Portable Proteomics Pipeline (P3) Labelled (iTRAQ4) Benchmark

Daniel Kristiyanto (daniel.kristiyanto@pnnl.gov)

May 14, 2016

The purpose of this document is to benchmark the iTRAQ quantification on the Portable Proteomics Pipelines (P3) containers.

Input

1. TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzid : The identification file (product of MSGF+)
2. TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzML : The mass spectometry dataset
3. TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01_ReporterIons.txt : The quantification results used to benchmark (calcullated by PNNL)

Methods

Identification

```
## The identification file
mzids.raw
```

```
## An mzID object
##
## Software used:   MS-GF+ (version: Beta (v10089))
##
## Rawfile:         /root/data/TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzML
##
## Database:        /root/data/ID_003632_9011437E.fasta
##
## Number of scans: 15696
## Number of PSM's: 16978
```

```
head(mzid.flat[order(mzid.flat[,"ms-gf:evalue"]),])[1:3,]
```

```
##                                      spectrumid scan number(s)
```

```
## 1 controllerType=0 controllerNumber=1 scan=17146          17146
## 2 controllerType=0 controllerNumber=1 scan=13662          13662
## 3 controllerType=0 controllerNumber=1 scan=14281          14281
##   acquisitionnum passthreshold rank calculatedmasstocharge
## 1          17146          TRUE    1               943.4951
## 2          13662          TRUE    1               727.1045
## 3          14281          TRUE    1               936.7944
##   experimentalmasstocharge chargestate ms-gf:denovoscore ms-gf:evalue
## 1                 943.8400           3               233 7.599588e-24
## 2                 727.0984           4               235 7.777001e-22
## 3                 936.7979           3               188 3.604979e-20
##   ms-gf:rawscore ms-gf:specevalue assumeddissociationmethod isotopeerror
## 1            227     6.677701e-31                      HCD            1
## 2            229     6.859839e-29                      HCD            0
## 3            180     3.179835e-27                      HCD            0
##   isdecoy post pre end start          accession length
## 1   FALSE    A   K 993   967    ref|NP_001840.3   1019
## 2   FALSE    C   R 494   470    ref|NP_004915.2    911
## 3   FALSE    C   R 475   451 ref|NP_001123476.1    914
##                                                                    description
## 1 collagen alpha-2(VI) chain isoform 2C2 precursor, gi|115527062 [Homo sapiens]
## 2                                  alpha-actinin-4, gi|12025678 [Homo sapiens]
## 3                       alpha-actinin-1 isoform a, gi|194097350 [Homo sapiens]
##                       pepseq modified modification
## 1 QNVVPTVLALGSDVDMDVLTTLSLGDR    FALSE         <NA>
## 2   VEQIAAIAQELNELDYYDSHNVNTR    FALSE         <NA>
## 3   VEQIAAIAQELNELDYYDSPSVNAR    FALSE         <NA>
##                                                idFile
## 1 TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzid
## 2 TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzid
## 3 TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzid
##                                             spectrumFile
## 1 TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzML
## 2 TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzML
## 3 TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzML
##              databaseFile
## 1 ID_003632_9011437E.fasta
## 2 ID_003632_9011437E.fasta
## 3 ID_003632_9011437E.fasta
```

Based on this data, peptide QNVVPTVLALGSDVDMDVLTTLSLGDR was listed as a good quality of identification with lowest e-value.

Spectrum Objects

```
msexp.raw
```

```
## Object of class "MSnExp"
##  Object size in memory: 111.6 Mb
## - - - Spectra data - - -
##  MS level(s): 2
##  Number of MS1 acquisitions: 1606
##  Number of MSn scans: 16058
##  Number of precursor ions: 16058
##  14573 unique MZs
##  Precursor MZ's: 300.18 - 1732.7
##  MSn M/Z range: 100 1999.87
##  MSn retention times: 1:2 - 98:59 minutes
## - - - Processing information - - -
## Data loaded: Fri May 13 21:17:10 2016
##  MSnbase version: 1.16.2
## - - - Meta data  - - -
## phenoData
##   rowNames: 1
##   varLabels: sampleNames
##   varMetadata: labelDescription
## Loaded from:
##   TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzML
## protocolData: none
## featureData
##   featureNames: X1.1 X10.1 ... X9999.1 (16058 total)
##   fvarLabels: spectrum
##   fvarMetadata: labelDescription
## experimentData: use 'experimentData(object)'
```

The spectrum data consists of 1606 percursor ions and with features ranged from X1.1 to x9999.1.

