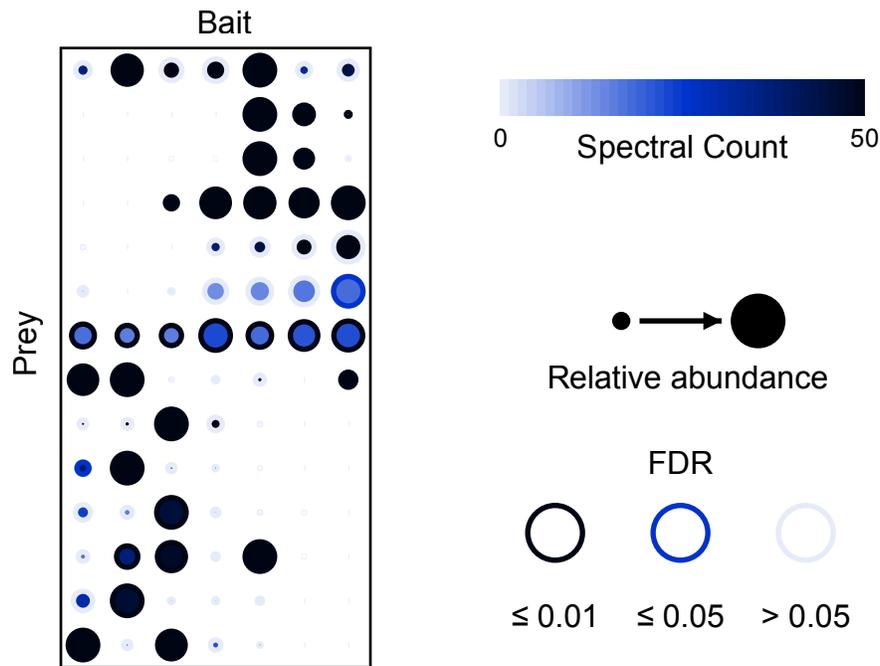




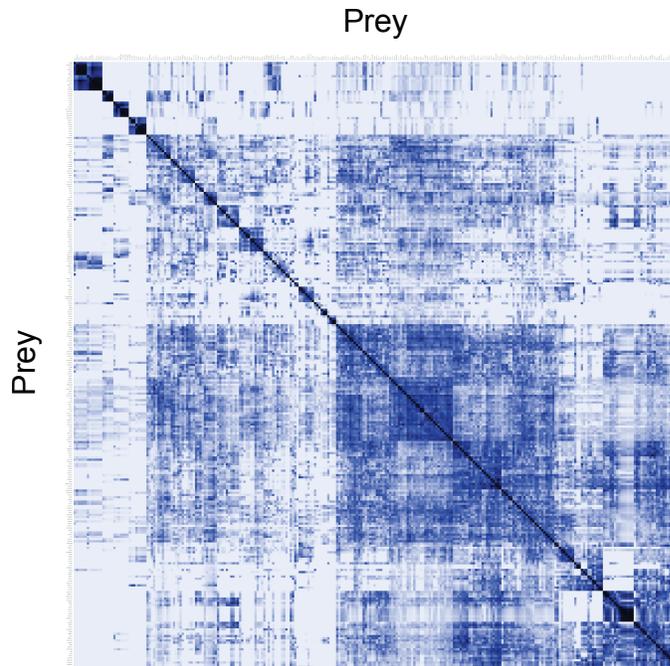
Dot plot Generator

This tool is designed to take a file with quantitative information on bait-prey interactions and visualize those interactions as either a dot plot (shown below) or a heat map. Dot plots have the advantage over heat maps in that they take up the same amount of space but visualize more information. In addition to raw quantitative values being displayed via coloured circles, dot plots display the relative abundance of preys between baits via circle size and confidence in the interaction via coloured edge. Heat maps, however, are better for presenting very large data sets. Regardless, a static dot plot and heat map will be generated for all input datasets.



Distance matrices are a by-product of generating clustered bait vs prey images and we use these distance matrices to create bait vs bait and prey vs prey heat maps as a way of showing similarity between them (as an alternative to comparing columns or rows in the bait vs prey images). We also generate Cytoscape compatible files for these matrices as network diagrams offer another alternative way of visualizing this similarity.

Dot plot generator



Static images, files compatible with our interactive viewer and Cytoscape formatted files for the bait-bait and prey-prey distance matrices will be available upon completion of the task. Alternatively, interactively images can be directly viewed by clicking on the appropriate link. Although a direct link is not provided for an interactive heat map, this can be viewed by selecting the dot plot, and changing the image type to "Heat map" under the display options of the interactive viewer.

