HHS Public Access

Author manuscript

*J Am Stat Assoc*. Author manuscript; available in PMC 2019 February 06.

Published in final edited form as:

*J Am Stat Assoc*. 2018 ; 113(524): 1733–1741. doi:10.1080/01621459.2017.1371024.

# A Powerful Bayesian Test for Equality of Means in High Dimensions

**Roger S. Zoh**,

Department of Epidemiology & Biostatistics, Texas A&M University, 1266 TAMU, College Station, TX 77843-1266, USA

**Abhra Sarkar**,

Department of Statistical Science, Duke University, Box 90251, Durham NC 27708-0251, USA

**Raymond J. Carroll**, and

Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA. School of Mathematical Sciences, University of Technology, Sydney, Broadway NSW 2007, Australia

**Bani K. Mallick**

Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA

## Abstract

We develop a Bayes factor based testing procedure for comparing two population means in high dimensional settings. In 'large-p-small-n' settings, Bayes factors based on proper priors require eliciting a large and complex $p \times p$ covariance matrix, whereas Bayes factors based on Jeffrey's prior suffer the same impediment as the classical Hotelling $T^2$ test statistic as they involve inversion of ill-formed sample covariance matrices. To circumvent this limitation, we propose that the Bayes factor be based on lower dimensional random projections of the high dimensional data vectors. We choose the prior under the alternative to maximize the power of the test for a fixed threshold level, yielding a restricted most powerful Bayesian test (RMPBT). The final test statistic is based on the ensemble of Bayes factors corresponding to multiple replications of randomly projected data. We show that the test is unbiased and, under mild conditions, is also locally consistent. We demonstrate the efficacy of the approach through simulated and real data examples.

### Some Key Words

Bayes factor; Random projection; Restricted most powerful Bayesian tests; Testing of hypotheses

Supplementary Material

Supplementary Material presents additional tables from the simulation study and a section presenting the derivation of a Bayes factor based on proper joint Normal-Inv-Wishart prior for the nuisance parameters. We have provided a table comparing power estimates between a deterministic projection based test and a random projection based test. We have also included some additional tables containing the results of a simulation comparing the power of RMPBT to that of a test obtained from a proper prior. A Julia code, implementing our approach, is also available as part of the Supplementary Material.

## 1 Introduction

High dimensional population mean testing is common in many application areas including genomics, where gene-set testing is often of more interest than individual gene tests (Ein-Dor *et al.*, 2006; Subramanian *et al.*, 2005). A natural high dimensional test is based on the distance between the sample mean vectors weighted by the inverse sample covariance matrix, also known as the Mahalanobis distance (Johnson and Wichern, 1992). However, the weight becomes undetermined when the dimension of the population mean vectors is larger than the total sample size minus 2.

To circumvent these limitations, two major approaches have emerged. The first approach is centered around constructing tests that eliminate the need to invert ill-formed covariance matrices. Bai and Saranadasa (1996) replaced the sample covariance matrix by a diagonal covariance matrix, for which the inverse exists. Srivastava (2007) substituted the inverse covariance matrix by its Moore-Penrose inverse, under the assumption that the groups have the same covariances. Wu *et al.* (2006) and Gregory *et al.* (2014) proposed tests based on the pooled squared univariate t-tests, eliminating the need to invert non-positive definitive matrices.

The latter approach centers around transforming the data, instead of the test statistics, so that existing tests could be applied to the transformed data. Random projection (RP) is one such method that works by projecting high dimensional data into lower dimensions while only slightly distorting the distances between the original vectors. See, for example, Dasgupta and Gupta (2003). RP has become a popular tool used extensively in machine learning literature where texts documents, imaging and MRI data are often high dimensional. Dasgupta (2000), for example, used RP to uncover the components of high dimensional mixtures of Gaussians. Fern and Brodley (2003) showed the improvement in clustering high dimensional data using RP over other standard approaches. Recently, Guhaniyogi and Dunson (2015) proposed a Bayesian compression regression approach in $n \ll p$ scenarios, where RP is used to reduce the covariate space. RP has entered the frequentist hypothesis testing literature where the $T^2$ statistics are based on the projected version of the data in 'large-p-small-n' setting. See, for example, Lopes *et al.* (2011) and Srivastava *et al.* (2016).

However, to our knowledge, no work has been done to extend Bayesian machineries to hypothesis testing in high dimensional group means testing. Bayesian hypothesis testing differs from its frequentist counterpart in that the decision to reject or accept a null hypothesis is based on the Bayes factor and a chosen evidence threshold (Jeffreys, 1961; Kass and Raftery, 1995). More precisely, the Bayes factor in favor of the alternative hypothesis $H_1$, denoted by $BF_{10}$, is defined as

$$BF_{10} = \frac{m(\mathbf{Y} \mid H_1)}{m(\mathbf{Y} \mid H_0)} = \frac{\int f(\mathbf{Y} \mid \mathbf{\Theta})\, \pi(\mathbf{\Theta} \mid H_1) d\mathbf{\Theta}}{\int f(\mathbf{Y} \mid \mathbf{\Theta})\, \pi(\mathbf{\Theta} \mid H_0) d\mathbf{\Theta}} \quad (1)$$

where $m(\boldsymbol{Y}|H_i)$ denotes the marginal distribution of $\boldsymbol{Y}$ under $H_i$; $\pi(\boldsymbol{\Theta}|H_i)$ denotes the prior distribution of $\boldsymbol{\Theta}$ under $H_i$, for $i = 0, 1$.

Equation (1) involves high dimensional integrals and the choice of $\pi(\boldsymbol{\Theta}|H_i)$ often focuses on distributions that lead to closed form expressions for the Bayes factor.

In this paper, we use the random projection approach to develop a Bayes factor based restricted most powerful Bayesian test (RMPBT) (Goddard and Johnson, 2016) for the high dimensional group means testing problem. In an RMPBT, the prior distribution under the alternative is chosen by maximizing the power of the test with respect to a restricted class of priors. The evidence threshold is selected to match the rejection region of its non-Bayesian counterpart. We show that our proposed test is unbiased and consistent.

The paper is organized as follows. In Section 2, we derive RMPBT for testing differences between two mean vectors. We establish some asymptotic properties of the test in Section 3. Section 4 provides a simulation study investigating the power of the proposed test. We apply the proposed test to the analysis of some real data sets in Section 5. Section 6 concludes with a discussion.

## 2 Bayes factor in high dimensions

Let $\boldsymbol{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ denote a $p$-dimensional multivariate normal density with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_{n_1} \in \mathbb{R}^p$ and $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_{n_2} \in \mathbb{R}^p$ be independent random draws from $\boldsymbol{N}_p(\boldsymbol{\mu}_1,\boldsymbol{\Sigma})$ and $\boldsymbol{N}_p(\boldsymbol{\mu}_2,\boldsymbol{\Sigma})$, respectively. Also, let $\boldsymbol{X}_{n_1\times p} = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_{n_1})^{\mathrm{T}}$ and $\boldsymbol{Y}_{n_2\times p} = (\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_{n_2})^{\mathrm{T}}$.

The minimal sufficient statistics are $\boldsymbol{D} = \bar{\boldsymbol{Y}} - \bar{\boldsymbol{X}}$, $\boldsymbol{A} = (n_1\bar{\boldsymbol{Y}} + n_2\bar{\boldsymbol{X}})/(n_1 + n_2)$ and $\boldsymbol{S} = \{(n_1 - 1)\sum_{i=1}^{n_1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^{\mathrm{T}} + (n_2 - 1)\sum_{i=1}^{n_2}(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})^{\mathrm{T}}\}/(n_1 + n_2 - 2)$. Also, $\boldsymbol{D} \sim \boldsymbol{N}_p(\boldsymbol{\delta}, n_0^{-1}\boldsymbol{\Sigma})$, $\boldsymbol{A} \sim \boldsymbol{N}_p(\boldsymbol{\mu}, n^{-1}\boldsymbol{\Sigma})$ and $(n - 2)\boldsymbol{S} \sim \boldsymbol{W}_p(n - 2,\boldsymbol{\Sigma})$, independently, where $\boldsymbol{\delta} = \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$, $\boldsymbol{\mu} = (n_1\boldsymbol{\mu}_1 + n_2\boldsymbol{\mu}_2)/n$, $n_0^{-1} = n_1^{-1} + n_2^{-1}$, $n = n_1 + n_2$ and $\boldsymbol{W}_p(n,\boldsymbol{\Sigma})$ denotes a Wishart distribution on the space of $p\times p$ dimensional positive definite matrices with degrees of freedom $n$ and mean $n\boldsymbol{\Sigma}$.

The problem is to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ against $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$. We will work with the reparametrization in terms of $\boldsymbol{\delta}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, which parametrize the distribution of the minimal sufficient statistics. The hypotheses of interest can accordingly be reformulated as

$$H_0 : \boldsymbol{\delta} = \boldsymbol{0} \text{ against } H_1 : \boldsymbol{\delta} \neq \boldsymbol{0}.$$

The generic form of the prior that we consider is given by $\pi(\boldsymbol{\delta},\boldsymbol{\mu},\boldsymbol{\Sigma}|H_i) = \pi(\boldsymbol{\mu},\boldsymbol{\Sigma})\ \pi(\boldsymbol{\delta}|\boldsymbol{\Sigma},H_i)$. For $\pi(\boldsymbol{\mu},\boldsymbol{\Sigma})$, we consider the Jeffrey's prior given by

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(p+1)/2}. \quad (2)$$

The choice $\pi(\boldsymbol{\delta}|\boldsymbol{\Sigma}, H_0) = 1\{\boldsymbol{\delta} = \mathbf{0}\}$ is trivially dictated by $H_0$. For $\pi(\boldsymbol{\delta}|\boldsymbol{\Sigma}, H_1)$, we consider the prior

$$\pi(\boldsymbol{\delta} \mid \boldsymbol{\Sigma}, \tau_0, H_1) \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}/\tau_0). \quad (3)$$

The choice of the hyper-parameter $\tau_0 \in (0, \infty)$ is crucial and is discussed in Section 2.1. Throughout the paper, we assume the same prior weight for each hypothesis, that is, $P(H_0) = P(H_1) = 0.5$.

When $1 < p < n-2$, the Bayes factor admits a closed form expression under the assumed priors, as shown by the following result.

**Lemma 1—** *With* $1 < p < n-2$ *and under the priors (2) and (3), we have*

$$BF_{10}(\boldsymbol{X}, \boldsymbol{Y}) = (1+\eta)^{-p/2} \left\{ \frac{1 + \frac{pf}{(1+\eta)(n-p-1)}}{1 + \frac{pf}{(n-p-1)}} \right\}^{-(n-1)/2}, \quad (4)$$

*where* $\eta = n_0/\tau_0$ *and* $f = \frac{(n-p-1)}{(n-2)p} n_0 (\bar{\boldsymbol{Y}} - \bar{\boldsymbol{X}})^{\mathrm{T}} \boldsymbol{S}^{-1} (\bar{\boldsymbol{Y}} - \bar{\boldsymbol{X}})$.

We show the derivation of $BF_{10}$ in Appendix A. Here, $f$ is the scaled Hotelling's $T^2$ statistic with $f \sim F_{p,n-p-1}$ under $H_0$, where $F_{\nu_1,\nu_2}$ denotes a central $F$ distribution with $\nu_1$ and $\nu_2$ degrees of freedom. When $p \geq n-2$, $\boldsymbol{S}$ is no longer positive definite, hence its inverse is not unique and (4) can not be employed to test $H_0$ against $H_1$. To handle the dimensionality problem, we project the data vectors to a lower dimensional subspace using a random projection matrix $\boldsymbol{R}_{p\times m}$ satisfying $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} = \boldsymbol{I}_{m\times m}$ where $1 < m < n-2$. The projected data for group 1 are then obtained as $\boldsymbol{X}_i^{\star} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{X}_i$, $i = 1, \cdots, n_1$. Likewise, the projected data for group 2 are $\boldsymbol{Y}_i^{\star} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Y}_i$, $i = 1, \cdots, n_2$. For a given projection matrix $\boldsymbol{R}_{p\times m}$, under the priors (2) and (3), using Lemma 1 and basic properties of multivariate normal distributions, $BF_{10}$ based on the projected data is given by

$$BF_{10}(\boldsymbol{X}^{\star}, \boldsymbol{Y}^{\star}) = (1+\eta)^{-m/2} \left\{ \frac{1 + \frac{mf^{\star}}{(1+\eta)(n-m-1)}}{1 + \frac{mf^{\star}}{(n-m-1)}} \right\}^{-(n-1)/2}, \quad (5)$$

where $f^{\star} = \frac{(n-m-1)}{(n-2)m} n_0 (\bar{\boldsymbol{Y}} - \bar{\boldsymbol{X}})^{\mathrm{T}} \boldsymbol{R} (\boldsymbol{R}^{\mathrm{T}} \boldsymbol{S} \boldsymbol{R})^{-1} \boldsymbol{R}^{\mathrm{T}} (\bar{\boldsymbol{Y}} - \bar{\boldsymbol{X}})$, $1 < m < n - 2$, $p \geq n - 2$, and $n = n_1 + n_2$. Also, $f^{\star} \sim F_{m,n-m-1}$ under $H_0$. The following result establishes some desirable asymptotic properties of $BF_{10}(\boldsymbol{X}^{\star}, \boldsymbol{Y}^{\star})$ when $\tau_0$ is fixed and also when $\tau_0$ is allowed to depend on $n$.

