



Review Article

A Review on Recent Statistical Models for RNA-Seq Data

Mohammad Samir Farooqi*, DC Mishra, KK Chaturvedi, Anil Rai, SB Lal, Sanjeev Kumar, Jyotika Bhati and Anu Sharma

Abstract

The next generation sequencing technology, RNA-sequencing (RNA-seq), has got growing acceptance in transcriptome analyses. Statistical methods used for gene expression analyses with RNA-seq provide meaningful inferences of gene expression using counts of reads. There are various statistical models with its pros and cons available for RNA-seq data analysis. There is a need for consistent statistical methods to explore the information from the developing sequencing technologies. The current article gives a review of the statistical methods with their limitations that can be useful for the RNA-seq analysis. The main emphasis is given to the parametric, nonparametric and hybrid models for identifying the genes with differential expression.

Keywords

RNA-seq; Statistical models; R packages; edgeR; DEseq

Introduction

Complete set of all the RNA molecules in a cell, makes up a transcriptome. Both the noncoding RNAs and protein coding mRNAs are included in the set. Understanding the transcriptome helps in deciphering the functional elements of the genome which can be further used to know the molecular constituents present in a cell and tissues [1]. Study of transcriptome, provides the basis for understanding the growth of an organism and also threats to its development from diseases and environment. The transcriptomic analysis is extensively used in deciphering the genomic function, and thereby quantifies the changes in genomic abundance, adaptation and helps in correlating these changes in different tissues at different time intervals [2]. High-throughput mRNA sequencing technique (also known as RNA-seq) has gained immense popularity among researchers and scientists for transcriptome analysis. This technique provides the ability to develop precise methodologies for a variety of RNA-seq applications, including gene expression quantification, novel transcript and exon discovery, differential expression (DE) analysis and splice variant detection [3].

The main concern related to transcriptome data analysis is in selecting informative genes from the transcriptome data and to develop a high-performance model with the selected genes, thus RNA-seq experiments must be analysed with robust, efficient and

statistically valid algorithms. In a standard RNA-seq experiment, a sample of RNA is converted to a library of complementary DNA fragments and then sequenced on a high-throughput sequencing platform [4]. From this sequencing, millions of short sequences (reads) are obtained which are then mapped to a reference genome to measure the abundance of each transcript/gene. The count of reads mapped to a given gene or transcript, measures the expression level of this gene/transcript. *De novo* method has been used to assemble the transcripts in case of non-availability of reference genome.

One of the primary goals for most RNA-seq experiments is to compare the gene expression levels across various experimental conditions, treatments, tissues, or time points. In most of the studies, researchers are particularly interested in detecting gene with differential expressions (DE). The study of determining which genes have changed significantly in terms of their expression across biological samples is referred to as DE analysis. Identifying which genes are differentially expressed between samples helps researchers to understand the functions of genes in response to a given condition. To conduct DE analysis, large number of statistical models and tools are devised [2,5,6]. This review discusses some of the recently developed statistical models applied for DE analysis.

Statistical models applied for DE analysis

In a RNA-seq study two types of variations are involved in the data, the technical variation, induced due to machines and the biological variation caused because of different biological samples. Because of these two kinds of variations, as well as biases that exist within and between treatments, it is difficult to precisely distinguish real biological differences between groups. The read counts obtained under case vs. control conditions are found to be consistently different which can be attributed to true biological differences between case vs. control conditions or experimental error. Statistical tests are required to distinguish between the two possibilities. For evaluating this expression change in RNA-seq data, different statistical models have been developed. The patterns of differential gene expression are approximated using appropriate statistical distributions. Initially, Bullard, et al. [7], obtained normalized expression counts for RNA-seq data using Fisher's exact test. All these statistical methods for DE analysis are categorized in parametric, nonparametric and hybrid (combination of both parametric and nonparametric) approach.

Parametric models

In a parametric model, 'shape' of the data is assumed and coefficients are estimated for the model. Thus a parametric model is based on the estimation of parameters and all the information about the data is obtained within its parameters and a future data value from the current state of the model is predicted based on the estimated parameters. These models help in predicting the value of unknown data. A parametric model has a fixed number of parameters and is computationally faster, but makes stronger assumptions about the data; the model may perform well if the assumptions are correct, but it may give misleading results if the assumptions are wrong. Differential gene expression analysis is performed under assumption that the expression values are normalized and mapped to a particular distribution [8].

