Xia, F., Canovas, P., Guadagno, T., and Altieri, D. (2008), "A survivin-ran complex regulates spindle formation in tumor cells," *Mol. Cell. Biol.*, 28, 5299–5311.

Yuan, M. (2010), "High dimensional inverse covariance matrix estimation via linear programming," *Journal of Machine Learning Research*, 11, 2261 –2286.

Yuan, M. and Lin, Y. (2007), "Model selection and estimation in the Gaussian graphical model," *Biometrika*, 94, 19–35.

Zou, H. (2006), "The adaptive lasso and its oracle properties," *Journal of American Statistical Association*, 101, 1418–1429.

Zou, H. and Li, R. (2008), "One-step sparse estimates in nonconcave penalized likelihood models," *Annals of Statistics*, 36, 1108–1126.

Supplementary Materials

Code

MATLAB code for simulation studies in this article are provided.

Comparison of JGGM and separate approaches

We compared AUC values from the JGGM and the separate approaches (Figure 3). 150, 134 and 149 pathways (out of 277) show higher AUC values with the JGGM approach for B-cells, fibroblasts and T-cells, respectively. We then identified the pathways that show the differences greater than 0.4 (marked in Figure 3 and listed in Table 2). The largest difference occurs in the fibroblasts, where the pathways 5, 6 and 7 show high AUC values only with the JGGM approach. Since these pathways have high AUC values in other cell types and are related to the immune responses, the result may be driven by the joint regularization. The pathways 8 and 9 show high

ACCEPTED MANUSCRIPT

AUC values only with the separate approach, but these pathways are metabolic pathways and not directly related to the fibroblasts. In addition, the pathways 11 and 12 show high AUC values in T-cells only from the separate analysis, but these pathways have remote relationships with the immune activity. Overall, the B-cells and the T-cells show higher AUC values with the JGGM approach than with the separate approach, while the fibroblasts do not.

Appendix

Proof of Proposition 1

Proof 1 *The proof for the* f_1 *penalty function is provided in Huang et al. (2009).*

We start to prove the proposition for the case of the f_2 *penalty function. When* $\sum_{t=1}^{T} |\omega_{j,j'}^t| > \epsilon$*, one can find that the solution of the derivative equation,* $\frac{\partial}{\partial \theta_{j,j'}} \tilde{PL}(\{\mathbf{\Omega}^t\}_{t=1}^T, \Theta) = 0$*, is* $\theta_{j,j'} = 1 - \log(\epsilon) + \log(\sum_{t=1}^{T} |\omega_{j,j'}^t|)$*. Hence,* $\sum_{t=1}^{T} |\omega_{j,j'}^t| > \epsilon$ *is equivalent to* $\theta_{j,j'} > 1$*. Pluggin this into* $\tilde{PL}(\{\mathbf{\Omega}^t\}_{t=1}^T, \Theta)$ *yields a profiled penalized likelihood of* $p\tilde{PL}(\{\mathbf{\Omega}^t\}_{t=1}^T) = L(\{\mathbf{\Omega}^t\}_{t=1}^T) + \tau \sum_{j \neq j'} (\log(\sum_{t=1}^{T} |\omega_{j,j'}^t|) - \log \epsilon + 1)$*. By taking* $\lambda = \tau$*, one can find that* $p\tilde{PL}(\{\mathbf{\Omega}^t\}_{t=1}^T) = PL(\{\mathbf{\Omega}^t\}_{t=1}^T)$*.*

When $\sum_{t=1}^{T} |\omega_{j,j'}^t| \leq \epsilon$*, the penalty form of* $PL(\{\mathbf{\Omega}^t\}_{t=1}^T)$ *becomes* $\frac{1}{\epsilon} \sum_{t=1}^{T} |\omega_{j,j'}^t|$*. This is equivalent to not assuming a common structure, which can be achieved by setting* $g(\theta_{j,j'})$ *to be a constant function* $\frac{1}{\epsilon}$ *when* $0 \leq \theta_{j,j'} \leq 1$*.*

