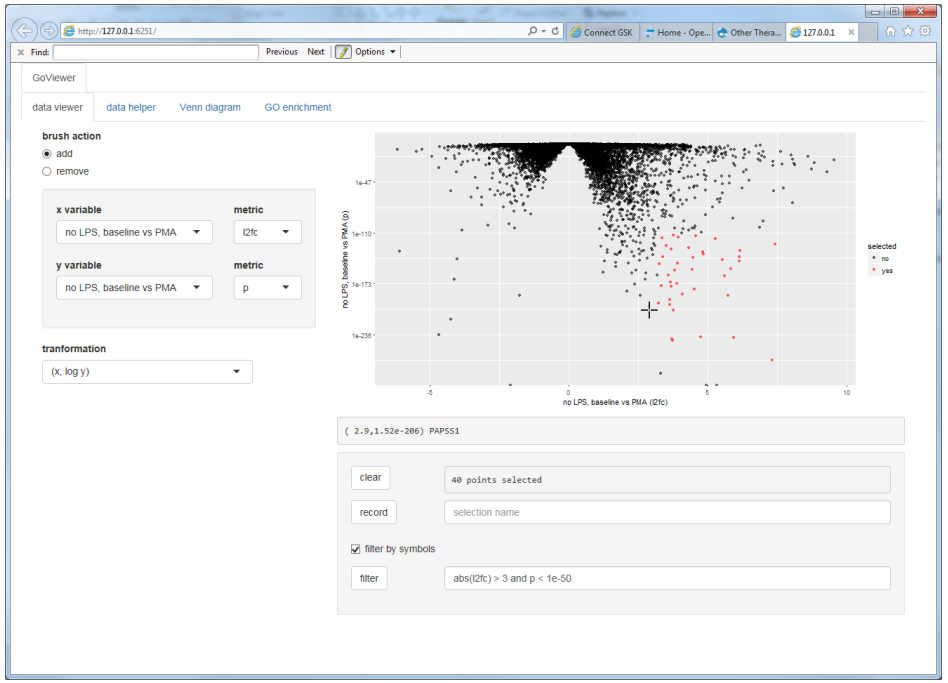
About

This is a top level guide to a developmental tool used to explore selections of genes using gene ontologies for the project two data.

Many of the widgets are developmental and will eventually be used to explain directions in PCA and cross platform tensor statistical outputs.

The following slides will take you through a simple analysis from left to right. Please let us know if you have any comments or suggestions.

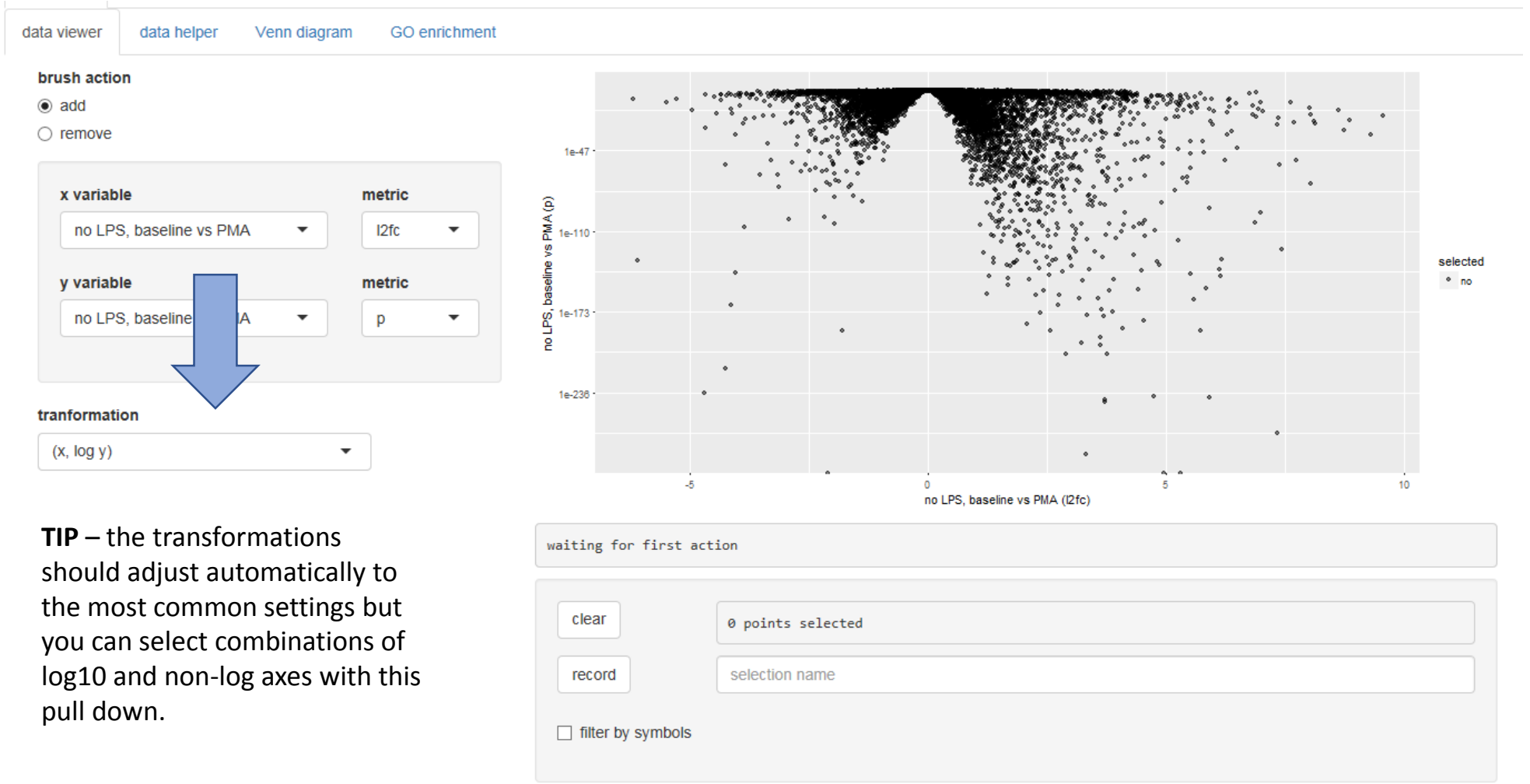
David Willé
GSK / Open Targets 05.05.2018



