Biometrics 000, S1–S10 DOI: 000
000 0000

Supplementary Materials for "A Significance Test for Graph-Constrained Estimation"

Sen Zhao

Department of Biostatistics, University of Washington, Seattle, Washington, U.S.A.

email: senz@u.washington.edu

and

Ali Shojaie

Department of Biostatistics, University of Washington, Seattle, Washington, U.S.A.

email: ashojaie@u.washington.edu

This paper has been submitted for consideration for publication in *Biometrics*

S1. Proof of Lemma 1

Proof. Given that $(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})$ is invertible and $h > 0$, we have

$$\begin{aligned}\mathbf{Bias}\big(\hat{\boldsymbol{\beta}}(h)\big|\boldsymbol{X}\big) &= \mathrm{E}\big(\hat{\boldsymbol{\beta}}(h)\big|\boldsymbol{X}\big) - \boldsymbol{\beta}^* \\ &= (n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1} n\hat{\boldsymbol{\Sigma}}\boldsymbol{\beta}^* - (n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1}(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})\boldsymbol{\beta}^* \\ &= -(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1} h\boldsymbol{L}\boldsymbol{\beta}^*,\end{aligned}$$

which is equal to $\mathbf{0}$ if and only if $\boldsymbol{L}\boldsymbol{\beta}^* = \mathbf{0}$. We know that

$$(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1} \preceq \frac{1}{\lambda_0(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})}\boldsymbol{I}.$$

Therefore,

$$\begin{aligned}\big\|\mathbf{Bias}(\hat{\boldsymbol{\beta}}(h))\big|\boldsymbol{X}\big\|_2 &= h\sqrt{(\boldsymbol{L}\boldsymbol{\beta}^*)^\top (n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-2}(\boldsymbol{L}\boldsymbol{\beta}^*)} \\ &\leqslant h\sqrt{(\boldsymbol{L}\boldsymbol{\beta}^*)^\top \frac{1}{\lambda_0(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^2}(\boldsymbol{L}\boldsymbol{\beta}^*)} \\ &= \frac{h\|\boldsymbol{L}\boldsymbol{\beta}^*\|_2}{\lambda_0(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})}.\end{aligned}$$

S2. Proof of Theorem 1

Proof. Under the null hypothesis $H_0 : \beta_j^* = 0$, we have

$$\begin{aligned}\big|\gamma_j^G\big| &= h\big|(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1}\boldsymbol{L}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\big|_j \\ &= h\Big|\sum_{i=1}^{p}\big[(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1}\boldsymbol{L}\big]_{(j,i)}(\tilde{\beta}_i - \beta_i^*)\Big| \\ &\leqslant h\Big|\sum_{i:i\neq j}\big[(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1}\boldsymbol{L}\big]_{(j,i)}(\tilde{\beta}_i - \beta_i^*)\Big| + h\Big|\big[(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1}\boldsymbol{L}\big]_{(j,j)}\tilde{\beta}_j\Big| \\ &\leqslant h\big\|\big[(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1}\boldsymbol{L}\big]_{(j,-j)}\big\|_\infty\big\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\big\|_1 + h\Big|\big[(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1}\boldsymbol{L}\big]_{(j,j)}\tilde{\beta}_j\Big|\end{aligned}$$

Based on Bühlmann and van de Geer (2011), Chapter 6.12, with Gaussian design, if the $\boldsymbol{\Sigma}$-compatibility condition is met for the set S_0 with compatibility constant $\phi_{\boldsymbol{\Sigma}}$, with probability tending to 1, the condition is also met for $\hat{\boldsymbol{\Sigma}}$ with compatibility constant $\phi_{\hat{\boldsymbol{\Sigma}}} > \phi_{\boldsymbol{\Sigma}}/2$.

Moroever, with $h_{Lasso} \asymp \sqrt{\log p/n}$ and the $\hat{\boldsymbol{\Sigma}}$-compatibility condition for the set S_0, with probability tending to 1, we have

$$\left\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_1 \leqslant 4\frac{h_{Lasso} s_0}{\phi^2_{\hat{\boldsymbol{\Sigma}}}}.$$

Then, because $s_0 = o([n/\log p]^{\xi})$ and $\liminf \phi^2_{\hat{\boldsymbol{\Sigma}}} > d/2 > 0$, we get

$$\left\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_1 = o_p\left(\left(\frac{\log p}{n}\right)^{\frac{1}{2}-\xi}\right).$$

On the other hand, by Assumption A4, $\left((n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1} h\boldsymbol{L}\right)_{(j,j)} = \mathcal{O}_p\big((n/\log p)^{1/2-\xi}\big)$. Thus

$$h\left|\left[(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1}\boldsymbol{L}\right]_{(j,j)}\tilde{\beta}_j\right| = \left|\left[(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1} h\boldsymbol{L}\right]_{(j,j)}\right|\left|\tilde{\beta}_j - \beta_j^*\right| = o_p(1),$$

and hence

$$Pr\left(\left|\gamma_j^G\right| \leqslant h\left\|[(n\hat{\boldsymbol{\Sigma}} + h\boldsymbol{L})^{-1}\boldsymbol{L}]_{(j,-j)}\right\|_{\infty}\left(\frac{\log p}{n}\right)^{\frac{1}{2}-\xi}\right) \to 1,$$

where the right hand side is Γ_j^G. We can thus write

$$\begin{aligned}\left|\hat{\boldsymbol{z}}_j^G\right| &= \left|Z_j^G + \gamma_j^G\right| \\ &\leqslant \left|Z_j^G\right| + \left|\gamma_j^G\right| \\ &\precsim^{asy.} \left|Z_j^G\right| + \Gamma_j^G.\end{aligned}$$

