

Portable Proteomics Pipeline (P3)

MSGF Benchmark

Daniel Kristiyanto (daniel.kristiyanto@pnnl.gov)

June 3, 2016

1. Files:

```
## [1] "From PNNL: TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f11_msgfdb_fht.txt"
```

```
## [1] "From Docker: TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f11.txt"
```

2. Identification:

Identification used the following switches:

```
# Command Line (2)
java -Xmx8000M -jar P3/MSGFPlus.jar -s \
TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f11.mzML -o \
TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f11.mzid \
-d ID_003632_9011437E.fasta \
-t 30ppm -m 0 -inst 1 -e 1 -ti -1,2 -ntt 2 -tda 1 -minLength 6 \
-maxLength 50 -minCharge 2 -maxCharge 5 -n 1 -thread 7 \
-mod MSGFDB_Mods.txt -minNumPeaks 5 -addFeatures 1
```

3. Differences. Spec eValue Threshold= 10^{-10}

a. From PNNL

```
## [1] "Remaining peptides: 4633"
```

b. From Docker

```
## [1] "Remaining peptides: 4781"
```

c. Differences

```
## [1] "Exact match: 3475"
```

```
## [1] "From PNNL: MSGFDB_SpecEValue Peptide . From Docker: Spec_Evalue PepSeq"
```

```
## Scan MSGFDB_SpecEValue Peptide Spec_Evalue PepSeq
## 1 75 1.782822e-12 DAHKSEVAHR 1.782822e-12 DAHKSEVAHR
## 2 94 7.398652e-15 AHATGAGPAGR 7.398652e-15 AHATGAGPAGR
```

```
## [1] "Different Spec eValue: 952"
```

```
## Scan MSGFDB_SpecEValue Peptide Spec_Evalue
## 4625 15044 1.040916e-24 VGATAAVYSAAILEYLTAEVLELAGNASK 8.448862e-25
## 4630 15064 4.561789e-25 VTVEPQDSGTSALPLVSLFFYVVTGDK 4.528295e-25
## 4632 15094 9.944752e-13 KIPNPDDFFEDLEPFR 9.955277e-13
## PepSeq
## 4625 VGATAAVYSAAILEYLTAEVLELAGNASK
## 4630 VTVEPQDSGTSALPLVSLFFYVVTGDK
## 4632 KIPNPDDFFEDLEPFR
```

```
## [1] "Different ID: 195"
```

```
## Scan MSGFDB_SpecEValue Peptide Spec_Evalue
## 4461 14385 2.573502e-14 LFCLETFLSDLIAHQK 2.427363e-16
## 4502 14488 3.130180e-22 AIMNLVTGVTSLTFNPTTEILAIASEK 1.234673e-07
## 4594 14819 2.155357e-18 EVYAAAAEVLGLILR 9.332659e-08
## PepSeq
## 4461 NAIDKLFVLFGAIEILK
## 4502 KHILQVSTFQMTILMLFNNREK
## 4594 GVNVPFLELIGLPDSVVSILK
```

```
## [1] "There are 11 of unmatched peptides"
```

```
## Scan MSGFDB_SpecEValue Peptide Spec_Evalue PepSeq
## 364 2691 3.089682e-11 AISSSAISK NA <NA>
## 379 2746 5.810974e-11 SAALQVTK NA <NA>
## 664 4028 2.473212e-13 IVSSAMEPDR NA <NA>
## 668 4040 2.516175e-11 AFGGPGAGCISEGR NA <NA>
## 732 4226 7.681084e-11 SWNDCLNK NA <NA>
## 905 4786 2.062452e-14 YDDMAACMK NA <NA>
```

