Biostatistics (2014), **0**, 0, *pp.* 1–24
doi:10.1093/biostatistics/BIOSTS˙14082˙SCPG˙supp˙v6

Statistical completion of a partially identified graph with applications for the estimation of gene regulatory networks: Supplementary Materials

DONGHYEON YU

Department of Statistics, Keimyung University, Daegu, Korea

WON SON

Department of Statistics, Seoul National University, Seoul, Korea

JOHAN LIM

Department of Statistics, Seoul National University, Seoul, Korea

GUANGHUA XIAO*

Department of Clinical Sciences, University of Texas Southwestern Medical Center, TX 75390, USA

guanghua.xiao@utsouthwestern.edu

APPENDIX

A. Modification of the active shooting algorithm

Let $\hat{\rho}^{ij,(m)}$ be the estimates of ρ^{ij} at the mth iteration in Step 1. The modified active shooting algorithm obtains the estimates for ρ by the following steps:

*To whom correspondence should be addressed.

- Step A1: (Initialization) From the one-dimensional lasso regression problems,

$$\hat{\rho}^{ij,(0)} = \begin{cases} \operatorname{sign}\big(\mathbf{Y}^T\tilde{\mathbf{X}}^{i,j}\big)\dfrac{\big(|\mathbf{Y}^T\tilde{\mathbf{X}}^{i,j}| - \lambda\big)_+}{(\tilde{\mathbf{X}}^{i,j})^T\tilde{\mathbf{X}}^{i,j}} & \text{for } (i,j) \notin \mathcal{K} \\ \dfrac{\mathbf{Y}^T\tilde{\mathbf{X}}^{i,j}}{(\tilde{\mathbf{X}}^{i,j})^T\tilde{\mathbf{X}}^{i,j}} & \text{for } (i,j) \in \mathcal{K} \end{cases},$$

  where $(x)_+ = \max(x, 0)$.

- Step A2: Define the active set $\mathcal{A} = \big\{(i,j) \mid \hat{\rho}^{ij,(m)} \neq 0\big\}$.

- Step A3: For each $(i,j) \in \mathcal{A}$, update $\hat{\rho}^{ij}$ by the following equations:

  for $(i,j) \notin \mathcal{K}$,

$$\begin{aligned} \hat{\rho}^{ij,(m)} &= \operatorname*{argmin}_{\rho^{ij}} \frac{1}{2}\|\mathbf{Y} - \sum_{(k,l)\neq(i,j)} \tilde{\mathbf{X}}^{k,l}\hat{\rho}^{kl,(m)} - \tilde{\mathbf{X}}^{i,j}\rho^{ij}\|_2^2 + \lambda|\rho^{ij}| \\ &= \operatorname{sign}\big(\mathbf{e}_{ij}^T\tilde{\mathbf{X}}^{i,j}\big)\frac{\big(|\mathbf{e}_{ij}^T\tilde{\mathbf{X}}^{i,j}| - \lambda\big)_+}{(\tilde{\mathbf{X}}^{i,j})^T\tilde{\mathbf{X}}^{i,j}}, \end{aligned} \tag{A.1}$$

  for $(i,j) \in \mathcal{K}$,

$$\hat{\rho}^{ij,(m)} = \operatorname*{argmin}_{\rho^{ij}} \frac{1}{2}\|\mathbf{Y} - \sum_{(k,l)\neq(i,j)} \tilde{\mathbf{X}}^{k,l}\hat{\rho}^{kl,(m)} - \tilde{\mathbf{X}}^{i,j}\rho^{ij}\|_2^2 = \frac{\mathbf{e}_{ij}^T\tilde{\mathbf{X}}^{i,j}}{(\tilde{\mathbf{X}}^{i,j})^T\tilde{\mathbf{X}}^{i,j}}, \tag{A.2}$$

  where $\mathbf{e}_{ij} = \mathbf{Y} - \sum_{(k,l)\neq(i,j)} \tilde{\mathbf{X}}^{k,l}\hat{\rho}^{kl,(m)}$ and $(x)_+ = \max(x, 0)$.

- Step A4: Repeat Step A3 until $\hat{\rho}^{ij,(m)}$ for $(i,j) \in \mathcal{A}$ converge.

- Step A5: For $1 \leqslant i < j \leqslant p$, update $\hat{\rho}^{ij,(m+1)}$ by (A.1) and (A.2).

- Step A6: Repeat Steps A2–A5 until $\hat{\rho}$ converges.

We remark that the modified equations above have inner products whose complexities are $O(p^2)$. However, the column vector $\tilde{\mathbf{X}}^{i,j}$ has many zero elements and, in addition, its non-zero elements are systematically allocated to allow for an efficient computation of the inner products. For example, $(\tilde{\mathbf{X}}^{i,j})^T(\tilde{\mathbf{X}}^{i,j}) = v_{ij}^2\sum_{k=1}^n (X_j^k)^2 + v_{ji}^2\sum_{k=1}^n (X_i^k)^2$ whose complexity is $O(n)$.

B. Preliminaries for the asymptotic properties

Assume we have observations from the model

$$Y_i = \mathbf{x}_i^T \beta + \epsilon_i, \quad i = 1, 2, \ldots, n$$

where $\mathbf{x}_i$ is a $p \times 1$ dimensional vector, $\beta = \big(\beta_1, \beta_2, \ldots, \beta_p\big)^T$ and ϵ_is are identically and independently distributed with mean 0 and variance σ^2. Let $\hat{\beta}_{\text{lasso}}(\lambda_n)$ be the solution to

$$\sum_{i=1}^{n} \big(Y_i - \mathbf{x}_i^T \beta\big)^2 + \lambda_n \sum_{j=1}^{p} \big|\beta_j\big|. \tag{B.1}$$

Suppose $\big(1/n\big) \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T$ converges to a non-singular matrix $\mathbf{S}$ with a size of $p \times p$ and $\lambda_n / \sqrt{n}$ approaches λ_0. Then, Knight and Fu (2000) shows that

$$\sqrt{n}\Big(\hat{\beta}_{\text{lasso}}(\lambda_n) - \beta\Big) \to \operatorname{argmin}\big(\mathrm{V}(\mathbf{u})\big) \tag{B.2}$$

in distribution, where

$$\mathrm{V}\big(\mathbf{u}\big) = -2\mathbf{u}^{\mathrm{T}}\mathbf{W} + \mathbf{u}^{\mathrm{T}}\mathbf{S}\mathbf{u} + \lambda_0 \sum_{j=1}^{p} \Big\{u_j \mathrm{sgn}\big(\beta_j\big)\mathrm{I}\big(\beta_j \neq 0\big) + \big|u_j\big|\mathrm{I}\big(\beta_j = 0\big)\Big\}, \tag{B.3}$$

and $\mathbf{W}$ is a multivariate normal random vector with mean 0 and variance $\sigma^2\mathbf{S}$.

The above results of Knight and Fu (2000) could not be directly applied to our linear model since the errors are not identically distributed. To be specific,

$$\mathrm{E} = \begin{pmatrix} \epsilon^1 \\ \vdots \\ \epsilon^n \end{pmatrix} \quad \text{with} \quad \epsilon^k = \begin{pmatrix} \epsilon_1^k \\ \vdots \\ \epsilon_p^k \end{pmatrix}, \; k = 1, 2, \ldots, n,$$

and ϵ_j^ks are uncorrelated (independent under normality) with each other but non-identically distributed for $j = 1, 2, \ldots, p$. However, the extension to the model with non-identical errors can be determined from the convergence of the law of

$$\begin{aligned} V_n\big(\mathbf{u}\big) \quad &= \quad \frac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{p} \Big(\big(\epsilon_j^k - \mathbf{u}^T \tilde{\mathbf{X}}_{p(k-1)+j} / \sqrt{np}\big)^2 - (\epsilon_j^k)^2 \Big) \\ &\quad + \lambda_n \sum_{1 \leqslant i < j \leqslant p} \big(|\rho^{ij} + u_{ij} / \sqrt{np}| - |\rho^{ij}| \big) \end{aligned}$$

with that of

$$-\mathbf{u}^T\mathbf{W} + \frac{1}{2}\mathbf{u}^T\mathbf{S}\mathbf{u} + \lambda_0 \sum_{1\leqslant i<j\leqslant p} \big(u_{ij}\mathrm{sign}(\rho^{ij})I(\rho^{ij} \neq 0) + |u_{ij}|I(\rho^{ij} = 0)\big),$$

