**Supplementary Text:** Highly Efficient Hypothesis Testing Methods for Regression-type Tests with Correlated Observations and Heterogeneous Variance Structure

Yun Zhang  Gautam Bandyopadhyay  David J. Topham  Ann R. Falsey  Xing Qiu

### S1 Mathematical Derivations

Using the notions introduced in the main manuscript, we first establish some useful identities to facilitate our subsequent mathematical derivations. For \( S \), the standardized structure matrix, and \( J \), a matrix of all 1’s with corresponding dimensionality, we have

\[
JS^{-1}J = 1 \left(1'S^{-1}1\right) 1'
\]

\[= J, \tag{S1}\]

and

\[
(S - J)S^{-1}(S - J) = (S - J)(I - S^{-1}J)
\]

\[= S - J + JS^{-1}J
\]

\[= S - J. \tag{S2}\]

#### S1.1 Centered data

First, we use the generalized least squares approach to estimate the weighted mean for the original data. In our case, simple sample mean is not efficient because we consider non i.i.d. data. For the mean estimation, we reformulate it into an equivalent regression model with only an intercept, \( Y = 1\mu + \epsilon, \epsilon \sim N(0, \Sigma) \), i.e. the null hypothesis case of Equation (3). For the regression model with non-i.i.d. error, weighted least square (WLS) estimator is more efficient than ordinary least square (OLS).

\[
\hat{\mu}(Y) = (1'S^{-1}1)^{-1}1'S^{-1}Y
\]

\[= (1'S^{-1}1)^{-1}1'S^{-1}Y
\]

\[= 1'S^{-1}Y. \]

The last equality is based on Equation (5) in the main manuscript.

Recall that we define the centered data as

\[
Y^{(1)} := Y - 1\hat{\mu} = (I - JS^{-1}) Y.
\]

*Department of Biostatistics and Computational Biology, University of Rochester, Rochester, New York, U.S.A. Tel: 001-585-275-0666; Email: xing_qiu@urmc.rochester.edu.
By construction, $\hat{\mu}$ is the best linear unbiased estimator (BLUE) of $\mu$ under $H_0$. Therefore, the joint distribution of $Y^{(1)}$ under $H_0$ is $Y^{(1)} \sim \mathcal{N}(0, \Sigma^{(1)})$, where

\[
\Sigma^{(1)} = (I - JS^{-1}) \sigma^2 S (I - JS^{-1})' \\
= \sigma^2 (S - 2J + JS^{-1}J) \\
= \sigma^2 (S - J).
\]

Under $H_1$, $Y^{(1)}$ has the same covariance matrix and the following expectation vector

\[
E(Y^{(1)}) = (I - JS^{-1}) E(Y) \\
= (I - JS^{-1}) (1\mu + x\beta) \\
= (I - JS^{-1}) x\beta.
\]

**S1.2 The B-map**

Based on the eigen-decomposition of $S - J$ (Equation (6)), we construct $Y^{(2)}$ as follows

\[
Y^{(2)} := \Lambda_{n-1}^{-1/2} T_{n-1}' Y^{(1)} = \Lambda_{n-1}^{-1/2} T_{n-1}' (I - JS^{-1}) Y \\
= \Lambda_{n-1}^{-1/2} T_{n-1}' (S - J) S^{-1} Y \\
= \underbrace{\Lambda_{n-1}^{1/2} T_{n-1}' S^{-1} Y}_{B} := BY \in \mathbb{R}^{n-1}.
\]

Note that $B$ is a linear map that maps from the $\mathbb{R}^n$ space to an $\mathbb{R}^{n-1}$ subspace.

**S1.3 Proof of Theorem 2.1**

We first start with a geometric discussion of the operator $P$. Let $L_z$ be the linear subspace that is spanned by vectors $1$ and $z$. By construction, columns in matrix $Q$ are the orthogonal basis of $L_z$. Matrix $P$ consists of two operations: $I - QQ'$ and $Q \text{Rot} Q'$. Note that $QQ'$ is an orthogonal projection onto the subspace $L_z$, so the first operation $I - QQ'$ is the “complementary” projection on $L_z^\perp$, the subspace that is orthogonal to $L_z$. Such a projection will keep any vector in $L_z^\perp$ invariant, which is Equation (11). The second operation is a rotation in $L_z$, that rotates vector $z$ to vector $1$ by construction, which is Equation (10). Consider an arbitrary vector $v \in \mathbb{R}^{n-1}$, we can always decompose it into $v = v_1 + v_2$, such that $v_1 \in L_z$ and $v_2 \perp v_1$. The $P$ transformation will rotate $v_1$ and keep $v_2$ invariant; therefore $P$ must be an orthogonal transformation, which is Equation (9).

Below we present a rigorous proof of Theorem 2.1 based on matrix algebra.

**Proof.** First, let us show that $P$ is an orthogonal matrix. Remember that $\text{Rot}$ is a rotation matrix such that $\text{Rot} \cdot \text{Rot}' = I_{2 \times 2}$, and $Q_{(n-1) \times 2}$ is a semi-orthogonal matrix such that $Q'Q = I_{2 \times 2}$. We have

\[
PP' = (I - QQ' + Q \cdot \text{Rot} \cdot Q') (I - QQ' + Q \cdot \text{Rot}' \cdot Q') \\
= I - QQ' + Q \cdot \text{Rot}' \cdot Q' - QQ' + QQ'QQ' - QQ'Q \cdot \text{Rot}' \cdot Q' \\
+ Q \cdot \text{Rot} \cdot Q' - Q \cdot \text{Rot} \cdot Q'QQ' + Q \cdot \text{Rot} \cdot Q'Q \cdot \text{Rot}' \cdot Q' \\
= I.
\]

Next, let us prove Equation (11). By construction, $u$ is orthogonal to both $1$ and $z$. Therefore, $A'u = 0$, which implies that $Q'u = 0$, based on the assumption that $A$ is of full-rank. If $A_{(n-1) \times 2} :=
\((1_{\bar{n}-1}|z)\) is not of full-rank, \(c\) must be already in the direction of \(1\) and we don’t need the \(P\) transformation in the first place. Hence, \(Pu = (I - QQ' + Q \cdot \text{Rot} \cdot Q')u = u\).

