Executable Analysis Document Supporting:

Proteomic Profiling of the Substantia Nigra to Identify Determinants of Lewy Body Pathology and Dopaminergic Neuronal Loss

Open Modification Search Analysis

Vladislav A. Petyuk[1], Lei Yu[2,3], Heather M. Olson[4], Fengchao Yu[5], Geremy Clair[1], Wei-Jun Qian[1], Joshua M. Shulman[6,7], and David A. Bennett[2,3]

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA
[2]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA
[3]Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA
[4]Enviromental and Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA
[5]Department of Pathology, University of Michigan, Ann Arbor, MI, USA
[6]Departments of Neurology, Molecular & Human Genetics, and Neuroscience, Baylor College of Medicine, Houston, TX, USA
[7]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, USA

March 5, 2021

Contents

1 Objective

The objective of this analysis is to discover peptide modifications, including biologically-relevant PTMs, and test if they are associated with any of the study design groups: LB+NL+, LB-NL+ or controls LB-NL-.

2 Reading MSFragger output

Two types of files were read as MSFragger output: 'psm.tsv' and '<dataset name>.tsv' files from the folders correspoding to each dataset.

```
library(tidyverse)

samples <- list.dirs(".", full.names = FALSE) %>%
   setdiff("")

psm_files <- map(samples, ~paste0("./",.,"/psm.tsv")) %>%
   map(~read_tsv(.x) %>% mutate(id = .x)) %>%
   bind_rows()

extra_files <- map(samples,
                   ~list.files(., pattern = "^Lewy.*tsv$", full.names = TRUE)) %>%
   map(~read_tsv(.) %>% mutate(id = .x)) %>%
   bind_rows()

# add sample_name, scannum to psm_files
psm_files <- psm_files %>%
   mutate(parsed = map(Spectrum, ~strsplit(.x, split = "\\."))) %>%
   mutate(dataset = map_chr(parsed, ~.x[[1]][1]),
          scannum = map_chr(parsed, ~.x[[1]][2]),
          scannum = as.numeric(scannum),
          charge = map_chr(parsed, ~.x[[1]][4]),
          charge = as.numeric(charge)) %>%
   mutate(sample_name = sub("\\./(.*)/psm.tsv","\\1",id)) %>%
   select(-c(parsed, Spectrum, `Spectrum File`, id)) %>%
   mutate(calibrated_delta_mass =
            `Delta Mass` - `Calibrated Observed Mass` + `Calculated Peptide Mass`) %>%
   mutate(calibrated_delta_mass = as.numeric(as.character(calibrated_delta_mass))) %>%
   filter(!is.na(`Assigned Modifications`))

extra_files <- extra_files %>%
   mutate(sample_name = sub("^([^/]*)/.*","\\1",id))

x <- inner_join(psm_files, extra_files)

x <- x %>%
   select(-c(id, protein, peptide, peptide_prev_aa, peptide_next_aa,
             `Ion Mobility`, `Number of Enzymatic Termini`,
             `Number of Missed Cleavages`, hyperscore,
             nextscore,  expectscore))

save(x,
     file = "msfragger_output.RData",
     # version=2,
```

```
    compress = "gzip")
```

3 Preprocessing

Essentially, during the preprocessing steps we removed peptides with no modification or with modification due to added iodoacetamide during the sample preparation procedure. Also, we added tentative annotation (linking to 'global.modsummary.tsv') based on the modification mass.

- Identification with no 'Assigned Modifications' were removed. Note, this allowed to pass promiscuous alkylation. It will be removed later.
- Modification masses were linked to the tentative annotations from 'global.modsummary.tsv' (**Supplementary Table 5** file using mass tolerance 0.8 mDa.
- Modifications corresponding to alkylation (57.0215 Da) were removed from 'Assigned Modifications'
- Peptides with no remained modifications were removed.
- We also required a modified peptide to be present in at least three samples.