S3. Proof of Theorem 2

Proof. Given (12), conditional on $\boldsymbol{X}$, the objective of $P_j^G \leqslant \alpha$ is satisfied if $\left|\hat{\boldsymbol{z}}_j^G\right| \geqslant \Gamma_j^G + q_{(1-\alpha/2)}\sqrt{\mathrm{Var}(Z_j^G|\boldsymbol{X})}$. According to Equation (6), this is equivalent of $\left|\beta_j^* + Z_j^G + \gamma_j^G\right| \geqslant \Gamma_j^G + q_{(1-\alpha/2)}\sqrt{\mathrm{Var}(Z_j^G|\boldsymbol{X})}$, which is satisfied if

$$\left|\beta_j^*\right| - \left|\gamma_j^G\right| - \left|Z_j^G\right| \geqslant \Gamma_j^G + q_{(1-\alpha/2)}\sqrt{\mathrm{Var}(Z_j^G|\boldsymbol{X})}.$$

This holds with probability at least ψ if

$$\left|\beta_j^*\right| - \left|\gamma_j^G\right| \geqslant \Gamma_j^G + q_{(1-\alpha/2)}\sqrt{\mathrm{Var}(Z_j^G|\boldsymbol{X})} + q_{(1-\psi/2)}.$$

We know that with probability tending to 1, $\left|\gamma_j^G\right| \leqslant \Gamma_j^G$. Therefore, conditional on $\boldsymbol{X}$, we

have $P_j^G \leqslant \alpha_L$ with probability tending to at least ψ, if

$$\left|\beta_j^*\right| > 2\Gamma_j^G + q_{(1-\alpha/2)}\sqrt{\mathrm{Var}(Z_j^G|\boldsymbol{X})} + q_{(1-\psi/2)}.$$

S4. Proof of Theorem 3

Proof. **a)** We note that $P_1^G/P_1^{GI} \leqslant 1$ is equivalent of

$$\frac{\left(\left|\hat{\boldsymbol{z}}_1^{GI}\right| - \Gamma_1^{GI}\right)_+ / \sqrt{\mathrm{Var}(Z_1^{GI}|\boldsymbol{X})}}{\left(\left|\hat{\boldsymbol{z}}_1^{G}\right| - \Gamma_1^{G}\right)_+ / \sqrt{\mathrm{Var}(Z_1^{G}|\boldsymbol{X})}} \leqslant 1.$$

We first write out those components for the Grace test:

$$\begin{aligned}
\hat{\boldsymbol{z}}_1^G &= \left((\boldsymbol{X}^\top\boldsymbol{X} + h_n^G\boldsymbol{L})^{-1}(\boldsymbol{X}^\top\boldsymbol{y} + h_n^G\boldsymbol{L}\tilde{\boldsymbol{\beta}})\right)_1 \\
&= \frac{(n+h_n^G)\boldsymbol{x}_1^\top\boldsymbol{y} - (n\rho + h_n^G l)\boldsymbol{x}_2^\top\boldsymbol{y} + h_n^G\tilde{\beta}_1(n + h_n^G - n\rho l - h_n^G l^2) + nh_n^G\tilde{\beta}_2(l-\rho)}{(n+h_n^G)^2 - (n\rho + h_n^G l)^2}; \\
\Gamma_1^G &= \left|h_n^G\left[(\boldsymbol{X}^\top\boldsymbol{X} + h_n^G\boldsymbol{L})^{-1}\boldsymbol{L}\right]_{(1,-1)}\right|\left(\frac{\log p}{n}\right)^{\frac{1}{2}-\xi} \\
&= \left|h_n^G\left[(\boldsymbol{X}^\top\boldsymbol{X} + h_n^G\boldsymbol{L})^{-1}\boldsymbol{L}\right]_{(1,2)}\right|\left(\frac{\log p}{n}\right)^{\frac{1}{2}-\xi} \\
&= \frac{|nh_n^G l - nh_n^G\rho|}{(n+h_n^G)^2 - (n\rho + h_n^G l)^2}\left(\frac{\log p}{n}\right)^{\frac{1}{2}-\xi}; \\
\mathrm{Var}(Z_1^G|\boldsymbol{X}) &= \sigma_{\boldsymbol{\epsilon}}^2\left[(\boldsymbol{X}^\top\boldsymbol{X} + h_n^G\boldsymbol{L})^{-1}\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X} + h_n^G\boldsymbol{L})^{-1}\right]_{(1,1)} \\
&= \sigma_{\boldsymbol{\epsilon}}^2\frac{(n^3 + 2h_n^G n^2)(1-\rho^2) + n(h_n^G)^2(1 + l^2 - 2l\rho)}{[(n+h_n^G)^2 - (n\rho + h_n^G l)^2]^2}.
\end{aligned}$$

We can also write out those components for the GraceI test likewise with $l = 0$.

In the proof of Theorem 1, we have shown that $Pr\left(\left\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_1 \leqslant 4h_{Lasso}s_0/\phi_{\hat{\boldsymbol{\Sigma}}}^2\right) \to 1$. With $h_{Lasso} = \mathcal{O}(\log p/n)$, $s_0 = o([n/\log p]^\xi)$ for some $0 \leqslant \xi < 1/2$, $\liminf \phi_{\hat{\boldsymbol{\Sigma}}} > d/2 > 0$, and $p = \mathcal{O}(\exp(n^\nu))$ for some $0 \leqslant \nu < 1$, we have $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 = o_p(1)$. Thus we get

$$\tilde{\beta}_1 = \beta_1^* + o_p(1), \qquad \tilde{\beta}_2 = \beta_2^* + o_p(1).$$

We also note that since our design matrix is scaled, we get

$$\boldsymbol{x}_1^\top\boldsymbol{y} = \boldsymbol{x}_1^\top\boldsymbol{x}_1\beta_1^* + \boldsymbol{x}_1^\top\boldsymbol{x}_2\beta_2^* + \boldsymbol{x}_1^\top\boldsymbol{\epsilon} = n\beta_1^* + n\rho\beta_2^* + nE,$$

$$\boldsymbol{x}_2^\top\boldsymbol{y} = \boldsymbol{x}_2^\top\boldsymbol{x}_1\beta_1^* + \boldsymbol{x}_2^\top\boldsymbol{x}_2\beta_2^* + \boldsymbol{x}_2^\top\boldsymbol{\epsilon} = n\rho\beta_1^* + n\beta_2^* + nE,$$

where $E \sim N(\mathbf{0}, \sigma_{\boldsymbol{\epsilon}}^2/n) = \mathcal{O}_p(1)$.

