Supplementary Information

A regularized Hotelling's T^2 test for pathway analysis in proteomic studies

Lin S. Chen[1], Debashis Paul[2], Ross L. Prentice[3], Pei Wang[3,*]

1. Department of Health Studies, The University of Chicago, IL.

2. Department of Statistics, University of California, Davis, CA.

3. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, WA.

* To whom correspondence should be addressed. Email:pwang@fhcrc.org

1 The distribution of F^+, T_D and T_{BS} statistics

When $p \geq n$, the sample covariance matrix, S, is singular and not invertible, and Hotelling's T^2 statistic not well-defined. No test is invariant under nonsingular linear transformation, the so-called MANOVA problem (Lehmann and Romano 2008).

Srivastava (2007) proposed a generalized Hotelling's T^2 statistics, which is defined as $T^{+2} = n\left(\bar{\boldsymbol{x}} - \boldsymbol{\mu_0}\right)^T S^+ \left(\bar{\boldsymbol{x}} - \boldsymbol{\mu_0}\right)$, where S^+ is the Moore-Penrose inverse of S. When using SVD to decompose S as $S = UDU^T$, S^+ is equivalent to $S^+ = HL^{-1}H^T$, where H is a $p \times r$ matrix consisting of the first r columns of U, and r is the rank of S. U is column orthogonal $UU^T = I$, so that $HH^T = I_p$. $L = \text{diag}(d_1, \ldots, d_r)$

is an $r \times r$ diagonal matrix, the diagonal elements of which are the first r non-zero eigenvalues of S. Note that T^{+2} is not invariant. When $p \geq n$ and $\Sigma = \phi I_p$ ($\phi > 0$ is a constant), $F^+ = (p - m + 1)/m^2 T^{+2}$, where $m = n - 1$, follows an F-distribution with m and $p - m + 1$ degree of freedom. When $\Sigma \neq \phi I_p$, $F^+ > F_{m,p-m+1}$ with probability one (Srivastava 2007). A suitably normalized F^+ statistic is asymptotically normally distributed under the null. The asymptotic distribution of F^+ is given by $\lim_{m,p\to\infty} \Pr\left[c_{m,p}\left(\frac{m}{2}\right)^{\frac{1}{2}}\left(\hat{b}F^+ - 1\right) \leq z_{1-\alpha}\right] = \Phi(z_{1-\alpha})$, where

$$\hat{b} = \frac{(m+2)(m-1)}{m^2} \frac{(\mathrm{tr}S/p)^2}{p^{-1}\left[\mathrm{tr}S^2 - \frac{1}{m}(\mathrm{tr}S)^2\right]}, \tag{1}$$

and $c_{m,p} = \left(\frac{p-m+1}{p+1}\right)^{1/2}$. The asymptotic power of the F^+ test is given by $\beta(F^+) \simeq \Phi\left(-z_{1-\alpha} + \left(\frac{m}{p}\right)\left(\frac{m}{2}\right)^{\frac{1}{2}} \frac{\boldsymbol{\mu}^T\Lambda\boldsymbol{\mu}}{\mathrm{tr}\Sigma^2/p}\right)$, where $\Lambda = \mathrm{diag}(\eta_1, \ldots, \eta_p)$ and η_i are the eigenvalues of the covariance matrix Σ.

Dempster (1958) and Dempster (1960) proposed a non-exact test that can be applied to the cases when $p \geq n$. Let H be an $n \times n$ orthogonal matrix with the first row being $\frac{1}{\sqrt{n}}(1, \ldots, 1)'$, and the rest m rows being orthogonal. Let $Y = HX = (\boldsymbol{y_1}, \ldots, \boldsymbol{y_n})'$, with each $\boldsymbol{y_i}$ being independent. Dempster proposed his test statistic as $T_D = Q_1/(\sum_{i=2}^n Q_i/m)$, where $Q_i = \boldsymbol{y_i}\boldsymbol{y_i}'$. When $p < n$ and $\Sigma = I_p$, the statistic reduces to $T_D = \frac{n\bar{\boldsymbol{x}}^T\bar{\boldsymbol{x}}}{\mathrm{tr}S/p}$, and the test is a uniformly most powerful invariant test. Under the null, it approximately follows an F-distribution, $T_D \sim F_{[p\hat{b}],[mp\hat{b}]}$ where $\hat{b}$ can be obtained by Equation (1), and $[\cdot]$ denotes the largest integer that is less than the variable. However, when $p \geq n$ or $\Sigma \neq I_p$, the test statistic is not invariant. The degree of freedoms of the asymptotic distribution of T_D depends on $\hat{b}$, and $\hat{b}$ depends on the covariance matrix.

Bai and Saranadasa (1996) proposed a standardized version of the T_D test, $T_{BS} = \frac{n\bar{\boldsymbol{x}}^T\bar{\boldsymbol{x}} - \mathrm{tr}S}{\sigma_{BS}}$ where $\sigma_{BS}^2 = \frac{2m(m+1)}{(m+2)(m-1)}\left[\mathrm{tr}S^2 - \frac{1}{m}(\mathrm{tr}S)^2\right]$. Note that T_{BS} is also not in-

variant. Asymptotically, $\lim_{m,p\to\infty} \Pr(T_{BS} \leq z_{1-\alpha}) = \Phi(z_{1-\alpha})$. Furthermore, T_D and T_{BS} have the same asymptotic power (Srivastava 2007). $\beta(T_D), \beta(T_{BS}) \simeq \Phi\left(-z_{1-\alpha} + \frac{m\boldsymbol{\mu}^T\boldsymbol{\mu}}{\sqrt{2\mathrm{tr}\Sigma^2}}\right)$. However, with severe missingness, these tests can be unreliable (see **Simulations** section in the main text).

