $L_1$  norm, which gives rise to LASSO regression [11]. Regularization with a combination of the  $L_1$  and  $L_2$  norms, i.e. elastic net regression, typically outperforms feature selection with the LASSO in settings with correlated features [18]. Many additional extensions and variations of the LASSO have been proposed to, for example, reduce over-shrinkage of larger coefficients (adaptive LASSO [19]) or to handle features with additional structure (e.g. group LASSO [15], fused LASSO [20], group LASSO with overlap [21], graph LASSO [22]). For definiteness, in this paper we focus on the elastic net, which includes the LASSO as a special

case. However, customized tuning parameters for each omic type can be used with any type of regularized regression.

The elastic net regularization penalty is a weighted mixture of the LASSO ( $L_1$  norm) and ridge (square of  $L_2$  norm) penalties given by:  $N(\beta) = (1-\alpha)\frac{1}{2}\|\beta\|_{L_2}^2 + \alpha \|\beta\|_{L_1}$ ,  $= (1-\alpha)\frac{1}{2}\sum_{j=1}^p \beta_j^2 + \alpha \sum_{j=1}^p |\beta_j|$ , where  $\alpha$ ,  $0 \leq \alpha \leq 1$ , is the weight given to the LASSO penalty and  $1 - \alpha$  the weight given to the ridge penalty. Both the LASSO ( $\alpha = 1$ ) and the ridge ( $\alpha = 0$ ) penalties are particular cases of the elastic net penalty.

Standard elastic-net logistic regression solves the penalized regression problem given by:

$$\min_{\beta \in \mathbb{R}^p} -l(y, X; \beta_0, X\beta) + \lambda N(\beta) \tag{1}$$

where  $l(y, X; \beta_0, \beta) = \sum_{i=1}^n \log(1 + \exp(\beta_0 + x_i^T \beta)) - \sum_{i=1}^n y_i (\beta_0 + x_i^T \beta)$  is the standard logistic log-likelihood function and the regularization parameter  $\lambda \geq 0$  controls the degree of penalization applied to the vector of regression coefficients  $\beta$  (except the intercept  $\beta_0$  which is typically not penalized). The single regularization parameter  $\lambda$  is common to all features and is usually tuned by cross-validation.

In this paper, we propose using separate tuning parameters for each omic type by solving the penalized regression problem given by,

$$\min_{\beta \in \mathbb{R}^p} -l(y, X, \beta_0; \beta) + \lambda_1 N(\beta^{(1)}) + \lambda_2 N(\beta^{(2)}) \tag{2}$$

where the vector of regression coefficients  $\beta = (\beta^{(1)}, \beta^{(2)})$  is partitioned according to the omic type conformably to  $X = [X^{(1)} | X^{(2)}]$ . The regularization parameters  $\lambda_1 \geq 0$  and  $\lambda_2 \geq 0$  are now specific to each omic type. Our hypothesis is that a ‘custom’ degree of regularization for each type can account for intrinsic differences between the data types and lead to a selected model with better prediction performance.

**Model fitting and parameter tuning**

Although the two-tuning parameter model (2) is non-standard, it can be fitted using elastic-net regression software provided it has an option to use a weighted elastic-net penalty of the form

$N_w(\beta) = \alpha \sum_{j=1}^p w_j |\beta_j| + (1-\alpha) \sum_{j=1}^p w_j \beta_j^2$ . Using a weighted penalty, the multi-tuning parameter elastic-net penalty in (2) can be equivalently written as

$$\lambda_1 N(\beta^{(1)}) + \lambda_2 N(\beta^{(2)}) = \lambda_1 N_w(\beta)$$

with a  $p$ -dimensional weight vector  $w = (1, 1, \dots, 1, \frac{\lambda_2}{\lambda_1}, \frac{\lambda_2}{\lambda_1})$  with 1 in its first  $p_1$  entries (corresponding to the  $p_1$

features of the first omic type and the dimension of  $\beta^{(1)}$ ) and a  $\kappa = \frac{\lambda_2}{\lambda_1}$  in the next  $p_2$  entries (corresponding to the  $p_2$  features of the second omic type and the dimension of  $\beta^{(2)}$ ). Thus, the two-penalty model can be alternatively specified using the tuning parameters  $\lambda = \lambda_1$ ,  $\kappa = \frac{\lambda_2}{\lambda_1}$ , where  $\lambda$  controls the overall shrinkage on both types of omic features, and  $\kappa$  controls the shrinkage ratio between the two types. For  $\kappa = 1$ , the model reduces to the standard elastic net with a single tuning parameter.

To investigate the performance of the multi-tuning parameter elastic net regression (MTP EN) for building predictive models based on multi-omic features we conduct a series of simulations under a range of scenarios. To fit the MTP EN we use the efficient and widely used elastic-net implementation in the glmnet R package [23], which allows for a user-specified weighted penalty via the ‘*penalty.factor*’ argument. Additional file 1 contains an R script implementing a complete and self-contained analysis example using MTP EN.