We then prove the proposition for the case of the f_3 *penalty function by using the same principle. We find that the solution* $\theta_{j,j'} = \nu\epsilon^{1-\nu} - (\sum_{t=1}^{T} |\omega_{j,j'}|)^{1-\nu}$*. Hence,* $\sum_{t=1}^{T} |\omega_{j,j'}^t| > \epsilon$ *is equivalent to* $\theta_{j,j'} > (\nu - 1)\epsilon^{1-\nu}$*. This yields a profiled likelihood of* $p\tilde{PL}(\{\mathbf{\Omega}^t\}_{t=1}^T,) = L(\{\mathbf{\Omega}^t\}_{t=1}^T) + \frac{\tau}{\nu} \sum_{j \neq j'} (\nu\epsilon^{1-\nu} - (\sum_{t=1}^{T} |\omega_{j,j'}|)^{1-\nu}) = PL(\{\mathbf{\Omega}^t\}_{t=1}^T)$ *by taking* $\lambda = \frac{\tau}{\nu}$*.*

Lemma 1 *If either* x *or* y *is greater than* $\tau(> 0)$*, then* $|x^\alpha - y^\alpha|\tau^{1-\alpha} \leq |x - y|$*, for* $0 < \alpha < 1$*.*

Proof 2 *Without loss of generality, we can assume that* $x \geq y$*.*

When $x > \tau > y$,

$$
\begin{aligned}
(x^{\alpha} - y^{\alpha})\tau^{1-\alpha} &\leq x - y^{\alpha}\tau^{1-\alpha} \\
&\leq x - y.
\end{aligned}
$$

When $x > y > \tau$,

$$x^{\alpha} - y^{\alpha} \leq \frac{1}{y^{1-\alpha}}(x - y) \leq \frac{1}{\tau^{1-\alpha}}(x - y).$$

Proof of Theorem 1

Proof 3 *Theorem 1 can be proved with a slight extension to the proof of Guo et al. (2011), which is similar to the proof of Theorem 1 of Rothman et al. (2008).*

Denote the objective function 1 as $Q(\mathbf{\Omega})$, where $\mathbf{\Omega} = \{\mathbf{\Omega}^t\}_{t=1}^T$ and we write the true precision matrices as $\mathbf{\Omega}_0 = \{\mathbf{\Omega}_0^t\}$. We would like to show $Q(\Omega)$ has the local minimum near $\mathbf{\Omega}_0$.

Specifically, we would like to show that $P(\tilde{Q}(\Delta) = Q(\mathbf{\Omega}_0 + \Delta) - Q(\mathbf{\Omega}_0) > 0)$ converges to 1, when $\Delta \in \partial\mathcal{A}$, where $\partial\mathcal{A} = \{\Delta : \sum_{t=1}^T \|\Delta^t\|_F = Mr_n\}$, and $\Delta^t = \hat{\mathbf{\Omega}}^t - \mathbf{\Omega}_0^t$, and M is a positive constant and $r_n = \sqrt{\frac{(p+q)\log p}{n}}$.

We will use the following notation: for a matrix $\mathbf{M} = [m_{j,j'}]_{p\times p}$, $|\mathbf{M}|_1 = \sum_{j,j'} |m_{j,j'}|$, $\mathbf{M}^+$ is a diagonal matrix with the same diagonal as $\mathbf{M}$, $\mathbf{M}^- = \mathbf{M} - \mathbf{M}^+$, and M_S is M with all elements outside an index set $\mathbf{S}$ replaced by zeros. Also, vec$(\mathbf{M})$ for the vectorized form of $\mathbf{M}$, and $\otimes$ for the Kronecker product of two matrices.