Following *Example A* in the main text, Figure S1 listed the distribution of Dempster's and Bai and Saranadasa's test statistics when $p < n$.

Example B. We simulated a case where $n = 10$ and $p = 20$ and so $p \geq n$. We then sample X from a multivariate normal distribution $N(\mu, \Sigma_{p\times p})$, and consider three different settings:

(1) *Independent Null*: $\mu_1 = \cdots = \mu_{20} = 0$; $\Sigma = I$;

(2) *Dependent Null*: $\mu_1 = \cdots = \mu_{20} = 0$; $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.5$ for $i \neq j$;

(3) *Dependent Alternative*: $\mu_1 = \cdots = \mu_{10} = 1$, $\mu_{11} = \cdots = \mu_{20} = 0$; $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.5$ for $i \neq j$.

Figure S2 shows the distribution of $RHT_{Bootstrap}$, RHT without bootstrap, F^+, Dempster's and Bai and Saranadasa's test statistics when $p \geq n$.

2 Simulation: Comparing `RHT` test with and without bootstrap as $\gamma = p/n$ increases

In this simulation, we compare the performance of `RHT` test using the proposed bootstrap-like procedure and without bootstrap when $p \geq n$. Specifically, we fix $n = 20$ and for each $p \in \{5, 10, 15, \ldots, 100\}$, we simulate $X \sim N(\mu, \Sigma)$, where $\mu = \mathbf{0}$ under the null and $\mu = (0.5, \ldots, 0.5)^T$ under the alternative, $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.2$ $(i \neq j)$. We also simulate 20% of the data missing completely at random. The type I error rate and power of the `RHT` test with and without bootstrap are calculated based on 1000 simulations. The results are illustrated in Figure S3. We can see that as

$\gamma = p/n$ increases, both can control type I error rate, while the power of RHT test with bootstrap exceeds that of RHT without bootstrap as γ increases. Under some other missing/correlation settings, RHT test without bootstrap may occasionally give anti-conservative type-I error rate, but RHT test with bootstrap always conservatively controls type-I error rate (data not shown).

3 Two-sample regularized Hotelling's T^2 statistic

Suppose $\boldsymbol{x_1}, \ldots, \boldsymbol{x_{n_x}} \overset{i.i.d}{\sim} N_p(\boldsymbol{\mu_X}, \Sigma)$ and $\boldsymbol{y_1}, \ldots, \boldsymbol{y_{n_y}} \overset{i.i.d}{\sim} N_p(\boldsymbol{\mu_Y}, \Sigma)$ with a common covariance matrix Σ, and that one wishes to test whether the mean vectors of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are equal, i.e., $H_0 : \boldsymbol{\mu_X} = \boldsymbol{\mu_Y}$.

The two-sample Hotelling's T^2 statistic is defined as

$$T^2 \;=\; \frac{n_x n_y}{n_x + n_y} \left(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}\right)^T S^{-1} \left(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}\right),$$

where $\bar{\boldsymbol{x}} = \frac{1}{n_x} \sum_{i=1}^{n_x} \boldsymbol{x_i}$, $\bar{\boldsymbol{y}} = \frac{1}{n_y} \sum_{i=1}^{n_y} \boldsymbol{y_i}$, and
$S = \frac{1}{n_x + n_y - 2} \left[\sum_{i=1}^{n_x} (\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^T + \sum_{i=1}^{n_y} (\boldsymbol{y_i} - \bar{\boldsymbol{y}})(\boldsymbol{y_i} - \bar{\boldsymbol{y}})^T\right].$

We decompose S as $S = UDU^T$, and $S^{-1} = UD^{-1}U^T$. The regularized Hotelling's T^2 statistics is defined correspondingly as

$$\texttt{RHT} \;=\; \frac{n_x n_y}{n_x + n_y} \left(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}\right)^T U(D + \lambda I)^{-1} U^T \left(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}\right)$$

When $p < n$ and Σ is nonsingular, $\texttt{RHT} = \frac{n_x n_y}{n_x + n_y} \left(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}\right)^T (S + \lambda I)^{-1} \left(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}\right)$.

Note that our proposed procedure to choosing tuning parameter only uses the eigenvalues from the sample covariance matrix S. Hence, it can also be applied to the common sample covariance estimator for two-sample RHT test to obtain a tuning parameter value for a conservative two-sample test.

4 Simulations for two-sample testing

Figure S4 compared the type I error rates and power of `RHT` test with the other competing tests in two-sample testing, for the cases in which $m = n_x + n_y - 2 = 38 > p = 20$ and the cases in which $m = 8 < p = 20$, with and without missing data. When there is missingness, `HT`/F^+ and T_{BS} tests fail to control type I error rate. `SAT` test does not control type I error rate under correlation. T_D test controls type I error rate but is not powerful when sample size is small and/or there are missingness in the data. Our `RHT` test conservatively controls type I error rate under correlation and/or missingness. When there is no missingness, in both $p < n$ and $p \geq n$, the `RHT` test is the most powerful test among those controlling type I error rates.

