Supplemental Material

Data Processing with Hardklör and Bullseye

To adjust the monoisotopic precursor mass for MS/MS spectra, the high resolution mass spectra were analyzed by Hardklör[1], a feature detection algorithm that identifies the monoisotopic masses of peptide isotope distributions (PIDs). Hardklör parameters used were a correlation threshold of 0.9, a maximum depth of 3, and a maximum charge state of 5. Hardklör results were then analyzed by *Bullseye* to identify PIDs that were persistent (Supplemental Figure 1). A PID was persistent if an isotope distribution was detected by Hardklör at the same charge and with a monoisotopic mass within 10 ppm in at least three of four consecutive MS scans. A MS/MS spectrum was matched to a persistent PID if the fragmentation event occurred in the mass range of the isotope distribution and within 30 seconds of the PID's elution profile (Supplemental Figure 2). A ± 30 second window was used because ions are rarely fragmented at the apex of their chromatographic elution and, thus, a MS/MS spectrum's corresponding PID might not be detected in the adjacent MS scans. For a matching persistent PID, the monoisotopic mass and charge state were assigned to the fragmentation spectrum. In cases where multiple matches were found, multiple precursor mass and charge entries were added to the data file as additional "Z-lines" with respect to the MS2 file format[2]. If no matching PID could be found for a fragmentation spectrum, the fragmentation m/z was assigned as the precursor mass with an approximate charge state given (+1 or +2/+3). Information about obtaining *Bullseye* can be found at http://proteome.gs.washington.edu/software/bullseye/.

The monoisotopic precursor mass for the MS/MS spectra were also determined by the embedded firmware during data acquisition. For this dataset, during the conversion of the RAW file to an MS2 file[2], the mass and charge state values for the MS/MS spectra were adjusted to

those values. If no monoisotopic m/z was available, the m/z selected for fragmentation by data-dependent acquisition was used.

Comparison between Bullseye and instrument data-system reported monoisotopic mass

Bullseye performance in assigning a correct monoisotopic mass was compared to existing values stored in the RAW data files during spectra acquisition. Charge state and monoisotopic mass information for many, but not all, MS/MS spectra were contained within RAW data files. This information was computed by the instrument firmware from the first ~100 milliseconds of the Fourier transform transient acquired during data acquisition. We compared the data from each set where a precursor mass was detected and assigned to the respective MS/MS spectrum.

Data acquired from the *Saccharomyces cerevisiae* sample contained 3,530 MS spectra and 17,650 MS/MS spectra. *Bullseye* was able to identify at least one precursor PID for 10,824 of the MS/MS spectra (61.3%). For MS/MS spectra with multiple possible precursor PIDs, a separate entry was created for each precursor PID in the MS2 file[2]. There were 7,666 MS/MS spectra assigned a single precursor PIDs and 3,158 MS/MS spectra were assigned multiple precursor PIDs. In total, 14,774 database searches were performed.

The dataset output by *Bullseye* was compared to monoisotopic m/z values determined by the instrument during data acquisition ("instrument-precursor" dataset). Of the 17,650 MS/MS spectra, the firmware running on the embedded computer assigned a precursor mass for 12,622 spectra (71.5%). Multiple precursor PIDs were not assigned. The onboard computer assigned a precursor mass to a greater number of spectra than *Bullseye*; however, by examining the scan events that are assigned a precursor mass we found that the *Bullseye* dataset only overlaps with 66% (8,344 of 12,622 spectra) of the instrument-precursor dataset (Supplemental Figure 3). The

Bullseye dataset contained 2,480 spectra that have no precursor mass reported in instrument-precursor. The instrument-precursor dataset contained 4,278 spectra not matched by *Bullseye*.

