



Published in final edited form as:

J Proteomics Bioinform. 2010 February 5; 3: 39–47. doi:10.4172/jpb.1000119.

Evaluation of the Consensus of Four Peptide Identification Algorithms for Tandem Mass Spectrometry Based Proteomics

Ruben K. Dagda¹, Tamanna Sultana², and James Lyons-Weiler^{2,3,*}

¹Department of Pathology, University of Pittsburgh, PA

²Bioinformatics Analysis Core, Genomics and Proteomics Core Laboratories, University of Pittsburgh, Pittsburgh, PA

³Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA

Abstract

The availability of different scoring schemes and filter settings of protein database search algorithms has greatly expanded the number of search methods for identifying candidate peptides from MS/MS spectra. We have previously shown that consensus-based methods that combine three search algorithms yield higher sensitivity and specificity compared to the use of a single search engine (individual method). We hypothesized that union of four search engines (Sequest, Mascot, X!Tandem and Phenyx) can further enhance sensitivity and specificity. ROC plots were generated to measure the sensitivity and specificity of 5460 consensus methods derived from the same dataset. We found that Mascot outperformed individual methods for sensitivity and specificity, while Phenyx performed the worst. The union consensus methods generally produced much higher sensitivity, while the intersection consensus methods gave much higher specificity. The union methods from four search algorithms modestly improved sensitivity, but not specificity, compared to union methods that used three search engines. This suggests that a strategy based on specific combination of search algorithms, instead of merely 'as many search engines as possible', may be key strategy for success with peptide identification. Lastly, we provide strategies for optimizing sensitivity or specificity of peptide identification in MS/MS spectra for different user-specific conditions.

Keywords

Proteomics; Peptide identification; Bioinformatics; Consensus approach; Consensus methods; Phenyx

Introduction

Peptide identification from tandem mass spectrometry (MS/MS) data using database search strategies requires the use of commercial or publicly available search algorithms that are used to match MS2 spectra against a protein desired database. The following is a brief workflow that most search engines perform for searching candidate peptides from MS2 spectra. A large raw file containing the experimental spectra obtained by mass spectrometry is submitted to a desired search algorithm (engine). The search algorithm then matches the

© 2010 Dagda RK, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

*Corresponding author: James Lyons-Weiler, 2Bioinformatics Analysis Core, Genomics and Proteomics Core Laboratories, University of Pittsburgh, Pittsburgh, PA, Tel: 412-728-8743, Fax: 412-648-1891, jim@bioinformatics.pitt.edu.

experimental mass spectra with the theoretical spectra generated from *in silico* (e.g., trypsin) digested peptides, a list of candidate peptides is then output with specific scores in the order of score rank. The reported scores depend on the mathematical and algorithmic strategies employed by a specific search engine.

For each search algorithm, a plethora of search parameters can be modified by the user. These parameters include, but are not limited to: the enzyme used for digesting peptides, the number of missed cleavages, amino acid modifications, peptide mass error tolerance, fragment ion mass error tolerance, charge of the parent peptide, and the minimum length of peptides used to match an experimental spectrum. In addition to these different search parameters, the models used to score the match between peptide and an MS2 spectrum vary widely among different search engines. These models include but are not limited to: statistic and probability models (Mascot), stochastic based search engines (Phenyx, SCOPE and Sherenga), descriptive models (Sequest) and interpretive models for database searching (PeptideSearch) reviewed in literature (Sadygov et al., 2004). Too many choices for the search engine and large number of options for the search algorithms can make user's decision making process cumbersome and complicated with respect to which strategies and protocols to employ for identifying peptides.

Mascot, Sequest, and X!Tandem (Perkins et al., 1999; Eng et al., 1994; Craig et al., 2004) are three of the most popular search algorithms available for identifying peptides from MS/MS data. While Sequest scores peptides solely on descriptive parameters and uses correlative matching of peptide fragments in a two-tiered stage process (homology matching), Mascot employs statistical and geometric probability- based scoring methods for scoring and ranking peptides. Unlike the other aforementioned search algorithms, X!Tandem generates a list of identified peptides upon searching for post-translational modifications on high confident peptide identifications in order to improve the confidence (Craig et al., 2004). Most peptide identification searches currently rely on the use of one search engine for a given single large dataset derived from one MS platform. Each search engine shows variability in accuracy, sensitivity and specificity, and performs differently under different conditions. For instance, receiver operating characteristic (ROC) method analysis for one particular study that compared two search algorithms, showed that X!Tandem is more robust for sensitivity and specificity, compared to Mascot, at different mass accuracy (Brosch et al., 2008). In addition, other studies have shown modest differences in the output generated by OMMSA and Sequest mediated searches (Balgley et al., 2007).

Consensus methods that employ more than one search engine are a recent and natural development for bioinformatics approaches to peptide identification. Consensus-based approaches have consistently identified peptides in complex biological mixtures with higher sensitivity and specificity, compared to the use of individual search algorithms. For instance, several studies have shown that consensus methods that employ two search algorithms greatly enhanced mass spectral coverage and specificity, compared to data analysis using one search algorithm (MacCoss et al., 2002; Moore et al., 2002; Chamrad et al., 2004; Resing et al., 2004; Kapp et al., 2005; Rudnick et al., 2005). A consensus method (CM) is a particular logical operation performed on protein lists from specific combinations of search engines after filtering with a specific combination of filter options. Individual methods (e.g., Mascot, Sequest), logically, are a type of CM with a set that contains one member. A consensus set is a set of proteins that results from a consensus method. The multitude of settings and search parameters, different logical operations (union, intersection or single), and a large array of available search engines can create a large number of possible combinations of search engines with different search filter settings. Evaluating the performance of each CM should lead us to improved understanding of the effects our

assumptions have on individual search engines, and creating CMs, thus leading to improved peptide identification.

We have recently reported on the utility of 2310 consensus methods based on three popular search algorithms (Sultana et al., 2009). We observed that the intersection of search engines greatly enhances specificity, but yields a much lower sensitivity. By comparison, consensus methods based on the union of the search engines increased accuracy, sensitivity, and specificity. Further, we noted the potential utility of different combinations for optimizing sensitivity and specificity, depending on the aims of a particular study (Sultana et al., 2009). In this study, we extend our previous studies by including a fourth search algorithm (Phenyx) in consensus based peptide searches in order to determine whether four consensus methods that employ more than three search engines further enhances sensitivity and specificity for the identification of peptides. Phenyx is a search algorithm that relies on the OLAV scoring method for identifying peptides; it is a modification of the heuristic and descriptive approaches employed by Sequest, Mascot or X!Tandem (Magnin et al., 2004; Colinge et al., 2003). We report the performance of these four search algorithms for reliability, sensitivity, and specificity of peptide identification.

