Comparison of normalization methods for differential gene expression analysis in RNA-Seq experiments
A matter of relative size of studied transcriptomes

Elie Maza1,2, Pierre Frasse1,2, Pavel Senin1,2, Mondher Bouzayen1,2 and Mohamed Zouine1,2,*

1Université de Toulouse; INP-ENSA Toulouse; Laboratoire Génomique et Biotechnologie des Fruits; Castanet-Tolosan, France; 2INRA; Laboratoire Génomique et Biotechnologie des Fruits; Auzeville, Castanet-Tolosan, France

Keywords: Computational methods, Differential gene expression analysis, High throughput sequencing, Normalization methods, Relative transcriptome size, RNA-Seq data

In recent years, RNA-Seq technologies became a powerful tool for transcriptome studies. However, computational methods dedicated to the analysis of high-throughput sequencing data are yet to be standardized. In particular, it is known that the choice of a normalization procedure leads to a great variability in results of differential gene expression analysis. The present study compares the most widespread normalization procedures and proposes a novel one aiming at removing an inherent bias of studied transcriptomes related to their relative size. Comparisons of the normalization procedures are performed on real and simulated data sets. Real RNA-Seq data sets analyses, performed with all the different normalization methods, show that only 50% of significantly differentially expressed genes are common. This result highlights the influence of the normalization step on the differential expression analysis. Real and simulated data sets analyses give similar results showing 3 different groups of procedures having the same behavior. The group including the novel method named “Median Ratio Normalization” (MRN) gives the lower number of false discoveries. Within this group the MRN method is less sensitive to the modification of parameters related to the relative size of transcriptomes such as the number of down- and upregulated genes and the gene expression levels. The newly proposed MRN method efficiently deals with intrinsic bias resulting from relative size of studied transcriptomes. Validation with real and simulated data sets confirmed that MRN is more consistent and robust than existing methods.

Introduction

RNA-Seq approaches, based on high-throughput sequencing technologies, are becoming an essential tool in transcriptomics studies (ref. 24). In the beginning, its ability to capture transcriptome dynamics (across different tissues and conditions) without sophisticated normalization of data sets was considered as a particular advantage over other techniques. Nevertheless, it was found that a normalization preprocessing step can significantly improve the quality of the analysis. Therefore, the effect of normalization was rigorously studied and many works related to this issue have been published in recent years. Notably, an evaluation of statistical methods for normalization in mRNA-Seq experiments stressed (ref. 2) the requirement for further research in the development of statistical and computational methods dedicated to the processing of mRNA-Seq data. In particular, Bullard, Purdom, Hansen and Dudoit (ref. 2) demonstrated the impact of the choice of the normalization procedure on the Differential Expression (DE) analysis.

Many of the normalization methods proposed in the literature are based on the correction of biases and “artifacts” directly related with RNA-Seq technology. For instance, the read (or fragment) number per lane and the gene (or RNA) length are trivial biases that have been pointed out very early (refs. 11, 12, and 25). Sequencing depth is another trivial and important bias highlighted in reference 19. Other technical biases, such as non-uniformity of the distributed reads within the transcripts, and the strong sample-specific GC-content effects on read counts, were later studied in references 15, 14, and 9.

The present paper focuses on another significant bias in RNA-Seq experiments that is not introduced by the technology, but is rather intrinsic to the studied transcriptomes. It is shown here that the relative size of studied transcriptomes represents an important bias and that a particular normalization procedure is needed to address this issue. The aim of the present study is therefore to propose an improved version of an existing normalization method and to evaluate its performance on simulated and...
real data sets. The newly established method named “Median Ratio Normalization” (MRN) is then compared with normalization methods that have been described in the literature (refs. 17, 1, and 2).

Bias Related to the Relative Size of Transcriptomes

Thereafter, we will refer to different biological states of a given studied transcriptome as conditions, and to independent biological samples of a given condition as replicates.

