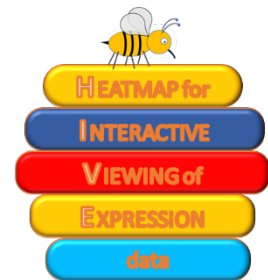


HEATMAP FOR INTERACTIVE VIEWING OF GENE EXPRESSION: HIVE BROWSER

A USER GUIDE

July, 2023



CONTENTS

Introduction	1
Accessing the HIVE browser	2
Loading gene expression data	2
Display options & Filtering by fold change, p value, cell type/condition	5
Selection of genes of interest	7
Submitting gene lists for secondary analysis	7
g:Profiler: a web server for functional enrichment analysis.....	8
STRING: functional protein association networks	8
DAVID Bioinformatics Resources	8
Creating a heatmap image for download.....	9
Data formats	9
Example datasets	12
Bibliography	13

To demonstrate how to upload expression data to the HIVE browser, explore it to identify genes of interest, and submit gene lists to secondary tools for further analysis, we use two exemplar single cell RNA-Seq datasets. The first has changes in expression in the cell types of the mouse retina in the Akita model of diabetes. The second is from single human islet cells from Type 2 diabetics and non-diabetics generated by Xin *et al* (Xin, Kim et al. 2016), with expression differences in diabetes according to gender. Further details of the example datasets are available in the '[Example datasets](#)' section.

[Back to Contents](#)

ACCESSING THE HIVE BROWSER

Quick start

Click on this [link](#) to open the browser tab with preloaded data. You can start to explore the example data and then input your own data.

Download HIVE browser file from GitHub (https://qub-simpson-lab.github.io/HIVE_browser/ or https://github.com/QUB-Simpson-lab/HIVE_browser).

Click on the file to run locally

[Back to Contents](#)

LOADING GENE EXPRESSION DATA

Download example datasets. There are 3 formats for you to explore. Example A is the most common starting point and comprises individual files for each comparison. It is also possible to import differential gene expression for multiple conditions from a single file (example B). Any gene expression data organized in columns separated by comma, tab or semi-colon can be imported into the HIVE browser. Once you have your data loaded you can create a custom HIVE format file (example C) suitable for data sharing by clicking on the 'Download your data as a single data file' link. For more information about file formats and two-level inputs see Combined multi-sample and multi-level input files section [below](#).

A. Individual differential expression files (eg output from Seurat)

Diabetic retina: Amacrine.txt, Bipolar.txt, Cone.txt, endothelial.txt, Microglia.txt, Muller.txt, Pericyte.txt, RGC.txt, Rod.txt

B. Single file

Diabetic islet cells: **File_name**

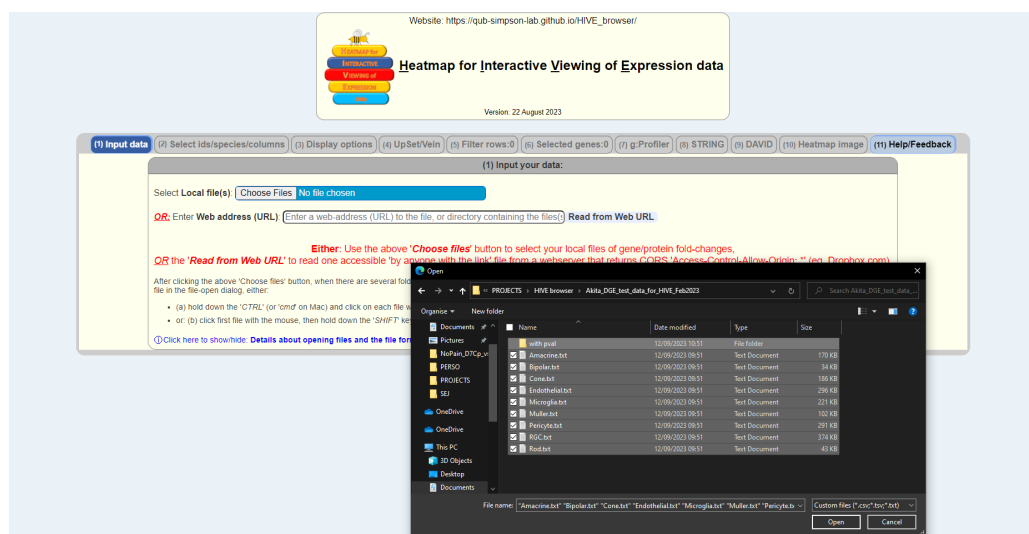
C. HIVE custom format

Diabetic islet cells: **File_name**

Choose files

Navigate to box (1) in the browser, entitled 'Input your data'. In this case we are using individual differential expression files (example dataset A)

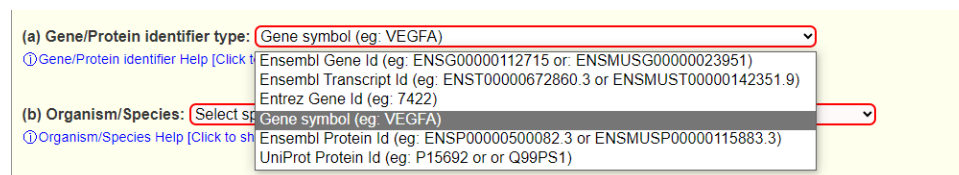
Click on 'Choose files' and select the individual files (hold down the 'CTRL' key ('cmd' on Mac) and click on each file with the mouse, or click on first file, then hold down 'SHIFT' key and click on the last file in list).



