

METHODOLOGY ARTICLE

Open Access



Sample size calculation while controlling false discovery rate for differential expression analysis with RNA-sequencing experiments

Ran Bi and Peng Liu*

Abstract

Background: RNA-Sequencing (RNA-seq) experiments have been popularly applied to transcriptome studies in recent years. Such experiments are still relatively costly. As a result, RNA-seq experiments often employ a small number of replicates. Power analysis and sample size calculation are challenging in the context of differential expression analysis with RNA-seq data. One challenge is that there are no closed-form formulae to calculate power for the popularly applied tests for differential expression analysis. In addition, false discovery rate (FDR), instead of family-wise type I error rate, is controlled for the multiple testing error in RNA-seq data analysis. So far, there are very few proposals on sample size calculation for RNA-seq experiments.

Results: In this paper, we propose a procedure for sample size calculation while controlling FDR for RNA-seq experimental design. Our procedure is based on the weighted linear model analysis facilitated by the *voom* method which has been shown to have competitive performance in terms of power and FDR control for RNA-seq differential expression analysis. We derive a method that approximates the average power across the differentially expressed genes, and then calculate the sample size to achieve a desired average power while controlling FDR. Simulation results demonstrate that the actual power of several popularly applied tests for differential expression is achieved and is close to the desired power for RNA-seq data with sample size calculated based on our method.

Conclusions: Our proposed method provides an efficient algorithm to calculate sample size while controlling FDR for RNA-seq experimental design. We also provide an R package *ssizeRNA* that implements our proposed method and can be downloaded from the Comprehensive R Archive Network (<http://cran.r-project.org>).

Keywords: RNA-seq, FDR, Experimental design, Sample size calculation, Power analysis

Background

During the past decade, next generation sequencing (NGS) technology has revolutionized genomic studies, and tremendous development has been made in terms of throughput, scalability, speed and sequencing cost. RNA-Sequencing (RNA-seq), also called Whole Transcriptome Shotgun Sequencing (WTSS), is a technology that uses the capabilities of NGS to study the entire transcriptome. Compared with microarray technologies that used to be the major tool for transcriptome studies, RNA-seq technologies have several advantages including a larger dynamic range of expression levels, less noise, higher

throughput, and more power to detect gene fusions, single nucleotide variants and novel transcripts. Hence, RNA-seq technologies have been popularly applied in transcriptomic studies.

In a typical RNA-seq experiment, messenger RNA (mRNA) molecules are extracted from samples, fragmented, and reverse transcribed to double-stranded complementary DNA (cDNA). The cDNA fragments are then sequenced on a high-throughput platform, such as HiSeq by Illumina or SOLiD by Applied Biosystems. After sequencing, millions of DNA fragment sequences, called reads, are recorded and aligned to a reference genome. The number of reads mapped to each gene measures the expression level for that gene. Thus, RNA-seq provides

*Correspondence: pliu@iastate.edu
Department of Statistics, Iowa State University, Snedecor Hall, Ames, Iowa 50011, USA

discrete count data serving as measurements of mRNA expression levels, which is different from the fluorescence intensity measurements from microarray technologies that have been considered as continuous variables after transformation. As a result of high frequency of low integers, the statistical methods developed for analyzing microarray data are not directly applicable for RNA-seq data.

In the statistical analysis of RNA-seq data, identifying differentially expressed (DE) genes across treatments or conditions is a major step or main focus. A gene is considered to be DE across treatments or conditions if the mean read counts differ across treatment groups. Otherwise, we say the gene is equivalently expressed (EE). Many statistical methods have been proposed for the detection of DE genes with RNA-seq data. Some popular methods, including *edgeR* [1–4], *DESeq* [5] and *DESeq2* [6], are based on the negative binomial (NB) distribution. *QuasiSeq* [7] presented quasi-likelihood methods with shrunken dispersion estimates. A more recently proposed method by the Smyth group [8] works with log-transformed count data and captures the mean-variance relationship of the log-count data through a precision weight for each observation (using a function called *voom* in their R package) and then applies the *limma* method [9] for differential expression analysis.

Due to the genetic complexity and high-dimensionality of the resulting datasets, RNA-seq experiments require complicated bioinformatic and statistical analysis in addition to the cost of experimental materials and sequencing. Many experiments only employ a small number of replicates, in which cases the power of statistical inference is limited. However, if the sample size is too large (which is rare), it is also a waste of experimental materials and manpower. For these reasons, one of the principal questions in designing an RNA-seq experiment is: how many biological replicates should be used to achieve a desired power? In other words, how large of the sample size do we need?

To answer this question, we need to determine a sample size that is required to achieve a desired power while controlling an appropriate error rate. When calculating sample size for a single test, type I error rate is commonly used. Fang and Cui [10] discussed a sample size formula for a single gene based on likelihood ratio test or Wald test. Hart et al. [11] and their associated R package *RNASeqPower* [12] proposed a sample size calculation method for any single gene based on score test while controlling type I error rate. However, for RNA-seq data analysis, tens of thousands of genes are simultaneously tested for differential expression, which requires the correction of multiple testing error, and false discovery rate (FDR) [13] has been the choice of error criterion in RNA-seq data analysis.

Several sample size calculation methods while controlling FDR have been proposed in microarray experiments. For example, Liu and Hwang [14] developed a method to calculate sample size given a desired power and a controlled level of FDR by finding the rejection region for the test procedure and hence power for each sample size. Hereafter, we call this sample size calculation method the LH method. Orr and Liu [15] assembled the *ssize.fdr* R package which implements the LH method.

However, sample size calculation for RNA-seq data analysis while controlling FDR is underdeveloped. Some earlier studies performed sample size and power estimation for RNA-seq experiments under Poisson distribution [16–18], but the additional biological variation across RNA-seq samples yields overdispersion, which means the equal mean-variance relationship for the Poisson distribution does not adapt to the variability present in RNA-seq data. To account for overdispersion, the negative binomial distribution is more flexible to use. Li et al. [19] proposed a sample size determination method while controlling FDR based on the exact test implemented in *edgeR* that tests for genes differentially expressed between two treatments or conditions. This method calculates a sample size based on the minimum fold change of DE genes, the minimum average read counts of DE genes in the control group, and the maximum dispersion of DE genes under negative binomial models. As expected, such a method would be very conservative and not practically informative. The *RnaSeqSampleSize* R package [20] provides an estimation of sample size based on single read count and dispersion which implements Li et al.'s method. Also, instead of using the minimum average read counts and the maximum dispersion, *RnaSeqSampleSize* gives an estimation of sample size based on the read count and dispersion distributions estimated from real data, together with the minimum fold change, which is much better than Li et al.'s method, but would still be conservative due to the usage of the minimum fold change. The LH method is applicable as long as we can compute the power and type I error rate given a rejection region. However, there are no closed-form formulae for power for the popularly applied NB based methods. Then we have to rely on a lot of simulation to figure out quantities such as power and type I error rate for each sample size and each simulation setting [10]. Ching et al. [21] provided a power analysis tool that calculates the power for a given budget constraint for each size of samples, and then determined the sample size for a desired power. Wu et al. [22] introduced the concepts of stratified targeted power and false discovery cost, and estimated sample size by the evaluation of statistical power over a range of sample sizes based on simulation studies. Both Ching et al. and Wu et al.'s methods are simulation-based, thus we need to do a lot

of simulations for power assessment for each sample size, which is time-consuming.