Let \( X_{grk} \) be the observed number of reads (or count) of gene \( g \) in condition \( k \) for replicate \( r \); \( \mu_{gk} \); the expectation of the true and unknown number of transcripts of gene \( g \) in condition \( k \); \( L_g \), the length of gene \( g \); and \( N_{gk} \), the total number of reads in condition \( k \) for replicate \( r \). As described by Robinson and Oshlack (ref. 17), among others, we can model the expected value of \( X_{grk} \) as:

\[
\mathbb{E}(X_{grk}) = \frac{\mu_{gk} L_g}{S_k} N_{gkr}
\]  

(1)

where \( S_k \) is the size of studied transcriptome in condition \( k \), that is:

\[
S_k = \sum_{g=1}^{G} g \mu_{gk}.
\]  

(2)

For each gene \( g \), an approximation of the expected value of the ratio between 2 conditions, say 1 and 2, is given by the delta method (see, for example, ref. 23) as:

\[
\frac{X_{g2r}}{X_{g1r}} = \frac{\mu_{g2} S_2}{\mu_{g1} S_1} \frac{N_{g2r}}{N_{g1r}}.
\]  

(3)

As our interest is gene expression fold change, it is obviously given in the above equation (3) by \( \mu_{g2}/\mu_{g1} \). Then, ratios \( N_{g2}/N_{g1} \), and \( S_2/S_1 \) are the biases introduced, respectively, by RNA-Seq technology and by the relative size of studied transcriptomes. The former bias can be easily corrected by division of the observed counts (in condition \( k \) for replicate \( r \)) by the total number of reads \( N_{gkr} \) before the ratio calculation. The latter bias is related to the relative size of studied transcriptomes and cannot be removed directly because the values referring to \( \mu_{g2} \) and \( \mu_{g1} \) are obviously unknown. It is shown here that, for a given experiment, this bias is not directly related to the technology, but rather to the studied transcriptomes. Robinson and Oshlack (ref. 17) also clearly show this bias and give a solution which seems nevertheless less robust than the one we propose. However, it is important to mention that the implementation of the MRN method requires taking into account a biological assumption based on (but less restrictive than) that made previously in reference 17. This assumption states the following: in transcriptomes under study, less than 50% of the genes are upregulated, whereas less than 50% are downregulated. We detail in the Methods section (subsection Median Ratio Normalization Method) the interest of such an assumption.

Evaluated Normalization Methods

The present paper evaluates a number of normalization methods designed to resolve (directly or indirectly) the bias induced by the relative size of studied transcriptomes.

First, methods directly addressing this issue are taken into consideration. The Trimmed Mean of M-values (TMM) method (developed in ref. 17) deals with estimating the “relative RNA production of 2 samples” under the assumption that the majority of genes are not differentially expressed. This method is implemented in the edgeR package available from the Bioconductor website at http://www.bioconductor.org (see refs. Sixteen and 7). Similarly, Anders and Huber (ref. 1), proposed a normalization method based on “size factors” that renders counts from different samples comparable. This method is called “Relative Log Expression” (RLE) and is implemented in edgeR and DESeq packages (available from the Bioconductor website).

Second, we evaluate 4 normalization methods that do not directly aim at removing this bias. Instead, they are dealing with normalization of the read counts in order “to adjust for varying lane sequencing depths and potentially other technical effects” (see ref. 2 and references therein). These are the Upper Quartile normalization (UpQu), the Median normalization (Medi), the Total Counts normalization (ToCo) and the FPKM normalization (refs. Two and 22). While in the UpQu normalization method, counts are normalized by division (in a given replicate) by the upper quartile of these counts, the Medi simply computes the median, and the ToCo uses the sum of all counts. Finally, the FPKM normalization method normalizes raw counts by both the length of the gene and the total counts. In addition, for this normalization method, the obtained quantities are multiplied by \(10^9\) in order to obtain “fragments per kilobase of transcript per million fragments mapped” abbreviated as FPKM (see refs. Eleven and 22).

We compare the performance of all methods mentioned above against “No Normalization” (NoNo) and against our proposed method called “Median Ratio Normalization” (MRN). The MRN method follows the idea behind TMM and RLE normalization methods aiming at removing the bias due to the relative size of studied transcriptomes.

