

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: 1665"
```

c. Differences

```
## [1] "Exact match: 1227"
```

##	Scan	MSGFDB_SpecEValue	Peptide	Spec_Evalue	PepSeq	
##	54	120	8.875677e-14	TSYAQQHQQVR	8.875677e-14	TSYAQQHQQVR
##	210	355	5.394193e-12	VGPVSHHR	5.394193e-12	VGPVSHHR

```
## [1] "Different Spec eValue: 325"
```

##	Scan	MSGFDB_SpecEValue	Peptide	Spec_Evalue
##	12813	14400	3.676064e-13	QLYNVEATSYALLALLQLK 3.904655e-13
##	13036	14646	1.950997e-15	TLNLAQNLLAQLPEELFHPLTSLQTLK 2.059538e-15
##	13049	14660	5.907738e-25	LLIPVVPGVDSLNSAMAASILLFEGK 5.911286e-25
##			PepSeq	
##	12813		QLYNVEATSYALLALLQLK	
##	13036		TLNLAQNLLAQLPEELFHPLTSLQTLK	
##	13049		LLIPVVPGVDSLNSAMAASILLFEGK	

```
## [1] "Different ID and Spec eValue: 3069"
```

##	Scan	MSGFDB_SpecEValue	Peptide	Spec_Evalue
##	13412	15084	8.427389e-11	GSFPWQAK 8.960810e-07
##	13421	15094	9.944752e-13	KIPNPDDFFEDLEPFR 3.318390e-06
##	13425	15100	1.462152e-17	DLYANTVLSGGTTMYPGIADR 2.336472e-06
##			PepSeq	
##	13412		IFQAENAK	
##	13421		HHTTRQQEPVTLTTSK	
##	13425		VKQCRCMSVNLSDSDK	

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

a. From PNNL

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

b. From Docker

```
## [1] "Remaining peptides: 767"
```

c. Differences

```
## [1] "Exact match: 572"
```

##	Scan	MSGFDB_SpecEValue	Peptide	Spec_Evalue
##	353	519	1.850634e-19	DTPRPDHPHDGHSPASR 1.850634e-19
##	825	1052	1.750968e-16	KKDLTQTASSTAR 1.750968e-16
##			PepSeq	
##	353		DTPRPDHPHDGHSPASR	
##	825		KKDLTQTASSTAR	

```
## [1] "Different Spec eValue: 170"
```

##	Scan	MSGFDB_SpecEValue	Peptide	Spec_Evalue
##	12702	14291	2.763854e-19	GISNMLDVNGLFTLSHITQLVLSHNK 3.272698e-19
##	12759	14354	4.520059e-19	SNIMTLFQCMQDSMPEVR 5.012071e-19
##	13037	14660	5.907738e-25	LLIPVVPGVDSLNSAMAASILLFEGK 5.911286e-25

```
##                               PepSeq
## 12702 GISNMLDVNGLFTLSHITQLVLSHNK
## 12759          SNIMTLFLQCMQDSMPEVR
## 13037 LLIPVVPGVDSLNSAMAASILLFEGK

## [1] "Different ID and Spec eValue: 1769"
```

```
##      Scan MSGFDB_SpecEValue      Peptide  Spec_Evalue
## 13383 15058      4.821639e-16      DLYANTVLSGGTMYPGIADR 5.005288e-07
## 13387 15064      4.561789e-25      VTVEPQDSGTSALPLVSLFFYVVTGDK 1.123928e-06
## 13413 15100      1.462152e-17      DLYANTVLSGGTMYPGIADR 2.336472e-06
##                               PepSeq
## 13383          QESGSVMCRYDFVPRIK
## 13387 QLGWAILLLGTNTINSCGGLLELVELGK
## 13413          VKQCRCMSVNLSDSDK
```

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

a. From PNNL

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

b. From Docker

```
## [1] "Remaining peptides: 223"
```

c. Differences

```
## [1] "Exact match: 165"
```

```
##      Scan MSGFDB_SpecEValue      Peptide  Spec_Evalue
## 908  1143      1.987852e-23      TGQCECQPGITGQH CER 1.987852e-23
## 2467 2903      4.178550e-21      DNCVTVPNSGQEDVDR 4.178550e-21
##                               PepSeq
## 908  TGQCECQPGITGQH CER
## 2467  DNCVTVPNSGQEDVDR
```

```
## [1] "Different Spec eValue: 54"
```

```
##      Scan MSGFDB_SpecEValue      Peptide  Spec_Evalue
## 12273 13817      5.721607e-21      MTDLTSSIPKPLLPVGNKPLIWYPLNLLER 5.471066e-21
## 12338 13889      1.390017e-21      SETCSVEIPELEFELGMAVLGGK 1.016951e-21
## 13037 14660      5.907738e-25      LLIPVVPGVDSLNSAMAASILLFEGK 5.911286e-25
##                               PepSeq
## 12273 MTDLTSSIPKPLLPVGNKPLIWYPLNLLER
## 12338      SETCSVEIPELEFELGMAVLGGK
## 13037      LLIPVVPGVDSLNSAMAASILLFEGK
```

```
## [1] "Different ID and Spec eValue: 717"
```

##	Scan	MSGFDB_SpecEValue	Peptide	Spec_Evalue
## 13372	15044	1.040916e-24	VGATAAVYSAAILEYLTAEVLELAGNASK	9.795364e-07
## 13378	15053	4.192276e-21	GEAIEAILAALEVSEPFRR	2.401887e-06
## 13387	15064	4.561789e-25	VTVEPQDSGTSALPLVSLFFYVVTGK	1.123928e-06
##			PepSeq	
## 13372			EKCTLFQLRCKTCRHIGKK	
## 13378			RKQEELQQLEQQRR	
## 13387			QLGWAILLLGTNTINSCGGLELVELGK	