Web-based Supplementary Materials for "Meta-Analysis Based Variable Selection for Gene Expression Data" by

Quefeng Li[1], Sijian Wang[1,2], Chiang-Ching Huang[3], Menggang Yu[2] and Jun Shao[1,*]

[1]Department of Statistics, University of Wisconsin, Madison, Wisconsin, U.S.A

[2]Department of Biostatistics & Medical Informatics, University of Wisconsin, Madison, Wisconsin, U.S.A

[3]Feinberg School of Medicine, Northwestern University, Illinois, U.S.A

**email:* shao@stat.wisc.edu

Web Appendix A: Proofs of Lemmas and Theorems

Proof of Lemma 1. We follow the spirit of the proof of Lemma 1 in Zhou and Zhu (2010). Let $Q_1(\lambda_g, \lambda_\zeta, \boldsymbol{\beta}_0, \boldsymbol{g}, \boldsymbol{\zeta})$ denote the objective function in (5) and $Q_2(\lambda, \boldsymbol{\beta}_0, \boldsymbol{g}, \boldsymbol{\zeta})$ denote the objective function in (7). Suppose $(\tilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{g}}, \tilde{\boldsymbol{\zeta}})$ is the local maximizer of $Q_1(\lambda_g, \lambda_\zeta, \boldsymbol{\beta}_0, \boldsymbol{g}, \boldsymbol{\zeta})$. We would like to show that $(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}}) = (\tilde{\boldsymbol{\beta}}_0, \lambda_g \tilde{\boldsymbol{g}}, \frac{1}{\lambda_g}\tilde{\boldsymbol{\zeta}})$ is a local maximizer of $Q_2(\lambda, \boldsymbol{\beta}_0, \boldsymbol{g}, \boldsymbol{\zeta})$.

Actually, since $(\tilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{g}}, \tilde{\boldsymbol{\zeta}})$ is a local maximizer of (5), there exists $\delta > 0$ such that for any $(\boldsymbol{\beta}_0, \boldsymbol{g}, \boldsymbol{\zeta})$ satisfying $\|\boldsymbol{\beta}_0 - \tilde{\boldsymbol{\beta}}_0\|_1 + \|\boldsymbol{g} - \tilde{\boldsymbol{g}}\|_1 + \|\boldsymbol{\zeta} - \tilde{\boldsymbol{\zeta}}\|_1 \le \delta$, $Q_1(\lambda_g, \lambda_\zeta, \boldsymbol{\beta}_0, \boldsymbol{g}, \boldsymbol{\zeta}) \le Q_1(\lambda_g, \lambda_\zeta, \tilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{g}}, \tilde{\boldsymbol{\zeta}})$.

Choose δ' such that $\frac{\delta'}{\min\left(\lambda_g, \frac{1}{\lambda_g}\right)} \le \delta$, then for any $(\boldsymbol{\beta}'_0, \boldsymbol{g}', \boldsymbol{\zeta}')$ satisfying $\|\boldsymbol{\beta}'_0 - \hat{\boldsymbol{\beta}}_0\|_1 + \|\boldsymbol{g}' - \hat{\boldsymbol{g}}\|_1 + \|\boldsymbol{\zeta}' - \hat{\boldsymbol{\zeta}}\|_1 \le \delta'$, it holds that

$$
\begin{aligned}
&\|\boldsymbol{\beta}'_0 - \tilde{\boldsymbol{\beta}}_0\|_1 + \|\frac{\boldsymbol{g}'}{\lambda_g} - \tilde{\boldsymbol{g}}\|_1 + \|\lambda_g \boldsymbol{\zeta}' - \tilde{\boldsymbol{\zeta}}\|_1 \\
\le\ & \frac{\|\boldsymbol{\beta}'_0 - \tilde{\boldsymbol{\beta}}_0\|_1 + \lambda_g \|\frac{\boldsymbol{g}'}{\lambda_g} - \tilde{\boldsymbol{g}}\|_1 + \frac{1}{\lambda_g}\|\lambda_g \boldsymbol{\zeta}' - \tilde{\boldsymbol{\zeta}}\|_1}{\min\left(\lambda_g, \frac{1}{\lambda_g}\right)} \\
=\ & \frac{\|\boldsymbol{\beta}'_0 - \hat{\boldsymbol{\beta}}_0\|_1 + \|\boldsymbol{g}' - \hat{\boldsymbol{g}}\|_1 + \|\boldsymbol{\zeta}' - \hat{\boldsymbol{\zeta}}\|_1}{\min\left(\lambda_g, \frac{1}{\lambda_g}\right)} \\
\le\ & \delta.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
Q_2(\lambda, \boldsymbol{\beta}'_0, \boldsymbol{g}', \boldsymbol{\zeta}') &= Q_1(\lambda_g, \lambda_\zeta, \boldsymbol{\beta}'_0, \boldsymbol{g}'/\lambda_g, \lambda_g \boldsymbol{\zeta}') \\
&\le Q_1(\lambda_g, \lambda_\zeta, \tilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{g}}, \tilde{\boldsymbol{\zeta}}) \\
&= Q_2(\lambda, \hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}}).
\end{aligned}
$$

Therefore, $(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}}) = (\tilde{\boldsymbol{\beta}}_0, \lambda_g \tilde{\boldsymbol{g}}, \frac{1}{\lambda_g}\tilde{\boldsymbol{\zeta}})$ is a local maximizer of $Q_2(\lambda, \boldsymbol{\beta}_0, \boldsymbol{g}, \boldsymbol{\zeta})$. Similarly, we can show the reverse.

Proposition 1 *Suppose $(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}})$ is a local maximizer of (7). For $j = 1, \ldots, p$, let $\hat{\beta}_{mj} = \hat{g}_j \hat{\zeta}_{mj}$, $\hat{\boldsymbol{\beta}}_{(j)} = (\hat{\beta}_{1j}, \ldots, \hat{\beta}_{Mj})^T$ and $\hat{\boldsymbol{\zeta}}_{(j)} = (\hat{\zeta}_{1j}, \ldots, \hat{\zeta}_{Mj})^T$.*

(a) If $\hat{g}_j = 0$, then $\hat{\boldsymbol{\beta}}_{(j)} = \mathbf{0}$;

(b) If $\hat{g}_j \neq 0$, then $\hat{\boldsymbol{\beta}}_{(j)} \neq \mathbf{0}$ and $|\hat{g}_j| = \sqrt{\lambda \|\hat{\boldsymbol{\beta}}_{(j)}\|_1}$, $|\hat{\boldsymbol{\zeta}}_{(j)}| = \frac{|\hat{\boldsymbol{\beta}}_{(j)}|}{\sqrt{\lambda\|\hat{\boldsymbol{\beta}}_{(j)}\|_1}}$, where the absolute value of $\hat{\boldsymbol{\zeta}}_{(j)}$ and $\hat{\boldsymbol{\beta}}_{(j)}$ are taken componentwise.

Proof of Proposition 1. Statement (a) is obvious. Similarly, if $\hat{\boldsymbol{\beta}}_{(j)} = \mathbf{0}$, then $\hat{g}_j = 0$.

