A Step-by-Step Guide to Using UMPIRE

Weiliang Huang, Ph.D.

1. Check system requirements:

a. Minimum: 64-bit Windows 7, 3 GHz octa-core CPU, 320 GB of free disk space, 32 GB of free memory. Minimum/subminimum hardware may often run at a slower speed.

b. Recommended: 64-bit Windows 10, 3.5 GHz 32-core CPU, 320 GB of free SSD space, 128 GB of free memory and above.

2. Check computer user account privilege:

a. If the computer is administered by an IT administrator, please check whether your user account has been granted with a privilege to install and run new software.

3. Install UMPIRE on your computer:

a. Please ensure that you have 142 GB of free disk space to download the installation files and 318 GB of free disk space to install UMPIRE.

b. Please visit http://umpire.umaryland.edu/ and save both the installer and the latest version of UMPIRE installation package to the same location and double-click the installer to install UMPIRE. Installation of UMPIRE to a fast solid-state drive is recommended.

c. The UMPIRE installation package is a standard 7z compression file and can be decompressed to a location of your choice directly using any file archiver software supporting the 7z format without installation using the installer.

4. Prepare MS raw data for analysis:

a. Please convert MS raw files into Mascot generic files (MGF) using MSConvert or vendor software before use.

5. Provide supplementary databases for inference (optional):

a. UMPIRE comes with all prokaryotic protein sequences (bacteria and archaea) of UniProt database. You may want to add protein sequences of host, diet, or eukaryotes in the environment.

b. Save the additional protein sequences of your choice in standard FASTA format to the "UserDB" folder in the UMPIRE program folder. Precompiled proteome databases from UniProt saved with a file name like "UP123456789.fasta" are recommended.

c. Databases of common laboratory contaminants such as keratin and albumin are included by default.

6. Launch UMPIRE in the wizard mode (recommended for new users):

a. Please double-click the icon of UMPIRE.exe.

b. Following the prompt on screen, please enter the path to a MGF file or a folder of MGF files to be analyzed, for example:

```
D:\MS_Data\MGF_Folder\Example.mgf
```

To analyze all MGF files in a folder:

```
D:\MS_Data\MGF_Folder
```

c. Following the prompt on screen, please enter the precursor mass tolerance in Da, e.g., 0.01.

d. Following the prompt on screen, please enter the fragment mass tolerance in Da, e.g., 0.02 for high-resolution MS2 data and 0.5 for low-resolution MS2 data.

e. The analysis will start. The time to completion will depend on the CPU speed, the memory available, the hard drive speed, the number of files, and the number of spectra per file. It may take a few hours to process a file. Please be patient and avoid using other software consuming substantial CPU and memory during analysis.

f. An optimized protein sequence database in FASTA format for standard target-decoy search and an optimized database for Percolator validated search will be saved to the same folder of your MGF file(s).

7. Run UMPIRE via the command line (for advanced users and automation):

a. Please launch Command Prompt via Windows Start menu -> Windows System -> Command Prompt.

b. If you have high-resolution MS2 data, please run the following command:

```
C:\UMPIRE\UMPIRE.exe /File:D:\Data\Test.mgf /MS1Err:0.01 /MS2Err:0.02
```

or

```
C:\UMPIRE\UMPIRE.exe /File:D:\My_MGF_Folder /MS1Err:0.01 /MS2Err:0.02
```

c. If you have low-resolution MS2 data, please run the following command:

```
C:\UMPIRE\UMPIRE.exe /File:D:\Data\Test.mgf /MS1Err:0.01 /MS2Err:0.5
```

or

```
C:\UMPIRE\UMPIRE.exe /File:D:\My_MGF_Folder /MS1Err:0.01 /MS2Err:0.5
```

d. Please remember to replace the paths to the UMPIRE program folder and the MGF file(s) in the examples above with the real paths in your computer.

e. Please feel free to adjust the MS1Err (precursor mass tolerance) and the MS2Err (fragment mass tolerance) parameters to match your MS data.

f. The analysis will start. The time to completion will depend on the CPU speed, the memory available, the hard drive speed, the number of files, and the number of spectra per file. It may take a few hours to process a file. Please be patient and avoid using other software consuming substantial CPU and memory during analysis.

g. Optimized databases will be saved to the same folder of your MGF file(s) for standard target-decoy search and target-decoy search validated by Percolator, respectively.

8. If you have any questions or suggestions:

a. Please don't hesitate to contact Dr. Weiliang Huang (Weiliang.huang@uqconnect.edu.au).