We set the elastic net penalty  $\alpha$  to  $\frac{1}{2}$  and tune the overall shrinkage,  $\lambda$ , and shrinkage ratio parameter,  $\kappa$ , by  $k$ -fold cross-validation (CV), with the area under the ROC curve, AUC, as the performance metric. The training data is randomly split into  $k$  equally-sized folds, each with (approximately) the same proportion of positive ( $y = 1$ ) and negative ( $y = 0$ ) samples as the full training data. The MTP EN is trained on  $k - 1$  folds for all values of  $(\lambda, \kappa)$  in a grid of possible tuning parameter values and the AUC is computed based on the validation held-off fold. This is repeated, using in turn each of the folds as validation held-off fold. The ‘optimal’ tuning parameters are the values  $\kappa = \kappa_{max}$  and  $\lambda = \lambda_{max}$  that maximize the average AUC across all folds. The optimal tuning parameter values are then applied to a fully independent test set to unbiasedly assess the model AUC. For model tuning we utilized both 5-fold and 10-fold CV, 5-fold for the analysis of real data because of a small sample size in one class and 10-fold for the simulation study.

**Simulation study**

We evaluate the MTP elastic-net through simulation, exploring varying proportions and effect sizes of the relevant features in each omic type, varying dimensionalities of the feature sets, and a range of correlation structures. Specifically, the binary outcome was generated based on a logistic regression model of the form:

$$y_i \mid x_i^{(1)}, x_i^{(2)} \sim \text{Bernoulli}(P_i)$$

$$P_i = \frac{\exp(\eta)}{1 + \exp(\eta)}$$

$$\eta_i = \beta_0 + x_i^{(1)T} \beta_{(1)} + x_i^{(2)T} \beta_{(2)}$$

where we set  $q_j$  features among the  $p_j$  in omic type  $j = 1, 2$  ( $q_j \ll p_j$ ) to be predictive by arranging the vector of regression coefficients  $\beta_{(j)}$  to be sparse, with  $q_j$  non-zero and  $p_j - q_j$  zero entries. We set the effect size of the predictive features, i.e. the non-zero entries in omic type  $j = 1, 2$  to a common value  $\beta_j$ .

We generated feature data  $X = [X^{(1)} | X^{(2)}]$  by sampling from a multivariate normal distribution  $X \sim \mathcal{N}(\mathbf{0}, \Sigma)$ , where  $\Sigma$  is a population covariance matrix with the following structure: i)  $dia(\Sigma) = \mathbf{1}$ , i.e. the  $X$ 's are already standardized and have marginal variances equal to one; ii)  $r_j$  features among the  $q_j$  informative ones in omic type  $j = 1, 2$  have a common pairwise correlation  $\rho_j$  and correlation  $\rho_{12}$  with the counterpart set of features in the other platform iii) all remaining correlations are zero. The simulation scenarios are summarized in Table 1. For each scenario, we simulated 400 replicate data sets, 200 training and 200 test sets. Every data set included 500 features on 200 samples, with an expected 100 samples in each class ( $\beta_0 = 0$ ).

Model tuning parameters  $(\lambda, \kappa)$  were selected using the training data, and applied to the test data set for estimating prediction performance. To further reduce the dimensionality of the selected features beyond what is achieved by maximizing the training data prediction performance, we set  $\lambda$  to  $\lambda_{1se}$ , the largest value of  $\lambda$  such that the AUC is within one standard error of the maximum (achieved at  $\lambda = \lambda_{max}$ ). This strategy by Friedman et al. [12] yields predictive performance similar to that achieved by setting  $\lambda = \lambda_{max}$  but with a more parsimonious model. Thus, the parameters  $\kappa = \kappa_{max}$  and  $\lambda = \lambda_{1se}$  are used to estimate AUC from the independent test data set.

**Table 1** Summary of simulation scenarios

Independent Features: $\rho_1 = \rho_2 = \rho_{12} = 0$					
Scenario #	$\beta_1$	$\beta_2$	$q_1$	$q_2$	Optimal penalty ratio ( $\hat{\kappa}$ )
1	0.6	0.8	5	20	0.55
2	0.6	0.6	5	20	0.70
3	0.8	0.6	5	20	0.75
4	0.8	0.8	5	5	1
5	0.8	0.6	5	5	1
Correlated Features: $\beta_1 = 0.8, \beta_2 = 0.6, q_1 = 5, q_2 = 20, r_1 = r_2 = 3$					
	$\rho_1$	$\rho_2$	$\rho_{12}$	Optimal penalty ratio ( $\kappa^*$ )	
6	0.4	0.2	0	0.85	
7	0.4	0.2	0.4	0.9	

Two hundred samples per data set, 250 features per omic type, 2 omic types. The performance of MTP EN is evaluated by varying effect sizes, number of informative features, and correlation structures between omic types. Specifically,  $\rho_1$  is the correlation between informative features in platform 1,  $\rho_2$  is the correlation between informative features in platform 2,  $\rho_{12}$  is the correlation between informative features from the different platforms,  $\beta_1$  and  $\beta_2$  are the effect sizes of informative features in platforms 1 and 2, respectively, while  $q_1$  and  $q_2$  are the numbers of informative features

In addition to the AUC, we also report accuracy (1-misclassification error) and sensitivity and specificity of feature selection. The accuracy was calculated as the number of samples correctly classified in testing dataset divided by the total number of samples. The sensitivity (specificity) was calculated as the number of informative (uninformative) features correctly selected in the final model divided by the total number of informative (uninformative) features.