*Corresponding authors: Mohammad Samir Farooqi, Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, Pusa, New Delhi 110012, India, Tel: +91-11-25841721; E-mail: samirfarooqi8@gmail.com

Received: January 07, 2019 Accepted: March 11, 2019 Published: March 29, 2019

Significant issues involved in RNA-seq data are represented as below:

1. RNA-seq data includes biases occurred during library preparation, the varying length of genes or transcripts and the effect of nucleotide composition causes biases of abundance measures and there is also variation due to the effects of difference in total number of mapped reads for different samples.
2. The effect of sequencing depth and the number of replicates are uncertain.
3. The data used, is count data because it is the number of counts aligned to a gene. Hence it is not continuous and therefore cannot be modeled as say a normal distribution.
4. Since Poisson distribution is designed for modeling count data, it assumes the first and second moments (mean and variance) as equal. This is not true for RNA-seq data. Lowly expressed genes have much higher variance than highly expressed genes.
5. To account for the variability, the negative-binomial (NB) model is used which is an extension of Poisson model. The NB model has an extra parameter to model for the variance. It has been proved as the variance approach the mean, the NB model becomes the Poisson model.

Hence models applied for rarely occurring events such as Poisson distribution, negative binomial distribution etc. is widely applied in this scenario [9,10]. The researchers have to judiciously select, which model to apply, looking into the advantages and limitations of these models.

Poisson model

Poisson distribution is widely used in modeling counting processes. When something is sampled and counted, then the inherent uncertainty that is present in that measurement is Poisson variance. Since this variance is based on the value of the count itself, it is not experiment-specific. A two-parameter generalized Poisson (GP) model based on likelihood ratio test (LRT) for identification of DE genes was given by Srivastava and Chen [11] and proved that this model showed the significant improvement over the traditional Poisson model. A two stage Poisson model (TSPM) was given by Auer and Doerge [12] here at first stage genes with small cumulative counts are filtered and classified as not over-dispersed and then at the second stage these genes are tested using a LRT from a Poisson model. Through simulation study they showed that TSPM has improved performance over negative binomial model and quasi likelihood approach in situations where both over-dispersed and non over-dispersed are present. Further, Poisson distribution based on a Wald test or a LRT for sample size calculation to model RNA-seq count data for single-gene differential expression analysis was proposed by Fang and Cui [13]. Since DE analysis involves examination and testing of large amount of genes simultaneously, therefore, this method has the limitation of hypothesis correction. This problem was further tackled by Li et al. [14] who gave formula for sample size calculation using the Wald test, Rao's score test and log transformation of these tests. Further, they also extended their method by using numerical approach [14] which was implemented as Web based calculator for sample size calculation, i.e., RNaseqPS [15].

Negative binomial model

RNA-seq data has been shown to have the problem of over dispersion as it exhibit greater variability [16]. Hence, Poisson

distribution models fail to counter this variability. To tackle this over dispersion problem, negative binomial models have been proposed by different authors to take care of mean – variance relationship [17-19]. The negative binomial distribution provides the exact test for comparison of two groups [17]. To analyse RNA-seq data using negative binomial distribution, DESeq [18], baySeq [19], DESeq2 [20], edgeR [21], PoissonSeq [22] and gfold [23] have been proposed. Further, negative binomial power test was developed to control the false discovery rate in DE analysis [24]. RNaseqPower in Bioconductor package was implemented as power analysis method based on ranking of DE genes [25].

Since number of genes in RNA-seq data is very large, identification of DE genes requires simultaneous testing of null hypothesis of all these genes that involves multiple hypothesis testing problems. In this case, false discoveries are inevitable and the error rate of this needs to be controlled. Several methods have been proposed to solve the problem of multiple gene comparison and a power analysis method has been developed to control the false discovery rate [26]. Calculation of sample size using minimum average read counts in control conditions, the minimum fold change and maximum dispersion was proposed [27] which was further implemented as an algorithm in Bioconductor package RnaSeqSampleSize [28]. Further, a Bioconductor package, PROPER has been developed and implemented for simulation based power analysis [29]. All these negative binomial based methods are developed to measure the DE between two conditions. A generalized linear model framework was given to perform power analysis for multiple group comparisons [30], giving due considerations to other packages DESeq [18], edgeR [21], DSS [31], DESeq2 [32], EBSeq [33] and SSeq [34]. In order to avoid complex mathematical approximations, Bioconductor packages edgeR, DESeq and DESeq2 used LRT for DE analysis. This was further modified using LRT under the generalized linear model approach to develop power analysis method [35]. This method was implemented through online user interface available at (<http://140.116.152.140/shiny/App/GLM/>).