Materials and Methods

Dataset

A sample mixture containing 49 human proteins (Sigma-Aldrich, St. Louis, MO) was processed in the Genomics and Proteomics Core Laboratories (GPCL) at the University of Pittsburgh prior to performing tandem mass spectrometry. In brief, the sample mixture was reduced with tris-2-carboxyethyl-phosphine (TCEP), alkylated with methylmethane-thiosulfonate (MMTS), and digested with trypsin (Promega). The ESI-MS and information dependent (IDA) MS/MS spectra were acquired at Thermo (by Research Scientist Tim Keefe) with an LTQ-XL coupled with a nano-LC system (Thermo Scientific, Waltham, MA). The IDA was set so that MS/MS was done on the top three intense peaks per cycle.

Database search

Experimental spectra obtained from the trypsin digested mixture of 49 proteins were searched against the human database, IPI Human v3.48 (71401 sequence entries), for identifying peptides using following four search algorithms: Sequest, Mascot, X!Tandem, and Phenyx. The search parameters employed in this study were kept consistent as in our previous study (Sultana et al., 2009). In brief, the search parameters for searching candidate peptides were: precursor ion tolerance: 2 Da, fragment ion tolerance: 1 Da, variable modification: MMTC on cysteine, and oxidation on methionine. All searches for the protein dataset were performed at GPCL at the University of Pittsburgh. The parameters are so chosen to be optimal from our previous experiences with LTQ-XL data sets.

Merging the data

All four search engines used in this study have different strategies for assigning scores and ranking candidate peptides. Due to the facility added in filtering methods under different filtering assumptions, we used Scaffold v.2.0.1, (Proteome Software Inc.) as a common platform; all search results derived from Mascot, Sequest, Phenyx and X!Tandem were merged into a single file, and assigned rankings and scores to candidate peptides, using one mathematical model. In brief, the files containing the database search results derived from Mascot, Sequest, X!Tandem, and Phenyx were imported into Scaffold. The software then merged the peptide lists identified by all the four search algorithms, re-scored, and re-ranked. Scaffold uses PeptideProphet and ProteinProphet, that employ Bayesian statistics to combine the probability of identifying spectra with the probability that all search methods

agree with each other (Keller et al., 2002; Nesvizhskii et al., 2003). We named this Scaffold-generated file as MSXP-50.

Generation of consensus methods

As described before, a specific combination of settings and search algorithms is defined in this study as a specific “consensus method”. Two-hundred and ten consensus methods were generated in Scaffold from MSXP-50 (MSXP-50 is the compiled results file from the same data set) by varying the settings of three different criteria, which filter candidate peptides based on specific user-defined threshold values. The three filter settings are defined as follows:

1. The minimum protein probability is defined as the probability that the identification of a specific candidate protein is correct. We varied this Scaffold filter setting from 20–99% in this study in order to generate consensus methods from MSXP-50.
2. The minimum number of peptides is defined as the number of unique peptide sequences that match a given candidate protein. We varied this Scaffold filter setting from 1–5 peptides in this study in order to generate consensus methods from MSXP-50.
3. Lastly, the minimum peptide probability is defined as the probability that a given candidate peptide matches at least one experimental MS/MS spectrum. This Scaffold filter setting was varied from 0–95% in order to generate consensus methods from MSXP-50.

Generation of consensus sets

First of all, twenty-six consensus sets, peptide lists from specific combinations of search engines, including individual search engines (O), intersections (I) and unions (U) of the search results among four search engines were generated using a program written in python. The consensus sets generated were: MO (Mascot only), SO, XO, PO, MSI (intersection between Mascot and Sequest), MXI, SXI, PSI, PMI, PXI, MSXI, PMXI, PMSI, PSXI, MXSPI, MSU (union between Mascot and Sequest), MXU, SXU, PSU, PMU, PXU, PXSU, PMSU, PMXU, MSXU, and MXSPU. Finally, 5,460 total consensus methods were created from 26 consensus sets and 210 consensus methods. For example, one specific consensus method is the union of Mascot, Phenyx, and Sequest (termed MSPU) containing a 20% minimum protein probability, 3 minimum peptides and 20% minimum peptide probability. This consensus method represents one out of 5,460 consensus methods analyzed. Peptide lists derived from each of these consensus methods are termed as ‘consensus sets’, therefore producing a total of 5460 consensus sets (results from 5460 consensus methods) for this study.

Calculating sensitivity and approximation of specificity

We modified and expanded the code of our Python program employed in our previous study (Sultana et al., 2009) to perform sensitivity (SN) and apparent specificity (SP*) calculations of consensus methods that contained up to four search algorithms. Prior to running calculations, the Python program asks users to input the following information: the number and type of search engines desired, the text file containing a “true positive” list of expected peptides, a text file containing a list of all the names of 210 text files that contain peptide lists (generated by Scaffold by altering each of three Scaffold filter settings as described above), and the total number of false positives identified (263 total false positives identified for this study under least stringent settings). In order to create a list of true positive peptides, the duplicate and false positive peptides were discarded from the MSXP peptide list, which was generated by applying the least stringent combination of filter setting values (20%

protein probability, 1 peptide and 0% peptide probability). Conversely, the total number of FP was calculated by counting the total number of false positive peptides that were initially discarded from the true positive list. These true positive and false positive values were employed by the Python program to specifically calculate the TP and FP rates for each specific consensus method. At the end of each calculation, Python program outputs the following result: total number true positives (TP), false positives and false negatives (FN), as well as sensitivity and apparent specificity for each consensus method.

It is important to note that the false positive rate is estimated as an indirect approximation of specificity (apparent specificity, SP^*), as previously described (Sultana et al., 2009). This approximation is useful because the total universe of true negatives is unknown, and will likely never be fully known. While our knowledge of which proteins and peptides do exist increases, as additional known protein sequences are added to the human database and the number of true negatives changes, the use of SP^* provides an approximation of specificity that is best interpreted in the context of each comparative evaluative study. The apparent specificity is dependent on the type of the biological mixture, the mass spectrometry used (and its settings), and the number and type of proteins in each dataset; thus it cannot be compared across different studies. Nevertheless, ROC can gauge the performance for SN and SP^* . ROC allow us to compare different methodologies and protocols in order to optimize peptide searches in a controlled setting where the protein complement of a sample is known.

