Journal of proteome research

pubs.acs.org/jpr

Article

Tailor: A Nonparametric and Rapid Score Calibration Method for Database Search-Based Peptide Identification in Shotgun Proteomics

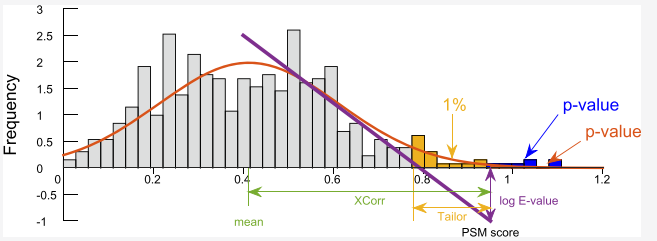
Pavel Sulimov and Attila Kertész-Farkas*

Cite This: https://dx.doi.org/10.1021/acs.jproteome.9b00736

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Peptide-spectrum-match (PSM) scores used in database searching are calibrated to spectrum- or spectrum-peptide-specific null distributions. Some calibration methods rely on specific assumptions and use analytical models (e.g., binomial distributions), whereas other methods utilize exact empirical null distributions. The former may be inaccurate because of unjustified assumptions, while the latter are accurate, albeit computationally exhaustive. Here, we introduce a novel, nonparametric, heuristic PSM score calibration method, called Tailor, which calibrates PSM scores by dividing them with the top 100-quantile of the empirical, spectrum-specific null distributions (i.e., the score with an associated *p*-value of 0.01 at the tail, hence the name) observed during database searching. Tailor does not require any optimization steps or long calculations; it does not rely on any assumptions on the form of the score distribution (i.e., if it is, e.g., binomial); however, it relies on our empirical observation that the mean and the variance of the null distributions are correlated. In our benchmark, we re-calibrated the match scores of XCorr from Crux, HyperScore scores from X!Tandem, and the *p*-values from OMSSA with the Tailor method and obtained more spectrum annotations than with raw scores at any false discovery rate level. Moreover, Tailor provided slightly more annotations than *E*-values of X!Tandem and OMSSA and approached the performance of the computationally exhaustive exact *p*-value method for XCorr on spectrum data sets containing low-resolution fragmentation information (MS2) around 20−150 times faster. On high-resolution MS2 data sets, the Tailor method with XCorr achieved state-of-the-art performance and produced more annotations than the well-calibrated residue-evidence (Res-ev) score around 50−80 times faster.



KEYWORDS: *score calibration, PSM scores, heuristic, fast, peptide assignment, spectrum identification, database search*

INTRODUCTION

Mass spectrometry (MS) is now the de facto method used for protein identification in complex biological samples.[1] Subsequent computational analysis of data involves the application of long pipelines of various algorithms to reveal the protein identification and structure.[2−4] Score functions are the workhorses in these pipelines. Each experimental spectrum obtained from an experiment is iteratively scored against a set of so-called candidate peptides (CPs), which are peptides that are selected from a peptide database and whose calculated masses are equal to the precursor mass of the observed spectrum up to a certain tolerance. The scoring typically provides a similarity-like score (i.e., a higher score indicates a better match) based on matching the peaks of the experimental and the theoretical spectrum generated from the peptide sequences in silico. The top-scoring peptide-spectrum match (PSM) is assigned to the experimental spectrum, and this assignment may be correct or incorrect. Uncalibrated, raw scores may indicate different match quality for different spectra. For instance, the distributions of the top-scoring PSMs of doubly and triply charged spectra shown in Figure 1A indicate that a raw score of 2.5 may imply a correct annotation for a doubly charged peptide molecule but an incorrect annotation for a triply charged peptide molecule.[5] Spectrum-specific score calibration methods aim to provide a sort of score normalization so that spectrum assignments become comparable with each other; therefore, a single threshold can be selected to accept or reject spectrum annotations. The calibration allows one to obtain many more spectrum annotations at any desired false discovery rate (FDR).[5] Score calibration methods involve a null distribution and calibrate a raw score to either the mean or the tail of the null distribution. In this article, we focus on spectrum-specific score calibration methods and we disregard methods aimed at (a) assessing the confidence of the correctness of PSMs at the

Received: November 1, 2019
Published: March 16, 2020

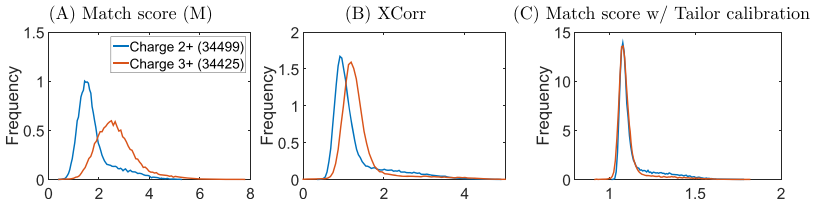


Figure 1. Distributions of top-scoring PSMs obtained on Yeast data for doubly (blue) and triply (red) charged precursor ions. The panels show the distribution of the scores obtained (A) with simple match score $M(.,.)$, (B) with XCorr, and (C) with match scores calibrated with the Tailor method. The density functions were obtained by kernel smoothing using the Gaussian smoother and normalized so that the area under the curve is equal to 1.0. The PSM scoring was carried out via concatenated target−decoy search.

whole-experiment level, such as q-values,[6] (b) improving protein identification,[7,8] (c) improving FDR estimation,[9] or (d) search protocols.[10]

Every experimental spectrum can be associated with a null distribution, which may depend on (a) the experimental spectrum itself, (b) its precursor mass and charge state, (c) the score function, (d) the precision of the discretization, and (e) the theoretical peptide generation procedure (i.e., the charge states and the types of the theoretical fragmentation ions such as b, y, a). For instance, the XCorr[11] score function implemented in SEQUEST,[12,13] in Comet[14] (an open-source reimplementation of SEQUEST[12]), and in Tide-search[15] generates y^{2+}, b^{2+} doubly charged theoretical fragmentation ion masses if the charge state of the precursor ion is higher than 2. Therefore, theoretical spectra contain twice more peaks and the matching results in higher scores simply by chance during searching. Another example is the score function of the Andromeda program,[16] which additionally generates water or ammonia loss peaks for peptides containing D, E, S, T or K, N, Q, R amino acids, respectively. Therefore, a single experimental spectrum can be associated with multiple null distributions with respect to the amino acid compositions of the theoretical peptides.

The XCorr score function is one of the most popular score functions. It was developed for the SEQUEST tandem mass spectrometry (MS/MS) database search tool,[11−13] and it calibrates the match scores to the mean of the null distribution. The XCorr is defined as

$$\mathrm{XCorr}(e, p) = M(e, p) - \frac{1}{151}\sum_{\tau=-75}^{+75} M(e, p[\tau]) \quad (1)$$

for a discretized experimental spectrum e and a theoretical spectrum p and consists of two parts. The first part simply quantifies the match between the experimental and theoretical spectra using the inner product of the corresponding vectors denoted $M(.,.)$. The score function M is referred to as the Match score in this article. The second part provides an estimation of the mean of the null distribution from 151 random matches obtained with shifted theoretical peptides $p[\tau]$ generated by shifting the components of vector p by $\tau \in [-75,75]$ steps. We note that the theoretical spectrum corresponds to a real peptide sequence while shifting its components by $\pm\tau \neq 0$ steps breaks its semantics, and it cannot be associated with any real peptide sequence of the original mass, hence resulting in a random match in scoring. Consequently, the XCorr score returns the signed difference between the match score and an estimated mean of the null distribution. Hence, XCorr incorporates a score calibration step, which is somewhat similar to the standardization (or Z-score normalization) defined as $z = (x - \mu)/\sigma$, where x, μ, and σ represent the match score and the mean and standard deviation (std) of the null distribution, respectively.