**Real data applications**

*Acute myeloid leukemia data*

The Acute myeloid leukemia data for 344 samples were obtained from NCBI Gene Expression Omnibus, with accession numbers for gene expression and methylation data as GSE14468 (HOVON-SAKK cohort) and GSE18700, respectively. The data was also used and described by Taskensen et al. [14]. Briefly, for each patient sample, Affymetrix HGU133 plus2.0 (Santa Clara, CA, USA) and HELP-assay [24] was used to measure the gene expression and DNA methylation data, respectively. We filtered the gene features with fewer than 10 unique expression values, resulting in 46,083 gene features. All of the 25,626 DNA methylation features were considered in this study. Groups of AML patients that are characterized by common cytogenetic or molecular abnormality are denoted as subtypes. From the 15 subtypes studied by Taskesen et al. [14], we focus on the 7q AML subtype, characterized by partial or complete deletion of the genome fragments on the long arm of chromosome 7. In the 344 patients, 35 have 7q AML, and the rest can be characterized as non-7q. The multi-tuning parameter Elastic net was used to build classifiers by combining gene expression and DNA methylation data to differentiate 7q AML from the rest of the subtypes.

*Prostate cancer*

The prostate adenocarcinoma data was obtained from the Genomic Data Commons (GDC) Data Portal (Project ID TCGA-PRAD) and assembled by our collaborators at USC. In total 444 samples with both RNA-Seq gene expression and HumanMethylation450 (HM450) array data available were considered. The goal is to classify tumor aggressiveness, defined by both grade and stage of prostate cancer. We defined the tumor samples with Gleason score 7 and below, and T category of T1 or T2 (T2a, T2b and T2c) as non-aggressive, and those with Gleason score 7 or higher and T category of T3 (T3a, T3b) and T4 as aggressive. This resulted in 143 samples defined as non-aggressive and 265 defined as aggressive. Thirty-six samples (8%) were omitted for being high grade/low stage (25) or low grade/high stage (11) and unknown aggressiveness. After filtering gene features with zero variance across all samples 20,216 gene features remained. For the

HM450 DNA methylation array, we excluded probes targeting SNPs, mapped to the X and Y chromosomes, in cross-reactive regions, and having missing values in at least one sample. This resulted in 371,513 remaining features. As customary, the gene expression data were log2 transformed to reduce the skewness of the gene expression distribution in its original scale.

**Data analysis**

We split the data into training and test sets to build a prediction model and assess its performance. Training was performed using a random 80% of the samples, stratified by outcome to preserve the original proportion of positive and negative cases in the full data. The remaining 20% of samples were used to compute the test AUC. In order to obtain stable estimates, the data splitting, model training and testing was repeated 50 times with the average performance from the independent test sets reported. In the training step, MTP EN penalties were tuned by 5-fold cross-validation over a grid of values ranging from 0.1 to 1.5 for the penalty ratio parameter  $\kappa$ , and a large grid of values for the overall penalty parameter  $\lambda$ . The 5-fold CV was repeated 10 times and the optimal  $\lambda$  for each ratio value  $\kappa$  was selected as the one yielding the largest mean AUC across the 10 CV repeats.