Poisson-lognormal distribution model

Gene expression can be modeled using log normal distribution because truncated normal distribution is found to approximate the distribution of log read counts [36]. Busby et al. [36] assumed that read counts follows Poisson distribution and gene expression follows log normal distribution. Thus, Poisson-log normal distribution was more appropriate to model gene expression data and further developed Scotty-Power analysis method for DE analysis.

Limma model

Limma is a linear model method designed for analysing complex experiments with variety of experimental conditions and predictors for gene expression analysis. It uses empirical Bayes method. Comparisons between many RNA targets can be simultaneously analysed using Limma. It uses empirical Bayes method to obtain information across genes and provide stability to the analysis with the smaller number of arrays [37]. This was implemented in Bioconductor package for RNA-seq data [38]. Further [39] applied precision weights to counter the mean – variance relationship of log count data and developed voom method based on Limma approach. This was implemented in the R package as ssizeRNA [40,41] which is a simulation based power analysis method to measure the differential expression between two conditions and was extended for paired-sample or multiple treatment comparisons procedures. Performance

of ssizeRNA was compared with RnaSeqSampleSize and PROPER. The ssizeRNA has shown the improved performance based on computational time and accuracy [42].

Selection of appropriate parametric model

Selection of appropriate model under different situations is a difficult task as the parametric models are based on certain assumptions, which must be fulfilled so that the data fits the distribution and the results are appropriate. However, for applying specific models under different situations, models proposed are enlisted in Table 1 below.

Non parametric model

Parametric approaches are based on the distributional assumption and are effective when these assumptions are satisfied. Violation of these distributional assumptions or a poor estimate of parameters gives results which are inappropriate. Both Poisson and Negative binomial models are heavily influenced by the presence of outliers in the data [43]. This may be due to presence of an expressed gene in one condition but not expressed in other conditions. Errors introduced during mapping also produce outliers. In order to tackle these issues, models based on nonparametric methods have also been developed. A nonparametric model uses a flexible number of parameters, and the number of parameters often grows as it learns from more data. A nonparametric model is computationally slower, but makes fewer assumptions about the data. A nonparametric model can capture more subtle aspects of the data distribution. A nonparametric model that predicts future data is dependent on not just the parameters alone but also on the present data characteristic that has been observed. Numerous nonparametric models have been proposed for this purpose. Fold Change based methods for the

selection of genes have been developed and shown that these methods give more reproducible results [44-46].

Log fold changes and absolute expression difference, test statistic has been used in NOISeq nonparametric approach [47]. SamSeq, a resampling based method was developed to consider the difference in sequencing depths [27]. This method can be applied with quantitative, survival, two-class or multiple class outcomes and was found to be more robust than parametric methods, edgeR, PoissonSeq, DESeq, when compared with real and simulated data. MRFSeq, Markov random field (MRF) approach was introduced which used gene co-expression data for improvement of prediction power [48]. LFCseq, a new data-driven non parametric model was proposed by Lin et al. [42]. Shi et al. [49], developed rSeqNP that uses a nonparametric approach to test for differential expression and splicing from RNA-seq data. rSeqNP uses estimated expression value as input to obtain the statistical significance of DE genes and isoforms. Since information across isoforms is combined in rSeqNP, it is able to detect more differentially expressed or spliced genes from RNA-seq data. BNPSeq, a stochastic process based approach, which does not require pre-processing of data and works in Bayesian nonparametric framework and models the gene count data using gamma-negative binomial process (GNBP) [50] but it cannot be extended to complex experimental design. Widely used software based on nonparametric methods are listed in Table 2.

Hybrid models

Efforts have also been made to combine the parametric and nonparametric approach and develop statistical models that are more robust and improve accuracy in terms of DE genes detection. Xiao et al. [51], developed a new ranking method for genes based on combination of biologically relevant expression change value (f value)

Table 1: Appropriate parametric models for specific condition.

