



Getting Started with Scaffold 4

Proteome Software Users Group 2013

Tutorial Session

Module 2: Viewing and Filtering Your Data

Getting Started with Scaffold 4

User's Tutorial - Proteome Software Users Group 2013

Many resources are available to help you get started using Scaffold. The Proteome Software website offers video tutorials, white papers and FAQ's on various topics at <http://proteome-software.wikispaces.com/Resource+Library>. The Scaffold User's Guide is found under the Help menu, and there is an online Help system built into the program which provides answers to many of your questions. You can find quick answers to many questions by clicking on the help icons scattered throughout the program. This tutorial is intended to supplement these resources. It does not give a complete introduction to Scaffold. Rather, it will focus on some of the features that we have found to cause the most confusion in new users.

The modules are designed to be independent, so you may choose to work on the units of greatest interest to you.

For these exercises, we will use the demo files available through Scaffold's Help menu under Open Demo Files.

Contents:

Getting Started with Scaffold 4.....	0
Module 2: Viewing and Filtering Your Data	1
Exercise 1: Setting Probability Thresholds	1
Exercise 2: Setting FDR Thresholds	3
Exercise 3: Using the FDR Browser	4
Exercise 4: Using Bulk Operations in the Samples View	5
Exercise 5: Adding Biological Information	7

Module 2: Viewing and Filtering Your Data

In the Samples View, it is easy to tell whether or not the data has been searched against a decoy database. If there are no decoys, the FDR Dashboard (the information box at the lower left of the Scaffold display) is blue. The FDR values presented are estimates based on the probabilities determined by the Prophet algorithms. In a decoy search, the dashboard is pink and the FDR's are calculated values based on the ratio of the counts of decoy to target identifications.

Depending on which type of search has been done, the options available for filtering your data to ensure quality are different. In a non-decoy search, you must set probability thresholds and a minimum number of peptides. In a decoy search, you may use probability thresholds, or you may specify FDR targets and allow Scaffold to determine the best settings to achieve them.

Thresholds are set through three dropdowns at the top of the screen, the Protein Threshold, the Min # Peptides and the Peptide Threshold.

Exercise 1: Setting Probability Thresholds



Open tutorial_4.

a) Is this a decoy or non-decoy search?

Begin with the thresholds set to 99%, 2 and 95%.

b) How many proteins are displayed?

c) What are the estimated Peptide and Protein FDR rates?

Experiment with different threshold settings and observe the effects on number of proteins, Peptide FDR and Protein FDR.

d) With Min # Peptides set to 1 and Protein Threshold at 20%, what is the minimum setting for Peptide Threshold that will keep the estimated Peptide FDR below 1%?

e) Keeping the other settings as they were in d), what is the minimum protein threshold that produces an estimated Protein FDR of less than 1%?

Please note that this is an exercise to familiarize you with the effects of the different threshold settings. In a probability-based experiment, generally you would choose fairly high probabilities for Protein and Peptide thresholds, e.g. $\geq 95\%$, and a minimum Peptide number of 2 or more for publication.

Exercise 1 Result:

- a) Non-decoy
- b) 32
- c) Estimated Peptide FDR is 0.31%, Estimated Protein FDR is 0.1%
- d) This is a trick question. If you limit yourself to the dropdown choices, the answer is 90%. It is possible to type other values directly into the threshold box, however, and if you do this you could reduce it to 85%. Try it.
- e) Assuming that the Peptide Threshold is set to either 90% or 85%, from the dropdown the Protein Threshold can be set at 80%. We could type in a lower value, however, and reduce it to 51% and still have an estimated protein FDR of 0.3%

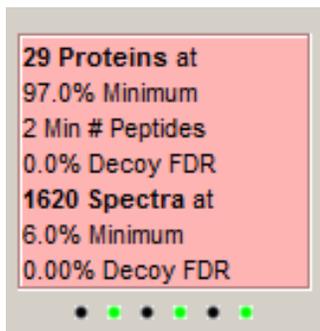
Exercise 2: Setting FDR Thresholds

When a decoy search has been performed, Scaffold can rely on direct calculations of the Peptide and Protein FDR's. Scaffold 4 allows you to set FDR targets for peptides and proteins, and then it finds the probability thresholds that maximize the number of proteins while keeping the FDR's at or below the targets.



Open tutorial_2. Set the Minimum number of Peptides to 1. Set the Peptide and Protein Thresholds to 1% FDR.

Look at the FDR Dashboard. It shows the protein and peptide probability thresholds that Scaffold picked in order to give the most proteins possible while still meeting the FDR targets set and respecting the minimum number of peptides specified.



