Published in final edited form as:

J Proteome Res. 2018 September 07; 17(9): 3325–3331. doi:10.1021/acs.jproteome.8b00594.

Spectral Library Search Improves Assignment of TMT Labeled MS/MS Spectra

Jianqiao Shen¹, Vishwajeeth Reddy Pagala², Alex Michael Breuer³, Junmin Peng^{2,4,5,*}, Bin Ma^{1,*}, and Xusheng Wang^{2,*}

¹Department of Computer Science, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

²St. Jude Proteomics Facility, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

³Department of Information Services, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

⁴Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

⁵Department of Developmental Neurobiology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

Abstract

Tandem mass tag (TMT)-based liquid chromatography-tandem mass spectrometry (LC-MS/MS) is a proven approach for large-scale multiplexed protein quantification. However, the identification of TMT-labeled peptides is compromised by the labeling during traditional sequence database search. In this study, we aim to use spectral library search to increase the sensitivity and specificity of peptide identification for TMT-based MS data. Compared to MS/MS spectra of unlabeled peptides, the spectra of TMT-labeled counterparts usually display intensified *b* ions, suggesting that TMT labeling can alter product ion patterns during MS/MS fragmentation. We compiled a human TMT spectral library of 401,168 unique peptides of high quality from millions of peptide-spectrum matches in tens of profiling projects, matching to 14,048 non-redundant proteins (13,953 genes). A mouse TMT spectral library of the similar size was also constructed. The libraries were subsequently appended with decoy spectra to evaluate false discovery rate, which was validated by a simulated null TMT dataset. The performance of library search was further optimized by removing TMT reporter ions and selecting an appropriate library construction method. Finally, we searched a human TMT dataset against the spectral library to demonstrate that the spectral library outperformed the sequence database. Both human and mouse TMT libraries were made publicly available to the research community.

*Corresponding authors: Xusheng Wang, xusheng.wang@stjude.org, Bin Ma, binma@uwaterloo.ca; Junmin Peng, junmin.peng@stjude.org.

CONFLICT OF INTEREST DISCLOSURE

The authors declare no competing financial interests.

Keywords

spectral library search; peptide identification; TMT; database search; target-decoy strategy; proteomics

INTRODUCTION

Peptide identification is one of the cornerstones in liquid chromatography-mass spectrometry (LC-MS/MS)-based proteomics. Three major peptide identification approaches have been developed over the past two decades: sequence database search, *de novo* derivation, and spectral library search. The sequence database search approach typically identifies peptides through a two-step process: (1) generating theoretical fragments from peptide sequences in the database; (2) comparing acquired spectra to theoretical fragments. The *de novo* approach derives sequence tags directly from the MS/MS spectra by considering possible amino acid combinations. The spectral library approach determines peptides by searching experimental MS/MS spectra against a collected spectral library, which includes high confidence peptide-spectrum matches (PSMs) from the sequence database searching approach. Recently, the spectral library search approach has gained increasing popularity because of its high sensitivity and rapid search speed.

The idea of spectral library searching was first proposed by Yates et al. in 1998¹. Later numerous spectral library tools have been developed, including Biblispes², X!Hunter³, SpectraST⁴, Pepitome⁵, Peptide Atlas⁶, spec2spec⁷, MSPLIT-DIA⁸, and Tremolo⁹. In addition, several spectral library tools were also developed for identification with peptide modifications, such as QuickMod¹⁰ and pMatch¹¹ for post-translational modification, Hybrid for unanticipated modifications¹², cross-linked peptide identification¹³, and GPQuest¹⁴ for N-linked glycopeptides. Owing to improved sensitivity and turnaround time, several sample-specific libraries have been built, such as the neuropeptide database¹⁵, and a database for cancer cell lines¹⁶.

In recent years, isobaric labeling-based relative quantification approaches, such as tandem mass tag (TMT), have emerged as the mainstream method for multiplexed protein quantification¹⁷. In the Orbitrap instruments, high energy collisional dissociation (HCD) is implemented, in which higher energy and shorter activation time are used to generate product ions, including low-mass regions, such as TMT reporter ions, y_1 , and b_1 ions^{18,19}. A recent study reported spectral library search for TMT-based data by converting between unlabeled and labeled peptide MS/MS spectra²⁰. However, the approach is confounded by distinctive fragmentation patterns between TMT and label-free spectra. Therefore, developing a genuine TMT library will be valuable for peptide identification from TMT-based MS experiments.

