Getting Started with Scaffold 4
Proteome Software Users Group 2013
Tutorial Session

Module 3: Quantitation and Statistics

Getting Started with Scaffold 4

Module 3: Quantitation and Statistics

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:

Module 3. Quantitation and Statistics

Exercise 1. Spectral Counting

Open demo file Label-free.sf3. For convenience, hide the GO terms by selecting unchecking "Show GO Annotations" in the View menu. Set the thresholds to 99%, 2, 95%.

First, it is important to understand the meaning of the various display options for spectral counting. Scaffold 4 uses different terminology for these options than did previous versions. Now that peptides may be shared among different proteins, there are more options, and it is important to be very clear about the meanings of the terms.

Display Options for Spectral Counting

Unique peptides – peptides with different amino acid sequences, regardless of any modifications

Unique spectra – spectra that differ in amino acid sequence, charge state or modifications

Exclusive – associated with a single protein group

Total – associated with a protein group, whether or not it is shared with other protein groups

Scaffold 4 Name	Scaffold 3 Name	Description
Exclusive Unique Peptide Count	Number of Unique Peptides	Number of different amino acid sequences that are associated **only** with this protein
Total Unique Peptide Count		Number of different amino acid sequences that are associated with this protein
Exclusive Unique Spectrum Count	Number of Unique Spectra	Number of distinct spectra associated **only** with this protein
Total Unique Spectrum Count		Number of distinct spectra associated with this protein
Exclusive Spectrum Count	Number of Assigned Spectra	Number of spectra associated only with this protein
Total Spectrum Count	Unweighted Spectrum Count	Total number of spectra associated with this protein including those shared with other proteins
Quantitative Value (Spectrum Counting)	Quantitative Value	Normalized value based on Inclusive Spectrum Count

Select protein 1.16 HBA_ORNAN and go to the Proteins View.

Using the definitions above, and the Spectrum Table at the upper right, calculate the following values:

Exclusive Unique Peptide Count _____

Total Unique Peptide Count _____

Exclusive Unique Spectrum Count _____

Total Unique Spectrum Count _____

Exclusive Spectrum Count _____

Total Spectrum Count _____

Confirm your answers by going to the Samples View and changing the Display Options setting.

Exercise 1 Result:

Exclusive Unique Peptide Count   __1__

Total Unique Peptide Count   __3__

Exclusive Unique Spectrum Count   __2__

Total Unique Spectrum Count   __5__

Exclusive Spectrum Count   __4__

Total Spectrum Count   __11__

Explanation:

Valid	Weight	Sequence	Other Prote...	Charge	Modifications	P
☑	1.0	(K)AAGHGEEYGAEALER(L)		2		9
☑	1.0	(K)AAGHGEEYGAEALER(L)		3		9
☑	1.0	(K)AAGHGEEYGAEALER(L)		2		9
☑	1.0	(K)AAGHGEEYGAEALER(L)		2		9
☑	0.0	(S)ALSALSDLHAHK(L)	HBA2_LEMV...	2		9
☑	0.0	(S)ALSALSDLHAHK(L)	HBA2_LEMV...	3		9
☑	0.0	(K)LRVDPVNFK(L)	HBA2_LEMV...	2		1(
☑	0.0	(K)LRVDPVNFK(L)	HBA2_LEMV...	2		1(
☑	0.0	(K)LRVDPVNFK(L)	HBA2_LEMV...	2		1(
☑	0.0	(K)LRVDPVNFK(L)	HBA2_LEMV...	2		1(
☑	0.0	(K)LRVDPVNFK(L)	HBA2_LEMV...	2		1(

Only the first four spectra are exclusive to this protein (weight 1, and no Other Proteins). These all have the same amino acid sequence, so they constitute one exclusive unique peptide, but there are two charge states, so they give two exclusive unique spectra and four exclusive spectra in all.

Disregarding exclusivity, there are three different amino acid sequences, hence total unique peptide count is 3. Two of the peptides have both charge 2 and charge 3, so we have five for total unique spectra. The total spectrum count is just the number of spectra shown in the table, which is eleven.

