

RESEARCH ARTICLE

# Performance Investigation of Proteomic Identification by HCD/CID Fragmentations in Combination with High/Low-Resolution Detectors on a Tribid, High-Field Orbitrap Instrument

Chengjian Tu<sup>1,2</sup>\*, Jun Li<sup>1,2</sup>, Shichen Shen<sup>1,2</sup>, Quanhu Sheng<sup>3</sup>, Yu Shyr<sup>3</sup>, Jun Qu<sup>1,2</sup>\*

**1** Department of Pharmaceutical Sciences, University at Buffalo, State University of New York, Buffalo, United States of America, **2** New York State Center of Excellence in Bioinformatics and Life Sciences, 701 Ellicott Street, Buffalo, United States of America, **3** Center for Quantitative Sciences, Vanderbilt University School of Medicine, Nashville, United States of America

\* These authors contributed equally to this work.

\* [junqu@buffalo.edu](mailto:junqu@buffalo.edu) (JQ); [ctu3@buffalo.edu](mailto:ctu3@buffalo.edu) (CT)



OPEN ACCESS

**Citation:** Tu C, Li J, Shen S, Sheng Q, Shyr Y, Qu J (2016) Performance Investigation of Proteomic Identification by HCD/CID Fragmentations in Combination with High/Low-Resolution Detectors on a Tribid, High-Field Orbitrap Instrument. PLoS ONE 11(7): e0160160. doi:10.1371/journal.pone.0160160

**Editor:** Haixu Tang, Indiana University, UNITED STATES

**Received:** April 14, 2016

**Accepted:** July 14, 2016

**Published:** July 29, 2016

**Copyright:** © 2016 Tu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All raw files of different MS2 acquisition methods are available at [ChorusProject.org](http://ChorusProject.org) under project name "Investigation of Fusion MS2 acquisition methods (ID: 1029)".

**Funding:** This work was supported by American Heart Association (AHA) award 12SDG9450036 (JQ), by a Center of Protein Therapeutics Industrial Award (JQ), by NIH grants U54HD071594 (JQ) and HL103411 (JQ). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

The recently-introduced Orbitrap Fusion mass spectrometry permits various types of MS2 acquisition methods. To date, these different MS2 strategies and the optimal data interpretation approach for each have not been adequately evaluated. This study comprehensively investigated the four MS2 strategies: HCD-OT (higher-energy-collisional-dissociation with Orbitrap detection), HCD-IT (HCD with ion trap, IT), CID-IT (collision-induced-dissociation with IT) and CID-OT on Orbitrap Fusion. To achieve extensive comparison and identify the optimal data interpretation method for each technique, several search engines (SEQUEST and Mascot) and post-processing methods (score-based, PeptideProphet, and Percolator) were assessed for all techniques for the analysis of a human cell proteome. It was found that divergent conclusions could be made from the same dataset when different data interpretation approaches were used and therefore requiring a relatively fair comparison among techniques. Percolator was chosen for comparison of techniques because it performs the best among all search engines and MS2 strategies. For the analysis of human cell proteome using individual MS2 strategies, the highest number of identifications was achieved by HCD-OT, followed by HCD-IT and CID-IT. Based on these results, we concluded that a relatively fair platform for data interpretation is necessary to avoid divergent conclusions from the same dataset, and HCD-OT and HCD-IT may be preferable for protein/peptide identification using Orbitrap Fusion.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

High resolution mass spectrometers play a pivotal role in the field of proteomics and recent improvements in sensitivity and scan rates enable them to detect ~90% of the expressed proteome for yeast[1]. The traditional collision induced dissociation (CID, or resonant excitation CID) and higher energy collisional dissociation (HCD, or beam-type CID) are the most popular fragmentation techniques for bottom-up proteomics studies[2, 3]. Tandem mass spectra (MS2) by CID with the detection of ion trap (CID-IT) and HCD with the Orbitrap detection (HCD-OT) are the two common MS2 acquisition methods, presenting the advantages of speed and sensitivity for the former and wide  $m/z$  range and accurate MS2 for the latter[2–4]. The comparative studies of HCD-OT and CID-IT have been investigated and divergent conclusions have been presented from different labs[2, 5–7], even for the same data set[2, 7]. This may be caused by the different parameters of MS and different MS data interpretation methods they have used. Moreover, the new MS instrument Orbitrap Fusion enables MS2 fragmentation by HCD with IT detection (HCD-IT)[1], combining the benefits of wide range of  $m/z$  products and speed. However, it is not clear which MS2 acquisition method provides better sensitivity, reproducibility and convenience for proteomics.

With the rapid development of MS instruments, database search engines such as SEQUEST [8], Mascot[9], OMSSA[10], MyriMatch[11], Andromeda[12], Morpheus[13], MS-GF+[14], and MS Amanda[15] are subsequently developed to interpret the different types of spectra generated by MS. To achieve confident and accurate MS-based identification, the target-decoy search strategy is widely applied because of its conceptual simplicity and easy implementation [16]. Moreover, besides the utilization of original score to calculate false discovery rate (FDR), many post-search algorithms have been developed to perform statistical classification between correct and incorrect using machine learning methods[17–24] such as PeptideProphet[17, 24] and Percolator[20]. The Percolator algorithm[20] initially identifies a subset of high-confidence target PSMs, and then learns to optimally separate correct and incorrect PSMs using the support vector machines (SVM)-based classifier. More peptide/protein identifications are achieved using Percolator while compared to other filtering methods[20]. The different combinations of search engines and post-processing approaches may also give different conclusions for comparative analysis of the same data sets.

In this work, we evaluated different MS2 acquisition methods, including HCD-OT, HCD-IT and CID-IT and CID-OT on one of the most recent instruments, Orbitrap Fusion. To select a relatively fair platform for data interpretation, popular search engines (e.g. SEQUEST and Mascot) and filtering approaches (e.g. score-based, PeptideProphet, and Percolator) were applied to analyze all the raw files in this study. Sensitivity, reproducibility and convenience for large-scale data analysis of these MS2 acquisition methods were investigated and discussed. Based on these results, the optimal MS2 strategy and corresponding data interpretation approach could be suggested for other proteomics researchers to achieve better proteome coverage from the Orbitrap Fusion or similar instruments.

