This supplementary file contains Sections 6 and 7. Technical proofs are provided in Section 6 and additional numerical results, Tables 6-10, are provided in Section 7.

As for the other supplementary materials,

code_simulation.R and main_functions.R were used for Section 3.1 and

code_data_analysis.R and main_functions_data_analysis.R were for Section 3.2.

6 Proofs of technical lemmas

Recall that $|S|$ and $|S \cup \{j\}|$ are less than or equal to K_n.

Proof of Lemma 1. First note that Assumption FY(1) implies that

$$C_1 < f_Y(y|(\boldsymbol{X}_S^T\boldsymbol{\beta}_S^*, X_j)^T) < C_2 \text{ on } (\boldsymbol{X}_S^T\boldsymbol{\beta}_S^* - D_hX_M - \delta_1, \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* + D_hX_M + \delta_1) \quad (58)$$

uniformly in S ($\mathcal{M} \not\subset S$) and $j \in S^c$. This is sufficient in this lemma.
(i) We have for some positive constant C_1,

$$\begin{aligned}
&|\mathrm{E}\{X_j\psi_\tau(Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^*) - X_j\psi_\tau(Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* - X_jh_{jS}^*)\}| \\
&\qquad = |\mathrm{E}[X_j\{I(Y \le \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* + X_jh_{jS}^*) - I(Y \le \boldsymbol{X}_S^T\boldsymbol{\beta}_S^*)\}]| \\
&\qquad \le C_1\mathrm{E}\{X_j^2\}|h_{jS}^*| = C_1|h_{jS}^*|.
\end{aligned}$$

We used (58) here. Hence the desired result follows from Assumption LB.
(ii) We employ Knight's identify :

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v \{I(u \le s) - I(u \le 0)\}ds. \quad (59)$$

With $u = Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* - X_jh_{jS}^*$ and $v = -X_jh_{jS}^*$ in (59), we have for some positive constant C_2,

$$\begin{aligned}
&\mathrm{E}\{\rho_\tau(Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^*) - \rho_\tau(Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* - X_jh_{jS}^*)\} \\
&\quad = \mathrm{E}\{X_jh_{jS}^*\psi_\tau(Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* - X_jh_{jS}^*)\} \\
&\qquad + \mathrm{E}\Big[\int_0^{-X_jh_{jS}^*} \{1(Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* - X_jh_{jS}^* \le s) - 1(Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* - X_jh_{jS}^* \le 0)\}ds\Big] \\
&\quad \ge C_2|h_{jS}^*|^2\mathrm{E}\{X_j^2\} = C_2|h_{jS}^*|^2.
\end{aligned}$$

The first term in the second line is 0 and we used (58) to evaluate the second term in the second line. Hence the desired result follows from (i) and Assumption LB. □

Remark 3 We can relax Assumption B on the boundedness of X_j a little by strengthening some other assumptions. For example, in (i) of Lemma 1,

$$f_Y(y|(\boldsymbol{X}_S^T\boldsymbol{\beta}_S^*, X_j)^T) < C_1$$

is sufficient and we do not have to use the boundedness of X_j. In (ii) of Lemma 1, we could do with (58) in the proof of Lemma 1 and

$$\mathrm{E}[X_j^2I\{|X_j| \le X_M\}] > C_2$$

uniformly in j. We used the boundedness of X_j in the proofs of Lemma 2 and Theorem 2. However, we can let X_M go to ∞ slowly enough there. Therefore we could use an assumption like

$$|X_j| \le X_{M1} \qquad \text{and} \qquad \mathrm{E}[X_j^2I\{|X_j| \le X_{M2}\}] > C_3$$

uniformly in j, where X_{M2} is fixed and we allow X_{M1} to go to ∞ slowly enough. If we do so, X_{M1} will appear in Lemmas 2-4 and some other assumptions. However, there will be no essential change in our main results.

Proof of Lemma 2. The proof consists of three steps and Lemma 5 at the end of this section.

First define $B_S(\delta) \in \mathbb{R}^{|S|}$ and $\varDelta_S(\delta)$ for $\delta \in \mathbb{R}$ by

$$B_S(\delta) := \{\boldsymbol{\beta}_S \in \mathbb{R}^{|S|} \mid \|\boldsymbol{\beta}_S - \boldsymbol{\beta}_S^*\| \le \delta\}$$

and

$$\varDelta_S(\delta) := \sup_{\boldsymbol{\beta}_S \in B_S(\delta)} |\{L_n(\boldsymbol{X}_S^T\boldsymbol{\beta}_S) - L_n(\boldsymbol{X}_S^T\boldsymbol{\beta}_S^*)\} - \{L_S(\boldsymbol{\beta}_S) - L_S(\boldsymbol{\beta}_S^*)\}|,$$

where δ is to be specified later in this proof.