Lastly, the proof of (10) is as follows. Notice that by definition of QR decomposition, we can write

\[
(1|z) = (Q_1|Q_2) \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix},
\]

which implies

\[Q_1 = \frac{1}{\sqrt{n-1}}, \quad R_{11} = \sqrt{n-1}.
\]

Furthermore,

\[z = Q \begin{pmatrix} R_{12} \\ R_{22} \end{pmatrix} = R_{12}Q_1 + R_{22}Q_2\]

implies that

\[R_{12} = \frac{\langle z, 1 \rangle}{\sqrt{n-1}} = \xi \cdot \|z\|, \quad R_{22} = \sqrt{\|z\|^2 - R_{12}^2} = \sqrt{1 - \xi^2 \cdot \|z\|}.
\]

Based on the following matrix calculations, we conclude that

\[(I - QQ')c = (I - QQ')Q \begin{pmatrix} R_{12} \\ R_{22} \end{pmatrix} = 0,
\]

\[Q \cdot \text{Rot} \cdot Q'z = Q \cdot \begin{pmatrix} \xi & \sqrt{1 - \xi^2} \\ -\sqrt{1 - \xi^2} & \xi \end{pmatrix} \begin{pmatrix} R_{12} \\ R_{22} \end{pmatrix} = (Q_1|Q_2) \begin{pmatrix} \xi & \sqrt{1 - \xi^2} \\ -\sqrt{1 - \xi^2} & \xi \end{pmatrix} \begin{pmatrix} \xi \\ \sqrt{1 - \xi^2} \end{pmatrix} \cdot \|z\| = Q_1 \cdot \|z\| = \frac{\|z\|}{\sqrt{n-1}} \cdot 1.
\]

Careful readers may notice that there are several choices needed to be made when constructing \(P\). For example, whether to use a clockwise or anti-clockwise rotation matrix, or whether to take positive or negative signs in some of the steps in the QR decomposition. These choices are irrelevant for two-sided tests; but they can affect the direction of one-sided tests.

In our implementation, we force the first row of \(\text{Rot}\) and the second column of \(R\) (highlighted in red boxes) be identical. In this way, \(Q \cdot \text{Rot} \cdot Q'z = Q_1 \cdot \|z\|\) is guaranteed, and the rotation direction and the signs can be determined solely by \(Q_1\). As mentioned before, we need to ensure that \(P\) rotates the coefficient vector \(z\) to the direction of \(+1\), so that the one-sided test after rotation is consistent with the original problem. Hence, if \(Q_1\) is negative, we will replace \(P\) by \(-P\).

### S1.4 Proof of Proposition 2.3

**Proof.** We claim that the linear transformation \(D\) and the two constants \(u, v\) in Equation (15) can be constructed as follows

\[
D = T_{n \times (n-1)}A^{1/2}_{(n-1) \times (n-1)}P^{-1}_{(n-1) \times (n-1)},
\]

\[u = \frac{\beta \|z\|}{\sqrt{n - 1}}, \quad v = \mu + 1_n'x.\]

3
Here $T$ and $\Lambda$ are defined in Equation (6), $z$ is defined in Equation (8), and $P$ is defined in Theorem 2.1.

Let $Y := D (\hat{Y} + u_{1_{n-1}}) + (Y^* + v)1_n$, then

$$E(Y) = TA^{1/2} (P^{-1} \cdot u_{1_{n-1}}) + v1_n = TA^{1/2} \cdot \frac{u\sqrt{n-1}}{\|z\|} \cdot z + v1_n$$

$$= \frac{u\sqrt{n-1}}{\|z\|} \cdot (S - J) S^{-1} x + v1_n$$

$$= \beta (1_n - 1_n \cdot 1_n' S^{-1}) x + v1_n$$

$$= \beta x + (v - \beta \cdot 1_n' S^{-1} x) 1_n$$

$$= \beta x + \mu 1_n.$$ 

And,

$$\text{cov}(Y) = \sigma^2 DD' + \sigma^2 1_n 1_n'$$

$$= \sigma^2 T\Lambda^{1/2} P^{-1} \cdot P\Lambda^{1/2} T' + \sigma^2 J = \sigma^2 S.$$ 

The second part of Proposition 2.3 claims that the PB-transformation can convert $Y$ back to a sequence random variables with zero mean if $\beta = 0$ and a nonzero mean otherwise. This is because

$$(S - J) S^{-1} (S - J) = S - J - J = S - J,$$

$$TAT' \cdot S^{-1}. TAT' = TAT', \quad T' S^{-1} T = \Lambda^{-1}.$$

$$\|B1_n\|^2 = 1_n' B'B1_n = 1_n' S^{-1} T\Lambda^{1/2} \cdot \Lambda^{1/2} T' S^{-1} 1_n$$

$$= 1_n' S^{-1} (S - J) S^{-1} 1_n = 1_n' S^{-1} 1_n - 1_n' S^{-1} JS^{-1} 1_n$$

$$= 1 - 1_n' S^{-1} 1_n 1_n' S^{-1} 1_n = 1 - 1^2 = 0.$$ 

Therefore,

$$\text{PBD} = PA^{1/2} T' S^{-1} \cdot T\Lambda^{1/2} P^{-1} = I_{(n-1) \times (n-1)},$$

$$\hat{Y} = \text{PBY} = \text{PB} (D (\hat{Y} + u1_{1_{n-1}}) + (Y^* + v)1_n)$$

$$= \hat{Y} + u1_{1_{n-1}} + (Y^* + v) PB1_n$$

$$= \hat{Y} + u1_{1_{n-1}} + \hat{Y} + \frac{\beta \|z\|}{\sqrt{n-1}}1_{1_{n-1}},$$

Hence, $E\hat{Y} = 0$ if $\beta = 0$ and $E\hat{Y} \neq 0$ otherwise.

\[\square\]