In this study, we first describe the difference in fragmentation patterns of both b and y ions between TMT and label-free MS data. We then construct the largest human and mouse TMT spectral libraries so far using ultra-deep TMT-LC/LC-MS/MS data with millions of PSMs, and evaluate the library with an additional TMT dataset. The decoy library is constructed with the target-decoy strategy to estimate false discovery rate (FDR).

Materials and Methods

MS data generation and database search with JUMP

A total of 3.7 million and 6.7 million filtered PSMs (at ~1% protein FDR) from 531 and 1,162 TMT-LC/LC-MS/MS runs were used to construct human and mouse libraries, respectively. These TMT datasets were from 11 human and 25 mouse TMT profiling projects (Table 1). The samples were processed with a previously optimized protocol^{21,22}. In general, quantified protein (~1 mg in the lysis buffer with 8 M urea) for each TMT channel were proteolyzed with Lys-C (Wako, 1:100 w/w) at 21°C for 2 h, diluted 4-fold to reduce urea to 2 M, and further trypsinized at 21°C overnight (Promega, 1:50 w/w). The insoluble debris was kept in the lysates, which allowed the recovery of insoluble proteins during digestion. The digestion was stopped with the addition of 1% trifluoroacetic acid followed by centrifugation. The supernatant was desalted with Sep-Pak C18 cartridge (Waters), and dried by speedvac. Each sample was resuspended in 50 mM HEPES, pH 8.5, labeled with TMT reagents, and equally mixed.

The pooled TMT labeled samples were fractionated by offline basic pH reverse phase LC followed by acidic pH reverse phase LC-MS/MS analysis²³. We performed the offline LC runs (~2 h gradient) on an XBridge C18 column (3.5 μ m particle size, 4.6 mm x 25 cm, Waters; buffer A: 10 mM ammonium formate, pH 8.0; buffer B: 95% acetonitrile, 10 mM ammonium formate, pH 8.0)²¹. In the LC-MS/MS analysis, each fraction was run sequentially on an LC column (75 μ m x ~40 cm for the whole proteome, 1.9 μ m C18 resin from Dr. Maisch GmbH, at 65°C to reduce backpressure) interfaced with a Q Exactive HF or Orbitrap Fusion (Thermo Fisher). Peptides were eluted by a 2–3 h gradient (buffer A: 0.2% formic acid, 5% DMSO; buffer B: buffer A plus 65% acetonitrile). MS settings included MS1 scans (60,000 resolution, 1×10^6 AGC and 100 ms maximal ion time) and 20 data-dependent MS2 scans (410–1600 m/z , 60,000 resolution, 1×10^5 AGC, 100–150 ms maximal ion time, HCD, 38% normalized collision energy, 1.0 m/z isolation window with 0.3 m/z offset, and ~15 s dynamic exclusion).

The generated MS data were analyzed by JUMP, a hybrid database search algorithm²⁴. The human protein reference database was generated by combining downloaded Swiss-prot, TrEMBL, and UCSC databases and removing redundancy (human: 83,955 entries; mouse: 59,423 entries), followed by the concatenation of the decoy database. JUMP is a tag-based database search tool for peptide identification. JUMP combines pattern scoring with *de novo* tag scoring, displaying a significant improvement in PSM scoring. Major parameters included precursor and product ion mass tolerance (± 15 ppm), full trypticity, static mass shift for TMT tag (+229.16293) and carbamidomethylation (+57.02146) of Cysteine, dynamic mass shift for Methionine oxidation (+15.99491), maximal missed cleavage ($n = 2$), and maximal modification sites ($n = 3$). Putative PSMs were filtered by mass accuracy and then grouped by precursor ion charge state and filtered by JUMP-based matching scores (Jscore and dJn) to reduce protein FDR below 1%. If one peptide could be generated from multiple homologous proteins, the peptide was assigned to the protein with the highest PSM based on the rule of parsimony. The identification results were kept in the format of .pepXML for subsequent analysis.