Once the files have been selected click on 'OPEN'. The files may take few seconds to load (depending on size)

Navigate to box (2) in the browser, entitled 'Select columns'.

a. Select the gene or protein identifier used in your data from the drop-down menu (it will normally be correctly pre-populated based on your input data)



b. Select the organism from the drop-down menu

(b) Organism/Species: Select species

[Organism/Species Help](#)

(c) Set column names

[Column Names Help](#)

Input	Species
(1) Am	Homo sapiens (Human) [9606 : Homo_sapiens : hsapiens]
(2) Bip	Mus musculus (Mouse) [10090 : Mus_musculus : mmusculus]
(3) Cone	Arabidopsis thaliana (Thale cress) [3702 : plants/Arabidopsis_thaliana : athaliana]
(4) End	Bos taurus (Cow) [9913 : Bos_taurus : btaurus]
(5) Micro	Caenorhabditis elegans (C. elegans Nematode worm) [6239 : Caenorhabditis_elegans : celegans]
(6) Mull	Danio rerio (Zebrafish) [7955 : Danio_rerio : drerio]
(7) Peric	Drosophila melanogaster (Fruit fly) [7227 : Drosophila_melanogaster : dmelanogaster]
(8) RGC	Gallus gallus (Chicken) [9031 : Gallus_gallus : ggallus]
(9) Rod	Oryza sativa Japonica group (Japanese rice) [39947 : plants/Oryza_sativa : osativa]
	Rattus norvegicus (Rat) [10116 : Rattus_norvegicus : rnorvegicus]
	Saccharomyces cerevisiae S288C (Baker's Yeast) [4932 : scerevisiae]
	Xenopus tropicalis (Western/Tropical clawed frog) [8364 : Xenopus_tropicalis : xtropicalis]
	Zea mays (Maize/corn) [4577 : plants/Zea_mays : zmais]
	OTHER

c. Set names **for columns in the heatmap**. These will be pre-populated based on your input files, but can be amended if required (eg of Main-name: CellType or TissueType; and of sub-name: TimePoint, Condition or Location). In this example we have just the main column for each file.

(c) Set column names (eg: CellType or TissueType; and eg: TimePoint, Condition or Location) for each input file:

[Column Names Help](#) [Click to show/hide](#)

Input filename	Column main-name (eg: Cell-type/cluster)	Column sub-name (optional) (eg: Time-point or Condition)
(1) Amacrine.txt	Amacrine	
(2) Bipolar.txt	Bipolar	
(3) Cone.txt	Cone	
(4) Endothelial.txt	Endothelial	
(5) Microglia.txt	Microglia	
(6) Muller.txt	Muller	
(7) Pericyte.txt	Pericyte	
(8) RGC.txt	RGC	
(9) Rod.txt	Rod	

d. Select the column data type for each column in your input files

(d) Select the column data type (eg: Fold-change, P-value, etc) for columns from input files:

[Data types Help](#) [Click to show/hide](#)

Column number in files	Column heading from file(s)	Select Type for this column
1 in files: All	[No Heading for this column]	Gene/Protein ID
2 in files: All	p_val	Don't use this column
3 in files: All	avg_log2FC	Fold-change
4 in files: All	pct.1	pct.1: percentage of cells where feature is detected in the first group
5 in files: All	pct.2	pct.2: percentage of cells where feature is detected in the second group
6 in files: All	p_val_adj	P-value/FDR/q-value

[Show](#)

Don't use this column
Gene/Protein ID
Fold-change
P-value/FDR/q-value
pct.1: percentage of cells where feature is detected in the first group
pct.2: percentage of cells where feature is detected in the second group

View the heatmap by clicking on the 'Show/Update heatmap' button

[Show/Update heatmap](#)

The heatmap is displayed below.

[Optionally, Download your data as single data file.](#)

You can use the above tabs:

- (3) **'Display options'** to show p-values/FDR on the heatmap, and set the row ordering, etc
- (4) **'UpSet/Vein'** to optionally view an overview UpSet/Vein-diagram-type table,
- (5) **'Filter rows'** to filter the heatmap rows on fold-change and p-value/FDR,
- (6) **'Selected genes'** to select genes/proteins of interest,
- (7) **'g:Profiler'** to send your selected genes to g:Profiler website for functional enrichment analysis,
- (8) **'STRING'** to send your selected genes to STRING website to display Interaction Networks,
- (9) **'DAVID'** to send your selected genes to DAVID website for functional interpretation,
- (10) **'Heatmap image'** to download the heatmap as an png or jpg image,
- (11) **'Help/Feedback'** for help links and sending feedback.

Fold-change Heatmap for: Gene, Amacrine, Bipolar, Cone, Endothelial, Microglia, Muller, Pericyte, RGC, Rod

Heatmap log2fc colour scale:

Clicking the Row-Id (Gene name) in the left or right columns will open a new browser tab showing the NCBI Entrez entry for that gene name in the Species selected above (eg. mouse: Mus musculus). Columns with no fold-changes have their column heading in light-grey (eg. "Time-1")

All None	Gene	Amacrine	Bipolar	Cone	Endothelial	Microglia	Muller	Pericyte	RGC	Rod	Gene
<input type="checkbox"/>	Rporip1	-0.71	-0.44	-0.46	-0.3			1.27	-0.22	-0.22	Rporip1
<input type="checkbox"/>	Ccdc126	-0.33		-0.32	0.3	0.27					Ccdc126
<input type="checkbox"/>	Klf21b	-0.48		-0.39		1.09	-0.27		-0.22		Klf21b
<input type="checkbox"/>	Cados	0.29	-0.15			0.41	0.52		-0.19		Cados
<input type="checkbox"/>	Adcy8	0.6	-0.35						0.29		Adcy8

[Back to Contents](#)

DISPLAY OPTIONS & FILTERING BY FOLD CHANGE, P VALUE, CELL TYPE/CONDITION

Navigate to box (3) at the top of the webpage entitled 'Display options'. You can toggle display of p-values and pct1 & 2 in the heatmap by selecting the boxes as required.

(1) Input data (2) Select ids/species/columns (3) **Display options** (4) UpSet/Vein (5) Filter rows:217 (6) Selected genes:0 (7) g:Profiler (8) STRING (9) DAVID (10) Heatmap image (11) Help/Feedback

(3) Display options:

☒ Show p-values/FDR on the heatmap below.

