

Portable Proteomics Pipeline (P3)

MSGF Benchmark

Daniel Kristiyanto (daniel.kristiyanto@pnnl.gov)

May 25, 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 (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

```
## [1] 12695
```

```
##      Scan MSGFDB_SpecEValue      Peptide
## 1 11612      1.481722e-36 IQDLTGLDVSTAEELQVANYGVGGQYEPHFDFAR
## 2 14441      8.834771e-36 LTPFIIQENLNLALNSASAIGCHVNVNIGAEDLR
```

```
##      Scan MSGFDB_SpecEValue Peptide
## 12947 3823      0.04856372 SPAGGG
## 12948 4099      0.05022644 PGGAGM
```

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
##
##      pepseq
## X9740.1 SHSTEPGLVLTGQGQDVGQLGLGENVMER
## X8340.1 DLYANTVLSGGTTMYPGIADR
## X10711.1 ILGGVISAISEAAAQYNPEPPPPR
```

```
##      scan number(s) ms-gf:denovoscore ms-gf:specevalue
## X13036.1     14441      381      4.861374e-08
## X10464.1     11612      279      5.246882e-08
##
##      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.SHSTEPGLVLTGQGQDVGQLGLGENVMER.K ref|NP_001041660.1 4.43e-30
## 2      10815 R.SHSTEPGLVLTGQGQDVGQLGLGENVMER.K ref|NP_001041659.1 4.43e-30
## 3      10815 R.SHSTEPGLVLTGQGQDVGQLGLGENVMER.K ref|NP_001041664.1 4.43e-30
```

5. Results with switches

```
##      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
##                                     pepseq
## X11197.1 VYLASLETLDNGKPFQESYALDLDEVIK
## X9050.1      CLCLPSYVGALCEQDTETCDYGWHK
## X9060.1      CLCLPSYVGALCEQDTETCDYGWHK

##          scan number(s) ms-gf:denovoscore ms-gf:specvalue      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 FILPNVSTPVSDAFKTMELLQAGLSRTPTR
## 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
##      SpecEValue
## 1      6.10e-34
## 2      4.55e-33
## 3      4.55e-33
```

6. Filtering (Docker → PNNL)

a Spec-Evalue 10^{-10}

From PNNL

```
## [1] "==Head=="

##      Scan MSGFDB_SpecEValue      Peptide
## 1 11612      1.481722e-36 IQDLTGLDVSTAEELQVANYGVGGQYEPHFDFAR
## 2 14441      8.834771e-36 LTPFIIQENLNLALNSASAIGCHVNVNIGAEDLR

## [1] "==Tail=="

##      Scan MSGFDB_SpecEValue      Peptide
## 4634 7324      9.821299e-11 VTGTQPITCTWMK
## 4635 12688      9.998422e-11 SFLFQLLK
```

There are 4633 rows remained.

From Docker

```
## [1] "==Head=="
```

```
##          scan number(s) ms-gf:specvalue      pepseq
## X11197.1          12418      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK
## X9050.1           10056      4.553317e-33  CLCLPSYVGALCEQDTETCDYGWHK
```

```
## [1] "==Tail=="
```

```
##          scan number(s) ms-gf:specvalue      pepseq
## X3022.1          3426      9.690560e-11  MKPEFEK
## X6566.1          7324      9.821299e-11  VTGTQPITCTWMK
```

There are 1665 rows remained.

Differences

Data from Docker is matched with from PNNL using scan number as ID. ms-gf:specvalue, pepseq are values from Docker. MSGFDB_SpecEValue, Peptide are the corresponding values from PNNL. There are 1652 of rows in total with 1227 exact match.

Aside from the 13 unmatched values, 425 have different spec-value, 101 of different peptide identification.