## Materials and Methods

### Sample preparation

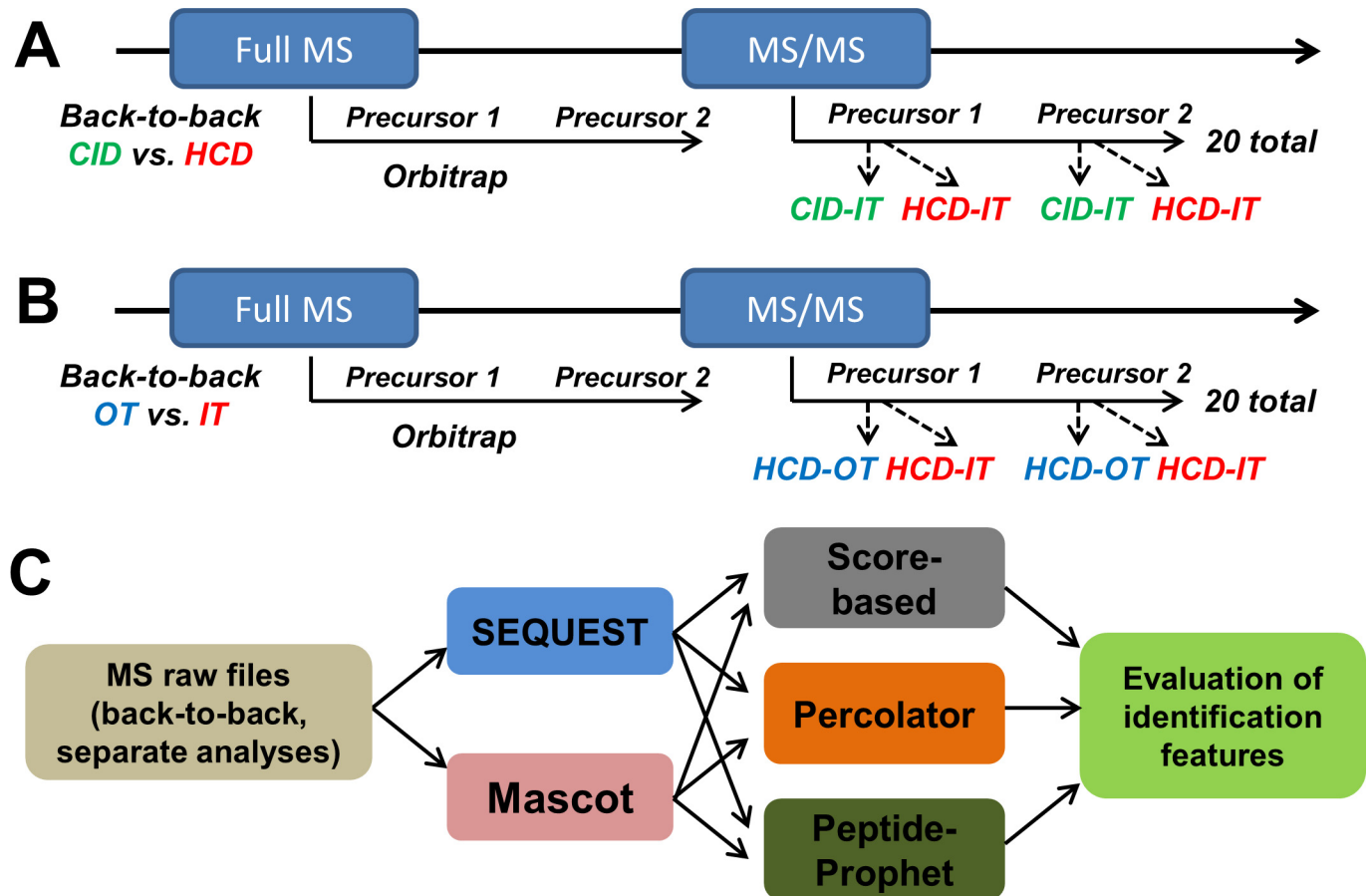
The PANC-1 cells (human pancreatic carcinoma cell line) were from Dr. William J. Jusko's lab in the Department of Pharmaceutical Sciences at University at Buffalo (Buffalo, USA). Cell samples were homogenized by ultra-sonication (Sonicatore XL-2000, Misonix, Inc., USA) in an ice-cold lysis buffer (50 mM Tris-formic acid, 150 mM NaCl, 0.5% sodium deoxycholate, 2% SDS, 2% NP-40 substitute, pH 8.0) for 3–5 second bursts, until the solution was clear.

Lysates were centrifuged at 20,000 g for 30 min at 4°C. The resulting supernatant was transferred to a new tube and BCA Protein Assay (Pierce, Rockford, IL) was used for protein concentration determination. All samples were stored at -80°C until further analysis. For each sample, 100 µg of protein was reduced with 10 mM DTT for 30 min at 56°C, and then alkylated with 30 mM IAM for 30 min in darkness at 37°C. In this study a precipitation/on-pellet-digestion procedure was used to perform precipitation and tryptic digestion as previously described [25, 26].

## NanoLC-MS/MS analysis

The peptide mixture of PANC-1 cells was analyzed using an ultra-high pressure Eksport nanoLC 425 system (Eksigent, Dublin, CA) coupled to a Orbitrap Fusion tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA). The mobile phase consisted of 0.1% formic acid in 2% acetonitrile (A) and 0.1% formic acid in 88% acetonitrile (B). Each 4 µg of peptide mixture was loaded onto a reversed-phase trap (300 µm ID x 0.5 cm, packed with Zorbax 3-µm C18 particles), with 1% mobile phase B at a flow rate of 10 µL/min, and the trap was washed for 3 min. A series of nanoflow gradients (flow rate, 250 nL/min) was used to back-flush the trapped samples onto the nano-LC column (75 µm ID x 100 cm, packed with Pepmap 3-µm C18 particles) for separation. The nano-LC column was heated at 52°C to greatly improve both chromatographic resolution and reproducibility. A 160-min gradient was applied in these analyses to achieve sufficient peptide separation. The gradient profile was as follows: 0 to 3% B over 3 min; 3 to 6% B over 5 min; 6 to 28% B over 118 min; 28 to 50% B over 10 min; 50 to 97% B over 1 min; and finally isocratic at 97% B for 23 min, and then the column was equilibrated with mobile phase A.

The data-dependent product ion mode was applied for all analyses. For precursor peptides fragmentation and detection, MS1 survey scans ( $m/z$  400 to 1500) were performed at a resolution of 120,000 with a  $5 \times 10^5$  AGC target. A resolution of 60,000 for MS1 was also investigated and a lower number of peptides was observed relative to a resolution of 120,000 we applied here (S1A Fig). Peptide precursors with charge state 2–7 were sampled for MS2. The instrument was run in top speed mode with a cycle time of 3 s. Dynamic exclusion was enabled with the following settings: repeat count = 1; exclusion duration = 60 s; mass tolerance =  $\pm 10$  ppm. Monoisotopic precursor selection was turned on. Tandem MS was performed by isolation at 1.0 Th with the quadrupole for HCD or CID fragmentation. The normalized collision energy (NCE) was 35% for both HCD and CID. Comparison between the NCE of 30% and 35% for HCD was performed, and we found that the use of 35% NCE for HCD gave similar or a little bit more protein identifications (S1B Fig). For OT detection, tandem mass spectra were analyzed with a resolution of 15,000, MS2 AGC target was set to  $5 \times 10^4$  and the max injection time was 50 ms. Maximum injection times of 20, 35 and 50 ms were also evaluated and the use of 50 ms gave more peptide/protein identifications (S1C Fig). For IT detection, tandem mass spectra were analyzed by ion trap with rapid scan rate, MS2 AGC target was set to  $1 \times 10^4$  and the max injection time was 35 ms. In back-to-back (B2B) analyses (Fig 1), the 20 most abundant ions (TOP20) detected in an Orbitrap full MS spectrum were selected for further MS/MS analyses. For comparison of CID-IT and HCD-IT, each precursor ion was first selected for CID-fragmentation with IT detection and then for HCD-fragmentation with IT detection; for comparison of HCD-OT and HCD-IT, each precursor ion was first selected for HCD-fragmentation with OT detection and then for HCD-fragmentation with IT detection. Other MS settings were the same as described above. Each type of MS2 strategies was analyzed in triplicate.



**Fig 1. Flowchart of evaluation of CID- and HCD-type fragmentation, OT- and IT-type detection, and optimal MS2 acquisition method in the Orbitrap Fusion mass spectrometer.** We used the back-to-back strategy to perform the comparisons: HCD versus CID, and OT versus IT. (A) The 20 most abundant ions from each full MS cycle in Orbitrap were subjected to sequential CID-IT (ion-trap detection) and HCD-IT. (B) The 20 most abundant ions from each full MS cycle in Orbitrap were subjected to sequential HCD-OT (Orbitrap detection) and HCD-IT. (C) The files of resulting back-to-back analyses and separate MS2 acquisition methods (HCD-OT, HCD-IT, CID-OT, CID-IT) were analyzed using different search engines (SEQUEST, Mascot) and post-processing approaches (Score-based, Percolator, PeptideProphet).

doi:10.1371/journal.pone.0160160.g001

### Database search and post-search filtering analyses

Proteome Discoverer (PD, version 1.4.1.14, Thermo-Scientific) was used to perform database searching against Swiss-Prot human protein database (20,166 entries, released January, 2015). Search engines: SEQUEST-HT and Mascot (version 2.4.0) implemented in Proteome Discoverer were individually applied for all MS raw files. The search parameters used were as follows: 20 ppm tolerance for precursor ion masses, 1.0 Da tolerance for fragment ion masses analyzed by ion trap, 0.02 Da tolerance for fragment ion masses analyzed by Orbitrap. We also evaluated 0.35 Da tolerance for fragment ion masses analyzed by HCD-IT as suggested by previous reports[27]. Lower peptide/protein identifications were achieved relative to the use of 1.0 Da tolerance when using SEQUEST-PeptideProphet as shown in S1D Fig and no difference was observed when using SEQUEST-Percolator as we previously studied [28]. Thus, 1.0 Da tolerance for fragment ion masses analyzed by ion trap was applied in this study as we previously reported[28]. Two missed cleavages were allowed for fully tryptic peptides. Carbamidomethylation of cysteines (+57 Da) was set as a fixed modification, and methionine oxidation (+16 Da) was set as a variable modification. The false discovery rate (FDR) was determined by using

the target-decoy search strategy[29]. The sequence database contains each sequence in both forward and reversed orientations, enabling FDR calculation. Here, peptide/protein FDR is calculated as the number of decoy peptides/proteins divided by the number of target peptides/proteins. The FDR was set to 0.01 for both peptide and protein identifications.

