Received: 8 November 2021 | Revised: 17 September 2022 | Accepted: 6 April 2023

DOI: 10.1111/sjos.12655

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

# Regularized *t* distribution: definition, properties, and applications

Zongliang Hu[1] | Yiping Yang[2] | Gaorong Li[3] | Tiejun Tong[4]

[1]College of Mathematics and Statistics, Shenzhen University, Shenzhen, China

[2]School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing, China

[3]School of Statistics, Beijing Normal University, Beijing, China

[4]Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

**Correspondence**

Yiping Yang, School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing, China.
Email: yeepingyang@foxmail.com

Tiejun Tong, Department of Mathematics, Hong Kong Baptist University, Hong Kong.
Email: tongt@hkbu.edu.hk

**Funding information**

National Natural Science Foundation of China, Grant/Award Numbers: 12001378, 12271046, 11871001, 12131006, 1207010822; Guangdong Basic and Applied Basic Research Foundation, Grant/Award Number: 2019A1515110449; Natural Science Foundation of Guangdong Province, Grant/Award Number: 2023A1515010027; Chongqing Natural Science Foundation, Grant/Award Number: cstc2021jcyj-msxmX0079; Humanities and Social Sciences Program of Chongqing Education Commission, Grant/Award Number: 21SIGH118; General Research Fund of Hong Kong, Grant/Award Numbers: HKBU12303918, HKBU12303421; Initiation Grant for Faculty Niche Research Areas, Grant/Award Number: RC-FNRA-IG/20-21/SCI/03

## Abstract

For gene expression data analysis, an important task is to identify genes that are differentially expressed between two or more groups. Nevertheless, as biological experiments are often measured with a relatively small number of samples, how to accurately estimate the variances of gene expression becomes a challenging issue. To tackle this problem, we introduce a regularized *t* distribution and derive its statistical properties including the probability density function and the moment generating function. The noncentral regularized *t* distribution is also introduced for computing the statistical power of hypothesis testing. For practical applications, we apply the regularized *t* distribution to establish the null distribution of the regularized *t* statistic, and then formulate it as a regularized *t*-test for detecting the differentially expressed genes. Simulation studies and real data analysis show that our regularized *t*-test performs much better than the Bayesian *t*-test in the "*limma*" package, in particular when the sample sizes are small.

**KEYWORDS**

Bayesian *t*-test, hypothesis testing, noncentral regularized *t* distribution, regularized *t* distribution, regularized *t*-test

# 1 | INTRODUCTION

With the advances in modern technologies, high-throughput omics data are becoming more and more popular in various areas of science. A typical example of such data includes gene expression data, which are frequently applied to detect the differentially expressed (DE) genes for exploring the biological mechanisms in the gene level. For simplicity, let us first consider the one-sample test. Let $X_{ij}$ be the expression level of the $j$th gene in the $i$th sample that is normally distributed with mean $\mu_j$ and variance $\sigma_j^2$, where $i = 1, \dots, n$ and $j = 1, \dots, G$. Then for the $j$th gene, if we want to test the hypothesis

$$H_{0j} : \mu_j = \mu_{j0} \qquad \text{versus} \qquad H_{1j} : \mu_j \neq \mu_{j0}, \tag{1}$$

Student's $t$ statistic is given as

$$T_j^{(S)} = \frac{\overline{X}_j - \mu_{j0}}{\sqrt{s_j^2/n}},$$

where $\overline{X}_j = \sum_{i=1}^{n} X_{ij}/n$ and $s_j^2 = \sum_{i=1}^{n} (X_{ij} - \overline{X}_j)^2/(n-1)$ are the sample mean and sample variance of the $j$th gene, respectively. Due to the high-dimensionality of gene expression data, the number of genes $p$ is often larger or much larger than the sample size $n$. In particular, when the sample size is small and the gene expression levels are similar, the sample standard deviation (SD) $s_j$ will be very close to zero so that the test statistic value can be arbitrarily large as long as the sample mean is nonzero. As a consequence, Student's $t$-test will tend to detect many weakly expressed genes as biologically significant.

To improve the performance of detecting DE genes, it is desired to have more accurate and more stable estimates for the gene-specific variances. To achieve this, one popular approach in the literature is borrowing information across genes to further improve their estimation accuracy. To name a few earlier works, Tusher et al. (2001), Efron et al. (2001), and Storey and Tibshirani (2003) added a small positive constant $c_0$ to the sample SD, the applied $\widetilde{s}_j = s_j + c_0$ as the denominator of the test statistic. Baldi and Long (2001), Smyth (2004), and Phipson et al. (2016) considered an inverse-gamma prior distribution for $\sigma_i^2$ with $\nu_0$ degrees of freedom and scale parameter $s_0^2$, which leads to the posterior estimates of the variances as $\widetilde{s}_j^2 = ((n-1)s_j^2 + \nu_0 s_0^2)/(n - 1 + \nu_0)$. In other words, they proposed to estimate the gene variance by a weighted average of the sample variance and the hyperparameter value. To test the hypothesis (1), they proposed a Bayesian $t$ statistic as

$$T_j^{(B)} = \frac{\overline{X}_j - \mu_{j0}}{\sqrt{\widetilde{s}_j^2/n}}. \tag{2}$$

They further showed by simulations that the Bayesian $t$-test is often more powerful and meanwhile provides a lower false discovery rate than Student's $t$-test. Since then, the Bayesian $t$-test has played an important role in multiple testing problems. For instance, most existing $t$-type statistics for detecting DE genes for microarray data or RNA-Seq data can be dated back to formula (2), including the well-known test statistics in the "*limma*" package (Ritchie et al., 2015) which have been extensively used in thousands of published biological studies.

Despite the huge popularity, we note however that the exact distribution of the test statistic (2) has not yet been studied analytically. As pointed by Robinson et al. (2010) and Patrick et al. (2013), the posterior estimate $\tilde{s}_j$ does not follow a chi-square distribution, and consequently, $T_j^{(B)}$ does not follow an exact $t$ distribution. In the literature, several approaches have been proposed to approximate the null distribution of the Bayesian $t$ statistic. For example, Tusher et al. (2001), Cui and Churchill (2003), and Robinson et al. (2010) approximated the null distribution of $T_j^{(B)}$ by the permutation methods. On the other side, Baldi and Long (2001), Smyth (2004), and Phipson et al. (2016) proposed to still use a Student's $t$ distribution but with an enlarged degree of freedom to approximate the null distribution of $T_j^{(B)}$. Nevertheless, when the sample size is small, those approximated null distributions may not perform well and often have a large deviation from their exact distributions, which consequently lead to a lower statistical power or an inflated type I error for the conducted tests.

Inspired by the above observations, we propose to study the exact distribution of $T_j^{(B)}$, and also investigate how it can be applied to multiple testing with real applications. To be more specific, we will reformulate the problem and introduce a regularized $t$ distribution, which includes two shape parameters for modeling the test statistic in (2). We then show that, with a proper selection of the two parameters, the regularized $t$ distribution will include the normal distribution, Student's $t$ distribution, and the scaled $t$ distribution as important special cases. While for the statistical power of the proposed test, we will also introduce the noncentral regularized $t$ distribution. Simulation studies show that the test associated with the regularized $t$ distribution is more powerful in multiple testing than those with the null distributions based on the permutation or approximation methods.

The rest of the paper is organized as follows. In Section 2, we introduce the regularized $t$ distribution and investigate its statistical properties. The noncentral regularized $t$ distribution is introduced in Section 3. In Section 4, we conduct simulation studies to illustrate the usefulness of the new proposed distribution in multiple testing. We further apply the new distribution to analyze a real data example in Section 5. Finally, we conclude the paper with some future works in Section 6, and provide the technical results in the Appendices.

