

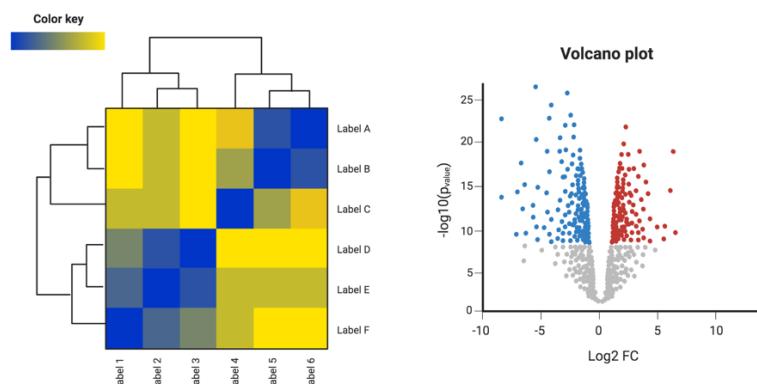
Guidelines for Final Project – Part I

****Please note this is only part of the guidelines, as the semester continues, you will be provided with even more information****

- I. **Overview:** Each student or student group will give an oral presentation describing their analysis of an NGS dataset. These presentations will be given by the student(s) on April 27th, May 2nd, or May 4th. All students are expected to be present for the entire session, provide feedback, and ask questions to the presenter. In addition, each student will be expected to attend an additional session and provide feedback to student groups anonymously. Student's will be asked to select a trail and rationalize why the trail was selected to Dr. Rodriguez. The trail selected will guide the overall goal of the analysis for the NGS dataset downloaded from GEO.

Trail 1: Green Mountain Trail

“Replicate figure(s) in a primary research article ... and change one parameter”



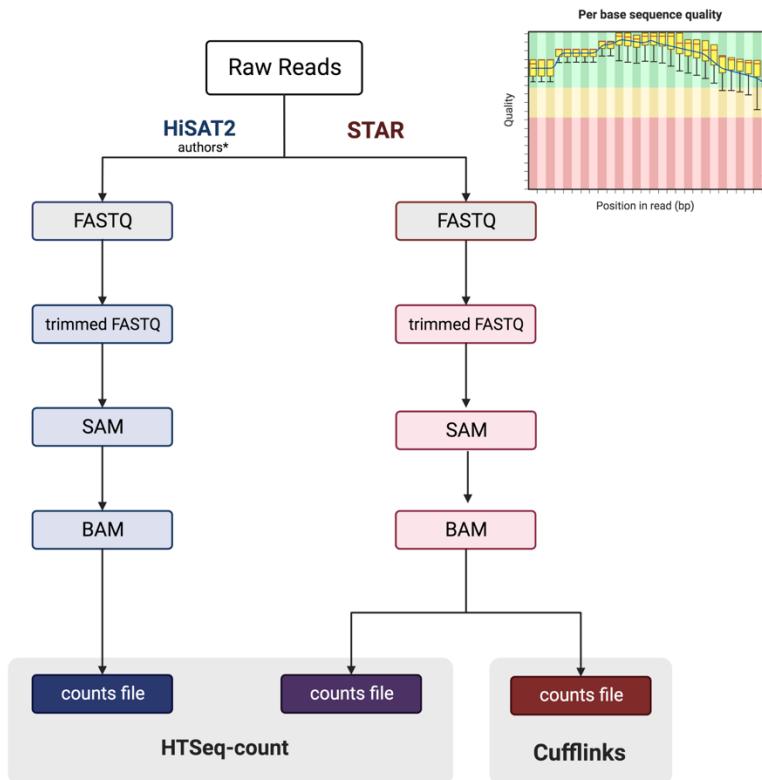
- Change the colors, scale, labels, and/or input



Trail 2: Blue Sky Trail

“Compare and Contrast select parameters within the NGS pipeline and describe how this impacts your final output”

"I'm ready to challenge myself"



- Requires an in-depth understanding of how the authors processed their dataset
- You will be required to give an in-depth overview of the entire analysis from raw reads to DE analysis + figures
- **We will not** cover all pipelines during class - so this does require *independence*



Trail 3: Black Diamond Trail

“Process and Download an NGS dataset to test an original hypothesis”



- Graduate students will be **required** to select this option
- The final figure(s) you generate should not be found within the published dataset
- **Original, creating your own path, you are the trailblazer**



II. General:

- a. Undergraduate students will be allowed to work in groups. No more than 3 students per group. This is an individual assignment for graduate students.
- b. Each student project will be allocated 15 minutes to present their findings and answer questions from the audience. The audience will be able to ask you questions **during** the presentation.
- c. All team members must speak for the same amount of time and be ready to answer questions.
- d.

III. **Presentation content & Grading Rubric:** This will be provided to students in the March 2nd class and will be posted on Blackboard and on the course website.

IV. For your consideration while selecting a primary research article:

- a. RNA-Seq: Not single-cell RNA-Seq or small-RNA-Seq or microarray
- b. Organism: I will provide the indexed genome for mm10, hg38, and hg19 for HISAT2 and STAR*

- c. Number of biological replicates available: at minimum 3 replicates per group required
- d. Date of publication: 2000 to 2023
 - i. Would prefer within the past 10-15 years but you will not be limited unless I determine that the dataset is not usable

V. Timeline for Part I :

	Selecting a dataset	Download dataset	Index Genome	Alignment
Estimated time to complete	1 week	24 hours	1hr – 3 days	3-7 days +
Comment		Per 5G = 1.5 hrs = one sample	Depends on how large the genome is Dependent on alignment strategy	Dependent on the number of samples
Homework Assignment	~50 points	~100 points Why dataset was selected FASTQC + interpretation		~100 points Alignment stats + interpretation Decision to be made on <i>how</i> to proceed based on interpretation

VI. Important Disclosures

- I do not know the quality of the dataset selected as I have not processed these myself. Therefore, depending on what we find we may need to pivot and select another trail to blaze!
- I am most comfortable with human and mouse, other organisms are completely fine to do, however you will be in charge of understanding if for example "...there are pathway analysis tools available for Drosophila..."
- We will hit some unforeseen hiccups – this is completely normal in the realm of bioinformatics!
- I do not have control over how fast or slow your data will process on the VACC. The alignment step is the most COMPUTATIONAL HEAVY STEP of the ENTIRE pipeline. Please do not leave this for the last minute as the VACC will have multiple users!