We note that we tested a variation of the XCorr, in which the scores were normalized with the variance, but, unfortunately, the variance can often be close to zero, and division by tiny numbers made the overall results worse. Figure 1B shows the distribution of the XCorr scores of the spectrum annotations separately for the doubly and triply charged spectra, and the figure suggests that the calibration mitigates the problem, albeit not entirely. We argue that the XCorr is suboptimal because it does not take into account either the variance or the tail of the null distribution.

The standard approach of score calibration is to assign a spectrum-specific statistical significance to a raw PSM score by estimating a probability of observing a random score equal to or greater than the observed PSM score. This is the p-value, which in fact has well-defined and accurate semantics[17] over various experimental protocols and diverse configurations of MS instruments. The success of the score calibration methods relies on how well they approximate the tail or the extreme tail of the null distribution to obtain a p-value estimation. Some methods employ analytical models, such as a binomial distribution in Andromeda[16] and MS Amanda,[18] Poisson distribution by open mass spectrometry search algorithm (OMSSA),[19] a Weibull distribution for the XCorr,[20] or a Gumbel distribution for spectrum-specific p-value (SSPV),[21] and rely on the assumption that the null distributions have the shape of a certain analytical distribution.

X!Tandem[22] and Comet[14] fit a linear regression line to the estimated survival function of the null model to calibrate the score for each experimental spectrum. Comet employs a log transformation of the survival function, fits a linear regression line, and calculates a calibrated score, an E-value, by extrapolating the linear regression model at the top-scoring PSM score. X!Tandem employs a similar approach; it fits a linear regression line to the empirical survival function of the log of the HyperScores.[22] Both approaches assume that the tail of the null distribution decays exponentially; however, this assumption has not been critically analyzed.

The drawbacks of score calibration methods based on fitting specific parametric models include that they cannot be straightforwardly generalized to other score functions and that the parametric distribution whose parameters are estimated using the overall distributions of PSM scores might not be accurate at the extreme tail.[21]

Other types of p-value estimation methods exploit the exact null distribution obtained from scoring all possible peptide sequences that have the same precursor mass as the observed spectrum.[23−27] The explicit enumeration of all sequences is computationally unfeasible; therefore, a dynamic programming technique is employed to count the peptides at each score in the null distribution. These methods, indeed, result in a perfect score calibration; however, they have several drawbacks. First, they require a proper estimation of the amino acid frequencies in

https://dx.doi.org/10.1021/acs.jproteome.9b00736
J. Proteome Res. XXXX, XXX, XXX−XXX

the peptide database. Second, the calculation of the elements of the dynamic programming table requires a significant amount of central processing unit (CPU) time. Third, the dynamic programming approach requires the score function to be additive.[26] Fourth, the dynamic programming method fails for peak-matching-based score functions (e.g., XCorr) used with data of high-resolution fragment mass accuracy.[27] To overcome many of the issues mentioned above, empirical *p*-values of PSMs can be estimated via scoring spectra against a large number, say 10k, of decoy peptide databases.[5] In this scenario, well-calibrated *p*-values can be obtained for any type of score function using high- or low-resolution MS2 data, albeit at the expense of CPU time. Figure 2 illustrates and compares the principles of the score calibration methods on a null distribution.

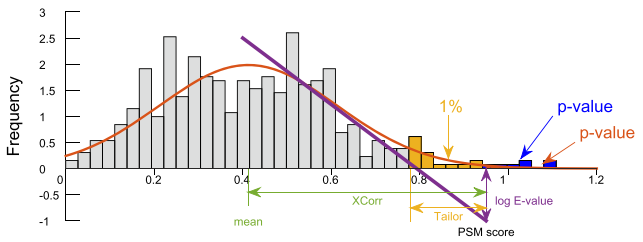


Figure 2. (Gray) Illustration of the principles of the PSM score calibration approaches on a null distribution. (Green) XCorr calibrates the matching score by measuring the difference between the PSM score and an approximation of the mean of the random matching scores (Comet, SEQUEST, Tide-search). (Purple) Regression-based methods fit a linear line on the empirical survival function based on the histogram of the random scores and extrapolate an *E*-value where a PSM score falls on this regression line (Comet, X!Tandem). (Blue) Empirical *p*-values are calculated from the exact null distribution obtained with dynamic programming methods (XCorr exact *p*-value (XPV) in Tide-search, MS-GF+) or with Monte-Carlo techniques.[5] (Red) *p*-Values are calculated using analytical probability density functions (OMSSA, Andromeda, Morpheus, SSPV, Weibull calibration of XCorr). (Yellow) Tailor method calibrates the score to the top 100-quantiles, i.e., relative to the score that has a *p*-value of 0.01 (introduced in this article).

In this article, we propose a new, heuristic score calibration method, called Tailor, which calibrates the score of a PSM to the last 100-quantile, Q100, at the tail of the empirical null distribution, which is constructed for a given experimental spectrum from the scores obtained during scoring it against the candidate peptides. The Tailor method exploits the tail of the observed null distribution, where random scores are observed during the database search step, but not the extreme tail, where samples are rare. This is in contrast to the exact *p*-value (XPV) methods (MS-GF+), which enumerate all random scores, including those at the extreme tail, at the expense of the CPU time to obtain an exact and accurate empirical null distribution. Therefore, Tailor is quick and works with any score function, albeit less accurate (i.e., a heuristic approach), whereas exact methods are accurate, albeit slow and require specific score functions.

Tailor is very easy to implement even in low-level programming languages such as C or C++, and it does not require specific statistical or optimization toolboxes. The Tailor method is nonparametric, and it does not require any specific form of the null distribution. The main benefit of Tailor is its simplicity while providing nearly perfect score calibration. For instance, the distributions of the spectrum assignment scores for doubly and triply charged experimental precursor ions calibrated with Tailor are nearly identical, as shown in Figure 1C. We emphasize that PSM scores ought to be calibrated for every experimental spectrum separately; however, we use the score distributions with respect to charge states only for illustrational purposes, as a sort of sanity checks.

This article is organized as follows. In the next section, we give a detailed definition of the Tailor method and explain the intuition behind it. The Data Sets and Methods section provides details on our experimental protocols and the data sets we used. This is followed by the Results and Discussion section in which we show that Tailor yields more spectrum annotations compared to raw and calibrated scores, such as *E*-values obtained with X!Tandem and OMSSA. We also show that Tailor performs comparably to the computationally intensive exact methods in spectrum annotation, although Tailor runs 20−160 times faster on our benchmark data sets. The last section presents the conclusions of our study.

