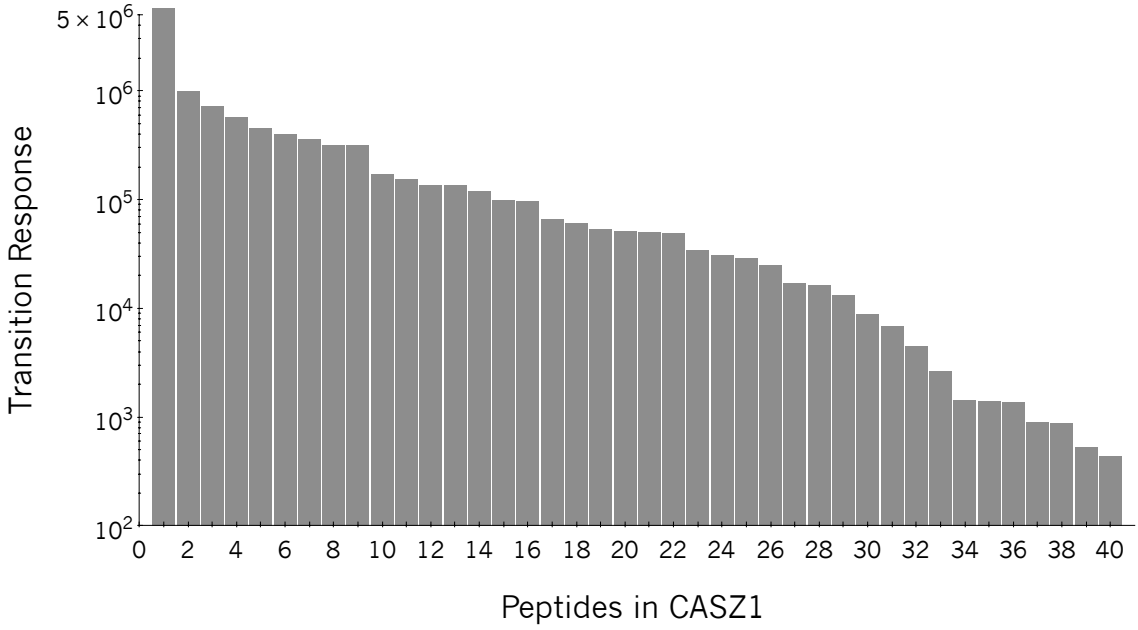
Supplementary Figures for "Using data independent acquisition to model high-responding peptides for targeted proteomics experiments"

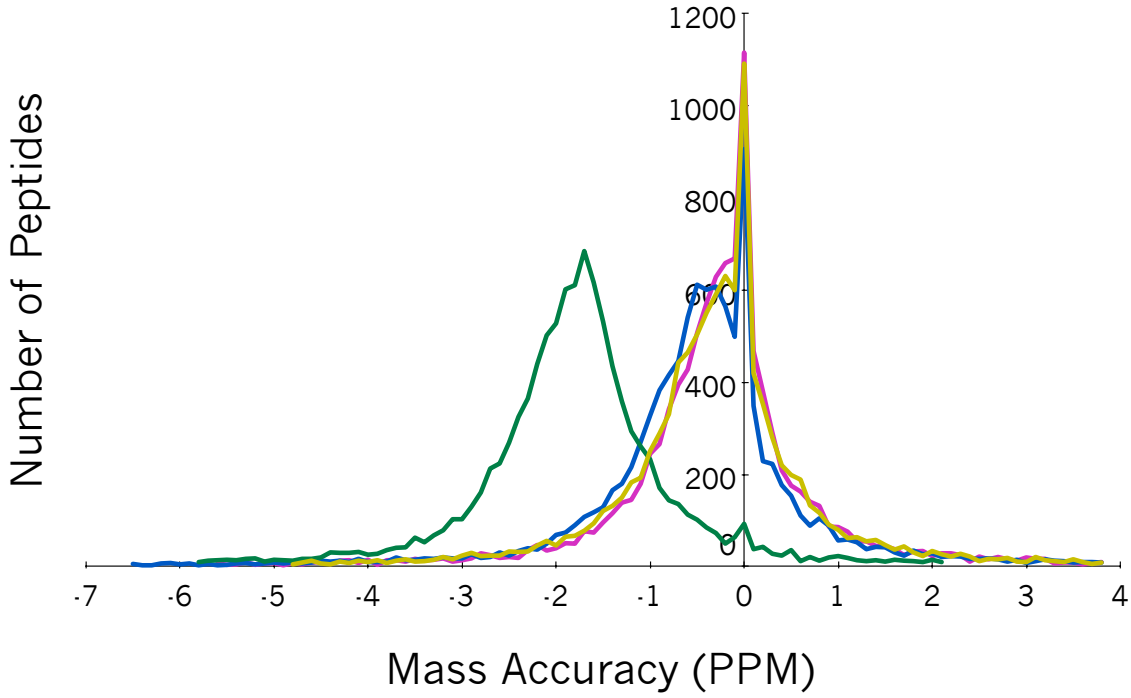
Brian C. Searle[1,2], Jarrett D. Egertson[1], James G. Bollinger[1], Andrew B. Stergachis[1], Michael J. MacCoss[1,*]
[1]Department of Genome Sciences, University of Washington, Seattle, WA
[2]Proteome Software Inc., Portland, OR
[*]Corresponding author