Exercise 2: Quantitative Values

The last Display Option in the dropdown is Quantitative Value. By default, this is set to Normalized Total Spectra, but this can be changed. The options are:

> Total Spectra
> Weighted Spectra
> Average TIC
> Total TIC
> Top 3 TIC
> Average Precursor Intensity
> Total Precursor Intensity
> Top 3 Precursor Intensity
> emPAI
> NSAF

The Precursor Intensity options are only available if the intensities were calculated by the search engine and reported in the files loaded into Scaffold. Currently, precursor intensity quantitation is available from:

> Agilent Spectrum Mill
> Thermo Proteome Discoverer
> MaxQuant
> Mascot Distiller.

To select the Quantitative Value type you would like Scaffold to display, open the Quantitative Analysis Setup either by clicking on the ▣ icon or by choosing Experiment>Quantitative Analysis Setup. The settings are at the lower left.
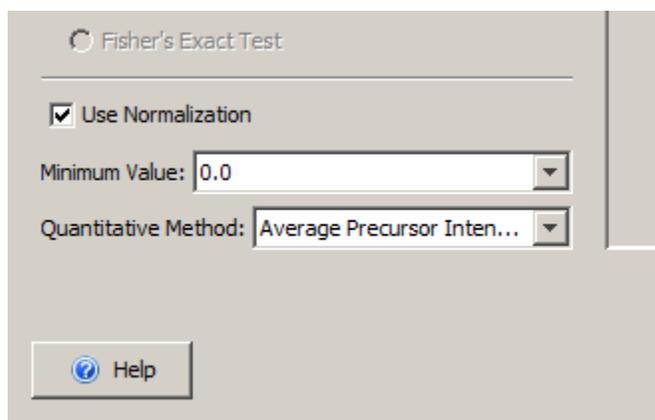
Select "Total TIC" from the dropdown in the Quantitative Analysis Setup. Uncheck "Use Normalization" and go to the Samples View. Select Quantitative Value from the Display Options dropdown.

Notice that the name of the value is shown in parentheses in the Display Options box. Also look at the values in the columns for BioSample1 and BioSample2.

Return to Quantitative Analysis Setup, and check "Use Normalization". The word Normalized is added to the Display Options box, and the values have been adjusted.

Now go back to the Setup once more, and fill in a Minimum Value. Since we are looking at Total TIC, we need to enter a large number, so select "Other" from the dropdown. Type "2.5E7" into the box and click "OK".

Notice that any values that were previously below 2.50E7 have now been replaced by this minimum value and they are shown in parentheses. Parentheses indicate that the displayed value has been adjusted by filling in the minimum value for a smaller or missing value.

The minimum value is designed to allow you to set a floor when calculating quantitative values to avoid identifying changes as significant when they may really only be a result of noise.

Experiment with other choices and settings.

Exercise 2 Result:

| Display Options: | Quantitative Value (Normalized Total TIC) ▼ | Req Mods: | No Filter ▼ | Search: | |