☐ Show pct.1 & pct.2 values (percentage of cells where the feature is detected in the first & second groups.)

Sort order of heatmap rows: Same as input file(s)

[Display Options Help](#) [Click to show/hide](#)

Fold-change Heatmap for: Gene, Amacrine, Bipolar, Cone, Endothelial, Microglia, Muller, Pericyte, RGC, Rod

Heatmap log2fc colour scale:

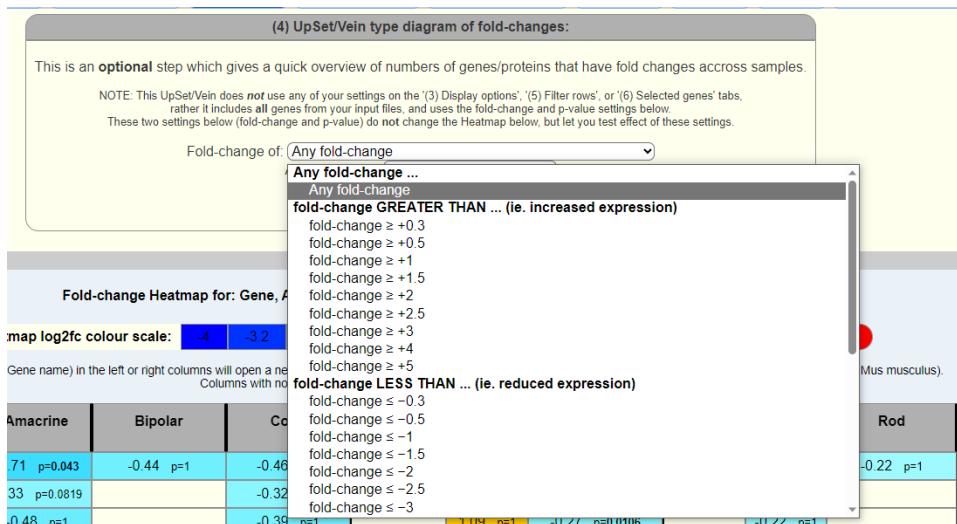
Clicking the Row-Id (Gene name) in the left or right columns will open a new browser tab showing the NCBI Entrez entry for that gene name in the Species selected above (eg. mouse: Mus musculus). Columns with no fold-changes have their column heading in light-grey (eg. "Time-1")

All None	Gene	Amacrine	Bipolar	Cone	Endothelial	Microglia	Muller	Pericyte	RGC	Rod	Gene
<input type="checkbox"/>	Rporip1	-0.71 p=0.043	-0.44 p=1	-0.46 p=1	-0.3 p=1			1.27 p=1	-0.22 p=1	-0.22 p=1	Rporip1
<input type="checkbox"/>	Ccdc126	-0.33 p=0.0619		-0.32 p=1	0.3 p=1	0.27 p=1					Ccdc126
<input type="checkbox"/>	Klf21b	-0.48 p=1		-0.39 p=1		1.09 p=1	-0.27 p=0.0106		-0.22 p=1		Klf21b
<input type="checkbox"/>	Cados	0.29 p=1	-0.15 p=0.34			0.41 p=1	0.52 p=1		-0.19 p=1		Cados
<input type="checkbox"/>	Adcy8	0.6 p=1	-0.35 p=0.321						0.29 p=1		Adcy8
<input type="checkbox"/>	Sag	-0.81 p=1		-0.21 p=1				2.16 p=1		-0.18 p=0.000149	Sag
<input type="checkbox"/>	Malat1	0.2 p=1	0.2 p=2.63e-11	0.22 p=1				0.27 p=1	0.21 p=1	0.15 p=1	Malat1
<input type="checkbox"/>	Gnb1	-0.64 p=1			-0.31 p=1	0.67 p=1	-0.35 p=0.487	2 p=1	-0.33 p=1		Gnb1
<input type="checkbox"/>	Pfkfb3	-0.5 p=1	-0.2 p=0.0128			-0.3 p=1					Pfkfb3
<input type="checkbox"/>	Slmn1	0.58 p=1	-0.51 p=1		-0.18 p=1		-0.48 p=0.189	-0.46 p=1	0.98 p=1		Slmn1

Navigate to box (4)

Default filter options are; one or more columns having fold-change values, $p < 0.5$ in at least one column in the row, and any fold change. The users can change these values as shown below.

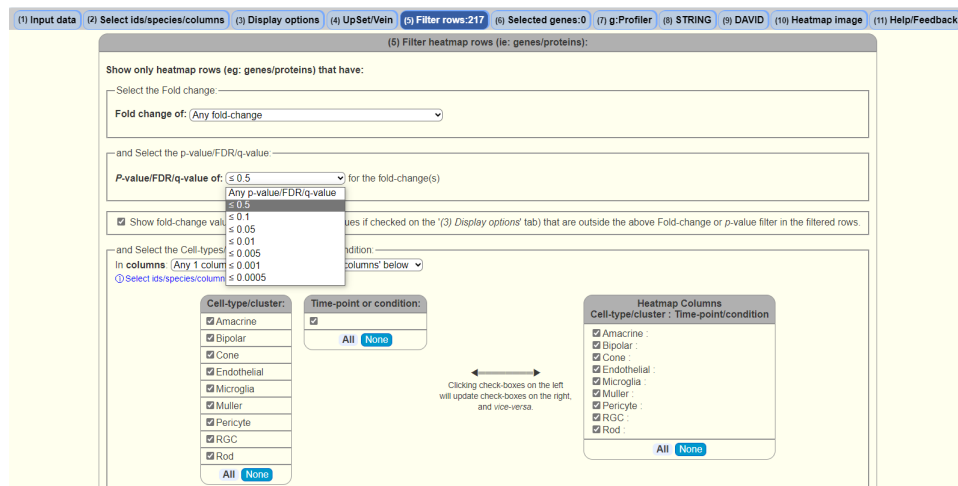
a. Select how many rows must have the chosen fold change



b. Select the desired fold change range and cell type of interest.

