Supporting Information File 1 for "Integrative Analysis of Genetical Genomics Data Incorporating Network Structures" by

Bin Gao[1,2†], Xu Liu[3†], Hongzhe Li[4], and Yuehua Cui[1*]

[1] *Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824*

[2] *Quantitative Sciences, Janssen Research & Development, LLC, Spring House, PA 19477*

[3] *School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China*

[4] *Department of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine, PA 19014*

[†] The first two authors contributed equally to this work.

[*] *e-mail*: cuiy@msu.edu (corresponding author)

This supporting information file contains 3 sections. The first section gives the proof of the theorem. The second section provides additional simulation results and the third section provides a list of genes in KEGG "Metabolism of Xenobiotics by Cytochrome P450" pathway.

1 Proofs

The proof of the theorem is given in this section. We closely follow the proof in Lin et al. (2015) to establish the theory. First, we prove Condition (C2) holds for the sample matrix of $\hat{\boldsymbol{X}}$ with a smaller α.

Lemma 1. *If the tuning parameters λ_j in the first step are selected to satisfy $\frac{16\phi}{\kappa^2} rs\lambda_{max}(2M_1 + \lambda_{max}) \leq \frac{\alpha}{2(4-\alpha)}$ and Condition (C2) holds, then with probability at least $1 - \sum_{j=1}^{p} q\exp(\frac{-n\lambda_j^2}{8\sigma_j^2})$,* *the matrix* $\hat{\boldsymbol{C}} = \frac{1}{n}(\hat{\boldsymbol{X}})^T\hat{\boldsymbol{X}} = \frac{1}{n}(\boldsymbol{G}\hat{\boldsymbol{\Gamma}})^T\boldsymbol{G}\hat{\boldsymbol{\Gamma}}$ *satisfies*

$$\|(\hat{\boldsymbol{C}}_{SS} + 2\mu_2\boldsymbol{L}_{SS})^{-1}\|_\infty \leq \frac{2(4-\alpha)}{8-3\alpha}\phi \tag{S1}$$

$$\|(\hat{\boldsymbol{C}}_{S^cS} + 2\mu_2\boldsymbol{L}_{S^cS})(\hat{\boldsymbol{C}}_{SS} + 2\mu_2\boldsymbol{L}_{SS})^{-1}\|_\infty \leq 1 - \frac{3}{4}\alpha. \tag{S2}$$

Proof. By applying (A.1) and (A.2) in the proof of theorem 1 in Lin et al. (2015), similar to their derivation, we have $\phi\|\hat{\mathbf{C}}_{SS} - \mathbf{C}_{SS}\|_\infty \leq \frac{\alpha}{2(4-\alpha)}$ and $\phi\|\hat{\mathbf{C}}_{S^cS} - \mathbf{C}_{S^cS}\|_\infty \leq \frac{\alpha}{2(4-\alpha)}$. Then, we have

$$\begin{aligned}
&\|(\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1} - (\mathbf{C}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty \\
\leq&\frac{\phi\|\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS} - \mathbf{C}_{SS} + 2\mu_2\mathbf{L}_{SS}\|_\infty}{1-\phi\|\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS} - \mathbf{C}_{SS} + 2\mu_2\mathbf{L}_{SS}\|_\infty}\phi \\
=&\frac{\phi\|\hat{\mathbf{C}}_{SS} - \mathbf{C}_{SS}\|_\infty}{1-\phi\|\hat{\mathbf{C}}_{SS} - \mathbf{C}_{SS}\|_\infty}\phi \\
\leq&\frac{\alpha}{8-3\alpha}\phi.
\end{aligned}$$

The triangle inequality implies

$$\begin{aligned}
&\|(\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty \\
\leq&\|(\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1} - (\mathbf{C}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty + \|(\mathbf{C}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty \\
\leq&\frac{\alpha}{8-3\alpha}\phi + \phi \\
=&\frac{2(4-\alpha)}{8-3\alpha}\phi.
\end{aligned}$$

Then we have

$$
\begin{aligned}
&\|(\hat{\mathbf{C}}_{S^cS} + 2\mu_2\mathbf{L}_{S^cS})(\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1} - (\mathbf{C}_{S^cS} + 2\mu_2\mathbf{L}_{S^cS}) * (\mathbf{C}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty \\
=&\|(\hat{\mathbf{C}}_{S^cS} - \mathbf{C}_{S^cS})(\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1} - (\mathbf{C}_{S^cS} + 2\mu_2\mathbf{L}_{S^cS}) \\
&* (\mathbf{C}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}(\hat{\mathbf{C}}_{SS} - \mathbf{C}_{SS})(\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty \\
\leq&\|(\hat{\mathbf{C}}_{S^cS} - \mathbf{C}_{S^cS})\|_\infty\|(\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty + \|(\mathbf{C}_{S^cS} + 2\mu_2\mathbf{L}_{S^cS})(\mathbf{C}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty \\
&* \|(\hat{\mathbf{C}}_{SS} - \mathbf{C}_{SS})\|_\infty * \|(\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty \\
\leq&\frac{\alpha}{2(4-\alpha)\phi}\frac{2(4-\alpha)}{8-3\alpha}\phi + (1-\alpha)\frac{\alpha}{2(4-\alpha)\phi}\frac{2(4-\alpha)}{8-3\alpha}\phi = \frac{2-\alpha}{8-3\alpha}\alpha \\
\leq&\frac{1}{4}\alpha.
\end{aligned}
$$