4 Creating MSnSet object for testing

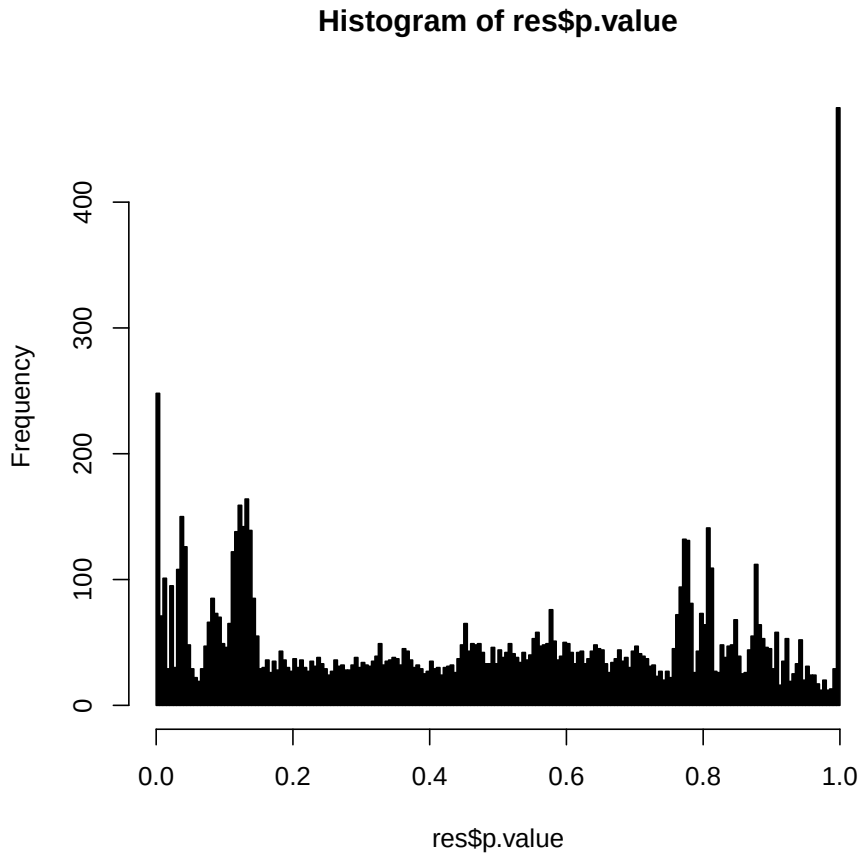
Total number of remained peptides 9877.

```
MSnSet (storageMode: lockedEnvironment)
assayData: 9877 features, 51 samples
  element names: exprs
protocolData: none
phenoData
  sampleNames: 10Cs 10Ct1 ... 9Ct2 (51 total)
  varLabels: sample_name match_id group_id
  varMetadata: labelDescription
featureData
  featureNames: AARS_KAEEIANEMIEAAK_79.966331 ABAT_ALLTGLLDLQAR_43.005814 ...
    LSAMP_EFEGEEEYLEILGITR_52.911464 (9877 total)
  fvarLabels: Peptide Theoretical Mass Shift ... feature_name (8 total)
  fvarMetadata: labelDescription
experimentData: use 'experimentData(object)'
Annotation:
- - - Processing information - - -
Subset [20379,51][9877,51] Wed Mar  3 00:53:33 2021
 MSnbase version: 2.0.2
```

5 Test of the differential abundance

Test for differential abundance is based on quasi-likelihood Poisson model.