TAILOR METHOD

Let us consider an experimental spectrum e that is matched to N different candidate peptide sequences during the database-searching step, resulting in the following positive PSM scores: s_1, s_2, ..., $s_N > 0$. Let us assume, for now, that N is large enough and

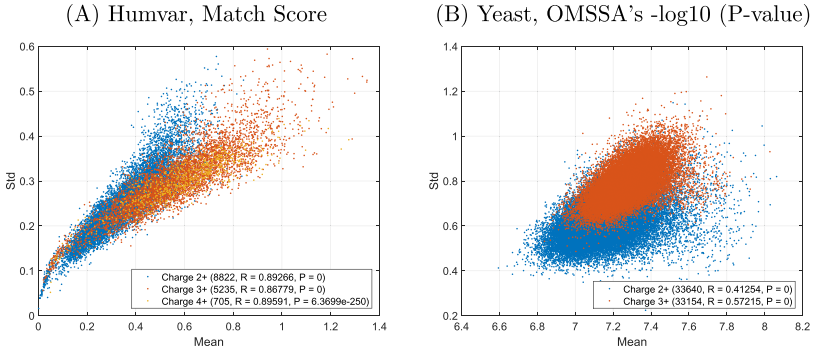


Figure 3. Correlation between the mean and the standard deviation (std) of the empirical null distributions. Each dot represents the mean and the std of the match scores of the candidate peptides of a single experimental spectrum. The scoring was carried out using the HumVar data set with (A) Match score and (B) Yeast with OMSSA's −log(*p*-values) against a decoy peptide data set. For more information about the data sets used, see the Data sets and Methods section. Colors indicate charge states; the numbers in parentheses show the number of the spectra, the correlation coefficients (R), and the *p*-values (P) for testing the hypothesis that there is no relationship between the mean and the std (null hypothesis). The PSM scoring was carried out via concatenated target−decoy search.

https://dx.doi.org/10.1021/acs.jproteome.9b00736
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 1. Summary of Mass Spectrometry Data Sets

name	instrument	#spectra	tolerance[a]	#proteins[b]	#peptides[c]	ACP[d]	MVM[e]	modifications[f]
HumVar	LTQ Orbitrap	15 057	50 ppm/0.02 Da	91 464	3 420 673	1353.7	2	O(V),TMT6-plex(V)
iPRG	MALDI 5600	14 141	10 ppm/0.02 Da	42 450	4 283 235	185.5	1	O(V)
HSPP2A	LTQ	29 583	50 ppm/1.0005079 Da	91 464	1 591 444	443.9	1	O(V)
Yeast	LTQ	69 705	3 Da/1.0005079 Da	6734	269 373	702.4	0	none

[a]Precursor/fragmentation ion tolerance. No isotope error was allowed. [b]In silico enzymatic digestion was performed using the trypsin digestion rule allowing one missed cleavage. [c]Includes modified, target, and decoy peptides. [d]Average number of candidate peptides per spectrum-charge combination (ACP). [e]Maximal variable modifications per peptide sequence. [f]Variable (V) and static (S) modifications, TMT labeling (229.162932 Da) on lysine (K) and on N-terminal (Nt) modifications, oxidation (O) of methionine (+15.9949 Da). Static carbamidomethylation modification of cysteine (+57.02) was used for all data sets.

that these scores are sorted in decreasing order; thus, the experimental spectrum e is to be annotated with the peptide sequence that produces the score s_1. These scores form the basis of an empirical null distribution for the spectrum e. The 100-quantiles define 99 cut points, dividing the range of the score distribution into 100, continuous intervals with equal probabilities. The last (99th) score of the 100-quantiles of the empirical null distribution, denoted $Q100$, is obtained here by selecting the PSM score at the position $i^* = [N/100]$, where $[\cdot]$ denotes the standard rounding operation. Therefore, $Q100 = s_{i^*}$ and the Tailor method calibrates the raw match scores by

$$\tilde{s}_i = \frac{s_i}{Q100} \tag{2}$$

for $i = 1, ..., N$, which are simply referred to as Tailor scores.

The Tailor score calibration is based on division instead of subtraction, and this relies on the following empirical observation. The mean of the null distribution is highly correlated with its standard deviation, as illustrated in Figure 3 for two experimental data sets from our benchmark. Consequently, a null distribution that has a larger mean also has a tail that decays slower compared to distributions that have smaller means. Therefore, a certain difference, say l, between the top score s_1 and $Q100$ ($l = s_1 - Q100$) might be significant for null distributions with a low mean but can be puny for those having a large mean. The ratio of s_1 to $Q100$ implicitly takes into account the width of the null distribution; that is, the wider the null distribution, the higher its mean, the higher the score $Q100$; therefore, the score s_1 is calibrated with a stronger factor $Q100$. Thus, the Tailor method incorporates the width (std) of the null distribution in this way. This is in contrast to the XCorr metric.

Implementation Details

The Tailor method is very simple to implement; we implemented it in the Tide-search[15] program in Crux[28] using 13 lines of C++ code, and it is freely available in Crux at http:crux.ms. The Tailor method in Tide-search can be activated using the parameter `--use-tailor-calibration T`, and the calibrated scores can be found in the column entitled `Tailor score` in the output.

We note that the number of candidate peptides N can be small, resulting in a degenerated null distribution. In this case, the Tailor method scores additional theoretical peptides that are the closest from the outside of the precursor mass window of the experimental spectrum so that $Q100$ can be calculated based on at least 30 PSM scores ($N > 30$). We emphasize that these PSMs are not taken into account in the peptide assignment. The following remarks should be noted: (1) X!Tandem performs via a similar approach, which generates additional peptide sequences with circular permutations to build better null distribution for its E-value calculation. (2) In a typical experiment, most spectra have more than 30 candidate peptides, and this is a rather rare event. (3) The position of $Q100$ is at least 3 (i.e., $Q100 = s_{i^*}$, where $i^* = \min\{3,[N/100]\}$) to avoid homologous or correlated peptides among the top-ranking PSMs.[29]

We also implemented the Tailor method in Python scripts to calibrate the score of the search results of X!Tandem and OMSSA. The Python scripts are provided in the Supporting Information. We note that in these scripts, Tailor eliminates any spectrum annotations that have fewer than 30 candidate peptides.

Negative Scores

Certain score functions (raw score from MS-GF+, XCorr from Comet) may produce negative values for some PSMs, and it is very unfortunate when the signs of the score s_1 and the $Q100$ are different. For such score functions, we add a constant value to every PSM score large enough to keep all scores positive and larger than 1.

DATA SETS AND METHODS

The scripts we used to run our experiments are available in the Supporting Information. The fasta files we used in our experiments are available in the Supporting Information.