Results

Our results show that among consensus methods employing upto four search engines, modest improvement in sensitivity was observed in union consensus methods while slight improvement in specificity was observed in intersection consensus based methods for MS/MS spectra derived from complex standard protein mixture.

ROC plots of all consensus methods analyzed-

By comparing the list of known proteins in the mixture with the list of true positive peptides identified, we found that a total 36 of 49 (73% of proteins) proteins were detected by combined four search engine results from the same LTQ-XL data set; thus limiting the universe of true positive peptides to 36 proteins in our study. ROC method plots, graphical representations of paired values of SN, and one minus apparent specificity ($1-SP^*$) were generated to study the performance of each of the 5460 consensus methods. Figure 1 shows ROC method plots for all 5,460 consensus methods, which were generated by combining 26 consensus sets (peptide lists from specific search engine combinations) with the 210 Scaffold consensus and filter setting combinations. In summary, Figure 1 shows that many union consensus methods, which use at least three search engines (MSXU and MSXPU) yielded the highest SN, and were followed closely by the union consensus methods MXPU, MXU, and MSU (Figure 1A and 1D).

Moreover, these results agree with our previously published observations that the union consensus methods, which contained two or more search engines, generally gave higher sensitivity values but lower SP^* scores, compared to the intersection consensus methods or to individual search engine methods (Sultana et al., 2009). Conversely, the intersection consensus methods gave higher specificity values than the union consensus methods (Figure 1C and 1D). A comparison of the performance of each individual search engine method shows that Mascot performed the best in terms of specificity and sensitivity, followed by Sequest as the second best while Phenyx performed the worst in terms of sensitivity, but showed comparable specificity with the other three search engines (Figure 1B). As a general trend, individual methods performed worse for sensitivity and specificity compared to the

unions based consensus methods, but only performed modestly better for sensitivity compared to the intersection consensus methods. Among intersection consensus methods, we found that the intersection of Mascot and Sequest (MSI) performed optimally for both SN and SP* (Figure 1A and 1C).

To detect overall trends in both SN and SP* of consensus methods, we averaged the SN and SP* values for all methods belonging to a specific consensus set, regardless of the Scaffold filter setting used. In Figure 2, ROC plots show the averages of all, individual, intersection, and the union for each of the twenty-six consensus sets. Figure 2 shows that the average of MXSPU and MSXU performed the best in terms of SN, compared to the average of the other 24 consensus sets. Moreover, the average of MSU and MSPU ranked second best for SN, followed by SXU and SPXU. Of all intersection consensus methods analyzed in this study, the average of MSI yielded highest SP* value. Interestingly, we observed that the union consensus methods that contained Phenyx showed a modest improvement in specificity, but gave much lower SN values compared to union consensus methods containing search engines that excluded Phenyx. This observation held true regardless of whether the intersections or unions were compared in this figure (Figure 2, compare MXU vs. MPU, MSU vs. MPU, or MSXU vs. MSPU, MSI vs. MPI).

Interestingly, we found that the MSXPI gave a similar SP* value compared to MSXI, suggesting that a maximum SP* is achieved when three or more search engines are used, at least as shown for this study. Overall, these observations suggest that a particular combination of search engines in consensus sets may enhance sensitivity and/or specificity, while other combinations may result in a counter-productive effect. One likely explanation for this observation is that the particular strengths for each algorithm are enhanced in consensus sets that employs more than one search engine. However, these strengths are overcome by the weaknesses of each engine's mathematical model used to search peptides when more than three search algorithms are used. To compare and understand the overall performance of consensus methods, we ranked the top 50% of consensus methods based on an aggregate score of the average specificity and sensitivity values as shown in Table 1. In summary, the union consensus methods that contained results from more than two search engines produced the highest aggregate scores, with MSXPU being the top ranked consensus method, followed by MSXU and MSPU. Interestingly, we observed that the union of Mascot and Sequest (MSU) scored higher than several union consensus methods that contained three search engines (i.e., SXPU and MXPU), suggesting that it is the specific combination, and not the number of search engines, that may provide optimal SN and SP* results.

The performance of all consensus methods that contained Phenyx is compared in Figure 3. In summary, we observed that the union consensus methods containing Phenyx produced higher average SN values compared to their corresponding intersection consensus methods. This result is intuitive, and confirms our previous observations that the union consensus methods yield higher SN values, regardless of the search engines combined in consensus-based searches (Figure 3A and C–D). However, considering overall accuracy, no consensus methods containing Phenyx with any two others improved on the MSXU method. This is due to a loss of specificity that can be attributed to the inclusion of Phenyx in the consensus trios or four way combinations relative to the MSXU result. In addition, a head-to-head comparison of the average SN and SP* of individual search engines produced the following ranking for best to worst sensitivity: Mascot, Sequest, X!Tandem, and Phenyx, which agrees with the observations seen in Figure 2. On the other hand, Phenyx and X!Tandem modestly performed better than Mascot and Sequest in terms of SP* (Figure 3B).

ROC plots of the average specificity and 1-SP* values of consensus methods with a specified minimum protein probability

Figure 4 shows ROC method plots of the average specificity and sensitivity for all, individual search engine, intersection, and union consensus methods at a certain average minimum protein probability. As a general trend, we observed that SN above 0.75 and 1-SP* below 0.25, an optimal SN/1-SP* range for identifying candidate peptides, was observed for the union consensus methods that contained a minimum protein probability of 80–99% (Figure 4D). Only a few methods of individual engines containing a minimum protein probability of 80–90% achieved this optimal range (Figure 4B). Conversely, the union consensus methods containing a 99.0% minimum protein probability considerably dropped in sensitivity while significantly gained in specificity (Figure 4D). The sensitivity and 1-SP* values of the intersection consensus methods performed similarly, regardless of the minimum protein probability used as a search criteria with only a few consensus methods at 80–90% minimum protein probability showing slightly higher sensitivity values above 0.50 (Figure 4C).