**Theorem 1—***Let* $n_{\min} = \min\{n_1, n_2\} \to \infty$ *and* $m \to \infty$ *with* $\lim_{n_{\min} \to \infty} m/n = \theta \in (0, 1)$.

**a.** *If* $\tau_0$ *is fixed, under* $H_0$, $\log\{BF_{10}(\boldsymbol{X}^{\star}, \boldsymbol{Y}^{\star})\} \xrightarrow{p} -\infty$, *and, under* $H_1$, $\log\{BF_{10}(\boldsymbol{X}^{\star}, \boldsymbol{Y}^{\star})\} \xrightarrow{p} \infty$.

**b.** *If* $n_0/\tau_0 \to 0$ *and* $mn_0/\tau_0 \to \infty$, *under* $H_0$, $\log\{BF_{10}(\boldsymbol{X}^{\star}, \boldsymbol{Y}^{\star})\} = \mathcal{O}_p(1)$. *For the corresponding sequence of* $H_1^n$, $\log\{BF_{10}(\boldsymbol{X}^{\star}, \boldsymbol{Y}^{\star})\} \xrightarrow{p} \infty$, *with* $f^{\star} \xrightarrow{p} \infty$ *as* $n_{\min} \to \infty$.

The proof of Theorem 1 is deferred to Appendix B. Part (a) of Theorem 1 states that for fixed $\tau_0$, the Bayes factors is consistent under $H_0$ and a fixed alternative $H_1$. However, if $\tau_0$ is allowed to depend on $n$ so that $n_0/\tau_0 \to 0$ at a slower rate than $1/n$, $BF_{10}(\boldsymbol{X}^{\star}, \boldsymbol{Y}^{\star})$ is not consistent under $H_0$, but is consistent for that sequence of local alternatives, provided that the F-statistic $f^{\star}$ is unbounded as $n_{\min} \to \infty$. Although the lack of consistency of the Bayes factors in part (b) of Theorem 1 under $H_0$ seems unsettling at first, this property is similar to that of frequentist tests, where, for a chosen significance level, the null hypothesis has a non-zero probability of being rejected regardless of the sample size when the null is actually true. We show below that the construction of the restricted most powerful Bayesian test satisfies the conditions enumerated above.

## 2.1 Restricted most powerful Bayesian tests

Recently, Johnson (2013b) introduced the idea of the uniformly most powerful Bayesian tests (UMPBTs) in the context of point hypothesis testing, providing a Bayesian parallel to the idea of uniformly most powerful tests (UMPTs) proposed by Neyman and Pearson (1928, 1933). UMPTs are defined as tests with the highest power among all possible tests of a given size. For a fixed size, UMPTs have the rejection region with the highest probability under the alternative. The rejection region refers to the range of values of the test statistics that leads to a rejection of the null.

In Bayesian hypothesis testing, the decision to reject the null hypothesis is based on the Bayes factor or evidence (log Bayes factor) for a given fixed alternative. Johnson (2013b) defines a UMPBT for testing a null against a fixed alternative as the test corresponding to the prior under the alternative that maximizes the probability of deciding in favor of the alternative for a fixed evidence level $\gamma$, among all possible data generating parameters. More precisely, we have a UMPBT for a given evidence threshold $\gamma$ if the Bayes factor in favor of an alternative hypothesis $H_1$ against a fixed null hypothesis $H_0$ satisfies

$$P_{\boldsymbol{\theta}}(BF_{10} > \gamma) \geq P_{\boldsymbol{\theta}}(BF_{20} > \gamma)$$

for all possible values of the data generating parameter $\boldsymbol{\theta}$ and all alternative hypotheses $H_2$. However, as noted in Johnson (2013b), UMPBTs exist in a limited number of relatively simple testing scenarios. Finding a UMPBT in our setting is also a daunting task. Recently, Goddard and Johnson (2016) introduced the idea of restricted most powerful Bayesian tests (RMPBTs). RMPBTs are obtained by restricting the choice of the alternatives to a smaller family of distributions. Here, we focus the search of alternatives to a narrow class of distributions, preferably to priors that lead to Bayes factors with closed forms, like the prior considered in (3). We can subsequently choose the hyper-parameter $\tau_0$ using the idea of RMPBT by maximizing the probability of deciding in favor of the alternative for a fixed evidence level $\gamma$. In other words, we choose $\tau_0 = \tau^\star$ so that

$$P_{\boldsymbol{\theta}}\{BF_{10}(\boldsymbol{X}^\star, \boldsymbol{Y}^\star, \tau^\star) > \gamma\} \geq P_{\boldsymbol{\theta}}\{BF_{10}(\boldsymbol{X}^\star, \boldsymbol{Y}^\star, \tau_0) > \gamma\},$$

for a chosen value of the evidence threshold $\gamma$, all possible values of $\tau_0$ and all data generating model parameters $\boldsymbol{\theta} = (\delta, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. That is, we choose $\tau_0$ so as to maximize the following probability

$$P_{\boldsymbol{\theta}}\left\{\frac{mf^\star}{mf^\star + n - m - 1} > \left(\frac{1 + n_0/\tau_0}{n_0/\tau_0}\right)[1 - \{\gamma(1 + n_0/\tau_0)^{m/2}\}^{-2/(n-1)}]\right\},$$

which is at its maximum when the quantity on the right-hand side of the inequality is at its minimum. The RMPBT is thus obtained with

$$\tau_0 = \arg\min\left(\frac{1 + n_0/\tau_0}{n_0/\tau_0}\right)[1 - \{\gamma(1 + n_0/\tau_0)^{m/2}\}^{-2/(n-1)}]. \quad (6)$$

The optimization in (6) requires a value of $\gamma$ which can be chosen according to the evidence threshold suggested in Kass and Raftery (1995). Alternatively, we can choose $\gamma$ by equating the rejection region of the Bayes factor to that of the classical $F$ statistic. In the non-Bayesian setting, a level $\alpha$ test would reject $H_0$ if $f^\star > F_{\alpha,m,n-m-1}$, where $F_{\alpha,m,n-m-1}$ is the upper $\alpha$ quantile of an $F$ distribution with $m$ and $n - m - 1$ degrees of freedom. The rejection region based on the Bayes factor in favor of the alternative can be expressed as

$$\left\{(\boldsymbol{X}, \boldsymbol{Y}): f^\star > \frac{C_n(n - m - 1)}{m(1 - C_n)}\right\}, \quad (7)$$

where $C_n = \{(1 + n_0/\tau_0)/(n_0/\tau_0)\}[1 - \{\gamma(1 + n_0/\tau_0)^{m/2}\}^{-2/(n-1)}]$. Setting

$$F_{\alpha, m, n-m-1} = \{C_n(n-m-1)\}/(m(1-C_n)), \quad (8)$$

we can then solve for $\gamma_\alpha$. This way, under $H_0$, we have $P\{BF_{10}(\boldsymbol{X}^\star, \boldsymbol{Y}^\star) > \gamma_\alpha\} = P(f^\star > F_{\alpha,m,n-m-1}) = \alpha$. We obtain an exact form for $\tau_0$, denoted $\tau_\alpha(n)$, that satisfies (6) given (8) as

$$\tau_\alpha(n) = n_0 \frac{m}{(n-1)} \frac{(1-C_n)}{(C_n - \frac{m}{n-1})} = \frac{n_0}{F_{\alpha, m, n-m-1} - 1}, \quad (9)$$

where, from (8), $C_n = (mF_{\alpha,m,n-m-1})/(mF_{\alpha,m,n-m-1} + n - m - 1)$.

We can then obtain the equivalent value of $\gamma$, denoted $\gamma_\alpha(n)$, as

$$\gamma_\alpha(n) = \{1 + n_0/\tau_\alpha(n)\}^{-m/2} \left[1 - \frac{n_0/\tau_\alpha(n)}{\{1 + n_0/\tau_\alpha(n)\}} C_n\right]^{-(n-1)/2}. \quad (10)$$

The plot of the evidence threshold $\gamma_\alpha(n)$ along with the associated $\tau_\alpha(n)$ value for various values of $\alpha$ is shown in Figure 1 for two different cases. In both cases, we note that the values of $\gamma_\alpha(n)$ above 20 (strong evidence) are associated with very small significance level $\alpha < 0.007$. This is consistent with the findings that the evidence reflected in Bayes factors often requires very strong evidence in classical settings (Johnson, 2013a).

### 2.2 Choice of *R* and *m*

We discuss the choice of $\boldsymbol{R}$ and $m$ here. We make no attempt to find an optimal projection matrix but are primarily motivated by practical convenience.

Intuitively, however, the projection matrix $\boldsymbol{R}$ should be selected so to only slightly perturb all pairwise distances between the sample vectors (Li *et al.*, 2006). One possible way to achieve this is to sample the entries of $\boldsymbol{R}$ from a distribution with mean zero and variance one. Since our test statistics involves the inversion of $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{R}$, which is positive definite if $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} = \boldsymbol{I}_m$ (see Lemma 1 of Srivastava *et al.*, 2016), we further restrict our choices to the family of semi-orthogonal matrices. We consider two constructions of the projection matrix. The first one, denoted $\boldsymbol{R}_1$, is similar to the one permutation + one random projection considered in Srivastava *et al.* (2016) and yields a sparse matrix with only $p$ non-zero elements. It is constructed as follows.

1. Start with a $p \times m$ matrix of zero in each entry.
2. Simulate $\{r_1, r_2, \cdots, r_p\}$ independently from a standard normal.
3. For each of the $m$ columns, iteratively select $\lfloor p/m \rfloor$ elements from $\boldsymbol{r} = \{r_1, r_2, \cdots, r_p\}$ without replacement and assign them respectively to the positions 1 to $\lfloor p/m \rfloor$

for column 1 vector, $\lfloor p/m \rfloor+1$ to $2\lfloor p/m \rfloor$ for column 2 vector, and elements $(m-1)\lfloor p/m \rfloor$ to $m\lfloor p/m \rfloor$ for column m vector. Finally, assign the remaining elements of $\boldsymbol{r}$, if any, one per column and per row in the remaining rows. Each row of $\boldsymbol{R}_1$ should now have exactly one non-zero element.

4. Randomly permute the row vectors of $\boldsymbol{R}_1$.

5. Finally, standardize the columns vectors so that they have length 1.

The second approach obtains $\boldsymbol{R}_2$ as the Q matrix of the QR decomposition of a $p \times m$ matrix with entries simulated independently from a standard normal distribution.

QR decomposition of a large matrix is computationally intensive.

Note, however, that any matrix $\boldsymbol{U} \in \mathbb{R}^{p\times m}$ admits a QR decomposition $\boldsymbol{U} = \boldsymbol{RB}$, where $\boldsymbol{R} \in \mathbb{R}^{p\times m}$ is an orthonormal matrix, that is, $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} = \boldsymbol{I}_m$ and $\boldsymbol{B} \in \mathbb{R}^{m\times m}$ is an upper triangular matrix with positive entries on the diagonal. This implies $\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{SU})^{-1}\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{RB}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{SRB})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}} = \boldsymbol{R}(\boldsymbol{R}^{\mathrm{T}}\boldsymbol{SR})^{-1}\boldsymbol{R}^{\mathrm{T}}$.

This suggests that we could simply replace $\boldsymbol{R}$ by $\boldsymbol{U}$ in the equation for $f^{\star}$ to speed up the computation.

As one reviewer pointed out, the matrix $\boldsymbol{R}$ could also be obtained from singular value decomposition (SVD) of the sample covariance matrix by ignoring the eigenvectors associated with small eigenvalues. However, since such a construction involves the data, the projected data would have a more complex distribution, adding another layer of complexity to the test. Simulation experiments, where we approximated the null distribution of the test statistic using Monte Carlo simulations, also suggest that this approach does not perform well in practice. The results are summarized in Section S.2 of the Supplementary Material.

