

# 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 10ppm -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: 4611"
```

#### c. Differences

```
## [1] "Exact match: 3429"
```

```
## [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: 911"
```

```
## 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: 282"
```

```
## Scan MSGFDB_SpecEValue Peptide
## 4600 14850 7.007801e-16 AYEPQELKPFVTAISSALVIAAVFDRDINCR
## 4610 14905 2.057563e-13 FPFIPMLDYILPELLK
## 4618 14987 1.283296e-29 ILDDDEAAQELMPVVAGAVFTLTAHLSQAVLTEQK
## Spec_Evalue PepSeq
## 4600 1.859416e-08 AEDKLKESDDKVLENRVLDGKLSSEK
## 4610 6.778596e-07 GRASLIFSLKNEVGGLIK
## 4618 1.299389e-07 DGRIKTGDILLNVDGVELTEVSRSEAVALLKR
```

```
## [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: 2480"
```

### c. Differences

```
## [1] "Exact match: 1835"

## [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: 541"

##   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 135"

##   Scan MSGFDB_SpecEValue      Peptide
## 2485 14819      2.155357e-18  EVYAAAAEVLGLILR
## 2488 14850      7.007801e-16  AYEPQELKPFVTAISSALVIAAVFDRDINCR
## 2499 14987      1.283296e-29  ILDDEAAQELMPVVAGAVFTLTAHLSQAVLTEQK
##   Spec_Evalue      Peptide
## 2485 9.332659e-08  GNVVVPFLELIGLPDSVVSILK
## 2488 1.859416e-08  AEDKLKESSDKVLENRVLDGKLSSEK
## 2499 1.299389e-07  DGRIKTGDILLNVDGVELTEVSRSEAVALLKR

## [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: 920"
```

### c. Differences

```
## [1] "Exact match: 686"

## [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: 209"

##      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: 41"

##      Scan MSGFDB_SpecEValue      Peptide
## 860 14345      5.187062e-28 TVLLSIQALLSAPNPDDPLANDVAEQWK
## 887 14488      3.130180e-22 AIMNLVTGVTSLTFNPTEILAIASEK
## 930 14987      1.283296e-29 ILDEEAAQELMPVVAGAVFTLTAHLSQAVLTEQK
##
##      Spec_Evalue      PepSeq
## 860 5.696890e-08      KEQQIAFRLSCLEKELELTHEK
## 887 1.234673e-07      KHILQVSTFQMTILMLFNNREK
## 930 1.299389e-07 DGRIKTDILLNVDGVELTEVSRSEAVALLKR

## [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).

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

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

```
## [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     VPVPPR
## 53 DDDSNFHMDFIVAASNL
```

## 7. Quantification Benchmark

```
## [1] "No of rows 13451 . Error measured by = |rslt_docker - rslt_pnnl|/rslt_docker"
```

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

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 0.0044  0.7995  0.9425  0.8562  0.9885  1.0000  2272
```

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

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 0.0001  0.7342  0.9240  0.8008  0.9869  1.0000  1565
```

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

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 0.0000  0.6196  0.8972  0.7306  0.9831  1.0000  1150
```

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

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 0.0035  0.7804  0.9398  0.8490  0.9884  1.0000  1982
```

