Supplementary Information For

NSI and NSMT: usages of MS/MS fragment ion intensity for sensitive differential proteome detection and accurate protein fold change calculation in relative label-free proteome quantification

Qi Wu, Qun Zhao, Zhen Liang, Yanyan Qu, Lihua Zhang and Yukui Zhang

1 Database search parameters for the Mix1-LTQ dataset

Carbamidomethyl of C (+57.0215 Da) is set as fixed modification, while oxidation of M (+15.99492 Da) is set as variable modification. The precursor and fragment mass tolerances are 2.0 and 1.0 Da respectively. The acceptable charge states range from 2+ to 4+, while the number of allowed missed cleavages is set to 2. The database is provided along with the dataset, consisting of a *H. influenzae* database appending the 18 standard proteins of interest, common contaminating proteins and trace level contaminants.

2 Protocol for choosing one protein to represent each protein group from pBuild result

If there is a protein that contains at least one unique peptide for a group (by “unique”, we mean that this peptide doesn’t belong to any other single protein in the searched database), this protein is chosen to represent this protein group; if not, all sameset proteins in a group are sorted by ascending alphabetic order of protein names and the first protein in the list is chosen.

3 Calculation of AMT, SMT and SI

SMT is rather easy to calculate since m/z and intensity of fragment ion are stored in mgf file in plain text. The intensities of all fragment ions in one MS/MS spectrum are summed to provide SMT of this particular spectrum. Then SMTs of all spectra assigned to one protein are summed to provide the SMT of this protein. AMT is calculated by dividing SMT of one protein by its SC.

For SI calculation, the algorithm should operate in a search engine independent way and runs in the following manner, similar to Trudigan *et al* in the main text:

- Retrieve MS/MS spectrum and filter fragment ions according to S/N ratio threshold of 10.
- Calculation of theoretical fragment ions:
  - ✓ Calculate theoretical b, y fragment ions and MH precursor ions for peptide at charge states retrieved from the search result.
  - ✓ Calculate loss of NH_3 and H_2O masses for MH ion.
  - ✓ Calculate loss of NH_3 masses for b and y fragment ions if they include R/K/N.
  - ✓ Calculate loss of H_2O masses for b and y fragment ions if they include S/T/E/D.
  - ✓ For any of the above mentioned fragment ion, the highest charge state is set to 3+.
  - ✓ Calculate masses of singly charged immonium ions (AA mass – CO + proton)
- Loop through filtered fragment ions in spectrum from most to least abundant. Check for match to theoretical ions within specified mass tolerance.
- Exclude the theoretical ion from further matches when a match occurs.

4 Data processing for LTQ-XL-Orbitrap@O86 dataset

12 RAW (4 samples, each got triplicate runs) data files were converted to mgf format using pExtract. Identification of 12 mgf files was performed with a local installation of MASCOT (Matrix Science, version 2.3.02) against a database consisting of the 48 UPS1 proteins sequences downloaded from Sigma-Aldrich product website and the *Saccharomyces cerevisiae* sequences from the Saccharomyces Genome Database (SGD)[1] downloaded from the SGD web site (http://downloads.yeastgenome.org/), along with the appended decoy database consisting of reversed form of all the sequences mentioned above. The search parameters were originated from dataset producer in the main text and listed as follows: precursor and fragment mass tolerances are 10ppm and 0.5Da respectively; only double- and triple-charged tryptic peptides with at most one missed cleavage site are taken into account; observed precursor masses are permitted to differ from the peptide monoisotopic mass by 1 or 2Da in case that a non-monoisotopic peak is chosen for fragmentation; no fixed modifications are considered and five variable modifications are set, which are acetylation of the N-terminus, oxidation of methionine, pyro-glutamate formation for N-terminal Gln, carbamidomethyl cysteine formation, and pyro-carbamidomethyl cysteine formation for N-terminal cysteine. The search results were filter by pBuild in exactly the same way as in main text. Also, the calculation of AMT, SMT and SI, along with NSMT and NSI, was implemented in the same way as mentioned above in Supplementary Information and in the main text.