Notice that the minimum number of peptides is 2, not 1. Scaffold tried all possible combinations of thresholds with one *or more* peptide minima, and reports the most stringent thresholds that satisfied all of the criteria while giving the maximum possible number of proteins.

Try changing the Min # of Peptides filter and see what happens to the number of proteins .

You should see that you can increase the number to 2 without changing any values in the dashboard, but if you increase it to 3, the number of proteins decreases.

Now, try setting the Peptide and Protein Thresholds to the values shown in the Dashboard. Observe what happens to the number of proteins and FDR rates.

You will need to type a number into each dropdown to enter a custom FDR value. You should see the results unchanged, as these settings are equivalent to those calculated by Scaffold.

Set the Peptide Threshold to 2.5% FDR. You must type % FDR after the number to set a custom FDR threshold.

Exercise 3: Using the FDR Browser



Open tutorial_2, set the thresholds to 99%, 2 and 95%. Go to the Statistics View. The FDR Browser is in the upper left quadrant.

Whenever a decoy search is loaded, Scaffold computes and displays an FDR Browser in the Statistics View. This is a heat map in which the color indicates the number of proteins identified as a function of the combination of peptide and protein probability thresholds. Actually, Scaffold makes a series of these heat maps, one for each possible value of the minimum number of peptides. You can step through the various maps by changing the Min # Peptides setting at the bottom. If you hover over any point in the browser, a tooltip displays the number of proteins and other information for that point.

The FDR Browser can help you understand your data better and make informed choices about your threshold settings. If you are doing a peptidomics experiment, for instance, you may be willing to settle for fewer protein identifications in order to get very high quality peptide assignments. You could explore the areas toward the right side of the graph. On the other hand, if you are primarily interested in identifying a large number of candidate proteins, but peptide quality is not as important, you might look along the top of the graph. Often there are many possible threshold combinations that give nearly equivalent numbers of protein identifications, and the best solution for your experiment may not be the one the program would have chosen.

When you look at the browser, you will see two spots marked with cross-hairs. The blue dot indicates the threshold settings that give the most proteins and meet the criteria specified in the controls under the browser. The pink dot represents the current setting in the program thresholds at the top of the screen. It is only seen if the Min # Peptides setting below the Browser matches the current Min # Peptides setting.

If it is impossible to meet the FDR cutoffs specified below the Browser with a pair of probability thresholds, a gray background is shown in the Browser.

Try adjusting the cutoff controls to find gray areas, and to see where the optimal solution falls for different values.

Now set the Peptide Threshold at the top of the screen to 1% FDR.

Notice that the Max Peptide FDR under the Browser also switches to 1% FDR. If the thresholds are set to FDR values, the Browser reflects those values.

Double-click on a point in the FDR Browser. Notice the Threshold settings at the top of the screen, and the effect on the FDR Dashboard.

You may set your thresholds directly through the Browser.

Exercise 4: Using Bulk Operations in the Samples View

In Scaffold 4, if you right-click in the Samples View to bring up the context menu, you will see a number of new options. Many of these give the opportunity to do bulk operations on the proteins in **the Samples View**.



Open the tutorial_2. Lower the thresholds to 20%, 1, 0%.

Right-click in the Samples View, go to Show/Hide and select Hide Decoys.

The pink decoys disappear, but notice that the FDR values in the Dashboard do not change. The proteins are only hidden, not discarded.

Select Show Decoys to bring them back.

Now highlight a group of proteins. To select multiple proteins, use one of the following methods:

- a) Hold down CTRL while clicking on the proteins to be selected**
- b) Hold down SHIFT while clicking on the first and last of a series of adjacent proteins**
- c) With the cursor in the Samples table, right-click and choose "Select All"**

Right-click, go to Show/Hide and choose Hide others.

Only the selected proteins remain.

Right-click again and choose Show All.

Through bulk operations, you can also expand or collapse all of the clusters, or apply different colors of stars to groups of proteins. You can combine bulk operations with filtering for even more power.



Using the Search box, filter to show only the proteins with BOVIN in their accession numbers. Right-click, Select All, then right-click and add blue stars. Clear the Search box.

Now filter the PIG proteins, Select All and add orange stars. Clear the Search box, then collapse all clusters. Sort on the Starred column., Select the proteins with stars, and hide others.

The result is that we have grouped the proteins into categories (protein clusters containing cow proteins, clusters containing pig proteins, clusters containing both and clusters with neither. Then we filtered out the ones with neither. This can be a handy way to categorize and filter data on various attributes, such as modifications, GO terms, taxonomy, etc.