Define $k_n^G \triangleq h_n^G/n$ and $k_n^{GI} \triangleq h_n^{GI}/n$. With some algebra, We get

$$\frac{\left(|\hat{\boldsymbol{z}}_1^G| - \Gamma_1^G\right)_+}{\sqrt{\mathrm{Var}(Z_1^G|\boldsymbol{X})}} = \frac{\sqrt{n}\left[|(k_n^G+1)^2 - (\rho + lk_n^G)^2 + \mathcal{O}_p(1)| \cdot |\beta_1^*| - (\log p/n)^{1/2-\xi} \cdot |k_n^G(l-\rho)|\right]_+}{\sigma_{\boldsymbol{\epsilon}}\sqrt{(1+2k_n^G)(1-\rho^2) + (k_n^G)^2(1+l^2-2l\rho)}}. \tag{S1}$$

Similarly for the GraceI, we get

$$\frac{\left(|\hat{\boldsymbol{z}}_1^{GI}| - \Gamma_1^{GI}\right)_+}{\sqrt{\mathrm{Var}(Z_1^{GI}|\boldsymbol{X})}} = \frac{\sqrt{n}\left[|(k_n^{GI}+1)^2 - \rho^2 + \mathcal{O}_p(1)| \cdot |\beta_1^*| - (\log p/n)^{1/2-\xi} \cdot |k_n^{GI}\rho|\right]_+}{\sigma_{\boldsymbol{\epsilon}}\sqrt{(1+2k_n^{GI})(1-\rho^2) + (k_n^{GI})^2}}. \tag{S2}$$

We observe that $k_n^{GI} + 1 > 1 \geqslant |\rho|$ and $k_n^G + 1 \geqslant |l|k_n^G + |\rho| \geqslant |\rho + lk_n^G|$. We plug in those two inequalities into Equation (S1) and (S2). Hence, conditional on the design matrix $\boldsymbol{X}$, $P_1^G/P_1^{GI} \leqslant 1$ with probability tending to 1 if

$$\begin{aligned}
&\lim_{n\to\infty} \frac{\left\{\left[(k_n^G+1)^2 - (\rho+lk_n^G)^2\right] \cdot |\beta_1^*| - (\log p/n)^{1/2-\xi} \cdot |k_n^G(l-\rho)|\right\}_+}{\sqrt{(1+2k_n^G)(1-\rho^2) + (k_n^G)^2(1+l^2-2l\rho)}} \\
\geqslant &\lim_{n\to\infty} \frac{\left\{\left[(k_n^{GI}+1)^2 - \rho^2\right] \cdot |\beta_1^*| - (\log p/n)^{1/2-\xi} \cdot |k_n^{GI}\rho|\right\}_+}{\sqrt{(1+2k_n^{GI})(1-\rho^2) + (k_n^{GI})^2}}.
\end{aligned}$$

Note that for any two real numbers f and g, $f \geqslant g$ implies $f_+ \geqslant g_+$. Thus, conditional on the design matrix $\boldsymbol{X}$, $P_1^G/P_1^{GI} \leqslant 1$ with probability tending to 1 if

$$\begin{aligned}
&\lim_{n\to\infty} \frac{\left[(k_n^G+1)^2 - (\rho+lk_n^G)^2\right] \cdot |\beta_1^*| - (\log p/n)^{1/2-\xi} \cdot |k_n^G(l-\rho)|}{\sqrt{(1+2k_n^G)(1-\rho^2) + (k_n^G)^2(1+l^2-2l\rho)}} \\
\geqslant &\lim_{n\to\infty} \frac{\left[(k_n^{GI}+1)^2 - \rho^2\right] \cdot |\beta_1^*| - (\log p/n)^{1/2-\xi} \cdot |k_n^{GI}\rho|}{\sqrt{(1+2k_n^{GI})(1-\rho^2) + (k_n^{GI})^2}}.
\end{aligned} \tag{S3}$$

If we assume $k_n^G = k_n^{GI} = k \to \infty$, Inequality (S3) is satisfied if

$$\begin{aligned}
&\lim_{n\to\infty} \frac{\left[(k+1)^2 - (\rho+lk)^2\right] \cdot |\beta_1^*| - (\log p/n)^{1/2-\xi} \cdot |k(l-\rho)|}{\left[(k+1)^2 - \rho^2\right] \cdot |\beta_1^*| - (\log p/n)^{1/2-\xi} \cdot |k\rho|} \\
\times &\frac{\sqrt{(1+2k)(1-\rho^2) + k^2}}{\sqrt{(1+2k)(1-\rho^2) + k^2(1+l^2-2l\rho)}} \\
= &\lim_{n\to\infty} \frac{\left[(1-l^2) + (2-2l\rho)/k + (1-\rho^2)/k^2\right] \cdot |\beta_1^*| - (\log p/n)^{1/2-\xi} \cdot |(l-\rho)/k|}{\left[1 + 2/k + (1-\rho^2)/k^2\right] \cdot |\beta_1^*| - (\log p/n)^{1/2-\xi} \cdot |\rho/k|} \\
\times &\frac{\sqrt{1 + (2-2\rho^2)/k + (1-\rho^2)/k^2}}{\sqrt{(1+l^2-2l\rho) + (2-2\rho^2)/k + (1-\rho^2)/k^2}} \\
= &\frac{(1-l^2)}{\sqrt{(1+l^2-2l\rho)}} \geqslant 1.
\end{aligned} \tag{S4}$$

The last equality holds because $p = \mathcal{O}(\exp(n^\nu))$ for some $0 \leqslant \nu < 1$ implies that $\log p/n \to 0$.