## 2 | REGULARIZED T DISTRIBUTION

Assume that $Z$ follows a standard normal distribution, $U$ follows a chi-square distribution with $\nu$ degrees of freedom, and that $Z$ and $U$ are independent of each other. In statistics, it is well known that the ratio

$$\frac{Z}{\sqrt{U/\nu}},$$

follows a Student's $t$ distribution with $\nu$ degrees of freedom. The $t$ distribution was first introduced in 1908 by William S. Gosset, who is better known under the pseudonym "Student" (Student, 1908).

To illustrate the usefulness of Student's $t$ distribution, we considered the one-sample hypothesis $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, where $\mu_0$ is a specified value. Let also $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean $\mu$ and variance $\sigma^2$. When $\sigma^2$ is known, the test statistic for the hypothesis is given as

$$T_Z = \frac{\overline{X} - \mu_0}{\sqrt{\sigma^2/n}}, \tag{3}$$

where $\overline{X} = \sum_{i=1}^{n} X_i/n$ is the sample mean. Under the null hypothesis, noting that $\overline{X} \sim N(\mu_0, \sigma^2/n)$, $T_Z$ follows a standard normal distribution. This test is well known as the $Z$-test.

In practice, however, the variance $\sigma^2$ is often unknown so that the $Z$-test is no longer applicable. To have a valid test statistic, one needs to replace the unknown $\sigma^2$ by a sample estimate in the denominator of $T_Z$ in (3). Say, for example, if we apply the sample variance, it will yield the test statistic as

$$T_S = \frac{\overline{X} - \mu_0}{\sqrt{s^2/n}} = \frac{(\overline{X} - \mu_0)/\sqrt{\sigma^2/n}}{\sqrt{s^2/\sigma^2}} = \frac{T_Z}{\sqrt{U_{n-1}/(n-1)}}, \tag{4}$$

where $U_{n-1} = (n-1)s^2/\sigma^2$ and $T_Z$ is the same as defined in (3). Under the null hypothesis, $T_Z$ follows a standard normal distribution, $U_{n-1}$ follows a chi-square distribution with $n-1$ degrees of freedom, and that $T_Z$ and $U_{n-1}$ are independent of each other. Then by the definition of Student's $t$ distribution, $T_S$ follows a Student's $t$ distribution with $n-1$ degrees of freedom.

## 2.1 | Definition

As mentioned in the introduction, for high-dimensional small sample size data, the sample variances may not provide reliable estimates for the true variances. And consequently, the associated $t$-tests using the sample variances are often at the risk of generating misleading results including the uncontrolled type I errors or resulting in higher false discovery rates.

Motivated by the success of Bayesian $t$-test, many researchers have devoted to develop new and novel estimates for the gene-specific variances (Cui et al., 2005; Opgen-Rhein & Strimmer, 2007; Pimentel et al., 2017; Tong et al., 2014; Tong & Wang, 2007). To summarize, many variance estimators for the $j$th gene can be formulated in a regularized form, $\breve{\sigma}_j^2 = \lambda s_j^2 + (1-\lambda)\sigma_{j0}^2$, where $\sigma_{j0}^2$ is a fixed value serving as the shrinkage target, and $\lambda \in (0,1)$ is the shrinkage parameter. Now if the regularized estimator $\breve{\sigma}_j^2$ is applied to estimate $\sigma^2$ in (3), we will have the regularized $t$ statistic as

$$T_j = \frac{\overline{X}_j - \mu_{j0}}{\sqrt{\breve{\sigma}_j^2/n}} = \frac{\overline{X}_j - \mu_{j0}}{\sqrt{\frac{1}{n}[\lambda s_j^2 + (1-\lambda)\sigma_{j0}^2]}} = \frac{Z_j}{\sqrt{\lambda \frac{U_{n-1}}{n-1} + (1-\lambda)\frac{\sigma_{j0}^2}{\sigma_j^2}}}, \tag{5}$$

where $Z_j = (\overline{X}_j - \mu_{j0})/\sqrt{\sigma_j^2/n}, \sqrt{\sigma^2/n}$ and $U_{n-1} = (n-1)s_j^2/\sigma_j^2$. Under the null hypothesis, $Z_j$ follows a standard normal distribution, $U_{n-1}$ follows a chi-square distribution with $n-1$ degrees of freedom, and $Z_j$ and $U_{n-1}$ are independent of each other. To the best of our knowledge, however, the exact distribution of this regularized $t$ statistic has never been studied in the literature. As alternatives, researchers often applied the approximation or permutation methods to establish the null distribution of $T_j$.

To systematically study the regularized $t$ statistic, we treat $a = \lambda$ and $b = (1-\lambda)\sigma_{j0}^2/\sigma_j^2$ as two regularization parameters in formula (5), and then define the regularized $t$ distribution as follows.

**Definition 1.** Assume that $Z$ follows a standard normal distribution, $U$ follows a chi-square distribution with $\nu$ degrees of freedom, and that $Z$ and $U$ are independent of each other. Then for any $a > 0$ and $b \geq 0$, the probability distribution of

$$T = \frac{Z}{\sqrt{a(U/\nu) + b}}, \tag{6}$$

defines a regularized $t$ distribution with $\nu$ degrees of freedom and regularization parameters $a$ and $b$. For convenience, we also refer to $T$ as a regularized $t$ random variable, and denote it by $T \sim t_\nu(a, b)$.

Following the definition, it is clear that the regularized $t$ distribution includes some important distributions as special cases. When $a = 1$ and $b = 0$, it gives a Student's $t$ distribution. When $a \neq 1$ and $b = 0$, it gives a scaled $t$ distribution; and more specifically by Praetz (1972), $a = \nu\sigma^2/(\nu - 2)$ with $\nu > 2$ is most frequently used. In addition, when $a = 0$ and $b > 0$, the regularized $t$ distribution reduces to a normal distribution with zero mean and variance $1/b$. And because of this, we only consider $a > 0$ in the definition of the regularized $t$ distribution.

Also following the definition, there is another interesting connection between the regularized $t$ distribution and the normal distribution, with the proof given in Appendix A.1 in Data S1.

**Theorem 1.** *Let $T = Z/\sqrt{a(U/\nu) + b}$ be a regularized t random variable with parameters $\nu > 0$, $a > 0$ and $b \geq 0$, where $Z$ and $U$ are two random variables as defined in (6). Then, $T \xrightarrow{D} N(0, 1/(a + b))$ as $\nu \to \infty$, where $\xrightarrow{D}$ denotes the convergence in distribution.*

Finally, it is noteworthy that the regularized $t$ distribution has no essential difference between $0 < a < 1$ and $a \geq 1$. To demonstrate it, we let $X$ follow the regularized $t$ distribution $t_\nu(a_1, b_1)$ with $0 < a_1 < 1$ and $b_1 \geq 0$. Then if $Y = X/\sqrt{\varpi_0}$ with $\varpi_0 \geq 1/a_1$ being a positive constant, by (6) we have $Y \sim t_\nu(\varpi_0 a_1, \varpi_0 b_1)$. This shows that a regularized $t$ distribution with $0 < a_1 < 1$ and $b_1 \geq 0$ can be easily transformed to another regularized $t$ distribution with $a_2 = \varpi_0 a_1 \geq 1$ and $b_2 = \varpi_0 b_1 \geq 0$. In view of this, we will focus on $0 < a \leq 1$ only in the remainder of the paper.

## 2.2 | Probability density function

In this section, we derive the probability density function (PDF) and the cumulative distribution function (CDF) of the regularized $t$ distribution.