Finally we have

$$
\begin{aligned}
&\|(\hat{\mathbf{C}}_{S^cS} + 2\mu_2\mathbf{L}_{S^cS})(\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty \\
\leq&\|(\hat{\mathbf{C}}_{S^cS} + 2\mu_2\mathbf{L}_{S^cS})(\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1} - (\mathbf{C}_{S^cS} + 2\mu_2\mathbf{L}_{S^cS}) \\
&* (\mathbf{C}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty + \|(\mathbf{C}_{S^cS} + 2\mu_2\mathbf{L}_{S^cS})(\mathbf{C}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty \\
\leq&\frac{1}{4}\alpha + 1 - \alpha \\
=&1 - \frac{3}{4}\alpha.
\end{aligned}
$$

□

Proof of the Theorem. Here we closely follow the proof in Lin et al. (2015). For an index set I, define $\boldsymbol{X}_I$ as the submatrix consisting of the jth columns of $\boldsymbol{X}$, where $j \in I$. By Karush-Kuhn-Tucker conditions, the solution $\hat{\boldsymbol{\beta}}$ of Equation (2.4) in the main content must satisfies

$$\frac{1}{n}\hat{\boldsymbol{X}}_{\hat{S}}^T(\boldsymbol{Y} - \hat{\boldsymbol{X}}\hat{\boldsymbol{\beta}}) - 2\mu_2\mathbf{L}_{\hat{S}}^T\hat{\boldsymbol{\beta}} = \mu_1 \operatorname{sign}(\hat{\boldsymbol{\beta}}_{\hat{S}}) \tag{S3}$$

$$\|\frac{1}{n}\hat{\boldsymbol{X}}_{\hat{S}^c}^T(\boldsymbol{Y} - \hat{\boldsymbol{X}}\hat{\boldsymbol{\beta}}) - 2\mu_2\mathbf{L}_{\hat{S}^c}^T\hat{\boldsymbol{\beta}}\|_\infty \leq \mu_1. \tag{S4}$$

Let $\hat{\boldsymbol{\beta}}_{S^c} = 0$. We first find $\hat{\boldsymbol{\beta}}_S$ from (S3), then we prove such $\hat{\boldsymbol{\beta}}$ also satisfies (S4). Besides, we prove such $\hat{\boldsymbol{\beta}}$ possesses the property of consistency.

Using the similar argument in Lin et al. (2015), we can find constants $c_0, c_1, c_2 > 0$ such that, if we select μ_1 as in the Theorem, then with probability at least $1 - c_1(pq)^{-c_2}$, we have

$$\|\frac{1}{n}\hat{\boldsymbol{X}}^T\boldsymbol{\eta} - \frac{1}{n}\hat{\boldsymbol{X}}^T(\hat{\boldsymbol{X}} - \boldsymbol{X})\boldsymbol{\beta})\|_\infty \leq \frac{\alpha}{2(4-\alpha)}\mu_1. \tag{S5}$$

From now on, we base our analysis on (S5). By using the equality $\boldsymbol{Y} = \boldsymbol{X}_S\boldsymbol{\beta}_S + \boldsymbol{\eta}$, we have $\boldsymbol{Y} - \hat{\boldsymbol{X}}\hat{\boldsymbol{\beta}} = \boldsymbol{\eta} - (\hat{\boldsymbol{X}}_S - \boldsymbol{X}_S)\boldsymbol{\beta}_S - \hat{\boldsymbol{X}}_S(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S)$. Replacing $\hat{S}$ with S, we write (S3) as

$$\hat{\mathbf{C}}_{SS}(\hat{\beta}_S - \beta_S) + 2\mu_2\mathbf{L}_{SS}\hat{\beta}_S + 2\mu_2\mathbf{L}_{S^cS}^T\hat{\beta}_{S^c} = \frac{1}{n}\hat{\mathbf{X}}_S^T\eta - \frac{1}{n}\hat{\mathbf{X}}_S^T(\hat{\mathbf{X}}_S - \mathbf{X}_S)\beta_S - \mu_1\,\mathrm{sign}(\hat{\beta}_S).$$

After some algebra, we have

$$\hat{\beta}_S - \beta_S = (\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\Big[\frac{1}{n}\hat{\mathbf{X}}_S^T\eta - \frac{1}{n}\hat{\mathbf{X}}_S^T(\hat{\mathbf{X}}_S - \mathbf{X}_S)\beta_S - \mu_1\,\mathrm{sign}(\hat{\beta}_S) - 2\mu_2\mathbf{L}_{SS}\beta_S\Big] \tag{S6}$$