Construction of the spectral library

The spectral library was built by the SpectraST⁴ program (version 5.0). The SpectraST used .pepXML files as input to extract experimental MS/MS spectra from the corresponding .mzXML files, and generated an initial spectral library in .msp format as output. The .msp file contained identified peptide sequences, MS/MS spectrum, and annotations. The SpectraST converted the spectra library in .msp into .splib format. The .splib file was kept as a binary to facilitate quick retrieval by SpectraST. All peaks of product ions and neutral losses were recorded in .splib file. We generated the consensus library and the best replicate library for both human and mouse, which are available at <https://www.stjuderesearch.org/site/lab/peng/tmt>.

Generation of decoy library

In this study, we used the precursor swap method in SpectraST to generate the decoy library²⁵. Briefly, the method generated a decoy for a spectrum i with precursor mass of M in following steps: (1) identifying all possible spectra with the same charge whose precursor is within a shifted mass d Da (i.e. between $M-d$ and $M+d$); (2) selecting the spectrum j with the highest precursor mass between $M-d$ and $M+d$; (3) swapping the precursor of the spectrum j with that of the spectrum i . Note that all MS/MS spectra between targets and decoys were the same. We used the default parameter of 8 Da mass shift for this study.

Spectral library search

We evaluated the TMT spectral library by searching an additional LC-MS/MS dataset, which was not used for the library construction. The LC-MS/MS dataset contained 56 fractions, including 3.3 million MS/MS spectra. The dataset was searched against the human target-decoy composite spectra library using SpectraST. The default search parameters of the SpectraST were used except for the following parameters: (1) precursor isolation window = 1.6 Th; (2) disable binning and performing peak-to-peak matching; (3) excluding TMT reporter ions. The FDR was estimated by $FDR = n_d / n_t$, where n_t and n_d were the numbers of target and decoy hits, respectively⁴⁰. The search results were then filtered by the SpectraST score (fval) to reduce the FDR to a certain level.

Construction of the semi-empirical TMT library

To build a semi-empirical TMT library, we first extracted a list of spectra from a label-free dataset. To make an unbiased comparison with the labeled data, only those spectra with the same peptide and charge state in both TMT labeled and label-free libraries were kept. The filtered spectra were then used to generate a semi-empirical spectral library²⁶ by adding the mass of TMT tag (+229.16293) to both N-terminus and lysine residue(s) using SpectraST with the parameter of -cx'n[230]K[357]'. In the semi-empirical library, all annotated peaks associated with both N-terminus and Lysine residue(s) were re-positioned to a new m/z location without changing their intensities. All other unannotated peaks were kept unchanged by assuming that they are noise, impurity, and unknown fragments. Similarly, an equal-size of decoys were generated using precursor swapping approach and appended to the library as a target-decoy composite semi-empirical spectral library.

RESULTS AND DISCUSSION

To develop and evaluate spectral library search for TMT-based MS data, we performed the analysis in three major steps (Figure 1): (1) constructing a consensus TMT spectral library with sequence database search; (2) creating and evaluating the decoy library for FDR estimation; (3) comparing the performance of spectral library search against sequence database search.

Pattern difference in MS/MS spectra between TMT and label-free data

MS/MS fragmentation patterns of peptides labeled with TMT tags (or other isobaric tags) are substantially different from that of unlabeled peptides. In a typical TMT experiment, the TMT reagents react with peptide amino-termini or lysine residues (Figure 2A,B), releasing TMT reporter ions in the MS/MS spectra. However, the labeling may affect proton affinities and gas-phase basicities, resulting in alterations of the fragmentation patterns (i.e. intensity and m/z of peaks)²⁷. For example, the fragmentation patterns of b and y ions in both K- and R-containing TMT peptides display difference when compared to corresponding non-labeled peptides (Figure 2C,D). The b_1 ion tends to be present in the labeled peptides, but not in the label-free peptides. The labeled peptides appear to have more and stronger b ions than the corresponding label-free peptides. In contrast, the intensities of y ions in TMT labeled peptides become weaker, especially in the high mass range. We then performed a comprehensive comparison of b and y ions from 148,615 label-free and labeled pairs. When summarizing the intensities of b_1 - b_{10} and y_1 - y_{10} , it shows similar patterns as seen in the examples (Figure 2E,F) for both K- and R-containing peptides, emphasizing the intensity shift from y ions to b ions in the TMT peptides. The analysis indicates that TMT-labeling changes the patterns of product ions during MS/MS fragmentation. Therefore it is important to build a TMT-specific library when searching TMT-labeled MS data.