S2 Correlation Estimation

S2.1 Correlation estimation for repeated measurements

In this section, we reframe our regression model in the following matrix form:

$$y = X\beta + \epsilon, \quad X = (1, x), \quad \beta = (\mu, \beta)'.$$ 

Suppose there are a priori weights available in a diagonal matrix $W$, define the residuals

$$\hat{\epsilon}_{\text{WLS}} = y - X\hat{\beta}_{\text{WLS}} = y - X(X'W^{-1}X)^{-1}X'W^{-1}y.$$
Careful readers may notice that the errors are also correlated due to the repeated measurements. In linear model theory, the above \( \hat{\beta}_{\text{WLS}} \) is the BLUE estimator if errors are uncorrelated, and it is an unbiased estimator of \( \beta \) if errors are correlated. Therefore, the above residual estimator is valid even with correlated errors in large samples. By construction, the variance-covariance structure of the above residuals preserves the structure in the original data, which consists of the weighting part and the correlation part by design. Before estimating the correlation, we first re-balance the weights and work on the following residual estimates

\[
\hat{\epsilon} = W^{-1} \hat{\epsilon}_{\text{WLS}}. \tag{S3}
\]

Further, we have the following assumptions on the correlation structure for repeated measurements:

- Asymptotic normality of the residuals \( \hat{\epsilon} = (\hat{\epsilon}_1, \cdots, \hat{\epsilon}_n)' \) such that
  
  \[
  \hat{\epsilon} \sim \mathcal{N}(0, \Sigma_{\epsilon}).
  \]

  for some \( \Sigma_{\epsilon} \neq I \).

- Common variance scaling factor \( \sigma^2 \), so that
  
  \[
  \Sigma_{\epsilon} = \sigma^2 \cdot \text{Cor},
  \]

  where \( \text{Cor} \) is the correlation matrix.

- With appropriate data rearrangement, we can write \( \text{Cor} \) as a block-diagonal matrix. Suppose there are \( l = 1, \cdots, L \) blocks,

  \[
  \text{Cor} = \begin{pmatrix}
  \text{Cor}_1 \\
  \vdots \\
  \text{Cor}_L
  \end{pmatrix}.
  \]

- Common correlation magnitude \( \rho \) for all the blocks, i.e. \( \text{Cor} = \text{Cor}(\rho) \). For each block, assume interchangeable correlation structure, so that

  \[
  \text{Cor}_l(\rho) = \begin{pmatrix}
  1 & \rho & \cdots & \rho \\
  \rho & 1 & \cdots & \rho \\
  \vdots & \cdots & \ddots & \cdots \\
  \rho & \cdots & \rho & 1
  \end{pmatrix}_{n_l \times n_l} = (1 - \rho) \cdot I_{n_l \times n_l} + \rho \cdot J_{n_l \times n_l},
  \]

  where \( n_l \) is the size of the \( l \)th block, and \( J_{n_l \times n_l} = 1_{n_l} \cdot 1_{n_l}' \). Note, \( n = \sum_{l=1}^L n_l \).

For the above specific structure of the correlation matrix, we derive the following lemma that facilitates our future mathematical computation.

**Lemma S2.1.** *Denote \( J = 11' \), the outer product of \( 1 \) and \( 1 \). Let \( \text{Cor} = aI + bJ \) for \( a, b \in \mathbb{R} \) and \( I, J \in \mathbb{R}^{n \times n} \). Then,*

\[
\det(\text{Cor}) = a^n \left( 1 + \frac{nb}{a} \right) \quad \text{and} \quad \text{Cor}^{-1} = \frac{1}{a} \cdot I - \frac{b}{a(a + nb)} \cdot J.
\]

*Proof. Write \( \text{Cor} = E + uv' \) with \( E = aI, u = b1 \) and \( v = 1 \). The above results follow immediately from the Sherman-Morrison formula.*
The maximum likelihood estimators (MLEs) are
\[ \hat{\rho}_{MLE}, \hat{\sigma}^2_{MLE} = \arg\max_{\rho, \sigma^2} -2\mathcal{L}(\rho, \sigma^2). \]

In the following subsections, we are going to derive explicitly the MLEs under the balanced design, i.e. \( n_1 = \cdots = n_L = n/L \). Under the unbalanced design, i.e. \( n_l \)'s are not all the same for \( l = 1, \cdots, L \), it is not straightforward to derive explicitly the MLEs. Though gradient descent algorithms may provide numerical solutions to the MLEs, we are going to provide a computationally efficient alternative method based on moment matching.

### S2.1.2 Balanced Design
WLOG, denote \( n_l = n_1 = n/L \) for all \( l \). Then
\[ -2\mathcal{L}(\rho, \sigma^2) = n \log(\sigma^2) + L \log((1 - \rho)^{n_l - 1}(1 - \rho + n_l \rho)) \]
\[ + \frac{1}{\sigma^2} \cdot \sum_{l=1}^{L} \hat{\epsilon}_l^2 \left( \frac{1}{1 - \rho} \right) \cdot \left( \frac{\rho}{(1 - \rho)(1 - \rho + n_l \rho)} \right) \cdot \hat{\epsilon}_l, \]
\[ = \left( \frac{1}{1 - \rho} \right) SS_1 - \left( \frac{1}{(1 - \rho)(1 - \rho + n_l \rho)} \right) SS_2, \]
where \( SS_1 = \sum_l \hat{\epsilon}_l \hat{\epsilon}_l \) and \( SS_2 = \sum_l \hat{\epsilon}_l^T J \hat{\epsilon}_l = \sum_l \sum_{i,j} \hat{\epsilon}_{li} \hat{\epsilon}_{lj} \) are two sufficient statistics.

Take partial derivative with respect to \( \sigma^2 \) and set to zero,
\[ \frac{\partial}{\partial \sigma^2} (-2\mathcal{L}(\rho, \sigma^2)) = \frac{n}{\sigma^2} - \frac{1}{\sigma^4} \left( \frac{1}{1 - \rho} SS_1 - \frac{\rho}{(1 - \rho)(1 - \rho + n_l \rho)} SS_2 \right) = 0 \]
\[ \Rightarrow \hat{\sigma}^2_{MLE} = \frac{1}{n} \left( \frac{1}{1 - \rho} SS_1 - \frac{\rho}{(1 - \rho)(1 - \rho + n_l \rho)} SS_2 \right). \]
Also, take partial derivative with respect to $\rho$ and set to zero,

$$
\frac{\partial}{\partial \rho} \left( -2\mathcal{L}(\rho, \sigma^2) \right) = L \frac{-n_1(n_1 - 1)\rho}{(1 - \rho)(1 - \rho + n_1\rho)} + \frac{1}{\sigma^2} \left( \frac{1}{(1 - \rho)^2} SS_1 - \frac{1 - \rho^2 - n_1\rho^2}{(1 - \rho)^2(1 - \rho + n_1\rho)^2} SS_2 \right) = 0.
$$