In total, 9 normalization methods, used for the DE analysis, were compared: TMM, TMM50 (TMM with 50% of trimmed M-values), RLE, UpQu, Medi, ToCo, FPKM, MRN and NoNo.

Materials and Methods

This section is divided into 4 independent parts. The first part is devoted to the description of the MRN procedure. We explicitly show, in this part, how the MRN procedure overcomes the bias due to the relative size of studied transcriptomes. The second part presents the pipeline used in this paper for the DE analysis. The third part describes the algorithm used to simulate RNA-Seq data sets that will be used to compare the normalization methods described above. Finally, the last part describes the tomato’s RNA-Seq data set studied in this paper.
Median ratio normalization method. The computational steps of the MRN method aiming at removing the bias due to the relative size of studied transcriptomes, are given below. An R program of the MRN method is provided within a supplementary file (see Section Supplementary Materials). For sake of simplicity, the method is here described for 2 conditions (k = 1 and k = 2) with the same number of replicates R. Obviously, the MRN method can be generalized to an arbitrary number of conditions K > 2 with different numbers of replicates.

1. Calculation of weighted means of gene expressions for both conditions k ∈ [1, 2] and all genes g ∈ [1, ..., G]:

\[
\bar{X}_{gk} = \frac{1}{R} \sum_{r=1}^{R} X_{gkr}.
\]

2. Calculation of gene expression ratios with condition 1 as a reference, for all genes g ∈ [1, ..., G]:

3. Calculation of the median of obtained ratios: τ = median g (τ g).

4. Calculation of normalization factors for both conditions k ∈ [1, 2] (taking into account factor τ) and for each replicate r ∈ [1, ..., R] (taking into account the sequencing depth Nkr):

   \[e_{kr} = \frac{1}{\tau Nkr},\]

   \[e_{2r} = \frac{\tau}{N2r}.
   \]

5. Calculation of adjusted normalization factors to multiply, for symmetry, to 1: fkr = ekr/ for k ∈ {1, K = 2} and r ∈ [1, ..., R] with


The theoretical bases of the above computational workflow are the followings. Step 1 and Equation (1) imply that the expected values of meansE(Xg1) are equal to μg1/Lg/S1. Hence, in Step 2, Equation (3) implies that the expected values of ratios E(τ g) are approximately equal to (μg2/μg1)(S2/S1) where the bias of the relative size of transcriptomes is given by S2/S1. By assuming that less than half the genes are upregulated, and less than half are downregulated, the median τ of all ratios (calculated in Step 3) is then an approximation of S1/S2. In Step 4, in order to remove this bias due to the relative size of transcriptomes, normalization factors are fixed to e1 for the reference condition (condition 1) and to e2 for condition 2. Finally, Step 5 only aims at having adjusted factors to multiply to 1 for symmetry (as in ref. 16).

Pipeline for the differential expression analysis. Since the objective of the present work is to evaluate the impact of normalization methods on the DE analysis, each of studied data sets was analyzed by the application of the same universal pipeline for DE analysis. However, the normalisation step was obviously different for each of the studied methods.

All computations were done with R environment (ref. 13). TMM, RLE, UpQu and Medi normalization methods were performed with function calcNormFactors() of the edgeR package. Computations of MRN, FPKM and ToCo normalization methods were implemented by us and provided within a supplementary file (see Supplementary Materials).

For the DE analysis, an error model based on the Negative Binomial distribution was used, with variance and mean linked by local regression. This model is well suited for modeling count data dispersion (refs. 3, 6, and 10) and is implemented in DESeq and DESeq2 packages. In this study, we used the 2 functions estimateDispersions() and nbinomTest() of the DESeq package. The R program for the DE analyses is also provided in a supplementary file (see Supplementary Materials).

Simulation of RNA-Seq data. Simulated RNA-Seq data sets for K = 2 different conditions and R = 3 biological replicates were generated by the following 2-steps procedure.