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

Bio View:
358 Proteins in 305 Clusters

							Control	Treatm...
#	Visible?	Starred?		Accession Number	Molecular Weight	Protein Grouping Ambiguity	BioSample 1	BioSample 2
1.1	☑	☆	Hemoglobin subunit alpha OS=Spalax leucodon ehrenbergi ... HBA_SPAEH		15 kDa	★	6.90E8	(2.50E7)
1.2	☑	☆	Hemoglobin subunit alpha OS=Mus musculus GN=Hba PE=1 ...HBA_MOUSE		15 kDa	★	6.02E8	(2.50E7)
1.3	☑	☆	Hemoglobin subunit alpha OS=Spermophilus citellus GN=HB... HBA_SPECI		15 kDa	★	5.01E8	(2.50E7)
1.4	☑	☆	Hemoglobin subunit alpha OS=Chalinolobus morio GN=HBA ... HBA_CHAMP		15 kDa	★	(2.50E7)	(2.50E7)
1.5	☑	☆	Hemoglobin subunit alpha OS=Semnopithecus entellus GN=... HBA_SEMEN		15 kDa	★	3.22E7	(2.50E7)
1.6	☑	☆	Hemoglobin subunit alpha OS=Balaenoptera acutorostrata ... HBA_BALAC		16 kDa	★	(2.50E7)	(2.50E7)
1.7	☑	☆	Hemoglobin subunit alpha OS=Spermophilus townsendii GN... HBA_SPETO		15 kDa	★	2.87E7	(2.50E7)
1.8	☑	☆	Hemoglobin subunit alpha OS=Ondatra zibethicus GN=HBA ... HBA_ONDZI		15 kDa	★	3.34E7	(2.50E7)
1.9	☑	☆	Hemoglobin subunit alpha OS=Suncus murinus GN=HBA PE=...HBA_SUNMU		15 kDa	★	2.81E7	(2.50E7)
1.10	☑	☆	Hemoglobin subunit alpha OS=Cynopterus sphinx GN=HBA ... HBA_CYNSP (+1)		15 kDa	★	3.17E7	(2.50E7)
1.11	☑	☆	Hemoglobin subunit alpha-1/2/3 OS=Macaca nemestrina P... HBA_MACNE		15 kDa	★	(2.50E7)	(2.50E7)
1.12	☑	☆	Hemoglobin subunit alpha OS=Procavia capensis habessinic...HBA_PROHA		16 kDa	★	(2.50E7)	(2.50E7)
1.13	☑	☆	Hemoglobin subunit alpha-1/2 OS=Oryctolagus cuniculus PE...HBA_RABIT		16 kDa	★	(2.50E7)	(2.50E7)
1.14	☑	☆	Hemoglobin subunit alpha OS=Macrotus californicus GN=HB... HBA_MACCA		15 kDa	★	(2.50E7)	(2.50E7)
1.15	☑	☆	Hemoglobin subunit alpha OS=Loxodonta africana GN=HBA ... HBA_LOXAF		16 kDa	★	(2.50E7)	(2.50E7)
1.16	☑	☆	Hemoglobin subunit alpha OS=Ornithorhynchus anatinus GN...HBA_ORNAN		15 kDa	★	(2.50E7)	(2.50E7)
1.17	☑	☆	Hemoglobin subunit alpha-1/2 OS=Tapirus terrestris PE=1 ... HBA_TAPTE		15 kDa	★	(2.50E7)	(2.50E7)
1.18	☑	☆	Hemoglobin subunit alpha-1/2 OS=Macaca speciosa PE=1 S... HBA_MACSP		15 kDa	★	2.99E7	(2.50E7)
1.19	☑	☆	Hemoglobin subunit alpha OS=Equus caballus przewalskii G... HBA_EQUPR		15 kDa	★	(2.50E7)	(2.50E7)
1.20	☑	☆	Hemoglobin subunit alpha OS=Cercocebus torquatus atys G...HBA_CERTO		15 kDa	★	2.85E7	(2.50E7)
1.21	☑	☆	Hemoglobin subunit alpha OS=Saguinus mystax GN=HBA PE...HBA_SAGMY		15 kDa	★	2.85E7	(2.50E7)
1.22	☑	☆	Hemoglobin subunit alpha-1 OS=Lemur variegatus PE=1 SV... HBA1_LEMVA (+1)		15 kDa	★	(2.50E7)	(2.50E7)
1.23	☑	☆	Hemoglobin subunit alpha OS=Lama vicugna GN=HBA PE=1 ...HBA_LAMVI		15 kDa	★	(2.50E7)	(2.50E7)
1.24	☑	☆	Hemoglobin subunit alpha OS=Dasypus novemcinctus GN=... HBA_DASNO		15 kDa	★	(2.50E7)	(2.50E7)
1.25	☑	☆	Hemoglobin subunit alpha OS=Saguinus fuscicollis GN=HBA ... HBA_SAGFU		15 kDa	★	2.79E7	(2.50E7)
1.26	☑	☆	Hemoglobin subunit alpha-1 OS=Tadarida brasiliensis PE=1 ...HBA1_TADBR		15 kDa	★	2.73E7	(2.50E7)
1.27	☑	☆	Hemoglobin subunit alpha-1 OS=Catharacta maccormicki P... HBA1_CATMA (+1)		15 kDa	★	(2.50E7)	(2.50E7)
1.28	☑	☆	Hemoglobin subunit alpha OS=Eulemur fulvus fulvus GN=HB... HBA_EULFU		15 kDa	★	(2.50E7)	(2.50E7)
1.29	☑	☆	Hemoglobin subunit alpha-B OS=Galago crassicaudatus GN... HBA2_GALCR		16 kDa	★	(2.50E7)	(2.50E7)
⊟ 2	☑		**Cluster of Hemoglobin subunit beta-1 OS=Mus musculus GN=... HBB1_MOUSE [24]**		16 kDa	★	6.78E8	(2.50E7)
2.1	☑	☆	Hemoglobin subunit beta-1 OS=Mus musculus GN=Hbb-b1 P...HBB1_MOUSE		16 kDa	★	4.03E8	(2.50E7)
2.2	☑	★	Hemoglobin subunit delta OS=Tarsius syrichta GN=HBD PE= HBD_TARSY		16 kDa	★	4.18E8	(2.50E7)

