

Visualization methods for RNA-sequencing data analysis

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Supplementary information: Supplementary data are available at *Bioinformatics* online.

[illegible]

In light of these complications, researchers should analyze RNA-seq data like they would any other biased multivariate data. Simply applying models to such data is problematic because models hold assumptions that they alone cannot call into question. Fortunately, data visualization enables researchers to see patterns and problems they may not otherwise detect with traditional modeling. As a result, the most effective approach to data analysis is to iterate between models and visuals, and enhance the appropriateness of applied models based on feedback from visuals (Shneiderman, 2002). With RNA-seq data, we primarily want to compare the variability between replicates and between treatment groups. This is visually best achieved by drawing the mapped read count distributions across all genes and samples. Unfortunately, the few plotting tools offered in popular RNA-seq packages do not allow users to effectively view their data in this manner.

In this paper, we strive to remedy this problem by publishing new and effective RNA-seq plotting tools. We use real RNA-seq data to show that our tools can detect normalization problems, DEG designation problems, and common errors in the analysis pipeline. We also show that our tools can identify genes of interest that cannot otherwise be obtained by models. We emphasize that interactive graphics should be an indispensable component of modern RNA-seq analysis: Researchers should be able to quickly flip through plots of genes that appear promising or problematic, and link between plots to swiftly obtain various perspectives of their data. Here, we do not propose that users drastically change their approach to RNA-seq analysis. Instead, we propose that users simply modify their approach to RNA-seq analysis by assessing the sensibility of their models with multivariate graphical tools, namely with parallel coordinate plots, scatterplot matrices, and replicate point plots.

Equation (??) Text. Figure 2 shows that the above method Text Text Text Text Text Text Text Text Text Text Text Text Text Text. ? might want to know about text text text text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text. Figure 2 shows that the above method Text
Text Text Text Text Text Text Text Text Text Text Text. ? might want to
know about text text text text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text. Figure 2 shows
that the above method Text Text Text Text Text Text Text Text Text Text
Text Text. ? might want to know about text text text text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text. Figure 2 shows that the above method Text Text Text Text
Text Text Text Text Text Text Text Text Text. ? might want to know about text
text text text

- "main" — 2018/2/5 — page 2 — #2

[illegible]

3.1.1 This is subsubheading

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text. Figure 2 shows that the above method
Text Text Text Text Text Text Text Text Text Text Text Text Text. ? might want to
know about text text text text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text. Figure 2 shows
that the above method Text Text Text Text Text Text Text Text Text Text
Text Text. ? might want to know about text text text text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text. Figure 2 shows that the above method Text Text Text Text Text
Text Text Text Text Text Text Text. ? might want to know about text text
text text

[illegible]

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text. Figure 2 shows that the above method
Text Text Text Text Text Text Text Text Text Text Text Text Text. ? might want
to know about text text text text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text. Figure 2
shows that the above method Text Text Text Text Text Text Text Text Text Text
Text Text Text. ? might want to know about text text text text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text. Figure 2 shows that the above method Text Text Text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text. Figure 2 shows that the above method Text Text Text Text Text Text Text Text Text Text Text. ? might want to know about text text text text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text. Figure 2 shows that the above method Text Text Text Text Text Text Text Text Text Text Text Text.

Text Text. Figure 2 shows that the above method Text Text Text Text Text Text Text Text Text Text Text Text. ? might want to know about text text text text

- 1. this is item, use enumerate
- 2. this is item, use enumerate
- 3. this is item, use enumerate

Text Text. Figure 2 shows that the

above method Text Text Text Text Text Text Text Text Text Text Text Text Text Text. ? might want to know about text text text text Text. Figure 2 shows that the above method Text Text Text Text Text Text Text Text Text Text Text Text. ? might want to know about text text text text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text.

Text Text. Figure 2 shows that the above method Text Text Text Text

Acknowledgements

Text Text Text Text Text Text Text Text. ? might want to know about text text text text

Funding

This work has been supported by the... Text Text Text Text.

References

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

Anders, S., Reyes, A., and Huber, W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome Research*, **22**, 2008–2017.

Baggerly, K.A. and Coombes, K.R. (2009) Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *The Annals of Applied Statistics*, **3**, 1309–1334.

Brown, A.V. and Hudson, K.A. (2015) Developmental profiling of gene expression in soybean trifoliate leaves and cotyledons. *BMC Plant Biology*, **15**, 169.

Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*, **11**, 94.

Cook, D., Hofmann, H., Lee, E.-K., Yang, H., Nikolau, B., and Wurtele, E. (2007) Exploring gene expression data, using plots. *Journal of Data Science*, **5**, 151–182.

Hansen, K.D., Brenner, S.E., and Dudoit, S. (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research*, **38**, e131.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., and Carvalho, B.S. (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, **12**, 115–121.

Ioannidis, J.P., Allison, D.B., Ball, C.A., Coulibaly, I., Cui, X., Culhane, A.C., Falchi, M., Furlanello, C., Game, L., Jurman, G., Mangion, J., Mehta, T., Nitzberg, M., Page, G.P., Petretto, E. and van Noort, V. (2009) Repeatability of published microarray gene expression analyses. *Nature Genetics*, **41**, 149–155.

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

Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008) RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, **18**, 1509–1517.

McIntyre, L.M., Lopiano, K.K., Morse, A.M., Amin, V., Oberg, A.L., Young, L.J., et al. (2011) RNAseq: Technical variability and sampling. *BMC Genomics*, **12**, 293.

MoranLauter, A.N., Peiffer, G.A., Yin, T., Whitham, S.A., Cook, D., and Shoemaker, R.C. (2014) Identification of candidate genes involved in early iron deficiency chlorosis signaling in soybean (glycine max) roots and leaves. *BMC Genomics*, **15**, 1–25.

Morin, R., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T., et al. (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques*, **45**, 81–94.

Oshlack, A., Robinson, M.D., and Young, M.D. (2010) From RNA-seq reads to differential expression results. *Genome Biology*, **11**, 220.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, **40**, 1413–1415.

Risso, D., Schwartz, K., Sherlock, G., Dudoit, S. (2011) GC-Content normalization for RNA-Seq data. *BMC Bioinformatics*, **12**, 480.

- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, **43**, e47.
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., et al. (2010) De novo assembly and analysis of RNA-seq data. *Nature Methods*, **7**, 909–912.
- Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, **11**, R25.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
- Shneiderman, B. (2002) Inventing discovery tools: Combining information visualization with data mining. *Information Visualization*, **1**, 5–12.
- Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., 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.
- Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, **31**, 46–53.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., and Kelley D.R. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, **7**, 562–578.
- Wang, Z., Gerstein, M., and Snyder, M. (2009) RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**, 57–63.
- Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., and Liu, X. (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE*, **9**, e78644.