In this paper, we propose a much less computationally intensive method, which only demands one-time simulation, for sample size calculation in designing RNA-seq experiments. First, we use the *voom* method to model the mean-variance relationship of the log-count data of RNA-seq and produce a precision weight for each observation. Second, based on the normalized log-counts and associated precision weights, we estimate the distribution of weighted residual standard deviation of expression levels. Then for two-sample experiments, we derive a formula of the *t* test statistic in the weighted least squares setting and estimate the distribution of effect sizes for differential expression. Next, we apply the LH method to calculate the required sample size for a given desired power and a controlled FDR level. Our simulation demonstrates that the desired power is reached for data with the sample size calculated from our method for several popular tests for differential expression.

The article is organized as follows. The ‘Methods’ section describes our proposed method illustrated with the two-sample *t*-test. In the ‘Results and discussion’ section, we present four simulation studies based on either negative binomial distributions or real RNA-seq dataset, and our method provide reliable sample sizes for all simulation studies. The ‘Conclusions’ section discusses our results and some future work.

Methods

In this section, we first review the *voom* method [8] and the LH method of sample size calculation. Then, we introduce our approach for calculating sample size while controlling FDR in designing RNA-seq experiments.

The *voom* method

Suppose that an RNA-seq experiment includes a total of N samples. Each sample has been sequenced, and the resulting reads are aligned with a reference genome. The number of reads mapped to each reference gene is recorded. The RNA-seq data then consist of a matrix of read counts r_{gij} , where $g = 1, 2, \dots, G$ denotes gene g , $i = 1, 2$ denotes group where $i = 1$ is for the control group and $i = 2$ is for the treatment group, and $j = 1, 2, \dots, n_i$ denotes replicates in each group with $N = n_1 + n_2$. The idea of the *voom* method proposed by Law et al. [8] is to use precision weights to account for the mean-variance relationship and apply weighted least square analysis to RNA-seq data.

The method of *voom* starts from transforming the RNA-seq count data to the log-counts per million (log-cpm) value calculated by

$$y_{gij} = \log_2 \left(\frac{r_{gij} + 0.5}{R_{ij} + 1} \times 10^6 \right),$$

where $R_{ij} = \sum_{g=1}^G r_{gij}$ is the library size for the i th treatment and j th replicate. As has been done in [9], Law et al. then fit a linear model to the transformed data according to the experimental design. For each gene g , the following linear model

$$\mathbf{y}_g = X\boldsymbol{\beta}_g + \boldsymbol{\varepsilon}_g$$

is fitted to $\mathbf{y}_g = (y_{g11}, \dots, y_{g1n_1}, y_{g21}, \dots, y_{g2n_2})'$, the vector of log-cpm values, where X is the design matrix with rows \mathbf{x}_{ij}^T , $\boldsymbol{\beta}_g$ is a vector of parameters that may be parameterized to include \log_2 -fold changes between experimental conditions, and $\boldsymbol{\varepsilon}_g$ is the error term with $E(\boldsymbol{\varepsilon}_g) = \mathbf{0}$.

Assuming that $E(y_{gij}) = \mu_{gij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}_g$, then by ordinary least squares, the above linear model is fitted for each gene g , which yields regression coefficient estimates $\hat{\boldsymbol{\beta}}_g$, fitted values $\hat{\mu}_{gij} = \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_g$, residual standard deviations η_g and fitted \log_2 -read counts

$$\hat{l}_{gij} = \hat{\mu}_{gij} + \log_2(R_{ij} + 1) - \log_2(10^6).$$

To obtain a smooth mean-variance trend, Law et al. fit a LOWESS curve to the square root of residual standard deviations $\eta_g^{1/2}$ as a function of average log-counts \tilde{r}_g , where $\tilde{r}_g = \bar{y}_g + \log_2(\bar{R} + 1) - \log_2(10^6)$ with \bar{y}_g being the average log-cpm value for each gene g and \bar{R} being the geometric mean of library sizes. Then for each observation y_{gij} , the predicted square root residual standard deviation $\hat{\eta}_{gij}^{1/2}$ is obtained to be the LOWESS fitted value corresponding to \hat{l}_{gij} .

Finally, the *voom* precision weights are defined as the inverse variances $w_{gij} = \frac{1}{\hat{\eta}_{gij}^2}$. Law et al. recommended analyzing the log-cpm data with weighted least squares, and the weights (w_{gij}) are used to account for the mean-variance relationship in the log-cpm values. Assuming normal distribution for residual errors ($\boldsymbol{\varepsilon}_g$), methods such as *t*-tests or moderated *t*-tests can then be applied for differential expression analysis.

The LH method of sample size calculation

In genomic studies, we simultaneously test a large number of hypotheses, each relating to a gene. Hence, multiple testing is commonly used in the analysis. Assume there are G genes in total and each gene is tested for the significance of differential expression. Table 1 summarizes the various outcomes that occur when testing G hypotheses, where V is the number of false positives, R is the number

Table 1 Outcomes when testing G hypotheses

	Accept null	Reject null	Total
True nulls	U	V	$\pi_0 G$
False nulls	T	S	$(1 - \pi_0)G$
Total	W	R	G

of rejections among the G tests, and π_0 is the proportion of non-differentially expressed genes.

False discovery rate (FDR), defined by Benjamini and Hochberg [13], is the expected proportion of false positives among the rejected hypothesis:

$$FDR = E\left(\frac{V}{R} \mid R > 0\right) Pr(R > 0),$$

while positive FDR (pFDR), proposed by Storey [23], is defined to be

$$pFDR = E\left(\frac{V}{R} \mid R > 0\right).$$

Both FDR and pFDR are widely used error rates to control in multiple testing encountered in genomic studies. In RNA-seq experiments, most often we end up detecting DE genes, i.e. $R > 0$. Hence, in this paper, we do not differentiate between FDR and pFDR.

Liu and Hwang [14] proposed a method for a quick sample size calculation for microarray experiments while controlling FDR. Let $H = 0$ represent no differential expression (null hypothesis is true) and $H = 1$ represent differential expression (null hypothesis is false). Based on the definition of pFDR and assumptions in [23] (all tests are identical, independent and Bernoulli distributed with $Pr(H = 0) = \pi_0$, where π_0 is the proportion of EE genes), they derived that

$$\frac{\alpha}{1 - \alpha} \frac{1 - \pi_0}{\pi_0} \geq \frac{Pr(T \in \Gamma | H = 0)}{Pr(T \in \Gamma | H = 1)}, \tag{1}$$

where α is the controlled level of FDR, T denotes the test statistic and Γ denotes the rejection region of the test. Then for each comparison, the LH method calculates the sample size as follows. First, for a fixed proportion of non-differentially expressed genes, π_0 , and the level of FDR to control, α , they find a rejection region Γ that satisfies (1) for each sample size. Then for the selected rejection region Γ for each sample size, the power is calculated by $Pr(T \in \Gamma | H = 1)$. According to the desired power, a sample size is determined.

The rejection region depends on the test applied for differential expression, and the method based on (1) can be applied to any multiple testing procedure where the same rejection region is used. This LH method can be implemented using an R package, *ssize.fdr*, developed by Orr and Liu [15], and applied for designing one-sample, two-sample, or multi-sample microarray experiments. The method would be applicable to RNA-seq experiments if we can calculate power and type I error rate given a rejection region.

Proposed method for RNA-seq experiments with two-sample comparison

For the popularly applied tests in RNA-seq differential expression analysis such as *edgeR* and *DESeq*, there are

no closed-form expressions to calculate the two quantities $Pr(T \in \Gamma | H = 0)$ and $Pr(T \in \Gamma | H = 1)$. Hence, the LH method cannot be directly applicable to these methods. However, the recently proposed *voom* and *limma* analysis for RNA-seq data [8, 24] is based on weighted linear models and we can obtain tractable formulae for power and type I error rate. In this paper, our idea is to derive formulae to calculate power and type I error rate based on *voom* and weighted linear model analysis, and then apply the LH method for sample size calculation. We will use two-sample t -tests to illustrate our idea. Similar methods can be derived for other designs such as paired-sample or multiple treatments comparison.

