

Amrita Vishwa Vidyapeetam

Biological Systems – 21BIO211

Lab Assignment – 3

NAME:- J Viswakse

ROLL.NO:- AM.EN.U4AIE21035

1. What are the available Gene expression databases

Solution:-

There are several gene expression databases available that provide valuable resources for studying gene expression patterns and related data. Here are some notable gene expression databases:

Gene Expression Omnibus (GEO): GEO, maintained by the National Center for Biotechnology Information (NCBI), is a public repository that stores a wide range of gene expression data, including microarray, RNA-seq, and other high-throughput experimental data. It hosts data from various organisms and provides tools for data retrieval and analysis.

The screenshot shows the GEO homepage with a search bar at the top right. Below the search bar are three main sections: 'Getting Started' (with links to Overview, FAQ, About GEO DataSets, About GEO Profiles, About GEO2R Analysis, How to Construct a Query, and How to Download Data), 'Tools' (with links to Search for Studies at GEO DataSets, Search for Gene Expression at GEO Profiles, Search GEO Documentation, Analyze a Study with GEO2R, Studies with Genome Data Viewer Tracks, Programmatic Access, FTP Site, and ENCODE Data Listings and Tracks), and 'Browse Content' (with statistics: DataSets: 4348, Series: 203181, Platforms: 25148, Samples: 5923684). At the bottom are sections for 'Information for Submitters' (Login to Submit, Submission Guidelines, Update Guidelines) and 'MIAME Standards' (Citing and Linking to GEO, Guidelines for Reviewers, GEO Publications).

ArrayExpress: ArrayExpress is a database of functional genomics experiments, including gene expression data. It is hosted by the European Bioinformatics Institute (EBI) and provides access to a diverse collection of high-throughput data, including microarray, RNA-seq, ChIP-seq, and more. ArrayExpress also offers data analysis tools and supports data deposition.

The screenshot shows the ArrayExpress in BioStudies website. At the top, there's a navigation bar with links to EMBL-EBI Training, On-demand training, Online tutorial, and ArrayExpress in BioStudies. The main header is 'ArrayExpress in BioStudies'. Below the header, there's a 'Quick tour' button and a 'Course overview' sidebar with links to 'Search within this course', 'What is the ArrayExpress collection?', 'Searching data in the ArrayExpress collection', 'Components of an ArrayExpress study', 'Getting data from the ArrayExpress collection', 'Submitting data to the ArrayExpress collection', 'Your feedback', and 'Get help and support on the ArrayExpress collection'. The main content area features a section titled 'Components of an ArrayExpress study' with a sub-section 'Bulk RNA-seq of cerebral organoids investigating the role of CHD8 in cortical development'. This section includes a table of data files (e.g., A2308_WT_rep1_ResultsPerformance.xls, A2308_WT_rep2_ResultsPerformance.xls, etc.) and a detailed description of the experiment. At the bottom, there's a footer with a Creative Commons license logo and a statement: 'All materials are free cultural works licensed under a'.

The Cancer Genome Atlas (TCGA): TCGA is a comprehensive cancer genomics database that includes gene expression data for various cancer types. It provides a wealth of genomic, transcriptomic, and clinical data for researchers studying cancer biology and personalized medicine. We can also get the gene database from the nci website which is officially maintained by US government.

The screenshot shows the official NCI website for the Center for Cancer Genomics. The main navigation bar includes links for Research, Access Data, Funding, News & Events, About CCG, Contacts & Help, and a search bar. Below this, a breadcrumb trail shows Home > Research > Genome Sequencing > The Cancer Genome Atlas Program (TCGA). The main content area features a section titled "The Cancer Genome Atlas Program (TCGA)" with a brief description of the program's history and impact. Two images are displayed: one showing a diverse group of people and another showing a globe with biological concepts like "Patterns", "Processes", and "Pathways".

Expression Atlas: This is the website for all organisms having the gene expressions . It consists all type of databases belong to organisms. Which is under the EBI. It integrates data from various resources and provides interactive visualization and analysis tools.

The screenshot shows the Expression Atlas homepage. It features a search bar at the top with the placeholder "Search across 65 species, 4,315 studies, 153,212 assays". Below the search bar are sections for "Gene set enrichment" and "Gene properties". A central grid displays icons for various organisms: Homo sapiens (1508 experiments), Mus musculus (1241 experiments), Rattus norvegicus (773 experiments), Drosophila melanogaster (104 experiments), Gallus gallus (3 experiments), and Caenorhabditis elegans (31 experiments). Navigation tabs for "Animals", "Plants", and "Fungi" are visible.