5 Proteins lists for the five significant pathways identified from HT proteomics study

1. HSA04610 Complement and coagulation cascades

A2M MASP1 MASP2 C1QC CFH CFI CFD F11 KNG1 F12 F10 F9 SERPING1 C8G PROC C1QA C1QB F5 SERPINF2 F2 TFPI PROS1 C7 MBL2 C9 C3 C6 C5 C1S F13B FGG FGA FGB KLKB1 SERPINC1 C2 CR2 CFB C4BPB C4BPA PLG VWF CD59 SERPIND1 CPB2

2. HSA04512 ECM receptor interaction

VTN GP5 CD44 COL1A1 TNC DAG1 COL6A3 THBS1 THBS4 FN1 HSPG2 VWF

3. HSA01430 Cell communication

VTN COL1A1 FN1 TNC COMP COL6A3 KRT1 THBS1 THBS4 VWF DSC3 DSC1

4. HSA04510 Focal adhesion

VEGFC FLNA COL6A3 THBS1 THBS4 IGF1 PDGFB PDGFA VTN COL1A1 FN1 TLN1 TNC COMP VWF

5.HSA04350 TGF-beta signaling

LTBP1 TGFB1 INHBE INHBC COMP THBS1 THBS4

6 Proofs

Comment on the decay rate of the signal size in Theorem 2. In order for the RHT test to be asymptotically consistent, by Theorem 2, we require the "effect size" $\| \delta_n \|$ (where $\delta_n = \Sigma^{-1/2}\mu$) to go to zero at rate $n^{-1/4}$ or slower. It is in contrast with a classical multivariate testing situation (i.e., p is fixed), where asymptotic consistency of the Hotelling's T^2 test is achieved as long as the effect size goes to zero at a rate $n^{-1/2}$ or slower. This phenomenon is an illustration of the differences between the fixed p and $p \sim n$ scenarios. This effect of dimensionality can be seen clearly in the expansion of the normalized RHT statistic in equation (26) in the main manuscript. Notice that, the second and third terms on the right hand side of (26) behave very differently when p is fixed versus when p is comparable to n. Specifically, when p is fixed and $n \to \infty$, the second term is $O_P(\sqrt{n} \| \delta_n \|)$ and the third term is bounded below by a term of the order $n \| \delta_n \|^2$. Therefore, $n \| \delta_n \|^2 \to \infty$ would be a sufficient condition for the asymptotic consistency of the test. However, as the arguments immediately after (26) show, when $p \sim n$, it is necessary to have $\sqrt{n} \| \delta_n \|^2 \to \infty$. Note also that under H_0, the RHT statistic is stochastically bounded for the fixed p setting, while it varies at a rate $\sqrt{p}$ around $p\Theta_1(\lambda) \sim p$ when $p \sim n$.

To illustrate this difference further, we consider an expository example. Let $\Sigma = I$

and $\mu = (c, \ldots, c)^T$ for some $c \neq 0$. When p is fixed and $n \to \infty$, we need c to be at least as big as $n^{-1/2}$, in order that $n \parallel \delta_n \parallel^2 \to \infty$, which ensures consistency. When $p \sim n$, for $\delta_n = \mu$ having the same form, it is enough to have $c \gg n^{-3/4}$, in order that $\sqrt{n} \parallel \delta_n \parallel^2 \to \infty$. Thus, when p is comparable to n, even if each of the coordinates of μ are smaller than $1/\sqrt{n}$, still the power can converge to 1, which is not possible in the fixed p setting. However, in the first setting $\parallel \delta_n \parallel \sim c$ and in the second $\parallel \delta_n \parallel \sim \sqrt{n}c$ showing that the overall signal strength is larger in the latter case.

Proof of Lemma 4. If $g_n|y_n \to_D Z$, it follows, for $\forall c, \Pr(g_n \leq c|y_n) \to_{a.s.} \Pr(Z \leq c)$. Then, $\Pr(g_n \leq c) = E_{y_n}(\Pr(g_n \leq c|y_n)) \to E(\Pr(Z \leq c)) = \Pr(Z \leq c)$. □

Proof of (31). We give the proof only for $\lambda > 0$, although a slight modification of the proof also works for $\lambda = 0$ when $\gamma < 1$. Then,

$$\parallel R_n(-\lambda) \parallel \leq 1/\lambda \quad \text{and} \quad \max_{1 \leq j \leq N} \parallel R_n^{(j)}(-\lambda) \parallel \leq 1/\lambda. \tag{2}$$

We make use of the following lemma (see for example, Paul (2007)).