Exercise 4 Results:

Step 1

Display Options: Protein Identification Probability Req Mods: No Filter Search: BOVIN

Probability Legend:
over 95%
80% to 94%
50% to 79%
20% to 49%
0% to 19%

Bio View:
343 Proteins in 334 Clusters
With 332 Filtered Out

#	Visible?	Starred?	Description	Accession Number	Molecular Weight	Protein Grouping Ambiguity	c1	c2	c3	c4	c5
1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Cluster of Serum albumin precursor (Allergen Bos d 6) (ALBU_BOVIN)	ALBU_BOVIN [5]	69 kDa	★	100%	100%	100%	100%	100%
1.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Serum albumin precursor (Allergen Bos d 6)	ALBU_BOVIN (+1)	69 kDa	★	100%	100%	100%	100%	100%
2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Cluster of Beta-lactoglobulin precursor (Beta-LG) (Allergen (LACB_BOVIN)	LACB_BOVIN [2]	20 kDa	★	100%	100%	100%	100%	100%
2.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Beta-lactoglobulin precursor (Beta-LG) (Allergen	LACB_BOVIN	20 kDa	★	100%	100%	100%	100%	100%
3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Superoxide dismutase [Cu-Zn] (EC 1.15.1.1)	SODC_BOVIN (+1)	16 kDa	★	100%	100%	100%	100%	100%
4	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Cluster of Cytochrome c (CYC_BOVIN)	CYC_BOVIN [2]	12 kDa	★	100%	100%	100%	100%	100%
4.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Cytochrome c	CYC_BOVIN (+1)	12 kDa	★	100%	100%	100%	100%	100%
5	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Creatine kinase, ubiquitous mitochondrial precurs	KCRU_BOVIN (+3)	47 kDa			93%			
6	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Transthyretin precursor (Prealbumin)	TTHY_BOVIN	16 kDa						95%
7	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Collagen alpha 3(IV) chain (Fragment)	CA34_BOVIN	48 kDa			93%			
8	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Gamma-aminobutyric-acid receptor beta-1 subunit p	GAB1_BOVIN (+9)	54 kDa			67%			
9	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Keratin, type I cytoskeletal 19 (Cytokeratin 19)	K1CS_BOVIN	44 kDa			66%			
10	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	NADH-ubiquinone oxidoreductase 39 kDa subunit, mi	NUEM_BOVIN	43 kDa			36%			
11	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Protein-tyrosine phosphatase-like N precursor (R-	PTPN_BOVIN	106 kDa			23%			

Final

Display Options: Protein Identification Probability Req Mods: No Filter Search:

Probability Legend:
over 95%
80% to 94%
50% to 79%
20% to 49%
0% to 19%

Bio View:
343 Proteins in 334 Clusters
With 1 Decoy and 327 Hidden

#	Visible?	Starred?	Description	Accession Number	Molecular Weight	Protein Grouping Ambiguity	c1	c2	c3	c4	c5
1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Trypsin precursor (EC 3.4.21.4)	TRYP_PIG	24 kDa		100%	100%	100%	100%	100%
2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	FAAH_PIG-DECOY	FAAH_PIG-DECOY	?			69%			
3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Muellerian inhibiting factor precursor (MIS) (Ant	MIS_PIG	62 kDa			65%			
4	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Cluster of Beta-lactoglobulin precursor (Beta-LG) (Allergen (LACB_BOVIN)	LACB_BOVIN [2]	20 kDa	★	100%	100%	100%	100%	100%
5	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Superoxide dismutase [Cu-Zn] (EC 1.15.1.1)	SODC_BOVIN (+1)	16 kDa	★	100%	100%	100%	100%	100%
6	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Cluster of Cytochrome c (CYC_BOVIN)	CYC_BOVIN [2]	12 kDa	★	100%	100%	100%	100%	100%
7	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Creatine kinase, ubiquitous mitochondrial precurs	KCRU_BOVIN (+3)	47 kDa			93%			
8	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Transthyretin precursor (Prealbumin)	TTHY_BOVIN	16 kDa						95%
9	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Collagen alpha 3(IV) chain (Fragment)	CA34_BOVIN	48 kDa			93%			
10	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Gamma-aminobutyric-acid receptor beta-1 subunit p	GAB1_BOVIN (+9)	54 kDa			67%			
11	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Keratin, type I cytoskeletal 19 (Cytokeratin 19)	K1CS_BOVIN	44 kDa			66%			
12	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	NADH-ubiquinone oxidoreductase 39 kDa subunit, mi	NUEM_BOVIN	43 kDa			36%			
13	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Protein-tyrosine phosphatase-like N precursor (R-	PTPN_BOVIN	106 kDa			23%			
14	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Cluster of Serum albumin precursor (Allergen Bos d 6) (ALBU_BOVIN)	ALBU_BOVIN [5]	69 kDa	★	100%	100%	100%	100%	100%

