



Ball Horticultural Company  
622 Town Road  
West Chicago, IL 60185-2698  
USA  
ballhort.com

630 231-3600  
Fax: 630 231-3605

## Ball Helix Bioinformatics Associate Scientist Coding Challenge

### Summary:

As a part of the Ball Helix Bioinformatics and Data Analytics team, we are frequently required to discuss complex data analysis results to audiences without scientific backgrounds.

In this scenario, you are working on a project to identify potential genes that are responsible for a flowering trait, and have quantified paired end RNA-sequencing data with the tool Salmon (Patro, et al., 2017). Once you have finalized your analysis, you are asked to present the results to the larger project team.

Other team members include a plant breeder, a product manager and two representatives from the supporting seed business who are involved in making financial decisions.

### Challenge:

Using the attached gene expression matrix (**test\_gene\_expression\_matrix.txt**) in your programming language of choice:

- 1) Identify the total number of differentially expressed genes between flower buds (Bud\_Rep\*) and flowers two days after anthesis (Day2\_Rep\*) and indicate the statistical threshold that was used to determine the number of differentially expressed genes.
- 2) With the list of differentially expressed genes, create a figure that you would use to explain to the team which genes are candidates for the trait and should be studied further.
- 3) How would you describe the results of a RNA-sequencing experiment to team members who are not experienced in genomics?

Upload your code, figure and response to question three to a GitHub repository and send the link to [mbowman@ballhort.com](mailto:mbowman@ballhort.com) no later than 5PM CDT on December 13<sup>th</sup>, 2019.