Selection of model	Analysis method/Criteria	Source
Limma	Power analysis method for appropriate sample size	[40]
Poisson lognormal model	Power Scott Analysis	[36]
Poisson model	single-gene comparison	[13]
Poisson model	multiple-gene comparison	[27]
Negative binomial model	two-group comparison	[41,27]
Negative binomial model	multiple-group comparison	[30,35]
edgeR	superior combination of true positive and false positive performances	[21]
DESeq	number of replicates is higher, then minimizing false positives is more important	[42]

Table 2: Software based on nonparametric methods for RNA-seq.

RNAseq software	Test statistic employed	Source
NOISeq	Log fold changes and absolute expression	[47]
SAMSeq	Wilcoxon test with a resampling scheme to compensate for sequencing depth	[27]
NPEBseq	Nonparametric empirical Bayesian-based procedure	[51]
MRFSeq	Markov random field (MRF) approach for improvement of prediction power by using log fold changes and co-expression data	[48]
LFCseq	Log fold changes	[43]
rSeqNP	Wilcoxin rank sum (two condition), wilcoxin singled rank (paired two condition comparison), Kruskal-Wallis statistic (Multiple comparison), Spearman's rank correlation coefficient (Quantitative outcomes), Score statistic of the Cox proportional hazard model (Survival outcomes)	[49]
BNP-Seq	Gamma-negative binomial process (GNBP) or beta-negative binomial process (BNBP)	[50]

and statistical significance (p value), for identification of DE genes from high-throughput gene expression measurements and showed that their combined method is more robust in selecting DE genes. For explicitly modelling both over-dispersed count data and genetic structure in RNAseq data, Sun et al. [16], gave Poisson mixed model (PMM). Further, to make this model applicable to large data sets, they combined the statistical method and the computational algorithm and gave a computationally efficient inference method MACAU (Mixed model Association for Count data via data Augmentation). It uses two random effects terms and controls for both independent over-dispersion and sample non-independence. Benefits of MACAU are that it achieves higher power than several other methods for DE analysis and can also accommodate continuous predictor variables and biological or technical covariates. Compound Poisson-gamma distribution model has been developed by Anjum et al. [52]. The joint likelihood density function and estimate of the parameters were obtained and it was shown that this compound model helps in identifying DE genes more effectively as it captures the extra variation. Farooqi et al. [53], combined models from parametric and nonparametric statistic and developed a hybrid model NBPFCROS for identification and ranking of DE genes. The combined model was developed using the negative binomial power (NBP) model given by Di et al. [24] as parametric statistic and nonparametric statistic given by Dembele et al. [54]. This model was compared with few existing models such as Fold change rank order statistic (FCROS), NBP, edgeR and DESeq2. Linear kernelized Support Vector Machine (SVM) was used to evaluate the accuracy of the model and showed that the hybrid models are more robust in identifying true DE genes.

Conclusion

Recent parametric, nonparametric and hybrid methods for DE genes have been reviewed. We have also discussed the pros and cons including features of these models. These methods are widely used in finding highly expressed DE genes from RNA-seq data. Parametric methods produce p values on the basis of approximate/asymptotic null distribution assumptions. When the genes are highly expressed, the performance of these models is found to be good. However, for lowly expressed genes, it is otherwise. Thus there is a bias in selection of DE genes. Use of parametric approach has preference over nonparametric approach as it provide the increase in power of detection but when the assumptions of these models are violated, there is significant increase in false positive rate. So, models based on nonparametric, as well as hybrid approaches have been developed to take care of such issues. However, still there is lack of consensus on the appropriateness of particular methodology in terms of robustness, accuracy and reproducibility of the results. This research topic is still developing and newer models and methodologies are being developed and explored for their robustness and accuracy.