If your data has two levels, for example several time-points, age groups or sexes at which each of the cell types were assessed, it is possible to select both (more than one group can be selected by holding down the CTRL button).

c. Select the P-value (this refers to the adjusted P-value or raw P-value as selected [above](#)). All rows that contain at least one value below this threshold will be displayed. The fold changes with p-values below this threshold can be hidden by checking the box.



[Back to Contents](#)

SELECTION OF GENES OF INTEREST

Navigate to box (6) 'Selected genes'

(1) Input data (2) Select ids/species/columns (3) Display options (4) UpSet/View (5) Filter rows:217 (6) Selected genes:0 (7) g:Profiler (8) STRING (9) DAVID (10) Heatmap image (11) Help/Feedback

Selected genes: 0

Select gene(s) using the check-boxes in left column of the heatmap table below.

You can submit these selected genes/proteins to (7) g:Profiler, (8) STRING or (9) DAVID on the above tabs, or use the 'Copy ...' button below to copy the above selected gene/protein ids to your Windows/Mac/Linux clipboard, to paste into a document or another website.

Copy selected genes to clipboard:

1. Select gene(s) using the check-boxes in left column of the heatmap table below.
2. Separate the genes in the list with:

Copy selected genes to Windows/Mac/Linux clipboard

All
None

Gene

<input checked="" type="checkbox"/>	Rpgrip1
<input checked="" type="checkbox"/>	Ccdc126
<input checked="" type="checkbox"/>	Kif21b
<input checked="" type="checkbox"/>	Cadps
<input checked="" type="checkbox"/>	Adcy8
<input checked="" type="checkbox"/>	Sag

Your genes of interest can be chosen simply by using the filtering criteria above and then clicking on 'All.' It is also possible to select a subset of genes by manually clicking on genes of interest. This is particularly helpful for preliminary investigation of genes of interest identified while browsing the heatmap.

	Amacrine	Bipolar	Cone	Endothelial	Microglia	Muller	Pericyte	RGC	Rod	Gene
<input type="checkbox"/>	-0.71 p=0.043	-0.44 p=1	-0.46 p=1	-0.3 p=1			1.27 p=1	-0.22 p=1	-0.22 p=1	Rpgrip1
<input type="checkbox"/>	-0.33 p=0.0819		-0.32 p=1	0.3 p=1	0.27 p=1					Ccdc126
<input type="checkbox"/>	-0.48 p=1		-0.39 p=1		1.09 p=1	-0.27 p=0.0106		-0.22 p=1		Kif21b
<input type="checkbox"/>	0.29 p=1	-0.15 p=0.34			0.41 p=1	0.52 p=1		-0.19 p=1		Cadps
<input type="checkbox"/>	0.6 p=1	-0.35 p=0.321						0.29 p=1		Adcy8
<input type="checkbox"/>	-0.81 p=1		-0.21 p=1				2.16 p=1		-0.18 p=0.000149	Sag
<input type="checkbox"/>	0.2 p=1	0.2 p=2.63e-11	0.22 p=1				0.27 p=1	0.21 p=1	0.15 p=1	Malat1
<input checked="" type="checkbox"/>	-0.64 p=1			-0.31 p=1	0.87 p=1	-0.35 p=0.487	2 p=1	-0.33 p=1		Gnb1
<input type="checkbox"/>	-0.5 p=1	-0.2 p=0.0128			-0.3 p=1					Prkcsb
<input checked="" type="checkbox"/>	0.58 p=1	-0.51 p=1		-0.18 p=1		-0.48 p=0.189	-0.46 p=1	0.98 p=1		Stmn1
<input type="checkbox"/>	-0.54 p=1	-0.38 p=9.38e-8			0.82 p=1		-1.33 p=1			Cd63
<input type="checkbox"/>	-0.15 p=1		0.82 p=1	-0.5 p=1	-0.27 p=1		0.7 p=1		0.68 p=2.95e-15	Hmgb2
<input type="checkbox"/>	-0.34 p=1	-0.24 p=0.0122	-0.56 p=1	-1.25 p=1			-0.36 p=1	-0.38 p=1		Nr4a1
<input type="checkbox"/>	0.17 p=1	-0.16 p=0.359	-0.41 p=1		0.26 p=1			-0.21 p=1		Slc35c2
<input type="checkbox"/>	0.32 p=1					0.34 p=0.000311		-0.51 p=1		Tpm3

As you type a gene name into the box at the top of the 'Gene' column in the heatmap, the genes will be filtered to retain those with a matching name.

When you have chosen your genes of interest they can be copied to the clipboard with the chosen separator.

[Back to Contents](#)

SUBMITTING GENE LISTS FOR SECONDARY ANALYSIS

To submit selected genes directly to an external tool for network analysis and detection of enriched gene categories and pathways navigate to box 6, 7 or 8.

(7) g:Profiler (8) STRING (9) DAVID

G:PROFILER: A WEB SERVER FOR FUNCTIONAL ENRICHMENT ANALYSIS

For g:Profiler, simply click on the blue button to open the g:Profiler web server with your list of selected genes

G:Profiler:

1. This uses the 'organism/species' selected on the '(2) Select ids/species/columns' tab.
2. And your Selected gene(s) (selected using the check-boxes in left column of the heatmap table below).

Organism: 10090 : Mus_musculus : mmusculus

Number of genes selected: 3

[Submit](#)

STRING: FUNCTIONAL PROTEIN ASSOCIATION NETWORKS

For StringDB you can choose some basic parameters for the network and then view the image within the HIVE browser by clicking the 'Show StringDB image here' button. For more in depth analyses you can view your selected gene list on the String server by clicking the 'Show StringDB in new tab' button.