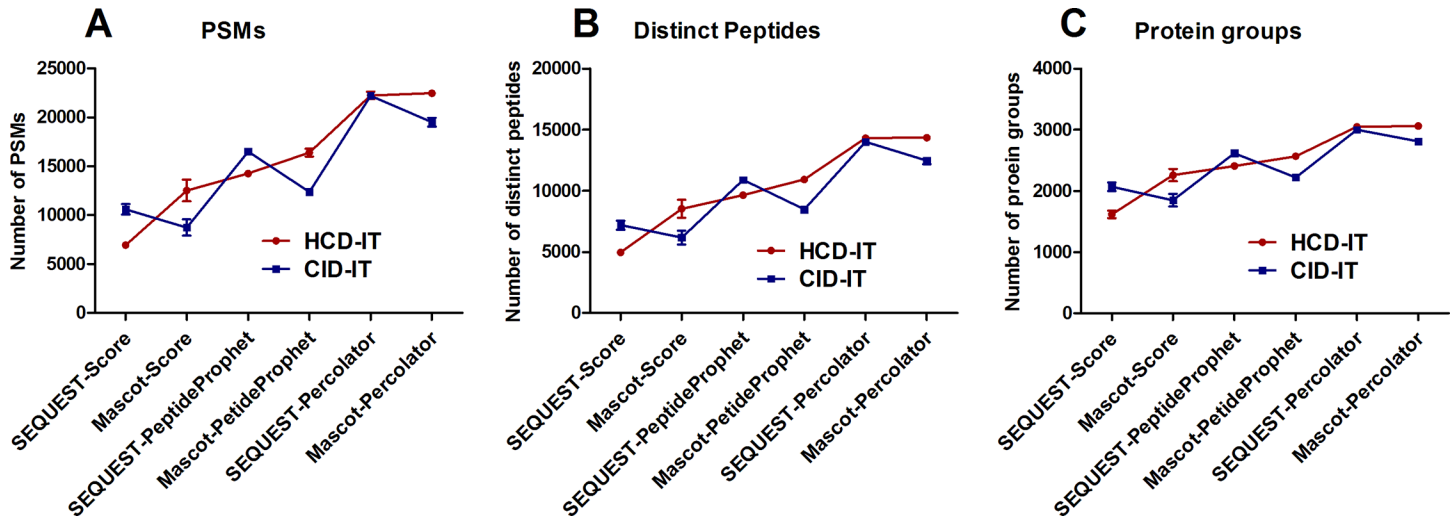
For different post-search filtering approaches, Scaffold version 4.3.2 (Proteome Software, Portland, OR) and custom software BuildSummary (v7.1.1)[30] were used in this study. For score-based filtering, PSMs were sorted and selected according to scores to achieve a protein FDR of 1.00%: ascending XCorr ( $\Delta Cn \geq 0.1$ ) for SEQUEST-HT and ascending ion score for Mascot; For PeptideProphet (with delta mass correction), ascending probability was used to select confident PSMs for SEQUEST-HT and Mascot; For Percolator, ascending SVM-score was used. The score threshold yielding the maximum target protein groups at less than or equal to 1% FDR was determined. Percolator (version 2.04), incorporated in PD, was used to generate q-values, SVM-scores and posterior error probabilities. The newest version of BuildSummary could be downloaded freely at <https://github.com/shengqh/RCPA.Tools/releases/>.

## Results and Discussion

To date, because of the prevalent use of hybrid MS, multiple MS2 acquisition methods are available on the same instrument, and suitable data interpretation approaches were correspondingly developed for maximizing the identifications. However, evaluation of different MS2 acquisition methods on Orbitrap Fusion using a relatively fair platform has not been comprehensively investigated. Because of the data-dependent MS2 analyses and different speed of HCD and CID acquisition, the back-to-back strategy is often performed for parallel comparison between the HCD-OT and CID-IT to ensure that identical precursors were respectively analyzed[2]. In previous studies, the mixed effect of HCD fragmentation and OT detection are commonly compared to CID-IT [2, 7], and more peptide-spectra-matches (PSMs) by HCD-OT than CID-IT is observed using the back-to-back strategy. Here we designed the back-to-back analyses of CID-IT *versus* HCD-IT, and HCD-OT *versus* HCD-IT to separately evaluate the different fragmentations (HCD *versus* CID) and detections (OT *versus* IT) for peptide identification on the Orbitrap Fusion mass spectrometer. The scheme of back-to-back experimental designs was shown in Fig 1A & 1B. Three LC-MS/MS replicates were analyzed for each comparison by two popular search engines (SEQUEST and Mascot) and three common filtering approaches (original score, PeptideProphet, and Percolator) as shown in Fig 1C. SEQUEST and Mascot are two classical search engines with different scoring strategies. SEQUEST utilizes a cross-correlation score to evaluate the similarity between experimental and theoretical mass spectra while Mascot emphasizes the rank of possible peptide matches that best fits the acquired MS2 via probabilistic modeling. The 1% of FDR at respective peptide level and protein level was applied in all analyses. Separate analyses of different MS2 acquisition methods (S2 Fig) were also investigated.

### Back-to-back comparison for HCD- and CID-type MS2 fragmentations

To identify a relatively fair platform for peptide/protein identification, a total of six MS data interpretation approaches (search engines + post-processing methods) were used to analyze these raw files (Fig 1C). As shown in Fig 2, the CID-IT mode achieved significantly more PSMs (10597 vs. 6944), distinct peptides (7184 vs. 4964) and protein groups (2069 vs. 1620) than HCD-IT when using SEQUEST-Score or SEQUEST-PeptideProphet interpretation approaches. However, when using SEQUEST-Percolator for the data interpretation, similar quantities (PSMs, distinct peptides and protein groups) were provided by CID-IT and HCD-IT modes (Fig 2). Interestingly, opposite results were obtained by using Mascot coupled with the