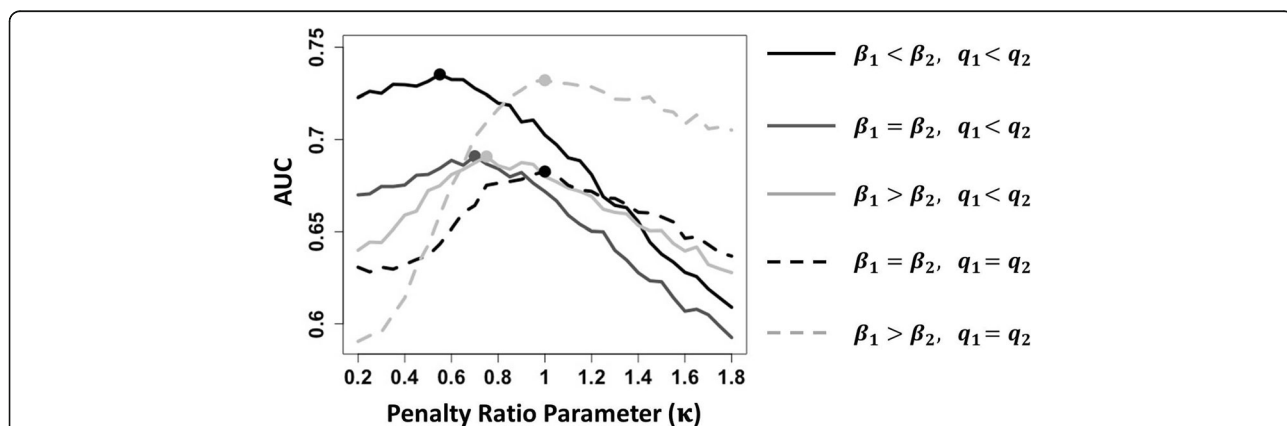
**Results**

**Simulation studies**

The main finding (see Figure 1) is that when the number of informative features differs between the two omic types ( $q_1 < q_2$ ), differential penalization can yield a model with better prediction performance in terms of AUC. Specifically, the optimal parameter tuning favors a larger penalty on the omic type with fewer informative

features ( $\kappa < 1$ ) (Table 1, scenarios 1–3). Differential penalization increased AUC the most when the effect sizes are smaller in the omic type with fewer informative features (Figure 1). By contrast, when the two omic types have the same number of informative features ( $q_1 = q_2$ ), the optimal penalty ratio parameter  $\kappa$  is close to 1, indicating that the standard EN yields the best predictive performance. For Scenarios 1–3 we also evaluated the performance of MTP-EN in terms of accuracy and sensitivity and specificity of feature selection (Additional file 2). Consistent with Figure 1, MTP-EN achieves higher rates of correct classification and better sensitivity in Scenarios 1–3. In terms of specificity, MTP-EN outperforms standard EN in Scenario 1. In Scenario 2 and 3, however, standard EN shows slightly higher specificity.

We explored the change of the optimal penalty ratio parameter as we varied characteristics of the true regression model one at a time. For the scenario with different numbers of informative features,  $q_1 < q_2$ , the optimal penalty ratio parameter  $\kappa$  is smallest (i.e. there is maximal differential penalization) when the effect sizes are smaller in the omic type with fewer informative features ( $\beta_1 < \beta_2$ ), and it increases monotonically to 1 (i.e. less differential penalization) as the effect size,  $\beta_1$ , in the first omic type increases (Figure 2a). For fixed effect sizes  $\beta_1$  and  $\beta_2$ , the optimal penalty ratio parameter  $\kappa$  becomes smaller (more differential penalization) as the number of informative features in the second omic type,  $q_2$ , increases (Figure 2b). In this case, it is advantageous to penalize the omic with more noise features more highly. As the overall proportion of informative features of both types,  $\frac{q_1+q_2}{p}$ , decreases relative to the total number of features, the optimal penalty ratio parameter  $\kappa$  approaches 1, i.e. less differential penalization is required to maximize the AUC (Figure 2c).



**Fig. 1** Mean testing AUC as a function of the penalty ratio parameter  $\kappa$  for different simulation settings. The effect sizes and numbers of informative features are given in Table 1, Scenarios 1–5. Dots indicate the  $\kappa$  resulting in the maximum of mean testing AUC. Each analysis includes 200 samples per data set with 250 features per data type ( $N = 200$  simulation replicates). When the number of informative features differs among the platforms (Scenarios 1–3), the multi-tuning parameter EN yields more predictive models comparing with the standard EN where  $\kappa = 1$ . Differential penalization increased AUC the most when the effect sizes are smaller in the omic type with fewer informative features (Scenario 1)

With the decrease of  $\frac{q_1+q_2}{p}$ , the AUC for both the standard and MTP EN decreases and also the difference in AUC between the two approaches decreases (results not shown).

