GEOsearch User Interface Manual

Zhicheng Ji, Hongkai Ji

July 10, 2017

1 GEOsearch Online GUI

GEOsearch Online GUI can be launched directly online: https://zhiji.shinyapps.io/GEOsearch/. The only needed software is the web browser and no additional software is required to be installed.

2 Installation of GEOsearch Software

GEOsearch software can be installed via Github. Users should first have R installed on their computer before installing GEOsearch. R can be downloaded here: http://www.r-project.org/. Users can then install the latest version of GEOsearch package via Github by running the following commands in R (the lastest version of R is recommended):

```
if (!require("devtools"))
  install.packages("devtools")
devtools::install_github("zji90/GEOsearch")
```

After that, one can launch the user interface by running the following commands in R:

```
library(GEOsearch)
GEOsearchui()
```

3 GEOsearch GUI Overview

The GEOsearch GUI consists of three main parts: the main menu on the top of the GUI where users can switch between different functions (Figure 1A), the sidebar panel on the left side where users can specify options (Figure 1B), and the main panel where the results will appear (Figure 1D).

GEOsearch has four main functions: searching GSE or GSM terms, key word enrichment analysis, obtaining details of all samples in GSE series, and generating a shell file to batch download samples in GSE series.

4 Performing Search

The first and most important step is to perform the search (Figure 1). GEOsearch provides two modes: traditional search and extended search, and GSM samples can also be searched in addition to GSE series (Figure 1B). Note that searching GSM samples could take a long time in some cases. To start the search one can simply enter the search term ("Oct4" in this example) in the text box "Enter Search Term" and click "Start Search" button (Figure 1C). By default GEOsearch will perform traditional search and will return exactly the same records as searching the same term in GEO. One can also switch to "Extended Search" using the radio button above the text box (Figure 1B). For extended search GEOsearch will split the term being searched into separate words and check whether each word is a gene name. For gene names GEOsearch will find all its gene aliases (e.g. Pou5f1, Oct3, Oct3/4, etc.) and perform additional searches by replacing the original gene name with each gene alias. The search results are then integrated and displayed in a compact table (Figure 1D). By default GEOsearch shows four columns that contain the most important information. One can use the "Column visibility" tab (Figure 1D) to display additional columns.

After an initial search, one can further narrow down the search results by performing a second-round search. For example, one can type in "pluripotency" in the text box above the "Title" column (Figure 1E) and type "array" in the text box above the "Type" column to only keep the search results that have the word "pluripotency" in title and the word "array" in type. This function allows one to conveniently focus on the GSE records of most interest without having to perform complicated programmatic search in GEO. One can also select the results that contain the keyword in any of the field by using the search box (Figure 1F).

One can change the order of the rows and columns using the computer mouse. To change the order of the columns, click and hold the title of a column and hover to a new place. To change the order of the rows, click and hold the ID of a row and hover to a new place. To sort the results according to a column, click on the title of a column and the results will be sorted in ascending or descending order. One can select the GSE records by clicking on the rows. The details of each samples for the selected GSE can be found in "Sample Details".

One can download the contents of the current page as CSV, Excel or PDF (Figure 1D). One can also download all search results using the download button (Figure 1G). Click "Download Selected Part" if only the selected rows are to be downloaded.

5 Key Word Enrichment

GEOsearch enables one to effectively summarize the biological contexts (cell types, tissues and diseases) of the search results. Lists of cell types, tissues and diseases commonly used in biological experiments were precompiled from ATCC (www.atcc.org). After one selects which lists to use (Figure 2A), GEOsearch displays the number of GEO records whose titles or descriptions contain each key word (Figure 2B). GEOsearch will also perform a simple enrichment analysis to test whether each key word is significantly enriched in the search results