Construction of comprehensive human and mouse TMT spectral libraries

To construct human and mouse TMT spectral libraries of high coverage, we used ultra-deep 10-plex TMT datasets with 531 and 1,162 LC-MS/MS runs generated from 11 human and 25 mouse sample sets (Table 1). The data were searched against the corresponding human or mouse target-decoy composite database with JUMP²⁴. By filtering peptide-spectrum matches (PSMs) with Jscore, we accepted 3.7 million and 6.7 million PSMs, corresponding to 455,921 and 507,702 peptides for human and mouse at protein FDR < 1%.

A peptide could be annotated by several similar MS/MS spectra. Therefore, it is required to identify the best spectrum for a query peptide. To resolve this issue, two library construction approaches, consensus and best replicate, have been previously proposed²⁸. Using the consensus approach, we constructed both human and mouse TMT spectral libraries, containing 392,537 and 363,138 consensus spectra respectively (Table 1). Similarly, we constructed human and mouse spectral libraries using the best replicate approach for subsequent database search and comparison.

Evaluation of TMT decoy library for FDR evaluation

Similar to sequence database search, an estimated FDR is needed for spectral library search. The FDR estimation approaches can be broadly grouped into two categories: target-decoy searching and empirical Bayes approaches²⁹. The target-decoy strategy is the most widely used approach for both sequence database search and spectral library search. In spectral library search, several methods have been proposed to construct a decoy spectral library, including precursor swap²⁵, precursor swap peak shift, peak shift, shuffle and reposition³⁰, and DeLiberator¹⁰. The precursor swap strategy has been shown to be superior to others in terms of efficiency and memory usage for constructing decoy libraries²⁵. In this study, we selected the precursor swap method for FDR estimation (see Methods).

To evaluate precursor swap method for TMT spectral library search, we generated a simulated null TMT dataset based on an LC-MS/MS run by adding 2 *Da* to each MS/MS precursor mass to ensure that the simulated precursor mass was not within a mass tolerance window. We searched this simulated dataset against the target-decoy composite spectral library with SpectraST. As shown in Figure 3, it showed that the score distributions of target and decoy hits are highly consistent, confirming that the generated TMT decoy library is suitable for FDR estimation.

Performance evaluation of TMT spectral library search

Previous studies have reported that modified and labeled data (e.g. TMT) can be searched against a spectral library converted from label-free²⁶ or other types of isobaric labeled data such as iTRAQ²⁰. To compare spectral library searches using a genuine TMT library and a converted semi-empirical TMT library, we first extracted a total of 435,310 PSMs with the same peptide and charge state from both TMT and label-free datasets. The label-free spectra were converted into a semi-empirical TMT library by adding the mass of the TMT tag to both the N-terminus and the lysine residue(s). By searching a genuine TMT library, we found that score distributions of target and decoy hits were well separated, and target hits showed a bias towards higher scores (Figure 4A). In contrast, the semi-empirical library search led to much less separation of target and decoy hits (Figure 4B). For example, at FDR of 5%, the genuine library detected 9,900 PSMs while the semi-empirical library identified only 1,036 PSMs (Figure 4C; black vs red). The marked difference in the number of identified PSMs could be attributed to the fact that distinct intensity patterns of *b* and *y* ions between TMT-labeled and label-free spectra with the same peptide, resulting in a low SpectraST score with the semi-empirical library, despite the present of the same *b* and *y* ions in both spectra. This analysis emphasizes that the genuine library results in a substantial increase of identification sensitivity than the converted semi-empirical TMT library.

Typically, the low mass region of TMT MS/MS spectra contains reporter ions, which might affect peak matching and scoring during the library search. We then performed spectral library searches with and without these TMT reporter ions. Interestingly, we found that both searches detected a similar number of PSMs at low FDR (e.g. 1%), but the reporter-excluding library detected more PSMs with a loose FDR threshold (Figure 4C; black vs dark red), suggesting that TMT reporter ions have a negative impact on library search, especially for those spectra of low quality.