By Lemma 1 and (S5), we have

$$\begin{aligned}
\|\hat{\beta}_S - \beta_S\|_\infty &\leq \|(\hat{\mathbf{C}}_{SS} + 2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty\Big[\|\frac{1}{n}\hat{\mathbf{X}}_S^T\eta - \frac{1}{n}\hat{\mathbf{X}}_S^T(\hat{\mathbf{X}}_S - \mathbf{X}_S)\beta_S\|_\infty \\
&\qquad + \|\mu_1\,\mathrm{sign}(\hat{\beta}_S)\|_\infty + \|2\mu_2\mathbf{L}_{SS}\beta_S\|_\infty\Big] \\
&\leq \frac{2(4-\alpha)}{8-3\alpha}\phi\Big[\frac{\alpha}{2(4-\alpha)}\mu_1 + \mu_1 + 2\mu_2C_L\Big] \\
&= \frac{2(4-\alpha)}{8-3\alpha}\phi\Big[\frac{8-\alpha}{2(4-\alpha)}\mu_1 + 2\mu_2C_L\Big] \\
&< b_0
\end{aligned}$$

which implies $\mathrm{sign}(\hat{\boldsymbol{\beta}}_S) = \mathrm{sign}(\boldsymbol{\beta}_S)$. Besides, we know $\hat{\beta}_{S^c} = 0$ by definition. Hence, we have $\hat{S} = S$. Let $\hat{\beta}_S$ be defined by (A6) with $\mathrm{sign}(\hat{\beta}_S)$ replaced by $\mathrm{sign}(\beta_S)$. Now, we need to check if (S4) holds. By using the equality $\boldsymbol{Y} - \hat{\boldsymbol{X}}\hat{\boldsymbol{\beta}} = \boldsymbol{\eta} - (\hat{\boldsymbol{X}}_S - \boldsymbol{X}_S)\boldsymbol{\beta}_S - \hat{\boldsymbol{X}}_S(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S)$

and (S6), we have

$$\begin{aligned}
&\frac{1}{n}\hat{\mathbf{X}}_{S^c}^T(Y-\hat{\mathbf{X}}\hat{\beta})-2\mu_2\mathbf{L}_{S^c}^T\hat{\beta}\\
=&\frac{1}{n}\hat{\mathbf{X}}_{S^c}^T\eta-\frac{1}{n}\hat{\mathbf{X}}_{S^c}^T(\hat{\mathbf{X}}_S-\mathbf{X}_S)\beta_S-(\hat{\mathbf{C}}_{S^cS}+2\mu_2\mathbf{L}_{S^cS})(\hat{\mathbf{C}}_{SS}+2\mu_2\mathbf{L}_{SS})^{-1}\\
&*\Big[\frac{1}{n}\hat{\mathbf{X}}_{S^c}^T\eta-\frac{1}{n}\hat{\mathbf{X}}_{S^c}^T(\hat{\mathbf{X}}_S-\mathbf{X}_S)\beta_S-\mu_1\operatorname{sign}(\hat{\beta}_S)-2\mu_2\mathbf{L}_{SS}\beta_S\Big]-2\mu_2\mathbf{L}_{S^cS}\beta_S.
\end{aligned}$$

By Lemma 1, (S5), and $\mu_2 C_L\le\frac{\alpha(16-3\alpha)}{4(4-\alpha)(8-3\alpha)}\mu_1$, we have

$$\begin{aligned}
&\|\frac{1}{n}\hat{\mathbf{X}}_{S^C}^T(Y-\hat{\mathbf{X}}\hat{\beta})-2\mu_2\mathbf{L}_{S^C}^T\hat{\beta}\|_\infty\\
\le&\|\frac{1}{n}\hat{\mathbf{X}}_{S^C}^T\eta-\frac{1}{n}\hat{\mathbf{X}}_{S^C}^T(\hat{\mathbf{X}}_S-\mathbf{X}_S)\beta_S\|_\infty+\|(\hat{\mathbf{C}}_{S^cS}+2\mu_2\mathbf{L}_{S^cS})*(\hat{\mathbf{C}}_{SS}+2\mu_2\mathbf{L}_{SS})^{-1}\|_\infty\\
&*\Big[\|\frac{1}{n}\hat{\mathbf{X}}_{S^C}^T\eta-\frac{1}{n}\hat{\mathbf{X}}_{S^C}^T(\hat{\mathbf{X}}_S-\mathbf{X}_S)\beta_S\|_\infty+\|\mu_1\operatorname{sign}(\hat{\beta}_S)\|_\infty+\|2\mu_2\mathbf{L}_{SS}\beta_S\|_\infty\Big]+\|2\mu_2\mathbf{L}_{S^CS}\beta_S\|_\infty\\
\le&\frac{\alpha}{2(4-\alpha)}\mu_1+(1-\frac{3}{4}\alpha)\Big[\frac{\alpha}{2(4-\alpha)}\mu_1+\mu_1+2\mu_2C_L\Big]+2\mu_2C_L\\
=&\frac{3\alpha^2-24\alpha+32}{8(4-\alpha)}\mu_1+\frac{8-3\alpha}{2}\mu_2C_L\\
\le&\mu_1.
\end{aligned}$$