Lemma S1 Suppose that $X \sim N(0, I_n)$, and let C be a symmetric $n \times n$ matrix with $\parallel C \parallel \leq L$. Then, for all $0 < t < L$,

$$\mathbb{P}(\frac{1}{n}|X^T C X - \mathtt{tr}(C)| > t) \leq 2\exp\left(-\frac{nt^2}{4L^2}\right). \tag{3}$$

Using Lemma S1 and (2) we have, for any $\varepsilon > 0$, for n sufficiently large so that $\sqrt{(1+\varepsilon)\log n/n} < 1/2$,

$$\mathbb{P}\left(\max_{1 \leq j \leq n} |Y_j^T \Sigma_p^{1/2} R_n^{(j)}(-\lambda) \Sigma_p^{1/2} Y_j - \frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p)| > \frac{2\tau_{1,p}}{\lambda}\sqrt{\frac{(1+\varepsilon)\log n}{n}}\right) \leq 2n^{-\varepsilon} \tag{4}$$

and

$$\mathbb{P}\left(\max_{1\le j\le n}|Y_j^T\Sigma_p^{1/2}R_n^{(j)}(-\lambda)\Sigma_pR_n^{(j)}(-\lambda)\Sigma_p^{1/2}Y_j-\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_pR_n^{(j)}(-\lambda)\Sigma_p)|\right.$$
$$\left.>\frac{2\tau_{1,p}^2}{\lambda^2}\sqrt{\frac{(1+\varepsilon)\log n}{n}}\right)\ \le\ 2n^{-\varepsilon}. \tag{5}$$

Next, from **(27)** of the paper, we have, for $j=1,\ldots,n$,

$$\begin{aligned}\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p)-\frac{1}{n}\mathrm{tr}(R_n(-\lambda)\Sigma_p) &= \frac{1}{n}\frac{Y_j^T\Sigma_p^{1/2}R_n^{(j)}(-\lambda)\Sigma_pR_n^{(j)}(-\lambda)\Sigma_p^{1/2}Y_j}{1+Y_j^T\Sigma_p^{1/2}R_n^{(j)}(-\lambda)\Sigma_p^{1/2}Y_j}\\ &\le \frac{\|\ \Sigma_p^{1/2}R_n^{(j)}(-\lambda)\Sigma_p^{1/2}\ \|}{n}\ \le\ \frac{\tau_{1,p}}{n\lambda}\ ,\end{aligned} \tag{6}$$

where the last inequality is due to (2). Now, recall that,

$$\begin{aligned}\delta_n^{(1)} &= \frac{1}{p}\sum_{j=1}^n\left[\frac{1}{1+\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p)}-\frac{1}{1+Y_j^T\Sigma_p^{1/2}R_n^{(j)}(-\lambda)\Sigma_p^{1/2}Y_j}\right]\\
&= \frac{1}{p}\sum_{j=1}^n\frac{Y_j^T\Sigma_p^{1/2}R_n^{(j)}(-\lambda)\Sigma_p^{1/2}Y_j-\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p)}{(1+\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p))^2}\\
&\quad-\frac{1}{p}\sum_{j=1}^n\frac{(Y_j^T\Sigma_p^{1/2}R_n^{(j)}(-\lambda)\Sigma_p^{1/2}Y_j-\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p))^2}{(1+\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p))^2(1+Y_j^T\Sigma_p^{1/2}R_n^{(j)}(-\lambda)\Sigma_p^{1/2}Y_j)}\\
&= \frac{1}{(1+\frac{1}{n}\mathrm{tr}(R_n(-\lambda)\Sigma_p))^2}\frac{1}{p}\sum_{j=1}^n(Y_j^T\Sigma_p^{1/2}R_n^{(j)}(-\lambda)\Sigma_p^{1/2}Y_j-\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p))\\
&\quad+\frac{1}{p}\sum_{j=1}^n(Y_j^T\Sigma_p^{1/2}R_n^{(j)}(-\lambda)\Sigma_p^{1/2}Y_j-\frac{1}{n}\mathrm{tr}(R_n(-\lambda)\Sigma_p))\\
&\quad\cdot\left(\frac{1}{(1+\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p))^2}-\frac{1}{(1+\frac{1}{n}\mathrm{tr}(R_n(-\lambda)\Sigma_p))^2}\right)\\
&\quad-\frac{1}{p}\sum_{j=1}^n\frac{(Y_j^T\Sigma_p^{1/2}R_n^{(j)}(-\lambda)\Sigma_p^{1/2}Y_j-\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p))^2}{(1+\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p))^2(1+Y_j^T\Sigma_p^{1/2}R_n^{(j)}(-\lambda)\Sigma_p^{1/2}Y_j)}\\
&= \frac{1}{(1+\frac{1}{n}\mathrm{tr}(R_n(-\lambda)\Sigma_p))^2}\frac{1}{p}\sum_{j=1}^n(Y_j^T\Sigma_p^{1/2}R_n^{(j)}(-\lambda)\Sigma_p^{1/2}Y_j-\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p))\\
&\quad+o_P(n^{-1/2}),\end{aligned} \tag{7}$$

where the last equality is due to (2), (4), (6) and assumptions A2 and A4.

Define,

$$V_j = Y_j^T \Sigma_p^{1/2} R_n^{(j)}(-\lambda) \Sigma_p^{1/2} Y_j - \frac{1}{n} \mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p), \quad j = 1, \ldots, n, \tag{8}$$

and

$$V_{jj'} = Y_j^T \Sigma_p^{1/2} R_n^{(jj')}(-\lambda) \Sigma_p^{1/2} Y_j - \frac{1}{n} \mathrm{tr}(R_n^{(jj')}(-\lambda)\Sigma_p), \quad 1 \leq j \neq j' \leq n, \tag{9}$$

where

$$R_n^{(jj')}(z) = (S_n - \Sigma_p^{1/2}(Y_j Y_j^T + Y_{j'} Y_{j'}^T)\Sigma_p^{1/2} - z I_p)^{-1} = R_n^{(j'j)}(z), \quad z \in \mathbb{C}. \tag{10}$$

