Portable Proteomics Pipeline (P3)
MSGF Benchmark

Daniel Kristiyanto (daniel.kristiyanto@pnnl.gov)

May 18, 2016

---

1. Files:

```
## [1] "TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f11.mzML"
```

```
## [1] "ID_003632_9011437E.fasta"
```

2. Identification:

`MSGFPlus.jar` was downloaded from https://omics.pnl.gov/software/ms-gf. Identification was run multiple times, using different computers (Linux, MacOS, within or without Docker), and resulted the same results.

Identification with different switches resulted different results.

```
# Command Line (Without Switch)
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_no_switch.mzid \
-d ID_003632_9011437E.fasta

# 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


## [1] "TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f11_no_switch.mzid"
## [2] "TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f11.mzid"
```

3. Result from PNNL

```
##   ResultID  Scan                                          Peptide             Protein
## 1        1 11612 R.IQDLTGLDVSTAEELQVANYGVGGQYEPHFDFAR.K ref|NP_001017962.1
## 2        2 14441   K.LTPFIIQENLNLALNSASAIGCHVVNIGAEDLR.A    ref|NP_005023.2
##   MSGFDB_SpecEValue
## 1      1.481722e-36
## 2      8.834771e-36
```

4. Result from MSGF+ without any switches (sorted by SpecE-Value)

```
##          scan number(s) ms-gf:denovoscore ms-gf:specevalue
## X9740.1           10815               277     4.434865e-30
## X8340.1            9275               234     2.157325e-27
## X10711.1          11883               157     1.386474e-25
##                                                                             accession
## X9740.1  ref|NP_001041660.1;ref|NP_001041659.1;ref|NP_001041664.1;ref|NP_001260.1
## X8340.1                         ref|NP_001092.1;ref|NP_001186883.1;ref|NP_001605.1
## X10711.1                                       ref|NP_001003962.1;ref|NP_001740.1
##                                pepseq
## X9740.1  SHSTEPGLVLTLGQGDVGQLGLGENVMER
## X8340.1          DLYANTVLSGGTTMYPGIADR
## X10711.1      ILGGVISAISEAAAQYNPEPPPPR
```

```
##          scan number(s) ms-gf:denovoscore ms-gf:specevalue     accession
## X13036.1          14441               381     4.861374e-08 ref|NP_683515.3
## X10464.1          11612               279     5.246882e-08 ref|NP_000368.1
##                                pepseq
## X13036.1     KETDLKQIQTLIQGTQTRLKYSQNELEMIKK
## X10464.1 AGISEAQLTDAETSKLIYDFIEDQGGLEAVRQEMR
```

Reading the MZID file manually

```
java -Xmx2000M -XX:+UseConcMarkSweepGC -cp ../P3/MSGFPlus.jar \
edu.ucsd.msjava.ui.MzIDToTsv -i \
TCGA_13-1489_42-2590_36-2529_117C_W_PNNL_B2S5_f11_no_switch.mzid \
-showQValue 1 -showDecoy 1 -unroll 1
```

```
##   ScanNum                              Peptide            Protein SpecEValue
## 1   10815 R.SHSTEPGLVLTLGQGDVGQLGLGENVMER.K ref|NP_001041660.1   4.43e-30
## 2   10815 R.SHSTEPGLVLTLGQGDVGQLGLGENVMER.K ref|NP_001041659.1   4.43e-30
## 3   10815 R.SHSTEPGLVLTLGQGDVGQLGLGENVMER.K ref|NP_001041664.1   4.43e-30
```

5. Results with switches (2)

```
##          scan number(s) ms-gf:denovoscore ms-gf:specevalue
## X11197.1          12418               321     6.104450e-34
## X9050.1           10056               258     4.553317e-33
## X9060.1           10067               288     9.340615e-33
##                                                                  accession
## X11197.1                                                   ref|NP_000683.3
## X9050.1  ref|NP_004376.2;ref|NP_001157569.1;ref|NP_001157570.1;ref|NP_001119808.1
## X9060.1  ref|NP_004376.2;ref|NP_001157569.1;ref|NP_001157570.1;ref|NP_001119808.1
##                              pepseq
## X11197.1 VYLASLETLDNGKPFQESYALDLDEVIK
## X9050.1       CLCLPSYVGALCEQDTETCDYGWHK
## X9060.1       CLCLPSYVGALCEQDTETCDYGWHK


##          scan number(s) ms-gf:denovoscore ms-gf:specevalue      accession
## X13036.1          14441               345     2.196462e-06 ref|NP_060250.2
## X10464.1          11612               279     2.280634e-06 ref|NP_000487.1
##                                       pepseq
## X13036.1 FILPNVSTPVSDAFKTQMELLQAGLSRTPTR
## X10464.1   GDYKDSSDFGAPHPQVQSVRRIRDMQWHQR
```

Reading the MZID file manually

```
##    ScanNum
## 1    12418
## 2    10056
## 3    10056
##                                                             Peptide
## 1                  R.+144.102VYLASLETLDNGK+144.102PFQESYALDLDEVIK+144.102.V
## 2 R.+144.102C+57.021LC+57.021LPSYVGALC+57.021EQDTETC+57.021DYGWHK+144.102.F
## 3 R.+144.102C+57.021LC+57.021LPSYVGALC+57.021EQDTETC+57.021DYGWHK+144.102.F
##            Protein SpecEValue
## 1    ref|NP_000683.3   6.10e-34
## 2    ref|NP_004376.2   4.55e-33
## 3 ref|NP_001157569.1   4.55e-33
```