5 PLGEM argument settings

Two parameters were set as following: trimAllZeroRows=TRUE and zeroMeanOrSD="trim" to reduce the side effect of missing values normally seen in proteomics dataset. The number of iterations of the permutation step was set to 2000 rather than the default 500 to stabilize *p* values from run to run.

6 References for Supplementary Information

1. A. I. Nesvizhskii, A. Keller, E. Kolker and R. Aebersold, A statistical model for identifying proteins by tandem mass spectrometry, *Anal. Chem.*, 2003, **75**, 4646-4658.

Table S1. The pooled ratios calculated from average abundance features of the triplicate analyses for the 5 normalized algorithms.

NSAF	Normalized SI divided by SC	NSI	Normalized SMT divided by SC	NSMT
0.59	0.70	0.97	0.74	1.03
0.66	0.99	1.03	0.98	1.08
0.89	0.99	1.34	1.06	1.35
0.98	1.03	1.35	1.06	1.36
1.08	1.10	1.42	1.13	1.46
1.10	1.13	1.61	1.16	1.52
1.10	1.16	1.85	1.21	1.52
1.11	1.23	1.88	1.26	1.58
1.19	1.25	1.98	1.29	1.64
1.29	1.45	2.10	1.29	1.80
1.40	1.48	2.12	1.31	1.82
1.47	1.57	2.15	1.35	1.91
1.48	1.63	2.46	1.36	1.95
1.48	1.64	2.55	1.40	2.01
1.50	1.65	2.70	1.45	2.05
1.51	1.73	2.71	1.51	2.19
1.53	2.04	2.74	1.54	2.30
1.53	2.06	3.06	1.58	2.54
1.56	2.08	3.25	1.63	2.59
1.58	2.22	3.32	1.69	2.61
1.59	2.23	3.38	1.73	2.65
1.59	2.25	3.41	1.79	2.67
1.60	2.29	3.48	1.88	2.69
1.69	2.32	3.53	1.88	2.81
1.70	2.38	3.65	1.94	2.94
1.86	2.40	3.65	1.98	2.99
1.87	2.41	3.68	2.07	2.99
2.01	2.74	3.71	2.14	3.37
2.06	2.77	3.83	2.38	3.38
2.11	2.84	4.14	2.38	3.44
2.15	2.85	4.16	2.48	3.49
2.17	2.89	4.17	2.51	3.77
2.21	3.06	4.31	2.55	3.91
2.44	3.08	4.31	3.06	4.19
2.64	3.15	4.35	3.09	4.20
2.93	3.21	4.55	3.11	4.24

NSAF	Normalized SI divided by SC	NSI	Normalized SMT divided by SC	NSMT
3.16	3.28	4.76	3.13	4.54
3.37	3.37	5.05	3.20	4.64
3.42	3.46	5.05	3.31	4.77
3.58	4.09	5.96	3.37	4.78
3.98	4.17	6.00	3.55	4.97
4.18	4.63	6.23	3.84	5.05
4.22	5.20	6.96	3.96	5.40
4.62	5.21	8.08	4.17	5.87
4.75	6.92	8.21	4.18	6.00
5.00	7.15	9.90	6.24	7.30
5.94	12.23	12.18	7.88	10.58
6.70	27.85	59.43	17.64	40.20
7.02	50.53	67.20	38.47	44.09

The values were sorted in ascending order. All values falling in the acceptable range (1.5-6) were marked in red.

Table S2. Number of significantly changed proteins determined by PLGEM at two fixed p=0.05 and 0.1, respectively.

Algorithm	p	UPS1 proteins	Yeast proteins	Sensitivity	FDR
NSAF	0.05	18	62	48.6%	9.5%
	0.1	24	105	64.9%	16.2%
NSI	0.05	20	74	54.1%	11.4%
	0.1	27	145	73.0%	22.3%
NSMT	0.05	18	73	48.6%	11.2%
	0.1	25	152	67.6%	23.4%

Sensitivity=number of UPS1 proteins/37, FDR=number of yeast proteins/650.