To effectively compare the performance of *Bullseye* with the instrument determined results, we examined the differences in the precursor mass values that were assigned by the two methods. MS/MS spectra were searched against a yeast peptide database using the SEQUEST algorithm with a precursor mass tolerance of ± 3 Da. Search results were separated into two groups. The first group contained the 8,344 spectra that are present in both datasets (Supplemental Figure 4A and 4B) and the second group contained those spectra that were exclusive to each dataset (Supplemental Figure 4C and 4D). As shown in Figure 4A and 4B, *Bullseye* was able to assign precursor masses with fewer instances of the M+1 and M+2 isotope peaks incorrectly assigned as the monoisotopic mass. We used XCorr values as a score metric to calculate false discovery rates (FDR) to estimate the number of positive PSMs that were misassigned. For the *Bullseye* dataset, in a mass window around the M+1 and M+2 mass differences, there were 12 PSMs at a FDR less than 0.01. For the instrument-precursor dataset, there were 31 PSMs (FDR 0.01) with an M+1 and M+2 mass assignment error.

Bullseye uses mass measurements averaged over the range of a peptide's elution profile to assign a precursor mass to the MS/MS spectrum. The potential benefit is that an averaged mass minimizes the effect of fluctuations in mass measurement that normally occur over the course of a peptide's elution. For this dataset, the 2 μg of unfractionated whole cell lysate injected onto the chromatography column is more than sufficient to fill the FTICR cell to the Automatic Gain Control limit (2×10^6 ions), thus minimizing the effect of mass measurement shifts due to total ion abundance. It has been documented previously that an ion's abundance can have an affect on ICR MS measurements independent of the total ion abundance[3, 4]. This may

explain in part the fluctuations in mass measurement that we observed over the course of a peptide's chromatographic elution.

A histogram was constructed of the mass differences between the exact peptide mass and the mass assigned by *Bullseye* or the mass present in the RAW file spectrum header (Supplemental Figure 5). Only MS/MS scan events where both methods assigned the same charge state and peptide ID at an FDR less than 0.01 were used to construct the histogram. For these spectra, *Bullseye* assigns precursor masses with more accuracy compared to the mass assigned by the instrument firmware. This improvement is not surprising because *Bullseye* uses the higher resolution data from the entire peptide chromatographic elution profile to perform its calculation.

A powerful method implemented in *Bullseye* is the use of the peptide isotope distribution as it persists chromatographically over time[5, 6]. In our analysis, we used the heuristic that a PID be detected in three of four sequential MS scans before being accepted. Besides improving mass accuracy by averaging the mass measurements of the scans, this acted as a filter to discard MS/MS scans of low quality. In panels C and D of Supplemental Figure 4, search results from spectra with precursor mass assignments made exclusively by *Bullseye* and the instrument are shown. The instrument-precursor dataset, shown in panel D, contained 70% more spectra than the *Bullseye* exclusive dataset (Panel C). However at a FDR of 0.01, the instrument-precursor dataset contained only 10 PSMs compared to 255 PSMs in the *Bullseye* exclusive dataset. The average XCorr score of the instrument-precursor dataset was also notably lower than the *Bullseye* exclusive dataset. *Bullseye* did not find a precursor isotope distribution for these MS/MS scans because they are likely low intensity signals or non-peptide signals. The MS/MS scans that *Bullseye* did not find a precursor mass for are placed in the “unmatched” dataset. In

this dataset, only 5 PSMs were found from 10,800 spectra at a FDR of 0.05 (Supplemental Table 1), indicating that *Bullseye'*s persistent PID detection worked well in removing poor scoring spectra.