For the ridge test, we assume $h_n^R = \mathcal{O}(1)$. Thus with some algebra we can similarly write out the ridge test objective:

$$\frac{|\hat{\boldsymbol{z}}_1^R|}{\sqrt{\mathrm{Var}(Z_1^R|\boldsymbol{X})}} = \frac{\sqrt{n}|1-\rho^2+o_p(1)|\cdot|\beta_1^*|}{\sigma_{\boldsymbol{\epsilon}}\sqrt{(1-\rho^2)+o(1)}}. \tag{S5}$$

b) Thus, conditional on $\boldsymbol{X}$, we get $P_1^G/P_1^R \leqslant 1$ with probability tending to 1 if

$$\lim_{n\to\infty} \frac{\left((k_n^G+1)^2-(\rho+lk_n^G)^2\right)\cdot|\beta_1^*| - (\log p/n)^{1/2-\xi}\cdot|k_n^G(l-\rho)|}{\sqrt{(1+2k_n^G)(1-\rho^2)+(k_n^G)^2(1+l^2-2l\rho)}} \geqslant \sqrt{1-\rho^2}\cdot|\beta_1^*|. \tag{S6}$$

c) We also have$P_1^{GI}/P_1^R \leqslant 1$ with probability tending to 1 if

$$\lim_{n\to\infty} \frac{\left((k_n^{GI}+1)^2-\rho^2\right)\cdot|\beta_1^*| - (\log p/n)^{1/2-\xi}\cdot|k_n^{GI}\rho|}{\sqrt{(1+2k_n^{GI})(1-\rho^2)+(k_n^{GI})^2}} \geqslant \sqrt{1-\rho^2}\cdot|\beta_1^*|. \tag{S7}$$

S5. Illustration of the Graph Structure in the Simulation Study

Figure S1 shows the graph structure used in the simulation study with 5 hub-satellite clusters. In the simulation study, we use 50 such hub-satellite clusters.

[Figure 1 about here.]

S6. Additional Details for Analysis of TCGA Data

S6.1 *Biological Evidence*

In this section, we summarize some of the biological evidences in support of the association between genes identified by the Grace and GraceR tests with the onset, progression and severity of prostate cancer, as well as PSA level.

As pointed out in the main paper, the Grace and GraceR tests identify a number of histone genes and histone deacetylase (HDAC) genes. Previous research indicates that histone genes are associated with the occurrence, clinical outcomes and recurrence of prostate cancer (Seligson et al., 2005; Ke et al., 2009). The pathological role of HDAC genes on the onset and progression of prostate cancer have also been previously studied (Halkidou et al., 2004; Chen et al., 2007; Abbas and Gupta, 2008).

In addition to the highly connected histone and HDAC genes, the GraceR test also identifies some disconnected genes. Prior works shows that the expression of ribonucleoside-diphosphate reductase subunit M2 (RRM2) is associated with higher Gleason scores, which correlate with the severity of prostate cancer (Huang et al., 2014). Protein arginine methyl-transferase 1 (PRMT1) may also have an effect on the proliferation of prostate cancer cells (Yu et al., 2009). Activation of olfactory receptors (OR) prevents proliferation of prostate cancer cells (Neuhaus et al., 2009). Interferon-γ (IFNG) plays a role in the differentiation of human prostate basal-epithelial cells (Untergasser et al., 2005). IFNG is connected to the interleukin receptor 22 α1 (IL22RA1), the role of which related to prostate cancer is unknown. However, several earlier studies point out the associations between prostate cancer and several other interleukin receptors in the Janus kinase and signal transducer and activator of transcription (JAK-STAT) activating family, including IL 6, 8, 11, 13 and 17 genes(Culig et al., 2005; Inoue et al., 2000; Campbell et al., 2001; Maini et al., 1997; Zhang et al., 2012). Cell-division cycle genes (CDC) may also be associated with various cancers. The association between collagen type 2 α1 (COL2A1) and prostate cancer is also not known, but other collagen genes, including type 1 $\alpha 2\beta 1$, type 4 $\alpha 5$ and $\alpha 6$, have been shown to be associated with prostate cancer progression (Hall et al., 2008; Dehan et al., 1997). Although the association between phosphate cytidylyltransferase 1 choline-α (PCYT1A) and prostate cancer or PSA level is not known, Vaezi et al. (2014) shows that PCYT1A is a prognostic factor in survival for patients with lung and head and neck squamous cell carcinomas.

S6.2 *Stability of the Grace Test to the Tuning Parameter*

Figure S2 shows the number of significant genes identified by the Grace test in the TCGA data against various values of h_G. The results indicate that the number of genes found by the Grace test is relatively stable for a range of tuning parameters including the CV choice. On the other hand, very few genes are identified when the tuning parameter is too small or too

large. This is because, with small tuning parameters, the variance is large and thus no gene is statistically significant. On the other hand, with large tuning parameters, the stochastic bound Γ_j dominates $\hat{z}_j$. Note that above results of power do not contradict Theorem 3, which shows the *asymptotic* power of the Grace test improves as we use larger h_G. A vital condition for Theorem 3 to hold is $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 = \mathcal{O}_p(1)$.

[Figure 2 about here.]

S6.3 *Stability of the Grace Test to the Network*

We examine whether the result of the Grace test on the TCGA data is sensitive to the KEGG network structure. To this end, we randomly change the connectivity of m node pairs in the KEGG network and form the new perturbed network $\tilde{G}$, $|E\Delta\tilde{E}| = m$, where Δ is the symmetric difference operator between two sets. In other words, for m randomly selected node pairs (a_i, b_i), $i = 1, ..., m$, if there is an edge (a_i, b_i) in the KEGG network, we remove it in the perturbed network; otherwise, we add an edge in the perturbed network. In our examination, m ranges from $10,000$ to $600,000$. Note that there are 38,541 edges in the original KEGG network. We counted the number of genes that are significant using both networks. The result shown in Figure S3 is an average of 50 independent replications.

[Figure 3 about here.]

S6.4 *Prediction Performance*

We also compare the prediction performance by Grace, GraceR, GraceI and lasso with tuning parameters chosen by 10-fold CV, as well as ridge with $h_2 = 1$. The result is shown in Table S1. GraceR produced the smallest CV prediction error, followed closely by GraceI and Grace. This result may indicate the KEGG network information is in fact informative in prediction.

