Basic Bioinformatics (BE304)

## Lab Assignment 5

## Total Marks: 40

* 1. Open the NCBI home page (<https://www.ncbi.nlm.nih.gov/>)
  2. Change the database option from ‘All database to PubMed’
  3. Enter the keyword for which you want to access the publications. For eg use “genomics”
  4. How many articles are available under the current keyword?

.Ans. 1,718,303 articles

* 1. Download the ‘Publication vs Year’ data and plot the graph for the same in a spreadsheet
  2. Download ‘Abstract’ for the first 10 ‘Review’ articles.

**(5)**

* 1. Open the NCBI home page (<https://www.ncbi.nlm.nih.gov/>)
  2. Change the database option from ‘All database to Gene’
  3. Enter the gene name “indy”
  4. Download the gene sequence (FASTA) of “indy” from *Drosophila melanogaster* genome.
  5. Report the number of exons found within the gene.

Ans. 11 exons

**(5)**

* 1. Open the NCBI home page (<https://www.ncbi.nlm.nih.gov/>)
  2. Change the database option from ‘All database to Protein’
  3. Search for accession “YP\_009724390.1”
  4. Which protein did you find?

Ans. surface glycoprotein

* 1. Download the protein sequence (FASTA)
  2. Open <https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>
  3. Goto CD-Search tool
  4. Paste the downloaded FASTA sequence as a query and submit the data
  5. Report the different conserved domains present within your protein sequence

Ans.

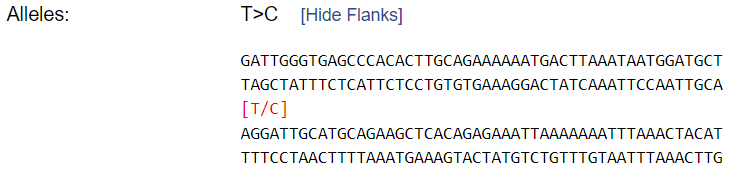
# 

**(5)**

* 1. Open the following link [www.ncbi.nlm.nih.gov/SNP](http://www.ncbi.nlm.nih.gov/SNP)
  2. Enter rs25 in the search box.
  3. Which organism’s sequence is present?

Ans 

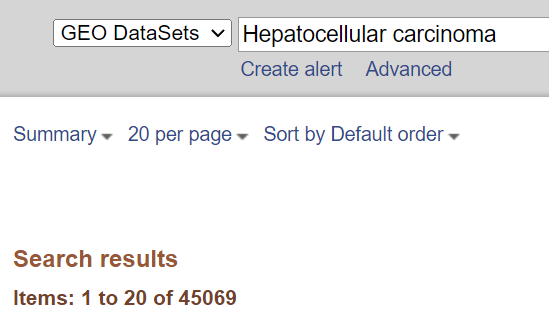
* 1. Mention the nucleotides responsible for this SNP

Ans 

**(5)**

* 1. Open the NCBI home page (<https://www.ncbi.nlm.nih.gov/>)
  2. Change the database option from ‘All database to GEO datasets’
  3. Enter “Hepatocellular carcinoma” in the search box.
  4. How many hits did you find?