By substituting $\sigma^2 = \hat{\sigma}^2_{\text{MLE}}$ from above, we get

$$
\hat{\rho}_{\text{MLE}} = \frac{SS_2 - SS_1}{(n_1 - 1)SS_1}.
$$

### S2.1.3 Unbalanced Design

If the blocks are not of the same size, i.e. $n_l$’s may be different, then the full log-likelihood becomes

$$
-2\mathcal{L}(\rho, \sigma^2) = n \log(\sigma^2) + \sum_{l=1}^{L} \log \left( (1 - \rho)^{n_l - 1}(1 - \rho + n_1\rho) \right)
+ \frac{1}{\sigma^2} \sum_{l=1}^{L} \hat{\epsilon}_l \left( \frac{1}{1 - \rho} \cdot I - \frac{\rho}{(1 - \rho)(1 - \rho + n_1\rho)} \cdot J \right) \hat{\epsilon}_l.
$$

Due to the different $n_l$’s in the denominator, $\hat{\rho}_{\text{MLE}}$ has no explicit form in terms of $\hat{\epsilon}$. Although, we may apply some gradient descent algorithm to obtain the numerical value at convergence, the method is time consuming and may fail to converge. Here, we provide an alternative method that is computationally efficient and has no convergence issue. By matching the following moments, we have

\[
\mathbb{E} SS_1 = n\sigma^2,
\]
\[
\mathbb{E} (SS_2 - SS_1) = \sigma^2 \rho \sum_{l=1}^{L} n_l(n_l - 1),
\]

where $SS_1$ and $SS_2$ are defined above. So,

$$
\hat{\sigma}^2_{\text{moment}} = \frac{SS_1}{n},
\hat{\rho}^2_{\text{moment}} = \frac{SS_2 - SS_1}{\sum_{l=1}^{L} n_l(n_l - 1)} SS_1.
$$

It is easy to see that $\hat{\rho}^2_{\text{moment}} = \hat{\rho}^2_{\text{MLE}}$ if balanced design.

### S2.2 Correlation estimation in simulations

The following table compares the correlation estimation based on Equation (13), which is $\hat{\rho}_{\text{PB}}$, or Equation (14), which is $\hat{\rho}_{\text{LMER}}$, in simulation studies. Overall, we find that our proposed correlation estimator has both smaller bias and smaller variance than its counterpart in LMER.
Table S1: Correlation estimation in simulation studies, mean (standard deviation)

|                  | SIM1: Group Comparison | SIM2: Regression Test |
|------------------|------------------------|-----------------------|
|                  | ρ=0.2 | ρ=0.4 | ρ=0.6 | ρ=0.8 | ρ=0.2 | ρ=0.4 | ρ=0.6 | ρ=0.8 |
| ̂ρ\text{PB}      | 0.192 (0.217) | 0.382 (0.194) | 0.577 (0.156) | 0.775 (0.096) | 0.178 (0.175) | 0.370 (0.174) | 0.566 (0.157) | 0.768 (0.123) |
| ̂ρ\text{LMER}    | 0.326 (0.340) | 0.496 (0.350) | 0.698 (0.311) | 0.889 (0.176) | 0.273 (0.211) | 0.481 (0.202) | 0.678 (0.150) | 0.847 (0.075) |

S3 Simulation Design

In our simulation studies, we design the covariance structure of an \( n \)-dimensional joint distribution based on Equation (12) in the main manuscript. Let \( \mathbf{W} = \text{diag}(w_1, \cdots, w_n) \), and the weights are \( w_i = 1/\sigma_i^2 \) for \( i = 1, \cdots, n \), where \( \sigma_i^2 \) is the variance of the \( i \)th marginal distribution. We specify \( \text{Cor} \) in the following two simulation scenarios.

SIM1: This simulation is designed for completely paired group comparison. Suppose there are two groups, namely A and B. The total sample size is \( n = 40 \), within which \( n_A = n_B = 20 \). All observations are paired in the two groups. The correlation matrix for SIM1 can be expressed as

\[
\text{Cor}(\rho) = \begin{pmatrix}
1 & \rho & \cdots & \rho & 1 \\
\rho & 1 & \cdots & \rho & 1 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\rho & \rho & \cdots & 1 & \rho \\
1 & \rho & \cdots & \rho & 1
\end{pmatrix}.
\]

The corresponding group indicator is \( \mathbf{1}_B = (0, 1, \cdots, 0, 1)' \). The weights, \( \sigma_1^2, \cdots, \sigma_n^2 \), are randomly drawn from \text{Unif}(1, 6). Effect size is 0.2.

SIM2: This simulation is designed for regression-type test with a continuous covariate. Let \( n = 80 \) and generate the covariate from \( \mathcal{N}(3, 1) \). The observations are mixed with singletons, doubles, and triplets, i.e. subjects with one, two, and three repeated measurements, respectively. Specifically, we generate 10 doubles and 10 triplets; and the rest are singletons. The correlation matrix for SIM2 can be expressed as

\[
\text{Cor}(\rho) = \begin{pmatrix}
1 & \rho & \cdots & \rho & 1 \\
\rho & 1 & \cdots & \rho & 1 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\rho & \rho & \cdots & 1 & \rho \\
1 & \rho & \cdots & \rho & 1
\end{pmatrix}.
\]

For the weights, \( \sigma_1^2, \cdots, \sigma_n^2 \) are random draws from \text{Unif}(1, 2). Effect size is 0.05.

For each simulation scenario, we consider two underlying distributions, i.e. normal distribution and double exponential distribution; as well as four correlation levels, i.e. \( \rho = 0.2, 0.4, 0.6 \) and 0.8. We generate the correlated observations using Proposition 2.3. In each simulation, we test 2,000 hypotheses simultaneously: 1,000 from the null distribution and 1,000 from the alternative distribution. We repeat 20 times for each simulation.

In the simulation studies, we investigated two versions of the PB-transformed tests, i.e. with \( \rho \) (oracle) or \( ̂\rho \) (estimation), and compared the proposed parametric and semiparametric tests to the following methods in each category.
• SIM1 & normal: weighted linear mixed-effects model with test for the fixed effect (weighted LMER), Rfast implementation of LMER without weights (Rfast LMER), paired $t$-test, weighted regression $t$-test, and Welch’s $t$-test.

