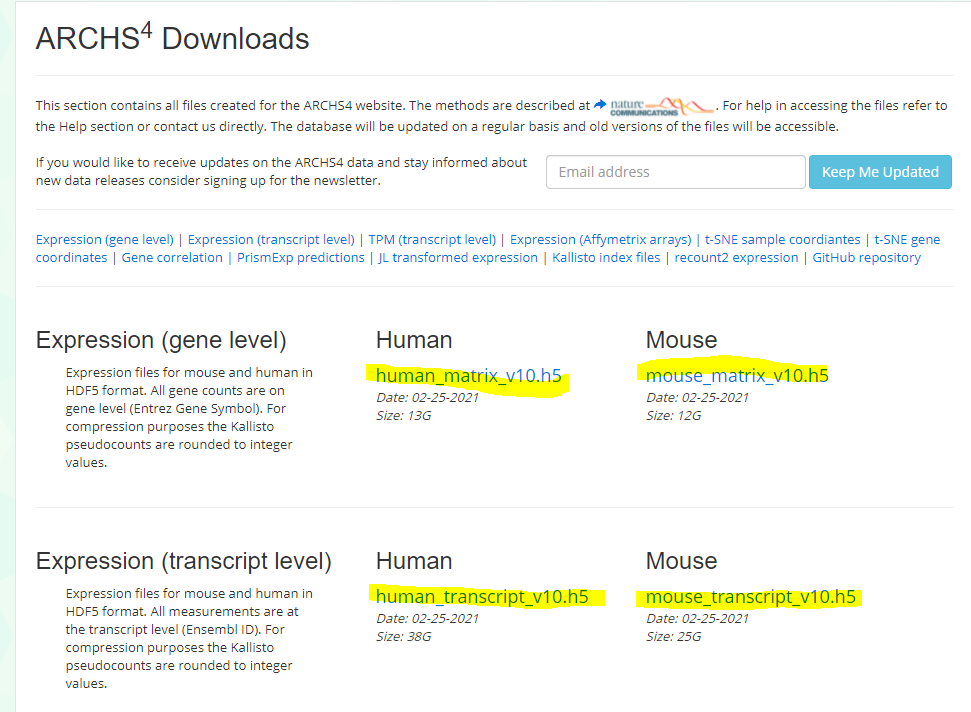
#Reads App

The reads app allows user to search for GSE ids and download their data sets. Download counts data for 27,215 human and mouse datasets from ARCHS4 v6 and 32,151 datasets from DEE2 for 9 model organisms.

##ARCHS4 data

Download four datasets from https://maayanlab.cloud/archs4/download.html

(Mouse and human matrices and transcripts)

Create a “readCounts” and a “data” folder in the idep repository:

...\idep-master\idep-master\data\readCounts

Save the archs4 files in the “readCounts” folder.

Next, run each line in the Reads\_DB\_ARCHS4\_DEE2.R file. This file builds the database GEO.db from metadata in the ARCHS4 files and metadata downloaded using the “getdee2” library. The database GEO.db contains two tables, GSEInfo and SampleInfo.

Graphical user interface

Description automatically generated

The table GSEInfo contains all 59,366 unique GSE IDs from both ARCHS4 and DEE2, in addition to their species, number of samples, title, and summary. The table sampleInfo contains 2,280118 rows with the fields in the figure above. Since some GSMs have multiple GSEs, there are only 1,039,005 unique GSMs. These tables are used in the UI of reads app. The user selects a GSEid from the GSEInfo table, and the app displays its corresponding GSMs as a “preview” of the counts and transcript data.

## Reads\_DB\_ARCHS4\_DEE2.R

The first section of the Reads\_DB\_ARCHS4\_DEE2.R code loads and reads the files downloaded from ARCHS4. The ARCHS4 data did not contain missing GSE ids, however there are GSMs with multiple GSEs. Thus, instead of using the sample\_series\_id field from the ARCHS4 file, the code takes the unique GSMS from the ARCHS4 “sample\_name” field and their matching GSEs from the “GEOmetadb” library. Running the command, “geometadb::getSQLiteFile()” downloads the file “GEOmetadb.sqlite”. This file has 3 tables with fields we’ll use.

Graphical user interface, application

Description automatically generated

One table is “gse\_gsm” with two fields, “gse” and “gsm”. The code finds every unique sample name found in the archs4 files that contains the letters “gsm” and retrieves their corresponding GSE ids from the gse\_gsm table.

Next, the “title” and “summary” fields from the “gse” table are added. These are the fields that will make up the “GSEinfo” table of “GEO.db”.

Similarly, the “archs4” table and the “gsm” table are joined to add the fields “source\_name\_ch1” and “title” from the “gsm” table. These fields will be in the “sampleInfo” table of “GEO.db”.

The process outlined above is repeated in a similar way for the dee2 data.

Diagram

Description automatically generated with medium confidence

The last section of code combines the GSEinfo tables and sampleInfo tables for archs4 and dee2, then writes them into a new database called “GEO.db”. Once GEO.db is created, you should be able to use ui.R or server.R to run the app.