Supplementary Information

For “DirecTag: accurate sequence tags from peptide MS/MS through statistical scoring”

David L. Tabb[1, 2, 3,4], Ze-Qiang Ma[4], Daniel B. Martin[5], Amy-Joan L. Ham[2,4], Matthew C. Chambers[2,3]

1 Corresponding author: (phone) 615-936-0380; (fax) 615-343-8372; (email) david.l.tabb@vanderbilt.edu
2 Mass Spectrometry Research Center, Vanderbilt University Medical Center, 37232-8575
3 Department of Biomedical Informatics, Vanderbilt University Medical Center, 37232
4 Department of Biochemistry, Vanderbilt University Medical Center, 37232
5 Institute for Systems Biology, Fred Hutchinson Cancer Research Center, 98103

URLs for data used in the paper

VUMC websites

- Tabb group website: source and binaries
  - http://fenchurch.mc.vanderbilt.edu/
- Mass Spectrometry Research Center
  - http://www.mc.vanderbilt.edu/msrc/

Sub-score evaluation data sets

- VUMC Human gastric vesicles
  - http://www.mc.vanderbilt.edu/msrc/bioinformatics/data.php
- UMich Aurum human protein digests
  - http://www.proteomecommons.org/data/show.jsp?id=90
- ORNL *R. palustris* lysates
  - https://compbio.ornl.gov/mspipeline/raw_files/biofilm_amd/

Algorithm reproducibility data sets

- FHCRC Depleted human serum digest replicates
  - http://www.proteomecommons.org/data/show.jsp?id=2938
- VUMC Yeast lysate instrument tests
  - http://www.mc.vanderbilt.edu/msrc/bioinformatics/data.php

Multiplatform algorithm performance data sets

- ISB 18-protein cross-platform replicates
  - http://regis-web.systemsbiology.net/PublicDatasets/

DirecTag configurations

LTQ/LTQ configuration:

```
FragmentMzTolerance = 0.5
ComplementMzTolerance=0.5
IsotopeMzTolerance = 0.25
PrecursorMzTolerance = 1.25
UseAvgMassOfSequences = true
TicCutoffPercentage = 1
MaxPeakCount = 100
MaxTagCount = 50
MaxTagScore = 0
UseChargeStateFromMS = 0
DeisotopingMode = 1
ContextScoreWeight = 0
UseMultipleProcessors = 1
OnLongestPathScoreWeight = 0
StaticMods = C 57.0215
ComplementScoreWeight   =     1
IntensityScoreWeight    =     1
MzFidelityScoreWeight   =     1
```

FTMS/LTQ or Orbi/LTQ changes:

```
FragmentMzTolerance = 0.5
ComplementMzTolerance=0.5
IsotopeMzTolerance = 0.25
PrecursorMzTolerance = 0.1
UseAvgMassOfSequences = false
```

Orbi/Orbi changes:

```
FragmentMzTolerance = 0.1
ComplementMzTolerance=0.1
IsotopeMzTolerance = 0.05
PrecursorMzTolerance = 0.1
UseAvgMassOfSequences = false
```

Quad/TOF and TOF/TOF changes:

```
FragmentMzTolerance = 0.25
ComplementMzTolerance=0.25
IsotopeMzTolerance = 0.125
PrecursorMzTolerance = 0.25
UseAvgMassOfSequences = false
```

MyriMatch configurations

LTQ/LTQ configuration:

```
CleavageRules          = [|[M|K|R . . ]
DynamicMods            = M * 15.994 (Q! % -17.03 C & 57.0215
FragmentMzTolerance    = 0.5
PrecursorMzTolerance   = 1.25
TicCutoffPercentage    = 0.95
UseAvgMassOfSequences  = 1
UseChargeStateFromMS   = 0
CalculateRelativeScores = 0
```

FTMS/LTQ or Orbi/LTQ changes:

```
FragmentMzTolerance    = 0.5
PrecursorMzTolerance   = 0.1
UseAvgMassOfSequences  = 0
AdjustPrecursorMass    = true
MinPrecursorAdjustment = -1.008665
MaxPrecursorAdjustment = 1.008665
PrecursorAdjustmentStep = 1.008665
NumSearchBestAdjustments = 3
```

Orbi/Orbi changes:

```
FragmentMzTolerance    = 0.1
PrecursorMzTolerance   = 0.1
UseAvgMassOfSequences  = 0
AdjustPrecursorMass    = true
MinPrecursorAdjustment = -1.008665
MaxPrecursorAdjustment = 1.008665
PrecursorAdjustmentStep = 1.008665
NumSearchBestAdjustments = 3
```

Quad/TOF and TOF/TOF changes:

```
FragmentMzTolerance    = 0.25
PrecursorMzTolerance   = 0.25
UseAvgMassOfSequences  = 0
AdjustPrecursorMass    = true
MinPrecursorAdjustment = -1.008665
MaxPrecursorAdjustment = 1.008665
PrecursorAdjustmentStep = 1.008665
NumSearchBestAdjustments = 3
```

Sequest configurations

Serum (LTQ/LTQ):

```
[SEQUEST]
database_name =/root/20080121-IPIHum3.37-Cntm-reverse.fasta
first_database_name =/root/20080121-IPIHum3.37-Cntm-reverse.fasta
second_database_name =
peptide_mass_tolerance = 2.5
create_output_files = 1                ; 0=no, 1=yes
ion_series = 0 1 1 0.0 1.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0
fragment_ion_tolerance = 0.0           ; leave at 0.0 unless you have real poor data
num_output_lines = 12                  ; # peptide results to show
num_description_lines = 12             ; # full protein descriptions to show for top N peptides
num_results = 500                      ; # of results to process
show_fragment_ions = 0                 ; 0=no, 1=yes
print_duplicate_references = 1         ; 0=no, 1=yes
enzyme_number = 1
diff_search_options = 57.0519 C 15.9994 M
term_diff_search_options = 0.000 0.000 ; c term, n term diff mods
max_num_differential_AA_per_mod = 4    ; max # of modified AA per diff. mod in a peptide
nucleotide_reading_frame = 0           ; 0=proteinDB, 1-6, 7=forward three, 8=reverse three, 9=all six
mass_type_parent = 0                   ; 0=average masses, 1=monoisotopic masses
mass_type_fragment = 1                 ; 0=average masses, 1=monoisotopic masses
remove_precursor_peak = 0              ; 0=no, 1=yes
ion_cutoff_percentage = 0.0            ; prelim. score cutoff % as a decimal number i.e. 0.30 for 30%
protein_mass_filter = 0 0              ; enter protein mass min & max value ( 0 for both = unused)
max_num_internal_cleavage_sites = 10   ; maximum value is 5; for enzyme search
match_peak_count = 0                   ; number of auto-detected peaks to try matching (max 5)
match_peak_allowed_error = 1           ; number of allowed errors in matching auto-detected peaks
match_peak_tolerance = 1.0             ; mass tolerance for matching auto-detected peaks
partial_sequence =
sequence_header_filter =

[No static mass shifts added]

[SEQUEST_ENZYME_INFO]
0.  No_Enzyme              0   -       -
1.  Trypsin_Strict         1   KR      -
2.  Trypsin                1   KRLNH   -
```

Yeast (Orbi/LTQ) changes:

```
database_name =/root/20070727-SGD-Cntms-reverse.fasta
first_database_name =/root/20070727-SGD-Cntms-reverse.fasta
peptide_mass_tolerance = 0.1
mass_type_parent = 1                   ; 0=average masses, 1=monoisotopic masses
mass_type_fragment = 1                 ; 0=average masses, 1=monoisotopic masses
```