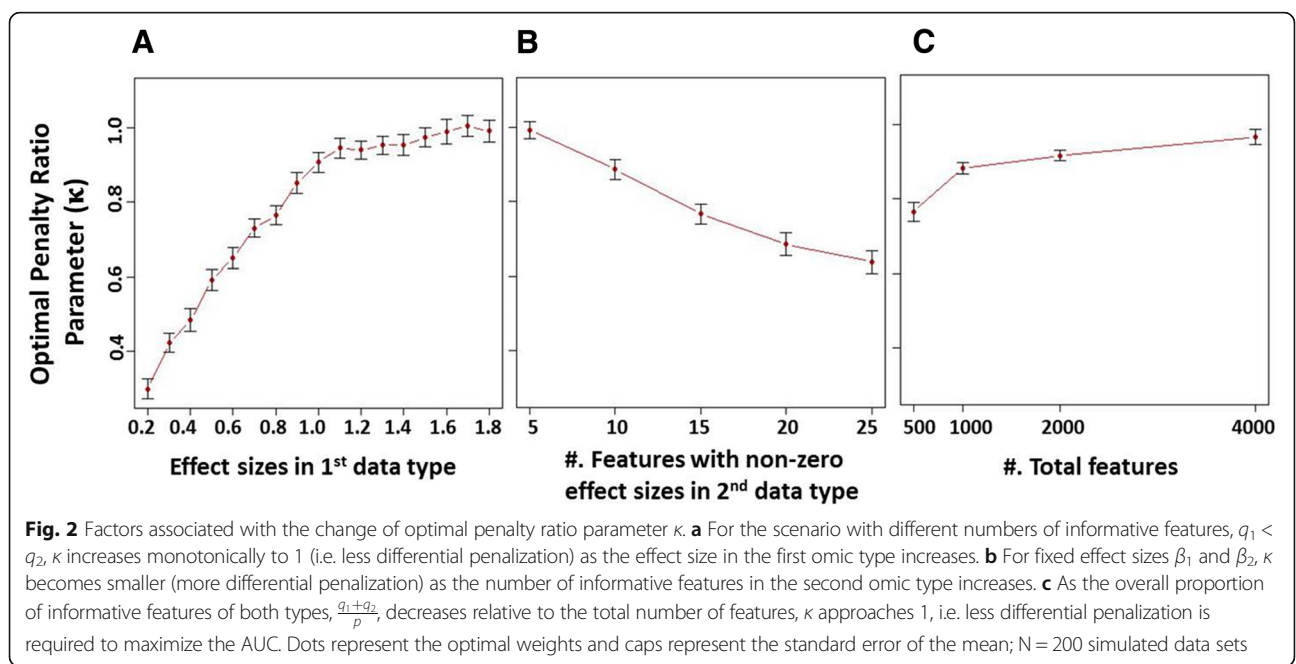
Because correlation among features both within- and between- omic types is common, we also investigated the effects of correlations on the performance of MTP EN. Across all settings we fixed the effect sizes and number of informative features as described in the third scenario in Figure 1 ( $\beta_1 > \beta_2, q_1 < q_2$ ). We compare the situations where the correlations are only between informative features within the individual platforms (Table 1, scenario 6), and correlations both within- and between- platforms (scenario 7). We find that the higher the correlations between informative features, the higher the AUC for a given weight parameter, and the closer the optimal weight is to 1. This can be explained by the fact that the higher correlations increase the chance the set of correlated informative features are selected by standard EN, improving prediction performance without requiring the help of weight parameters. As a matter of fact, the advantage of MTP EN comparing with the standard EN is more obvious when there are fewer correlated features (Additional file 3A), or the correlation coefficients are smaller (Additional file 3B and C). Further, the correlations between the informative features from different platforms have an even larger impact on the performance of MTP EN than the correlations between the features from a single platform (Additional file 3B and C).

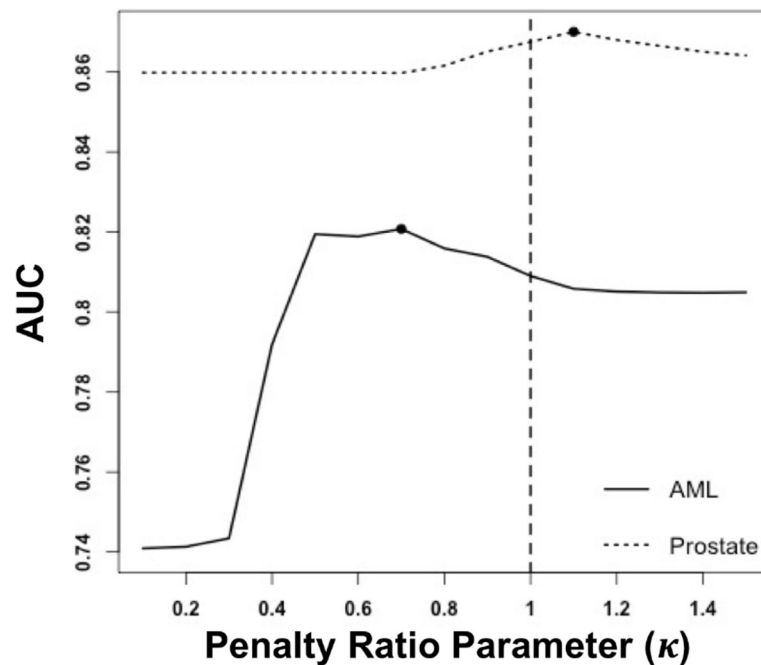
**Real data analysis**

We also applied the MTP EN to two real cancer data sets, combining gene expression and DNA methylation

data for outcome prediction. For the AML data, the goal is to discriminate the 7q subtype from the rest. Classifying patients as 7q is clinical significance as the subtype is characterized susceptibility to infection, quick aggravation, treatment resistance and poor prognosis [25, 26]. We found that the MTP EN achieved the best performance at the weight of 0.7 (solid line in Figure 3), indicating that by penalizing less the methylation features, we can obtain a better classifier than using standard elastic-net. This is consistent with previous finding that the molecular subtypes involving chromosomal abnormalities such as -7/7q- could not be correctly predicted using gene expression profiling alone [27], while DNA methylation signatures have been shown predictive in classifying 7q subtype [28, 29]. Using MTP-EN we have a better chance to keep informative methylation features in the final model, which in turn yields improved prediction performance over using a standard single EN penalty regression.

