**Gene Set Enrichment with Enrichr (R)**

**Purpose**

This repository provides an R workflow for performing gene set enrichment analysis using the Enrichr API.  
The workflow reads a list of gene symbols from an Excel file (single column, no header), runs enrichment across selected Enrichr libraries, and saves the results in CSV format.  
It is intended for reproducibility and transparency as supplementary material to a manuscript.

**Requirements**

* R (version 4.0 or later recommended)
* R packages: enrichR, readxl

The script will automatically attempt to install any missing packages.

**Input format**

* Input file: Excel .xlsx
* Structure: one gene symbol per row, **no header row**
* Example:

TP53

BRCA1

EGFR

**Databases used**

The following Enrichr libraries are requested by default (names must match Enrichr exactly):

* WikiPathways\_2024\_Human
* GO\_Cellular\_Component\_2023
* GO\_Biological\_Process\_2023
* GO\_Molecular\_Function\_2023
* ChEA\_2022
* DrugMatrix
* Jensen\_TISSUES
* Jensen\_COMPARTMENTS
* HMDB\_Metabolites

If Enrichr updates or renames libraries, users should verify the current list of available libraries and adjust accordingly.

**Output**

* One CSV file is generated per database.
* Files are named Enrichr\_<DatabaseName>.csv and saved in the working directory.
* Each CSV contains the Enrichr results table (terms, overlap, adjusted p-value, and enriched genes).

**Reproducibility notes**

* Database availability may vary over time; confirm library names using Enrichr directly.
* Enrichment results depend on the version/date of Enrichr queried.

**Troubleshooting**

* If no results are saved: verify internet connection and that the requested databases match Enrichr’s available libraries.
* If the script reports “None of the requested databases matched”: run listEnrichrDbs() in R to view exact library names.
* Ensure the input Excel file has no header and only one column of valid gene symbols.