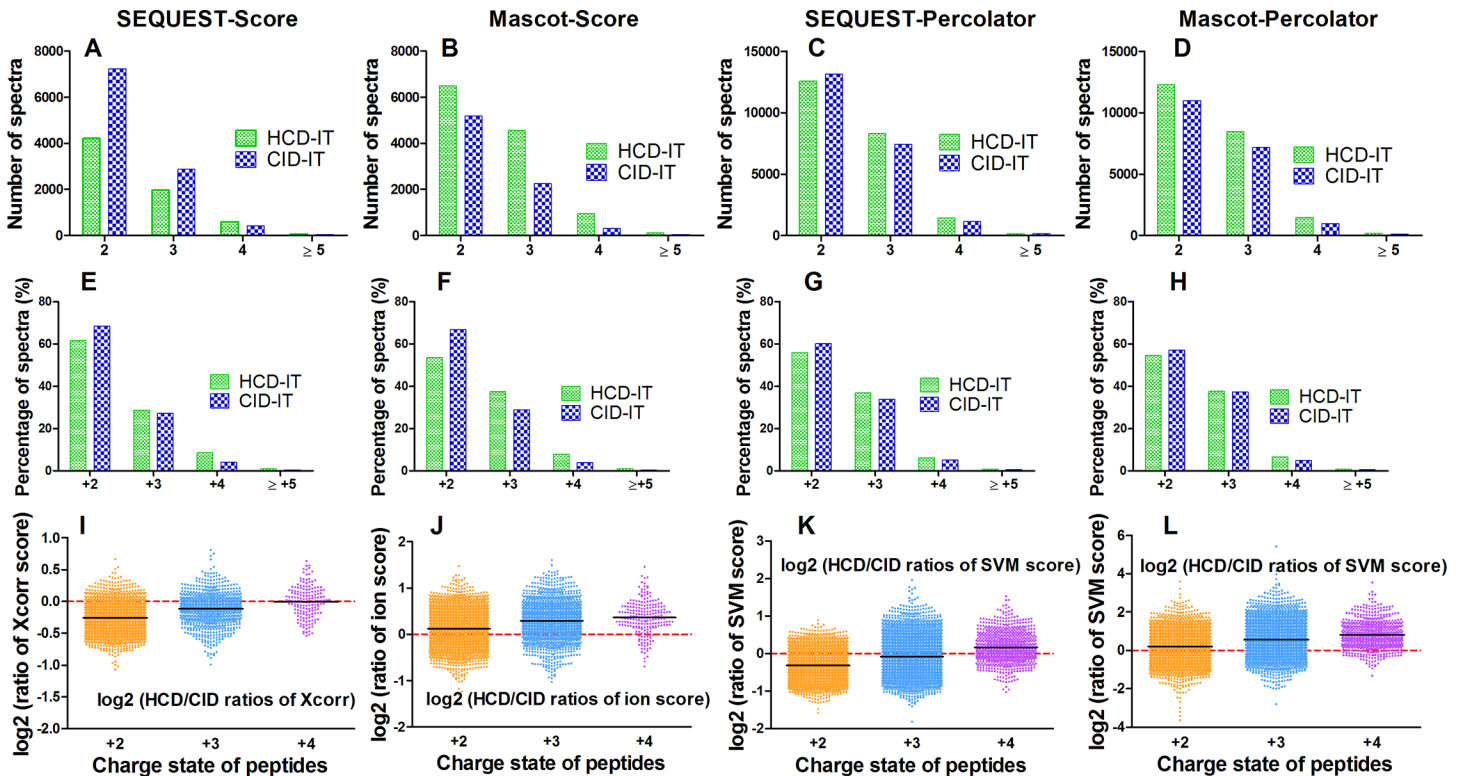


**Fig 2. (A) Peptide spectra matches (PSMs), (B) distinct peptides and (C) protein groups from the raw files for the comparison of HCD- and CID-type fragmentation using the back-to-back strategy (HCD-IT vs. CID-IT).** Six different MS data interpretation approaches were used in this study by combining SEQUEST or Mascot with different post-processing methods such as score-based, PeptideProphet, and Percolator.

doi:10.1371/journal.pone.0160160.g002

three data interpretation approaches. For instance, more PSMs (12531 vs. 8742), distinct peptides (8530 vs. 6169) and protein groups (2260 vs. 1849) were significantly achieved from the HCD-IT mode than the CID-IT mode using Mascot-Score (Fig 2). Therefore, the results of comparison between the HCD and CID modes were greatly biased by the MS data interpretation approaches. In this study, SEQUEST-Percolator and Mascot-Percolator obtained the highest number of quantities for the CID-IT and HCD-IT modes, respectively. However, there was no significant difference ( $p > 0.05$ ) between them as shown in S3A Fig. As expected, the overlap between CID-IT data (interpreted by SEQUEST-Percolator) and HCD-IT data (by Mascot-Percolator) from the same LC-MS replicate was excellent and more than 84.5% of peptides were co-identified (S3B Fig). We further analyzed the overlapped and unique peptides for each method, and the overlapped peptides tend to have higher identification scores and more amino acids as shown in S3C and S3D Fig.

The distributions of peptide charge and ratio of searching score (i.e. Xcorr for SEQUEST, ion score for Mascot, and SVM score for Percolator) between HCD-IT and CID-IT for co-identified peptides were further investigated. One replicate of the back-to-back LC-MS analyses of HCD-IT versus CID-IT was randomly selected for this purpose. As shown in Fig 3A, 3B & 3D, more PSMs for each charge state in one type of fragmentation were identified when that type of fragmentation achieved more total PSMs by using the specific combination of search engine and post-processing method. For instance, Mascot-Score identified more PSMs (12531 vs. 8742) from HCD-IT than CID-IT (Fig 2), for each charge state more PSMs (i.e. +2 charged peptides, 6506 vs. 5190) were still got from HCD-IT than CID-IT (Fig 3B). Interestingly, though similar PSMs (or a little bit more PSMs from HCD-IT) were achieved from HCD-IT (22240) and CID-IT (22227) when using SEQUEST-Percolator (Fig 2), HCD-IT achieved fewer doubly charged (+2) PSMs (12556 vs. 13157) and more highly charged PSMs ( $\geq +3$ ) (i.e. +3 charged peptides, 8313 vs. 7413) as shown in Fig 3C. The percentage distribution of different charged peptides further proved it. As shown in Fig 3E–3H, no matter how many PSMs were identified from HCD-IT or CID-IT by different interpretation approaches, HCD-IT achieved higher percentages of highly charged PSMs while CID-IT leaned towards to doubly charged PSMs. We also investigated searching scores of peptides between HCD-IT and



**Fig 3. The comparison of HCD- and CID-type fragmentation using the back-to-back strategy (HCD-IT vs. CID-IT).** (A-D) Number of spectra for +2, +3, +4 and  $\geq 5$  charged ions using different data interpretation methods. (E-H) Percentage of spectra for for +2, +3, +4 and  $\geq 5$  charged ions using different data interpretation methods. (I-L) Log<sub>2</sub> (HCD/CID ratio of score) distributions for +2, +3 and +4 charged peptides. Xcorr for SEQUEST, ion score for Mascot, SVM score for Percolator related combinations were presented. One replicate of the back-to-back LC-MS analyses of HCD-IT versus CID-IT was randomly selected.

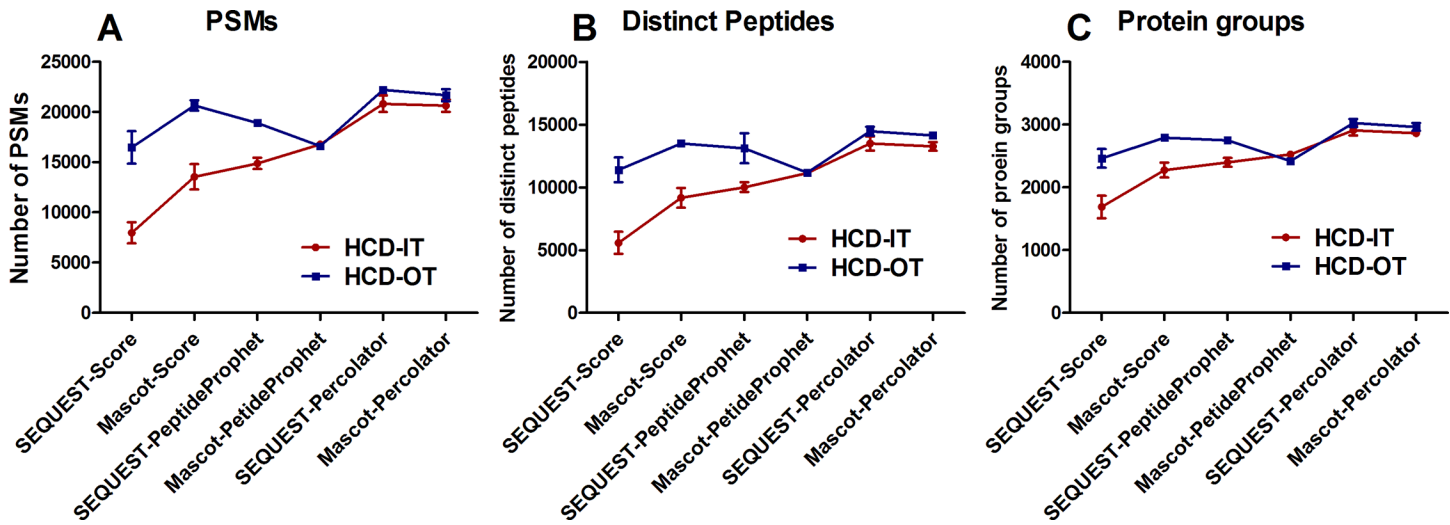
doi:10.1371/journal.pone.0160160.g003

CID-IT. When using Mascot (coupled with score-based or Percolator), HCD-IT achieved higher search score than CID-IT (Fig 3J & 3L); for SEQUEST-Score or -Percolator, CID-IT achieved higher search score for doubly charged peptides and triply charged peptides but not quadruply charged peptides (Fig 3I & 3K).