Since $\hat{S}=S$, we see $\hat{\beta}$ also satisfies (A4). Lastly, we have

$$\begin{aligned}
\|\hat{\beta}_S-\beta_S\|_\infty&\le\frac{2(4-\alpha)}{8-3\alpha}\phi\Big[\frac{8-\alpha}{2(4-\alpha)}\mu_1+2\mu_2C_L\Big]\\
&\le\frac{2(4-\alpha)}{8-3\alpha}\phi\Big[\frac{8-\alpha}{2(4-\alpha)}\mu_1+2\frac{\alpha(16-3\alpha)}{4(4-\alpha)(8-3\alpha)}\mu_1\Big]\\
&=\frac{16(4-\alpha)\phi C_0}{(8-3\alpha)^2\kappa}\sqrt{\frac{r(\log p+\log q)}{n}}.
\end{aligned}$$

□

This completes the proof of the theorem.

2 Additional simulations

2.1 Comparison of IVGC with IV when the signal to noise ratio is reduced

We follow the same model setup reported in Table 3 of the manuscript, while reducing the effect size of the β coefficients to check the selection performance of the method. Here we reported the case with $p = 600$, $q = 600$ and $n = 300$ to show the impact of reducing the size of the regression coefficients. Similar performance was observed for other combination of p, q and n and hence are omitted. The nonzero $\boldsymbol{\beta}$ coefficients were replaced by $\beta_1 = \cdots = \beta_5 = 0.2$, $\beta_6 = \cdots = \beta_{10} = 0.5$, and $\beta_k \sim U(0.2, 0.5)$, $k = 1, \cdots, 10$. The results are reported in Table S1. It can be seen that the true positive (TP) and MCC become slightly lower, the false positive (FP) becomes slightly higher compared to the results in Table 3 in the main context when signal becomes weaker. This implies that the variable selection performance can be affected by weak regression signals. On the other hand, our proposed method still performs better than the IV method does, indicating the relative gain by incorporating network information.

2.2 Comparison of IVGC with 1-stage LASSO

In this scenario, we compared the performance of IVGC with a one-stage LASSO method without considering instrumental variables and network information. In Lin et al. (2015), the authors have shown the advantage of IV regression against the method without instrumental variables. Here reported the results under the scenario of $n = 300$, $p = 600$, and $q = 600$. Similar results were observed for other settings by varying n, p and q, hence are omitted. Table S2 summarizes the results. It is shown that IVGC actually has slightly higher estimation loss and model error compared to the one-stage LASSO method. However, IVGC has much smaller false positive rates and higher MCC values compared to the LASSO method, under different network conditions, indicating the relative gain of the proposed met-

Table S1: Comparison of IVGC with IV when the signal to noise ratio is reduced for $p = 600$, $q = 600$ and $n = 300$. The numbers in the parentheses are the empirical standard errors.

Method	numSNP	Estimation Loss	Model Error	True Positive	False Positive	MCC
			$\beta_1 = \cdots = \beta_5 = 0.2, \beta_6 = \cdots = \beta_{10} = 0.5$			
	3	1.50 (0.72)	0.66 (0.19)	9.66 (0.73)	11.18 (8.66)	0.70 (0.13)
	4	1.53 (0.77)	0.74 (0.22)	9.41 (0.97)	8.60 (7.71)	0.73 (0.14)
	5	2.44 (1.39)	1.06 (0.41)	8.20 (1.84)	7.87 (7.98)	0.68 (0.18)
IVGC			$\beta_k \sim U(0.2, 0.5), k = 1, \cdots, 10$			
	3	1.63 (0.73)	0.68 (0.19)	9.91 (0.37)	11.12 (8.70)	0.71 (0.13)
	4	1.74 (0.77)	0.74 (0.24)	9.59 (0.82)	8.64 (7.57)	0.74 (0.14)
	5	2.51 (1.30)	0.99 (0.38)	8.41 (1.74)	7.94 (7.96)	0.69 (0.16)
			$\beta_1 = \cdots = \beta_5 = 0.2, \beta_6 = \cdots = \beta_{10} = 0.5$			
	3	2.36 (0.91)	0.89 (0.20)	8.80 (0.88)	12.77 (10.93)	0.64 (0.14)
	4	2.68 (0.91)	1.03 (0.21)	8.40 (0.97)	11.53 (10.97)	0.64 (0.15)
	5	3.73 (1.07)	1.39 (0.28)	6.74 (1.08)	10.08 (10.94)	0.57 (0.16)
IV			$\beta_k \sim U(0.2, 0.5), k = 1, \cdots, 10$			
	3	2.63 (0.97)	0.97 (0.21)	9.59 (0.61)	14.15 (10.98)	0.66 (0.14)
	4	2.68 (0.81)	1.02 (0.21)	8.99 (0.85)	9.81 (8.83)	0.69 (0.14)
	5	3.73 (0.91)	1.33 (0.25)	7.18 (1.11)	10.13 (9.89)	0.59 (0.15)

hod against a nave LASSO method without considering instrumental variables and network information.