[Table 1 about here.]

S7. Additional Simulation Studies with Extended NPE

We performed simulation studies with extended NPE $\in$ {-225, -165, -70, -10, 0, 15, 135, 350, 600, 900, 1250, 1650, 2050, 3150}. These perturbations in the network correspond to the spectral norm of perturbations $\|\boldsymbol{L} - \boldsymbol{L}^*\|_2/\|\boldsymbol{L}^*\|_2$ equal 0.85, 0.75, 0.50, 0.25, 0, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00 and 2.65, respectively. The power and type-I error rates are summarized in Figure S4, Table S2 and Table S3. Our conclusions on the simulation study stated in the main paper do not change with this expanded version of simulation study.

[Figure 4 about here.]

[Table 2 about here.]

[Table 3 about here.]

References

Abbas, A. and Gupta, S. (2008). The role of histone deacetylases in prostate cancer. *Epigenetics* **3,** 300–309.

Bühlmann, P. and van de Geer, S. (2011). *Statistics for High-dimensional Data: Methods, Theory and Applications. Springer Series in Statistics*. Springer.

Campbell, C., Jiang, Z., Savarese, D., and Savarese, T. (2001). Increased expression of the interleukin-11 receptor and evidence of STAT3 activation in prostate carcinoma. *The American Journal of Pathology* **158,** 25–32.

Chen, C.-S., Wang, Y.-C., Yang, H.-C., Huang, P.-H., Kulp, S., Yang, C.-C., Lu, Y.-S., Matsuyama, S., Chen, C.-Y., and Chen, C.-S. (2007). Histone deacetylase inhibitors sensitize prostate cancer cells to agents that produce DNA double-strand breaks by targeting Ku70 acetylation. *Cancer Research* **67,** 5318–5327.

Culig, Z., Steiner, H., Bartsch, G., and Hobisch, A. (2005). Interleukin-6 regulation of prostate cancer cell growth. *Journal of Cellular Biochemistry* **95,** 497–505.

Dehan, P., Waltregny, D., Beschin, A., Noel, A., Castronovo, V., Tryggvason, K., De Leval, J., and Foidart, J.-M. (1997). Loss of type IV collagen $\alpha 5$ and $\alpha 6$ chains in human invasive prostate carcinomas. *The American Journal of Pathology* **151,** 1097–1104.

Halkidou, K., Gaughan, L., Cook, S., Leung, H., Neal, D., and Robson, C. (2004). Upregulation and nuclear recruitment of HDAC1 in hormone refractory prostate cancer. *The Prostate* **59,** 177–189.

Hall, C., Dubyk, C., Riesenberger, T., Shein, D., Keller, E., and van Golen, K. (2008). Type I collagen receptor ($\alpha 2\beta 1$) signaling promotes prostate cancer invasion through RhoC GTPase. *Neoplasia* **10,** 797–803.

Huang, Y., Liu, X., Wang, Y.-H., Yeh, S.-D., Chen, C.-L., Nelson, R., Chu, P., Wilson, T., and Yen, Y. (2014). The prognostic value of ribonucleotide reductase small subunit M2 in predicting recurrence for prostate cancers. *Urologic Oncology* **32,** 51.e9–51.e19.

Inoue, K., Slaton, J., Eve, B., Kim, S., Perrotte, P., Balbay, M., Yano, S., Bar-Eli, M., Radinsky, R., Pettaway, C., and Dinney, C. (2000). Interleukin 8 expression regulates tumorigenicity and metastases in androgen-independent prostate cancer. *Clinical Cancer Research* **6,** 2104–2119.

Ke, X.-S., Qu, Y., Rostad, K., Li, W.-C., Lin, B., Halvorsen, O., Haukaas, S., Jonassen, I., Petersen, K., Goldfinger, N., Rotter, V., Akslen, L., Oyan, A., and Kalland, K.-H. (2009). Genome-wide profiling of histone H3 lysine 4 and lysine 27 trimethylation reveals an epigenetic signature in prostate carcinogenesis. *PlOS ONE* **4,** e4687.

Maini, A., Hillman, G., Haas, G., Wang, C., Montecillo, E., Hamzavi, F., Pontes, E., Leland, P., Pastan, I., Debinski, W., and Puri, R. (1997). Interleukin-13 receptors on human prostate carcinoma cell lines represent a novel target for a chimeric protein composed of IL-13 and a mutated form of Pseudomonas exotoxin. *The Journal of Urology* **158,** 948–953.

Neuhaus, E., Zhang, W., Gelis, L., Deng, Y., Noldus, J., and Hatt, H. (2009). Activation of an olfactory receptor inhibits proliferation of prostate cancer cells. *The Journal of Biological Chemistry* **284,** 16218–16225.

Seligson, D., Horvath, S., Shi, T., Yu, H., Tze, S., Grunstein, M., and Kurdistani, S. (2005). Global histone modification patterns predict risk of prostate cancer recurrence. *Nature* **435,** 1262–1266.

Untergasser, G., Plas, E., Pfister, G., Heinrich, E., and Berger, P. (2005). Interferon-γ induces neuroendocrine-like differentiation of human prostate basal-epithelial cells. *The Prostate* **64,** 419–429.

Vaezi, A. E., Bepler, G., Bhagwat, N. R., Malysa, A., Rubatt, J. M., Chen, W., Hood, B. L., Conrads, T. P., Wang, L., Kemp, C. E., and Niedernhofer, L. J. (2014). Choline phosphate cytidylyltransferase-α is a novel antigen detected by the anti-ercc1 antibody 8f1 with biomarker value in patients with lung and head and neck squamous cell carcinomas. *Cancer* **120,** 1898–1907.

Yu, Z., Chen, T., Hebert, J., Li, E., and Richard, S. (2009). A mouse PRMT1 null allele defines an essential role for arginine methylation in genome maintenance and cell proliferation. *Molecular and Cellular Biology* **29,** 2982–2996.

