High Quality Top-Down Proteomics Analysis Using R

Christian Heckendorf, Roger Theberge, Jean L. Spencer, Catherine E. Costello, Mark E. McComb
Cardiovascular Proteomics Center, Boston University School of Medicine, Boston, MA

Overview

- Top-down proteomics has emerged as a technique that preserves labile post-translational modifications and offers full protein sequence coverage.
- We previously developed BUPID Top-Down, a web-based analysis pipeline for processing top-down proteomics data, to assist with the assignment of fragments from potentially unknown intact proteins.
- In order to expand the flexibility of the results output, we have now developed an R package, named BTDR, suitable for quickly producing publication-quality tables, figures, and file format conversions using the results from BUPID Top-Down.

Introduction to Top Down MS/MS

Top down proteomics involves introducing intact protein ions into the mass spectrometer and fragmenting them using ion-activation methods such as CID, ECD, and ETD. This has the potential for complete protein sequence and PTM identification without having to spend time digesting the protein. Making use of top-down data is very computationally taxing and the availability of software that can do this effectively is limited.

Following the development of BUPID Top-Down, it became apparent that the web interface, while highly accessible and convenient for job submission, did not provide the best platform for data manipulation. For this reason, an R package was designed to supplement the web-based results page and give greater control over the results output.

Filtering Results

Results can be imported into R either through a link to the web server or a file containing previously downloaded results.

```
library(BTDR)
results <- read.bupid(url="http://bupid.bumc.bu.edu/...") # abbrev. as example
```

The results file will often contain more than one candidate protein or other extraneous information. This is particularly true for LC-MS/MS data, but also with any data processed using the database search module. This package contains functions which allow the results to be filtered out until only the target results remain for further analysis.

```
head(results, "overview", n=3L)
##                               protein.name top.rank scan.count
## 1          P68871|Hemoglobin subunit beta        6          6
## 2         P02042|Hemoglobin subunit delta        0          6
## 3            Q6B0K9|Hemoglobin subunit mu        0          6

hbb <- subset(results, grepl("P68871", protein.name), "overview")
head(hbb, "protein", n=3L)
##                      protein.name tag.coverage tag.score tag.rank scan.num
## 1-0 P68871|Hemoglobin subunit beta     26.02740  60.25642        0       60
## 2-0 P68871|Hemoglobin subunit beta     25.34246  50.55006        0       61
## 3-0 P68871|Hemoglobin subunit beta     26.71233  51.80438        0       62
```

Exporting Results

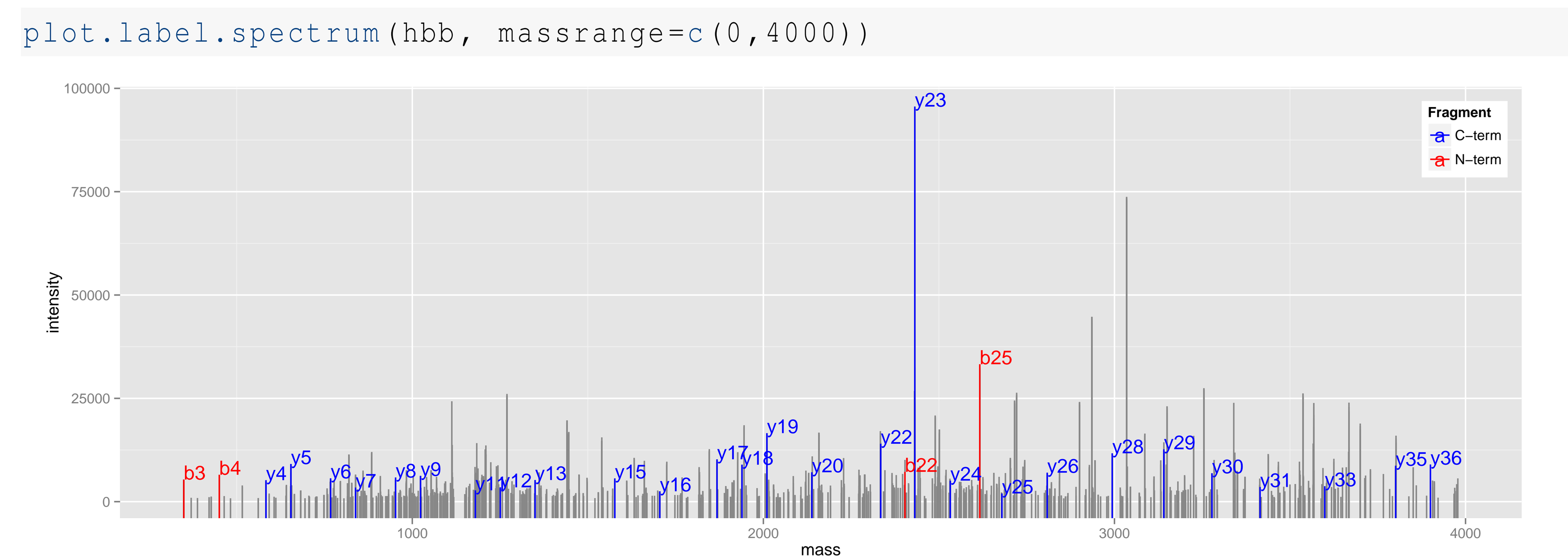
Information about protein fragments and their associated isotopic peaks in the raw spectra can easily be inspected:

```
fragment.matched.clusters(hbb)
##   scanIDs   name intensity ppmMassError monoisotopicMZ z
## 1      60 b[56]  3484.623       9.9711      1017.5315 6
## 2      60 b[25] 31528.448       3.4968      1309.1883 2
## 3      60  b[4]  6386.199      -0.5175       451.2651 1
## 4      60  b[3]  5338.280      -0.4868       350.2176 1
## 5      60  y[4]  4852.765      -0.0343       584.2929 1
## 6      60  y[5]  7314.591      -0.1803       655.3299 1

fragment.matched.ions(hbb)
##    name mods start end     massE     massT  intensity ppmMassError
## 1 b[56]          1  56 6099.1454 6099.0880 0.01836304       9.9711
## 2 b[25]          1  25 2616.3620 2616.3548 0.16614653       3.4968
## 3  b[4]          1   4  450.2579  450.2591 0.03365357      -0.5175
## 4  b[3]          1   3  349.2103  349.2114 0.02813132      -0.4868
## 5  y[4]        143 146  583.2856  565.2761 0.02557278      -0.0343
## 6  y[5]        142 146  654.3226  636.3132 0.03854595      -0.1803
```