Data Sets

HumVar. The human variation data set was derived from lymphoblastoid cell lines from 95 HapMap individuals, including 53 Caucasians, 33 Yorubans, 9 eastern Asians, and 1 Japanese.[30] The complete data set is available at (http://www.peptideatlas.org/PASS/PASS00230), and it consists of 9 092 380 fragmentation spectra in 560 RAW files with a total size of 224 GB. In our study, we used only the Linfeng_012511_HapMap39_3.mzML file, which contained 15 057 experimental spectra. Protein lysates were subjected to detergent cleanup, cysteine alkylation, trypsin digestion, and isobaric tandem mass tag (TMT) labeling. Digested peptides were labeled with sixplex TMT, and the six TMT-labeled samples were equally mixed to generate the final digest mixture. All digest mixtures were analyzed on an LTQ Orbitrap Velos (Thermo Scientific) equipped with an online 2D nanoACQUITY UPLC system (Waters). During data acquisition, the full MS scan was performed in the orbitrap in the range of 400−1800 m/z at a resolution of 60 000, followed by the selection of the 10 most intense ions for HCD-MS2 fragmentation using a precursor isolation window width of 1.5 m/z. Ions with a singly charged state or unassigned charge states were rejected for MS2. Ions within a 10 ppm m/z window around the ions selected for MS2 were excluded from further selection for fragmentation for 60 s. This data set was searched against the IPI.Human database version 3.87, which contains 91 464 protein sequences. Peptides

https://dx.doi.org/10.1021/acs.jproteome.9b00736
J. Proteome Res. XXXX, XXX, XXX−XXX

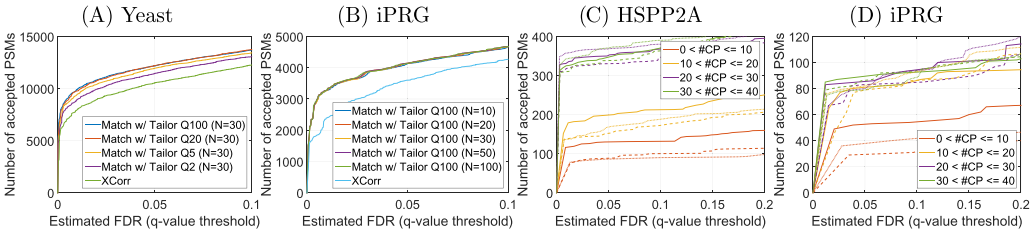


Figure 4. Robustness of the Tailor method to (A) quantile parameter, (B) required minimal number of candidate peptides, and to data sparsity with (C) high-resolution and (D) low-resolution MS2 information. In panels (C, D), the solid, dashed, and dotted lines stand for the Tailor, the XCorr, and the XPV methods, respectively. The lines indicated with the same color correspond to the same category.

were generated using the trypsin cleavage rule, allowing a minimum length of seven amino acids and up to one missed cleavages. Static carbamidomethylation (57.02146 Da) modification of cysteine, TMT labeling (229.16293 Da) of lysine (L) and N-terminal (Nt) amino acids, and oxidation (O) (15.995 Da) on methionine were included as variable (V) modifications. The database-searching parameters for this data set are summarized in Table 1.

Proteome Informatics Research Group (iPRG)2012. The Proteome Informatics Research Group (iPRG) at the Association of Biomolecular Resource Facilities (ABRF) released an MS/MS data set (in this paper called the iPRG data set) to conduct a competition on detecting modified peptides in a complex mixture.[31] Further details and the evaluation of submissions from participants are available at ABRF's website at https://abrf.org/research-group/proteome-informatics-research-group-iprg. The MS/MS data set was obtained from tryptic digestion of yeast proteins and, in addition, was spiked with 69 synthetic modified peptides (called spike peptides). A candidate set of protein sequences (in fasta format) was also provided. Peptides were generated via the trypsin digestion rule. The static carbamidomethylation modification of cysteine (57.02146 Da) and one oxidation of methionine (15.9949 Da) per peptide were used as variable modifications. The database-searching parameters for this data set are summarized in Table 1.

HSPP2A. This data set was generated from the human protein phosphatase 2A system[32] using an LTQ mass spectrometer, resulting in 29 583 spectra, and the data were downloaded from www.peptideatlas.org/repository/publications/Glatter2008. This link is broken; therefore, we provide this data in Supporting File 2. This data set was searched against the IPI.Human database version 3.87, which contains 91 464 protein sequences. Peptides were generated using the trypsin cleavage rule, allowing a minimum length of seven amino acids and up to one missed cleavages. Static carbamidomethylation (57.02146 Da) modification of cysteine and one variable oxidation (15.9949 Da) modification were included on methionine per peptide. The database-searching parameters for this data set are summarized in Table 1.

Yeast. This data set was generated from yeast (*Saccharomyces cerevisiae* strain S288C) samples using a μLC-MS/MS instrument, resulting in 69 705 spectra. This data set was downloaded from https://noble.gs.washington.edu/proj/percolator/. Details on the sample preparation can be found in Section S1.3 of the Supporting Information of ref 33. The yeast protein sequence file was also provided, containing 6734 protein sequences in the fasta format. The peptide sequences were generated with trypsin cleavage using only the default carbamidomethylation modification of cysteine (+57.02146 Da) and allowing up to one missed cleavage. The database-searching parameters for this data set are summarized in Table 1.

Database-Searching Protocols and Programs

The protein sequence (fasta) files were predigested in silico using the Pyteomics Python toolbox,[34,35] allowing one missed cleavage. Decoy peptides were generated with the peptide-reverse approach, in which nonterminal amino acids were reversed. The peptides were exported in a fasta file format in which the labels "decoy" and "target" were appended to the header of the peptides. Therefore, database-searching programs were not allowed to perform (a) any decoy peptide generation, (b) in silico protein digestion, and (c) FDR control. Each database-searching program used the same peptide data set with respect to the input data set. The results were evaluated using the concatenated target−decoy approach[36] on the output of the database search programs. The FDR was estimated as $\hat{\mathrm{FDR}}(t) = (\#\{\mathrm{decoy} > t\} + 1)/\#\{\mathrm{target} > t\}$, as proposed recently by refs 37 and 38, and the q-values[39] of the PSMs were calculated. The number of the accepted spectra was plotted at various q-value thresholds. In the resulting plots, the higher curves stand for better performance.

In our benchmark, we used (a) the Tide program from Crux v3.8.2, (b) X!Tandem v2015, and (c) OMSSA v2.1.9. The parameterization of these programs can be found in Supporting Note S1. We emphasize that the Tailor method was implemented in the Tide-search program, and the experiments were run under the same searching conditions (same search parameters, filtering, preprocessing steps, peptide data sets, etc.). Consequently, the results obtained with different score calibration methods are directly comparable, and the gain or loss can be straightforwardly attributed to the score calibration methods, unless stated otherwise. For X!Tandem and OMSSA, Tailor was implemented as a Python script and run on the output to re-calibrate results. Therefore, the gain or loss in the search results within the same data sets can be directly compared and attributed to the score calibration methods. We note that the results obtained with the same data sets but using different database-searching programs cannot be directly compared because gain or loss in performance can arise from, for instance, different preprocessing steps or from the discriminative power of the different underlying score functions.

All programs were run on an Ubuntu Linux (v18.04.1) PC equipped with 64-bit Intel(R) Xeon(R) CPU E5-2640 v4@2.40 GHz, 128 Gb RAM. Scripts for plotting the results were written in Matlab.

https://dx.doi.org/10.1021/acs.jproteome.9b00736
J. Proteome Res. XXXX, XXX, XXX−XXX

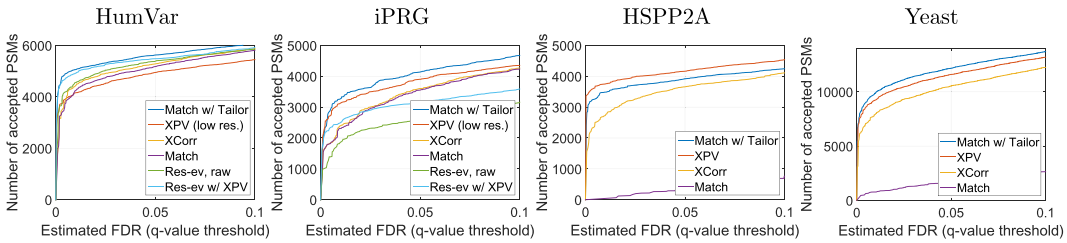


Figure 5. Results of spectrum annotation with various scoring and calibration methods using the Tide-search program. Each plot corresponds to a different data set, as indicated on top.