**Theorem 2.** *Let $\Gamma(\cdot)$ be the gamma function. For $t \in (-\infty, \infty)$, the PDF of the regularized t distribution is*

*(i) when $b = 0$,*

$$f_T(t) = \sqrt{\frac{a}{\nu\pi}} \frac{\Gamma((\nu + 1)/2)}{\Gamma(\nu/2)} \left(1 + \frac{at^2}{\nu}\right)^{-\frac{\nu+1}{2}};$$

*(ii) when $b > 0$,*

$$f_T(t) = \sqrt{\frac{a}{\nu\pi}} \left(\frac{b\nu}{2a}\right)^{\frac{\nu+1}{2}} U\left(\frac{\nu}{2}, \frac{\nu + 3}{2}, \frac{\nu b}{2a} + \frac{bt^2}{2}\right) e^{-\frac{bt^2}{2}}, \tag{7}$$

*where $U(x,y,z) = \int_0^\infty e^{-zt} t^{x-1}(1+t)^{y-x-1}dt/\Gamma(x)$ is the confluent hypergeometric function of the second kind, and the real parts of $x$ and $z$ are required to be positive values (Abramowitz & Stegun, 1972; Prosser, 1994).*

The proof of Theorem 2 is given in Appendix A.2 in Data S1. Accordingly, the CDF of the regularized $t$ distribution can be written as

$$F_T(t) = \int_{-\infty}^{t} f_T(u)du.$$

As mentioned before, when $a = 1$ and $b = 0$, the regularized $t$ distribution is equivalent to Student's $t$ distribution. In this case, the PDF is given as

$$f_T(t) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < t < \infty.$$

And accordingly, the CDF of Student's $t$ distribution is $F_T(t) = I_{\nu/(\nu+t^2)}(\nu/2, 1/2)/2$ for $t \le 0$, or $F_T(t) = 1 - I_{\nu/(\nu+t^2)}(\nu/2, 1/2)/2$ for $t > 0$, where $I_x(y,z)$ is the regularized incomplete beta function (Dutka, 1981). For more details, one may refer to Definition A1 in Appendix A in Data S1.

**Theorem 3.** *For the PDF of the regularized t distribution in (7), we have*

*(i)* $f_T(t)$ is an even function, that is, $f_T(t) = f_T(-t)$ for any $t > 0$.
*(ii)* $f_T(t)$ is a unimodal distribution such that its first derivative is positive when $t < 0$, negative when $t > 0$, and zero when $t = 0$.
*(iii)* $f_T(t)$ is a convex function when $t$ is close to zero, and is a concave function when the absolute value of $t$ goes to large.
*(iv)* $f_T(t)$ is infinitely differentiable for any $-\infty < t < \infty$.

Theorem 3 provides some basic properties for the PDF of the regularized $t$ distribution, and its proof is given in Appendix A.3 in Data S1. Specifically by Theorem 3(i) and 3(ii), both the median and mode of the regularized $t$ distribution are located at the origin point. For more properties of $f_T(t)$ including its effects on the parameters, one may refer to Theorem B1 in Appendix B.5 in Data S1. In addition for visualization purposes, we have also plotted the PDFs of the regularized $t$ distribution with different values of $\nu$, $a$, and $b$ in Figure 1.

Figure 1a shows the PDFs of the regularized $t$ distribution with $a = 1$, $b = 0$ and varying $\nu$. From Definition 1, those PDFs degenerate to the PDFs of Student's $t$ distribution with $\nu$ degrees of freedom. Figure 1b shows the PDFs of the regularized $t$ distribution with $a = 0.5$, $b = 0.5$ and varying $\nu$. When $\nu$ becomes larger, the tails of the PDFs become shorter and the peaks of the PDFs turn to be higher. Besides, the PDFs are close to each other when $\nu$ increases from 10 to 30. This coincides with Theorem 1 that, as $\nu \to \infty$, the regularized $t$ distribution converges to a normal distribution with mean 0 and variance $1/(a+b)$. Figure 1c shows the PDFs of the regularized $t$ distribution with $b = 0.5$ and $\nu = 1$ but with varying $a$. When $a$ becomes larger, the tails of the PDFs are shorter and the peaks of the PDFs are higher. Note that this also coincides with Theorem

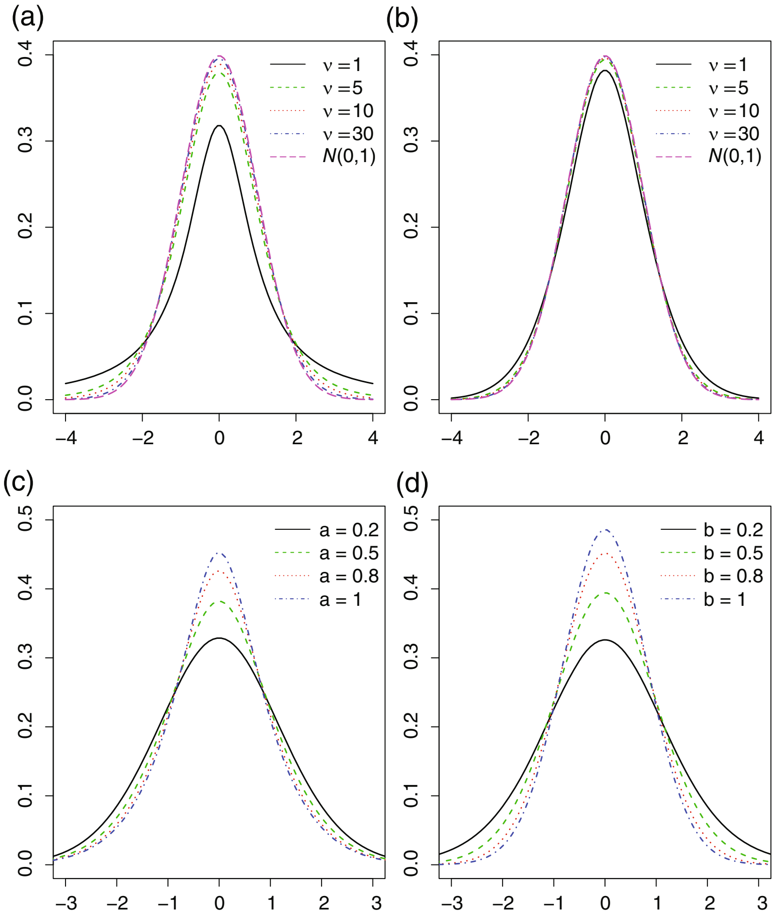

**FIGURE 1** The probability density functions of the regularized $t$ distribution, $t_\nu(a, b)$, with different values of $\nu$, $a$, and $b$. (a) $a = 1$ and $b = 0$; (b) $a = 0.5$ and $b = 0.5$; (c) $b = 0.5$ and $\nu = 1$; (d) $a = 0.5$ and $\nu = 1$.

B1 in Appendix B.5 in Data S1. Figure 1d shows the PDFs of the regularized $t$ distribution with $a = 0.5$ and $\nu = 1$ but with varying $b$. Similar to the effects on $a$, when $b$ becomes larger, the tails of the PDFs also become shorter and the peaks of the PDFs will be higher, which again coincides with Theorem B1 in Appendix B.5 in Data S1.

## 2.3 | Moments and moment generating function

In this section, we derive the finite-order moments of the regularized $t$ distribution, followed by its moment generating function (MGF) which is also essential in learning the statistical behaviors of a certain distribution. For ease of notation, we let $k!! = k(k-2)\cdots 2$ for even $k$, and $k!! = k(k-2)\cdots 1$ for odd $k$.

