POSTER NOTE

New Method for Label-Free Quantification in the Proteome Discoverer Framework

David M Horn,¹ Torsten Ueckert,² Kai Fritzemeier,² Katja Tham,² Carmen Paschke,² Frank Berg,² Hans Pfaff,² Xiaoyue Jiang,¹ Shijun Li,¹ and Dani Lopez-Ferrer,¹ ¹Thermo Fisher Scientific, San Jose, CA, USA; ²Thermo Fisher Scientific, Bremen, Germany

ABSTRACT

A new label-free quantification method based on the Minora algorithm is presented and compared to pre-existing label free quantification methods in the Thermo Scientific[™] Proteome Discoverer[™] software framework. The results of the new algorithm were significantly more accurate across a wide dynamic range compared to spectral counting and "Top N" quantification. The new algorithm was also run on a subset of the Akhilesh Pandey human proteome dataset to identify proteins specific to specific tissue types.

INTRODUCTION

Proteome Discoverer software is a node-based workflow engine and study management platform for analysis of mass spectrometry-based proteomics datasets. The latest released version 2.1 fully supports isotopically-labeled quantitative workflows, such as TMTTM reporter ion-based quantification and SILAC precursor ion quantification, but the supported label-free quantification methods are significantly less sophisticated. Currently, spectral counting is possible but not recommended when quantitative accuracy is required. The only supported label-free quantification workflow produces an average abundance of the top "N" most abundant peptides and this has been shown to be accurate for even highly complex datasets. However, "Top N" quantification results cannot be used to create ratios, scaled abundance values, or to be used as replicates to generate standard errors. Here we present a new workflow for untargeted label-free quantification using a new feature detection approach that provides the full suite of quantitative compared to the two aforementioned label-free quantification. The workflow will be compared to the two for isotopically-labeled quantification. The workflow will be compared to the two aforementioned label-free quantification workflow savailable within Proteome Discoverer 2.1 software.

MATERIALS AND METHODS

A standard dataset of Arabidopsis proteasome proteins spiked into a background of *E. coli* proteins (PXD003002) was downloaded from the PRoteomics IDEntifications (PRIDE) repository. This dataset was originally used to evaluate a spectral counting algorithm and is described in reference 1. The Pandey human proteome dataset² was also downloaded from PRIDE and a portions of the dataset to demonstrate untargeted label free quantification of data with a multi-dimensional separation.

For quantification using spectral counts, each of the datasets with the different levels of *Arabidopsis* proteasome proteins was run separately in batch mode using a standard Sequest™ HT-Percolator workflow and a basic consensus workflow. Subsequently, all Processing results were reprocessed using a single Consensus workflow with the "Merge Mode" parameter in the MSF files node set to Do Not Merge. With this setting, the number of unique peptides and PSMs for each of the datasets will be represented as a separate column. The SequestHT search was performed against the entire *Arabidopsis thaliana* and *Escherichia coli* databases. The table with PSM values for each sample was exported to Microsoft Excel format and ratios were calculated manually.

For the "Top N" quantification workflow, a Precursor Ion Area Detector node was incorporated in the Processing workflow used for spectral counting above. The default "CWE_Comprehensive_Enhanced_Annotation_Quan" template was used for the Consensus workflow. In the Peptide and Protein Quantifier node, the "Top N Peptides Used for Quantification" parameter was set to 3. Like for spectral counting, the table with the reported Top N protein abundances was exported to Excel and ratios were calculated manually.

New Method for Feature Detection

The new feature detection algorithm is an extension of the Minora algorithm, which had already been used for precursor ion quantification since the release of Proteome Discoverer 1.2 software. Minora had always detected all isotopic peaks in a given data set, but up to now only those LC/MS peaks associated with peptide spectral matches (PSMs), and their associated isotopic forms in the case of SILAC, were used for quantification. In this prerelease version of Proteome Discoverer 2.2 software, the Minora algorithm has been modified to detect and quantify isotopic clusters regardless of whether or not they are associated with a PSM.

A typical Processing workflow for Minora feature detection is shown in **Figure 1**. The new label-free quantification workflow can be invoked by simply attaching the "Minora Feature Detector" to the Spectrum Files node. This new feature detector will also be used for the isotopically-labeled precursor quantification method such as SILAC.

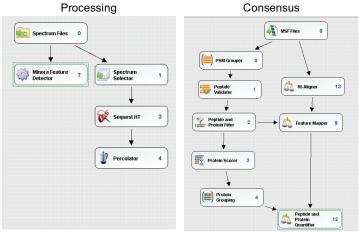


Figure 1. Typical Processing and Consensus workflows for untargeted label-free quantification. The Minora Feature Detector, Rt-Aligner and Feature Mapper are new nodes created for the untargeted label-free quantification workflow. The Minora Feature Detector will also replace the old Precursor lons Quantifier node used for SILAC and other precursor ion quantification workflows.



A typical Consensus workflow for label free quantification is also shown in **Figure 1**. There are two new nodes added to this workflow that perform retention time alignment and feature mapping. The feature mapper groups features detected from the Processing runs into "Consensus Features" that are mapped and quantified across all raw files and performs gap filling to find features that were not initially detected in the processing workflows. The Peptide and Protein Quantifier node works as previously, with improvements to scaling and normalization that benefit all quantification workflows.