• SIM1 & double exponential: test performed by \texttt{svyranktest()}, Friedman test, Wilcoxon signed rank test, and Wilcoxon rank-sum test.

• SIM1 & logistic: test performed by \texttt{svyranktest()}, Friedman test, Wilcoxon signed rank test, and Wilcoxon rank-sum test.

• SIM2 & normal: weighted LMER, Rfast LMER, weighted regression $t$-test, and Welch’s $t$-test.

• SIM2 & double exponential: Spearman rank correlation tests with and without B-transformation. One important purpose of this simulation is to show the advantage of the P-map.

• SIM2 & logistic: Spearman rank correlation tests with and without B-transformation.

To our best knowledge, the above tests are the most widely used practical solutions for each category.
S3.1 Additional simulation results of $\rho = 0.4$ and $\rho = 0.6$

**Figure S1:** ROC curves for group comparison tests. In SIM1, seven parametric methods and six rank-based methods are compared. AUC values are reported in the legend. Plot A is zoomed to facilitate the view of curves that overlay on top of each other. When curves are severely overlayed, line widths are slightly adjusted. For both $\rho = 0.4$ and $\rho = 0.6$, the PB-transformed parametric and rank-based tests outperform all other tests.
Figure S2: ROC curves for regression tests. In SIM2, six parametric methods and four rank-based methods are compared. AUC values are reported in the legend. Plot A is zoomed to facilitate the view of curves that overlay on top of each other. When curves are severely overlayed, line widths are slightly adjusted. For both $\rho = 0.4$ and $\rho = 0.6$, the PB-transformed parametric and rank-based tests outperform all other tests.
Table S2: **Type-I error and power comparison for group comparison tests.** At the 5% significance level, mean and standard deviation (in brackets) of the type-I error rate and power over 20 sets of SIM1 data are reported.

|               | $\rho = 0.4$ |         | $\rho = 0.6$ |         |
|---------------|--------------|---------|--------------|---------|
|               | Type-I error | Power   | Type-I error | Power   |
| Normal        |              |         |              |         |
| PB.t (oracle) | 0.036 (0.005)| 0.817 (0.012) | 0.036 (0.005)| 0.831 (0.011)|
| PB.t (estimation) | 0.035 (0.005)| 0.812 (0.012) | 0.033 (0.006)| 0.822 (0.011)|
| Weighted LMER | 0.031 (0.007)| 0.734 (0.013) | 0.024 (0.004)| 0.650 (0.014)|
| Rfast LMER    | 0.059 (0.005)| 0.515 (0.017) | 0.061 (0.005)| 0.497 (0.015)|
| Weighted regression $t$-test | 0.019 (0.004)| 0.710 (0.016) | 0.010 (0.003)| 0.559 (0.015)|
| Paired $t$-test | 0.047 (0.006)| 0.473 (0.014) | 0.048 (0.004)| 0.452 (0.014)|
| Welch’s $t$-test | 0.016 (0.004)| 0.330 (0.011) | 0.006 (0.002)| 0.207 (0.012)|
| Double Exponential |              |         |              |         |
| PB.wilcox (oracle) | 0.031 (0.007)| 0.880 (0.012) | 0.031 (0.007)| 0.892 (0.012)|
| PB.wilcox (estimation) | 0.041 (0.008)| 0.843 (0.016) | 0.043 (0.007)| 0.837 (0.017)|
| svyranktest    | 0.115 (0.012)| 0.751 (0.013) | 0.108 (0.010)| 0.683 (0.015)|
| Friedman       | 0.058 (0.006)| 0.475 (0.015) | 0.058 (0.006)| 0.475 (0.011)|
| Wilcoxon signed rank | 0.059 (0.008)| 0.541 (0.018) | 0.057 (0.008)| 0.537 (0.018)|
| Wilcoxon rank-sum | 0.023 (0.004)| 0.475 (0.016) | 0.008 (0.002)| 0.340 (0.017)|

Table S3: **Type-I error and power comparison for regression tests.** At the 5% significance level, mean and standard deviation (in brackets) of the type-I error rate and power over 20 sets of SIM2 data are reported.

|               | $\rho = 0.4$ |         | $\rho = 0.6$ |         |
|---------------|--------------|---------|--------------|---------|
|               | Type-I error | Power   | Type-I error | Power   |
| Normal        |              |         |              |         |
| PB.t (oracle) | 0.046 (0.007)| 0.722 (0.014) | 0.045 (0.007)| 0.696 (0.013)|
| PB.t (estimation) | 0.045 (0.007)| 0.720 (0.014) | 0.042 (0.008)| 0.686 (0.013)|
| Weighted LMER | 0.049 (0.008)| 0.706 (0.014) | 0.047 (0.007)| 0.653 (0.014)|
| Rfast LMER    | 0.065 (0.010)| 0.671 (0.013) | 0.066 (0.010)| 0.634 (0.012)|
| Weighted regression $t$-test | 0.049 (0.007)| 0.669 (0.012) | 0.055 (0.007)| 0.557 (0.015)|
| Welch’s $t$-test | 0.052 (0.008)| 0.604 (0.011) | 0.053 (0.007)| 0.496 (0.015)|
| Double Exponential |              |         |              |         |
| PB.wilcox (oracle) | 0.042 (0.007)| 0.777 (0.015) | 0.042 (0.008)| 0.749 (0.014)|
| PB.wilcox (estimation) | 0.076 (0.008)| 0.684 (0.017) | 0.075 (0.011)| 0.659 (0.017)|
| B.spearman (estimation) | 0.082 (0.010)| 0.664 (0.015) | 0.086 (0.006)| 0.614 (0.014)|
| Spearman test  | 0.076 (0.009)| 0.567 (0.011) | 0.075 (0.006)| 0.472 (0.016)|
S3.2 Logistic distribution simulation results