In summary, the similar identifications including PSMs, distinct peptides and proteins groups were achieved from HCD-IT (Mascot- or SEQUEST-Percolator) and CID-IT (SEQUEST-Percolator) when using the relatively fair platform of MS data interpretation for each MS2 acquisition method. These results indicated that HCD- and CID-type MS2 fragmentation on Orbitrap Fusion mass spectrometer may contribute almost equally for the peptide/protein identification. Moreover, the percentage of highly charged peptides appeared higher in HCD-IT data than that in CID-IT.

### Back-to-back comparison for OT- and IT-type MS2 detections

MS data interpretation approaches described above were also applied to the back-to-back analyses of HCD-IT versus HCD-OT to perform comparison between OT and IT detections. As shown in Fig 4, Percolator also achieved higher identifications than score-based and Peptide-Prophet methods, which was search engine-independent. More PSMs, distinct peptides and protein groups were identified from MS data generated in HCD-OT mode than HCD-IT mode when the data was analyzed with selected MS data interpretation approaches except Mascot-PeptideProphet (Fig 4A–4C). For instance, 20659 PSMs, 13496 distinct peptides and 2791



**Fig 4. (A) Peptide spectra matches (PSMs), (B) distinct peptides and (C) protein groups from the raw files for the comparison of OT- and IT-type detection using the back-to-back strategy (HCD-IT vs. HCD-OT).** Six different MS data interpretation approaches were used in this study by combining SEQUEST or Mascot with different post-processing methods such as score-based, PeptideProphet, and Percolator.

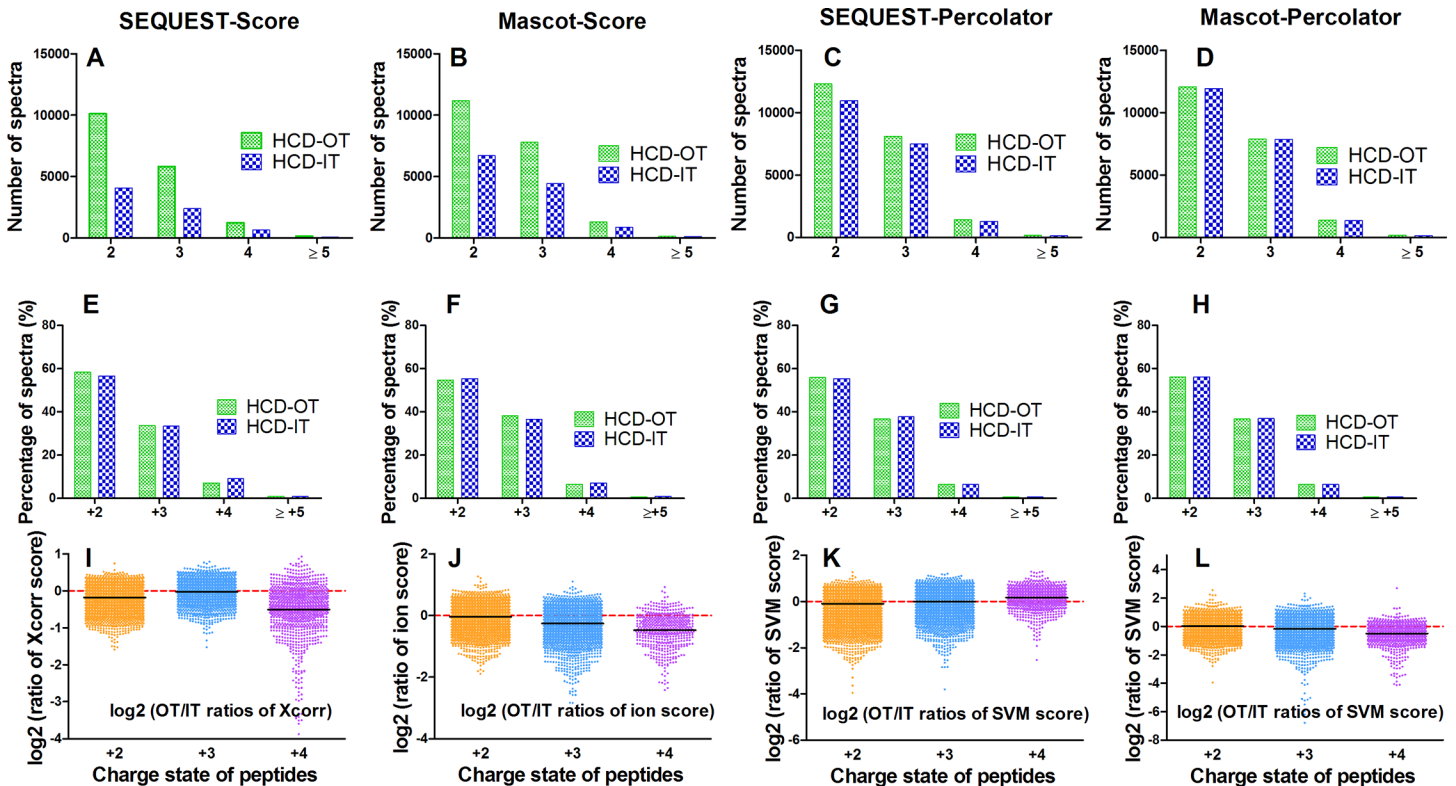
doi:10.1371/journal.pone.0160160.g004

protein groups were identified in the HCD-OT mode using Mascot-Score, while only 13549 PSMs, 9173 distinct peptides and 2275 protein groups in the HCD-IT mode. In this study, SEQUEST-Percolator achieved the highest number of identifications for both HCD-OT and HCD-IT modes. Though the difference between HCD-IT and HCD-OT by SEQUEST-Percolator was not significant ( $p > 0.05$ ), HCD-OT obtained a slightly higher number of identifications than HCD-IT (Fig 4). However, the advantage of speed for IT detection was not fully utilized here because of the back-to-back strategy, while the advantage of accuracy for OT still contributes to improve identification.

The distributions of peptide charge and ratio of searching scores between HCD-IT and HCD-OT for co-identified peptides were also investigated. One replicate of the back-to-back LC-MS analyses of HCD-IT versus HCD-OT was randomly selected. As shown in Fig 5A–5D, because of the fact that HCD-OT identified more total PSMs than HCD-IT (Fig 4), more PSMs for each charge state were still obtained in HCD-OT mode than HCD-IT mode. Different from the results of HCD-IT/CID-IT comparison described above, the percentage distributions of PSMs identified by HCD-IT and HCD-OT were similar, though the numbers were significantly different because of the use of different interpretation approaches (Fig 5E–5H). For instance, more PSMs (20659 vs. 13549) were identified from HCD-OT than HCD-IT using Mascot-Score, but the percentages of different charged peptides were similar between HCD-OT and HCD-IT (54.7% vs. 55.4% for +2 charged peptides, 38.2% vs 36.6% for +3) as shown in Fig 5F. Therefore, the use of OT- and IT-type detection did not affect the identification of peptides with different charge state. For the searching score, since the scoring systems of SEQUEST and Mascot are not optimized for high-resolution MS2 spectra, the scores from these two engines turned out to be higher for peptides from IT than those from OT, and thus the majority of log<sub>2</sub> transformed OT/IT ratios of scores were below zero (Fig 5I & 5J). Percolator, on the other hand, has considered many features including the accuracy of MS2 spectra[22], which improves the scoring system of these two search engines and more similar scores were given for the peptides co-identified by HCD-OT and HCD-IT as expected (Fig 5K & 5L).

