Supporting Information

Assessment of MS/MS search algorithms with parent-protein profiling

Miin S. Lin[1,3], Justin J Cherny[1,3], Claire T. Fournier[1], Samuel J. Roth[1], Danny Krizanc[2] and Michael P. Weir[1,*]

[1]Department of Biology, Wesleyan University Middletown CT 06459.
[2]Department of Mathematics and Computer Science, Wesleyan University Middletown CT 06459.

**Supplementary Table 1.** Detection of peptides by multiple OMSSA parameter sets.

A b/y ion screen

OMSSA parameter set	Number of OMSSA parameter sets[1]					
	All (3,922)	1 (198)	2 (667)	3 (276)	4 (1,097)	5 (1,684)
0	3060	10	74	197	1095	1684
1	3644	90	597	194	1079	1684
2	2134	56	16	131	247	1684
3	3295	38	578	120	875	1684
4	3035	4	69	186	1092	1684

B a/b/y ion screen

| OMSSA parameter set	Number of OMSSA parameter sets[1]					
	All (3,662)	1 (286)	2 (766)	3 (255)	4 (1,059)	5 (1,296)
0	2583	12	53	163	1059	1296
1	3393	136	714	200	1047	1296
2	1702	75	21	129	181	1296
3	3065	50	695	129	895	1296
4	2556	13	49	144	1054	1296

[1] Distinct peptides are counted once per LC-MS/MS Run and also counted once even if seen in multiple parameter sets. The table partitions matched peptides according to the number of OMSSA parameter sets in which they were detected.
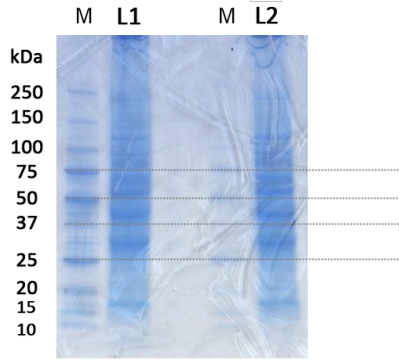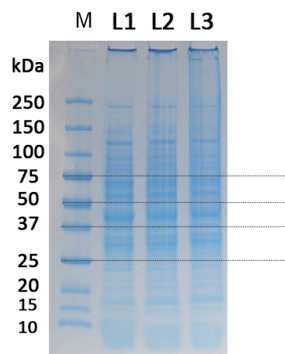
GelSlice_Wes_6

Lane	MS/MS experiment	Size Range (kDa)	
		min	max
L1	1	25	37
L2	2	25	37
L1	3	37	50
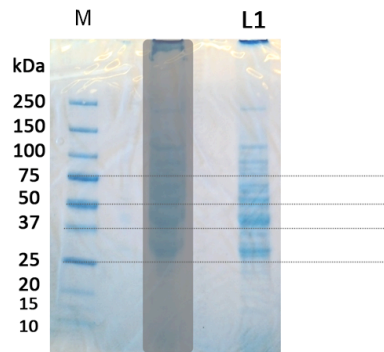
GelSlice_Wes_10

Lane	MS/MS experiment	Size Range (kDa)	
		min	max
L1	4	37	50
L1	5	25	37
L1	6	50	75
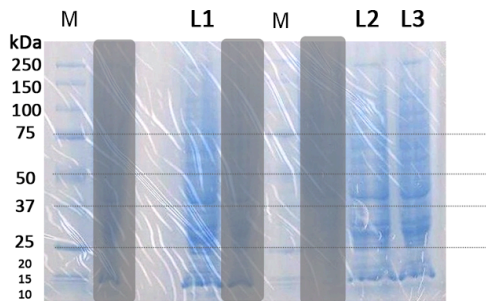
GelSlice_Wes_12

Lane	MS/MS experiment	Size Range (kDa)	
		min	max
L2	7	25	37
L2	8	37	50
L2	9	50	75
L1	10	25	37
L1	11	37	50
L3	12	25	37
L3	13	37	50

GelSlice_Wes_14

| Lane	MS/MS experiment	Size Range (kDa)	
		min	max
L1	14	25	37
L2	15	25	37
L3	16	25	37
L1	17	37	50
L2	18	37	50
L3	19	37	50
L1	20	50	75
L2	21	50	75
L3	22	50	75