Now, we discuss how to choose $m$. Intuitively, small values of $m$ will tend to ignore dependence in the data and the value $m = 1$ completely ignore any correlation. Large values of $m$ close to $n_1 + n_2 - 2$, on the other hand, will lead to tests with low power as the sample covariance matrix is getting closer to a degenerate matrix with small eigenvalues as noted by Bai and Saranadasa (1996). It is expected that the best value of $m$ will tend to depend on the form of the true unknown covariance matrix - while smaller values of $m$ may perform well when the true covariance matrix is diagonal, larger values will be appealing in more complex cases. Here, we present a heuristic approach to obtain $m$ in general settings.

For a significance level $\alpha$ and a random projection matrix $\boldsymbol{R}(m)$, we have $P_m[BF_{10}\{\boldsymbol{X}^{\star}(m),\boldsymbol{Y}^{\star}(m),m\} > \gamma_{\alpha} \mid H_1,\boldsymbol{R}(m)] = P_m\{f^{\star}(m) > F_{\alpha,m,n-m-1} \mid H_1,\boldsymbol{R}(m)\}$. Ideally, we would want to choose a value of $m$ that maximizes this probability and hence optimizes the power. This, however, is a difficult problem since $m$ is also involved in the construction of $f^{\star}$ itself and in its distribution as well. We obtain an approximate solution instead by minimizing the value of the threshold $F_{\alpha,m,n-m-1}$ with respect to $m$. The values of $m$ hence obtained are similar to the values of $m$ obtained by Srivastava *et al.* (2016) using a similar argument. We show in the proof of Theorem 2 in Appendix C that such a choice of $m$ satisfies the assumptions of Theorem 1.

Numerical experiments also suggest the empirical power of the test based on such a choice of $m$ to be very close to the optimal power.

## 3 Test based on Bayes factor and random projections

### 3.1 Single random projection

We derived the Bayes factor in Section 2 after we apply a single random projection $\boldsymbol{R}$ to the data (see Equation 5). Given the sample sizes $n_1$ and $n_2$ and a choice of $\alpha$, we choose $m(n)$, $\tau_\alpha(n)$, and $\gamma_\alpha(n)$ as discussed in Sections 2.1 and 2.2. A test based on the resulting Bayes factor is then obtained as

$$\phi(\boldsymbol{R}) = \begin{cases} 1 & \text{if } BF_{10}(\boldsymbol{X}^\star, \boldsymbol{Y}^\star) > \gamma_\alpha(\eta), \\ 0 & \text{Otherwise,} \end{cases} \quad (11)$$

where $\phi(\boldsymbol{R}) = 1$ signifies rejection of $H_0$ in favor of $H_1$, and $\phi(\boldsymbol{R}) = 0$ accept $H_0$. We make the following observations about the test in (11).

**Theorem 2—***For a given significance level $\alpha \in (0, 1)$, we have*

- **a.** *Under $H_0$, for fixed $n_1$, $n_2$, and $m(n)$, $E\{\phi(\boldsymbol{R}) \mid H_0\} = \alpha$.*
- **b.** *Under $H_1$, for fixed $n_1$, $n_2$, and $m(n)$, $E\{\phi(\boldsymbol{R}) \mid H_1\} \geq \alpha$.*
- **c.** *Let the assumptions in Theorem 1 part (b) be satisfied such that $n_1$, $n_2$, $p \to \infty$, with $m(n)$, $\tau_\alpha(n)$, $\gamma_\alpha(n)$ chosen as described in Section 2.2. Then, $\lim_{n_{\min} \to \infty} E\{\phi(\boldsymbol{R}) \mid H_1^n\} = 1$, where $m(n)/n \to \theta \in (0, 1)$, $n_0 m(n)/\tau_\alpha(n) \to \infty$.*

We show the proof of Theorem 2 in Appendix C. Note that (*a*) shows that the test described in (11) has size $\alpha$; (b) shows that the test is unbiased; finally (c) shows that the power converges to 1 with increasing sample size. In part (c) of Theorem 2, we impose that $m(n)/n \to \theta \in (0, 1)$ and $n_0 m(n)/\tau_\alpha(n) \to \infty$ which are satisfied by our construction suggested in Section 2.2.

### 3.2 Multiple random projections

A test based on a Bayes factor obtained from a single random projection may lead to different decisions for two different random projection matrices. To avoid that, we consider a multitude of Bayes factors computed using many different random projections. Subsequently, we define our test statistic based on the ensemble of Bayes factors and study its power.

Let $\boldsymbol{R}_1, \cdots, \boldsymbol{R}_N$ be a collection of independently and identically distributed random projection matrices. For a choice of $n_1$, $n_2$, and $\alpha$, the values of $m(n)$, $\tau_\alpha(n)$, and $\gamma_\alpha(n)$ are obtained as discussed in Sections 2.1 and 2.2. We then define $\bar{\phi}(N)$ as

$$\bar{\phi}(N) = \frac{1}{N}\sum_{i=1}^{N}\phi(\boldsymbol{R}_i) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{BF_{10}(\boldsymbol{R}_i) > \gamma_\alpha(n)\}, \quad (12)$$

where $\mathbf{1}\{A\} = 1$ if $A$ is true and 0 otherwise. Clearly, $BF_{10}(\boldsymbol{R}_i)$ depends on $\tau_\alpha(n)$. Note that $\bar{\phi}(N)$ represents the proportion of Bayes factors based on projected data that yield Bayes factor larger than the specified evidence threshold $\gamma_\alpha(n)$, for a choice of $\alpha$ and $m(n)$. We then define RMPBT as

$$\begin{cases} \text{Reject } H_0 & \text{if } \bar{\phi}(N) > \phi_\alpha^0, \\ \text{Accept } H_0 & \text{Otherwise}, \end{cases} \quad (13)$$

where $\phi_\alpha^0$ is the upper $\alpha$ quantile of the distribution of $\bar{\phi}(N)$ under $H_0$, which depends on $m(n)$, $n_1$, $n_2$, $p$ and $\alpha$.

**Theorem 3—***Suppose the assumptions of Theorems 1 and 2 hold. Given a collection* $\boldsymbol{R}_1, \cdots, \boldsymbol{R}_N$ *of random projections matrices, where* $\boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i = \boldsymbol{I}$ *for all* $i = 1, \cdots, N$, *then* $\lim_{n_{\min}\to\infty} P\{\bar{\phi}(N) > \phi_\alpha^0\} = 1$ *under the sequence* $H_1^n$ *of alternatives.*

We show the proof of Theorem 3 in Appendix D. For fixed $m(n)$, $p$, $n_1$, $n_2$ and $\alpha$, RMPBT in (13) requires that we first compute $\phi_\alpha^0$. Under $H_0$, $\boldsymbol{\delta} = \mathbf{0}$, but $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown. Fortunately, the asymptotic null distribution of $\bar{\phi}(N)$ is independent of the nuisance parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, providing a simple way of finding $\phi_\alpha^0$. The result is formalized in the following theorem.

**Theorem 4—***Under* $H_0$, *the distribution of* $\bar{\phi}(N)$ *as* $N \to \infty$ *is independent of* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$ *for any fixed* $n_1$, $n_2$, $m \in (1, n_1 + n_2 - 2)$ *and* $p \geq n_1 + n_2 - 2$.

We show the proof of Theorem 4 in Appendix E. Theorem 4 suggests that for large values of $N$ we can approximate the null distribution of $\bar{\phi}(N)$ by simulating data assuming $\boldsymbol{\Sigma} = \boldsymbol{I}$ and $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$.

## 4 Simulation study

We designed a simulation study aimed at investigating the power of the test proposed in (13) with respect to the proportion of true elements of $\boldsymbol{\delta}$ that are actually zero for various choices of covariance matrices and different scenarios or conditions. We consider two simulation settings. In the first, we assume $p = 200$ and $n_1 = n_2 = 50$. Using the approach described in Section 2.2, we find $m = 43$. In the second setting, $p = 1000$, $n_1 = n_2 = 70$ and we get $m = 62$. We denote by $p_0$ the proportion of entries of the vector $\boldsymbol{\delta}$ that are exactly zero. We choose $p_0 = 0.5, .75, .80, 0.95, 0.99, 1.00$. In each setting, the values of $\tau_\alpha$ and $\gamma_\alpha$ are

chosen according to our discussion in Section 2.1. We consider two types of random projections matrices, $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, as described in Section 2.2.

We consider the following choices for $\boldsymbol{\Sigma} = ((\sigma_{ij}))$.

1. $\boldsymbol{\Sigma}_1 = \boldsymbol{I}_{p\times p}$ is the identity matrix.
2. $\boldsymbol{\Sigma}_2$ is a diagonal matrix where the first 20 elements are set at 1 to 20 and the remaining are set exactly 1.
3. $\boldsymbol{\Sigma}_3$ is an AR(1) covariance matrix with $\sigma_{ij} = \sigma^2\rho^{|i-j|}$. We chose $\sigma^2 = 1$ and $\rho = 0.4$.
4. $\boldsymbol{\Sigma}_4$ is block diagonal matrix, with block $\boldsymbol{B} = 0.85\boldsymbol{I}_{25\times 25} + 0.15\boldsymbol{J}_{25\times 25}$. $\boldsymbol{J}$ denotes a matrix with 1 in all of its entries.
5. $\boldsymbol{\Sigma}_5$ is an ARIMA(1,1) covariance matrix with $\sigma_{ij} = \sigma^2\gamma^{1\{|i-j|>0\}}\rho^{|i-j|1\{|i-j|\geq 2\}}$. We chose $\sigma^2 = 1$, $\gamma = 0.5$, $\rho = 0.9$.

We also consider two possible alternatives.

a. We simulate $\boldsymbol{\mu}_2 \sim \boldsymbol{N}_p(\mathbf{1}, \mathbf{I})$, set $p_0$ randomly selected elements to zero, and scale $\boldsymbol{\mu}_2$ so that $(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) = 2$.
b. We simulate $\boldsymbol{\mu}_2 \sim \boldsymbol{N}_p(\mathbf{1}, \mathbf{I})$ and set $p_0$ of its elements to zero and rescale $\boldsymbol{\mu}_2$ so that $\frac{\|\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\|^2}{\sqrt{tr\left(\boldsymbol{\Sigma}^2\right)}} = 0.1$.

$\boldsymbol{\mu}_1$ is chosen to be a vector of zeros. The two alternatives described above were also considered by Srivastava *et al.* (2016). We include in these comparisons the following competitors.

A. The approach of Srivastava *et al.* (2016) referred to as RAPTT.
B. The approach of Bai and Saranadasa (1996) referred to as BS96.
C. The approach of Srivastava and Du (2008) referred to as SD08.
D. The approach of Chen *et al.* (2010) referred to as CQ10.

To estimate the power of our test, we use $N = 5000$ random projections for each of the 1000 independently simulated data sets under each of the alternatives considered. Recall that when $p_0 = 1.0$, the null hypothesis is true and the power represents the type-I error rate estimate. When the true covariance matrix $\boldsymbol{\Sigma}$ is diagonal ($\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$) (see Table 1 and Table S.1 in the Supplemental Material), the tests that do not rely on random projections, namely BS96, SD08 and CQ10, perform slightly better compared to the tests based on random projections, namely RAPTT and the proposed RMPBT. However, RMPBT tended to have higher power when compared to RAPTT for $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. Also, for $\boldsymbol{\Sigma}_3$, an AR(1) covariance matrix with $\rho = 0.4$, RMPBT performed better than all its competitors, especially for sparse alternatives (Table S.2 in the Supplemental Material). Also, the random projection based tests tended to be slightly more conservative when $\Sigma = \boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$, or $\boldsymbol{\Sigma}_3$. For more complex covariance matrices ($\boldsymbol{\Sigma}_4$, $\boldsymbol{\Sigma}_5$), the non-random projection based approaches tended to have lower overall power (Table 2, Table 3), and, in some cases, significantly lower power than

their random projection based competitors, especially when $n_1 = n_2 = 50$. However, random projection based approaches tended to have slightly higher or similar estimated type-I error rates for complex true covariance matrices.

This performance differences can be explained as follows. All three non-projection approaches attempt to get around the issue related to inverting an ill-formed sample covariance matrix. In doing so, BS96 based their test statistic on a quadratic norm of the vector sample means, ignoring any possible weighing for the vectors entries. CQ10 develop their test around the cross-product of the sample vectors, also ignoring possible weighing. SD08 also based their test statistic around the square differences of the group sample means, but used the diagonal sample covariance as weights. Consequently, all three approaches tended to only perform well for diagonal or near diagonal structure covariance matrices. On the other hand, because the random projection based approaches make no assumption about the data covariance matrix, they can use the additional information provided in the dependency relationships to improve power. Both random projection based approaches seem to perform similarly under less sparse alternatives, that is, smaller value of $p_0$.