There are three new tabs for feature detection results in the consensus report: Consensus Features, LCMS Features, and LCMS Peaks. The LCMS features are isotopic clusters grouped together for a given raw dataset and consist of multiple LCMS Peaks. Ultimately, the release may not include the LCMS Peaks list given that as much as 10°s of millions peaks could be detected in complex datasets. The consensus features link directly to the associated peptide group as well as the list of LCMS features detected from each data files (**Figure 2**). Also, when a consensus feature is selected, the traces for each of the features are shown in the chromatogram traces view. When a single LCMS Feature is selected, the chromatographic profile for only that individual feature is displayed.

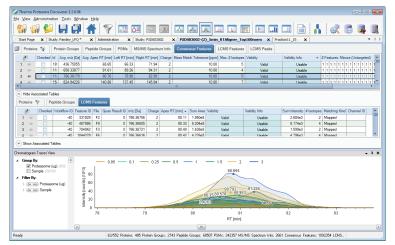


Figure 2. The Consensus Features table is linked to the collection of LCMS Features from each raw file. The chromatographic profiles for each LCMS Feature are shown in the Chromatogram Traces View at the bottom.

Like the other quantification workflows in Proteome Discoverer 2.1 software, the peptide group abundances from the new label-free quantification method are calculated as the sum of the abundances of the individual PSMs for a given study factor that pass a quality threshold. The protein abundance is calculated as the sum of the peptide group abundances associated with that protein.

RESULTS AND DISCUSSION

The database searches produced a total of 55 *Arabidopsis* proteins and 423 *E. coli* proteins. This is less than would be expected given the relatively high protein concentration and long gradient length, but the chromatography used for these data analyses was suboptimal with peaks up to 5 minutes wide (**Figure 3**). Also, as the amount of the *Arabidopsis* proteins added to the sample increased past 1 µg, the peptides from these proteasome proteins dominate the chromatogram and thus the number of *E. coli* proteins decreases dramatically with increasing *Arabidopsis* protein concentration (data not shown).

0.05 μg proteasomes 1 μg *E coli*

74.56 75.44 78.82 87.21 89.69 96.08 101.63 110.18 111.37 113.72 120.25 122.23 130.84 133.26 140.42 146.94 151.05 155.67 157.59

0.5 μg proteasomes 1 μg *E coli*

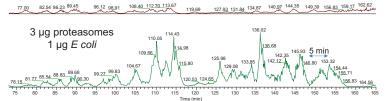


Figure 3. Base peak chromatograms for three of the LC/MS runs, each scaled to 2e7 intensity. The dataset at the bottom is dominated by *Arabidopsis* peptides, leading to significant suppression the *E coli* peptides. Also, the typical chromatographic peak in this chromatogram can be up to 5 minutes wide, also decreasing the number of peptides and proteins that can be identified.

The spectral counts-based quantification results correctly indicate the direction of expression for the *Arabidopsis* proteasome proteins, but the ratios are inaccurate for the more extreme ratios. The response is also relatively non-linear, with the average ratio for the 0.1 µg/1 µg samples showing a lower value than the 0.5 µg/1 µg samples and the 3 µg/1 µg ratio measuring lower than the 2 µg/1 µg ratio. These results are not a surprise given that it is widely known that this type of spectral counting is not expected to produce accurate quantification results. Normalized spectral counting algorithms are a significant improvement over the basic spectral counting method shown here and reference 1 from which these data were obtained describes a such a method. Implementation of such a method using emPAI values is planned for the individual study factors is being considered for a future Proteome Discoverer software release. However, all spectral counting-based quantification methods usually provide poorer sensitivity and dynamic range than other quantitative techniques due to the requirement for multiple PSMs for any given protein. As can be seen in Table 1, less than half of the *Arabidopsis* proteins could be quantified

The "Top N" protein quantification results are shown in the second set of columns in Table 1. The accuracy of the ratios is noticeably improved compared to spectral counting, producing a response that is closer to linear. However, there are fewer quantified proteins in the "Top N" method than for spectral counting, primarily due to the requirement that the same three peptides need to be identified across all of the datasets. This is in effect even more stringent than the spectral counting method above and as a result even fewer proteins are quantified across the samples. Also, while the accuracy of the ratios is improved, the precision of the measurements are not much improved over spectral counting.

For feature detection-based quantification, the calculated ratios were significantly closer to the theoretically expected values at the lowest *Arabidopsis* concentrations. The precision of the ratios was also significantly improved in almost all cases for the feature detection results. The use of feature mapping led to a significantly increased number of quantified proteins given that only a single PSM is required for a given peptide across all raw files. The accuracy and precision of this method also benefits from the use normalization based on the *E. coli* proteins, which are known to be equally abundant across all samples.

A screen shot of the *Arabidopsis* protein identification results with untargeted label-free quantification is shown in **Figure 4**. The ratios and the scaled abundances for the identified proteins and peptide groups are color coded based on the level of expression. Scaled abundances were originally introduced in Proteome Discoverer 2.1 software primarily for the TMT quantification workflow and are now available in the preliminary version of Proteome Discoverer 2.2 software for feature detection-based label free quantification. The samples can be sorted by scaled abundance for any given sample type, as seen in **Figure 4** for the highlighted 0.05 sample group. It can be easily seen that each of the proteasome proteins exhibit a similar trend by simply looking at the pattern of blue and red boxes. Also, since the scaled abundances exhibit the same profile as the ratios, the need for the calculation of ratios is somewhat obviated.

