Supplementary Materials of "Gene-Set Integrative Analysis of Multi-Omics Data Using Tensor-based Association Tests"

Sheng-Mao Chang [1**], Meng Yang [2**], Wenbin Lu [2],
Yu-Jyun Huang [3], Yueyang Huang [4], Hung Hung [3], Jeff Miecznikowski [5],
Tzu-Pin Lu [3] and Jung-Ying Tzeng [1,2,3,4*]

[1]: Department of Statistics, National Cheng Kung University, Tainan, Taiwan

[2]: Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA

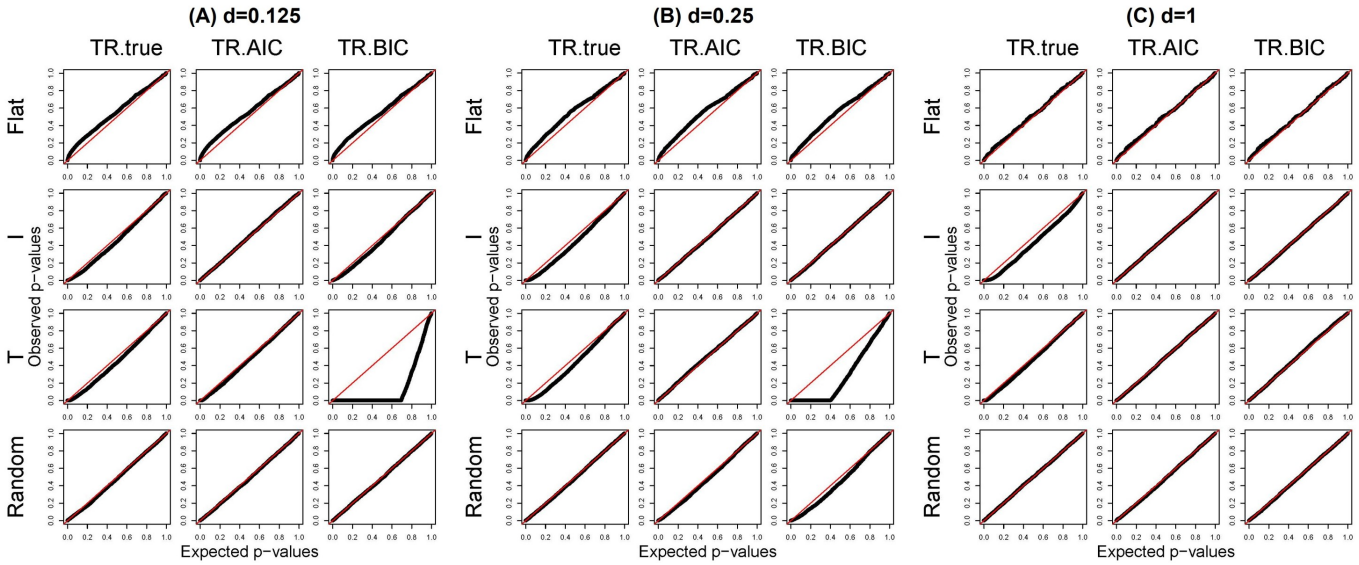
[3]: Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan

[4]: Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, USA

[5]: Department of Biostatistics, University at Buffalo, Buffalo New York, 14214, USA

** Equal contribution

* Correspondence authors: Jung-Ying Tzeng, E-mail: jytzeng@ncsu.edu



Supplementary Figure S1. Quantile-quantile plots of the null p-values of the tensor association test, obtained from TR.true (TR model evaluated at true rank), TR.AIC (TR model evaluated at AIC-selected rank) and TR.BIC (TR model evaluated at BIC-selected rank) under different $\mathbf{B}$ shapes. For a given $\mathbf{B}$ shape, the null p-values are obtained from those omics variables with $B_{pg} = 0$ when causal omics variables have effect strength $d = 0.125$, 0.25 and 1. The x axis represents the expected p-values and the y axis represents the observed p-values. The simulation is conducted using TCGA Pathway M13087 data, which retains correlation structure among omics variables.

Supplementary Table S1. Performance of selecting causal omics variables of tensor regression evaluated at true rank (TR.true), at AIC selected rank (TR.AIC), and at BIC selected rank (TR.BIC), as well as linear regression model (LM) and lasso on vectorized omics variables, based on 10^5 replications and effect strength $d = 0.25$. For TR and LM, two different selection rules are used: (a) p-value<0.05 and (b) Benjamini-Hochberg false discovery rate (BH-FDR)<0.05. TPR = true positive rate; FDR = false discovery rate. The best performed methods among TR.AIC, LM and LASSO, judged by F measures, are highlighted in shaded cells. The simulation is based on TCGA Pathway M13087 data, which retain the correlation structure among omics variables. The results show that using either selection rule, TR.AIC has higher F measures than LM and LASSO in almost all **B** shapes, except for "Flat" with Rule (b), where LASSO has the highest F measure.

B shape = Flat	TR.true	TR.AIC	TR.BIC	LM	LASSO
p-value < 0.05					
TPR	0.6486	0.6485	0.6486	0.5647	0.7566
FDR	0.0366	0.0383	0.0366	0.1460	0.3198
F measure	0.7714	0.7707	0.7715	0.6781	0.7153
BH-FDR < 0.05					
TPR	0.5226	0.5226	0.5227	0.4018	0.7566
FDR	0.0054	0.0064	0.0054	0.0329	0.3198
F measure	0.6835	0.6833	0.6836	0.5653	0.7153
B shape = I	TR.true	TR.AIC	TR.BIC	LM	LASSO
p-value < 0.05					
TPR	0.8095	0.8737	0.8754	0.5563	0.8736
FDR	0.2272	0.1188	0.1176	0.1614	0.4103
F measure	0.7894	0.8757	0.8772	0.6667	0.7036
BH-FDR < 0.05					
TPR	0.7606	0.8406	0.8433	0.3146	0.8736
FDR	0.1468	0.0385	0.0379	0.0340	0.4103
F measure	0.8002	0.8928	0.8950	0.4690	0.7036
B shape = T	TR.true	TR.AIC	TR.BIC	LM	LASSO
p-value < 0.05					
TPR	0.8657	0.8975	0.9723	0.8078	0.9326
FDR	0.1469	0.0601	0.2860	0.0634	0.3269
F measure	0.8587	0.9165	0.8085	0.8670	0.7817
BH-FDR < 0.05					
TPR	0.8433	0.8695	0.9667	0.7600	0.9326
FDR	0.1110	0.0276	0.2730	0.0262	0.3269
F measure	0.8645	0.9152	0.8124	0.8533	0.7817
B shape = Random	TR.true	TR.AIC	TR.BIC	LM	LASSO
p-value < 0.05					
TPR	0.8607	0.8742	0.9679	0.7125	0.9118
FDR	0.0304	0.0341	0.0401	0.0321	0.1845
F measure	0.9116	0.9166	0.9632	0.8204	0.8608
BH-FDR < 0.05					
TPR	0.8387	0.8539	0.9622	0.6491	0.9118
FDR	0.0187	0.0220	0.0300	0.0162	0.1845
F measure	0.9042	0.9100	0.9652	0.7815	0.8608

[Supplementary Figure S2 and Supplementary Table S2 are in Section S5.]

Supplementary Table S3. Effect estimates of omics variables on the drug sensitivity of Vandetanib in the CCLE analysis, **excluding negative control genes**. Values in parentheses are the corresponding p-values. Shaded cells indicate the selected important omics variables affecting Vandetanib sensitivity. For tensor regression (TR.AIC) and linear model (LM), a variable is selected as important if p-value<0.05. For LASSO, a variable is selected if it has non-zero coefficient. Omics variables with BH-FDR< 0.05 are shown in bold.

	CNV			Methylation			mRNA Expression		
	TR.AIC	LM	LASSO	TR.AIC	LM	LASSO	TR.AIC	LM	LASSO
EGFR	0.2054 (0.0267)	0.2818 (0.0127)	0.1321	**-0.2778 (0.0006)**	-0.2671 (0.0112)	-0.1669	0.1389 (0.1191)	0.0987 (0.3151)	0.1034
EREG	-0.0317 (0.4046)	0.1070 (0.1874)	0	0.0429 (0.4092)	0.1373 (0.1293)	0.0035	-0.0214 (0.4695)	0.0464 (0.5562)	0
HRAS	-0.0292 (0.5130)	-0.0927 (0.2938)	-0.1047	0.0395 (0.5347)	-0.0349 (0.6840)	0	-0.0197 (0.5860)	-0.0059 (0.9454)	0
KRAS	-0.0317 (0.4106)	0.1196 (0.2546)	0	0.0429 (0.4030)	0.0008 (0.9938)	0	-0.0214 (0.4168)	-0.1871 (0.0913)	-0.048
PTPN11	0.0668 (0.1933)	0.0297 (0.7338)	0	-0.0903 (0.0973)	-0.1089 (0.2701)	0	0.0452 (0.2208)	0.1280 (0.1919)	0
STAT3	-0.0460 (0.3088)	-0.0824 (0.3954)	0	0.0622 (0.2474)	0.1721 (0.0664)	0.0113	-0.0311 (0.3458)	0.0272 (0.7901)	0
TGFA	0.0338 (0.4992)	0.1023 (0.2373)	0	-0.0458 (0.4671)	0.0509 (0.5125)	0	0.0229 (0.4912)	0.0184 (0.8460)	0.0296

S1. Constrained Parameterization and Identifiablility of $\mathrm{B} = \mathrm{B}_1\mathrm{B}_2^\top$

The applied rank-R tensor model considers a pair of matrices $(\boldsymbol{B}_1, \boldsymbol{B}_2)$ to form the target matrix $\boldsymbol{B}$ by setting $\boldsymbol{B} = \boldsymbol{B}_1\boldsymbol{B}_2^\top$ where $\mathbf{B}_1 \in \mathbb{R}^{P\times R}$ and $\mathbf{B}_2 \in \mathbb{R}^{G\times R}$. This model implicitly represents that both $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ have full column rank, otherwise the rank should be less than R. This tensor model is non-identifiable because, for any two distinct invertible matrices, say $\boldsymbol{O}_1$ and $\boldsymbol{O}_2$, with conformable dimension, two pairs of matrices $(\boldsymbol{B}_1\boldsymbol{O}_1, \boldsymbol{B}_2\boldsymbol{O}_1^{-\top})$ and $(\boldsymbol{B}_1\boldsymbol{O}_2, \boldsymbol{B}_2\boldsymbol{O}_2^{-\top})$ result in the same $\boldsymbol{B}$. To fix this disadvantage, we apply the constraint that

$$\boldsymbol{B}_1 = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{B}_{12} \end{bmatrix} \tag{1}$$

with $\boldsymbol{C} \in \mathbb{R}^{R\times R}$ invertible and $\boldsymbol{B}_{12} \in \mathbb{R}^{(P-R)\times R}$. Note that, given $\boldsymbol{B}_1$, this $\boldsymbol{C}$ always exists with suitable row permutations. Without loss of generality, hereafter, we assume that $\boldsymbol{B}_1$ is permuted by row so that its upper $R \times R$ submatrix is invertible.