where $\mathbf{u} = (u_{12}, u_{13}, \dots, u_{(p-1)p})^T$ is a $\big(p(p-1)/2\big)$-dimensional vector, $\tilde{\mathbf{X}}_i^T$ is the ith row vector of $\tilde{\mathbf{X}}$, $\mathbf{W} \sim N(0, \mathbf{V})$, $\mathbf{S}$ and $\mathbf{V}$ are non-singular matrices with a size of $\big(p(p-1)/2\big) \times \big(p(p-1)/2\big)$ and $\lambda_0 \geqslant 0$. To achieve this, we assume that

$$\frac{\lambda_n}{\sqrt{np}} \to \lambda_0 \ \text{ and } \ \frac{1}{np}\sum_{k=1}^{n}\sum_{j=1}^{p} \tilde{\mathbf{X}}_{p(k-1)+j}\tilde{\mathbf{X}}_{p(k-1)+j}^T \to \mathbf{S} \ \text{ as } n \to \infty$$

as in Knight and Fu (2000). In addition, to take care of non-identical errors, we assume that

$$\frac{1}{np}\sum_{k=1}^{n}\sum_{j=1}^{p} (\sigma^{jj})^{-1}\tilde{\mathbf{X}}_{p(k-1)+j}\tilde{\mathbf{X}}_{p(k-1)+j}^T \to \mathbf{V} \ \text{ as } n \to \infty, \tag{B.4}$$

where $(\sigma^{jj})^{-1}$s are variances of ϵ_j^ks for $j = 1, 2, \dots p$.

Another important result to be used in this section is the inequality from Anderson (1955). The following lemma is from Anderson (1955) and will be used to compare error probabilities between the SPACE and SCPG methods. In the lemma, a set A is denoted to be symmetric (about the origin) if and only if $\forall x \in S$, $-x \in S$.

Lemma B.1 Let D be a convex set in n-space, symmetric about the origin. Let $f(x) \geqslant 0$ be a function such that (i) $f(x) = f(-x)$, (ii) $\{x \mid f(x) \geqslant u\} = K_u$ is convex for every u $(0 < u < \infty)$, and (iii) $\int_D f(x)\, \mathrm{d}x < \infty$ (in Lebesgue sense). Then

$$\int_D f(x + ky)\, dx \geqslant \int_D f(x + y)\, dx,$$

for $0 \leqslant k \leqslant 1$.

C. Proofs for the asymptotic properties

C.1 *Proof of Theorem 1*

The proof is performed using the asymptotic properties of the lasso estimate by Knight and Fu (2000) and Anderson (1955)'s inequality. We recall that the SPACE model without the pre-information $\eta \neq 0$ solves

$$\frac{1}{2}\sum_{i=1}^{np}\big(y_i - \mathbf{b}_i^T\gamma - \mathbf{c}_i^T\eta\big)^2 + \lambda_n\Bigg(\sum_{j=1}^{|\mathcal{G}_0|}\big|\gamma_j\big| + \sum_{k=1}^{|\mathcal{K}|}\big|\eta_k\big|\Bigg), \tag{C.1}$$

whereas the SCPG model with the pre-information $\eta \neq 0$ solves

$$\frac{1}{2}\sum_{i=1}^{np}\big(y_i - \mathbf{b}_i^T\gamma - \mathbf{c}_i^T\eta\big)^2 + \lambda_n\Bigg(\sum_{j=1}^{|\mathcal{G}_0|}\big|\gamma_j\big|\Bigg), \tag{C.2}$$

where $\mathbf{b}_i^T$ and $\mathbf{c}_i^T$ are the ith rows of matrices $\mathbf{B}$ and $\mathbf{C}$, respectively.

Lemma C.1

$$\mathrm{P}\big(\hat{\gamma} = 0\big) \quad = \quad \mathrm{P}\big(-\lambda_0 1_{|\mathcal{G}_0|\times 1} < W_\gamma - S_{\gamma\eta}S_{\eta\eta}^{-1}\big(W_\eta - \lambda_0\mathrm{sign}(\eta)\big) < \lambda_0 1_{|\mathcal{G}_0|\times 1}\big),$$

where $W_\gamma - S_{\gamma\eta}S_{\eta\eta}^{-1}\big(W_\eta - \lambda_0\mathrm{sign}(\eta)\big) \sim N(\lambda_0 S_{\gamma\eta}S_{\eta\eta}^{-1}\mathrm{sign}(\eta), V_{\gamma\gamma\cdot\eta})$.

Proof. Let $\hat{\rho} = \big(\hat{\gamma}^T, \hat{\eta}^T\big)^T$ and $\hat{\rho}_{\mathcal{K}} = \big(\hat{\gamma}_{\mathcal{K}}^{\mathrm{T}}, \hat{\eta}_{\mathcal{K}}^T\big)^T$ be the solutions to the SPACE and SCPG models, respectively. Using the result of Knight and Fu (2000), we have

$$\sqrt{np}\Big(\hat{\rho}(\lambda_n) - \rho\Big) \rightarrow \underset{\mathbf{u}}{\mathrm{argmin}}\,\big(V(\mathbf{u})\big)$$

in distribution, where for $\mathbf{u} = \big(u_\gamma^T, u_\eta^T\big)^T$ and $\mathbf{W} \sim N\big(0, \mathbf{V}\big)$,

$$V\big(\mathbf{u}\big) = -\mathbf{u}^{\mathrm{T}}\mathbf{W} + \frac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{S}_{[\gamma,\eta]}\mathbf{u} + \lambda_0\Bigg(\sum_{j=1}^{|\mathcal{G}_0|}|u_{\gamma,j}| + \sum_{k=1}^{|\mathcal{K}|}u_{\eta,k}\mathrm{sign}(\eta_k)\Bigg). \tag{C.3}$$

In the above, $\mathbf{S}_{[\gamma,\eta]}$ is the limit of

$$\frac{1}{np}\begin{pmatrix} \mathbf{B}^T\mathbf{B} & \mathbf{B}^T\mathbf{C} \\ \mathbf{C}^T\mathbf{B} & \mathbf{C}^T\mathbf{C} \end{pmatrix}$$

and $\mathbf{V}$ is defined as the limit of a quadratic form of the design matrix $\tilde{X}$ of the model (3.11) in the main article as in (B.4), and they are partitioned into

$$\mathbf{S}_{[\gamma,\eta]} = \begin{pmatrix} S_{\gamma\gamma} & S_{\gamma\eta} \\ S_{\eta\gamma} & S_{\eta\eta} \end{pmatrix} \quad \text{and} \quad \mathbf{V} = \begin{pmatrix} V_{\gamma\gamma} & V_{\gamma\eta} \\ V_{\eta\gamma} & V_{\eta\eta} \end{pmatrix}.$$

We let $V_{\gamma\gamma\cdot\eta} = V_{\gamma\gamma} - S_{\gamma\eta} S_{\eta\eta}^{-1} V_{\eta\eta} S_{\eta\eta}^{-1} S_{\eta\gamma}$.

Let the solution to the minimum of (C.3) be $\mathbf{u}^* = \big((u_\gamma^*)^T, (u_\eta^*)^T\big)^T$. The solution $\mathbf{u}^*$ satisfies the optimality condition of $V(\mathbf{u})$, defined as

$$\mathbf{S}_{[\gamma,\eta]}\mathbf{u}^* - \mathbf{W} = -\lambda_0 \begin{pmatrix} \partial\|u_\gamma\|_1 \\ \text{sign}(\eta) \end{pmatrix},$$

where $\partial\|\mathbf{u}\|_1$ is a subdifferential of $\|\mathbf{u}\|_1$ such that $(\partial\|\mathbf{u}\|_1)_j = \text{sign}(u_j)$ if $u_j \neq 0$ and $(\partial\|\mathbf{u}\|_1)_j \in [-1,1]$ if $u_j = 0$. From the optimality condition, the event $\big\{u_\gamma^* = 0\big\}$ occurs if and only if the following coordinate-wise inequalities are satisfied

$$-\lambda_0 1_{|\mathcal{G}_0|\times 1} < W_\gamma - S_{\gamma\eta} S_{\eta\eta}^{-1}\big(W_\eta - \lambda_0 \text{sign}(\eta)\big) < \lambda_0 1_{|\mathcal{G}_0|\times 1}. \tag{C.4}$$

Hence, in asymptotic true negative probability, the true negative probability of the lasso estimate without the pre-information on η can be stated as

$$\begin{aligned} \text{P}\big(\hat{\gamma} = 0\big) &= \text{P}\big(u_\gamma^* = 0\big) \\ &= \text{P}\big(-\lambda_0 1_{|\mathcal{G}_0|\times 1} < W_\gamma - S_{\gamma\eta} S_{\eta\eta}^{-1}\big(W_\eta - \lambda_0 \text{sign}(\eta)\big) < \lambda_0 1_{|\mathcal{G}_0|\times 1}\big), \end{aligned} \tag{C.5}$$

where $W_\gamma - S_{\gamma\eta} S_{\eta\eta}^{-1}\big(W_\eta - \lambda_0 \text{sign}(\eta)\big) \sim N(\lambda_0 S_{\gamma\eta} S_{\eta\eta}^{-1} \text{sign}(\eta), V_{\gamma\gamma\cdot\eta})$. □

Lemma C.2

$$\text{P}\big(\hat{\gamma}_{\mathcal{K}} = 0\big) = \text{P}\big(-\lambda_0\ 1_{|\mathcal{G}_0|\times 1} < W_\gamma - S_{\gamma\eta} S_{\eta\eta}^{-1} W_\eta < \lambda_0\ 1_{|\mathcal{G}_0|\times 1}\big),$$

where $W_\gamma - S_{\gamma\eta} S_{\eta\eta}^{-1} W_\eta \sim N(0, V_{\gamma\gamma\cdot\eta})$.