**Theorem 4.** *Let $T$ be a regularized $t$ random variable with $\nu > 0$, $a > 0$ and $b \geq 0$. Then for the kth moment of $T$, we have* *(i) when $b = 0$,*

$$E(T^k) = \begin{cases} 0, & k \text{ is odd}, \ k < \nu, \\ (k-1)!! \cdot (\frac{\nu}{2a})^{k/2} \frac{\Gamma((\nu-k)/2)}{\Gamma(\nu/2)}, & k \text{ is even}, \ k < \nu, \\ \text{undefined}, & k \geq \nu. \end{cases} \tag{8}$$

*(ii) when $b > 0$,*

$$E(T^k) = \begin{cases} 0, & k \text{ is odd}, \\ (k-1)!! \cdot \left(\frac{\nu}{2a}\right)^{\frac{\nu}{2}} b^{\frac{\nu-k}{2}} \cdot U(\frac{\nu}{2}, \frac{\nu-k+2}{2}, \frac{b\nu}{2a}), & k \text{ is even}. \end{cases} \tag{9}$$

The proof of Theorem 4 is given in Appendix A.4 in Data S1. This theorem shows that the mean of the regularized $t$ distribution is zero if it exists, which can also be derived from Theorem 3(i) that the regularized $t$ distribution has a symmetric PDF about $t = 0$. Also by (8) and (9), we can derive the variance of $T$ as

$$\text{Var}(T) = \begin{cases} \frac{\nu}{a(\nu-2)}, & b = 0, \nu > 2, \\ b^{\frac{\nu}{2}-1}(\frac{\nu}{2a})^{\frac{\nu}{2}} U(\frac{\nu}{2}, \frac{\nu}{2}, \frac{b\nu}{2a}), & b > 0, \nu > 0, \\ \text{undefined}, & \text{otherwise}. \end{cases}$$

**Theorem 5.** *Let $T$ be a regularized $t$ random variable with $\nu > 0$, $a > 0$ and $b \geq 0$. The MGF of $T$ exists if and only if $b > 0$. When $b = 0$, $T$ is a heavy tailed random variable and its MGF does not exist.*

The proof of Theorem 5 is given in Appendix A.5 in Data S1. This theorem shows that the existence of the MGF for the regularized $t$ distribution is fully determined by the regularized parameter $b$. In other words, the regularized $t$ distribution behaves like Student's $t$ distribution when $b = 0$, and it behaves like the normal distribution when $b > 0$. In view of this, for the regularized $t$ statistic in (5), it may not be accurate to approximate its null distribution by Student's $t$ distribution, especially when the sample size is small. Nevertheless, we also note that such a heavy tailed approximation has be commonly applied in the literature including, for example, Baldi and Long (2001), Smyth (2004), Cui et al. (2005), Law et al. (2014), and Pimentel et al. (2017). Moreover, to visualize the tails of the regularized $t$ distribution, we have also plotted several PDFs with $\nu = 5$ under the setting of $a + b = 1$ in Figure 2. From the right panel of Figure 2, it clearly shows that the tails of the regularized $t$ distribution will quickly converge to the tail of the standard normal distribution rather than Student's $t$ distribution when $b$ increases from 0 to 1.

Finally, for reference, we have also followed the Wikipedia style and provided the summary statistics of the regularized $t$ distribution in Table 1.

## 3 | NONCENTRAL REGULARIZED T DISTRIBUTION

In the power analysis of Student's $t$-test, the noncentral $t$ distribution is a generalization of Student's $t$ distribution with a shifted mean in the numerator. Specifically, if we consider the

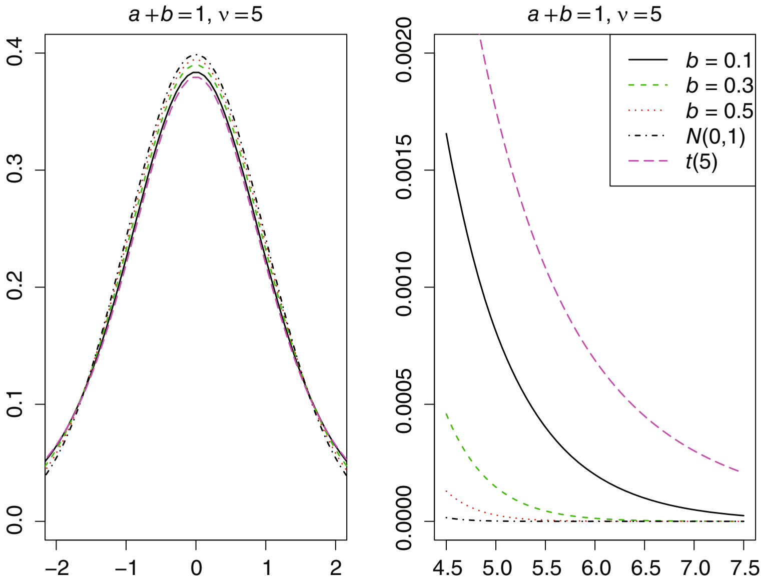

**FIGURE 2** The probability density functions of the regularized $t$ distribution, $t_\nu(a,b)$, with $\nu = 5$ and $a+b=1$.

following ratio

$$\frac{Z+\mu}{\sqrt{U/\nu}}, \tag{10}$$

it will lead to a noncentral $t$ distribution with $\nu$ degrees of freedom and noncentrality parameter $\mu \neq 0$. Inspired by this, we define the noncentral regularized $t$ distribution as follows.

**Definition 2.** Assume that $Z$ follows a standard normal distribution, $U$ follows a chi-square distribution with $\nu$ degrees of freedom, and that $Z$ and $U$ are independent of each other. For any given $a > 0$, $b \geq 0$ and $\mu \neq 0$, the probability distribution of

$$W = \frac{Z+\mu}{\sqrt{a(U/\nu)+b}}, \tag{11}$$

defines a noncentral regularized $t$ distribution with $\nu$ degrees of freedom, regularization parameters $a$ and $b$, and noncentrality parameter $\mu$. We also refer to $W$ as a noncentral regularized $t$ random variable, and denote it by $W \sim t_\nu(a,b;\mu)$.

In the special case when $a=1, b=0$ and $\mu \neq 0$, the noncentral regularized $t$ distribution reduces to a noncentral $t$ distribution. In addition, as $\nu \to \infty$, the noncentral regularized $t$ distribution will converge to a normal distribution with mean $\mu$ and variance $1/(a+b)$. The following theorem gives the PDF, the moments and the MGF of the noncentral regularized $t$ distribution.

