HHS Public Access

Identifying gene-gene interactions using penalized tensor regression[†]

Mengyun Wu[a,b], **Jian Huang**[c], and **Shuangge Ma**[b,*]

[a]School of Statistics and Management, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai 200433, China

[b]Department of Biostatistics, School of Public Health, Yale University, 60 College Street, New Haven, CT 06520, U.S.A

[c]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Abstract

Gene-gene (G×G) interactions have been shown to be critical for the fundamental mechanisms and development of complex diseases beyond main genetic effects. The commonly adopted marginal analysis is limited by considering only a small number of G factors at a time. With the "main effects, interactions" hierarchical constraint, many of the existing joint analysis methods suffer from prohibitively high computational cost. In this study, we propose a new method for identifying important G × G interactions under joint modeling. The proposed method adopts tensor regression to accommodate high data dimensionality and the penalization technique for selection. It naturally accommodates the strong hierarchical structure without imposing additional constraints, making optimization much simpler and faster than in the existing studies. It outperforms multiple alternatives in simulation. The analysis of TCGA data on lung cancer and melanoma demonstrates that it can identify markers with important implications and better prediction performance.

Keywords

1. Introduction

For the risk, development, and prognosis of many diseases, gene-gene (G×G) interactions have been shown to play an important role beyond main genetic effects. For example, it has been suggested that the interaction between genes MUC1 and β-Catenin affects the

---

prognosis of colorectal carcinomas [1]. For obesity, researchers have found that the interaction between genes PPARγ2 and ADRβ3 increases risk in children and adolescents [2]. There are many challenges in identifying and characterizing G×G interactions using traditional statistical methods because of the high dimensionality of genetic measurements, computational burden, multiple-testing correction problem, and unstable parameter estimation. Reviews of challenges and machine learning and statistical approaches for identifying G×G interactions are available in the literature [3, 4].

There are two main families of statistical methods for genetic interaction analysis. The first family conducts marginal analysis and analyzes one or a small number of genes and/or G×G factors at a time. Such analysis is often based on statistical tests. Two-stage methods are often used in which multiple tests are performed and then the p-values are adjusted by multiple comparison adjustment, using for example the false discovery rate (FDR) approach [5, 6]. Different marginal analyses may have minor differences in terms of statistical models, hypothesis testing methods, and multiple comparison adjustment techniques, but share the same spirit. Such methods are simple and have low computational cost but may miss factors with weak marginal but strong joint effects. Many biomedical studies have shown that complex diseases are usually attributable to the joint effects of multiple G factors and their interactions. As a result, many of the recent studies have focused on joint analysis which accommodates the joint effects of all genes and their interactions in one single model and more accurately describes the underlying biological processes.

For joint analysis, the number of parameters is usually much larger than the sample size. In addition, among all of the candidate main effects and interactions, we expect only a small subset to be associated with the disease outcome. Therefore, the procedure of identifying important interactions can be regarded as the statistical problem of variable selection and estimation. Multiple families of variable selection approaches have been developed [7]. In the literature, a popular choice relies on the notion of regression with penalization, which has satisfactory performance in data analyses and simulations [8]. Among the existing methods, a Lasso-based penalization model with a hierarchical constraint is perhaps the most representative [9]. It considers a linear regression model for an outcome variable y, predictors $x_1, \cdots, x_p$, and pairwise interactions between these predictors, that is,

$$y = \alpha_0 + \sum_j \alpha_j x_j + \sum_{j \neq k} \theta_{jk} x_j x_k + \varepsilon, \tag{1}$$

where $\varepsilon \sim N(0, \sigma^2)$, and α_j's and θ_{jk}'s represent main effects and interactions, respectively. Two types of constraints have been defined to only allow an interaction in the model if the corresponding main effects are also in the model. Specifically,

$$\text{Strong hierarchy:} \theta_{jk} \neq 0 \Rightarrow \alpha_j \neq 0 \text{ and } \alpha_k \neq 0;$$
$$\text{Weak hierarchy:} \theta_{jk} \neq 0 \Rightarrow \alpha_j \neq 0 \text{ or } \alpha_k \neq 0.$$

It has been suggested that this hierarchy facilitates both estimation and interpretation [9], and that violating the hierarchy may lead to serious problems in modeling [3]. Beyond the aforementioned approach, other relevant developments include the hierarchical group-Lasso regularization for learning pairwise interactions [10], the joint Cox model with a modified adaptive Lasso [11], and others. Optimization with the presence of hierarchical constraints is extremely difficult and incurs prohibitively high computational cost, posing challenges to practical applications of these methods.

In this article, we conduct G×G interaction analysis under joint modeling. The overall marker selection and regularized estimation framework of this study is similar to that of some recently published studies. Our goal is to develop a practically useful G×G interaction analysis method that is applicable to various types of outcomes and has lower computational cost and better numerical performance. The proposed method is based on tensor which is a multidimensional array. The tensor technique represents data by multiple views and has been widely applied in image processing, video recognition, and also genetic data analysis, for instance, the comparison of global mRNA expressions from multiple organisms [12], integrative analysis of DNA microarray data from different studies [13], and integrative analysis of weighted co-expression networks [14]. This study advances from the existing ones along the following directions. First, we consider the joint effects of a large number of G factors, and the analysis is more informative than marginal analysis and may better reflect disease biology. Second, we provide an alternative approach that treats G factors and their interactions in the form of second-order tensors (as opposed to taking vectors as covariates) and substantially reduce data dimensionality, which in turn leads to lower computational cost and more effective estimation. Third, a penalization approach is adopted for selection which naturally respects the strong "main effects, interactions" hierarchy without imposing additional constraints, resulting in much simpler optimization. Fourth, different from the previous applications of tensor, we conduct interaction analysis with a symmetrical second-order tensor. With both methodological and numerical advancements, the proposed method can provide a useful alternative to the existing literature and is warranted.

The rest of the article is organized as follows. In Section 2, we formulate the penalized tensor regression model and present the computational algorithm. The hierarchy property, convergence of the algorithm, and selection of tuning parameters are also discussed. Performance of the proposed method is examined using simulation for both continuous and censored survival outcomes in Section 3. Analysis of two cancer studies with gene expression measurements is conducted in Section 4. Finally, discussions are provided in Section 5. Additional numerical results are provided in the Supplementary Files.

2. Methods

2.1. Modeling

Let y be the disease outcome or phenotype, which can be a continuous marker, categorical disease status, or survival time. $\boldsymbol{x} = [x_1 \ x_2 \cdots x_p]$ denotes the p-dimensional vector of genes (SNPs, or other genetic functional units). Although clinical and environmental factors are not considered here for the simplicity of notations, generalization to the analysis with

additional factors is straightforward. Consider the joint regression model with all main effects and their interactions

$$E(y) = \phi \left(\alpha_0 + \sum_j \alpha_j x_j + \sum_{j,k} \theta_{jk} x_j x_k \right),$$

(2)

where α_0 is the intercept, α_j's and θ_{jk}'s represent main effects and interactions, respectively, the functional form of ϕ is known, and $E(\cdot)$ denotes expectation. Formulation (2) can accommodate many types of data and models, such as the linear model for continuous data, logistic model and probit model for binary data, Poisson model for count data, and Cox model and accelerated failure time (AFT) model for survival data. Assume n iid observations $\{(y_i, x_i), i = 1, \cdots, n\}$. Denote y as the n-dimensional vector composed of y_i's and X as the design matrix composed of x_i's. Under model (2), the unknown parameters $\alpha = \{\alpha_0, \alpha_1, \cdots, \alpha_p\}$ and $\theta = \{\theta_{jk}, j, k = 1, \cdots, p\}$ can be estimated by minimizing $L(\alpha, \theta, y, X)$ which is the lack-of-fit measure such as the negative log-likelihood function. For flexibility, the proposed approach does not make distributional assumptions on y and can accommodate lack-of-fit measures other than likelihood.