Step I. This step deals with the simulation of gene expression values μg, for both conditions 1 and 2. The input consists of various parameters, some of which are not fixed (underlined parameters) and allow the simulation of different data types by changing their values. The aim of this step is to start from simulated M and A-values (from Log-Normal and Normal distributions respectively) and then to resolve the expression values. (Simulations of Gaussian laws are performed with the rnorm() function of R.)

1. Number of simulated genes: G = 50000.
2. Percentages of differentially expressed genes: Pu % of upregulated genes and Pd % of downregulated genes.
3. Simulation of M-values from Log-Normal distributions for upregulated genes M̂u g and downregulated ones M̂d g:

   \[M̂u g \sim 1 + \text{LN}(m_u, s_u)\] and \[s_u = 0.7,\]

   \[M̂d g \sim 1 - \text{LN}(m_d, s_d)\] and \[s_d = 0.7,\]

   Where -1 (resp. 1) implies a 2-fold downregulation (resp. upregulation).
4. Simulation of A-values from a Normal distribution: \(A_g \sim N(m_A, s_A)\) with \(m_A = 7\) and \(s_A = 3\).
5. Calculation of gene expression values for conditions 1 and 2:

   \[g1 = 2^{(\frac{1}{2}(A_g - M̂_1))},\]

   and where \(\lceil x \rceil\) is the integer immediately higher than \(x\).
6. Calculation of simulated M and A-values:

   \[M̂_g = \log_2 \left( \frac{g1}{g2} \right) - \log_2 \left( \frac{g1}{g2} \right),\]

   \[A_g = \left( \log_2 \left( \frac{g1}{g2} \right) + \log_2 \left( \frac{g1}{g2} \right) \right)/2.\]

Step II. This step is dedicated to the simulation of observed counts Xgkr from previously computed gene expression values and Negative Binomial distributions. Simulated Negative Binomial random counts are computed with the nbinom() function of R which depends upon 2 parameters: mu for the mean and size for the dispersion. (In this case, the variance is linked to the mean by the equation \(\text{mu + mu}^2/\text{size}\).)

1. It is assumed that all genes have the same length: \(L_g = 1000\) bp.
2. Calculation of transcriptome sizes for conditions 1 and 2 from Equation (2):
underlined parameters in Step I of the above algorithm. Some of these simulated data sets are performed with quite extreme situations leading to some challenging data sets in order to really discriminate the performance of studied normalization methods. The full parameter details of these 4 simulations are given in Table 1. Simulation types 1, 2 and 3 differ from each other only by the amounts of upregulated and downregulated genes: \( P_u \) and \( P_d \). While having the same amount of upregulated and downregulated genes, simulation type 4 has a higher mean expression: \( \mu > m_d \). Theoretically, simulation types 1, 3 and 4 should result on transcriptomes with different sizes. Indeed, for simulation types 1 and 3, the amount of upregulated genes \( P_u \) is greater than the amount of downregulated ones \( P_d \) and, as a consequence of independent and identical distributions of simulated expressions, \( S_2 \) should be greater than \( S_1 \). For the simulation type 4, while the percentages of upregulated and downregulated genes are the same, the mean expression of upregulated genes \( \mu \) is greater than the mean expression of downregulated genes \( m_d \) and, in turn, \( S_2 \) should also be greater than \( S_1 \). Only simulation type 2 should have equal transcriptome sizes.

An example of a data set from simulation type 1 is shown in Figure 1. Panel A shows the MA-plot of gene expression values (simulated from Step I of the simulation algorithm) with green dots for upregulated genes, orange dots for non DE genes and red dots for downregulated genes. Panels B, C, and D show the same MA-plot of means of observed values (simulated from Step II of the simulation algorithm) with black dots for all genes and, respectively, green dots for upregulated genes in panel B, orange dots for non DE genes in panel C and red dots for downregulated genes in panel D.