**Theorem 6.** *Let $W$ be a noncentral regularized $t$ random variable with $\nu > 0$, $a > 0$, $b \geq 0$ and $\mu \neq 0$. Then,*

**TABLE 1** Regularized *t* distribution.

| Notation | $t_\nu(a,b)$ | |
|---|---|---|
| | $\nu > 0$ — degrees of freedom | |
| Parameters | $a > 0$ — regularized shape parameter | |
| | $b \geq 0$ — regularized shape parameter | |
| Support | $t \in (-\infty, \infty)$ | |
| PDF | $f(t) = \sqrt{\frac{a}{\nu\pi}}\left(\frac{b\nu}{2a}\right)^{\frac{\nu+1}{2}} U\left(\frac{\nu}{2}, \frac{\nu+3}{2}, \frac{b\nu}{2a} + \frac{bt^2}{2}\right) e^{-\frac{bt^2}{2}}$ | for $b > 0$ |
| | $f(t) = \sqrt{\frac{a}{\nu\pi}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\left(1 + \frac{at^2}{\nu}\right)^{-\frac{\nu+1}{2}}$ | for $b = 0$ |
| CDF | $F(t) = \int_{-\infty}^{t} f(u)du$ | |
| Mean | 0 | for $b > 0$ |
| | 0 | for $b = 0$ and $\nu > 1$ |
| | Undefined | for $b = 0$ and $0 < \nu \leq 1$ |
| Median | 0 | |
| Mode | 0 | |
| Variance | $b^{\frac{\nu}{2}-1}\left(\frac{\nu}{2a}\right)^{\frac{\nu}{2}} U\left(\frac{\nu}{2}, \frac{\nu}{2}, \frac{b\nu}{2a}\right)$ | for $b > 0$ |
| | $\frac{\nu}{a(\nu-2)}$ | for $b = 0$ and $\nu > 2$ |
| | Undefined | for $b = 0$ and $0 < \nu \leq 2$ |
| Skewness | 0 | for $b > 0$ |
| | 0 | for $b = 0$ and $\nu > 3$ |
| | Undefined | for $b = 0$ and $0 < \nu \leq 3$ |
| MGF | $\sqrt{\frac{a}{\nu\pi}}\left(\frac{b\nu}{2a}\right)^{\frac{\nu+1}{2}} \int_0^\infty U\left(\frac{\nu}{2}, \frac{\nu+3}{2}, \frac{\nu b}{2a} + \frac{bt^2}{2}\right) e^{\xi t-\frac{bt^2}{2}} dt$ | for $b > 0$ |
| | Undefined | for $b = 0$ |

*(i) the PDF of W, defined on $w \in (-\infty, \infty)$, is given as*

$$f_W(w) = \frac{\sqrt{2/\pi}}{\Gamma(\nu/2)}\left(\frac{\nu}{2a}\right)^{\frac{\nu}{2}} e^{\frac{\nu b}{2a}-\frac{\mu^2}{2}} \int_{\sqrt{b}}^{\infty} s^2(s^2 - b)^{\frac{\nu}{2}-1} e^{-s^2(\nu/a+w^2)/2+\mu ws} ds;$$

*(ii) the kth moment of W is*

$$E(W^k) = \begin{cases} \text{undefined}, & b = 0, k \geq \nu, \\ \left(\frac{\nu}{2a}\right)^{\frac{k}{2}} \frac{\Gamma((\nu-k)/2)}{\Gamma(\nu/2)\sqrt{\pi}} \cdot \sum_{r\in\mathcal{S}} \binom{k}{r} 2^{\frac{r}{2}} \mu^{k-r} \Gamma\left(\frac{r+1}{2}\right), & b = 0, k < \nu, \\ \frac{b^{-\frac{k}{2}}}{\sqrt{\pi}}\left(\frac{b\nu}{2a}\right)^{\frac{\nu}{2}} U\left(\frac{\nu}{2}, \frac{\nu+2-k}{2}, \frac{b\nu}{2a}\right) \cdot \sum_{r\in\mathcal{S}} \binom{k}{r} 2^{\frac{r}{2}} \mu^{k-r} \Gamma\left(\frac{r+1}{2}\right), & b > 0, \end{cases}$$

*where $\mathcal{S} = \{r : r \text{ is even}, r \leq k\}$;*

*(iii) the MGF of W exists if and only if $b > 0$; and when $b = 0$, W is a heavy tailed random variable and its MGF does not exist.*

The proof of Theorem 6 is given in Appendix A.6 in Data S1. This theorem indicates that the existence of the MGF for the noncentral regularized $t$ distribution is fully determined by the regularized parameter $b$ and is irrelevant to the shift in $\mu$. For more details about the noncentral regularized $t$ distribution, one may refer to the additional PDF plots in Appendix B.1 in Data S1.

## 4 | MONTE CARLO SIMULATION STUDIES

In this section, we carry out simulation studies to evaluate the performance of the regularized test based on the regularized $t$ distribution. For ease of presentation, we present the simulations for the two-sample test only, where one is the treatment group and the other is the control group. We also compare our regularized $t$-test (R-$t$) to the Bayesian $t$-test with the null distribution approximated by Student's $t$ distribution (B-$t$) or by the permutation method (P-$t$), respectively.

### 4.1 | Type I error rate

We first evaluate the gene-specific type I error rate for each test. To start with, we generate data for $G = 1000$ independent genes. For the $j$th gene, let $X_{1j},\ldots,X_{n_1j}$ be an independent random sample of size $n_1$ from $N(\mu_{1j},\sigma_j^2)$, $Y_{1j},\ldots,Y_{n_2j}$ be an independent random sample of size $n_2$ from $N(\mu_{2j},\sigma_j^2)$, and that the two samples are independent of each other. For computing the type I error rate, we further let $\mu_{1j} = \mu_{2j} = 0$ for all genes. The variance $\sigma_j^2$ is randomly sampled from a scaled inverse chi-square distribution, $\zeta_{0j}^2 d_{0j}/\chi^2(d_{0j})$, with $d_{0j} = 100$ degrees of freedom and location parameter $\zeta_{0j}^2$ (Opgen-Rhein & Strimmer, 2007; Smyth, 2004). Also to account for different levels of heterogeneity, we let $\zeta_{0j}^2 = 0.25$ for the first 500 genes and $\zeta_{0j}^2 = 4$ for the last 500 genes.

To test whether the genes are differentially expressed, we consider the hypotheses $H_{0j} : \mu_{1j} = \mu_{2j}$ versus $H_{1j} : \mu_{1j} \neq \mu_{2j}$ for $j = 1,\ldots,G$. The regularized $t$ statistic for the $j$th gene is then given as

$$\widetilde{T}_j = \frac{n_1 n_2}{N}\frac{\overline{X}_j - \overline{Y}_j}{\widetilde{s}_{\text{pool},j}}, \tag{12}$$

where $N = n_1 + n_2$, $\overline{X}_j$ and $\overline{Y}_j$ are the two sample means of the $j$th gene, and $\widetilde{s}^2_{\text{pool},j} = [(N-2)s^2_{\text{pool},j} + \nu_0 s_0^2]/(N-2+\nu_0)$ with $s^2_{\text{pool},j}$ being the pooled sample variance and $\nu_0$ and $s_0^2$ being two hyperparameters. By the definition in Section 2, under the null hypothesis we have

$$\widetilde{T}_j \sim t_{N-2}(a_j, b_j), \tag{13}$$

with $a_j = (N-2)/(N-2+\nu_0)$ and $b_j = \nu_0 s_0^2/[(N-2+\nu_0)\sigma_j^2]$. We further specify the values of $\nu_0$ and $s_0^2$ by following the same procedure as in Ritchie et al. (2015), and estimate the unknown $1/\sigma_j^2$ by its unbiased estimate $(N-4)/[(N-2)s^2_{\text{pool},j}]$. For more details, see Appendix B.2 in Data S1. Finally, to assess the performance of R-$t$, B-$t$ and P-$t$, we consider the sample sizes as $(n_1, n_2) = (3,3), (5,5)$ or (30,30), respectively.