References

1. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* 10: 57-63.
2. Fang Z, Martin J, Wang Z (2012) Statistical methods for identifying differentially expressed genes in RNA-Seq experiments. *Cell Biosci* 2: 26.
3. Eteleeb AM, Flight RM, Harrison BJ, Petruska JC, Rouchka EC (2013) An Island-based approach for differential expression analysis. *ACM Conf Bioinform Comput Biol Biomed Inform* 419-429.
4. Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26: 1135-1145
5. Reeb PD, Steibel JP (2013) Evaluating statistical analysis models for RNA sequencing experiments. *Front Genet* 4: 178.
6. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, et al. (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol* 17: 13
7. Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11: 94.
8. Costa-Silva J, Domingues D, Lopes FM (2017) RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS One* 12: e0190152.
9. Baggerly KA, Deng L, Morris JS, Aldaz CM (2003) Differential expression in SAGE: accounting for normal between-library variation. *Bioinformatics* 19: 1477-1483.
10. Lu J, Tomfohr JK, Kepler TB (2005) Identifying differential expression in multiple SAGE libraries: An overdispersed log-linear model approach. *BMC Bioinformatics* 6: 165.
11. Srivastava S, Chen L (2010) A two-parameter generalized Poisson model to improve the analysis of RNA-seq data. *Nucleic Acids Res* 38: e170.
12. Auer PL, Doerge RW (2011) A two-stage Poisson model for testing RNA-Seq data. *Stat Appl Genet Mol* 10.
13. Fang Z, Cui X (2011) Design and validation issues in RNA-Seq experiments. *Brief Bioinform* 12: 280-287.
14. Li CI, Su PF, Guo Y, Shyr Y (2013) Sample size calculation for differential expression analysis of RNA-Seq data under Poisson distribution. *Inter J Comput Biol Drug Design* 6: 358-375.
15. Guo Y, Zhao S, Li CI, Sheng Q, Shyr Y (2014) RNAseqPS: a web tool for estimating sample size and power for RNAseq experiment. *Cancer Informatics* 13: CIN-S17688.
16. Sun S, Hood M, Scott L, Peng Q, Mukherjee S, et al. (2017) Differential expression analysis for RNAseq using Poisson mixed models. *Nucleic Acids Res* 45: e106.
17. Robinson MD, Smyth GK (2007) Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* 23: 2881-2887
18. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11: R106.
19. Hardcastle TJ, Kelly KA (2010) BaySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* 11: 422.
20. Wang L, Feng Z, Wang X, Wang X, Zhang X (2010) DEGseq: An R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26: 136-138.
21. Robinson MD, McCarthy DJ, Smyth GK (2010) EdgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140.
22. Li J, Witten DM, Johnstone I, Tibshirani R (2011) Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics* 13: 523-538
23. Feng J, Meyer CA, Wang Q, Liu JS, Shirley LX, et al. (2012) GFOLD: A generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics* 28: 2782-2788.
24. Di Y, Schafer DW, Cumbie JS, Chang JH (2011) The NBP negative binomial model for assessing differential gene expression from RNA-Seq. *Stat Appl Genet Mol Biol* 10: 28.
25. Hart SN, Therneau TM, Zhang Y, Poland GA, Kocher JP (2013) Calculating sample size estimates for RNA sequencing data. *J Comput Biol* 20: 970-978.
26. Guo Y, Li J, Li CI, Shyr Y, Samuels DC (2013) MitoSeek: Extracting mitochondria information and performing high-throughput mitochondria sequencing analysis. *Bioinformatics* 29: 1210-1211.
27. Li J, Tibshirani R (2013) Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-Seq data. *Stat Methods Med Res* 22: 519-536.
28. Zhao S, Zhang Y, Gamini R, Zhang B, Schack D (2018) Evaluation of two main RNA-Seq approaches for gene quantification in clinical RNA sequencing: PolyA+ selection versus rRNA depletion. *Sci Rep* 8: 4781.