Corresponding Addresses:
Michael J. MacCoss, Ph.D.
Department of Genome Sciences
University of Washington
Seattle, WA
E-mail: maccoss@uw.edu
Telephone: (206) 616-7451

Supplementary Figure 1: SRM transition responses for peptides in CASZ1. The most intense y-type ion fragments intensities (a proxy for the best transition to monitor for each peptide) for the 40 tryptic peptides measured in CASZ1 span 4.1 orders of magnitude. CASZ1 represents a typical protein in the Stergachis *et al* data set, where transition responses of peptides in most proteins span an average of 3.4 orders of magnitude.

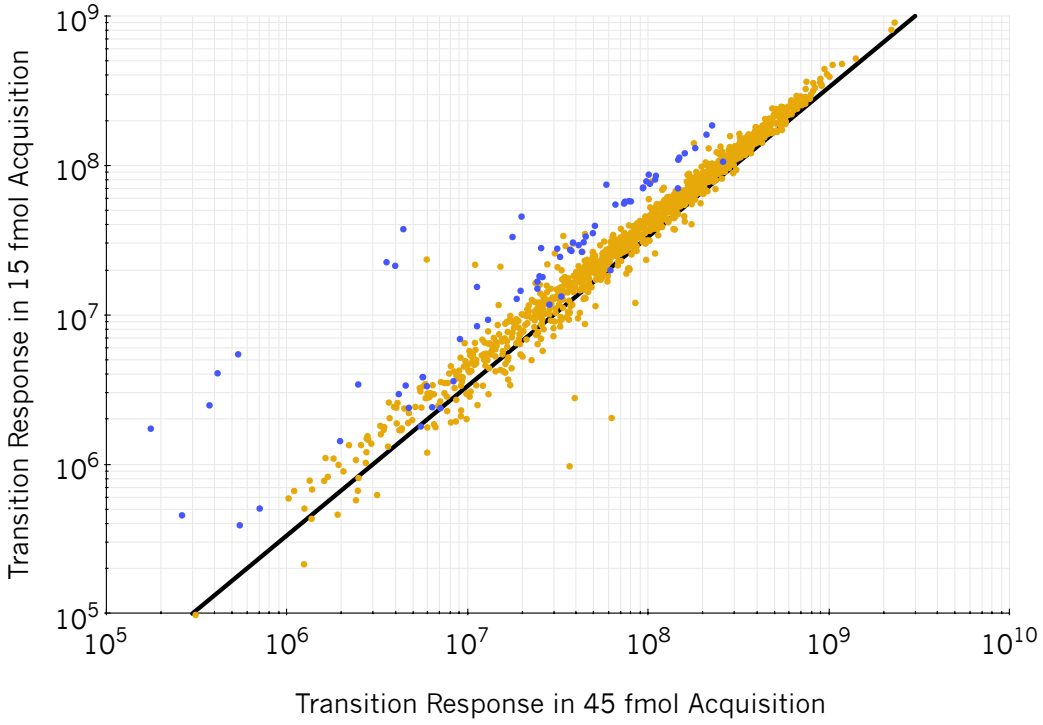


Supplementary Figure 2: Mass accuracy of training set fragment ions. The red histogram corresponds to QEP_2014_0930_RJ_15_DIA, the blue histogram corresponds to QEP_2014_0930_RJ_18_DIA, the green histogram corresponds to QEP_2014_0930_RJ_22_DIA, and the yellow histogram corresponds to QEP_2014_0930_RJ_25_DIA. The interquartile ranges for each distribution are approximately 