(1) We prove that we have for some positive constant $D_{(1)}$,

$$L_S(\boldsymbol{\beta}_S) - L_S(\boldsymbol{\beta}_S^*) \ge D_{(1)}\|\boldsymbol{\beta}_S - \boldsymbol{\beta}_S^*\|^2 \tag{60}$$

if $X_M K_n^{1/2}\|\boldsymbol{\beta}_S - \boldsymbol{\beta}_S^*\| \le \gamma_n \to 0$, where $\{\gamma_n\}$ is a suitable sequence tending to 0 slowly enough.

Applying Knight's identity in (59) with $u = Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^*$ and $v = \boldsymbol{X}_S^T(\boldsymbol{\beta}_S - \boldsymbol{\beta}_S^*)$, we have for some positive constants C_1 and $D_{(1)}$,

$$\begin{aligned}
&L_S(\boldsymbol{\beta}_S) - L_S(\boldsymbol{\beta}_S^*)\\
&\qquad = \mathrm{E}\{\boldsymbol{X}_S^T(\boldsymbol{\beta}_S^* - \boldsymbol{\beta}_S)\psi_\tau(Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^*)\}\\
&\qquad\quad + \mathrm{E}\Big[\int_0^{\boldsymbol{X}_S^T(\boldsymbol{\beta}_S - \boldsymbol{\beta}_S^*)} \{1(Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* \le s) - 1(Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* \le 0)\}ds\Big]\\
&\qquad \ge C_1(\boldsymbol{\beta}_S - \boldsymbol{\beta}_S^*)^T\mathrm{E}\{\boldsymbol{X}_S\boldsymbol{X}_S^T\}(\boldsymbol{\beta}_S - \boldsymbol{\beta}_S^*)\\
&\qquad \ge D_{(1)}\|\boldsymbol{\beta}_S - \boldsymbol{\beta}_S^*\|^2.
\end{aligned}$$

We used Assumptions FY(1)(2) and X(1) here. Note that Assumption FY(1) in a neighborhood of $\boldsymbol{X}_S^T\boldsymbol{\beta}_S^*$ is sufficient here.

(2) We prove that

$$\mathrm{P}(\|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S^*\| > \delta) \le \mathrm{P}(\varDelta_S(\delta) \ge D_{(1)}\delta^2/4), \tag{61}$$

where

$$\delta = \frac{16|S|^{1/2}}{D_{(1)}n^{1/2}}(1 + \eta_S) \quad \text{and} \quad \eta_S = D_{(1)}(1 + X_M\sqrt{3|S|\log p_n}).$$

Define $\boldsymbol{\beta}_{\alpha S}$ by

$$\boldsymbol{\beta}_{\alpha S} := \alpha\widehat{\boldsymbol{\beta}}_S + (1-\alpha)\boldsymbol{\beta}_S^*,$$

where $\alpha = \delta/(\delta + \|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S^*\|)$. Then we have

$$\|\boldsymbol{\beta}_{\alpha S} - \boldsymbol{\beta}_S^*\| = \alpha\|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S^*\| \le \delta. \tag{62}$$

By the convexity of $L_n(\boldsymbol{X}_S^T\boldsymbol{\beta}_S)$ w.r.t. $\boldsymbol{\beta}_S$, we have

$$L_n(\boldsymbol{X}_S^T\boldsymbol{\beta}_{\alpha S}) \le \alpha L_n(\boldsymbol{X}_S^T\widehat{\boldsymbol{\beta}}_S) + (1-\alpha)L_n(\boldsymbol{X}_S^T\boldsymbol{\beta}_S^*) \le L_n(\boldsymbol{X}_S^T\boldsymbol{\beta}_S^*). \tag{63}$$

We will exploit (63) to give an upper bound of $L_S(\boldsymbol{\beta}_{\alpha S}) - L_S(\boldsymbol{\beta}_S^*)$. Note that it is written as

$$\begin{aligned} L_S(\boldsymbol{\beta}_{\alpha S}) - L_S(\boldsymbol{\beta}_S^*) &= [\{L_n(\boldsymbol{X}_S^T\boldsymbol{\beta}_S^*) - L_n(\boldsymbol{X}_S^T\boldsymbol{\beta}_{\alpha S})\} - \{L_S(\boldsymbol{\beta}_S^*) - L_S(\boldsymbol{\beta}_{\alpha S})\}] \\ &\quad + \{L_n(\boldsymbol{X}_S^T\boldsymbol{\beta}_{\alpha S}) - L_n(\boldsymbol{X}_S^T\boldsymbol{\beta}_S^*)\} \\ &\le \Delta_S(\delta) \end{aligned} \tag{64}$$

We used (62) and (63) here.

Now notice that

$$\|\boldsymbol{\beta}_{\alpha S} - \boldsymbol{\beta}_S^*\| \le \frac{\delta}{2} \Rightarrow \|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S^*\| \le \delta.$$

Hence we have

$$\|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S^*\| > \delta \Rightarrow \|\boldsymbol{\beta}_{\alpha S} - \boldsymbol{\beta}_S^*\| > \frac{\delta}{2}. \tag{65}$$

(65), (60) with $\boldsymbol{\beta}_S = \boldsymbol{\beta}_{\alpha S}$, (62), and (64) imply (61).