2.3 Comparison of IVGC with IV by mimicking real situations

To mimic the real situation, we did a real data guided simulation. We picked 10,000 SNPs from the real data located consecutively on chromosome 1. By using the real data SNPs, the nature of the linkage disequilibrium (LD) structure is well preserved. We then randomly sample individuals by treating the original data as the pseudo population. Then we followed the steps as stated in the original manuscript for the follow up analysis. For this simulation, we considered the scenario with $n = 600$, $p = 100$ and $q = 10,000$. The results are summarized in Table S3. Again, the results show smaller estimation loss and model error of the IVGC method compared to the IV method without network constraint, although the TP, TP and MCC are quite similar between the two methods.

Table S2: Comparison of IVGC with 1-stage LASSO for $p = 600$, $q = 600$ and $n = 300$. The numbers in the parentheses are the empirical standard errors.

Method	numSNP	Estimation Loss	Model Error	True Positive	False Positive	MCC
			$\beta_k = 0.5, k = 1, \cdots, 10$			
IVGC	3	1.71 (0.93)	0.74 (0.25)	9.99 (0.1)	11.46 (9.04)	0.71 (0.13)
	4	1.78 (0.95)	0.86 (0.28)	9.89 (0.43)	8.47 (8.09)	0.76 (0.14)
	5	2.9 (1.64)	1.23 (0.47)	9.15 (1.37)	8.16 (8.62)	0.73 (0.15)
			$\beta_k \sim U(0.5, 1), k = 1, \cdots, 10$			
	3	2.8 (1.20)	1.20 (0.33)	9.96 (0.18)	11.49 (9.09)	0.71 (0.13)
	4	2.75 (1.16)	1.40 (0.34)	9.96 (0.27)	8.75 (8.12)	0.76 (0.13)
	5	3.92 (2.06)	1.83 (0.59)	9.30 (1.26)	6.86 (6.12)	0.76 (0.14)
			$\beta_k = 0.5, k = 1, \cdots, 10$			
1-stage LASSO	3	1.11 (0.48)	0.32 (0.08)	10 (0)	37.32 (20.84)	0.48 (0.11)
	4	0.97 (0.40)	0.29 (0.07)	10 (0)	29.73 (17.40)	0.52 (0.11)
	5	0.91 (0.36)	0.28 (0.06)	10 (0)	25.85 (15.58)	0.55 (0.12)
			$\beta_k \sim U(0.5, 1), k = 1, \cdots, 10$			
	3	1.02 (0.44)	0.31 (0.07)	10 (0)	33.01 (19.01)	0.50 (0.12)
	4	1.01 (0.46)	0.30 (0.07)	10 (0)	32.12 (18.80)	0.51 (0.11)
	5	0.93 (0.43)	0.28 (0.07)	10 (0)	27.00 (17.97)	0.55 (0.12)

Table S3: Comparison of IVGC with IV by mimicking real situations for $p = 100$, $n = 300$ and $q = 10,000$. The numbers in the parentheses are the empirical standard errors.

Method	numSNP	Estimation Loss	Model Error	True Positive	False Positive	MCC
			$\beta_k = 0.5, k = 1, \cdots, 10$			
IVGC	3	0.77 (0.17)	0.49 (0.10)	10 (0)	3.00 (2.73)	0.88 (0.09)
	4	0.92 (0.19)	0.69 (0.10)	10 (0)	2.33 (2.73)	0.90 (0.09)
	5	1.11 (0.52)	0.73 (0.16)	9.97 (0.22)	1.75 (2.07)	0.92 (0.08)
			$\beta_k \sim U(0.5, 1), k = 1, \cdots, 10$			
	3	1.46 (0.25)	0.88 (0.12)	10 (0)	2.67 (2.46)	0.89 (0.09)
	4	1.70 (0.28)	1.06 (0.13)	10 (0)	2.19 (2.09)	0.90 (0.08)
	5	1.86 (0.65)	1.16 (0.22)	9.95 (0.35)	2.04 (2.04)	0.91 (0.08)
			$\beta_k = 0.5, k = 1, \cdots, 10$			
IV	3	1.18 (0.26)	0.64 (0.11)	10 (0)	3.18 (2.45)	0.87 (0.08)
	4	1.44 (0.3)	0.83 (0.12)	10 (0)	2.64 (2.66)	0.89 (0.09)
	5	3.05 (0.7)	1.29 (0.23)	9.16 (0.76)	1.88 (2.24)	0.87 (0.10)
			$\beta_k \sim U(0.5, 1), k = 1, \cdots, 10$			
	3	1.65 (0.36)	1.03 (0.16)	10 (0)	3.19 (2.77)	0.87 (0.09)
	4	1.95 (0.42)	1.13 (0.15)	9.99 (0.07)	2.79 (2.15)	0.88 (0.08)
	5	4.34 (0.92)	1.88 (0.28)	9.13 (0.82)	2.37 (2.43)	0.85 (0.10)