Proof. On the other hand, for the SCPG model with pre-information $\eta \neq 0$,

$$\sqrt{np}\Big(\hat{\rho}_{\mathcal{K}}(\lambda_n) - \rho\Big) \to \underset{\mathbf{u}}{\text{argmin}}\big(V_{\mathcal{K}}(\mathbf{u})\big) \tag{C.6}$$

in distribution, where, for $\mathbf{u} = \big(u_\gamma^T, u_\eta^T\big)^T$ and $\mathbf{W} \sim N\big(0, \mathbf{V}\big)$,

$$V_{\mathcal{K}}\big(\mathbf{u}\big) = -\mathbf{u}^{\mathrm{T}}\mathbf{W} + \frac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{S}_{[\gamma,\eta]}\mathbf{u} + \lambda_0\bigg(\sum_{j=1}^{|\mathcal{G}_0|}|u_{\gamma,j}|\bigg).$$

Let the solution to the minimum of (C.6) be $\mathbf{u}^*_{\mathcal{K}} = \big((u^*_{\mathcal{K},\gamma})^T, (u^*_{\mathcal{K},\eta})^T\big)^T$. The solution $\mathbf{u}^*_{\mathcal{K}}$ satisfies the optimality condition of $V_{\mathcal{K}}(\mathbf{u})$, defined as

$$\mathbf{S}_{[\gamma,\eta]}\mathbf{u}^*_{\mathcal{K}} - \mathbf{W} = -\lambda_0 \begin{pmatrix} \partial\|u_\gamma\|_1 \\ 0 \end{pmatrix},$$

where $\partial\|\mathbf{u}\|_1$ is a subdifferential of $\|\mathbf{u}\|_1$ such that $(\partial\|\mathbf{u}\|_1)_j = \mathrm{sign}(u_j)$ if $u_j \neq 0$ and $(\partial\|\mathbf{u}\|_1)_j \in [-1, 1]$ if $u_j = 0$. From the optimality condition, the event $\big\{u^*_{\mathcal{K},\gamma} = 0\big\}$ occurs if and only if the following coordinate-wise inequalities are satisfied

$$-\lambda_0 1_{|\mathcal{G}_0|\times 1} < W_\gamma - S_{\gamma\eta}S_{\eta\eta}^{-1}W_\eta < \lambda_0 1_{|\mathcal{G}_0|\times 1}. \tag{C.7}$$

The asymptotic true negative probability of the lasso estimate with the pre-information on $\eta \neq 0$ is

$$\mathrm{P}\big(\hat{\gamma}_{\mathcal{K}} = 0\big) = \mathrm{P}\big(u^*_{\mathcal{K},\gamma} = 0\big) = \mathrm{P}\big(-\lambda_0\ 1_{|\mathcal{G}_0|\times 1} < W_\gamma - S_{\gamma\eta}S_{\eta\eta}^{-1}W_\eta < \lambda_0\ 1_{|\mathcal{G}_0|\times 1}\big), \tag{C.8}$$

where $W_\gamma - S_{\gamma\eta}S_{\eta\eta}^{-1}W_\eta \sim \mathrm{N}(0, V_{\gamma\gamma\cdot\eta})$. □

We now compare the probabilities (C.5) and (C.8) using Lemma B.1 in Appendix B. Note that they are the probabilities that two multivariate normal variables, having different mean vectors and the same covariance matrix, are in the same rectangle centered at the origin. To be specific, let $\mathbf{Z} = (Z_1, Z_2, \dots, Z_{|\mathcal{G}_0|})^T \equiv W_\gamma - S_{\gamma\eta}S_{\eta\eta}^{-1}W_\eta \sim \mathrm{N}(0, V_{\gamma\gamma\cdot\eta})$. For notational simplicity, we let $|\mathcal{G}_0|$, the dimension of Z, be m. Suppose $f(\mathbf{z})$ is the probability density function (pdf) of the normal random variable $\mathbf{Z}$, which is nonnegative and symmetric about the origin.

To apply Anderson's lemma, we consider the level set $K_u = \{\mathbf{z} \mid f(\mathbf{z}) \geqslant u\}$ and will show it is convex for $0 < u < \infty$. Recall that the pdf of $\mathbf{Z}$ is

$$f(\mathbf{z}) = c\exp\big\{\frac{1}{2}\mathbf{z}^T V_{\gamma\gamma\cdot\eta}^{-1}\mathbf{z}\big\},$$

where $c = (2\pi)^{-m/2}|V_{\gamma\gamma\cdot\eta}|^{-1/2}$ and is strictly positive. In the above, since $V_{\gamma\gamma\cdot\eta}^{-1}$ is a non-negative definite, we find that for $\mathbf{z}_1, \mathbf{z}_2 \in K_u$ and $0 \leqslant q \leqslant 1$,

$$\begin{aligned}&q\mathbf{z}_1^T V_{\gamma\gamma\cdot\eta}^{-1}\mathbf{z}_1 + (1-q)\mathbf{z}_2^T V_{\gamma\gamma\cdot\eta}^{-1}\mathbf{z}_2 - \big(q\mathbf{z}_1 + (1-q)\mathbf{z}_2\big)^T V_{\gamma\gamma\cdot\eta}^{-1}\big(q\mathbf{z}_1 + (1-q)\mathbf{z}_2\big)\\ &= q(1-q)(\mathbf{z}_1 - \mathbf{z}_2)^T V_{\gamma\gamma\cdot\eta}^{-1}(\mathbf{z}_1 - \mathbf{z}_2) \geqslant 0.\end{aligned}$$

Thus,

$$\begin{aligned} f\big(q\mathbf{z}_1 + (1-q)\mathbf{z}_2\big) &\geqslant f(\mathbf{z}_1)^q f(\mathbf{z}_2)^{1-q}\\ &\geqslant \min\{f(\mathbf{z}_1), f(\mathbf{z}_2)\} \geqslant u \Rightarrow q\mathbf{z}_1 + (1-q)\mathbf{z}_2 \in K_u,\end{aligned}$$

which implies that K_u is convex for $0 < u < \infty$. In the lemma, we consider the rectangular set $D = \{\mathbf{z} \mid |z_j| < \lambda_0,\ \ j = 1, 2, \ldots, m\}$. The set D is symmetric about the origin and convex in $\mathcal{R}^m$. Finally, using Anderson's lemma with $k = 0$, $x = \mathbf{z}$, $y = -\lambda_0 S_{\gamma\eta}S_{\eta\eta}^{-1}\text{sign}(\eta)$ and $D = \{\mathbf{z} \mid |z_j| < \lambda_0 \text{ for } j = 1, 2, \ldots, m\}$, we show that

$$\mathrm{P}\big(\hat\gamma_{\mathcal{K}} = 0\big) = \int_D f(x)\ dx \geqslant \int_D f(x+y)\ dx = \mathrm{P}\big(\hat\gamma = 0\big).$$

C.2 *Proof of Theorem 2*

Suppose we have knowledge on $\eta \neq 0$. We compare the asymptotic probabilities

$$\mathrm{P}\big(\hat\alpha \neq 0\big) \quad \text{and} \quad \mathrm{P}\big(\hat\alpha_{\mathcal{K}} \neq 0\big), \tag{C.9}$$

where $\hat\alpha$ and $\hat\alpha_{\mathcal{K}}$ are the lasso estimates without/with information on $\eta \neq 0$.