2.2. Background on tensor

Since tensor plays an essential role in the proposed method, here we first provide a brief review of notations and some tensor operations which are used throughout the article. Extensive discussions are available in the literature [15, 16].

A tensor is a multidimensional array. For example, a first-order tensor is a vector, and a second-order tensor is a matrix. The elements of a tensor can be rearranged to facilitate easier computation and interpretation. Let $A \in \mathbb{R}^{p_1 \times p_2 \cdots \times p_D}$ be a Dth-order tensor with elements $a_{i_1, i_2, \cdots, i_D}$'s. $Vec(A)$ denotes the column vector mapped from tensor A. The mode-d matricization (unfolding) of tensor A is denoted by $A_{(d)}$ and arranges the tensor into a $p_d \times \prod_{d' \neq d}^{D} p_{d'}$ matrix such that the $(i_1, i_2, \cdots, i_D)$th element of A maps to the (i_d, j)th element of $A_{(d)}$, where $j = 1 + \sum_{d' \neq d, d'=1}^{D} (i_{d'} - 1) \prod_{d'' \neq d', d''=1}^{d'} p_{d''}$.

There are multiple types of products for tensors. The inner product of two same-sized tensors $A, B \in \mathbb{R}^{p_1 \times p_2 \cdots \times p_D}$ is the sum of the products of their entries, that is,

$$\langle A, B \rangle = \sum_{i_1=1}^{p_1} \cdots \sum_{i_D=1}^{p_D} a_{i_1, i_2, \cdots, i_D} b_{i_1, i_2, \cdots, i_D}.$$

Based on the above definitions, we have

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \langle Vec(\boldsymbol{A}), Vec(\boldsymbol{B}) \rangle = \langle \boldsymbol{A}_{(d)}, \boldsymbol{B}_{(d)} \rangle, d = 1, \cdots, D. \quad (3)$$

In addition, for second-order tensors $\boldsymbol{A} = (a_{ij})_{k \times q} \in \mathbb{R}^{k \times q}$, $\boldsymbol{B} = (b_{ij})_{k \times p} \in \mathbb{R}^{k \times p}$, and $\boldsymbol{C} = (c_{ij})_{p \times q} \in \mathbb{R}^{p \times q}$, we have

$$\langle \boldsymbol{A}, \boldsymbol{BC} \rangle = \left\langle \boldsymbol{A}, \left(\sum_{r=1}^{p} b_{ir} c_{rj} \right)_{k \times q} \right\rangle = \sum_{i=1}^{k} \sum_{j=1}^{q} a_{ij} \left(\sum_{r=1}^{p} b_{ir} c_{rj} \right) = \sum_{i=1}^{k} \sum_{r=1}^{p} \left(\sum_{j=1}^{q} a_{ij} c_{rj} \right) b_{ir} = \langle \boldsymbol{AC}', \boldsymbol{B} \rangle.$$

$$(4)$$

For $d = 1, \dots, D$, let $\boldsymbol{a}_d = [a_{d,1} \cdots a_{d,p_d}]$ be a p_d-dimensional vector. Then the outer product of $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_D$ is a Dth-order tensor represented by

$$\boldsymbol{A} = \boldsymbol{a}_1 \circ \cdots \circ \boldsymbol{a}_D,$$

whose elements are $a_{i_1, i_2, \cdots, i_D} = a_{1, i_1} a_{2, i_2} \cdots a_{D, i_D}$. For matrices $\boldsymbol{A} \in \mathbb{R}^{k \times q}$ and $\boldsymbol{B} \in \mathbb{R}^{p \times q}$, their Khatri-Rao product $\boldsymbol{A} \odot \boldsymbol{B}$ is a matrix of size $(kp) \times q$ defined by

$$\boldsymbol{A} \odot \boldsymbol{B} = [\boldsymbol{a}_1 \otimes \boldsymbol{b}_1 \cdots \boldsymbol{a}_q \otimes \boldsymbol{b}_q],$$

where $\boldsymbol{a}_i \otimes \boldsymbol{b}_i = [a_{i1} b_{i1} \; a_{i2} b_{i1} \; \cdots \; a_{ik} b_{i1} \; a_{i1} b_{i2} \; \cdots \; \cdots a_{i1} b_{ip} \; \cdots \; a_{ik} b_{ip}]'$.

A Dth-order tensor $\boldsymbol{A} \in \mathbb{R}^{p_1 \times p_2 \cdots \times p_D}$ is called rank-one tensor if it can be written as the outer product of D vectors, that is, $\boldsymbol{A} = \boldsymbol{a}_1 \circ \cdots \circ \boldsymbol{a}_D$. CANDECOMP/PARAFAC (CP) decomposition is one of the most important decomposition methods for high-order tensors. It decomposes a tensor into a sum of rank-one tensors, that is,

$$\boldsymbol{A} = \sum_{r=1}^{R} \boldsymbol{a}_1^{(r)} \circ \cdots \circ \boldsymbol{a}_D^{(r)},$$

where R is a positive integer referred to as the rank of the CP decomposition, and $\boldsymbol{a}_d^{(r)}$ is a p_d-dimensional vector [15]. There are various CP decomposition algorithms, among which the most popular is the alternating least squares (ALS) algorithm [17, 18]. CP decomposition has been commonly adopted in published studies, such as those on image compression and neuroscience, because of its uniqueness properties and computational efficiency.

2.3. The penalized tensor regression method

With the proposed method, instead of using vectors as covariates, we use a symmetric matrix (second-order tensor) $\boldsymbol{W}_{(p+1)\times(p+1)} = [1\ \boldsymbol{x}] \circ [1\ \boldsymbol{x}]$ to describe the main effects and interactions, where the main effects are in the first column and first row, and the other elements represent interactions. Model (2) can be reorganized as

$$E(y)=\phi\left(\langle \boldsymbol{W}, \boldsymbol{B}\rangle\right)=\phi\left(b_{11}+\sum_{j=2}^{p+1}(b_{1j}+b_{j1})x_{j-1}+\sum_{j=2}^{p+1}\sum_{k=2}^{p+1}(b_{kj}+b_{jk})x_{j-1}x_{k-1}\right),\tag{5}$$

where $\boldsymbol{B} = (b_{jk})_{(p+1)\times(p+1)}$ is a second-order tensor with unknown b_{jk}'s. $\boldsymbol{B}$ can be asymmetric, as we will take $b_{1j} + b_{j1}$'s as main effects and $b_{kj} + b_{jk}$'s as interactions, which automatically leads to symmetry.

The CP decomposition factorizes a tensor as a sum of rank-one tensors and plays a central role in the proposed method [15]. Specifically, we can write

$$\boldsymbol{B}=\sum_{r=1}^{R}\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}=\left(\sum_{r=1}^{R}\beta_{1j}^{(r)}\beta_{2k}^{(r)}\right)_{(p+1)\times(p+1)},\tag{6}$$

where $\boldsymbol{\beta}_d^{(r)}=\left[\beta_{d1}^{(r)}\ \beta_{d2}^{(r)}\ \cdots\ \beta_{d(p+1)}^{(r)}\right]$ and $b_{jk}=\sum_{r=1}^{R}\beta_{1j}^{(r)}\beta_{2k}^{(r)}$. Then model (5) can be rewritten as

$$E(y)=\phi\left(\left\langle \boldsymbol{W}, \sum_{r=1}^{R}\boldsymbol{\beta}_1^{(r)}\circ\boldsymbol{\beta}_2^{(r)}\right\rangle\right),\tag{7}$$

where $\sum_{r=1}^{R}\beta_{11}^{(r)}\beta_{21}^{(r)}$ is the constant, $\sum_{r=1}^{R}\left(\beta_{11}^{(r)}\beta_{2j}^{(r)}+\beta_{1j}^{(r)}\beta_{21}^{(r)}\right)$ is the coefficient for x_{j-1}, j $= 2, \cdots, p + 1$, and $\sum_{r=1}^{R}\left(\beta_{1j}^{(r)}\beta_{2k}^{(r)}+\beta_{1k}^{(r)}\beta_{2j}^{(r)}\right)$ is the coefficient for $x_{j-1}x_{k-1}$, $k, j = 2, \cdots, p + 1$. Let $\boldsymbol{B}_d$ denote the combination of the vectors from the rank-one components, that is, $\boldsymbol{B}_d=\left[\boldsymbol{\beta}_d^{(1)}\ \boldsymbol{\beta}_d^{(2)}\ \cdots\ \boldsymbol{\beta}_d^{(R)}\right]\in\mathbb{R}^{(p+1)\times R}$, $d = 1, 2$.