Then $V_j = V_{jj'} + \frac{1}{n} D_{jj'}$, for $1 \leq j \neq j' \leq n$, where

$$D_{jj'} = -\frac{n(Y_j^T \Sigma_p^{1/2} R_n^{(jj')}(-\lambda) \Sigma_p^{1/2} Y_{j'})^2 - Y_{j'}^T \Sigma_p^{1/2} R_n^{(jj')}(-\lambda) \Sigma_p R_n^{(jj')}(-\lambda) \Sigma_p^{1/2} Y_{j'}}{1 + Y_{j'}^T \Sigma_p^{1/2} R_n^{(jj')}(-\lambda) \Sigma_p^{1/2} Y_{j'}}. \tag{11}$$

Since Y_j and $Y_{j'}$ are independent, and both are independent of $R_n^{(jj')}(-\lambda)$, the following holds.

$$\mathbb{E}(V_{12}) = \mathbb{E}(V_{21}) = \mathbb{E}(V_{12}V_{21}) = \mathbb{E}(D_{12}) = \mathbb{E}(D_{21}) = 0. \tag{12}$$

Moreover,

$$\mathrm{Var}(V_{12}) = \mathbb{E}(V_{12}^2) = \mathbb{E}(\mathbb{E}(V_{12}^2 | R_n^{(12)}(-\lambda))) = \frac{2}{n^2} \mathbb{E}(\mathrm{tr}(R_n^{(12)}(-\lambda)\Sigma_p R_n^{(12)}(-\lambda)\Sigma_p)) = O(n^{-1}), \tag{13}$$

where the last equality is due to (2) and assumption A4. Similarly,

$$\begin{aligned}
\mathrm{Var}(D_{12}) &= \mathbb{E}(D_{12}^2) = \mathbb{E}(\mathbb{E}(D_{12}^2|Y_2, R_n^{(12)}(-\lambda))) \\
&= 2\mathbb{E}\left[\mathbb{E}\left(\frac{(Y_2^T\Sigma_p^{1/2}R_n^{(12)}(-\lambda)\Sigma_p R_n^{(12)}(-\lambda)\Sigma_p^{1/2}Y_2)^2}{(1+Y_2^T\Sigma_p^{1/2}R_n^{(12)}(-\lambda)\Sigma_p^{1/2}Y_2)^2} \mid R_n^{(12)}(-\lambda)\right)\right] \\
&\leq 2\mathbb{E}\left[\mathbb{E}\left((Y_2^T\Sigma_p^{1/2}R_n^{(12)}(-\lambda)\Sigma_p R_n^{(12)}(-\lambda)\Sigma_p^{1/2}Y_2)^2|R_n^{(12)}(-\lambda)\right)\right] \\
&= \frac{4}{n}\mathbb{E}(\mathrm{tr}((R_n^{(12)}(-\lambda)\Sigma_p)^4)) + \frac{2}{n^2}(\mathbb{E}(\mathrm{tr}((R_n^{(12)}(-\lambda)\Sigma_p)^2)))^2 \;=\; O(1). \qquad (14)
\end{aligned}$$

From (14) and (14), we have $\mathrm{Var}(V_1) = O(n^{-1})$. Since the random variables $\{V_j\}$ are exchangeable, we have

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{n}\sum_{j=1}^{n} V_j\right) \\
=\;& \frac{1}{n}\mathrm{Var}(V_1) + \left(\frac{n-1}{n}\right)\mathrm{Cov}(V_1, V_2) \\
=\;& \frac{1}{n}\mathrm{Var}(V_1) + \left(\frac{n-1}{n}\right)\mathrm{Cov}\left[(V_{12} + \frac{1}{n}D_{12})(V_{21} + \frac{1}{n}D_{21})\right] \\
=\;& \frac{1}{n}\mathrm{Var}(V_1) + \frac{1}{n}(\mathrm{Cov}(V_{12}, D_{21}) + \mathrm{Cov}(V_{21}, D_{12})) + \frac{1}{n^2}\mathrm{Cov}(D_{12}, D_{21}) \quad \text{(by (14))} \\
\leq\;& \frac{1}{n}\mathrm{Var}(V_1) + \frac{2}{n}\sqrt{\mathrm{Var}(V_{21})}\sqrt{\mathrm{Var}(D_{12})} + \frac{1}{n^2}\mathrm{Var}(D_{12}) \\
=\;& O(n^{-2}) + O(n^{-3/2}) + O(n^{-2}) \;=\; O(n^{-3/2}), \qquad (15)
\end{aligned}$$

where the fourth step is due to (12) and Cauchy-Schwarz inequality, and the last step is due to (13) and (14). Now, combining this with (7) we conclude that

$$\sqrt{n}\delta_n^{(1)} = o_P(1). \qquad (16)$$

Next,

$$\begin{aligned}\delta_n^{(2)} &= \frac{n}{p}\left[\frac{1}{1+\frac{1}{n}\mathrm{tr}(R_n(-\lambda)\Sigma_p)} - \frac{1}{n}\sum_{j=1}^{n}\frac{1}{1+\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p)}\right] \\ &= \frac{1}{p}\sum_{j=1}^{n}\frac{(1/n)(\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p) - \mathrm{tr}(R_n(-\lambda)\Sigma_p))}{(1+\frac{1}{n}\mathrm{tr}(R_n^{(j)}(-\lambda)\Sigma_p))(1+\frac{1}{n}\mathrm{tr}(R_n(-\lambda)\Sigma_p))}\ .\end{aligned}$$

