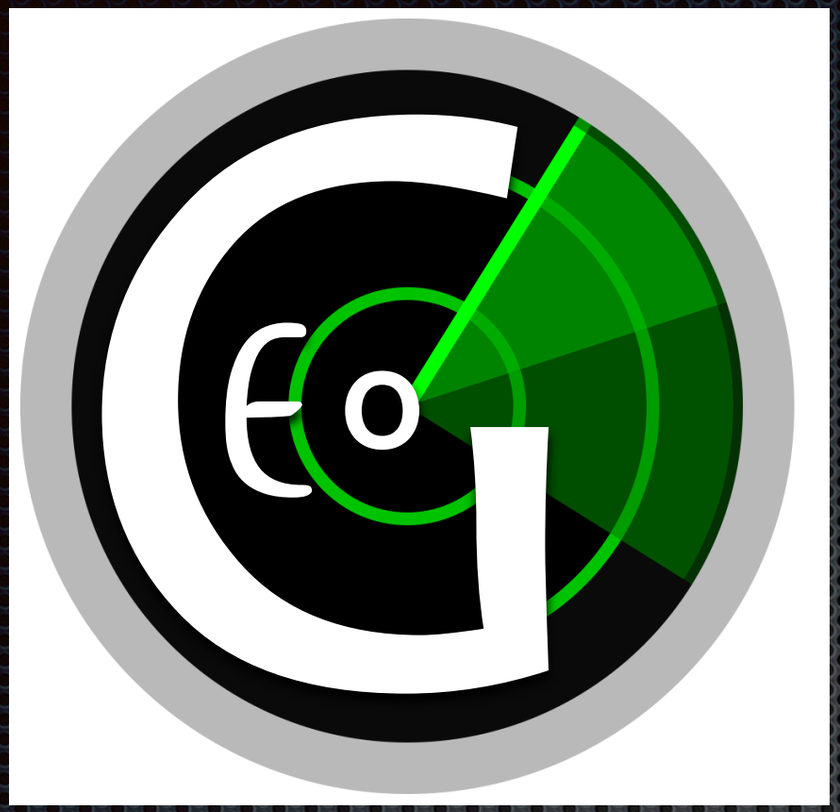


ScanGEO

parallel mining of high-throughput gene expression data

Katja Koeppen, Bruce A. Stanton, Thomas H. Hampton
The Geisel School of Medicine at Dartmouth

User Guide for ScanGEO Shiny App

ScanGEO Highlights

- A simple, user friendly Shiny app that searches for differentially expressed genes across multiple NCBI gene expression omnibus (GEO) data sets
- Search can be limited to a particular keyword
- Uses a custom list of genes and/or a KEGG pathway to be tested for differential gene expression
- Outputs include summary tables of all selected GEO data sets, significant genes and studies, PDF files with plots of differentially expressed genes, and CSV files with expression values, greatest fold-change and p-values

