Programming for Biomedical Informatics

Lecture 10 "Differential Gene Expression"

https://github.com/tisimpson/pbi

Ian Simpson
ian.simpson@ed.ac.uk

Coding

Differential Gene Expression Analysis Using NCBI-GEO & PyDESeq2

https://www.ncbi.nlm.nih.gov/geo/

https://github.com/owkin/PyDESeq2

Retrieving Data from NCBI GEO

Programmatic Access to NCBI GEO

https://www.ncbi.nlm.nih.gov/geo/info/geo_paccess.html

Download Options
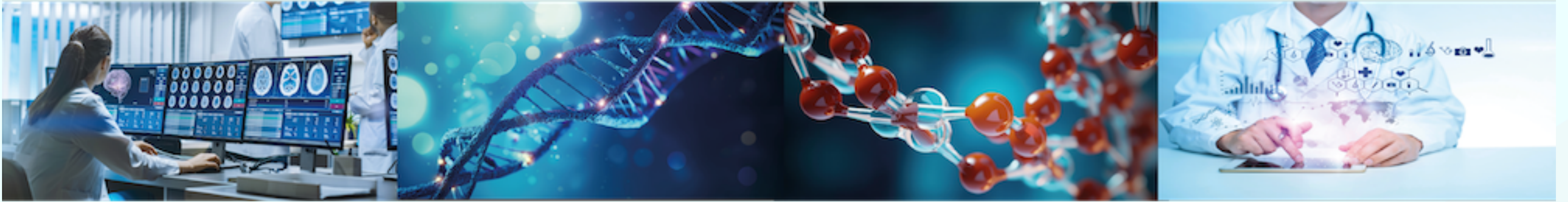
https://www.ncbi.nlm.nih.gov/geo/info/download.html

Workflow

• Use the qualifier fields in Entrez GEO DataSets to fine-tune a search

• Construct the appropriate eSearch query in your script/program

• Run the query, retrieve the results in the form of UIDs or history parameters (query_key and WebEnv) as needed

• Run eSummary or eFetch and/or eLink depending on your needs to retrieve the final metadata or accessions.

• If you need to download full records or supplementary files, use the accession information to construct an FTP URL and download the data.

Programming for Biomedical Informatics

Next Lecture - Tuesday 29th October
"Biological Networks"

Ask Questions on the EdStem Discussion Board

https://github.com/tisimpson/pbi

Background