start here: use the x and y variables to select your end points (e.g. DE outputs) and metric to select the outputs you need. “p” are p-values, “l2fc” log2 fold changes.

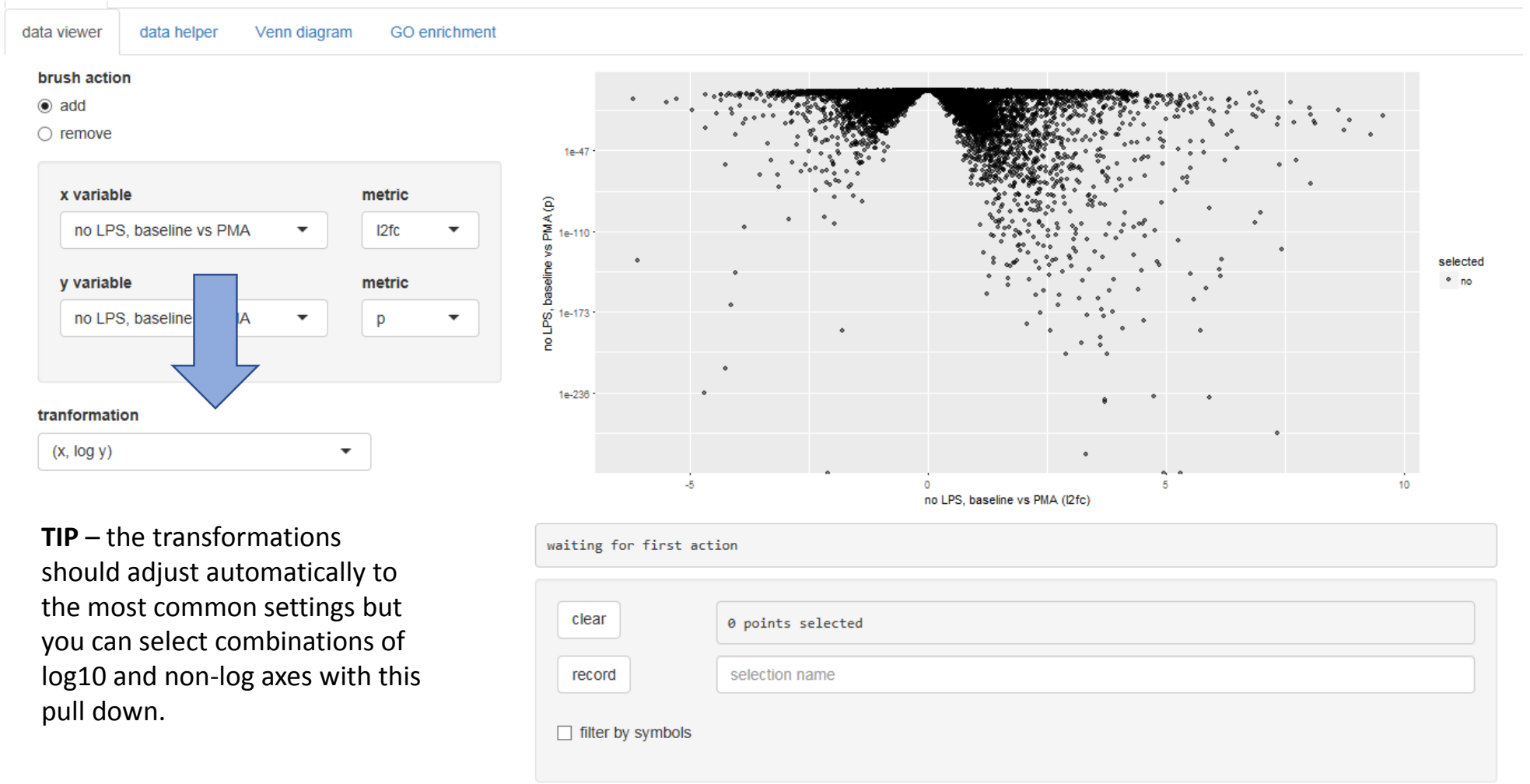
l2fc.se and padj are the associated standard errors and adjusted p-values but you are unlikely to need these.

