**#-- Find a Study:**

Early Transcriptomic Changes in the Ileal Pouch Provide Insight into the Molecular Pathogenesis of Pouchitis and Ulcerative Colitis.

<https://pubmed.ncbi.nlm.nih.gov/28221248/>

**#- Search for Data deposited**

*The raw reads and processed data were deposited in NCBI Gene Expression Omnibus (GEO) database with accession number* ***GSE81266****, and Sequence Read Archive (SRA) database with accession number* ***SRP074739***

#-- There are multiple ways to look for the data and download it

1. Search SRR id in google,

- <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81266>

- Go to the BioProject

- SRA Experiments

- Click any one the experiment

- Click on all runs

2. How to download it

<https://sra-explorer.info>

paste your SRP074739 ID and see what it is showing

Select all, add to the collections, then check saved datasets to get links to download the data.

#- Download one file and check how it look like and you are ready to go----

#-- A quick way to get counts and generate a Jupyter notebook file to analyze your data.

**#---------- Steps to Generate Jupiter Notebook:**

BioJupies: <https://pubmed.ncbi.nlm.nih.gov/30447998/>

**Step1:** Please visit the following Link: <https://maayanlab.cloud/biojupies/>

Graphical user interface, application, Teams

Description automatically generated

**Step2:** Click on Get Started

Graphical user interface, application

Description automatically generated

We don’t have our own data. So, we can start working with the already published RNA-seq data.

**Step3:** Click on the Published data

In this case we will use data generated in this Paper: Early Transcriptomic Changes in the Ileal Pouch Provide Insight into the Molecular Pathogenesis of Pouchitis and Ulcerative Colitis. <https://pubmed.ncbi.nlm.nih.gov/28221248/>

Graphical user interface, application

Description automatically generated

**Step4:** Click on GEO

Can you please look into this paper and let me know what is the GEO number? Put that ID and search. You will able to see this.

Graphical user interface, text, application, email

Description automatically generated

**Step5:** Click on the button Analyze:

On this page you can see lot of different type of analysis which you can add. However, this exercise we are doing to generate count file and meta file. Which we will use in R to do all downstream analysis in our next Lecture.

Graphical user interface, application

Description automatically generated

**Step6:** Click on the button Continue:

Then all options very carefully. You should no every possible detail about your experiment. In case of any confusion please read the paper which you are going to analyze. For us this is a Pouchitis paper.

Graphical user interface, application

Description automatically generated

**Step7:** Click on the Generate Notebook: This process will take little time. Let’s understand the data structure which is generated by BioJupies.

Table

Description automatically generated

* Understand the structure of Data
* Download the Dataset and Metadata file. (**Take length of Genes also from Pouchitis- Mask mapped file**). So we will have three files here.
* Save these files in a Specific folder and note the location
* Download Rstudio – I will show you how to do some meaningful analysis on this Data. Such as PCA, Differential expression, Volcano plots, and Pathway enrichment in R

**#---- Material to be distributed for both lectures:**

1. Fastq files both R1 and R2
2. Fastq quality file
3. Count table, Metadata file, and gene length file.
4. Requirement.txt file which includes steps to install Putty, R and all R packages

**Lecture 2:** In the first lecture, I showed you how to generate count data for downstream analysis. In this tutorial we will learn how to generate count data by your own. You will have more flexibility to play with parameters.

1. **Fastq file** and quality assessment file along with the command and tool by which we can access the quality – Demo
2. **Trimmed files:** - Trimmomatic command to show how to remove low quality bases or reads.
3. **Mapping command:** Here, I am taking an example of Kallisto – Kalisto mapping command.
4. Merged all counts to make a final count table.

Once you have the Final count table, and Metadata file. We are not going to do it here. These files you have already generated and saved in the Folder.

**Open R**

Load count data and metadata

Generate TPM counts

Use count data to get VSD data

Use count data to get Differential expression profiles- export expression fold change and p-values and let me know top 5 genes which are enriched in Healthy and top 5 which are enriched in Pouchitis.

Use this object to generate PCA, Heatmap and/or a Cladogram

Pathway analysis: Use any method such as metascape: <https://metascape.org/gp/index.html#/main/step1>