STRAP Tutorial

A Guide to Using the Software Tool for Rapid Annotation of Proteins

STRAP saves you time by automatically annotating a protein list with information that helps you meaningfully interpret your mass spectrometry data. This tutorial will teach you how to use it.



Installation

STRAP runs on Windows XP or higher and requires the Microsoft .NET Framework. The STRAP program and sample files can be downloaded as zip files (*STRAP.zip* and *STRAP-Sample-Files.zip*) from <u>CPC Tools</u>. There are directions on the Download Page for installing STRAP and for uncompressing the directory *STRAP Sample Files*. Installation of STRAP should automatically install the latest version of the .NET Framework.

Opening and Saving Protein Lists

The starting point for any analysis with STRAP is a protein list in a variety of forms, usually originating from search results on raw spectra from mass spectrometry

experiments. Sample files are provided in the *STRAP Sample Files* directory for exploring the capabilities of STRAP. The sample data set originates from proteins that were overexpressed in each of three different B-cell proliferative states from a lymphomagenesis biomarker study.

STRAP can open protXML files generated from the <u>Trans-Proteomic Pipeline (TPP)</u>, DAT files from <u>Mascot</u> searches, and plain-text files containing one <u>UniProtKB</u> protein name per line. To open the sample data, go to File->Open, select "Protein List (*.txt)" as the file type, and select set1.txt, set2.txt, and set3.txt.

All protein information is retrieved from the UniProt database, but additional gene ontology (GO) annotation can be imported from the <u>EBI database</u> by clicking Edit->Add Annotations from EBI.



After pressing the Open button, STRAP will open three tabs and retrieve information about the proteins in your lists over several minutes.

🔛 STR	AP by Vivek Bhati	a (viv@bu.edu)						[
File	Edit View Help								
0 💕		🖹 🍬 🧹 E. 🔇 🖉							
set1	set2 set3								
	SwissProt Accession	Uniprot ID Accession Number	Name	Primary Gene Name	Taxonomy	Length	Function	Catalytic Activity	GO:Biologi Process
•	O08749	<u>008749</u>	Dihydroli	DId	Mus musculus	509	Lipoamide d	Protein N(6)-(dihy	GO:004545
	O89053	<u>089053</u>	Coronin-1A	Coro 1a	Mus musculus	461	May be a cr		GO:000681
	P17182	P17182	Alpha-en	Eno1	Mus musculus	434	Multifunctio	2-phospho-D-glyc	GO:000605
	P18760	P18760	Cofilin-1	Cfl1	Mus musculus	166	Controls rev		GO:000091
	P26443	P26443	Glutamat	Glud1	Mus musculus	558	May be inv	L-glutamate + H(2	GO:000652
	P40124	P40124	Adenylyl	Cap1	Mus musculus	474	Directly reg		GO:003003
	P45376	P45376	Aldose re	Akr1b1	Mus musculus	316	Catalyzes th	Alditol + NAD(P)(GO:005511
	P62827	P62827	GTP-bin	Ran	Mus musculus	216	GTP-bindin		GO:000660
	Q8BG05	Q8BG05	Heteroge	Hnmpa3	Mus musculus		Plays a role		GO:000635
	Q8R081	Q8R081	Heteroge	Hnmpl	Mus musculus	586	This protein		GO:000635
	Q99020	<u>Q99020</u>	Heteroge	Hnmpab	Mus musculus	285	Transcriptio	5 I I I I I I I I I I I I I I I I I I I	GO:000183
<			-						>
Retriev	ing 088569 (entry 12 o	f 22)	3						

Since it takes time for STRAP to annotate all proteins, it's important that you save these lists so that you don't waste time downloading annotations every time you want to look at a data set. Once the protein annotation table is completely filled, you can save it as a STRAP Annotation Text File (*.atxt).

If you need to close the currently open tab, right click on it, and click "Close Active Tab." You can also click the "X" button in the toolbar to achieve the same thing.

The Protein Annotation Table

The protein annotation table has many categories. Each category can be sorted if you click on its header at the top of the window. The columns in STRAP's annotation table include columns for general protein information (e.g., gene name) as well as for GO terms. The GO categories used in STRAP are a simplification of those released by the <u>Gene Ontology Consortium</u>.