4. Differences. Spec eValue Threshold = 10^{-15}

a. From PNNL

```
## [1] "Remaining peptides: 2511"
```

b. From Docker

```
## [1] "Remaining peptides: 2562"
```

c. Differences

```
## [1] "Exact match: 1858"

## [1] "From PNNL: MSGFDB_SpecEValue Peptide . From Docker: Spec_Evalue PepSeq"

##   Scan MSGFDB_SpecEValue      Peptide  Spec_Evalue
## 1  519      1.850634e-19 DTPRPDHPHDGHSPASR 1.850634e-19
## 6  747      1.638325e-19  CENPCHAGSCQPCPR 1.638325e-19
##                               PepSeq
## 1 DTPRPDHPHDGHSPASR
## 6  CENPCHAGSCQPCPR

## [1] "Different Spec eValue: 563"

##   Scan MSGFDB_SpecEValue      Peptide  Spec_Evalue
## 2505 15043      1.685906e-24 VGATAAVYSAAILEYLTAEVLELAGNASK 1.704616e-24
## 2506 15044      1.040916e-24 VGATAAVYSAAILEYLTAEVLELAGNASK 8.448862e-25
## 2510 15064      4.561789e-25 VTVEPQDSGTSALPLVSLFFYVVTGDK 4.528295e-25
##                               PepSeq
## 2505 VGATAAVYSAAILEYLTAEVLELAGNASK
## 2506 VGATAAVYSAAILEYLTAEVLELAGNASK
## 2510 VTVEPQDSGTSALPLVSLFFYVVTGDK

## [1] "Different ID 90"

##   Scan MSGFDB_SpecEValue      Peptide  Spec_Evalue
## 2365 14354      4.520059e-19 SNIMTLFQCMQDSMPEVR 2.971317e-22
## 2406 14488      3.130180e-22 AIMNLVTGVTSITFNPTTEILAIASEK 1.234673e-07
## 2485 14819      2.155357e-18 EVYAAAAEVLGLILR 9.332659e-08
##                               PepSeq
## 2365 LVAAPLFELYDNAPGYGPIISSLPQLLSR
## 2406 KHILQVSTFQMTILMLFNNREK
## 2485 GVNVPFLELIGLPDSVVSILK

## [1] "There are 0 of unmatched peptides"

## [1] Scan      MSGFDB_SpecEValue Peptide      Spec_Evalue
## [5] PepSeq
## <0 rows> (or 0-length row.names)
```

5. Differences. Spec eValue Threshold = 10^{-20}

a. From PNNL

```
## [1] "Remaining peptides: 936"
```

b. From Docker

```
## [1] "Remaining peptides: 943"
```

c. Differences

```
## [1] "Exact match: 691"

## [1] "From PNNL: MSGFDB_SpecEValue Peptide . From Docker: Spec_Evalue PepSeq"

##      Scan MSGFDB_SpecEValue      Peptide  Spec_Evalue
## 3  858      3.118871e-22 ADAGKEGNNPAENGDAKTDQAQK 3.118871e-22
## 4 1103      4.194526e-27 GHQDPSQATGTTGSSVSCTEEK 4.194526e-27
##                               PepSeq
## 3 ADAGKEGNNPAENGDAKTDQAQK
## 4 GHQDPSQATGTTGSSVSCTEEK

## [1] "Different Spec eValue: 217"

##      Scan MSGFDB_SpecEValue      Peptide  Spec_Evalue
## 933 15043      1.685906e-24 VGATAAVYSAAILEYLTAEVLELAGNASK 1.704616e-24
## 934 15044      1.040916e-24 VGATAAVYSAAILEYLTAEVLELAGNASK 8.448862e-25
## 936 15064      4.561789e-25 VTVEPQDSGTSALPLVSLFFYVVTDGK 4.528295e-25
##                               PepSeq
## 933 VGATAAVYSAAILEYLTAEVLELAGNASK
## 934 VGATAAVYSAAILEYLTAEVLELAGNASK
## 936 VTVEPQDSGTSALPLVSLFFYVVTDGK

## [1] "Different ID: 28"

##      Scan MSGFDB_SpecEValue      Peptide
## 705 13489      9.813857e-21 YETVIMPVFGIATPFHIATIK
## 811 13965      2.702549e-23 LPSGLGCSTVLSPEGSAQFAAQIFGLSNHLVWSK
## 887 14488      3.130180e-22 AIMNLVTGVTSLTFNPTTEILAIASEK
##      Spec_Evalue      PepSeq
## 705 3.804641e-07      EEKPKHTEAKKSLSFR
## 811 8.912736e-08 ISHDSFGQTDIMVDAEDLVFVRIKTLDILK
## 887 1.234673e-07      KHILQVSTFQMTILMLFNNREK

## [1] "There are 0 of unmatched peptides"

## [1] Scan      MSGFDB_SpecEValue Peptide      Spec_Evalue
## [5] PepSeq
## <0 rows> (or 0-length row.names)
```