1.0 parts per million (PPM). Most of the samples follow the same trend, although the green distribution (QEP_2014_0930_RJ_22_DIA) is shifted -1.8 ppm away from the expected median.

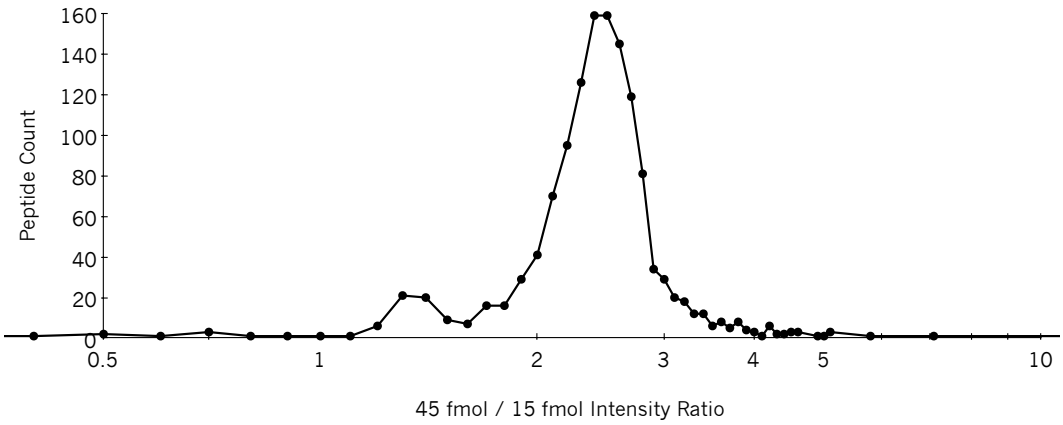


Supplementary Figure 3: Distribution of transition responses in training data set. (a) A scatter plot comparing the 45 fmol injections with 15 fmol injections. The majority of peptides (orange dots) produce responses that fall near the expected 3:1 ratio (black line). Peptides with retention times earlier than 30 minutes (blue) fall off that line (average of 1.35:1) and were excluded from the training data set. (b) The distribution of 45 fmol to 15 fmol peptide transition response ratios. The sample concentrations were generated simply by injecting 3x by volume using the relatively imprecise HPLC autosampler. It is not surprising that the true ratio of peptide transitions falls around 2.45:1. Consequently we used this ratio to normalize the sample intensities rather than 3:1.