However, RMPBT tended to have, in most cases, much higher power when compared to RAPTT for cases where the true mean differences are very sparse, especially in the scenario $n_1 = n_2 = 50$ and $p = 200$. Note that both RMPBT and RAPTT depend on the $F$-statistics. We think that the difference observed between the power of RMPBT and RAPTT are not due to the $F$-statistic and the choice of m per se. Instead, the differences found between RMPBT and RAPTT are due to the way each test quantifies the evidence contained in each $F$-statistic computed from each projected copies of a data set. For example, consider two arbitrary values, 20 and 1, for the $F$-statistic with 2 and 3 degrees of freedom. The RMPBT statistic is 0.5, and the equivalent test statistic for the RAPTT is 0.256. Suppose now that the values of the $F$-statistics are 10 and 0.01 instead. The test statistic for the RMPBT remains unchanged, but the test statistic for RAPTT become 0.518. Because in sparse alternatives scenarios and small sample setting, a large number of $F$-statistics are reported large (small p-values), and a smaller number is reported small (large p-values), the test statistic in RAPTT would tend to be affected by these few less significant p-values and causing non-rejection. RMPBT does not suffer this problem since the test statistic only relies on a 0–1 decision. In less sparse alternatives however, this discrepancy between RMPBT and RAPTT is less severe, which is reflected in very similar power reported by both approaches.

The Bayes factor based test proposed in this article assumed a non-informative prior for the nuisance parameters with a single scalar hyper-parameter. Other possibilities include proper priors, like a joint normal-inverse-Wishart prior for the nuisance parameters. Specifying such a prior, however, requires a practitioner to carefully choose high dimensional prior hyper-parameters, including a large $p \times p$ covariance matrix, a very difficult exercise for most applications. We found the power of the resulting test to be highly sensitive to the choice of the covariance matrix hyper-parameter of the normal-inverse-Wishart prior. Results are deferred to Section S.3 of the Supplemental Material.

# 5 Applications

## 5.1 Colon organoids data

Stem cells have some unique regenerative abilities and offer new potential to treating chronic diseases. But stem cells are often modulated by many factors, like the aryl hydrocarbon receptors (AhR), which are not well understood. A study was designed to examine the effect of the AhR on intestinal stem cells. Intestinal crypts were isolated from one mouse, plated, cultivated, separated in 4 sets of 3 plates and each set was treated with one of the following 4 treatments: TCDD only, Indole only, TCDD+Indole, DMSO (control). TCDD is a cancer inducing agent and has the effect of changing the expressions of many genes by activating the AhR. Indole has the role of modulating the effect of TCDD whereas DMSO has an anti-inflammatory property. Finally, RNA is isolated from each of the 12 organoids and sequenced. A gene-by-gene comparison between the TCDD only group versus TCDD+Indole group resulted in only 6 differentially expressed (DE) genes after adjusting for multiple comparisons (McCarthy *et al.*, 2012). We use RMPBT to compare the expression of p = 2000 genes simultaneously between TCDD only and TCDD +Indole groups, with $n_1 = 3$ and $n_2 = 3$. Before we apply the tests, we take a $\log_2$ transformation of the gene expressions (after adding one to avoid issues with $\log_2$ of 0). In this set of 2000 genes, all the 6 DE genes previously found are included. For RMPBT, we use $m = 2$, $N = 100000$ random projections, $\phi_\alpha^0 = 0.054$, $\tau_a = 0.175$, and $\gamma_a = 4.302$. Based on the results in Table 4, RMPBT is the only test that reports significance, when the random projection matrix is $\boldsymbol{R}_1$. This is consistent with the findings in the simulation where RMPBT showed high power for sparse alternative when the projection matrix was $\boldsymbol{R}_1$. The next smallest p-value is also produced by RMPBT with the projection matrix $\boldsymbol{R}_2$.

## 5.2 Breast cancer data

We apply the proposed method to the analysis of a breast cancer data set reported by Gravier, Eleonore *et al.* (2010). The study investigates the involvement of small, invasive ductal carcinoma without lymph (T1T2N0) in predicting metastasis of small node-negative breast carcinoma. Gene expression levels of around 2905 genes were reported for 168 patients over five years. Of the 168, $n_1 = 111$ patients with no event after diagnosis were labelled good and the remaining $n_2 = 57$ with early metastasis were labelled poor. We performed three gene-set comparisons between the two groups, good and poor. The gene-sets were similar to the sets compared in Thulin (2014). The first gene set had $p = 374$ genes located on Chromosome 1. The second set had $p = 233$ genes located on Chromosome 2 and the third set had $p = 191$ genes located on Chromosome 12. A restricted most powerful test 15 is obtained by choosing $\tau_a = 86.880$ and evidence threshold $\gamma_a = 3.852$ with $m = 75$ and $N = 10000$ random projections. The values of $\phi_\alpha^0$, the cutoff values for the test statistic in Chromosome 1, 2 and 12 are 0.127, 0.165, and 0.175 respectively. We also compared both groups using RAPTT, BS96, SD08 and CQ10. The results reported in Table 4 indicate that RMPBT and RAPTT found significance in each gene set, but the non-random projection based approaches failed to find significance for the genes on Chromosome 1.

To investigate the impact of smaller sample sizes on the performance of each test, we compare again the two groups, now by only considering one-third of the total samples ($n_1$ = 37 and $n_2$ = 19). We run the tests on 100 independently sampled data sets from the original data set. For all three chromosomes, the median (over the 100 p-values) p-value for RMPBT ($\boldsymbol{R}_1$), RMPBT ($\boldsymbol{R}_2$), RAPTT ($\boldsymbol{R}_1$), RAPTT ($\boldsymbol{R}_2$) are all highly significant, whereas the p-values for BS96, SD08 and CQ10 are highly insignificant (see Table 4). We see that the approaches based on random projection are able to detect differences between both groups even for relatively smaller sample sizes. However, the other approaches fail to detect difference between both groups for smaller samples sizes.

### 5.3 SRBCT data

We finally apply our approach to the small round blue cell tumors (SRBCTs) data set which is available at http://www.biolab.si/supp/bi-cancer/projections/info/SRBCT.htm. SRBCT's are comprised of 4 different childhood tumors. In this exercise, we would like to test for equality of expression mean of the genes between the neuroblastoma(NB) and the Burkitt's lymphoma(BL) tumors group. The data contain $p$ = 2308 gene expression for both NB and BL tumors with sample sizes 11 and 18 respectively. We estimate $\phi_\alpha^0 = 0.0604$, $\tau_a = 4.836$, $\gamma_a = 3.720$, and $N = 10000$. We report the p-values for each of these tests (see Table 4 SRBCT part). All the tests rejected the null hypothesis with high significance.

## 6 Conclusion

In this article, we proposed a Bayes factor based test for differences between group means in high dimensions. We transformed the data points to lower dimensional spaces using random projections. Using the transformed data, we obtained a closed form Bayes factor by carefully choosing the priors for the model parameters, involving a single scalar hyper-parameter. The hyper-parameter was chosen to obtain a restricted most powerful Bayesian test (RMPBT). Our final test was based on an ensemble of Bayes factors obtained from multiple projected copies. We showed unbiasedness and consistency of the proposed test under mild conditions. We illustrated the efficacy of the test in real and simulated examples.

An ongoing extension of the proposed test also considers non-local priors of Johnson and Rossell (2010) for the distribution of $\boldsymbol{\delta}$ under the alternative and relaxes the assumption of equal covariance matrices between the two groups.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Carroll and Mallick's research was supported by grant U01-CA057030 and by grant R01-CA194391, both from the National Cancer Institute. We are grateful to Dr. Robert Chapkin for sharing the organoids data with us. We also thank the Associate Editor and the anonymous referees for their comments and suggestions which helped greatly improve the paper's presentation.

## References


Bai ZD, Saranadasa H. 1996; Effect of high dimension: by an example of a two sample problem. Statistica Sinica. 6:311–329.

Chen SX, Qin YL, et al. 2010; A two-sample test for high-dimensional data with applications to gene-set testing. The Annals of Statistics. 38:808–835.

Dasgupta, S. Experiments with random projection. Uncertainty in Artificial Intelligence: Proceedings of the Sixteenth Conference (UAI-2000); Morgan Kaufmann; 2000. 143–151.

Dasgupta S, Gupta A. 2003; An elementary proof of a theorem of Johnson and Lindenstrauss. Random Structures and Algorithms. 22:60–65.

Ein-Dor L, Zuk O, Domany E. 2006; Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. Proceedings of the National Academy of Sciences. 103:5923–5928.

Fern, XZ; Brodley, CE. Random projection for high dimensional data clustering: A cluster ensemble approach. Proceedings of the 20th International Conference on Machine Learning; 2003. 186–193.

Goddard SD, Johnson VE. 2016; Restricted most powerful Bayesian tests for linear models. Scandinavian Journal of Statistics. 43(4):1162–1177.

Gravier, Eleonore; Pierron, G; Vincent-Salomon, A; gruel, N; Raynal, V; Savignoni, A; De Rycke, Y; Pierga, JY; Lucchesi, C; Reyal, F; Fourquet, A; Roman-Roman, S; Radvanyi, F; Sastre-Garau, X; Asselain, B; Delattre, O. 2010; A prognostic DNA signature for T1T2 node-negative breast cancer patients. Genes, Chromosomes and Cancer. 49:1125–1125. [PubMed: 20842727]

Gregory KB, Carroll RJ, Baladandayuthapani V, Lahiri SN. 2014; A two-sample test for equality of means in high dimension. Journal of the American Statistical Association. 110:837–849.

Guhaniyogi R, Dunson DB. 2015; Bayesian compressed regression. Journal of the American Statistical Association. 110(512):1500–1514.

Jeffreys, H. Theory of Probability. 3. Oxford University Press; New York, NY: 1961.

Johnson, RA, Wichern, DW. Applied Multivariate Statistical Analysis. Vol. 4. Prentice Hall; Englewood Cliffs, NJ: 1992.

Johnson VE. 2005; Bayes factors based on test statistics. Journal of the Royal Statistical Society: Series B. 67:689–701.

Johnson VE. 2013a; Revised standards for statistical evidence. Proceedings of the National Academy of Sciences. 110:19313–19317.

Johnson VE. 2013b; Uniformly most powerful Bayesian tests. Annals of Statistics. 41:1716. [PubMed: 24659829]

Johnson VE, Rossell D. 2010; On the use of non-local prior densities in Bayesian hypothesis tests. Journal of the Royal Statistical Society: Series B. 72:143–170.

Kass RE, Raftery AE. 1995; Bayes factors. Journal of the American Statistical Association. 90:773–795.

Li, P; Hastie, TJ; Church, KW. Very sparse random projections. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining; 2006. 287–296.

Lopes M, Jacob L, Wainwright MJ. 2011A more powerful two-sample test in high dimensions using random projection. Advances in Neural Information Processing Systems. :1206–1214.

McCarthy D, Chen Y, Smyth G. 2012; Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. Nucleic Acids Research. 40(10):4288–4297. [PubMed: 22287627]

Neyman J, Pearson E. 1933; On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. 231:289–337.

Neyman J, Pearson ES. 1928; On the use and interpretation of certain test criteria for purposes of statistical inference: Part ii. Biometrika. 20A:263–294.

Srivastava MS. 2007; Multivariate theory for analyzing high dimensional data. Journal of the Japan Statistical Society. 37:53–86.

Srivastava MS, Du M. 2008; A test for the mean vector with fewer observations than the dimension. Journal of Multivariate Analysis. 99:386–402.

Srivastava R, Li P, Ruppert D. 2016; RAPTT: An exact two-sample test in high dimensions using random projections. Journal of Computational and Graphical Statistics. 25(3):954–970.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005; Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 102(43):15545–15550.

Thulin M. 2014; A high-dimensional two-sample test for the mean using random subspaces. Computational Statistics & Data Analysis. 74:26–38.