Thus, using (6), and assumptions A2 and A4 we have

$$\sqrt{n}\delta_n^{(2)} = o_P(1). \tag{17}$$

Next, we write

$$\delta_n^{(3)} = (\frac{1}{p}\mathrm{tr}(R_n(-\lambda)) - m_{\widetilde{F}_{n,p}}(-\lambda)) + (m_{\widetilde{F}_{n,p}}(-\lambda) - m_F(-\lambda)), \tag{18}$$

where, for $z \in \mathbb{C}$, $m_{\widetilde{F}_{n,p}}(z)$ is the unique solution of the equation

$$m = \int \frac{dH_p(\tau)}{\tau(1-(p/n)-(p/n)zm) - z}\ .$$

Then, by assumption A2, the second term on the RHS of (18) is $o(n^{-1/2})$, and using Theorem 1.1 of Bai and Silverstein (2004), the first term on the RHS is $O_P(n^{-1})$, so that we have

$$\sqrt{n}\delta_n^{(3)} = o_P(1). \tag{19}$$

Combining (16), (17) and (19) we get **(31)** of the paper.

Proof of (34)

In order to prove **(34)** , first by using argument similar to (but simpler than) those used in proving (16) and (17), we conclude that

$$\max\{|\delta_n^{(4)}|, |\delta_n^{(5)}|\} = o_P(1). \tag{20}$$

Next, define $\widetilde{R}_n(\lambda) = R_n(-\lambda)$ and notice that the functions $\mathrm{tr}(\widetilde{R}_n(\lambda)\Sigma_p)$ and $\Theta_1(\lambda,\gamma)$ are differentiable for $\lambda > 0$ and their derivatives are uniformly bounded over $\lambda \in (0,\infty)$ (by assumptions (ii), (iii) and (iv)). Thus, given $\varepsilon > 0$, we can find $\eta > 0$ such that

$$\mathbb{P}\left(|\frac{1}{\eta}(\mathrm{tr}(\widetilde{R}_n(\lambda+\eta)\Sigma_p) - \mathrm{tr}(\widetilde{R}_n(\lambda)\Sigma_p)) - \frac{d}{d\lambda}\mathrm{tr}(\widetilde{R}_n(\lambda)\Sigma_p)| > \varepsilon/3\right) \to 0 \tag{21}$$

and

$$|\frac{1}{\eta}(\Theta_1(\lambda+\eta,\gamma) - \Theta_1(\lambda,\gamma)) - \frac{\partial}{\partial\lambda}\Theta_1(\lambda,\gamma)| < \varepsilon/3. \tag{22}$$

Now, using the first part of Lemma 2 (equation **(19)** in the paper) separately for λ and $\lambda+\eta$, we have

$$\mathbb{P}\left(\frac{1}{\eta}|(\mathrm{tr}(\widetilde{R}_n(\lambda+\eta)\Sigma_p) - \Theta_1(\lambda+\eta,\gamma)) + (\mathrm{tr}(\widetilde{R}_n(\lambda)\Sigma_p) - \Theta_1(\lambda,\gamma))| > \varepsilon/3\right) \to 0 \tag{23}$$

Combining (21), (22) and (23) we have that

$$\mathbb{P}(|\delta_n^{(6)}| > \varepsilon) \to 0 \ \text{ as } n \to \infty.$$

Since $\varepsilon > 0$ is arbitrary, we conclude **(34)** by combining with (20).

References

Bai, Z. and H. Saranadasa (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica 6*, 311–329.

Bai, Z. D. and J. W. Silverstein (2004). Clt for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab. 32(1A)*, 553–605.

Dempster, A. P. (1958). A high dimensional two sample significance test. *Ann.*

Math. Statist. 29, 995–1010.

Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics 16*, 41–50.

Lehmann, E. L. and J. P. Romano (2008). *Testing Statistical Hypotheses* (3rd ed.). Springer.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica 17*, 1617–1642.

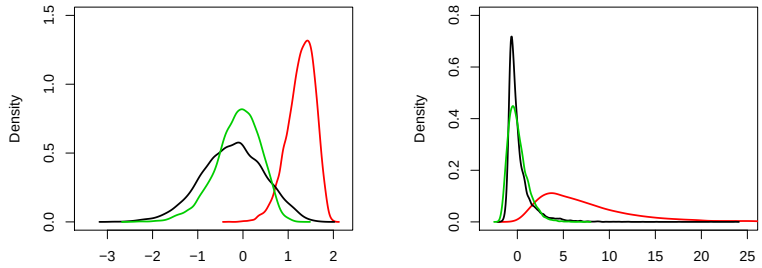
Srivastava, M. S. (2007). Multivariate theory for analyzing high dimensional data. *Journal of the Japan Statistical Society*, 53–86.

TABLES

Table S1: The number of significant pathways at FWER of 0.05 and 0.10 identified by different tests. The numbers in the parenthesis are the percentages of the significant ones out of all pathways (18 in total).