Proposition 1 *When $\boldsymbol{B}_1$ is defined as (1) with arbitrary, fixed and invertible $\boldsymbol{C}$, the factorization $\boldsymbol{B} = \boldsymbol{B}_1\boldsymbol{B}_2^\top$ is unique.*

Proof. Assume that $\boldsymbol{B}_1\boldsymbol{B}_2^\top = \boldsymbol{B} = \boldsymbol{B}_1^*(\boldsymbol{B}_2^*)^\top$ and both $\boldsymbol{B}_1$ and $\boldsymbol{B}_1^*$ satisfy (1). In this sequel,

$$\mathbf{0} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{B}_{12} \end{bmatrix} \boldsymbol{B}_2^\top - \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{B}_{12}^* \end{bmatrix} (\boldsymbol{B}_2^*)^\top = \begin{bmatrix} \boldsymbol{C}\boldsymbol{B}_2^\top - \boldsymbol{C}(\boldsymbol{B}_2^*)^\top \\ \boldsymbol{B}_{12}\boldsymbol{B}_2^\top - \boldsymbol{B}_{12}^*(\boldsymbol{B}_2^*)^\top \end{bmatrix}.$$

Because $\boldsymbol{C}$ is invertible, the first row of above equation implies $\boldsymbol{B}_2 = \boldsymbol{B}_2^*$. On the other hand, the second row implies

that

$$\begin{aligned} \mathbf{0} &= (\boldsymbol{B}_{12} - \boldsymbol{B}_{12}^*)\boldsymbol{B}_2^\top \\ \Rightarrow \mathbf{0} &= (\boldsymbol{B}_{12} - \boldsymbol{B}_{12}^*)\boldsymbol{B}_2^\top \boldsymbol{B}_2 \\ \Rightarrow \mathbf{0} &= (\boldsymbol{B}_{12} - \boldsymbol{B}_{12}^*) \end{aligned}$$

where the last line is because $\boldsymbol{B}_2^\top \boldsymbol{B}_2$ is invertible. Thus, we conclude that $\boldsymbol{B}_1^* = \boldsymbol{B}_1$ and $\boldsymbol{B}_2^* = \boldsymbol{B}_2$. In other words, with the constraint (1), the factorization $\boldsymbol{B} = \boldsymbol{B}_1\boldsymbol{B}_2^\top$ is uniquely determined. The proposed constraint (1) makes the tensor model identifiable.

S2. Equivalence between Linear Model and Tensor Regression When $R = min(P, G)$

Without loss of generality, assume that $R = min(P, G) = P$, and then $\boldsymbol{B}_1 = \boldsymbol{C}$ and $\boldsymbol{B} = \boldsymbol{B}_1\boldsymbol{B}_2^\top = \boldsymbol{C}\boldsymbol{B}_2^\top$. Also let $[\cdot]_{pg}$ be the operator that extracts the (p, g) element of a matrix. The proposed tensor model of rank P can be expressed as

$$y_i = \boldsymbol{z}_i^\top \boldsymbol{\beta} + \langle \mathbf{X}_i, \boldsymbol{B} \rangle + \boldsymbol{\epsilon}_i = \boldsymbol{z}_i^\top \boldsymbol{\beta} + \langle \mathbf{X}_i, \boldsymbol{C}\boldsymbol{B}_2^\top \rangle + \boldsymbol{\epsilon}_i = \boldsymbol{z}_i^\top \boldsymbol{\beta} + \sum_{p=1}^{P}\sum_{g=1}^{G} [\boldsymbol{X}_i]_{pg}[\boldsymbol{C}\boldsymbol{B}_2^\top]_{pg} + \boldsymbol{\epsilon}_i,$$

which corresponds to the linear model on vectorized $\boldsymbol{X}_i$; the result follows because $\boldsymbol{C}$ is invertable, there is a unique mapping between $\boldsymbol{B}_2$ and $\boldsymbol{B}$.

S3. Asymptotic Normality of $\hat{\mathbf{B}}$

S3.A. Asymptotic Normality of $\hat{\mathbf{B}}$ with a Constant Invertible Matrix $\boldsymbol{C}$ in $\boldsymbol{B}_1$

Before moving forward, we fix notation. Define $\boldsymbol{X}_{i1}^\top \in \mathbb{R}^{R\times G}$ and $\boldsymbol{X}_{i2}^\top \in \mathbb{R}^{(P-R)\times G}$ such that $\boldsymbol{X}_i^\top = [\boldsymbol{X}_{i1}^\top, \boldsymbol{X}_{i2}^\top]$, for $i = 1, \ldots, n$. Under low-rank assumption, we define $\boldsymbol{B} = \boldsymbol{B}_1\boldsymbol{B}_2^\top$ and $\boldsymbol{B}_1^\top = [\boldsymbol{C}, \boldsymbol{B}_{12}^\top]$, for some positive integer $R \leq P$ and $F \in \mathbb{R}^{R\times R}$ fixed. Thus, the tensor regression is $y_i = \mathbf{z}_i^\mathsf{T}\boldsymbol{\beta} + \langle \mathbf{X}_i, \mathbf{B}_1\mathbf{B}_2^\mathsf{T} \rangle + \epsilon_i$ and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. Define $\boldsymbol{b}_{12} = vec(\boldsymbol{B}_{12})$, $\boldsymbol{c} = vec(\boldsymbol{C})$, and $\boldsymbol{b}_2 = vec(\boldsymbol{B}_2)$. Further, the parameter vector is defined as $\boldsymbol{f}^\top = (\boldsymbol{b}_{12}^\top, \boldsymbol{b}_2^\top)$. Denote $\hat{\boldsymbol{f}}$ as the maximum likelihood estimator of the low-rank tensor regression. Hereafter, matrix or vector with a *hat* means the elements consisting of $\boldsymbol{f}$ are replaced by $\hat{\boldsymbol{f}}$. Finally, for $\mathbf{A} \in \mathbb{R}^{m_1\times n_1}$ with entries A_{ij} and $\mathbf{D} \in \mathbb{R}^{m_2\times n_2}$ with entries $D_{\ell k}$, their box product, which can be viewed as a permuted Kronecker product, denoted by $\mathbf{E} = \mathbf{A} \boxtimes \mathbf{D}$, is a $m_1m_2 \times n_1n_2$ matrix with its entries:

$$\mathbf{E}_{(i-1)m_2+\ell,(k-1)n_1+j} = \mathbf{A}_{ij}\mathbf{D}_{\ell k}.$$

Some quick comparisons of box product and the Kronecker product in inverse, transpose and distribution are:

$$\begin{aligned} (\mathbf{A}\otimes\mathbf{B})^{-1} &= \mathbf{A}^{-1}\otimes\mathbf{B}^{-1} & \qquad (\mathbf{A}\boxtimes\mathbf{B})^{-1} &= \mathbf{B}^{-1}\boxtimes\mathbf{A}^{-1} \\ (\mathbf{A}\otimes\mathbf{B})^\mathsf{T} &= \mathbf{A}^\mathsf{T}\otimes\mathbf{B}^\mathsf{T} & \qquad (\mathbf{A}\boxtimes\mathbf{B})^\mathsf{T} &= \mathbf{B}^\mathsf{T}\boxtimes\mathbf{A}^\mathsf{T} \\ (\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) &= \mathbf{AC}\otimes\mathbf{BD} & \qquad (\mathbf{A}\boxtimes\mathbf{B})(\mathbf{C}\boxtimes\mathbf{D}) &= \mathbf{AC}\boxtimes\mathbf{BD} \end{aligned}$$

In the following proposition, we summarize the asymptotic result of $\hat{\boldsymbol{B}}$. Without loss of generality, we assume that $\boldsymbol{\beta}$ and σ^2 are known.

Proposition 2 *Under the low-rank assumption, the tensor regression* $y_i = \langle \boldsymbol{X}_i, \boldsymbol{B}_1\boldsymbol{B}_2^{\intercal}\rangle + \epsilon_i$ *for* $i = 1, \ldots, n$ *and* $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ *has Fisher information matrix*

$$\Im(\boldsymbol{f}) = E\left(\frac{\partial l}{\partial \boldsymbol{f}}\frac{\partial l}{\partial \boldsymbol{f}}^{\top}\right) = \sum_{i=1}^{n}\frac{1}{\sigma^2}\begin{bmatrix} \boldsymbol{K}_{i2}\boldsymbol{b}_2 \\ \boldsymbol{K}_{i2}^{\top}\boldsymbol{b}_{12} + \boldsymbol{K}_{i1}^{\top}\boldsymbol{c}\end{bmatrix}\left[\boldsymbol{b}_2^{\top}\boldsymbol{K}_{i2}^{\top}, \boldsymbol{b}_{12}^{\top}\boldsymbol{K}_{i2} + \boldsymbol{c}^{\top}\boldsymbol{K}_{i1}\right]$$

where $\boldsymbol{K}_{ij} = \boldsymbol{I}_R \otimes \boldsymbol{X}_{ij}$, $j = 1, 2$. *As the result, asymptotically,*

$$\sqrt{n} \times vec\left(\hat{\boldsymbol{B}}^{\top} - \boldsymbol{B}^{\top}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \Sigma(\boldsymbol{C})\right)$$

where

$$\Sigma(\boldsymbol{C}) = \begin{bmatrix} \boldsymbol{A}_4\boldsymbol{A}_1\Im^{-1}(\boldsymbol{f})\boldsymbol{A}_1^{\top}\boldsymbol{A}_4^{\top} & \boldsymbol{A}_4\boldsymbol{A}_1\Im^{-1}(\boldsymbol{f})\boldsymbol{A}_1^{\top}\boldsymbol{A}_3^{\top} \\ \boldsymbol{A}_3\boldsymbol{A}_1\Im^{-1}(\boldsymbol{f})\boldsymbol{A}_1^{\top}\boldsymbol{A}_4^{\top} & \boldsymbol{A}_2\Im^{-1}(\boldsymbol{f})\boldsymbol{A}_2^{\top}\end{bmatrix},$$

$\boldsymbol{A}_1 = \left[\boldsymbol{0}_{GR\times(P-R)R}, \boldsymbol{I}_{GR}\right]$, $\boldsymbol{A}_2 = [\boldsymbol{I}_{P-R} \boxtimes \boldsymbol{B}_2, \boldsymbol{B}_{12} \otimes \boldsymbol{I}_G]$, $\boldsymbol{A}_3 = \boldsymbol{B}_{12} \otimes \boldsymbol{I}_G$ *and* $\boldsymbol{A}_4 = \boldsymbol{C} \otimes \boldsymbol{I}_G$.