We adopt penalization for marker selection. Specifically, we propose the Penalized Tensor regression (PTensor) method, which has objective function

$$l(\boldsymbol{B}_1, \boldsymbol{B}_2; \boldsymbol{y}, \boldsymbol{W})=L(\boldsymbol{B}_1, \boldsymbol{B}_2; \boldsymbol{y}, \boldsymbol{W})+\sum_{r=1}^{R}\sum_{d=1}^{2}\sum_{j=2}^{p+1}P\left(\beta_{dj}^{(r)};\lambda\right),\tag{8}$$

where $P\left(\beta_{dj}^{(r)};\lambda\right)=\lambda\left|\beta_{dj}^{(r)}\right|$ is the Lasso penalty with a data-dependent tuning parameter λ, and the first element of each vector $\beta_d^{(r)}$ is not penalized. As in other penalization studies, the proposed estimate is defined as the minimizer of (8). Interactions and main effects corresponding to the nonzero components of the estimate are identified as associated with the outcome.

The proposed method has been motivated by the following considerations. We adopt the penalized estimation and selection framework, which has been well tested in the literature. On the other hand, the proposed method significantly advances from the existing ones along multiple directions. First, by adopting the CP decomposition, the total number of unknown parameters reduces from $(p+1)^2$ to $2R(p+1)$. When B is sparse, R is usually small. With a moderate to large λ, for some r's, almost all of the elements of $\beta_d^{(r)}(d=1,2)$ are shrunken to zero, rendering little information in the rth rank. In a wide spectrum of numerical studies, it is observed that R is usually smaller than ten. Overall, the resulted number of parameters is much smaller than that of model (2), leading to more effective variable selection and parameter estimation. Second, the proposed method accommodates the strong hierarchical structure without additional constraints. Specifically, if the interaction between genes j and k is selected as an important variable, then

$b_{(k+1)(j+1)}+b_{(j+1)(k+1)}=\sum_{r=1}^R\left(\beta_{1(j+1)}^{(r)}\beta_{2(k+1)}^{(r)}+\beta_{1(k+1)}^{(r)}\beta_{2(j+1)}^{(r)}\right)\neq 0$, where at least one product $\beta_{d(k+1)}^{(r)}\beta_{d'(j+1)}^{(r)}(d\neq d')$ is nonzero. If there is only one nonzero product, for example $\beta_{1(j+1)}^{(1)}\beta_{2(k+1)}^{(1)}\neq 0$, since $\beta_{21}^{(1)}\neq 0$ and $\beta_{11}^{(1)}\neq 0$ without penalization,

$b_{1(j+1)}+b_{(j+1)1}=\beta_{1(j+1)}^{(1)}\beta_{21}^{(1)}\neq 0$ and $b_{1(k+1)}+b_{(k+1)1}=\beta_{11}^{(1)}\beta_{2(k+1)}^{(1)}\neq 0$. The corresponding main effects are also identified. For other cases where more than one nonzero product exist, for example, $\beta_{1(j+1)}^{(1)}\beta_{2(k+1)}^{(1)}\neq 0$, and $\beta_{1(k+1)}^{(2)}\beta_{2(j+1)}^{(2)}\neq 0$, then $\beta_{1(j+1)}^{(1)}\beta_{21}^{(1)}+\beta_{11}^{(2)}\beta_{2(j+1)}^{(2)}$ and $\beta_{11}^{(1)}\beta_{2(k+1)}^{(1)}+\beta_{1(k+1)}^{(2)}\beta_{21}^{(2)}$ are nonzero with probability one as all four coefficients have continuous nonzero values. For the situation with more than two nonzero products, the coefficients can be analyzed in the same manner. In all simulations and real data analyses, the strong hierarchical structure is satisfactorily respected. Other types of penalties can also be used in (8), including the bridge penalty [19], minimax concave penalty (MCP) [20], smoothly clipped absolute deviation (SCAD) penalty [21], and others. We adopt the Lasso penalty here due to its computational simplicity and satisfactory numerical performance. In addition, adopting the Lasso penalty facilitates a direct comparison with the existing popular Lasso-related methods such as HierNet and glinternet (more details below). In this article, we mostly focus on methodological and numerical development. The proposed method is based on commonly adopted statistical models, penalized estimation, and tensor operations, all of which are based on sound statistical principles. It is thus reasonable to conjecture that the proposed method has sound statistical properties. Because of the extreme challenges involved, we postpone theoretical investigations to future study.

Remarks—The CP decomposition has been shown to be unique under very weak conditions, without regard to the elementary indeterminacies of scaling and permutation [15]. The scaling indeterminacy refers to that the decomposition (6) can be rewritten as

$$\boldsymbol{B} = \sum_{r=1}^{R} \left(\eta_1^{(r)} \boldsymbol{\beta}_1^{(r)} \right) \circ \left(\eta_2^{(r)} \boldsymbol{\beta}_2^{(r)} \right),$$

where $\eta_1^{(r)} \eta_2^{(r)} = 1$ for $r = 1, \cdots, R$. The permutation indeterminacy refers to that the vectors $\boldsymbol{\beta}_d^{(1)}, \boldsymbol{\beta}_d^{(2)}, \cdots, \boldsymbol{\beta}_d^{(R)}$ of $\boldsymbol{B}_d$ can be reordered arbitrarily. It has been shown that the sufficient condition uniqueness for the CP decomposition is

$$\sum_{d=1}^{D} k_{\boldsymbol{B}_d} \geq 2R + (D-1),$$

where, for a matrix $\boldsymbol{A}$, k_A is the maximum value k such that any k columns are linearly independent [22]. In addition, Liu and Sidiropoulos [23] have established the necessary condition for the uniqueness of the CP decomposition as

$$\min_{d=1,\cdots,D} \prod_{d' \neq d, d=1}^{D} \tilde{R}(\boldsymbol{B}_d),$$

where $\tilde{R}(A)$ is the rank of matrix $\boldsymbol{A}$. In numerical studies, we only focus on the indeterminacies of scaling and permutation. Specifically, we tackle the scaling and permutation indeterminacy problem using a special constrained parameterization similar to that in the literature [24]. It has been suggested that the optimal value of R for CP decomposition may not exist [15]. However, as suggested in [24], the selection of R can be formulated as a model selection problem, for which we adopt the Extended Bayesian Information Criteria (EBIC) [25] in numerical studies. Rigorous investigation on the uniqueness of the CP decomposition and model selection procedure is postponed to future study.