a

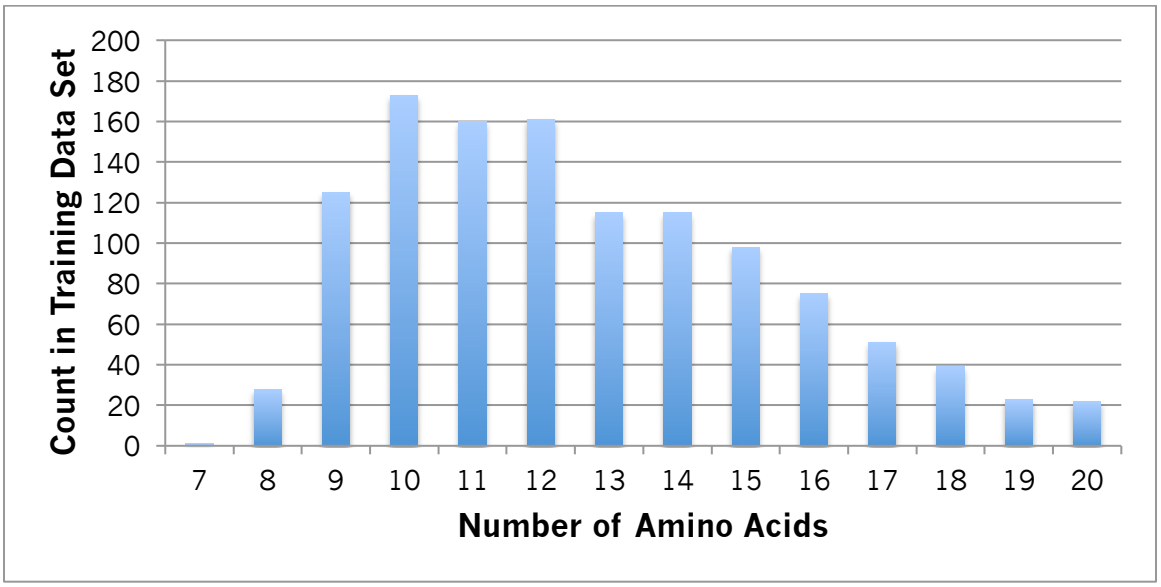


b

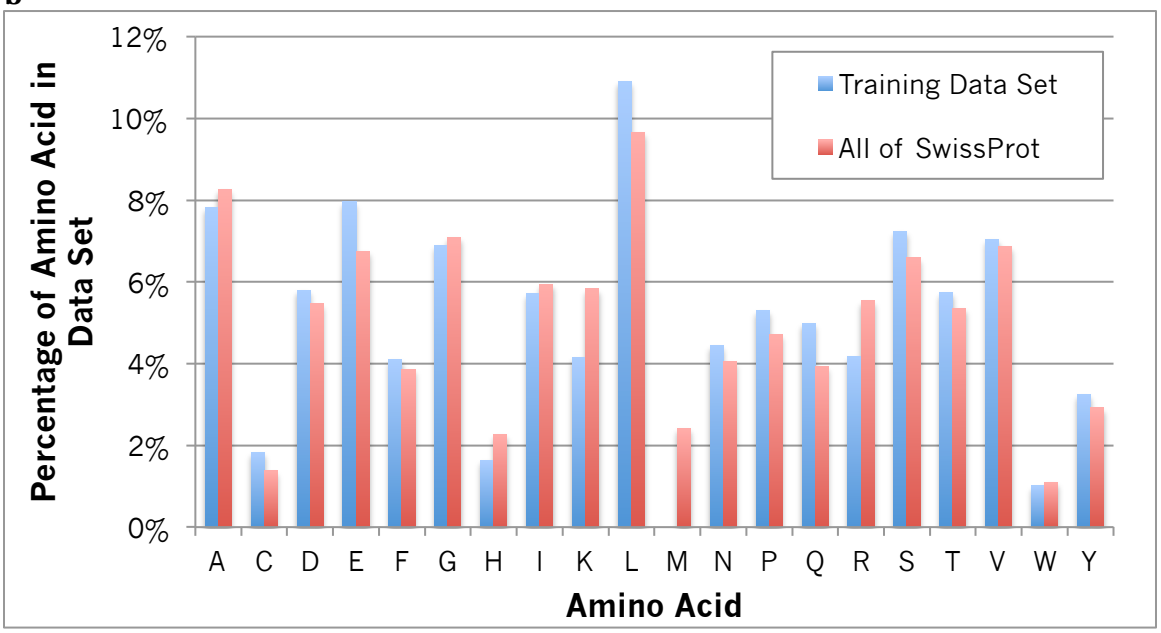


Supplementary Figure 4: Global properties of the training data set. (a) The distribution of peptide lengths in the final 1186 peptide DIA training data set. In Skyline, chromatograms were extracted for the +2 and/or +3 charged precursor of each peptide that fell within the analyzed 500-900 *m/z* range. Consequently peptides primarily contain between 9 and 18 amino acids. (b) The frequency of amino acids in the training data set as compared to all of UniProtKB Swiss-Prot (Release 2015_06, 548,586 entries). The frequencies match quite closely to Swiss-Prot with the exception of methionine residues, which are not represented in the training data set.