RESULTS AND DISCUSSION

Robustness

We first show that the Tailor method is robust to parameterization. We executed the Tailor score calibration method using various quantiles as reference points on the Yeast data set. This data set was chosen because Yeast experimental spectra are associated with a large number of candidate peptides (702 candidate peptides per spectrum-charge combination on average) because of the low-resolution precursor information; thus, quantiles could be determined on a large population. The following quantiles were studied: (a) the top 100-quantile denoted $Q100$ and corresponding to a p-value of 0.01, (b) the top 20-quantile defined as $Q20 = s_{\min\{3,[N/20]\}}$ (corresponding to a p-value of 0.05), (c) the top 5-quantile $Q5 = s_{\min\{3,[N/5]\}}$ (corresponding to a p-value of 0.2), and (d) the 2-quantile $Q2 = s_{\min\{3,[N/2]\}}$ (corresponding to a p-value of 0.5). The results are shown in Figure 4A and indicate that the quantile chosen as a reference point has a little effect on the overall results, although there is a clear tendency. The closer the quantile is to the tail of the null distribution, the better the results are. The match scores calibrated to $Q2$ yield the lowest number of annotations, shown by the purple line in this comparison; however, they still yield an improved spectrum annotation compared to the standard XCorr score function (marked with the green line). The match scores calibrated to $Q100$ yield the highest number of annotations, shown with the blue line, although $Q20$ closely follows (red line). Based on this experiment, we propose and use $Q100$ in the Tailor method. We note that the selection of the quantile does not have any effect on the run time.

Next, we studied the effect of the threshold values (N) for the required minimal number of PSMs used to estimate the quantile score. This test was done using the iPRG data set because experimental spectra have relatively few candidate peptides because of the high-resolution precursor information; more specifically, there are 335, 572, 761, 1165, and 2337 experimental spectra, which have fewer than 10, 20, 30, 50, and 100 candidate peptides, respectively, out of a total 14 141 experimental spectra. The number of spectrum annotations shown in Figure 4B indicates that different thresholds result in a small variance in the outcome. Moreover, one could expect longer calculation time for higher threshold values; however, in practice, this has a negligible impact on the CPU time. The run times were 18.3, 17.4, 19.4, 17.4, and 17.5 s for N = 10, 20, 30, 50, and 100, respectively. We note that the run time for standard XCorr is 21.7 s. We conclude that the value of the threshold N has a minimal impact on the search time and the results.

Finally, we show that Tailor is robust to data sparsity, i.e., when the numbers of the candidate peptides (#CP) are low to obtain accurate estimations for the $Q100$ quantiles. We re-run the database searching on the HSPP2A, containing low-resolution MS2 information, and on the iPRG, containing high-resolution MS2 information, data sets using a very narrow, 5 ppm precursor tolerance window. Then, we categorized the spectra into four groups, depending on the number of their candidate peptides. For the HSPP2A data set, this tight tolerance resulted in (a) 2402 spectra having at least 1 but at most 10 candidate peptides, (b) 3543 spectra having more than 10 but at most 20 candidate peptides, (c) 3497 spectra with more than 20 but at most 30 candidate peptides, and (d) 3380 spectra with more than 30 but at most 40 candidate peptides. The rest of the spectra were excluded from this experiment. Figure 4C shows the numbers of the annotated spectra as a function of q-values for each category separately. The dashed, solid, and the dotted lines stand for the spectrum annotations obtained with XCorr, Tailor, and XPV, respectively. The colors indicate the categories. For instance, the red dashed line shows the number of the annotated spectra having at most 10 candidate peptides, annotated with XCorr. Search results marked with different colors cannot be compared. The results show that Tailor calibration works well for spectra having a small set of candidate peptides; Tailor fairly outperforms XCorr and XPV when the numbers of candidate peptides are less than 20 (see red and yellow lines). Tailor lightly outperforms XCorr, but XPV lightly outperforms Tailor when the number of the candidate peptides is more than 20 but maximum 40 (see purple and green lines).

For the iPRG data set, the narrow tolerance window resulted in (a) 589 spectra having maximum 10 candidate peptides, (b) 404 spectra having more than 10 but maximum 20 candidate peptides, (c) 447 spectra with more than 20 but maximum 30 candidate peptides, and (d) 409 spectra with more than 30 but maximum 40 candidate peptides. The rest of the spectra were excluded from this experiment. The spectrum annotation results shown in Figure 4D indicate that Tailor works well with spectra having few candidate peptides in the case of high-resolution MS2 information as well. We re-run the database searching on the Yeast data set using a 10 ppm precursor tolerance window, and the results shown in Figure S4 along with details indicate a similar outcome. Overall, we conclude that the Tailor method remains robust when less data is available to estimate $Q100$ for score calibration.

XCorr

The Tailor score calibration method was compared against the XCorr method[11] and the XCorr's exact p-value (XPV) methods.[26] The underlying score function of XCorr and XPV is the Match score. We obtained the results of the Match score by removing the `void SubtractBackground(double*, int)` procedure call (this procedure is responsible for performing the cross-correlation step) from the XCorr function in the spectrum_preprocess2.cc source file of Tide-search. The numbers of the annotated spectra as a

https://dx.doi.org/10.1021/acs.jproteome.9b00736
J. Proteome Res. XXXX, XXX, XXX−XXX

function of the FDR are shown in Figure 5, and the run times are shown in Table 2 separately for all spectrum data sets. The

Table 2. Database Search Time in Seconds for Various Scoring and Calibration Methods

name	HumVar	iPRG	HSPP2A	Yeast
Match	45.6	19.3	13.5	29.9
Match w/Tailor	48.4	19.4	13.6	30.4
XCorr	49.5	21.7	13.2	28.2
XCorr w/XPV	318.0	138.0	309.0	4940.0
Res-ev[a]	3400.0	308	n.a.	n.a.

[a]Residue-evidence (Res-ev) was designed for high-resolution MS2 data; thus, it was not run on the HSPP2A and Yeast data sets, which contain low-resolution fragmentation ion information.

Match score, as expected, performs poorly on low-resolution MS2 data (HSPP2A, Yeast), as shown by the purple lines, but when it is calibrated with the Tailor method, it consistently and significantly yields more peptides at any FDR than when it is calibrated with the cross-correlation method (i.e., the XCorr). These methods run very fast; however, the Tailor method is slightly faster than XCorr because it does not require a cross-correlation step. The Tailor method approaches the exact and computationally expensive XPV score calibration method in spectrum annotation and even outperforms it on the Yeast data set. The Tailor method is around 20 and 160 times faster than XPV on the HSPP2A and Yeast data sets, respectively.

