

**CMPSC 301
Data Analytics
Summer 2021**

Quiz 1: Exploring Gene Expression Data With R

Objectives

To enhance the understanding code for data gene expression analysis in Bioinformatics. In this lab, you will learn to read basic plots and to understand the code that produced the plots.

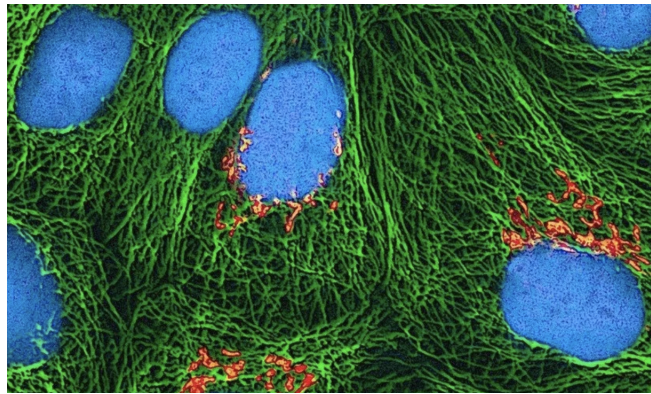


Figure 1: The NCI60 data set was developed by the National Cancer Institute in the 1980's to allow for the study of potential anti-cancer drugs. Concerns probes, which are used to detect molecular compounds, monitor cellular activity, and to deliver therapy in cells, the data allows researchers to investigate potential anti-cancer drugs in a wide variety of cell lines which have been organized (i.e., arrogated) by types of cancer

Reading Assignment

This quiz is open notes and open book. Please read Chapters assigned for this week's lessons which you will find in the class slides, in addition to reviewing your notes. Please take some time to gain experience with using Markdown to write your work. See *Mastering Markdown* <https://guides.github.com/features/mastering-markdown/> for more details about Markdown.

Part of the motivation for this study may be found in the YouTube video that discusses the gene expressions of Planarian worms at the following link: <https://www.youtube.com/watch?v=roZeOBZAa2Q>. Much of the knowledge that we have about the curious genetics of the stem cell function in these organisms was the result of gene expression analysis. Interested readers are invited to learn more about the biology of stem cell growth and regeneration in Planarians from Sánchez *et al* [1].

GitHub Starter Link

<https://classroom.github.com/a/J7cIJ0Y5>

To use this link, please follow the steps below.

- Click on the link and accept the assignment.
- Once the importing task has completed, click on the created assignment link which will take you to your newly created GitHub repository for this deliverable.
- Clone this repository (bearing your name) and work on the deliverable locally.
- As you are working on your deliverable, you are to commit and push regularly. You can use the following commands to add a single file, you must be in the directory where the file is located (or add the path to the file in the command):

```
- git commit <nameOfFile> -m ‘‘Your notes about commit here’’  
- git push
```

Alternatively, you can use the following commands to add multiple files from your repository:

```
- git add -A  
- git commit -m ‘‘Your notes about commit here’’  
- git push
```

Understanding the R Code and Making Sense of the Output

Imagine that you are discussing the code and its output with someone who is new to the project. It is your task to explain to them what each part of the code is doing and what knowledge it brings. In this deliverable you are to explain in clear and meaningful language *what the code is doing and how to make some sense of its output*.

You have been given an R source code file `src/geneExpAnalysis.r` which you are to run using rStudio. **Although the questions to which you are to respond have been included in this source code, you are to enter your responses to the file `writing/report.md` that will be graded.**

Required Deliverables

1. File `writing/report.md`: Your **labeled** answers to the questions (included in this assignment sheet and also in the source code itself `src/geneExpAnalysis.r`)

When you have finished, please ensure that the GitHub web site has your pushed work by visiting your repository at the site. Please see the instructor if you have any questions about assignment submission.

References

- [1] P. W. Reddien and A. S. Alvarado, “Fundamentals of planarian regeneration,” *Annu. Rev. Cell Dev. Biol.*, vol. 20, pp. 725–757, 2004.