Proof. First of all, because

$$\begin{aligned}\langle \boldsymbol{X}_i, \boldsymbol{B}\rangle =& tr(\boldsymbol{X}_i^{\top}\boldsymbol{B}_1\boldsymbol{B}_2^{\top}) = tr(\boldsymbol{B}_2^{\top}\boldsymbol{X}_i^{\top}\boldsymbol{B}_1) \\ =& tr\left(\boldsymbol{B}_2^{\top}[\boldsymbol{X}_{i1}^{\top}, \boldsymbol{X}_{i2}^{\top}]\begin{bmatrix}\boldsymbol{C} \\ \boldsymbol{B}_{12}\end{bmatrix}\right) = tr(\boldsymbol{B}_2^{\top}\boldsymbol{X}_{i1}^{\top}\boldsymbol{C} + \boldsymbol{B}_2^{\top}\boldsymbol{X}_{i2}^{\top}\boldsymbol{B}_{12}) \\ =& \langle \boldsymbol{X}_{i1}^{\top}\boldsymbol{C}, \boldsymbol{B}_2\rangle + \langle \boldsymbol{X}_{i2}\boldsymbol{B}_2, \boldsymbol{B}_{12}\rangle = \langle \boldsymbol{X}_{i1}^{\top}\boldsymbol{C}, \boldsymbol{B}_2\rangle + \langle \boldsymbol{X}_{i2}^{\top}\boldsymbol{B}_{12}, \boldsymbol{B}_2\rangle\end{aligned}$$

we have

$$\frac{\partial\langle \boldsymbol{X}_i, \boldsymbol{B}\rangle}{\partial \boldsymbol{b}_{12}} = vec(\boldsymbol{X}_{i2}\boldsymbol{B}_2\boldsymbol{I}_R) = (\boldsymbol{I}_R \otimes \boldsymbol{X}_{i2})\boldsymbol{b}_2$$

and

$$\frac{\partial\langle \boldsymbol{X}_i, \boldsymbol{B}\rangle}{\partial \boldsymbol{b}_2} = vec(\boldsymbol{X}_{i1}^{\top}\boldsymbol{C}\boldsymbol{I}_R) + vec(\boldsymbol{X}_{i2}^{\top}\boldsymbol{B}_{12}\boldsymbol{I}_R) = (\boldsymbol{I}_R \otimes \boldsymbol{X}_{i1})^{\top}\boldsymbol{c} + (\boldsymbol{I}_R \otimes \boldsymbol{X}_{i2})^{\top}\boldsymbol{b}_{12}$$

Define $l = -\sum_{i=1}^{n}(y_i - \langle \boldsymbol{X}_i, \boldsymbol{B}\rangle)^2/\sigma^2$. Then, according to the derivatives above, we have

$$\frac{\partial l}{\partial \boldsymbol{f}} = \sum_{i=1}^{n}\begin{bmatrix}(\boldsymbol{I}_R \otimes \boldsymbol{X}_{i2})\boldsymbol{b}_2 \\ (\boldsymbol{I}_R \otimes \boldsymbol{X}_{i2})^{\top}\boldsymbol{b}_{12} + (\boldsymbol{I}_R \otimes \boldsymbol{X}_{i1})^{\top}\boldsymbol{c}\end{bmatrix}\frac{(y_i - \langle \boldsymbol{X}_i, \boldsymbol{B}\rangle)}{\sigma^2}$$

which implies that the Fisher information matrix with respect to $\boldsymbol{f}$ is

$$\Im(\boldsymbol{f}) = E\left(\frac{\partial l}{\partial \boldsymbol{f}}\frac{\partial l}{\partial \boldsymbol{f}}^{\top}\right) = \sum_{i=1}^{n}\frac{1}{\sigma^2}\begin{bmatrix} \boldsymbol{K}_{i2}\boldsymbol{b}_2 \\ \boldsymbol{K}_{i2}^{\top}\boldsymbol{b}_{12} + \boldsymbol{K}_{i1}^{\top}\boldsymbol{c}\end{bmatrix}\left[\boldsymbol{b}_2^{\top}\boldsymbol{K}_{i2}^{\top}, \boldsymbol{b}_{12}^{\top}\boldsymbol{K}_{i2} + \boldsymbol{c}^{\top}\boldsymbol{K}_{i1}\right]$$

where $\boldsymbol{K}_{i1} = \boldsymbol{I}_R \otimes \boldsymbol{X}_{i1}$ and $\boldsymbol{K}_{i2} = \boldsymbol{I}_R \otimes \boldsymbol{X}_{i2}$. Consequently, by central limit theorem

$$\sqrt{n}\left(\hat{\boldsymbol{f}} - \boldsymbol{f}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \Im^{-1}(\boldsymbol{f})\right)$$

For convenience, we consider the asymptotic distribution of

$$\hat{\boldsymbol{B}}^{\top} = \left(\begin{bmatrix}\boldsymbol{C} \\ \hat{\boldsymbol{B}}_{12}\end{bmatrix}\hat{\boldsymbol{B}}_2^{\top}\right)^{\top} = \left[\hat{\boldsymbol{B}}_2\boldsymbol{C}^{\top}, \hat{\boldsymbol{B}}_2\hat{\boldsymbol{B}}_{12}^{\top}\right]$$

instead of $\hat{\boldsymbol{B}}$. Because $vec(\boldsymbol{B}_2) = \boldsymbol{A}_1\boldsymbol{f}$, we have

$$\sqrt{n} \times vec\left(\hat{\boldsymbol{B}}_2\boldsymbol{C}^\top - \boldsymbol{B}_2\boldsymbol{C}^\top\right) = \sqrt{n}\boldsymbol{A}_4\boldsymbol{A}_1\left(\hat{\boldsymbol{f}} - \boldsymbol{f}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{A}_4\boldsymbol{A}_1\Im^{-1}(\boldsymbol{f})\boldsymbol{A}_1^\top\boldsymbol{A}_4^\top\right) \tag{2}$$

Additionally,

$$\begin{aligned}
&\sqrt{n} \times vec\left(\hat{\boldsymbol{B}}_2\hat{\boldsymbol{B}}_{12}^\top - \boldsymbol{B}_2\boldsymbol{B}_{12}^\top\right) = \sqrt{n} \times vec\left(\hat{\boldsymbol{B}}_2\hat{\boldsymbol{B}}_{12}^\top - \boldsymbol{B}_2\hat{\boldsymbol{B}}_{12}^\top + \boldsymbol{B}_2\hat{\boldsymbol{B}}_{12}^\top - \boldsymbol{B}_2\boldsymbol{B}_{12}^\top\right) \\
=&\sqrt{n} \times vec\left(\boldsymbol{I}_G(\hat{\boldsymbol{B}}_2 - \boldsymbol{B}_2)\hat{\boldsymbol{B}}_{12}^\top + \boldsymbol{B}_2(\hat{\boldsymbol{B}}_{12} - \boldsymbol{B}_{12})^\top\boldsymbol{I}_{P-R}\right) \\
=&\sqrt{n}\left\{\left(\hat{\boldsymbol{B}}_{12} \otimes \boldsymbol{I}_G\right) vec\left(\hat{\boldsymbol{B}}_2 - \boldsymbol{B}_2\right) + \left(\boldsymbol{I}_{P-R} \boxtimes \boldsymbol{B}_2\right) vec\left(\hat{\boldsymbol{B}}_{12} - \boldsymbol{B}_{12}\right)\right\} \\
=&\sqrt{n}\left[\boldsymbol{I}_{P-R} \boxtimes \boldsymbol{B}_2, \hat{\boldsymbol{B}}_{12} \otimes \boldsymbol{I}_G\right]\left(\hat{\boldsymbol{f}} - \boldsymbol{f}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{A}_2\Im^{-1}(\boldsymbol{f})\boldsymbol{A}_2^\top\right)
\end{aligned} \tag{3}$$

by Slutsky's theorem. Last,

$$\begin{aligned}
&Cov(vec(\hat{\boldsymbol{B}}_2\boldsymbol{C}^\top), vec(\boldsymbol{I}_G\hat{\boldsymbol{B}}_2\hat{\boldsymbol{B}}_{12}^\top)) = Cov\left(\left(\boldsymbol{C} \otimes \boldsymbol{I}_G\right)\hat{\boldsymbol{b}}_2, \left(\hat{\boldsymbol{B}}_{12} \otimes \boldsymbol{I}_G\right)\hat{\boldsymbol{b}}_2\right) \\
&= Cov\left(\left(\boldsymbol{C} \otimes \boldsymbol{I}_G\right)\hat{\boldsymbol{b}}_2, \left[\left(\hat{\boldsymbol{B}}_{12} - \boldsymbol{B}_{12}\right) \otimes \boldsymbol{I}_G\right]\hat{\boldsymbol{b}}_2\right) + Cov\left(\left(\boldsymbol{C} \otimes \boldsymbol{I}_G\right)\hat{\boldsymbol{b}}_2, \left(\boldsymbol{B}_{12} \otimes \boldsymbol{I}_G\right)\hat{\boldsymbol{b}}_2\right) \\
&\xrightarrow{\mathcal{P}} \boldsymbol{A}_4 var\left(\hat{\boldsymbol{b}}_2\right)\boldsymbol{A}_3^\top = \boldsymbol{A}_4\boldsymbol{A}_1\Im^{-1}(\boldsymbol{f})\boldsymbol{A}_1^\top\boldsymbol{A}_3^\top
\end{aligned} \tag{4}$$

