Bioinformatics Bite #3: Intro to PCA Plots

Worksheet
Student Materials

OVERVIEW

This activity introduces students to data visualizations relevant to omics studies. A Principal Component Analysis (PCA) plot is a common method of processing large data sets.

KEY CONCEPTS

- Traditional graphs are an example of one method for visualizing data.
- Principal Component Analysis (PCA) is another method of visualizing data, especially when pertinent to large data sets.

OBJECTIVES

- Students will be able to define what a PCA plot is and justify its role as a data analysis tool.
- Students will be able to read a PCA plot and evaluate the utility of information from it.
- Students will be able to navigate through the NASA GeneLab database and use one tool from the Galaxy platform.

BIOINFORMATICS BITE #3: PCA PLOTS

Part 1: Brainstorm

Brainstorm: "If you measure the expression of 15 genes from 60 mice, and the data come back as a 15×60 table, how do you make sense of all that?" (Ngo, 2018).

What would be the challenges of trying to make sense of that?

Part 2: PCA Plots--Video Introduction

Watch StatQuest: PCA main ideas in only 5 minutes!!! and take notes on the main ideas related to PCA plots.

Be sure to sketch a PCA plot when he starts explaining them, your teacher will look for that as part of your notes!

Video Notes (including PCA plot):

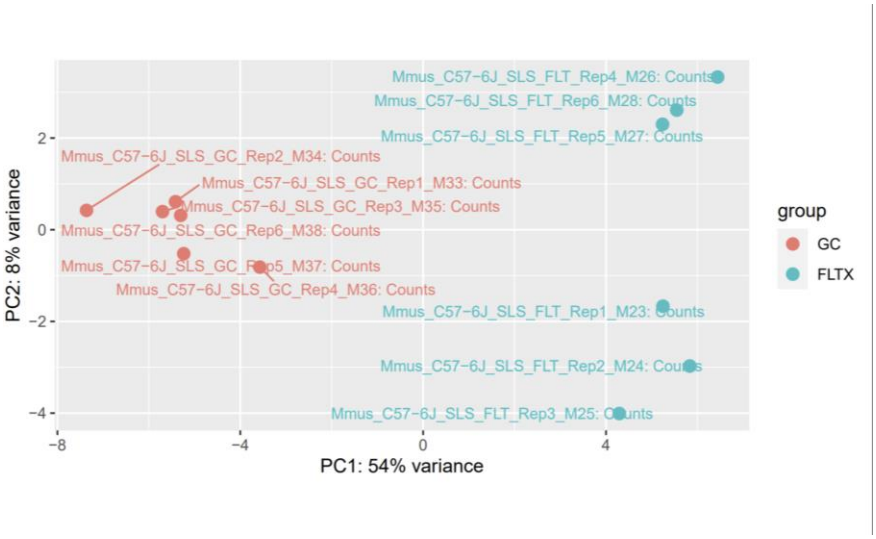
Part 3: PCA Plots--Reading & Notes

Your teacher will assign you a section of the chart to take notes in the highlighted portion of the table below (unless you are assigned *Introduction/1, which will use the article). Don't be afraid to sketch a PCA plot in your notes if it is relevant, it is encouraged! When you are done, you will collaborate with your classmates to fill in the rest of the table.

Section	Notes
Introduction Principal component analysis explained simply	
1. Principal components capture the most variation in a dataset Principal component analysis explained simply	
2. PCA deals with the curse of dimensionality by capturing the essence of data into a few principal components. Principal component analysis explained simply	
3. Dimensions vary in the weights they have on each principal component. Principal component analysis explained simply	

4. How to read a PCA plot **Principal component analysis explained simply**	
***Introduction/1. A PCA plot shows clusters of samples based on their similarity.** **How to read PCA biplots and scree plots**	
Principal Component Analysis **Principal Component Analysis (Excerpt)**	

Part 4: Practice--reviewing the GLDS-104 PCA Plot



Navigate your web browser to https://genelab.nasa.gov.

Click on the Data Repository button

Then search "GLDS-104" in the provided search bar.

Then click on GLDS-104: Rodent Research-1 (RR1) NASA Validation Flight: Mouse soleus muscle transcriptomic and epigenomic data

Use the instructions above to answer the first two questions and enhance your understanding in general! Use the plot for all the questions!

1. **What is *mus musculus* (Mmmus)? (Go to the organisms tab).**

2. **What were the two differing conditions *mus musculus* was exposed to in this experiment? (Go to the study description tab).**

3. **Match the conditions to their color/acronym on the PCA Plot.**

4. **What is the first dimension (PC1) separating?**

5. **And the second dimension (PC2)?**

6. **What can we conclude about the DESeq design (factors, levels) we choose?**

Part 5: Database Practice--Generating the PCA Plot

Just for fun, let's practice briefly navigating the database to see what it looks like to take the final step towards a PCA Plot. We will check off the steps as we go and take a snip and insert a picture of the final product!

Completed? (Add checkmark)	Task to be completed
	Go to https://genelab.nasa.gov
	Click the Analyze Data button.
	Sign in with google into your account. If you do not have an account, use your school gmail to set one up.
	Navigate to Shared Data -> Histories.
	Select GLDS-104: DESeq2 DGE Files.
	Click on the + on the far-right corner to add these files to your history.
	In the pop-up, keep the name as is and click Import.
	Now the files, will appear in your current history.
	Select the file labeled 'DESeq2 plots on data…'. This file will enable us to visualize the results. Remember to select the eye icon to view the results.
	Let's look at the PCA plot. Does it resemble the GLDS-104 PCA plot from the previous step?

Part 6: Learning Aims and Evaluation

Please rate where you personally are at, with regards to the learning aims, at the end of the lesson and why.

Rating Scale

1- I do not understand it at all yet.

2-I understand parts of it, but I need my teacher and/or classmates' support to answer questions.

3-I understand it and can complete an assignment by myself.

4-I understand it so well I can teach others and apply my knowledge to new situations.

Learning Aim #1: Students will be able to define what a PCA plot is and why it is a useful tool for data analysis.

My Evaluation of Learning Aim #1 and Explanation:

Learning Aim #2: Students will be able to read a PCA plot and pull useful information from it.

My Evaluation of Learning Aim #2 and Explanation:

Learning Aim #3: Students will build confidence and experience in navigating the NASA GeneLab database.

My Evaluation of Learning Aim #3 and Explanation:

References

Ngo, L. (2018). How to read PCA biplots and scree plots. BioTuring.com. https://blog.bioturing.com/2018/06/18/how-to-read-pca-biplots-and-scree-plots/

Ngo, L. (2018). Principal component analysis explained simply. BioTuring.com. https://blog.bioturing.com/2018/06/14/principal-component-analysis-explained-simply/

GL4HS Manual: GeneLab for High School Bioinformatics Manual. Blaber, Elizabeth. 2021.

Stamer, J. (2017). StatQuest: PCA main ideas in only 5 minutes!!!!. YouTube.com. https://www.youtube.com/watch?v=HMOI_lkzW08

AUTHOR
Catherine Boileau, Bloomington High School South (Bloomington, Indiana)
Edited by GL4HS Staff