Zhang, Q., Liu, S., Ge, D., Zhang, Q., Xue, Y., Xiong, Z., Abdel-Mageed, A., Myers, L., Hill, S., Rowan, B., Sartor, O., Melamed, J., Chen, Z., and You, Z. (2012). Interleukin-17 promotes formation and growth of prostate adenocarcinoma in mouse models. *Cancer Research* **72,** 2589–2599.

Figure S1. An illustration of the graph structure with 5 hub-satellite clusters.

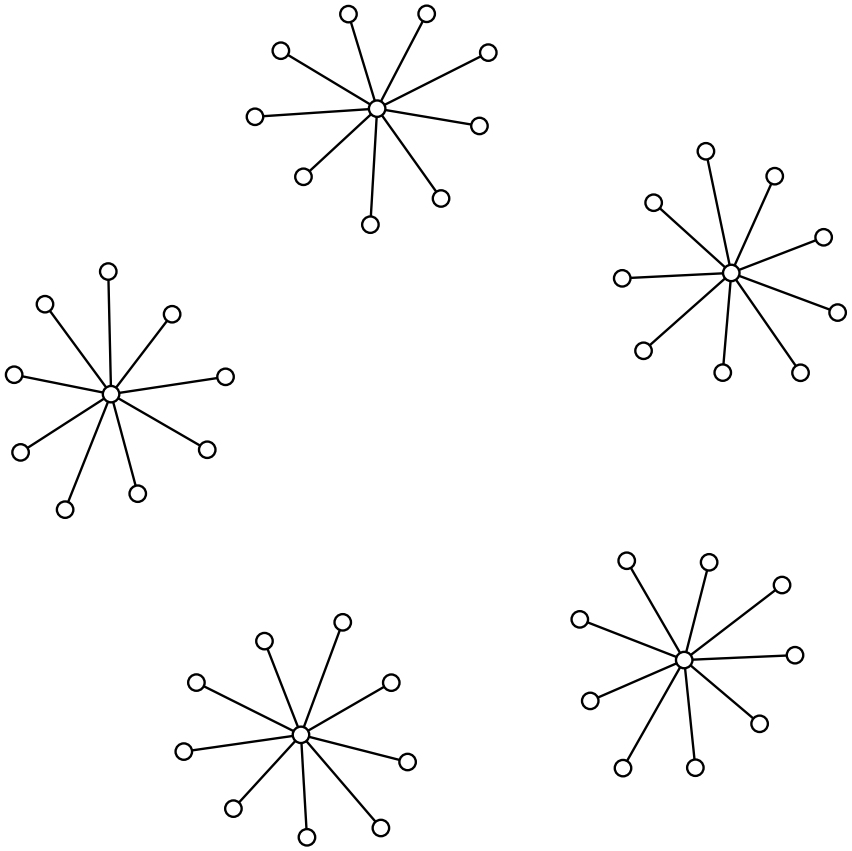

Figure S2. Number of genes identified by the Grace test in the TCGA data against the tuning parameter of the Grace test, h_G. The red dashed line corresponds to the choice made by 10-fold CV ($h_G = \exp(14.2)$).

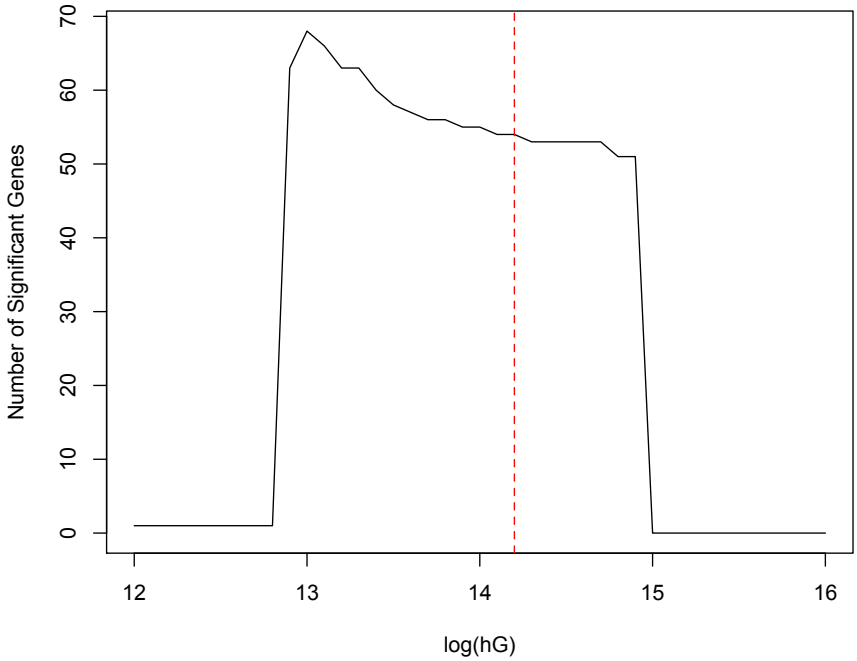

Figure S3. Number of genes that are significant using both the KEGG network and the perturbed network against the number of perturbed edges. The red dashed line represents the number of genes identified by the Grace test with the KEGG network.

