

# Supplementary Note

## **GameTag: A New Sequence Tag Generation Algorithm Based on Cooperative Game Theory**

Zhengcong Fei<sup>1,2</sup>, Kaifei Wang<sup>1,2</sup>, Si-Min He<sup>1,2</sup>, Hao Chi<sup>1,2\*</sup>

<sup>1</sup>Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, *Beijing*, China

<sup>2</sup>University of Chinese Academy of Sciences, *Beijing*, China

\*To whom correspondence should be addressed:

Hao Chi: [chihao@ict.ac.cn](mailto:chihao@ict.ac.cn)

# 1. Dataset Annotation

In this section, we provide the detailed parameter settings of Open-pFind, PEAKS, and MSFragger, for dataset pre-processing and annotating. The intersection of identification results from the three can be considered as ground-truth PSMs. In practice, we employed `data_label_union.py` for intersection operation and result statistics.

## 1.1 Open-pFind

- 1) Tool download: <http://pfind.ict.ac.cn/software/pFind3/index.html>
- 2) Parameter setting interface:  
(Please note that no screenshots are all set by default)

MS Data

Property	Value
Format	RAW
Instrument	HCD-FTMS
Data File List	W:\liuchao_Mann\raw_MichalskiA\QExactive\velos\raw\20100825_Velos2_AnMi_QC_wt_HCD_iso4_swG.raw W:\liuchao_Mann\raw_MichalskiA\QExactive\velos\raw\20100825_Velos2_AnMi_QC_wt_HCD_iso4_swG_2.raw W:\liuchao_Mann\raw_MichalskiA\QExactive\velos\raw\20100826_Velos2_AnMi_SA_HeLa_4Da.raw
Mixture Spectra	True
Decimal Places of M/Z	5
Decimal Places of Intensity	1
Model	Normal
Threshold	-0.5

Search

Property	Value
Database	uniprot_contaminants_AILL_con
Enzyme	Trypsin KR _ C
Enzyme Specificity	Full-Specific
Number of Missed Cleavages	3
Precursor Tolerance	±20 ppm
Fragment Tolerance	±20 ppm
Open Search	True
Fixed Modifications	
Variable Modifications	

Filter

Property	Value
FDR	Less than 1% at Peptides Level
Peptide Mass	[600 , 10000]
Peptide Length	[6 , 100]
Number of Peptides Per Protein	At least 1
Protein FDR	1%

MS1 Quantitation

Property	Value
Quantitation	Labeling_None
Multiplicity	1
Label	None;

## 1.2 PEAKS

- 1) Tool download: <https://www.bioinfor.com/download-peaks-studio/>
- 2) Parameter setting interface:

Peptides  $-10\lg P \geq 11.2$   Proteins  $-10\lg P \geq 20$  and  $\geq 0$  unique peptides

De novo only ALC (%)  $\geq 50$  and  $-10\lg P \leq 11.2$

**PEAKS Search**

PEAKS Search Predefined parameters

**Error Tolerance**  
Precursor mass: 0.1 Da using monoisotopic mass Fragment ion: 0.2 Da

**Enzyme**  
Specified by each sample   
Allow non-specific cleavage at one end of the peptide.  
Maximum missed cleavages per peptide: 3

**PTM**  
  
  
  