Suppose our interest is to identify the differentially expressed (DE) genes between a treatment and a control group. Assuming that for gene g , group i and replicates j , we observe the RNA-seq data read counts r_{gij} , where the mean for gene g in group i is $\lambda_{gij} = d_{ij}\gamma_{gi}$. Here, d_{ij} stands for a normalization factor or effective library size that adjusts the sequencing depth for sample j in group i , γ_{gi} stands for the normalized mean expression level of gene g in group i . Then for each gene g , to test for differential expression means to test the hypothesis:

$$H_0^g : \gamma_{g1} = \gamma_{g2} \quad vs. \quad H_1^g : \gamma_{g1} \neq \gamma_{g2}.$$

As reviewed in the first part of the ‘Methods’ section, when applying the *voom* method, the RNA-seq read counts r_{gij} are transformed to log-cpm values y_{gij} with associated weights w_{gij} and mean μ_{gi} for each sample j in group i . With this parameterization, testing for DE means testing

$$H_0^g : \mu_{g1} = \mu_{g2} \quad vs. \quad H_1^g : \mu_{g1} \neq \mu_{g2},$$

where μ_{g1} and μ_{g2} are the expectation of log-cpm values of g th gene for control and treatment group, respectively.

For each individual gene g , the weighted linear model

$$y_g = X\beta_g + \sigma_g W_g^{-\frac{1}{2}} \epsilon$$

can be fitted to log-cpm values

$$y_g = (y_{g11}, \dots, y_{g1n_1}, y_{g21}, \dots, y_{g2n_2})$$

with design matrix

$$X = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix},$$

coefficients vector

$$\beta_g = \begin{pmatrix} \beta_{g1} \\ \beta_{g2} \end{pmatrix},$$

unknown gene-specific standard deviation σ_g , and associated *voom* precision weights

$$W_g = \text{diag}(w_{g11}, \dots, w_{g1n_1}, w_{g21}, \dots, w_{g2n_2}).$$

Assuming $\epsilon \sim \text{MVN}(\mathbf{0}, I_{n_1+n_2})$, where MVN stands for multivariate normal distribution, the *t*-test statistic for gene *g* is

$$T_g = \frac{\hat{\beta}_{g2}}{S.E.(\hat{\beta}_{g2})}, \tag{2}$$

where the estimated \log_2 -fold change between treatment and control group $\hat{\beta}_{g2}$ and its standard error $S.E.(\hat{\beta}_{g2})$ could be obtained through weighted least squares estimation.

To make the *t*-test based method more straightforward to apply, we reparameterize the formula (2) to

$$T_g = \frac{\Delta_g}{s_g \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{3}$$

where

$$s_g = \sqrt{\frac{(\mathbf{y}_g - X\beta_g)' W_g (\mathbf{y}_g - X\beta_g)}{n_1 + n_2 - 2}}$$

can be viewed as the pooled sample standard deviation, which is an estimator of σ_g , and

$$\Delta_g \equiv \hat{\beta}_{g2} \sqrt{\frac{\bar{w}_{g1} \cdot \bar{w}_{g2}}{\bar{w}_{g\cdot}}}. \tag{4}$$

can be viewed as the scaled effect size which is defined by weighted mean difference of log-cpm values. Here, $\bar{w}_{g1\cdot} = \frac{1}{n_1} \sum_{j=1}^{n_1} w_{g1j}$, $\bar{w}_{g2\cdot} = \frac{1}{n_2} \sum_{j=1}^{n_2} w_{g2j}$ and $\bar{w}_{g\cdot} = \frac{1}{n_1+n_2} \sum_{i=1}^2 \sum_{j=1}^{n_i} w_{gij}$. Details of the derivation for (3) is provided in the Appendix A.

After generating the effect size Δ_g , and the standard deviation σ_g for each gene *g*, we could assume, as in [14], that the effect size follows a normal distribution

$$\Delta_g \sim N(\mu_\Delta, \sigma_\Delta^2),$$

and the variance of log-cpm values for each gene follows an inverse gamma distribution

$$\sigma_g^2 \sim \text{Inv} - \text{Gamma}(a, b)$$

with mean $\frac{b}{a-1}$. Then we apply the LH method to calculate the optimal sample size given desired power and controlled FDR level. See Appendix B for a brief review of the calculations in the LH method involving in choosing the rejection region Γ satisfying formula (1).

Our proposed method requires the estimation of hyperparameters μ_Δ , σ_Δ , *a*, and *b*. If a relatively large pilot dataset is available, these parameters can be estimated based on the pilot data. Otherwise, we can simulate data to obtain the values for these hyperparameters. It has been

shown that the NB model fits real RNA-seq data well [5]. In addition, many popularly applied tests for differential expression analysis of RNA-seq data are based on NB models. Hence, we suggest to simulate data according to NB models, and then use such simulated data to obtain the estimates of μ_Δ , σ_Δ , *a*, and *b*, which are then used to calculate sample size. We outline our proposed procedure for sample size calculation as follows:

1. For a given RNA-seq experiment, specify the following parameters:
G: total number of genes for testing;
 π_0 : proportion of non-DE genes;
 α : FDR level to control;
pow: desired average power to achieve;
 λ_g : average read count for gene $g = 1, \dots, G$ in control group (without loss of generality, we assume that the normalization factors d_{ij} are equal to 1 for all samples);
 ϕ_g : dispersion parameter for gene *g*;
 δ_g : fold change for gene *g*.
 Note that λ_g and ϕ_g could be estimated from real data using methods such as *edgeR*.
2. Simulate RNA-seq read count data from a NB distribution with given parameters in step 1.
3. Use the *voom* and *limma* method to obtain the log-cpm value and the associated precision weight for each count, and then estimate effect size Δ_g according to (4) for each gene *g* and parameters *a*, *b* for the prior of σ_g .
4. Estimate μ_Δ and σ_Δ by fitting

$$\Delta_g \sim N(\mu_\Delta, \sigma_\Delta^2).$$
5. Use the LH method to determine the sample size *n* to achieve desired power and controlled FDR level.

Results and discussion

In this section, we present four simulation studies to evaluate our proposed method for sample size calculation for RNA-seq experiments. In the first three simulation studies, we set the total number of genes to be $G = 10,000$ and the desired average power to be 80 %. The last simulation is real data-based.

Simulation 1. Same set of parameters

We start from the simplest simulation setting where all genes share the same set of parameters for the NB distribution. Although such cases are unrealistic, they allow the method of Li et al. [19] to perform best because this method uses a single set of NB parameters (mean, dispersion, fold change) when calculating sample size. Hence, we use this simulation setting to study the performance of our method and compare it to the method of Li et al. We refer to the parameter settings from Table 1 in [19], and

compare the resulting sample size and power calculated by both Li et al.'s method and our proposed method.

In the main manuscript, we present results for one of those parameter settings as an example: the proportion of non-DE genes $\pi_0 = 0.99$, the mean read counts for control group $\lambda = 5$ with normalization factors $d_{ij} = 1$, dispersion parameter $\phi = 0.1$, FDR controlling at level 0.05, and fold change $\delta = 2$ for differentially expressed genes. Suppose r_{gij} denotes the read count for gene g , group i and replicate $j = 1, 2, \dots, n_i$ in each group with $n_1 = n_2 = n$. Then, for EE genes, both r_{g1j} and r_{g2j} were drawn from $NB(5, 0.1)$; for DE genes, r_{g1j} were drawn from $NB(5, 0.1)$ and r_{g2j} were drawn from $NB(10, 0.1)$ or $NB(2.5, 0.1)$.