For the prostate cancer data, the goal is to classify tumor aggressiveness, defined by both grade and stage of prostate cancer. In contrast to the AML data, in prostate cancer data the MTP EN achieved the best performance at the weight of 1.1 (dashed line in Figure 3), indicating MTP-EN yields almost equivalent prediction performance over standard EN. In fact, either DNA methylation or gene expression data alone achieve good prediction, which is consistent with previous findings [30, 31]. Given that the two platforms have similar prediction performance, the observation of little difference between MTP-EN and standard EN is within expectation.





**Fig. 3** The AUC as a function of the penalty ratio parameters  $\kappa$  in cancer data sets. Solid line: AML data set; dashed line: Prostate cancer data set. The dark dots represent the  $\kappa$  that resulted in the maximum of mean test set AUC. For the AML data, a better classifier is obtained by adding relatively less penalty on the methylation features. For the prostate cancer data, there was very little difference in prediction performance between using data from a single platform or combining it

We also investigated the difference in feature selection between MTP EN and standard EN for the two cancer data sets. Differential feature selection was defined by the numbers of times a variable was selected across the 50 repetitions of cross-validation. Since the optimal weight parameter in AML data is less than 1, we expect the MTP EN model to select more methylation features than the standard EN. Indeed, six methylation loci, seldom selected by standard EN, were often selected by MTP EN. Furthermore, the majority of those six features are associated with blood cancer (Additional file 4). Moreover, the genes that were selected by standard EN were also selected by MTP EN. In the PRAD data the opposite occurred. The optimal weight parameter was slightly larger than 1 and MTP EN selected more gene expression features than standard EN (Additional file 4). Four of the five genes, which were not selected by standard EN, have previously been identified as associated with cancer.

## Discussion

We propose a multi-tuning parameter elastic net (MTP EN) for the classification of samples with data from multiple -omic platforms. The MTP EN yields more predictive models in several scenarios, including when the proportion of predictive features is larger in one omic type. In all other scenarios, the predictive

performance of MTP EN matches that of the standard single penalty EN, so there is no performance downside for using MTP EN. Importantly, MTP EN can be fitted using standard EN software like the 'glmnet' R package.

However, the MTP EN requires the tuning of one penalty parameter per omic type, and the computational effort for penalty-parameter-tuning using cross-validation grows exponentially in the number of tuning parameters when using a grid search. This should not prove a limitation for the majority of current studies that collect data on a few omic types only. Alternatively, much less costly parameter tuning using a random search of the tuning parameter space has been shown to be effective [32].

As is the case with penalized regression approaches in general, the benefit of MTP EN decreases with the increase of the number of noise features, so it is important to consider feature pre-selection before applying MTP EN. Excluding noise features and appropriately reducing dimensionality can maximize the performance of MTP EN.

The acute myeloid leukemia (AML) data set we used to illustrate the MTP EN has been previously analyzed by Taskesen et al. [14] using logistic regression with Lasso regularization to predict AML subtypes in 344 samples. They evaluated three different classification strategies,

including early, late and no integration. In early integration, they combined gene expression profile (GEP) and DNA methylation profile (DMP) features first and then applied Lasso regression on all features to predict AML subtypes ('concatenation-based' integration). In late integration, they first used Lasso regression on GEP and DMP individually, and then trained a nearest mean classifier with the posterior probabilities of the GEP and DMP logistic regression as predictors (a two-layer classifier). They showed that early integration improved the predictive power compared to classifiers trained on GEP or DMP alone, and that in turn late integration outperformed early integration. Our results are consistent with Taskesen et al. regarding the performance of combined data versus the individual data; we observed that the maximum AUC was obtained at the weight of 0.7 (combined data) rather using methylation data only (equivalent to using a small weight in the MTP EN; left tail of the solid line in Figure 3) or using gene expression data only (equivalent to using a large weight; right tail of the solid line in Figure 3). More importantly, however, we noticed that the mean AUC for MTP EN is 0.82 for the optimal weight parameter, which is higher than the AUC of 0.80 obtained from the best method in Taskesen et al. These results further support the view that when we consider the inter-correlations between different platforms when setting up the prediction model, we can better utilize the complementary information across different platforms and obtain a model with better prediction performance.

We also applied our method on prostate adenocarcinoma (PRAD) TCGA data, which utilized the HM450 DNA methylation array and contained many more features than the AML data set. The change of the mean testing AUCs with the change of weight parameter was very slight in the PRAD data. Still, we identified several genes that had a much larger chance to be selected in MTP EN (under the weight parameter of 1.1) than in standard EN (weight = 1). Specifically, *ABCC5* has been reported to support osteoclast formation and promote breast cancer metastasis to bone [33], *ITGA11* has been identified to regulate cancer stromal stiffness and promote metastasis in non-small cell lung cancer [34], and *ZNF706* has been associated with tumor progression in head and neck cancer [35].