For statement (b), suppose there exists j' such that $\hat{g}_{j'} \neq 0$ and $|\hat{g}_{j'}| \neq \sqrt{\lambda \|\hat{\boldsymbol{\beta}}_{(j')}\|_1}$. Let $\frac{\sqrt{\lambda\|\hat{\boldsymbol{\beta}}_{(j')}\|_1}}{|\hat{g}_{j'}|} = c$. Without loss of generality, we assume $c > 1$.

Let $\tilde{g}_j = \hat{g}_j$ and $\tilde{\boldsymbol{\zeta}}_{(j)} = \hat{\boldsymbol{\zeta}}_{(j)}$ for $j \neq j'$ and $\tilde{g}_{j'} = \delta' \hat{g}_{j'}$ and $\tilde{\boldsymbol{\zeta}}_{(j')} = \frac{1}{\delta'}\hat{\boldsymbol{\zeta}}_{(j')}$, where $1 < \delta' < c$ such that $|\tilde{g}_{j'} - \hat{g}_{j'}| + \|\tilde{\boldsymbol{\zeta}}_{(j')} - \hat{\boldsymbol{\zeta}}_{(j')}\|_1 < \delta$ for some $\delta > 0$. Then, we have

$$\begin{aligned}
Q_2(\lambda, \hat{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{g}}, \tilde{\boldsymbol{\zeta}}) - Q_2(\lambda, \hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}}) &= -\delta' |\hat{g}_{j'}| - \frac{1}{\delta'}\lambda \|\hat{\boldsymbol{\zeta}}_{(j')}\|_1 + |\hat{g}_{j'}| + \lambda\|\hat{\boldsymbol{\zeta}}_{(j')}\|_1 \\
&= \left(-\frac{\delta'}{c} - \frac{c}{\delta'} + \frac{1}{c} + c \right) \sqrt{\lambda \|\hat{\boldsymbol{\beta}}_{(j')}\|_1} \\
&= \frac{1}{c}(\delta' - 1)\left(\frac{c^2}{\delta'} - 1 \right) \sqrt{\lambda \|\hat{\boldsymbol{\beta}}_{(j')}\|_1} \\
&> 0.
\end{aligned}$$

Therefore, for any $\delta > 0$, we can find $\tilde{\boldsymbol{g}}, \tilde{\boldsymbol{\zeta}}$ such that $|\tilde{\boldsymbol{g}} - \hat{\boldsymbol{g}}| + \|\tilde{\boldsymbol{\zeta}} - \hat{\boldsymbol{\zeta}}\|_1 < \delta$ and $Q_2(\lambda, \hat{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{g}}, \tilde{\boldsymbol{\zeta}}) > Q_2(\lambda, \hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}})$. This contracts the fact that $(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}})$ is a local maximizer.

Proof of Lemma 2. Let $Q_3(\lambda, \boldsymbol{\beta})$ be the objective function in (8) and $Q_2(\lambda, \boldsymbol{\beta}_0, \boldsymbol{g}, \boldsymbol{\zeta})$ be the objective function in (7).

First, we show that if $(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}})$ is a local maximizer of $Q_2(\lambda, \boldsymbol{\beta}_0, \boldsymbol{g}, \boldsymbol{\zeta})$, then the vector $\hat{\boldsymbol{\beta}} = (\hat{\beta}_{10}, \hat{\beta}_{11}, \ldots, \hat{\beta}_{Mp})^T$, where $\hat{\beta}_{mj} = \hat{g}_j \hat{\zeta}_{mj}$, for $j = 1, \ldots, p$, is a local maximizer of $Q_3(\lambda, \boldsymbol{\beta})$.

Denote $\Delta\boldsymbol{\beta} = \Delta\boldsymbol{\beta}^{(1)} + \Delta\boldsymbol{\beta}^{(2)}$, where $\Delta\boldsymbol{\beta}^{(1)}_{(j)} = (\Delta\boldsymbol{\beta}^{(1)}_{1j}, \ldots, \Delta\boldsymbol{\beta}^{(1)}_{Mj}) \neq \mathbf{0}$ if and only if $\hat{\boldsymbol{\beta}}_{(j)} \neq \mathbf{0}$; $\Delta\boldsymbol{\beta}^{(2)}_{(j)} = (\Delta\boldsymbol{\beta}^{(2)}_{1j}, \ldots, \Delta\boldsymbol{\beta}^{(2)}_{Mj}) \neq \mathbf{0}$ if and only if $\hat{\boldsymbol{\beta}}_{(j)} = 0$. Then we have $\|\Delta\boldsymbol{\beta}\|_1 = \|\Delta\boldsymbol{\beta}^{(1)}\|_1 + \|\Delta\boldsymbol{\beta}^{(2)}\|_1$.

First, we show that there exists $\delta' > 0$ such that for any $\|\Delta\boldsymbol{\beta}^{(1)}\|_1 < \delta'$, $Q_3(\lambda, \hat{\boldsymbol{\beta}} + \Delta\boldsymbol{\beta}^{(1)}) \leq$

$Q_3(\lambda, \hat{\boldsymbol{\beta}})$. By Proposition 1, we have, for $j = 1, \ldots, p$, $|\hat{g}_j| = \sqrt{\lambda \|\hat{\boldsymbol{\beta}}_{(j)}\|_1}$, $|\hat{\boldsymbol{\zeta}}_{(j)}| = |\hat{\boldsymbol{\beta}}_{(j)}| / \sqrt{\lambda \|\hat{\boldsymbol{\beta}}_{(j)}\|_1}$ if $\hat{g}_j \neq 0$ and $\hat{\boldsymbol{\zeta}}_{(j)} = \mathbf{0}$, if $\hat{g}_j = 0$. Now, let

$$\hat{g}'_j = \operatorname{sgn}(\hat{g}_j) \sqrt{\lambda(\|\hat{\boldsymbol{\beta}}_{(j)} + \Delta \boldsymbol{\beta}^{(1)}_{(j)}\|_1)},$$

$$\hat{\boldsymbol{\zeta}}'_{(j)} = \operatorname{sgn}(\hat{\boldsymbol{\zeta}}_{(j)}) \frac{\hat{\boldsymbol{\beta}}_{(j)} + \Delta \boldsymbol{\beta}^{(1)}_{(j)}}{\sqrt{\lambda(\|\hat{\boldsymbol{\beta}}_{(j)} + \Delta \boldsymbol{\beta}^{(1)}_{(j)}\|_1)}},$$

if $\hat{g}_j \neq 0$ and $\hat{g}'_j = 0$, $\hat{\boldsymbol{\zeta}}'_{(j)} = \mathbf{0}$ if $\hat{g}_j = 0$. Then, it holds that,

$$Q_3(\lambda, \hat{\boldsymbol{\beta}}) = Q_2(\lambda, \hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}}),$$

$$Q_3(\lambda, \hat{\boldsymbol{\beta}} + \Delta \boldsymbol{\beta}^{(1)}) = Q_2(\lambda, \hat{\boldsymbol{\beta}}'_0, \hat{\boldsymbol{g}}', \hat{\boldsymbol{\zeta}}'),$$

Therefore, it suffices to show $Q_2(\lambda, \hat{\boldsymbol{\beta}}'_0, \hat{\boldsymbol{g}}', \hat{\boldsymbol{\zeta}}') \leq Q_2(\lambda, \hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}})$.