2.4. Computation

With fixed λ and R, optimization in (8) can be solved using a block relaxation algorithm to update each $\boldsymbol{B}_d$ alternately, which proceeds as follows: (i) Randomly initialize $\boldsymbol{B}_d$, $d = 1, 2$; (ii) For $d, d' = 1, 2$, compute $\hat{\boldsymbol{B}}_d = \operatorname{argmin} l(\boldsymbol{B}_d; \boldsymbol{B}_{d'}, d, \boldsymbol{y}, \boldsymbol{W})$ alternately; (iii) Arrange the estimates of $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$; (iv) Iterate Steps (ii)–(iii) until convergence; (v) Compute the final estimate $\hat{\boldsymbol{B}} = \sum_{r=1}^{R} \hat{\boldsymbol{\beta}}_1^{(r)} \circ \hat{\boldsymbol{\beta}}_2^{(r)}$. The difference between two consecutive $l(\boldsymbol{B}_1, \boldsymbol{B}_2; \boldsymbol{y}, \boldsymbol{W})$ smaller than 10^{-4} is used as the convergence criterion. Convergence is achieved in all of our numerical studies within 30 overall iterations. Convergence properties of the block relaxation algorithm have been examined in the literature [24].

In this algorithm, the most challenging step is (ii). It is shown in [15] that the mode-d matricization $\boldsymbol{B}_{(d)}$ has the following relationship with its rank-R decomposition:

$$\boldsymbol{B}_{(d)}=\boldsymbol{B}_d(\boldsymbol{B}_D \odot \cdots \odot \boldsymbol{B}_{d+1} \odot \boldsymbol{B}_{d-1} \odot \cdots \odot \boldsymbol{B}_1)'. \quad (9)$$

Based on this result and equations (3) and (4), the tensor regression model (7) with $D = 2$ can be rewritten as

$$\begin{aligned} E(y)&=\phi\left(\langle\boldsymbol{W}, \boldsymbol{B}\rangle\right)=\phi\left(\langle\boldsymbol{W}_{(1)}, \boldsymbol{B}_{(1)}\rangle\right)=\phi\left(\langle\boldsymbol{W}_{(1)}, \boldsymbol{B}_1\boldsymbol{B}_2'\rangle\right), \\ &=\phi\left(\langle\boldsymbol{W}_{(1)}\boldsymbol{B}_2, \boldsymbol{B}_1\rangle\right) \\ &=\phi\left(\langle Vec(\boldsymbol{W}_{(1)}\boldsymbol{B}_2), Vec(\boldsymbol{B}_1)\rangle\right). \end{aligned} \quad (10)$$

Similarly,

$$E(y)=\phi\left(\langle Vec(\boldsymbol{W}_{(2)}\boldsymbol{B}_1), Vec(\boldsymbol{B}_2)\rangle\right).$$

Thus, for estimating $\boldsymbol{B}_d$, with the other $\boldsymbol{B}_{d'}$ $_d$ fixed, the tensor regression model turns into an ordinary regression with $Vec(\boldsymbol{B}_d)$ as the $(p + 1)R$-dimensional unknown parameter and $Vec(\boldsymbol{W}_{(d)}\boldsymbol{B}_{d'})$ as the new predictor. Penalized optimization with a much smaller number of parameters is more stable and faster. Many methods are available for optimizing the objective function constructed from (10). For the linear model and Cox model which are examined in our numerical study, we use glmnet [26], which is a coordinate-wise algorithm and more efficient than some other methods such as least angle regression (lars) [27]. It takes less than one second to run each optimization constructed from (10) in our numerical examples with $p = 1000$ and R 10.

A rearrangement needs to be conducted on the estimates in step (iii) to tackle the nonidentifiability issue of the CP decomposition after every optimization of $\boldsymbol{B}_2$. First, $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are scaled such that $|\beta_{11}^{(r)}|=|\beta_{21}^{(r)}|$ and $\beta_{11}^{(r)} \geq \beta_{21}^{(r)}$. Second, both $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are rearranged such that $\beta_{11}^{(1)}>\beta_{11}^{(2)}>\cdots>\beta_{11}^{(R)}$. This rearrangement is feasible since the first element of $\beta^{(r)}$ is always nonzero. There are also several alternative constraints on the parameters. However, as the estimate of tensor $\boldsymbol{B}$ is of ultimate interest, the choice of the particular rearrangement is not crucial.

The proposed method has two tuning parameters, namely the rank R and penalty parameter λ. Their optimal values are obtained using the EBIC approach, which is commonly adopted in published studies, and a grid search. Since the block relaxation algorithm may converge to a local minimum, we run the entire algorithm multiple times with different initial values of $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$, and choose the result that gives the lowest value of EBIC.

3. Simulation

In this section, we consider data with continuous outcomes under linear regression models and survival outcomes (subject to right censoring) under Cox models. For data with continuous outcomes and linear regression models, the sum of squared errors is adopted as the lack-of-fit. For survival data and Cox models, the negative partial log-likelihood is adopted. The analysis of categorical data and count data under generalized linear models is very similar to the analysis of continuous data and is not presented. To better gauge performance of the proposed method, we also compare with multiple competing alternatives. For the linear model case, the following four state-of-the-art alternatives are considered. (**a**) The all-pairwise Lasso (APLasso) method directly applies the Lasso penalization to both main effects and all second-order interactions, without considering the hierarchical structure, and is implemented using the R package *glmnet*. (**b**) A Lasso for hierarchical interactions (HierNet) has been developed [9], which can be realized using the R package *hierNet*. Since it is not computationally feasible to enforce the strong hierarchical structure using HierNet in our large-scale simulations ($p = 500$ and 1000), the weak hierarchical structure option is adopted for HierNet. (**c**) Another interaction learning method based on the hierarchical group-Lasso regularization (glinternet) has been proposed [10]. It selects pairwise interactions in linear regression and logistic regression models in a manner that satisfies the strong hierarchy. This method is realized using the R package *glinternet*. (**d**) Researchers have also proposed two screening algorithms [28], iFORT and iFORM, to identify interactions in a greedy forward fashion while maintaining the hierarchical structure. Since iFORM has better finite sample performance, we choose iFORM as a competing alternative. It can select at most n nonzero parameters.

The above four methods, except for APLasso, have not been developed for the analysis of survival data. Thus two other methods are adopted for comparison. Specifically, for the analysis of survival data, the following alternatives are considered. (**a**) The APLasso method is again considered and implemented using the R package *glmnet*. (**b**) The second method is the Cox model with modified adaptive Lasso penalization (HierLasso) [11]. (**c**) A boosting interaction screening model based on random forests and the Cox model [29] is adopted and realized using R package *sprinter*. This method does not account for the hierarchical structure. There may be other G×G interaction learning methods that are suitable for comparison. The above alternative methods are chosen due to their popularity and competitive performance.

The summary of simulation settings is presented in Table A.1 (Supplementary Material A). Two types of G factors are considered, mimicking gene expression and SNP data, respectively. The continuous G variables are simulated from a multivariate normal distribution with marginal means 0 and marginal variances 1. Two correlation structures are considered. The first is the AR (auto-regressive) structure where the jth and kth G variables have correlation coefficient $0.3^{|j-k|}$. The second is the Band (banded) structure where the jth and kth G variables have correlation coefficient 0.33 if $|j - k| = 1$ and 0 otherwise. For the discrete G variables, we further dichotomize the above continuous variables at the 1st and 3rd quartiles and generate 3-level measurements. We consider the following two models. The first is the linear model,

$$y = \alpha_0 + \sum_j \alpha_j x_j + \sum_{j,k} \theta_{jk} x_j x_k + \varepsilon, \tag{11}$$

where e follows a standard normal distribution. The second is the Cox model where the covariate effect has the same form as in (11). The baseline hazard function is constant. Additionally, we generate censoring times from an exponential distribution, where the parameter is adjusted to make the censoring rate around 20%. We acknowledge that the structures of simulated data are simpler compared to real data. The above data generation mechanisms comprehensively cover different models, distributions, and correlation structures, and are similar to those adopted in the literature.