String DB:

1. This uses the 'organism/species' selected on the '(2) Select columns' tab.
2. And your Selected gene(s) (selected using the check-boxes in left column of the heatmap table below).
3. You can then set the 'Required-score', 'Additional network nodes', and 'Network flavor' options below.
4. For the 'Require score' option: a lower score includes more interactions.
5. After clicking this "Show Stringdb image" button, if the String image doesn't appear within a couple of seconds, it's probably because the gene(s) is not in in StringDB for the selected Organism/species.
6. Some genes (eg: 'Malat1') are in Human but not in Mouse STRING-db.
7. Click on a gene in the String image to show its abstract.
8. Click the 'STRING' logo to go to the String webpage in a new browser tab (eg. to view the interaction details).

Organism/Species: 10090 : Mus_musculus : mmusculus

Number of rows (genes) selected: 3

Required score: (500) Additional network nodes: (0) Network flavor: Confidence

[Show String-DB image here](#) [Show String-DB in new browser tab](#)

DAVID BIOINFORMATICS RESOURCES

For DAVID Bioinformatics Resources copy the selected genes to the clipboard using the tool provided, follow the DAVID link, and paste in your gene list.

DAVID Bioinformatics Resources (at LHR):

To submit your selected gene list to DAVID:

[Send selected gene list to the DAVID webserver](#)

However: for 'OFFICIAL_GENE_SYMBOL' identifiers (eg: VEGFA), the above button may fail as DAVID may be unable to match the gene symbol or one symbol can often apply to multiple species. (Eg. This does work for well established symbols, eg: EGR1, MALAT1, VEGFA, EGFR, but not for less established symbols. The above button doesn't have a way to specify the species to DAVID in the API). However you can manually submit your selected genes as explained below.

1. Select your gene(s)/proteins (using the check-boxes in left column of the heatmap table below)
2. Copy your list of selected genes/proteins using the 'Copy selected genes to Windows/Mac/Linux clipboard' button (below)
3. then go to: [DAVID Bioinformatics Resources - NIAID/NIH \(at LHR\)](#) (opens in new tab in your browser)
4. and paste your list into 'Step 1: Enter gene list'
5. then on then DAVID website 'Step 2: Select identifier type', etc.

Number of genes/proteins selected: 3

[Copy selected genes to Windows/Mac/Linux clipboard](#)

[Back to Contents](#)

CREATING A HEATMAP IMAGE FOR DOWNLOAD

Navigate to box 9

The screenshot shows a web interface with a tabbed menu at the top. The active tab is 'Heatmap image'. Below the menu is a dialog box titled 'Create heatmap image (to download):'. Inside the dialog, there are several settings:

- Rows to plot: (As selected by the 'Filter genes' above)
- Columns to plot (not implemented): (All columns)
- Margin width: (7) pixels
- Column width: (120) pixels
- Font height: (16) pixels
- Space around text: (5) pixels
- Background colour: (white)
- ☐ Draw fold-change numbers on the heatmap coloured boxes.

Below these settings is a note: 'NOTE: This currently ignores the "Show fold-change values ..." that are above this p-value filter" setting on the '4) Filter rows' tab, and draws all cells with fold-changes.'

At the bottom of the dialog are two buttons: 'Preview the Heatmap image below' and 'View or Download your heatmap image: Show'.

Choose options for customizing the output

Preview

Select format

[Back to Contents](#)

DATA FORMATS

The only requirement of your data is that it contains gene id and differential expression ("avg_log2FC") columns, which can be selected during the input process. It should be in comma, semicolon, or tab, separated format and can include optional p value ("p_val", "p_val_adj") and percent of cells ("pct.1" "pct.2") columns (see example input files). You can import the expression data from multiple individual files or from a single file containing multiple measurements.

Individual input files

The input data must be in tabular format as shown above. This is the typical column order, but specific columns can be selected during import. The first column is expected to contain gene/protein identifiers, such as gene symbols. It is possible to use any gene identifier, but for the 'additional gene information' links in the HIVE browser to be functional, it must be one of those types available from the drop-down list. Subsequent columns contain

information about the differential expression of each gene. The 'avg_log2FC' or 'avg_logFC' column is the fold-change value. The 'p_val_adj' is the P-value (if there is no 'p_val_adj' column, then 'p_val' [unadjusted p-value] column is used instead). The 'pct' is the percentage of cells in which the gene is detected.

A full dataset follows the format generated by analysis of scRNA-Seq data using the Seurat pipeline and contains the following columns: "" "p_val" "avg_log2FC" "pct.1" "pct.2" "p_val_adj". Only the "avg_log2FC" is essential, but omission of the other columns will reduce the functionality of the browser. Whatever the source of your data, the column headings must have these names enclosed within double quotation marks (fold change can be "avg_log2FC" or "avg_logFC").

The columns in the input file(s) should be separated by either: comma, semicolon, tab, or space. Typically an individual fold-change file is provided for each comparison eg differential expression in a specific cell type.

Example of data in the required format:

```
" " "p_val" "avg_log2FC" "pct.1" "pct.2" "p_val_adj"
"Rpgr1p1" 2.49e-06 -0.71 0.025 0.186 0.043
"Ccdc126" 4.74e-06 -0.33 0.006 0.137 0.081
"Epha6" 3.56e-05 0.78 0.46 0.288 0.615
...etc...
```

These columns typically have the following meaning (from: [Differential expression testing in Seurat](#)):

- *p_val*: p_val (unadjusted)
- *avg_logFC*: log fold-change of the average expression between the two groups. Positive values indicate that the feature is more highly expressed in the first group
- *pct.1*: The percentage of cells in which the feature is detected in the first group
- *pct.2*: The percentage of cells in which the feature is detected in the second group
- *p_val_adj* : Adjusted p-value, based on Bonferroni correction using all features in the dataset

Each filename (without any '.csv' or '.tsv' ending) will be used as the Heatmap column heading.