The SPACE model without information on $\eta \neq 0$ solves

$$\frac{1}{2}\sum_{i=1}^{n}\big(y_i - \mathbf{a}_i^T\alpha - \mathbf{c}_i^T\eta\big)^2 + \lambda_n\bigg(\sum_{j=1}^{|\mathcal{G}_1|}\big|\alpha_j\big| + \sum_{j=1}^{|\mathcal{K}|}\big|\eta_j\big|\bigg), \tag{C.10}$$

and the SCPG model with the pre-identified information solves

$$\frac{1}{2}\sum_{i=1}^{n}\big(y_i - \mathbf{a}_i^T\alpha - \mathbf{c}_i^T\eta\big)^2 + \lambda_n\bigg(\sum_{j=1}^{|\mathcal{G}_1|}\big|\alpha_j\big|\bigg), \tag{C.11}$$

where $\mathbf{a}_i^T$ and $\mathbf{c}_i^T$ are the ith row vectors of $\mathbf{A}$ and $\mathbf{C}$, respectively. Let $\hat\rho = \big(\hat\alpha^T, \hat\eta^T\big)^T$ and $\hat\rho_{\mathcal{K}} = \big(\hat\alpha_{\mathcal{K}}^{\mathrm{T}}, \hat\eta_{\mathcal{K}}^T\big)^T$ be the solutions to the SPACE and SCPG models, respectively. Similar to the above method, using Knight and Fu (2000),

$$\sqrt{np}\Big(\hat{\rho}(\lambda_n) - \rho\Big) \to \operatorname*{argmin}_{\mathbf{u}} \big(V(\mathbf{u})\big) \tag{C.12}$$

in distribution, where, for $\mathbf{u} = (u_\alpha^T, u_\eta^T)^{\mathrm{T}}$ and $\mathbf{W} = \big(W_\alpha^T, W_\eta^T\big)^T \sim N\big(0, \mathbf{V}\big)$,

$$V\big(\mathbf{u}\big) = -\mathbf{u}^{\mathrm{T}}\mathbf{W} + \frac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{S}_{[\alpha,\eta]}\mathbf{u} + \lambda_0\bigg(\sum_{j=1}^{|\mathcal{G}_1|} u_{\alpha,j}\mathrm{sign}\big(\alpha_j\big) + \sum_{k=1}^{|\mathcal{K}|} u_{\eta,k}\mathrm{sign}(\eta_k)\bigg). \tag{C.13}$$

In the above, $\mathbf{S}_{[\alpha,\eta]}$ is the limit of

$$\frac{1}{np}\left(\begin{array}{cc} \mathbf{A}^T\mathbf{A} & \mathbf{A}^T\mathbf{C} \\ \mathbf{C}^T\mathbf{A} & \mathbf{C}^T\mathbf{C} \end{array}\right)$$

and $\mathbf{V}$ is again defined as in (B.4). Both matrices, $\mathbf{S}_{[\alpha,\eta]}$ and $\mathbf{V}$, are partitioned into

$$\mathbf{S}_{[\alpha,\eta]} = \left(\begin{array}{cc} S_{\alpha\alpha} & S_{\alpha\eta} \\ S_{\eta\alpha} & S_{\eta\eta} \end{array}\right) \quad \text{and} \quad \mathbf{V} = \left(\begin{array}{cc} V_{\alpha\alpha} & V_{\alpha\eta} \\ V_{\eta\alpha} & V_{\eta\eta} \end{array}\right).$$

Let $\mathbf{u}^* = \big((u_\alpha^*)^T, (u_\gamma^*)^T\big)^T$ be the solution to the minimum of (C.13). Since $V(\mathbf{u})$ is differentiable with respect to $\mathbf{u}$, the asymptotic distribution of $\sqrt{np}\Big(\hat{\rho}(\lambda_n) - \rho\Big)$ can be explicitly described as

$$\mathbf{u}^* = \mathbf{S}_{[\alpha,\eta]}^{-1}\left(\left(\begin{array}{c} W_\alpha \\ W_\eta \end{array}\right) - \lambda_0\left(\begin{array}{c} \mathrm{sign}(\alpha) \\ \mathrm{sign}(\eta) \end{array}\right)\right) \tag{C.14}$$

In comparison, for the SCPG model with pre-information $\eta \neq 0$,

$$\sqrt{np}\Big(\hat{\rho}_{\mathcal{K}}(\lambda_n) - \rho\Big) \to \operatorname*{argmin}_{\mathbf{u}} \big(V_{\mathcal{K}}(\mathbf{u})\big) \tag{C.15}$$

in distribution, where for $\mathbf{u} = \big(u_\alpha^T, u_\eta^T\big)$ and $\mathbf{W} \sim N\big(0, \mathbf{V}\big)$,

$$V_{\mathcal{K}}\big(\mathbf{u}\big) = -\mathbf{u}^{\mathrm{T}}\mathbf{W} + \frac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{S}_{[\alpha,\eta]}\mathbf{u} + \lambda_0\sum_{j=1}^{|\mathcal{G}_1|} u_{\alpha,j}\mathrm{sign}\big(\alpha_j\big).$$

Let the solution to the minimum of (C.6) be $\mathbf{u}_{\mathcal{K}}^* = \big((u_{\mathcal{K},\alpha}^*)^T, (u_{\mathcal{K},\gamma}^*)^T\big)^T$. Since $V_{\mathcal{K}}(\mathbf{u})$ is also differentiable with respect to $\mathbf{u}$, the asymptotic distribution of $\sqrt{np}\Big(\hat{\rho}_{\mathcal{K}}(\lambda_n) - \rho\Big)$ can be explicitly described as

$$\mathbf{u}_{\mathcal{K}}^* = \mathbf{S}_{[\alpha,\eta]}^{-1}\left(\left(\begin{array}{c} W_\alpha \\ W_\eta \end{array}\right) - \lambda_0\left(\begin{array}{c} \mathrm{sign}(\alpha) \\ 0 \end{array}\right)\right) \tag{C.16}$$

To describe the events $\{\hat{\alpha} \neq 0\}$, let $\mathbf{u}_\alpha^{(n)} = \sqrt{np}(\hat{\alpha}(\lambda_n) - \alpha)$. The event $\{\hat{\alpha} = 0\}$ is equivalent to the event $\{\mathbf{u}_\alpha^{(n)} = -\sqrt{np}\alpha\}$. From the previous result that $\mathbf{u}_\alpha^{(n)}$ converges with $\mathbf{u}_\alpha^*$ in distribution, which follows a Gaussian distribution, and the assumption that α is a non-zero vector, the false negative probability $\mathrm{P}(\mathbf{u}_\alpha^{(n)} = -\sqrt{np}\alpha)$ for α of the SPACE model converges to zero as $n \to \infty$. This implies the asymptotic true positive probability $\mathrm{P}(\{\hat{\alpha} \neq 0\})$ for α of the SPACE model converges to one as $n \to \infty$. Using the same step in the SPACE model, can show that the asymptotic true positive probability $\mathrm{P}(\{\hat{\alpha}_{\mathcal{K}} \neq 0\})$ for α of the SCPG model also converges to one as $n \to \infty$.

D. Simulated networks in Numerical study

Let p denote the number of nodes in graphs.

- **(N1)** AR(1) network: The AR(1) network is also known as a chain graph having an edge set $E = \{(j-1, j) \mid j = 2, 3, \ldots, p\}$. We define a concentration matrix as

$$\Omega = (\sigma^{ij})_{1 \leqslant i,j \leqslant p} = \begin{cases} 1 & \text{if } i = j \\ 0.45 & \text{if } |i-j| = 1 \\ 0 & \text{otherwise} \end{cases}$$

- **(N2)** AR(2) network: The AR(2) network contains the AR(1) network and additionally has edges between the jth node and $(j-2)$th node for $j = 3, 4, \ldots, p$. We define a concentration matrix as

$$\Omega = (\sigma^{ij})_{1 \leqslant i,j \leqslant p} = \begin{cases} 1 & \text{if } i = j \\ 0.5 & \text{if } |i-j| = 1 \\ 0.4 & \text{if } |i-j| = 2 \\ 0 & \text{otherwise} \end{cases}$$

- **(N3)** Hub network: We consider a hub network as described in Peng et al. (2009). For $p = 100$, a hub network consists of three hub nodes whose degrees are each around 15, and 97 non-hub nodes whose degrees lie between 1 and 3. Edges in a hub network are randomly selected according to the above conditions. We generate a concentration matrix corresponding to a given edge set E by the following steps. At step 1, element σ^{ij} of a concentration matrix is defined as

$$\Omega = (\sigma^{ij})_{1 \leqslant i,j \leqslant p} = \begin{cases} 1 & \text{if } i = j \\ \sim \mathcal{U}([-1,-0.5] \cup [0.5,1]) & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases},$$

where $\mathcal{U}(A)$ denotes a uniform distribution over a set A. To assure the positive definiteness of the concentration matrix, the off-diagonal elements σ^{ij} are divided by $1.5 \sum_{k \neq i} |\sigma^{ik}|$ at step 2. Finally, the concentration matrix Ω is replaced into $(\Omega + \Omega^T)/2$ to be symmetric.