For each model, two different cases of the strong hierarchical structure are considered. The first case (I) has 15 nonzero main effects and ten dense nonzero interactions, that is, for genes G_1, G_2, G_3, if there are two interactions $G_1 \times G_2$ and $G_2 \times G_3$, then interaction $G_1 \times G_3$ is also present. The second case (II) has 10 nonzero main effects and 15 sparse nonzero interactions, which have patterns different from case I. Since glinternet cannot identify self-interactions, we set all nonzero interactions as between different genes. All of the nonzero parameters a_0, a_j, θ_{jk} are generated from Uniform (0.6, 1). We set $n = 150$ and $p = 500$ or 1000.

When evaluating identification performance, we note that although the proposed method has been developed for interaction identification, the identification of main effects is also of interest. Thus, we evaluate the identification of both main effects and interactions, and the measures include: (a) TP30 and TP50 which are the number of true positives under model size 30 and 50 (including both main effects and interactions), respectively. (b) TP.EBIC and FP.EBIC, which are the number of true positives and false positives when the tuning parameters R and λ are selected using EBIC. (c) For each simulated dataset, an independent testing set is generated in the same manner, and prediction performance is assessed using the mean squared error (MSE.EBIC) and rank correlation (RC.EBIC) for the linear models and C-statistic (Cstat.EBIC) for the Cox models, based on the estimates generated under the EBIC. Rank correlation is closely tied to the AUC (area under the receiver operating characteristic curve) measure [30, 31]. The C-statistic is the time-integrated AUC under the time-dependent ROC framework to measure the overall adequacy of prediction with censored survival data. It is computed using the R package *survAUC* [32]. For both rank correlation and C-statistic, larger values indicate better prediction.

There are a total of 32 simulation settings. Under each setting, we simulate 200 replicates. Summary statistics for Scenarios 1 and 2 are graphically presented in Figures 1 and 2, respectively. The detailed numerical results are presented in Tables A.2–A.9 in Supplementary Material A. It is observed that across the whole spectrum of simulation, the proposed method has competitive performance. It is able to identify the majority of true positives, while having a small number of false positives, and has higher prediction accuracy. For example in Table A.2 for continuous data and linear models, with 10 truly important main effects and 15 interactions and $p = 1000$, the proposed method has TP30 equal to 9.7

for interactions, compared to 2.3 (APLasso), 4.1 (HierNet), 6.7 (glinternet), and 2.1 (iFORM) of the alternatives. It also has good performance in terms of false positives. Under this specific setting, the FP.EBIC values are 129.4 (APLasso), 2.4 (HierNet), 2.3 (glinternet), 1.5 (iFORM), and 3.5 (proposed), respectively. Although the number of false positives of the proposed method is slightly larger, it is still very small and can be well "compensated" by the improvement in true positives. Similar competitive performance is observed for the identification of main effects. The proposed method also has satisfactory prediction performance. For example in Table A.2 with 15 main effects, 10 interactions, and $p = 500$, the prediction MSE values are 23.12 (APLasso), 10.69 (HierNet), 6.24 (glinternet), 10.59 (iFORM), and 2.54 (proposed), respectively. Similar patterns are observed for rank correlation. Under this specific setting, the proposed method has RC.EBIC=0.90, compared to 0.67 (APLasso), 0.79 (HierNet), 0.86 (glinternet), and 0.81 (iFORM). For the analysis of survival data under Cox models, for example as in Table A.3, similar competitive performance of the proposed method is observed. When the G variables have discrete distributions, similar patterns are observed. It is noted that APLasso may have especially inferior performance by missing almost all of the important main effects. We have also examined scenarios with a larger number of interactions and main effects and found that the proposed method "scales" well (details omitted).

The "standard" definition of genetic interactions is between two different genes. The statistical definition of interactions also includes self-interactions (that is, the squared terms of genes). To be thorough, we also consider the scenario with self-interactions. Specifically, we consider data with a continuous outcome under the linear regression model, with continuously distributed G factors and the AR correlation structure. There are 15 main effects, 10 interactions between different genes, and 5 self-interactions. The results with p=500 are shown in Table A.10 (Supplementary Material A). The superior identification and prediction performance of the proposed method is again clearly observed.

As described in Section 1, a strong motivation for the proposed method is the high computational cost of the existing methods. Simulation provides strong evidence that the proposed method is computationally much affordable. In Table A.11 (Supplementary Material A), we provide the average computer time for analyzing one simulated dataset with $p = 1000$ using a laptop with standard configurations. APLasso takes the least amount of time as it does not account for the hierarchical structure. The proposed method has satisfactory computational efficiency compared to the alternatives that respect the hierarchical constraint. HierNet takes over 3 hours to compute a single solution path (50 λ values) under the strong hierarchical constraint. The screening method iFORM takes about 10 minutes. The proposed method and glinternet take less than 5 minutes. It is noted that the computational core of glinternet is written in C, while our computer code is written entirely in R. The computer time of the proposed method can be further reduced if also written in C. For survival data, the proposed method also takes a few minutes, compared to 20–30 minutes for HierLasso and sprinter, although sprinter does not respect the hierarchical structure.

4. Data analysis

We analyze TCGA (https://cancergenome.nih.gov/) data on cutaneous melanoma (SKCM) and lung adenocarcinoma (LUAD) and search for important G×G interactions (and main G effects) that are associated with cancer phenotypes/outcomes. The TCGA data have been recently collected, have a high quality, and serve as an ideal testbed for the proposed method. For G variables, we consider mRNA gene expressions, which have been collected using the IlluminaHiseq RNAseq V2 platform. We analyze the processed level 3 data, which have been lowess-normalized, log-transformed, and median-centered, and are downloaded from the TCGA Provisional using R package *cgdsr* [33].

4.1. Cutaneous melanoma (SKCM) data

We focus on metastatic samples of the Whites. The outcome variable of interest is the (log-transformed) Breslow's thickness, which is a continuous variable and has been widely used as a prognostic indicator for melanoma. Data are available on 259 subjects. For mRNA gene expressions, measurements on 18,934 Z-scores, which quantify the relative expressions of tumor samples with respect to normal, are available for analysis. As the number of cancer-related genes is not expected to be large, to improve stability, we conduct a simple prescreening via marginal analysis and select the top 1,000 genes for downstream analysis.

The proposed method identifies 25 main G effects and 76 G×G interactions. The detailed estimation results are provided in Supplementary Material B. Genes identified as having interactions are also presented in Figure A.1 (Supplementary Material A), where two genes are connected if they have an interaction. It is interesting that the identified interactions naturally lead to two gene clusters, a big one with 17 genes and a small one with 5 genes. Literature search suggests that the identified genes may have important implications. For example, gene CDKN2A has been suggested as a major gene associated with the high risk inherited in melanoma prone families and in multiple primary melanoma patients [34]. Gene FMNL2 is dependent on the ERK MAPK signaling pathway which is commonly overly activated in melanoma, and an over expression of FMNL2 in localized cutaneous melanoma has been shown to be associated with an increased risk of recurrence during follow-up [35]. Gene PTMA has been suggested as a possible molecular signature underlying melanoma in vivo growth rate, for which increased levels have been found in primary and metastatic melanoma tissues [36]. Inherited mutations in gene CDK4 have been documented in some families with hereditary melanomas and confer a 60%–90% lifetime risk of cutaneous melanoma [37]. Gene SLC24A5 has been shown to cause a reduction in the quantity of melanin in melanosomes, contributing to lighter skin pigmentation for some people, and be associated with cutaneous melanoma [38]. IFI44 is a cytoplasmic protein in human melanoma cell lines and has been found to be overexpressed in cutaneous melanoma [39]. The absence of gene NLRC4, which is an important regulator of key inflammatory signaling pathways in macrophages, has been shown to play a critical role in suppressing tumor growth in cutaneous melanoma [40]. Published research on G×G interactions in melanoma is limited. To better comprehend the identified interactions, we further consider genes' functional and biological connections by conducting Gene Ontology (GO) enrichment analysis, which is implemented using DAVID 6.8 [41, 42]. Our analysis suggests that genes

identified as having interactions share common GO terms, and the small p-values suggest that they are functionally and biologically connected. For example, genes SLC17A7, SLC8A1, and SLC24A5 share the term GO:0035725, with a p-value of 0.003.