As in Guo et al. (2011), $\tilde{Q}$ is the sum of the following components:

$$
\begin{aligned}
I_1 &= \sum_{t=1}^{T} trace((\mathbf{S}^t - \mathbf{\Sigma}_0^t)\Delta^k) \\
I_2 &= \sum_{t=1}^{T} \tilde{\Delta}^{t\prime} \int_0^1 (1 - v)(\mathbf{\Omega}_0^t + v\Delta^t)^{-1} \otimes (\mathbf{\Omega}_0^t + v\Delta^t)^{-1} dv \tilde{\Delta}^t \\
I_3 &= \lambda \sum_{(j,j')\in\mathbf{E}^c} f_i(\sum_{t=1}^{T}(|\delta_{j,j'}^t|))
\end{aligned}
$$

$$I_4 \quad = \quad \lambda \sum_{(j,j')\in\mathbf{E}} (f_i(\sum_{t=1}^{T} |\omega^t_{j,j'}|) - f_i(\sum_{t=1}^{T} |\omega_0{}^t_{j,j'}|))$$

The bound for the likelihood part can be found in Guo et al. (2011), where

$$|I_1| \leq C_1 \sqrt{\frac{\log p}{n}} \sum_{t=1}^{T} |\Delta^{t-}|_1 + C_2 \sqrt{\frac{p \log p}{n}} \sum_{t=1}^{R} \|\Delta^{t+}\|_F,$$

$$I_2 \geq \frac{1}{4\xi_2^2} \sum_{t=1}^{T} \|\Delta^t\|_F^2,$$

for some constants C_1 and C_2 with probability tending to 1.

When $(p+q)(\log p)/n$ is small,

$$I_3 \quad \geq \quad \lambda \sum_{t=1}^{T} |\Delta^{t-}_{\mathbf{E}^c}|_1,$$

due to the concavity of the penalty functions.

Also,

$$I_4 \quad \leq \quad \lambda \sum_{j\neq j':(j,j')\in\mathbf{E}} \left| f_i(\sum_{t=1}^{T} |\omega^t_{j,j'}|) - f_i(\sum_{t=1}^{T} |\omega_0{}^t_{j,j'}|) \right|.$$

For the f_1 function, by using Lemma 1,

$$\begin{aligned} I_4 \quad &\leq \quad \frac{\lambda}{\xi_3^\nu} \sum_{j\neq j':(j,j')\in\mathbf{E}} |\sum_{t=1}^{T} |\omega^t_{j,j'}| - \sum_{t=1}^{T} |\omega_0{}^t_{j,j'}|| \\ &\leq \quad \frac{\lambda}{\xi_3^\nu} \sum_{t=1}^{T} \sum_{j\neq j':(j,j')\in\mathbf{E}} |\omega^t_{j,j'} - \omega_0{}^t_{j,j'}| \\ &\leq \quad \frac{\lambda}{\xi_3^\nu} \sqrt{q} \sum_{t=1}^{T} \|\Delta^t\|_F \\ &\leq \quad \frac{\Lambda_2}{\xi_3^\nu} \sqrt{\frac{(p+q)\log p}{n}} \sum_{t=1}^{T} \|\Delta^t\|_F. \end{aligned}$$

For the functions f_2 *and* f_3*,*

$$\begin{aligned}
|I_4| &\leq \lambda \sum_{j\neq j':(j,j')\in\mathbf{E}} |f_i(\sum_{t=1}^{T} |\omega^t_{j,j'}|) - f_i(\sum_{t=1}^{T} |\omega_0{}^t_{j,j'}|)| \\
&\leq \lambda \sum_{j\neq j':(j,j')\in\mathbf{E}} f_i'(\xi_3 - \sum_{t=1}^{T} ||\Delta^t||_F) \sum_{t=1}^{T} |\omega^t_{j,j'} - \omega_0{}^t_{j,j'}| \\
&\leq \lambda \sum_{j\neq j':(j,j')\in\mathbf{E}} f'(\xi_3/2) \sum_{t=1}^{T} |\omega^t_{j,j'} - \omega_0{}^t_{j,j'}| \text{ for sufficiently small } r_n, \\
&\leq \lambda \sum_{j\neq j':(j,j')\in\mathbf{E}} \frac{\nu-1}{(\xi_3/2)^\nu} \sum_{t=1}^{T} |\omega^t_{j,j'} - \omega_0{}^t_{j,j'}| \\
&\leq \frac{(\nu-1)\lambda}{(\xi_3/2)^\nu} \sqrt{q} \sum_{t=1}^{T} ||\Delta^t||_F \\
&\leq \frac{(\nu-1)\Lambda_2}{(\xi_3/2)^\nu} \sqrt{\frac{(p+q)\log p}{n}} \sum_{t=1}^{T} ||\Delta^t||_F,
\end{aligned}$$