In addition, we also compared the two library construction approaches (consensus and best replicate) by searching the same test dataset. The best replicate library detected slightly more PSMs at a low FDR (e.g. $<1\%$), but less PSMs at a high FDR (e.g. $>5\%$), compared to the consensus library (Figure 4C; black vs blue). We decided to select the consensus approach for the subsequent analysis.

We next sought to compare the performance of the spectral library search with SpectraST to sequence database search with JUMP. We used another 10-plex TMT dataset containing 56 LC-MS/MS runs, which was not used for the library construction. The dataset was searched against the human consensus spectral library. SpectraST fval score was used to separate the targets from decoys (Figure 5A), resulting in 653,954 PSMs at FDR of 1% (Figure 5B). By searching the same 56 LC-MS/MS runs with JUMP, we identified 451,666 PSMs at FDR $<1\%$ (Figure 5B), $\sim 30\%$ (202,288/653,954) less than library search (Figure 5C). Spectral library search covered the vast majority ($\sim 82\%$; 371,336/451,666) of the PSMs detected by JUMP. For those PSMs that are missed by the spectral library approach, it is largely because JUMP and library search assign spectra to different peptide sequences.

CONCLUSION

In summary, we studied the effect of TMT labeling on MS/MS spectra, and constructed comprehensive human and mouse TMT spectral libraries from ultra-deep TMT-LC/LC-MS/MS data. To the best of our knowledge, the two libraries are the largest TMT libraries currently available to the research community. We assessed the performance of the spectral library search with and without TMT reporter ions, and two library construction approaches (i.e. consensus and best replicate). The target-decoy strategy was also evaluated using a simulated TMT dataset for FDR estimation. With the consensus spectral library and the target-decoy strategy, we found that the spectral library approach is superior to the traditional sequence database search. In addition, the TMT spectral library is particularly useful for multiple batches of TMT experiments as it increases the sensitivity of identifications across the batches. This study provides the first application of spectral library identification for TMT experiments.

ACKNOWLEDGMENTS

The authors thank Bing Bai, Haiyan Tan, and Zhiping Wu for datasets and helpful discussion. This work was partially supported by National Institutes of Health grant R01AG047928, R01GM114260, R01AG053987, and ALSAC (American Lebanese Syrian Associated Charities). The MS analysis was performed in the St. Jude Children's Research Hospital Proteomics Facility and high performance computing (HPC) shared resources, partially supported by NIH Cancer Center Support Grant (P30CA021765). JS and BM are supported by an NSERC discovery grant.

REFERENCES

- (1). Yates JR, 3rd; Morgan SF; Gatlin CL; Griffin PR; Eng JK Method to compare collision-induced dissociation spectra of peptides: potential for library searching and subtractive analysis. *Anal. Chem* 1998, 70, 3557–3565. [PubMed: 9737207]
- (2). Frewen BE; Merrihew GE; Wu CC; Noble WS; MacCoss MJ Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Anal. Chem* 2006, 78, 5678–5684. [PubMed: 16906711]