29. Wu H, Wang C, Wu Z (2015) PROPER: Comprehensive power evaluation for differential expression using RNA-Seq. *Bioinformatics* 31: 233-241.
30. Ching T, Huang S, Garmire LX (2014) Power analysis and sample size estimation for RNA-Seq differential expression. *RNA* 20: 1684-1696.
31. Wu H, Wang C, Wu Z (2012) A new shrinkage estimator for dispersion improves differential expression detection in RNA-Seq data. *Biostatistics* 14: 232-243.
32. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol* 15: 550.
33. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, et al. (2013) EBSeq: An empirical Bayes hierarchical model for inference in RNA-Seq experiments. *Bioinformatics* 29: 1035-1043.
34. Yu D, Huber W, Vitek O (2013) Shrinkage estimation of dispersion in Negative Binomial models for RNA-Seq experiments with small sample size. *Bioinformatics* 29: 1275-1282.
35. Li CI, Shyr Y (2016) Sample size calculation based on generalized linear models for differential expression analysis in RNA-Seq data. *Stat Appl Genet Mol* 15: 491-505.
36. Busby MA, Stewart C, Miller CA, Grzeda KR, Marth GT (2013) Scotty: A web tool for designing RNA-Seq experiments to measure differential gene expression. *Bioinformatics* 29: 656-657.
37. Smyth GK, Michaud J, Scott HS (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21: 2067-2075.
38. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, et al. (2015) limma powers differential expression analyses for RNA-Sequencing and microarray studies. *Nucleic Acids Res* 43: e47.
39. Law CW, Chen Y, Shi W, Smyth GK (2014) voom: Precision weights unlock linear model analysis tools for RNA-Seq read counts. *Genome Biol* 15: R29.
40. Bi R, Liu P (2016) Sample size calculation while controlling false discovery rate for differential expression analysis with RNA-Sequencing experiments. *BMC Bioinformatics* 17: 146.
41. Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, et al. (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* 22: 839-851.
42. Lin B, Zhang LF, Chen X (2014) LFCseq: A nonparametric approach for differential expression analysis of RNA-Seq data. *BMC Genomics* 15: S7.
43. Guo L, Lobenhofer EK, Wang C, Shippy R, Harris SC, et al. (2006) Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nat Biotech* 24: 1162.
44. Shi L, Tong W, Fang H, Scherf U, Han J, et al. (2005) Cross-platform comparability of microarray technology: intra-platform consistency and appropriate data analysis procedures are essential. *BMC Bioinformatics* 6: S12.
45. Chen JJ, Hsueh HM, Delongchamp RR, Lin CJ, Tsai CA (2007) Reproducibility of microarray data: A further analysis of microarray quality control (MAQC) data. *BMC Bioinformatics* 8: 412.
46. Tarazona S, García F, Ferrer A, Dopazo J, Conesa A (2012) NOseq: A RNA-Seq differential expression method robust for sequencing depth biases. *EMBnet J* 17: 18.
47. Yang EW, Girke T, Jiang T (2013) Differential gene expression analysis using coexpression and RNA-Seq data. *Bioinformatics* 29: 2153-2161.
48. Shi Y, Chinnaiyan AM, Jiang H (2015) rSeqNP: A non-parametric approach for detecting differential expression and splicing from RNA-Seq data. *Bioinformatics* 31: 2222-2224.
49. Dadaneh SZ, Qian X, Zhou M (2018) Bnp-seq: Bayesian nonparametric differential expression analysis of sequencing count data. *J Am Stat Assoc* 113: 81-94.
50. Bi Y, Davuluri RV (2013) NPEBseq: Nonparametric empirical bayesian-based procedure for differential expression analysis of RNA-Seq data. *BMC Bioinformatics* 14: 262.
51. Xiao Y, Hsiao TH, Suresh U, Chen HI, Wu X, et al. (2014) A novel significance score for gene selection and ranking. *Bioinformatics* 30: 801-807.
52. Anjum A, Jaggi S, Varghese E, Lal S, Bhowmik A, Rai A (2016) Identification of differentially expressed genes in RNA-seq data of Arabidopsis thaliana: A Compound Distribution Approach. *J Comput Biol* 23: 239-247.
53. Farooqi S, Kumar D, Mishra DC, Rai A, Singh NK (2018) A hybrid method for differentially expressed genes identification and ranking from RNA-Seq data. *Inter J Bioinform Res App*. (In press)
54. Dembele D, Kastner P (2014) Fold change rank ordering statistics: A new method for detecting differentially expressed genes. *BMC Bioinformatics* 15: 14.

Author Affiliation

Top

Centre for Agricultural Bioinformatics, Indian Agricultural Statistics Research Institute, Library Avenue, Pusa, New Delhi-110012

Submit your next manuscript and get advantages of SciTechnol submissions

- ❖ 80 Journals
- ❖ 21 Day rapid review process
- ❖ 3000 Editorial team
- ❖ 5 Million readers
- ❖ More than 5000 
- ❖ Quality and quick review processing through Editorial Manager System

Submit your next manuscript at • www.scitechnol.com/submission