Although we only investigated the simple convex EN penalty, using multiple penalties is likely to be beneficial for more sophisticated convex penalties such as group- and fused-LASSO penalties but it remains an open question whether this would extend to non-convex penalties such as SCAD, or the minimax concave penalty [36].

## Conclusions

We proposed a multi-tuning parameter elastic net (MTP EN) model for the classification of samples with data from

multiple -omic platforms, with separate tuning parameters for each omic type that can be fitted using existing software. We found that MTP EN yields a more predictive model than ordinary EN where a single penalty parameter is used for all features in different platforms, particularly when the proportion of informative features differs between platforms and when there is no notable correlation between the informative features.

## Additional files

**Additional file 1:** Self-contained R script for MTP-EN with full example. (R 5 kb)

**Additional file 2:** The accuracy in testing dataset, sensitivity and specificity of feature selection from Standard EN and MTP-EN for different simulation settings. MTP-EN achieves better classification and sensitivity in Scenarios 1–3. (PNG 214 kb)

**Additional file 3:** The optimal penalty ratio parameters versus the change of (A) number of correlated features in the second data type; (B) correlations among features in the second data type; and (C) correlations among features between different platforms. Dots represent the mean of optimal weights and caps represent the standard error of the mean;  $N = 200$  simulation replicates. (PNG 100 kb)

**Additional file 4:** Listing of features that have more chance to be selected by MTP EN in AML and PRAD datasets. (DOCX 29 kb)

## Abbreviations

AML: Acute myeloid leukemia; CV: Cross-validation; DMP: DNA methylation profile; EN: Elastic net regression; GDC: Genomic Data Commons; GEO: Gene Expression Omnibus; GEP: Gene expression profile; HM450: HumanMethylation450; MTP EN: Multi-tuning parameter elastic net regression; PRAD: Prostate adenocarcinoma; TCGA: The Cancer Genome Atlas

## Funding

This work was supported by NHGRI grant number R01 HG006705, NCI grant numbers 5P30 CA014089 and 5P01 CA196569 and NIEHS grant number P30 ES07048. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funding body, NIH, did not play any roles in the design of the study, the collection, analysis, and interpretation of data, or in writing the manuscript.

## Availability of data and materials

The AML datasets analyzed during the current study are available in NCBI Gene Expression Omnibus, with accession numbers for gene expression and methylation data as GSE14468 and GSE18700, respectively. The data was also used and described by Taskesen et al. [14]. The prostate adenocarcinoma data was obtained from GDC Data Portal (Project ID TCGA-PRAD).

## Authors' contributions

The study was conceived by JPL. JL conducted the simulation study and performed the data analysis. The manuscript was written by JL, JPL, and KDS. GNL proposed the use of and provided the assembled PRAD data. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interest.



## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>Department of Preventive Medicine, USC Keck School of Medicine, 2001 N Soto Street, Los Angeles, CA 90089, USA. <sup>2</sup>USC Institute of Urology and the Catherine & Joseph Aresty Department of Urology, Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90089, USA.