In summary, a slight increase of identifications were achieved from HCD-OT than HCD-IT in the back-to-back LC-MS analyses when SEQUEST-Percolator, a relatively fair platform, was





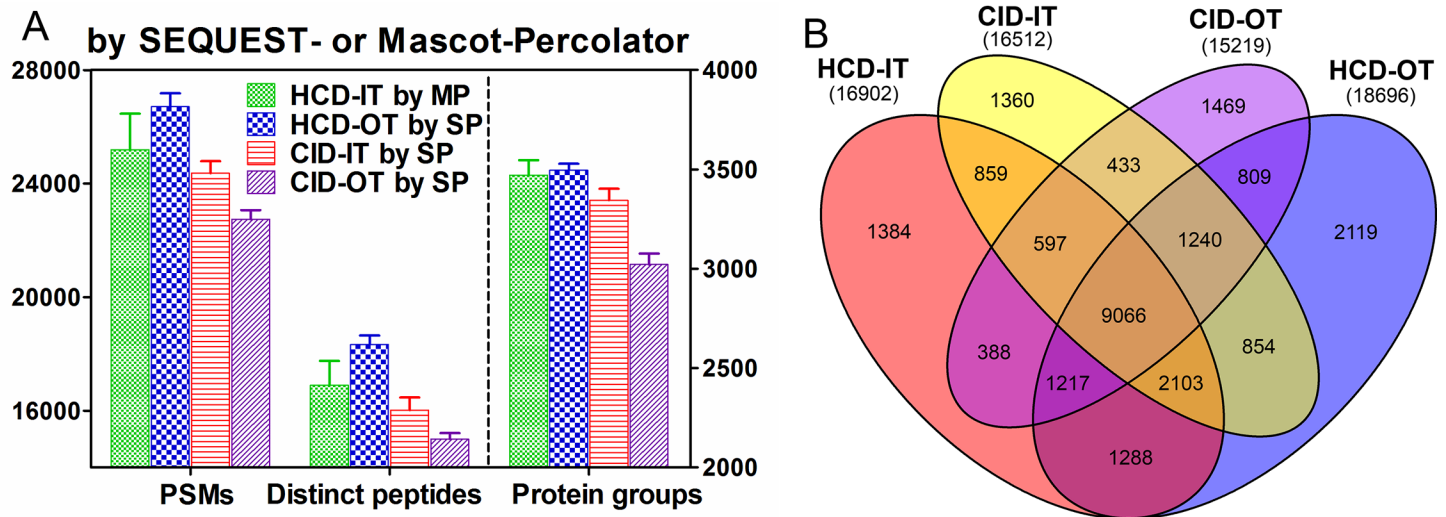
**Fig 5. The comparison of OT- and IT-type detection using the back-to-back strategy (HCD-IT vs. HCD-OT).** (A-D) Number of spectra for +2, +3, +4 and  $\geq 5$  charged ions using different data interpretation methods. (E-H) Percentage of spectra for for +2, +3, +4 and  $\geq 5$  charged ions using different data interpretation methods. (I-L) Log<sub>2</sub> (OT/IT ratio of score) distributions for +2, +3 and +4 charged peptides. Xcorr for SEQUEST, ion score for Mascot, SVM score for Percolator related combinations were presented. One replicate of the back-to-back LC-MS analyses of HCD-IT versus HCD-OT was randomly selected.

doi:10.1371/journal.pone.0160160.g005

used for MS data interpretation. Moreover, lower SEQUEST-Xcorr or Mascot-ion scores were achieved for the peptides analyzed in the OT mode than that in the IT mode because of the sub-optimized scoring system. It is important to note that, because of the different mechanisms of ion detection by OT and IT, higher AGC target and maximum injection were used for OT than IT in routine analysis.

### Evaluating different MS2 acquisition strategies: HCD-IT, HCD-OT, CID-IT, and CID-OT

To further address which combination of fragmentation and detection would give the best performance, the human cell proteome (PANC-1 cells) was also used to evaluate different MS2 strategies in Orbitrap Fusion. Four MS2 acquisition methods were theoretically available in Orbitrap Fusion as shown in S2 Fig. Of these four methods, HCD-IT, HCD-OT, and CID-IT were commonly used and applied in various proteomics studies. Here three replicates for each of them were alternatively analyzed. As shown in S4 Fig, except Mascot-PeptideProphet, HCD-OT achieved the best performance in PSMs, distinct peptides and protein groups in other five combinations. More IDs were given when analyzing HCD-IT by Mascot than SEQUEST; on the contrary, more IDs were given when analyzing CID-IT by SEQUEST than Mascot as described above, no matter which post-processing approach was used. To perform the relatively fair comparison between these three MS2 pipelines, the best interpretation



**Fig 6. Comparison of four MS2 acquisition methods: HCD-OT, HCD-IT, CID-IT and CID-OT in Orbitrap Fusing using suitable data interpretation approaches.** (A) Peptide spectra matches (PSMs), distinct peptides and protein groups from separate analyses of human cell proteome. (B) The unique peptide overlaps between these four MS2 acquisition methods. MP: Mascot-Percolator; SP: SEQUEST-Percolator.

doi:10.1371/journal.pone.0160160.g006

methods for each pipeline were selected. Here HCD-OT and CID-IT by SEQUEST-Percolator (SP) and HCD-IT by Mascot-Percolator (MP) were analyzed. As shown in Fig 6A, these three MS2 pipelines achieved similar results in PSMs, distinct peptides and protein groups. Statistical analysis between any two of them showed that significant difference ( $p < 0.05$ ) was only observed between HCD-OT (an average of 26708 PSMs, 18332 distinct peptides, 3495 protein groups) and CID-IT (24367, 16022, 3346). For HCD-IT by MP, an average of 25192 PSMs, 16901 distinct peptides and 3471 protein groups were observed. Moreover, the file size of HCD-OT here is around 1.4 gigabyte (GB), smaller than HCD-IT (~3.1 GB) and CID-IT (~2.8 GB). All the MS/MS data is collected in the centroid mode. The smaller file size of HCD-OT enabled multiplex quantitative analyses (i.e. 40-plex) for algorithms such as Thermo SIEVE and reduced the amount of time of identification and quantification procedures. Therefore, for Orbitrap Fusion, the MS2 pipelines of HCD-OT were suggested as the first choice for identification and multiplex quantification analysis.

In addition, the MS2 acquisition method of CID-OT is theoretically available on Fusion Orbitrap, though it has not been discussed in previous studies. In this study three replicates of LC-MS/MS runs by the CID-OT pipeline were performed by using the same LC column and peptide mixture. Because of the comparable IDs from HCD-IT and CID-IT, the similar result may be achieved in the comparison of HCD-OT and CID-OT if the efficiency of ion transfer is still similar between them. However, CID-OT, which performs the fragmentation in CID and MS2 detection in Orbitrap, is not an effective MS2 acquisition method because of the long distance between the fragmentation cell and Orbitrap as shown in S2 Fig. Surprisingly, CID-OT achieved more than 3,000 protein groups (by SEQUEST-Percolator) from one-shot LC-MS/MS run (Fig 6A), only a decrease of 10%-14% than other three MS2 pipelines.

No significant difference ( $p > 0.05$ ) were observed between CID-IT and HCD-IT (Fig 6A), consistent with the results from the back-to-back analyses. To further validate the finding described above that more highly charged peptides were achieved from HCD than CID, the charge state of peptides were further analyzed on the separate LC-MS/MS runs here. The respective runs with highest identifications for each method were selected for comparison. As shown in S5A Fig, since HCD-OT achieved more total identifications than HCD-IT and

CID-IT, more identifications at each charge state were also observed. However, the percentage of doubly charged peptides from HCD-OT or HCD-IT was still lower than that from CID-IT (S5B Fig), agree well with the results from the back-to-back analyzes above.