On high-resolution MS2 data (HumVar, iPRG), the raw Match score performs as well as the XCorr function, as shown by the purple and yellow lines in the top plots of Figure 5. This suggests that cross-correlation calibration for high-resolution MS2 data is futile because, in our opinion, the mean of the null distribution (estimated by the cross-correlation part and then subtracted from the raw score) is close to zero. However, Tailor calibration can further increase the number of the annotated spectra at any FDR level significantly (compare the dark blue and yellow lines) because Tailor calibrates to the top percentile of the null distribution, which is possibly larger than zero.

We executed the XPV method as well, although with a wide bin (bin_width = 1.0005079) because a narrow bin (bin_width = 0.02) would break down XPV.[27] The results show that the Match scores that were corrected by the heuristic Tailor method executed with high-resolution settings yielded more spectrum annotations at any FDR level than the exact XPV methods with low-resolution settings in around 5−20 times less time. The residue-evidence (Res-ev) score function was designed specifically for high-resolution MS2 data to overcome the shortcomings of XPV,[27] and it can be calibrated with exact methods using a dynamic programming approach. The calibrated Res-ev score (p-value) function yielded slightly fewer annotations on the HumVar data set and significantly fewer annotations on the iPRG data set, at any FDR level than the Tailor method; compare the light and dark blue lines in the top plots of Figure 5. Furthermore, Tailor runs around 20−70 times faster, depending on the data set. We note that the results obtained with the Match score with Tailor on high-resolution MS2 data cannot be directly compared to the results obtained with XPV using low-resolution MS2 data or to the Res-ev score using high-resolution MS2 data because the gain or loss in spectrum annotation might not arise from the score calibration but from the discriminative capability of different score functions (i.e., the ability of separating correct annotations from incorrect ones[17]). However, these experiments show that the Match score with Tailor correction can achieve state-of-the-art results on high-resolution MS2 data. Moreover, the fact that the heuristic Tailor method outperformed the well-calibrated Res-ev method may suggest that the Res-ev score might be less discriminative than XCorr, so there still is a room for improvement in data sets of high resolution.

Next, we plotted the score distributions of the top-scoring PSMs for all experimental spectra with respect to their charge states. The distributions are shown in the top row for the XCorr method and in the bottom row for the Tailor method in Figure S1 for all data sets. The plots show that the modes of the XCorr score distributions have different locations with respect to the charge states. This is particularly well visible for the HumVar and Yeast data sets. Furthermore, the location of the modes varies over the data sets as well. Figure 6A illustrates this for the

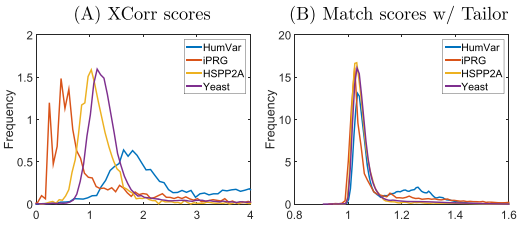


Figure 6. Distributions of the (A) XCorr and (B) Tailor scores obtained using only experimental spectra that have an associated triply charged precursor ion. The distributions were smoothed using a Gaussian kernel. The PSM scoring was carried out via concatenated target−decoy search.

distributions of the top-scoring PSMs of triply charged experimental spectra. The mode obtained for the HumVar data sets is located around 1.8, while the mode corresponding to the iPRG data set is located around 0.5. The Tailor method does a better job in score calibration with respect to charge states shown in the bottom row of Figure S1 and across different data sets shown in Figure 6B.

HyperScore

The Tailor score calibration was applied to the HyperScore metric of the X!Tandem search engine. The numbers of annotated experimental spectra obtained with raw HyperScore scores, E-values, and HyperScore scores calibrated with the Tailor method are shown in Figure 7. The results show that the Tailor method yields an improved number of spectrum annotations compared to the raw HyperScore, but Tailor is on par with the E-value method; however, it yields more spectrum annotations than E-value on the Yeast data set. The score distributions smoothed with kernel smoothing with respect to charge states are shown in the top, middle, and bottom row of Figure S2 for the raw HyperScore, the E-value, and the Tailor scores, respectively. These plots show that raw HyperScore results in different modes with respect to the charge states, whereas the E-values and the Tailor method result in modes roughly at the same location. The plots in Figure 8A,B show the score distributions of the triply charged experimental spectra with respect to the data set for the raw HyperScore and the HyperScore with the Tailor method, respectively. These plots suggest that the Tailor calibration improves the raw HyperScore metric.

https://dx.doi.org/10.1021/acs.jproteome.9b00736
J. Proteome Res. XXXX, XXX, XXX−XXX

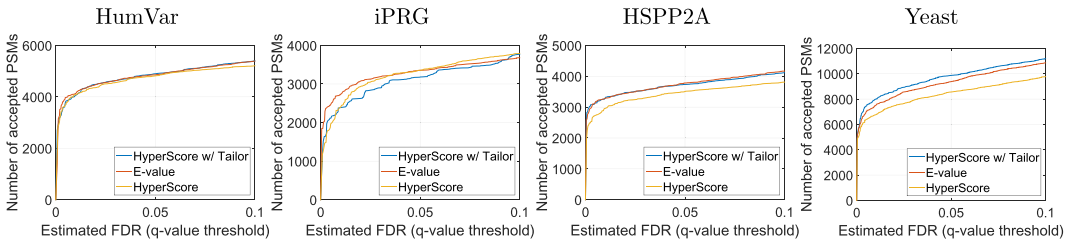


Figure 7. Spectrum annotation results using the X!Tandem program. Each plot corresponds to a different data set indicated on top.

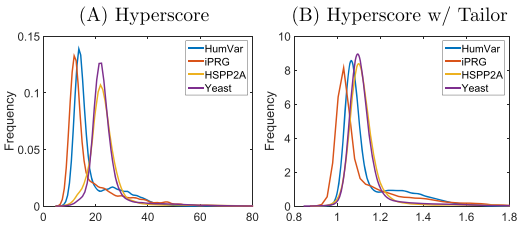


Figure 8. Distributions of the (A) HyperScore and (B) Tailor scores obtained using only experimental spectra having an associated triply charged precursor ion. Distributions were smoothed using a Gaussian kernel. The PSM scoring was carried out via concatenated target−decoy search.

OMSSA Score

We examined the *p*-values obtained with the OMSSA software and re-calibrated them with the Tailor method. We note that OMSSA does not report any raw match scores but only *p*-values and *E*-values of the corresponding PSMs, and the Tailor method was run to re-calibrate the minus-log transformation of the OMSSA's *p*-values (i.e., $-\log_{10}(p\text{-value})$). The transformation does not affect the FDR calculation but converts the *p*-values into higher-the-better-like scores for Tailor and enables more appealing and informative visualization.

The results plotted in Figure 9 show that the Tailor method managed to yield additional spectrum annotations by re-calibrating OMSSA's *p*-values. This outcome suggests that the *E*-values could be further improved and OMSSA's assumptions regarding its *p*-value calculation methodology might not be profoundly justified in practice. Figure S3 plots the distribution of the minus-log transformation of the *p*-values (top row) and their calibration with the Tailor method (bottom row) with respect to the charge states. These plots indicate that the modes of these distributions are located around the same place for *E*-value and Tailor, except for the Yeast data sets, for which both methods produced different modes for doubly and triply charged spectra. However, we assume that the Tailor method "inherited" the different modes from OMSSA's *p*-values.