where $\nu \geq 1$ *and* $f_i'(a)$ *denotes* $\frac{\partial f_i(x)}{\partial x}|_{x=a}$

The second inequality comes from the application of the mean value theorem and the fact that f' *is decreasing function as well as* $|\omega_0{}^t_{j,j'}| > \xi_3$.

Combining all the results,

$$\begin{aligned}
\tilde{Q}(\Delta) &\geq -|I_1| + I_2 + I_3 - |I_4| \\
&\geq -C_1\sqrt{\frac{\log p}{n}} \sum_{t=1}^{T}(|\Delta^{t-}_{\mathbf{E}}|_1 + |\Delta^{t-}_{\mathbf{E}^c}|_1) - C_2\sqrt{\frac{p\log p}{n}}||\Delta^{t+}||_F + \frac{1}{4\xi_2^2}\sum_{t=1}^{T}||\Delta^t||_F^2 \\
&\quad +\Lambda_1\sqrt{\frac{\log p}{n}}\sum_{t=1}^{T}|\Delta^{t-}_{\mathbf{E}^c}|_1 - \frac{\Lambda_2}{g(\xi_3)}\sqrt{\frac{(p+q)\log p}{n}}\sum_{t=1}^{T}||\Delta^t||_F \\
&\geq (\Lambda_1 - C_1)\sqrt{\frac{\log p}{n}}\sum_{t=1}^{T}|\Delta^{t-}_{\mathbf{E}^c}|_1 - (C_1+C_2)\sqrt{\frac{(p+q)\log p}{n}}\sum_{t=1}^{T}||\Delta^t||_F \\
&\quad +\frac{1}{4\xi_2^2}\sum_{t=1}^{T}||\Delta^t||_F^2 - \frac{\Lambda_2}{g(\xi_3)}\sqrt{\frac{(p+q)\log p}{n}}\sum_{t=1}^{T}||\Delta^t||_F \\
&\geq \frac{1}{4\xi_2^2 T}(\sum_{t=1}^{T}||\Delta^t||_F)^2 - (C_1+C_2)\sqrt{\frac{(p+q)\log p}{n}}\sum_{t=1}^{T}||\Delta^t||_F - \frac{\Lambda_2}{g(\xi_3)}\sqrt{\frac{(p+q)(\log p)}{n}}\sum_{t=1}^{T}||\Delta^t||_F
\end{aligned}$$

ACCEPTED MANUSCRIPT

$$\text{for } \Lambda_1 > C_1$$
$$= \ (\sum_{t=1}^{T} \|\Delta^t\|_F)^2 \left(\frac{1}{4T\xi_2^2} - \frac{C_1 + C_2 + \Lambda/g(\xi_3)}{\sum_{t=1}^T \|\Delta^t\|_F / \sqrt{\frac{(p+q)(\log p)}{n}}} \right),$$

where $g(\xi) = \xi^\nu$ *for* f_1 *penalty function, and* $g(\xi) = (\nu-1)(\xi/2)^{-\nu}$ *for* f_2 *and* f_3 *penalty functions. Thus, for sufficiently large* M*, we have* $\tilde{Q}(\Delta) > 0$ *for any* $\Delta \in \partial\mathcal{A}$*.*

Proof of Theorem 2

Proof 4 *Define* $\mathbf{E}_n = \mathbf{E}_{n,1} \cup \ldots \cup \mathbf{E}_{n,T}$*, where* $\mathbf{E}_{n,t} = \{(j, j') : j \neq j', \hat{\omega}^t_{j,j'} \neq 0\}$*.*