Beyond the proposed method, we also analyze data using the alternatives. The summary comparison results are provided in Table 1. More detailed estimation results using the alternatives are available from the authors. It is observed that different methods identify quite different sets of main effects and interactions. APLasso identifies a large number of interactions but no main effects. Compared to HierNet and glinternet, the proposed method identifies fewer main effects but more interactions. Interactions identified using the alternatives are also graphically presented in Supplementary Material A. The clustering structures also vary significantly across methods. With practical data, it is difficult to evaluate identification accuracy.

To complement identification analysis, we also evaluate prediction performance, which may provide an indirect support to identification. Without a second independent dataset, we use a resampling-based evaluation approach [43]. With 100 resamplings, we compute the mean prediction MSEs, which are 3.366 (APLasso), 2.374 (HierNet), 2.783 (glinternet), 2.981 (iFORM), and 1.924 (proposed), respectively.We observe considerable improvement in prediction with the proposed method.

4.2. Lung adenocarcinoma (LUAD) data

We focus on primary tumor samples of the Whites. The goal is to identify genes and their interactions that are associated with overall survival (described using a Cox model). Data are available for 387 subjects, among whom 144 died during followup. The observed survival times range from 0.13 to 238.11 months, with median equal to 21.39 months. For gene expressions, measurements of 18,893 Z-scores are available. Again we conduct a marginal screening and keep 1,000 gene expressions for downstream analysis.

The proposed method identifies 63 main effects and 32 G×G interactions. The detailed estimation results are provided in Supplementary Material B. The identified interactions are graphically presented in Figure A.2 in Supplementary Material A, where we observe two interaction clusters. For the identified genes, we search the literature for independent evidences of their associations with lung cancer. Among the identified genes, gene TLE1 has been suggested as a putative lung-specific oncogene and shown to be overexpressed in human lung tumors [44]. The DNA hypermethylation of gene EVX1 at precancerous stages has been observed to be strengthened during progression to lung adenocarcinomas and significantly correlated with tumor aggressiveness [45]. Gene ZNF185 has been found to inhibit growth and invasion of lung adenocarcinoma cells through inhibition of the akt/gsk3β pathway and suggested as a potential therapeutic target for the treatment of lung adenocarcinoma [46]. Gene XRCC5 is a DNA repair gene and overexpressed in lung cancers [47]. The loss of gene GPRC5A in lung has been found to influence clinical outcomes of lung adenocarcinoma [48]. Published analysis has also suggested the strong association between gene BRCA2 and lung adenocarcinoma [49]. Gene NOX1 has been demonstrated as a mediator of growth factor-induced proliferation and transformation of cells, and for anchorage-independent growth, a characteristic of tumor cells including lung

adenocarcinoma [50]. The suppression of gene RAD51, which can protect lung cancer cells from cytotoxic effects induced by gefitinib, has been identified as a novel and additive therapeutic modality in lung adenocarcinoma [51]. The encoded protein of gene PSPN signals through the RET receptor tyrosine kinase and a GPI-linked coreceptor, and promotes the survival of neuronal populations, where the rearrangements in RET have been recently described as new driver mutations in lung adenocarcinoma [52]. We also conduct GO enrichment analysis for the LUAD data and again observe that genes identified as having interactions often share common GO terms. For example, in Figure A.2, genes in the first cluster (left) share the term GO:0008022 (protein C-terminus binding) with p-value 0.0004, and those in the second cluster (right) share the term GO:0010628 (positive regulation of gene expression) with p-value 0.041.

Analysis is also conducted using the alternatives. The summary comparison results are shown in Table 1, and detailed estimation results using the alternatives are available from the authors. Again, APLasso identifies a large number of interactions but not main effects. HierLasso and sprinter have quite different identification results from the proposed method. Different clustering structures are also observed in the figures in Supplementary Material A. We conduct resampling-based prediction evaluation and compute the mean C-statistics as 0.556 (APLasso), 0.584 (HierLasso), 0.572 (sprinter), and 0.661 (proposed). Although all C-statistics are moderate, the proposed method has slightly better prediction.

5. Discussion

The analysis of G×G interactions is challenging because of the high data dimensionality and the need to respect the "main effects, interactions" hierarchical structure. Although there are a few recent methods, the development is still relatively limited compared to other high-dimensional genetic data analysis problems. In this study, we have developed a penalized tensor regression method. The proposed method takes advantage of the tensor technique to significantly reduce dimensionality. Unlike with other penalization methods, the "main effects, interactions" strong hierarchy is automatically satisfied. Without additional constraints (which are common with the existing alternatives), the proposed method has a significant computational advantage. The proposed tensor technique and penalized selection are relatively independent of the data and model forms. Thus the proposed method enjoys broad applicability and is applicable to other data types and models. Simulation suggests that the proposed method has superior identification and prediction performance. Its computational advantage is also clearly observed. In data analysis, it identifies interactions and main effects different from the alternatives. Interestingly, for both datasets, the identified interactions naturally form two gene clusters. Literature search suggests that the identified genes have important implications, and the identified interactions are meaningful, as supported by the shared GO terms. The improved prediction performance provides an indirect support to the identification validity. Overall, the proposed method provides a practically useful alternative venue for analyzing G×G interactions.

This study can be extended in multiple directions. For practical applications, it will be of interest to combine the proposed tensor technique and penalized selection with other data types and models. In penalization, the Lasso penalty can be replaced by other penalties. The

proposed method respects the strong hierarchy. As no gene is "special", it may be sensible to include both main effects corresponding to an identified interaction. However, from a methodological perspective, it may be of interest to explore extending the proposed method to respect the weak hierarchy. Establishing statistical properties, which is highly challenging, is postponed to future research. For the proposed and alternative methods, the patterns of the identified genes and interactions vary across datasets. More functional examination of the analysis results will be needed to better comprehend the differences across methods/data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Baldus SE, Mönig SP, Huxel S, Stephanie L, Franz-Georg H, Katja E, Paul MS, Thiele Jürgen, Arnulf HH, Hans PD. MUC1 and nuclear β-catenin are coexpressed at the invasion front of colorectal carcinomas and are both correlated with tumor prognosis. Clinical Cancer Research. 2004; 10(8):2790–2796. [PubMed: 15102686]

2. Ochoa MC, Marti A, Azcona C, Chueca M, Oyarzabal M, Pelach R, Martinez JA. Gene-gene interaction between PPARγ2 and ADRβ3 increases obesity risk in children and adolescents. International Journal of Obesity. 2004; 28:S37–S41. [PubMed: 15543217]

3. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. Nature Reviews Genetics. 2009; 10(6):392–404.