After setting these simulation parameters in step 1, we follow steps 2-4 to simulate data and obtain the values of hyperparameters. To investigate the effect of this simulation step, we tried different sizes of simulated data, $m = 50, 100, 200, 500, 1000$, where m is the sample size for each group in step 2 of our procedure. For each m , we generated read counts r_{g1j} (control group) and r_{g2j} (treatment group) from independent NB distributions for every gene g and sample $j, g = 1, \dots, G, j = 1, \dots, m$. After using *voom* and *lmFit* in the R package *limma* [9] to produce weights w_{gij} for each observation, we then obtained effect size Δ_g for each gene and parameters a, b for the prior distribution of σ_g^2 . The fitted inverse gamma distributions of σ_g^2 for each m are shown as in Fig. 1, with vertical lines indicating the modes. It seems that the mode doesn't change much, and the distribution of σ_g^2 shrinks towards the center as sample size gets larger.

After obtaining the fitted parameters, we calculated sample size according to our proposed method described in the third part of the 'Methods' section to achieve a desired power of 80 %. We then simulated data according to each calculated sample size and checked whether the desired power was achieved. In Table 2, the first three columns listed our simulation results corresponding to

this simulation setting. As m increased from 50 to 100 to 1000, the calculated sample size dropped from 35 to 34 and 32, respectively. This decrease is expected because the parameters were estimated more precisely with larger m . For example, the distribution of σ_g^2 shrank as m increased as shown in Fig. 1. The effect on the resulting sample size is not big, at most with a difference of 3 (35 vs. 32).

We now choose a sample size $n = 32$ and demonstrate this sample size indeed reaches the desired power 0.8. At $n = 32$, we simulated 100 datasets and performed several popularly applied tests such as the *edgeR* exact test, the *voom* and *limma* method, *DESeq*, *DESeq2* and *QuasiSeq* using the corresponding R packages. Desired power (0.8) was achieved for all testing methods when controlling FDR at 0.05 using q -value procedure [25], and the observed FDR was controlled successfully under all the five methods. The results are shown in Fig. 2. For the *voom* and *limma* pipeline method, the observed power curves while FDR was controlled using the Benjamini and Hochberg's method [13] and the q -value procedure [25] and the power curve based on our calculation are shown in Fig. 3. The observed power was obtained by averaging actual power over 100 simulated datasets for each sample size. The observed power and the power calculated by our method are close with our calculation being a little conservative. Hence, our proposed method provides an accurate estimate of power, and the sample size calculated by our method is reliable.

Finally we would like to compare our method with other existing sample size calculation methods, including Li et al.'s approach [19, 20] and Wu et al.'s approach [22]. Li et al. proposed to calculate the sample size by "using a common $\rho^* = \operatorname{argmin}_{g \in M_1} \{|\log_2(\rho_g)|\}$ minimum fold change", where ρ_g in their paper denotes the fold change and is equivalent to δ_g in this paper. However, we found that the direction of fold change does matter when applying their code. If we set $\rho_g = 2$, the sample size calculated by their method is $n = 20$, as presented in their Table 1. The plot of average power vs. nominal FDR for their method is shown in Fig. 4, from which we notice that the desired power (0.8) is not achieved at sample size $n = 20$ when controlling FDR at 0.05. In fact, the observed power is 0.6166 when using the *edgeR* exact test based on which they derived their method. When applying the *voom* and *limma* pipeline, the observed power is 0.4608 for sample size 20. If we set $\rho_g = 0.5$, then the sample size will be 32, same as our proposed method, and we get power of 0.8988 using the *edgeR* exact test and 0.8149 using the *voom* and *limma* pipeline for differential expression analysis. Wu et al. (*PROPER*) provided a simulation-based power evaluation tool, which requires a lot of simulations to assess the power for each sample size. Table 3 presents the computation time needed for the calculation. It took *PROPER* 6.5 hours to get the resulting sample size while

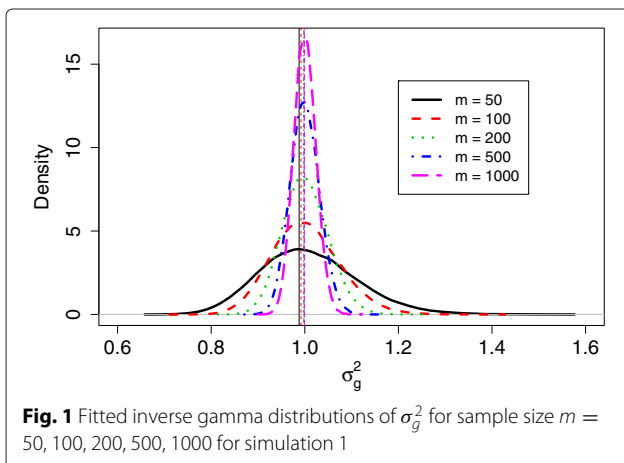


Table 2 Sample size and anticipated average power calculated by our method and observed average power by the *voom* and *limma* pipeline while controlling FDR using *q*-value procedure based on parameters at different *m* for three simulation settings

m	Simulation 1			Simulation 2			Simulation 3				
	Our Method			Our Method			<i>RnaSeqSampleSize</i>		Our Method		
	Sample size	Anticipated power	Observed power	Sample size	Anticipated power	Observed power	Sample size	Estimated power	Sample size	Anticipated power	Observed power
50	35	0.802	0.858	13	0.810	0.876	9	0.780	22	0.800	0.801
100	34	0.814	0.846	13	0.814	0.876	9	0.724	22	0.803	0.801
200	34	0.815	0.846	13	0.823	0.876	9	0.765	22	0.804	0.801
500	33	0.817	0.833	13	0.827	0.876	9	0.769	22	0.805	0.801
1000	32	0.800	0.804	13	0.826	0.876	9	0.764	22	0.805	0.801

We also present the comparison between our method and *RnaSeqSampleSize* R package for simulation 2, where the right two columns are sample size and power calculated by the *RnaSeqSampleSize* R package

