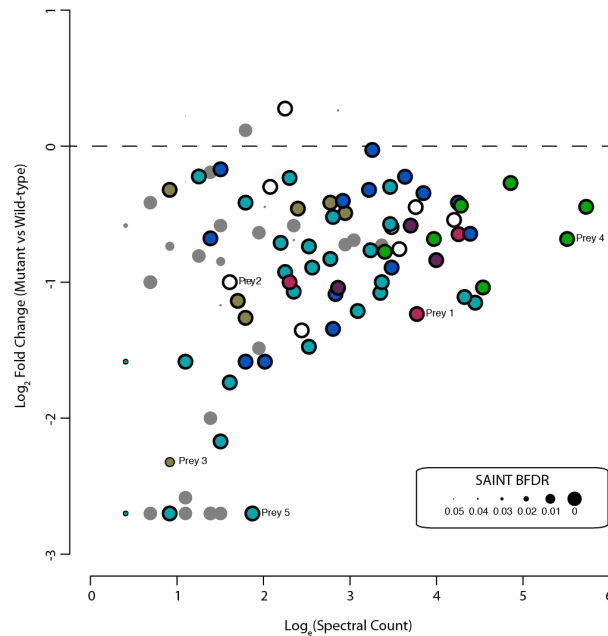


Bait vs Bait Comparison Tool User's Guide **v1.0 (September 3rd, 2014)**

This manual provides usage information for the bait vs bait comparison tool at <http://prohibitools.mshri.on.ca>. This tool will take quantitative protein-protein interaction for two baits and generate a scatterplot summarizing the spectral count, fold change and confidence information, similar to the example below. For more details about the tool, see JDR Knight, G Liu, JP Zhang, A Pasculescu, H Choi and AC Gingras (Submitted) "A web-tool for visualizing quantitative protein-protein interaction data," *Proteomics*.



Inputting data

Data can be input in one of two formats: SAINTexpress or generic, but both must be tab-delimited text. The file should only contain information on the two baits of interest and data for each bait must be found in continuous blocks (meaning, all preys for bait 1 should be listed before preys for bait 2, etc.). The first bait found in the file will be used as the control/default condition. Spectral count and confidence/FDR information will refer to this bait. Fold change data will refer to the fold change of the second bait found in the file relative to this control. Example formats are available for download from the "Sample Input Files" hyperlink on the main page.

1. SAINTexpress (default): After statistically analyzing MS data using SAINTexpress [1], the output is contained in a file called "list.txt". This output can be used directly with the dot plot tool without any modification. Again, this file should only contain information on the two baits of interest; additional rows for other baits analyzed at the same time can be deleted.

2. Generic: For those using an older version of SAINT or other statistical tools, a generic format can be input instead. Here is an example of the generic format:

Bait	PreyGene	AvgSpec	BFDR
Bait1	AHNAK	399.5	0.87
Bait1	FLNA	937.5	0.87
Bait1	ACACA	392	0.87
Bait1	MKI67	165	0.86
Bait1	RANBP2	116	0.86
Bait1	PC	1277.5	0.87
Bait1	CHD4	105	0.86
Bait1	TPR	76	0.86
Bait1	NUMA1	71	0.86
Bait1	CEP170	106	0
Bait1	TCOF1	133.5	0.86
Bait1	TOP2A	87	0.84
Bait1	DLG5	73	0
Bait1	PCCA	120	0.86
Bait2	ACACA	823	0.87
Bait2	AHNAK	172	0.87
Bait2	TPR	153.5	0.86
Bait2	PC	1766	0.87
Bait2	TRIP11	114	0
Bait2	LRPPRC	164	0.86
Bait2	NUMA1	98.5	0.86
Bait2	FLNA	139	0.87
Bait2	GOLGA3	92	0
Bait2	MKI67	74.5	0.87
Bait2	CPS1	98.5	0.86
Bait2	PCCA	322	0.86
Bait2	HSPA9	119.5	0.86
Bait2	HSPD1	88.5	0.86
Bait2	UBR4	35.5	0
Bait2	USO1	93	0

When creating this file, it is important to use the column headers exactly as shown here (Bait, PreyGene, AvgSpec, BFDR). Entries for each bait need to be entered in contiguous blocks (meaning, all preys for bait 1 should be listed before preys for bait 2, etc.). If the spectral sum across replicates is being used as a quantitative measure instead of average spectral count across replicates, the third column should still be labelled as AvgSpec. If some other kind of statistical score is being used instead of an FDR, the available score should be converted, bearing in mind that for FDR zero is the best score and one is the worst. For example, if the statistical score available was the inverse of this, with one being a perfect score, the scores could simply be inverted to give values compatible with this tool.

Parameter Options

- 1. FDR (default 0.05 or 5%):** All preys that satisfy this cutoff for the control bait (i.e. the first bait found in the input file) will be displayed in the scatter plot. The circle size for the prey on the scatter plot indicates the strictest FDR it satisfied, with zero plotted as the largest size and 5% the smallest.
- 2. Normalization (none by default):** No normalization is applied by default but when baits in the same dataset have been run on instruments with varying sensitivity or dynamic range, normalization should be applied. Currently, normalization based on total spectral counts is the only available option, but others will be added in the future.

Output

After the data has been processed, the user will be prompted to download a .zip file that contains the results in a folder. There are several files in this folder.

- 1. baitvbait.pdf:** This is the scatter plot and the file can be opened and edited in Adobe Illustrator or a similar program. In some cases you may get the following warning on opening the file in Illustrator: "The

font AdobePiStd is missing. Affected text will be displayed using a substitute font.” If this occurs, the scatter plot will not display correctly.

To fix this issue on a Mac, copy the file AdobePiStd.otf from /Library/Application/Support/Adobe/PDFL/*Current Version*/Fonts/ and transfer it to the folder /Library/Fonts/. The *Current Version* folder refers to your version of Adobe. On Windows, the font file is located in C:\Program Files\Common Files\Adobe\PDFL*Current Version*\Fonts\ and needs to be placed in C:\Program Files\Adobe\Adobe Illustrator CS5\Support Files\Required\Fonts\. If the AdobePiStd.otf file is missing, it can be downloaded from a number of sites on the web for free in either Mac or Windows format.

2. legend.pdf: This is an FDR legend for the scatter plot and it can be opened and edited in Adobe Illustrator or a similar program.

3. BaitVsBait_data.txt: This file is a tab-delimited table that contains information for each point on the scatter plot. Often in the case of crowded scatter plots, data points and labels may overlap. This table is provided to assist the user in finding preys of interest in such cases, and to help resolve issues as to which points and labels belong to one another.

4. process.log: This file contains information on the input parameters that were selected for the user’s future reference (e.g. to assist with writing the Methods section for a manuscript).

References

[1] Teo, G., Liu, G., Zhang, J., Nesvizhskii, A. I., *et al.*, SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. *Journal of proteomics* 2014, *100*, 37-43. (Available for download from <http://saint-apms.sourceforge.net/Main.html>).

Troubleshooting

Problems generally result from errors in the input file format, and we encourage users to compare their input files against the examples provided on the web page.

Any other issues should be sent to jknight@lunenfeld.ca.