```
## [1] "==Head=="
```

```
##          scan number(s) ms-gf:specvalue      pepseq
## 1580          12418      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK
## 1431          10056      4.553317e-33  CLCLPSYVGALCEQDTETCDYGWHK
##          MSGFDB_SpecEValue      Peptide
## 1580      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK
## 1431      4.553317e-33  CLCLPSYVGALCEQDTETCDYGWHK
```

```
## [1] "==Tail=="
```

```
##          scan number(s) ms-gf:specvalue      pepseq MSGFDB_SpecEValue
## 240          3426      9.690560e-11  MKPEFEK      9.690560e-11
## 1029          7324      9.821299e-11  VTGTQPITCTWMK      9.821299e-11
##          Peptide
## 240          MKPEFEK
## 1029 VTGTQPITCTWMK
```

```
## [1] "==Unmatched=="
```

```
##      scan number(s) ms-gf:specvalue
## 1090      7622      4.957978e-18
## 1554      12009     3.005655e-16
##
##                                pepseq MSGFDB_SpecEValue Peptide
## 1090                                EQQEAIIEHIDEVQNEIDR      NA    <NA>
## 1554 SGINPFDSQEAKPYKNDPEILAMTNLVSAYQNNDITEFEK      NA    <NA>
```

```
## [1] "=="Different Spec Values=="
```

```
##      scan number(s) ms-gf:specvalue                                pepseq
## 1510      11078      7.059070e-31      MGLDPACQVDIICGDHLLLEQYQTLR
## 1393      9602      1.788561e-30 WSGPLSLQEVDQPPHPLHVTYAGAAVDELGK
##      MSGFDB_SpecEValue                                Peptide
## 1510      5.495878e-31      MGLDPACQVDIICGDHLLLEQYQTLR
## 1393      2.682754e-30 WSGPLSLQEVDQPPHPLHVTYAGAAVDELGK
```

```
## [1] "=="Different Peptides identification=="
```

```
##      scan number(s) ms-gf:specvalue                                pepseq
## 1543      11669      1.659842e-23 DMETIGFAVYEVPPPELVGQPAVHLK
## 448      4656      3.368141e-21      QFEHLDPQNPQHTFEAR
##      MSGFDB_SpecEValue                                Peptide
## 1543      1.541134e-06 EKGSYGRIYNANIYTSLSDTVNK
## 448      2.303560e-07      GDPECHLFYNEQQEAR
```

b. Spec-Evalue 10^{-15}

From PNNL

```
## [1] "=="Head=="
```

```
##      Scan MSGFDB_SpecEValue                                Peptide
## 1 11612      1.481722e-36 IQDLTGLDVSTAEELQVANYGVGGQYEPHFDFAR
## 2 14441      8.834771e-36 LTPFIIQENLNLALNSASAIGCHVNVNIGAEDLR
```

```
## [1] "=="Tail=="
```

```
##      Scan MSGFDB_SpecEValue                                Peptide
## 2510 13805      9.825662e-16 VLSEIFSPMLFR
## 2511 8246      9.964967e-16 TMFSLDTYSTK
```

There are 2511 rows remained.

From Docker

```
## [1] "==Head=="
```

```
##          scan number(s) ms-gf:specvalue      pepseq
## X11197.1          12418      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK
## X9050.1           10056      4.553317e-33  CLCLPSYVGALCEQDTETCDYGWHK
```

```
## [1] "==Tail=="
```

```
##          scan number(s) ms-gf:specvalue      pepseq
## X5748.1           6424      9.527297e-16   QDCETFGMVVK
## X7929.1           8823      9.711903e-16  ITNQVIYLNPPIEECR
```

There are 767 rows remained.

Differences

Data from Docker is matched with from PNNL using scan number as ID. ms-gf:specvalue, pepseq are values from Docker. MSGFDB_SpecEValue, Peptide are the corresponding values from PNNL. There are 764 of rows in total with 572 exact match.

Aside from the 3 unmatched values, 192 have different spec-evalue, 23 of different peptide identification.

```
## [1] "==Head=="
```

```
##          scan number(s) ms-gf:specvalue      pepseq
## 715          12418      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK
## 629          10056      4.553317e-33  CLCLPSYVGALCEQDTETCDYGWHK
##          MSGFDB_SpecEValue      Peptide
## 715      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK
## 629      4.553317e-33  CLCLPSYVGALCEQDTETCDYGWHK
```

```
## [1] "==Tail=="
```

```
##          scan number(s) ms-gf:specvalue      pepseq MSGFDB_SpecEValue
## 295          6424      9.527297e-16   QDCETFGMVVK      9.527297e-16
## 535          8823      9.711903e-16  ITNQVIYLNPPIEECR      9.711903e-16
##          Peptide
## 295      QDCETFGMVVK
## 535  ITNQVIYLNPPIEECR
```