Since $(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}})$ is a local maximizer of $Q_2(\lambda, \boldsymbol{\beta}_0, \boldsymbol{g}, \boldsymbol{\zeta})$, there exists $\delta > 0$ such that for any $\hat{\boldsymbol{\beta}}_0$, $\hat{\boldsymbol{g}}'$, $\hat{\boldsymbol{\zeta}}'$ satisfying $\|\hat{\boldsymbol{\beta}}'_0 - \hat{\boldsymbol{\beta}}_0\|_1 + \|\hat{\boldsymbol{g}}' - \hat{\boldsymbol{g}}\|_1 + \|\hat{\boldsymbol{\zeta}}' - \hat{\boldsymbol{\zeta}}\|_1 < \delta$, it holds that $Q_2(\lambda, \hat{\boldsymbol{\beta}}'_0, \hat{\boldsymbol{g}}', \hat{\boldsymbol{\zeta}}') \leq Q_2(\lambda, \hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}})$.

Note that,

$$\begin{aligned}
|\hat{g}'_j - \hat{g}_j| &= \left| \sqrt{\lambda \|\hat{\boldsymbol{\beta}}_{(j)} + \Delta \boldsymbol{\beta}^{(1)}_{(j)}\|_1} - \sqrt{\lambda \|\hat{\boldsymbol{\beta}}_{(j)}\|_1} \right| \\
&\leq \frac{1}{2} \frac{\lambda \|\Delta \boldsymbol{\beta}^{(1)}_{(j)}\|_1}{\sqrt{\lambda \|\hat{\boldsymbol{\beta}}_{(j)}\|_1 - \lambda \|\Delta \boldsymbol{\beta}^{(1)}_{(j)}\|_1}} \\
&\leq \frac{1}{2} \frac{\lambda \|\Delta \boldsymbol{\beta}^{(1)}_{(j)}\|_1}{\sqrt{\lambda l - \lambda \delta'}} \\
&\leq \frac{1}{2} \frac{\lambda \|\Delta \boldsymbol{\beta}^{(1)}_{(j)}\|_1}{\sqrt{\lambda l / 2}},
\end{aligned}$$

where $l = \min_{j : \hat{\boldsymbol{\beta}}_{(j)} \neq \mathbf{0}} \|\hat{\boldsymbol{\beta}}_{(j)}\|_1$ and $\delta' < l/2$.

Meanwhile,

$$
\begin{aligned}
\|\hat{\boldsymbol{\zeta}}_{(j)}' - \hat{\boldsymbol{\zeta}}_{(j)}\|_1 &= \left\| \frac{\hat{\boldsymbol{\beta}}_{(j)} + \Delta\boldsymbol{\beta}_{(j)}^{(1)}}{\sqrt{\lambda\|\hat{\boldsymbol{\beta}}_{(j)} + \Delta\boldsymbol{\beta}_{(j)}^{(1)}\|_1}} - \frac{\hat{\boldsymbol{\beta}}_{(j)}}{\sqrt{\lambda\|\hat{\boldsymbol{\beta}}_{(j)}\|_1}} \right\|_1 \\
&\leq \left\| \frac{\hat{\boldsymbol{\beta}}_{(j)} + \Delta\boldsymbol{\beta}_{(j)}^{(1)}}{\sqrt{\lambda\|\hat{\boldsymbol{\beta}}_{(j)} + \Delta\boldsymbol{\beta}_{(j)}^{(1)}\|_1}} - \frac{\hat{\boldsymbol{\beta}}_{(j)}}{\sqrt{\lambda\|\hat{\boldsymbol{\beta}}_{(j)} + \Delta\boldsymbol{\beta}_{(j)}^{(1)}\|_1}} \right\|_1 \\
&\quad + \left\| \frac{\hat{\boldsymbol{\beta}}_{(j)}}{\sqrt{\lambda\|\hat{\boldsymbol{\beta}}_{(j)} + \Delta\boldsymbol{\beta}_{(j)}^{(1)}\|_1}} - \frac{\hat{\boldsymbol{\beta}}_{(j)}}{\sqrt{\lambda\|\hat{\boldsymbol{\beta}}_{(j)}\|_1}} \right\|_1 \\
&\leq \frac{\|\Delta\boldsymbol{\beta}_{(j)}^{(1)}\|_1}{\sqrt{\lambda l/2}} + \|\hat{\boldsymbol{\beta}}_{(j)}\|_1 \cdot \frac{\left| \sqrt{\lambda\|\hat{\boldsymbol{\beta}}_{(j)} + \Delta\boldsymbol{\beta}_{(j)}^{(1)}\|_1} - \sqrt{\lambda\|\hat{\boldsymbol{\beta}}_{(j)}\|_1} \right|}{\sqrt{\lambda\|\hat{\boldsymbol{\beta}}_{(j)} + \Delta\boldsymbol{\beta}_{(j)}^{(1)}\|_1}\sqrt{\lambda\|\hat{\boldsymbol{\beta}}_{(j)}\|_1}} \\
&\leq \|\Delta\boldsymbol{\beta}_{(j)}^{(1)}\|_1 \left(\frac{1}{\sqrt{\lambda l/2}} + \frac{L}{l\sqrt{\lambda l}} \right),
\end{aligned}
$$

where $L = \max_{j:\hat{\boldsymbol{\beta}}_{(j)}\neq\mathbf{0}} \|\hat{\boldsymbol{\beta}}_{(j)}\|_1$. In addition, $\|\hat{\boldsymbol{\beta}}_0' - \hat{\boldsymbol{\beta}}_0\|_1 = \|\Delta\boldsymbol{\beta}_0^{(1)}\|_1$. Therefore, by choosing a proper δ', we have $\|\hat{\boldsymbol{\beta}}_0' - \hat{\boldsymbol{\beta}}_0\|_1 + \|\hat{\boldsymbol{g}}' - \boldsymbol{g}\|_1 + \|\hat{\boldsymbol{\zeta}}' - \boldsymbol{\zeta}\|_1 < \delta$. Then,

$$
Q_3(\lambda, \hat{\boldsymbol{\beta}} + \Delta\boldsymbol{\beta}^{(1)}) = Q_2(\lambda, \hat{\boldsymbol{\beta}}_0', \hat{\boldsymbol{g}}', \hat{\boldsymbol{\zeta}}') \leq Q_2(\lambda, \hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}}) = Q_3(\lambda, \hat{\boldsymbol{\beta}}).
$$