Wu Y, Genton MG, Stefanski LA. 2006; A multivariate two-sample mean test for small sample size and missing data. Biometrics. 62:877–885. [PubMed: 16984331]

# Appendix

## Appendix A Lemma 1 - Derivation of the Bayes factor

We recall that the minimal sufficient statistics are $\boldsymbol{D} = \bar{\boldsymbol{Y}} - \bar{\boldsymbol{X}}$, $\boldsymbol{A} = (n_1\boldsymbol{Y} + n_2\boldsymbol{X})/(n_1+n_2)$ and $\boldsymbol{S} = \{(n_1-1)\sum_{i=1}^{n_1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^{\mathrm{T}} + (n_2-1)\sum_{i=1}^{n_2}(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})^{\mathrm{T}}\}/(n_1+n_2-2)$. Also, $\boldsymbol{D} \sim \boldsymbol{N}_p(\boldsymbol{\delta}, n_0^{-1}\boldsymbol{\Sigma})$, $\boldsymbol{A} \sim \boldsymbol{N}_p(\boldsymbol{\mu}, n^{-1}\boldsymbol{\Sigma})$ and $(n-2)\boldsymbol{S} \sim \boldsymbol{W}_p(n-2, \boldsymbol{\Sigma})$, independently, where $\boldsymbol{\delta} = \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$, $\boldsymbol{\mu} = (n_1\boldsymbol{\mu}_1 + n_2\boldsymbol{\mu}_2)/n$, $n_0^{-1} = n_1^{-1} + n_2^{-1}$, $n = n_1 + n_2$. Under our assumed framework, the joint distribution of the data in terms of the minimal sufficient statistics $\boldsymbol{D}$, $\boldsymbol{A}$ and $\boldsymbol{S}$ is given by

$$P(Data \mid \boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Sigma}) = \boldsymbol{N}_p(\boldsymbol{D} \mid \boldsymbol{\delta}, n_0^{-1}\boldsymbol{\Sigma})\, \boldsymbol{N}_p(\boldsymbol{A} \mid \boldsymbol{\mu}, n^{-1}\boldsymbol{\Sigma})\, \boldsymbol{W}_p\{(n-2)\boldsymbol{S} \mid n-2, \boldsymbol{\Sigma}\}.$$

We assume the following joint prior for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(p+1)/2}$. Under $H_1$, we choose the following conditional prior for $\boldsymbol{\delta}$ as $\boldsymbol{\delta} \mid \boldsymbol{\Sigma} \sim \boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}/\tau_0)$. Under $H_1$, we have

$$P(\boldsymbol{D} \mid \boldsymbol{\Sigma}) = \int \boldsymbol{N}_p(\boldsymbol{D} \mid \boldsymbol{\delta}, n_0^{-1}\boldsymbol{\Sigma})\boldsymbol{N}_p(\boldsymbol{\delta} \mid \boldsymbol{0}, \boldsymbol{\Sigma}/\tau_0)d\boldsymbol{\delta} = \boldsymbol{N}_p(\boldsymbol{D} \mid \boldsymbol{0}, n_\tau^{-1}\boldsymbol{\Sigma}),$$

where $n_\tau^{-1} = \tau_0^{-1} + n_0^{-1}$. If we denote the marginal distribution of the data under $H_1$ by $m_1(Data)$, then we have

$$\begin{aligned} m_1(Data) &\propto \int \boldsymbol{N}_p(\boldsymbol{D} \mid \boldsymbol{0}, n_\tau^{-1}\boldsymbol{\Sigma})\boldsymbol{N}_p(\boldsymbol{A} \mid \boldsymbol{\mu}, n^{-1}\boldsymbol{\Sigma})\boldsymbol{W}_p\{(n-2)\boldsymbol{S} \mid n-2, \boldsymbol{\Sigma}\} \mid \boldsymbol{\Sigma} \mid^{-(p+1)/2} d\boldsymbol{\mu} d\boldsymbol{\Sigma} \\ &\propto \int \boldsymbol{N}_p(\boldsymbol{D} \mid \boldsymbol{0}, n_\tau^{-1}\boldsymbol{\Sigma})\boldsymbol{W}_p\{(n-2)\boldsymbol{S} \mid n-2, \boldsymbol{\Sigma}\} \mid \boldsymbol{\Sigma} \mid^{-(p+1)/2} d\boldsymbol{\Sigma}, \end{aligned}$$

where the second line is obtained by taking the integral with respect to $\boldsymbol{\mu}$. This integral only involves $\boldsymbol{N}_p(\boldsymbol{\mu} \mid \boldsymbol{A}, n^{-1}\boldsymbol{\Sigma})$ and evaluates to 1. With a little bit of algebra, we get that

$$m_1(Data) = A_0 n_\tau^{p/2}\left(1+\frac{n_\tau}{n-2}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{D}\right)^{-(n-1)/2},$$

where $A_0$ represents a quantity independent of the data. Similarly, under $H_0$, $\boldsymbol{\delta}=\mathbf{0}$ and we can show that the marginal distribution of the data, denoted $m_0(Data)$, is

$$m_0(Data) = A_0 n_0^{p/2}\left(1+\frac{n_0}{n-2}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{D}\right)^{-(n-1)/2}.$$

Therefore, the Bayes factor in favor of the alternative is

$$BF_{10} = \left(\frac{n_\tau}{n_0}\right)^{p/2}\left(\frac{1+\frac{n_\tau}{n-2}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{D}}{1+\frac{n_0}{n-2}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{D}}\right)^{-(n-1)/2}.$$

Setting $\eta = n_0/\tau_0$ and $f = \frac{(n-p-1)}{(n-2)p} n_0(\bar{\boldsymbol{Y}}-\bar{\boldsymbol{X}})^{\mathrm{T}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{Y}}-\bar{\boldsymbol{X}})$, we get Equation (4).

## Appendix B Proof of Theorem 1

**Part(a)** For $1 < m < n-2$, we integrate out the parameters with respect to the conjugate priors to obtain the Bayes factor in favor of the alternative as

$$BF_{10}(\boldsymbol{X}^\star, \boldsymbol{Y}^\star) = (1+\eta)^{-m/2}\left\{1-\frac{mf^\star\eta/(1+\eta)}{mf^\star+n-m-1}\right\}^{-(n-1)/2},$$

where

$$f^\star = \frac{n-m-1}{(n-2)m} n_0(\bar{\boldsymbol{Y}}-\bar{\boldsymbol{X}})^{\mathrm{T}}\boldsymbol{R}(\boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^{\mathrm{T}})^{-1}\boldsymbol{R}^{\mathrm{T}}(\bar{\boldsymbol{Y}}-\bar{\boldsymbol{X}}).$$

Recall that $n = n_1+n_2$, $1/n_0 = 1/n_1+1/n_2$, $\eta = n_0/\tau_0$, and $n_{\min} = \min\{n_1, n_2\}$. Since $\tau_0$ is fixed, $\eta \to \infty$ as $n_{\min} \to \infty$. For a randomly chosen $\boldsymbol{R}$, under $H_0$, $f^\star \sim F_{m,n-m-1}$ with $m$ and $n-m-1$ degrees of freedom. Thus, $f^\star = O_p(1)$. Also, from well-known properties of the F distribution, we have that

$$U = \frac{mf^\star/(n-m-1)}{\{mf^\star/(n-m-1)+1\}} = \frac{mf^\star}{(mf^\star+n-m-1)} \sim \mathrm{Beta}\{m/2, (n-m-1)/2\},$$

where Beta($a$, $b$) denotes a Beta distribution. Therefore, $\{\eta/(1+\eta)\}U = O_p(1)$. Hence, $\log\{1-\eta U/(1+\eta)\} = O_p(1)$ as $n_{\min} \to \infty$. We then get

$$-\frac{m}{2n}\log(1+\eta)-\frac{(n-1)}{2n}\log\left\{1-\eta U/(1+\eta)\right\}\xrightarrow{p}-\infty,$$

since $\log(1+\eta)\to\infty$ as $n_{\min}\to\infty$ and $\lim_{n_{\min}\to\infty} m/n=\theta\in(0,1)$. We conclude that $\log\left\{BF_{10}(\boldsymbol{R})\right\}\xrightarrow{p}-\infty$ under the null hypothesis.

Under the alternative, $\boldsymbol{\mu}_1\neq\boldsymbol{\mu}_2$ and $\boldsymbol{\delta}\sim N_p(\mathbf{0},\boldsymbol{\Sigma}/\tau_0)$. Then, $f^\star\mid\lambda\sim F_{m,n-m-1}(\lambda)$ with non-centrality $\lambda=n_0\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{R}(\boldsymbol{R}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{R})^{-1}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{\delta}$. Since $\boldsymbol{\delta}\sim N_p(\mathbf{0},\boldsymbol{\Sigma}/\tau_0)$, $\lambda\sim n_0\chi^2_m/\tau_0$, where $\chi^2_m$ denotes a $\chi^2$ distribution with m degrees of freedom. The non-centrality parameter depends on $n$ through $n_0$. We can show that the unconditional distribution of $f^\star/(1+\eta)\sim F_{m,n-m-1}$ (see Johnson, 2005, page 704). If we denote $f^0=f^\star/(1+\eta)$, we have $f^0=O_p(1)$, and $mf^0/n=O_p(1)$, as $n_{\min}\to\infty$. We have that

$$\frac{\eta U}{(1+\eta)}=\frac{mf^0\eta}{mf^0(1+\eta)+n-m-1}$$

From the above equation, we get

$$-\log\left\{1-\frac{\eta U}{(1+\eta)}\right\}=\log\left\{\frac{mf^0(1+\eta)/(n-m-1)+1}{\{mf^0/(n-m-1)\}+1}\right\}.$$

Since $f^0=O_p(1)$, and $m/(n-m-1)$ converges, we have

$$-\log\left\{1-\frac{\eta U}{(1+\eta)}\right\}\xrightarrow{p}\infty.$$

We conclude that $\log\{BF_{10}(\boldsymbol{R})\}\xrightarrow{p}\infty$, under the alternative hypothesis.

**Part(b)** We now assume that $\eta\to 0$ and $m\eta\to\infty$. We have

$$\log\{BF_{10}(\boldsymbol{X}^\star,\boldsymbol{Y}^\star)\}=\frac{n}{2}\left(1-\frac{m}{n}\right)\log(1+\eta)-\frac{n}{2}\log\{1+\eta(1-U)\}+\frac{1}{2}\log\left\{1-\frac{\eta U}{1+\eta}\right\},$$

where $U\sim Beta\{m/2,(n-m-1)/2\}$ under $H_0$. For large $n$, none of the terms with $n$ dominates and their difference converges. The distribution $\log\{BF_{10}(\boldsymbol{X}^\star,\boldsymbol{Y}^\star)\}$ then depends on that of $U$, which is bounded in probability. Therefore, under $H_0$, $\log\{BF_{10}(\boldsymbol{X}^\star,\boldsymbol{Y}^\star)\}=\mathcal{O}_p(1)$.

Under $H_1^n$, again we have

$$\log\{BF_{10}(\boldsymbol{X}^\star,\boldsymbol{Y}^\star)\}=-\frac{m}{2}\log(1+\eta)-\frac{(n-1)}{2}\log\left\{1-\frac{\eta U}{1+\eta}\right\},$$

where $U \xrightarrow{p} 1$ with $f^* \xrightarrow{p} \infty$. Since $\log(1+\eta)\{(n-1)/2 - m/2\} \to \infty$, we conclude that $\log\{BF_{10}(\boldsymbol{X}^\star, \boldsymbol{Y}^\star)\} \xrightarrow{p} \infty$.

## Appendix C Proof of Theorem 2

In this proof, we denote $m(n)$, $\tau_\alpha(n)$ and $\gamma_\alpha(n)$ simply as $m_n$, $\tau_n$ and $\gamma_n$, respectively. That the projection matrix $\boldsymbol{R}$, the Bayes factor, the variable $f^\star$, their distributions etc. all depend on $m_n$ is implicitly understood and hence $m_n$ is suppressed in their notation.

**Part(a)** Under $H_0$, and some chosen values of $\alpha$ and $m_n$, we have

$$E\{\phi(\boldsymbol{R})\} = E_{\boldsymbol{R}}\left[P_{\boldsymbol{X},\boldsymbol{Y}}\{BF_{10}(\boldsymbol{XR},\boldsymbol{YR}) > \gamma_n \mid \boldsymbol{R}\}\right] = E_{\boldsymbol{R}}(\alpha) = \alpha,$$

making use of the results in Section 2.1 and noting that $P_{\boldsymbol{X},\boldsymbol{Y}}\{BF_{10}(\boldsymbol{XR},\boldsymbol{YR}) > \gamma_n\}$ is computed over the data generating model under the null hypothesis.

