Introduction

3D RNA-seq App
Easy RNA-seq analysis

3D RNAseq App: differential expression and alternative splicing

Alternative link on Bilibili: https://www.bilibili.com/video/BV1mS4y1g7uB/ (https://www.bilibili.com/video/BV1mS4y1g7uB/)

Licence

3D RNA-seq is currently under a dual-licensing model.

- Open source under GPLV3.0. For academic and non-commercial use, it is free.
- For commercial use, please get in touch to obtain commercial licenses. Contact us

Citation

- Guo,W. et al. (2020) 3D RNA-seq: a powerful and flexible tool for rapid and accurate differential expression and alternative splicing analysis of RNA-seq data for biologists. RNA Biol., DOI: 10.1080/15476286.2020.1858253.
- Calixto,C.P.G., Guo,W., James,A.B., Tzioutziou,N.A., Entizne,J.C., Panter,P.E., Knight,H., Nimmo,H.G., Zhang,R., and Brown,J.W.S. (2018) Rapid and Dynamic Alternative Splicing Impacts the Arabidopsis Cold Response Transcriptome. Plant Cell, 30, 1424-1444.

Report issues & feedback

If you have questions to raise or are experiencing difficulties using the 3D RNA-seq, please use the 3D RNA-seq user group. (https://github.com/wyguo/ThreeDRNAseq/blob/master/vignettes/user_manuals/3D_RNA-seq_App_manual.md)

Table of contents

Description

The ThreeDRNAseq (3D RNA-seq) R package provides an interactive graphical user interface (GUI) for RNA-seq data analysis using accurate quantification of RNA-seq reads. It allows users to perform differential expression (DE), differential alternative splicing (DAS) and differential transcript usage (DTU) (3D) analyses based on limma (Ritchie et al., 2015). The 3D RNA-seq GUI is based on R shiny App and enables a complete RNA-seq analysis to be done within only 3 Days (3D).

Prior to 3D RNA-seq

RNA-seq raw read data is organised by sample. Different samples are either generated by different experimental conditions (multiple genotypes, tissue types or treatments, time-of day circadian experiments) or by biological replication (sometimes also sequencing replication) and each of them is used for the construction of individual libraries. Next, quality control and trimming of the RNA-seq reads should be performed to ensure that high quality RNA-seq reads are used for downstream analysis. The read data are then used for RNA-seq quantification. This step can be performed using many different pipelines, and the type of pipeline determines whether you can use 3D RNA-seq for your downstream expression analyses or not. 3D RNA-seq is only compatible with transcript quantification data derived from Salmon (Patro et al., 2017) or Kallisto (Bray et al., 2016) with the use of a reference transcriptome or Reference Transcript Dataset which contains a list of the known genes and transcripts for the organism under study. The Salmon/Kallisto output file contains the TPM values for each transcript organised by biological repeat and treatment(s). Depending on the size of the dataset, the transcript quantification procedure might take up to 1-2 days.

Inputs to 3D RNA-seq

Three input files are required for downstream analysis:

1. Gather the meta-data of the experimental design in a csv spreadsheet, the columns of which must include the following information (Figure A):
    - A column of the factor or multiple columns of the factors of the experimental design.
    - A column of the biological replicate labels.
    - A column of the sequencing replicate labels if they exist.
    - A column of the file names of transcript quantifications.

2. A folder that contains the transcript quantification files. Each file contains transcript quantification data of a single sample. Read counts and TPMs for 3D analysis will be generated from the "quant.sf" objects if these files are generated by Salmon (Patro et al., 2017) and the "abundance.h5"/"abundance.tsv" objects if these files are generated by Kallisto (Bray et al., 2016) (Figure B).