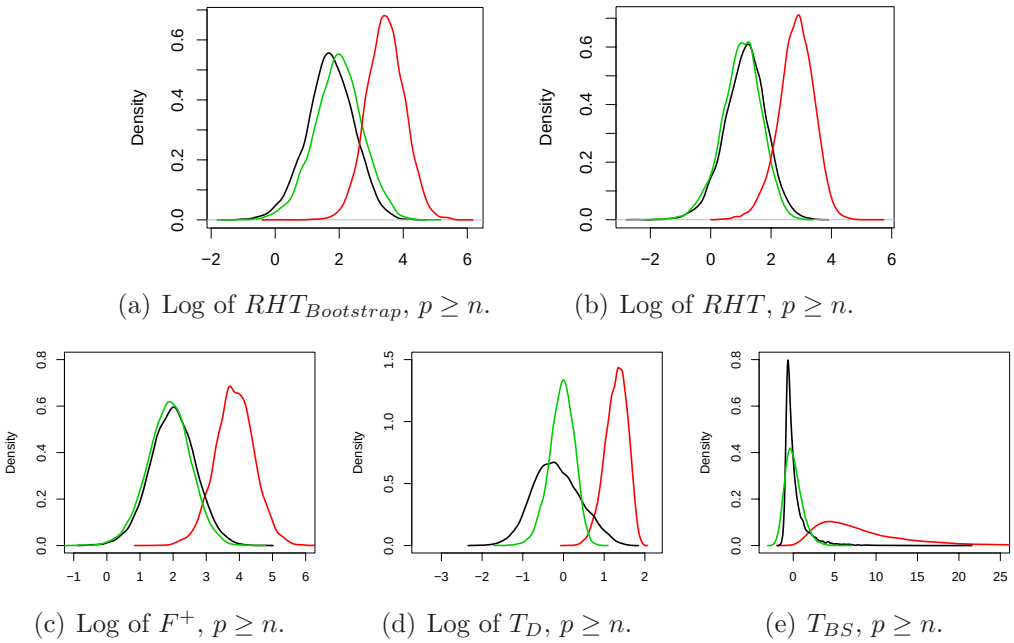
FWER cutoff	RHT	T_D	SAT	HT/F^+	T_{BS}
0.05	4 (22.2%)	5 (27.8%)	7 (38.9%)	8 (44.4%)	16 (88.9%)
0.10	5 (27.8%)	6 (33.3%)	8 (44.4%)	9 (50.0%)	17 (94.4%)

FIGURES



(a) Log of T_D, $p < n$. (b) T_{BS}, $p < n$.

Figure S1: Distribution of T_D and T_{BS} statistics on a simulated example with $p = 8, n = 10$. The green curves are the densities of the test statistics from *Independent Null*. The black curves are the densities under *Dependent Null*. And the red curves are the densities under *Dependent Alternative*.



(a) Log of $RHT_{Bootstrap}$, $p \geq n$. (b) Log of RHT, $p \geq n$.

(c) Log of F^+, $p \geq n$. (d) Log of T_D, $p \geq n$. (e) T_{BS}, $p \geq n$.

Figure S2: Distribution of F^+, T_D and T_{BS} statistics on a simulated example with $p = 20, n = 10$. The green curves are the densities of the test statistics from *Independent Null*. The black curves are the densities under *Dependent Null*. And the red curves are the densities under *Dependent Alternative*.

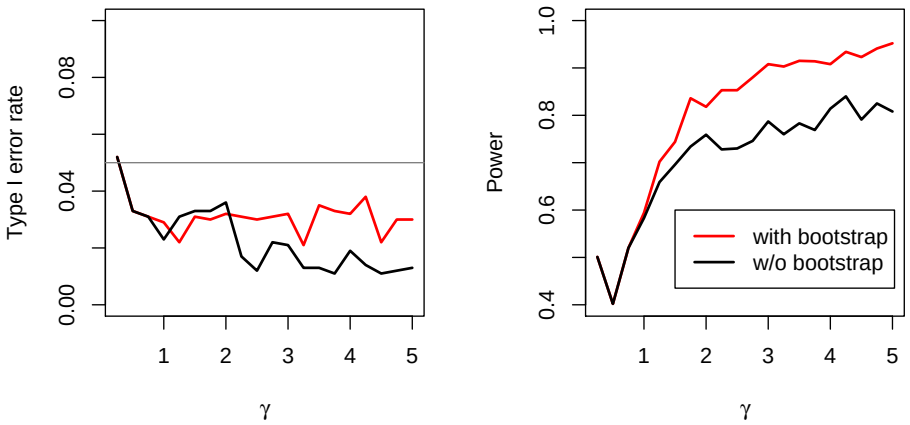
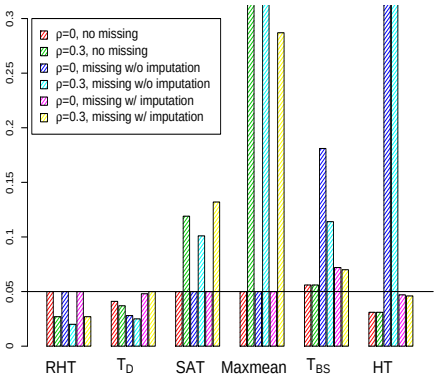
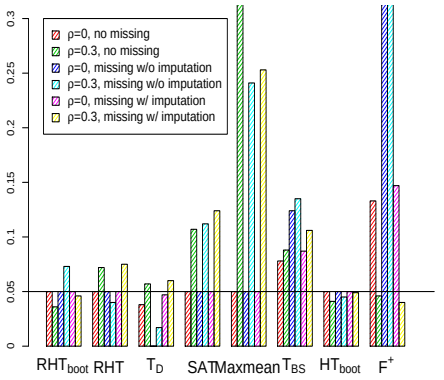