```
## [1] "==Unmatched=="
```

```
##      scan number(s) ms-gf:specvalue
## 448          7622      4.957978e-18
## 702          12009     3.005655e-16
##
##                                pepseq MSGFDB_SpecEValue Peptide
## 448                                EQQEAIIEHIDEVQNEIDR      NA    <NA>
## 702 SGINPFDSQEAKPYKNDPEILAMTNLVSAYQNNDDITEFEK      NA    <NA>
```

```
## [1] "=="Different Spec Values=="
```

```
##      scan number(s) ms-gf:specvalue                                pepseq
## 677          11078      7.059070e-31      MGLDPACQVDIICGDHLLEQYQTLR
## 606          9602      1.788561e-30 WSGPLSLQEVDQPPQHPLHVTYAGAAVDELGK
##      MSGFDB_SpecEValue                                Peptide
## 677      5.495878e-31      MGLDPACQVDIICGDHLLEQYQTLR
## 606      2.682754e-30 WSGPLSLQEVDQPPQHPLHVTYAGAAVDELGK
```

```
## [1] "=="Different Peptides identification=="
```

```
##      scan number(s) ms-gf:specvalue                                pepseq
## 696          11669      1.659842e-23 DMETIGFAVYEVPPPELVGQPAVHLK
## 148          4656      3.368141e-21      QFEHLDPQNHQHTFEAR
##      MSGFDB_SpecEValue                                Peptide
## 696      1.541134e-06 EKGSYGRIYNANIYTSLSDTVNK
## 148      2.303560e-07      GDPECHLFYNEQQEAR
```

c. Spec-Evalue 10^{-20}

From PNNL

```
## [1] "=="Head=="
```

```
##      Scan MSGFDB_SpecEValue                                Peptide
## 1 11612      1.481722e-36 IQDLTGLDVSTAEELQVANYGVGGQYEPHFDFAR
## 2 14441      8.834771e-36 LTPFIIQENLNLALNSASAIGCHVNVNIGAEDLR
```

```
## [1] "=="Tail=="
```

```
##      Scan MSGFDB_SpecEValue                                Peptide
## 935 13489      9.813857e-21 YETVIMPVFGIATPFHIATIK
## 936 13601      9.996277e-21 IAAGLPMAGIPFLTTDLTYR
```

There are 936 rows remained.

From Docker

```
## [1] "==Head=="

##          scan number(s) ms-gf:specvalue          pepseq
## X11197.1          12418      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK
## X9050.1           10056      4.553317e-33  CLCLPSYVGALCEQDTETCDYGWHK

## [1] "==Tail=="

##          scan number(s) ms-gf:specvalue          pepseq
## X7456.1           8303      8.950565e-21  TPALVFEHVNNTDFK
## X7761.1           8638      9.647618e-21  LILEQMQKDPQALSEHLK
```

There are 223 rows remained.

Differences

Data from Docker is matched with from PNNL using scan number as ID. ms-gf:specvalue, pepseq are values from Docker. MSGFDB_SpecEValue, Peptide are the corresponding values from PNNL. There are 223 of rows in total with 165 exact match.

Aside from the 0 unmatched values, 58 have different spec-evalue, 5 of different peptide identification.

```
## [1] "==Head=="

##          scan number(s) ms-gf:specvalue          pepseq
## 199          12418      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK
## 163          10056      4.553317e-33  CLCLPSYVGALCEQDTETCDYGWHK
##          MSGFDB_SpecEValue          Peptide
## 199          6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK
## 163          4.553317e-33  CLCLPSYVGALCEQDTETCDYGWHK

## [1] "==Tail=="

##          scan number(s) ms-gf:specvalue          pepseq MSGFDB_SpecEValue
## 110          8303      8.950565e-21  TPALVFEHVNNTDFK      6.808255e-21
## 116          8638      9.647618e-21  LILEQMQKDPQALSEHLK      7.600654e-21
##          Peptide
## 110          TPALVFEHVNNTDFK
## 116          LILEQMQKDPQALSEHLK