Received: 1 March 2018 Accepted: 26 September 2018

Published online: 10 October 2018

### References

- Mariani M, He S, McHugh M, Andreoli M, Pandya D, Sieber S, et al. Integrated multidimensional analysis is required for accurate prognostic biomarkers in colorectal cancer. *PLoS One*. 2014;9:e101065.
- Chari R, Coe BP, Vucic EA, Lockwood WW, Lam WL. An integrative multi-dimensional genetic and epigenetic strategy to identify aberrant genes and pathways in cancer. *BMC Syst Biol* [Internet]. 2010;4:67. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2880289&tool=pmcentrez&rendertype=abstract>
- Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. *Genomics* [Internet]. 2011;98:288–95 Available from: <https://doi.org/10.1016/j.ygeno.2011.07.007>.
- Serra A, Fratello M, Fortino V, Raiconi G, Tagliaferri R, Greco D. MVDA: a multi-view genomic data integration methodology. *BMC Bioinformatics* [Internet]. 2015;16:261. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4539887&tool=pmcentrez&rendertype=abstract>
- Zhao Q, Shi X, Xie Y, Huang J, Shia B, Ma S. Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA. *Brief Bioinform* [Internet]. 2014;16:bbu003 Available from: <https://www.bib.oxfordjournals.org/content/early/2014/03/09/bib.bbu003.short?rss=1%5Cnhttps://www.bib.oxfordjournals.org/cgi/doi/10.1093/bib/bbu003%5Cnhttps://www.ncbi.nlm.nih.gov/pubmed/24632304>.
- Costello JC, Heiser LM, Georgii E, Gönen M, Menden MP, Wang NJ, et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol* [Internet]. 2014;32:1–103 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24880487>.
- Kim D, Li R, Dudek SM, Ritchie MD. Predicting censored survival data based on the interactions between meta-dimensional omics data in breast cancer. *J Biomed Inform*. 2015;56:220–8.
- Fan J, Lv J. Sure independence screening for ultrahigh dimensional feature space. *J R Stat Soc Ser B Stat Methodol*. 2008;70:849–911.
- Jolliffe IT. *Principal component analysis*. 2nd ed. New York: Springer; 2002. p. 150–66.
- Abdi H. Partial least squares regression (PLS-regression). In: Lewis-Beck M, Bryman A, Futing T, editors. *Encyclopedia for research methods for the social sciences*. Thousand Oaks (CA): Sage; 2003. p. 792–5.
- Tibshirani R, Society RS. Regression and shrinkage via the lasso. *J R Stat Soc Ser B*. 1996;58:267–88.
- Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. New York:Springer; 2009. p. 61–73.
- Fan J, Li R. Variable selection via nonconcave penalized likelihood and its Oracle properties. *J Am Stat Assoc*. 2001;96:1348–60.
- Taskesen E, Babaei S, Reinders MM, De Ridder J. Integration of gene expression and DNA- methylation profiles improves molecular subtype classification in acute myeloid leukemia. *BMC Bioinformatics*. 2015;16:55.
- Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *J R Stat Soc Ser B Stat Methodol*. 2006;68:49–67.
- Ma S, Song X, Huang J. Supervised group lasso with applications to microarray data analysis. *BMC Bioinformatics*. 2007;8:60.
- Sun H, Wang S. Penalized logistic regression for high-dimensional DNA methylation data with case-control studies. *Bioinformatics*. 2012;28:1368–75.
- Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol*. 2005;67:301–20.
- Zou H. The adaptive lasso and its Oracle properties. *J Am Stat Assoc*. 2006;101:1418–29.
- Tibshirani R, Saunders M, Rosset S, Zhu J, Knight K. Sparsity and smoothness via the fused lasso. *J R Stat Soc Ser B Stat Methodol*. 2005;67:91–108.
- Jacob L, Obozinski G, Vert J-P. Group lasso with overlaps and graph lasso. *ICML*. 2009;433–40.
- Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*. 2008;9:432–41.
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* [Internet]. 2010;33:1–22 Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2929880&tool=pmcentrez&rendertype=abstract>.
- Shaknovich R, Figueroa ME, Melnick A. HELP (HpaII tiny fragment enrichment by ligation-mediated PCR) assay for DNA methylation profiling of primary normal and malignant B lymphocytes. *Methods Mol Biol*. 2010;632:191–201.
- Nevill TJ, Fung HC, Shepherd JD, Horsman DE, Nantel SH, Klingemann HG, et al. Cytogenetic abnormalities in primary myelodysplastic syndrome are highly predictive of outcome after allogeneic bone marrow transplantation. *Blood*. 1998;92:1910–7.
- Kumar CC. Genetic abnormalities and challenges in the treatment of acute myeloid leukemia. *Genes Cancer*. 2011;2:95–107.
- Verhaak RGW, Wouters BJ, Erpelinck CAJ, Abbas S, Beverloo HB, Lugthart S, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica*. 2009;94:131–4.
- Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell*. 2010;17:13–27.
- Christiansen DH, Anderson MK, Pedersen-Bjerggaard J. Methylation of p15INK4B is common, is associated with deletion of genes on chromosome arm 7q and predicts a poor prognosis in therapy-related myelodysplasia and acute myeloid leukemia. *Leukemia*. 2003;17:1813–9.
- Glinisky GV, Gliniskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. *J Clin Invest*. 2004;113:913–23.
- Mundbjerg K, Chopra S, Alemozaffar M, Duymich C, Lakshminarasimhan R, Nichols PW, et al. Identifying aggressive prostate cancer foci using a DNA methylation classifier. *Genome Biol*. 2017;18:3.
- Bergstra JAMESBERGSTRAJ, Yoshua Bengio YOSHUBENGIU. Random search for hyper-parameter optimization. *J Mach Learn Res*. 2012;13:282–305.
- Mourskaia A a, Amir E, Dong Z, Tiedemann K, Cory S, Omeroglu A, et al. ABC5 supports osteoclast formation and promotes breast cancer metastasis to bone. *Breast Cancer Res*. [Internet]. 2012;14:R149 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23174366>.
- Navab R, Strumpf D, To C, Pasko E, Kim KS, Park CJ, et al. Integrin  $\alpha 11\beta 1$  regulates cancer stromal stiffness and promotes tumorigenicity and metastasis in non-small cell lung cancer. *Oncogene* [Internet]. 2015;35:1899–908. Available from: <http://www.nature.com/doi/doi/10.1038/ncr.2015.254>.
- Belbin TJ, Singh B, Smith R V, Socci ND, Wreesmann VB, Sanchez-Carbayo M, et al. Molecular profiling of tumor progression in head and neck cancer. *Arch Otolaryngol Head Neck Surg* [Internet]. 2005;131:10–18. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15655179>.
- Zhang CH. Nearly unbiased variable selection under minimax concave penalty. *Ann Stat*. 2010;38:894–942.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