Figure S3: ROC curves for group comparison tests. In SIM1, six rank-based methods are compared. AUC values are reported in the legend. For all correlation levels, the PB-transformed parametric and rank-based tests outperform all other tests.
| Simulation | Logistic, ρ | AUC value | PB.wilcox (oracle) | PB.wilcox (estimation) | B.spearman (estimation) | Spearman |
|------------|-------------|-----------|---------------------|------------------------|-------------------------|----------|
| A          | ρ = 0.2     |           |                     | 0.953                  | 0.935                   | 0.928    |
| B          | ρ = 0.8     |           |                     | 0.937                  | 0.917                   | 0.891    |
| C          | ρ = 0.4     |           |                     | 0.942                  | 0.927                   | 0.913    |
| D          | ρ = 0.6     |           |                     | 0.935                  | 0.918                   | 0.904    |

**Figure S4:** ROC curves for regression tests. In SIM2, four rank-based methods are compared. AUC values are reported in the legend. For all correlation levels, the PB-transformed parametric and rank-based tests outperform all other tests.
Table S4: **Type-I error and power comparison for group comparison tests.** At the 5% significance level, mean and standard deviation (in brackets) of the type-I error rate and power over 20 sets of SIM1 data are reported.

|                     | $\rho = 0.2$                  | $\rho = 0.8$                  |
|---------------------|-------------------------------|-------------------------------|
|                     | Type-I error | Power   | Type-I error | Power   |
| Logistic            |               |         |               |         |
| PB.wilcox (oracle)  | 0.050 (0.007) | 0.817 (0.015) | 0.047 (0.007) | 0.897 (0.013) |
| PB.wilcox (estimation) | 0.042 (0.007) | 0.819 (0.018) | 0.035 (0.005) | 0.886 (0.011) |
| svyranktest         | 0.104 (0.013) | 0.811 (0.018) | 0.096 (0.011) | 0.597 (0.014) |
| Friedman            | 0.045 (0.005) | 0.430 (0.018) | 0.045 (0.005) | 0.437 (0.015) |
| Wilcoxon signed rank | 0.052 (0.006) | 0.528 (0.011) | 0.051 (0.007) | 0.520 (0.014) |
| Wilcoxon rank-sum   | 0.036 (0.006) | 0.543 (0.014) | 0.001 (0.001) | 0.143 (0.009) |

|                     | $\rho = 0.4$                  | $\rho = 0.6$                  |
|---------------------|-------------------------------|-------------------------------|
|                     | Type-I error | Power   | Type-I error | Power   |
| Logistic            |               |         |               |         |
| PB.wilcox (oracle)  | 0.050 (0.007) | 0.791 (0.014) | 0.050 (0.006) | 0.810 (0.014) |
| PB.wilcox (estimation) | 0.040 (0.006) | 0.790 (0.014) | 0.040 (0.006) | 0.798 (0.017) |
| svyranktest         | 0.101 (0.011) | 0.743 (0.011) | 0.100 (0.011) | 0.667 (0.018) |
| Friedman            | 0.046 (0.005) | 0.407 (0.015) | 0.045 (0.007) | 0.404 (0.017) |
| Wilcoxon signed rank | 0.050 (0.006) | 0.497 (0.018) | 0.051 (0.007) | 0.494 (0.016) |
| Wilcoxon rank-sum   | 0.020 (0.004) | 0.421 (0.016) | 0.008 (0.003) | 0.284 (0.014) |

Table S5: **Type-I error and power comparison for regression tests.** At the 5% significance level, mean and standard deviation (in brackets) of the type-I error rate and power over 20 sets of SIM2 data are reported.

|                     | $\rho = 0.2$                  | $\rho = 0.8$                  |
|---------------------|-------------------------------|-------------------------------|
|                     | Type-I error | Power   | Type-I error | Power   |
| Logistic            |               |         |               |         |
| PB.wilcox (oracle)  | 0.046 (0.007) | 0.784 (0.012) | 0.044 (0.007) | 0.719 (0.009) |
| PB.wilcox (estimation) | 0.057 (0.007) | 0.755 (0.014) | 0.047 (0.008) | 0.671 (0.010) |
| B.spearman (estimation) | 0.058 (0.007) | 0.733 (0.014) | 0.059 (0.007) | 0.629 (0.014) |
| Spearman test       | 0.056 (0.008) | 0.688 (0.012) | 0.061 (0.008) | 0.362 (0.016) |

|                     | $\rho = 0.4$                  | $\rho = 0.6$                  |
|---------------------|-------------------------------|-------------------------------|
|                     | Type-I error | Power   | Type-I error | Power   |
| Logistic            |               |         |               |         |
| PB.wilcox (oracle)  | 0.046 (0.007) | 0.744 (0.010) | 0.045 (0.007) | 0.717 (0.009) |
| PB.wilcox (estimation) | 0.053 (0.007) | 0.719 (0.012) | 0.052 (0.006) | 0.689 (0.012) |
| B.spearman (estimation) | 0.063 (0.007) | 0.696 (0.013) | 0.062 (0.007) | 0.662 (0.013) |
| Spearman test       | 0.061 (0.008) | 0.603 (0.012) | 0.059 (0.009) | 0.507 (0.012) |
S4  Real Data Analyses

In this study, we illustrate the PB-transformed tests in the context of differential expression analysis for next-generation sequencing data.

RNA-seq (Wang et al., 2009) is a deep-sequencing technology that is widely used in gene expression experiments. Depending on different experiment choices, e.g. instrument platforms, sequencing libraries preparation, etc., the number of reads sequenced per sample varies largely ($10^7 - 10^{12}$ (Sims et al., 2014)) from study to study. In addition to the technical variation, the number of read sequenced is also strongly related with the true biological variation within the specimen that has been sampled. Formally, *sequencing depth* is defined as the total number of reads sequenced in each sample. Considerations of sequencing depth in genomic analyses are reviewed in Sims et al. (2014). Here, we use the PB-transformed test to adjust the varying sequencing depth in RNA-seq data analyses.

We download RNA-seq data of breast invasive carcinoma from The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Network, 2012). Primarily, breast cancers can be classified into subtypes based on human epidermal growth factor receptor 2 (HER2) gene (Onitilo et al., 2009; Carey et al., 2006). In this study, we select 35 samples from tissue source site “BH”, which have the following clinical traits: HER2-positive, white race and female sex. In this tissue source site, a subset of the tumor specimens have companion normal tissue specimens.