Exercise 3: Fold Change

Open demo file tutorial_6. Set filters to 99%, 2, 95%. Hide GO terms by unchecking "Show GO Annotations" in the View menu. Open the Quantitative Analysis Setup .

a) Fold Change by Sample

Remove Samples Int-2 and Un-2 from the Selected Samples list. Now that there are only two samples, the Fold Change by Sample option should be available for selection. Click the radio button and click Apply.

In the Samples View, you should see a new column called Fold Change by Sample. The two samples being compared are highlighted in color, with the purple as the numerator and peach as the denominator. If there are multiple MS Samples in a BioSample, the values are summed and the comparison is done at the BioSample level.

#	Visible?	Starred?	Bio View: Identified Proteins (11)		Accession Number	Molecular Weight	Protein Grouping Ambiguity	Fold Change by Sample (Fold Change)	Int-1	Int-2	Un-1	Un-2
1	☑	☆	Keratin, type I cytoskeletal 10 (C...	gi|547749 ...		60 kDa		2.1	17	16	8	(1)
2	☑	☆	Keratin, type II cytoskeletal 1 (Cy...	gi|1346343...		66 kDa	★	4.5	18	13	4	(5)
3	☑	☆	Chain E, Leech-Derived Tryptase I...	gi|3318722...		23 kDa		1.6	13	9	8	4
4	☑	☆	Keratin, type II cytoskeletal 2 epi...	gi|547754 ...		66 kDa	★	INF	15	6	(0)	(1)
5	☑	☆	Keratin, type I cytoskeletal 9 (Cyt...	gi|8117517...		62 kDa		INF	(6)	3	(0)	(1)
6	☑	☆	RecName: Full=Actin-2, muscle-s...	gi|1168320...		42 kDa		INF	2	(1)	(0)	(4)
7	☑	☆	PREDICTED: similar to protein tyr...	gi|50728376		622 kDa		0.0	(0)	2	(4)	(0)
8	☑	☆	Chain A, Crystal Structure Of Hum...	gi|2894874...		76 kDa		INF	(0)	(3)	(0)	(0)
9	☑	☆	Chain A, Crystal Structure Of Hum...	gi|5655458...		18 kDa		INF	(0)	(3)	(0)	(0)
10	☑	☆	Lysozyme C precursor (1,4-beta-...	gi|2497776...		17 kDa		INF	(0)	(2)	(0)	(0)
11	☑	☆	myosin VI [Homo sapiens].	gi|5058254...		149 kDa		INF	(2)	(0)	(0)	(0)

Probability Legend:
- over 95% (green)
- 80% to 94% (yellow)
- 50% to 79% (orange)
- 20% to 49% (red)
- 0% to 19% (white)

Fold changes that are more than twofold, i.e. the ratio is >2 or < 0.5 are colored in green. INF indicates that the denominator was missing. These can be eliminated by setting a minimum value through Quantitative Analysis Setup.

b) Fold Change by Category

Return to the Quantitative Analysis Setup. Add samples Int-2 and Un-2 to the Selected Samples list. Select Fold Change by Category, and set a Minimum Value of 0.5. Click Apply.