ROC plots of the average specificity and 1-SP* values of consensus methods with a specified minimum number of peptides

Figure 5 shows ROC plots of the average sensitivity and specificity for all, individual, intersection and union consensus methods containing a specified minimum number of peptides that matched a candidate protein. In agreement with our previous study (Sultana et al., 2009), consensus methods that contained 1 minimum peptide as a search filter gave the highest sensitivity but the least specificity. On the other hand, we observed that the individual search engine, the intersection and the union consensus methods containing 2 peptides gave an optimal balance for both specificity and sensitivity, whereas consensus methods containing 3–5 minimum peptides caused both a significant loss in specificity and sensitivity (Figure 5 A–D).

ROC plots of the average specificity and 1-SP* values of consensus methods with a specified minimum peptide probability

Figure 6 shows ROC plots for the average sensitivity and specificity values of consensus methods that contained a specified minimum peptide probability. In general, the individual search engine or the union consensus methods that contained a 0–50% minimum peptide probability gave an optimal performance for both sensitivity and specificity (Figure 6A–B and D), while a decrease in performance was observed when the minimum peptide probability was increased beyond 80%. As expected, the intersection consensus methods fared much worst in sensitivity compared to the union consensus methods and varying minimum peptide probability did not have much effect on the sensitivity and the specificity of the intersection consensus methods (Figure 6C).

Discussion

There are many factors to consider for obtaining a highly reliable and accurate high-throughput analysis of MS/MS data derived from complex biological mixtures. Among some of these factors include a proper experimental design, the quality of samples and of the metadata. More importantly, it is crucial to use appropriate bioinformatics tools to analyze MS/MS data in order to increase the reproducibility of data analysis. In the past, bioinformatics analysts have relied on the use of one of many available search algorithms/engines that use a variety of mathematical strategies to match experimental spectra with a database of theoretical spectra, to rank candidate peptides and assess significance of a matched peptide. We (and others) have shown that consensus-based approaches for identifying peptides in MS/MS data provides improvement over individual methods;

moreover, the selection of a specific combination of a set of search engines can increase the flexibility of the strategies available for peptide searches. We have shown that consensus methods that harness the computational power of multiple search engines enhances the sensitivity and specificity of peptide identification in MS/MS data derived from proteins found in complex biological mixtures. Consensus methods that employ the union based approach that employ the output from a combination of at least three search engines yielded optimal sensitivity and specificity results regardless of the size of the dataset or sample size (10 proteins vs. 49 proteins in a complex mixture) (Sultana et al., 2009).

Prior to this study, we hypothesized that consensus methods that employ four search engines would enhance sensitivity and specificity of peptide identification compared to consensus methods that use fewer search algorithms. However, contrary to our initial hypothesis, we found that the union of four search algorithms (Mascot, Sequest, X-Tandem! and Phenyx) does not enhance specificity and only modestly improves sensitivity compared to consensus methods that use two to three search engines (Table 1, Figure 1 and Figure 2). It is conceivable that the union of more than three search engines that employ multiple statistical strategies for scoring and ranking peptides may not necessarily translate into a lower false positive rate or higher identification of true positives. One of our important findings is that that a specific combination of search algorithms should be considered as opposed to the 'highest possible number of search engines' as a strategy for maximizing the accuracy of peptide identification while minimizing false positives. It is conceivable that the advantages of a particular scoring method used by a particular search engine outweighs the disadvantages of other search engines in consensus methods while increasing the number of search algorithms to four becomes counter-productive.

Interestingly, a head-to-head comparison of the four individual search engines analyzed in this study showed that Mascot outperformed all the four search engines, while Phenyx fared the worst. This was an unexpected result since Phenyx employs the OLAV algorithmic scoring strategy, a statistical method considered to be a significant upgrade of search engines that employ stochastic/probability models (i.e., SCOPE). In addition, one bioinformatics study showed that Phenyx outperformed Mascot in sensitivity regardless of the MS/MS techniques used in the study (Ion-Trap and Q-TOF) (Colinge et al., 2003). Another study showed Sequest is superior to Mascot or X!Tandem for identification of proteins in blood samples. Some possible explanations for this apparent discrepancy in results across studies is that the performance for each search engine depends on a variety of conditions including the search parameters, the MS/MS technique, the abundance of proteins in a biological mixture, the database used for comparing the experimental MS/MS spectra, and search parameters used (Kapp et al., 2005).

Overall, this study supports our previous report findings that the union but not the intersection of peptide searches derived from multiple search engines in consensus methods yield better sensitivity while the intersections produce better specificity. It was found in this study that intersection consensus method of four search engines modestly improved apparent specificity. As a "prescriptive" strategy for performing peptide searches using consensus based methods, our studies give further credence to the notion that consensus methods that employ a minimum number of 2 peptides, a minimum protein probability of 80–99.9% percent or a minimum peptide probability of 0–50% may yield the best sensitivity and specificity results regardless of the search engines used in a given study. In our case, the inclusion of Phenyx in our consensus based searches does not alter this strategy. Ideally, we recommend consensus methods, or other protocols for that matter, that yield an optimal specificity and sensitivity results within a range of SN equal or more than 0.75 and 1-SP* values of less than 0.25. Our studies identified only two consensus methods (MXSPU and MSXU) produced very high sensitivity and apparent specificity within this range (SN \geq 0.75

and $1-SP^* \leq 0.25$) which underscores the need to find more protocols and consensus methods using other search engines that may fulfill these criteria (Table 1). Indeed, further studies using different MS/MS techniques (e.g. MALDI-TOF or Ion-Trap), complex biological mixtures (e.g., plasma, cerebrospinal fluid), and more search engines should be carried out in order to determine whether the consensus methods that we identified in this study that gave high SN and SP^* is applicable in other experimental conditions. Alternatively, the findings in this study call for more future studies of consensus methods that employ other search engines such as Spectrum Mill (Agilent Technologies, Santa Clara, California), SCOPE, a two stage stochastic technique for scoring peptides, Paragon, a novel search engine that employs temperature sequence values and other probability estimates, which have been previously shown to perform well in a very large search space (Bafna et al., 2001; Shilov et al., 2007).

Limitations and future directions

Overall, our results validate our previous findings that consensus methods that combine multiple search engines produce optimal ROC characteristics compared to single search engines (Sultana et al., 2009). However, the unions of consensus methods that include Phenyx as a fourth search engine only modestly improved sensitivity scores while having no effect on specificity compared to consensus methods containing three search engines. Based on our results of this study, we devise and report on strategies for obtaining optimal sensitivity and specificity of peptide identification.

