**Student projects**

Each group must submit a final report. **Any evidence of plagiarism more than 20 %  
 will** **immediately result in a grade of 0.**

***Divide yourself into groups of 3-4 members. Each group should choose a different topic.  Once you have picked your topic, write down the name of your chosen project and group members in the appropriate column in this Google Doc:*** [***https://nileuniversity-my.sharepoint.com/:x:/g/personal/mrashad\_nu\_edu\_eg/EebK0TQRUQdPmkrHrBc4U8IBEEA0nIJ4l6y9lU0\_F1gHcw?e=NzdPeU***](https://nileuniversity-my.sharepoint.com/:x:/g/personal/mrashad_nu_edu_eg/EebK0TQRUQdPmkrHrBc4U8IBEEA0nIJ4l6y9lU0_F1gHcw?e=NzdPeU)

Projects

1. **Diabetes** is one of the metabolic disorder diseases and its **type 2** is considered the most prevalent among 90% of diabetic patient. Accordingly, we need to determine the enriched **KEGG pathways between diabetic patients and healthy control** in the following **transcriptomic** **single** data project PRJNA358470.

Steps:

* Download the project data
* Convert it to fastq files using sra-tools
* To easily used, minimize it using sra-tools to 10000 reads

**[The above 3 steps: found in the uploaded tutorial]**

* Apply the quality step on the minimized data using FastQC
* According to the previous step, you will decide if you need to shuffle the data or not [using seqkit tool]
* Trimming for shuffled and unshuffled (optional)
* Pseudoalignment by Kallisto against chromosome 22  
  **Note: there’s a little difference between using of kallisto in paired and single reads. Each single file should have a kallisto ouput folder “little trick need little search”.**
* Deferential expression using deseq2 using **tximport.r script uploaded for you**
* Functional enrichment analysis using Gprofiler GUI tool or R package
* Interpretation for the analysis [Literature review]

1. **Alzheimer is the most common cause of dementia,** a general term for memory loss and other cognitive abilities serious. Alzheimer's disease accounts for 60-80% of dementia cases. Accordingly, we need to determine the enriched **GO terms between healthy and Late-onset Alzheimer disease (LOAD)** in the following **transcriptomic** **single** data PRJNA714081.

Steps:

* Download the project data
* Convert it to fastq files using sra-tools
* To easily used, minimize it using sra-tools to 10000 reads

**[The above 3 steps: found in the uploaded tutorial]**

* Apply the quality step on the minimized data using FastQC
* According to the previous step, you will decide if you need to shuffle the data or not [using seqkit tool]
* Trimming for shuffled and unshuffled (optional)
* Pseudoalignment by Kallisto against chromosome 22  
  **Note: there’s a little difference between using of kallisto in paired and single reads. Each single file should have a kallisto ouput folder “little trick need little search”.**
* Deferential expression using deseq2 using **tximport.r script uploaded for you**
* Functional enrichment analysis using Gprofiler GUI tool or R package
* Interpretation for the analysis [Literature review]

1. **Parkinson's** disease is a brain disorder that leads to shaking, stiffness, and [difficulty with walking, balance, and coordination](https://www.nia.nih.gov/health/topics?field_a_to_z_topics_index_target_id=Movement%20disorders%20and%20problems%20%283762%29). Accordingly, we need to determine the enriched **KEGG pathways between male and female** in the following **transcriptomic** **paired** data PRJNA732831.

Steps:

* Download the project data
* Convert it to fastq files using sra-tools
* To easily used, minimize it using sra-tools to 10000 reads

**[The above 3 steps: found in the uploaded tutorial]**

* Apply the quality step on the minimized data using FastQC
* According to the previous step, you will decide if you need to shuffle the data or not [using seqkit tool]
* Trimming for shuffled and unshuffled (optional)
* Pseudoalignment by Kallisto against chromosome 22  
  **Note: there’s a little difference between using of kallisto in paired and single reads. Each single file should have a kallisto ouput folder “little trick need little search”.**
* Deferential expression using deseq2 using **tximport.r script uploaded for you**
* Functional enrichment analysis using Gprofiler GUI tool or R package
* Interpretation for the analysis [Literature review]

1. **Breast cancer** is the second most common cancer in women after skin cancer. Breast cancer develops as a result of genetic mutations or damage to DNA. Accordingly, we need to determine the variants (mutations)in the following **genomic** **single** data [PRJNA188274](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA188274)  
     
   Steps:

* Download the project data
* Convert it to fastq files using sra-tools
* To easily used, minimize it using sra-tools to 10000 reads

**[The above 3 steps: found in the uploaded tutorial]**

* Apply the quality step on the minimized data using FastQC
* According to the previous step, you will decide if you need to shuffle the data or not [using seqkit tool]
* Trimming for shuffled and unshuffled (optional)
* Alignment using BWA
* Variant calling using BCFtools
* Visualization using IGV
* Interpretation for the analysis [Literature review]

1. **Leukemia** is cancer of the body's blood-forming tissues, including the bone marrow and the lymphatic system. Accordingly, we need to determine the available variantsin the following **genomic** **paired** data PRJNA668367

Steps:

* Download the project data
* Convert it to fastq files using sra-tools
* To easily used, minimize it using sra-tools to 10000 reads

**[The above 3 steps: found in the uploaded tutorial]**

* Apply the quality step on the minimized data using FastQC
* According to the previous step, you will decide if you need to shuffle the data or not [using seqkit tool]
* Trimming for shuffled and unshuffled (optional)
* Alignment using BWA
* Variant calling using BCFtools
* Visualization using IGV
* Interpretation for the analysis [Literature review]

**Requests:   
A- Your report (ranges 6-9 pages) [IEEE template] should contain:**

1. Abstract (1/2 page)
2. An introduction that provides a brief background and explains the problem (1-2 page).
3. A methods section that describes in detail the exact steps you took with screenshots (2-3 pages).
4. A results/discussion section that presents and discusses the results (2-3 pages).
5. Conclusion (1/2 page)
6. References using Footnotes or Mendeley
7. A contributions section that details what each member of the group did.

(This includes everything, as well as preparing and giving the presentation, and writing the report).

**B-** **On the Presentation Day**, you will get to present your work in progress to the rest of the groups. Each group will have 15 -20 minutes for the presentation and 10 minutes for questions for instructor, TA and anyone of other groups.

**Note: Automatic bash script, bonus maximum 10 %  
Deadline:**   
Report submission: 6 February 2022, 23:59 pm  
Presentation: 7 February 2022, 12:30 pm