Together with (2), (3), and (4), we have

$$\sqrt{n} \times vec\left(\hat{\boldsymbol{B}}^\top - \boldsymbol{B}^\top\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \Sigma(\boldsymbol{C})\right),$$

where

$$\Sigma(\boldsymbol{C}) = \begin{bmatrix} \boldsymbol{A}_4\boldsymbol{A}_1\Im^{-1}(\boldsymbol{f})\boldsymbol{A}_1^\top\boldsymbol{A}_4^\top & \boldsymbol{A}_4\boldsymbol{A}_1\Im^{-1}(\boldsymbol{f})\boldsymbol{A}_1^\top\boldsymbol{A}_3^\top \\ \boldsymbol{A}_3\boldsymbol{A}_1\Im^{-1}(\boldsymbol{f})\boldsymbol{A}_1^\top\boldsymbol{A}_4^\top & \boldsymbol{A}_2\Im^{-1}(\boldsymbol{f})\boldsymbol{A}_2^\top \end{bmatrix}. \tag{5}$$

Last, the variance estimator is defined as

$$\widehat{\Sigma}(\boldsymbol{C}) = \begin{bmatrix} \hat{\boldsymbol{A}}_4\boldsymbol{A}_1\Im^{-1}(\hat{\boldsymbol{f}})\boldsymbol{A}_1^\top\hat{\boldsymbol{A}}_4^\top & \hat{\boldsymbol{A}}_4\boldsymbol{A}_1\Im^{-1}(\hat{\boldsymbol{f}})\boldsymbol{A}_1^\top\hat{\boldsymbol{A}}_3^\top \\ \hat{\boldsymbol{A}}_3\boldsymbol{A}_1\Im^{-1}(\hat{\boldsymbol{f}})\boldsymbol{A}_1^\top\hat{\boldsymbol{A}}_4^\top & \hat{\boldsymbol{A}}_2\Im^{-1}(\hat{\boldsymbol{f}})\hat{\boldsymbol{A}}_2^\top \end{bmatrix}, \tag{6}$$

where $\hat{\boldsymbol{A}}_2 = \left[\boldsymbol{I}_{P-R} \boxtimes \hat{\boldsymbol{B}}_2, \hat{\boldsymbol{B}}_{12} \otimes \boldsymbol{I}_G\right]$, $\hat{\boldsymbol{A}}_3 = \hat{\boldsymbol{B}}_{12} \otimes \boldsymbol{I}_G$, and $\hat{\boldsymbol{A}}_4 = \boldsymbol{C} \otimes \boldsymbol{I}_G$. It can be shown that $\widehat{\Sigma}(\boldsymbol{C})$ is a consistent estimator of $\Sigma(\mathbf{C})$.

S3.B. Extending Results of Section S3.A to a Data-Dependent $\boldsymbol{C}_n$ in $\boldsymbol{B}_1$

In Section S3.A, we obtain the normality of $\hat{\boldsymbol{B}}$ for an arbitrary constant invertible matrix $\boldsymbol{C}$. Here we extend the results to a constant invertible matrix $\boldsymbol{C}_n$, which may be data dependent. Let $\hat{\boldsymbol{B}}_{12,n}$ and $\hat{\boldsymbol{B}}_{2,n}$ be the estimators based on the representation in Proposition 1 (i.e., $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{C}_n \\ \boldsymbol{B}_{12} \end{bmatrix} \boldsymbol{B}_2^\top$) and let $\hat{\boldsymbol{B}}_n = \begin{bmatrix} \boldsymbol{C}_n \\ \hat{\boldsymbol{B}}_{12,n} \end{bmatrix} \hat{\boldsymbol{B}}_{2,n}^\top$. Also recall that $\hat{\boldsymbol{B}} = \hat{\boldsymbol{B}}_1\hat{\boldsymbol{B}}_2^\top$ with $\hat{\boldsymbol{B}}_1 = \begin{bmatrix} \boldsymbol{C} \\ \hat{\boldsymbol{B}}_{12} \end{bmatrix}$ and $\boldsymbol{C}$ invertible. Below we show that

(i) $\hat{\boldsymbol{B}}_n = \hat{\boldsymbol{B}}$, and

(ii) $\widehat{\Sigma}_n(\boldsymbol{C}_n) = \widehat{\Sigma}(\boldsymbol{C})$,

where $\widehat{\Sigma}_n(\boldsymbol{C}_n)$, defined in a similar fashion as $\widehat{\Sigma}(\boldsymbol{C})$ of Equation (6), is the variance estimator of $\hat{\boldsymbol{B}}$ evaluated at $(C_n, \hat{\boldsymbol{B}}_{12,n}, \hat{\boldsymbol{B}}_{2,n})$. Then with (i), we have that $\sqrt{n} \times \left(\hat{\boldsymbol{B}}_n - \boldsymbol{B}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0}, \Sigma(\boldsymbol{C}))$; and (ii) says that the variance estimators of $\hat{B}$ computed based on $(C, \hat{\boldsymbol{B}}_{12}, \hat{\boldsymbol{B}}_2)$ or $(C_n, \hat{\boldsymbol{B}}_{12,n}, \hat{\boldsymbol{B}}_{2,n})$ are the same, i.e., $\widehat{\Sigma}(\boldsymbol{C})$ is invariant to $\boldsymbol{C}$. (i) and (ii) together justify our inference procedure based on the estimator $\hat{\boldsymbol{B}}_n$ obtained using the alternating least square (ALS) algorithm given in Section S4.

To show (i), rewrite

$$\hat{\boldsymbol{B}} = \hat{\boldsymbol{B}}_1 \hat{\boldsymbol{B}}_2^\top = \hat{\boldsymbol{B}}_1 \boldsymbol{C}^{-1} \boldsymbol{C}_n \boldsymbol{C}_n^{-1} \boldsymbol{C} \hat{\boldsymbol{B}}_2^\top = \begin{bmatrix} \boldsymbol{C}_n \\ \hat{\boldsymbol{B}}_{12} \boldsymbol{C}^{-1} \boldsymbol{C}_n \end{bmatrix} \left[\boldsymbol{C}_n^{-1} \boldsymbol{C} \hat{\boldsymbol{B}}_2^\top \right].$$

The last term satisfies Proposition 1 and hence is unique. Therefore, we have $\hat{\boldsymbol{B}}_{12,n} = \hat{\boldsymbol{B}}_{12} \boldsymbol{C}^{-1} \boldsymbol{C}_n$ and $\hat{\boldsymbol{B}}_{2,n}^\top = \boldsymbol{C}_n^{-1} \boldsymbol{C} \hat{\boldsymbol{B}}_2^\top$. In other words, $\hat{\boldsymbol{B}} = \hat{\boldsymbol{B}}_n$.

We show (ii) in the following proposition.

Proposition 3 *$\widehat{\Sigma}(\boldsymbol{C})$ is invariant to $\boldsymbol{C}$, i.e., $\widehat{\Sigma}_n(\boldsymbol{C}_n) = \widehat{\Sigma}(\boldsymbol{C})$.*

Proof. Define $\hat{\boldsymbol{f}}_n = \left(vec^\top(\hat{\boldsymbol{B}}_{12,n}), vec^\top(\hat{\boldsymbol{B}}_{2,n})\right)^\top$ and $\hat{\boldsymbol{f}}$ is the same as defined in Section S3.A. Since $vec(\hat{\boldsymbol{B}}_{12,n}) = vec(\hat{\boldsymbol{B}}_{12} \boldsymbol{C}^{-1} \boldsymbol{C}_n) = \left\{ \boldsymbol{C}_n^\top \boldsymbol{C}^{-\top} \otimes \boldsymbol{I}_{P-R} \right\} vec(\hat{\boldsymbol{B}}_{12})$ and $vec(\widehat{\boldsymbol{B}}_{2,n}) = vec(\hat{\boldsymbol{B}}_2 \boldsymbol{C}^\top \boldsymbol{C}_n^{-\top})$
$= \left\{ \boldsymbol{C}_n^{-1} \boldsymbol{C} \otimes \boldsymbol{I}_G \right\} vec(\hat{\boldsymbol{B}}_2)$, we have $\hat{\boldsymbol{f}}_n = \boldsymbol{L} \hat{\boldsymbol{f}}$ where

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{C}_n^\top \boldsymbol{C}^{-\top} \otimes \boldsymbol{I}_{P-R} & 0 \\ 0 & \boldsymbol{C}_n^{-1} \boldsymbol{C} \otimes \boldsymbol{I}_G \end{bmatrix}.$$

In other words, $\hat{\boldsymbol{f}}_n$ is a linear transformation of $\hat{\boldsymbol{f}}$. So $Var\left(\hat{\boldsymbol{f}}_n\right) = \boldsymbol{L} Var\left(\hat{\boldsymbol{f}}\right) \boldsymbol{L}^\top = \boldsymbol{L} \Im^{-1}(\boldsymbol{f}) \boldsymbol{L}^\top$. This implies that

$$\begin{aligned} \widehat{\Sigma}_n(\boldsymbol{C}_n) &= \begin{bmatrix} \hat{\boldsymbol{A}}_{4,n} \boldsymbol{A}_1 \Im^{-1}(\hat{\boldsymbol{f}}_n) \boldsymbol{A}_1^\top \hat{\boldsymbol{A}}_{4,n}^\top & \hat{\boldsymbol{A}}_{4,n} \boldsymbol{A}_1 \Im^{-1}(\hat{\boldsymbol{f}}_n) \boldsymbol{A}_1^\top \hat{\boldsymbol{A}}_{3,n}^\top \\ \hat{\boldsymbol{A}}_{3,n} \boldsymbol{A}_1 \Im^{-1}(\hat{\boldsymbol{f}}_n) \boldsymbol{A}_1^\top \hat{\boldsymbol{A}}_{4,n}^\top & \hat{\boldsymbol{A}}_{2,n} \Im^{-1}(\hat{\boldsymbol{f}}_n) \hat{\boldsymbol{A}}_{2,n}^\top \end{bmatrix} \\ &= \begin{bmatrix} \hat{\boldsymbol{A}}_{4,n} \boldsymbol{A}_1 \boldsymbol{L} \Im^{-1}(\hat{\boldsymbol{f}}) \boldsymbol{L}^\top \boldsymbol{A}_1^\top \hat{\boldsymbol{A}}_{4,n}^\top & \hat{\boldsymbol{A}}_{4,n} \boldsymbol{A}_1 \boldsymbol{L} \Im^{-1}(\hat{\boldsymbol{f}}) \boldsymbol{L}^\top \boldsymbol{A}_1^\top \hat{\boldsymbol{A}}_{3,n}^\top \\ \hat{\boldsymbol{A}}_{3,n} \boldsymbol{A}_1 \boldsymbol{L} \Im^{-1}(\hat{\boldsymbol{f}}) \boldsymbol{L}^\top \boldsymbol{A}_1^\top \hat{\boldsymbol{A}}_{4,n}^\top & \hat{\boldsymbol{A}}_{2,n} \boldsymbol{L} \Im^{-1}(\hat{\boldsymbol{f}}) \boldsymbol{L}^\top \hat{\boldsymbol{A}}_{2,n}^\top \end{bmatrix}, \end{aligned}$$