**Part(b)** Under the alternative hypothesis, $\boldsymbol{\delta} = \boldsymbol{\delta}_1 \neq \boldsymbol{0}$, and $f^\star \mid \lambda \sim F_{m_n, n-m_n-1}(\lambda)$ distribution, with degrees of freedoms $m_n$ and $n - m_n - 1$ and non-centrality parameter $\lambda = n_0 \boldsymbol{\delta}_1^{\mathrm{T}} \boldsymbol{R}(\boldsymbol{R}^{\mathrm{T}} \sum \boldsymbol{R})^{-1} \boldsymbol{R}^{\mathrm{T}} \boldsymbol{\delta}_1$. Hence, under the alternative hypothesis $P\{BF_{10}(\boldsymbol{XR},\boldsymbol{YR}) > \gamma_\alpha\} = P(f^\star > F_{\alpha, m_n, n-m_n-1}) \geq \alpha$, where $F_{\alpha, m_n, n-m_n-1}$ is the $\alpha$ upper quantile of a $F_{m_n, n-m_n-1}$ distribution, as seen in Section 2.1. Marginalizing over $\boldsymbol{\delta}_1$ under the alternative, $f^\star/(1+\eta_n) \sim F_{m_n, n-m_n-1}$, where $\eta_n = n_0/\tau_n$. Therefore, $\alpha = P\{f^\star/(1+\eta_n) > F_{\alpha, m_n, n-m_n-1}\} = P\{f^\star > (1+\eta_n) F_{\alpha, m_n, n-m_n-1}\} \leq P(f^\star > F_{\alpha, m_n, n-m_n-1})$. We conclude $E_{\boldsymbol{R}}\{\phi(\boldsymbol{R}) \mid H_1\} \geq \alpha$.

**Part(c)** First, we show that our construction of $m_n$ satisfies $m_n/n \to \theta \in (0, 1)$. For chosen $n$, $F_{\alpha, m, n-m-1}$ is a convex function over the range of possible values of $m$, suggesting that $m_n$ and $n - m_n$ both diverge. See Figure 3.

The mean $\mu_n$ and variance $\sigma_n^2$ of an $F_{m_n, n-m_n-1}$ distribution are given by $\mu_n = \dfrac{n - m_n - 1}{n - m_n - 3} > 1$ and $\sigma_n^2 = \dfrac{2(n - m_n - 1)^2 (n - 3)}{m_n (n - m_n - 5)(n - m_n - 3)^2}$. For large $m_n$ and $n - m_n$, we have $F_{\alpha, m_n, n-m_n-1} \approx \mu_n + \Sigma_n \Phi^{-1}(\alpha)$, where $\Phi^{-1}(\alpha)$ is the upper $\alpha$-quantile of the standard normal distribution. So $F_{\alpha, m_n, n-m_n-1}$ is at its minimum when $\Sigma_n$ is at its minimum, which happens when $m_n \approx \dfrac{n-5}{2}$. Hence, $m_n/n \to \theta = 1/2 \in (0, 1)$ as $n_{\min} \to \infty$. Second, we show that $n\eta_n \to \infty$ as $n_{\min} \to \infty$. From equation (8), we have

$$C_n = \frac{m_n F_{\alpha, m_n, n-m_n-1}}{m_n F_{\alpha, m_n, n-m_n-1} + n - m_n - 1}. \quad \text{(A.1)}$$

We have

$$n\eta_n = \frac{nn_0}{\tau_n} = \frac{n(n-1)\{C_n - m_n/(n-1)\}}{m_n(1-C_n)} = n(n-1)\left\{\frac{C_n}{m_n(1-C_n)} - \frac{1}{(n-1)(1-C_n)}\right\}$$
$$= n(n-1)\left\{\frac{F_{\alpha, m_n, n-m_n-1}}{(n-m_n-1)} - \frac{m_n F_{\alpha, m_n, n-m_n-1}{}^{+n-m_n-1}}{(n-1)(n-m_n-1)}\right\}$$
$$= \frac{n}{(n-m_n-1)}\{(n-m_n-1)F_{\alpha, m_n, n-m_n-1} - (n-m_n-1)\}$$
$$= n(F_{\alpha, m_n, n-m_n-1} - 1).$$

The variance of a central $F$ distribution with $m_n$ and $n-m_n-1$ degrees of freedom is $2(n-m_n-1)^2(n-3)/\{m_n(n-m_n-3)^2(n-m_n-5)\} = O(1/m_n)$. The convergence of $F_{\alpha,m_n,n-m_n-1}$ is thus slower than $1/\sqrt{n}$. We conclude that $n\eta_n$ and hence $m_n\eta_n \to \infty$. For a chosen $\alpha$ and the sequence of alternatives $H_1^n$, we have

$$\lim_{n_{\min}\to\infty} E_{\boldsymbol{R}}\phi(\boldsymbol{R} \mid H_1^n) = E_{\boldsymbol{R}}\left[\lim_{n_{\min}\to\infty} P_{\boldsymbol{X},\boldsymbol{Y}}\{BF_{10}(\boldsymbol{XR},\boldsymbol{YR}) > \gamma_\alpha(n) \mid \boldsymbol{R}, H_1^n\}\right]$$
$$= E_{\boldsymbol{R}}\left[\lim_{n_{\min}\to\infty} P_{\boldsymbol{X},\boldsymbol{Y}}\{f^\star > F_{\alpha, m_n, n-m_n-1} \mid \boldsymbol{R}, H_1^n\}\right].$$

Since $f^\star \xrightarrow{p} \infty$ and $F_{\alpha,m_n,n-m_n-1} \to 1$, we conclude that $\lim_{n_{\min}\to\infty} E_{\boldsymbol{R}}\phi(\boldsymbol{R} \mid H_1^n) = 1$.

## Appendix D Proof of Theorem 3

The power of RMBPT is $P\{\bar{\phi}(N) > \phi_\alpha^0 \mid H_1^n\}$. Henceforth, we make it explicit that $\phi_\alpha^0$ depends on $(n_1, n_2)$ and write $\phi_\alpha^0(n_1, n_2)$ instead.

For given $n_1$, $n_2$ and $\alpha$, we choose $\phi_\alpha^0(n_1, n_2)$ so that $P\{\bar{\phi}(N) > \phi_\alpha^0(n_1, n_2) \mid H_0\} = \alpha$. Since $0 < \phi_\alpha^0(n_1, n_2) < 1$, for $0 < \alpha < 1$, we have that $P\{\sum_{i=1}^N \phi(\boldsymbol{R}_i) \geq 0 \mid H_1^n\} \geq P\{\sum_{i=1}^N \phi(\boldsymbol{R}_i) > N\phi_\alpha^0(n_1, n_2) \mid H_1^n\} \geq P\{\sum_{i=1}^N \phi(\boldsymbol{R}_i) \geq N \mid H_1^n\}$.

We have that $P\{\phi(\boldsymbol{R}_i) = 1 \mid H_1^n\} \to 1$ as $n_{\min} \to \infty$, under the alternative for $i = 1, \cdots, N$. So, $P\{\sum_{i=1}^N \phi(\boldsymbol{R}_i) \geq 0 \mid H_1^n] = 1 - \prod_{i=1}^N P\{\phi(\boldsymbol{R}_i) = 0 \mid H_1^n\} \to 1$. Additionally, $P\{\sum_{i=1}^N \phi(\boldsymbol{R}_i) \geq N \mid H_1^n\} = P\{\sum_{i=1}^N \phi(\boldsymbol{R}_i) = N \mid H_1^n\} = \prod_{i=1}^N P\{\phi(\boldsymbol{R}_i) = 1 \mid H_1^n\} \to 1$ for fixed $N$ as $n_{\min} \to \infty$. We conclude that $P\{\bar{\phi}(N) \geq \phi_\alpha^0 \mid H_1^n\} \to 1$ as $n_{\min} \to \infty$.

## Appendix E Proof of Theorem 4

The proof is similar to that of Theorem 2 of Srivastava *et al.* (2016) except that we do not rely on the cumulative distribution function of the F distribution.

Suppose $\boldsymbol{R}_1, \boldsymbol{R}_2, \cdots, \boldsymbol{R}_N$ is a collection independently sampled projection matrices. Let $\phi_i = \phi(\boldsymbol{R}_i) = \mathbf{1}\{BF_{10}(\boldsymbol{R}_i, \tau_a) > \gamma_a\}$ for $i = 1, 2, \cdots, N$. Recall that $\bar{\phi}(N) = \sum_{i=1}^{N} \phi_i/N$. Evaluating the conditional probability that $\bar{\phi}(N) < x$ over the distribution of the random projection matrices given $\boldsymbol{X}$ and $\boldsymbol{Y}$ and then taking the expectation over the data, we get

$$P\{\bar{\phi}(N) < x\} = E_{\boldsymbol{X},\boldsymbol{Y}}[P_{\boldsymbol{R}}\{\bar{\phi}(N) < x \mid \boldsymbol{X}, \boldsymbol{Y}\}]. \quad \text{(A.2)}$$

We have

$$P_{\boldsymbol{R}}\{\bar{\phi}(N) < x \mid \boldsymbol{X}, \boldsymbol{Y}\} = P_{\boldsymbol{R}}\left\{\frac{\bar{\phi}(N) - E_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X}, \boldsymbol{Y})}{\sqrt{\mathrm{var}_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X}, \boldsymbol{Y})/N}} < \frac{x - E_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X}, \boldsymbol{Y})}{\sqrt{\mathrm{var}_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X}, \boldsymbol{Y})/N}} \mid \boldsymbol{X}, \boldsymbol{Y}\right\},$$

where $E_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X},\boldsymbol{Y})$ and $\mathrm{var}_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X},\boldsymbol{Y})$ are respectively the conditional mean and variance of $\phi_1$. Given $\boldsymbol{X}$ and $\boldsymbol{Y}$, the binary variables $\phi_i$, $i = 1, \cdots, N$, are independent and identically distributed with finite mean and variance. By the Central Limit Theorem,

$$\lim_{N \to \infty}\left[P_{\boldsymbol{R}}\{\bar{\phi}(N) < x \mid \boldsymbol{X}, \boldsymbol{Y}\} - \Phi\left\{\frac{x - E_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X}, \boldsymbol{Y})}{\sqrt{\mathrm{var}_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X}, \boldsymbol{Y})/N}}\right\}\right] = 0 \quad \text{(A.3)}$$

where $\Phi(a)$ is the standard normal cumulative distribution function evaluated at $a$. We need to show that both $E_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X},\boldsymbol{Y})$ and $\mathrm{var}_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X},\boldsymbol{Y})$ have a distribution independent of $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ under $H_0$. This is equivalent to showing that $E_{\boldsymbol{X},\boldsymbol{Y}}\{E_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X},\boldsymbol{Y})\}^r$ is independent of $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$, and $\boldsymbol{\Sigma}$, for $r = 1, 2, \cdots$. From (7), we have

$$E_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X}, \boldsymbol{Y}) = P_{\boldsymbol{R}}\{f(\boldsymbol{R}_1 \mid \boldsymbol{X}, \boldsymbol{Y}) > F_{\alpha, m_n, n - m_n - 1}\},$$

where $f(\boldsymbol{R} \mid \boldsymbol{X}, \boldsymbol{Y}) = \frac{n - m_n - 1}{(n-2)m_n} n_0 (\bar{\boldsymbol{Y}} - \bar{\boldsymbol{X}})^{\mathrm{T}} \boldsymbol{R}(\boldsymbol{R}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{R})^{-1}\boldsymbol{R}^{\mathrm{T}}(\bar{\boldsymbol{Y}} - \bar{\boldsymbol{X}})$. $P_{\boldsymbol{R}}$ is a probability measure associated with the random projection matrices. Thus, the $r^{th}$ moment of $E_{\boldsymbol{R}}\{\phi_1 \mid \boldsymbol{X},\boldsymbol{Y}\}$ is expressed as

$$E_{\boldsymbol{X}\boldsymbol{Y}}\{E_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X}, \boldsymbol{Y})\}^r = \int\left[\int \mathbf{1}\{f(\boldsymbol{R} \mid \boldsymbol{X}, \boldsymbol{Y}) > F_{\alpha, m_n, n - m_n - 1}\} dP_{\boldsymbol{R}}\right]^r dP_{\boldsymbol{X},\boldsymbol{Y}}. \quad \text{(A.4)}$$

Since $0 \le E_{\boldsymbol{R}}(\boldsymbol{\phi}_1 \mid \boldsymbol{X},\boldsymbol{Y}) \le 1$, we have

$$\int\left(\int [\mathbf{1}\{f(\boldsymbol{R} \mid \boldsymbol{X},\boldsymbol{Y}) > F_{\alpha, m_n, n-m_n-1}\}]^r dP_{\boldsymbol{R}}\right) dP_{\boldsymbol{X},\boldsymbol{Y}} = \int\left(\int [\mathbf{1}\{f(\boldsymbol{R} \mid \boldsymbol{X},\boldsymbol{Y}) > F_{\alpha, m_n, n-m_n-1}\}]^r dP_{\boldsymbol{X},\boldsymbol{Y}}\right) dP_{\boldsymbol{R}}, \quad \text{(A.5)}$$

where we can safely interchange the order of integration by Fubini's Theorem. Under $H_0$, $f(\boldsymbol{R}\,/\boldsymbol{X},\boldsymbol{Y}) \sim F_{m_n,n-m_n-1}$ and $P_{\boldsymbol{X},\boldsymbol{Y}}\{f(\boldsymbol{R}\,/\boldsymbol{X},\boldsymbol{Y}) > F_{\alpha,m_n,n-m_n-1}\} = \alpha$. We conclude that (A.5) is independent of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and $\boldsymbol{\Sigma}$ for any positive integer $r$.