Supplementary Figure 1. Partitioning of parent proteins via SDS-PAGE. Illustrated are the 22 gel slices utilized in this analysis. The gel slices fell in three size ranges: 25-37 kDa (9 slices), 37-50 kDa (8 slices) and 50-75 kDa (5 slices). Our earlier experiments focused on the two lower size ranges. The

three size ranges encompass 55.4% (62.5% if not including dubious ORF assignments) of the predicted molecular weights of annotated proteins without post-translational modifications (using 10% mass tolerance: 22.5 kDA to 82.5 kDa).
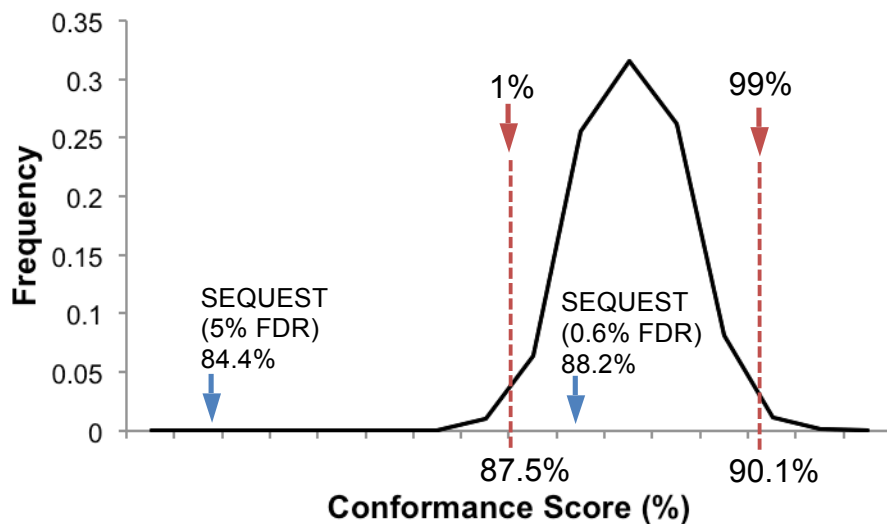