data viewer



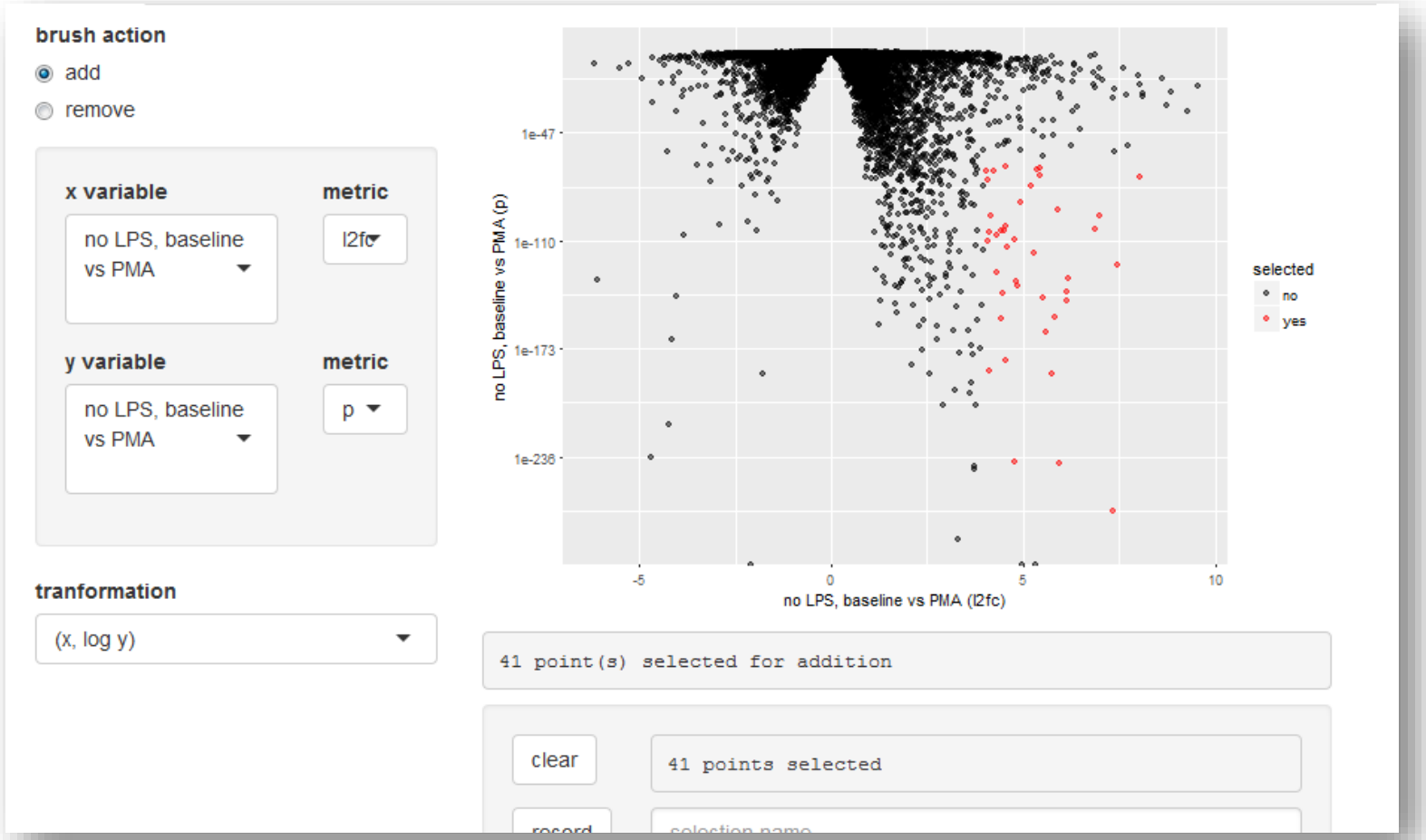
TIP – the transformations should adjust automatically to the most common settings but you can select combinations of log10 and non-log axes with this pull down.