					<u>.</u>																				
1.00	X Study: F	unders_UPQ	X Administration	K Study: PID000002 K	PXD000002-{2	0_Smin_RTABs	per_Tep100r	arm X Predical1_	5 X																
	E Prot	ein Groupe	Peptide Groups PSR	4s MS/MS Spectrum Info	Consensus Fe	alures LCMS	3 Feetures	CH3 Peaks																	
												undances (Grouped						8						
0	lecked Mash	v Accession	Description		Coverage	#Pepides #	PS - #Uniqu	Peptides # Protein Gro	os FAA	Mr/ MDel	calc pl #1	azor Peptides	÷								Bundance F	nice			
	_								-			11	-	8	10	5	-	2	- 14		_				-
• L		001146		toe-G-A (CG-Ambidges theler toe-2-A (CG-Ambidges theler		23		1	1 246		5.05	11	5.4	127	30.5	55.5	100.5	123.7	245.2	2014	6.054	6 126	0.300	0.563	+
		Q8L4A7		top-2-8 (05-Arabidopes theler top-2-8 (05-Arabidopes theler	205		1803	2	1 235	25.7		13	42	100	31.2	40.5	102.6	1081	196.6	100	0.003	6 107	0.105	0.548	÷
- 1		P34066		type-1-A I25+Arabidopais thater	215		1710	8	1 278	30.5		13			25.1	29.7	01.8	1047	104.0	100		6.117	0.307	0.406	÷
		001149		r ype-1-x (US+Arabidopais thatar	955		1697	4	1 200	25.9		15	30		201	42.2	01.5	1047	1215	1001	000	6117	0.307	0.406	٠
		G42124		type-5-B 125-Jubidopia thatar			1625	2	1 227	26.0		0				27.6	79.7	124.9	126.5	111		1.000	0.555	0.471	÷
-		081147-1		type 6-8 JOS=Arabidopsis Ihalian	203		1538	12	1 245	27.5			43	111	26.4	45.2	02.8	126.2	112.2		2.040	6 129	0.257	0.495	÷
		023715		type 3105-Anabidopeis thalianal	6N		1552	21	1 249	27.4		0	56	55	265	12.2	05.2	122.6	201.5	206.1		6111	0.304	0.591	Ŧ
		004533		hoe-3-6 105-Anbidoosis thelian	285		1400	12	1 122	22.0		11	55	12.0	29.7	05	47.7	122.4		24.4	1002	6125	0.304	0.539	t
	n v	9704.89		toe-HCS-Antidopis trailanal	6355		1267	12	1 246	27.6			2.9	10	25.5	12.5	124.5	125.2	1917	222	5.007	5.008	0.244	0.512	
		024616	Protessome subunit alpha	type-7-8 105-Arabidopsis thatas	825	25	1358	6	1 250	27.3	8.63	14	3.0	87	28.4	83.3	97.1	1380	100.5	280.1	6.082	6 100	0.282	0.884	t
0		P30186-1		type 7 A 108-Arabidopsis thatian	825		1254	6	1 250	27.8	7.97	0	41	56	26.7	51.5	195.7	138.4	126.7	334.3	6.055	6.055	0.245	0.474	
9		P42742-1	protresome subunit beta	pe-1 (05+Arabidopsis theliana)	265	19	1246	19	1 223	24.6	7.40	0	\$7	169	25.6	51.0	85.5	120.5	296.2	200.7	6.065	6125	0.296	0.590	T
		061148	Protessome subunit alpha	type-1-A (05=Arabidopsis thallar	80%	19	1244	19	1 250	27.5		0			28.3	65.3	95.3	120.7		270.5	6.069	6 300	0.297	0.534	
40		081148 023712		type-1-8 (35=Arabidopsis thallar type-1-8 (35=Arabidopsis thallar	10% 74%		1244	19	1 250		5.07	0	5.6	55	26.5	55.3 20.1	96.3 143.5	1207	154.6	299.3	E.010 8.610	6.010	6.297	0.458	t
40 40 40			Protessome subunit sight			16					5.07				263 61 32.6		143.5	120.7	214.0 154.6 540.3	2953 2863 2863	£000 £000 £000		6.297 6.297		
to to to the Assess in Group	ieled Tables	023712 022714	Protessome suburit sight Protessome suburit beta PSNs M5/M8 Spectra	type 18 (35-Anthologuis thele type 2 A (35-Anthologuis thele) "	74% 50%	14	1236	2	1 277	30.4 22.5	5.07	0	54	84	81 326	20.1 63.8	143.5	145.0	540.3	2993 2993 2993	£.010 £.610 £.050	6.010	6.057	0.458	
er er in Grou	eled Tables	O23712 O22714	Protessome suburit sight Protessome suburit beta PSNs M5/M8 Spectra	nye 18 (35-Antidosis Hale ye 2A (35-Antidosis Hale an Info Consense Features	74% 58%	16	1236	2 6 mility P x Qnality quali	1 277	30.4 22.5	5.07	0	54	84	81 326	20.1 63.8	143.5	145.0	540.3	201	E 010 E 010 E 010	6.010	6.057	0.458	
e Aaeoo	ieled Tables	023712 022714 n Occupa n Ukage Nat Ukad	Protessome subunit sight Protessome subunit bett PBMs M&M&MS Spectry Confidence & Conse	rige 14 (35-Arabidosis Halen type 2 A (35-Arabidosis Halen ministric Conservus Features neut Features, Sequence 0 SQV/TGHX1MI	245 585	16	1236	2 6 nity P x Quality quali 1.22e-15	1 277	30.4 22.5	5.07	0 0 0 0 0 0 0 0 0 0	B2 58	84	81 326	20.1 53.3 Decempos	143.5 139.7 Theo: Mile (Day 3555 77455	145.0	9401 e Ratios	2003	600	6 010	6.297	0.458	
er er er er Annord er er er er er er er er er er er er er	ieled Tables	O23712 O22714	Protessome subunit sight Protessome subunit beta PSMs M&MS/HS Spectry Confidence # Conse ©	nge 18 (35-Arabidopis hele type 2 A (35-Arabidopis hele) * * * * * * * * *	NAN SES PLOVARAMEDI THEFTIK	16	1236	2 6 mility P x Qnality quali	1 277 1 204	30.4 22.5 min Groups 1	5.07 6.39 # Prosies	0 0	22 58 Protein Acco	84	81 326	20.1 63.8	143.5 109.7 Treo: Mile [Do	145.0	540.3	2303 2353 296.3 0.254 0.254	0.000 0.000 0.000 0.436 0.436	6.010	6.057	0.458	