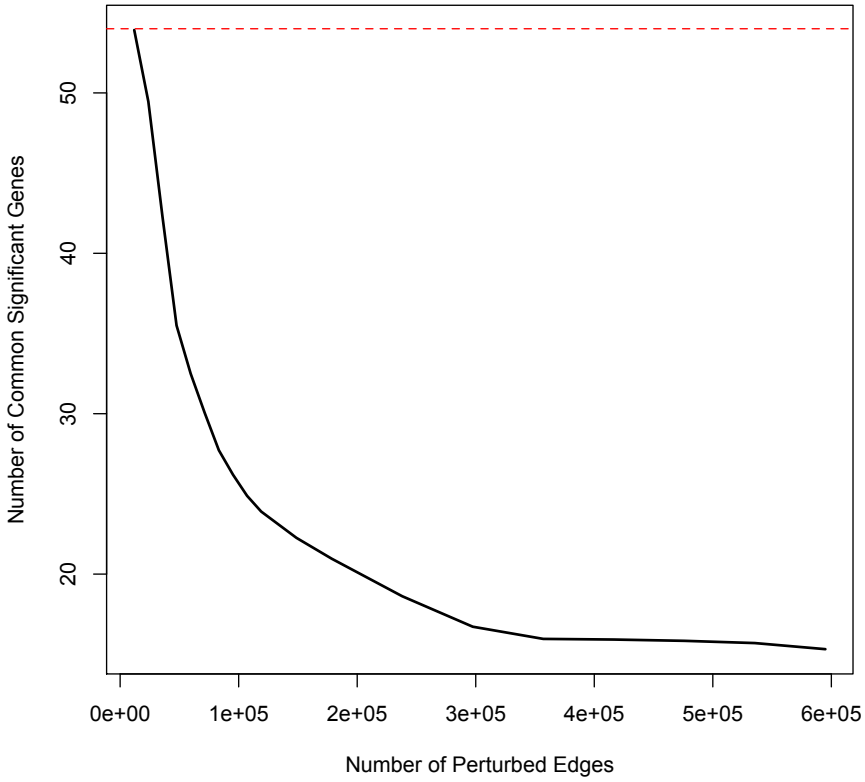

Figure S4. Comparison of power and type-I error rates of different testing methods with their 95% confidence bands. Testing methods include LDPE, ridge, GraceI, Grace and GraceR. Filled circles (•) show powers, whereas crosses (×) are type-I error rates. Numbers on x-axis for Grace and GraceR tests refer to the number of perturbed edges (NPE).

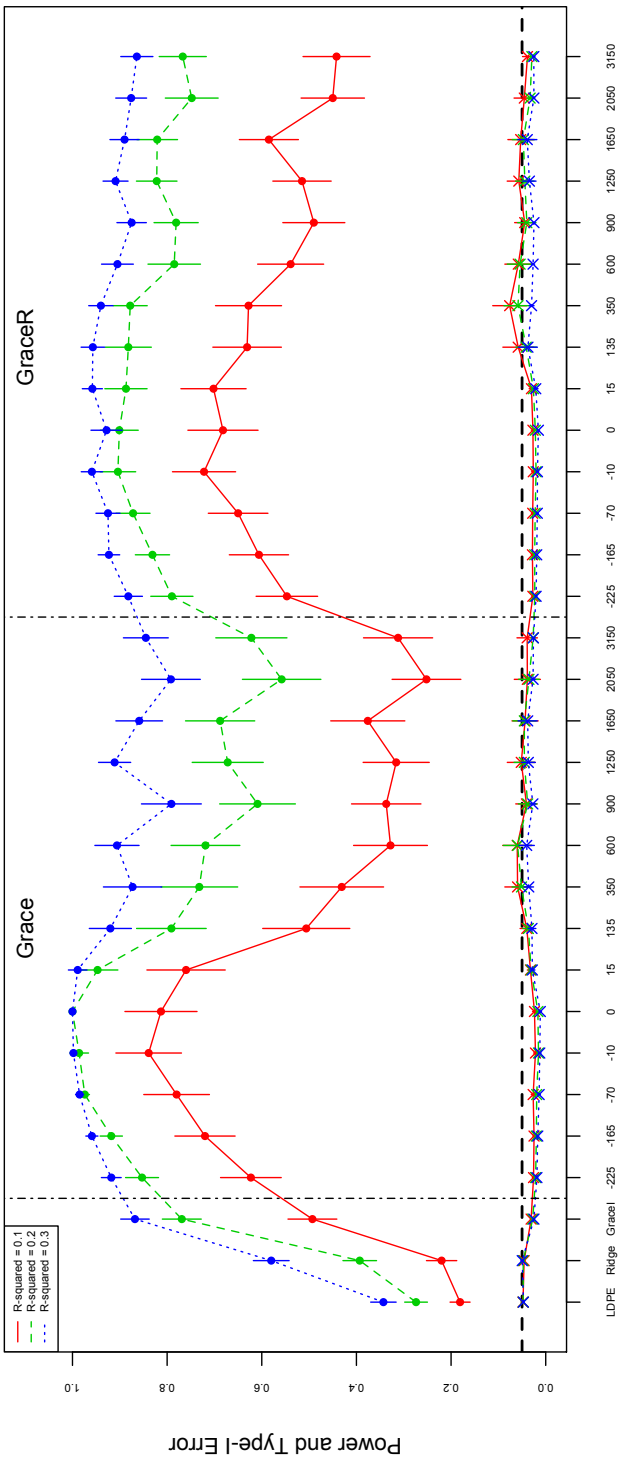

Table S1

Prediction performance of the Grace, GraceR, GraceI(ridge regression with tuning parameter chosen by CV), ridge ($h_2 = 1$) and lasso. The performance metric is the sum of 10-fold CV prediction error (CVER).

	Grace	GraceR	GraceI	Ridge	Lasso
CVER	3473	3411	3418	3917	3546

Table S2
Mean power and the standard error for the LDPE test, ridge test, GraceI, Grace and GraceR tests with different R^2 values.

