DESeq2 Normalization for Aneuploidy

Samuel Hunter, 2023

The files are located in the github repo (git@github.com:Dowell-Lab/sr2023.git). Clone to your local computer (or update your branch with git pull if you already have the repository). The project files for today are located in sr2023/day05/ds3\_work\_only

Open a new R script and follow along (the completed script is also included in that repo, but it’s good practice to type it yourself!)

Required libraries: tidyverse and DESeq2.

1. Load in your files. Change the path to wherever you stored your files:

Graphical user interface, text

Description automatically generated

1. Load in and clean up your counts file to be DESeq2-compatible. Run DESeq2 with a batch correction:

Text

Description automatically generated

1. Generate your results table (shrinking log2FoldChange is optional). Also calculate a per-chromosome median fold change:

Graphical user interface, text, application, email

Description automatically generated

1. Generate some plots for visualization:

Graphical user interface, text, application

Description automatically generated

1. You’ll notice that for chr21, most genes are at 1.5x expression (as expected), but a subset are closer to a fold-change of 1.0, and are not considered significant. What can we say about these genes?

**Normalization Method 1: Adjusting Counts By Ploidy**

1. Our metadata table lists the ploidy status for each sample. We can define a function which will print the copy number of the gene based on the sample’s ploidy status. We use this to generate a matrix corresponding to the copy number for each gene in each sample:

Graphical user interface, text

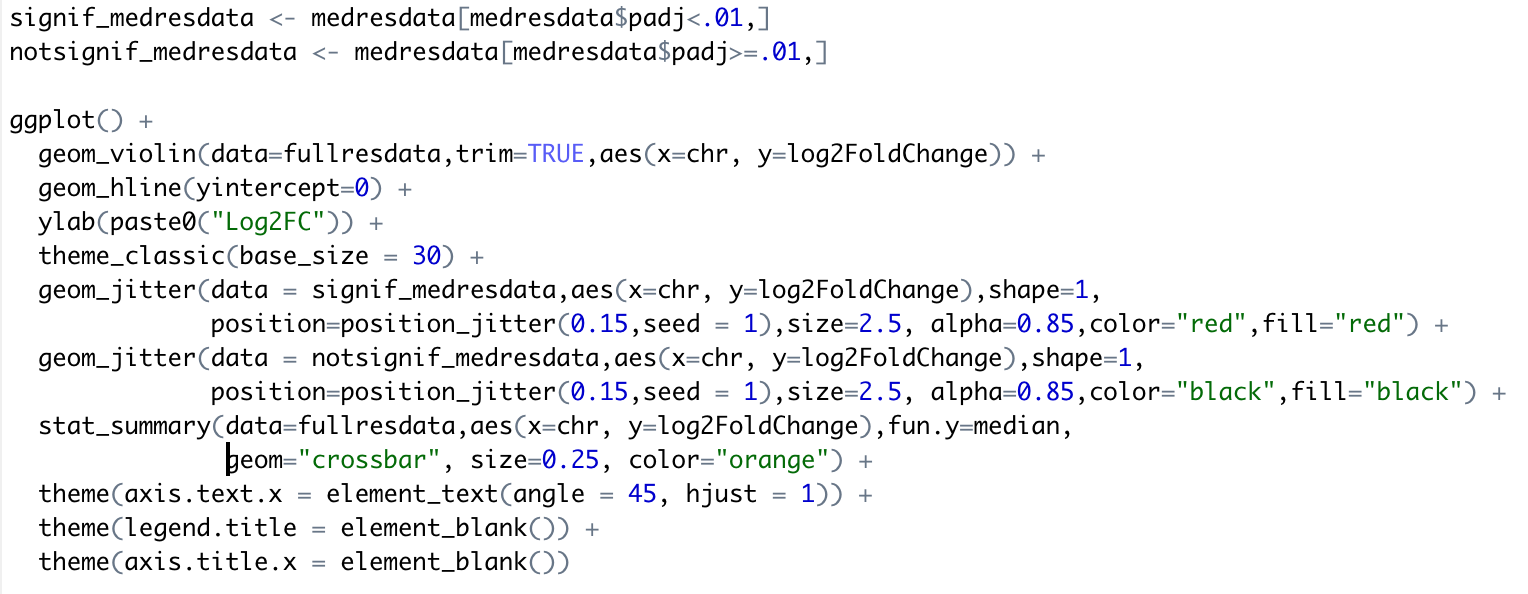
Description automatically generated

1. Re-run DESeq2 with the new normalization matrix:

Graphical user interface, text, application

Description automatically generated

1. Visualize our results



**Method 2: Change the alternative hypothesis**

1. Re-run DESeq2, using an adjusted alternative hypothesis

Graphical user interface, text, application, email

Description automatically generated

1. Visualize the results:

Graphical user interface, text, application

Description automatically generated