data viewer



TIP – the transformations should adjust automatically to the most common settings but you can select combinations of log10 and non-log axes with this pull down.

data viewer



You can select genes (points) by pulling out rectangles with your cursor. They will be highlighted in the plot and summarised in the information boxes below.

Clicking on a point will give its co-ordinates in the plot and the corresponding gene name.

To remove points, change the brush action in the top left. Alternatively, press clear to clear all selections.

data viewer

You can save a selection of points (genes) for later analysis using the record button. First enter in a name for your selection, and then type record.

ALTERNATIVELY …

data viewer

After clicking the filter by symbols checkbox, you can select points by entering in a symbolic filter. Mathematically terms like abs, < and > are understood as are the words ‘and’ and ‘or’. Everything else is ignored.

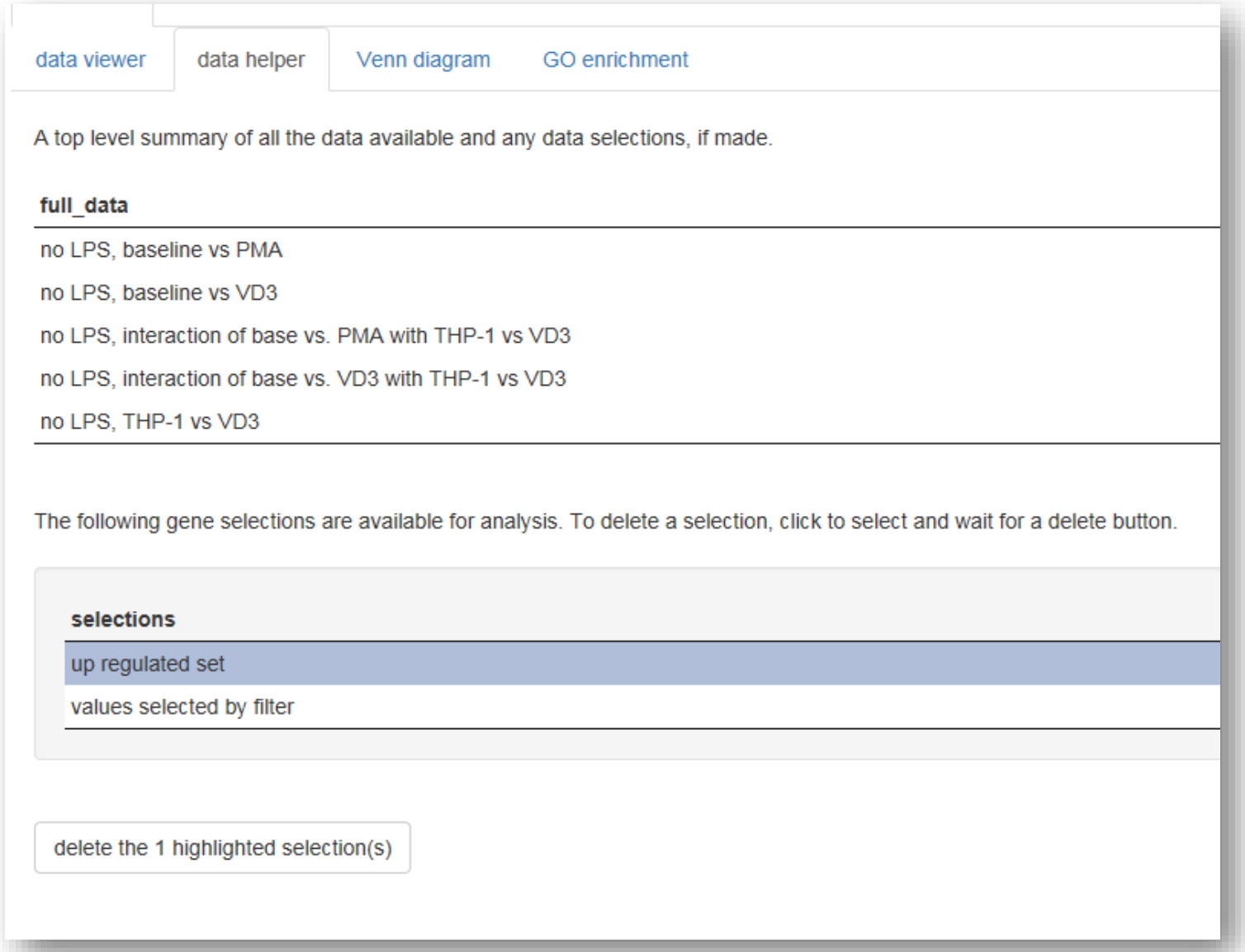
Typing in conditions rather than selecting on the graph is recommended if you need to apply a consistent gene selection filter.