For each scenario, we repeat the simulations for 1000 times and compute the gene-specific type I error rates for each test. Figure 3 plots the gene-specific type I error rates for the three tests at the nominal level $\alpha = 0.05$. We note that R-$t$ is able to control the gene-specific type I error rates in most settings. While for B-$t$, when the sample sizes are small, it either provides an inflated or

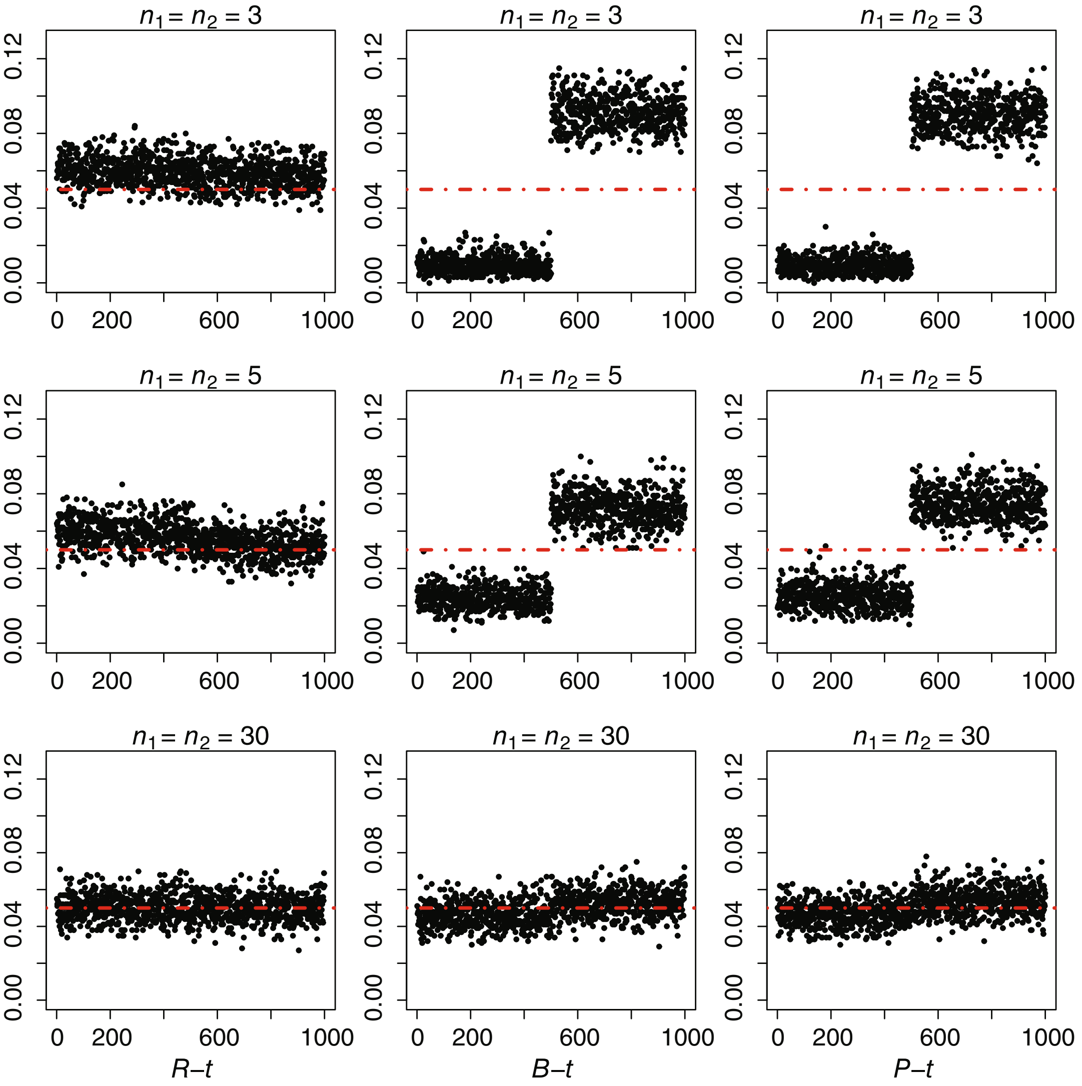

**FIGURE 3** Type I error rates of each gene among R-*t*, B-*t* and P-*t*. The horizontal dashed red lines represent the nominal level of $\alpha = 0.05$. The gene-specific type I error rate is plotted for each test versus the index number of genes.

a conservative type I error rate. More specifically, it exhibits a conservative type I error rate for the genes with smaller variances, and a significantly inflated type I error rate for the genes with larger variances, mainly because the null distribution built by B-*t* (always set as $t_{N-2+\nu_0}$, regardless of the unknown gene-specific variances) can be far away from the exact null distribution, and so provides a less satisfactory control for the type I error rate. In addition, we note that P-*t* performs very similarly as B-*t*, with the reasons explained in Appendix B.3 in Data S1. Finally, when the sample sizes are large, the three tests perform similarly and all keep the type I error rates around the nominal level. This coincides with Theorem 1 that the regularized *t* distribution converges to a normal distribution when the degrees of freedom is large, showing that the three tests are all asymptotically equivalent.

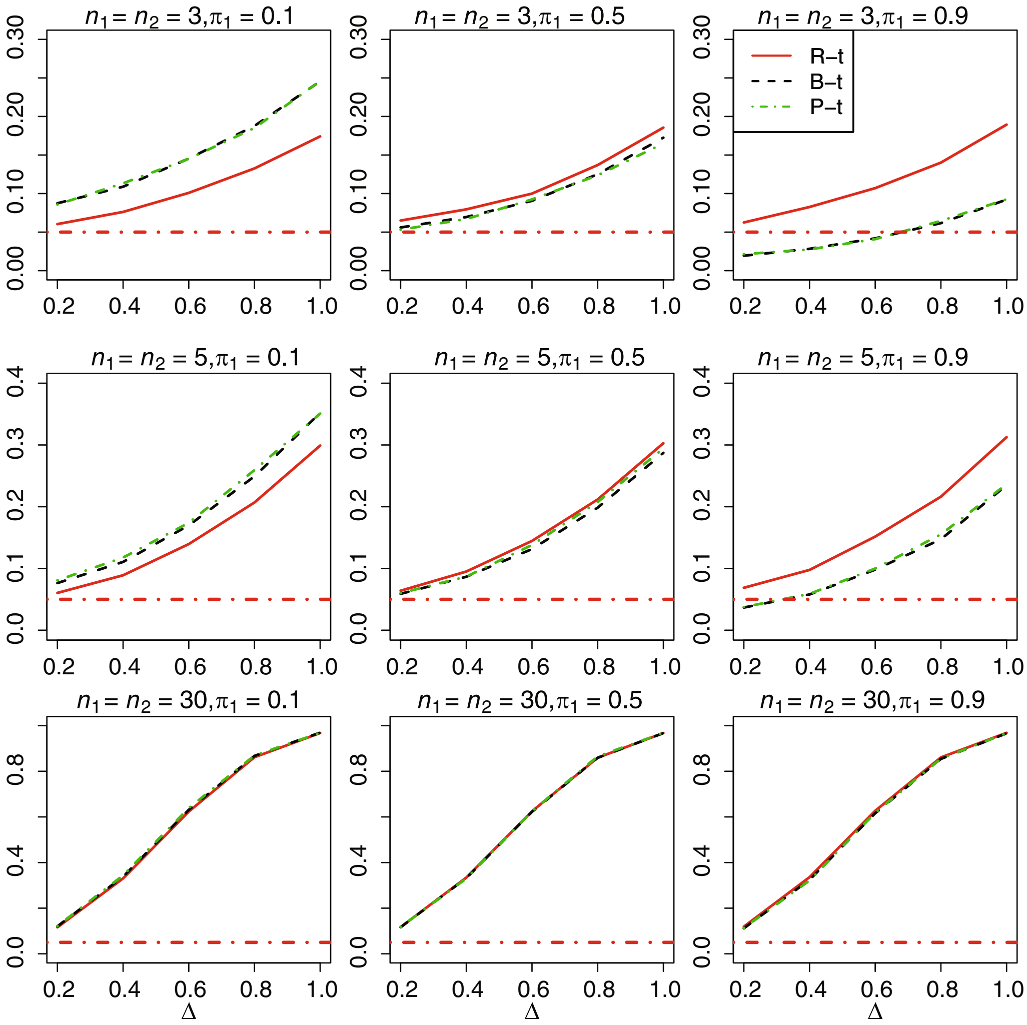

**FIGURE 4** Empirical power comparisons among R-$t$, B-$t$ and P-$t$. The horizontal dashed red lines represent the nominal level of $\alpha = 0.05$. The power is plotted for each test versus the effect size $\Delta$.