S4.1 Data preprocessing

Raw RNA-seq count data are normalized by reads per kilobase million (RPKM) (Mortazavi et al., 2008). Figure S5 shows data preprocessing criteria. Sample filtering is based on between-sample correlations (Figure S5a). Samples with very low similarities with other samples, which is defined as median of $(\text{corr}_{i,j})_j \leq 0.6$, where $\text{corr}_{i,j}$ is the correlation between sample $i$ and sample $j$, are removed. After this step, 28 samples are kept; and the mean correlation across these remaining samples is 0.72.

Next, we apply nonspecific filtering based on log2-transformed gene expression values (Figure S5b). The log2-transformation is defined as $\log_2(X + 1)$, where $X$ are the normalized counts. This histogram shows that the distribution of mean expression level for individual genes can be roughly divided into “low” and “high” expression clusters. Low expression genes (i.e. mean expression value after log2-transformation is $\leq 1.5$) are filtered, and 11,453 genes remain in the working data.
S4.2 Weights for RNA-seq samples

To justify the choice of weights in Equation (18), we consider a simple count model based on the binomial distribution. The same principle is extendable to more advanced models. Borrowing the notations in Robinson et al. (2010), let $Y_{ij}$ be the raw count for gene $i$ in sample $j$. We assume that $Y_{ij} \sim \text{Binom}(M_j, p_i)$, where $M_j$ is the sequencing depth (or library size), and $p_i$ is the relative abundance of gene $i$. The null hypothesis is $H_0: p_{i,\text{tumor}} = p_{i,\text{normal}}$. By large sample theory,

$$\sqrt{M_j} (\hat{p}_i - p_i) \to N(0, \sigma^2),$$

where $\sigma^2$ is the variance for the binomial distribution. Equation (S4) shows that the sampling variability of the relative abundance $p_i$ is inversely proportional to $M_j$. The same large sample property can be applied to more complex models such as Negative Binomial model (Robinson & Smyth, 2007; Anders & Huber, 2010; Robinson et al., 2010; Love et al., 2014).

S4.3 Criteria for differentially expressed gene (DEG)

In our real data analyses, Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) is applied to adjust the raw $p$-values so that we can control the false discovery rate (FDR) at 0.05 level. We also threshold the fold change (FC), which is defined as

$$FC_i = \frac{\sum_{j=1}^{n_A} x_{ij}/n_A}{\sum_{j=1}^{n_B} y_{ij}/n_B}$$

for gene $i$,

where $x_{ij}$ and $y_{ij}$ are expression values, and $n_A$ and $n_B$ are sample sizes for the tumor and normal groups, respectively. FC is a popular gene selection criterion among practitioners, as it reflects the actual mean expression difference between the groups. Using both criteria, we define DEG as the gene with BH-adjusted $p$-value $< 0.01$, and FC $< 1/2$ (downregulated) or $> 2$ (upregulated).
S4.4 Biology of DEG lists

The eleven genes uniquely identified by the PB-transformed t-test are known to be involved in cell survival, proliferation and migration. The CXCR4-CXCL12 chemokine signaling pathway is one of the deregulated signaling pathway uniquely identified by PB-transformed t-test in HER2+ breast cancer cells. This pathway is known to play a crucial role in promoting breast cancer metastasis and has been reported to be associated with poor prognosis (Sun et al., 2014; Müller et al., 2001). Metastatic breast tumor cells have been shown to express high levels of CXCR4 (Xu et al., 2015). After binding to its ligand CXCL12, CXCR4 activates downstream PI3K and MAP kinase signaling pathways that regulate cell proliferation, survival and chemotaxis (Xu et al., 2015; Andreou et al., 2012). CXCL12 is constitutively expressed by different cell types in tissues such as lung, liver and bone marrow (Xu et al., 2015; Cojoc et al., 2013). Secreted CXCL12 from these healthy tissues drives metastatic invasion of CXCR4+ tumor cells into normal tissues (Xu et al., 2015; Sun et al., 2010; Sarvaiya et al., 2013). The PB-transformed t-test also identifies increased PI3KR2 (the regulatory p85 subunit of PI3K) and MAP kinase 13 (p38 δ-subunit). Hyper-activation of both PI3K p85 and p38 were shown to be associated with different types of metastases (Cuenda & Sanz-Ezquerro, 2017; Martini et al., 2014). Both of these intracellular molecules can be activated by CXCR4 signaling (Xu et al., 2015; Cojoc et al., 2013; Sun et al., 2002). Our results demonstrate that these HER2+ breast cancer cells not only had increased CXCR4 transcripts, but also had elevated transcriptional levels of its downstream signaling molecules to support CXCR4-triggered metastasis. Another critical gene annotated by PB-transformed t-test is GAB2. GAB2 has been reported to be upregulated in primary breast tumors as well as in breast cancer cell lines (Bentires-Alj et al., 2006). Co-expression of GAB2 along with HER2 is pivotal to acquire metastatic property by HER2+ cells (Bentires-Alj et al., 2006). Activated GAB2 can in turn stimulate several intracellular signaling molecules such as SHC and p85 subunit of PI3K (Bentires-Alj et al., 2006; Ding et al., 2015), both uniquely annotated by our PB-transformed t-test (SHC2 and PIK3R2).

In contrast, the weighted LMER uniquely identifies six important genes from the same set of data. Three of those genes (COX6B1, COX4I2 and COX5A) are different subunits of cytochrome C oxidase. Another gene, UQCR10, is a cytochrome c reductase. These enzymes are essential for ATP generation during mitochondrial respiration. Since tumor cells need more energy than most of the normal cells, it is expected that those breast cancer cells had over-expression of those genes. Another gene identified by LMER is HSP70 family member HSPA5. The HSP70 family proteins are known to transduce anti-apoptotic signals in many types of cancer cells (Murphy, 2013). Multiple HSP70 inhibitors are being tested for anti-tumor activity (Murphy, 2013).

S4.5 Functional annotation terms from DAVID gene set analyses

Table S6 shows the functional annotation terms identified by DAVID gene set analysis based on the unique genes exclusively belonging to one of the top 2,000 DEG lists of the PB-transformed t-test and the weighted LMER test.