2.4 Comparison of IVGC with one-stage GC without considering instrumental variables

To check the model misspecification of ignoring instrumental variables, we simulated data considering instrumental variables, then analyzed the simulated data using IVGC and a one-stage variable selection method imposing a graph constrained penalty ignoring instrumental variables (denoted by 1-stage GC). We reported results under the case of $p = 600$, $q = 600$ and $n = 300$ in Table S4. Although the estimation loss and model error were slightly larger for the IVGC method than the one-stage GC does, the one stage GC method has substantially larger false positive and lower MCC values. When false positive rate is a great concern, IVGC method should be preferred in practice.

Table S4: Comparison of IVGC with 1-stage GC for $p = 600$, $q = 600$ and $n = 300$. The numbers in the parentheses are the empirical standard errors.

Method	numSNP	Estimation Loss	Model Error	True Positive	False Positive	MCC
		$\beta_k = 0.5, k = 1, \cdots, 10$				
IVGC	3	1.71 (0.93)	0.74 (0.25)	9.99 (0.1)	11.46 (9.04)	0.71 (0.13)
	4	1.78 (0.95)	0.86 (0.28)	9.89 (0.43)	8.47 (8.09)	0.76 (0.14)
	5	2.9 (1.64)	1.23 (0.47)	9.15 (1.37)	8.16 (8.62)	0.73 (0.15)
		$\beta_k \sim U(0.5, 1), k = 1, \cdots, 10$				
	3	2.8 (1.20)	1.20 (0.33)	9.96 (0.18)	11.49 (9.09)	0.71 (0.13)
	4	2.75 (1.16)	1.40 (0.34)	9.96 (0.27)	8.75 (8.12)	0.76 (0.13)
	5	3.92 (2.06)	1.83 (0.59)	9.30 (1.26)	6.86 (6.12)	0.76 (0.14)
		$\beta_k = 0.5, k = 1, \cdots, 10$				
1-stage GC	3	0.86 (0.39)	0.26 (0.07)	10 (0)	34.33 (16.76)	0.49 (0.10)
	4	0.68 (0.35)	0.23 (0.07)	10 (0)	26.8 (15.84)	0.54 (0.12)
	5	0.62 (0.34)	0.21 (0.07)	10 (0)	24.55 (15.51)	0.56 (0.12)
		$\beta_k \sim U(0.5, 1), k = 1, \cdots, 10$				
	3	0.94 (0.41)	0.29 (0.07)	10 (0)	29.57 (17.01)	0.52 (0.11)
	4	0.84 (0.32)	0.27 (0.06)	10 (0)	25.59 (13.97)	0.55 (0.11)
	5	0.83 (0.31)	0.27 (0.06)	10 (0)	22.5 (13.64)	0.58 (0.11)

2.5 The impact on false positive control by imposing a network structure on null genes

In practice, one has no idea of whether a network of genes has any effect on the response Y. If a group of genes have no effect on Y, will imposing a network structure increase the false positive selection (FPS)? In our simulation setup, we have one group (the 3rd group) with a network structure, but no effect at all. Here we reported the false positive of this group. As a comparison, we randomly picked variable 16-20 (no network structure) and reported the false positives on these 5 genes (as group 4). The results are summarized in Table S5. We reported the case with $n = 300$, $p = 600$ and $q = 600$ and with β $U(0.5, 1)$ for the first 10 variables. For the IV method, there is very little difference on FPS between the two groups. For the IVGC method, we can see that imposing a network structure actually increases the FPS a little bit, although the difference is not striking. Also the standard errors of FPS for group 3 is a little larger than group 4. This actually fits to our intuition, and also raises our attention in real applications: one should always be careful to borrow network information in gene selection. Applying wrong network information can be harmful than imposing no network structure at all.

Table S5: The impact on false positive control by imposing a network structure on null genes. The numbers in the parentheses are the empirical standard errors.