From our results described above, the significantly different results will be achieved when analyzing the same CID-IT data with SEQUEST and Mascot. This may explain the divergent conclusions recently reported by Mann's group [7] and Gygi's group [2] who used different search engines and post-processing algorithms to evaluate HCD- and CID-type fragmentation (actually HCD-OT versus CID-IT). Though the same dataset from Mann's group was individually analyzed by these two groups, the opposite conclusions were achieved. SEQUEST- linear discriminant analysis (LDA) and MaxQuant (Mascot v2.2) were respectively used in Gygi's group and Mann's group. PeptideProphet originally uses the LDA classifier to separate correct and incorrect PSMs in an unsupervised fashion (i.e., without decoy information) [17] and is improved by the semi-supervised approach using decoy PSMs to estimate probabilities from the discriminant scores [24]. Therefore, the MS data interpretation methods they used are respectively similar as SEQUEST-PeptideProphet and Mascot-Score we tested here. In this study similar IDs were achieved from HCD-OT and CID-IT analysis of the PANC-1 cell sample when using SEQUEST-PeptideProphet (S4B Fig), but significantly more PSMs (~3-fold) were achieved from HCD-OT than CID-IT (S4D Fig) using Mascot-Score. Because of this fact, it is not surprising that different conclusions could be made from the same dataset when using different interpretation methods. Therefore, to achieve a fair comparison, respective optimal data interpretation methods were required when performing comparison between different MS2 techniques. In this study, Percolator associated MS data interpretation methods were suggested based on our results.

## Overlaps between different MS2 pipelines

Further investigation of peptide identification overlap between HCD-IT, HCD-OT, CID-IT and CID-OT was performed by using the results from SEQUEST-Percolator. The raw files with highest identifications by SEQUEST-Percolator for each MS2 acquisition method were selected. As shown in Fig 6B, 15219, 16512, 16902 and 18696 distinct peptides were respectively identified from data generated by CID-OT, CID-IT, HCD-IT and HCD-OT. A total of 25186 unique peptides were detected from these four raw files, and 71%-77% overlaps (double shared peptides over the sum of peptides) were observed between any two of them. The highest number of unique peptides (21945) was achieved from the combination of HCD-OT and CID-IT among the combinations of any two of them (S6A Fig). Moreover, we further investigated the overlaps between two technical replicates using these four MS2 strategies respectively. As shown in S6B Fig, slight higher overlaps (80%-83%) between technical replicates were observed. The highest number of unique peptides (22169) is from the combination of two HCD-OT replicates with an overlap of 80%. Therefore, the combination of different MS2 pipelines (i.e. HCD-OT & CID-IT) may be not necessary for identification here.

## Conclusions

Here we have investigated multiple MS2 acquisition methods on the Orbitrap Fusion, a mass spectrometry that incorporates quadrupole, Orbitrap, and dual-cell ion trap analyzers and enables to enhance proteome coverage and peptide identification rates. Percolator associated data interpretation approach was selected as a fair platform to perform the respective comparisons of different methods of fragmentation, detection, and MS2 acquisition. Different charged peptides were observed to be preferred to CID or HCD fragmentation events. HCD-OT and HCD-IT these two MS2 acquisition methods were suggested for peptide identification using

Orbitrap Fusion. Further MS2 acquisition methods, such as ETD-OT, ETD-IT, can also be incorporated and applied for highly charged peptides. Moreover, the combination of CID for doubly charged peptides and HCD for highly charged peptides is possible when both of their collision cells are close to the Orbitrap analyzer or with respective ones, or the same precursor derived fragmentation ions from CID and HCD are stored in the same cell and then detected together. All these strategies are possible in principle and improvements in mass spectrometry, data interpretation techniques, and clinical application are interesting for the proteomics community.

## Supporting Information

**S1 Fig. Evaluations of some parameters for MS instrument and database search.**

(PDF)

**S2 Fig. The four MS2 acquisition methods in the Orbitrap Fusion mass spectrometry analyzed in this study.**

(PDF)

**S3 Fig. The comparison between HCD-IT by MP and CID-IT by SP for separate LC-MS/MS analyses (N = 3).**

(PDF)

**S4 Fig. Peptide spectra matches (PSMs), distinct peptides and protein groups from the separate LC-MS/MS analyses for the comparison of HCD-OT, HCD-IT, and CID-IT.**

(PDF)

**S5 Fig. The comparison of HCD-OT by SP, HCD-IT by MP, and CID-IT by SP for separate LC-MS/MS analyses.**

(PDF)

**S6 Fig.** Respective overlaps between (A) different MS2 acquisition methods and (B) duplicate analyses of each MS2 acquisition methods.

(PDF)

## Acknowledgments

This work was supported by American Heart Association (AHA) award 12SDG9450036 (JQ), by a Center of Protein Therapeutics Industrial Award (JQ), by NIH grants U54HD071594 (JQ) and HL103411 (JQ).

## Author Contributions

Conceived and designed the experiments: CT JQ. Performed the experiments: CT JL. Analyzed the data: CT QS. Contributed reagents/materials/analysis tools: QS YS JQ. Wrote the paper: CT JQ. The software BuildSummary used in analysis: QS YS.

## References

1. Hebert AS, Richards AL, Bailey DJ, Ulbrich A, Coughlin EE, Westphall MS, et al. The one hour yeast proteome. *Molecular & cellular proteomics: MCP*. 2014; 13(1):339–47. Epub 2013/10/22. doi: [10.1074/mcp.M113.034769](https://doi.org/10.1074/mcp.M113.034769) PMID: [24143002](https://pubmed.ncbi.nlm.nih.gov/24143002/); PubMed Central PMCID: PMC3879625.
2. Jedrychowski MP, Huttlin EL, Haas W, Sowa ME, Rad R, Gygi SP. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Molecular & cellular proteomics: MCP*. 2011; 10(12):M111 009910. Epub 2011/09/16. doi: [10.1074/mcp.M111.009910](https://doi.org/10.1074/mcp.M111.009910) PMID: [21917720](https://pubmed.ncbi.nlm.nih.gov/21917720/); PubMed Central PMCID: PMC3237076.

3. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods*. 2007; 4(9):709–12. Epub 2007/08/28. doi: [10.1038/nmeth1060](https://doi.org/10.1038/nmeth1060) PMID: [17721543](https://pubmed.ncbi.nlm.nih.gov/17721543/).
4. McAlister GC, Phanstiel DH, Brumbaugh J, Westphall MS, Coon JJ. Higher-energy collision-activated dissociation without a dedicated collision cell. *Molecular & cellular proteomics: MCP*. 2011; 10(5):O111009456. Epub 2011/03/12. doi: [10.1074/mcp.O111.009456](https://doi.org/10.1074/mcp.O111.009456) PMID: [21393638](https://pubmed.ncbi.nlm.nih.gov/21393638/); PubMed Central PMCID: PMC3098599.
5. Second TP, Blethrow JD, Schwartz JC, Merrihew GE, MacCoss MJ, Swaney DL, et al. Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. *Analytical chemistry*. 2009; 81(18):7757–65. Epub 2009/08/20. doi: [10.1021/ac901278y](https://doi.org/10.1021/ac901278y) PMID: [19689114](https://pubmed.ncbi.nlm.nih.gov/19689114/); PubMed Central PMCID: PMC2810160.
6. Olsen JV, Schwartz JC, Griep-Raming J, Nielsen ML, Damoc E, Denisov E, et al. A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Molecular & cellular proteomics: MCP*. 2009; 8(12):2759–69. Epub 2009/10/16. doi: [10.1074/mcp.M900375-MCP200](https://doi.org/10.1074/mcp.M900375-MCP200) PMID: [19828875](https://pubmed.ncbi.nlm.nih.gov/19828875/); PubMed Central PMCID: PMC2816009.
7. Nagaraj N, D'Souza RC, Cox J, Olsen JV, Mann M. Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *Journal of proteome research*. 2010; 9(12):6786–94. Epub 2010/09/30. doi: [10.1021/pr100637q](https://doi.org/10.1021/pr100637q) PMID: [20873877](https://pubmed.ncbi.nlm.nih.gov/20873877/).
8. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*. 1994; 5(11):976–89. Epub 1994/11/01. doi: [10.1016/1044-0305\(94\)80016-2](https://doi.org/10.1016/1044-0305(94)80016-2) PMID: [24226387](https://pubmed.ncbi.nlm.nih.gov/24226387/).
9. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999; 20(18):3551–67. Epub 1999/12/28. doi: [10.1002/\(SICI\)1522-2683\(19991201\)20:18<3551::AID-ELPS3551>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2) PMID: [10612281](https://pubmed.ncbi.nlm.nih.gov/10612281/).
10. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, et al. Open mass spectrometry search algorithm. *Journal of proteome research*. 2004; 3(5):958–64. Epub 2004/10/12. doi: [10.1021/pr0499491](https://doi.org/10.1021/pr0499491) PMID: [15473683](https://pubmed.ncbi.nlm.nih.gov/15473683/).
11. Tabb DL, Fernando CG, Chambers MC. MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *Journal of proteome research*. 2007; 6(2):654–61. Epub 2007/02/03. doi: [10.1021/pr0604054](https://doi.org/10.1021/pr0604054) PMID: [17269722](https://pubmed.ncbi.nlm.nih.gov/17269722/); PubMed Central PMCID: PMC2525619.
12. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research*. 2011; 10(4):1794–805. Epub 2011/01/25. doi: [10.1021/pr101065j](https://doi.org/10.1021/pr101065j) PMID: [21254760](https://pubmed.ncbi.nlm.nih.gov/21254760/).
13. Wenger CD, Coon JJ. A proteomics search algorithm specifically designed for high-resolution tandem mass spectra. *Journal of proteome research*. 2013; 12(3):1377–86. Epub 2013/01/18. doi: [10.1021/pr301024c](https://doi.org/10.1021/pr301024c) PMID: [23323968](https://pubmed.ncbi.nlm.nih.gov/23323968/); PubMed Central PMCID: PMC3586292.
14. Kim S, Pevzner PA. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nature communications*. 2014; 5:5277. Epub 2014/11/02. doi: [10.1038/ncomms6277](https://doi.org/10.1038/ncomms6277) PMID: [25358478](https://pubmed.ncbi.nlm.nih.gov/25358478/).
15. Dorfer V, Pichler P, Stranzl T, Stadlmann J, Taus T, Winkler S, et al. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *Journal of proteome research*. 2014; 13(8):3679–84. Epub 2014/06/10. doi: [10.1021/pr500202e](https://doi.org/10.1021/pr500202e) PMID: [24909410](https://pubmed.ncbi.nlm.nih.gov/24909410/); PubMed Central PMCID: PMC4119474.
16. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature methods*. 2007; 4(3):207–14. Epub 2007/03/01. doi: [10.1038/nmeth1019](https://doi.org/10.1038/nmeth1019) PMID: [17327847](https://pubmed.ncbi.nlm.nih.gov/17327847/).
17. 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. *Analytical chemistry*. 2002; 74(20):5383–92. Epub 2002/10/31. PMID: [12403597](https://pubmed.ncbi.nlm.nih.gov/12403597/).
18. Sadygov RG, Liu H, Yates JR. Statistical models for protein validation using tandem mass spectral data and protein amino acid sequence databases. *Analytical chemistry*. 2004; 76(6):1664–71. Epub 2004/03/17. doi: [10.1021/ac035112y](https://doi.org/10.1021/ac035112y) PMID: [15018565](https://pubmed.ncbi.nlm.nih.gov/15018565/).
19. Ulintz PJ, Zhu J, Qin ZS, Andrews PC. Improved classification of mass spectrometry database search results using newer machine learning approaches. *Molecular & cellular proteomics: MCP*. 2006; 5(3):497–509. Epub 2005/12/03. doi: [10.1074/mcp.M500233-MCP200](https://doi.org/10.1074/mcp.M500233-MCP200) PMID: [16321970](https://pubmed.ncbi.nlm.nih.gov/16321970/).
20. Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature methods*. 2007; 4(11):923–5. Epub 2007/10/24. doi: [10.1038/nmeth1113](https://doi.org/10.1038/nmeth1113) PMID: [17952086](https://pubmed.ncbi.nlm.nih.gov/17952086/).

21. Klammer AA, Reynolds SM, Bilmes JA, MacCoss MJ, Noble WS. Modeling peptide fragmentation with dynamic Bayesian networks for peptide identification. *Bioinformatics*. 2008; 24(13):i348–56. Epub 2008/07/01. doi: [10.1093/bioinformatics/btn189](https://doi.org/10.1093/bioinformatics/btn189) PMID: [18586734](https://pubmed.ncbi.nlm.nih.gov/18586734/); PubMed Central PMCID: PMC2665034.
22. Spivak M, Weston J, Bottou L, Kall L, Noble WS. Improvements to the percolator algorithm for Peptide identification from shotgun proteomics data sets. *Journal of proteome research*. 2009; 8(7):3737–45. Epub 2009/04/24. doi: [10.1021/pr801109k](https://doi.org/10.1021/pr801109k) PMID: [19385687](https://pubmed.ncbi.nlm.nih.gov/19385687/); PubMed Central PMCID: PMC2710313.
23. Granholm V, Noble WS, Kall L. A cross-validation scheme for machine learning algorithms in shotgun proteomics. *BMC bioinformatics*. 2012; 13 Suppl 16:S3. Epub 2012/11/28. doi: [10.1186/1471-2105-13-S16-S3](https://doi.org/10.1186/1471-2105-13-S16-S3) PMID: [23176259](https://pubmed.ncbi.nlm.nih.gov/23176259/); PubMed Central PMCID: PMC3489528.
24. Choi H, Nesvizhskii AI. Semisupervised model-based validation of peptide identifications in mass spectrometry-based proteomics. *Journal of proteome research*. 2008; 7(1):254–65. Epub 2007/12/28. doi: [10.1021/pr070542g](https://doi.org/10.1021/pr070542g) PMID: [18159924](https://pubmed.ncbi.nlm.nih.gov/18159924/).
25. Tu C, Li J, Jiang X, Sheflin LG, Pfeffer BA, Behringer M, et al. Ion-current-based proteomic profiling of the retina in a rat model of Smith-Lemli-Opitz syndrome. *Molecular & cellular proteomics: MCP*. 2013; 12(12):3583–98. Epub 2013/08/28. doi: [10.1074/mcp.M113.027847](https://doi.org/10.1074/mcp.M113.027847) PMID: [23979708](https://pubmed.ncbi.nlm.nih.gov/23979708/); PubMed Central PMCID: PMC3861709.
26. Duan X, Young R, Straubinger RM, Page B, Cao J, Wang H, et al. A straightforward and highly efficient precipitation/on-pellet digestion procedure coupled with a long gradient nano-LC separation and Orbitrap mass spectrometry for label-free expression profiling of the swine heart mitochondrial proteome. *Journal of proteome research*. 2009; 8(6):2838–50. Epub 2009/03/18. doi: [10.1021/pr900001t](https://doi.org/10.1021/pr900001t) PMID: [19290621](https://pubmed.ncbi.nlm.nih.gov/19290621/); PubMed Central PMCID: PMC2734143.
27. Richards AL, Hebert AS, Ulbrich A, Bailey DJ, Coughlin EE, Westphall MS, et al. One-hour proteome analysis in yeast. *Nature protocols*. 2015; 10(5):701–14. Epub 2015/04/10. doi: [10.1038/nprot.2015.040](https://doi.org/10.1038/nprot.2015.040) PMID: [25855955](https://pubmed.ncbi.nlm.nih.gov/25855955/).
28. Tu C, Sheng Q, Li J, Ma D, Shen X, Wang X, et al. Optimization of Search Engines and Postprocessing Approaches to Maximize Peptide and Protein Identification for High-Resolution Mass Data. *Journal of proteome research*. 2015; 14(11):4662–73. Epub 2015/09/22. doi: [10.1021/acs.jproteome.5b00536](https://doi.org/10.1021/acs.jproteome.5b00536) PMID: [26390080](https://pubmed.ncbi.nlm.nih.gov/26390080/).
29. Elias JE, Haas W, Faherty BK, Gygi SP. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nature methods*. 2005; 2(9):667–75. PMID: [16118637](https://pubmed.ncbi.nlm.nih.gov/16118637/).
30. Sheng Q, Dai J, Wu Y, Tang H, Zeng R. BuildSummary: using a group-based approach to improve the sensitivity of peptide/protein identification in shotgun proteomics. *Journal of proteome research*. 2012; 11(3):1494–502. Epub 2012/01/06. doi: [10.1021/pr200194p](https://doi.org/10.1021/pr200194p) PMID: [22217156](https://pubmed.ncbi.nlm.nih.gov/22217156/).