

MzIdentMLFormatConverter

This pdf file describes and depicts the input method for both mzid and fasta files for this tool. The input method can look complicated but is actually quite easy to create. First, create a main text file that will be used as the input for the -mzid argument. This main text file should consist of one or more pathways to other text files. These other text files should contain the pathways to the mzid file of each sample. See figure 1 for a more detailed overview. In this example, the 1D50CM.txt is the main text file. This text file contains a file path to a second text file called 1D50CM_Individual_mzid.txt. The second text file contains the file path to the mzid files. It is possible to add another text file with data to the main text file to process multiple datasets.

When creating the text file containing the mzid files make sure that each file path has the same dataset and sample type name. For this input file the dataset name is 1D50CM and the sample type name is Individual. Be sure to have a sample number ready in the sample file. This sample number is indicated by the lower spaces. Your sample number should be separated from the rest of the sample name by underscores, as shown in figure 1.

```
./Entry
├── 1D50CM.txt
└── 1D50CM_Individual_mzid.txt

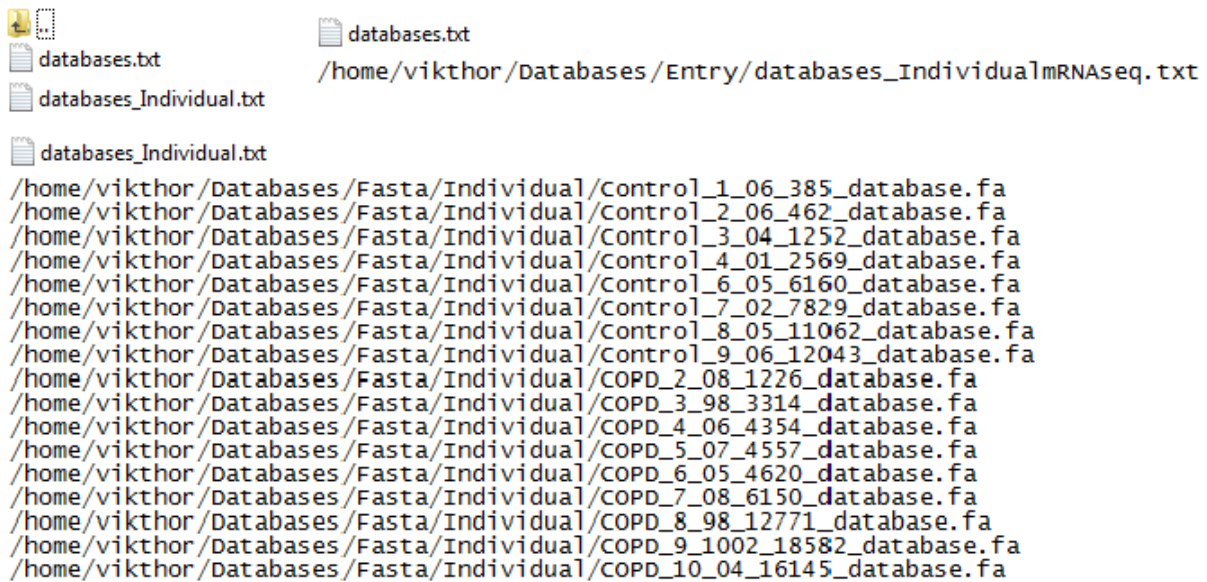
/home/vikthor/Entry/1D50CM.txt

/home/vikthor/mzIdentML/1D50cm/Individual/150210_1_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_2_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_3_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_4_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_6_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_7_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_8_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_9_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_11_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_13_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_14_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_15_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_16_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_17_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_18_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_19_01.mzid
/home/vikthor/mzIdentML/1D50cm/Individual/150210_20_01.mzid
```

Figure 1. MzIdentML format input.

The MzIdentML input is done by using text files that contain paths to the mzid files. The file used for the -mzid input parameter is the 1D50CM.txt file. This text file should contain a pathway to the text file of each set of samples used for this data set. In this case only samples of the individual data set are present. The contents of this file should only consist of pathways to the mzid files of each sample. It is possible to add other entries to the main 1D50CM.txt file to analyze multiple sets of samples.

The database text file is created in a similar matter. Once more a main text file is created that will contain file paths to other text files. These other text files will contain the file paths to the protein sequence fasta files. When creating the text file containing the fasta files make sure that each file path has the same sample type name. For this input file the sample type name is Individual. Be sure to have a sample number ready in the sample file. This sample number is indicated by the lower spaces. Your sample number should be separated from the rest of the sample name by underscores, as shown in figure 1.



```

databases.txt
/home/vikthor/Databases/Entry/databases_IndividualmRNAseq.txt

databases_Individual.txt
/home/vikthor/Databases/Fasta/Individual/Control_1_06_385_database.fa
/home/vikthor/Databases/Fasta/Individual/Control_2_06_462_database.fa
/home/vikthor/Databases/Fasta/Individual/Control_3_04_1252_database.fa
/home/vikthor/Databases/Fasta/Individual/Control_4_01_2569_database.fa
/home/vikthor/Databases/Fasta/Individual/Control_6_05_6160_database.fa
/home/vikthor/Databases/Fasta/Individual/Control_7_02_7829_database.fa
/home/vikthor/Databases/Fasta/Individual/Control_8_05_11062_database.fa
/home/vikthor/Databases/Fasta/Individual/Control_9_06_12043_database.fa
/home/vikthor/Databases/Fasta/Individual/COPD_2_08_1226_database.fa
/home/vikthor/Databases/Fasta/Individual/COPD_3_98_3314_database.fa
/home/vikthor/Databases/Fasta/Individual/COPD_4_06_4354_database.fa
/home/vikthor/Databases/Fasta/Individual/COPD_5_07_4557_database.fa
/home/vikthor/Databases/Fasta/Individual/COPD_6_05_4620_database.fa
/home/vikthor/Databases/Fasta/Individual/COPD_7_08_6150_database.fa
/home/vikthor/Databases/Fasta/Individual/COPD_8_98_12771_database.fa
/home/vikthor/Databases/Fasta/Individual/COPD_9_1002_18582_database.fa
/home/vikthor/Databases/Fasta/Individual/COPD_10_04_16145_database.fa

```

Figure 2. Database input layout.

This picture shows the D:/Input/Databases/ map which contains the main databases.txt. This file contains the pathways to the individual text file. The contents of this individual text file consist of pathways to protein fasta files of each sample. These fasta files function as the protein sequence database of each individual sample.