ScanGEO App Flowchart

Select organism

Enter keyword

Specify genes and/or KEGG pathway

Summary tables with # of significant genes and studies

PDF files with dot plots of significant genes

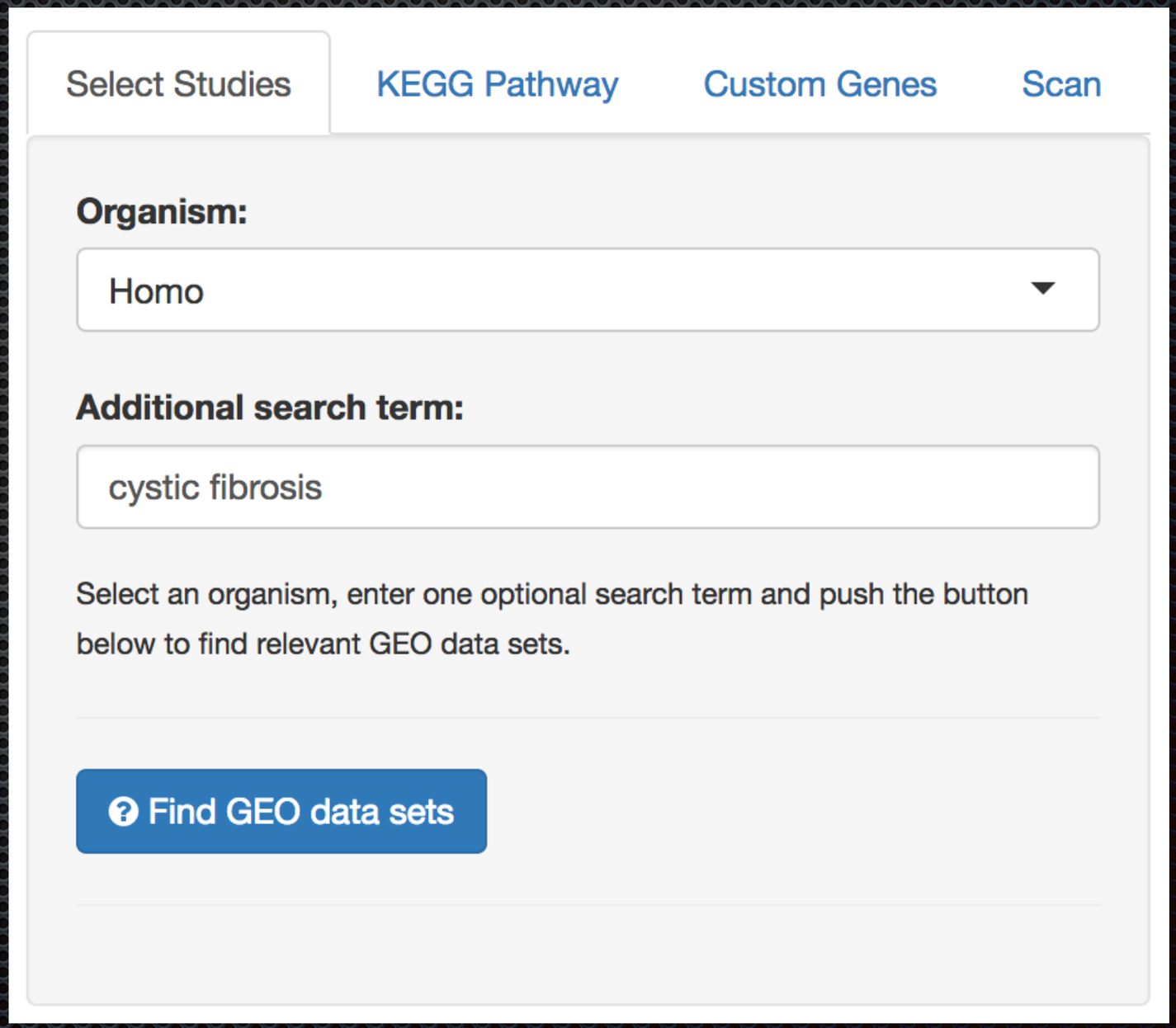
CSV files with expression values, p-values & max FC

Step 1 - Select Studies

1. Select organism →

2. Enter keyword (optional, limited to one search term) →

3. Find matching data sets →



Step 2 - Obtain Table with Relevant Studies

Select Studies | KEGG Pathway | Custom Genes | Scan

Organism:

Homo

Additional search term:

cystic fibrosis

Select an organism, enter one optional search term and push the button below to find relevant GEO data sets.

Find GEO data sets

Number of studies that match a given organism and search term. If 0, try modifying the search term.

```
8 studies found
 searching for " cystic fibrosis "
 in Homo
```