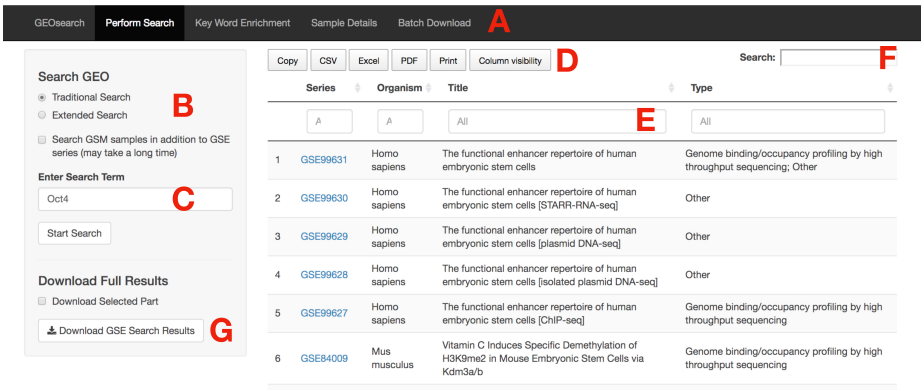

Figure 1: Screenshot demonstrating: Perform Search

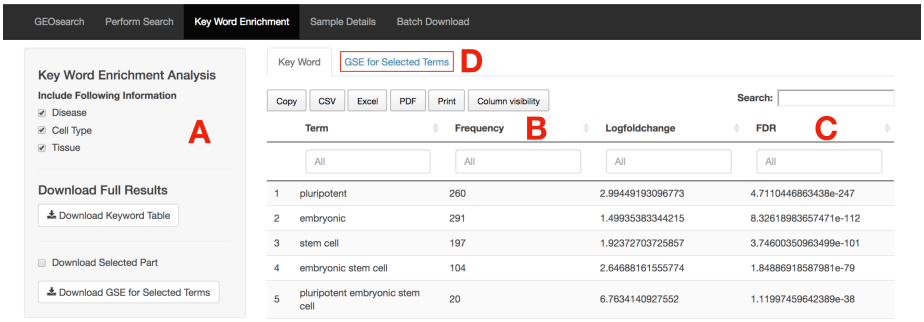

Figure 2: Screenshot demonstrating: Key Word Enrichment

compared to all samples in GEO (Figure 2C). In this example the top enriched key words include "embryonic stem cell" and "pluripotent embryonic stem cell", which are well-known contexts related to Oct4. Thus the key word enrichment analysis can provide a quick overview of related biological contexts. One can select one or multiple biological contexts in the table and GEOsearch will return all GSE records whose titles or descriptions contain the selected contexts (Figure 2D). Again, one can select GSE records and explore the sample details in "Sample Details".

6 Sample Details

For GSE records selected in "Perform Search" or "Key Word Enrichment", GEOsearch will return the details of all samples related to these GSE records. In this way one can conveniently compare samples from different GSE records, which could be quite tedious in GEO. Users can choose selected GSE from "Perform Search" or "Key Word Enrichment" or enter new GSE ID in the text box (Figure 3A). The GSE ID can be entered in the text box (Figure 3B) if users have chosen "Enter new GSE". Details of all samples in the given GSE will be shown in Figure 3C.

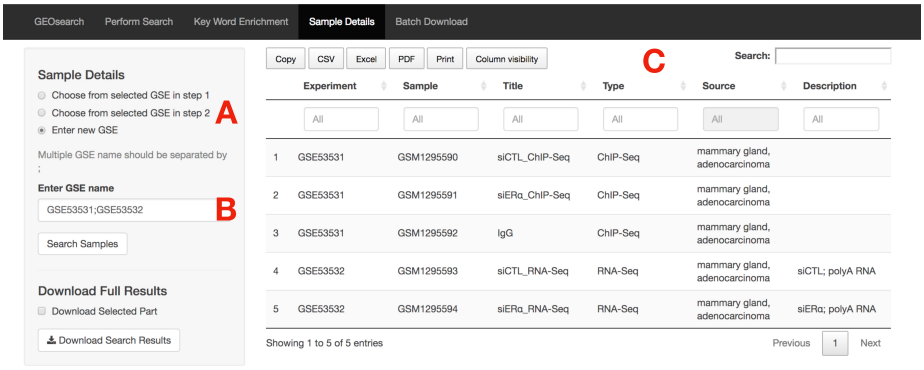

Figure 3: Screenshot demonstrating: Sample Details

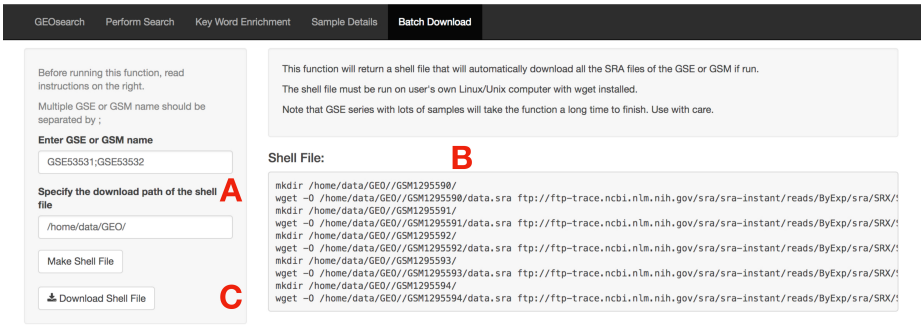

Figure 4: Screenshot demonstrating: Batch Download

7 Batch Download

GEOsearch provides a function to generate a shell file for downloading all the raw files corresponding to the GSE from GEO. Enter GSE or GSM name for which the raw data will be downloaded in Figure 4A. Download path can also be designated. The shell file generated by GEOsearch to download the SRA raw data for the given GSE or GSM will be shown in Figure 4B. The generated shell file can be downloaded in Figure 4C. Users can run the shell file on their own computer to download all the files.

8 Contact

To report bugs and provide suggestions for the GEOsearch GUI as well as the GEOsearch package, please contact the maintainer Zhicheng Ji (zji4@jhu.edu).