We note that we did not re-calibrate the OMSSA's *E*-values with the Tailor method because OMSSA's *p*-values and *E*-values are highly correlated and would result in the same outcome.

COMBINATION WITH POSTPROCESSING METHODS

The Tailor method, similarly to any other score calibration methods, can be combined with any postprocessing methods such as Percolator[33] and PeptideProphet[40] to obtain additional spectrum annotations at a given FDR. We tested Percolator with calibrated scores and displayed the results in Figure 10. We found that combining digestion- and charge-state-specific information with scores calibrated with XPV and Tailor using Percolator can result in an additional number of PSMs over the entire 0−10% FDR range. On the high-resolution data sets (HumVar and iPRG), Match + Tailor + Percolator combination identified 3.2 and 11.6% more PSMs than Match + Tailor alone (from 5059 to 5222 on HumVar and from 3264 to 3645 on iPRG) at 1% FDR, respectively; however, Match + Tailor alone identified 8.7 and 8.1% more PSMs than XCorr + Percolator (from 4653 to 5059 on HumVar and from 3020 to 3264 on iPRG) at 1% FDR, respectively. On the low-resolution data sets (HSPP2A and Yeast), Percolator can further improve the number of annotations obtained with Tailor or XPV in our experiments. Overall, the Tailor calibration of the Match score is on par with exact calibration methods such as XPV and MS-GF+ and outperforms other methods such as OMSSA and X!Tandem on our benchmark, shown in Figure 10. Furthermore, Percolator can provide additional PSMs via combining digestion- and charge-state-specific information with calibrated match scores. Table 3 shows the number of annotations obtained at 1% FDR level using various methods and the performance change relative to the Match + Tailor + Percolator method combination.

CONCLUSIONS

In this article, we introduced a new, heuristic score calibration method, called Tailor, for score functions employed in database searching for tandem mass spectrum identification. The Tailor method calibrates the PSM scores relative to the tail of the empirical null distribution of any experimental spectrum, and this is based on our empirical observation that the mean and the standard deviation of the null distribution are correlated. The Tailor method does not require additional computational steps and can be performed based on the observed PSM scores, and its

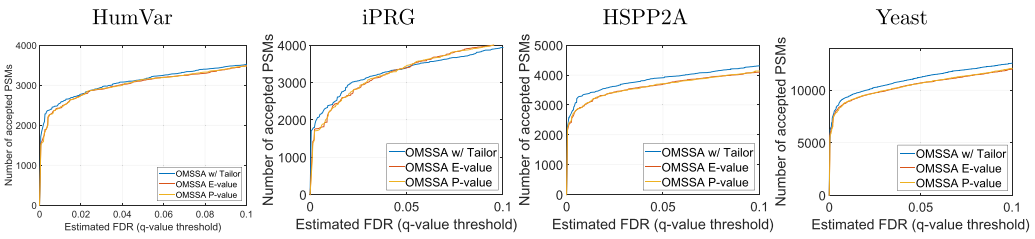


Figure 9. Spectrum annotation results obtained with the OMSSA program. Each plot corresponds to a different data set indicated on top.

https://dx.doi.org/10.1021/acs.jproteome.9b00736
J. Proteome Res. XXXX, XXX, XXX−XXX

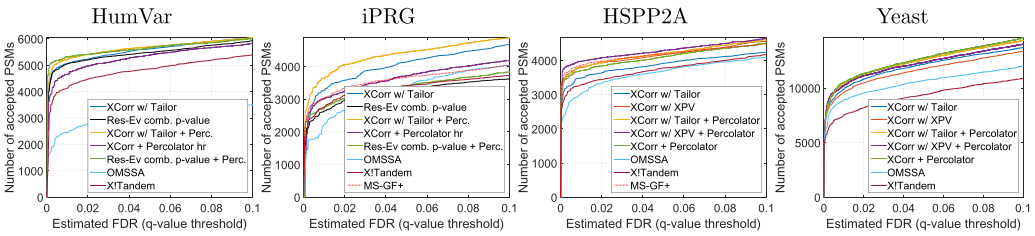


Figure 10. Overall results. Each plot corresponds to a different data set indicated on top. Note that MS-GF+ did not run on HumVar and Yeast data sets.

Table 3. Number of Spectrum Annotations at an FDR Level of 1%[a]

name	HumVar	(%)[d]	iPRG	(%)	HSPP2A	(%)	Yeast	(%)
Match + Tailor + Percolator	5222	100	**3645**	**100**	3818	100	10 335	100
XPV[d] + Percolator	**5320**	**102**	2713	74	**3979**	**104**	10 203	99
XCorr + Percolator	4653	89	3020	83	3705	97	**10 503**	**102**
Match + Tailor	5059	97	3264	90	3483	91	9891	96
XPV[d]	5051	97	2637	72	3788	99	9586	93
XCorr	4433	85	2391	66	2726	71	7957	77
OMSSA	2450	47	2210	61	3025	79	8883	86
X!Tandem	4125	79	2745	75	3282	86	7473	72
MS-GF+[d]	n.a.	n.a.	3036	83	3864	101	n.a.	n.a.

[a]The best performance is marked with boldface. [b]The relative change of the number of annotations compared to the Tailor + Percolator method combination. [c]XPV method was applied to the Res-ev score for high-resolution MS2 data (HumVar, iPRG) and to the XCorr score for the low-resolution MS2 data (HSPP2A, Yeast). [d]MS-GF+ did not run on HumVar and Yeast data sets.

performance is on par with the computationally expensive exact methods; however, Tailor runs around 100 times faster and could save a substantial amount of CPU time and reduce carbon emission. Unfortunately, because Tailor is a heuristic method, it may perform poorly with some score functions in some cases. However, we hope that its simplicity and robustness will make the Tailor method a standard option in many database-searching software programs. The Tailor method is freely available in the open-source Crux mass spectrometry toolkit (http://crux.ms).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.9b00736.

Detailed description of the methods used (Supporting Note S1); additional figures about score distribution (Supporting Note S2) (PDF)

Scripts used to run all methods in our tests (Supporting File 1) (ZIP)

HSPP2A data set (Supporting File 2) (ZIP)

Fasta files used in our experiments (Supporting File 3) (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Attila Kertész-Farkas − *Department of Data Analysis and Artificial Intelligence, Faculty of Computer Science, National Research University Higher School of Economics (HSE), Moscow 109028, Russian Federation*; orcid.org/0000-0001-8110-7253; Phone: +7 (499) 152-07-41; Email: akerteszfarkas@hse.ru

Author

Pavel Sulimov − *Department of Data Analysis and Artificial Intelligence, Faculty of Computer Science, National Research University Higher School of Economics (HSE), Moscow 109028, Russian Federation*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.9b00736

Author Contributions

A.K.-F. designed the experiments, performed data analysis, and wrote the manuscript. P.S. developed scripts and programs and performed data analysis. All authors read and approved the final manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank William Stafford Noble and the Crux development team for hosting the Tailor method in the Crux mass spectrometry toolkit at http://crux.ms.

■ REFERENCES

(1) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198.

(2) Nesvizhskii, A. I.; Aebersold, R. Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. *Drug Discovery Today* **2004**, *9*, 173−181.

(3) Kertész-Farkas, A.; Reiz, B.; Myers, M. P.; Pongor, S. Database searching in mass spectrometry based proteomics. *Curr. Bioinf.* **2012**, *7*, 221−230.