where $\hat{\boldsymbol{A}}_{2,n} = \left[\boldsymbol{I}_{P-R} \boxtimes \hat{\boldsymbol{B}}_{2,n}, \hat{\boldsymbol{B}}_{12,n} \otimes \boldsymbol{I}_G \right]$, $\hat{\boldsymbol{A}}_{3,n} = \hat{\boldsymbol{B}}_{12,n} \otimes \boldsymbol{I}_G$, and $\hat{\boldsymbol{A}}_{4,n} = \boldsymbol{C}_n \otimes \boldsymbol{I}_G$. Thus, the proof is complete if $\hat{\boldsymbol{A}}_{4,n} \boldsymbol{A}_1 \boldsymbol{L} = \hat{\boldsymbol{A}}_4 \boldsymbol{A}_1$, $\hat{\boldsymbol{A}}_{3,n} \boldsymbol{A}_1 \boldsymbol{L} = \hat{\boldsymbol{A}}_3 \boldsymbol{A}_1$, and $\hat{\boldsymbol{A}}_{2,n} \boldsymbol{L} = \hat{\boldsymbol{A}}_2$ hold.

Note that $\boldsymbol{A}_1 \boldsymbol{L} = [\boldsymbol{O}, \boldsymbol{C}_n^{-1} \boldsymbol{C} \otimes \boldsymbol{I}_G]$. Then, the first equation holds because

$$\hat{\boldsymbol{A}}_{4,n} \boldsymbol{A}_1 \boldsymbol{L} = (\boldsymbol{C}_n \otimes \boldsymbol{I}_G) [\boldsymbol{O}, \boldsymbol{C}_n^{-1} \boldsymbol{C} \otimes \boldsymbol{I}_G] = [\boldsymbol{O}, \boldsymbol{C}_n \boldsymbol{C}_n^{-1} \boldsymbol{C} \otimes \boldsymbol{I}_G] = [\boldsymbol{O}, \boldsymbol{C} \otimes \boldsymbol{I}_G] = \hat{\boldsymbol{A}}_4 \boldsymbol{A}_1.$$

Similarly, the second equation holds because

$$\hat{\boldsymbol{A}}_{3,n} \boldsymbol{A}_1 \boldsymbol{L} = (\hat{\boldsymbol{B}}_{12,n} \otimes \boldsymbol{I}_G) \boldsymbol{A}_1 \boldsymbol{L} = (\hat{\boldsymbol{B}}_{12} \boldsymbol{C}^{-1} \boldsymbol{C}_n \otimes \boldsymbol{I}_G) \boldsymbol{A}_1 \boldsymbol{L} = [\boldsymbol{O}, \hat{\boldsymbol{B}}_{12} \otimes \boldsymbol{I}_G] = \hat{\boldsymbol{A}}_3 \boldsymbol{A}_1.$$

Last, we have

$$\hat{\boldsymbol{A}}_{2,n} \boldsymbol{L} = [(\boldsymbol{I}_{P-R} \boxtimes \hat{\boldsymbol{B}}_{2,n})(\boldsymbol{C}_n^\top \boldsymbol{C}^{-\top} \otimes \boldsymbol{I}_{P-R}),\ \hat{\boldsymbol{B}}_{12,n} \boldsymbol{C}_n^{-1} \boldsymbol{C} \otimes \boldsymbol{I}_G]. \tag{7}$$

Define $\boldsymbol{e}_k$ as the kth column of an $R(P-R) \times R(P-R)$ identity matrix, and $\boldsymbol{E}_k$ as a $(P-R) \times R$ matrix with $vec(\boldsymbol{E}_k) = \boldsymbol{e}_k$. Then, the kth column of the first element of (7) can be expressed as

$$(\boldsymbol{I}_{P-R} \boxtimes \hat{\boldsymbol{B}}_{2,n})(\boldsymbol{C}_n^\top \boldsymbol{C}^{-\top} \otimes \boldsymbol{I}_{P-R})\boldsymbol{e}_k = (\boldsymbol{I}_{P-R} \boxtimes \hat{\boldsymbol{B}}_{2,n})vec(\boldsymbol{E}_k \boldsymbol{C}^{-1}\boldsymbol{C}_n) = vec(\hat{\boldsymbol{B}}_{2,n}\boldsymbol{C}_n^\top \boldsymbol{C}^{-\top}\boldsymbol{E}_k^\top)$$
$$= vec(\hat{\boldsymbol{B}}_2 \boldsymbol{E}_k^\top) = (\boldsymbol{I}_{P-R} \boxtimes \hat{\boldsymbol{B}}_2)\boldsymbol{e}_k.$$

Thus, we have $(\boldsymbol{I}_{P-R} \boxtimes \hat{\boldsymbol{B}}_{2,n})(\boldsymbol{C}_n^\top \boldsymbol{C}^{-\top} \otimes \boldsymbol{I}_{P-R}) = \boldsymbol{I}_{P-R} \boxtimes \hat{\boldsymbol{B}}_2$. The second element of (7) can be rewritten as $\hat{\boldsymbol{B}}_{12}\boldsymbol{C}^{-1}\boldsymbol{C}_n\boldsymbol{C}_n^{-1}\boldsymbol{C} \otimes \boldsymbol{I}_G = \hat{\boldsymbol{B}}_{12} \otimes \boldsymbol{I}_G$. Therefore, $\hat{\boldsymbol{A}}_{2,n}\boldsymbol{L} = [\boldsymbol{I}_{P-R} \boxtimes \hat{\boldsymbol{B}}_2, \hat{\boldsymbol{B}}_{12} \otimes \boldsymbol{I}_G] = \hat{\boldsymbol{A}}_2$. So the proof is complete.

S4. Parameter Estimation: the Alternating Least Square (ALS) Algorithm

For estimating $\mathbf{B}_1$ and $\mathbf{B}_2$, the term involved is $\langle \mathbf{X}_i, \mathbf{B}_1\mathbf{B}_2^\mathsf{T}\rangle$. By the cyclic property of the trace estimator, we have $\langle \mathbf{X}_i, \mathbf{B}_1\mathbf{B}_2^\mathsf{T}\rangle = \langle \mathbf{X}_i\mathbf{B}_2, \mathbf{B}_1\rangle = \langle \mathbf{X}_i^\mathsf{T}\mathbf{B}_1, \mathbf{B}_2\rangle$. That is, the response vector $\boldsymbol{y}$ is linear in $\mathbf{B}_1$ given $\mathbf{B}_2$ and linear in $\mathbf{B}_2$ given $\mathbf{B}_1$ for Gaussian response variable.

$$\begin{aligned} E(y_i \mid \mathbf{X}_i, \boldsymbol{z}_i) &= \boldsymbol{z}_i^\mathsf{T}\boldsymbol{\beta} + \langle \mathbf{X}_i, \mathbf{B}_1\mathbf{B}_2^\mathsf{T}\rangle \\ &= \boldsymbol{z}_i^\mathsf{T}\boldsymbol{\beta} + \langle \mathbf{X}_i\mathbf{B}_2, \mathbf{B}_1\rangle = \boldsymbol{z}_i^\mathsf{T}\boldsymbol{\beta} + \langle [\mathcal{A}(\mathbf{B}_2^{(j)})]_i, \mathbf{B}_1\rangle \qquad (8) \\ &= \boldsymbol{z}_i^\mathsf{T}\boldsymbol{\beta} + \langle \mathbf{X}_i^\mathsf{T}\mathbf{B}_1, \mathbf{B}_2\rangle = \boldsymbol{z}_i^\mathsf{T}\boldsymbol{\beta} + \langle [\mathcal{B}(\mathbf{B}_1^{(j+1)})]_i, \mathbf{B}_2\rangle, \qquad (9) \end{aligned}$$

where $[\bullet]_i$ is the i-th row of $\bullet$, $\mathcal{A}(\mathbf{B}_2) = \begin{bmatrix} vec(\mathbf{X}_1\mathbf{B}_2)^\mathsf{T} \\ vec(\mathbf{X}_2\mathbf{B}_2)^\mathsf{T} \\ \vdots \\ vec(\mathbf{X}_n\mathbf{B}_2)^\mathsf{T} \end{bmatrix}$, $\mathcal{B}(\mathbf{B}_1) = \begin{bmatrix} vec(\mathbf{X}_1^\mathsf{T}\mathbf{B}_1)^\mathsf{T} \\ vec(\mathbf{X}_2^\mathsf{T}\mathbf{B}_1)^\mathsf{T} \\ \vdots \\ vec(\mathbf{X}_n^\mathsf{T}\mathbf{B}_1)^\mathsf{T} \end{bmatrix}$, $\mathbf{B}_\bullet^{(j)}$ is the estimates of $\mathbf{B}_\bullet$ in the jth iteration. (8) and (9) show that we can update either $\mathbf{B}_1$ or $\mathbf{B}_2$ by solving a simple least square problem. Consequently, we obtain the following estimation algorithm:

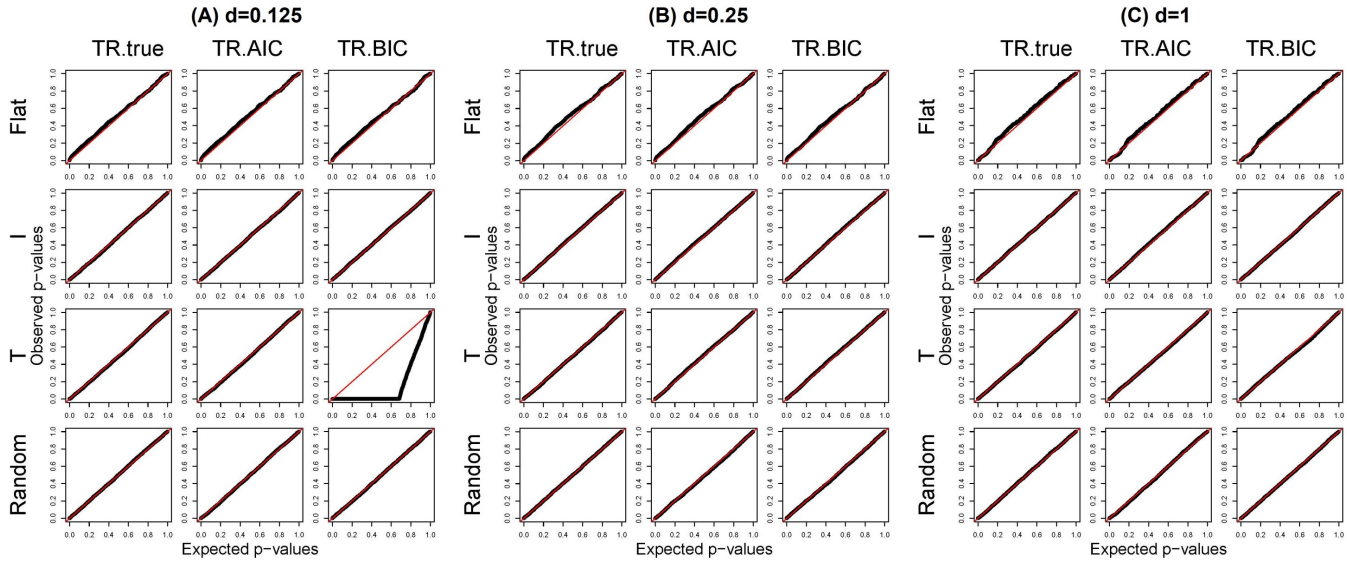
Algorithm 1 Tensor Regression in order-3 case for Gaussian outcome variables (TR-G)

procedure TR-G($\mathbf{B}_1^{(0)}$, $\mathbf{B}_2^{(0)}$, $\boldsymbol{\beta}^{(0)}$) ▷ Initialize $\mathbf{B}_1^{(0)}$, $\mathbf{B}_2^{(0)}$ and $\boldsymbol{\beta}^{(0)}$
  repeat
    $\boldsymbol{\beta}^{(j+1)} = LS(\boldsymbol{\beta}|\boldsymbol{y}, \mathbf{Z}, \mathbf{B}_1^{(j)}, \mathbf{B}_2^{(j)})$
    $\mathbf{B}_1^{(j+1)} = LS(\mathbf{B}_1|\boldsymbol{y}, \{\mathbf{Z}, \mathcal{A}(\mathbf{B}_2^{(j)})\}, \boldsymbol{\beta}^{(j+1)})$
    $\mathbf{B}_2^{(j+1)} = LS(\mathbf{B}_2|\boldsymbol{y}, \{\mathbf{Z}, \mathcal{B}(\mathbf{B}_1^{(j+1)})\}, \boldsymbol{\beta}^{(j+1)})$
  until convergence
end procedure

In Algorithm 1, $LS(\mathbf{B}_1 \mid \boldsymbol{y}, \mathbf{X}, \mathbf{A}, \ldots)$ indicates a least square problem with respect to $\mathbf{B}_1$ given a response vector $\boldsymbol{y}$ and a design matrix $\mathbf{X}$, conditioning on $\mathbf{A}$ or other variables. The estimation procedure of the tensor regression coefficients is a series of least square problems.

S5. Additional Simulations based on "Gene De-correlation" Data Tensor.

To make sure the results are robust to correlations among omics variables, we repeat the simulation analyses of Section 3.1 but using omics data tensors that remove the correlation among genes. We generate the "*gene de-correlation*" omics data tensors, denoted as $\mathcal{X}^{**}$ with dimension $(P, G, n) = (3, 74, 530)$, by resampling the original data tensor $\mathcal{X}^{*}$ across subjects for each gene as follows. Rewrite the ith slice of $\mathcal{X}^{**}$ as $\mathbf{X}_i^{**} \equiv [\boldsymbol{x}_{\cdot 1i}^{**}, \cdots, \boldsymbol{x}_{\cdot Gi}^{**}]$, where $\boldsymbol{x}_{\cdot gi}^{**}$ is the $P \times 1$ (and $P = 3$ here) multi-platform design vector of gene g for individual i. Similarly, the ith slice of $\mathcal{X}^{*}$ is $\mathbf{X}_i^{*} \equiv [\boldsymbol{x}_{\cdot 1i}^{*}, \cdots, \boldsymbol{x}_{\cdot Gi}^{*}]$, $g = 1, \cdots, 74$, $i = 1, \cdots, 530$. For subject i, we generate the multi-platform data of gene g by randomly selecting a vector from $\boldsymbol{x}_{\cdot g,1}^{*}, \cdots, \boldsymbol{x}_{\cdot g,530}^{*}$, and repeat this process for $g = 1, \cdots, G = 74$ to get $\mathbf{X}_i^{**}$. This sampling process is repeated n times with replacement to obtain $\mathbf{X}_1^{**}, \cdots, \mathbf{X}_{530}^{**}$. Finally, given an individual's omics data $\mathbf{X}_i^{**}$, we follow the same procedure and settings to simulate the outcome value y_i from the model $y_i = \boldsymbol{z}_i^{\top}\boldsymbol{\beta} + \langle \mathbf{X}_i^{**}, \mathbf{B} \rangle + \epsilon_i$. Results are summarized in Supplementary Figure S2 and Supplementary Table S2 (A)-(C); the results show similar findings as reported in Section 3.1 using correlated data tensor.



Supplementary Figure S2. Quantile-quantile plots of the null p-values of the tensor association test, obtained from TR.true (TR model evaluated at true rank), TR.AIC (TR model evaluated at AIC-selected rank) and TR.BIC (TR model evaluated at BIC-selected rank) under different $\mathbf{B}$ shapes. For a given $\mathbf{B}$ shape, the null p-values are obtained from those omics variables with $B_{pg} = 0$ when causal omics variables have effect strength $d = 0.125$, 0.25 and 1. The x axis represents the expected p-values and the y axis represents the observed p-values. The simulation is conducted using "gene de-correlated" data, i.e., the between-gene correlation are removed by resampling data across subjects, per gene.

Supplementary Table S2(A). Model rank determined using AIC and BIC for tensor regression (TR) model, based on the "gene de-correlated" simulation, i.e., the between-gene correlation are removed by resampling data across subjects, per gene. The table shows the proportion of a certain rank value is selected by AIC or BIC. For a given **B** shape, results of true rank are shown in shaded bold; d indicates the effect strength of causal omics variables.

	TR.AIC Selected Rank			TR.BIC Selected Rank		
B shape = Flat	**1**	2	3	**1**	2	3
$d = 0.125$	**0.990**	0.010	0	**1.000**	0	0
$d = 0.25$	**1.000**	0	0	**1.000**	0	0
$d = 1$	**0.995**	0.005	0	**1.000**	0	0
B shape = I	**1**	2	3	**1**	2	3
$d = 0.125$	**0.990**	0.010	0	**1.000**	0	0
$d = 0.25$	**1.000**	0	0	**1.000**	0	0
$d = 1$	**1.000**	0	0	**1.000**	0	0
B shape = T	1	**2**	3	1	**2**	3
$d = 0.125$	0	**0.940**	0.060	1.000	**0**	0
$d = 0.25$	0	**0.940**	0.060	0	**1.000**	0
$d = 1$	0	**0.900**	0.100	0	**1.000**	0
B shape = Random	1	**2**	3	1	**2**	3
$d = 0.125$	0.165	**0.790**	0.045	1.000	**0**	0
$d = 0.25$	0	**1.000**	0	1.000	**0**	0
$d = 1$	0	**0.950**	0.050	0	**1.000**	0

Supplementary Table S2(B). Performance of selecting causal omics variables under different **B** shapes for different methods, based on the "gene de-correlated" simulation, i.e., the between-gene correlation are removed by resampling data across subjects, per gene. Methods include tensor regression evaluated at true rank (TR.true), at AIC determined rank (TR.AIC), and at BIC determined rank (TR.BIC), as well as linear regression model (LM) and LASSO on vectorized omics variables, based on 200 replications. TPR = true positive rate; FDR = false discovery rate; d indicates the effect strength of causal omics variables. For TR and LM, a variable is selected as important if p-value < 0.05. The best performed methods among TR.AIC, LM and LASSO, judged by F-measures, are shown in shades.