Next, from (A.4), we have

$$E_{\boldsymbol{XY}}\{E_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X},\boldsymbol{Y})\}^r = \int\cdots\int\left[\int \prod_{i=1}^{r} \mathbf{1}\{f(\boldsymbol{R}_i \mid \boldsymbol{X},\boldsymbol{Y}) > F_{\alpha, m_n, n-m_n-1}\}\, dP_{\boldsymbol{X},\boldsymbol{Y}}\right]\prod_{i=1}^{r} dP_{\boldsymbol{R}_i}, \quad \text{(A.6)}$$

where we can again safely exchange the order of integration using Fubini's Theorem. In (A. 6), since $\boldsymbol{R}_i$, $i = 1, \cdots, N$, are identically and independently distributed with respect to the probability measure $P_{\boldsymbol{R}}$, we get that

$$\int \prod_{i=1}^{r} \mathbf{1}\{f(\boldsymbol{R}_i \mid \boldsymbol{X},\boldsymbol{Y}) > F_{\alpha, m_n, n-m_n-1}\} dP_{\boldsymbol{X},\boldsymbol{Y}}$$

is also independent of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and $\boldsymbol{\Sigma}$ based on the result obtained in (A.5).

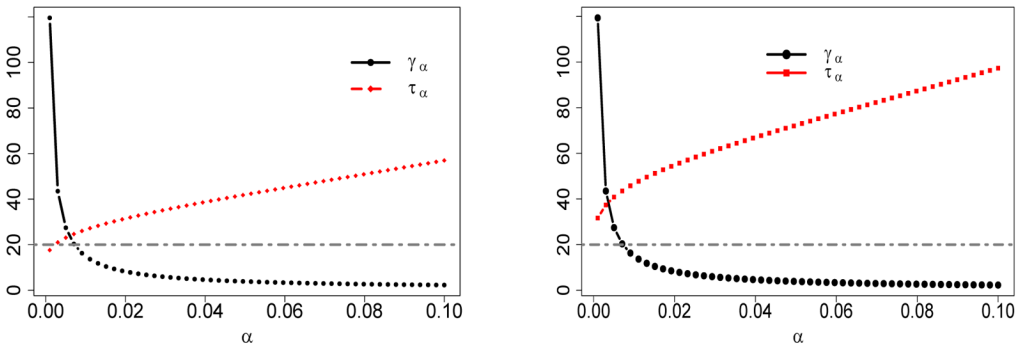
Also, we have

$$\left|P_{\boldsymbol{R}}\{\bar{\phi}(N) < x \mid \boldsymbol{X},\boldsymbol{Y}\} - \Phi\left\{\frac{x - E_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X},\boldsymbol{Y})}{\sqrt{\mathrm{var}_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X},\boldsymbol{Y})/N}}\right\}\right| \le 2. \quad \text{(A.7)}$$

Using (A.2), (A.3), (A.7) and the bounded convergence theorem, we have

$$\lim_{N\to\infty}\left(P\{\bar{\phi}(N) < x\} - E_{\boldsymbol{X},\boldsymbol{Y}}\left[\Phi\left\{\frac{x - E_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X},\boldsymbol{Y})}{\sqrt{\mathrm{var}_{\boldsymbol{R}}(\phi_1 \mid \boldsymbol{X},\boldsymbol{Y})/N}}\right\}\right]\right) = 0 \quad \text{(A.8)}$$

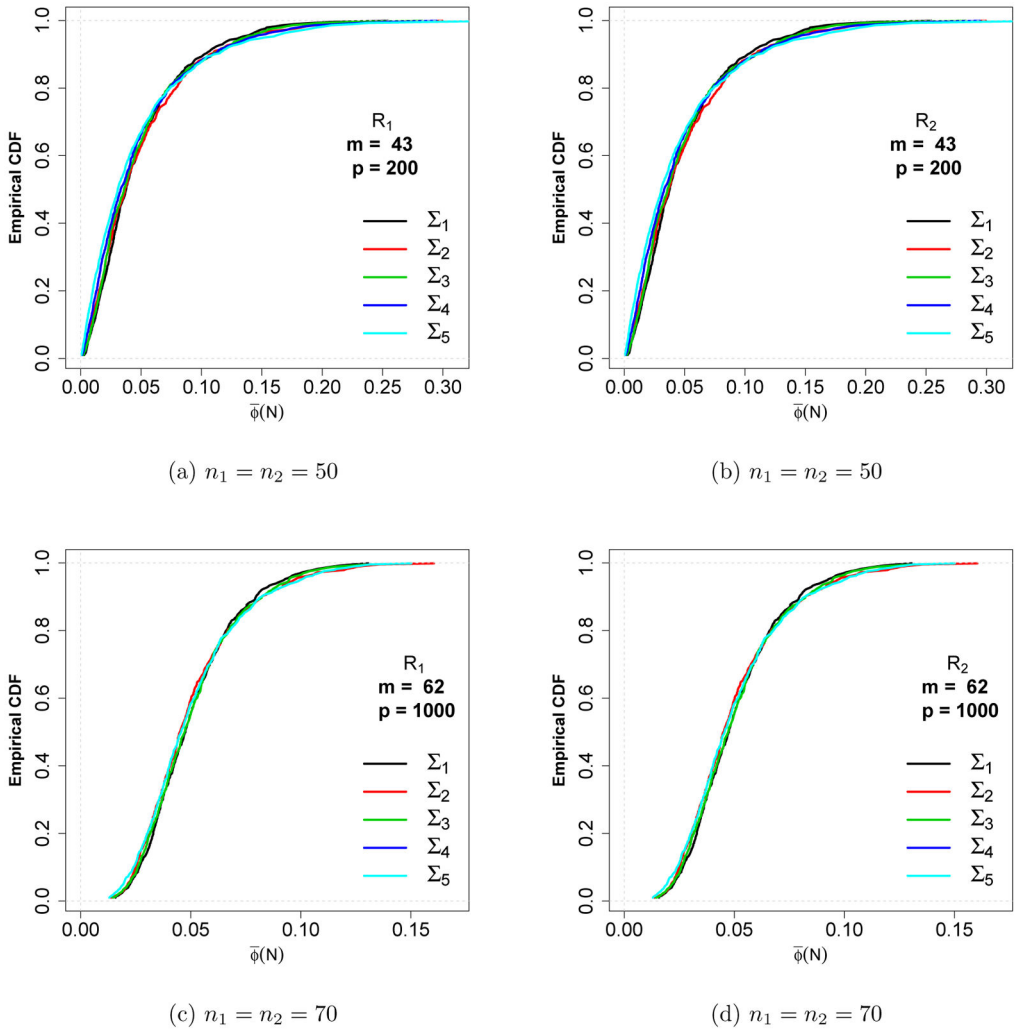
We conclude that for $n_1$, $n_2$, $m_n$, and $\alpha$, the asymptotic distribution of $\bar{\phi}(N)$ as $N \to \infty$ does not depend on the true parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and $\boldsymbol{\Sigma}$ under $H_0$.




(a) $n = n_1 + n_2 = 100$, $m = 43$ (b) $n = n_1 + n_2 = 140$, $m = 62$

**Figure 1.**

Plot of the estimated Bayes factor threshold $\gamma_a(n)$ and the value of $\tau_a(n)$ for various values of $a$, the significance level. The values of Bayes factor above the horizontal line at 20 denotes the ($\gamma_a$, $\tau_a$, $a$) triplet that represents strong evidence against the null hypothesis according to Kass and Raftery (1995).




(a) $n_1 = n_2 = 50$ (b) $n_1 = n_2 = 50$

(c) $n_1 = n_2 = 70$ (d) $n_1 = n_2 = 70$

**Figure 2.**

Empirical distribution function of $\bar{\phi}(N)$ under the null hypothesis for 5 different covariance matrices based on $N = 50000$ random projections and 1000 data sets.

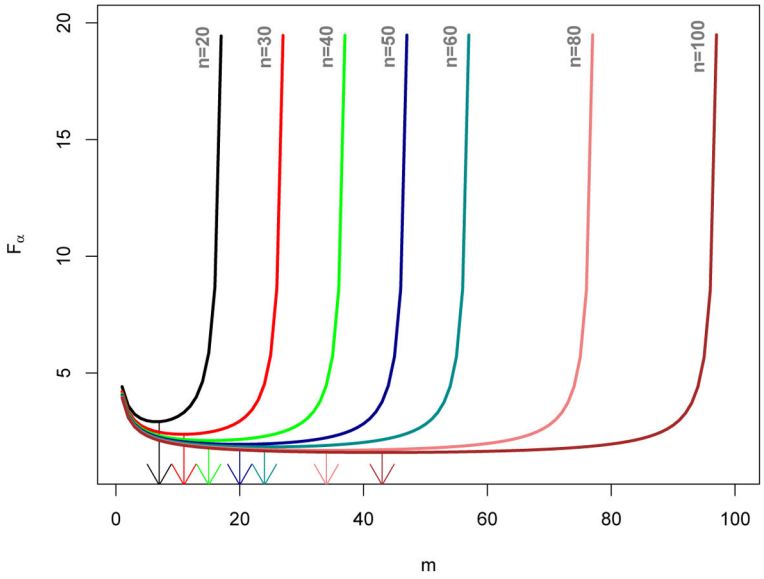



**Figure 3.**

Plot of $F_{\alpha,m,n-m-1}$ against $m = 1, 2, \cdots, n-2$ for different values of $n = n_1 + n_2$. The arrows point to values of $m_n$ obtained by our method for different $n$.

**Table 1**

Power analysis of 5 tests assuming the true covariance matrix is $\mathbf{\Sigma}_1 = \boldsymbol{I}_{p\times p}$. We chose the significant level at $\alpha = 0.05$. For the case $n_1 = n_2 = 50$, $p = 200$, we have $m = 43$, $\tau_\alpha = 41.918$, and $\gamma_\alpha = 3.841$. For the case, $n_1 = n_2 = 70$, $p = 1000$, we have $m = 62$, $\tau_\alpha = 72.318$, and $\gamma_\alpha = 3.850$. RMPBT is our approach. RAPTT is the approach of Srivastava *et al.* (2016). BS96 is the approach of Bai and Saranadasa (1996). SD08 is the approach of Srivastava and Du (2008). CQ10 is the approach of Chen *et al.* (2010).