![Figure 1](image-url)

**Figure 1.** MA-plots of a simulated data set. This figure corresponds to a simulated data set from simulation type 1 of Table 1. Panel A shows MA-plot of gene expression values (simulated from Step I of the simulation algorithm) with green dots for upregulated genes, orange dots for non DE genes and red dots for downregulated genes. Panels B, C, and D show the same MA-plot of means of observed values (simulated from Step II of the simulation algorithm) with black dots for all genes and, respectively, green dots for upregulated genes in panel B, orange dots for non DE genes in panel C and red dots for downregulated genes in panel D.

### Table 1. Parameters of 4 simulation types

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_u )</td>
<td>40%</td>
<td>30%</td>
<td>15%</td>
<td>10%</td>
</tr>
<tr>
<td>( P_d )</td>
<td>20%</td>
<td>30%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>( \mu )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>( m_d )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

This table contains parameters of the 4 simulation types studied in the paper. Parameters correspond to underlined parameters of Step I of the simulation algorithm described in the Methods section (subsubsection Simulation of RNA-Seq Data).

Tomato’s RNA-Seq data set. To investigate the tomato transcriptome dynamics of fruit set, RNA were isolated from flower buds and flowers at Anthesis and Post-Anthesis stages. For each stage, cDNA libraries were generated from 3 biological replicates and subjected to Illumina mRNA-Seq technology sequencing. We generated 30–40 million high-quality sequence reads for each replicate. A total of 287.5 millions of 101 bp paired-end
As described below, obtained results are remarkably stable for the MRN method from one repetition to another and, consequently, no more than 5 repetitions seem to be needed. Figure 2 shows the number of false discoveries produced by each of the normalization methods for each of the simulation types. Figure 3 shows the Mean Squared Error (MSE) of each normalization method for each of the simulation types, that is, the mean of squared differences between ratios of gene expressions obtained by a normalization method and the true ratios of simulated data. Results presented in these figures can be obtained with the R program of a given supplementary file (see Supplementary Materials). Table 2 contains, for each method, some interesting indicators as power and false positive and negative rates.

Number of false discoveries. Globally, as we can see on Figure 2, the total number of false discoveries (maximum ordinate of each graph) is decreasing starting from the simulation type 1 toward the simulation type 4. That is obviously due to the decreasing amount of DE genes (see Table 1). For any of the simulation types, the greatest amount of errors is always associated to the FPKM normalization method. Another global trend is that for all 4 simulation types, the MRN has the smallest number of errors closely followed by UpQu, Medi, TMM, TMM50, and RLE methods. Moreover, the MRN gives also more robust results with almost the same number of false discoveries for each of the 5 repetitions for each of the simulation types.

For simulation type 2, the differences between normalization methods (excepted FPKM) are less important. That is mostly due to the fact that, as described previously, the sizes of the 2 transcriptomes are nearly the same, and consequently, no normalization is needed. For simulation types 1, 3, and 4, only MRN seems to produce symmetric errors (although less significant for simulation type 1). The other methods, at least for some of the 5 repetitions, produce more false downregulated genes. This drawback is due to the non-symmetry of ratios around value 1 for simulation types 1, 3, and 4 as quoted in the Methods section (subsection Simulation of RNA-Seq Data). Also, as expected, this drawback is more important for NoNo and ToCo methods where no normalization for the relative size of transcriptomes is applied.

Quantitative results. Table 2 summarizes, for each method, the results obtained from the 20 simulations (from the 4 simulation types with 5 independent repetitions for each one). The
Obviously, these results do not intend to validate one or another of the studied normalization methods. Nevertheless, this study shows interesting links between the different normalization methods leading to different groups of methods having the same behavior. Moreover, these groups can be compared with those obtained in the previous simulation study.

Hierarchical classifications. The 2 upper graphs of Figure 4 are hierarchical classifications of all studied normalization methods for 5 independent simulation repetitions. For each simulation type, each repetition has the same symbol for each normalization method.

