Portable Proteomics Pipeline (P3) Labelled (iTRAQ4) Benchmark

Daniel Kristiyanto (daniel.kristiyanto@pnnl.gov)

April 29, 2016

This document is to benchmark and as a sanity check to the iTRAQ4 method for the Portable Proteomics Pipelines (P3) containers.

Results from PNNL

For this first purpose, 1 first file TCGA calculated by PNNL used as the comparasion:

```
show(mzml.sp)
```

```
## [1] "TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f01_ReporterIons.txt"
```

```
dim(msexp.sp) # The size of table
```

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

```
head(msexp.sp)[1:3,]
```

```
##   Dataset ScanNumber Collision.Mode ParentIonMZ BasePeakIntensity
## 1  322257         48            hcd      329.45          13848.50
## 2  322257         49            hcd      438.94           6742.62
## 3  322257         50            hcd      412.25           6918.29
##   BasePeakMZ ReporterIonIntensityMax Ion_114 Ion_115 Ion_116 Ion_117
## 1   390.9035                7211.01 2909.30 3358.80 7211.01 3741.59
## 2   585.8510                5977.44 2457.31 2793.67 5977.44 2964.12
## 3   412.2520                2242.89 1554.30 2242.89 1474.75 1375.35
##   Weighted.Avg.Pct.Intensity.Correction
## 1                                   3.0
## 2                                   3.2
## 3                                   3.2
```

Result from P3 container

For comparation, 1 file were also loaded and then merged with the identification data using the MSnBase package. A similar filter also applied: (1) unidentified spectrum with msgf e-value < `evalue_treshold` and (2) spectrum with multiple identification were removed

```
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
```

```
dim(fData(msexp.id)) # before filter
```

```
## [1] 16058    30
```

```
head(fData(msexp.id)[order(fData(msexp.id)[,"scan number(s)"]),])[1:3,cols]
```

```
##                                accession       pepseq ms-gf:evalue
## X1.1                     ref|NP_775799.2  DKGKLLIQRSR     5.595577
## X2.1 ref|NP_775808.2;ref|NP_001003795.1  LSSIEKIKQLR    61.405174
## X3.1    ref|NP_006010.2;ref|NP_006044.1 HRRRLRRRPADR    17.495434
```

```
dim(fData(msexp.filter2)) # After Filter
```

```
## [1] 8836   30
```

```
head(fData(msexp.id)[order(fData(msexp.filter2)[,"scan number(s)"]),])[1:3,cols]
```

```
##               accession      pepseq ms-gf:evalue
## X1.1    ref|NP_775799.2 DKGKLLIQRSR     5.595577
## X2354.1 ref|NP_055625.4    RFNMEKRR     9.515514
## X2918.1 ref|NP_065868.1    NSLSGLRR     2.306805
```

Comparing the results

Here is the results from the P3 container. Quantification used "MAX".

```
head(qnt.id[order(qnt.id$accession),])[1:3,cols.p3]
```

```
##                   accession            pepseq ms-gf:evalue iTRAQ4.114
## 5445 Contaminant_ALBU_BOVIN FGERALKAWSVARLSQK    0.7894632   27540.73
## 8719 Contaminant_ALBU_BOVIN  LCVLHEKTPVSEKVTK    1.6668929   94342.88
## 404         ref|NP_000005.2      GVPIPNKVIFIR   11.7619490  266568.03
##      iTRAQ4.115 iTRAQ4.116 iTRAQ4.117
## 5445   27540.73   27540.73   27540.73
## 8719   94342.88   94342.88   94342.88
## 404   266568.03  266568.03  266568.03
```

To put in contrast, the data aboved is merged with the data from PNNL. `iTRAQ4.114 - iTRAQ4.114` indicates results from P3, while `Ion_114 - iTRAQ4.117` indicate results from PNNL.

```
head(qnt.merged[,c(cols.p3, "Ion_114", "Ion_115", "Ion_116", "Ion_117")])[1:3,]
```

```
##         accession            pepseq ms-gf:evalue iTRAQ4.114 iTRAQ4.115
## 1 ref|NP_775799.2       DKGKLLIQRSR     5.595577 13848.4980 13848.4980
## 2 ref|NP_002211.1          RPWEEGNR     3.909789  1169.3190  1169.3190
## 3 ref|NP_056025.2 VDSITAAGGEGPFPTSR    56.255220   686.2188   686.2188
##   iTRAQ4.116 iTRAQ4.117 Ion_114 Ion_115 Ion_116 Ion_117
## 1 13848.4980 13848.4980 2909.30 3358.80 7211.01 3741.59
## 2  1169.3190  1169.3190  711.88  665.36 1191.37  745.25
## 3   686.2188   686.2188   52.28   55.44  177.77   41.19
```

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.