Supplementary Figure 2. Bootstrap analysis of parent protein conformance. 3,060 PSMs were sampled with replacement 1,000 times using output from OMSSA parameter set 0 after filtering at an FDR of 5%. The conformance rate for SEQUEST filtered at 5% FDR is significantly below the 1% distribution threshold whereas conformance of SEQUEST filtered at 0.6% FDR is within the 1% to 99% confidence thresholds range.
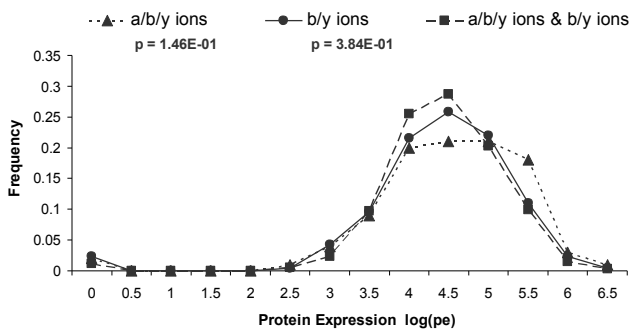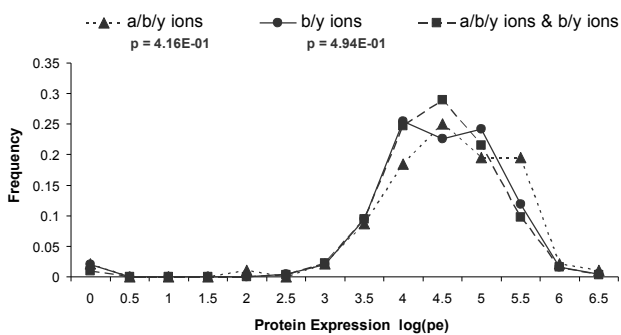
A OMSSA Parameter Set 1
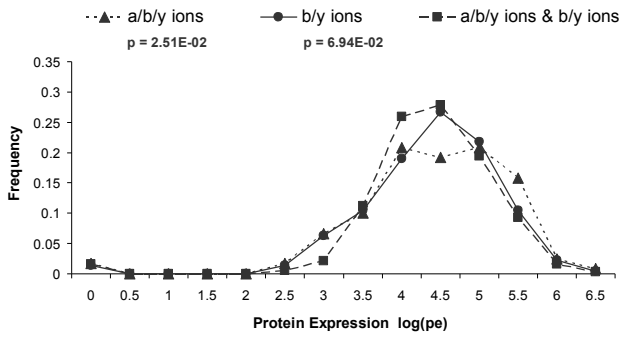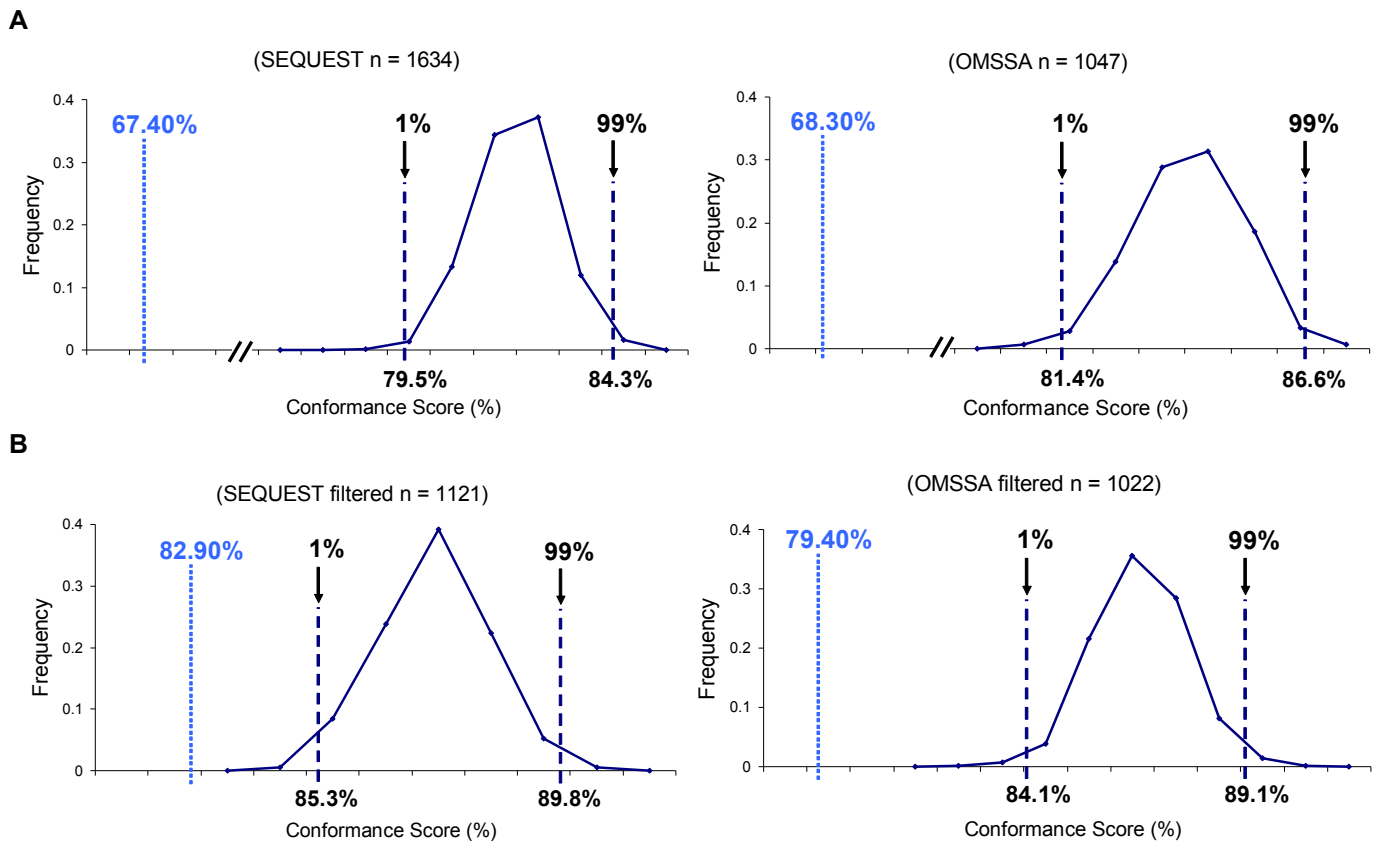