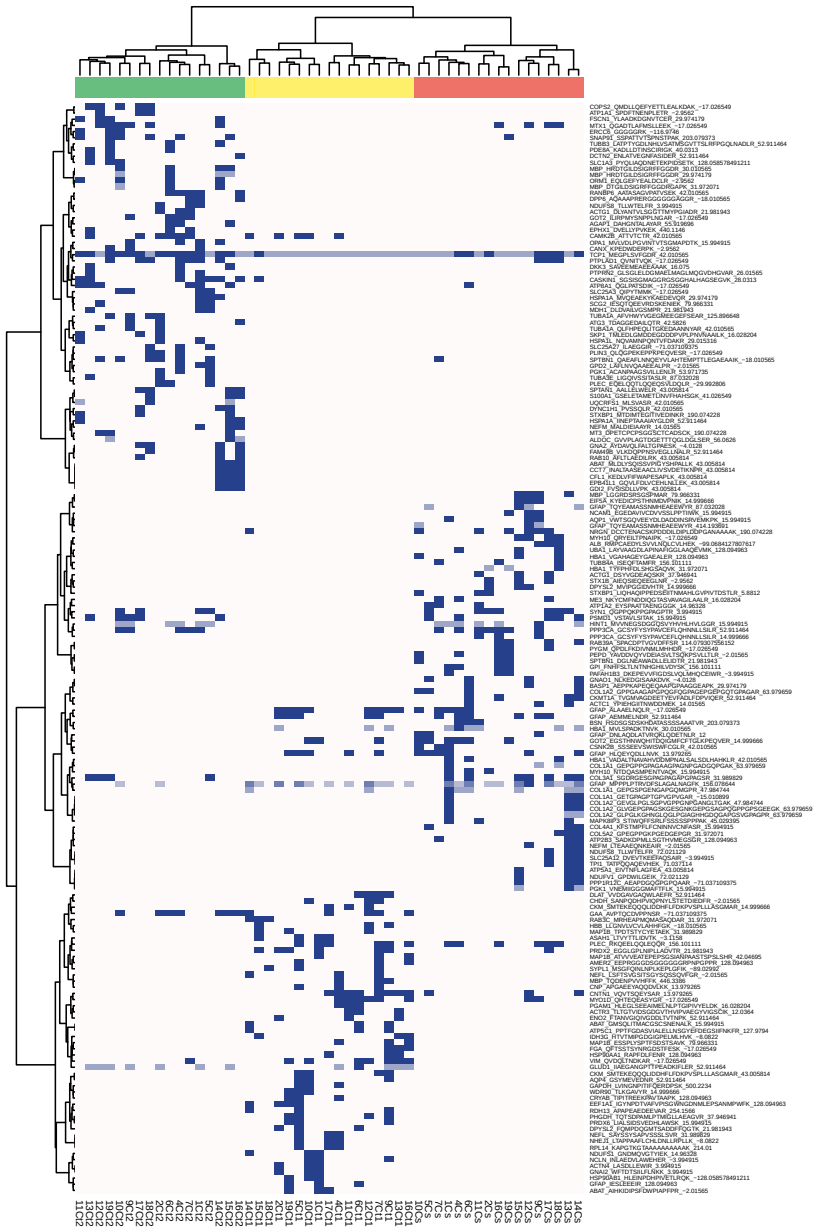
```
alt.f <- "y ~ group_id + 1"
null.f <- "y ~ 1"
div <- colSums(exprs(m2)) # normalization factor
res <- msms.glm.qlll(m2, alt.f, null.f, div=div)
res$p.val.adj <- p.adjust(res$p.value, "BH")
```



Spectral counts and results of the statistical test are saved as **Supplementary Table 6**.

6 Cluster represenation of the significant modified peptides

Clustering of 177 peptides that passed adjusted p-value < 0.05 threshold.



Corresponds to **Figure 5** in the main text. Table with clusters is saved as **Supplementary Table 7**.

7 Session information

All software and respective versions used in this document, as returned by sessionInfo() are detailed below.