Combined multi-sample and two-level input files









It is possible to upload a single CSV file containing differential (eg. between test and control samples) fold changes for all the samples. The HIVE browser requires a column with an identifier (gene or protein) and a column for log₂fold change for each condition (additional columns can include p value, adjusted p value and percent of cells with expression). If the

column headings have the standard format (ie. P_val, avg_log2FC...) they will be imported automatically. If your file has different column names you can select manually which contains each value). You can also load individual gene expression files by selecting them manually – once these are loaded you can download the input data as a single multi-sample file using the 'Optionally: Download your data as single data file' link. This single file makes subsequent loading of data easier.

Combined file

If your data has multiple levels, for example differential expression for each cell type was recorded in several conditions, groups or timepoints, it is possible to reflect this in either the title of individual file or the column name in a multi-sample file. If a '-' is included in the filename of an individual file, for example between the CellType and Condition this is reflected in the heatmap column titles. Alternatively, the same samples within a single file could be specified using a colon in the column names as follows CellType:Condition.

Add image of file with Celltype:Condition...

 alpha-female.txt	
 alpha-male.txt	
 beta-female.txt	
 beta-male.txt	
 delta-female.txt	
 delta-male.txt	
 PP-female.txt	
 PP-male.txt	

	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
54453	5.65E-24	1.33	0.996	0.919	9.13E-20
5967	0.063374	1.21	0.353	0.496	1

Gene	alpha		beta		delta		PP	
	female	male	female	male	female	male	female	male
54453	1.33		1.22	0.19	0.23	1.13	0.82	-0.44

OR:
(2)

Gene, Amacrine:EARLY, Amacrine:MID, Amacrine:LATE, Bipolar:EARLY, Bipolar:MID, Bipolar:LATE, Cone:EARLY, Cone:MID, Cone:LATE
Aak1, 0, 0.6, 0, 0, -2.1, 0, 0, 0, 0.4
Aamp, 0, 0.3, 0, 0, -0.6, 0, 0, 0, 0.3, 0

```
Aars, 0, 0.4, 0, 0, -0.3, 0, 1.2, 0, 0  
...etc....
```

where:

- first column is 'Gene' name
- then several columns with heading format: CellType:TimePoint (or CellType:Condition/Location, etc)
- then gene/protein fold-change rows in format: GeneName, Fold-change:Optional p-value:Optional pct.1: Optional pct2,

Multilevel Explain format:'s names

Eg Islet

Can be used for other datatypes

For an explanation of the format required for combined multi-sample and multi-level input files see [below](#).

[Back to Contents](#)

EXAMPLE DATASETS

Dataset 1: Mouse retina. This was generated by our group from the Akita mouse model of diabetes and has been deposited in GEO, accession number [xxxx](#). The major retinal cell types were identified and differential expression between Akita and control mice performed using the Seurat package v4 (Hao, Hao et al. 2021). The output files used for analysis in the HIVE browser are available from as 'Example dataset 1' from https://github.com/QUB-Simpson-lab/HIVE_browser.

Dataset 2: (Xin, Kim et al. 2016) 4 Human islet cells. The original data was generated by Xin et al (Xin, Kim et al. 2016), and is available from GEO, accession number GSE81608. The samples were divided into male and female and differential expression for each cell type and sex determined using the Seurat package v4 (Hao, Hao et al. 2021). The output files used for analysis in the HIVE browser are available as a single multi-sample file (Example dataset 2) and in the custom HIVE format (Example dataset 3) from https://github.com/QUB-Simpson-lab/HIVE_browser.

Dataset 2: Human islet cells. The original data was generated by Xin et al (Xin, Kim et al. 2016), and is available from GEO, accession number GSE81608. The samples were divided into male and female and differential expression for each cell type and sex determined using

the Seurat package v4 (Hao, Hao et al. 2021). The output files used for analysis in the HIVE browser are available as a single multi-sample file (Example dataset 2) and in the custom HIVE format (Example dataset 3) from https://github.com/QUB-Simpson-lab/HIVE_browser.

[Back to Contents](#)

BIBLIOGRAPHY

Hao, Y., et al. (2021). "Integrated analysis of multimodal single-cell data." Cell **184**(13): 3573-3587 e3529.

Xin, Y., et al. (2016). "RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes." Cell Metab **24**(4): 608-615.

.

Extra stuff

Combined file

Optionally a '-' in the filename between the CellType and TimePoint (or Condition, etc), eg:
"Amacrine-EARLY.csv"

OR:

(2) One CSV file containing differential (eg. between test and control samples) fold changes for all the samples,

which can be downloaded using the 'Optionally: Download your data as single data file' link after selecting your differential expression files. This single file can make future loading of data easier, as just need to open one file then.

It's format is, eg:

```
Gene,Amacrine:EARLY,Amacrine:MID,Amacrine:LATE,Bipolar:EARLY,Bipolar:MID,Bipolar:LATE,Cone:EARLY,Cone:MID,Cone:LATE
Aak1,0,0.6,0,0,-2.1,0,0,0,0.4
Aamp,0,0.3,0,0,-0.6,0,0,0.3,0
Aars,0,0.4,0,0,-0.3,0,1.2,0,0
...etc....
```

where:

- first column is 'Gene' name
- then several columns with heading format: CellType:TimePoint (or CellType:Condition/Location, etc)
- then gene/protein fold-change rows in format: GeneName, Fold-change:Optional p-value:Optional pct.1:Optional pct2,

Multilevel Explain format:'s names

Eg Islet

Can be used for other datatypes

For this vignette we will use a dataset of single cell RNA-Seq from human islet cells from diabetic and non-diabetic individuals [REF] We have

