

# 21BIO211 BIOLOGICAL SYSTEMS

## ASSIGNMENT - 3

Name – Gagan Dwaz

Roll no – AM.EN.U4AIE21028

### Q1) What are the available Gene expression databases

- 1) NCBI Sequence Read Archive (SRA): This public repository, maintained by the National Center for Biotechnology Information (NCBI), archives raw sequencing data, including gene expression data generated through RNA-seq experiments. Researchers can access and download this raw sequencing data from various studies, making it a valuable resource for gene expression analysis.
- 2) GTEx Portal: The Genotype-Tissue Expression (GTEx) Portal provides a comprehensive collection of gene expression data across multiple human tissues. It includes data from post-mortem donors and enables researchers to study gene expression patterns in various tissues and cell types. The GTEx project is an ongoing effort to obtain a comprehensive understanding of gene expression regulation across different tissues and cell types, making it an important resource for understanding human biology and disease.
- 3) ArrayExpress: ArrayExpress is a database hosted by the European Bioinformatics Institute (EBI). It provides a comprehensive repository of gene expression data across multiple organisms, including humans. ArrayExpress accepts data generated by diverse high-throughput technologies and offers curated metadata and analysis tools. This makes it a valuable resource for researchers who want to access and analyze gene expression data from multiple sources.
- 4) Expression Quantitative Trait Loci (eQTL) Databases: eQTL databases compile information on the relationship between genetic

variation and gene expression levels. Some popular eQTL databases include GTEx eQTL Portal, BIOS QTL Browser, and Genotype-Tissue Expression eQTL Browser. These databases are helpful for researchers who want to study the genetic basis of gene expression regulation and its implications for human disease.

- 5) Human Protein Atlas: The Human Protein Atlas (HPA) is a database that focuses on protein expression patterns in human tissues and cells. It incorporates gene expression data obtained through immunohistochemistry and antibody-based assays. This makes it a valuable resource for researchers who want to study protein expression patterns in different tissues and cell types.
- 6) European Nucleotide Archive (ENA): ENA is a database managed by EBI that encompasses a wide range of molecular data, including gene expression data. It accepts data from different sequencing technologies and provides access to a diverse array of biological datasets. This makes it a valuable resource for researchers who want to access and analyze gene expression data from multiple sources.
- 7) Gene Expression Omnibus (GEO): GEO is a widely used public repository managed by NCBI. It houses a vast collection of gene expression data from various high-throughput technologies, such as microarrays and RNA sequencing (RNA-seq). This makes it a valuable resource for researchers who want to access and analyze gene expression data from multiple sources.
- 8) Single-cell Expression Atlas: Single-cell Expression Atlas, hosted by EBI, focuses on gene expression at the single-cell level. It provides data from single-cell RNA sequencing (scRNA-seq) experiments, allowing researchers to explore gene expression patterns in specific cell types or during different developmental stages. This makes it a valuable resource for researchers who want to study gene expression patterns at a high resolution.
- 9) The Cancer Genome Atlas (TCGA): TCGA is a project that has generated comprehensive genomic and gene expression data for various cancer types. It comprises a rich resource for cancer researchers, housing data from thousands of patient samples. This

makes it a valuable resource for researchers who want to study the genetic basis of cancer and develop new therapeutic approaches.

## **Q2) How do you retrieve the Gene expression Data related to SARS COV2**

1. Go to a relevant gene expression database's website, such as GEO or ArrayExpress.
2. Use the search function or browse the available datasets to find studies or experiments related to SARS or COVID-19.
3. Combine keywords such as "SARS-CoV-2," "COVID-19," or "coronavirus" with "gene expression" or "RNA-seq" to narrow down the search.
4. Check the metadata and experimental information of any identified datasets to confirm that they include gene expression data specifically related to SARS.
5. Pay attention to sample types, experimental conditions, and relevant keywords mentioned in dataset descriptions.
6. Once you've identified the dataset(s) you need, download them directly from the database's user interface.
7. Data files can be downloaded in various formats, including raw data, processed expression values, and normalized expression data.
8. If necessary, access additional information such as experimental protocols and sample characteristics, as well as related publications.
9. Remember to cite the appropriate sources and datasets when using the retrieved gene expression data in your research or analysis.