- (3). Craig R; Cortens JC; Fenyo D; Beavis RC Using annotated peptide mass spectrum libraries for protein identification. *J. Proteome Res* 2006, 5, 1843–1849. [PubMed: 16889405]
- (4). Lam H; Deutsch EW; Eddes JS; Eng JK; King N; Stein SE; Aebersold R Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* 2007, 7, 655–667. [PubMed: 17295354]
- (5). Dasari S; Chambers MC; Martinez MA; Carpenter KL; Ham AJ; Vega-Montoto LJ; Tabb DL Pepitome: evaluating improved spectral library search for identification complementarity and quality assessment. *J. Proteome Res* 2012, 11, 1686–1695. [PubMed: 22217208]
- (6). Desiere F; Deutsch EW; King NL; Nesvizhskii AI; Mallick P; Eng J; Chen S; Eddes J; Loevenich SN; Aebersold R The PeptideAtlas project. *Nucleic Acids Res* 2006, 34, D655–658. [PubMed: 16381952]
- (7). Yen CY; Houel S; Ahn NG; Old WM Spectrum-to-spectrum searching using a proteome-wide spectral library. *Mol. Cell. Proteomics* 2011, 10, M111 007666.
- (8). Wang J; Tucholska M; Knight JD; Lambert JP; Tate S; Larsen B; Gingras AC; Bandeira N MSPLIT-DIA: sensitive peptide identification for data-independent acquisition. *Nat. Methods* 2015, 12, 1106–1108. [PubMed: 26550773]
- (9). Wang M; Bandeira N Spectral library generating function for assessing spectrum-spectrum match significance. *J. Proteome Res* 2013, 12, 3944–3951. [PubMed: 23808827]
- (10). Ahrne E; Nikitin F; Lisacek F; Muller M QuickMod: A tool for open modification spectrum library searches. *J. Proteome Res* 2011, 10, 2913–2921. [PubMed: 21500769]
- (11). Ye D; Fu Y; Sun RX; Wang HP; Yuan ZF; Chi H; He SM Open MS/MS spectral library search to identify unanticipated post-translational modifications and increase spectral identification rate. *Bioinformatics* 2010, 26, i399–406. [PubMed: 20529934]
- (12). Burke MC; Mirokhin YA; Tchekhovskoi DV; Markey SP; Heidbrink Thompson J; Larkin C; Stein SE The Hybrid Search: A Mass Spectral Library Search Method for Discovery of Modifications in Proteomics. *J. Proteome Res* 2017, 16, 1924–1935. [PubMed: 28367633]
- (13). Schweppe DK; Chavez JD; Navare AT; Wu X; Ruiz B; Eng JK; Lam H; Bruce JE Spectral Library Searching To Identify Cross-Linked Peptides. *J. Proteome Res* 2016, 15, 1725–1731. [PubMed: 27089058]
- (14). Toghi Eshghi S; Shah P; Yang W; Li X; Zhang H GPQuest: A Spectral Library Matching Algorithm for Site-Specific Assignment of Tandem Mass Spectra to Intact N-glycopeptides. *Anal. Chem* 2015, 87, 5181–5188. [PubMed: 25945896]
- (15). Kim Y; Bark S; Hook V; Bandeira N NeuroPedia: neuropeptide database and spectral library. *Bioinformatics* 2011, 27, 2772–2773. [PubMed: 21821666]
- (16). Faktor J; Bouchal P Building Mass Spectrometry Spectral Libraries of Human Cancer Cell Lines. *Klin. Onkol*, 29 Suppl 4, 54–58. [PubMed: 27846721]
- (17). Tan H; Yang K; Li Y; Shaw TI; Wang Y; Blanco DB; Wang X; Cho JH; Wang H; Rankin S; Guy C; Peng J; Chi H Integrative Proteomics and Phosphoproteomics Profiling Reveals Dynamic Signaling Networks and Bioenergetics Pathways Underlying T Cell Activation. *Immunity* 2017, 46, 488–503. [PubMed: 28285833]
- (18). Olsen JV; Macek B; Lange O; Makarov A; Horning S; Mann M Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* 2007, 4, 709–712. [PubMed: 17721543]
- (19). Shao C; Zhang Y; Sun W Statistical characterization of HCD fragmentation patterns of tryptic peptides on an LTQ Orbitrap Velos mass spectrometer. *J. Proteomics* 2014, 109, 26–37. [PubMed: 24981973]
- (20). Zhang Z; Yang X; Mirokhin YA; Tchekhovskoi DV; Ji W; Markey SP; Roth J; Neta P; Hizal DB; Bowen MA; Stein SE Interconversion of Peptide Mass Spectral Libraries Derivatized with iTRAQ or TMT Labels. *J. Proteome Res* 2016, 15, 3180–3187. [PubMed: 27386737]
- (21). Bai B; Tan H; Pagala VR; High AA; Ichhaporia VP; Hendershot L; Peng J Deep Profiling of Proteome and Phosphoproteome by Isobaric Labeling, Extensive Liquid Chromatography, and Mass Spectrometry. *Methods Enzymol* 2017, 585, 377–395. [PubMed: 28109439]