## [1] "==Unmatched=="

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



```
## [1] "=="Different Spec Values=="
```

```
##      scan number(s) ms-gf:specvalue      pepseq
## 185      11078      7.059070e-31      MGLDPACQVDIICGDHLLLEQYQTLR
## 151      9602      1.788561e-30 WSGPLSLQEVDQPHPLHVTYAGAAVDELGK
##      MSGFDB_SpecEValue      Peptide
## 185      5.495878e-31      MGLDPACQVDIICGDHLLLEQYQTLR
## 151      2.682754e-30 WSGPLSLQEVDQPHPLHVTYAGAAVDELGK
```

```
## [1] "=="Different Peptides identification=="
```

```
##      scan number(s) ms-gf:specvalue      pepseq
## 194      11669      1.659842e-23 DMETIGFAVYEVPPPELVGQPAVHLK
## 24      4656      3.368141e-21      QFEHLDPPQNQHTFEAR
##      MSGFDB_SpecEValue      Peptide
## 194      1.541134e-06 EKGSYGRIYNANIYTSLSDTVNK
## 24      2.303560e-07      GDPECHLFYNEQQEAR
```

7. Filtering (PNNL -> Docker)

a. Spec-Evalue 10^{-10}

From PNNL

```
## [1] "=="Head=="
```

```
##      Scan MSGFDB_SpecEValue      Peptide
## 1 11612      1.481722e-36 IQDLTGLDVSTAEELQVANYGVGGQYEPHFDFAR
## 2 14441      8.834771e-36 LTPFIIQENLNLALNSASAIGCHVVNIGAEDLR
```

```
## [1] "=="Tail=="
```

```
##      Scan MSGFDB_SpecEValue      Peptide
## 4634 7324      9.821299e-11 VTGTQPITCTWMK
## 4635 12688      9.998422e-11 SFLFQLLK
```

There are 4633 rows remained.

From Docker

```
## [1] "=="Head=="
```

```
##      scan number(s) ms-gf:specvalue      pepseq
## X11197.1      12418      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK
## X9050.1      10056      4.553317e-33 CLCLPSYVGALCEQDTETCDYGWHK
```

```
## [1] "=="Tail=="
```

```
##          scan number(s) ms-gf:specvalue      pepseq
## X3022.1          3426      9.690560e-11      MKPEFEK
## X6566.1          7324      9.821299e-11 VTGTQPITCTWMK
```

There are 1665 rows remained.

Differences

Data from Docker is matched with from PNNL using scan number as ID. MSGFDB_SpecEValue, Peptide are values from PNNL. ms-gf:specvalue, pepseq are the corresponding values from Docker Container. There are 4621 of rows in total with 1227 exact match.

Aside from the 12 unmatched values, 3394 have different spec-evalue, 3069 of different peptide identification.

```
## [1] "=="Head=="
```

```
##          Scan MSGFDB_SpecEValue      Peptide ms-gf:specvalue
## 3641 12418      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK      6.104450e-34
## 2717 10056      4.553317e-33  CLCLPSYVGALCEQDTETCDYGWHK      4.553317e-33
##
##          pepseq
## 3641 VYLASLETLDNGKPFQESYALDLDEVIK
## 2717  CLCLPSYVGALCEQDTETCDYGWHK
```

```
## [1] "=="Unmatched=="
```

```
##          Scan MSGFDB_SpecEValue      Peptide ms-gf:specvalue pepseq
## 2291 8856      1.119096e-15 CSEGVFLLTTPR      NA      <NA>
## 1416 6555      6.590530e-15 CQSLQEELDFRK      NA      <NA>
## 905 4786      2.062452e-14  YDDMAACMK      NA      <NA>
```

```
## [1] "=="Tail=="
```

```
##          Scan MSGFDB_SpecEValue      Peptide ms-gf:specvalue
## 2508 9513      2.871564e-13  NIIGSSPVADFS AIK      8.381110e-05
## 223 1927      5.243567e-25  CCYDGACVNNDETCEQR      9.379988e-05
##
##          pepseq
## 2508 LVDETEGQCGETDPNSEMPR
## 223  HECCFRYHCTGCCFR
```

```
## [1] "=="Unmatched=="
```

```
##          Scan MSGFDB_SpecEValue      Peptide ms-gf:specvalue pepseq
## 2291 8856      1.119096e-15 CSEGVFLLTTPR      NA      <NA>
## 1416 6555      6.590530e-15 CQSLQEELDFRK      NA      <NA>
```

```
## [1] "=="Different Spec Values=="
```

```
##      Scan MSGFDB_SpecEValue      Peptide
## 3127 11078      5.495878e-31      MGLDPACQVDIICGDHLLEQYQTLR
## 2548 9602      2.682754e-30 WSGPLSLQEVDQPHPLHVTYAGAAVDELGK
##      ms-gf:specvalue      pepseq
## 3127      7.059070e-31      MGLDPACQVDIICGDHLLEQYQTLR
## 2548      1.788561e-30 WSGPLSLQEVDQPHPLHVTYAGAAVDELGK
```

```
## [1] "=="Different Peptides identification=="
```

```
##      Scan MSGFDB_SpecEValue      Peptide ms-gf:specvalue
## 2595 9730      6.005831e-22      FFDHSGTLVMDAYEPEISR      2.231272e-20
## 4204 13789      4.811341e-14 NAQMAQSPILLGGAASTLLQNR      6.118720e-19
##      pepseq
## 2595 DLYANTVLSGGTTMYPGIADR
## 4204 IQTQLNLIHPDIFPLTSFR
```