(3) We verify

$$\sum_S \mathrm{P}(\Delta_S(\delta) \ge D_{(1)}\delta^2/4) \le \sum_{q=1}^{\infty} n^{-2q}. \tag{66}$$

The RHS of (66) goes to 0 as $n \to 0$. We should obtain a suitable upper bound of $\mathrm{P}(\Delta_S(\delta) \ge D_{(1)}\delta^2/4)$.

By Massart's concentration theorem (see Theorem 14.2 of Bühlmann and van de Geer (2011)) and Lemma 5 at the end of this section, we have with $R^2 = 4X_M^2|S|\delta^2$,

$$\mathrm{P}\Big[\Delta_S(\delta) \ge \mathrm{E}\{\Delta_S(\delta)\} + R\sqrt{\frac{8t}{n}}\Big] \le \exp(-t) \tag{67}$$

and

$$\begin{aligned} \mathrm{E}\{\Delta_S(\delta)\} + R\sqrt{\frac{8t}{n}} &\le 4\delta\Big(\frac{|S|}{n}\Big)^{1/2} + 2\delta X_M\Big(\frac{|S|}{n}\Big)^{1/2}\cdot 2\sqrt{2t} \\ &\le 4\delta\Big(\frac{|S|}{n}\Big)^{1/2}(1 + X_M\sqrt{2t}) \le D_{(1)}\frac{\delta^2}{4}. \end{aligned} \tag{68}$$

In (68), we took $t = 3|S|\log p_n$ and used the definition of δ in (61).

Finally we have by (67) and (68),

$$\sum_S \mathrm{P}\{\Delta_S(\delta) \ge D_{(1)}\delta^2/4\} \le \sum_{q=1}^{\infty} p_n^q \exp(-3q\log p_n) \le \sum_{q=1}^{\infty} n^{-2q}.$$

Hence we have established (66).

The desired result follows from (61) and (66) and the proof of Lemma 2 is complete. □

Proof of Lemma 3. We have

$$\begin{aligned} L_n(\boldsymbol{X}_S^T\widehat{\boldsymbol{\beta}}_S) - L_S(\boldsymbol{\beta}_S^*) &= \{L_n(\boldsymbol{X}_S^T\widehat{\boldsymbol{\beta}}_S) - L_n(\boldsymbol{X}_S^T\boldsymbol{\beta}_S^*)\} + \{L_n(\boldsymbol{X}_S^T\boldsymbol{\beta}_S^*) - L_S(\boldsymbol{\beta}_S^*)\} \\ &:= A + B. \end{aligned}$$

First we deal with A. Then

$$A = [\{L_n(\boldsymbol{X}_S^T \widehat{\boldsymbol{\beta}}_S) - L_n(\boldsymbol{X}_S^T \boldsymbol{\beta}_S^*)\} - \{L_S(\widehat{\boldsymbol{\beta}}_S) - L_S(\boldsymbol{\beta}_S^*)\}] + \{L_S(\widehat{\boldsymbol{\beta}}_S) - L_S(\boldsymbol{\beta}_S^*)\}$$
$$:= A_1 + A_2.$$

As in the proof of Lemma 2 , we have

$$|A_1| \leq \Delta_S(\delta)$$

with the same δ as in Lemma 2 if $\|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S^*\| \leq \delta$, which holds uniformly in S with probability tending to 1. Thus as we have evaluated $\Delta_S(\delta)$ in the proof of Lemma 2, we have for some positive constant C_1,

$$|A_1| \leq C_1 \delta^2 \tag{69}$$

uniformly in S with probability tending to 1.

As in the proof of Lemma 2, we have for some positive constant C_2,

$$|A_2| \leq \|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S^*\|^2 \leq C_2 \delta^2 \tag{70}$$

uniformly in S with probability tending to 1.

As for B, repeated use of Bernstein's inequality yields

$$|B| \leq C_3 \sqrt{\frac{|S| \log p_n}{n}} \tag{71}$$

uniformly in S with probability tending to 1 for some positive constant C_3.

The desired result follows from (69)-(71). □

Proof of Lemma 4. We have

$$\begin{aligned} L_n(\boldsymbol{X}_S^T \widehat{\boldsymbol{\beta}}_S + X_j h_{jS}^*) &- L_{S\cup\{j\}}((\boldsymbol{\beta}_S^{*T}, h_{jS}^*)^T) \\ &= \{L_n(\boldsymbol{X}_S^T \widehat{\boldsymbol{\beta}}_S + X_j h_{jS}^*) - L_n(\boldsymbol{X}_S^T \boldsymbol{\beta}_S^* + X_j h_{jS}^*)\} \\ &\quad + \{L_n(\boldsymbol{X}_S^T \boldsymbol{\beta}_S^* + X_j h_{jS}^*) - L_{S\cup\{j\}}((\boldsymbol{\beta}_S^{*T}, h_{jS}^*)^T\} \\ &:= A + B. \end{aligned}$$