4. Van Steen K. Travelling the world of gene-gene interactions. Briefings in Bioinformatics. 2012; 13(1):1–19. [PubMed: 21441561]

5. Dai H, Bhandary M, Becker M, Leeder JS, Gaedigk R, Motsinger-Reif AA. Global tests of p-values for multifactor dimensionality reduction models in selection of optimal number of target genes. BioData Mining. 2012; 5:3. [PubMed: 22616673]

6. Sung PY, Wang YT, Yu YW, Chung RH. An efficient gene-gene interaction test for genome-wide association studies in trio families. Bioinformatics. 2016; 32(12):1848–1855. [PubMed: 26873927]

7. Wakefield J, De Vocht F, Hung RJ. Bayesian mixture modeling of gene-environment and gene-gene interactions. Genetic Epidemiology. 2010; 34(1):16–25. [PubMed: 19492346]

8. Liu J, Huang J, Zhang Y, Lan Q, Rothman N, Zheng T, Ma S. Identification of gene-environment interactions in cancer studies using penalization. Genomics. 2013; 102(4):189–194. [PubMed: 23994599]

9. Bien J, Taylor J, Tibshirani R. A Lasso for hierarchical interactions. Annals of statistics. 2013; 41(3):1111–1141. [PubMed: 26257447]

10. Lim M, Hastie T. Learning interactions via hierarchical group-Lasso regularization. Journal of Computational and Graphical Statistics. 2015; 24(3):627–654. [PubMed: 26759522]

11. Wang L, Shen J, Thall PF. A modified adaptive Lasso for identifying interactions in the Cox model with the heredity constraint. Statistics and Probability Letters. 2014; 93:126–133. [PubMed: 25071299]

12. Ponnapalli SP, Saunders MA, Van Loan CF, Alter O. A higher-order generalized singular value decomposition for comparison of global mRNA expression from multiple organisms. PloS One. 2010; 6(12):e28072.

13. Omberg L, Golub GH, Alter O. A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies. Proceedings of the National Academy of Sciences. 2007; 104(47):18371–18376.

14. Li W, Liu CC, Zhang T, Li H, Waterman MS, Zhou XJ. Integrative analysis of many weighted co-expression networks using tensor computation. PLoS Computational Biology. 2011; 7(6):e1001106. [PubMed: 21698123]

15. Kolda TG, Bader BW. Tensor decompositions and applications. SIAM review. 2009; 51(3):455–500.

16. McCullagh, P. Tensor methods in statistics. London: Chapman and Hall; 1987.

17. Carroll JD, Chang JJ. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika. 1970; 35:283–319.

18. Nion D, De Lathauwer L. An enhanced line search scheme for complex-valued tensor decompositions application in DS-CDMA. Signal Process. 2008; 88:749–755.

19. Fu WJ. Penalized regressions: the bridge versus the lasso. Journal of Computational and Graphical Statistics. 1998; 7(3):397–416.

20. Zhang CH. Nearly unbiased variable selection under minimax concave penalty. The Annals of Statistics. 2010; 38(2):894–942.

21. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association. 2001; 96(456):1348–1360.

22. Ten Berge JMF, Sidiriopolous ND. On uniqueness in CANDECOMP/PARAFAC. Psychometrika. 2002; 67:399–409.

23. Liu X, Sidiropoulos N. Cramer-Rao lower bounds for low-rank decomposition of multidimensional arrays. IEEE Transactions on Signal Processing. 2001; 49:2074–2086.

24. Zhou H, Li L, Zhu H. Tensor regression with applications in neuroimaging data analysis. Journal of the American Statistical Association. 2013; 108(502):540–552. [PubMed: 24791032]

25. Chen J, Chen Z. Extended Bayesian information criteria for model selection with large model spaces. Biometrika. 2008; 95(3):759–771.

26. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software. 2010; 33(1):1. [PubMed: 20808728]

27. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. The Annals of Statistics. 2004; 32(2):407–499.

28. Hao N, Zhang HH. Interaction screening for ultrahigh-dimensional data. Journal of the American Statistical Association. 2014; 109(507):1285–1301. [PubMed: 25386043]

29. Sariyar M, Hoffmann I, Binder H. Combining techniques for screening and evaluating interaction terms on high-dimensional time-to-event data. BMC Bioinformatics. 2014; 15:58. [PubMed: 24571520]

30. Sherman RP. The limiting distribution of the maximum rank correlation estimator. Econometrica. 1993; 61(1):123–137.

31. Liu C, White M, Newell G. Measuring and comparing the accuracy of species distribution models with presence-absence data. Ecography. 2001; 34(2):232–243.

32. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Statistics in Medicine. 2011; 30(10):1105–1117. [PubMed: 21484848]

33. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Cerami E. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science Signaling. 2013; 6(269):pl1. [PubMed: 23550210]

34. Badenas C, Aguilera P, Puig-Butillë JA, Carrera C, Malvehy J, Puig S. Genetic counseling in melanoma. Dermatologic Therapy. 2012; 25(5):397–402. [PubMed: 23046018]

35. Peippo M, Gardberg M, Lamminen T, Kaipio K, Carpën O, Heuser VD. FHOD1 formin is upregulated in melanomas and modifies proliferation and tumor growth. Experimental Cell Research. 2017; 350(1):267–278. [PubMed: 27919746]

36. Fortis SP, Anastasopoulou EA, Voutsas IF, Baxevanis CN, Perez SA, Mahaira LG. Potential prognostic molecular signatures in a preclinical model of melanoma. Anticancer Research. 2017; 37(1):143–148. [PubMed: 28011484]

37. Tsao H, Atkins MB, Sober AJ. Management of cutaneous melanoma. New England Journal of Medicine. 2004; 351(10):998–1012. [PubMed: 15342808]

38. Zeng Z, Richardson J, Verduzco D, Mitchell DL, Patton EE. Zebrafish have a competent p53-dependent nucleotide excision repair pathway to resolve ultraviolet B-induced DNA damage in the skin. Zebrafish. 2009; 6(4):405–415. [PubMed: 20047468]

39. Hallen LC, Burki Y, Ebeling M, Broger C, Siegrist F, Oroszlan-Szovik K, Foser S. Antiproliferative activity of the human IFN-α-inducible protein IFI44. Journal of Interferon & Cytokine Research. 2007; 27(8):675–680. [PubMed: 17784819]

40. Janowski AM, Colegio OR, Hornick EE, McNiff JM, Martin MD, Badovinac VP, Sutterwala FS. NLRC4 suppresses melanoma tumor progression independently of inflammasome activation. The Journal of Clinical Investigation. 2016; 126(10):3917–3928. [PubMed: 27617861]

41. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols. 2009; 4(1):44–57. [PubMed: 19131956]

42. Huang DW, Sherman BT, Lempicki R. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research. 2009; 37(1):1–13. [PubMed: 19033363]

43. Jiang Y, Shi X, Zhao Q, Krauthammer M, Rothberg BE, Ma S. Integrated analysis of multidimensional omics data on cutaneous melanoma prognosis. Genomics. 2016; 107(6):223–230. [PubMed: 27141884]

44. Yao X, Jennings S, Ireland SK, Pham T, Temple B, Davis M, Biliran H. The anoikis effector Bit1 displays tumor suppressive function in lung cancer cells. PloS one. 2014; 9(7):e101564. [PubMed: 25003198]

45. Sato T, Arai E, Kohno T, Tsuta K, Watanabe SI, Soejima K, Kanai Y. DNA methylation profiles at precancerous stages associated with recurrence of lung adenocarcinoma. PLoS One. 2013; 8(3):e59444. [PubMed: 23544068]

46. Wang J, Huang HH, Liu FB. ZNF185 inhibits growth and invasion of lung adenocarcinoma cells through inhibition of the akt/gsk3β pathway. Journal of Biological Regulators and Homeostatic Agents. 2016; 30(3):683–691. [PubMed: 27655485]

47. Zhong L, Coe SP, Stromberg AJ, Khattar NH, Jett JR, Hirschowitz EA. Profiling tumor-associated antibodies for early detection of non-small cell lung cancer. Journal of Thoracic Oncology. 2006; 1(6):513–519. [PubMed: 17409910]

48. Kadara H, Fujimoto J, Men T, Ye X, Lotan D, Lee JS, Lotan R. A Gprc5a tumor suppressor loss of expression signature is conserved, prevalent, and associated with survival in human lung adenocarcinomas. Neoplasia. 2010; 12(6):499–505. [PubMed: 20563252]