Table S6: Functional annotations based on the unique genes in each of the top 2,000 gene lists.

| GO terms or KEGG pathways |
|----------------------------|
| **PB-transformed t-test**   |
| 1  GO:0007059 chromosome segregation |
| 2  GO:0007155 cell adhesion |
| 3  hsa04670 Leukocyte transendothelial migration |
4 hsa04914: Progesterone-mediated oocyte maturation
5 hsa04923: Regulation of lipolysis in adipocytes
6 GO:0030198: extracellular matrix organization
7 GO:0007165: signal transduction
8 hsa04222: cGMP-PKG signaling pathway
9 GO:0018105: peptidyl-serine phosphorylation
10 hsa04666: Fc gamma R-mediated phagocytosis
11 GO:0007204: positive regulation of cytosolic calcium ion concentration
12 hsa04015: Rap1 signaling pathway
13 hsa04810: Regulation of actin cytoskeleton
14 hsa04664: Fc epsilon RI signaling pathway
15 hsa04370: VEGF signaling pathway
16 hsa04510: Focal adhesion
17 hsa05230: Central carbon metabolism in cancer
18 GO:0035556: intracellular signal transduction
19 hsa04062: Chemokine signaling pathway
20 GO:0070527: platelet aggregation
21 GO:0008015: blood circulation
22 hsa04917: Prolactin signaling pathway
23 GO:0051056: regulation of small GTPase mediated signal transduction
24 GO:0050920: regulation of chemotaxis
25 hsa04611: Platelet activation
26 GO:0048015: phosphatidylinositol-mediated signaling
27 hsa05214: Glioma
28 GO:0010715: regulation of extracellular matrix disassembly
29 GO:0031100: organ regeneration
30 hsa05132: Salmonella infection
31 GO:0006110: regulation of glycolytic process
32 hsa04071: Sphingolipid signaling pathway
33 GO:0009615: response to virus
34 GO:0010038: response to metal ion
35 hsa04224: cAMP signaling pathway
36 GO:0006260: DNA replication
37 hsa05220: Chronic myeloid leukemia
38 hsa05100: Bacterial invasion of epithelial cells
39 hsa04014: Ras signaling pathway
40 GO:0019064: fusion of virus membrane with host plasma membrane
41 hsa04068: FoxO signaling pathway

**Weighted LMER**

1 GO:1902600: hydrogen ion transmembrane transport
2 hsa04932: Non-alcoholic fatty liver disease (NAFLD)
3 hsa05100: Alzheimer’s disease
4 GO:0006379: mRNA cleavage
5 GO:1990440: positive regulation of transcription from RNA polymerase II promoter in response to endoplasmic reticulum stress
6 hsa04260: Cardiac muscle contraction
7 hsa00190: Oxidative phosphorylation
References

S. Anders & W. Huber (2010). ‘Differential expression analysis for sequence count data’. *Genome biology* 11(10):R106.

K. Andreou, et al. (2012). ‘Regulation of CXCR4 gene expression in breast cancer cells under diverse stress conditions’. *International journal of oncology* 41(6):2253–2259.

Y. Benjamini & Y. Hochberg (1995). ‘Controlling the false discovery rate: a practical and powerful approach to multiple testing’. *Journal of the royal statistical society. Series B (Methodological)* pp. 289–300.

M. Bentires-Alj, et al. (2006). ‘A role for the scaffolding adapter GAB2 in breast cancer’. *Nature medicine* 12(1):114.

T. Cancer Genome Atlas Network (2012). ‘Comprehensive molecular portraits of human breast tumours’. *Nature* 490(7418):61–70.

L. A. Carey, et al. (2006). ‘Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study’. *Jama* 295(21):2492–2502.

M. Cojoc, et al. (2013). ‘Emerging targets in cancer management: role of the CXCL12/CXCR4 axis’. *OncoTargets and therapy* 6:1347.

A. Cuenda & J. J. Sanz-Ezquerro (2017). ‘p38γ and p38δ: From Spectators to Key Physiological Players’. *Trends in biochemical sciences* 42(6):431–442.
C.-B. Ding, et al. (2015). ‘Structure and function of Gab2 and its role in cancer’. *Molecular medicine reports* **12**(3):4007–4014.

M. I. Love, et al. (2014). ‘Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2’. *Genome biology* **15**(12):550.

M. Martini, et al. (2014). ‘PI3K/AKT signaling pathway and cancer: an updated review’. *Annals of medicine* **46**(6):372–383.

A. Mortazavi, et al. (2008). ‘Mapping and quantifying mammalian transcriptomes by RNA-Seq’. *Nature methods* **5**(7):621–628.

A. Müller, et al. (2001). ‘Involvement of chemokine receptors in breast cancer metastasis’. *Nature* **410**(6824):50.

M. E. Murphy (2013). ‘The HSP70 family and cancer’. *Carcinogenesis* **34**(6):1181–1188.

A. A. Onitilo, et al. (2009). ‘Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival’. *Clinical medicine & research* **7**(1-2):4–13.

M. D. Robinson, et al. (2010). ‘edgeR: a Bioconductor package for differential expression analysis of digital gene expression data’. *Bioinformatics* **26**(1):139–140.

M. D. Robinson & G. K. Smyth (2007). ‘Moderated statistical tests for assessing differences in tag abundance’. *Biostatistics* **8**(2):2881–2887.

P. J. Sarvaiya, et al. (2013). ‘Chemokines in tumor progression and metastasis’. *Oncotarget* **4**(12):2171.

D. Sims, et al. (2014). ‘Sequencing depth and coverage: key considerations in genomic analyses’. *Nature Reviews Genetics* **15**(2):121–132.

X. Sun, et al. (2010). ‘CXCL12/CXCR4/CXCR7 chemokine axis and cancer progression’. *Cancer and Metastasis Reviews* **29**(4):709–722.

Y. Sun, et al. (2002). ‘β-Arrestin2 is critically involved in CXCR4-mediated chemotaxis, and this is mediated by its enhancement of p38 MAPK activation’. *Journal of Biological Chemistry* **277**(51):49212–49219.

Y. Sun, et al. (2014). ‘CXCL12-CXCR4 axis promotes the natural selection of breast cancer cell metastasis’. *Tumor Biology* **35**(8):7765–7773.

Z. Wang, et al. (2009). ‘RNA-Seq: a revolutionary tool for transcriptomics’. *Nature reviews genetics* **10**(1):57–63.

C. Xu, et al. (2015). ‘CXCR4 in breast cancer: oncogenic role and therapeutic targeting’. *Drug design, development and therapy* **9**:4953.