Table | Significant Genes | Significant Studies | Documentation

Show 25 entries

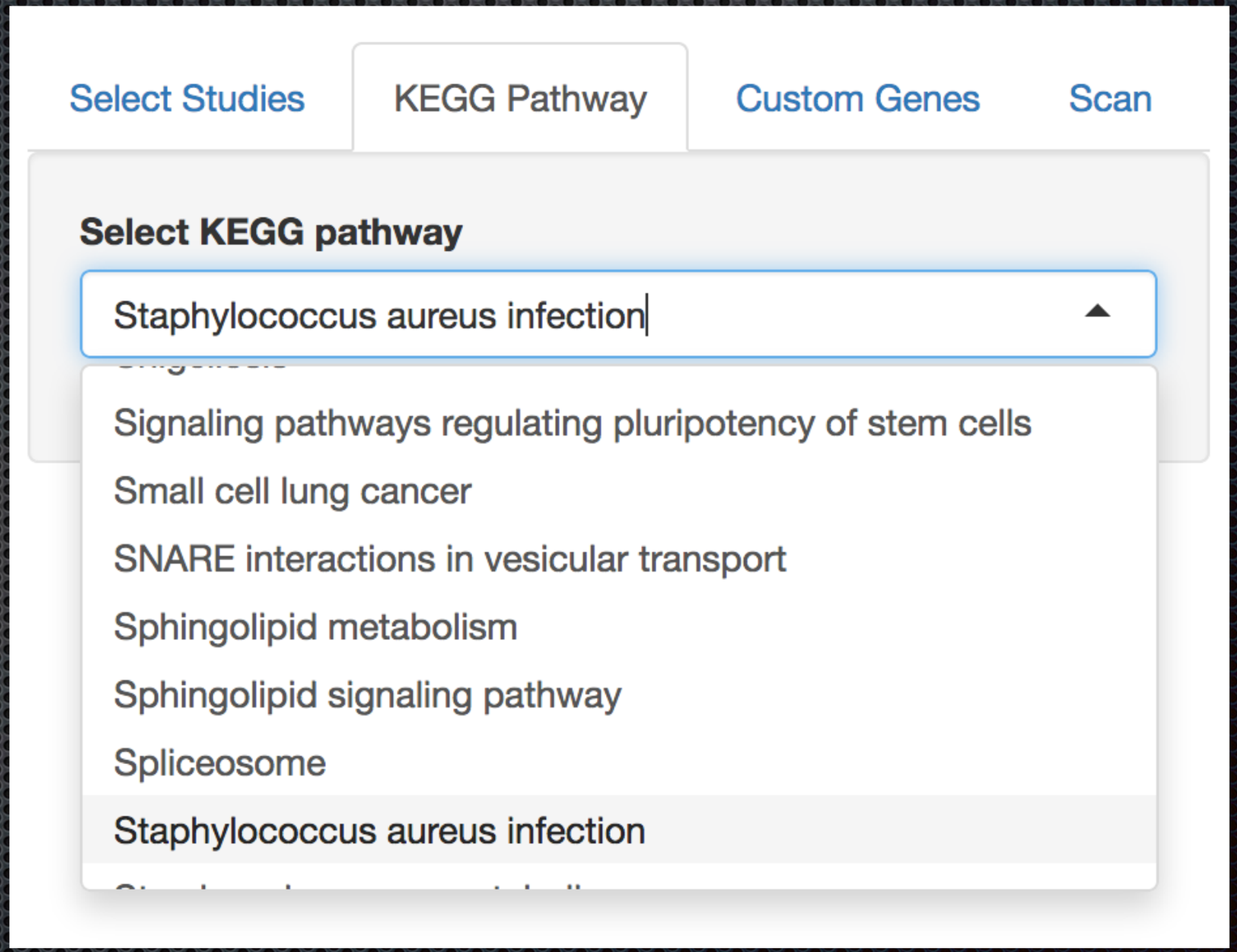
title	gds	pubmed_id	type	platform_organism	update_date
Cystic fibrosis pathology and 4-phenylbutyrate (HG-U133A)	GDS493	14583596	Expression profiling by array	Homo sapiens	2003-12-04
Cystic fibrosis pathology and 4-phenylbutyrate (HG-U133B)	GDS494	14583596	Expression profiling by array	Homo sapiens	2003-12-04
Lung pneumocyte response to Pseudomonas aeruginosa type III secretion system mutants	GDS1022	16207250	Expression profiling by array	Homo sapiens	2005-02-08
Cystic fibrosis patients with mild and severe lung disease: nasal respiratory epithelium (HG-U133A)	GDS2142	16614352	Expression profiling by array	Homo sapiens	2007-03-14
Cystic fibrosis patients with mild and severe lung disease: nasal respiratory epithelium (HG-U133B)	GDS2143	16614352	Expression profiling by array	Homo sapiens	2007-03-14
Cystic fibrosis bronchial epithelial cells exposure to Pseudomonas aeruginosa PA01 biofilms	GDS4252	22821996	Expression profiling by array	Homo sapiens	2013-04-23
Cystic fibrosis transmembrane conductance regulator expression in airway epithelial cells	GDS4255	22853952	Expression profiling by array	Homo sapiens	2013-04-23
Cystic fibrosis: rectal epithelia	GDS4844	24105369	Expression profiling by array	Homo sapiens	2014-05-22
title	gds	pubmed_id	type	platform_organism	update_date

Showing 1 to 8 of 8 entries

Step 3 - Select KEGG Pathway

Selection of an organism-specific KEGG pathway is **optional**.