We also offer sophisticated visual exploration utilities:

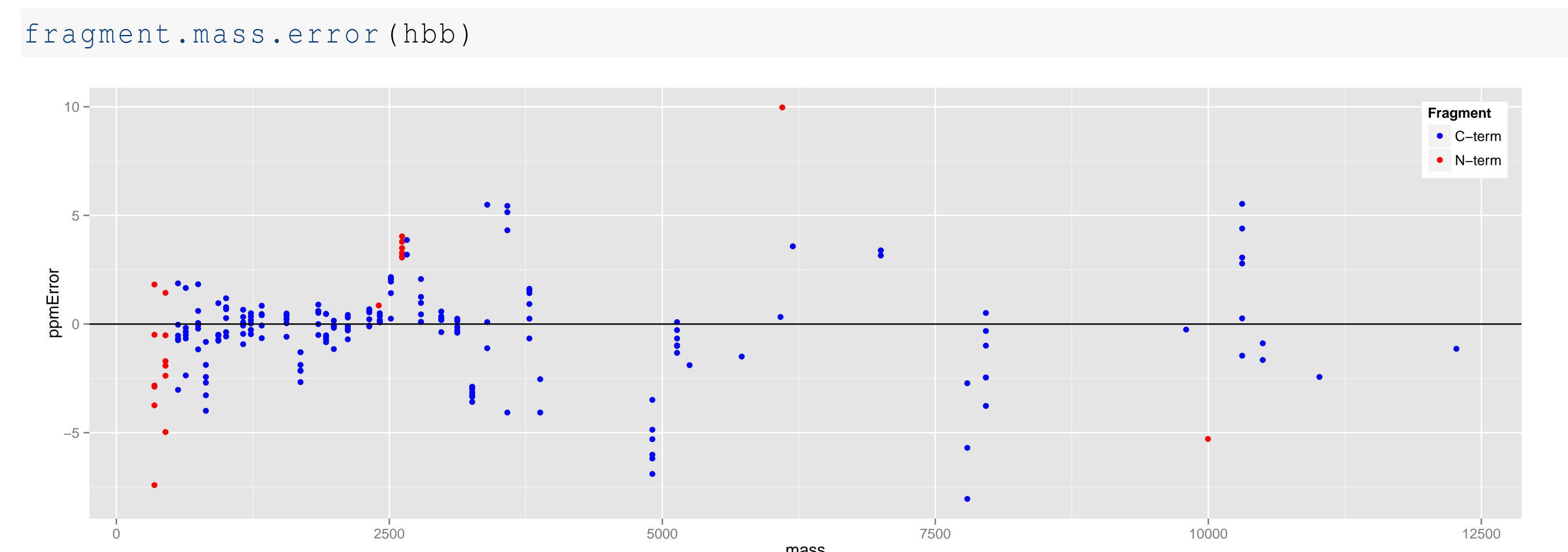
```
plot.label.spectrum(hbb, massrange=c(0,4000))
```




```
fragment.coverage(hbb, file="hbb_coverage.svg", columns=25)
```

V H L T P E E K S A V T A L W G K V N V D E V G G
E A L G R L L V V Y P W T Q R F F E S F G D L S T
P D A V M G N P K V K A H G K K V L G A F S D G L
A H L D N L K G T F A T L S E L H C D K L H V D P
E N F R L L G N V L V C V L A H H F G K E F T P P
V Q A A Y Q K V V A G V A N A L A H K Y H