## 4.2 | Statistical power

To compare the statistical power for the three tests, we follow the same simulation setting as in Section 4.1 except that $m_1 = 200$ genes are now differentially expressed (DE). We first randomly sample $\pi_1 m_1$ and $(1 - \pi_1)m_1$ genes from the first 500 and the last 500 genes, respectively. We further set $\mu_{1j} = 0$ and $\mu_{2j} = \Delta\gamma_j\sigma_j$ for those DE genes, where $\Delta$ is the effect size and $\gamma_j$ is a binary random variable with $P(\gamma_j = 1) = P(\gamma_j = -1) = 0.5$. Consequently, a smaller $\pi_1$ indicates that the DE genes tend to have a larger variance. Finally, with $\pi_1 = 0.1, 0.5, 0.9$, we repeat the simulations for 1000 times, and compute the empirical power as the proportion of the number of identified DE genes to the total number of true DE genes.

Figure 4 presents the simulated power of R-$t$, B-$t$ and P-$t$ versus the effect size $\Delta$ ranging from 0.2 to 1. We note that B-$t$ and P-$t$ will exaggerate the statistical power when most DE genes tend

**TABLE 2** The detected numbers of differentially expressed (DE) genes by R-*t*, B-*t* and P-*t*, respectively, where $\alpha$ is the nominal level.

| | $\alpha = 0.01$ | | | $\alpha = 0.025$ | | | $\alpha = 0.05$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | R-*t* | B-*t* | P-*t* | R-*t* | B-*t* | P-*t* | R-*t* | B-*t* | P-*t* |
| DE genes | 225 | 221 | 210 | 333 | 324 | 294 | 432 | 428 | 381 |

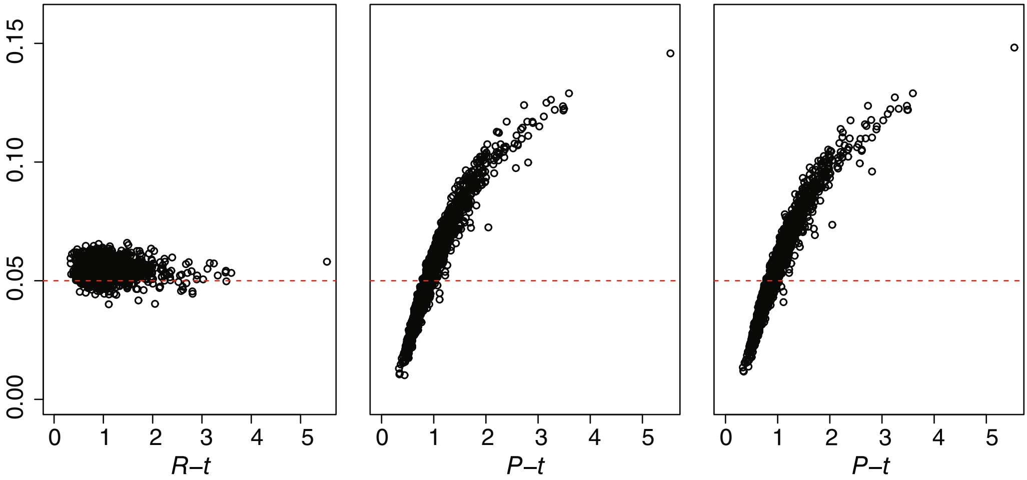

**FIGURE 5** False positive rate (FPR) comparisons among R-*t*, B-*t* and P-*t*. The gene-specific FPR is plotted for each test versus the corresponding gene variance. The horizontal dashed red lines represent the nominal level of $\alpha = 0.05$.

to have a larger variance (e.g., $\pi = 0.1$), since B-*t* and P-*t* significantly inflate the gene-specific type I error rates of those DE genes with larger variances. Moreover, B-*t* and P-*t* will reduce the statistical power when most DE genes tend to have a smaller variance (e.g., $\pi = 0.9$), since B-*t* and P-*t* exhibit a conservative gene-specific type I error rates of those DE genes with smaller variances. On the contrary, it is evident that R-*t* always provides a stable power across different values of $\pi_1$, mainly because R-*t* is the only test among the three that is able to control the gene-specific type I error rates around the nominal level. Finally, when the sample sizes are large, the power curves of B-*t*, R-*t* and P-*t* are nearly overlapped, showing that the three tests are asymptotically equivalent, which is also consistent with Theorem 1 and Figure 3.

## 5 | REAL DATA ANALYSIS

We consider the colon data from Alon et al. (1999), which contains gene expression levels of 40 tumor and 22 normal colon tissues samples for 6500 human genes obtained from an Affymetrix oligonucleotide array. The dataset is publicly available in the R package "*datamicroarray*" (John, 2016). We select the top 2000 genes with the highest minimal intensity across all 62 samples as in Alon et al. (1999) and further take the log2 transformation for the raw expression data. Let $X_{1,j}, \ldots, X_{40,j}$ be the expression levels of gene $j$ from the 40 tumor samples, and $Y_{1,j}, \ldots, Y_{22,j}$ be those from the normal samples. Assume that each gene from different tumor (or normal) samples are independently normally distributed with mean $\mu_{1j}$ or $\mu_{2j}$, respectively. Then to test the hypotheses $\mu_{1j} = \mu_{2j}$ versus $\mu_{1j} \neq \mu_{2j}$ for $j = 1, \ldots, 2000$, we compute the $p$-values for each gene based on R-*t*, B-$t$ and P-$t$, respectively. With the detected DE genes in Table 2 at the nominal level $\alpha = 0.01, 0.025$ or 0.05, we observe that R-$t$ detects more DE genes than B-$t$ and P-$t$ in all the settings.

To further compare the performance of the three tests, we now evaluate the gene-specific false positive rate (FPR). To achieve this goal, we apply the bootstrap method to generate two artificial groups for the 2000 genes that mimic the null and alternative hypotheses, respectively. Specifically, we randomly sample two distinct classes of size 10 with replacement from the tumor group. Since both classes are partitioned from the tumor data, the null hypothesis can be regarded as the truth. We repeat the procedure for 4000 times, and perform the three tests at the nominal level $\alpha = 0.05$. The rejection rate is then computed to represent the FPR for each gene, which is further plotted in Figure 5 along with the corresponding gene variance. From the figure, it is evident that the gene-specific FPRs of R-$t$ are all around the nominal level and not affected by the gene-specific variances. Nevertheless, both of B-$t$ and P-$t$ display an inflated FPR for genes with a larger variance and a conservative FPR for genes with a smaller variance, showing that they may not be valid tests for practical use.

## 6 | CONCLUSION

For high-dimensional data, traditional statistical tests often suffer from low powers in multiple testing due to the small number of collected samples. Borrowing information across variables help to stabilize the test statistics, and consequently, improve their detection ability. In this paper, motivated by the framework of empirical Bayesian and shrinkage estimation, we propose a new distribution family, the regularized $t$ distribution, and also demonstrate that most of existing test statistics based on Bayesian and shrinkage techniques fall into this new distribution family. We derive the density function of the regularized $t$ distribution and also thoroughly investigate its statistical properties. We further apply the regularized $t$ distribution to gene expression data analysis and compare its performance with the Bayeisan $t$-test, which is the most popular method in the "*limma*" package. Simulation studies and real data analysis lend further support to our proposed regularized $t$-test.