to to to to Association to to to to to to to to to to to to to	aled Tables	C23712 C22714 C22714 C22714	Protessome suburit sight Protessome suburit beta PBMs M5595 Spectra Confidence # Conse 0 0	rge-18335-Asbidges Belle type 2.4305-Asbidges Belle minist Conservus Features must Features Segurate 0 SOVYTGHUDSSV	NAN SEN PLOVARMMON THEFTEK ARE DK	16	1236	2 6 122=15 1.75=12	1 277 1 204 0 0	30.4 22.5 min Groups 1 1	5.07 6.39 # Possies 1 1	0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	22 58 Protein Acco	84	81 326	20.1 53.8 Decreges 0 2	143.5 108.7 2605. MH+ (26 3555.77453 2206.20254	145.0	9401 e Ratios		0.426	E 010 E 134 1.320	0.007	0.458 0.491 2.854	
to co le Association of Oreco to co co co co co co co co co co co co co	aled Tables re Profile Decked Out	CI23712 CI23714 CI23714 CI23714 CI23714 CI23714 CI23714 CI23712 CI23712 CI23712 CI23712 CI23712 CI23712 CI23712 CI23712 CI23712 CI23714	Protessome subunit sight Protessome subunit beta PSMs M&MS/HS Spectry Confidence # Conse ©	type-14.325-Aubidopsis Balan type-24.325-Aubidopsis Balan minfe Conservus Features must Features Sequence 0 SQV/TGALDQSSV 2 KVTGALDQSSV 3 JX554/GMXCQSCV	NIN SIS PLONAMMON THLIPPTK AFLDK SERD	16	1236	2 6 122+15 123+15 1736+12	1 277 1 204 0 0 0 0	30.4 22.5 min Groups 1 1 2	5.07 6.39 # Possies 1 1	0 0 0 1 PSMs Master 40 00114 137 00114 32 00116	22 58 Protein Acco	84	81 326	20.1 53.8 Descripts 0 2 2 2	143.5 108.7 740. MH+ (24) 3595.77459 2266.29554 1361.00590	145.0	9401 e Farios 0.155 0.254	0.236	0.000 0.426 0.426	6.010 © 134 1.220 1.422	0.005 0.297 1.303 0.434	0.458 0.401 0.814 2.954	
to co te Association in Oreco te co te co te co	aled Tables re Profile Decked Out	CI23712 CI23714 CI23714 CI23714 CI23714 CI23714 CI23714 CI23712 CI23712 CI23712 CI23712 CI23712 CI23712 CI23712 CI23712 CI23712 CI23714	Potessome suburit sight Potessome suburit beta PBMs M5505 Speech Confidence # Conse 0 0	type 14.305-Anthoness Roller type 24.305-Anthoness Roller model Features, Sequence 0 50/VTGHAMMO 2 KNYRELDOSSV 3 ATAGAMCODA 2 ALTERIBERT, A	24% 58% TULTITK AFLDX SER0 888	16	1236	2 6 1.22e-15 1.73e-12 7.65e-10 7.65e-10	5 277 1 204 0 0 0 0 0 0	30.4 22.5 1 1 2 1	5.07 6.39 # Possies 1 1	0 0 0 1 PSMs Master 40 00114 137 00114 32 00114 313 00116	22 58 Protein Acco	84	81 326	20.1 53.8 0 2 2 1	143.5 109.7 7460 Mile [26 3555 77455 2266 29554 1961 00590 2957 00392	145.0	9401 e Farios 0.155 0.254	0.236	0.650 0.636 0.636 0.636	6010 0134 1.350 1.452 1.342	6.007 0.297 1.325 6.034	0.438 0.491 2.814 2.915 3.943	
	aled Tables re Profile Decked Out	CO23712 CO23714 CO23714 CO23714 CO23714 CO23714 CO23714 CO23714 CO23712 CO23712 CO23712 CO23712 CO23712 CO23712 CO23712 CO23712 CO23714 CO2371	Potessore suburit split Potessore suburit bets PBMs M5555 Spectry Confidence # Conse 0 0 0	type 16 (25)-Lobitopis holes (see 2.4,05)-Valideois holes (see 2.4,05)-Valideois holes (see 16) (see 1	24% 56% PLONIAMIMON THERPTIK AFEDXX SER0 SER TK KIMMPLOVIM	18 17	1236	2 6 1226-15 1256-15 1256-15 7.656-10 7.656-10 7.656-10 7.656-10 7.656-10 7.656-10	5 277 1 204 0 0 0 0 0 0	30.4 22.5 1 1 2 1	5.07 6.39 # Possies 1 1		02 54 Protein Acco	E4 947	81 326	20.1 50.8 0 2 2 2 1 0	143.5 109.7 7460 Mile [26 3595 77450 2266 29554 1961 00590 2967 00392 1941 97638	145.0	1401 e Ratios 0.158 0.394 0.123 0.394	0.236	0 439 0 439 0 466 0 438 0 534	6.010 6.134 1.320 1.462 1.182 1.182	6.007 0.207 1.303 6.034 1.343 1.542	0.438 0.491 2.814 2.152 3.843 5.844	
to to to the Assess in Group	ated Tables re Rocki Decket Custon	023712 022714 e Groups n Ukage Nat Used Used Used Used	Potessome suburit sjøle Prøtessome suburit bela PBMs M&MS Spectra Confidence / # Conse • • • •	rype 14 (201-Joshidopsis Halles pe 24,005-Velidedesis Halles pris http://conserves/Features (550/YT04-Velidedesis 250/YT04-Velidedesis 24,075568-00020 24,075568-00000 24,075568-00000 24,075568-000000000 2	24% 56% PLONIAMIMON THERPTIK AFEDXX SER0 SER TK KIMMPLOVIM	18 17	1236	2 6 122+5 1.156-12 1.26+10 7.26+10 7.26+10 7.26+10 7.26+10 7.26+10 7.26+10 7.26+10	1 277 1 204 0 0 0 0 0 0 0 0	30.4 22.5 1 1 1 2 1 1 1	5.07 6.39 # Possies 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	02 54 Protein Acco	E4 947	81 326	20.1 50.8 0 2 2 1 0 0 0 0	143.5 138.7 266.7 2555.77455 2565.25554 1361.00520 2967.00332 1941.97438 1956.54250	145.0	1401 e Ratios 0.158 0.394 0.123 0.394	0.236	0 439 0 439 0 466 0 438 0 534	6.010 6.134 1.320 1.462 1.182 1.182	6.007 0.207 1.303 6.034 1.343 1.542	0.438 0.491 2.814 2.152 3.843 5.844	