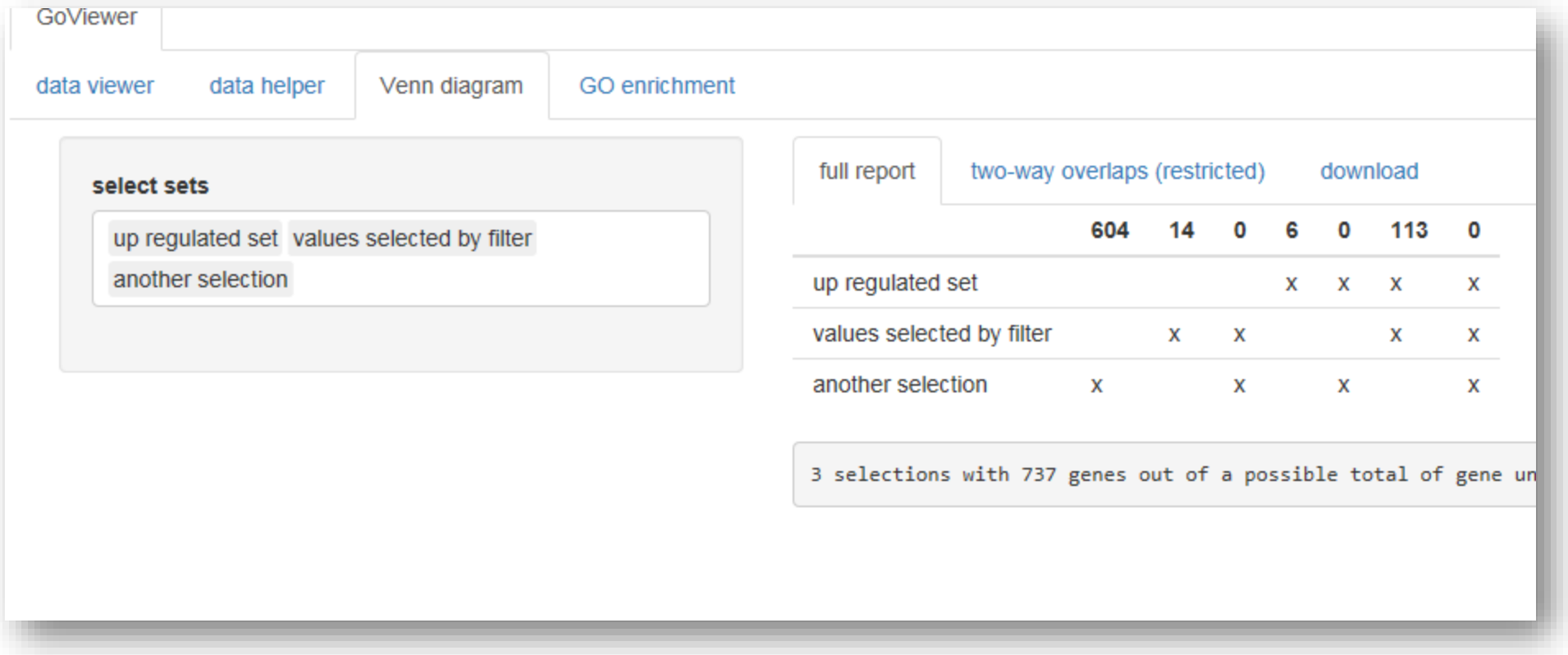
data viewer

The “data helper” tab gives a list of all the input datasets (typically comparisons from another program) and a list of all the gene selections you have made in the data viewer.