This data then paired with the identification file using `addIdentificationData` tool from MSnBase, resulting the table below:

```
### The spectra objects enrinched with identification file
msexp.id
```

```
## Object of class "MSnExp"
##  Object size in memory: 118.25 Mb
## - - - Spectra data - - -
##  MS level(s): 2
##  Number of MS1 acquisitions: 1606
##  Number of MSn scans: 16058
##  Number of precursor ions: 16058
##  14573 unique MZs
```

```
##  Precursor MZ's: 300.18 - 1732.7
##  MSn M/Z range: 100 1999.87
##  MSn retention times: 1:2 - 98:59 minutes
## - - - Processing information - - -
## Data loaded: Fri May 13 21:17:10 2016
##  MSnbase version: 1.16.2
## - - - Meta data  - - -
## phenoData
##   rowNames: 1
##   varLabels: sampleNames
##   varMetadata: labelDescription
## Loaded from:
##   TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzML
## protocolData: none
## featureData
##   featureNames: X1.1 X10.1 ... X9999.1 (16058 total)
##   fvarLabels: spectrum scan number(s) ... npsm.pep (30 total)
##   fvarMetadata: labelDescription
## experimentData: use 'experimentData(object)'
```

```
idSummary(msexp.id)
```

```
##                                                spectrumFile
## 1 TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzML
##                                                      idFile coverage
## 1 TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzid    0.977
```

```
#head(fData(msexp.id))[1:3,]
```

Table focusing on the peptide QNVVPTVLALGSDVDMDVLTTLSLGDR:

```
zoomData <- fData(msexp.id)[fData(msexp.id)$pepseq==pep & !is.na(fData(msexp.id)$pepseq),]
zoomData
```

```
##          spectrum scan number(s) passthreshold rank calculatedmasstocharge
## X15545.1     6164          17146          TRUE    1               943.4951
##          experimentalmasstocharge chargestate ms-gf:denovoscore
## X15545.1                   943.84           3               233
##          ms-gf:evalue ms-gf:rawscore ms-gf:specevalue
## X15545.1 7.599588e-24            227     6.677701e-31
##          assumeddissociationmethod isotopeerror isdecoy post pre end start
## X15545.1                       HCD            1   FALSE    A   K 993   967
##                accession length
## X15545.1 ref|NP_001840.3   1019
##                                                                          description
## X15545.1 collagen alpha-2(VI) chain isoform 2C2 precursor, gi|115527062 [Homo sapiens]
```

```
##                                pepseq modified modification
## X15545.1 QNVVPTVLALGSDVDMDVLTTLSLGDR    FALSE         <NA>
##                                                     idFile
## X15545.1 TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01.mzid
##                       databaseFile nprot npep.prot npsm.prot npsm.pep
## X15545.1 ID_003632_9011437E.fasta     1         1         2        1
```

From this table, we know that this peptide was identified from features X15545.1.

Using the `quantify` function from MSnBase, with parameters `method=max`, `verbose=FALSE`, `strict=FALSE`, resulting:

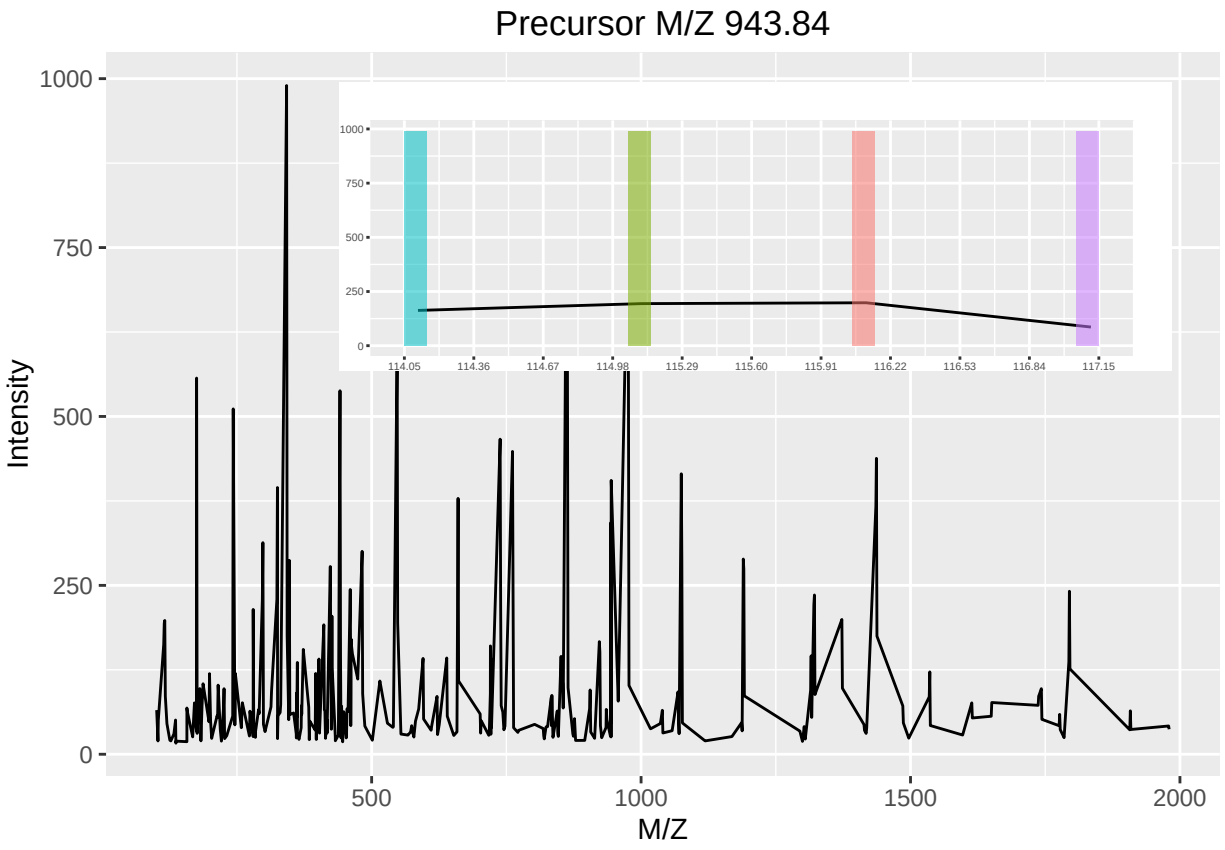
```
## Features
zoomData[,cols]
```

```
##                accession                          pepseq ms-gf:evalue
## X15545.1 ref|NP_001840.3 QNVVPTVLALGSDVDMDVLTTLSLGDR 7.599588e-24
```

```
exprs(qnt)[row.names(exprs(qnt))==row.names(zoomData),]
```

```
## iTRAQ4.114 iTRAQ4.115 iTRAQ4.116 iTRAQ4.117
##  162.69118  194.13533  197.98782   86.02748
```

```
plot(msexp.id[["X15545.1"]],reporters=iTRAQ4, full=T)
```



Quantification

This quantification then compared to the quantification data resulted from PNNL, by matching the scan number.

```
zoomData$`scan number(s)`
```

```
## [1] 17146
```

```
msexp.sp[msexp.sp$ScanNumber==zoomData$`scan number(s)`,]
```

```
##       Dataset ScanNumber Collision.Mode ParentIonMZ BasePeakIntensity
## 15545  322257      17146            hcd      943.84            989.79
##       BasePeakMZ ReporterIonIntensityMax Ion_114 Ion_115 Ion_116 Ion_117
## 15545   342.1776                  198.61  170.97  192.86  198.61   83.31
##       Weighted.Avg.Pct.Intensity.Correction
## 15545                                     2
```

Comparasion and error

The differences between the calculation by MSnBase and by PNNL:

```
zoomData$`scan number(s)`
```

```
## [1] 17146
```

```
from.msnbase  <- exprs(qnt)[row.names(zoomData),]
from.pnnl     <- msexp.sp[msexp.sp$ScanNumber==zoomData$`scan number(s)`,c("Ion_114", "Ion_115"
show(from.msnbase)
```

```
## iTRAQ4.114 iTRAQ4.115 iTRAQ4.116 iTRAQ4.117
##  162.69118  194.13533  197.98782   86.02748
```

```
show(from.pnnl)
```

```
##       Ion_114 Ion_115 Ion_116 Ion_117
## 15545  170.97  192.86  198.61   83.31
```

```
calc.error    <- from.msnbase - from.pnnl
calc.error
```

```
##         Ion_114 Ion_115    Ion_116  Ion_117
## 15545 -8.278823 1.27533 -0.6221765 2.717481
```