While the resolution and accuracy of the precursor mass measurement in Fourier transform mass analyzers is very high, the large isolation width typically used for MS/MS on the LTQ can result in multiple molecular species contained within a single spectrum. The presence of multiple peptides in a single fragmentation spectrum is a function of the complexity of the sample and the separation of peptides prior to detection. *Bullseye* and software tools described by Luethy *et al.*[7] and Scherl *et al*.[8] are capable of generating separate MS/MS entries for multiple detected precursor masses found within the isolation window. In the dataset used in this work, *Bullseye* assigned multiple precursor ions to 3,158 (29%) fragmentation spectra. Of those, 43 spectra were found to have multiple unique peptide identifications with FDRs less than 0.05 (data not shown). The percentage of spectra that contained multiple precursor ions where multiple peptides can be assigned with confidence is low (1.4%), although it is important to note that the SEQEUST search algorithm used was not able to take into account the possibility of multiple peptides when scoring MS/MS spectra.

References

1. Hoopmann, M. R.; Finney, G. L.; MacCoss, M. J., High-speed data reduction, feature detection, and MS/MS spectrum quality assessment of shotgun proteomics data sets using high-resolution mass spectrometry. *Anal Chem* **2007,** 79, (15), 5620-32.
2. McDonald, W. H.; Tabb, D. L.; Sadygov, R. G.; MacCoss, M. J.; Venable, J.; Graumann, J.; Johnson, J. R.; Cociorva, D.; Yates, J. R., 3rd, MS1, MS2, and SQT-three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Commun Mass Spectrom* **2004,** 18, (18), 2162-8.
3. Williams, D. K., Jr.; Kovach, A. L.; Muddiman, D. C.; Hanck, K. W., Utilizing artificial neural networks in MATLAB to achieve parts-per-billion mass measurement accuracy with a fourier transform ion cyclotron resonance mass spectrometer. *J Am Soc Mass Spectrom* **2009,** 20, (7), 1303-10.

4. Wong, R. L.; Amster, I. J., Experimental Evidence for Space-Charge Effects between Ions of the Same Mass-to-Charge in Fourier-Transform Ion Cyclotron Resonance Mass Spectrometry. *Int J Mass Spectrom* **2007,** 265, (2-3), 99-105.
5. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **2008,** 26, (12), 1367-72.
6. Bakalarski, C. E.; Elias, J. E.; Villen, J.; Haas, W.; Gerber, S. A.; Everley, P. A.; Gygi, S. P., The impact of peptide abundance and dynamic range on stable-isotope-based quantitative proteomic analyses. *J Proteome Res* **2008,** 7, (11), 4756-65.
7. Luethy, R.; Kessner, D. E.; Katz, J. E.; Maclean, B.; Grothe, R.; Kani, K.; Faca, V.; Pitteri, S.; Hanash, S.; Agus, D. B.; Mallick, P., Precursor-ion mass re-estimation improves peptide identification on hybrid instruments. *J Proteome Res* **2008,** 7, (9), 4031-9.
8. Scherl, A.; Tsai, Y. S.; Shaffer, S. A.; Goodlett, D. R., Increasing information from shotgun proteomic data by accounting for misassigned precursor ion masses. *Proteomics* **2008,** 8, (14), 2791-7.

Supplemental Table 1

Spectra Set	Database Search Mass Tolerance (precursor mass type)	Positive PSM (0.001 FDR)	Positive PSM (0.01 FDR)	Positive PSM (0.05 FDR)	Number of Spectra Searched[a]	Average Number of Peptides Searched per Spectrum[a]	Search Time (min)[ab]
Bullseye Unmatched	± 3 Da (Average)	5	5	5	10800	159966	39.45
Bullseye Matched	± 5 ppm (Monoisotopic)	1963	2543	3093	14774	1563	7.16
Instrument-precursor	± 5 ppm (Monoisotopic)	1403	2121	2525	21928	988	8.92

[a] Results from target search.

[b] Average of three replicate searches. SEQUEST processes were distributed to 4 cores of an AMD 64-bit 2.5 Ghz processor.