Help is available directly on the page by selecting the "help" links next to the section headers. Any additional questions jknight@lunenfeld.ca. For more information, see [PMID: 25422071](#).

Inputting data

Specify the file for upload and the file type, then hit "Process". Explicit support is provided for data files output from SAINT and CRAPome (CRAPome matrix format is not supported, however). Datasets from other tools or pipelines can be input by selecting the "Other" option. Files must be in tabular format as tab-delimited text. At a minimum, the file must contain four columns specifying the bait/condition, prey/readout, abundance measure (spectral count, intensity, etc) and a confidence metric (e.g. FDR). Sample input files are available for download by clicking the "help" link and contain BioID published in [PMID: 24255178](#).

For detailed information on tools that generate compatible input for ProHits-viz, see the references and links below.

References: [SAINT - PMID:21131968](#), [SAINTexpress - PMID:24513533](#), [SAINT-MS1 - PMID: 22352807](#), [ProHits - PMID:20944583](#), [ProHits Protocol - PMID:22948730](#), [ProHits 4.0 - PMID:27132685](#), [CRAPome - PMID:23921808](#)

Links: [SAINT-APMS.sourceforge.net](#), [ProHits.com](#), [CRAPome.org](#), [GalaxyP](#)

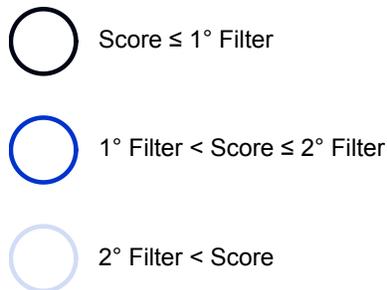
Data columns

Up to five columns will be used when processing the input file: a column listing the 1) baits, 2) preys, 3) abundance measure (spectral count, intensity or simply a non-negative number), 4) score for filtering (for unknown filter types, the direction of the filter must be specified, i.e. is a smaller number better, or vice versa), and 5) control values for performing control subtraction, i.e. subtracting the average control value from the observed prey abundance (this is optional and can either be a pipe-separated list of values or a single averaged number).

Parameters

There are several options available for processing and visualizing the data. Defaults are suggested on the input page but these can be changed as needed.

1. Primary filter: All preys that satisfy this score filter for at least one bait will be displayed in the dot plot. If a prey satisfies this filter for at least one bait, all the quantitative values for this prey – even those that did not satisfy the cutoff in particular bait-prey pairs - will be returned. Prey confidence be indicated by the circle edge as shown below:



2. Secondary filter: Interactions that do not pass the primary score filter but pass this secondary filter will be marked with an intermediate intensity in the dot plot. Interactions that do not pass either filter will be marked with a low intensity edge. The secondary filter can be adjusted depending on the dataset to allow a greater or lesser number of interactions into this “medium” confidence range.

3. Minimum abundance value: In addition to the above filter requirement, a prey must have an abundance value (e.g. spectral count) above this minimum to be included in the analysis. As with the primary filter cutoff, once a prey passes this threshold for one bait, all the values for this prey are returned for all baits, whether or not they pass the threshold for this specific pair.

4. Maximum abundance value: Any preys with an abundance value (e.g. spectral count) above this cutoff will be capped at this maximum in the output images. This is to give greater visual dynamic range for lower values when outlier preys with very high abundance are present. This cutoff will be dependent on the instrument and interaction method used and should be selected based on the data set.

5. Control subtraction: The average value of a prey across control samples will be subtracted from the detected value for the bait if this is set to "yes". The quantitative value

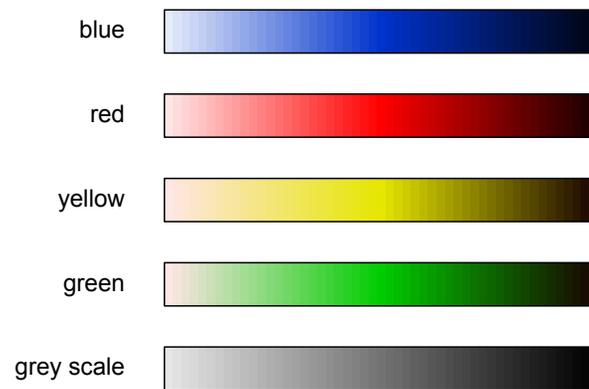
Dot plot generator

for the prey becomes the value above and beyond what is seen across the control samples. Specify the column to use for controls in the adjacent "Control column" field.