You can also use this tool to remove data selections you do not need. Click the no-longer required selections with your mouse, and press the delete button to remove.

data helper

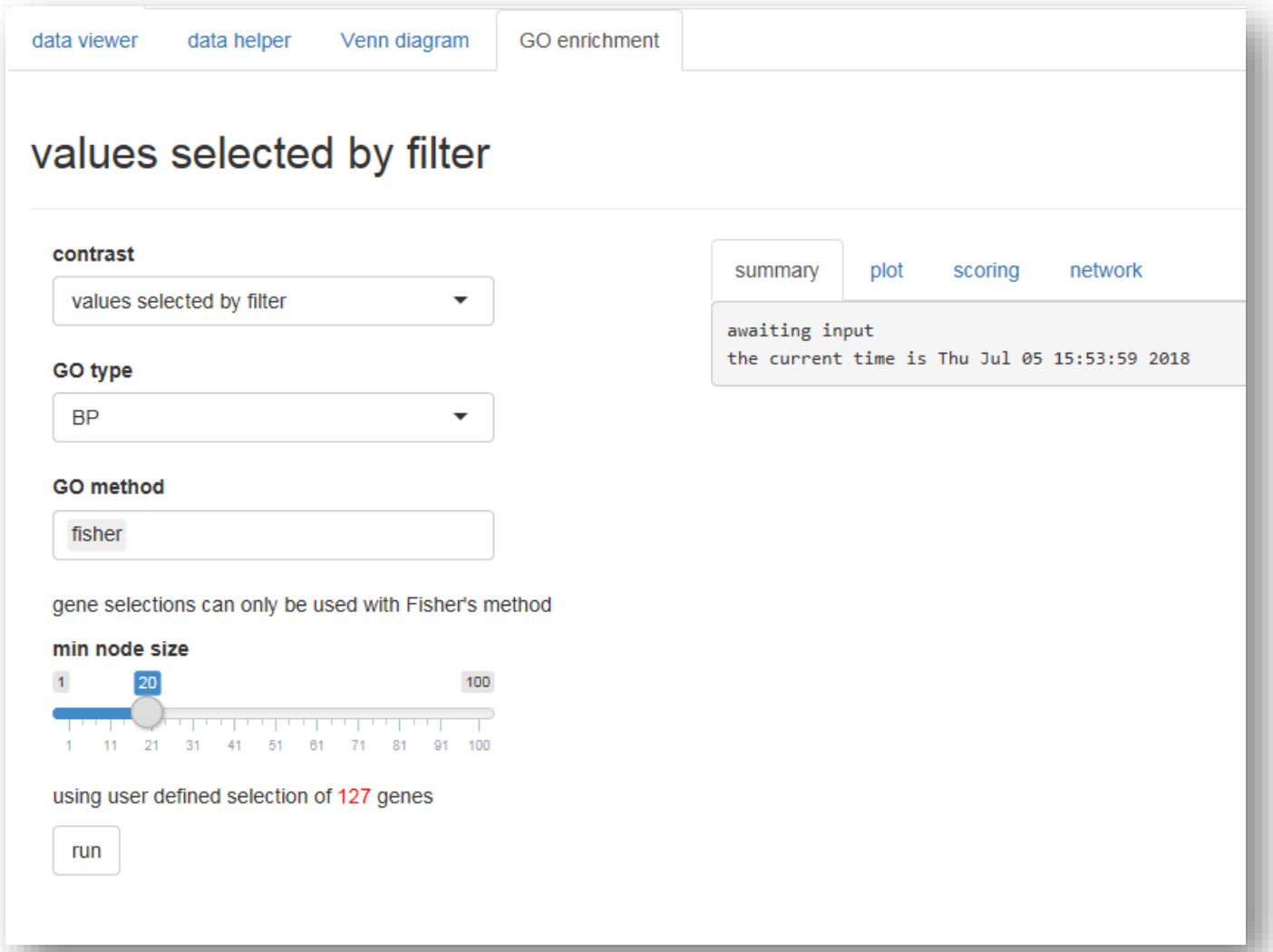




The Venn diagram tab gives summaries of the overlaps between sets of selected genes. A limited use but simpler to use table of two way intersects is also given but this is of lesser general use. You can also download information from the download tab.