b. Spec-Evalue 10^{-15}

From PNNL

```
##      Scan MSGFDB_SpecEValue      Peptide
## 1 11612      1.481722e-36 IQDLTGLDVSTAEELQVANYGVGGQYEPHFDFAR
## 2 14441      8.834771e-36 LTPFIQENLNLALNSASAIGCHVFNIGAEDLR

##      Scan MSGFDB_SpecEValue      Peptide
## 2510 13805      9.825662e-16 VLSEIFSPMLFR
## 2511 8246      9.964967e-16 TMFSLDTYSTK
```

There are 2511 rows remained.

From Docker

```
## [1] "=="Head=="
```

```
##      scan number(s) ms-gf:specvalue      pepseq
## X11197.1      12418      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK
## X9050.1      10056      4.553317e-33      CLCLPSYVGALCEQDTETCDYGWHK
```

```
## [1] "=="Tail=="
```

```
##      scan number(s) ms-gf:specvalue      pepseq
## X5748.1      6424      9.527297e-16      QDCETFGMVVK
## X7929.1      8823      9.711903e-16 ITNQVIYLNPPIEECR
```

There are 767 rows remained.

Differences

Data from Docker is matched with from PNNL using scan number as ID. MSGFDB_SpecEValue, Peptide are values from PNNL. ms-gf:specvalue, pepseq are the corresponding values from Docker Container. There are 2511 of rows in total with 572 exact match.

Aside from the 0 unmatched values, 1939 have different spec-evalue, 1769 of different peptide identification.

```
## [1] "==Head=="
```

```
##      Scan MSGFDB_SpecEValue      Peptide ms-gf:specvalue
## 1763 12418      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK      6.104450e-34
## 1170 10056      4.553317e-33 CLCLPSYVGALCEQDTETCDYGWHK      4.553317e-33
##
##      pepseq
## 1763 VYLASLETLDNGKPFQESYALDLDEVIK
## 1170 CLCLPSYVGALCEQDTETCDYGWHK
```

```
## [1] "==Unmatched=="
```

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

```
## [1] "==Tail=="
```

```
##      Scan MSGFDB_SpecEValue      Peptide ms-gf:specvalue
## 49 1876      1.449447e-17 QNEAAVGNCAEHMR      3.858078e-05
## 50 1927      5.243567e-25 CCYDGACVNNDTCEQR      9.379988e-05
##
##      pepseq
## 49 DSFMRHRGH CER
## 50 HECCFRYHCTGCCFR
```

```
## [1] "==Unmatched=="
```