First we deal with A, which is written as

$$\begin{aligned} A = [\{&L_n(\boldsymbol{X}_S^T \widehat{\boldsymbol{\beta}}_S + X_j h_{jS}^*) - L_n(\boldsymbol{X}_S^T \boldsymbol{\beta}_S^* + X_j h_{jS}^*)\} \\ &- \{L_{S\cup\{j\}}((\widehat{\boldsymbol{\beta}}_S^T, h_{jS}^*)^T) - L_{S\cup\{j\}}((\boldsymbol{\beta}_S^{*T}, h_{jS}^*)^T)\}] \\ &+ \{L_{S\cup\{j\}}((\widehat{\boldsymbol{\beta}}_S^T, h_{jS}^*)^T) - L_{S\cup\{j\}}((\boldsymbol{\beta}_S^{*T}, h_{jS}^*)^T)\} \\ :=& A_1 + A_2 \end{aligned}$$

As in the proof of Lemmas 2 and 3, we have

$$|A_1| \leq \Delta_{S\cup\{j\}}(\delta) \leq C_1 \delta^2 \tag{72}$$

uniformly in S and $j \in S^c$ with probability tending to 1. When we verify (72), we should take $(\boldsymbol{\beta}_S^{*T}, h_{jS}^*)^T$ as the center of $B_{S\cup\{j\}}(\delta)$ in the proof of Lemma 2 and we compare it with $(\widehat{\boldsymbol{\beta}}_S^T, h_{jS}^*)^T$

Next we evaluate A_2. We apply Knight's identity with $u = Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* - X_j h_{jS}^*$ and $v = \boldsymbol{X}_S^T\widehat{\boldsymbol{\beta}}_S^T - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^*$ and obtain

$$\begin{aligned}A_2 &= \mathrm{E}\{\boldsymbol{X}_S^T(\boldsymbol{\beta}_S^* - \boldsymbol{\beta}_S)\psi_\tau(Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* - X_j h_{jS}^*)\}|_{\boldsymbol{\beta}_S=\widehat{\boldsymbol{\beta}}_S}\\ &\quad + \mathrm{E}\Big[\int_0^{\boldsymbol{X}^T(\boldsymbol{\beta}_S-\boldsymbol{\beta}_S^*)}\{I(Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* - X_j h_{jS}^* \le s) - I(Y - \boldsymbol{X}_S^T\boldsymbol{\beta}_S^* - X_j h_{jS}^* \le 0)\}ds\Big]\Big|_{\boldsymbol{\beta}_S=\widehat{\boldsymbol{\beta}}_S}\\ &:= A_{21} + A_{22}.\end{aligned}$$

By Assumption X(1) and the Cauchy-Schwarz inequality, we have

$$|A_{21}| \le C_2\|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S^*\| \tag{73}$$

uniformly in S and $j \in S^c$ for some positive constant C_2.

By Assumptions FY(1) and X(1), we have

$$|A_{22}| \le C_3\|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S^*\|^2 \tag{74}$$

uniformly in S and $j \in S^c$ for some positive constant C_3.

As for B, repeated use of Bernstein's inequality yields

$$|B| \le C_4\sqrt{\frac{|S|\log p_n}{n}} \tag{75}$$

uniformly in S and $j \in S^c$ with probability tending to 1 for some positive constant C_4.

Hence the desired result follows from Lemma 2 and (72)-(75). □

Lemma 5 is employed in the proof of Lemma 2 and the same result is proved in the proof of lemma C of Kong et al. (2019). The proof is based on the standard symmetrization and contraction argument. See Sections 14.7-8 of Bühlmann and van de Geer (2011). Thus we omit the proof.

Lemma 5 *Under the same assumptions as in Lemma 2, we have*

$$\mathrm{E}\{\Delta_S(\delta)\} \le 4\delta\Big(\frac{|S|}{n}\Big)^{1/2}.$$

7 Additional results for simulations and data analysis

This section contains simulation results under $(n, p) = (400, 4000)$ for Examples 1–3 in Section 3.1. The simulation results are presented in Tables 6–8 and a correlation of selected genes is demonstrated in Table 10. From Tables 6–8, it can be seen that all results deteriorates due to the increasing of p, but the pattern is almost same as that we concluded in Section 3.1.

Table 6 Simulation results for Example 1 with $(n, p) = (400, 4000)$.