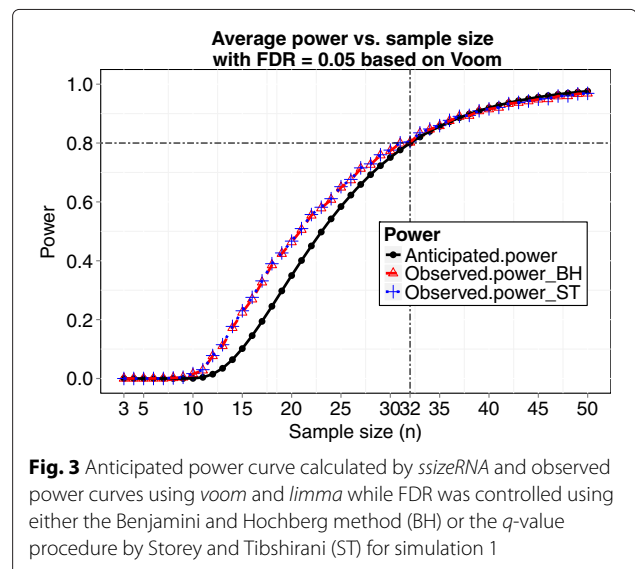
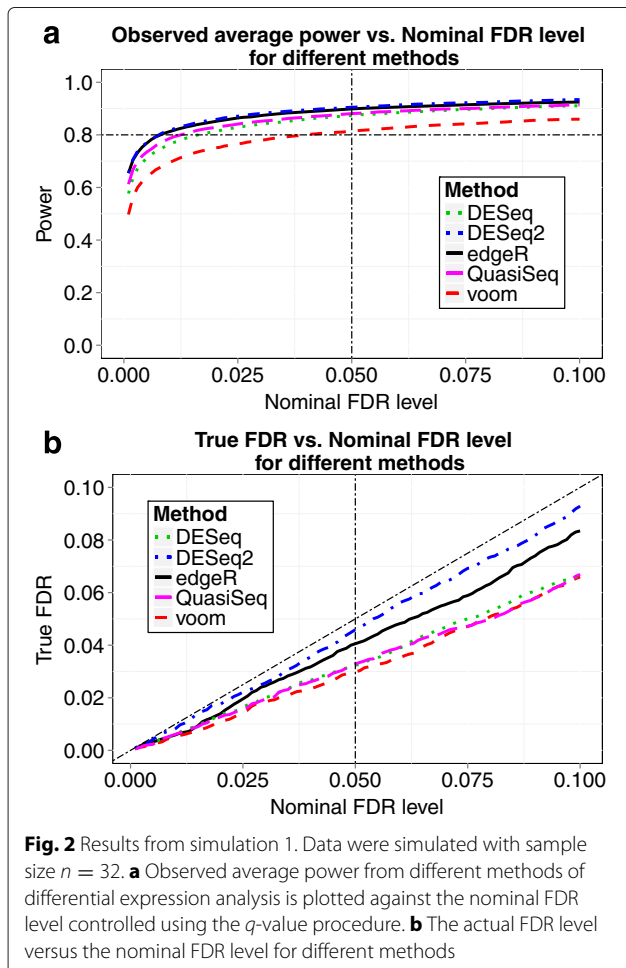
the other two methods only needed seconds. *PROPER* is more than 1,300 fold time-consuming than our proposed method. The resulting sample size from *PROPER* is 25, less than our proposed method. This is because *PROPER* is based on *edgeR* exact test, which tends to be more powerful than the *voom* and *limma* pipeline.

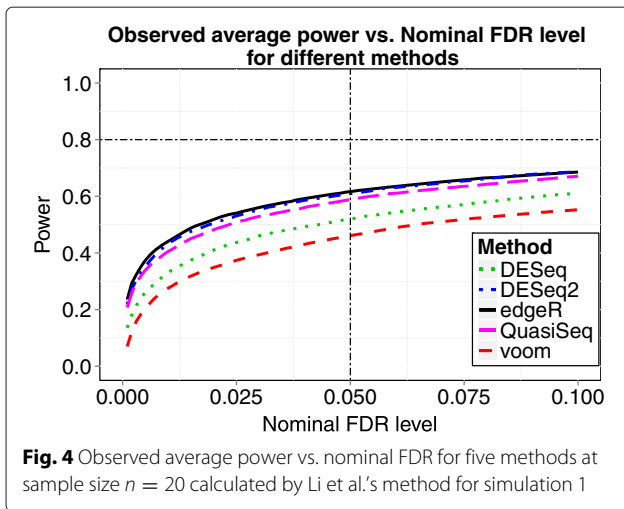
Results for other parameter settings under *m* = 200 are presented in the Additional file 1, with Li et al.'s results in the first row, and our results in the second row.

Simulation 2. Gene-specific mean and dispersion with fixed fold change

In the second simulation setting, we used a real RNA-seq dataset to generate gene-specific mean and dispersion parameters. A maize dataset was obtained from a study by Tausta et al. [26], who compared gene expression between bundle sheath and mesophyll cells of corn plants.

Similar to simulation 1, we generated 10,000 genes from $NB(\lambda_g, \phi_g)$, with fold change $\delta = 2$ for DE genes, λ_g and ϕ_g from the means and dispersions estimated for each gene in the maize dataset. For EE genes, both r_{g1j} and r_{g2j} were drawn from $NB(\lambda_g, \phi_g)$; for DE genes, r_{g1j} were drawn from $NB(\lambda_g, \phi_g)$ and r_{g2j} were drawn from $NB(2\lambda_g, \phi_g)$ or $NB(0.5\lambda_g, \phi_g)$. The proportion of non-DE genes was $\pi_0 = 0.8$.





The fitted inverse gamma distributions of σ_g^2 for $m = 50$ and 1000 are very similar, as shown as in Fig. 5, where vertical lines indicate the modes. The middle three columns in Table 2 give the sample size and average power calculated by our *ssizeRNA* package. As shown in Table 2, the resulting sample sizes are all 13 when m ranges from 50 to 1000. This is expected because Fig. 5 indicates that the estimated distributions of σ_g^2 are very close using different m values for this dataset.

At $n = 13$, we checked the plots of average power vs. nominal FDR and true FDR vs. nominal FDR, and the results were similar to those obtained in simulation 1. More specifically, the desired power (0.8) was achieved, and FDR was controlled successfully. Actually, the desired power can be reached at sample size $n = 11$. Figure 6(a) gives the power curve calculated by our method based on hyperparameters estimated at $m = 1000$ together with observed power curves with FDR controlled by the Benjamini and Hochberg's method and the q -value procedure, respectively. The anticipated power curve based on $m = 1000$ is close to the other two observed power curves.

The *RnaSeqSampleSize* R package [20] could give an estimation of sample size and power by prior real data.

They first use user-specified number of genes to estimate the gene read count and dispersion distribution, then *sample_size_distribution* and *est_power_distribution* functions will be used to determine sample size and actual power. When we used the same real dataset as our simulation setting 2, the sample size calculated by their method was 7, with actual power 0.774, which did not reach the desired power 0.8. We also tried to apply their method using our simulated data (with different m), the resulting sample size is larger ($n = 9$). The power estimated by their method at $n = 9$ are shown in Table 2, and all their estimated power were actually smaller than 0.8. *PROPER* started from an estimation of mean and dispersion parameters, which is similar to our method. The sample size calculated by their method is 10, with power 0.804 based on DE detection method *edgeR*. The comparison results of our proposed method and these three approaches are shown in the middle three columns of Table 3. Still, *PROPER* is much more time-consuming than the other two methods.

Simulation 3. Gene-specific mean and dispersion with different fold change

In this simulation, the setting is the same as the second simulation study, except that the fold change δ_g was simulated from a log-normal distribution for differentially expressed genes. For EE genes, both r_{g1j} and r_{g2j} were drawn from $NB(\lambda_g, \phi_g)$; for DE genes, r_{g1j} were drawn from $NB(\lambda_g, \phi_g)$ and r_{g2j} were drawn from $NB(\lambda_g \delta_g, \phi_g)$ or $NB(\lambda_g / \delta_g, \phi_g)$ where

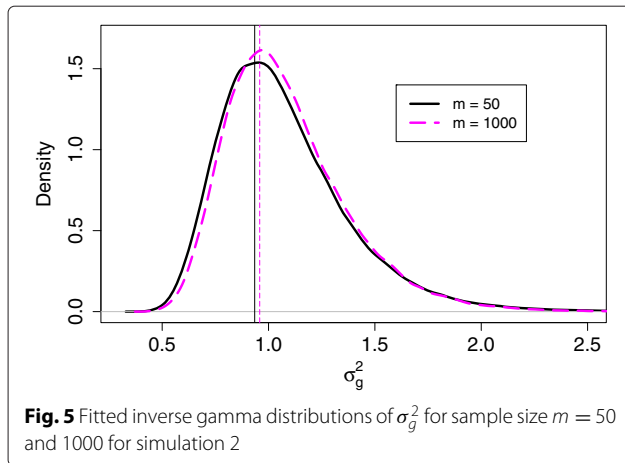
$$\delta_g \sim \log - normal(\log(2), 0.5\log(2)).$$

The last three columns in Table 2 give the sample size and power calculated by our method. As in simulation 2, varying the size of simulated data (m) did not result in different sample sizes. Anticipated and observed power curves are presented in Fig. 6(b), from which we notice that the three curves are almost indistinguishable after power reaches 60 %. This more realistic simulation demonstrates that our proposed method provides accurate power and sample size.