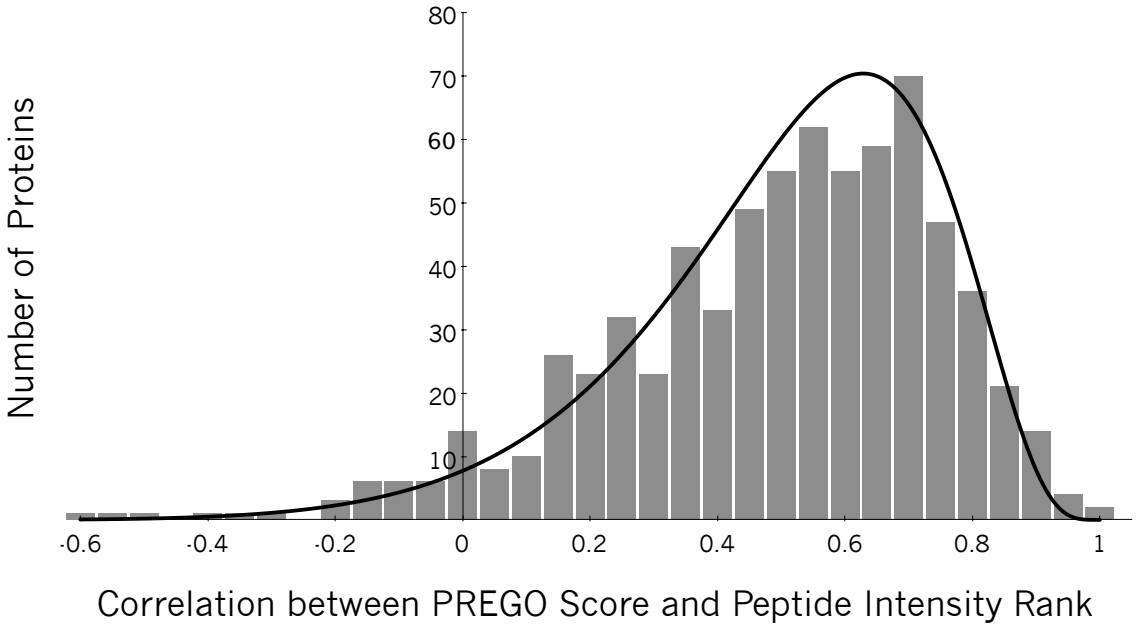
a



b

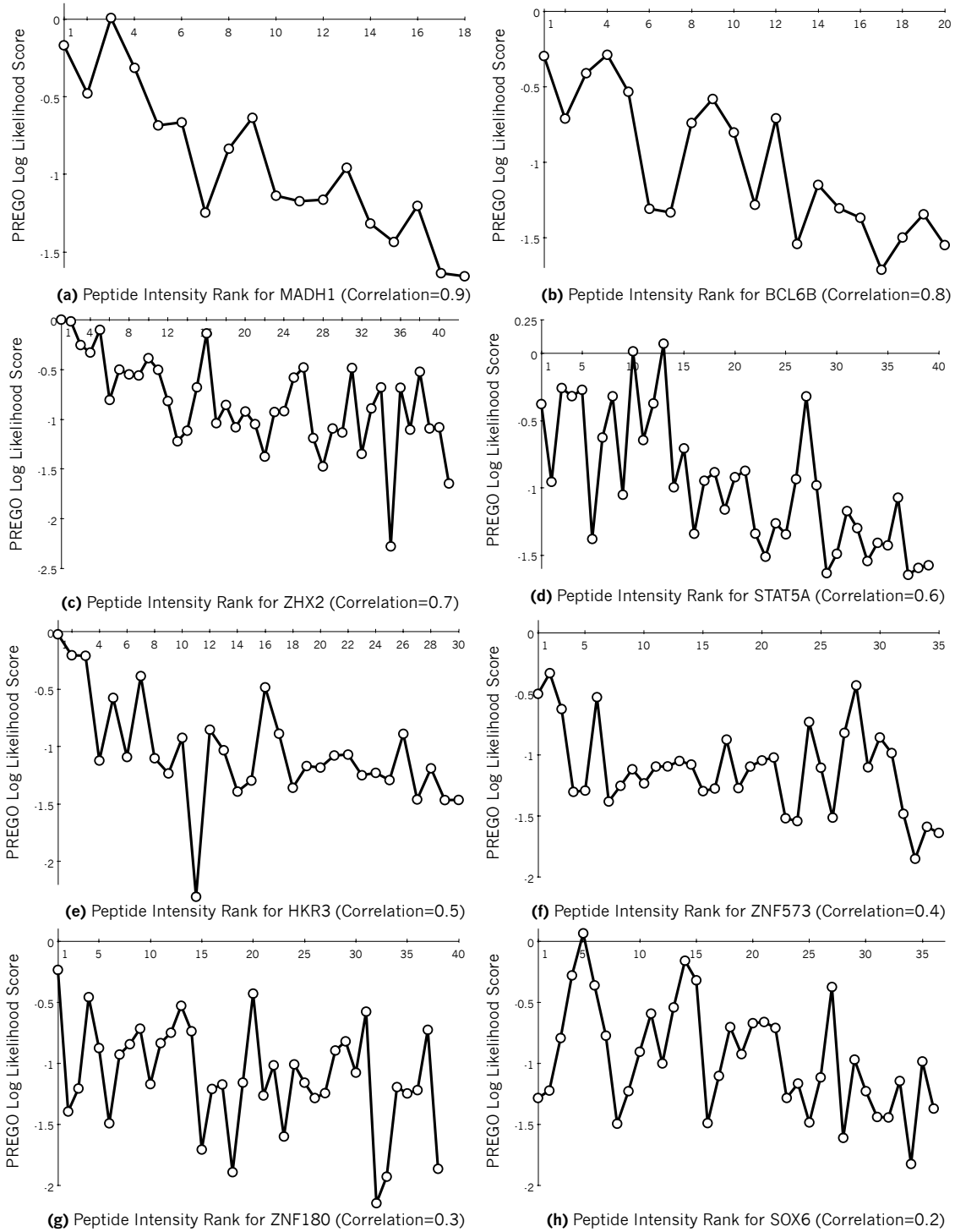


Supplementary Figure 5: The distribution of correlation values between PREGO Scores and ranked peptide intensities. In general Pearson's correlation coefficients center between 0.4 and 0.8, but there is significant spread and the ranked peptide intensities of some proteins negatively correlate with PREGO Scores. There is some expectation that poor correlation should occur by chance, since only a few peptides per protein are considered. The black line indicates the distribution of correlation coefficients predicted* given a sample size of 10 (the mode of the number of peptides per protein) and a true correlation coefficient of 0.65.



* Fisher, R.A. (1915) "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population". *Biometrika 10 (4): 507–521*.

Supplementary Figure 6: PREGO Scores for peptides in proteins with a variety of correlation values. Peptides are ranked on their experimentally acquired transition fragment intensity where the peptide with the strongest response is awarded a rank of 1. The proteins in this figure were chosen methodically: they were picked for having the highest number of peptides as long as the protein correlation value was within +/- 0.01 to the target correlation.



(a) Peptide Intensity Rank for MADH1 (Correlation=0.9)

(b) Peptide Intensity Rank for BCL6B (Correlation=0.8)

(c) Peptide Intensity Rank for ZHX2 (Correlation=0.7)

(d) Peptide Intensity Rank for STAT5A (Correlation=0.6)

(e) Peptide Intensity Rank for HKR3 (Correlation=0.5)