	X_2	X_7	X_{13}	X_{16}	X_{21}	Sure	TP	FP	time
				$\tau = 0.3, \mathcal{M} = \{2, 7, 13, 16, 21\}$					
CQU	1	100	99	100	69	0	3.69	63.31	11.87
Lasso	24	100	99	100	100	24	4.23	43.44	55.82
ALasso	11	100	99	100	100	11	4.10	0.28	56.26
SCAD	4	100	99	100	99	4	4.02	0.05	62.09
MCP	12	100	99	100	99	12	4.10	0.11	79.69
gSC(1)+T_1	0	100	98	100	94	0	3.92	0.04	14.58
gSC(1)+T_2	32	100	100	100	100	32	4.32	0.64	18.14
gSC(1)+T_3	41	100	100	100	100	41	4.41	1.55	21.68
gSC(1)	46	100	100	100	100	100	4.00	26.00	116.47
gSC(25)+T_1	1	100	99	100	93	1	3.93	0.03	14.50
gSC(25)+T_2	41	100	100	100	100	41	4.41	0.55	18.14
gSC(25)+T_3	51	100	100	100	100	51	4.51	1.45	21.75
gSC(25)	59	100	100	100	100	100	4.00	26.00	118.44
gSC(m_n)+T_1	1	100	99	100	93	1	3.93	0.03	14.62
gSC(m_n)+T_2	41	100	100	100	100	41	4.41	0.55	18.31
gSC(m_n)+T_3	51	100	100	100	100	51	4.51	1.45	21.91
gSC(m_n)	59	100	100	100	100	100	4.00	26.00	123.89
FR+T_1	1	100	99	100	93	1	3.93	0.03	16.44
FR+T_2	41	100	100	100	100	41	4.41	0.55	21.27
FR+T_3	51	100	100	100	100	51	4.51	1.45	26.29
FR	58	100	100	100	100	100	4.00	26.00	271.30
FR-PQU+QBIC	55	100	100	100	100	55	4.55	1.09	355.28
FR-PQU	69	100	100	100	100	100	4.00	26.00	355.28
				$\tau = 0.5, \mathcal{M} = \{7, 13, 16, 21\}$					
CQU	1	100	100	100	71	71	3.71	63.29	11.87
Lasso	3	100	99	100	100	99	3.99	33.42	30.00
ALasso	0	100	99	100	100	99	3.99	0.05	30.44
SCAD	0	100	99	100	100	99	3.99	0.06	34.07
MCP	0	100	99	100	100	99	3.99	0.03	41.03
gSC(1)+T_1	0	100	100	100	97	97	3.97	0.19	15.34
gSC(1)+T_2	0	100	100	100	100	100	4.00	1.16	18.86
gSC(1)+T_3	1	100	100	100	100	100	4.00	2.16	22.44
gSC(1)	1	100	100	100	100	100	3.00	27.00	119.25
gSC(25)+T_1	0	100	100	100	97	97	3.97	0.19	15.23
gSC(25)+T_2	0	100	100	100	100	100	4.00	1.16	18.95
gSC(25)+T_3	0	100	100	100	100	100	4.00	2.16	22.58
gSC(25)	3	100	100	100	100	100	3.00	27.00	121.22
gSC(m_n)+T_1	0	100	100	100	97	97	3.97	0.19	15.42
gSC(m_n)+T_2	0	100	100	100	100	100	4.00	1.16	19.04
gSC(m_n)+T_3	0	100	100	100	100	100	4.00	2.16	22.81
gSC(m_n)	1	100	100	100	100	100	3.00	27.00	125.50
FR+T_1	0	100	100	100	97	97	3.97	0.19	17.98
FR+T_2	0	100	100	100	100	100	4.00	1.16	23.01
FR+T_3	1	100	100	100	100	100	4.00	2.16	28.45
FR	2	100	100	100	100	100	3.00	27.00	286.11
FR-PQU+QBIC	1	100	100	100	100	100	4.00	0.26	359.46
FR-PQU	3	100	100	100	100	100	3.00	27.00	359.46
				$\tau = 0.7, \mathcal{M} = \{2, 7, 13, 16, 21\}$					
CQU	7	100	100	100	65	65	3.65	63.35	11.85
Lasso	27	100	99	100	100	99	3.99	44.85	68.73
ALasso	16	100	99	100	100	99	3.99	0.66	69.18
SCAD	11	100	99	100	99	98	3.98	0.35	74.00
MCP	16	100	99	100	99	98	3.98	0.36	84.29
gSC(1)+T_1	1	100	100	100	91	91	3.91	0.03	14.57
gSC(1)+T_2	30	100	100	100	100	100	4.00	0.94	18.09
gSC(1)+T_3	39	100	100	100	100	100	4.00	1.94	21.61
gSC(1)	42	100	100	100	100	100	4.00	26.00	112.51
gSC(25)+T_1	2	100	100	100	91	91	3.91	0.03	14.55
gSC(25)+T_2	37	100	100	100	100	100	4.00	0.94	18.13
gSC(25)+T_3	44	100	100	100	100	100	4.00	1.94	21.76
gSC(25)	53	100	100	100	100	100	4.00	26.00	113.98
gSC(m_n)+T_1	2	100	100	100	91	91	3.91	0.03	14.55
gSC(m_n)+T_2	38	100	100	100	100	100	4.00	0.94	18.28
gSC(m_n)+T_3	44	100	100	100	100	100	4.00	1.94	21.98
gSC(m_n)	54	100	100	100	100	100	4.00	26.00	118.66
FR+T_1	2	100	100	100	91	91	3.91	0.03	17.02
FR+T_2	38	100	100	100	100	100	4.00	0.94	21.98
FR+T_3	44	100	100	100	100	100	4.00	1.94	27.16
FR	53	100	100	100	100	100	4.00	26.00	268.49
FR-PQU+QBIC	50	100	100	100	100	100	4.00	1.61	352.10
FR-PQU	61	100	100	100	100	100	4.00	26.00	352.10

Table 7 Simulation results for Example 2 with $(n, p) = (400, 4000)$.

