

Nov 21, 2024

Coding-free differential TF activity estimation from *C. elegans* transcriptomic data with the CelEst R Shiny app

DOI

<https://dx.doi.org/10.17504/protocols.io.5qpvok6exl4o/v1>



Marcos Francisco Perez¹

¹IBMB-CSIC



Marcos Francisco Perez

IBMB-CSIC

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.5qpvok6exl4o/v1>

Protocol Citation: Marcos Francisco Perez 2024. Coding-free differential TF activity estimation from *C. elegans* transcriptomic data with the CelEst R Shiny app. protocols.io <https://dx.doi.org/10.17504/protocols.io.5qpvok6exl4o/v1>

Manuscript citation:

Perez MF, 2024. CelEst: a unified gene regulatory network for estimating transcription factor activities in *C. elegans*. GENETICS



License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: September 24, 2024

Last Modified: November 21, 2024

Protocol Integer ID: 108240

Keywords: transcription factor, *C. elegans*, activity, RNA-seq, enrichment, gene ontology, unified gene regulatory network, differential expression of target gene, transcription factor activity, free differential tf activity estimation, transcriptomic experiment, transcriptomic data, development of the celest network, celest network, rna, target gene, unified gene, regulatory network, gene, celest, differential expression, seq experiment

Funders Acknowledgements:

Ministerio de Economía y Competividad (Spain)

Grant ID: RYC2021-034496-I

Abstract

Differential TF activity can be estimated from differential expression of target genes in *C. elegans* transcriptomic experiments. This allow for greater mechanistic interpretation of your RNA-seq experiments.

The development of the CelEst network used in this analysis is described in full in:

Perez MF, 2024. **CelEst: a unified gene regulatory network for estimating transcription factor activities in *C. elegans*. GENETICS**

Download R Shiny app and install necessary packages

- 1 Go to github.com/IBMB-MFP/CeEst-app to download the Shiny app on to your local machine. Click on the green button marked '<> Code' to open a drop-down menu. Select 'Download ZIP' and wait for the file to download. Unzip the CeEst-app-main.zip file and move the folder to your desired location.
- 2 To run the app in R, the 'shiny' package needs to be installed and loaded. If the 'shiny' package is not installed, install it by running the code `install.packages("shiny")`. To load the package, run the code `library(shiny)`.

Once 'shiny' is loaded, change the working directory to the CeEst-app-main folder using the function `setwd("path/to/folder/CeEst-app-main")`. You can then run the app with command `runApp()`.

To run the Shiny app in R Studio, open the app.R file in the CeEst-app-main folder and click the 'Run App' button with the green play icon.

- 3 The first time that CeEst is run, it will install all the necessary packages for the analysis to be performed. This may take a while to complete; please be patient.

Estimating differential TF activity from existing differential expression stats

- 4 If you wish to perform an analysis starting from an existing differential expression analysis, click on the second tab on the left-hand menu, labelled 'TF act from DE stats'.

The input file needs to be correctly formatted. It should be a two-column tab-separated text file. The first column should contain the gene identifiers. The second column should be the differential expression statistics. The table should have column headers; the first column should be called 'geneid'. Any NA or NaN values should be set to 0. An example dataset can be downloaded through a blue link ('from DE stats') on the CeEst landing page.

The gene identifiers must be WormBase Gene IDs (e.g. WBGene00020142), sequence names (e.g. T01C8.1) or entrezGene IDs (e.g. 181727). The app should automatically detect the type of gene identifier and convert them as necessary.

- 5 Click the 'Browse' button and navigate to your file. If desired, choose an alternative network for the analysis; the CeEst GRN is recommended and used by default. When the file is selected, click on 'Compute TF activities' and wait until the analysis is complete. It should take a few minutes at most. Progress messages will inform you of the current stage of the analysis. When complete the results will appear below.

Expected result

You should see a volcano plot appear, which plots the estimated TF activity score against the p-value ($-\log_{10}$). If you wish to save this image, you can right-click to save it.

You should also see an interactive sortable data table with TF activity scores and p-values. The table is annotated with different TF identifiers and with any mode-of-regulation (i.e. activator/repressor) annotation from either WormBase or UniProt. To save this table, click on the green button 'Download Data Table'.

Estimating differential TF activity from raw counts

- 6 If you wish to perform an analysis starting from raw gene-level RNA-seq read counts, navigate to the third tab on the left-hand menu, labelled 'DE & TF act from counts'.

The input data file needs to be correctly formatted. It should be a tab-separated file. The first column should contain gene identifiers, which must be WormBase Gene IDs (e.g. WBGene00020142), sequence names (e.g. T01C8.1) or entrezGene IDs (e.g. 181727). Remaining columns should be gene-level read counts for samples (one column per sample). The first column header should be 'geneid'; remaining columns should be headed with sample identifiers. Avoid sample names that contain white space (e.g. use 'Sample_1' instead of 'Sample 1').

- 7 Click the 'Browse' button and navigate to your file. When it is selected, the sample identifiers from the column names should appear just below. The app needs to be told which are the control samples and which are the treatment samples. In the box 'Control Samples', enter the sample identifiers for the control samples, separated by a comma and without quotation marks. e.g.

'Control_1, Control_2, Control_3'

Do the same for the treatment samples.

- 8 If you wish to correct for differences in developmental age in your DE analysis, click on the tick box labelled 'Correct for developmental speed using RAPToR?'. Otherwise, skip to step 9.

When the tick box is marked, a drop-down button selection will appear for you to select an appropriate age reference from the 'wormRef' package. Click on the button whose age range matches what you expect based on your sample collection.

- 9 Click 'Compute DE & TF activities' only once. Wait patiently until the analysis is complete. Progress messages will appear but may not appear immediately. If you have selected to correct for developmental age, the analysis will take longer.



Expected result

You should see a volcano plot appear, which plots the estimated TF activity score against the p-value ($-\log_{10}$). If you wish to save this image, you can right-click to save it.

Below, you should also see an interactive sortable data table with TF activity scores and p-values. The table is annotated with different TF identifiers and with any mode-of-regulation (i.e. activator/repressor) annotation from either WormBase or UniProt. To save this table, click on the green button 'Download Data Table'.

The app will also return the results from the DE analysis as an interactive and sortable data table. To save this table, click on the green button 'Download Data Table'.

Lastly, if you elected to correct for developmental age, a table entitled 'RAPToR age estimations (hours post-hatch @ 20C)' will return the estimated ages for your samples. As the box title suggests, these timings will be scaled for development at 20C and so may not match your chronological timings if experiments were conducted at a different temperature.



Protocol references

If you use CelEsT in a publication, don't forget to cite:

Perez, MF (2024) CelEsT: a unified gene regulatory network for estimating transcription factor activity in *C. elegans*. bioRxiv [doi: 10.1093/bioadv/vbac016](https://doi.org/10.1093/bioadv/vbac016)

&

Badia-i-Mompel, P et al. (2022) decoupleR: ensemble of computational methods to infer biological activities from omics data. Bioinformatics Advances, [doi: 10.1093/bioadv/vbac016](https://doi.org/10.1093/bioadv/vbac016)

If you use the app to conduct your DE analysis, cite:

Love, MI et al. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, [doi: 10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8)

If you choose to control for developmental age in your analysis, cite:

Bulteau, R & Francesconi, M. (2022) Real age prediction from the transcriptome with RAPToR. Nature Methods, [doi: 10.1038/s41592-022-01450-0](https://doi.org/10.1038/s41592-022-01450-0)