Some peptides do not have complete (all ions) quantified. Only peptides with complete ions measurements are retained.

```
## Before filtered
dim(exprs(qnt))
```

```
## [1] 11  4
```

```
summary(exprs(qnt))
```

```
##    iTRAQ4.114        iTRAQ4.115        iTRAQ4.116      iTRAQ4.117
##  Min.   :  162.7   Min.   :  194.1   Min.   :  198   Min.   :   86.03
##  1st Qu.:  360.7   1st Qu.:  482.3   1st Qu.:  603   1st Qu.:  457.46
##  Median : 1094.5   Median : 1665.4   Median : 2610   Median : 1408.85
##  Mean   : 3332.5   Mean   : 4209.8   Mean   : 6789   Mean   : 3488.27
##  3rd Qu.: 2334.6   3rd Qu.: 3329.7   3rd Qu.: 4737   3rd Qu.: 2412.15
##  Max.   :12085.2   Max.   :17090.8   Max.   :29537   Max.   :12464.21
##  NA's   :2         NA's   :1         NA's   :1       NA's   :2
```

```
## After filtered
dim(exprs(qnt.f))
```

```
## [1] 9 4
```

```
summary(exprs(qnt.f))
```

```
##    iTRAQ4.114        iTRAQ4.115        iTRAQ4.116        iTRAQ4.117
##  Min.   :  162.7   Min.   :  194.1   Min.   :  198.0   Min.   :   86.03
##  1st Qu.:  360.7   1st Qu.:  734.4   1st Qu.:  691.7   1st Qu.:  457.46
##  Median : 1094.5   Median : 1925.8   Median : 2915.7   Median : 1408.85
##  Mean   : 3332.5   Mean   : 4633.3   Mean   : 7480.1   Mean   : 3488.27
##  3rd Qu.: 2334.6   3rd Qu.: 3523.5   3rd Qu.: 5063.6   3rd Qu.: 2412.15
##  Max.   :12085.2   Max.   :17090.8   Max.   :29537.0   Max.   :12464.21
```

Repeating the same process for all of peptides with e-value < 10^{-10}.

```
# Original data
dim(qnt.id)          # From MSXML File < e-value treshold
```

```
## [1]  9 47
```

```
dim(msexp.sp)        # From PNNL (all)
```

```
## [1] 16058    12
```

```
qnt.merged <- merge(qnt.id, msexp.sp, by.x = "scan number(s)", by.y = "ScanNumber", all.x = T,
dim(qnt.merged)     # Merged Data (left Join)
```

```
## [1]  9 58
```

```
head(qnt.merged[,c(cols.sp, cols.p3)])
```

```
##    Ion_114  Ion_115  Ion_116  Ion_117 iTRAQ4.114 iTRAQ4.115 iTRAQ4.116
## 1 12721.22 13357.10 23012.69 11474.32 12085.1553  13781.009  22496.141
## 2 11833.90 16758.86 30405.51 12003.47 11328.8691  17090.799  29536.990
## 3  2437.82  3491.33  5161.18  2357.97  2334.5640   3523.520   5063.642
## 4   882.51  1931.75  2976.80  1379.16   858.4903   1925.761   2915.677
## 5  1148.67  1370.74  2386.69   501.30  1094.5299   1405.068   2304.417
## 6  1641.77  2746.74  3799.68  2199.70  1580.1387   2748.293   3755.990
##   iTRAQ4.117
## 1  11639.726
## 2  12464.210
## 3   2412.149
## 4   1408.853
## 5    571.472
## 6   2204.060
```