We first show that $P(\mathbf{E} \subseteq \mathbf{E}_n)$ *converges to 1.*

$P(\mathbf{E} \subseteq \mathbf{E}_n) = P(|\hat{\omega}^t_{j,j'}| > 0$ *for some* $t \in 1, \ldots, T$ *for all* $(j, j') \in \mathbf{E})$*. Since* $\sum_{t=1}^T \|\hat{\mathbf{\Omega}}^t_{j,j'} - \mathbf{\Omega}^t_{0,j,j'}\|_F = O_p(\sqrt{\frac{(p+q)\log p}{n}})$ *by Theorem 1, one can see that* $P(|\hat{\omega}^t_{j,j'}| > 0$ *for some* $t \in 1, \ldots, T$ *for all* $(j, j') \in \mathbf{E})$ $\rightarrow P(|\omega^t_{0,j,j'}| > 0$ *for some* $t \in 1, \ldots, T$ *for all* $(j, j') \in \mathbf{E})$ *which should be 1 due to the fact that* $|\omega^t_{0,j,j'}| > \xi_3 > 0$ *for some* t *for all* $(j, j') \in \mathbf{E}$*.*

In order to show that $P(\mathbf{E}_n \subseteq \mathbf{E})$ *converges to 1, we will show* $P(\mathbf{E}^c \subseteq \mathbf{E}^c_n)$ *converges to 1. For this, we need to show that for any* $(j, j') \in \mathbf{E}^c$*, the derivative* $\frac{\partial Q}{\partial \omega^t_{j,j'}}$ *has the same sign as* $\hat{\omega}^t_{j,j'}$ *for all* $1 \leq t \leq T$ *with probability tending to 1.*

We first discuss the f_1 *penalty function. The derivative of the objective function can be written as*

$$\frac{\partial Q}{\partial \omega^t_{j,j'}} = W_1(t, j, j') + W_2 sign(\omega^t_{j,j'}),$$

where $W_1(t, j, j') = \mathbf{S}^t_{j,j'} - \mathbf{\Sigma}^t_{j,j'}$ *and* $W_2 = \lambda(1-\nu)(\sum_{t=1}^T |\omega^t_{j,j'}|)^{-\nu}$*, where* $0 < \nu < 1$*.*

Arguing as in Theorem 2 of Lam and Fan (2009), one can show that $\max_{t,j,j'} W_1(t, j, j') = O_p((\frac{\log p}{n})^{1/2} + \eta_n^{1/2})$*.*

For $(j, j') \in \mathbf{E}^c$*,* $\sum_{t=1}^T |\hat{\omega}^t_{j,j'}| = O_p(\eta_n)$ *and* $\eta_n^{-\nu}$ *goes to* ∞*, and* $(\frac{\log p}{n})^{1/2} + \eta_n^{1/2} = O(\lambda)$*,* W_2 *dominates* $\max_{(t,j,j')} W_1(t, j, j')$*.*

For the functions f_2 *and* f_3, $W_2 = \frac{\lambda}{\max((\sum_{t=1}^{T}|\omega_{j,j'}^{t}|),\epsilon)}$ *and* $W_2 = \frac{\lambda(\nu-1)}{\max((\sum_{t=1}^{T}|\omega_{j,j'}^{t}|)^{\nu},\epsilon^{\nu})}$*, respectively, and* $\nu > 1$.

For $(j, j') \in \mathbf{E}^c$, $\sum_{t=1}^{T}|\hat{\omega}_{j,j'}^{t}| = O_p(\eta_n)$*. Then,* $W_2 = O_p(\lambda \min(\eta_n^{-\nu}, \epsilon^{-\nu}))$, $\nu \geq 1$ *and* W_2 *dominates* $\max_{t,j,j'} W_1(t, j, j')$*, by taking sufficiently small* ϵ*.*

ACCEPTED MANUSCRIPT