e Aaeoo	ated Tables re Rocki Decket Custon	C C23113 C C23114 C C23214 C C	Pritescore solurit spile Pritescore solurit bett PBMs M5.NS Spectra Confidence # Conse 0 0 0 0 0	rge-16 (201-katologis) holizon 2 / 201-katologis) holizon 2 / 201-katologis holizon 2 / 201-katologis holizon 2 / 201-katologis holizon 2 / 201-katologis / 20	TAX SIS SIS NELEXX SERE SER TX MULTINE SERE TX MULTINE MULTINE MULTINE MULTINE	18 17	1236	2 6 1226-15 1366-10 7.66-10 2.166-00 2.166-00 2.166-00 2.166-00 2.166-00 2.166-00 2.166-00	1 277 1 204 0 0 0 0 0 0 0 0	30.4 22.5 1 1 1 2 1 1 1 1	5.07 6.39 # Possies 1 1	6 PSMs Master 46 00114 137 00114 32 00116 33 00116 34 00116 39 00116 130 00116 130 00116 42 00116	02 54 Protein Acco	E4 947	81 326	201 508 0 2 2 1 0 0 0 1 1 1	143.5 109.7 100.7	145.0	9401 • Ratios 0.155 0.229 0.153 0.153 0.153 0.153	0.236 0.281 0.287 0.288 0.288 0.288	0.436 0.436 0.436 0.334 0.486 0.334 0.480 0.480	1.105 0.134 1.105 1.105 0.827 1.105 0.827 1.106 0.869	0.297 0.297 1.325 0.404 1.543 1.642 1.226 0.477	0.458 0.491 2.814 2.152 3.383 5.584 2.399 2.322 2.382	
e Association	ated Tables re Rocki Decket Custon	CI23713 CI22714 CI22714 CI22714 n Using NatUsed Used Used Used Used Used	Pritesore advit ajže Pritesore advit bes PSMs M3545 Spectri Conferon / Conse 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	spe-1 8 101-Laborations Malanias A 101-Laborations Malanias 2 A 101-Laborations Malanias Malanias A 101-Laboration Malania Researce Seguration (Conserve) (Conserv	NIS SES DUTIENTS AFLENS SER TX AFLEN AFLEN AFLEX AFLEX AFLEX AFLEX A	18 17 VEENOPLLY	1236	2 6 122+5 122+5 126+5 126+5 156+10 226+00 226+00 226+00 246+00 246+00 226+00	1 277 1 204 0 0 0 0 0 0 0 0	30.4 22.5 1 1 1 2 1 1 1 1	5.07 6.39 # Possies 1 1		02 54 Protein Acco	E4 947	81 326	201 638 0 2 2 1 0 0 1 1 1	1435 1097 1097 1097 206 2054 1041 0059 207 0032 207 0032 1041 9763 1085 5425 4209 0059 1412 9763 1429 1059 1429 1059 1429 1059 1429 1059 1429 1059	145.0	9401 • Ratios 0 155 0 2094 0 153 0 153	0.236 0.281 0.287 0.286 0.227	0439 0439 0465 0466 0374 0460 0374	1.250 1.250 1.462 1.182 1.105 0.827 1.106	0.297 0.297 1.325 0.8434 1.543 1.543 1.543 1.226	0.458 0.491 2.814 2.152 3.383 5.544 2.399 2.322	
	ated Tables re Rocki Decket Custon	C23113 C23114 C2	Potessore advit ajór Pritessore advit bez PBMs M5565 Spectry Conderce / Conse 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	hge 1 8 (3) - Joind lases in data og 2 4 7.02 - Joind dass in halen and finance (separat og 2 7.02 - Joind Separat og 2 7.	NIS SES DUTIENTS AFLENS SER TX AFLEN AFLEN AFLEX AFLEX AFLEX AFLEX A	16 17 17 17 17 17 17 17 17 17 17 17 17 17	1236 1920 Modifications C	2 6 122e-15 125e-15 125e-15 125e-15 125e-16 125e-10 226e-00 236e-00 236e-00 236e-00 236e-00 236e-07 236e-05	1 277 1 204 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	384 225 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	5.07 6.39 # Possies 1 1	6 0 8 PSMs Haeser 40 00114 313 00114 313 00114 314 00114 315 00114 115 00114 115 00114 12 00114 31 00114	9.2 5.4 Protein Accor 0.001147-1	E4 H2 Haiore	81 326	201 508 0 2 2 1 0 0 0 1 1 1	1435 1997 2005 71455 2005 71455 2005 2005 1991 00592 2977 00392 1991 9409 1995 9425 4205 00595 1985 9425 4205 00595 1985 9425 4205 00595 1985 9425 4205 1994 1985 9425 1985 9425 1985 9425 1985 9425 1985 945 1985 945 1	145.0	9403 e Ratios 0.155 0.155 0.153 0.153 0.153 0.153 0.153	0.238 0.281 0.287 0.288 0.288 0.288 0.288 0.284 0.376	0.426 0.426 0.436 0.438 0.430 0.430 0.427 0.427 0.427 0.427	1.500 1.580 1.580 1.582 1.582 1.582 1.582 1.582 1.595 0.327	0.007 0.207 1.325 0.449 1.543 1.543 1.543 1.226 9.227 0.477 2.373	0.458 0.491 2.814 2.152 3.843 5.844 2.209 2.222 2.852 3.465	