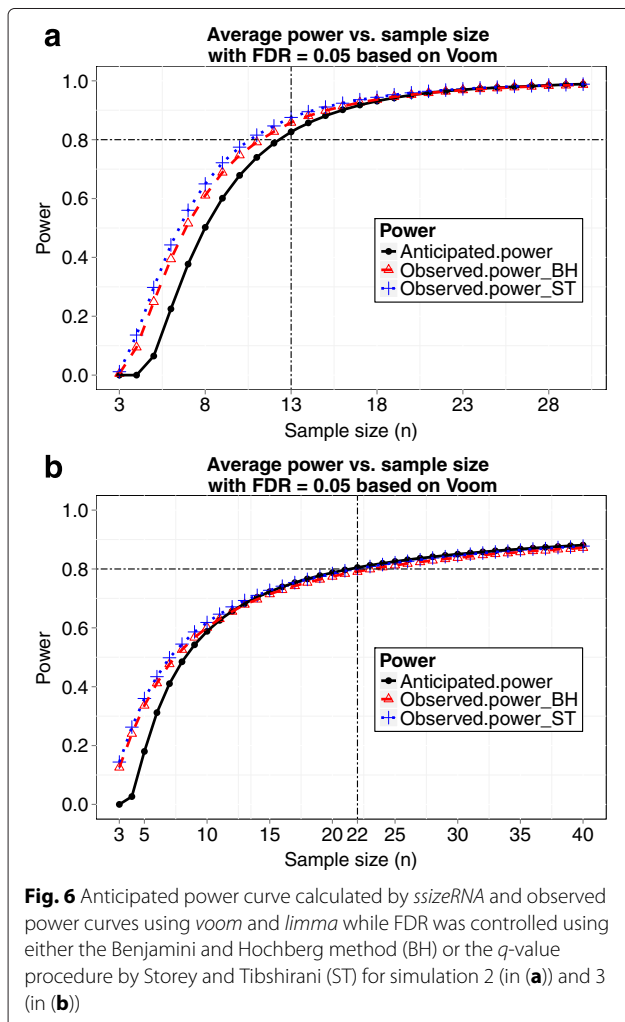
Table 3 Comparison of sample size calculation methods, including the proposed method in this paper, Zhao et al.'s approach (*RnaSeqSampleSize*) and Wu et al.'s approach (*PROPER*).

Method	Simulation 1			Simulation 2			Simulation 3		
	Sample size	Power	Computation time	Sample size	Power	Computation time	Sample size	Power	Computation time
Our method	34	0.815	17.3 sec	13	0.823	16.8 sec	22	0.804	15.5 sec
<i>RnaSeqSampleSize</i>	32	0.810	0.3 sec	9	0.765	54.6 sec	74	0.742	82.8 sec
<i>PROPER</i>	25	0.806	6.5 h	10	0.804	1.5 h	19	0.805	3.5 h

Results determined by our method were based on parameters estimated at $m = 200$. Power was evaluated based on the *voom* and *limma* pipeline for our method, while *edgeR* for *RnaSeqSampleSize* and *PROPER*. The computation time for each simulation was calculated on a MacAir laptop with 1.3 GHz i7 CPU and 4GB RAM



We also applied *RnaSeqSampleSize* to this simulation setting. Since their method is based on minimum fold change, such results will be conservative due to the variability of fold change, especially as in this case, the minimum fold change is close to 1. When we used the 10th



percentile of fold change of DE genes as the “minimum” fold change, the sample size calculated by their method was 74, which is still much larger than what we actually need, but the power calculated by their method based on the “minimum” fold change was less than the desired power 0.8. *PROPER* gave a result of sample size 19 with power 0.805 based on DE detection method *edgeR*. The comparison results of our proposed method and these two approaches are shown in the last three columns of Table 3.

Based on results from simulations, our proposed method and *RnaSeqSampleSize* provided answers much faster than *PROPER*, and our proposed method and *PROPER* provided good sample size estimation. Overall, our proposed method worked the best while both accuracy and computation time are considered.

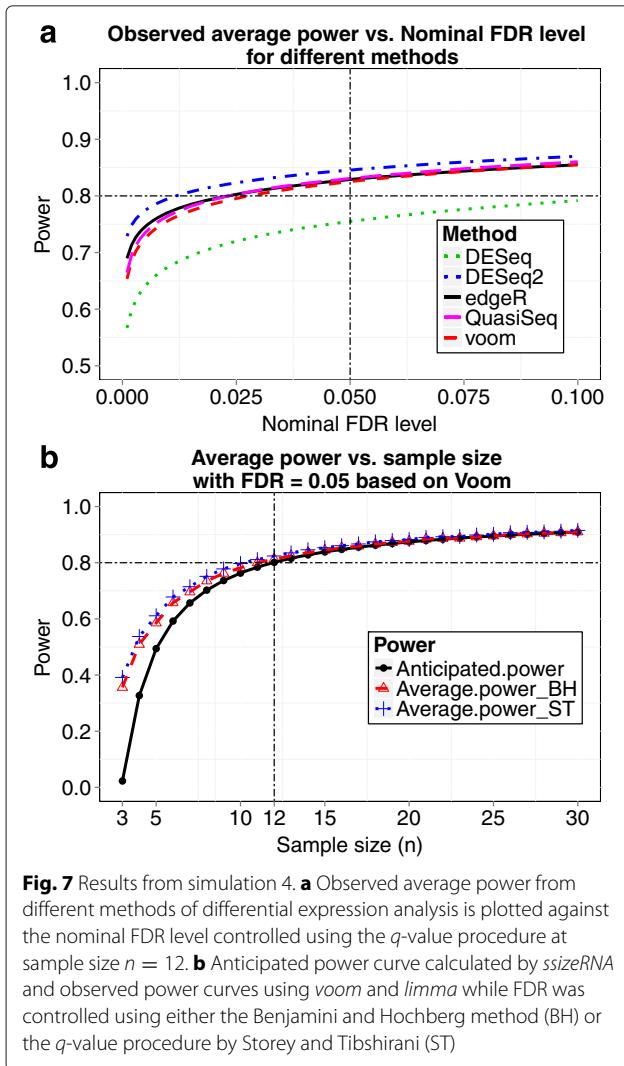
Simulation 4. Real data-based simulation

Our method involves simulating data based on negative binomial distributions. To check the robustness of our method, we conducted a simulation based on a real RNA-seq dataset from [27], which was upon an RNA-seq experiment that sequenced 69 lymphoblastoid cell lines (LCL) derived from unrelated Nigerian individuals. We used the genes with minimum read counts across all individuals larger than 10, which results in 9154 genes. First, we estimated the mean and dispersion across all 69 individuals for each gene. Assume that fold change comes from a log-normal distribution as in simulation 3,

$$\delta_g \sim \log - normal(\log(2), 0.5\log(2)),$$

the proportion of non-DE genes being 80 %, to reach a desired power 0.8 while controlling FDR at 0.05, the sample size calculated by our method is 12 at $m = 200$.