One must bear in mind that the universe of false positives and true negatives increases as more proteins are added to the human database on a daily basis which eventually alters the apparent specificity in any given study. While this caveat makes comparison of the false positive rate less tractable across experiments, it is still plausible to compare the sensitivity of different consensus methods that employ the same number of search engines. Further, we note that this is one dataset based on a sample of the entire proteome; further, we note that ours was not a complex mixture, making the sample more ideal than realistic. We note that we used the default search settings within each individual search engine. It is possible and perhaps likely that optimization of individual search engines prior to the consideration in the consensus setting may further improve the accuracy, sensitivity, specificity and lability of individual consensus methods. Finally, we note that weighted consensus methods may allow the inclusion of results from search engines that may be weak in, say, SP^* but add significant value in terms of increased SN, or vice versa. A good deal of further empirical comparative evaluation of consensus methods for peptide identification is needed.

Acknowledgments

This research was made possible by Grant Number 1 UL1 RR024153 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health.

References

1. Bafna V, Edwards N. SCOPE: a probabilistic model for scoring tandem mass spectra against a peptide database. *Bioinformatics* (Oxford, England). 2001; 17:S13–S21. » CrossRef » PubMed » Google Scholar.
2. Balgley BM, Laudeman T, Yang L, Song T, Lee CS. Comparative Evaluation of Tandem MS Search Algorithms Using a Target-Decoy Search Strategy. *Mol Cell Proteomics*. 2007; 6:1599–1608. » CrossRef » PubMed » Google Scholar. [PubMed: 17533222]
3. Brosch M, Swamy S, Hubbard T, Choudhary J. Comparison of Mascot and X!Tandem performance for low and high accuracy mass spectrometry and the development of an adjusted mascot threshold.

- Mol Cell Proteomics. 2008; 7:962–970. » CrossRef » PubMed » Google Scholar. [PubMed: 18216375]
4. Chamrad DC, Körting G, Stühler K, Meyer HE, Klose J, et al. Evaluation of algorithms for protein identification from sequence databases using mass spectrometry data. *Proteomics*. 2004; 4:619–628. » CrossRef » PubMed » Google Scholar. [PubMed: 14997485]
 5. Colinge J, Masselot A, Giron M, Dessingy T, Magnin J. OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics*. 2003; 3:1454–1463. » CrossRef » PubMed » Google Scholar. [PubMed: 12923771]
 6. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*. 2004; 20:1466–1467. » CrossRef » PubMed » Google Scholar. [PubMed: 14976030]
 7. Eng JK, McCormack AL, Yates JR III. 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. » CrossRef » PubMed » Google Scholar.
 8. Kapp EA, Schutz F, Connolly LM, Chakel JA, Meza JE, et al. An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity of analysis. *Proteomics*. 2005; 5:3475–3490. » CrossRef » PubMed » Google Scholar. [PubMed: 16047398]
 9. Keller A, Nesvizhskii AI, 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. » CrossRef » PubMed » Google Scholar. [PubMed: 12403597]
 10. MacCoss MJ, Wu CC, Yates JR III. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal Chem*. 2002; 74:5593–5599. » CrossRef » PubMed » Google Scholar. [PubMed: 12433093]
 11. Magnin J, Masselot A, Menzel C, Colinge J. OLAV-PMF: a novel scoring scheme for high-throughput peptide mass fingerprinting. *J Proteome Research*. 2004; 3:55–60. » CrossRef » PubMed » Google Scholar. [PubMed: 14998163]
 12. Moore RE, Young MK, Lee TD. Qscore: an algorithm for evaluating SEQUEST database search results. *J Am Soc Mass Spectrom*. 2002; 12:378–386. » CrossRef » PubMed » Google Scholar. [PubMed: 11951976]
 13. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem*. 2003; 75:4646–4658. » CrossRef » PubMed » Google Scholar. [PubMed: 14632076]
 14. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999; 20:3551–3567. » CrossRef » PubMed » Google Scholar. [PubMed: 10612281]
 15. Resing K, Meyer-Arendt K, Mendoza AM, Aveline-wolf LD, Jonscher KR, et al. Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. *Anal Chem*. 2004; 76:3556–3568. » CrossRef » PubMed » Google Scholar. [PubMed: 15228325]
 16. Rudnick PA, Wang Y, Evans E, Lee CS, Balgley BM. Large scale analysis of MASCOT results using a Mass Accuracy-based Threshold (MATH) effectively improves data interpretation. *J Proteome Research*. 2005; 4:1353–1360. » CrossRef » PubMed » Google Scholar. [PubMed: 16083287]
 17. Sadygov RG, Cociorva D, Yates JR III. Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nature Methods*. 2004; 1:195–202. » CrossRef » PubMed » Google Scholar. [PubMed: 15789030]
 18. Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, et al. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol Cell Proteomics*. 2007; 6:1638–1655. » CrossRef » PubMed » Google Scholar. [PubMed: 17533153]
 19. Sultana T, Jordan R, Lyons-Weiler J. Optimization of the use of consensus methods for the detection and putative identification of peptides via mass spectrometry using protein standard mixtures. *J Proteomics Bioinform*. 2009; 2:263–273. » CrossRef » PubMed » Google Scholar.