Next, we show that $Q_3(\lambda, \hat{\boldsymbol{\beta}} + \Delta\boldsymbol{\beta}^{(1)} + \Delta\boldsymbol{\beta}^{(2)}) \leq Q_3(\lambda, \hat{\boldsymbol{\beta}} + \Delta\boldsymbol{\beta}^{(1)})$. By mean-value theorem,

$$
\begin{aligned}
&Q_3(\lambda, \hat{\boldsymbol{\beta}} + \Delta\boldsymbol{\beta}^{(1)} + \Delta\boldsymbol{\beta}^{(2)}) - Q_3(\lambda, \hat{\boldsymbol{\beta}} + \Delta\boldsymbol{\beta}^{(1)}) \\
&= [\nabla\ell(\tilde{\boldsymbol{\beta}})]^T \Delta\boldsymbol{\beta}^{(2)} - 2\sqrt{\lambda}\sum_{j=1}^{p} \sqrt{\|\Delta\boldsymbol{\beta}^{(2)}\|_1},
\end{aligned} \tag{A.1}
$$

where $\tilde{\boldsymbol{\beta}}$ lies in the line segment connecting $\hat{\boldsymbol{\beta}} + \Delta\boldsymbol{\beta}^{(1)}$ and $\hat{\boldsymbol{\beta}} + \Delta\boldsymbol{\beta}^{(1)} + \Delta\boldsymbol{\beta}^{(2)}$. When $\|\Delta\boldsymbol{\beta}^{(2)}\|_1 < \delta'$ is small enough, the second term in (A.1) is larger than the first term. Hence, it holds that $Q_3(\lambda, \hat{\boldsymbol{\beta}} + \Delta\boldsymbol{\beta}^{(1)} + \Delta\boldsymbol{\beta}^{(2)}) \leq Q_3(\lambda, \hat{\boldsymbol{\beta}} + \Delta\boldsymbol{\beta}^{(1)})$.

In conclusion, we have shown that there exists a small enough δ' such that if $\|\Delta\boldsymbol{\beta}\|_1 \leq \delta'$, then $Q_3(\lambda, \hat{\boldsymbol{\beta}} + \Delta\boldsymbol{\beta}) \leq Q_3(\lambda, \hat{\boldsymbol{\beta}})$, namely $\hat{\boldsymbol{\beta}}$ is a local maximizer of $Q_3(\lambda, \boldsymbol{\beta})$. Similarly, we can show the reverse.

Proof of Theorem 1. We follow the technique in the proof of Theorem 1 in Fan and Lv (2011). Firstly we show the necessary condition. For the log-likelihood $\ell(\boldsymbol{\beta}) = \boldsymbol{Y}^T\boldsymbol{X}\boldsymbol{\beta} - \mathbf{1}^T\boldsymbol{b}(\boldsymbol{X}\boldsymbol{\beta})$, we have

$$\nabla\ell(\boldsymbol{\beta}) = \boldsymbol{X}^T\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\beta}) \quad \text{and} \quad \nabla^2\ell(\boldsymbol{\beta}) = -\boldsymbol{X}^T\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta})\boldsymbol{X}.$$

By the classical Karush-Kuhn-Tucker (KKT) condition, if $\hat{\boldsymbol{\beta}} = (\hat{\beta}_{11}, \hat{\beta}_{12}, \ldots, \hat{\beta}_{Mp})^T$ is a local maximizer of the regularized likelihood (9), there exists gradient $\nabla\rho(\hat{\boldsymbol{\beta}})$ and one sub-gradient $\partial\rho(\hat{\boldsymbol{\beta}})$ of $\rho(\cdot)$ such that

$$\boldsymbol{X}_I^T\boldsymbol{Y} - \boldsymbol{X}_I^T\boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) - n\lambda_n\nabla\rho(\hat{\boldsymbol{\beta}}_I) = \mathbf{0}$$

$$\boldsymbol{X}_{II}^T\boldsymbol{Y} - \boldsymbol{X}_{II}^T\boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) - n\lambda_n\partial\rho(\hat{\boldsymbol{\beta}}_{II}) = \mathbf{0}$$

$$\boldsymbol{X}_{III}^T\boldsymbol{Y} - \boldsymbol{X}_{III}^T\boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) - n\lambda_n\partial\rho(\hat{\boldsymbol{\beta}}_{III}) = \mathbf{0}$$

where

$$\nabla\rho(\hat{\beta}_{mj}) = \frac{1}{2}\text{sgn}(\hat{\beta}_{mj})\|\hat{\boldsymbol{\beta}}_{(j)}\|_1^{-1/2} \qquad \text{for } \hat{\beta}_{mj} \in \hat{\boldsymbol{\beta}}_I, \tag{A.2}$$

$$\partial\rho(\hat{\beta}_{mj}) \begin{cases} \in [-\frac{1}{2}\|\hat{\boldsymbol{\beta}}_{(j)}\|_1^{-1/2}, \frac{1}{2}\|\hat{\boldsymbol{\beta}}_{(j)}\|_1^{-1/2}] & \text{for } \hat{\beta}_{mj} \in \hat{\boldsymbol{\beta}}_{II}, \quad \text{(A.3)} \\ \in (-\infty, +\infty) & \text{for } \hat{\beta}_{mj} \in \hat{\boldsymbol{\beta}}_{III}. \quad \text{(A.4)} \end{cases}$$

where $\hat{\boldsymbol{\beta}}_I$ and $\hat{\boldsymbol{\beta}}_{II}$ are defined in Theorem 1 and $\hat{\boldsymbol{\beta}}_{III} = \{\hat{\beta}_{mj} | \hat{\beta}_{mj} = 0, \hat{\boldsymbol{\beta}}_{(j)} = \mathbf{0}\}$. In view of (A.4), $\boldsymbol{X}_{III}^T\boldsymbol{Y} - \boldsymbol{X}_{III}^T\boldsymbol{\mu}(\boldsymbol{X}^T\hat{\boldsymbol{\beta}}) = \lambda_n\partial\rho(\hat{\boldsymbol{\beta}}_{III})$ always holds. Hence, necessary conditions only require (10) and (11) hold for $\partial\rho(\cdot)$ in (A.2) and (A.3), respectively. Moreover, since $\hat{\boldsymbol{\beta}}$ is also a local maximizer of (9) constrained on the $|I|$-dimensional subspace $\mathcal{S}_1 = \{\boldsymbol{\beta} \in \mathcal{R}^{Mp} : \boldsymbol{\beta}_{II\cup III} = \mathbf{0}\}$ of $\mathcal{R}^{Mp}$, where $\boldsymbol{\beta}_{II\cup III}$ denotes the subvector of $\boldsymbol{\beta}$ formed by coordinates in $II \cup III$. By the second order condition,

$$\lambda_{\min}(\boldsymbol{X}_I^T\boldsymbol{\Sigma}(\boldsymbol{X}\hat{\boldsymbol{\beta}})\boldsymbol{X}_I) \geq n\lambda_n\kappa(\rho, \hat{\boldsymbol{\beta}}_I),$$

where $\kappa(\rho; \hat{\boldsymbol{\beta}}_I)$ is given in Theorem 1.