GTEx Portal: The Genotype-Tissue Expression (GTEx) Portal provides gene expression data from multiple human tissues across different individuals. It includes RNA-seq data and accompanying clinical information, offering insights into gene expression regulation and tissue-specific patterns.

The screenshot shows the GTEx Portal homepage. At the top, there is a banner about samples mapped to the HuBMAP Common Coordinate Framework. Below the banner, the "Resource Overview" section highlights the "Current Release (V8)". It includes links for "Tissue & Sample Statistics", "Tissue Sampling Info (Anatomogram)", "Access & Download Data", "Release History", and "About GTEx". A detailed description of the GTEx project is provided, mentioning its goal to build a comprehensive public resource to study tissue-specific gene expression and regulation. The "Explore GTEx" section contains several tabs: "Browse" (By gene ID, By variant or rs ID), "By Tissue", "Histology Viewer", "Single Cell" (Data Overview, Multi-Gene Single Cell Query), "Expression" (Multi-Gene Query, Transcript Browser), and "QTL" (Locus Browser (Gene-centric)). A search bar at the top right allows users to search for genes or SNP IDs.

2) How do you retrieve the Gene expression Data related to SARS COV2 .

Approach:-

We can download and access the data related to SARS COV2 from the above mentioned any database websites.

1. From ncbi.nlm.nih.gov website about National Library of Medicine(ncbi)
2. Go to the website and search for SARS COV2(syndrome coronavirus 2)

Results found in 25 databases

TAXONOMY

Severe acute respiratory syndrome coronavirus 2

Severe acute respiratory syndrome coronavirus 2 is a below-species classification of Severe acute respiratory syndrome-related coronavirus

Taxonomy ID: 2697049

NCBI SARS-CoV-2 resources

NCBI Virus Browse and download

Was this helpful?

BLAST

Use our new Betacoronavirus database for SARS-CoV-2 genome sequence analysis

3. You can get the results like the above figure.
4. By scrolling down the ui you can see the Genome. From there you can download the Genomes or datasets or curl

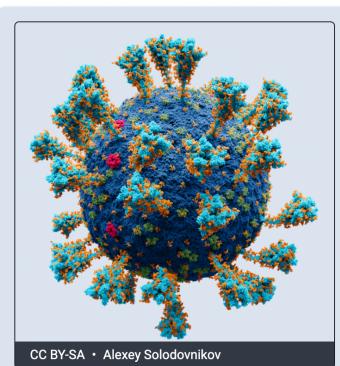
Severe acute respiratory syndrome coronavirus 2

Severe acute respiratory syndrome coronavirus 2 is a below-species classification of *Severe acute respiratory syndrome-related coronavirus*

[Browse taxonomy](#)

Current scientific name	Severe acute respiratory syndrome coronavirus 2
Acronym	SARS-CoV-2
Genome type	ssRNA(+)
NCBI Taxonomy ID	2697049

For more details see [NCBI Taxonomy](#)
View the legacy [Genome page](#)



CC BY-SA • Alexey Solodovnikov

Image may not have been verified for accuracy by NCBI Taxonomy.

Genome

[Browse all genomes in NCBI Virus](#)

[Download all genomes](#)

[datasets](#)

[curl](#)

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome
RefSeq NC_045512.2

[Download](#)

External links

[Encyclopedia of Life](#)

[Wikipedia](#)

Genome

Browse all genomes in NCBI Virus

[Download all genomes](#) [datasets](#) [curl](#)

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome
RefSeq NC_045512.2

[Download](#)

Genome size	29.9 kb
Viral segments	1
Genes	11

Annotation from NCBI RefSeq

SARS-CoV-2 reference genome (NC_045512.2)

5. You can download the FASTA file containing the accordingly type of sequences by clicking the Download button.

Download

Download the SARS-CoV-2 reference genome

File types

- Genomic sequences (FASTA)
- Protein sequences (FASTA)
- Coding sequences (FASTA)

Estimated download size is 976.6 kB or less

Genome count: 1

Name your file

SARS-CoV-2.zip

Your selected data and a detailed data report will be downloaded as a ZIP file

[Cancel](#) [Download](#)

Note:- There are so many other websites for retrieve the SARS COV2 websites that mentioned in Question-1.