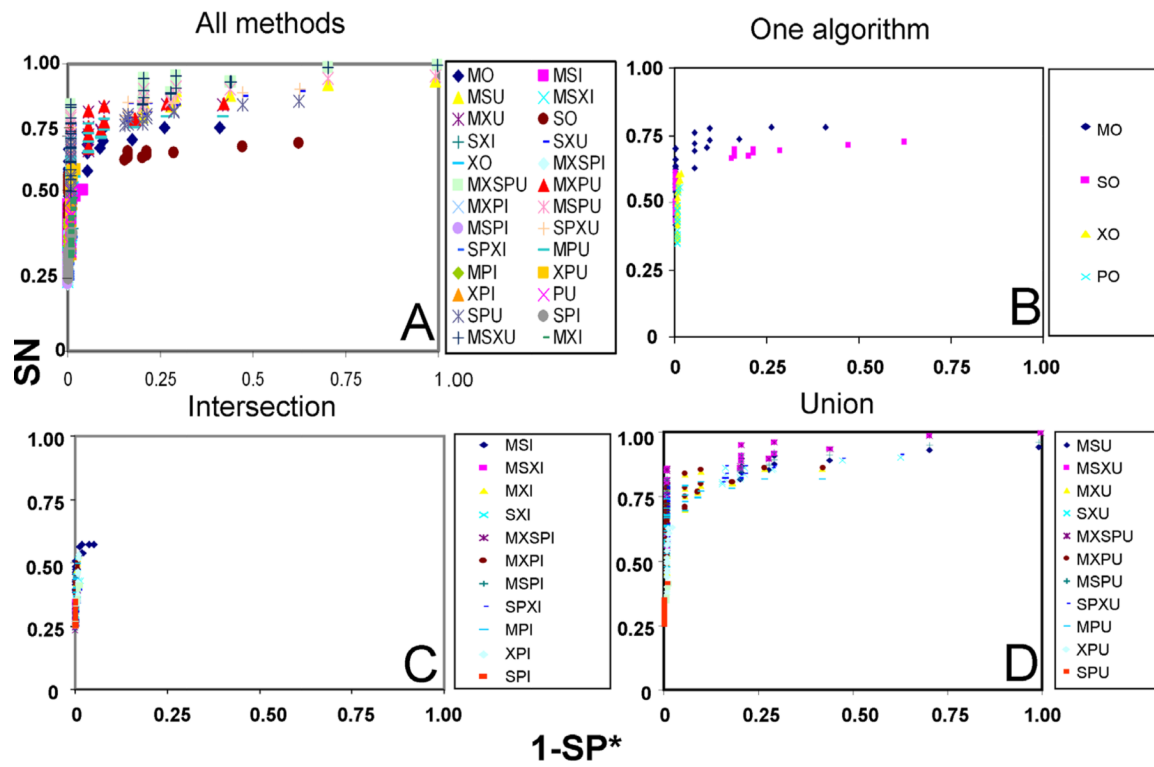


Figure 1. ROC method plots of all (5460) consensus methods analyzed in this study. ROC methods (SN vs. 1-SP*) were plotted for A) all consensus methods, B) individual (O), C) intersection (I), and D) union (U) consensus methods analyzed. The following search engines were employed: Mascot (M), Sequest (S), X!Tandem (X) and Phenyx (P).

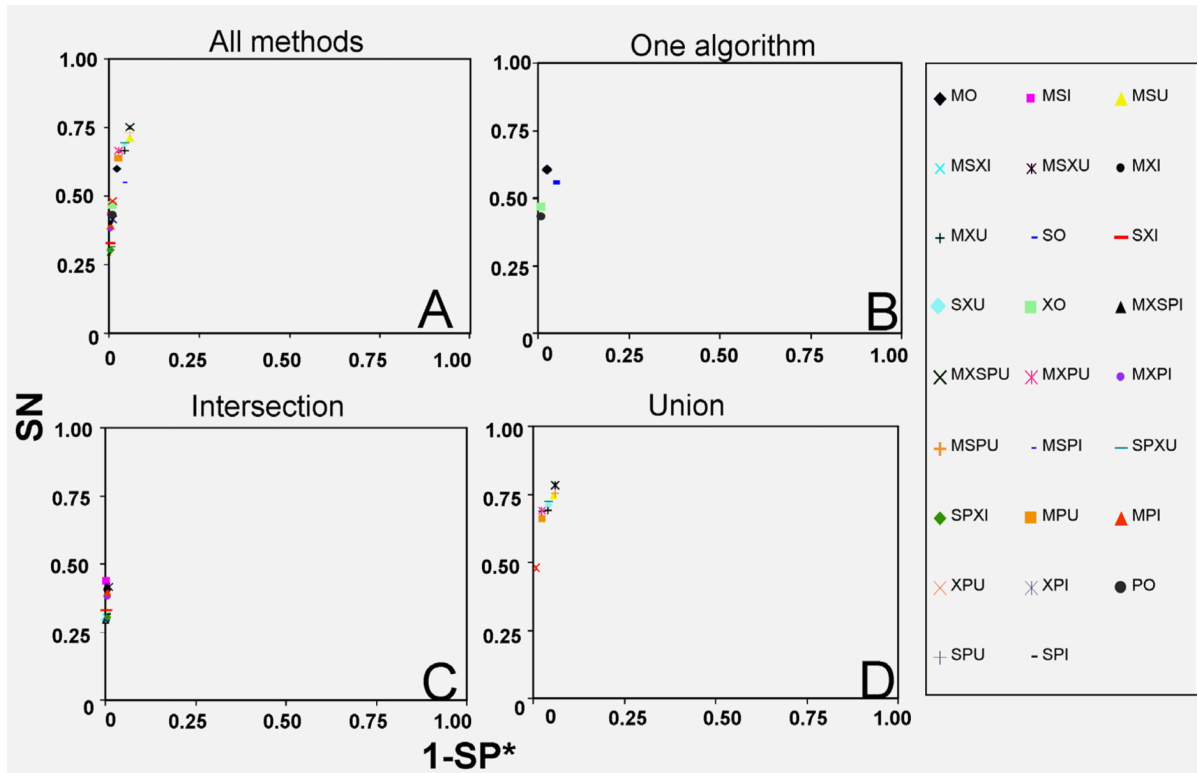


Figure 2. ROC method plots of the average SN and SP* of all consensus methods. ROC method plots (SN vs. 1-SP*) of average of A) all consensus methods, B) individual (O), C) intersection (I), and D) union (U) consensus methods that employed a combination of the following search engines: Mascot (M), Sequest (S), X!Tandem (X) and Phenyx (P).