3. Transcript-gene mapping file in one of the following formats:

    - "csv" spreadsheet with first column of transcript IDs and second column of gene IDs (Figure C).
    - Or a "fa" file of the transcriptome for quantification. Transcript names and gene IDs will be extracted from the "transcript" (if it exists) or ">" tags and "gene" tags in the description lines of sequences, respectively (Figure D).
    - Or a "gtf" file of the transcriptome for quantification. Transcript names and gene IDs will be extracted from the "transcript_id" and "gene_id" tags in the last column, respectively (Figure E).

Note:

1. Transcript-gene mapping in "csv" file is recommended. Depending on the size, it may take a while to generate the table from a "fa" or a "gtf" file and any missing tags for transcript name and gene ID extraction in these files may lead to errors.

2. The 3D analysis is executable in a computer with normal memory and CPU size. If the App is running on docker image, it is recommended to reduce the data size to upload to our server. Users can exclude all the files in sub-folders of transcript quantifications, except the files of "quant.sf" from Salmon. If the transcript quantifications are generated using Kallisto, users can keep either "abundance.h5" or "abundance.tsv" in the sub-folders with smaller size (Figure B).

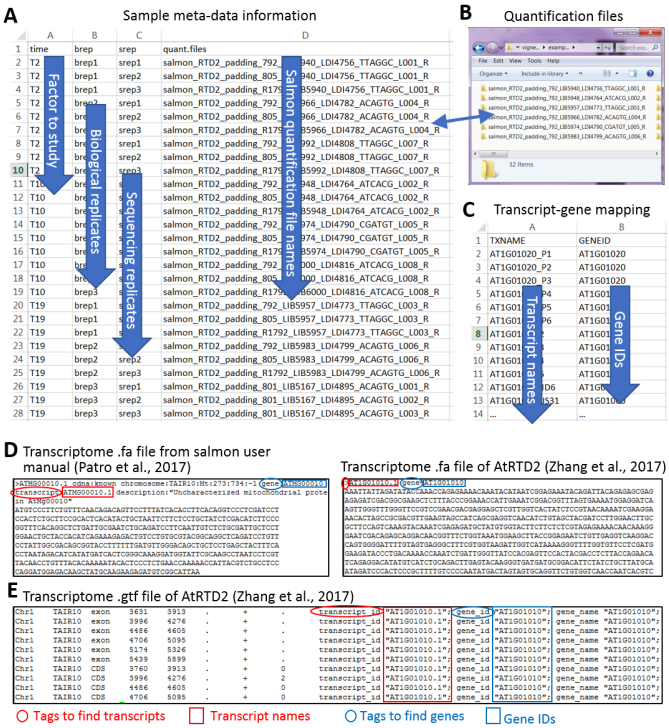

Figure 1: Input files of 3D RNA-seq App. The example is from a RNA-seq study of Arabidopsis in respond to cold (Calixto et al., 2018). (A) Meta-data table of sample information in csv file. (B) The folder contains transcript quantifications. The input of transcript-gene mapping information can be a "csv" spreadsheet with first column of transcript names and second column of gene IDs (C), a ".fa" file (D) or a ".gtf" file (E) of the transcriptome. If a ".fa" or a ".gtf" file is provided, the App will extract transcript-gene association information with specific tags.

Capability of 3D RNA-seq

The 3D RNA-seq analysis pipeline starts with a number of steps to pre-process the data and reduce noise (e.g. low expression filters based on expression mean-variance trend, visualise data variation, batch effect estimation, data normalisation, etc.). The optimal parameters for each step are determined by quality control plots (e.g. mean-variance trend plots, PCA plots, data distribution plots, etc.). Stringent filters are applied to statistical testing of 3D genes/transcripts to control false positives. After the analysis, publication quality plots (e.g. expression profile plots, heatmap, GO annotation plots, etc.) and reports can be generated. The entire 3D RNA-seq analysis takes only 1 Day or less and all actions are performed by simple mouse clicks on the App.