- **(N4)** Scale-free network: In the scale-free network, degrees of nodes follow the power law distribution having a form

  $$P(k) \propto k^{-\alpha},$$

  where $P(k)$ is a fraction of nodes having k connections and α is a preferential attachment parameter. The preferential attachment parameter generally lies between 2 and 3. We set $\alpha = 2.3$ and generate a scale-free network using the Barabasi and Albert (BA) model (Barabasi and Albert, 1999). This model is also implemented in the R package "`igraph`". With a given edge set E from the BA model, we generate a concentration matrix using the procedure described in **(N3)**.

E. Details of identification of survival-related genes

We used univariate Cox regression to identify the probe sets that are significantly correlated with a patient's overall survival time after adjusting for clinical sites, age, gender, and stages. In order to deal with the multiple comparison issue, we calculated the FDR, proposed by Benjamini and Hochberg (1995), by fitting p-values using a Beta-Uniform model (Pounds and Morris, 2003). We identified a subgroup of 983 probes whose expression levels had been shown to be strongly associated with patients' overall survival time (p-value < 0.00655, with estimated FDR $<$0.10). In total, there were 794 annotated genes from this probe set, and their expression values were calculated using the average values of related probe sets. Each array was normalized by subtracting its mean and dividing its median absolute deviation (MAD) before applying the estimation procedure.

F. Details of lung cancer genes identified by both the SPACE and SCPG methods

Weir et al. (2007) studied the cancer genome in lung adenocarcinoma and characterized copy-number alterations in a large collection of primary tumors ($n = 371$) using single nucleotide polymorphism (SNP) arrays. The study identified NKX2-1 as the driver gene of lung adenocarcinoma. The homeobox containing gene HOP (Homeodomain Only Protein) is a downstream gene of NKX2-1 in the regulation of pulmonary gene expression (Chen et al., 2007) and is a tumor suppressor gene in lung cancer (Yin et al., 2006). In addition, the plasma pro-surfactant protein B (pro-SFTPB) was identified as an independent predictor of lung cancer risk using baseline plasma samples in the large Pan-Canadian Early Detection of Lung Cancer Study ($n = 2,485$) (Sin et al., 2013).

G. Details of lung cancer genes only identified by the SCPG method

In comparing the two methods, we noted that the SCPG method identified nine genes that were missed by the SPACE method, including CTNNB1, CSNK2A1, ESR1, NEDD9, FYN, BRCA1, PTPN13, PIK3R1 and SLC34A2. Seven of these nine genes (except PTPN13 and SLC34A2) had been reported to play important roles in lung cancer. For example, the beta-catenin gene (CTNNB1) has been known to be genetically mutated in non-small-cell lung cancer (NSCLC) (Shigemitsu et al., 2001), and recent studies show its mRNA expression is significantly associated with the clinical outcome of NSCLC patients (Woenckhaus et al., 2008). The CSNK2A1 (sometimes referred to as CK2A1), Casein kinase 2, alpha 1 polypeptide, is overexpressed in many cancer types (such as lung cancer) (Ruzzene and Pinna, 2010), and CK2 inhibitors are currently being studied as promising treatments of lung cancer (Lin et al., 2011). ESR1 (EStrogen Receptor 1), is well known to be associated with risk, progression and treatment response in breast cancer (Holst et al., 2007), as well as in lung cancer (Stabile and Siegfried , 2004). The NEDD9 gene has been reportedly associated with lung cancer metastasis (Jin et al., 2014) and progression (Feng et al., 2012). The FYN gene (Semba et al., 1986) encodes the Proto-oncogene tyrosine-protein kinase Fyn and plays an important role in formation and treatment response in lung cancer (Kim et al., 2011). BRCA1 is the most important biomarker for the early onset of breast cancer, and recently many studies have shown that BRCA1 gene expression is a biomarker for the progression (Kang et al., 2010; Reguart et al., 2008) and response to chemotherapy (Kim et al., 2008) in lung cancer. Finally, the PIK3R1 gene encodes $p85\alpha$, which is the inhibitory subunit of PI3K, while the Akt/PI3K signaling pathway is one of the most important pathways in lung cancer development, progression and treatment (Gadgeel and Wozniak, 2013; Gustafson et al., 2010).

In addition, the SCPG method identified the PTPN13 gene, which had not been previously reported as a lung cancer related gene. To further study this gene, we have downloaded the mRNA expression together with the clinical annotation from four public lung cancer datasets, including 1. Tomida et al. (2009) ($n = 117$), 2. Bhattacharjee et al. (2001) ($n = 203$), 3. Raponi et al. (2006) ($n = 129$), 4. Jones et al. (2004) ($n = 80$). These four datasets were selected because they were published in high-profile journals, contained relatively large sample sizes (at least 80 samples), and were measured from different microarray platforms. Interestingly, the under-expression of the PTPN13 gene is consistently associated with the poor prognosis of lung cancer patients in the four independent datasets, which were measured using different platforms (see Fig. G.1). The results show that the mRNA expression of the PTPN13 gene is a novel and robust prognostic biomarker of potential clinical importance.

H. Comparison of the SCPG and Naive method

One heuristic way to use the pre-identified information would be to directly add (or delete) the pre-identified edges to (or from) the estimated network by the SPACE method. We denote this procedure as the "naive method" and, in this section, we compare the performances of SCPG (and also SPACE) to the naive method. We compare the SCPG and naive methods in all the numerical studies used in Section 4 of the main paper: four network structures (AR1, AR2, hub network and scale-free network), three network sizes (100, 250, 500) and two sets of pre-identified edges (10% and 30%). For each numerical study, we simulated 50 datasets. For both SCPG and naive methods, the FDRs averaged across 50 datasets for each study are summarized in Table H.1 . This shows that the SCPG method clearly performs better than the naive method.

I. Sensitivity analysis on false positives in the pre-identified edges

In this paper, we assume that all pre-identified edges are truly connected and nomisspecification occurs in the pre-identified information. However, as we remark in the Conclusion, this assumption may not always be true. In this section, we numerically investigate how the SCPG method is sensitive to the false positives in the pre-identified information.

In the study, the same four networks with two sets of pre-identified edges (10% and 30%) from Section 4 are considered. In the two sets of the pre-identified edges (10% and 30%), we assume 0%, 10%, 15%, 20%, and 30% of the pre-identified edges are falsely identified (they are not connected in truth). We refer to these rates as the misspecification rates. Tables I.1 and I.2 report the averages of $|\hat{E}|$, TPR, TNR, FDR, MISR, and MCC for the four networks for the case $p = 500$ and $n = 250$.

It is not surprising that all performance measures become worse as the misspecification rate increases. However, the tables show that the SCPG method still performs well (performs better than SPACE) if the misspecification rate is not very high. In comparison with SPACE, the findings for the 0% misspecification rate are still true unless the percentage of false positives is not larger than 15%.

J. Application to the construction of a larger network

The proposed SCPG method is a modification of the SPACE method and does not have any difficulty in estimating networks with a couple of thousands of nodes, like the SPACE method (Peng et al., 2009). In this section, to show the scalability of the SCPG, we additionally conduct a numerical study for hub and scale-free networks with a thousand nodes ($p = 1000$) and samples sizes $n = 200, 300, 500$ as in Peng et al. (2009). Here, the hub and scale-free networks are the two most popular large-scale networks used in many other applications. Again, the averages of $|\hat{E}|$, TPR, TNR, FDR, MISR, and MCC are reported in Table J.1. The findings apparent in the table are very similar to what we have in Section 4 for $p = 500$.

References

Anderson, T. W., (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society*, **6**(2), 170–176.

Barabasi, A. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, **286**(5439), 509–512.

Bejamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society- Series B*, **57**(1), 289–300.

Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M. *and others* (2001). Classification of human lung carcinomas by mrna expression profiling reveals distinct adenocarcinoma subclasses. *Proceedings of the National Academy of Sciences of the United States of America*, **98**(24), 13790–13795.