All genes on a given KEGG pathway can be included instead of or in addition to custom genes.

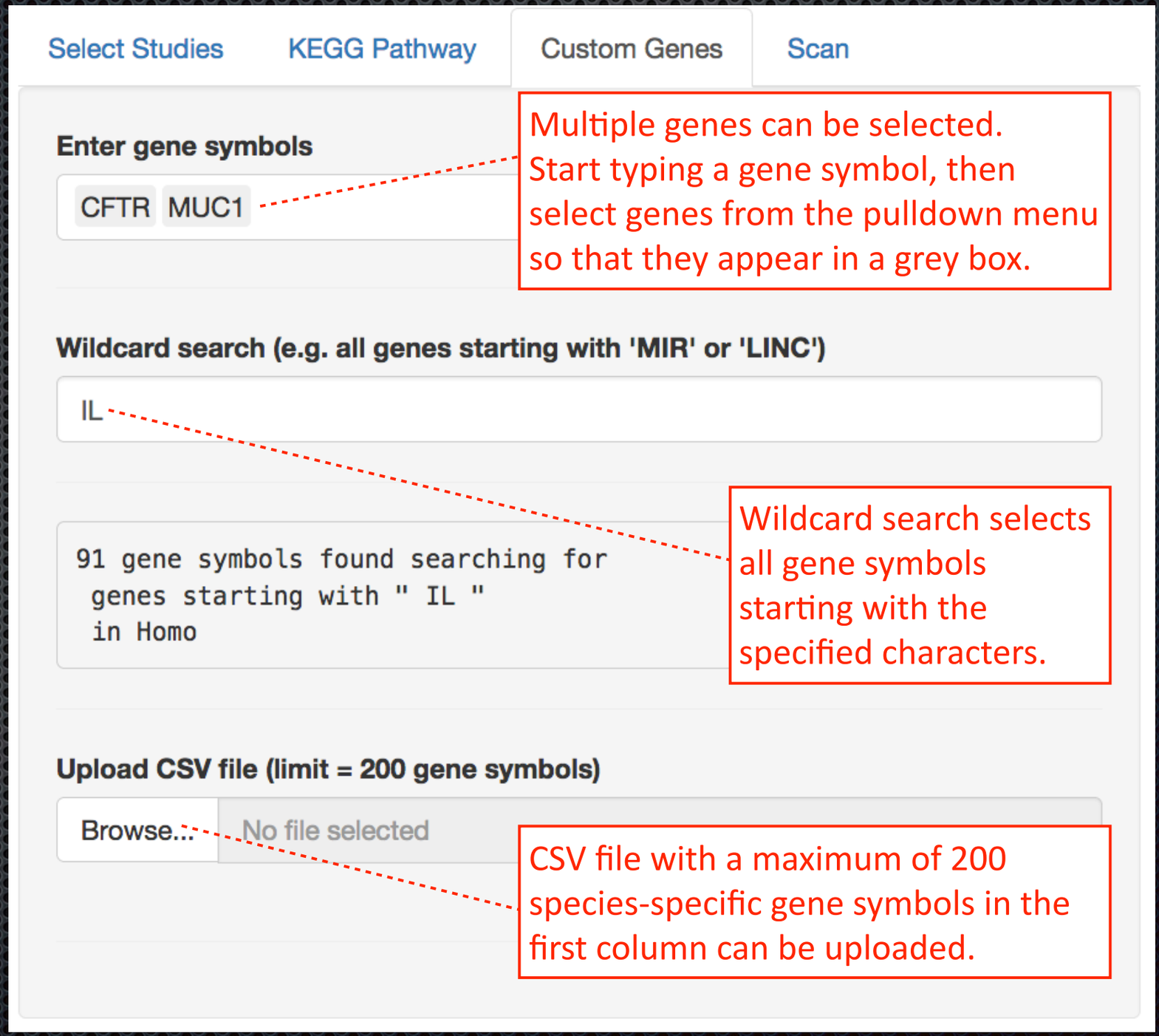


Step 4 - Select Custom Genes

Select genes
(pull-down menu
with choices appears
as you start typing)