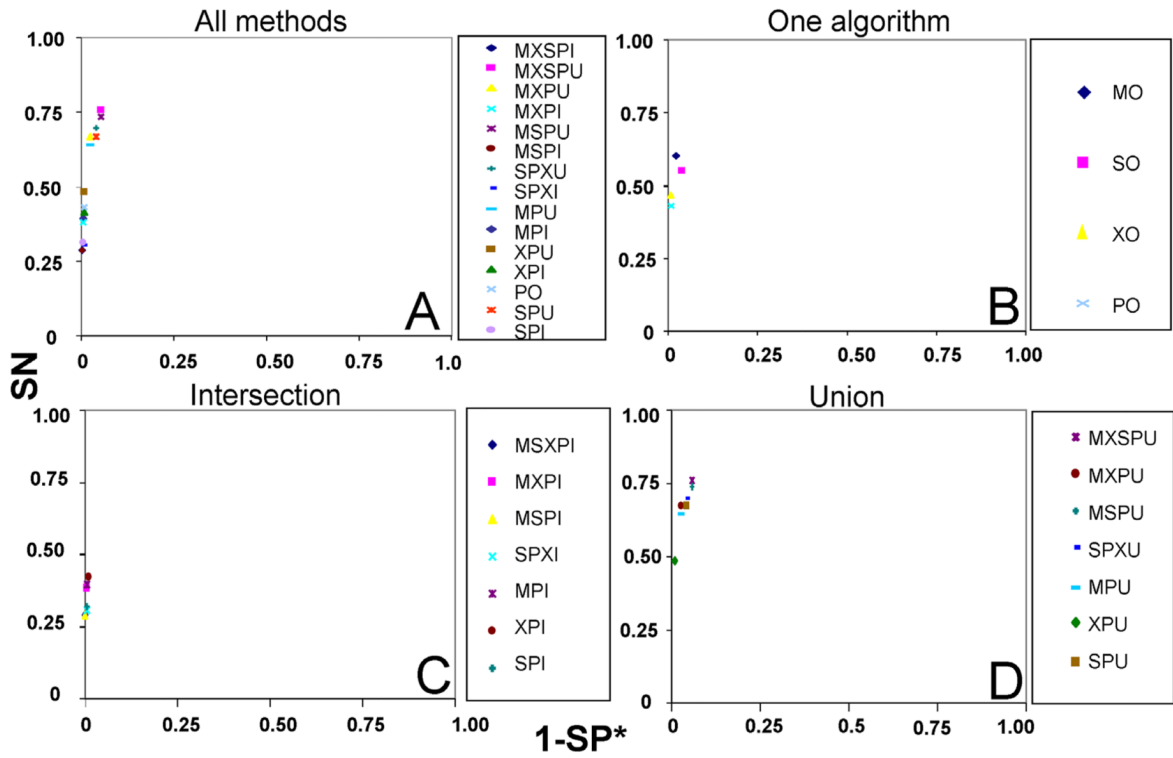


Figure 3. ROC method plots of average SN and SP* of all consensus methods containing Phenyx (P). ROC method plots (SN vs. 1-SP*) show the average of A) all consensus methods, B) individual (O), C) intersection (I), and D) union (U) consensus methods that employed a combination of search engines that contained Phenyx.

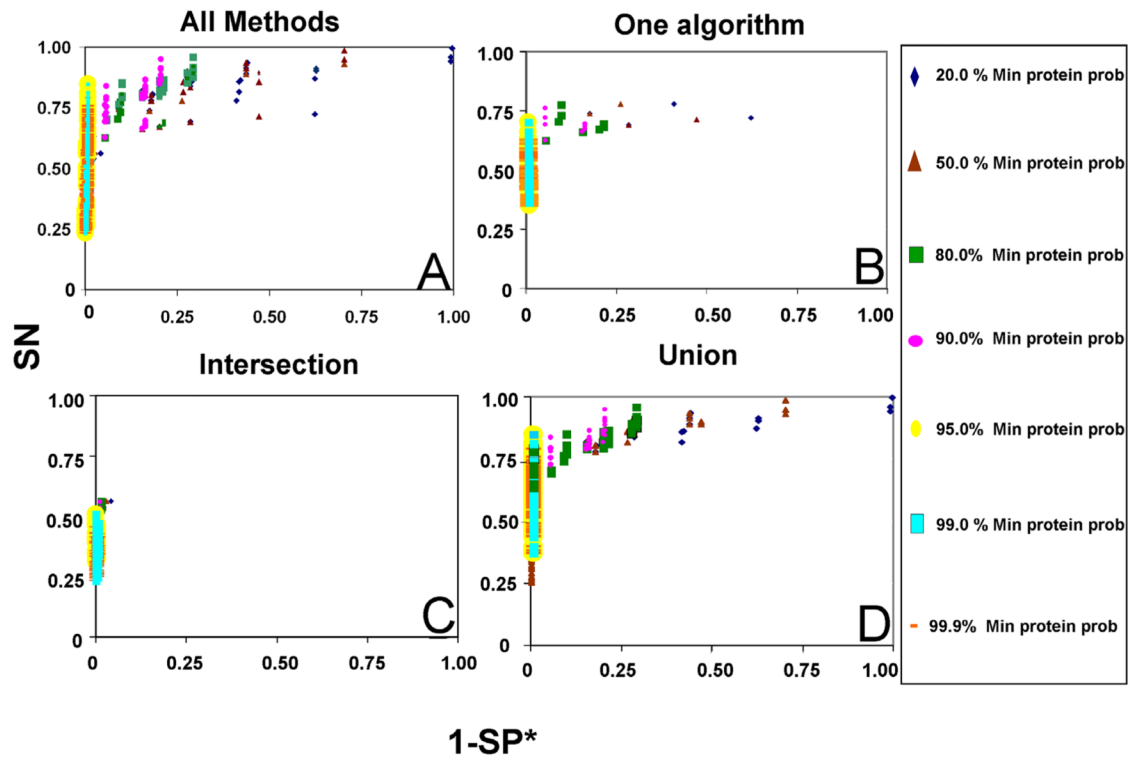


Figure 4.

ROC method plots (SN vs. 1-SP*) of average minimum protein probability of all consensus methods tested. ROC characteristics were plotted for A) all consensus methods, B) individual (O), C) the intersection (I), and D) the union (U) consensus methods filtered at a specified average minimum protein probability. The minimum protein is the probability that a protein's identification is correct. The minimum protein probability values used were 20%, 50%, 80%, 90%, 95%, 99%, and 99.9%.