Chen, Y., Pacyna-Gengelbach, M., Deutschmann, N., Niesporek, S. and Petersen, I. (2007). Homeobox gene hop has a potential tumor suppressive activity in human lung cancer. *International Journal of Cancer*, **121**(5), 1021–1027.

Feng, Y., Wang, Y., Wang, Z., Fang, Z., Li, F., Gao, Y., Liu, H., Xiao, T., Li, F., Zhou, Y. *and others* (2012). The crtc1-nedd9 signaling axis mediates lung cancer progression caused by lkb1 loss. *Cancer Research*, **72**(24), 6502–6511.

Gadgeel, S. M. and Wozniak, A. (2013). Preclinical rationale for pi3k/akt/mtor pathway inhibitors as therapy for epidermal growth factor receptor inhibitor-resistant non-small-cell lung cancer. *Clinical Lung Cancer*, **14**(4), 322–332.

Gustafson, A. M., Soldi, R., Anderlind, C., Scholand, M. B., Qian, J., Zhang, X., Cooper, K., Walker, D., McWilliams, A., Liu, G. *and others* (2010). Airway pi3k pathway activation is an early and reversible event in lung cancer development. *Science Translational Medicine*, **2**(26), 26ra25.

Holst, F., Stahl, P. R., Ruiz, C., Hellwinkel, O., Jehan, Z., Wendland, M., Lebeau, A., Terracciano, L., Al-Kuraya, K., Janicke, F. *and others* (2007). Estrogen receptor alpha (ESR1) gene amplification is frequent in breast cancer. *Nature Genetics*, **39**(5), 655–660.

Jin, Y., Li, F., Zheng, C., Wang, Y., Fang, Z., Guo, C., Wang, X., Liu, H., Deng, L., Li, C. *and others* (2014). Nedd9 promotes lung cancer metastasis through epithelial-mesenchymal transition. *International Journal of Cancer*, **134**(10), 2294–2304.

Jones, M. H., Virtanen, C., Honjoh, D., Miyoshi, T., Satoh, Y., Okumura, S., Nakagawa, K., Nomura, H. and Ishikawa, Y. (2004). Two prognostically significant subtypes of high-grade lung neuroendocrine tumours independent of small-cell and large-cell neuroendocrine carcinomas identified by gene expression profiles. *The Lancet*, **363**(9411), 775–781.

Kang, C. H., Jang, B. G., Kim, D. W., Chung, D. H., Kim, Y. T., Jheon, S., Sung, S. W. and Kim, J. H. (2010). The prognostic significance of ercc1, brca1, xrcc1, and betaiii-tubulin expression in patients with non-small cell lung cancer treated by platinum- and taxane-based neoadjuvant chemotherapy and surgical resection. *Lung Cancer*, **68**(3), 478–483.

Kim, H. T., Lee, J. E., Shin, E. S., Yoo, Y. K., Cho, J. H., Yun, M. H., Kim, Y. H., Kim, S. K., Kim, H. J., Jang, T. W. *and others* (2008). Effect of brca1 haplotype on survival of non-small-cell lung cancer patients treated with platinum-based chemotherapy. *Journal of Clinical Oncology*, **26**(36), 5972–5979.

Kim, A. N., Jeon, W. K., Lim, K. H., Lee, H. Y., Kim, W. J. and Kim, B. C. (2011). Fyn mediates transforming growth factor-beta1-induced down-regulation of E-cadherin in human A549 lung cancer cells. *Biochemical and Biophysical Research Communications*, **407**(1), 181–184.

Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics*, **28**(5), 1356–1378.

Lin, Y. C., Hung, M. S., Lin, C. K., Li, J. M., Lee, K. D., Li, Y. C., Chen, M. F., Chen, J. K. and Yang, C. T. (2011). Ck2 inhibitors enhance the radiosensitivity of human non-small cell lung cancer cells through inhibition of stat3 activation. *Cancer Biotherapy and Radiopharmaceuticals*, **26**(3), 381–388.

Peng, J., Wang, P., Zhou, N. and Zhu, J. (2009). Partial Correlation Estimation by Joint Sparse Regression Models. *Journal of the American Statistical Association*, **104**, 735–746.

Pounds, S. and Morris, S. W. (2003). Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *Bioinformatics*, **19**(10), 1236–1242.

Raponi, M., Zhang, Y., Yu, J., Chen, G., Lee, G., Taylor, J. M., Macdonald, J., Thomas, D., Moskaluk, C., Wang, Y. and Beer, D. G. (2006). Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Research*, **66**(15), 7466–7472.

Reguart, N., Cardona, A. F., Carrasco, E., Gomez, P., Taron, M. and Rosell, R. (2008). Brca1: a new genomic marker for non-small-cell lung cancer. *Clinical Lung Cancer*, **9**(6), 331–339.

Ruzzene, M. and Pinna, L. A. (2010). Addiction to protein kinase ck2: a common denominator of diverse cancer cells? *Biochimica et Biophysica Acta*, **1804**(3), 499–504.

Semba, K., Nishizawa, M., Miyajima, N., Yoshida, M. C., Sukegawa, J., Yamanashi, Y., Sasaki, M., Yamamoto, T. and Toyoshima, K. (1986). yes-related protooncogene, syn, belongs to the protein-tyrosine kinase family. *Proceedings of the National Academy of Sciences of the United States of America*, **83**(15), 5459–5463.

Shigemitsu, K., Sekido, Y., Usami, N., Mori, S., Sato, M., Horio, Y., Hasegawa, Y., Bader, S. A., Gazdar, A. F., Minna, J. D. *and others* (2001). Genetic alteration of the beta-catenin gene (CTNNB1) in human lung cancer and malignant mesothelioma and identification of a new 3p21.3 homozygous deletion. *Oncogene*, **20**(31), 4249–4257.

Sin, D. D., Tammemagi, C. M., Lam, S., Barnett, M. J., Duan, X., Tam, A., Auman, H., Feng, Z., Goodman, G. E., Hanash, S. and Taguchi, A. (2013). ProSurfactant Protein B as a Biomarker for Lung Cancer Prediction. *Journal of Clinical Oncology*, doi:10.1200/JCO.2013.50.6105.

Stabile, L. P. and Siegfried, J. M. (2004). Estrogen receptor pathways in lung cancer. *Current Oncology Reports*, **6**(4), 259–267.

Tomida, S., Takeuchi, T., Shimada, Y., Arima, C., Matsuo, K., Mitsudomi, T., Yatabe, Y. and Takahashi, T. (2009). Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. *Journal of Clinical Oncology*, **27**(17), 2793–2799.

Weir, B. A., Woo, M. S., Getz, G., Perner, S., Ding, L., Beroukhim, R., Lin, W. M., Province, M. A., Kraja, A., Johnson, L. A. *and others* (2007). Characterizing the cancer genome in lung adenocarcinoma. *Nature*, **450**(7171), 893–898.

Woenckhaus, M., Merk, J., Stoehr, R., Schaeper, F., Gaumann, A., Wiebe, K., Hartmann, A., Hofstaedter, F. and Dietmaier, W. (2008). Prognostic value of FHIT, CTNNB1, and MUC1 expression in non-small cell lung cancer. *Human Pathology*, **39**(1), 126–136.

Yin, Z., Gonzales, L., Kolla, V., Rath, N., Zhang, Y., Lu, M. M., Kimura, S., Ballard, P. L., Beers, M. F., Epstein, J. A. and Morrisey, E. E. (2006). Hop functions downstream of nkx2.1 and gata6 to mediate hdac-dependent negative regulation of pulmonary gene expression. *American Journal of Physiology - Lung Cellular and Molecular Physiology*, **291**(2), L191–L199.