This set of tools will need at least two gene selections to use. If you have several gene selections you can select or deselect them by dragging their names out of the set selection box. This will update your view in the obvious way.

gene intersections



The Gene Ontology enrichment tool allows you to identify sets of genes, processes or locations that figure prominently in your gene selections.

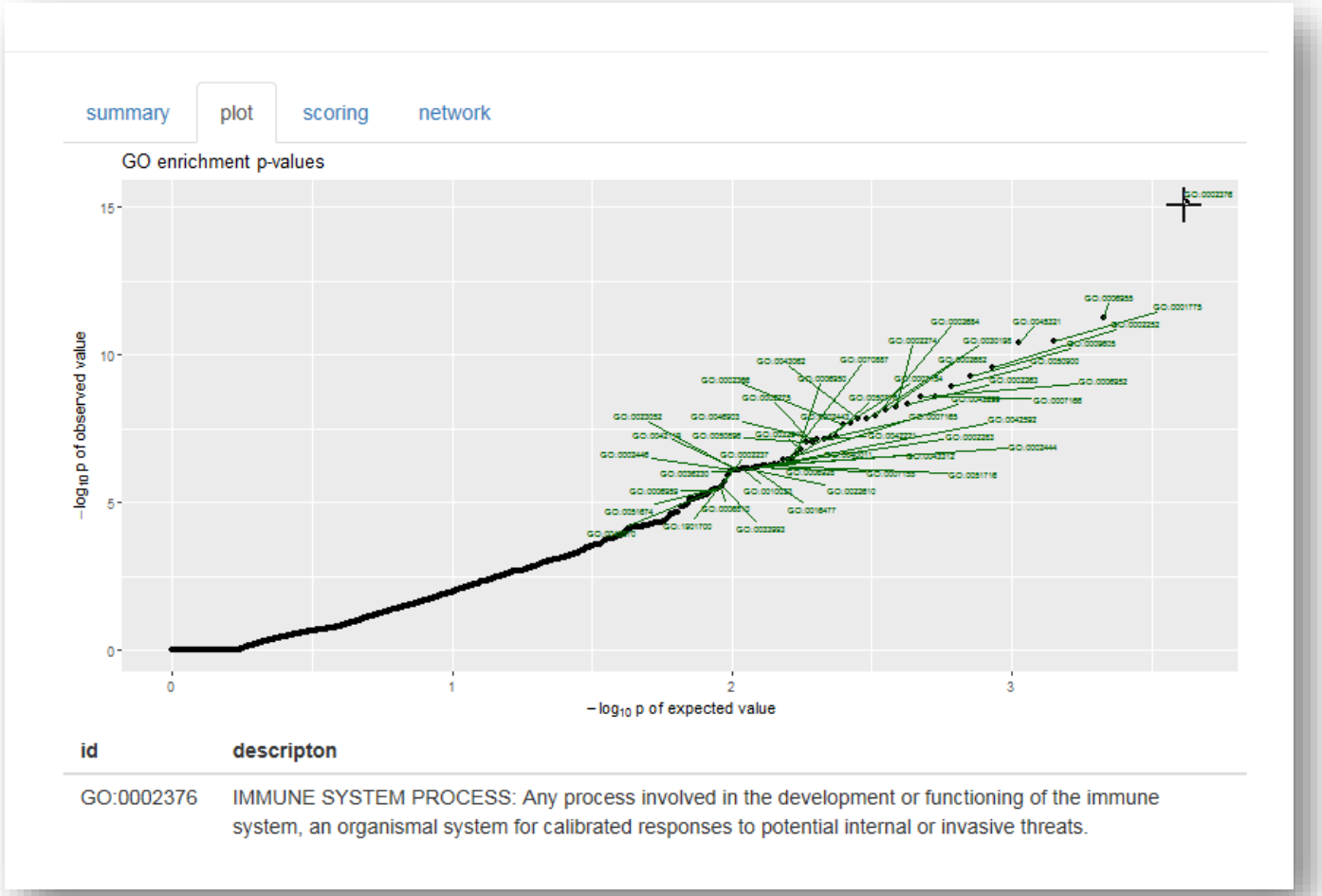
You can select which contrast you require using the pull down contrast menu.