#	Visible?	Starred?	Bio View: Identified Proteins (11)	Accession Number	Molecular Weight	Protein Grouping Ambiguity	Fold Change by Category (Fold Change)	Int		Un	
								Int-1	Int-2	Un-1	Un-2
1	☑	☆	Keratin, type I cytoskeletal 10 (C...	gi|547749 ...	60 kDa		0.3	17	16	8	(2)
2	☑	☆	Keratin, type II cytoskeletal 1 (Cy...	gi|1346343...	66 kDa	★	0.3	18	13	4	(6)
3	☑	☆	Chain E, Leech-Derived Tryptase I...	gi|3318722...	23 kDa		0.6	13	9	8	(5)
4	☑	☆	Keratin, type II cytoskeletal 2 epi...	gi|547754 ...	66 kDa	★	0.1	15	6	(1)	(2)
5	☑	☆	Keratin, type I cytoskeletal 9 (Cyt...	gi|8117517...	62 kDa		0.3	(7)	3	(1)	(2)
6	☑	☆	RecName: Full=Actin-2, muscle-s...	gi|1168320...	42 kDa		1.4	(3)	(2)	(1)	(5)
7	☑	☆	PREDICTED: similar to protein tyr...	gi|50728376	622 kDa		1.8	(1)	2	(5)	(1)
8	☑	☆	Chain A, Crystal Structure Of Hum...	gi|2894874...	76 kDa		0.4	(1)	(4)	(1)	(1)
9	☑	☆	Chain A, Crystal Structure Of Hum...	gi|5655458...	18 kDa		0.4	(1)	(4)	(1)	(1)
10	☑	☆	Lysozyme C precursor (1,4-beta-...	gi|2497776...	17 kDa		0.6	(1)	(3)	(1)	(1)
11	☑	☆	myosin VI [Homo sapiens].	gi|5058254...	149 kDa		0.6	(3)	(1)	(1)	(1)

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

Now the ratio is calculated at the category level. Note, however, that for the category comparison, Scaffold takes the AVERAGE of the values of all MS Samples in the category, and compares the averages. The reason is that categories may have different numbers of samples within them, and this should not affect the ratio.

Exercise 4: Statistical Testing

Scaffold provides a variety of statistical tests designed to assess the significance of any observed differences in expression. Different tests are available depending on the number of samples in your dataset and how they are organized.

T-Test compares two categories of samples and tells whether there is a significant difference in the protein levels between them. It is only available if your Selected Samples list contains exactly two categories, and each of them contains at least two BioSamples, because it is not reliable for smaller datasets.

Fisher's Exact Test is like a T-Test but is accurate for small numbers of samples. As a result, Scaffold lets you do a Fisher's Exact Test if you have exactly two categories in your Selected Samples List, even if you only have one BioSample in each.

ANOVA requires three or more categories, with at least two BioSamples in each. It tells whether there is significant variation among the categories, but does not tell you which of the samples is different. To see the differences, you can check the Quantitative Value Bar Chart in the Quantify View.

Coefficient of Variance (or Variation) is useful if you have three or more BioSamples, but do not have replicates within them. You would like to have some idea which proteins exhibit meaningful differences in expression among the samples, but you cannot do a formal test of statistical significance. A large CV suggests that at least one of the samples is showing differential expression as compared to another.

> **Open Demo file tutorial_6. Set thresholds to 99%, 2, 95%. Open the Quantitative Analysis Setup .**

This file contains four BioSamples, organized into two categories. Each BioSample consists of two MS Samples.

> **a) Apply the T-Test with Normalization.**
> Which protein demonstrates significantly different levels in the two categories?

> **b) Go to Load Data, select the Un-2 tab, right click and edit the BioSample. Change the Category name to "New". Open the Quantitative Analysis Setup. Now that there are three samples, the T-Test and Fisher's Exact Test are unavailable, but the ANOVA is allowed. Apply Anova with Normalization.**
> 1) Which protein shows significant variation among the categories?
> 2) Which category is most responsible for this variation?

Exercise 2 Result:

a)gi|1346343(+1)

b) 1)gi|1168320(+106) shows significant variation

2)Un-2 is significantly lower for this protein
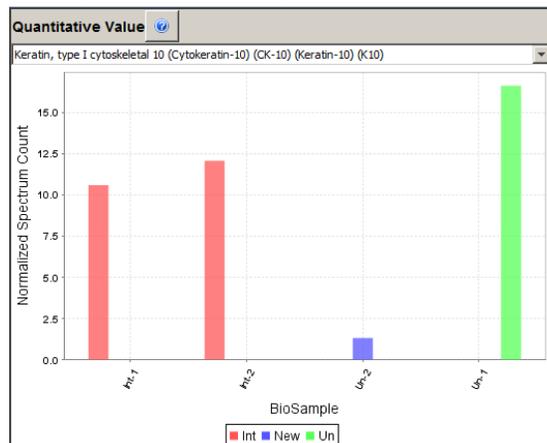
Release Information

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

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