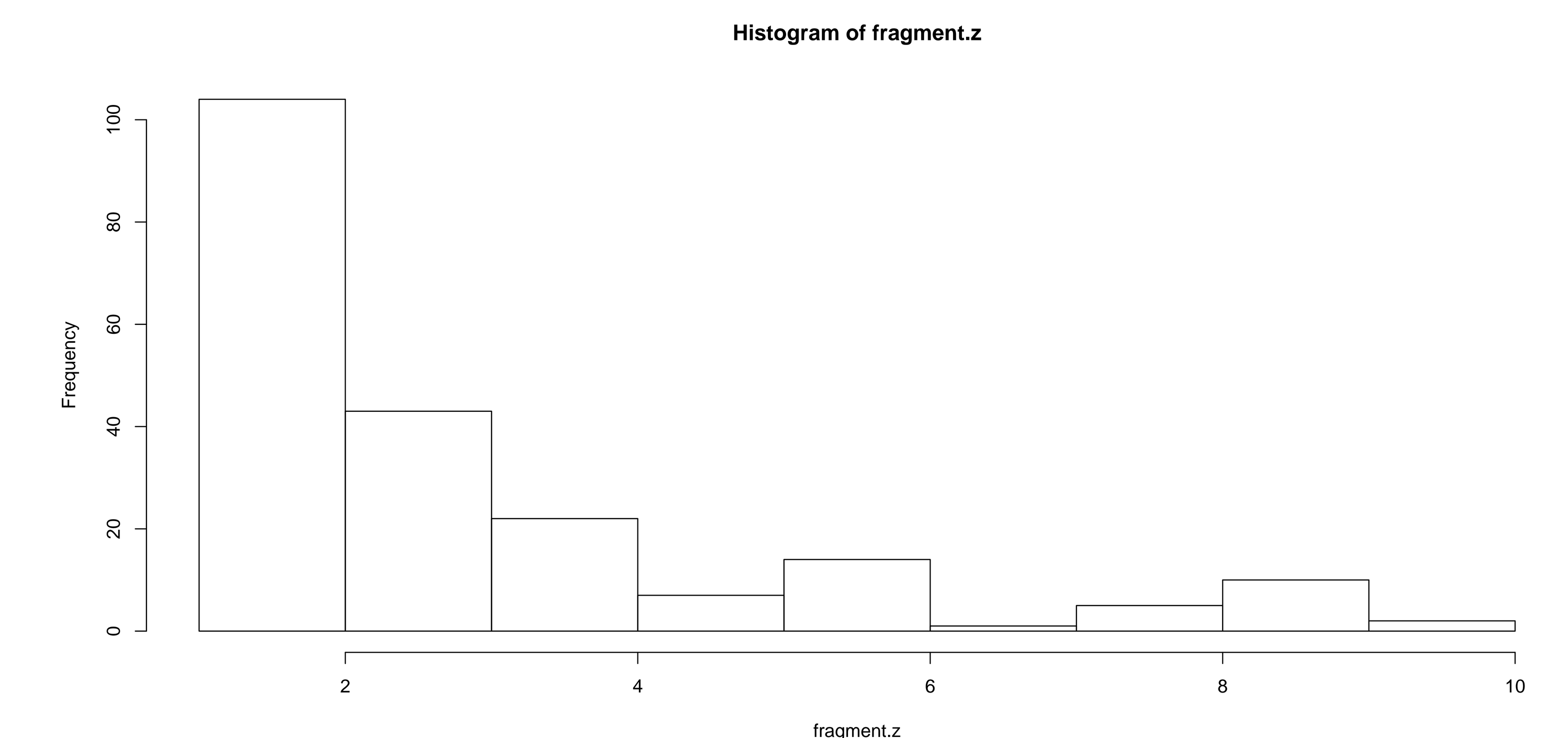
```
fragment.mass.error(hbb)
```



Discussion

- The data returned from nearly all of the functions contained in this package are standard R objects, allowing users to extend, customize, or export the results to fit their needs using familiar R tools.

```
fragment.z <- fragment.matched.clusters(hbb)$z
hist(fragment.z)
```



- An exception to this is the fragment coverage map which is written to a file as SVG, an XML-based format. These figures can be edited through many popular SVG editors such as Inkscape, Adobe Illustrator, or any text editor.
- Using BTDR, the BUPID Top-Down results can easily be exported to the standard HUPO-PSI format, mzIdentML, if one desires to view the them using other proteomics software.

Conclusions

- BTDR provides a flexible interface for viewing and customizing the results of experiments processed with BUPID Top-down
- BTDR can be installed using the devools package:

```
devtools::install_github("heckendorfc/BTDR")
```

- The BUPID Top-Down web service can be accessed at: http://bupid.bumc.bu.edu/

Acknowledgements

This project was funded by NIH grants R21 HL107993, P41 RR010888/GM104603, S10 OD010724, S10 RR020946, and S10 RR025082 and NIH-NHLBI contract HHSN268201000031C.
Special thanks to the members of the CBMS laboratories for their help.