Next, we show the sufficient condition. Firstly, we constrain $Q_n(\boldsymbol{\beta})$ in a $|I|$-dimensional subspace $\mathcal{S}_1$ of $\mathcal{R}^{Mp}$. It follows from condition (10) and (12) that $\hat{\boldsymbol{\beta}}$ is the unique maximizer of $Q_n(\boldsymbol{\beta})$ in a neighborhood $\mathcal{N}_1 \subset \mathcal{S}_1$. Next, we show that there exists a neighborhood $\mathcal{N}_2$ in a $(|I \cup II|)$-dimensional space $\mathcal{S}_2$, such that $\mathcal{S}_1 \subset \mathcal{S}_2 \subset \mathcal{R}^{Mp}$ and $\hat{\boldsymbol{\beta}}$ is the unique local maximizer of $Q_n(\boldsymbol{\beta})$ constrained in $\mathcal{S}_2$.

Take a sufficiently small L_1-ball $\mathcal{N}_2$ in $\mathcal{S}_2$ centered at $\hat{\boldsymbol{\beta}}$ such that $\mathcal{N}_2 \cap \mathcal{S}_1 \subset \mathcal{N}_1$. We next show that $Q_n(\hat{\boldsymbol{\beta}}) > Q_n(\boldsymbol{\eta}_2)$ for any $\boldsymbol{\eta}_2 \in \mathcal{N}_2 \backslash \mathcal{N}_1$. Let $\boldsymbol{\eta}_1$ be the projection of $\boldsymbol{\eta}_2$ onto the subspace $\mathcal{S}_1$. Then we have $\boldsymbol{\eta}_1 \in \mathcal{N}_1$, which entails that $Q_n(\hat{\boldsymbol{\beta}}) > Q_n(\boldsymbol{\eta}_1)$ if $\boldsymbol{\eta}_1 \neq \hat{\boldsymbol{\beta}}$. It then suffices to show that $Q_n(\boldsymbol{\eta}_1) > Q_n(\boldsymbol{\eta}_2)$.

By the mean-value theorem, we have

$$Q_n(\boldsymbol{\eta}_2) - Q_n(\boldsymbol{\eta}_1) = [\nabla Q_n(\boldsymbol{\eta}_0)]^T (\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1), \tag{A.5}$$

where $\boldsymbol{\eta}_0$ lies in the line connecting $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$. Since the coordinate of $\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1$ are zero for indices in I and $\text{sgn}(\eta_{0,mj}) = \text{sgn}(\eta_{2,mj})$ where $\eta_{0,mj}$ and $\eta_{2,mj}$ are the (m,j)th coordinate of $\boldsymbol{\eta}_0$ and $\boldsymbol{\eta}_2$, respectively. Therefore, the right hand side of (A.5) equals to

$$\begin{aligned}
&[\boldsymbol{X}_{II}^T\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\eta}_0)\}]^T \boldsymbol{\eta}_{2,1} - n\lambda_n \sum_{(m,j)\in II} \nabla\rho(\eta_{0,mj})\eta_{2,mj} \\
&= n\lambda_n \sum_{(m,j)\in II} z_{mj}(\boldsymbol{\eta}_0)\eta_{2,mj} - n\lambda_n \sum_{(m,j)\in II} \nabla\rho(|\eta_{0,mj}|)|\eta_{2,mj}| \\
&\leq n\lambda_n \sum_{(m,j)\in II} |z_{mj}(\boldsymbol{\eta}_0)| \cdot |\eta_{2,mj}| - n\lambda_n \sum_{(m,j)\in II} \nabla\rho(|\eta_{0,mj}|)|\eta_{2,mj}| \\
&= n\lambda_n \sum_{(m,j)\in II} |z_{mj}(\boldsymbol{\eta}_0)| \cdot |\eta_{2,mj}| - n\lambda_n \sum_{(m,j)\in II} \frac{1}{2}(\sum_{m=1}^{M} |\eta_{0,mj}|)^{-1/2}|\eta_{2,mj}|,
\end{aligned} \tag{A.6}$$

where $z_{mj}(\boldsymbol{\eta}_0)$ denotes the (m,j)th coordinate of $\boldsymbol{X}_{II}^T\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\eta}_0)\}$.

By continuity of $b'(\cdot)$ and $\nabla\rho(\cdot)$ and (11), there exists $\delta > 0$ such that for any $\boldsymbol{\eta}$ in an L_1-ball in $\mathcal{S}_2$ centered at $\hat{\boldsymbol{\beta}}$ with radius δ, $|z_{mj}(\boldsymbol{\eta})| < \frac{1}{2}(\|\hat{\boldsymbol{\beta}}_{(j)}\|_1 + \delta)^{-1/2}$. Let $\mathcal{N}_2$ be that ball. Therefore, (A.6) could be strictly bounded by

$$n\lambda_n \sum_{(m,j)\in II} \frac{1}{2}(\|\hat{\boldsymbol{\beta}}_{(j)}\|_1+\delta)^{-1/2}|\eta_{2,mj}| - n\lambda_n \sum_{(m,j)\in II} \frac{1}{2}(\sum_{m=1}^{M}|\eta_{0,mj}|)^{-1/2}|\eta_{2,mj}| \le 0,$$

since $\sum_{m=1}^{M}|\eta_{0,mj}| = \sum_{m=1}^{M}|\eta_{0,mj}-\hat{\beta}_{mj}+\hat{\beta}_{mj}| \le \sum_{m=1}^{M}|\eta_{0,mj}-\hat{\beta}_{mj}| + \sum_{m=1}^{M}|\hat{\beta}_{mj}| \le \|\hat{\boldsymbol{\beta}}_{(j)}\|_1+\delta$, because $\boldsymbol{\eta}_0$ is within the L_1-ball. This shows that there exists a neighborhood of $\hat{\boldsymbol{\beta}}$, namely $\mathcal{N}_2$, in the space of $\mathcal{S}_2$ such that $\hat{\boldsymbol{\beta}}$ constrained on $\mathcal{S}_2$ is the unique maximizer in that neighborhood.

Applying the same projection technique, we can show that $\hat{\boldsymbol{\beta}}$ is indeed a local maximizer in $\mathcal{R}^{Mp}$ by noting the fact that $\partial\rho(\hat{\beta}_{mj}) \in (-\infty,+\infty)$ for any $\hat{\beta}_{mj} \in \hat{\boldsymbol{\beta}}_{III} = \{\hat{\beta}_{mj}|\hat{\beta}_{mj}=0, \hat{\boldsymbol{\beta}}_{(j)}=\mathbf{0}\}$, so the third KKT condition $\boldsymbol{X}_{III}^T\boldsymbol{Y} - \boldsymbol{X}_{III}^T\boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) = \lambda_n\partial\rho(\hat{\boldsymbol{\beta}}_{III})$ always holds.

Proof of Theorem 2. Let $\boldsymbol{\xi} = (\xi_{11},\xi_{12},\ldots,\xi_{Mp})^T = \boldsymbol{X}^T\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{\mu}(\boldsymbol{\theta}^*)$. Consider events

$$E_1 = \{\|\boldsymbol{\xi}_I\|_\infty \le \sqrt{2^{-1}n\log n}\} \text{ and } E_2 = \{\|\boldsymbol{\xi}_{II\cup III}\|_\infty \le n^{1-\alpha_p}\sqrt{2^{-1}\log n}\},$$

where $\boldsymbol{\xi}_I$ and $\boldsymbol{\xi}_{II\cup III}$ are the sub-vectors of $\boldsymbol{\xi}$ with indices in I and $II\cup III$, respectively.