Wildcard search

File upload

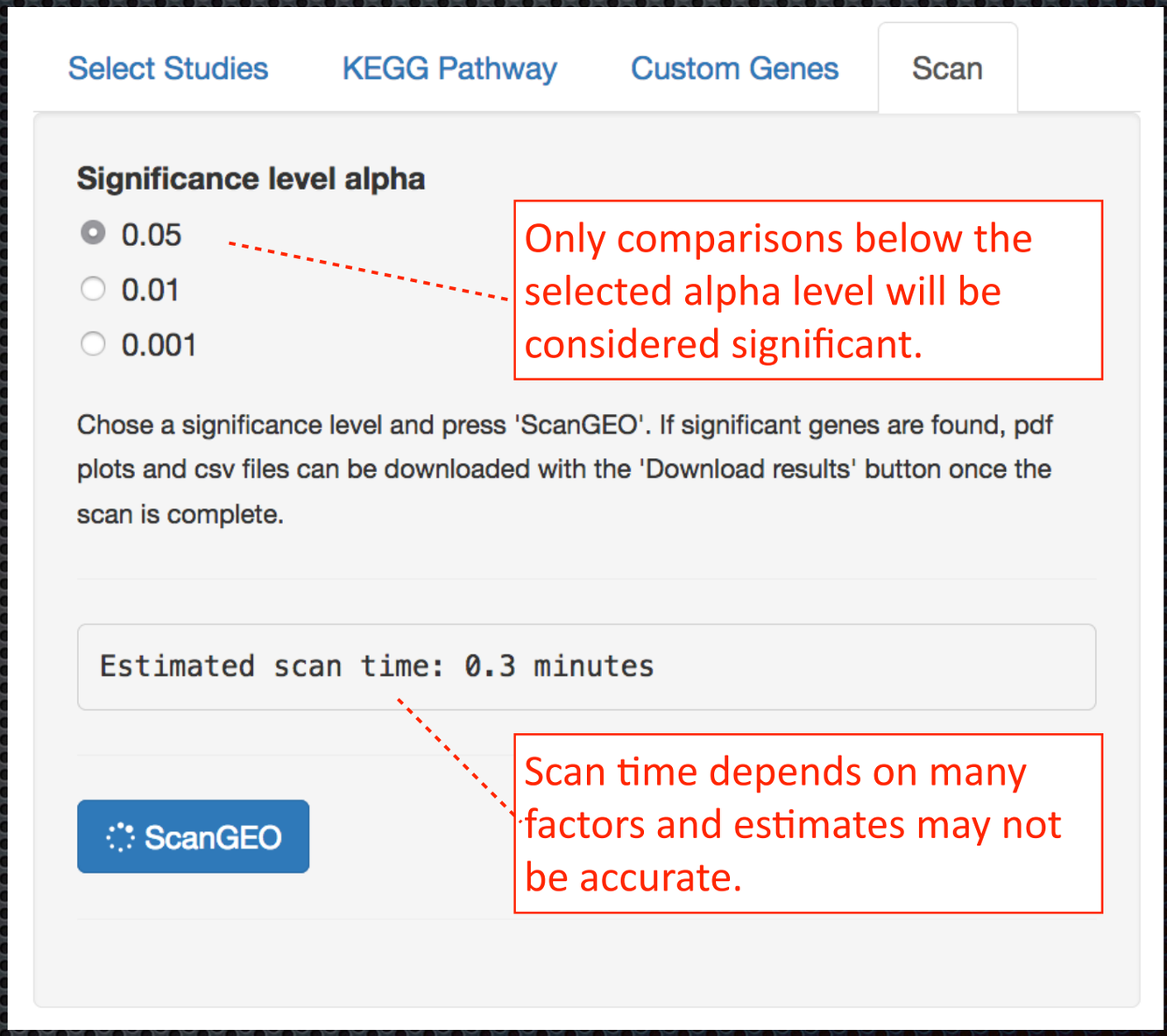


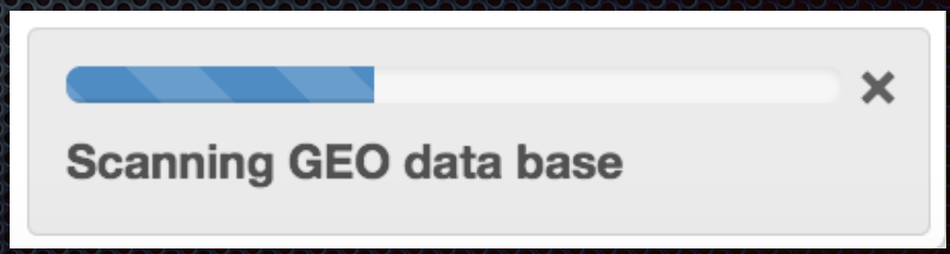
Step 5 - Scan Selected Genes

1. Select alpha

2. Scan GEO studies for differential expression of selected genes

3. A progress bar appears in the lower right while scan is running





Step 6a - View Results

Summary table with number of studies in which a gene of interest was differentially expressed