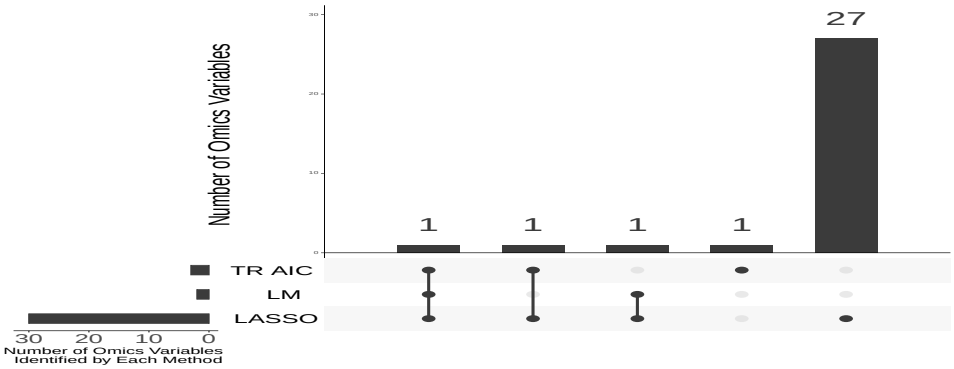
	B shape = Flat					B shape = I					B shape = T					B shape = Random				
	TR. true	TR. AIC	TR. BIC	LM	LASSO	TR. true	TR. AIC	TR. BIC	LM	LASSO	TR. true	TR. AIC	TR. BIC	LM	LASSO	TR. true	TR. AIC	TR. BIC	LM	LASSO
$d = 0.125$																				
TPR	0.667	0.667	0.667	0.552	0.758	0.905	0.903	0.905	0.525	0.781	0.944	0.936	0.982	0.819	0.948	0.847	0.860	0.962	0.691	0.863
FDR	0.046	0.049	0.046	0.144	0.352	0.110	0.111	0.110	0.168	0.358	0.058	0.061	0.440	0.064	0.291	0.029	0.031	0.025	0.035	0.144
F measure	0.780	0.779	0.780	0.669	0.698	0.896	0.895	0.896	0.641	0.704	0.942	0.937	0.713	0.873	0.811	0.905	0.910	0.968	0.805	0.859
$d = 0.25$																				
TPR	0.853	0.853	0.853	0.806	0.899	0.917	0.917	0.917	0.895	0.934	0.982	0.981	0.982	0.972	0.989	0.953	0.953	0.995	0.908	0.967
FDR	0.037	0.038	0.037	0.111	0.374	0.101	0.101	0.101	0.110	0.400	0.055	0.057	0.055	0.053	0.308	0.026	0.026	0.023	0.025	0.172
F measure	0.900	0.900	0.900	0.845	0.738	0.907	0.907	0.907	0.892	0.731	0.963	0.961	0.963	0.959	0.814	0.964	0.964	0.986	0.940	0.892
$d = 1$																				
TPR	0.988	0.988	0.988	0.987	1.000	0.919	0.919	0.919	0.918	0.940	0.982	0.982	0.982	0.982	0.989	0.968	0.968	0.968	0.960	0.984
FDR	0.060	0.062	0.060	0.091	0.406	0.101	0.101	0.101	0.108	0.434	0.054	0.057	0.054	0.055	0.329	0.026	0.026	0.026	0.025	0.194
F measure	0.954	0.953	0.954	0.946	0.745	0.908	0.908	0.908	0.905	0.706	0.963	0.962	0.963	0.963	0.799	0.971	0.971	0.971	0.967	0.886

Supplementary Table S2(C). Performance of selecting causal omics variables of tensor regression evaluated at true rank (TR.true), at AIC selected rank (TR.AIC), and at BIC selected rank (TR.BIC), as well as linear regression model (LM) and lasso on vectorized omics variables, based on 10^5 replications and effect strength $d = 0.25$. For TR and LM, two different selection rules are used: (a) p-value<0.05 and (b) Benjamini-Hochberg false discovery rate (BH-FDR)<0.05. TPR = true positive rate; FDR = false discovery rate. The best performed methods among TR.AIC, LM and LASSO, judged by F measures, are highlighted in shaded cells. The simulation is based on the "gene de-correlated" simulation, i.e., the between-gene correlation are removed by resampling data across subjects, per gene. The results show that using either selection rule, TR.AIC has higher F measures than LM and LASSO for different **B** shapes.

B shape = Flat	TR.true	TR.AIC	TR.BIC	LM	LASSO
p-value < 0.05					
TPR	0.8455	0.8453	0.8455	0.7858	0.8932
FDR	0.0413	0.0435	0.0413	0.1107	0.3803
F measure	0.8933	0.8922	0.8933	0.8334	0.7313
BH-FDR < 0.05					
TPR	0.7910	0.7907	0.7910	0.7054	0.8932
FDR	0.0132	0.0143	0.0132	0.0332	0.3803
F measure	0.8758	0.8752	0.8758	0.8146	0.7313
B shape = I	TR.true	TR.AIC	TR.BIC	LM	LASSO
p-value < 0.05					
TPR	0.9174	0.9173	0.9174	0.8708	0.9343
FDR	0.1051	0.1064	0.1051	0.1109	0.3952
F measure	0.9049	0.9042	0.9049	0.8791	0.7339
BH-FDR < 0.05					
TPR	0.9143	0.9140	0.9143	0.8050	0.9343
FDR	0.0345	0.0354	0.0345	0.0345	0.3952
F measure	0.9388	0.9382	0.9388	0.8771	0.7339
B shape = T	TR.true	TR.AIC	TR.BIC	LM	LASSO
p-value < 0.05					
TPR	0.9816	0.9799	0.9818	0.9618	0.9866
FDR	0.0542	0.0569	0.0548	0.0541	0.3278
F measure	0.9631	0.9609	0.9628	0.9536	0.7994
BH-FDR < 0.05					
TPR	0.9809	0.9780	0.9810	0.9478	0.9866
FDR	0.0272	0.0290	0.0278	0.0263	0.3278
F measure	0.9766	0.9744	0.9764	0.9604	0.7994
B shape = Random	TR.true	TR.AIC	TR.BIC	LM	LASSO
p-value < 0.05					
TPR	0.9498	0.9448	0.9952	0.8859	0.9634
FDR	0.0256	0.0267	0.0272	0.0261	0.1781
F measure	0.9618	0.9587	0.9837	0.9276	0.8869
BH-FDR < 0.05					
TPR	0.9426	0.9367	0.9936	0.8647	0.9634
FDR	0.0171	0.0178	0.0194	0.0162	0.1781
F measure	0.9622	0.9587	0.9869	0.9202	0.8869

S6. Omics Biomarkers for Paclitaxel Using CCLE Data

In this application, we aim at identifying important omics variables that affect the drug sensitivity of paclitaxel, one of the most commonly used chemotherapy for many different types of cancer. We focus on P=2 expression platforms (i.e., mRNA expression and protein expression) and the pan-cancer cell lines (including lung, haematopoietic and lymphoid tissue, skin, breast, ovary, pancreas, central nervous system, large intestine, endometrium and liver cancers). We consider G=55 genes that are from 5 KEGG pathways related to cell cycle and cell death (i.e., cell cycle pathway, apoptosis pathway, p53 signaling pathway, MAPK signaling pathway and PI3K-Akt signaling pathway) and have expression data in the two platforms. From the CCLE website (https://portals.broadinstitute.org/ccle/data), we download the paclitaxel sensitivity data (`CCLE_NP24.2009_Drug_data_2015.02.24.csv`) and the protein expression data (`CCLE_RPPA_20181003.csv`). We download the mRNA expression data from Gene Expression Omnibus (GEO accession GSE36133; `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE36133`). After removing cell lines with missing values, there are n =340 cell lines for further analysis.



Supplementary Figure S3. Number of important omics variables identified by TR.AIC, LM and LASSO for paclitaxel drug sensitivity, displayed by the UpSet plot. For TR.AIC and LM, a variable is selected as important if BH-FDR<0.05.

The TR model of rank 1 has the smallest AIC values. By controlling the FDR level at 0.05 using the Benjamini-Hochberg procedure, TR.AIC identifies 3 variables and LM identifies 2 variables to be important for paclitaxel sensitivity. LASSO again identifies many (i.e., 30) important variables, among which 3 variables overlap with TR.AIC, LM or both, but 2 variables are from housekeeping gene *ACTB*. Supplementary Figure S3 shows the number of important variables identified by different methods, and Supplementary Table S4 lists the coefficient estimates (and p-values if applicable) of different methods. All methods identify *BCL2L1* mRNA as biomarkers for paclitaxel; TR.AIC identifies 2 additional variables, i.e., *BID* mRNA (also identified by LASSO) and *BCL2L1* protein. Both LM and LASSO identify *RB1* protein as a biomarker for paclitaxel, and similar findings has been reported in breast cancer (e.g., Jones et al., 2016; Marangoni et al., 2018), especially the triple-negative breast cancer. Although *RB1* protein expression is not identified by TR.AIC, it has a TR.AIC p-value <0.05 (i.e., 0.0117).

BCL2L1 mRNA is identified as a biomarker for the efficacy of paclitaxel by all methods, and such association has

been widely reported in many cancers, e.g., gastric cancer (Park et al., 2015), colorectal cancer (Sillars-Hardebol et al., 2012), ovarian cancer (Stover et al., 2019), and liver cancer (Chun and Lee, 2004). Although only TR.AIC identifies the protein expression of *BCL2L1* as an important biomarker, several papers demonstrated that *BCL2L1* protein expression is associated with the paclitaxel resistance. For examples, the low protein expression level of *BCL2L1* in the Hep3B cell line led to its high sensitivity to paclitaxel (Chun and Lee, 2004), and this dysregulation of *BCL2L1* is observed in colorectal cancer patients with chromosome 20q gain (Sillars-Hardebol et al., 2012).

TR.AIC also identifies *BID* mRNA expression as important. Similar to *BCL2L1*, mRNA expression of *BID* was inhibited in the cells being treated by paclitaxel (Serbes et al., 2016). A previous study of the treatments of dexamethasone and paclitaxel showed that the expression of pro-apoptotic *BID* was inhibited and thus the chemotherapy-induced apoptosis was inhibited (Pang et al., 2006), suggesting the important role of *BID* in the response of paclitaxel.

References

Chun E, Lee KY. (2004) Bcl-2 and Bcl-xL are important for the induction of paclitaxel resistance in human hepatocellular carcinoma cells. *Biochemical and biophysical research communications*, 315:771-9.

Jones RA, Robinson TJ, Liu JC, Shrestha M, Voisin V, Ju Y, et al. (2016) RB1 deficiency in triple-negative breast cancer induces mitochondrial protein translation. *The Journal of clinical investigation*, 126:3739-57.

Marangoni E, Laurent C, Coussy F, El-Botty R, Chateau-Joubert S, Servely JL, et al. (2018) Capecitabine Efficacy Is Correlated with TYMP and RB1 Expression in PDX Established from Triple-Negative Breast Cancers. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 24:2605-15.

Pang D, Kocherginsky M, Krausz T, Kim SY, Conzen SD. (2016) Dexamethasone decreases xenograft response to Paclitaxel through inhibition of tumor cell apoptosis. *Cancer biology and therapy*, 5:933-40.

Park H, Cho SY, Kim H, Na D, Han JY, Chae J, et al. (2015) Genomic alterations in BCL2L1 and DLC1 contribute to drug sensitivity in gastric cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 112:12492-7.

Serbes U, Ozsoylemez OD, Ozcan G. Evaluation of Paclitaxel Effects in The Pattern of Expression of Survival and Apoptotic Genes Regulators in HeLa Cells. *Current pharmaceutical biotechnology*, 17:1058-67.