6. Differences between results from PNNL and DTA file from PNNL.

MSGF+ was run against TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f11_dta.txt using the same command (10 ppm, without Docker).

a.

```
## [1] "Peptides: 25782 with 13087 duplicated scan numbers."

## [1] "x = PNNL y=MSGF+"

## [1] "Exact same: 23911 Different Spec eValue: 0 Different Peptides: 1871"

## [1] "Different Spec eValue"

## [1] "Different Peptides"

##      Scan MSGFDB_SpecEValue.x      Peptide.x MSGFDB_SpecEValue.y
## 25542 14952      4.810757e-07      NDPEPVKKDK      4.810757e-07
## 25691 15066      1.450531e-06 LADQFLEEKQKETQKIQSNDGK      1.450531e-06
## 25699 15067      1.271352e-06      IKKTLK      1.271352e-06
## 25713 15070      8.152290e-07 EVDYRSDDRREDDYDIQLSIK GK      8.152290e-07
## 25740 15095      8.242609e-08 EQNYKQQVNCTVLSTPTSEVKK      8.242609e-08
## 25742 15095      8.242609e-08 EQNYKQQVNCTVLSTPTSEVKK      8.242609e-08
##      Peptide.y
## 25542      KFQEESKFK
## 25691 WYAVLFEFQPGLDELEIVLHFGDK
## 25699      LKKTLK
## 25713 RRNDLDDPERGMIFVCSATHKTK
## 25740 SEPKPGLPEDLQKLMKDYSSR
## 25742 SEPKPGLPEDLQKLMKDYSSR
```

a. Duplicates removed

```
## [1] "Peptides: 25782 with 13087 duplicated scan numbers."

## [1] "After removing duplicates 12695 peptides remained."

## [1] "x = PNNL y=MSGF+"

## [1] "Exact same: 12355 Different Spec eValue: 0 Different Peptides: 340"

## [1] "Different Spec eValue"

## [1] "Different Peptides"

##      Scan MSGFDB_SpecEValue.x      Peptide.x MSGFDB_SpecEValue.y
## 5      55      2.070157e-06      DPEKCDGRCR      2.070157e-06
## 12     72      1.255958e-05      VGSAPK      1.255958e-05
## 16     79      4.965787e-05      ILEEMR      4.965787e-05
## 29     103     2.321208e-06      NNLASCYLK      2.321208e-06
## 33     109     9.826282e-03      VPPVPR      9.826282e-03
## 53     154     4.760844e-06 CFLTAMRESGAHCPLCR      4.760844e-06
##      Peptide.y
## 5      QKCPPCWYR
## 12     TAAAPK
## 16     LLEEMR
## 29     ESCPKHAVR
## 33     VPPVPR
## 53     DDNSNFHMDFIVAASNL
```

7. Quantification Benchmark

```
## [1] "No of rows 13473 . Error measured by = |rslt_docker - rslt_pnnl|/rslt_docker"
```

```
## [1] "Ion 114"
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	0.0044	0.7989	0.9425	0.8560	0.9885	1.0000	2277

```
## [1] "Ion 115"
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	0.0001	0.7335	0.9238	0.8006	0.9869	1.0000	1570

```
## [1] "Ion 116"
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	0.0000	0.6196	0.8972	0.7305	0.9831	1.0000	1155

```
## [1] "Ion 117"
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	0.0035	0.7801	0.9398	0.8487	0.9884	1.0000	1987

Log2 relative error