(4) Noble, W. S.; MacCoss, M. J. Computational and statistical analysis of protein mass spectrometry data. *PLoS Comput. Biol.* **2012**, *8*, No. e1002296.

(5) Keich, U.; Noble, W. S. On the importance of well-calibrated scores for identifying shotgun proteomics spectra. *J. Proteome Res.* **2015**, *14*, 1147−1160.

(6) Granholm, V.; Käll, L. Quality assessments of peptide-spectrum matches in shotgun proteomics. *Proteomics* **2011**, *11*, 1086−1093.

https://dx.doi.org/10.1021/acs.jproteome.9b00736
J. Proteome Res. XXXX, XXX, XXX−XXX

(7) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646−4658.
(8) Nesvizhskii, A. I.; Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* **2005**, *4*, 1419−1440.
(9) Keich, U.; Kertesz-Farkas, A.; Noble, W. S. Improved false discovery rate estimation procedure for shotgun proteomics. *J. Proteome Res.* **2015**, *14*, 3148−3161.
(10) Kertesz-Farkas, A.; Keich, U.; Noble, W. S. Tandem mass spectrum identification via cascaded search. *J. Proteome Res.* **2015**, *14*, 3027−3038.
(11) Eng, J. K.; Fischer, B.; Grossmann, J.; MacCoss, M. J. A fast SEQUEST cross correlation algorithm. *J. Proteome Res.* **2008**, *7*, 4598−4602.
(12) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.
(13) Yates, J. R., 3rd; Eng, J. K.; McCormack, A. L.; Schieltz, D. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **1995**, *67*, 1426−1436.
(14) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **2013**, *13*, 22−24.
(15) Diament, B. J.; Noble, W. S. Faster SEQUEST searching for peptide identification from tandem mass spectra. *J. Proteome Res.* **2011**, *10*, 3871−3879.
(16) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **2011**, *10*, 1794−1805.
(17) Granholm, V.; Noble, W. S.; Käll, L. On using samples of known protein content to assess the statistical calibration of scores assigned to peptide-spectrum matches in shotgun proteomics. *J. Proteome Res.* **2011**, *10*, 2671−2678.
(18) Dorfer, V.; Pichler, P.; Stranzl, T.; Stadlmann, J.; Taus, T.; Winkler, S.; Mechtler, K. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J. Proteome Res.* **2014**, *13*, 3679−3684.
(19) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open mass spectrometry search algorithm. *J. Proteome Res.* **2004**, *3*, 958−964.
(20) Klammer, A. A.; Park, C. Y.; Noble, W. S. Statistical calibration of the SEQUEST XCorr function. *J. Proteome Res.* **2009**, *8*, 2106−2113.
(21) Spirin, V.; Shpunt, A.; Seebacher, J.; Gentzel, M.; Shevchenko, A.; Gygi, S.; Sunyaev, S. Assigning spectrum-specific p-values to protein identifications by mass spectrometry. *Bioinformatics* **2011**, *27*, 1128−1134.
(22) Fenyö, D.; Beavis, R. C. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal. Chem.* **2003**, *75*, 768−774.
(23) Kim, S.; Gupta, N.; Pevzner, P. A. Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J. Proteome Res.* **2008**, *7*, 3354−3363.
(24) Kim, S.; Mischerikow, N.; Bandeira, N.; Navarro, J. D.; Wich, L.; Mohammed, S.; Heck, A. J.; Pevzner, P. A. The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. *Mol. Cell. Proteomics* **2010**, *9*, 2840−2852.
(25) Kim, S.; Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* **2014**, *5*, No. 5277.
(26) Howbert, J. J.; Noble, W. S. Computing exact p-values for a cross-correlation shotgun proteomics score function. *Mol. Cell. Proteomics* **2014**, *13*, 2467−2479.
(27) Lin, A.; Howbert, J. J.; Noble, W. S. Combining High-Resolution and Exact Calibration To Boost Statistical Power: A Well-Calibrated Score Function for High-Resolution MS2 Data. *J. Proteome Res.* **2018**, *17*, 3644−3656.
(28) McIlwain, S.; Tamura, K.; Kertesz-Farkas, A.; Grant, C. E.; Diament, B.; Frewen, B.; Howbert, J. J.; Hoopmann, M. R.; Käll, L.; Eng, J. K.; et al. Crux: rapid open source protein tandem mass spectrometry analysis. *J. Proteome Res.* **2014**, *13*, 4488−4491.
(29) Danilova, Y.; Voronkova, A.; Sulimov, P.; Kertesz-Farkas, A. Bias in false discovery rate estimation in mass-spectrometry-based peptide identification. *J. Proteome Res.* **2019**, *18*, 2354−2358.
(30) Pease, B. N.; Huttlin, E. L.; Jedrychowski, M. P.; Talevich, E.; Harmon, J.; Dillman, T.; Kannan, N.; Doerig, C.; Chakrabarti, R.; Gygi, S. P.; Chakrabarti, D. Global analysis of protein expression and phosphorylation of three stages of Plasmodium falciparum intra-erythrocytic development. *J. Proteome Res.* **2013**, *12*, 4028−4045.
(31) Chalkley, R. J.; Bandeira, N.; Chambers, M. C.; Clauser, K. R.; Cottrell, J. S.; Deutsch, E. W.; Kapp, E. A.; Lam, H. H.; McDonald, W. H.; Neubert, T. A.; et al. Proteome informatics research group (iPRG) _2012: a study on detecting modified peptides in a complex mixture. *Mol. Cell. Proteomics* **2014**, *13*, 360−371.
(32) Glatter, T.; Wepf, A.; Aebersold, R.; Gstaiger, M. An integrated workflow for charting the human interaction proteome: insights into the PP2A system. *Mol. Syst. Biol.* **2009**, *5*, 237.
(33) Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **2007**, *4*, 923.
(34) Goloborodko, A. A.; Levitsky, L. I.; Ivanov, M. V.; Gorshkov, M. V. Pyteomics-a Python framework for exploratory data analysis and rapid software prototyping in proteomics. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 301−304.
(35) Levitsky, L. I.; Klein, J. A.; Ivanov, M. V.; Gorshkov, M. V. Pyteomics 4.0: five years of development of a Python proteomics framework. *J. Proteome Res.* **2019**, *18*, 709−714.
(36) Elias, J. E.; Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, *4*, 207.
(37) He, K.; Fu, Y.; Zeng, W.-F.; Luo, L.; Chi, H.; Liu, C.; Qing, L.-Y.; Sun, R.-X.; He, S.-M. A Theoretical Foundation of the Target-Decoy Search Strategy for False Discovery Rate Control in Proteomics. 2015, arXiv:physics/1501.00537. arXiv.org e-Print archive. https://arxiv.org/abs/1501.00537.
(38) Levitsky, L. I.; Ivanov, M. V.; Lobas, A. A.; Gorshkov, M. V. Unbiased false discovery rate estimation for shotgun proteomics based on the target-decoy approach. *J. Proteome Res.* **2016**, *16*, 393−397.
(39) Storey, J. D.; Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9440−9445.
(40) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002**, *74*, 5383−5392.

https://dx.doi.org/10.1021/acs.jproteome.9b00736
J. Proteome Res. XXXX, XXX, XXX−XXX