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

```
## [1] "==Different Spec Values=="
```

```
##      Scan MSGFDB_SpecEValue      Peptide
## 1437 11078      5.495878e-31 MGLDPACQVDIICGDHLEQYQTLR
## 1057 9602      2.682754e-30 WSGPLSLQEVDEQPQHPLHVTYAGAAVDELGK
##
##      ms-gf:specvalue      pepseq
## 1437      7.059070e-31 MGLDPACQVDIICGDHLEQYQTLR
## 1057      1.788561e-30 WSGPLSLQEVDEQPQHPLHVTYAGAAVDELGK
```

```
## [1] "=="Different Peptides identification=="
```

```
##      Scan MSGFDB_SpecEValue      Peptide ms-gf:specvalue
## 1091 9730      6.005831e-22 FFDHSGTLVMDAYEPEISR      2.231272e-20
## 123 3378      2.191136e-20  ASGNYATVISHNPETKK      2.189761e-14
##
##      pepseq
## 1091 DLYANTVLSGGTTMYPGIADR
## 123  NKPGPYSSVPPPSAPPPK
```

c. Spec-Evalue 10^{-20}

From PNNL

```
##      Scan MSGFDB_SpecEValue      Peptide
## 1 11612      1.481722e-36 IQDLTGLDVSTAEELQVANYGVGGQYEPHFDFAR
## 2 14441      8.834771e-36 LTPFIIQENLNLALNSASAIGCHVVNIGAEDLR
```

```
##      Scan MSGFDB_SpecEValue      Peptide
## 935 13489      9.813857e-21 YETVIMPVFGIATPFHIATIK
## 936 13601      9.996277e-21 IAAGLPMAGIPFLTTLTYR
```

There are 936 rows remained.

From Docker

```
## [1] "=="Head=="
```

```
##      scan number(s) ms-gf:specvalue      pepseq
## X11197.1      12418      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK
## X9050.1      10056      4.553317e-33  CLCLPSYVGALCEQDTETCDYGWHK
```

```
## [1] "=="Tail=="
```

```
##      scan number(s) ms-gf:specvalue      pepseq
## X7456.1      8303      8.950565e-21  TPALVFEHVNNNTDFK
## X7761.1      8638      9.647618e-21 LILEQMOKDPQALSEHLK
```

There are 223 rows remained.

Differences

Data from Docker is matched with from PNNL using scan number as ID. MSGFDB_SpecEValue, Peptide are values from PNNL. ms-gf:specvalue, pepseq are the corresponding values from Docker Container. There are 936 of rows in total with 165 exact match.

Aside from the 0 unmatched values, 771 have different spec-evalue, 717 of different peptide identification.

```
## [1] "==Head=="
```

```
##      Scan MSGFDB_SpecEValue      Peptide ms-gf:specvalue
## 571 12418      6.104450e-34 VYLASLETLDNGKPFQESYALDLDEVIK      6.104450e-34
## 327 10056      4.553317e-33 CLCLPSYVGALCEQDTETCDYGWHK      4.553317e-33
##
##                pepseq
## 571 VYLASLETLDNGKPFQESYALDLDEVIK
## 327 CLCLPSYVGALCEQDTETCDYGWHK
```

```
## [1] "==Unmatched=="
```

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

```
## [1] "==Tail=="
```

```
##      Scan MSGFDB_SpecEValue      Peptide ms-gf:specvalue
## 19 3273      1.342643e-21 ESCNPQDCVPEVDENCKDK      2.671039e-05
## 12 1927      5.243567e-25 CCYDGACVNNDTCEQR      9.379988e-05
##
##                pepseq
## 19 AEGAYYDNHSVVYLDHGGENGEK
## 12      HECCFRYHCTGCCFR
```

```
## [1] "==Unmatched=="
```

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

```
## [1] "==Different Spec Values=="
```

```
##      Scan MSGFDB_SpecEValue      Peptide
## 450 11078      5.495878e-31      MGLDPACQVDIICGDHLLQYQTLR
## 284 9602      2.682754e-30 WSGPLSLQEVDEQPQHPLHVTYAGAAVDELGK
##
##      ms-gf:specvalue      pepseq
## 450      7.059070e-31      MGLDPACQVDIICGDHLLQYQTLR
## 284      1.788561e-30 WSGPLSLQEVDEQPQHPLHVTYAGAAVDELGK
```

[1] "==Different Peptides identification=="

##	Scan	MSGFDB_SpecEValue	Peptide	ms-gf:specvalue
## 299	9730	6.005831e-22	FFDHSGTLVMDAYEPEISR	2.231272e-20
## 401	10627	4.148874e-23	GMSLNLEPDNVGVVFGNDK	7.877905e-10
##		pepseq		
## 299			DLYANTVLSGGTTMYPGIADR	
## 401			GNAEPDPPVRLFQIHGNDK	