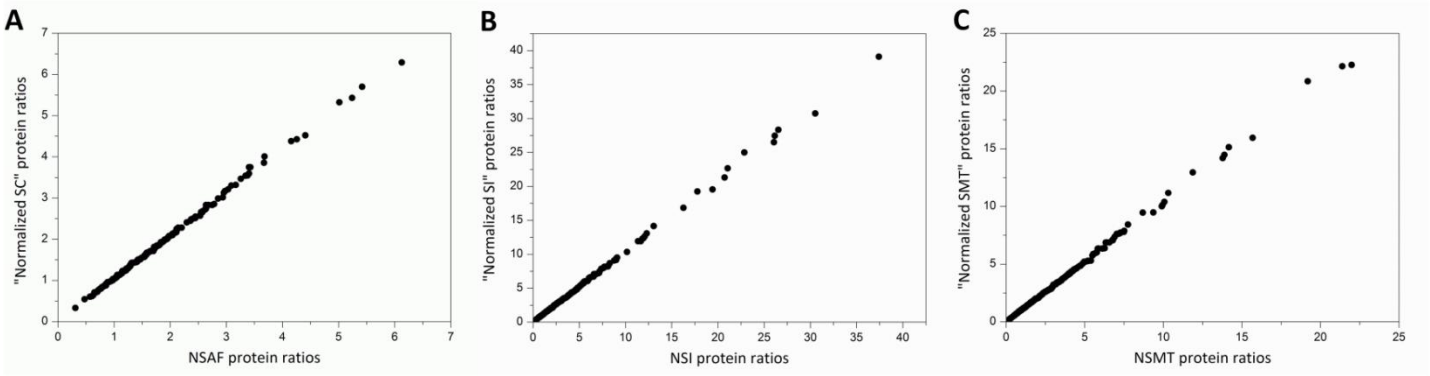

Figure S1. Scatter plot of UPS1 protein ratios obtained with and without adjustment for protein length. Results of (A) SC-based algorithms, (B) SI-based algorithms, and (C) SMT-based algorithms.

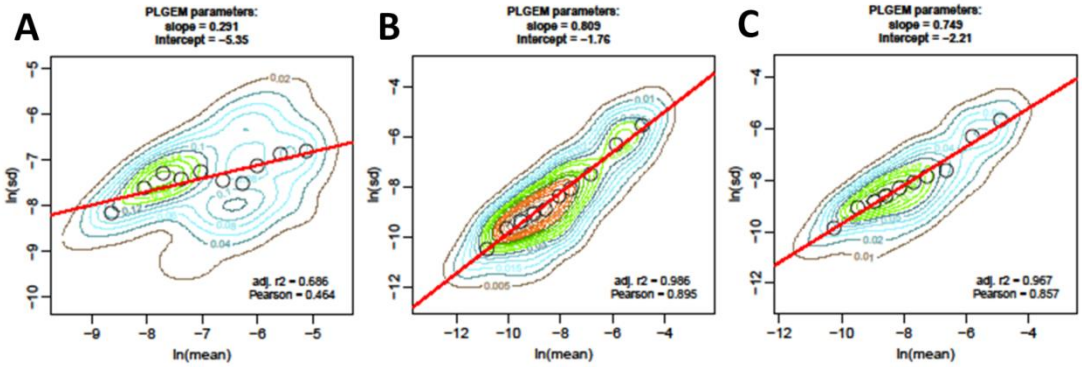

Figure S2. PLGEM diagnostic plot with variance-versus-mean trend line. The slope, r^2 and Pearson correlation coefficient demonstrate the goodness of fit to this model, all of which are the closer to 1 the better. (A) NSAF, (B) NSI, and (C) NSMT.