Supplemental Figure Legends

Supplemental Figure 1. Data processing scheme used to assign accurate precursor mass to an MS/MS spectrum. High resolution MS and MS/MS scan data are extracted from RAW data files. MS data is analyzed with Hardklör to identify peptide isotope distributions. *Bullseye* attempts to match a persistent peptide isotope distribution to each MS/MS scan. MS/MS scans are separated into two data files based on whether a matching PID is found. MS/MS scans with matching PIDs are searched with a narrow mass tolerance window. MS/MS scans without a matching PID are searched with a wide mass tolerance window.

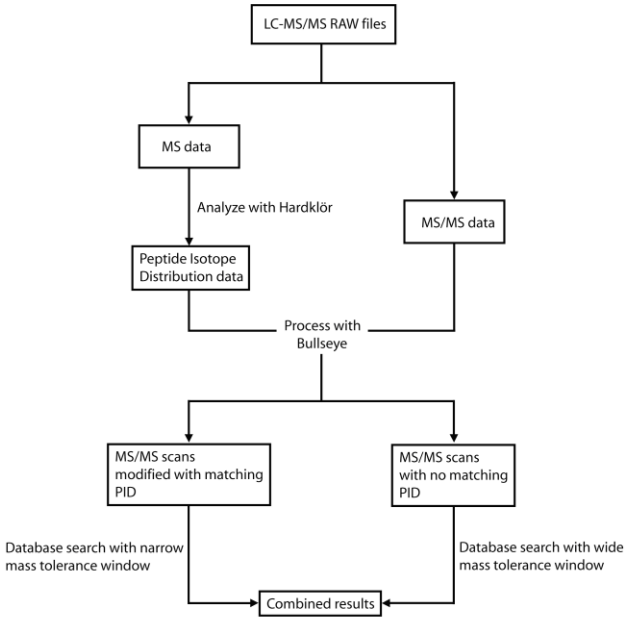
Supplemental Figure 2. An example of the complexity of mapping individual persistent peptide isotope distributions (PPIDs) to a MS/MS spectrum. The MS full scan data is plotted and the center of a MS/MS scan event is indicated by the blue circle. Two separate PPIDs are identified by *Bullseye* (red and green highlighted signals) and are assigned to the MS/MS scan event. *Bullseye* generates a target box (red and green boxes) around each PPID and MS/MS scans that lie within the target box have their precursor mass values reassigned to the monoisotopic mass of the PPID.

Supplemental Figure 3. Overlap of MS/MS spectra with precursor ion monoisotopic masses identified by *Bullseye* or LTQ-FT Ultra firmware. MS/MS spectra that have monoisotopic precursor masses identified by *Bullseye* or present in the RAW file header were compared against each other. *Bullseye* and the instrument firmware both identify monoisotopic parental masses for 8344 spectra. *Bullseye* identifies a monoisotopic precursor mass for an addition 2480 spectra that are not reported by the instrument firmware, and there were an additional 4278

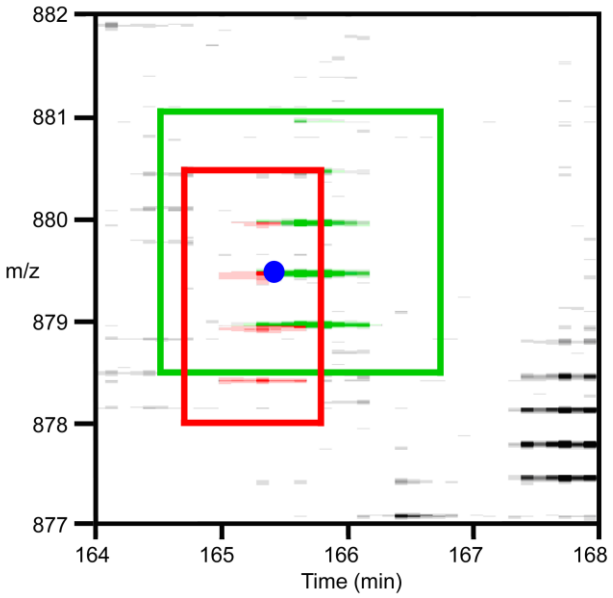
spectra with monoisotopic precursor masses in the RAW file header where *Bullseye* did not identify a monoisotopic parental mass.