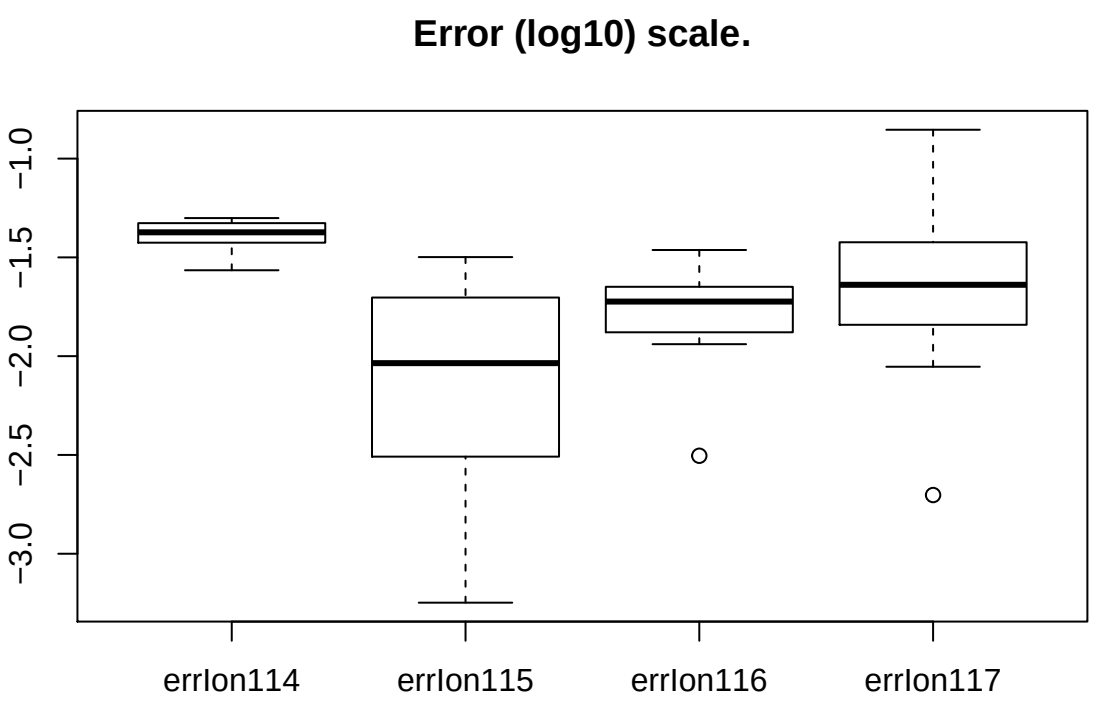
The error/differences (results from MsnBase - results from PNNL) / results from PNNL:

```
q.Error <- qnt.merged[,c("scan number(s)", "pepseq", "ms-gf:evalue")]
q.Error <- cbind(q.Error, "errIon114"= abs(qnt.merged$iTRAQ4.114 - qnt.merged$Ion_114)/qnt.merg
#q.Error <- cbind(q.Error, "errIon114"= apply(qnt.merged[,c("iTRAQ4.114","Ion_114") ],2,relErr

head(q.Error[order(q.Error$`ms-gf:evalue`),])
```

```
##   scan number(s)                       pepseq ms-gf:evalue  errIon114
## 9          17146  QNVVPTVLALGSDVDMDVLTTLSLGDR 7.599588e-24 0.04842266
## 5          13662    VEQIAAIAQELNELDYYDSHNVNTR 7.777001e-22 0.04713285
## 6          14281    VEQIAAIAQELNELDYYDSPSVNAR 3.604979e-20 0.03753956
## 4          13635    VEQIAAIAQELNELDYYDSHNVNTR 6.314509e-19 0.02721749
## 7          15079 NQAFIEMNTEEAANTMVNYYTSVTPVLR 1.025382e-16 0.03074266
## 8          16257 LMSANASDLPLSIECFMNDVDVSGTMNR 7.345664e-14 0.04061521
##      errIon115   errIon116   errIon117
## 9 0.0066127253 0.003132655 0.032618906
## 5 0.0250431351 0.034471594 0.139980022
## 6 0.0005652973 0.011498339 0.001982004
## 4 0.0031004943 0.020533204 0.021529615
## 7 0.0197597980 0.013630522 0.008840650
## 8 0.0009693562 0.013209565 0.037716650
```

```
boxplot(log10(q.Error[,c("errIon114","errIon115","errIon116","errIon117")]), main="Error (log1
```




```
## Error (Real Values)
summary(log10(q.Error[,c("errIon114","errIon115","errIon116","errIon117")]))
```

```
##    errIon114         errIon115         errIon116         errIon117
##  Min.   :-1.565   Min.   :-3.248   Min.   :-2.504   Min.   :-2.7029
##  1st Qu.:-1.426   1st Qu.:-2.509   1st Qu.:-1.879   1st Qu.:-1.8412
##  Median :-1.373   Median :-2.035   Median :-1.724   Median :-1.6387
##  Mean   :-1.398   Mean   :-2.166   Mean   :-1.806   Mean   :-1.6759
##  3rd Qu.:-1.327   3rd Qu.:-1.703   3rd Qu.:-1.649   3rd Qu.:-1.4235
##  Max.   :-1.301   Max.   :-1.498   Max.   :-1.463   Max.   :-0.8539
```

References

1. Domon B, Aebersold R: **Mass spectrometry and protein analysis**. *science* 2006, **312**:212–217.

2. Deutsch EW: **Mass spectrometer output file format mzML**. *Proteome Bioinformatics* 2010:319–331.

3. Gatto L, Gibb S: **MSnbase: Labelled and label-free mS2 data pre-processing, visualisation and quantification.** 2016.