Table | Significant Genes | Significant Studies | Documentation

Show 10 entries Search:

Gene	Significant.Studies
IL1RN	5
IL4R	5
IL6ST	5
CFD	3
DSG1	3
CFH	3
HLA-DMA	3
KRT10	3
PTAFR	3
CFB	3

Gene Significant.Studies

Showing 1 to 10 of 148 entries

Previous 1 2 3 4 5 … 15 Next

Step 6b - View Results

Summary table with number of genes that were differentially expressed in each study of interest

Table | Significant Genes | Significant Studies | Documentation

Show 25 entries Search:

Title	GDS	Significant_Genes
Cystic fibrosis bronchial epithelial cells exposure to Pseudomonas aeruginosa PA01 biofilms	GDS4252	45
Cystic fibrosis patients with mild and severe lung disease: nasal respiratory epithelium (HG-U133A)	GDS2142	30
Lung pneumocyte response to Pseudomonas aeruginosa type III secretion system mutants	GDS1022	20
Cystic fibrosis transmembrane conductance regulator expression in airway epithelial cells	GDS4255	20
Cystic fibrosis: rectal epithelia	GDS4844	17
Cystic fibrosis pathology and 4-phenylbutyrate (HG-U133A)	GDS493	12
Cystic fibrosis patients with mild and severe lung disease: nasal respiratory epithelium (HG-U133B)	GDS2143	5
Cystic fibrosis pathology and 4-phenylbutyrate (HG-U133B)	GDS494	1