Exercise 5: Adding Biological Information

Adding Gene Ontology information to your Scaffold experiment can give insight into the biological significance of your results. In Scaffold 4, a new interface makes it easier to add GO terms from a variety of sources. To add new Go Databases, use Edit>Edit GO Term Options, and then go to the GO Annotation Databases tab. You will see a list of all of the GO sources you have indexed in Scaffold. You can add new sources by choosing New Database.

Scaffold automatically provides a web link to look up proteins in the NCBI database online. It also has built-in links to download the GOA human and the GOA All Proteomes databases. If you would like to download other species-specific database, you can use the Other Web Site option. Locate the file you wish to download on the web, right click and choose “Copy Link Location”. Paste that string into the Source box, give your new database a name, and click Add.

If you are on a Windows machine, you may get a message telling you that the program does not have privileges to write to its own parameters folder. You can choose a different location that is not in a protected area by clicking Choose by the Save to box.

If you have already downloaded a GOA file, or have created a custom one, you can also choose Other File and index a file on your computer.

Once you have your GO sources, click on one, then click Select, and that database becomes your current source. When you choose Experiment>Add GO Annotations (or Add NCBI Annotations) it will look up your proteins in that source.



Open tutorial_6. Because of time and memory constraints, we will begin with GO terms already applied.

Notice the GO terms which appear as colored dots in columns with various GO terms in their headers. The GO terms are grouped into the three ontologies, Biological Process, Cellular Component and Molecular Function.

Now got to the Quantify View. The GO terms are displayed in a series of Pie Charts, one for each ontology. Each wedge of the pie represents the number of proteins with that GO term. Note that a protein may have more than one GO term associated with it, so the number of proteins in the chart may be more than the total number of proteins in the experiment.

Double-click on the pie section labeled Developmental process.

Scaffold shows you the Samples View with the proteins filtered so that only proteins which have a dot in the developmental process column appear.

Return to the Quantify View. Double-click on the Metabolic process section.

You have now done a hierarchical filtering and are displaying only proteins which have both developmental process and metabolic process.

Return to the Quantify View and Double-click outside the Pie Chart.

The filtering is removed.

More complicated filtering can be done through the Advanced Protein filter, accessed through the button with the magnifying glass next to the Search box in the Samples View.

Return to the Quantify View and click the tab “Bar Charts (by Category)”.

This is a new display that compares the number of proteins with a certain GO term in each category.

Release Information

This document is applicable for Scaffold, Release 4.0 or greater, and is current until replaced.

Copyright © 2013. Proteome Software, Inc., All rights reserved.

The information contained herein is proprietary and confidential and is the exclusive property of Proteome Software, Inc.. It may not be copied, disclosed, used, distributed, modified, or reproduced, in whole or in part, without the express written permission of Proteome Software, Inc.

Limit of Liability Proteome Software, Inc.. has used their best effort in preparing this guide. Proteome Software, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this guide and specifically disclaims any implied warranties of merchantability or fitness for a particular purpose. Information in this document is subject to change without notice and does not represent a commitment on the part of Proteome Software, Inc. or any of its affiliates. The accuracy and completeness of the information contained herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every user.

The software described herein is furnished under a license agreement or a non-disclosure agreement. The software may be copied or used only in accordance with the terms of the agreement. It is against the law to copy the software on any medium except as specifically allowed in the license or the non-disclosure agreement.

Trademark - The name *Proteome Software*, the Proteome Software logo, *Scaffold*, *Scaffold Q+*, *Scaffold Q+S*, and the Scaffold, Scaffold Q+, and Scaffold Q+S logos are trademarks or registered trademarks of Proteome Software, Inc. All other products and company names mentioned herein may be trademarks or registered trademarks of their respective owners.

Customer Support - Customer support is available to organizations that purchase *Scaffold*, *Scaffold Q+* or Scaffold Q+S and that have an annual support agreement.

Contact Proteome Software at:

Proteome Software, Inc.
1340 SW Bertha Blvd, Suite 10
Portland, OR 97219
1-800-944-6027 (Toll Free)
1-503-245-4910 (Fax)
www.proteomesoftware.com

Document Version Number Scaffold 4.0-Tutorial-001

Document Release Date June 5, 2013