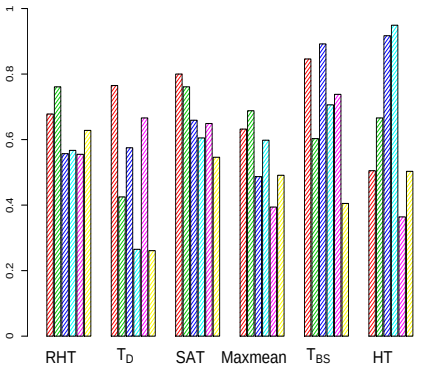
Figure S3: Comparing type I error rate and power of RHT test with and without bootstrap, as $\gamma = p/n$ increases (we fix sample size at $n = 20$ and increases p).



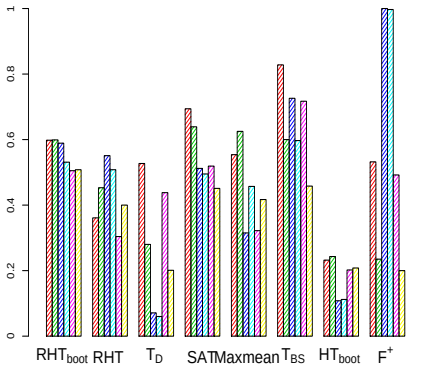
(a) Type I error rate, when $p < n$ ($n_x = n_y = 20, p = 20$)



(b) Type I error rate, when $p \geq n$ ($n_x = n_y = 5, p = 20$)



(c) Power, when $p < n$



(d) Power, when $p \geq n$

Figure S4: Comparing type I error rates and power of `RHT`, `SAT`, `HT`/F^+, T_D and T_{BS} tests at 0.05 p-value cutoff in two-sample tests.

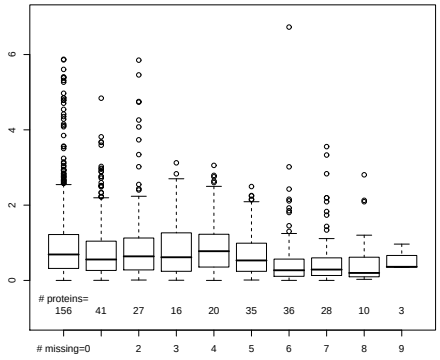


Figure S5: Boxplots of pooled absolute protein log 2 ratios stratified by # missing observation from the hormone therapy proteomics study.

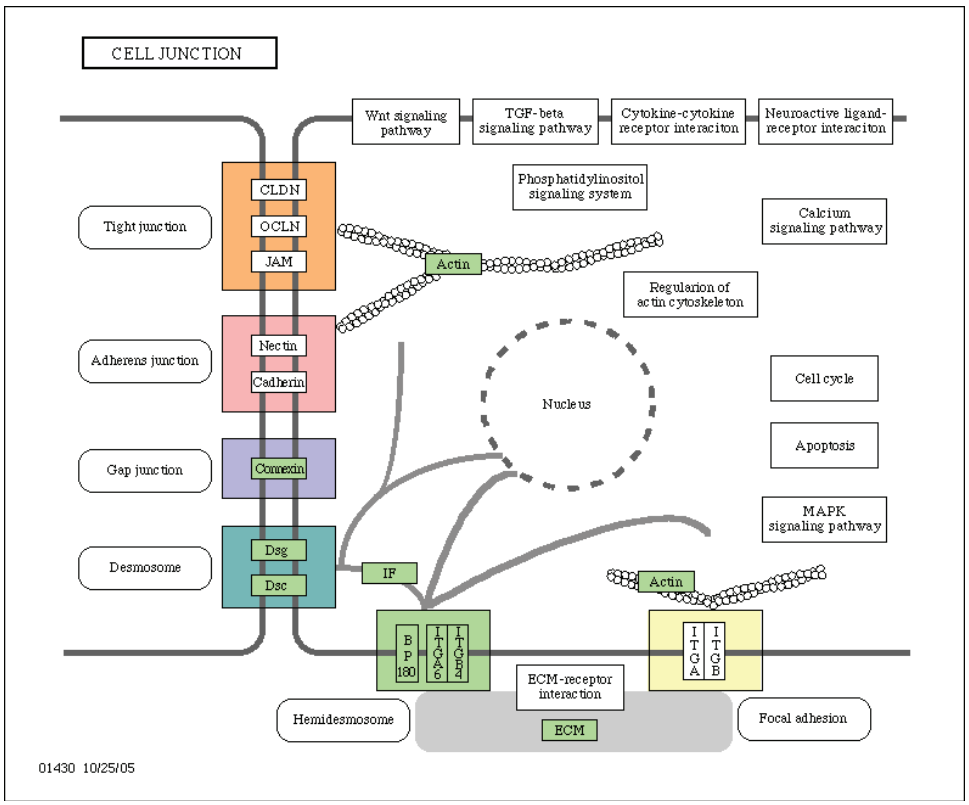


Figure S6: The cell communication pathway from KEGG. This figure shows that four out of the five identified pathways, cell communication, ECM receptor interaction, focal adhesion and TGF-beta signaling pathways, are inter-related.