To check whether desired power can be achieved at the calculated sample size, we simulated 100 datasets. For each simulation, we randomly picked 24 out of the 69 individuals and randomly assigned 12 individuals to the control group and the remaining 12 individuals to the treatment group. Consider all 9154 genes among the 24 individuals as EE since the samples were randomly selected from the same population. Then we randomly generated 20 % of the 9154 genes to be DE, and their counts in the treatment group were multiplied by fold change δ_g which were drawn from a $\log - normal(\log(2), 0.5\log(2))$ distribution. The scaled counts were rounded to the nearest integers. This strategy likely results in more realistic data because all counts come from real dataset and no distributional assumptions were imposed. The plot of average power vs. nominal FDR at $n = 12$ is shown in Fig. 7(a), where desired power (0.8) was achieved for most testing methods, including *edgeR*, *DESeq2*, *QuasiSeq*, *voom* and *limma* methods, when con-



trolling FDR at 0.05 using q -value procedure. Figure 7(b) gives the power calculated by our method based on hyper-parameters estimated at $m = 200$. It also presents the observed average power curves when FDR was controlled by either the Benjamini and Hochberg’s method or the q -value procedure. The anticipated power curve based on $m = 200$ is close to the other two observed power curves. Hence, our proposed method also provides a reliable estimation of sample size and power in the most realistic simulation study.

Conclusions

In recent years, RNA-seq technology has become a major platform to study gene expression. With large sample size, RNA-seq experiments would be rather costly; while insufficient sample size may result in unreliable statistical inference. Thus sample size calculation is a crucial issue when designing an RNA-seq experiment. Although

we could use a lot of simulations for each sample size and determine the one that reach our desired power as suggested in [10, 21, 22], this requires generous calculation and lacks efficiency. Our method provides a quick calculation for sample size, which only demands one-time simulation. From the simulation studies in the section of Results and discussion, we demonstrate that our proposed method offers a reliable approach for sample size calculation for RNA-seq experiments.

For each gene g , when we use a two-sample t -test to do differential expression analysis, the effect size Δ_g in formula (4) depends on the simulated sample size m . Larger m may lead to better estimation of the prior distributions and hence a more accurate sample size. Based on our simulation studies, the effect of m on the resulting sample size is not big, and $m = 200$ should be enough for providing a relatively precise sample size.

The ordinary t -test instead of the moderated t -test [9] was used in *ssizeRNA* R package. Because the ordinary t -test is a bit less powerful than the moderated t -test, it tends to overestimate the sample size which might be the reason why our calculated sample size in simulation 2 is a little bit larger than what we actually need according to the observed power curves using *voom* and *limma*. However, the overestimation is not dramatic and far less than the method of Li et al. [19].

In this article, we illustrate our idea using a method for two-sample comparison with the t -test, because detecting differentially expressed genes between two treatment groups is the most common case in RNA-seq analysis. Our idea could be applied to multi-sample comparison with an F-test or tests for linear contrasts of treatment means as well.

The R package *ssizeRNA* implements our proposed sample size calculation method for RNA-seq experiments and it is freely available on the Comprehensive R Archive Network (<http://cran.r-project.org>). To install this package, start R and enter:

```
source("http://bioconductor.org/biocLite.R")
biocLite("ssizeRNA")
```

Appendices

Appendix A: Derivation of Eq. (3)

For each individual gene g , the weighted linear model

$$y_g = X\beta_g + \sigma_g W_g^{-\frac{1}{2}} \epsilon$$

can be fitted to log-cpm values

$$Y_g = (y_{g11}, \dots, y_{g1n_1}, y_{g21}, \dots, y_{g2n_2})'$$

with design matrix

$$X = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix},$$

coefficients vector

$$\beta_g = \begin{pmatrix} \beta_{g1} \\ \beta_{g2} \end{pmatrix},$$

unknown gene-specific standard deviation σ_g , associated *voom* precision weights

$$W_g = \text{diag}(w_{g11}, \dots, w_{g1n_1}, w_{g21}, \dots, w_{g2n_2}),$$

and error

$$\epsilon \sim \text{MVN}(\mathbf{0}, I_{n_1+n_2}).$$

Thus we could obtain the coefficient estimators

$$\hat{\beta}_g = (X^T W_g X)^{-1} X^T W_g \mathbf{y}_g$$

with variance-covariance matrix

$$\text{Var}(\hat{\beta}_g) = \sigma_g^2 (X^T W_g X)^{-1},$$

where σ_g^2 is estimated by s_g^2

$$s_g^2 = \frac{(\mathbf{y}_g - X\beta_g)' W_g (\mathbf{y}_g - X\beta_g)}{n - p}$$

with

$$p = \text{rank}(X) = 2.$$

Let v_{gk} be the k th diagonal element of $(X^T W_g X)^{-1}$, where

$$\begin{aligned} & (X^T W_g X)^{-1} \\ &= \begin{pmatrix} \sum_{i=1}^2 \sum_{j=1}^{n_i} w_{gij} & \sum_{j=1}^{n_2} w_{g2j} \\ \sum_{j=1}^{n_2} w_{g2j} & \sum_{j=1}^{n_2} w_{g2j} \end{pmatrix}^{-1} \\ &= \frac{\begin{pmatrix} \sum_{j=1}^{n_2} w_{g2j} & -\sum_{j=1}^{n_2} w_{g2j} \\ -\sum_{j=1}^{n_2} w_{g2j} & \sum_{i=1}^2 \sum_{j=1}^{n_i} w_{gij} \end{pmatrix}}{\sum_{j=1}^{n_1} w_{g1j} \sum_{j=1}^{n_2} w_{g2j}}. \end{aligned}$$

Under the assumptions as made in Smyth (2004),

$$\hat{\beta}_{gk} | \beta_{gk}, \sigma_g^2 \sim N(\beta_{gk}, v_{gk} \sigma_g^2)$$

and

$$s_g^2 | \sigma_g^2 \sim \frac{\sigma_g^2}{d_g} \chi_{d_g}^2,$$

where d_g is the residual degrees of freedom for the linear model of gene g , the ordinary t -test statistic will be

$$t_{gk} = \frac{\hat{\beta}_{gk}}{s_g \sqrt{v_{gk}}},$$

which follows an approximate t -distribution with d_g degrees of freedom.

Assuming equal variance between treatment and control group, then the statistic for testing

$$H_0^g : \mu_{g1} = \mu_{g2} \quad \text{vs.} \quad H_1^g : \mu_{g1} \neq \mu_{g2}$$

for the g th gene is

$$\begin{aligned} T_g &= \frac{\hat{\beta}_{g2}}{S.E.(\hat{\beta}_{g2})} \\ &= \frac{\hat{\beta}_{g2}}{s_g \sqrt{v_{g2}}} \\ &= \frac{\hat{\beta}_{g2}}{s_g \sqrt{\frac{\sum_{i=1}^2 \sum_{j=1}^{n_i} w_{gij}}{\sum_{j=1}^{n_1} w_{g1j} \sum_{j=1}^{n_2} w_{g2j}}}} \\ &= \frac{\hat{\beta}_{g2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{\sum_{j=1}^{n_1} w_{g1j} \sum_{j=1}^{n_2} w_{g2j}}{\sum_{i=1}^2 \sum_{j=1}^{n_i} w_{gij}}}}{s_g \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ &= \frac{\hat{\beta}_{g2} \sqrt{\frac{\sum_{j=1}^{n_1} w_{g1j} / n_1 \sum_{j=1}^{n_2} w_{g2j} / n_2}{\sum_{i=1}^2 \sum_{j=1}^{n_i} w_{gij} / (n_1 + n_2)}}}{s_g \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ &= \frac{\hat{\beta}_{g2} \sqrt{\frac{\bar{w}_{g1} \cdot \bar{w}_{g2}}{\bar{w}_{g..}}}}{s_g \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ &\equiv \frac{\Delta_g}{s_g \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \end{aligned}$$

where

$$\Delta_g \equiv \hat{\beta}_{g2} \sqrt{\frac{\bar{w}_{g1} \cdot \bar{w}_{g2}}{\bar{w}_{g..}}}$$

with $\bar{w}_{g1} = \frac{1}{n_1} \sum_{j=1}^{n_1} w_{g1j}$, $\bar{w}_{g2} = \frac{1}{n_2} \sum_{j=1}^{n_2} w_{g2j}$ and $\bar{w}_{g..} = \frac{1}{n_1+n_2} \sum_{i=1}^2 \sum_{j=1}^{n_i} w_{gij}$.