How to get help

3D RNA-seq App "easy-to-use" manual:

https://github.com/wyguo/ThreeDRNAseq/blob/master/vignettes/user_manuals/3D_RNA-seq_App_manual.md (https://github.com/wyguo/ThreeDRNAseq/blob/master/vignettes/user_manuals/3D_RNA-seq_App_manual.md)

Tooltips:

In the GUI, users can click tooltips ? in specific steps for help information.

Contact us:

3D RNA-seq App is developed and maintained by Dr. Wenbin Guo from Plant Sciences Division, School of Life Sciences, University of Dundee. If you have any questions and suggestions, please contact:

- Dr. Wenbin Guo: wenbin.guo@hutton.ac.uk (mailto:wenbin.guo@hutton.ac.uk)
- Dr. Runxuan Zhang: runxuan.zhang@hutton.ac.uk (mailto:runxuan.zhang@hutton.ac.uk)

Pipeline of 3D RNA-seq App

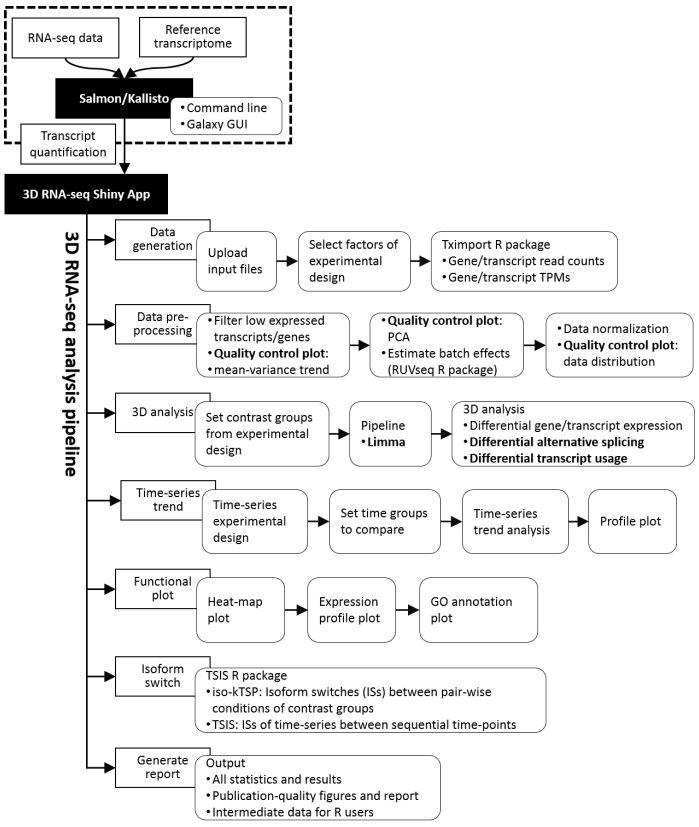

Figure 2: 3D RNA-seq analysis workflow on this App.

References

- Guo,W. et al. (2020) 3D RNA-seq: a powerful and flexible tool for rapid and accurate differential expression and alternative splicing analysis of RNA-seq data for biologists. RNA Biol., DOI: 10.1080/15476286.2020.1858253.
- Bray,N.L., Pimentel,H., Melsted,P., and Pachter,L. (2016) Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol., 34, 525-527.
- Calixto,C.P.G., Guo,W., James,A.B., Tzioutziou,N.A., Entizne,J.C., Panter,P.E., Knight,H., Nimmo,H.G., Zhang,R., and Brown,J.W.S. (2018) Rapid and Dynamic Alternative Splicing Impacts the Arabidopsis Cold Response Transcriptome. Plant Cell, 30, 1424-1444.
- Guo,W. et al. (2020) 3D RNA-seq: a powerful and flexible tool for rapid and accurate differential expression and alternative splicing analysis of RNA-seq data for biologists. RNA Biol., 00, 1–14.
- Patro,R., Duggal,G., Love,M.I., Irizarry,R.A., and Kingsford,C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods, 14, 417-419.
- 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 Res, 43, e47.
- 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, 139-40.

Data generation →