	ated Tables re Rocki Decket Custor	C23713 C223714 C223714 C223714 C223714 C223714 C223714 C223714 C223714 C223714 C223714 C223714 C223713 C223713 C223713 C223713 C223713 C223714	Potescire soluri della Potescire soluri della PBMs MSMS Spectra Conferenci i Conse O	sign 14 (20)-calcilises in shale log 24 (20)-calcilises in shale log 24 (20)-calcilises in shale in bits (Conversion Features (Source)) (Conversion) (Source)) (Conversion) (Source)) (Conversion) (C	NIS SES DUTIENTS AFLENS SER TX AFLEN AFLEN AFLEX AFLEX AFLEX AFLEX A	18 17 DEEHOPLUN MARO,DEEH	1236 1801 Modifications C	2 6 1220-15 1360-10 1360-10 1360-10 1360-10 1360-10 1360-00 2360-00 2360-00 2360-00 2360-00 3320-05 3320-05	1 277 1 204 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	384 225 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 1 2 1	5.07 6.39 # Possies 1 1	6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	52 54 Promin Accord 001147-1 0001147-1	E4 H2 Haiore	81 326	20.1 53.8 0 2 2 1 0 0 0 1 1 1 1 1 1 2 2 1	1435 1997 1997 1997 1997 1997 1997 1997 199	145.0	9403 e Partice 0.155 0.328 0.329 0.329 0.329 0.329 0.329 0.329 0.329 0.329 0.329 0.329 0.329 0.329	0.238 0.281 0.287 0.288 0.288 0.288 0.284 0.376 0.384 0.376	0.426 0.426 0.426 0.426 0.427 0.427 0.427 0.427 0.532 0.552	1.220 1.321 1.320 1.462 1.385 0.327 1.335 0.327 1.336 0.327 1.336 0.345	0.207 0.207 1.325 0.449 1.543 1.642 1.226 0.477 2.875 1.244	0.458 0.491 2.834 2.152 3.843 5.844 2.259 2.252 2.852 3.465 3.369	
	ated Tables re Rocki Decket Custor	CI23713 CI23714 CI23714 CI23714 CI2007 NatUsed Used Used Used Used Used Used Used	Prisence school app Prisence school app Paternere school best PBMs MSMS Spectre Confedence / I Come Confedence / I Come Confedence / I Come	1) pp - 12 (Di-utoritype) in their loge 2.7 (DI-utoritype) in their loge 2.7 (DI-utoritype) in their loge 2.7 (DI-utoritype) in the loge 3.7 (DI-utority	245 575 575 575 575 575 575 575 575 575 5	18 17 DEEHOPLUN MARO,DEEH	1236 1821 Modifications C	2 6 122e-15 1.25e-15 1.25e-15 1.25e-10 2.56e-06 2	1 277 1 204 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	384 225 1 2 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 2 1 1 1 2 2 1 1 2 2 2 1 2 2 5	5.07 6.39 # Possies 1 1	6 75Ma Masar 40 00156 137 00168 33 00156 33 00156 33 00156 130 00156 130 00156 3 00156 3 00156 3 00156	52 54 Promin Accord 001147-1 0001147-1	E4 H2 Haiore	81 326	20.1 53.8 0 2 2 2 1 1 0 0 0 1 1 1 1 1 0 0 2 1 1 0 0 0 1 1 1 0 0 0 1 0 0 0 0	1435 108-7 2005 2005 77450 2005 2005 2005 2005 2007 20020 2007 20020 2007 20020 2007 20020 2007 20020 2007 20020 2007 20020 1000 54250 1000 54250 100000000000000000000000000000000000	Abundani Abundani 0.942 0.942 0.944 0.954 0.954 0.954 0.954	940 3 e Ratios 0 150 0 150 0 153 0 155 0 155	0.228 0.281 0.287 0.286 0.227 0.286 0.227 0.384 0.376 0.386	0.426 0.426 0.436 0.334 0.460 0.460 0.460 0.460 0.460 0.460 0.460 0.460 0.460	6 000 6 134 1 320 1 42 1 42 1 42 1 42 1 42 1 42 1 42 1 42	0.207 0.207 1.303 0.434 1.543 1.543 1.543 1.543 0.477 2.375 1.244 1.565	0.458 0.491 2.894 2.153 3.894 2.253 3.884 2.899 2.852 2.852 3.865 3.960 2.852	
	ated Tables re Rocki Decket Custor	C23713 C223714 C223714 C223714 C223714 C223714 C223714 C223714 C223714 C223714 C223714 C223714 C223713 C223713 C223713 C223713 C223713 C223714	Potescire soluri della Potescire soluri della PBMs MSMS Spectra Conferenci i Conse O	sign 14 (20)-calcilises in shale log 24 (20)-calcilises in shale log 24 (20)-calcilises in shale in bits (Conversion Features (Source)) (Conversion) (Source)) (Conversion) (Source)) (Conversion) (C	245 575 575 575 575 575 575 575 575 575 5	18 17 DEEHOPLUN MARO,DEEH	1234 1301 Modifications C	2 6 1220-15 1360-10 1360-10 1360-10 1360-10 1360-10 1360-00 2360-00 2360-00 2360-00 2360-00 3320-05 3320-05	1 277 1 204 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	384 225 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 1 2 1	5.07 6.39 # Possies 1 1	6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	92 93 94 001167-0 001167-0 001167-0 0001167-0	E4 H2 Haiore	81 326	20.1 53.8 0 2 2 1 0 0 0 1 1 1 1 1 1 2 2 1	1435 1997 1997 1997 1997 1997 1997 1997 199	145.0	9403 e Partice 0.155 0.328 0.329 0.329 0.329 0.329 0.329 0.329 0.329 0.329 0.329 0.329 0.329 0.329	0.238 0.281 0.287 0.288 0.288 0.288 0.288 0.284 0.376 0.384	0.426 0.426 0.426 0.426 0.427 0.427 0.427 0.427 0.532 0.552	1.220 1.321 1.320 1.462 1.385 0.327 1.335 0.327 1.336 0.327 1.336 0.345	0.207 0.207 1.325 0.449 1.543 1.642 1.226 0.477 2.875 1.244	0.458 0.491 2.834 2.152 3.843 5.844 2.259 2.252 2.852 3.465 3.369	