Table 2. Quantitative results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NoNo</td>
<td>0.3279</td>
<td>0.7898</td>
<td>0.0195</td>
<td>0.2102</td>
</tr>
<tr>
<td>ToCo</td>
<td>0.3411</td>
<td>0.8159</td>
<td>0.0241</td>
<td>0.1841</td>
</tr>
<tr>
<td>FPKM</td>
<td>0.1956</td>
<td>0.4794</td>
<td>0.0062</td>
<td>0.5206</td>
</tr>
<tr>
<td>UpQu</td>
<td>0.3237</td>
<td>0.7858</td>
<td>0.0151</td>
<td>0.2142</td>
</tr>
<tr>
<td>Medi</td>
<td>0.3243</td>
<td>0.7878</td>
<td>0.0149</td>
<td>0.2122</td>
</tr>
<tr>
<td>TMM</td>
<td>0.3241</td>
<td>0.7899</td>
<td>0.0131</td>
<td>0.2101</td>
</tr>
<tr>
<td>TMM50</td>
<td>0.3243</td>
<td>0.7917</td>
<td>0.0123</td>
<td>0.2083</td>
</tr>
<tr>
<td>RLE</td>
<td>0.3247</td>
<td>0.7902</td>
<td>0.0139</td>
<td>0.2098</td>
</tr>
<tr>
<td>MRN</td>
<td>0.3412</td>
<td>0.8348</td>
<td>0.0117</td>
<td>0.1652</td>
</tr>
</tbody>
</table>

This table contains, for each method, the rate of significantly DE genes (Signif. = Number of signif. DE genes / Total number of simulated genes), the power to detect a DE gene (Power = Number of rightly signif. DE genes / Total number of DE genes), the false positive rate (F. Pos. = Number of false positive genes / Total number of non DE genes) and the false negative rate (F. Neg. = Number of false negative genes / Total number of DE genes). We notice that F. Neg. = 1 - Power. For Signif. and Power indicators (resp. F. Pos. and F. Neg. indicators), the maximum (resp. minimum) value is bolded in a grayed cell, the second one is only bolded, and the minimum (resp. the maximum) value is underlined.

Summary was done by adding all genes of all simulations (working then with a sample size of 20 × 30000 = 600000 genes). Excepting the FPKM method, all methods give between 32% and 34% of significantly DE genes. The MRN method maximizes this indicator with 34%. The FPKM method gives about 20% of significant DE genes. These values are consistent with the power to detect a DE gene: the MRN method holds the best power value (about 83%) and the FPKM method holds the worst one (about 48%). In terms of false positive rate, the FPKM has the best result (about 0.62%) followed by the MRN method (about 1.17%). Globally, the MRN method seems to give the best results.

Results on the tomato fruit set data set. Results of DE analyses performed on the tomato RNA-Seq data with the various normalization methods described above are shown on Figure 4. Panels A and B are related, respectively, to flower Buds to Anthesis transition and to Anthesis to Post-Anthesis transition.
we made is less restrictive than the one made in reference 17, which assumes that “the majority of genes are non DE.” The same issue has been also addressed by the RLE normalization method (ref. 1), which uses also another computational workflow. The processing procedure of RNA-Seq data described in the present study clearly outperforms comparatively to previous methods.

Of particular interest, the notion of relative size is also considered for high throughput metabonomics and proteomics data analyses where count data are commonly used (refs. Four and 8).

Based on poor benchmark performance, it clearly emerges that the FPKM normalization method should be largely avoided in DE analysis. This conclusion is also in agreement with the evaluation made by Bullard, Purdom, Hansen and Dudoit (ref. 2). Two other normalization methods, UpQu and Medi, assessed in our study display poor and very variable performances. On the other hand, our study indicated that the use of TMM or RLE methods lead to good performance on simulated data sets, though giving rise to some variability from one repetition to another. Finally, the MRN proposed method is found to be consistent and robust, producing globally better results that are remarkably stable from one repetition to another.

The study also shows that normalization methods globally behave similarly when trained with the simulated data sets and with the real data sets, which somehow tends to validate our simulation approach. In that regard, while, to our knowledge, a standard simulation procedure is still lacking, our current work provides a unique benchmark simulation procedure that could be useful for future researches on transcriptomics, metabonomics and proteomics data analyses.