- R version 3.3.2 (2016-10-31), x86_64-pc-linux-gnu
- Locale: LC_CTYPE=en_US.UTF-8, LC_NUMERIC=C, LC_TIME=en_US.UTF-8, LC_COLLATE=en_US.UTF-8, LC_MONETARY=en_US.UTF-8, LC_MESSAGES=C, LC_PAPER=en_US.UTF-8, LC_NAME=C, LC_ADDRESS=C, LC_TELEPHONE=C, LC_MEASUREMENT=en_US.UTF-8, LC_IDENTIFICATION=C
- Base packages: base, datasets, graphics, grDevices, grid, methods, parallel, stats, stats4, utils
- Other packages: AnnotationDbi 1.36.2, Biobase 2.34.0, BiocGenerics 0.20.0, BiocParallel 1.8.2, BiocStyle 2.2.1, biomaRt 2.30.0, bnlearn 4.0, Category 2.40.0, clusterProfiler 3.2.14, dendsort 0.3.3, doParallel 1.0.10, DOSE 3.0.10, dplyr 0.5.0, dynamicTreeCut 1.63-1, fastcluster 1.1.20, foreach 1.4.3, gelnet 1.2.1, ggbeeswarm 0.5.0, ggplot2 2.2.1, GOstats 2.40.0, graph 1.52.0, igraph 1.0.1, IRanges 2.8.2, iterators 1.0.8, knitr 1.14, lme4 1.1-12, lubridate 1.6.0, MASS 7.3-45, Matrix 1.2-7.1, msmsEDA 1.12.0, msmsTests 1.12.0, MSnbase 2.0.2, multcomp 1.4-5, mvtnorm 1.0-6, mzR 2.8.1, org.Hs.eg.db 3.4.0, pcaMethods 1.66.0, pheatmap 1.0.8, ProtGenerics 1.6.0, purrr 0.2.2, qvalue 2.6.0, RColorBrewer 1.1-2, Rcpp 0.12.9, ReactomePA 1.18.1, readr 1.0.0, reshape2 1.4.1, S4Vectors 0.12.2, scales 0.4.1, survival 2.40-1, TH.data 1.0-8, tibble 1.2, tidyr 0.6.1, vp.misc 0.1, WGCNA 1.51, xtable 1.8-2
- Loaded via a namespace (and not attached): acepack 1.4.1, ade4 1.7-4, affy 1.52.0, affyio 1.44.0, annotate 1.52.1, AnnotationForge 1.16.1, assertthat 0.1, backports 1.0.5, base64enc 0.1-3, beeswarm 0.2.3, BiocInstaller 1.24.0, bitops 1.0-6, Boruta 5.0.0, caTools 1.17.1, checkmate 1.8.2, cluster 2.0.5, codetools 0.2-15, colorspace 1.3-2, data.table 1.10.4, DBI 0.5-1, digest 0.6.9, DO.db 2.9, edgeR 3.16.5, evaluate 0.10, fastmatch 1.0-4, FField 0.1.0, fgsea 1.0.2, foreign 0.8-67, formatR 1.4, Formula 1.2-1, gdata 2.17.0, genefilter 1.56.0, glmnet 2.0-5, GO.db 3.4.0, GOSemSim 2.0.4, gplots 3.0.1, graphite 1.20.1, gridExtra 2.2.1, GSEABase 1.36.0, gtable 0.2.0, gtools 3.5.0, highr 0.6, Hmisc 4.0-2, htmlTable 1.9, htmltools 0.3.5, htmlwidgets 0.8, impute 1.48.0, KernSmooth 2.23-15, lattice 0.20-34, latticeExtra 0.6-28, lazyeval 0.2.0, limma 3.30.13, locfit 1.5-9.1, magrittr 1.5, MALDIquant 1.14, matrixStats 0.50.2, memoise 1.0.0, minqa 1.2.4, munsell 0.4.3, mzID 1.12.0, nlme 3.1-128, nloptr 1.0.4, nnet 7.3-12, outliers 0.14, plyr 1.8.4, preprocessCore 1.36.0, R6 2.2.0, randomForest 4.6-12, ranger 0.4.0, rappdirs 0.3.1, RBGL 1.50.0, RCurl 1.95-4.8, reactome.db 1.58.0, ROCR 1.0-7, rpart 4.1-10, RSQLite 1.1-2, sandwich 2.3-4, splines 3.3.2, stringi 1.1.2, stringr 1.2.0, tools 3.3.2, varSelRF 0.7-5, vipor 0.3.2, vsn 3.42.3, XML 3.98-1.5, zlibbioc 1.20.0, zoo 1.7-14