Supplemental Figure 4. Comparison of *Bullseye* and the LTQ-FT Ultra instrument reported precursor monoisotopic mass values. MS/MS data with matched precursor monoisotopic masses from *Bullseye* and the instrument's onboard processor were searched against a yeast protein database with SEQUEST. The SEQUEST XCorr values and mass differences were plotted using the top scoring peptide match for each spectrum. Spectra were plotted separately depending on whether a precursor mass was detected in one or both methods. A) *Bullseye* assigned masses on the data subset where a precursor mass was detected in both methods. B) Instrument assigned masses on the data subset where a precursor mass was detected in both methods. C) Spectra with *Bullseye* assigned masses that were not detected by the instrument. D) Spectra with instrument assigned masses that were not detected by *Bullseye*.

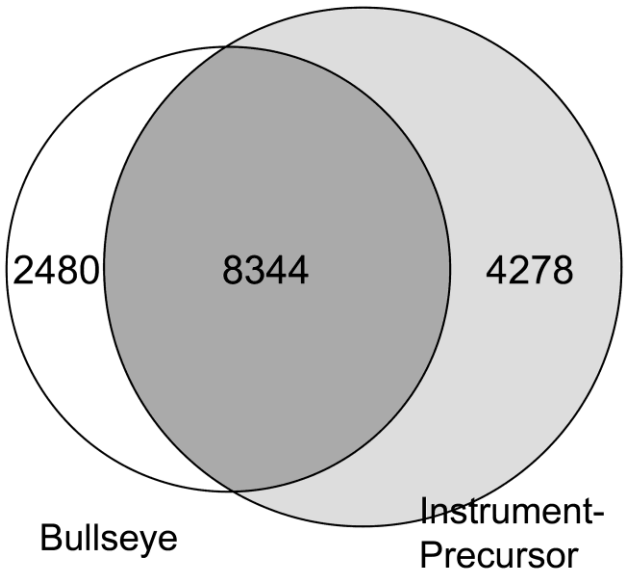
Supplemental Figure 5. Histogram of the deviation from theoretical mass for precursor mass assignments made by *Bullseye* and the LTQ-FT Ultra firmware. The mass differences (instrument assigned - calculated) for peptide spectrum matches with a false discovery rate less than 0.01 were calculated from the instrument-precursor search results and plotted as a histogram (dashed line). The mass assignments made by *Bullseye* for the same set of MS/MS scans was used to calculate mass difference and plotted as a separate histogram (solid line). A bin width of 0.2 ppm was used.