	X_2	X_3	X_4	X_5	X_6	X_{21}	Sure	TP	FP	time
				$\tau = 0.3, \mathcal{M} = \{2, 3, 4, 5, 6, 21\}$						
CQU	64	64	64	59	67	0	0	3.18	63.82	12.30
Lasso	91	92	89	91	87	3	3	4.53	32.97	75.41
ALasso	69	69	73	64	65	0	0	3.40	0.09	76.90
SCAD	50	41	44	47	47	0	0	2.29	0.56	131.50
MCP	46	50	53	44	50	1	1	2.44	0.61	193.96
gSC(1)+T_1	52	45	49	49	53	0	0	2.48	0.84	12.91
gSC(1)+T_2	62	61	61	60	65	2	0	3.11	1.21	16.72
gSC(1)+T_3	70	71	69	69	73	9	5	3.61	1.71	20.61
gSC(1)	85	84	87	82	86	38	27	4.62	25.38	116.20
gSC(25)+T_1	67	59	57	56	66	0	0	3.05	0.30	13.01
gSC(25)+T_2	80	81	79	76	81	1	0	3.98	0.37	16.93
gSC(25)+T_3	89	90	85	88	89	23	14	4.64	0.71	20.89
gSC(25)	96	96	94	96	96	61	55	5.39	24.61	117.14
gSC(m_n)+T_1	67	59	57	56	66	0	0	3.05	0.30	13.31
gSC(m_n)+T_2	80	81	78	77	81	1	0	3.98	0.37	17.28
gSC(m_n)+T_3	89	90	84	88	89	23	14	4.63	0.72	21.19
gSC(m_n)	97	96	94	96	96	61	56	5.40	24.60	119.95
FR+T_1	67	59	57	56	66	0	0	3.05	0.30	15.10
FR+T_2	80	81	78	77	81	1	0	3.98	0.37	20.17
FR+T_3	89	90	84	88	89	23	14	4.63	0.72	25.51
FR	97	96	94	96	96	62	56	5.41	24.59	279.64
FR-PQU+QBIC	38	35	42	42	34	0	0	1.91	1.11	454.71
FR-PQU	90	88	84	81	80	45	31	4.68	25.32	454.71
				$\tau = 0.5, \mathcal{M} = \{2, 3, 4, 5, 6\}$						
CQU	69	67	66	66	70	0	13	3.38	63.62	12.20
Lasso	94	96	94	97	90	1	89	4.71	34.91	52.58
ALasso	88	89	86	87	86	0	80	4.36	0.02	53.28
SCAD	82	83	78	80	77	1	64	4.00	0.05	62.09
MCP	82	81	76	83	79	0	67	4.01	0.09	97.90
gSC(1)+T_1	83	77	76	81	75	0	0	3.92	0.07	18.60
gSC(1)+T_2	99	99	99	97	98	0	93	4.92	0.07	23.14
gSC(1)+T_3	99	100	100	100	100	1	99	4.99	1.00	27.94
gSC(1)	100	100	100	100	100	3	100	5.00	25.00	118.95
gSC(25)+T_1	89	74	79	77	77	0	1	3.96	0.05	18.87
gSC(25)+T_2	100	99	98	98	100	0	96	4.95	0.06	23.58
gSC(25)+T_3	100	100	99	100	100	1	99	4.99	1.02	28.18
gSC(25)	100	100	100	100	100	2	100	5.00	25.00	120.27
gSC(m_n)+T_1	89	74	79	77	77	0	1	3.96	0.05	19.03
gSC(m_n)+T_2	100	99	98	98	100	0	96	4.95	0.06	23.84
gSC(m_n)+T_3	100	100	99	100	100	1	99	4.99	1.02	28.54
gSC(m_n)	100	100	100	100	100	2	100	5.00	25.00	123.40
FR+T_1	89	74	79	77	77	0	1	3.96	0.05	21.71
FR+T_2	100	99	98	98	100	0	96	4.95	0.06	28.00
FR+T_3	100	100	99	100	100	1	99	4.99	1.02	34.76
FR	100	100	100	100	100	2	100	5.00	25.00	286.61
FR-PQU+QBIC	68	70	66	70	66	0	51	3.40	0.80	457.61
FR-PQU	99	100	98	99	98	3	97	4.94	25.06	457.61
				$\tau = 0.7, \mathcal{M} = \{2, 3, 4, 5, 6, 21\}$						
CQU	48	53	51	49	52	13	0	2.66	64.34	11.03
Lasso	89	84	88	87	84	53	48	4.85	51.35	70.29
ALasso	73	67	70	73	74	12	8	3.69	0.31	72.40
SCAD	52	48	55	51	49	7	3	2.62	0.91	134.84
MCP	49	45	55	44	42	4	2	2.39	1.12	197.56
gSC(1)+T_1	59	54	59	59	57	4	0	2.92	0.91	16.45
gSC(1)+T_2	72	66	78	78	79	8	0	3.81	1.02	20.67
gSC(1)+T_3	85	78	87	85	84	23	18	4.42	1.41	24.94
gSC(1)	98	92	95	94	93	45	40	5.17	24.83	113.34
gSC(25)+T_1	72	63	63	66	63	8	0	3.35	0.55	16.85
gSC(25)+T_2	85	81	85	85	81	10	0	4.27	0.63	21.19
gSC(25)+T_3	91	86	92	91	88	47	39	4.95	0.95	25.56
gSC(25)	98	95	98	96	97	60	55	5.44	24.56	114.57
gSC(m_n)+T_1	72	63	63	66	63	8	0	3.35	0.55	17.11
gSC(m_n)+T_2	85	81	85	85	81	10	0	4.27	0.63	21.49
gSC(m_n)+T_3	91	86	92	91	88	47	39	4.95	0.95	25.78
gSC(m_n)	98	94	98	96	97	61	55	5.44	24.56	117.32
FR+T_1	72	63	63	66	63	8	0	3.35	0.55	19.70
FR+T_2	85	81	85	85	81	10	0	4.27	0.63	25.45
FR+T_3	91	86	92	91	88	47	39	4.95	0.95	31.51
FR	98	95	97	96	97	62	56	5.45	24.55	271.31
FR-PQU+QBIC	36	40	40	35	39	0	0	1.90	1.62	400.49
FR-PQU	90	79	81	86	82	37	26	4.55	25.45	400.49