| | $p_0$ | RMPBT ($R_1$) | RMPBT ($R_2$) | RAPTT ($R_1$) | RAPTT ($R_2$) | BS96 | SD08 | CQ10 | RMPBT ($R_1$) | RMPBT ($R_2$) | RAPTT ($R_1$) | RAPTT ($R_2$) | BS96 | SD08 | CQ10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Alternative 1** | | | | | | | | **Alternative 2** | | | | | | |
| **$n_1 = n_2 = 50$, $p = 200$** | 0.500 | 0.674 | 0.669 | 0.675 | 0.650 | **0.754** | 0.723 | **0.754** | 0.469 | 0.435 | 0.453 | 0.408 | **0.542** | 0.513 | **0.542** |
| | 0.750 | 0.684 | 0.663 | 0.669 | 0.654 | **0.769** | 0.726 | **0.769** | 0.474 | 0.463 | 0.455 | 0.414 | **0.535** | 0.503 | **0.535** |
| | 0.800 | 0.680 | 0.651 | 0.652 | 0.634 | **0.749** | 0.715 | **0.749** | 0.441 | 0.413 | 0.425 | 0.379 | **0.517** | 0.481 | **0.517** |
| | 0.950 | 0.716 | 0.698 | 0.649 | 0.622 | **0.756** | 0.720 | **0.756** | 0.480 | 0.439 | 0.425 | 0.384 | **0.525** | 0.482 | **0.525** |
| | 0.975 | 0.707 | 0.680 | 0.572 | 0.555 | **0.721** | 0.681 | **0.721** | 0.471 | 0.450 | 0.382 | 0.351 | **0.497** | 0.453 | **0.497** |
| | 0.990 | **0.793** | 0.771 | 0.561 | 0.553 | 0.748 | 0.726 | 0.748 | 0.553 | 0.525 | 0.367 | 0.326 | **0.554** | 0.506 | **0.554** |
| | 1.000 | 0.046 | 0.041 | 0.037 | 0.032 | 0.042 | 0.039 | 0.042 | 0.046 | 0.038 | 0.037 | 0.025 | 0.042 | 0.039 | 0.042 |
| **$n_1 = n_2 = 70$, $p = 1000$** | 0.500 | 0.372 | 0.343 | 0.347 | 0.311 | **0.473** | 0.422 | **0.473** | 0.677 | 0.588 | 0.644 | 0.559 | **0.767** | 0.727 | **0.767** |
| | 0.750 | 0.348 | 0.294 | 0.316 | 0.271 | **0.470** | 0.401 | **0.470** | 0.695 | 0.616 | 0.673 | 0.578 | **0.789** | 0.746 | **0.789** |
| | 0.800 | 0.337 | 0.304 | 0.314 | 0.267 | **0.448** | 0.389 | **0.448** | 0.660 | 0.581 | 0.634 | 0.552 | **0.762** | 0.717 | **0.762** |
| | 0.950 | 0.388 | 0.339 | 0.349 | 0.304 | **0.474** | 0.423 | **0.474** | 0.694 | 0.612 | 0.646 | 0.559 | **0.775** | 0.741 | **0.775** |
| | 0.975 | 0.332 | 0.309 | 0.298 | 0.253 | **0.450** | 0.384 | **0.450** | 0.685 | 0.612 | 0.619 | 0.532 | **0.761** | 0.722 | **0.761** |
| | 0.990 | 0.377 | 0.348 | 0.316 | 0.280 | **0.453** | 0.401 | **0.453** | 0.719 | 0.640 | 0.610 | 0.527 | **0.774** | 0.719 | **0.774** |
| | 1.000 | 0.031 | 0.030 | 0.028 | 0.023 | 0.063 | 0.040 | 0.063 | 0.031 | 0.022 | 0.028 | 0.020 | 0.063 | 0.040 | 0.063 |

*J Am Stat Assoc*. Author manuscript; available in PMC 2019 February 06.

**Table 2**

Power analysis of 5 tests assuming the true covariance matrix is $\mathbf{\Sigma}_4$. We chose the significant level at $\alpha = 0.05$. For the case $n_1 = n_2 = 50$, $p = 200$, we have $m = 43$, $\tau_\alpha = 41.918$, and $\gamma_\alpha = 3.841$. For the case, $n_1 = n_2 = 70$, $p = 1000$, we have $m = 62$, $\tau_\alpha = 72.318$, and $\gamma_\alpha = 3.850$. RMPBT is our approach. RAPTT is the approach of Srivastava *et al.* (2016). BS96 is the approach of Bai and Saranadasa (1996). SD08 is the approach of Srivastava and Du (2008). CQ10 is the approach of Chen *et al.* (2010).

| | $p_0$ | RMPBT ($R_1$) | RMPBT ($R_2$) | RAPTT ($R_1$) | RAPTT ($R_2$) | BS96 | SD08 | CQ10 | RMPBT ($R_1$) | RMPBT ($R_2$) | RAPTT ($R_1$) | RAPTT ($R_2$) | BS96 | SD08 | CQ10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Alternative 1 | | | | | | | Alternative 2 | | |
| $n_1 = n_2 = 50, p = 200$ | 0.500 | 0.668 | 0.653 | **0.682** | 0.669 | 0.644 | 0.602 | 0.644 | 0.534 | 0.504 | **0.539** | 0.489 | 0.508 | 0.465 | 0.508 |
| | 0.750 | **0.642** | 0.621 | 0.639 | 0.606 | 0.558 | 0.526 | 0.558 | **0.565** | 0.539 | 0.561 | 0.522 | 0.497 | 0.447 | 0.497 |
| | 0.800 | **0.658** | 0.649 | 0.645 | 0.617 | 0.576 | 0.535 | 0.576 | **0.591** | 0.569 | 0.577 | 0.532 | 0.528 | 0.485 | 0.528 |
| | 0.950 | **0.659** | 0.627 | 0.592 | 0.567 | 0.521 | 0.481 | 0.521 | **0.644** | 0.605 | 0.580 | 0.537 | 0.513 | 0.469 | 0.513 |
| | 0.975 | **0.688** | 0.667 | 0.564 | 0.538 | 0.531 | 0.478 | 0.531 | **0.678** | 0.656 | 0.556 | 0.521 | 0.523 | 0.473 | 0.523 |
| | 0.990 | **0.731** | 0.706 | 0.513 | 0.492 | 0.511 | 0.464 | 0.511 | **0.729** | 0.694 | 0.511 | 0.474 | 0.509 | 0.462 | 0.509 |
| | 1.000 | 0.049 | 0.049 | 0.049 | 0.048 | 0.057 | 0.044 | 0.057 | 0.049 | 0.046 | 0.049 | 0.042 | 0.057 | 0.044 | 0.057 |
| $n_1 = n_2 = 70, p = 1000$ | 0.500 | **0.414** | 0.379 | 0.407 | 0.378 | 0.401 | 0.353 | 0.401 | **0.774** | 0.717 | 0.767 | 0.704 | 0.761 | 0.720 | 0.761 |
| | 0.750 | 0.327 | 0.294 | 0.323 | 0.291 | **0.331** | 0.278 | 0.331 | **0.790** | 0.727 | 0.783 | 0.711 | 0.764 | 0.718 | 0.764 |
| | 0.800 | **0.348** | 0.318 | 0.347 | 0.304 | 0.343 | 0.285 | 0.343 | **0.796** | 0.734 | 0.788 | 0.713 | 0.775 | 0.728 | 0.775 |
| | 0.950 | **0.335** | 0.307 | 0.319 | 0.287 | 0.311 | 0.270 | 0.311 | **0.827** | 0.776 | 0.806 | 0.742 | 0.782 | 0.730 | 0.782 |
| | 0.975 | **0.315** | 0.278 | 0.295 | 0.266 | 0.294 | 0.245 | 0.294 | **0.836** | 0.776 | 0.786 | 0.720 | 0.755 | 0.716 | 0.755 |
| | 0.990 | **0.335** | 0.308 | 0.298 | 0.261 | 0.294 | 0.248 | 0.294 | **0.871** | 0.819 | 0.769 | 0.714 | 0.775 | 0.738 | 0.775 |
| | 1.000 | 0.060 | 0.052 | 0.056 | 0.050 | 0.063 | 0.045 | 0.063 | 0.060 | 0.040 | 0.056 | 0.038 | 0.063 | 0.045 | 0.063 |

*J Am Stat Assoc*. Author manuscript; available in PMC 2019 February 06.

**Table 3**

Power analysis of 5 tests assuming the true covariance matrix is $\mathbf{\Sigma}_5$. We chose the significant level at $\alpha = 0.05$. For the case $n_1 = n_2 = 50$, $p = 200$, we have $m = 43$, $\tau_\alpha = 41.918$, and $\gamma_\alpha = 3.841$. For the second case, $n_1 = n_2 = 70$, $p = 1000$, $m = 62$, $\tau_\alpha = 72.318$, and $\gamma_\alpha = 3.850$. RMPBT is our approach. RAPTT is the approach of Srivastava *et al.* (2016). BS96 is the approach of Bai and Saranadasa (1996). SD08 is the approach of Srivastava and Du (2008). CQ10 is the approach of Chen *et al.* (2010).

| | $p_0$ | RMPBT ($R_1$) | RMPBT ($R_2$) | RAPTT ($R_1$) | RAPTT ($R_2$) | BS96 | SD08 | CQ10 | RMPBT ($R_1$) | RMPBT ($R_2$) | RAPTT ($R_1$) | RAPTT ($R_2$) | BS96 | SD08 | CQ10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Alternative 1 | | | | | | | Alternative 2 | | | |
| $n_1 = n_2 = 50, p = 200$ | 0.500 | 0.526 | 0.508 | **0.541** | 0.520 | 0.255 | 0.216 | 0.255 | 0.861 | 0.845 | **0.870** | 0.850 | 0.481 | 0.408 | 0.481 |
| | 0.750 | 0.490 | 0.468 | **0.504** | 0.489 | 0.214 | 0.180 | 0.214 | **0.921** | 0.914 | 0.916 | 0.900 | 0.480 | 0.417 | 0.480 |
| | 0.800 | 0.505 | 0.497 | **0.519** | 0.502 | 0.226 | 0.168 | 0.226 | **0.928** | 0.923 | 0.918 | 0.908 | 0.500 | 0.409 | 0.500 |
| | 0.950 | **0.491** | 0.469 | 0.468 | 0.445 | 0.172 | 0.128 | 0.172 | **0.969** | 0.957 | 0.929 | 0.909 | 0.466 | 0.390 | 0.466 |
| | 0.975 | **0.537** | 0.517 | 0.455 | 0.437 | 0.192 | 0.139 | 0.192 | **0.969** | 0.959 | 0.894 | 0.882 | 0.484 | 0.410 | 0.484 |
| | 0.990 | **0.587** | 0.559 | 0.426 | 0.415 | 0.203 | 0.155 | 0.203 | **0.990** | 0.983 | 0.841 | 0.835 | 0.525 | 0.452 | 0.525 |
| | 1.000 | 0.060 | 0.055 | 0.075 | 0.067 | 0.064 | 0.047 | 0.064 | 0.060 | 0.054 | 0.075 | 0.061 | 0.064 | 0.047 | 0.064 |
| $n_1 = n_2 = 70, p = 1000$ | 0.500 | 0.295 | 0.273 | **0.326** | 0.304 | 0.175 | 0.132 | 0.175 | 0.950 | 0.930 | **0.954** | 0.934 | 0.763 | 0.695 | 0.763 |
| | 0.750 | 0.258 | 0.231 | **0.290** | 0.265 | 0.135 | 0.089 | 0.135 | **0.967** | 0.953 | 0.966 | 0.951 | 0.755 | 0.681 | 0.755 |
| | 0.800 | 0.273 | 0.242 | **0.288** | 0.264 | 0.133 | 0.097 | 0.133 | 0.966 | 0.950 | **0.969** | 0.954 | 0.772 | 0.708 | 0.772 |
| | 0.950 | 0.260 | 0.246 | **0.289** | 0.266 | 0.146 | 0.107 | 0.146 | **0.986** | 0.973 | 0.971 | 0.962 | 0.783 | 0.699 | 0.783 |
| | 0.975 | 0.229 | 0.207 | **0.248** | 0.225 | 0.109 | 0.078 | 0.109 | **0.986** | 0.978 | 0.970 | 0.956 | 0.767 | 0.692 | 0.767 |
| | 0.990 | 0.246 | 0.229 | **0.268** | 0.243 | 0.122 | 0.092 | 0.122 | **0.988** | 0.984 | 0.961 | 0.939 | 0.768 | 0.687 | 0.768 |
| | 1.000 | 0.096 | 0.078 | 0.112 | 0.103 | 0.059 | 0.034 | 0.059 | 0.096 | 0.071 | 0.112 | 0.089 | 0.059 | 0.034 | 0.059 |

**Table 4**

Summary of the analysis of 3 data sets: ORGNDS=organoids, BC=breast cancer, and SRBCT=small round blue cell tumors. In each case, we assume a significance level of $\alpha = 0.05$ and report the probability of exceeding the test statistic obtained based on the data under the null (p-value). RMPBT is our approach. RAPTT is the approach of Srivastava *et al.* (2016). BS96 is the approach of Bai and Saranadasa (1996). SD08 is the approach of Srivastava and Du (2008). CQ10 is the approach of Chen *et al.* (2010).

| Data Set | $(n_1, n_2)$ | Subset | p | RMPBT ($R_1$) | RMPBT ($R_2$) | RAPTT ($R_1$) | RAPTT ($R_2$) | BS96 | SD08 | CQ10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **ORGNDS** | (3, 3) | | 2000 | 0.0058 | 0.212 | 0.1392 | 0.510 | 0.6309 | - | 0.6309 |
| **BC** | (111, 57) | Chromosome 1 | 374 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | 0.19 | 0.11 | 0.23 |
| | | Chromosome 2 | 233 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | 0.012 | 0.043 | 0.022 |
| | | Chromosome 12 | 191 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | 0.0004 | 0.026 | 0.003 |
| | (37, 19) | Chromosome 1 | 374 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | 0.416 | 0.371 | 0.466 |
| | | Chromosome 2 | 233 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | 0.223 | 0.273 | 0.327 |
| | | Chromosome 12 | 191 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | 0.135 | 0.266 | 0.304 |
| **SRBCT** | (11, 18) | | 2308 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 |

*J Am Stat Assoc*. Author manuscript; available in PMC 2019 February 06.