Since $y_{mi}\in\{0,1\}$, by Hoeffding's inequality,

$$P(|\xi_{mj}|\ge t) \le 2\exp(-\frac{2t^2}{n}).$$

Then, it follows from Bonferroni's inequality that

$$\begin{aligned} P(E_1\cap E_2) &\ge 1 - \sum_{(m,j)\in I} P(|\xi_{mj}| \ge \sqrt{2^{-1}n\log n}) \\ &\quad - \sum_{(m,j)\in II\cup III} P(|\xi_{mj}| \ge n^{1-\alpha_p}\sqrt{2^{-1}\log n}) \\ &\ge 1 - 2\{s_p n^{-1} + (Mp-s_p)e^{-n^{1-2\alpha_p}\log n}\}. \end{aligned}$$

Next, we will show that, in event $E_1\cap E_2$, there exists a solution to (9) that achieve the weak oracle properties in (a) and (b).

Step 1: Existence of a solution to equation (10). We prove that, when n is sufficiently large, there exists a solution to (10) in the hypercube

$$\mathcal{N} = \{\boldsymbol{\delta} \in \mathcal{R}^{s_p} : \|\boldsymbol{\delta} - \boldsymbol{\beta}_I^*\|_\infty = n^{-\gamma}\}.$$

Let $\boldsymbol{\eta} = n\lambda_n \nabla \rho(\boldsymbol{\delta})$, where $\eta_{mj} = n\lambda_n \frac{2^{-1}\text{sgn}(\delta_{mj})}{\sqrt{\sum_{m=1}^M |\delta_{mj}|}}$. We have, for any $(m,j) \in I$,

$$\begin{aligned}
|\eta_{mj}| &\le \frac{2^{-1} n\lambda_n}{(\sum_{m=1}^M |\delta_{mj}|)^{1/2}} \\
&\le \frac{2^{-1} n\lambda_n}{(\sum_{m=1}^M |\beta_{mj}^*| - \sum_{m=1}^M |\delta_{mj} - \beta_{mj}^*|)^{1/2}} \\
&\le \frac{2^{-1} n\lambda_n}{(\sum_{m=1}^M |\beta_{mj}^*| - \frac{1}{2}\sum_{m=1}^M |\beta_{mj}^*|)^{1/2}} \\
&\le \frac{n\lambda_n}{\sqrt{2} l_p},
\end{aligned}$$

because under (C2), for sufficiently large n, $|\beta_{mj}^*| > d_p > n^{-\gamma} \ge |\delta_{mj} - \beta_{mj}^*|$. Clearly, $\frac{n\lambda_n}{\sqrt{2} l_p} \le n\lambda_n (2Md_p)^{-1/2}$. Hence, it holds that

$$\|\boldsymbol{\eta}\|_\infty \le \frac{n\lambda_n}{\sqrt{2} l_p} \le n\lambda_n (2Md_p)^{-1/2}. \tag{A.7}$$

Then, in event E_1,

$$\|\boldsymbol{\xi}_I - \boldsymbol{\eta}\|_\infty \le \|\boldsymbol{\xi}_I\|_\infty + \|\boldsymbol{\eta}\|_\infty \le \sqrt{2^{-1} n \log n} + n\lambda_n (2Md_p)^{-1/2}.$$

Define

$$\boldsymbol{\Psi}(\boldsymbol{\delta}) = \boldsymbol{X}_I^T \{\boldsymbol{\mu}(\boldsymbol{X}_I \boldsymbol{\delta}) - \boldsymbol{\mu}(\boldsymbol{X}_I \boldsymbol{\beta}_I^*)\} - (\boldsymbol{\xi}_I - \boldsymbol{\eta}). \tag{A.8}$$

Note that, (10) is equivalent to $\boldsymbol{\Psi}(\boldsymbol{\delta}) = 0$. For the first term in (A.8). By a second order Taylor expansion, we obtain,

$$\boldsymbol{X}_I^T \{\boldsymbol{\mu}(\boldsymbol{X}_I \boldsymbol{\delta}) - \boldsymbol{\mu}(\boldsymbol{X}_I \boldsymbol{\beta}_I^*)\} = \boldsymbol{X}_I^T \boldsymbol{\Sigma}(\boldsymbol{\theta}^*) \boldsymbol{X}_I (\boldsymbol{\delta} - \boldsymbol{\beta}_I^*) + \boldsymbol{r},$$

where the Lagrange reminder term can be expressed as $\boldsymbol{r} = (r_{mj}, (m,j) \in I)^T$ that

$$r_{mj} = \frac{1}{2} (\boldsymbol{\delta} - \boldsymbol{\beta}_I^*)^T \boldsymbol{R}(\tilde{\boldsymbol{\delta}}_{mj}) (\boldsymbol{\delta} - \boldsymbol{\beta}_I^*),$$

where $\boldsymbol{R}(\tilde{\boldsymbol{\delta}}_{mj}) = \boldsymbol{X}_I^T\{\mathrm{diag}(|\boldsymbol{X}_{mj}| \circ |\boldsymbol{\mu}''(\boldsymbol{X}_I\tilde{\boldsymbol{\delta}}_{mj})|\}\boldsymbol{X}_I$ and $\tilde{\boldsymbol{\delta}}_{mj}$ being some vector lying on the line segment joining $\boldsymbol{\delta}$ and $\boldsymbol{\beta}_I^*$. By condition (C5) and a similar argument as (43) of Fan and Lv (2011),

$$\|\boldsymbol{r}\|_\infty = O(s_p n^{1-2\gamma}). \tag{A.9}$$

Let

$$\bar{\boldsymbol{\Psi}}(\boldsymbol{\delta}) = [\boldsymbol{X}_I^T\boldsymbol{\Sigma}(\boldsymbol{\theta}^*)\boldsymbol{X}_I]^{-1}\boldsymbol{\Psi}(\boldsymbol{\delta}) = \boldsymbol{\delta} - \boldsymbol{\beta}_I^* + \boldsymbol{u}, \tag{A.10}$$

where $\boldsymbol{u} = -[\boldsymbol{X}_I^T\boldsymbol{\Sigma}(\boldsymbol{\theta}^*)\boldsymbol{X}_I]^{-1}(\boldsymbol{\xi}_I - \boldsymbol{\eta} - \boldsymbol{r})$. Then, it follows from (C2)-(C3) and the choice of λ_n in (13) that

$$\begin{aligned}\|\boldsymbol{u}\|_\infty &\le \|[\boldsymbol{X}_I^T\boldsymbol{\Sigma}(\boldsymbol{\theta}^*)\boldsymbol{X}_I]^{-1}\|_\infty(\|\boldsymbol{\xi}_I - \boldsymbol{\eta}\|_\infty + \|\boldsymbol{r}\|_\infty)\\ &= O(b_s n^{-1/2}\sqrt{\log n} + b_s\lambda_n d_p^{-1/2} + b_s s_p n^{-2\gamma})\\ &= o(n^{-\gamma}).\end{aligned}$$