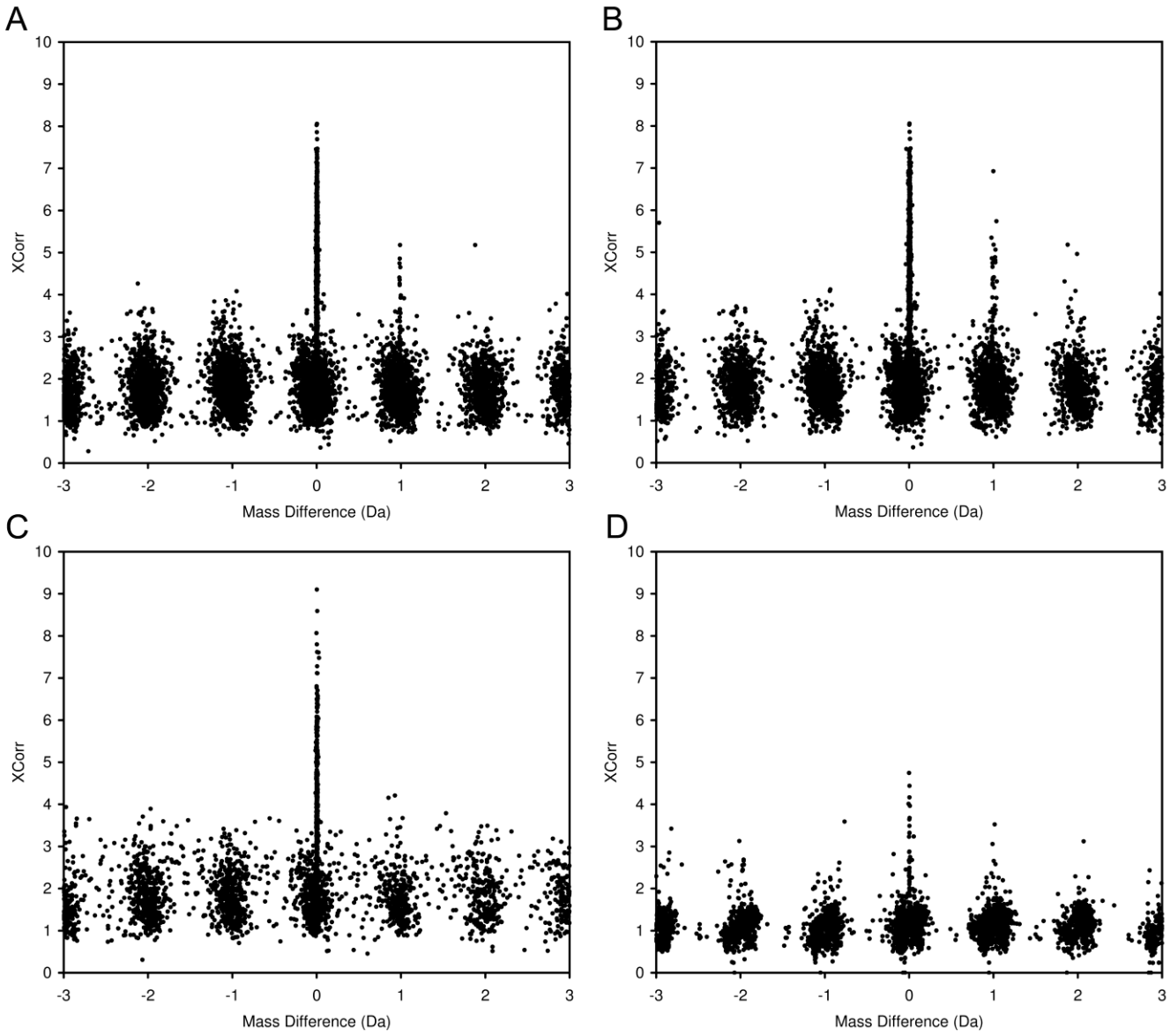
Supplemental Figure 1



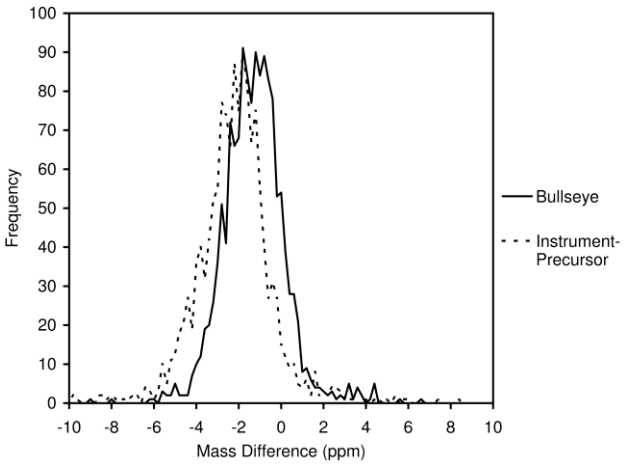
Supplemental Figure 2



Supplemental Figure 3



Supplemental Figure 4



Supplemental Figure 5