Clicking on the link in the Uniprot ID column will take you to the protein's entry in the UniProtKB database. For cases in which you'd like to create a custom FASTA database containing only proteins in a particular data set, you can export a FASTA file by clicking on File->Export to FASTA.



The sort feature can also be used to group proteins with similar GO terms:

If you'd like to manually change a protein's annotation, you can do so by double-clicking the corresponding GO category cell to toggle its value:

P STR	ар Бу	y Vivek Bhatia (v	riv@bu.edu)					_ 🗆 🗙
<u>F</u> ile	Edit	<u>V</u> iew <u>H</u> elp						
0 📂			🍬 🖌 E. 🔇	3 4 1 3				
set1	set2	set3						
		Macromolecular complex	Cytoplasm	Other	Binding	Catalytic Activity	Enzyme regulator activity	Other
•								
		1			<u> </u>	<u> </u>		
			Change Anno	tation Confirm	ation			
			You're about to c	hange a protein's g Yes	ene annotation. Are yo No on't ask me again	u sure you want to do	this?	
								
			•					
								
								~
<								
Ready								.::

If you're curious about the lineage of a particular protein's GO terms, you can investigate further through the Gene Ontology Term Browser. This browser can be accessed by right-clicking on any protein and pressing "Browse GO terms..." to display a window such as this:

P45376 - Gene Ontology Te	rm Browser	
-Biological Process- 0055114 metabolic process -Cellular Component- 0005737 (cytoplasm, intracellular) -Molecular Function- 0004032 (catalytic activity, oxidore	 cellular_component (GO:0005575) cell (GO:0005623) cell part (GO:0044464) intracellular (GO:0005622) intracellular part (GO:0044424) cytoplasm (GO:0005737) 	

Charts for GO Terms

You can create pie charts and bar charts to quickly interpret the biology of data sets via their associated GO information. Pie charts illustrate the variety of Biological Process, Cellular Component, and Molecular Function GO terms in *one* data set, whereas bar charts illustrate the differences in associated GO terms between *several* data sets.

To create a GO pie chart, first select the tab that contains the fully annotated data set of interest. Then, click the pie button on the toolbar. The resulting pie chart can be saved as an image by right-clicking on the pie chart itself and clicking "Save Image As...".



To create a GO bar chart, first make sure you have at least two data sets open and fully annotated. Then, click on the bar chart icon in the toolbar and configure the bar chart to your liking in this dialog box:

Select the GO term class to compare: Biological Process Cellular Component Molecular Function How would you like to compare your data? Table (recommended for >4 data sets) Bar Chart Please select the protein lists you would like to include in the bar chart: I set1 I act 2
✓ set2 ✓ set3

Press *OK* to see the resulting graph and/or table. To adjust the *y*-axis of the bar graph to see all possible GO terms, right click anywhere on the bar graph and select *Set Scale to Default*.

For comparisons that involve more than 4 files, it may be more useful to see the comparison in tabular form. If you have more than 4 files, check the "Table" checkbox to see your comparison in tabular form.

There are three major types of GO terms: biological process, cellular component, and molecular function. Comparisons between data sets can only be made within one of these major GO categories at a time. These images can be further edited or saved via the context menu as well. Here is an example of a comparison of molecular function GO terms between the three protein lists in the sample data:



Sending Data to the Gaggle

The Institute for Systems Biology has created a tool called <u>the Gaggle</u> which connects to many other data analysis programs (most of which are gene-oriented). STRAP can export genes associated with each protein to the Gaggle via <u>Firegoose</u>. To do this, open an annotated protein list and click the goose icon in the STRAP toolbar. This will open up an HTML page in Firefox, and the data contained in the HTML page can be broadcast to other Geese (e.g., Cytoscape).

Conclusion

This tutorial gives you an idea of what STRAP can do and how it can fit into your data processing workflow. In summary, STRAP can:

• Annotate a list of proteins given a DAT, protXML, or plain text file

- Group proteins in a list by their gene ontology (GO) terms
 Modify GO annotations of proteins
 Use a GO pie chart to interpret a protein list
 Use a GO bar chart to compare protein lists
 Export proteins in lists to tools in the Gaggle

If you have any questions or suggestions, please Contact Us.