**Discussion**

In this paper, we identified and discussed an important issue related to RNA-Seq data normalization for differential gene expression analysis. Simulated data analysis revealed that the bias due to the relative size of transcriptomes leads to poor estimations of ratios of gene expressions, and consequently to biased DE analysis. To address this issue, we benchmarked the performance of the most widespread normalization methods together with a novel proposed method named “Median Ratio Normalization” (MRN).

The newly established method is a modified version of the normalization method already proposed by Robinson and Oshlack (ref. 17), where authors also stressed the need to remove the bias due to the relative size of transcriptomes, but proposed a different computational workflow. Moreover, the assumption we made is less restrictive than the one made in reference 17, which assumes that “the majority of genes are non DE.” The same issue has been also addressed by the RLE normalization method (ref. 1), which uses also another computational workflow. The processing procedure of RNA-Seq data described in the present study clearly outperforms comparatively to previous methods.

Of particular interest, the notion of relative size is also considered for high throughput metabonomics and proteomics data analyses where count data are commonly used (refs. Four and 8).

Based on poor benchmark performance, it clearly emerges that the FPKM normalization method should be largely avoided in DE analysis. This conclusion is also in agreement with the evaluation made by Bullard, Purdom, Hansen and Dudoit (ref. 2). Two other normalization methods, UpQu and Medi, assessed in our study display poor and very variable performances. On the other hand, our study indicated that the use of TMM or RLE methods lead to good performance on simulated data sets, though giving rise to some variability from one repetition to another. Finally, the MRN proposed method is found to be consistent and robust, producing globally better results that are remarkably stable from one repetition to another.

The study also shows that normalization methods globally behave similarly when trained with the simulated data sets and with the real data sets, which somehow tends to validate our simulation approach. In that regard, while, to our knowledge, a standard simulation procedure is still lacking, our current work provides a unique benchmark simulation procedure that could be useful for future researches on transcriptomics, metabonomics and proteomics data analyses.
We notice here that, independently to the normalization method used, the issue of isofrom switch can influence the expression results, leading to erroneous DE analysis. Indeed, read counts calculation may change depending on the data treatment applied prior to the normalization step. This is obviously an important issue but not considered in our paper since it is prior to the normalization step.

Finally, we wanted to draw attention on the recent article written by Dillies, Rau, Aubert, Hennequot-Antier, Jeanmougin et al., (ref. 5), which appears when we were writing the present article. In this paper, the authors also give an interesting evaluation of normalization methods for RNA-Seq data analysis. We find that many of the results described in this article are consistent with our own results. Obviously, the MRN method that we propose here is not evaluated.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Acknowledgments
This work was supported by the Laboratoire d’Excellence, LabEx, TULIP (ANR-10-LABX-41). This work benefited from the networking activities within the European funded COST ACTION FA1106 “QualityFruit.”

Supplementary Materials
The reader is referred to the online Supplementary Materials for the 3 R programs (ref. 15) described below:

- **Simulation of RNA-Seq data.** Supplementary file 1 contains an R program for the simulation algorithm described and used in this paper. Parameters actually correspond to simulation type 1 (see Table 1), but can easily be changed in order to carry out other simulation types.

- **Normalizations and differential expression analyses.** Supplementary file 2 contains an R program for differential expression analyses with all the 9 normalization methods studied in this paper. This program can be directly executed with the output of Supplementary file 1.

- **Some graphical results.** Supplementary file 3 contains an R program providing some graphical results of differential expression analyses performed with the various normalization methods studied in this paper. This program can be directly executed with the output of Supplementary file 2.

Supplementary material may be found here:
http://www.landesbioscience.com/journals/cib/article/25849/

References


4. Dieterle F, Ross A, Schlotterbeck G, Senn H. Transcriptome Analysis. In this paper, the authors also give an interesting evaluation of normalization methods for RNA-Seq data analysis. We find that many of the results described in this article are consistent with our own results. Obviously, the MRN method that we propose here is not evaluated.