- (22). Pagala VR; High AA; Wang X; Tan H; Kodali K; Mishra A; Kavdia K; Xu Y; Wu Z; Peng J Quantitative protein analysis by mass spectrometry. *Methods Mol. Biol* 2015, 1278, 281–305. [PubMed: 25859956]
- (23). Wang H; Yang Y; Li Y; Bai B; Wang X; Tan H; Liu T; Beach TG; Peng J; Wu Z Systematic optimization of long gradient chromatography mass spectrometry for deep analysis of brain proteome. *J. Proteome Res* 2015, 14, 829–838. [PubMed: 25455107]
- (24). Wang X; Li Y; Wu Z; Wang H; Tan H; Peng J JUMP: a tag-based database search tool for peptide identification with high sensitivity and accuracy. *Mol. Cell. Proteomics* 2014, 13, 3663–3673. [PubMed: 25202125]
- (25). Cheng CY; Tsai CF; Chen YJ; Sung TY; Hsu WL Spectrum-based method to generate good decoy libraries for spectral library searching in peptide identifications. *J. Proteome Res* 2013, 12, 2305–2310. [PubMed: 23560440]
- (26). Hu Y; Li Y; Lam H A semi-empirical approach for predicting unobserved peptide MS/MS spectra from spectral libraries. *Proteomics* 2011, 11, 4702–4711. [PubMed: 22038894]
- (27). Dongre AR; Jones JL; Somogyi Á; Wysocki VH Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *J. Am. Chem. Soc* 1996, 118, 8365–8374.
- (28). Lam H; Deutsch EW; Eddes JS; Eng JK; Stein SE; Aebersold R Building consensus spectral libraries for peptide identification in proteomics. *Nat. Methods* 2008, 5, 873–875. [PubMed: 18806791]
- (29). Nesvizhskii AI; Vitek O; Aebersold R Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat. Methods* 2007, 4, 787–797. [PubMed: 17901868]
- (30). Lam H; Deutsch EW; Aebersold R Artificial decoy spectral libraries for false discovery rate estimation in spectral library searching in proteomics. *J. Proteome Res* 2010, 9, 605–610. [PubMed: 19916561]

Figure 1.
Flowchart for TMT spectral library search.

Figure 2. Fragmentation pattern difference between TMT and non-TMT data.

(A) Chemical structure of the TMT₀ reagent. The dotted line indicates the reaction site. (B) Diagram showing a peptide labeled with TMT. (C–D) MS/MS spectra showing the experimental fragmentation patterns of K and R containing peptides. (E–F) Box plots showing the intensity distribution of b_1 – b_{10} and y_1 – y_{10} ions of K and R containing peptides.

Figure 3.
Evaluation of decoy database generated by precursor swap method.

Figure 4. Performance comparison of the spectral library search with the genuine TMT library and the semi-empirical TMT library.

(A) SpectraST score (fval) distribution of targets and decoys when searching against the genuine TMT library. (B) SpectraST score (fval) distribution of targets and decoys when searching against the semi-empirical TMT library converted from label-free data. (C) Sensitivity analysis of spectral library searches with the genuine TMT library and the semi-empirical TMT library. At the FDR of 5%, the genuine TMT library detects 9,900 PSMs, whereas the semi-empirical TMT library only identifies 1,036 PSMs.

Figure 5. Performance comparison between the spectral library search and JUMP.

(A) Distribution of SpectraST score (fval) between targets and decoys. (B) FDR curve showing the number of PSMs identified library search (SpectraST) and sequence database search (JUMP). (C) Comparison of the number of identified PSMs between SpectraST and JUMP at PSM FDR of 1%.

Table 1.

Summary of human and mouse TMT spectral libraries

	Human	Mouse
No. of sample sets	11	25
No. of runs	531	1,162
No. of MS/MS spectra identified by JUMP	3,705,797	6,689,387
No. of consensus spectra in the library	546,841	829,323
No. of unique peptides in the library	401,168	432,649
No. of proteins in the library	74,566	55,503
No. of unique proteins in the library and match to core protein database *	14,048	13,721
No. of unique genes in the library and match to core protein database *	13,953	13,439

* Note: Human and mouse protein core databases contain 20,386 and 16,992 unique proteins, corresponding 20,080 and 16,564 unique genes, respectively. Both databases were downloaded from the UniProt website at <https://www.uniprot.org/>.