Appendix B: Choice of rejection region Γ satisfying formula (1)

For the two-sample comparison with t -test statistics T_g as in Eq. (3), we assume as in LH method that the effect size follows a normal distribution

$$\Delta_g \sim N(\mu_\Delta, \sigma_\Delta^2),$$

and the variance of log-cpm values for each gene follows an inverse gamma distribution

$$\sigma_g^2 \sim \text{Inv} - \text{Gamma}(a, b)$$

with mean $\frac{b}{a-1}$, then formula (1) becomes

$$\begin{aligned} & \frac{\alpha}{1-\alpha} \frac{1-\pi_0}{\pi_0} \\ & \geq \frac{\Pr(T \in \Gamma|H=0)}{\Pr(T \in \Gamma|H=1)} \\ & = \frac{\Pr(T \in \Gamma|H=0)}{\int \int \Pr(T \in \Gamma|H=1, \Delta_g, \sigma_g) \pi_1(\Delta_g) \pi_2(\sigma_g) d\Delta_g d\sigma_g} \\ & = \frac{\Pr(|T_g| > c|H=0)}{\int \int \Pr(|T_g| > c|H=1, \Delta_g, \sigma_g) \pi_1(\Delta_g) \pi_2(\sigma_g) d\Delta_g d\sigma_g}, \end{aligned} \tag{5}$$

where $\pi_1(\Delta_g)$ and $\pi_2(\sigma_g)$ denote the probability distribution function (p.d.f.) of Δ_g and σ_g respectively. The numerator in (5) equals

$$2 \cdot T_{n_1+n_2-2}(-c),$$

where $T_d(\cdot)$ denotes the cumulative distribution function (c.d.f.) of a central t -distribution with d degrees of freedom, and the denominator in (5) equals

$$\begin{aligned} & 1 - \int \int T_{n_1+n_2-2}(c|\theta_g) \pi_1(\Delta_g) \pi_2(\sigma_g) d\Delta_g d\sigma_g \\ & + \int \int T_{n_1+n_2-2}(-c|\theta_g) \pi_1(\Delta_g) \pi_2(\sigma_g) d\Delta_g d\sigma_g. \end{aligned} \tag{6}$$

Here, $T_d(\cdot|\theta_g)$ denotes the c.d.f. of a non-central t -distribution with d degrees of freedom and non-centrality parameter

$$\theta_g = \frac{\Delta_g}{\sigma_g \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

The integration in (6) with respect to Δ_g could be avoided through mathematical derivation, and the integration with respect to σ_g is approximated using static quadrature rules, which allows a stable calculation to get the root of c . Details of derivation could be found in Appendix B of [14].

Once the choice of c has been made for each sample size, power would be calculated accordingly by integrating over the prior distributions on effect size and residual variance. Hence based on the desired power, sample size is finally determined.

Additional file

Additional file 1: This excel file contains comparison of resulting sample size and power between Li et al.'s method [19] and our proposed method for simulation 1, with parameter settings from Table 1 in [19]. The results are obtained under $m = 200$, with Li's result in the first row from each parameter setting, and our result in the second row. (XLS 49.2 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

RB and PL developed the method. RB implemented the *ssizeRNA* R package. RB and PL drafted the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

This research was partially supported by the National Science Foundation Plant Genome Research Program under Award Number IOS-1127017.

Received: 5 November 2015 Accepted: 20 March 2016

Published online: 31 March 2016

References

- Robinson MD, Smyth GK. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*. 2007;23:2881–87.
- Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*. 2008;9:321–32.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–140.
- McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 2012;40:4288–97.
- Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11:R106.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
- Lund SP, Nettleton D, McCarthy DJ, Smyth GK. Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Stat Appl Genet Mol Biol*. 2012;11:Article 8.
- Law CW, Chen Y, Shi W, Smyth GK. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15:R29.
- Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3:Article 3.
- Fang Z, Cui X. Design and validation issues in RNA-seq experiments. *Brief Bioinform*. 2011;12:280–87.
- Hart SN, Therneau TM, Zhang Y, Poland GA, Kocher J-P. Calculating sample size estimates for RNA sequencing data. *J Comput Biol*. 2013;20:970–78.
- Therneau T, Hart S, Kocher J-P. Calculating sample size estimates for RNA Seq studies. R package version 1.10.0. <https://bioconductor.org/packages/release/bioc/html/RNASeqPower.html>.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57:289–300.
- Liu P, Hwang JTG. Quick calculation for sample size while controlling false discovery rate with application to microarray analysis. *Bioinformatics*. 2007;23(6):739–46.
- Orr M, Liu P. Sample size estimation while controlling false discovery rate for microarray experiments using *ssize.fdr* package. *The R J*. 2009;1(1, May 2009):47–53.
- Chen Z, Liu J, Ng HKT, Nadarajah S, Kaufman HL, Yang JY, Deng Y. Statistical methods on detecting differentially expressed genes for RNA-seq data. *BMC Syst Biol*. 2011;5(Suppl 3):S1.
- Busby MA, Stewart C, Miller CA, Grzeda KR, Marth GT. Scotty: a web tool for designing RNA-Seq experiments to measure differential gene expression. *Bioinformatics*. 2013;29:656–57.
- Li CI, Su PF, Guo Y, Shyr Y. Sample size calculation for differential expression analysis of RNA-seq data under poisson distribution. *Int J Comput Biol Drug Des*. 2013;6:358–75.
- Li CI, Su PF, Shyr Y. Sample size calculation based on exact test for assessing differential expression analysis in RNA-seq data. *BMC Bioinforma*. 2013;14(1):357.
- Zhao S, Li C, Guo Y, Sheng Q, Shyr Y. *RnaSeqSampleSize*: R package version 1.2.0. <https://www.bioconductor.org/packages/release/bioc/html/RnaSeqSampleSize.html>.

21. Ching T, Huang S, Garmire LX. Power analysis and sample size estimation for RNA-Seq differential expression. *RNA*. 2014;20(11):1684–96.
22. Wu H, Wang C, Wu Z. PROPER: comprehensive power evaluation for differential expression using RNA-seq. *Bioinformatics*. 2015;31:233–41.
23. Storey JD. A direct approach to false discovery rates. *J R Stat Soc B*. 2002;64:479–98.
24. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
25. Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous rates: a unified approach. *J R Stat Soc B*. 2004;66:187–205.
26. Tausta SL, Li P, Si Y, Gandotra N, Liu P, Sun Q, Brutnell TP, Nelson T. Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C4-related processes. *J Exp Bot*. 2014;65:3543–55.
27. Pickrell J, Marioni J, Pai A, Degner J, Engelhardt B, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*. 2010;464:768–72.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