By (A.10), for sufficiently large n, if $(\boldsymbol{\delta} - \boldsymbol{\beta}_I^*)_{mj} = n^{-\gamma}$, we have

$$\bar{\boldsymbol{\Psi}}_{mj}(\boldsymbol{\delta}) \ge n^{-\gamma} - \|\boldsymbol{u}\|_\infty \ge 0,$$

and if $(\boldsymbol{\delta} - \boldsymbol{\beta}_I^*)_{mj} = -n^{-\gamma}$, we have

$$\bar{\boldsymbol{\Psi}}_{mj}(\boldsymbol{\delta}) \le -n^{-\gamma} + \|\boldsymbol{u}\|_\infty \le 0,$$

where $(\boldsymbol{\delta} - \boldsymbol{\beta})_{mj}$ is the (m, j)th element of $\boldsymbol{\delta} - \boldsymbol{\beta}$ and $\bar{\boldsymbol{\Psi}}_{mj}(\boldsymbol{\delta})$ is the (m, j)th element of $\bar{\boldsymbol{\Psi}}$. By the continuity of $\bar{\boldsymbol{\Psi}}(\boldsymbol{\delta})$, an application of Miranda's existence theorem shows that equation $\bar{\boldsymbol{\Psi}}(\boldsymbol{\delta}) = \boldsymbol{0}$ has a solution $\hat{\boldsymbol{\beta}}_I$ in $\mathcal{N}$. In view of (A.10), $\hat{\boldsymbol{\beta}}_I$ is also a solution to $\boldsymbol{\Psi}(\boldsymbol{\delta}) = \boldsymbol{0}$. Hence, we have shown that there exits a solution $\hat{\boldsymbol{\beta}}_I$ inside $\mathcal{N}$.

Step 2: Verify equation (11). Let $\hat{\boldsymbol{\beta}} \in \mathcal{R}^{Mp}$ that $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_I, \boldsymbol{0})$ where $\hat{\boldsymbol{\beta}}_I$ is the solution to (10) as shown above. Next, we prove that $\hat{\boldsymbol{\beta}}$ satisfies (11) for the choice of λ_n in (13). Indeed, (11) requires that

$$|\boldsymbol{X}_{mj}^T\boldsymbol{Y} - \boldsymbol{X}_{mj}^T\boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}})| < \frac{1}{2}n\lambda_n\left(\sum_{m':(m',j)\in I}|\hat{\beta}_{m'j}|\right)^{-1/2},$$

for any $(m, j) \in II$, where $\boldsymbol{X}_{mj}$ denotes the column of $\boldsymbol{X}$ corresponding to the expression of jth gene in the mth dataset.

Since

$$\begin{aligned}\sum_{m':(m',j)\in I}|\hat{\beta}_{m'j}| &\leq \sum_{m':(m',j)\in I}|\hat{\beta}_{m'j} - \beta^*_{m'j}| + |\beta^*_{m'j}| \\ &\leq 2\sum_{m':(m',j)\in I}|\beta^*_{m'j}| \\ &\leq 2L_p^2,\end{aligned}$$

it follows that,

$$\min_{(m,j)\in II}\frac{1}{2}\left(\sum_{m':(m',j)\in I}|\hat{\beta}_{m'j}|\right)^{-1/2} \geq \frac{1}{2\sqrt{2}L_p}.$$

Then, it suffices to show that

$$\|\boldsymbol{X}_{II}^T\boldsymbol{Y} - \boldsymbol{X}_{II}^T\boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}})\|_\infty < \frac{n\lambda_n}{2\sqrt{2}L_p}. \tag{A.11}$$

Note that,

$$\begin{aligned}&\boldsymbol{X}_{II}^T\boldsymbol{Y} - \boldsymbol{X}_{II}^T\boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) \\ &= \boldsymbol{X}_{II}^T\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\beta}^*)\} + \boldsymbol{X}_{II}^T\{\boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\beta}^*) - \boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}})\}.\end{aligned} \tag{A.12}$$

In event E_2, $\|\boldsymbol{X}_{II}^T\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\beta}^*)\}\|_\infty = O(n^{1-\alpha_p}\sqrt{\log n})$. Then, by the choice of λ_n as in (13),

$$(n\lambda_n)^{-1}\|\boldsymbol{X}_{II}^T\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\beta}^*)\}\|_\infty = o(1). \tag{A.13}$$

For the second term in (A.12), by Taylor expansion,

$$\begin{aligned}\boldsymbol{X}_{II}^T\{\boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) - \boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\beta}^*)\} &= \boldsymbol{X}_{II}^T\{\boldsymbol{\mu}(\boldsymbol{X}_I\hat{\boldsymbol{\beta}}_I) - \boldsymbol{\mu}(\boldsymbol{X}_I\boldsymbol{\beta}_I^*)\} \\ &= \boldsymbol{X}_{II}^T\{\boldsymbol{\Sigma}(\boldsymbol{\theta}^*)\boldsymbol{X}_I(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_I^*)\} + \boldsymbol{w},\end{aligned}$$

where $\boldsymbol{w} = (w_{mj}, (m,j) \in II)^T$ that $w_{mj} = \frac{1}{2}(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_I^*)^T \boldsymbol{R}(\bar{\boldsymbol{\delta}}_{mj})(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_I^*)$, in which $\bar{\boldsymbol{\delta}}_{mj}$ are some vector lying on the line segment joining $\hat{\boldsymbol{\beta}}_I$ and $\boldsymbol{\beta}_I^*$. By (C5), $\hat{\boldsymbol{\beta}}_I \in \mathcal{N}$ and a similar argument as (43) in Fan and Lv (2011), we have

$$\|\boldsymbol{w}\|_\infty = O(s_p n^{1-2\gamma}). \tag{A.14}$$

Since $\hat{\boldsymbol{\beta}}_I$ solves $\bar{\boldsymbol{\Psi}}(\boldsymbol{\delta}) = \mathbf{0}$ in (A.10), we have,

$$\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_I^* = [\boldsymbol{X}_I^T \boldsymbol{\Sigma}(\boldsymbol{\theta}^*) \boldsymbol{X}_I]^{-1}(\boldsymbol{\xi}_I - \boldsymbol{\eta} - \boldsymbol{r}).$$