Ans. 45069 hits

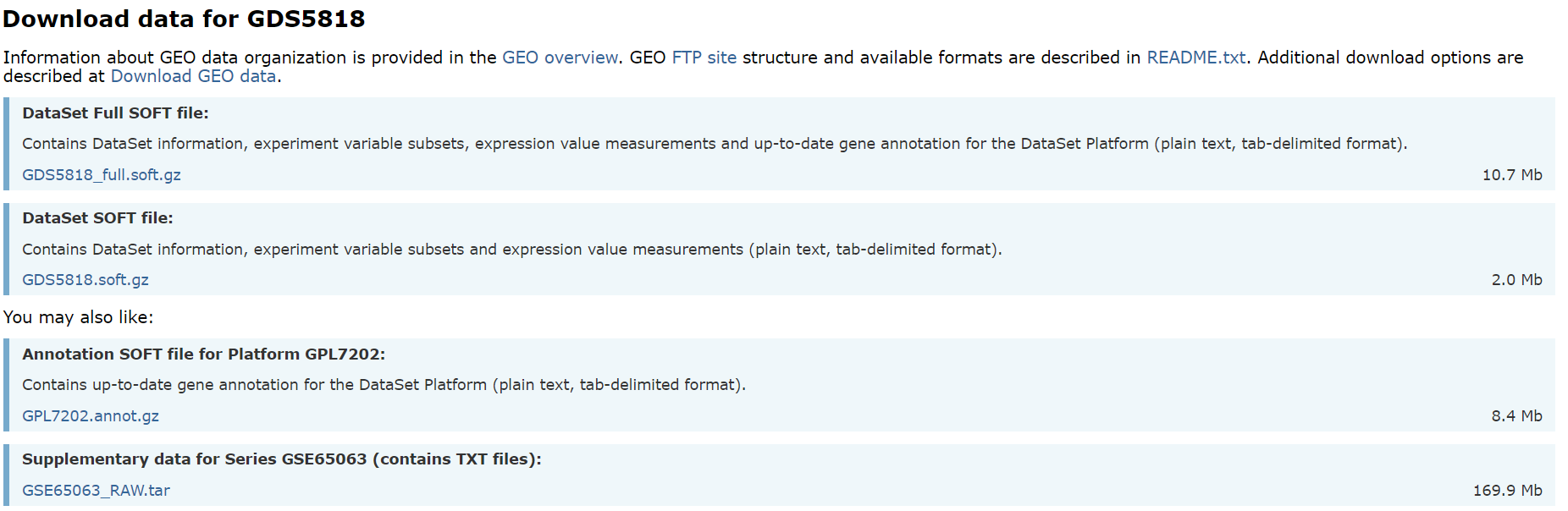


* 1. Download data using Accession: GDS5818.

Ans. file downloaded [GDS5818.soft.gz](https://ftp.ncbi.nlm.nih.gov/geo/datasets/GDS5nnn/GDS5818/soft/GDS5818.soft.gz)

* 1. How many file formats are available for download.

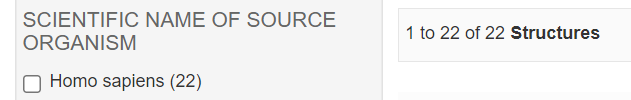
Ans. 3 file formats are available



**(5)**

* 1. Open the <http://www.rcsb.org/pdb/home/home.do>
  2. Enter the keyword “apoptosome”
  3. Restrict your search to *Homo sapiens*
  4. How many protein structures are available for *Homo sapiens* apoptosome?

*Ans. 22 protien structures*

**

* 1. Select any one of the structure and report the following;

Ans. 

* + 1. Method of discovery: ELECTRON MICROSCOPY
    2. Resolution of structure: 4.10 Å
    3. DOI of the publication: 19/10/2016
  1. Download the PDB file of the selected structure.

**(5)**

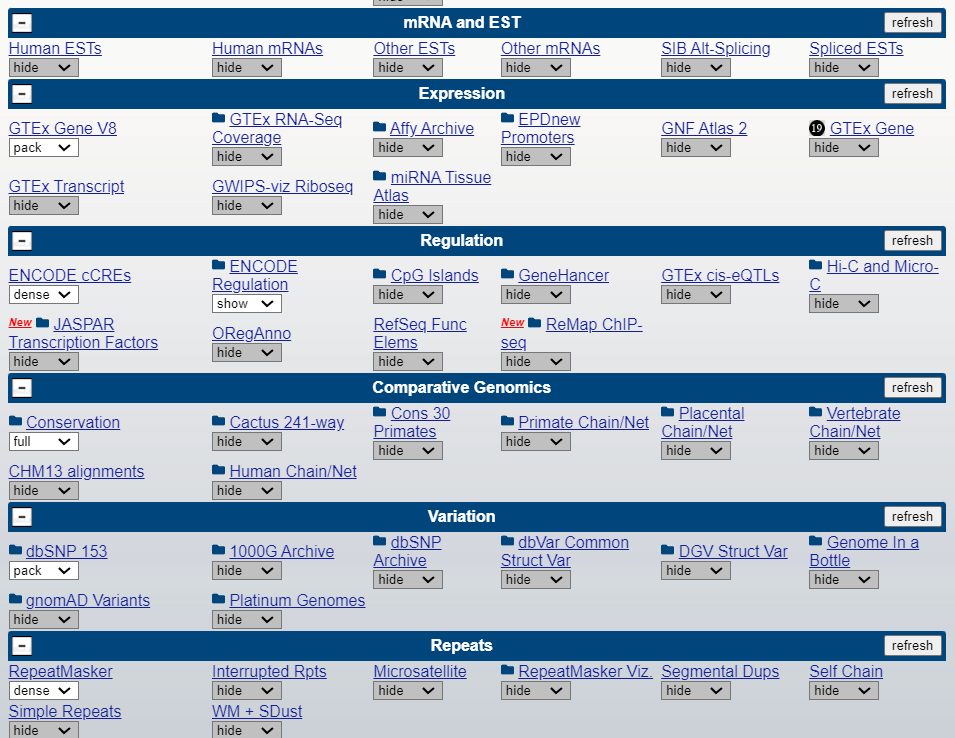
* 1. Open the following link <https://www.genome.jp/kegg/>
  2. Click on KEGG Pathway
  3. Search for “Carbon metabolism”
  4. Download the KEGG Map for the above query.
  5. Map the pathway for TCA cycle within the obtained Carbon Metabolism Map

Done, check files

**(5)**

* 1. Open the following link <https://genome.ucsc.edu>
  2. Go to genome browser.
  3. Select the appropriate server (preferably Asia)
  4. Click on GO option.
  5. Explore the sections Mapping and Sequencing, Genes and Gene Predictions, Phenotype and Literature, mRNA and EST, Expression, Regulation, Comparative Genomics, Variation and Repeats.





**(5)**