B OMSSA Parameter Set 2
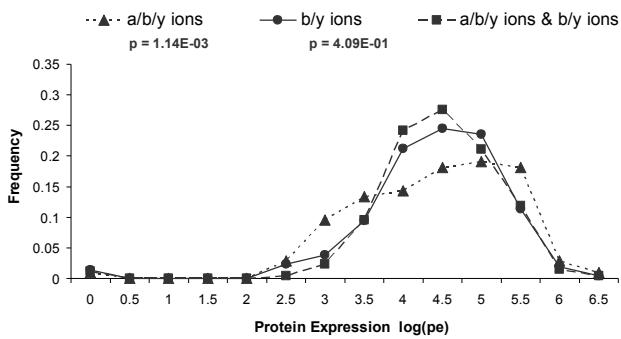
C OMSSA Parameter Set 3
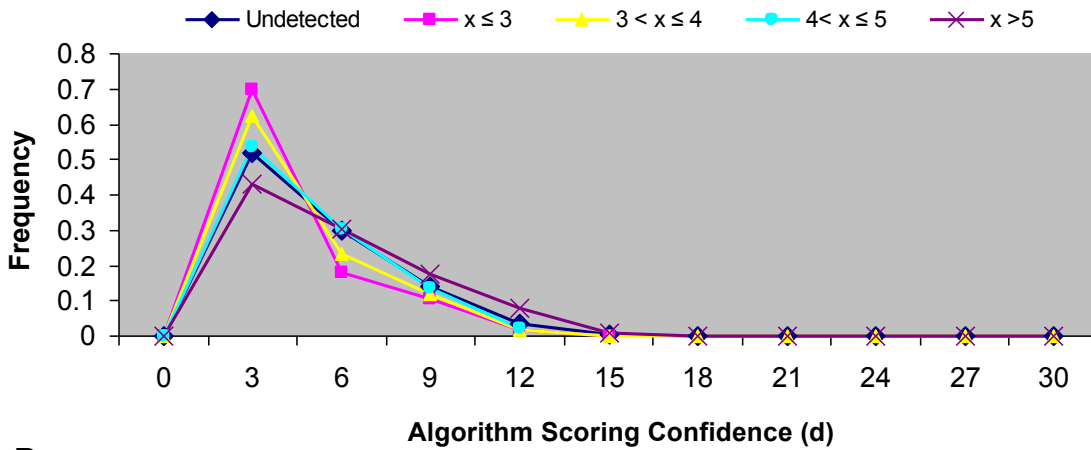

D OMSSA Parameter Set 4

Supplementary Figure 3. Comparison of protein expression for a/b/y and b/y ion screens. Subsets of peptides detected by the b/y ion screen alone, or the a/b/y ion screen alone, had similar protein expression to those detected by both the b/y and a/b/y ion screens. Also see Fig. 4 in main paper.

Supplementary Figure 4. Bootstrap analysis shows significantly depressed conformance scores for SEQUEST-only and OMSSA-only matches – both before and after application of algorithm-specific filters. Dotted lines and percentages represent the conformance score based on distinct peptides detected by SEQUEST or OMSSA alone. Dashed lines and percentages represent the 1st and 99th percentiles. **A.** Left panel: Distribution of 1,000 conformance scores calculated after random sampling with replacement of 1,634 samples from a full set of 5,016 distinct peptides per LC-MS/MS run detected by SEQUEST. Right panel: Distribution of 1,000 conformance scores calculated after random sampling with replacement of 1,047 samples from a full set of 4,429 distinct peptides per LC-MS/MS run detected by OMSSA. **B.** Left panel: Distribution of 1,000 conformance scores calculated after random sampling with replacement of 1,121 samples from a full set of 4,221 distinct peptides per LC-MS/MS run detected by both b/y and a/b/y ion screens of SEQUEST. Right panel: Distribution of 1,000 conformance scores calculated after random sampling with replacement of 1,022 samples from a full set of 4,122 distinct peptides per LC-MS/MS run detected by more than one parameter set of OMSSA.
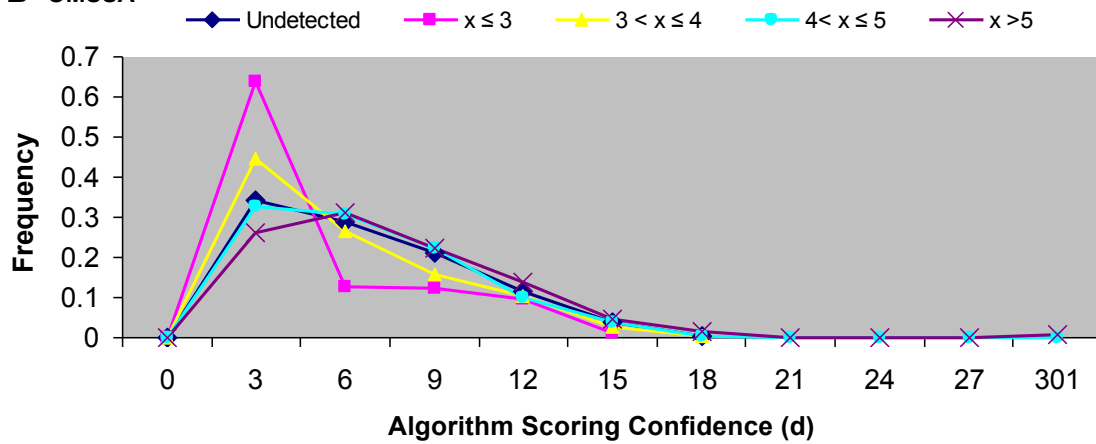
Supplementary Figure 5. Protein expression and algorithm score confidence are related. For both SEQUEST and OMSSA, low expression proteins (e.g. pink line) tend to have lower-confidence algorithm peptide-spectrum match (PSM) scores whereas high expression proteins (e.g. purple line) tend to have both lower- or higher-confidence PSM scores.