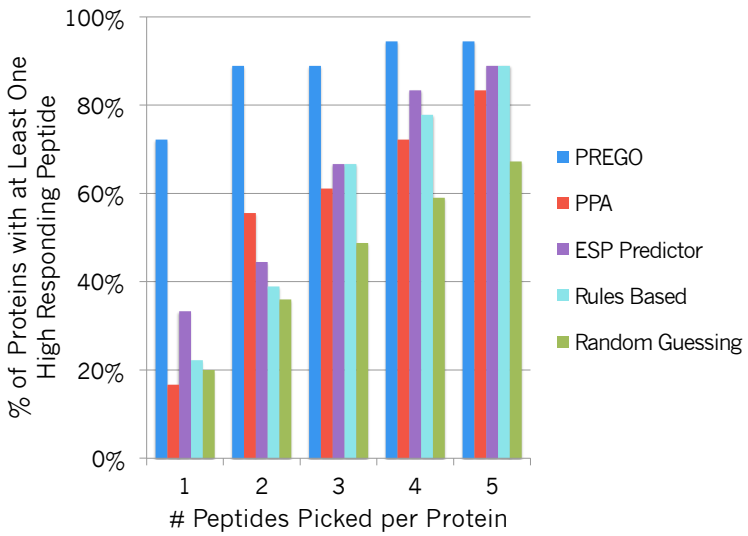
(f) Peptide Intensity Rank for ZNF573 (Correlation=0.4)

(g) Peptide Intensity Rank for ZNF180 (Correlation=0.3)

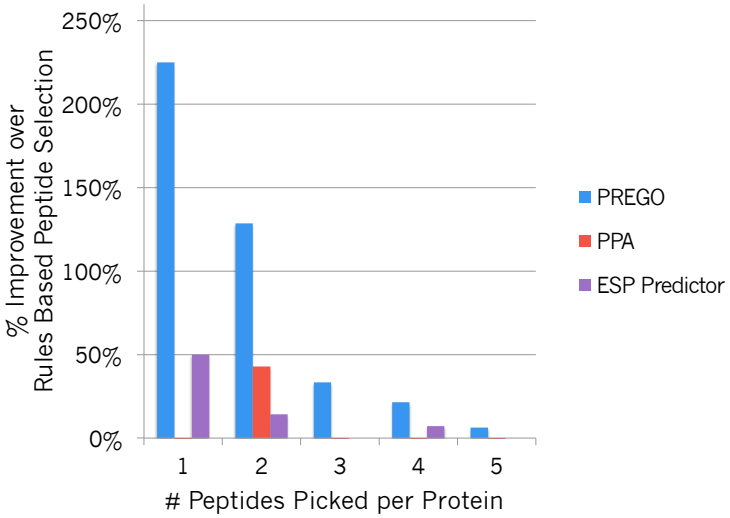
(h) Peptide Intensity Rank for SOX6 (Correlation=0.2)

Supplementary Figure 7: Percentage of proteins with at least one high-responding peptide, given N peptides picked. (a) PREGO (blue), PPA (red), and ESP Predictor (purple) machine learning-based scorers are compared to randomly guessing to select peptides (green) and the simple scoring function described in Equation 2 (cyan) based on common rules in the literature. The data presented here were exhaustively collected on 18 proteins using a similar methodology to that presented by Stergachis *et al*. Scorers are graded similar to Figure 5. (b) The same three learning-based scorers as a percentage improvement over rules based peptide selection. CONSeQuence results were omitted due to unexplainably poor performance (worse than randomly guessing).

a


100%
80%
60%
40%
20%
0%
% of Proteins with at Least One High Responding Peptide
1
2
3
4
5
Peptides Picked per Protein
PREGO
PPA
ESP Predictor
Rules Based
Random Guessing

b


250%
200%
150%
100%
50%
0%
% Improvement over Rules Based Peptide Selection
1
2
3
4
5
Peptides Picked per Protein
PREGO
PPA
ESP Predictor

Supplementary Figure 8: Scatterplot demonstrating the ratio of precursor to fragment intensities in the Stergachis *et al* SRM data set. While precursor intensity is generally a reasonable predictor for fragment intensity, for precursors of any given intensity there is an order of magnitude range between the highest optimal y-ion transition response and the lowest optimal y-ion transition response.