[*Submitted April 29, 2014*]

Table H.1. The FDRs of the SCPG and naive methods for AR(1), AR(2), hub network and scale-free network under various network sizes (100, 250, 500) and percentages of pre-identified edges (10% and 30%). The sample size n =250. Each performance was averaged over 50 simulated datasets. “Info.” stands for a percentage of pre-identified information and “Naive(k%)” denotes the naive method that incorporates the pre-identified edges whose percentage of total connected edges is k% into the estimated network from the SPACE. The numbers in parentheses denote the standard errors of measures.

p	Info.	Network			
		AR(1)	AR(2)	Hub	Scale-free
100	None	20.45 (0.32)	47.59 (0.14)	12.15 (0.49)	16.58 (0.42)
	Naive(10%)	20.45 (0.32)	46.87 (0.15)	11.08 (0.46)	15.85 (0.41)
	10%	18.7 (0.29)	42.29 (0.20)	10.56 (0.44)	14.04 (0.45)
	Naive(30%)	20.45 (0.32)	45.28 (0.18)	9.68 (0.42)	14.84 (0.40)
	30%	14.68 (0.36)	32.56 (0.25)	5.61 (0.35)	9.72 (0.38)
250	None	18.09 (0.30)	46.75 (0.19)	15.43 (0.37)	15.87 (0.42)
	Naive(10%)	18.09 (0.30)	46.75 (0.19)	15.26 (0.37)	15.73 (0.42)
	10%	16.30 (0.33)	42.14 (0.21)	13.39 (0.31)	14.31 (0.41)
	Naive(30%)	18.09 (0.30)	46.75 (0.19)	14.94 (0.37)	15.48 (0.42)
	30%	12.52 (0.28)	31.49 (0.19)	10.68 (0.29)	10.37 (0.39)
500	None	16.90 (0.30)	44.61 (0.23)	15.57 (0.33)	14.61 (0.34)
	Naive(10%)	16.90 (0.30)	44.61 (0.23)	15.55 (0.33)	14.60 (0.34)
	10%	14.61 (0.26)	40.09 (0.23)	14.31 (0.28)	13.18 (0.34)
	Naive(30%)	16.90 (0.30)	44.61 (0.23)	15.49 (0.32)	14.51 (0.34)
	30%	11.83 (0.25)	29.69 (0.18)	10.92 (0.29)	10.33 (0.33)

Table I.1. The averages of $|\hat{E}|$, TPR, TNR, FDR, MISR, and MCC for AR(1) and AR(2) networks over 50 datasets ($p = 500, n = 250$). "Info." and "FP(Info.)" stand for the percentage of pre-identified edges and the percentage of false positives in the pre-identified edges, respectively. $|\hat{E}|$ denotes the number of estimated edges. All values except for $|\hat{E}|$ are multiplied by 100. The numbers in parentheses denote the standard errors of measures.

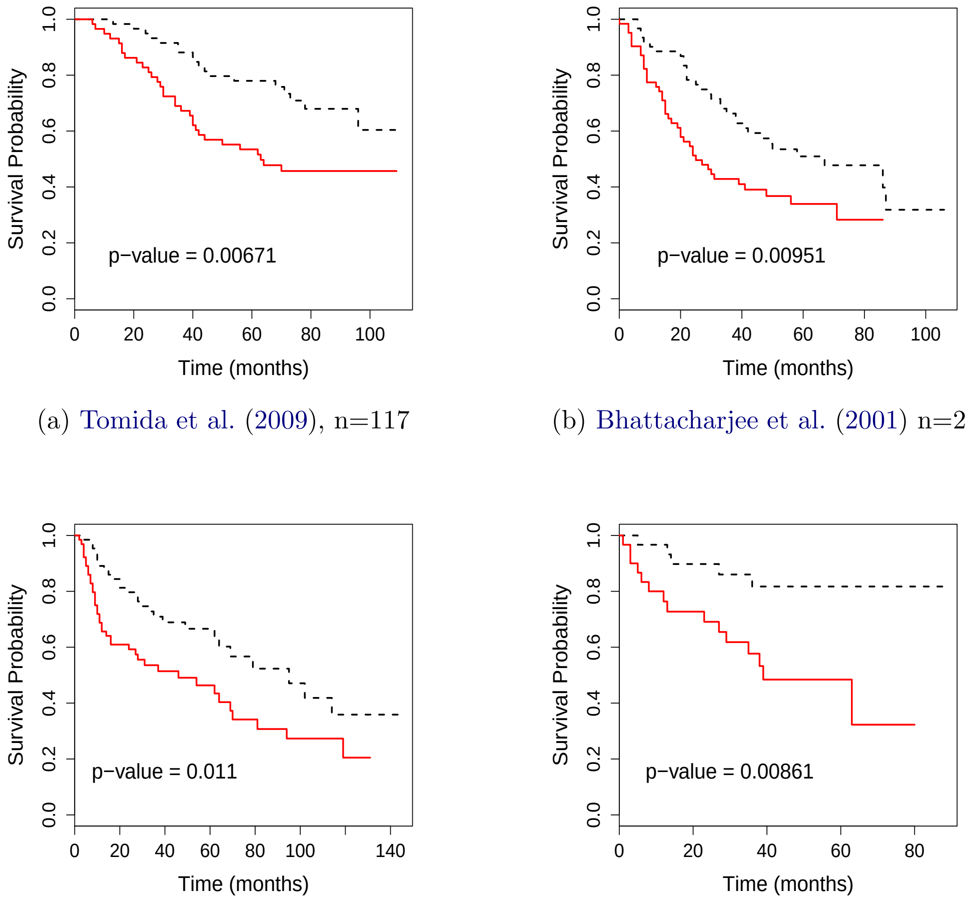
Network	Info.	FP(Info.)	$\lvert\hat{E}\rvert$	TPR	TNR	FDR	MISR	MCC
AR(1) ($\lvert E\rvert = 499$)	None		609.58	100	99.91	18.09	0.09	90.46
			(2.24)	(0.00)	(0.00)	(0.30)	(0.00)	(0.17)
	10%	0%	596.62	100	99.92	16.3	0.08	91.44
			(2.40)	(0.00)	(0.00)	(0.33)	(0.00)	(0.18)
		10%	599.46	100	99.92	16.71	0.08	91.22
			(2.09)	(0.00)	(0.00)	(0.29)	(0.00)	(0.16)
		15%	605.10	100	99.91	17.49	0.09	90.79
			(2.13)	(0.00)	(0.00)	(0.28)	(0.00)	(0.16)
		20%	608.64	100	99.91	17.97	0.09	90.52
			(2.00)	(0.00)	(0.00)	(0.27)	(0.00)	(0.15)
		30%	613.92	100	99.91	18.67	0.09	90.13
			(2.12)	(0.00)	(0.00)	(0.28)	(0.00)	(0.16)
	30%	0%	570.74	100	99.94	12.52	0.06	93.5
			(1.87)	(0.00)	(0.00)	(0.28)	(0.00)	(0.15)
		10%	590.46	100	99.93	15.43	0.07	91.92
			(2.17)	(0.00)	(0.00)	(0.31)	(0.00)	(0.17)
		15%	599.26	100	99.92	16.68	0.08	91.24
			(2.20)	(0.00)	(0.00)	(0.30)	(0.00)	(0.17)
		20%	607.30	100	99.91	17.77	0.09	90.63
			(2.50)	(0.00)	(0.00)	(0.33)	(0.00)	(0.18)
		30%	627.58	100	99.90	20.44	0.10	89.14
			(2.13)	(0.00)	(0.00)	(0.27)	(0.00)	(0.15)
AR(2) ($\lvert E\rvert = 997$)	None		1873.38	100	99.29	46.75	0.7	72.71
			(6.57)	(0.00)	(0.01)	(0.19)	(0.01)	(0.13)
	10%	0%	1724.14	100	99.41	42.14	0.58	75.84
			(6.22)	(0.00)	(0.01)	(0.21)	(0.00)	(0.14)
		10%	1746.64	100	99.39	42.89	0.60	75.34
			(5.89)	(0.00)	(0.00)	(0.19)	(0.00)	(0.13)
		15%	1757.10	100	99.39	43.23	0.61	75.11
			(5.73)	(0.00)	(0.00)	(0.18)	(0.00)	(0.12)
		20%	1770.48	100	99.37	43.66	0.62	74.82
			(5.29)	(0.00)	(0.00)	(0.17)	(0.00)	(0.11)
		30%	1811.62	100	99.34	44.94	0.65	73.95
			(5.22)	(0.00)	(0.00)	(0.16)	(0.00)	(0.11)
	30%	0%	1455.84	100	99.63	31.49	0.37	82.61
			(4.00)	(0.00)	(0.00)	(0.19)	(0.00)	(0.11)
		10%	1534.22	100	99.57	34.97	0.43	80.46
			(5.92)	(0.00)	(0.00)	(0.25)	(0.00)	(0.16)
		15%	1556.14	100	99.55	35.90	0.45	79.88
			(5.28)	(0.00)	(0.00)	(0.22)	(0.00)	(0.14)
		20%	1603.54	100	99.51	37.79	0.49	78.67
			(5.17)	(0.00)	(0.00)	(0.20)	(0.00)	(0.13)
		30%	1668.40	100	99.46	40.22	0.54	77.11
			(4.81)	(0.00)	(0.00)	(0.17)	(0.00)	(0.11)

Table I.2. The averages of $|\hat{E}|$, TPR, TNR, FDR, MISR, and MCC for hub and scale-free networks over 50 datasets ($p = 500, n = 250$). "Info." and "FP(Info.)" stand for the percentage of pre-identified edges and the percentage of false positives in the pre-identified edges, respectively. $|\hat{E}|$ denotes the number of estimated edges. All values except for $|\hat{E}|$ are multiplied by 100. The numbers in parentheses denote the standard errors of measures.