Sillars-Hardebol AH, Carvalho B, Belien JA, de Wit M, Delis-van Diemen PM, Tijssen M, et al. (2012) BCL2L1 has a functional role in colorectal cancer and its protein expression is associated with chromosome 20q gain. *The Journal of pathology*, 226:442-50.

Stover EH, Baco MB, Cohen O, Li YY, Christie EL, Bagul M, et al. (2019) Pooled Genomic Screens Identify Anti-apoptotic Genes as Targetable Mediators of Chemotherapy Resistance in Ovarian Cancer. *Molecular cancer research:MCR*, 17:2281-93.

Supplementary Table S4. Effect estimates of omics variables on the drug sensitivity of paclitaxel in the CCLE analysis; parentheses indicate the corresponding p-values. Shaded cells indicate important omics variables affecting paclitaxel sensitivity. For tensor regression (TR.AIC) and linear model (LM), a variable is selected as important at the FDR level of 0.05 using the Benjamini-Hochberg procedure. For LASSO, a variable is selected if it has non-zero coefficient. Omics variables with p-values<0.05 are shown in bold.

	mRNA Expression						Protein Expression					
	TR.AIC		LM		LASSO		TR.AIC		LM		LASSO	
ACTB	0.1071	(0.1469)	0.1974	(0.0760)	0.1086		ACTB	0.0803	(0.1547)	0.0031	(0.9808)	0.0018
ARAF	-0.0790	(0.1682)	-0.1386	(0.1439)	0		ARAF	-0.0592	(0.1865)	0.0585	(0.5274)	0
ATM	0.0655	(0.3226)	-0.1864	(0.1699)	0		ATM	0.0491	(0.3149)	0.0700	(0.3904)	0.0006
BAD	**0.1587**	**(0.0215)**	0.1823	(0.0674)	0		BAD	**0.1190**	**(0.0339)**	-0.0253	(0.7832)	0
BAX	0.0013	(0.9846)	0.0898	(0.3802)	0.0742		BAX	0.0090	(0.9846)	0.1829	(0.1337)	0
BCL2	-0.0459	(0.5336)	-0.1039	(0.5164)	0		BCL2	-0.0344	(0.5372)	-0.1122	(0.3186)	0
BCL2L1	**-0.2302**	**(0.0013)**	**-0.4334**	**(0.0004)**	-0.2742		BCL2L1	**-0.1726**	**(0.0011)**	0.0772	(0.4489)	0
BCL2L11	0.1282	(0.0816)	0.0884	(0.5054)	0		BCL2L11	0.0961	(0.0902)	0.1180	(0.3231)	0.0067
BID	**0.2506**	**(0.0003)**	**0.2457**	**(0.0030)**	0.1411		BID	**0.1879**	**(0.0030)**	0.2240	(0.1811)	0
BIRC2	**-0.1680**	**(0.0269)**	0.0036	(0.9726)	0		BIRC2	**-0.1259**	**(0.0458)**	-0.1447	(0.3855)	-0.011
BRAF	0.0205	(0.7250)	0.0428	(0.6712)	0		BRAF	0.0154	(0.7272)	-0.0065	(0.9419)	0
CASP7	-0.0765	(0.2172)	-0.0092	(0.9206)	0		CASP7	-0.0574	(0.2194)	-0.0379	(0.7381)	0
CASP8	-0.0509	(0.4663)	-0.0690	(0.5294)	0		CASP8	-0.0382	(0.4679)	-0.0405	(0.7743)	0
CCNB1	0.0413	(0.5575)	0.1533	(0.1458)	0		CCNB1	0.0309	(0.5563)	-0.1188	(0.3230)	0
CCND1	-0.0295	(0.6886)	-0.0329	(0.7572)	-0.0987		CCND1	-0.0221	(0.6869)	0.1541	(0.1471)	0.0092
CCNE1	-0.1159	(0.0847)	-0.2113	(0.0681)	0		CCNE1	-0.0869	(0.0880)	-0.0850	(0.5340)	0
CDK1	0.1383	(0.0564)	0.0393	(0.7269)	0		CDK1	0.1037	(0.0644)	0.0260	(0.7871)	0.0491
CDKN1A	0.1133	(0.0870)	0.1125	(0.2945)	0		CDKN1A	0.0849	(0.0851)	0.1735	(0.1301)	0
CDKN1B	-0.1027	(0.0809)	-0.1148	(0.2319)	0		CDKN1B	-0.0770	(0.0913)	-0.1442	(0.1156)	0
CDKN2A	0.0789	(0.3031)	-0.0943	(0.5112)	0		CDKN2A	0.0592	(0.3014)	0.0478	(0.7124)	0
CHEK1	0.0949	(0.1970)	0.1378	(0.275)	0.1282		CHEK1	0.0711	(0.2086)	-0.0959	(0.3598)	0
DIABLO	-0.0128	(0.8468)	-0.0653	(0.5433)	0		DIABLO	-0.0096	(0.8476)	0.0942	(0.4549)	0.0773
EIF4E	-0.0036	(0.9560)	-0.0123	(0.9167)	0		EIF4E	-0.0027	(0.9561)	0.0582	(0.5596)	0
EIF4EBP1	0.0777	(0.3081)	0.1344	(0.2968)	0.0466		EIF4EBP1	0.0582	(0.2878)	0.1370	(0.2350)	0
FOXO3	0.0752	(0.2592)	0.0556	(0.5705)	0		FOXO3	0.0564	(0.2634)	0.0884	(0.4966)	0.0025
GSK3B	-0.0371	(0.5686)	0.0496	(0.6507)	0		GSK3B	-0.0278	(0.5619)	-0.0292	(0.7431)	0
IRS1	-0.1065	(0.1249)	-0.0588	(0.6355)	0		IRS1	-0.0798	(0.1307)	-0.0985	(0.4178)	-0.0025
JUN	0.0060	(0.9340)	0.0042	(0.9707)	-0.027		JUN	0.0045	(0.9339)	0.0276	(0.8110)	0
MAP2K1	0.0930	(0.1413)	-0.1805	(0.1131)	0		MAP2K1	0.0697	(0.1548)	**0.1791**	**(0.0319)**	0
MAPK1	-0.0296	(0.6212)	-0.1276	(0.1201)	0		MAPK1	-0.0222	(0.62)	-0.0035	(0.9680)	0
MAPK14	-0.0036	(0.9564)	0.1361	(0.2641)	0.0417		MAPK14	-0.0027	(0.9563)	-0.1008	(0.3387)	-0.0215
MAPK8	0.0607	(0.3470)	0.0901	(0.2826)	0.0006		MAPK8	0.0455	(0.3534)	0.0680	(0.5269)	0
MDM2	-0.0993	(0.1549)	-0.0507	(0.676)	0		MDM2	-0.0744	(0.1523)	0.1989	(0.0802)	0
MDM4	-0.0576	(0.4820)	-0.0590	(0.6721)	0		MDM4	-0.0432	(0.4839)	-0.0411	(0.7862)	0
MTOR	-0.0070	(0.9087)	0.0424	(0.6659)	0		MTOR	-0.0052	(0.9087)	**-0.2084**	**(0.0304)**	-0.0362
MYC	0.1041	(0.1922)	0.0538	(0.6201)	0.0613		MYC	0.0780	(0.2028)	0.1501	(0.2912)	0.0086
NFKB1	0.0519	(0.4388)	0.1333	(0.1957)	0.0705		NFKB1	0.0389	(0.4298)	0.0794	(0.3885)	0
PCNA	0.1407	(0.0551)	0.1655	(0.1711)	0.0184		PCNA	**0.1054**	**(0.0466)**	**0.1995**	**(0.0365)**	0
PIK3CA	-0.0395	(0.5427)	-0.2081	(0.0503)	0.0153		PIK3CA	-0.0296	(0.5557)	-0.2118	(0.0826)	0
PRKAA1	0.1125	(0.0670)	**0.1963**	**(0.0338)**	0		PRKAA1	0.0843	(0.0807)	0.1607	(0.1548)	0
PRKCA	0.0294	(0.6975)	-0.0684	(0.6073)	0		PRKCA	0.0220	(0.6937)	0.1467	(0.3588)	0
PTEN	0.0939	(0.2445)	0.1970	(0.0952)	0		PTEN	0.0704	(0.2495)	-0.0340	(0.8452)	0
RAF1	**0.1638**	**(0.0273)**	0.0493	(0.6851)	0.0092		RAF1	**0.1228**	**(0.0365)**	0.0649	(0.5919)	0.0662
RB1	**0.2039**	**(0.0040)**	0.0626	(0.5219)	0		RB1	**0.1529**	**(0.0117)**	**0.4653**	**(0.0003)**	0.0932
RPS6	0.0126	(0.8610)	0.0782	(0.4426)	0.0454		RPS6	0.0094	(0.8616)	0.1157	(0.3858)	0
RPS6KB1	-0.0230	(0.7380)	-0.2350	(0.1015)	0		RPS6KB1	-0.0172	(0.7386)	0.0388	(0.7121)	0
RPTOR	-0.0414	(0.4528)	-0.1813	(0.0541)	0		RPTOR	-0.0310	(0.4498)	0.0558	(0.6729)	0
SERPINE1	-0.0309	(0.7031)	0.0523	(0.7598)	0		SERPINE1	-0.0232	(0.7030)	-0.0126	(0.9090)	0
SMAD4	0.0807	(0.2420)	0.0206	(0.835)	0		SMAD4	0.0605	(0.2533)	0.2266	(0.0586)	0
STMN1	-0.0078	(0.9181)	0.0084	(0.9448)	0		STMN1	-0.0058	(0.9181)	0.1349	(0.3328)	0
SYK	-0.1248	(0.0991)	0.2163	(0.304)	0		SYK	-0.0935	(0.0908)	-0.2226	(0.0647)	0
TP53	**-0.1376**	**(0.0498)**	**-0.2581**	**(0.0221)**	0		TP53	-0.1032	(0.0909)	-0.0287	(0.8544)	0
TSC1	-0.0047	(0.9456)	0.0897	(0.3802)	0		TSC1	-0.0035	(0.9457)	-0.2718	(0.1739)	0
TSC2	0.0945	(0.1386)	0.1207	(0.2136)	0		TSC2	0.0708	(0.1420)	-0.0629	(0.5284)	0
TUBA1B	-0.1061	(0.1551)	-0.0711	(0.5445)	0		TUBA1B	-0.0795	(0.1443)	-0.1881	(0.1355)	0