Title | GDS | Significant_Genes

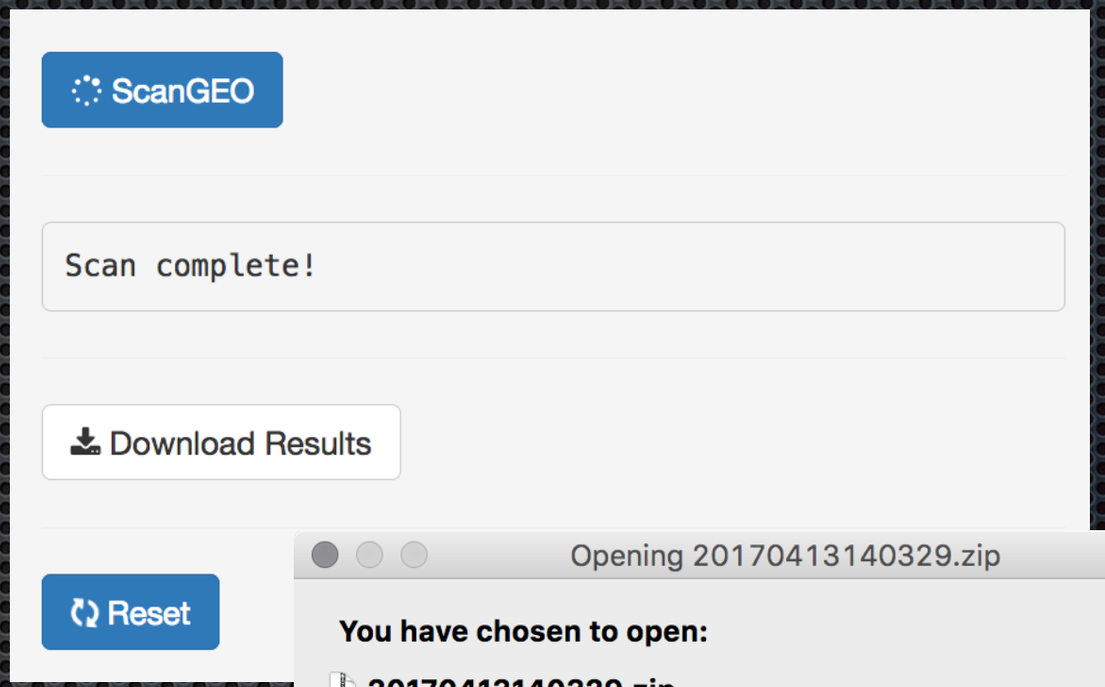
Showing 1 to 8 of 8 entries

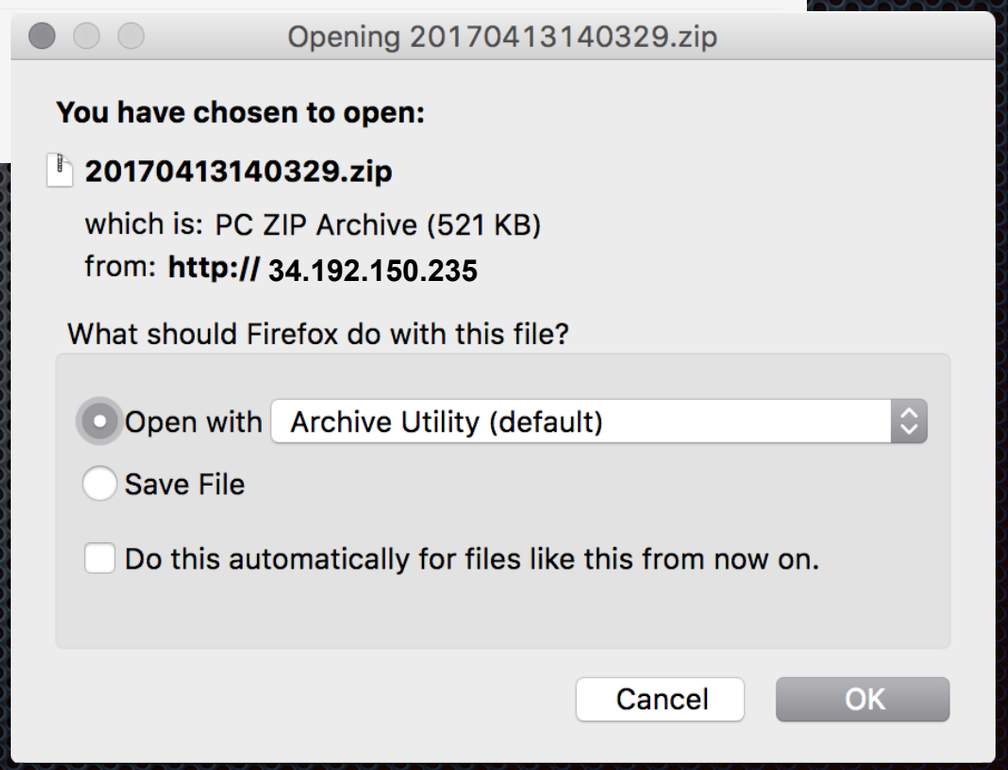
Previous 1 Next

Step 7 - Download Results

Download all scan results as a ZIP file

Reset analysis settings to begin a new search





Output Files

01_README.pdf

02_Summary_Table_*organism*.csv: GDS summary table for selected organism and keyword.

03_Results_sig_genes.csv: Summary table with number of significant studies per gene.

04_Results_sig_studies.csv: Summary table with number of significant genes per study.