6. 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. The options for normalization are to divide by the total abundance for all proteins identified in the run or normalize based on a specific prey). Normalization will be applied after control subtraction if both are specified.

7. Log transform (default no): If desired, data can be log-transformed by base 2, base 10 or natural. Log transformation will be performed after control subtraction and/or normalization if these are also specified.

8. Color scale (default blue-black): Various options are available for the fill and edge color, and can be applied independently to each.



Clustering options

1. Agglomerative Hierarchical (Hierarchical): This is the default clustering option and is executed using R. There are several options for calculating the distance metric and for the clustering criterion. Canberra is the default distance metric and Ward's method the default clustering type. From our own experience we have found Canberra to be a very good metric to use for protein-protein interaction data, although other options are available and may produce more desirable results. The defaults clustering type is Ward's, which acts to minimize variance within clusters, although many of the types available will produce comparable results.

Available distance metrics: binary, Canberra, Euclidean, Manhattan, maximum and Minkowski

Available clustering criteria: average, centroid, complete, McQuitty, median, single and Ward's

2. Nested Clustering (Biclustering): Another and more sophisticated clustering approach is available within ProHits-viz. This is a probabilistic biclustering approach termed "Nested Clustering", details of which can be found here: [PMID: 20571534](https://pubmed.ncbi.nlm.nih.gov/20571534/). This approach first clusters baits based on the similarity of their prey interaction profiles, then pools preys with

Dot plot generator

similar abundances within these clusters to form a nested cluster. Note that this clustering approach will take significantly longer than the hierarchical clustering option, especially for large data sets. Data sets must have at least 3 baits to use this clustering option.

3. No-clustering (Clustering: none): The user can generate dot plots without clustering if desired. In this case, a list of bait and prey genes in the desired display order must be supplied in the text boxes. Only baits and preys entered in the text box will be included in the dot plot. Bait and prey names must be entered as they appear in the input file name and are case sensitive. In some cases the user may want to control which baits are shown in the dot plot and their ordering, while wanting to show and cluster all preys, or vice versa. This can be done as well by selecting “cluster all” under the bait or prey options menu.

Output

After the data has been processed, the user will be prompted to download a Dotplot_results.zip file that contains the results in three subfolders. By default the parent unzipped folder will be named "RESULTS" but this name can be modified as desired. The subfolders are:

1. PDFs: This folder contains the dot plot and heat map images, as well as a legend. The output bait vs bait and prey vs prey images are the clustered distance matrices visualized as heat maps. In contrast to similar images generated by our Correlation Tool, baits (or preys) in these images have not been subjected to correlation analysis prior to clustering. Instead, the metric used is the distance metric calculated using the clustering option specified above. These files can be opened and edited in Adobe Illustrator or a similar program. In some cases the following warning may appear on opening the file in Illustrator: “The font AdobePiStd is missing. Affected text will be displayed using a substitute font.” If this occurs, the image 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 freely downloaded from a number of sites on the web in either Mac or Windows format.

2. InteractiveFiles: Contains plain text files formatted for our interactive viewer found at [ProHits-viz](#) that allows navigation and analysis of images. These files should be loaded directly into that tool without modification. Their extension is _df.txt.

3. CytoscapeFiles: This folder contains plain text files for importing to Cytoscape. The file 'bait-prey_cytoscape.txt' contains only bait-prey pairs that have passed the primary filter, and with the abundance equal to its post-processing value. For example, if control subtraction and normalization have been selected, the abundance value in this file will be the value after control subtraction and normalization have been performed. This allows users to import only desired bait-prey pairs into Cytoscape with an abundance value suitable for edge-weighting that has been processed according to the user's need. Two additional files are also present in this folder that contain all bait-bait and prey-prey pairs from the distance matrices with an additional column containing the distance metric itself that can be used for filtering which pairs will be displayed in Cytoscape. These files are generated so that users can view bait-bait and prey-prey information in a network format as an alternative the heat map format we provide here.

Dot plot generator

4. OtherFiles: This folder contains bait-bait and prey-prey dendrograms and will contain additional files generated by the Nested Clustering option, if selected.

Finally, the "RESULTS" folder will also contain a process.log file the 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).

Troubleshooting

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

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