49. Lee MN, Tseng RC, Hsu HS, Chen JY, Tzao C, Ho WL, Wang YC. Epigenetic inactivation of the chromosomal stability control genes BRCA1, BRCA2, and XRCC5 in non-small cell lung cancer. Clinical Cancer Research. 2007; 13(3):832–838. [PubMed: 17289874]

50. Malec V, Gottschald OR, Li S, Rose F, Seeger W, Hänze J. HIF-1α signaling is augmented during intermittent hypoxia by induction of the Nrf2 pathway in NOX1-expressing adenocarcinoma A549 cells. Free Radical Biology and Medicine. 2010; 48(12):1626–1635. [PubMed: 20347035]

51. Ko JC, Hong JH, Wang LH, Cheng CM, Ciou SC, Lin ST, Lin YW. Role of repair protein Rad51 in regulating the response to gefitinib in human non-small cell lung cancer cells. Molecular Cancer Therapeutics. 2008; 7(11):3632–3641. [PubMed: 19001445]

52. Bos M, Gardizi M, Schildhaus HU, Buettner R, Wolf J. Activated RET and ROS: two new driver mutations in lung adenocarcinoma. Translational Lung Cancer Research. 2013; 2(2):112–121. [PubMed: 25806222]
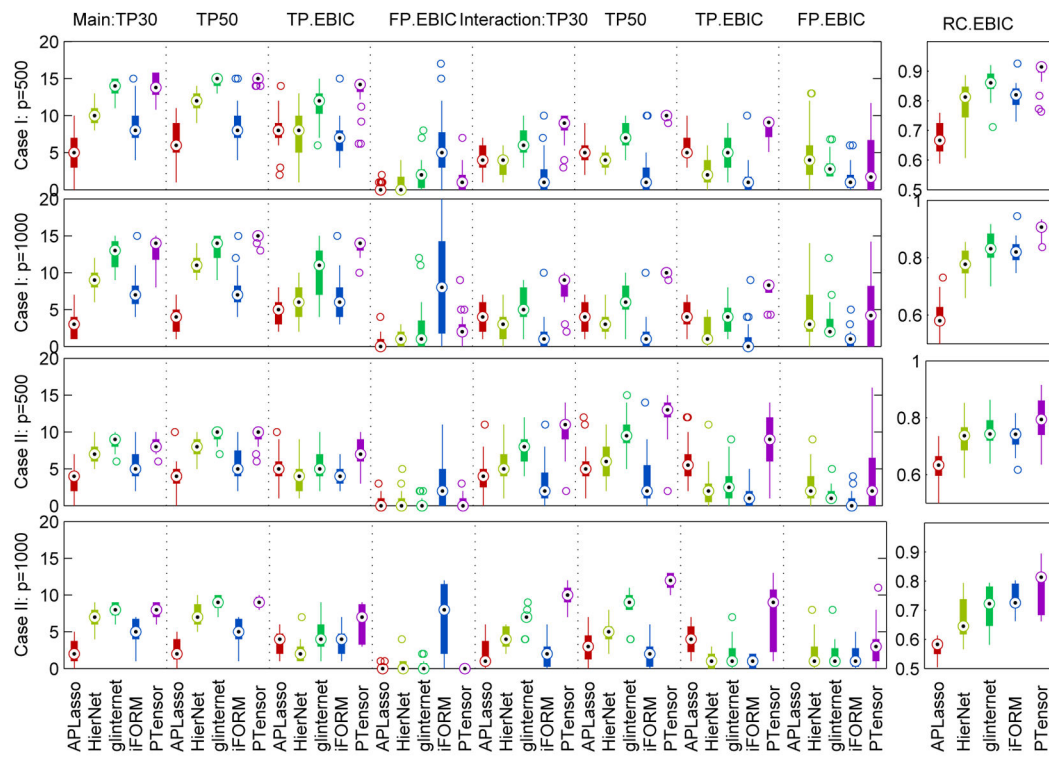
Figure 1.
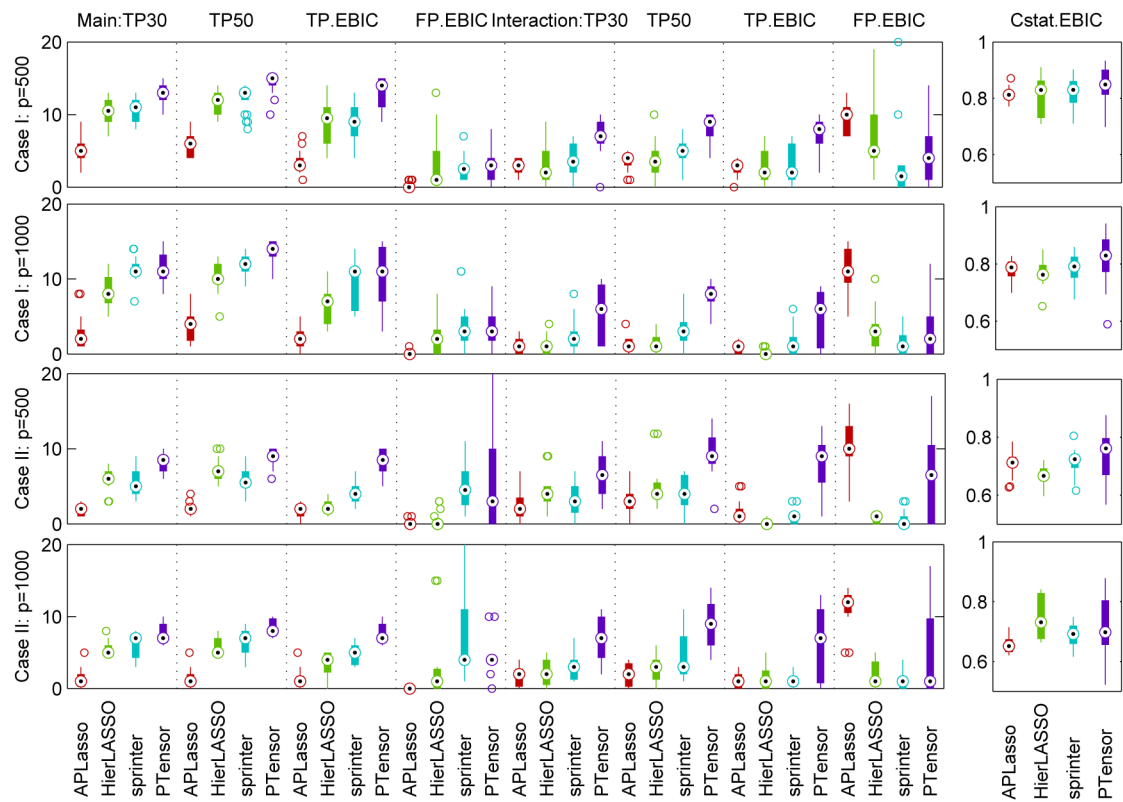Simulation Scenario 1: summary comparison of different methods.

Figure 2.
Simulation Scenario 2: summary comparison of different methods.

Table 1

Data analysis: numbers of main effects and interactions identified by different methods and their overlaps.

SKCM	Main					Interaction				
	APLasso	HierNet	glinternet	iFORM	PTensor	APLasso	HierNet	glinternet	iFORM	PTensor
APLasso	0	0	0	0	0	260	0	7	2	0
HierNet		14	4	0	1		14	0	0	0
glinternet			73	17	20			41	0	0
iFORM				61	8				29	0
PTensor					25					76

LUAD	Main				Interaction			
	APLasso	HierLasso	sprinter	PTensor	APLasso	HierLasso	sprinter	PTensor
APLasso	0	0	0	0	238	4	13	2
HierLasso		31	5	11		29	7	2
sprinter			9	6			110	2
PTensor				63				32