Alternatively you can pick a list of p-values in which case you can then indicate how many you would wish to regard as significant. In this mode, you can also select an alternative G) method based on a KS test.

Once you have selected an analysis, press run and wait.

Note: this is a complicated calculation and may take up to a minute to compute. Wait for the progress message to go and the screen to update.

gene ontologies



The first tab is a graphical summary of the data's suggested GO terms. Again some computation is needed when the tab is selected. Points higher and further to the right are more significant and so potentially more interesting.

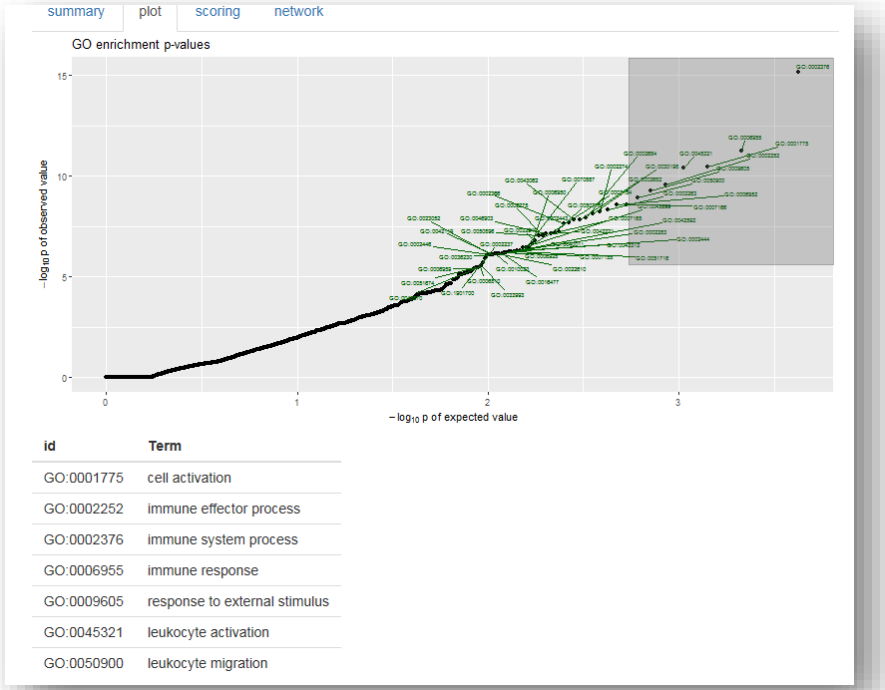
If the data was 'random' the points might lie in a straight line from the origin. Any deviations from this are interesting.

Clicking on a single point will give a summary of the corresponding GO term.

Alternatively ….

gene ontologies

… selecting a set of points will give a listing of the all the corresponding ontology names …



id	Term
GO:0001775	cell activation
GO:0002252	immune effector process
GO:0002376	immune system process
GO:0006955	immune response
GO:0009605	response to external stimulus
GO:0045321	leukocyte activation
GO:0050900	leukocyte migration

summary plot scoring network

GO.ID	Term	Annotated	Significant	Expected	fisher
GO:0002376	immune system process	2304	59	20.75	6.7e-16
GO:0006955	immune response	1576	43	14.19	5.7e-12
GO:0001775	cell activation	1130	35	10.18	3.4e-11
GO:0045321	leukocyte activation	1012	33	9.11	3.8e-11
GO:0002252	immune effector process	968	31	8.72	2.8e-10
GO:0009605	response to external stimulus	1667	41	15.01	5.7e-10
GO:0050900	leukocyte migration	337	18	3.04	1.2e-09
GO:0007166	cell surface receptor signaling pathway	2215	47	19.95	2.5e-09
GO:0006952	defense response	1192	33	10.74	2.7e-09
GO:0002263	cell activation involved in immune respo...	614	23	5.53	4.7e-09
GO:0002274	myeloid leukocyte activation	570	22	5.13	6.3e-09
GO:0002682	regulation of immune system process	1177	32	10.60	7.9e-09
GO:0002684	positive regulation of immune system pro...	814	26	7.33	1.1e-08
GO:0030198	extracellular matrix organization	265	15	2.39	1.5e-08
GO:0043062	extracellular structure organization	266	15	2.40	1.5e-08

… or opening the scoring tab will give you a more detailed tabular view of the same data.

A network tab is also provided with a graphical representation of the selected ontology hierarchy. This is currently in development and does not work correctly.

gene ontologies