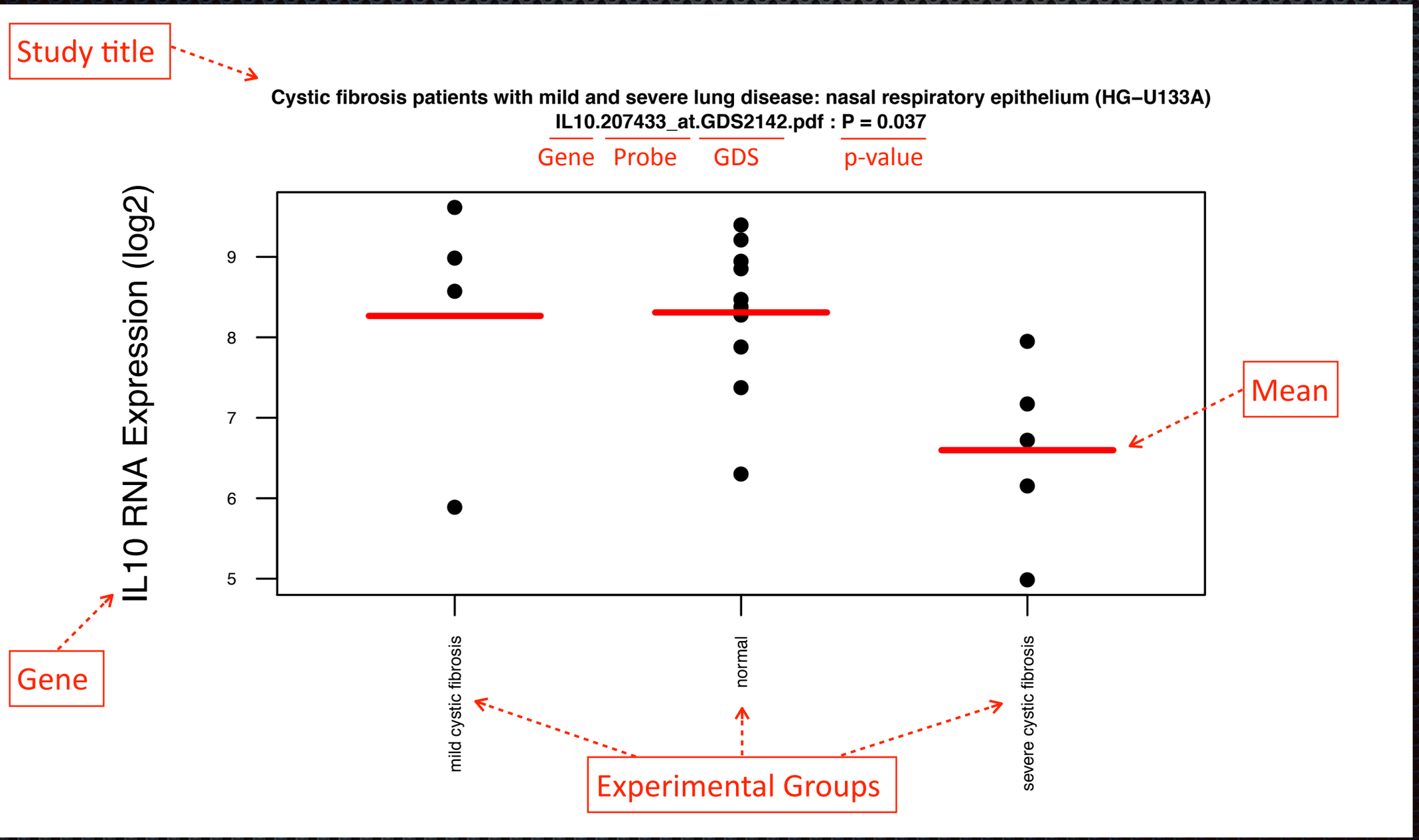
05_pValues_summary.csv: Unadjusted p-values for all mapped genes in studies with at least two samples per group. For genes with multiple probes the lowest p-value is shown. NA = gene was not mapped to a GDS or GDS had fewer than 2 Ns per group.

06_max_log2FC_summary.csv: Summary of largest absolute log2 fold change between the experimental groups for a mapped gene in a study with at least two samples per group. NA = gene was not mapped to a GDS or GDS had fewer than 2 Ns per group.

PDF files: Dot plots for probes that reached significance in ANOVA based on user-specified alpha level (default = 0.05). Red lines = mean.

CSV files: Expression values that were used to generate dot plots for significant probes.

Example Output Plot



Status Messages

Select Studies | KEGG Pathway | Custom Genes | Scan

Significance level alpha

- (•) 0.05
- ( ) 0.01
- ( ) 0.001

Chose a significance level and press 'ScanGEO'. If significant genes are found, pdf plots and csv files can be downloaded with the 'Download results' button once the scan is complete.

ScanGEO

Make sure to select at least one gene or a KEGG pathway.

No genes selected

ScanGEO

At least one of the selected genes was differentially expressed and there are downloadable results.

Scan complete!

Download results

ScanGEO

The selected genes were not differentially expressed in any of the selected studies at the chosen alpha level.

No significant genes!

Acknowledgements

- John Wallace (Dartmouth Research Computing)
- Dean Attali
- Funding sources:

National Heart, Lung, and Blood Institute

National Institute of Environmental Health Sciences

CYSTIC FIBROSIS FOUNDATION
ADDING TOMORROWS

Contact: Katja.Koeppen@Dartmouth.edu