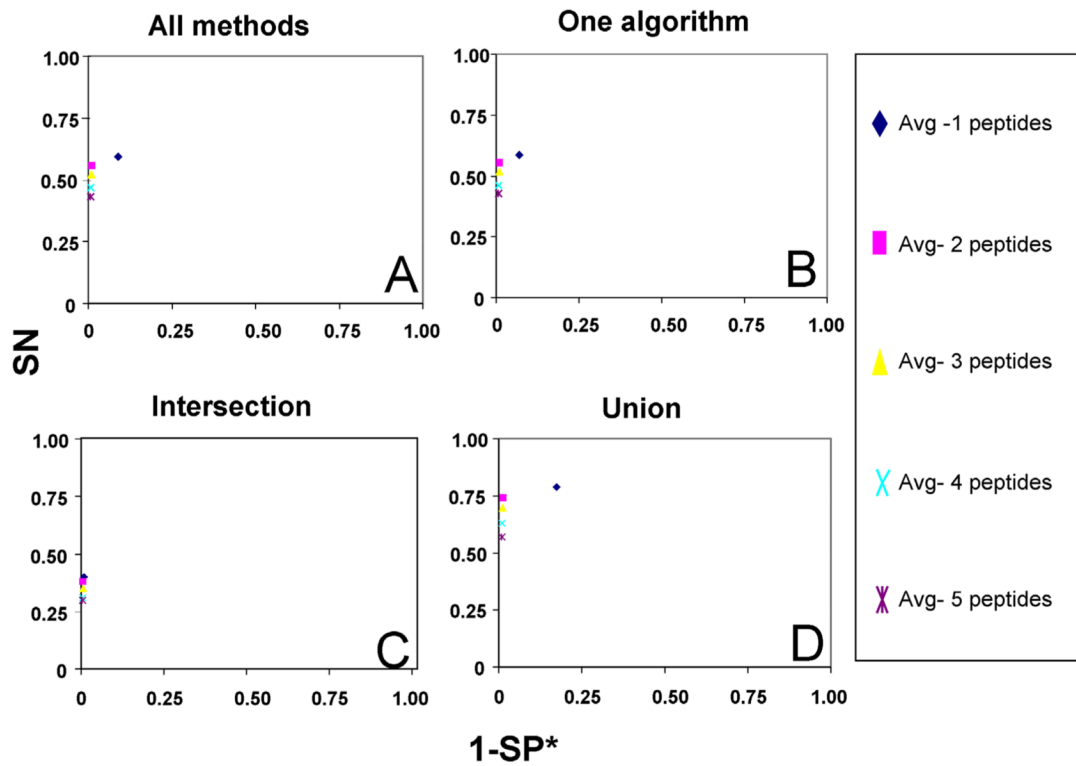


Figure 5.

ROC method plots (SN vs. 1-SP*) of average minimum number of peptides for all consensus methods tested. ROC characteristics were plotted for A) all consensus methods, B) individual (O), C) intersection (I), and D) the union (U) consensus methods filtered at a set specified average number of peptides. The minimum number of peptides values used as a search criteria were 1, 2, 3, 4, and 5 peptides.

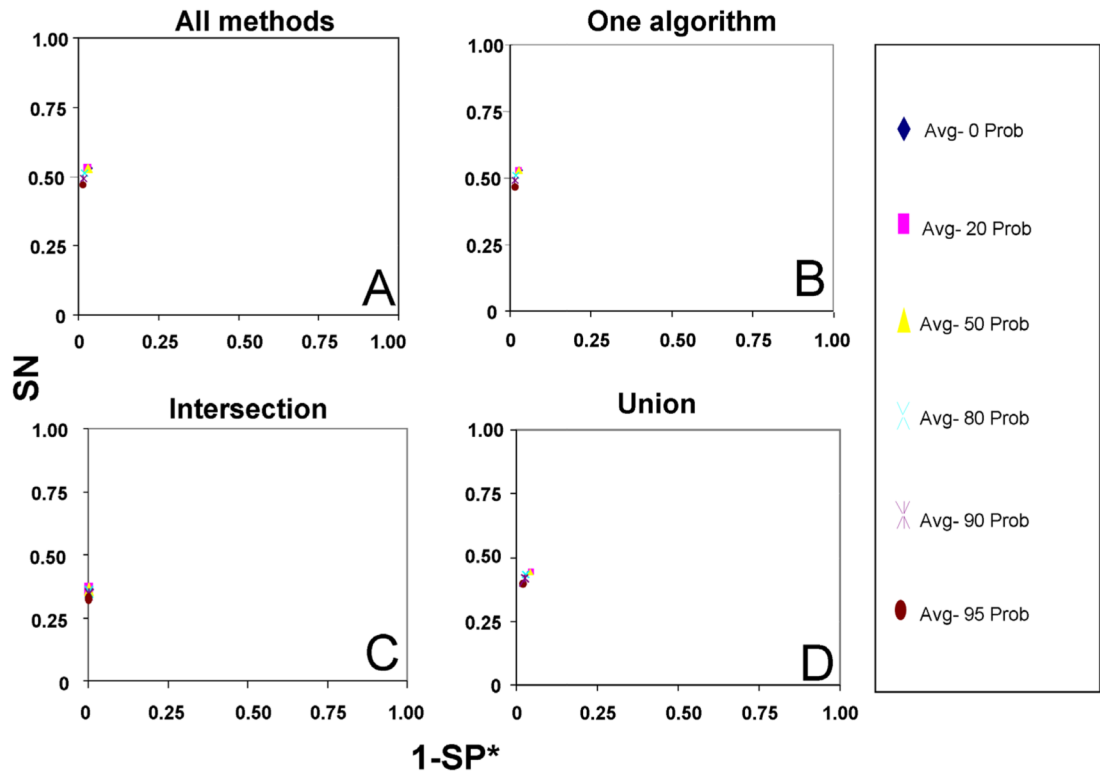


Figure 6.

ROC method plots (SN vs. 1-SP*) of average minimum peptide probability of all consensus methods tested. ROC characteristics were plotted for A) all consensus methods, B) individual (O), C) the intersection (I), and D) the union (U) consensus methods filtered at a set average minimum protein probability used as a search constraint. The minimum peptide probability values used were 0%, 20%, 50%, 80%, 90%, and 95%.

Table 1

An overall performance of consensus methods used in this study. The top 50% of consensus methods were ranked based on an aggregate score of SN and SP*. Note that MXSPU gave the highest aggregate score closely followed by MXSU.

Method	SN	SP	SN + SP
MXSPU	0.755	0.946	1.703
MSXU	0.755	0.946	1.702
MSPU	0.734	0.946	1.680
MSU	0.719	0.946	1.665
SXPU	0.696	0.961	1.957
SXU	0.690	0.961	1.651
MXPU	0.669	0.976	1.644
MXU	0.664	0.976	1.640
SPU	0.660	0.960	1.620
MPU	0.640	0.976	1.616
MO	0.601	0.980	1.580
SO	0.550	0.960	1.510
XPU	0.480	0.990	1.470
XO	0.460	0.990	1.450
PO	0.431	0.997	1.428