

# Portable Proteomics Pipeline (P3)

## MSGF Benchmark

*Daniel Kristiyanto (daniel.kristiyanto@pnnl.gov)*

*May 27, 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 = $10^{-10}$

#### a. From PNNL

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

#### b. From Docker

```
## [1] "Remaining peptides: 4611"
```

#### c. Differences

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

##	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	
##	13405	15044	1.040916e-24	VGATAAVYSAAILEYLTAEVLELAGNASK	8.448862e-25
##	13420	15064	4.561789e-25	VTVEPQDSGTSALPLVSLFFYVVDGK	4.528295e-25
##	13445	15094	9.944752e-13	KIPNPDDFEDLEPFR	9.955277e-13

  

##	Peptide	PepSeq
##	13405	VGATAAVYSAAILEYLTAEVLELAGNASK
##	13420	VTVEPQDSGTSALPLVSLFFYVVDGK
##	13445	KIPNPDDFEDLEPFR

```
## [1] "Different ID and Spec eValue: 277"
```

##	Scan	MSGFDB_SpecEValue	Peptide	
##	13238	14850	7.007801e-16	AYEPQELKPFVTAISSALVIAAVFDRDINCR
##	13288	14905	2.057563e-13	FPFIPMLDYILPELLK
##	13359	14987	1.283296e-29	ILDDEAAQELMPVVAGAVFTLTAHLSQAVLTEQK

  

##	Spec_Evalue	PepSeq	
##	13238	1.859416e-08	AEDKLKESDQVLENRVLDGKLSSEK
##	13288	6.778596e-07	GRASLIFSLKNEVGGLIK
##	13359	1.299389e-07	DGRIKTGDILLNVDGVELTEVSRSEAVALLKR

### 3. Differences. Spec eValue = $10^{-15}$

#### a. From PNNL

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

#### b. From Docker

```
## [1] "Remaining peptides: 2480"
```

#### c. Differences

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

##	Scan	MSGFDB_SpecEValue	Peptide	Spec_Evalue	
##	352	519	1.850634e-19	DTPRPDHPHDGHSPASR	1.850634e-19
##	551	747	1.638325e-19	CENPCHAGSCQPCPR	1.638325e-19

  

##	Peptide	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	VTVEPQDSGTSALPLVSLFFYVVDGK	4.528295e-25

```
##                                     PepSeq
## 13393 VGATAAVYSAAILEYLTAEVLELAGNASK
## 13394 VGATAAVYSAAILEYLTAEVLELAGNASK
## 13409 VTVEPQDSGTSALPLVSLFFYVVTDGK

## [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      GVNVPFLELIGLPDSVVSILK
## 13227 1.859416e-08      AEDKLKESSDKVLENRVLDGKLSSEK
## 13348 1.299389e-07      DGRIKTGDILLNVDGVELTEVSRSEAVALLKR
```

### 3. Differences. Spec eValue = $10^{-20}$

#### a. From PNNL

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

#### b. From Docker

```
## [1] "Remaining peptides: 920"
```

#### c. Differences

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

```
##      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      VTVEPQDSGTSALPLVSLFFYVVTDGK 4.528295e-25
##                                     PepSeq
## 13393 VGATAAVYSAAILEYLTAEVLELAGNASK
## 13394 VGATAAVYSAAILEYLTAEVLELAGNASK
## 13409 VTVEPQDSGTSALPLVSLFFYVVTDGK
```

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

##	Scan	MSGFDB_SpecEValue	Peptide
## 12769	14345	5.187062e-28	TVLLSIQALLSAPNPDDPLANDVAEQWK
## 12899	14488	3.130180e-22	AIMNLVTGVTSLTFTNPTEILAIASEK
## 13348	14987	1.283296e-29	ILDDEAAQELMPVVAGAVFTLTAHLSQAVLTEQK
##		Spec_EValue	PepSeq
## 12769		5.696890e-08	KEQQIAFRLSCLEKELELTEHK
## 12899		1.234673e-07	KHILQVSTFQMTILMLFNNREK
## 13348		1.299389e-07	DGRIKTGDILLNVDGVELTEVSRSEAVALLKR