Figure 4. Untargeted label-free quantification results within the Proteome Discoverer software framework. Both the ratios and the scaled abundance values are color-coded to display significantly under- or over-expressed proteins.

Multidimensional LC profiling

The new untargeted label-free quantification algorithm also supports multi-dimensional label-free data. The processing step works as described previously for such data with the feature mapping and retention alignment steps only applied to the same fraction from other datasets. Fractions 11-15 for 14 of the samples from the Pandey group human proteome data were run using the same workflows as shown in the previous section. For these data, a total of 5116 proteins and 60616 unique peptides were identified. Unlike the previous version of Proteome Discoverer software where these data were run using "Top 3" protein quantification results, the pre-release Proteome Discoverer 2.2 software enables scaled abundance visualization of the various samples. Figure 5 shows the proteins most overrepresented in the frontal cortex relative to the other samples by sorting by decreasing scaled abundance. Many of the most overrepresented proteins are all known to be neural proteins, including synapsin-1, synapsin-2, synapsin-3, neuromodulin, and microtubule-associated protein tau. Also, some of these neuronal proteins show no signals for any of the other samples in Figure 5 by showing gray boxes indicating that there were no quantification values for these proteins. This is an indication that the Feature Mapping is actually working correctly by not associating random features from the other datasets. Also, this shows the value of the scaled abundance compared to ratio calculations. If any of the other samples were used as the denominator for the ratio calculations, it might be missed that the selected protein is found only in the frontal cortex sample due to the undefined ratios that would be produced.