Therefore,

$$\begin{aligned}
&(n\lambda_n)^{-1} \boldsymbol{X}_{II}^T \{\boldsymbol{\mu}(\boldsymbol{X}_I \hat{\boldsymbol{\beta}}_I) - \boldsymbol{\mu}(\boldsymbol{X}_I \boldsymbol{\beta}_I^*)\} \\
\le\ & (n\lambda_n)^{-1} \|\boldsymbol{X}_{II}^T \boldsymbol{\Sigma}(\boldsymbol{\theta}^*) \boldsymbol{X}_I [\boldsymbol{X}_I^T \boldsymbol{\Sigma}(\boldsymbol{\theta}^*) \boldsymbol{X}_I]^{-1}\|_\infty \cdot (\|\boldsymbol{\xi}_I - \boldsymbol{\eta}\|_\infty + \|\boldsymbol{r}\|_\infty) \\
&+ (n\lambda_n)^{-1} \|\boldsymbol{w}\|_\infty \\
\le\ & (n\lambda_n)^{-1} O(\|\boldsymbol{\xi}_I\|_\infty + \|\boldsymbol{r}\|_\infty) + (n\lambda_n)^{-1} \|\boldsymbol{w}\|_\infty \\
&+ (n\lambda_n)^{-1} \|\boldsymbol{X}_{II}^T \boldsymbol{\Sigma}(\boldsymbol{\theta}^*) \boldsymbol{X}_I [\boldsymbol{X}_I^T \boldsymbol{\Sigma}(\boldsymbol{\theta}^*) \boldsymbol{X}_I]^{-1}\|_\infty \cdot \|\boldsymbol{\eta}\|_\infty,
\end{aligned}$$

because by (C4), $\|\boldsymbol{X}_{II}^T \boldsymbol{\Sigma}(\boldsymbol{\theta}^*) \boldsymbol{X}_I [\boldsymbol{X}_I^T \boldsymbol{\Sigma}(\boldsymbol{\theta}^*) \boldsymbol{X}_I]^{-1}\|_\infty = O(1)$.

It follows from (13), (A.7) and (A.9) that $(n\lambda_n)^{-1} O(\|\boldsymbol{\xi}_I\|_\infty + \|\boldsymbol{r}\|_\infty) = o(1)$. Meanwhile, by (13) and (A.14), $(n\lambda_n)^{-1}\|\boldsymbol{w}\|_\infty = o(1)$. By (A.7) and (C4),

$$(n\lambda_n)^{-1} \|\boldsymbol{X}_{II}^T \boldsymbol{\Sigma}(\boldsymbol{\theta}^*) \boldsymbol{X}_I [\boldsymbol{X}_I^T \boldsymbol{\Sigma}(\boldsymbol{\theta}^*) \boldsymbol{X}_I]^{-1}\|_\infty \cdot \|\boldsymbol{\eta}\|_\infty < (2\sqrt{2} L_p)^{-1}.$$

Therefore, (A.11) holds. At this moment, we have shown that $\hat{\boldsymbol{\beta}}$ satisfies (11).

Next, by the choice of λ_n, (12) holds for sufficiently large n. Therefore, by Theorem 1, we have shown that, in event $E_1 \cap E_2$, $\hat{\boldsymbol{\beta}}$ is a local maximizer of (9) that $\|\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_I^*\|_\infty \le n^{-\gamma}$ and $\hat{\boldsymbol{\beta}}_{II \cup III} = \mathbf{0}$. This completes the proof.

Web Appendix B: Additional Results in Data Analysis in Section 5.2

Table B1. Correlations between CD14 and (IFNA4, STAT1, TLR8) in each study

	GSE12288	GSE16561	GSE20129	GSE22255	GSE28829
			CD14		
IFNA4	-0.075	0.046	-0.013	0.449	-0.359
STAT1	0.339	0.062	0.044	0.055	0.507
TLR8	0.602	0.437	0.077	0.686	0.818

Table B2. Correlations between selections by meta lasso and separate lasso in GSE16561 and GSE28829

GSE16561

	CD14	CD86	CHUK	MAPK11	MAPK14	PIK3CG	PIK3R1	RAC1	STAT1	TLR2
STAT1	0.062	0.266	0.171	0.018	0.297	0.139	0.181	0.103	1.000	0.169
TLR8	0.437	0.185	0.674	0.118	0.679	0.017	0.161	0.481	0.251	0.741
	TLR7	TLR8	TNF	TRAF3						
STAT1	0.306	0.251	-0.014	-0.081						
TLR8	0.238	1.000	-0.113	-0.004						

GSE28829

	CD14	IFNAR2	IRF5	MAPK9
IFNA4	-0.359	-0.112	0.297	-0.208
STAT1	0.507	0.547	0.058	-0.169
TLR8	0.818	0.619	0.343	-0.583

Table B3. Gene selections of eight methods in four cardiovascular studies (excluding GSE20129)

Selections by meta lasso and separate lasso in each dataset

Datasets	meta lasso	separate lasso
GSE12288	CD40 CD86 CHUK IFNA2 IFNA21 IFNA4 IFNA8 IFNB1 IRF5 JUN LBP MAPK13 MAPK14 STAT1 TLR2 TLR7 TNF	none
GSE16561	CD14 CHUK JUN LBP MAPK11 PIK3CG PIK3R1 TLR7 TLR8	CD14 CD86 CHUK MAPK11 MAPK14 PIK3CG PIK3R1 RAC1 STAT1 TLR2 TLR7 TLR8 TNF TRAF3
GSE22255	IFNA2 JUN LBP MAPK14 PIK3R1 TLR8	none
GSE28829	CD14 IFNAR2 IRF5 PIK3CG	CD14 IFNAR2 IRF5 MAPK9

Selections by other methods in all datasets

Method	Gene list
stack lasso	none
group lasso	CD86 FOS IFNAR2 MAPK14 MAPK9 PIK3CA STAT1 TLR2 TLR7 TLR8 TNF
AW	AKT1 AKT3 CASP8 CCL5 CD14 CD40 CD80 CD86 CHUK FOS IFNAR1 IFNAR2 IKBKE IL1B IL8 IRAK1 IRF5 IRF7 JUN LBP LY96 MAP2K4 MAP3K7 MAP3K8 MAPK1 MAPK11 MAPK13 MAPK14 MAPK9 MYD88 PIK3CA PIK3CD PIK3CG PIK3R1 PIK3R5 RAC1 SPP1 STAT1 TBK1 TLR1 TLR2 TLR4 TLR5 TLR6 TLR7 TLR8 TNF TRAF3 TRAF6
Fisher	AKT1 AKT3 CASP8 CCL5 CD14 CD40 CD80 CD86 CHUK FOS IFNAR1 IFNAR2 IKBKE IL1B IL8 IRAK1 IRF5 IRF7 JUN LBP LY96 MAP2K3 MAP2K4 MAP3K7 MAP3K8 MAPK1 MAPK11 MAPK13 MAPK14 MAPK9 MYD88 PIK3CA PIK3CD PIK3CG PIK3R1 PIK3R5 RAC1 SPP1 STAT1 TBK1 TLR1 TLR2 TLR4 TLR5 TLR6 TLR7 TLR8 TNF TRAF3
FEM	AKT1 CD86 IFNA4 LBP MAP3K7 MAP3K8 MYD88 NFKB2 STAT1 TLR2 TLR4 TLR5 TLR7 TLR8
REM	none

References

Fan, J. and Lv, J. (2011). Nonconcave penalized likelihood with np-dimensionality. *Information Theory, IEEE Transactions* **57,** 5467–5484.

Zhou, N. and Zhu, J. (2010). Group variable selection via a hierarchical lasso and its oracle property. *Statistics and Its Interface* **3,** 557–574.