Table 8 Simulation results for Example 3 with $(n, p) = (400, 4000)$.

	X_2	X_3	X_4	Sure	TP	FP	time
			$\tau = 0.3, \mathcal{M} = \{2, 3, 4\}$				
CQU	100	2	0	0	1.02	65.98	12.25
Lasso	100	99	1	1	2.00	79.91	68.58
ALasso	100	99	0	0	1.99	12.83	69.41
SCAD	100	63	25	25	1.88	6.99	207.01
MCP	100	100	0	0	2.00	11.11	257.40
gSC(1)+T_1	100	35	0	0	1.35	1.64	11.45
gSC(1)+T_2	100	99	0	0	1.99	2.00	15.22
gSC(1)+T_3	100	100	0	0	2.00	2.99	19.06
gSC(1)	100	100	0	0	2.00	28.00	116.27
gSC(25)+T_1	100	46	15	15	1.61	1.54	12.13
gSC(25)+T_2	100	100	15	15	2.15	2.00	15.94
gSC(25)+T_3	100	100	15	15	2.15	3.00	19.88
gSC(25)	100	100	15	15	2.15	27.85	117.48
gSC(m_n)+T_1	100	57	27	27	1.84	1.43	12.82
gSC(m_n)+T_2	100	100	30	30	2.30	2.03	16.90
gSC(m_n)+T_3	100	100	31	31	2.31	3.05	20.99
gSC(m_n)	100	100	31	31	2.31	27.69	120.75
FR+T_1	100	61	31	31	1.92	1.39	14.52
FR+T_2	100	100	61	61	2.61	2.30	22.77
FR+T_3	100	100	100	100	3.00	3.54	31.72
FR	100	100	100	100	3.00	27.00	278.48
FR-PQU+QBIC	100	100	0	0	2.00	13.39	508.39
FR-PQU	100	100	1	1	2.01	27.99	508.39
			$\tau = 0.5, \mathcal{M} = \{2, 3, 4\}$				
CQU	100	0	0	0	1.00	66.00	12.15
Lasso	100	100	15	15	2.15	101.46	68.24
ALasso	100	100	10	10	2.10	13.95	69.43
SCAD	100	76	63	63	2.39	2.91	159.46
MCP	100	99	9	9	2.08	9.69	214.04
gSC(1)+T_1	100	41	2	2	1.43	1.58	13.84
gSC(1)+T_2	100	100	2	2	2.02	1.99	18.42
gSC(1)+T_3	100	100	2	2	2.02	2.99	23.05
gSC(1)	100	100	2	2	2.02	27.98	119.21
gSC(25)+T_1	100	54	23	23	1.77	1.45	14.89
gSC(25)+T_2	100	100	24	24	2.24	2.00	19.60
gSC(25)+T_3	100	100	24	24	2.24	3.00	24.22
gSC(25)	100	100	24	24	2.24	27.76	120.62
gSC(m_n)+T_1	100	60	31	31	1.91	1.39	15.50
gSC(m_n)+T_2	100	100	40	40	2.40	2.08	21.00
gSC(m_n)+T_3	100	100	41	41	2.41	3.10	25.75
gSC(m_n)	100	100	41	41	2.41	27.59	123.64
FR+T_1	100	61	32	32	1.93	1.38	17.14
FR+T_2	100	100	61	61	2.61	2.28	26.73
FR+T_3	100	100	100	100	3.00	4.00	40.75
FR	100	100	100	100	3.00	27.00	284.64
FR-PQU+QBIC	100	100	1	1	2.01	13.23	497.19
FR-PQU	100	100	3	3	2.03	27.97	497.19
			$\tau = 0.7, \mathcal{M} = \{2, 3, 4\}$				
CQU	100	3	0	0	1.03	65.97	10.85
Lasso	100	99	5	5	2.04	85.44	72.53
ALasso	100	99	3	3	2.02	13.59	73.30
SCAD	98	68	23	23	1.89	7.18	191.10
MCP	100	100	0	0	2.00	10.58	226.89
gSC(1)+T_1	100	27	0	0	1.27	1.68	12.70
gSC(1)+T_2	100	99	0	0	1.99	1.96	17.00
gSC(1)+T_3	100	100	0	0	2.00	2.95	21.23
gSC(1)	100	100	0	0	2.00	28.00	115.62
gSC(25)+T_1	100	36	9	9	1.45	1.60	13.20
gSC(25)+T_2	100	100	9	9	2.09	1.96	17.48
gSC(25)+T_3	100	100	9	9	2.09	2.96	21.64
gSC(25)	100	100	9	9	2.09	27.91	116.79
gSC(m_n)+T_1	100	47	22	22	1.69	1.49	13.77
gSC(m_n)+T_2	100	100	23	23	2.23	1.97	18.13
gSC(m_n)+T_3	100	100	23	23	2.23	2.97	22.51
gSC(m_n)	100	100	23	23	2.23	27.77	119.60
FR+T_1	100	53	28	28	1.81	1.43	15.76
FR+T_2	100	100	53	53	2.53	2.21	24.33
FR+T_3	100	100	100	100	3.00	3.55	36.02
FR	100	100	100	100	3.00	27.00	274.81
FR-PQU+QBIC	100	100	1	1	2.01	13.66	433.54
FR-PQU	100	100	1	1	2.01	27.99	433.54