Description	Exp. q+alue	Add. Adversigland, Gal. Ella, 42.4	Asis, Boolis, Gol, Elles, 76.1	Askt_CD6Toels_Gel_Else_4L1	Add Color, Gel Elec 51 J	Add. Freedores, Gol Elle, 30 Jo	Aduk Galibladder Gel Elve S2 4	Add, Heart, Gel, Else, 54, J	Mah, Kaliney, Gol, Eller, 55, J	Adult Liver_Gold Elles_33_f	Add, Lung, Gel, Eller, 58, f	Add, Panovar, Gel, Else, 01, f	Aduk, Planslets, Gel, Elin, 48. f	Aduk, Retina, Gol, Elia, 55. /	Mith Spirakoord Gal Else ET J	Sun
Isoform IIb of Synapsin-2 (OS=Homo sapiens) Isoform Tau-D of Microtubule-associated protein tau (OS=Homo sapien	0.000		_			1400.0										_
Isoform 1au-D or Inicrosubule-associated protein tau (US+homo septen Isoform 2 of Hepatocyte cell adhesion molecule IOS+Homo sapiens)	0.000					1403.0										
Isotom 2 of Hepatocyte cell adhesion molecule [US+Homo sapiens] Neuromodulin (OS+Homo sapiens)	0.000		12.9	14	5.8	14010	16	0.1	_		14.7	2.8	0.5	5.2	15.1	-
Isoform 4 of Laucine-rich repeat fightless-interacting protein 1 IDS+Hor	0.000	77.7	12.9	1.4	5.6	1340.0	1.0	0.1			94,7	2.8	0.5	5.2	19.1	
Syntaxin-binding protein 1 (OS+Homo sapiens)	0.000	13.8				1303.3	0.6	_			2.5	14.4	17	31.4	32.3	-
syntaxin-bending protein 1 [US+Homo sapiens] protein lifeguard 2 [OS+Homo sapiens]	0.000	13.8				1303.3	46				2.5	6.9	38	31.4	32.3	
Isoform 2 of Proline-rich transmembrane protein 2 (OS=Homo sapiens)	0.002	3.6	16.4	5.0	_	1209.5	40	27.1	12	5.8		37.4	3.0	22.0	11.8	_
Isoform 2 of Provine-non transmer/brane protein 2 [US=Homo sapiens] Isoform Tau-B of Microtubule-associated protein tau [OS=Homo sapiens]		2.6	16.4	12.7				27.1	12	5.8		3/.4	0.6		99.3	
DnaJ homolog subfamily C member 5 IOS+Homo sagiensi	0.000	1.2	_	127		1266.1	2.8	_				19.0	49	21.2 78.9	28.8	_
DriaJ homotog subtarnity C member 5 [US+Homo sapiens] Isoform 3 of Regulator of G-protein signaling 6 [DS+Homo sapiens]	0.000	1.7					2.8					19.0	49		28.8	-
		0.2	82.0		0.2	1256.9	7.3	37			58	_	54.1	143.1	13	_
N(G).N(G)-dimethylarginine dimethylarrinohydrolase 2 (OS=Homo sapi																
Synaptosomal-associated protein 25 [OS=Homo sapiens]	0.000	3.8	53.5	14.7	15.0	1229.2	23.9	7.0		23.4			5.2	1.4	3.1	-
wiskott-Aldrich syndrome protein family member 1 [OS+Homo sapiens]		7,4	41.8	17.8	6.9	1224.0		25.4	_				3.4	15.0	32.6	
Calcium/calmodulin-dependent protein kinase type II subunit alpha (DS		2.6	4.9	8.4	2.2	1220.3	6.2	91.2	7.0	4.0	4.1	9.8	11.6	15.9	11.8	-
Afamin (OS+Homo sapiens)	0.000	0.6	4.1	0.5	23.2	1219.4	66.2	4.0			1.5	5.2	0.1	31.2	43.9	
tyrosine-protein phosphatase non-receptor type substrate 1 [DS+Homo	0.000	9.3	13.1	41	8.9	1201.1	13.5	34.4	0.6	39.0	15.0	13.9	10.3	18.8	38.0	-
Synapsin-1 [OS=Homo sapiens]	0.000	6.8	17.4	5.9	4.7	1197.8	6.3	5.6	29.5	20.7	19.0	33.3	2.0	19.5	31.4	
Isoform 2 of Syntaxin-binding protein 1 [OS+Homo sepiens]	0.000	7,8	8.5	4.9	4.3	1194.2	2.9	3.3	18.9	11.2	8.5	11.7	7.6	65.2	51.0	

Figure 5. Minora feature detection results for the subset of the Pandey human proteome dataset. The features are sorted by decreasing scaled abundance for the frontal cortex sample.

CONCLUSIONS

A new untargeted label-free quantification workflow based on the Minora algorithm has been demonstrated on a dataset with proteins at known concentration and is shown to be more accurate and sensitive than the previously available label-free quantification approaches from previous versions of Proteome Discoverer software. The combination of the label-free quantification workflow integrated into the scaling, normalization, and study management features of Proteome Discoverer software provide a powerful means for analyzing highly complex proteomics data.

REFERENCES

- 1. Gemperline, DC et al, Proteomics, 16, pp 920-924.
- 2. Kim, M.-S. et al, Nature, 509, 575-581

For Research Use Only. Not to be used in diagnostic procedures.

Find out more at thermofisher.com

© 2016 Thermo Fisher Scientific Inc. All rights reserved. All rights reserved. TMT is a trademark of Proteome Sciences plc.SEQUEST is a registered trademark of the University of Washington in the United States. Microsoft and Excel are trademarks of Microsoft Corporation.All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified.