Apart from micro-array data analysis, the regularized $t$ distribution can also be applied to more modern type of data including, for example, the detection of DE genes in RNA-Seq data. Law et al. (2014), Kvam et al. (2012), and Zhang et al. (2019) considered the nature of discreteness of the RNA-Seq data. They performed a data transformation for the read counts in RNA-Seq after normalization by sequencing depth. Given that their transformed data is about normal, our proposed regularized $t$-test can still be directly used. In addition, we note that the negative binomial (NB) model-based method (Conesa et al., 2016) is another popular method for detecting DE genes in RNA-Seq data, which can also be traced back to the Bayesian $t$-test. To be more specific, we let $\phi_j$ be the dispersion parameter of the NB distribution for the $j$th gene, and $\theta_j$ be the corresponding mean expression level after normalization by sequencing depth. Then to test whether $\theta_j = 0$ for each gene, the authors provided a $t$-type statistic as

$$T_j^{(1)} = \frac{\widetilde{\theta}_j}{\breve{s}_j}, \tag{14}$$

where $\widetilde{\theta}_j$ is the sample estimate of $\theta_j$, and $\breve{s}_j^2 = a + b\phi_j$ is the variance of $\widetilde{\theta}_j$. Note also that $\breve{s}_j^2$ is a linear function of the dispersion parameter $\phi_j$. Hence, an improved estimate of $\phi_j$ may further improve the detection power of $T_j^{(1)}$.

In a similar vein as Bayesian $t$-test, many methods have been proposed to improve the estimation accuracy of $\phi_j$. For example, edgeR moderates the dispersion estimate for each gene toward a common estimate across all genes (McCarthy et al., 2012; Robinson et al., 2010). DESeq corrects the dispersion estimates that are too low through modeling of the dependence of the dispersion on the average expression strength over all samples (Anders & Huber, 2010). DESeq2 estimates the dispersions based on the empirical Bayesian method, which provides an automatic determination for the shrinkage intensities based on the available information in the data (Law et al., 2014; Pimentel et al., 2017). Nevertheless, similarly as the Bayesian $t$ statistic in (2), we note that the null distribution of $T_j^{(1)}$ in those methods were, once again, either approximated by a normal distribution, or approximated by permutation tests. When the sample sizes are small, the approximated null distributions may have a large deviation from the true distribution of $T_j^{(1)}$. In view of this, we expect that further research may also be needed to derive the exact distribution of $T_j^{(1)}$ for both theoretical interest and practical use.

## ACKNOWLEDGMENTS

The authors sincerely thank the editor, the associate editor, and one reviewer for their constructive comments that have led to a substantial improvement of this paper. Zongliang Hu's research was supported by National Natural Science Foundation of China (12001378), Guangdong Basic and Applied Basic Research Foundation (2019A1515110449), and Natural Science Foundation of Guangdong Province (2023A1515010027). Yiping Yang's research was supported by Chongqing Natural Science Foundation (cstc2021jcyj-msxmX0079) and Humanities and Social Sciences Program of Chongqing Education Commission (21SIGH118). Gaorong Li's research was supported by National Natural Science Foundation of China (12271046, 11871001, and 12131006). Tiejun Tong's research was supported by General Research Fund of Hong Kong (HKBU12303918 and HKBU12303421), Initiation Grant for Faculty Niche Research Areas (RC-FNRA-IG/20-21/SCI/03) of Hong Kong Baptist University, and National Natural Science Foundation of China (1207010822).

## ORCID

*Gaorong Li* https://orcid.org/0000-0002-1784-3472
*Tiejun Tong* https://orcid.org/0000-0003-0947-3990

## REFERENCES

Abramowitz, M., & Stegun, I. A. (1972). *Handbook of mathematical functions*. Dover Publications.

Alon, U., Barkai, N., Notterman, D., Gish, K., Ybarra, S., Mack, D., & Levine, A. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 6745–6750.

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*, R106.

Baldi, P., & Long, A. D. (2001). A Bayesian framework for the analysis of microarray expression data: regularized $t$-test and statistical inferences of gene changes. *Bioinformatics*, *17*, 509–519.

Conesa, A., Madrigal, P., Tarazona, S., & Zhang, X. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*, 13.

Cui, X., & Churchill, G. A. (2003). Statistical tests for differential expression in cDNA microarray experiments. *Genome Biology*, *4*, 210.

Cui, X., Hwang, J. G., Qiu, J., Blades, N. J., & Churchill, G. A. (2005). Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics*, *6*, 59–75.

Dutka, J. (1981). The incomplete beta function—A historical profile. *Archive for History of Exact Sciences*, *24*, 11–29.

Efron, B., Tibshirani, R., Storey, J. D., & Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association*, *96*, 1151–1160.

John, A. R. (2016). *datamicroarray: Collection of data sets for classification*. https://github.com/ramhiser/datamicroarray

Kvam, V. M., Liu, P., & Si, Y. (2012). A comparison of statistical methods for detecting differentially expressed genes from RNA-Seq data. *American Journal of Botany*, *99*, 248–256.

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*, R29.

McCarthy, D., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, *40*, 4288–4297.

Opgen-Rhein, R., & Strimmer, K. (2007). Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Statistical Applications in Genetics and Molecular Biology*, *6*, 9.

Patrick, E., Buckley, M., Lin, D. M., & Yang, Y. H. (2013). Improved moderation for gene-wise variance estimation in RNA-Seq via the exploitation of external information. *BMC Genomics*, *14*(Suppl 1), S9.

Phipson, B., Lee, S., Majewski, I. J., & Alexander, W. S. (2016). Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *The Annals of Applied Statistics*, *10*, 946–963.

Pimentel, H., Bray, N. L., Puente, S., Melsted, P., & Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods*, *14*, 687–690.

Praetz, P. D. (1972). The distribution of share price changes. *Journal of Business*, *45*, 49–55.

Prosser, R. T. (1994). On the Kummer solutions of the hypergeometric equation. *The American Mathematical Monthly*, *101*, 535–543.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*, e47.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*, 139–140.

Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiment. *Statistical Applications in Genetics and Molecular Biology*, *3*, 3.

Storey, J. D., & Tibshirani, R. (2003). *SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays*. In G. Parmigiani, E. S. Garrett, R. A. Irizarry, & S. L. Zeger (Eds.), *The analysis of gene expression data: methods and software*. Springer.

Student. (1908). The probable error of a mean. *Biometrika*, *6*, 1–25.

Tong, T., Wang, C., & Wang, Y. (2014). Estimation of variances and covariances for high-dimensional data: a selective review. *WIREs Computational Statistics*, *6*, 255–264.

Tong, T., & Wang, Y. (2007). Optimal shrinkage estimation of variances with applications to microarray data analysis. *Journal of the American Statistical Association*, *102*, 113–122.

Tusher, V. G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 5116–5121.

Zhang, Z., Yu, D., Seo, M., Hersh, C. P., Weiss, S. T., & Qiu, W. (2019). Novel data transformations for RNA-seq differential expression analysis. *Scientific Reports*, *9*, 4820–4831.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Hu, Z., Yang, Y., Li, G., & Tong, T. (2023). Regularized *t* distribution: definition, properties, and applications. *Scandinavian Journal of Statistics*, *50*(4), 1884–1900. https://doi.org/10.1111/sjos.12655