

Portable Proteomics Pipeline (P3)

MSGF Benchmark

Daniel Kristiyanto (daniel.kristiyanto@pnnl.gov)

May 31, 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: 4635"
```

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
## 21   75      1.782822e-12  DAHKSEVAHR 1.782822e-12  DAHKSEVAHR
## 35   94      7.398652e-15  AHATGAGPAGR 7.398652e-15  AHATGAGPAGR
```

```
## [1] "Different Spec eValue: 911"
```

```
##      Scan MSGFDB_SpecEValue      Peptide  Spec_Evalue
## 13407 15044      1.040916e-24  VGATAAVYSAAILEYLTAEVLELAGNASK 8.448862e-25
## 13422 15064      4.561789e-25  VTVEPQDSGTSALPLVSLFFYVVDGK 4.528295e-25
## 13447 15094      9.944752e-13      KIPNPDDFFEDLEPFR 9.955277e-13
##
##      PepSeq
## 13407 VGATAAVYSAAILEYLTAEVLELAGNASK
## 13422 VTVEPQDSGTSALPLVSLFFYVVDGK
## 13447      KIPNPDDFFEDLEPFR
```

```
## [1] "Different ID and Spec eValue: 279"
```

```
##      Scan MSGFDB_SpecEValue      Peptide
## 13240 14850      7.007801e-16  AYEPQELKPFVTAISSALVIAAVFDRDINCR
## 13290 14905      2.057563e-13      FPFIPMPLDYILPELLK
## 13361 14987      1.283296e-29  ILDDAAQELMPVVAGAVFTLTAHLSQAVLTEQK
##
##      Spec_Evalue      PepSeq
## 13240 1.859416e-08      AEDKLKESDDKVLNRLVDGKLSSEK
## 13290 6.778596e-07      GRASLIFSLKNEVGGLIK
## 13361 1.299389e-07  DGRIKTGDILLNVDGVELTEVSRSEAVALLKR
```

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
## 352   519      1.850634e-19  DTPRPDHPHDGHSPASR 1.850634e-19
## 551   747      1.638325e-19  CENPCHAGSCQPCPR 1.638325e-19
##
##      PepSeq
## 352 DTPRPDHPHDGHSPASR
## 551 CENPCHAGSCQPCPR
```

```
## [1] "Different Spec eValue: 541"
```

```
##          Scan MSGFDB_SpecEValue          Peptide  Spec_Evalue
## 13393 15043      1.685906e-24 VGATAAVYSAAILEYLTAEVLELAGNASK 1.704616e-24
## 13394 15044      1.040916e-24 VGATAAVYSAAILEYLTAEVLELAGNASK 8.448862e-25
## 13409 15064      4.561789e-25 VTVEPQDSGTSALPLVSLFFYVVTGDK 4.528295e-25
##
##          PepSeq
## 13393 VGATAAVYSAAILEYLTAEVLELAGNASK
## 13394 VGATAAVYSAAILEYLTAEVLELAGNASK
## 13409 VTVEPQDSGTSALPLVSLFFYVVTGDK
```

```
## [1] "Different ID and Spec eValue: 133"
```

```
##          Scan MSGFDB_SpecEValue          Peptide
## 13200 14819      2.155357e-18          EVYAAAAEVLGLILR
## 13227 14850      7.007801e-16      AYEPQELKPFVTAISSALVIAAVFDRDINCR
## 13348 14987      1.283296e-29 ILDDDEAAQELMPVVAGAVFTLTAHLSQAVLTEQK
##
##          Spec_Evalue          PepSeq
## 13200 9.332659e-08          GNVVVPFLELIGLPDSVVSILK
## 13227 1.859416e-08      AEDKLKESSDKVLNRLVDGKLSSEK
## 13348 1.299389e-07 DGRIKTGDILLNVDGVELTEVSRSEAVALLKR
```

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
## 652 858      3.118871e-22 ADAGKEGNNPAENGDAKTDQAQK 3.118871e-22
## 872 1103      4.194526e-27 GHQDPSQATGTTGSSVSCTEEK 4.194526e-27
##
##          PepSeq
## 652 ADAGKEGNNPAENGDAKTDQAQK
## 872 GHQDPSQATGTTGSSVSCTEEK
```

```
## [1] "Different Spec eValue: 209"
```

```
##          Scan MSGFDB_SpecEValue          Peptide  Spec_Evalue
## 13393 15043      1.685906e-24 VGATAAVYSAAILEYLTAEVLELAGNASK 1.704616e-24
## 13394 15044      1.040916e-24 VGATAAVYSAAILEYLTAEVLELAGNASK 8.448862e-25
## 13409 15064      4.561789e-25 VTVEPQDSGTSALPLVSLFFYVVTGDK 4.528295e-25
```

```
##                                     PepSeq
## 13393 VGATAAVYSAAILEYLTAEVLELAGNASK
## 13394 VGATAAVYSAAILEYLTAEVLELAGNASK
## 13409 VTVEPQDSGTSALPLVSLFFYVVTGDK

## [1] "Different ID and Spec eValue: 40"
```

##	Scan	MSGFDB_SpecEValue	Peptide
## 12769	14345	5.187062e-28	TVLLSIQALLSAPNPDDPLANDVAEQWK
## 12899	14488	3.130180e-22	AIMNLVTGVTSLTFNPTTEILAIASEK
## 13348	14987	1.283296e-29	ILDDEAAQELMPVVAGAVFTLTAHLSQAVLTEQK

##	Spec_Evalue	PepSeq
## 12769	5.696890e-08	KEQQIAFRLSCLEKELELTEHK
## 12899	1.234673e-07	KHILQVSTFQMTILMLFNNREK
## 13348	1.299389e-07	DGRIKTDILLNVDGVELTEVSRSEAVALLKR

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).

```
## reading TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f11_msgfplus.mzid... DONE!

## [1] "Exact match: 24379"

## [1] "From Docker: ms-gf:specvalue pepseq . From PNNL: MSGFDB_SpecEValue Peptide"

##      Scan ms-gf:specvalue pepseq MSGFDB_SpecEValue Peptide
## 6894 4099      0.05022644 PGGAGM      0.05022644 PGGAGM
## 6312 3823      0.04856372 SPAGGG      0.04856372 SPAGGG

## [1] "Different Spec eValue: 0"

## [1] Scan      ms-gf:specvalue pepseq      MSGFDB_SpecEValue
## [5] Peptide
## <0 rows> (or 0-length row.names)

## [1] "Different ID and Spec eValue: 1399 with 253 distinct Scan numbers."
```

##	Scan	ms-gf:specvalue	pepseq	MSGFDB_SpecEValue	Peptide
## 23193	12583	2.226537e-21	YDVDTLDMVFLDHWK	5.690908e-13	FLPSVFMDAMR
## 23195	12583	2.226537e-21	YDVDTLDMVFLDHWK	5.690908e-13	FLPSVFMDAMR
## 23197	12583	2.226537e-21	YDVDTLDMVFLDHWK	5.690908e-13	FLPSVFMDAMR