Import data into HIVE

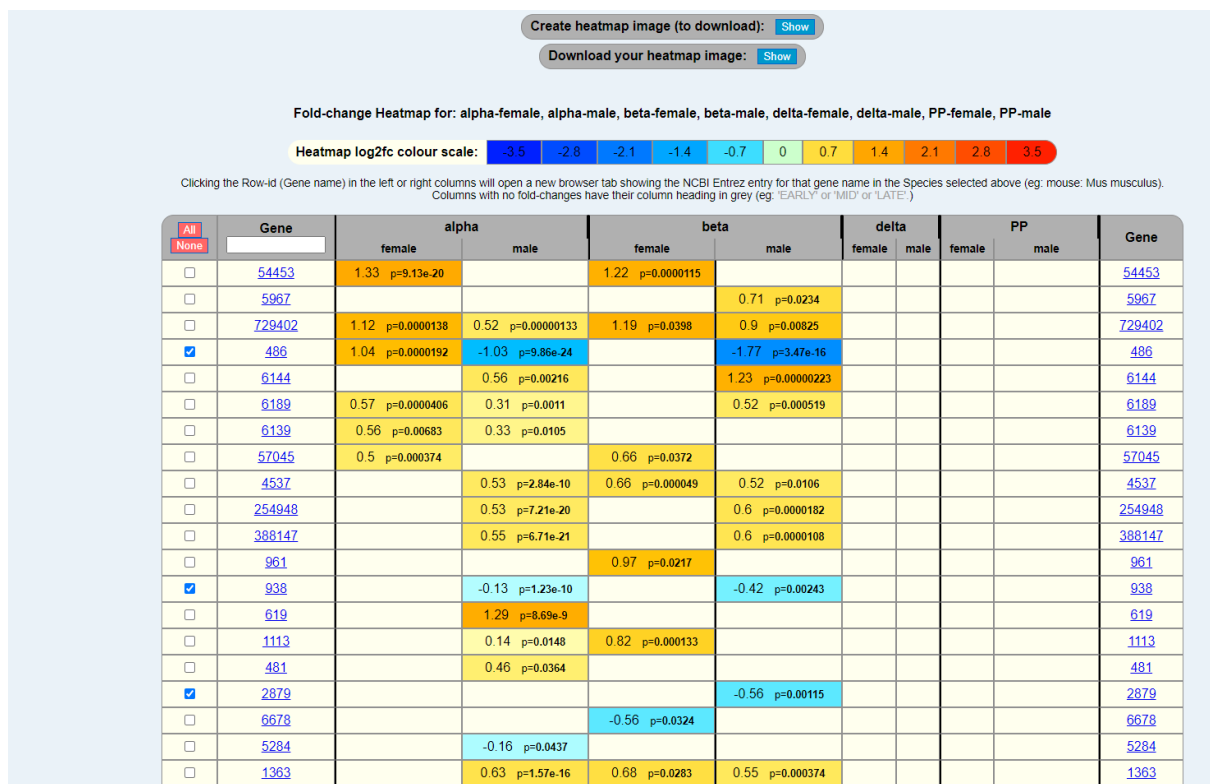
Select genes of interest

For example, genes which are not significantly changed in females but go down in males

Enriched in secretory vesicles

Other way round – genes high in females, unchanged in males - Chromogranin A (in vesicle) and INS-IGF2 readthrough (723961) and RIN2 Ras and Rab interactor 2, a small GTPase involved in membrane trafficking in the early endocytic pathway.

Similarly in alpha cells – genes higher in females (including down in male) enriched in extracellular vesicle



G profiler

Selected genes: 15 [Hide](#)

486, 938, 2879, 1191, 100533181, 2495, 7276, 3326, 2950, 5502, 28958, 100873254, 6319, 5950, 146456

Select gene(s) using the check-boxes in left column of the heatmap table below.

Copy selected genes to clipboard: [Show](#)

G:Profiler: [Hide](#)

1. Select the organism/species (above)
2. Select gene(s) (using the check-boxes in left column of the heatmap table below).

Organism: 9606 : Homo_sapiens : hsapiens

Number of genes selected: 15

[Open G:Profiler with this selected gene-list](#)

Redirecting to G:Profiler in a new tab ...

String DB: [Hide](#)

1. Select the 'Organism/Species' from the drop-down menu above.
2. Select gene(s) (using the check-boxes in left column of the heatmap table below).
3. The Stringdb button, Species and Required-score selection will then be enabled (red).
4. For 'Require score' option: lower score includes more interactions.
5. After clicking this "Show Stringdb image" button, if the String image doesn't appear within a couple of seconds, it's probably because the gene(s) is not be in StringDB for the selected Organism/species.
6. Some genes (eg: 'Malat1') are in Human but not in Mouse Stringdb.
7. Click on a gene in the String image to show its abstract.
8. Click the 'STRING' logo to go to the String webpage in a new browser tab (eg. to view the interaction details).

Organism/Species: 9606 : Homo_sapiens : hsapiens

Number of rows (genes) selected: 15

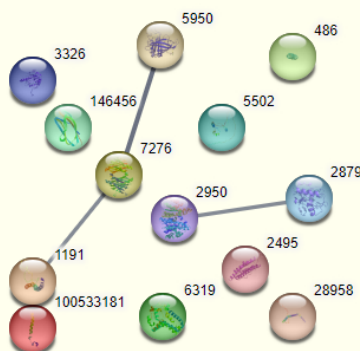
Required score: Additional network nodes: Network flavor:

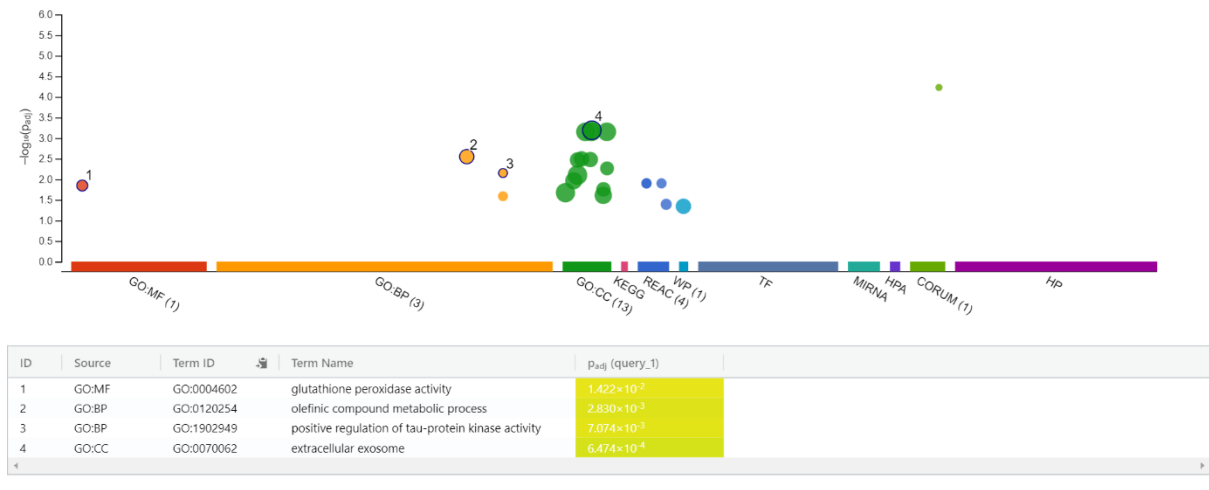
[Show StringDB image here](#)

[Show StringDB in new browser tab](#)

post onload(): Retrieved the StringDB link okay.

If your browser doesn't automatically open String in a new tab, then click this link: <https://string-db.org/cgi/link?to=744075EE90BB817B>





version e108_eg55_p17_0254fbf
date 10/03/2023, 18:15:05
organism hsapiens

g:Profiler

GO:CC		stats																	
Term name ↑	Term ID	P _{adj}	-log ₁₀ (P _{adj})	≤16	486	109	10033181	2485	7276	2950	5562	28958	6279	5950	146435				
cytoplasmic vesicle lumen	GO:0060205	3.325×10 ⁻³	<div><div></div></div>																
extracellular exosome	GO:0070062	6.474×10 ⁻⁴	<div><div></div></div>																
extracellular membrane-bounded organelle	GO:0065010	7.077×10 ⁻⁴	<div><div></div></div>																
extracellular organelle	GO:0043230	7.077×10 ⁻⁴	<div><div></div></div>																
extracellular space	GO:0005615	2.125×10 ⁻²	<div><div></div></div>																
extracellular vesicle	GO:1903561	7.052×10 ⁻⁴	<div><div></div></div>																
ficolin-1-rich granule	GO:0101002	1.761×10 ⁻²	<div><div></div></div>																
ficolin-1-rich granule lumen	GO:1904813	5.456×10 ⁻³	<div><div></div></div>																
secretory granule	GO:0030141	1.087×10 ⁻²	<div><div></div></div>																
secretory granule lumen	GO:0034774	3.206×10 ⁻³	<div><div></div></div>																
secretory vesicle	GO:0099503	2.468×10 ⁻²	<div><div></div></div>																
vesicle	GO:0031982	7.928×10 ⁻³	<div><div></div></div>																
vesicle lumen	GO:0031983	3.406×10 ⁻³	<div><div></div></div>																

1 to 13 of 13 |< < Page 1 of 1 > >|

For STRING set additional network nodes to 0 default (keep option to increase)

When have selected genes and looked at STRING DB make easier to select different genes and rerun String. Way to clear the String input?

Create heatmap image (to download): [Hide](#)

Rows to plot: As selected by the 'Filter genes' above
 Columns to plot: As selected by the 'Filter genes' above
 Columns to plot: As chosen by the checkboxes in left column below (so are in the genelist above)
 Columns to plot: All rows

Column width: 120 pixels.

Font height: 16 pixels.

Space around text: 5 pixels.

☐ Draw fold-change numbers on the heatmap coloured boxes.

NOTE: This currently ignores the "Show all p-values in rows with at least one p-value under the above p-value filter" setting above and therefore draws cells with p-value above the p-value filter, but I can fix this.

[Preview the Heatmap image below](#)

Create heatmap image (to download): [Hide](#)

Rows to plot: As selected by the 'Filter genes' above
 Columns to plot: As selected by the 'Filter genes' above
 Columns to plot: As chosen by the checkboxes in left column below (so are in the genelist above)
 Columns to plot: All rows

Column width: 120 pixels.

Font height: 16 pixels.

Space around text: 5 pixels.

☐ Draw fold-change numbers on the heatmap coloured boxes.

NOTE: This currently ignores the "Show all p-values in rows with at least one p-value under the above p-value filter" setting above and therefore draws cells with p-value above the p-value filter, but I can fix this.

[Preview the Heatmap image below](#)

Download your heatmap image: [Hide](#)

Select Image file type: PNG = Higher quality (Usually the best option)
 then: [Download](#) JPEG/JPG = Larger file size for this heatmap (although is smaller for photos, etc)
 WebP = Newer google (more compact) file for webpages

Gene	alpha female	alpha male	beta female	beta male	delta female	delta male	PP female	PP male
938								
2879								
1191								
100533181								
2495								
7276								
3326								
2950								
5502								
28958								
100873254								
6319								
5950								
146456								

Analysis by age

In beta cells:

Higher in over 45 – mostly mitochondria tRNA genes

Higher in under 45 – various including ribosomal and INS(!)

Down in under 45 mostly mitochondria tRNA genes

Down >0.5 and significant <0.01 in B cell at either age group: Extracellular vesicles

GEO submission

Retinas were dissected from xxx month old control and Akita mice, cells dissociated [REF] and processed on the 10X Genomics platform (gex v2). Sequence data was processed