Method	numSNP	FP of group 3	FP of group 4
IVGC	3	0.14 (0.51)	0.10 (0.29)
	4	0.06 (0.28)	0.05 (0.22)
	5	0.05 (0.26)	0.05 (0.22)
IV	3	0.12 (0.37)	0.12 (0.34)
	4	0.08 (0.27)	0.07 (0.29)
	5	0.06 (0.21)	0.08 (0.30)

2.6 Simulation results with high correlation between the X variables

In this section, we reported the results by assuming a high correlation between the multivariate X variables. The estimation procedure is the same as described in the main content. We assumed $cov(X) = \Sigma$, where $\Sigma_{ij} = \rho^{|i-j|}$. We reported the results for $\rho = 0.8$. The purpose of the simulation is to check if the first stage LASSO estimation has an impact on the second stage gene selection if there are high correlations between the X variables. Table S6 lists the result. Compared to Table 2 in the main content, we did not see much difference between the two tables. Table 2 is for the case with $\rho = 0.2$. This results show that it is generally safe to apply the LASSO algorithm at the first stage without considering the correlation information between the X variables. Note that after regressing each X variable with the G variable, the correlation between the fitted values is mainly determined by the number of SNPs they share in common. The original correlation structure has little impact on the correlation of the fitted value.

Table S6: Simulation results with $\rho = 0.8$ for $p = 100$, $q = 100$, $n = 200 \sim 1400$ and numSNP=4. The numbers in the parentheses are the empirical standard errors.

n	Method	Estimation Loss	Model Error	True Positive	False Positive	MCC
200	IVGC	1.33 (0.76)	0.74 (0.28)	9.93 (0.44)	4.35 (3.39)	0.83 (0.11)
	IV	3.32 (0.92)	1.33 (0.31)	9.15 (0.8)	4.2 (3.56)	0.78 (0.12)
400	IVGC	0.81 (0.46)	0.46 (0.18)	9.98 (0.12)	3.77 (3.59)	0.85 (0.11)
	IV	2.51 (0.62)	1 (0.21)	9.68 (0.53)	3.89 (3.32)	0.83 (0.11)
600	IVGC	0.63 (0.35)	0.34 (0.15)	9.99 (0.14)	3.82 (3.21)	0.85 (0.1)
	IV	2.09 (0.5)	0.83 (0.19)	9.8 (0.49)	4.82 (3.73)	0.81 (0.12)
800	IVGC	0.51 (0.26)	0.29 (0.11)	9.95 (0.71)	3.68 (3.15)	0.85 (0.12)
	IV	1.59 (0.44)	0.65 (0.16)	9.93 (0.72)	4 (3.15)	0.84 (0.12)
1000	IVGC	0.46 (0.2)	0.26 (0.09)	10 (0)	3.89 (3.29)	0.85 (0.1)
	IV	2.04 (0.47)	0.76 (0.17)	9.55 (0.5)	4.24 (3.48)	0.81 (0.11)
1200	IVGC	0.39 (0.2)	0.22 (0.08)	10 (0)	3.85 (3.41)	0.85 (0.1)
	IV	1.84 (0.49)	0.72 (0.18)	9.82 (0.41)	4.47 (3.82)	0.82 (0.11)
1400	IVGC	0.41 (0.2)	0.22 (0.08)	9.99 (0.07)	3.68 (3.06)	0.85 (0.1)
	IV	1.65 (0.45)	0.59 (0.14)	9.86 (0.36)	4.33 (3.6)	0.82 (0.11)

3 Gene list in "Metabolism of Xenobiotics by Cytochrome P450" pathway (hsa00980)

Table S7: List of genes in KEGG "Metabolism of Xenobiotics by Cytochrome P450" pathway

CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1 [KO:K07408]
CYP2C9	cytochrome P450, family 2, subfamily C, polypeptide 9 [KO:K17719]
CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4 [KO:K17689]
CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1 [KO:K07410]
GSTA5	glutathione S-transferase alpha 5 [KO:K00799]
GSTA2	glutathione S-transferase alpha 2 [KO:K00799]
GSTA4	glutathione S-transferase alpha 4 [KO:K00799]
GSTO2	glutathione S-transferase omega 2 [KO:K00799]
GSTM4	glutathione S-transferase mu 4 [KO:K00799]
GSTT2	glutathione S-transferase theta 2 (gene/pseudogene) [KO:K00799]
GSTT1	glutathione S-transferase theta 1 [KO:K00799]
GSTM3	glutathione S-transferase mu 3 (brain) [KO:K00799]
MGST1	microsomal glutathione S-transferase 1 [KO:K00799]
MGST3	microsomal glutathione S-transferase 3 [KO:K00799]
GSTP1	glutathione S-transferase pi 1 [KO:K00799]
GSTM1	glutathione S-transferase mu 1 [KO:K00799]
GSTM5	glutathione S-transferase mu 5 [KO:K00799]
MGST2	microsomal glutathione S-transferase 2 [KO:K00799]
GSTA1	glutathione S-transferase alpha 1 [KO:K00799]
GSTM2	glutathione S-transferase mu 2 (muscle) [KO:K00799]
GSTA3	glutathione S-transferase alpha 3 [KO:K00799]
GSTO1	glutathione S-transferase omega 1 [KO:K00799]
GSTT2B	glutathione S-transferase theta 2B (gene/pseudogene) [KO:K00799]
GSTK1	glutathione S-transferase kappa 1 [KO:K13299]
EPHX1	epoxide hydrolase 1, microsomal (xenobiotic) [KO:K01253]
CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6 [KO:K17709]
SULT2A1	sulfotransferase family, cytosolic, 2A, dehydroepiandrosterone (DHEA)-preferring, member 1 [KO:K11822]
CYP1A2	cytochrome P450, family 1, subfamily A, polypeptide 2 [KO:K07409]
CYP2A6	cytochrome P450, family 2, subfamily A, polypeptide 6 [KO:K17683]
CYP2E1	cytochrome P450, family 2, subfamily E, polypeptide 1 [KO:K07415]
CYP2F1	cytochrome P450, family 2, subfamily F, polypeptide 1 [KO:K07416]
CYP2S1	cytochrome P450, family 2, subfamily S, polypeptide 1 [KO:K07420]
AKR1C2	aldo-keto reductase family 1, member C2 [KO:K00089 K00212]
AKR1C4	aldo-keto reductase family 1, member C4 [KO:K00037 K00089 K00092 K00212]
AKR1C1	aldo-keto reductase family 1, member C1 [KO:K00089 K00212]
DHDH	dihydrodiol dehydrogenase (dimeric) [KO:K00078 [EC:1.1.1.179 1.3.1.20]
CYP2A13	cytochrome P450, family 2, subfamily A, polypeptide 13 [KO:K17685]
CYP2D6	cytochrome P450, family 2, subfamily D, polypeptide 6 [KO:K17712]
HSD11B1	hydroxysteroid (11-beta) dehydrogenase 1 [KO:K15680]
CBR1	carbonyl reductase 1 [KO:K00079]
CBR3	carbonyl reductase 3 [KO:K00084]
UGT2A1	UDP glucuronosyltransferase 2 family, polypeptide A1, complex locus [KO:K00699]
UGT2A3	UDP glucuronosyltransferase 2 family, polypeptide A3 [KO:K00699]
UGT2B17	UDP glucuronosyltransferase 2 family, polypeptide B17 [KO:K00699]
UGT2B11	UDP glucuronosyltransferase 2 family, polypeptide B11 [KO:K00699]

(cont'd)

UGT1A1	UDP glucuronosyltransferase 1 family, polypeptide A1 [KO:K00699]
UGT1A3	UDP glucuronosyltransferase 1 family, polypeptide A3 [KO:K00699]
UGT2B10	UDP glucuronosyltransferase 2 family, polypeptide B10 [KO:K00699]
UGT1A9	UDP glucuronosyltransferase 1 family, polypeptide A9 [KO:K00699]
UGT2B7	UDP glucuronosyltransferase 2 family, polypeptide B7 [KO:K00699]
UGT1A10	UDP glucuronosyltransferase 1 family, polypeptide A10 [KO:K00699]
UGT1A8	UDP glucuronosyltransferase 1 family, polypeptide A8 [KO:K00699]
UGT1A5	UDP glucuronosyltransferase 1 family, polypeptide A5 [KO:K00699]
UGT2B15	UDP glucuronosyltransferase 2 family, polypeptide B15 [KO:K00699]
UGT1A7	UDP glucuronosyltransferase 1 family, polypeptide A7 [KO:K00699]
UGT2B4	UDP glucuronosyltransferase 2 family, polypeptide B4 [KO:K00699]
UGT2A2	UDP glucuronosyltransferase 2 family, polypeptide A2 [KO:K00699]
CYP3A5	cytochrome P450, family 3, subfamily A, polypeptide 5 [KO:K17690]
AKR7A2	aldo-keto reductase family 7, member A2 (aflatoxin aldehyde reductase) [KO:K15303]
AKR7A3	aldo-keto reductase family 7, member A3 (aflatoxin aldehyde reductase) [KO:K15303]
ALDH3B1	aldehyde dehydrogenase 3 family, member B1 [KO:K00129]
ALDH3B2	aldehyde dehydrogenase 3 family, member B2 [KO:K00129]
ALDH1A3	aldehyde dehydrogenase 1 family, member A3 [KO:K00129]
ALDH3A1	aldehyde dehydrogenase 3 family, member A1 [KO:K00129]
ADH1A	alcohol dehydrogenase 1A (class I), alpha polypeptide [KO:K13951]
ADH1B	alcohol dehydrogenase 1B (class I), beta polypeptide [KO:K13951]
ADH1C	alcohol dehydrogenase 1C (class I), gamma polypeptide [KO:K13951]
ADH7	alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide [KO:K13951]
ADH4	alcohol dehydrogenase 4 (class II), pi polypeptide [KO:K13980]
ADH5	alcohol dehydrogenase 5 (class III), chi polypeptide [KO:K00121]
ADH6	alcohol dehydrogenase 6 (class V) [KO:K13952]