	$R^2 = 0.1$	$R^2 = 0.2$	$R^2 = 0.3$
LDPE	0.181 (0.011)	0.274 (0.012)	0.343 (0.014)
Ridge	0.220 (0.016)	0.393 (0.018)	0.580 (0.019)
GraceI	0.493 (0.026)	0.769 (0.021)	0.868 (0.015)
Grace NPE = -225	0.623 (0.033)	0.853 (0.018)	0.918 (0.011)
Grace NPE = -165	0.720 (0.032)	0.918 (0.012)	0.959 (0.007)
Grace NPE = -70	0.780 (0.035)	0.974 (0.005)	0.985 (0.004)
Grace NPE = -10	0.839 (0.035)	0.986 (0.010)	0.998 (0.001)
Grace NPE = 0	0.813 (0.039)	1.000 (0.000)	1.000 (0.000)
Grace NPE = 15	0.760 (0.042)	0.947 (0.022)	0.989 (0.010)
Grace NPE = 135	0.506 (0.047)	0.791 (0.038)	0.920 (0.023)
Grace NPE = 350	0.431 (0.045)	0.732 (0.041)	0.873 (0.031)
Grace NPE = 600	0.328 (0.040)	0.719 (0.037)	0.906 (0.024)
Grace NPE = 900	0.337 (0.037)	0.609 (0.041)	0.791 (0.032)
Grace NPE = 1250	0.316 (0.036)	0.672 (0.038)	0.911 (0.017)
Grace NPE = 1650	0.376 (0.040)	0.688 (0.037)	0.859 (0.025)
Grace NPE = 2050	0.252 (0.037)	0.558 (0.042)	0.792 (0.032)
Grace NPE = 3150	0.312 (0.037)	0.622 (0.038)	0.845 (0.024)
GraceR NPE = -225	0.547 (0.033)	0.790 (0.023)	0.882 (0.015)
GraceR NPE = -165	0.606 (0.032)	0.831 (0.018)	0.923 (0.012)
GraceR NPE = -70	0.650 (0.032)	0.872 (0.018)	0.925 (0.013)
GraceR NPE = -10	0.722 (0.034)	0.904 (0.019)	0.959 (0.011)
GraceR NPE = 0	0.682 (0.038)	0.901 (0.020)	0.928 (0.017)
GraceR NPE = 15	0.702 (0.035)	0.887 (0.023)	0.958 (0.011)
GraceR NPE = 135	0.631 (0.037)	0.882 (0.025)	0.957 (0.013)
GraceR NPE = 350	0.628 (0.036)	0.878 (0.018)	0.940 (0.013)
GraceR NPE = 600	0.539 (0.036)	0.785 (0.028)	0.905 (0.017)
GraceR NPE = 900	0.490 (0.033)	0.781 (0.024)	0.875 (0.016)
GraceR NPE = 1250	0.515 (0.031)	0.822 (0.022)	0.909 (0.013)
GraceR NPE = 1650	0.585 (0.032)	0.821 (0.022)	0.890 (0.016)
GraceR NPE = 2050	0.450 (0.034)	0.748 (0.028)	0.876 (0.017)
GraceR NPE = 3150	0.442 (0.036)	0.767 (0.025)	0.864 (0.017)

Table S3

Mean type-I error rate and the standard error for the LDPE test, ridge test, GraceI, Grace and GraceR tests with different R^2 values.

	$R^2 = 0.1$	$R^2 = 0.2$	$R^2 = 0.3$
LDPE	0.048 (0.0010)	0.048 (0.0010)	0.047 (0.0010)
Ridge	0.046 (0.0012)	0.048 (0.0013)	0.050 (0.0012)
GraceI	0.031 (0.0010)	0.027 (0.0009)	0.025 (0.0008)
Grace NPE = -225	0.026 (0.0013)	0.021 (0.0012)	0.019 (0.0010)
Grace NPE = -165	0.025 (0.0014)	0.020 (0.0013)	0.017 (0.0012)
Grace NPE = -70	0.027 (0.0021)	0.019 (0.0017)	0.014 (0.0013)
Grace NPE = -10	0.022 (0.0021)	0.015 (0.0017)	0.013 (0.0015)
Grace NPE = 0	0.024 (0.0021)	0.017 (0.0017)	0.011 (0.0013)
Grace NPE = 15	0.032 (0.0034)	0.031 (0.0031)	0.028 (0.0028)
Grace NPE = 135	0.040 (0.0073)	0.037 (0.0059)	0.029 (0.0042)
Grace NPE = 350	0.059 (0.0137)	0.051 (0.0102)	0.036 (0.0052)
Grace NPE = 600	0.060 (0.0156)	0.059 (0.0155)	0.040 (0.0083)
Grace NPE = 900	0.041 (0.0115)	0.038 (0.0101)	0.027 (0.0033)
Grace NPE = 1250	0.052 (0.0151)	0.045 (0.0111)	0.037 (0.0075)
Grace NPE = 1650	0.044 (0.0141)	0.045 (0.0125)	0.038 (0.0104)
Grace NPE = 2050	0.039 (0.0141)	0.035 (0.0112)	0.027 (0.0023)
Grace NPE = 3150	0.039 (0.0110)	0.027 (0.0024)	0.026 (0.0015)
GraceR NPE = -225	0.027 (0.0012)	0.023 (0.0011)	0.020 (0.0009)
GraceR NPE = -165	0.028 (0.0013)	0.023 (0.0011)	0.019 (0.0010)
GraceR NPE = -70	0.028 (0.0014)	0.022 (0.0014)	0.018 (0.0012)
GraceR NPE = -10	0.026 (0.0018)	0.020 (0.0015)	0.017 (0.0014)
GraceR NPE = 0	0.027 (0.0018)	0.022 (0.0016)	0.015 (0.0013)
GraceR NPE = 15	0.030 (0.0025)	0.026 (0.0025)	0.021 (0.0025)
GraceR NPE = 135	0.058 (0.0165)	0.041 (0.0112)	0.038 (0.0103)
GraceR NPE = 350	0.076 (0.0182)	0.059 (0.0152)	0.030 (0.0027)
GraceR NPE = 600	0.058 (0.0145)	0.054 (0.0139)	0.027 (0.0016)
GraceR NPE = 900	0.044 (0.0109)	0.040 (0.0099)	0.025 (0.0010)
GraceR NPE = 1250	0.057 (0.0125)	0.044 (0.0100)	0.034 (0.0071)
GraceR NPE = 1650	0.053 (0.0138)	0.047 (0.0122)	0.039 (0.0104)
GraceR NPE = 2050	0.045 (0.0111)	0.033 (0.0038)	0.025 (0.0009)
GraceR NPE = 3150	0.039 (0.0053)	0.029 (0.0017)	0.025 (0.0012)