Network	Info.	FP(Info.)	$\lvert\hat{E}\rvert$	TPR	TNR	FDR	MISR	MCC
Hub ($\lvert E\rvert = 569$)	None		586.38	87.05	99.93	15.43	0.13	85.71
			(3.34)	(0.18)	(0.00)	(0.37)	(0.00)	(0.17)
	10%	0%	574.96	87.45	99.94	13.39	0.12	86.96
			(2.74)	(0.18)	(0.00)	(0.31)	(0.00)	(0.14)
		10%	584.94	87.57	99.93	14.73	0.13	86.34
			(3.12)	(0.18)	(0.00)	(0.33)	(0.00)	(0.14)
		15%	591.66	87.72	99.93	15.55	0.13	85.99
			(3.30)	(0.19)	(0.00)	(0.35)	(0.00)	(0.15)
		20%	594.20	87.54	99.92	16.10	0.13	85.62
			(2.99)	(0.18)	(0.00)	(0.31)	(0.00)	(0.13)
		30%	599.06	87.27	99.92	17.05	0.14	85.00
			(2.80)	(0.17)	(0.00)	(0.29)	(0.00)	(0.14)
	30%	0%	564.5	88.56	99.95	10.68	0.1	88.88
			(2.54)	(0.17)	(0.00)	(0.29)	(0.00)	(0.13)
		10%	582.18	88.34	99.94	13.59	0.12	87.30
			(2.81)	(0.19)	(0.00)	(0.28)	(0.00)	(0.11)
		15%	585.70	87.74	99.93	14.71	0.13	86.43
			(2.76)	(0.19)	(0.00)	(0.26)	(0.00)	(0.11)
		20%	597.48	87.85	99.92	16.28	0.13	85.68
			(2.71)	(0.18)	(0.00)	(0.28)	(0.00)	(0.13)
		30%	617.58	87.83	99.91	19.04	0.15	84.24
			(2.46)	(0.18)	(0.00)	(0.22)	(0.00)	(0.11)
Scale − free ($\lvert E\rvert = 495$)	None		526.1	89.28	99.93	15.87	0.11	86.59
			(3.30)	(0.18)	(0.00)	(0.42)	(0.00)	(0.20)
	10%	0%	518.12	89.57	99.94	14.31	0.1	87.54
			(3.12)	(0.17)	(0.00)	(0.41)	(0.00)	(0.19)
		10%	524.40	89.68	99.94	15.25	0.11	87.11
			(2.99)	(0.16)	(0.00)	(0.39)	(0.00)	(0.18)
		15%	528.24	89.79	99.93	15.77	0.11	86.90
			(2.87)	(0.16)	(0.00)	(0.37)	(0.00)	(0.17)
		20%	532.54	89.79	99.93	16.44	0.11	86.54
			(2.99)	(0.17)	(0.00)	(0.38)	(0.00)	(0.18)
		30%	537.84	89.75	99.92	17.29	0.12	86.08
			(3.15)	(0.16)	(0.00)	(0.39)	(0.00)	(0.18)
	30%	0%	500.22	90.47	99.96	10.37	0.08	89.99
			(2.90)	(0.18)	(0.00)	(0.39)	(0.00)	(0.17)
		10%	516.78	90.45	99.94	13.27	0.09	88.51
			(2.92)	(0.19)	(0.00)	(0.36)	(0.00)	(0.15)
		15%	526.12	90.39	99.94	14.86	0.10	87.66
			(3.02)	(0.19)	(0.00)	(0.36)	(0.00)	(0.16)
		20%	532.84	89.89	99.93	16.40	0.11	86.62
			(2.90)	(0.18)	(0.00)	(0.35)	(0.00)	(0.17)
		30%	547.14	89.63	99.92	18.82	0.12	85.22
			(3.09)	(0.19)	(0.00)	(0.34)	(0.00)	(0.16)

Table J.1. The averages of $|\hat{E}|$, TPR, TNR, FDR, MISR, and MCC for hub and scale-free networks over 50 datasets. $|\hat{E}|$ denotes the number of estimated edges. All values except for $|\hat{E}|$ are multiplied by 100. The numbers in parentheses denote the standard errors of measures.

Network	n	Info.	$\lvert\hat{E}\rvert$	TPR	TNR	FDR	MISR	MCC
Hub ($\lvert E\rvert = 1167$)	200	None	988.70	74.60	99.98	11.87	0.08	81.03
			(5.63)	(0.21)	(0.00)	(0.30)	(0.00)	(0.10)
		10%	975.18	75.29	99.98	9.82	0.08	82.34
			(5.69)	(0.21)	(0.00)	(0.31)	(0.00)	(0.10)
		30%	993.24	79.14	99.99	6.96	0.06	85.77
			(5.05)	(0.23)	(0.00)	(0.24)	(0.00)	(0.09)
	300	None	1199.18	88.55	99.97	13.77	0.06	87.34
			(5.41)	(0.13)	(0.00)	(0.30)	(0.00)	(0.12)
		10%	1182.06	88.89	99.97	12.19	0.05	88.31
			(4.92)	(0.14)	(0.00)	(0.27)	(0.00)	(0.11)
		30%	1162.38	90.42	99.98	9.19	0.04	90.59
			(3.77)	(0.12)	(0.00)	(0.20)	(0.00)	(0.08)
	500	None	1311.30	96.80	99.96	13.80	0.04	91.32
			(4.91)	(0.07)	(0.00)	(0.31)	(0.00)	(0.16)
		10%	1288.88	96.76	99.97	12.34	0.04	92.07
			(4.76)	(0.08)	(0.00)	(0.29)	(0.00)	(0.15)
		30%	1262.20	97.24	99.97	10.05	0.03	93.50
			(4.20)	(0.07)	(0.00)	(0.28)	(0.00)	(0.15)
Scale-free ($\lvert E\rvert = 990$)	200	None	947.78	81.92	99.97	14.37	0.06	83.71
			(4.54)	(0.15)	(0.00)	(0.30)	(0.00)	(0.12)
		10%	955.48	83.45	99.97	13.50	0.06	84.92
			(3.30)	(0.12)	(0.00)	(0.23)	(0.00)	(0.11)
		30%	935.04	85.12	99.98	9.84	0.05	87.58
			(3.22)	(0.11)	(0.00)	(0.23)	(0.00)	(0.10)
	300	None	1033.74	89.73	99.97	14.03	0.05	87.80
			(3.62)	(0.12)	(0.00)	(0.24)	(0.00)	(0.12)
		10%	1026.56	90.29	99.97	12.89	0.05	88.65
			(3.37)	(0.11)	(0.00)	(0.24)	(0.00)	(0.12)
		30%	999.42	91.08	99.98	9.74	0.04	90.65
			(3.36)	(0.11)	(0.00)	(0.23)	(0.00)	(0.10)
	500	None	1103.76	96.41	99.97	13.47	0.04	91.31
			(4.02)	(0.07)	(0.00)	(0.29)	(0.00)	(0.14)
		10%	1092.44	96.50	99.97	12.52	0.03	91.86
			(3.18)	(0.07)	(0.00)	(0.21)	(0.00)	(0.10)
		30%	1055.60	96.57	99.98	9.40	0.03	93.52
			(2.99)	(0.07)	(0.00)	(0.23)	(0.00)	(0.11)



(a) Tomida et al. (2009), n=117

(b) Bhattacharjee et al. (2001) n=203

(c) Raponi et al. (2006), n=129

(d) Jones et al. (2004), n=80

Fig. G.1. Kaplan-Meier curves for the PTPN13 gene from four datasets (Bhattacharjee et al., 2001; Jones et al., 2004; Raponi et al., 2006; Tomida et al., 2009). For each dataset, we divide patients into two groups, "High" and "Low", by their PTPN13 gene expression levels. We consider a patient to have a high expression level of PTPN13 if the expression level is greater than or equal to a median of expression levels. Red solid lines (—) denote the "Low" group and black dashed lines (- -) denote the "High" group. The p-values in the figures are from the log-rank test to compare two survival distributions.