Table 9 Prediction analysis for the gene expression dataset using the forward-type algorithms without stopping rules, followed (or not followed) by a regularized method. Values in parentheses are estimated standard deviation.

Screen	Regularization	τ = 0.3		τ = 0.5		τ = 0.7	
		PE	Size	PE	Size	PE	Size
gSC(1)	-	0.277(0.075)	30(0.0)	0.307(0.061)	30(0.0)	0.254(0.056)	30(0.0)
	Lasso	0.257(0.054)	4(3.1)	0.246(0.054)	4(3.1)	0.211(0.044)	4(3.0)
	SCAD	0.265(0.062)	4(3.1)	0.258(0.054)	4(3.1)	0.221(0.043)	4(3.1)
	MCP	0.257(0.055)	4(3.1)	0.254(0.053)	4(3.1)	0.221(0.043)	4(3.1)
gSC(m_n)	-	0.285(0.064)	30(0.0)	0.306(0.061)	30(0.0)	0.295(0.068)	30(0.0)
	Lasso	0.243(0.060)	5(9.0)	0.255(0.071)	5(9.0)	0.194(0.056)	5(9.0)
	SCAD	0.247(0.060)	5(6.4)	0.241(0.069)	5(6.4)	0.199(0.050)	5(6.4)
	MCP	0.247(0.064)	4(6.4)	0.243(0.068)	4(6.4)	0.202(0.048)	4(6.4)
FR	-	0.307(0.085)	30(0.0)	0.330(0.070)	30(0.0)	0.330(0.094)	30(0.0)
	Lasso	0.243(0.067)	6(10.6)	0.268(0.061)	6(10.6)	0.206(0.064)	6(10.6)
	SCAD	0.248(0.063)	5(9.0)	0.255(0.057)	5(9.0)	0.213(0.061)	5(9.0)
	MCP	0.243(0.071)	5(8.3)	0.252(0.059)	5(8.4)	0.196(0.057)	5(8.4)
FR-PQU	-	0.301(0.063)	30(0.0)	0.298(0.074)	30(0.0)	0.298(0.062)	30(0.0)
	LASSO	0.283(0.065)	2(2.7)	0.306(0.058)	2(2.7)	0.258(0.057)	2(2.7)
	SCAD	0.263(0.070)	3(3.2)	0.290(0.056)	3(3.2)	0.249(0.062)	3(3.2)
	MCP	0.258(0.070)	4(2.8)	0.283(0.057)	4(2.8)	0.234(0.062)	4(2.8)

Table 10 Pairwise correlation between genes in {YXLC, YXLD, YXLE, YXLJ}, genes in {XHLA, XHLB, XTRA} and genes in {ARGF, ARGJ}.

	YXLC	YXLD	YXLE	YXLJ	XHLA	XHLB	XTRA	ARGF	ARGJ
YXLC	1.00	0.98	0.97	0.87					
YXLD	0.98	1.00	0.98	0.90					
YXLE	0.97	0.98	1.00	0.87					
YXLJ	0.87	0.90	0.87	1.00					
XHLA					1.00	0.98	0.76		
XHLB					0.98	1.00	0.71		
XTRA					0.76	0.71	1.00		
ARGF								1.00	0.92
ARGJ								0.92	1.00