Maximum allowed variable PTM per peptide 3

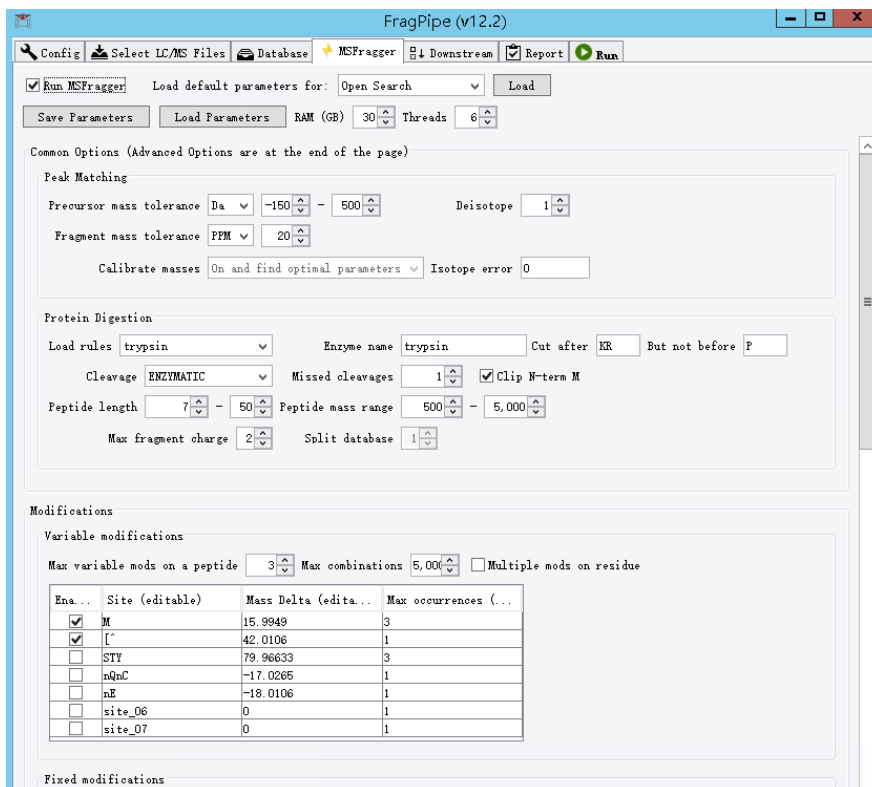
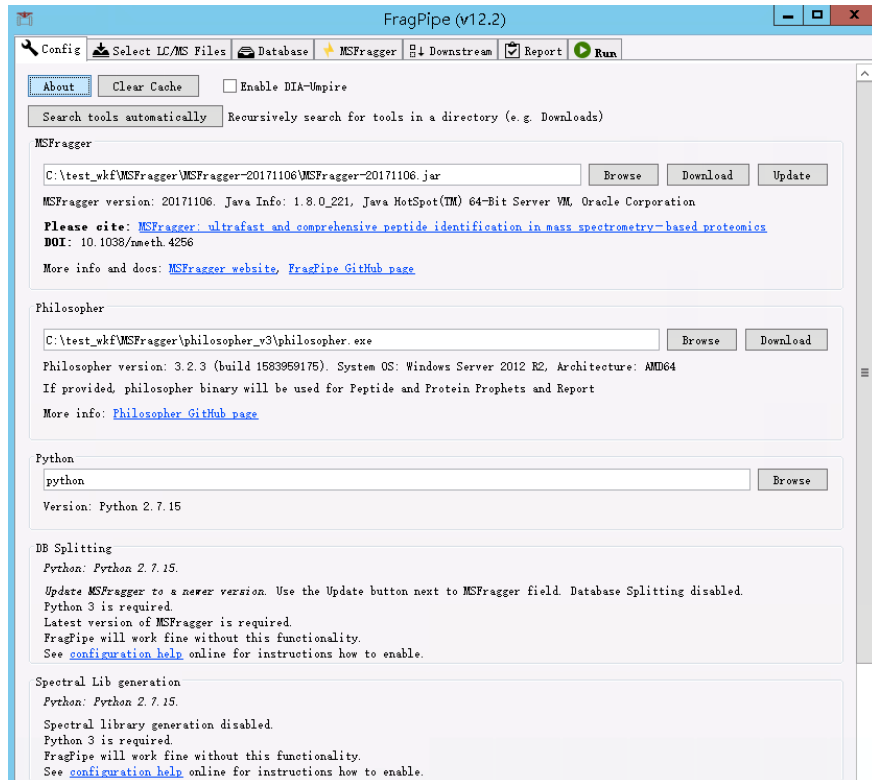
**Database**  
☒ Select database Database: human\_demo   
☐ Paste sequence Taxa: all species 

**De Novo Tag Options**  
Available de novo tags: de novo with current parameter

**General Options**  
☒ Estimate FDR with decoy-fusion.   
☒ Find unspecified PTMs and common mutations with PEAKS PTM  
☐ Find more mutations with SPIDER

### 1.3 MSFragger

- 1) Tool download: <http://msfragger.nesvilab.org/>
- 2) Parameter setting interface:



## 2. Baseline Setting

In this section, we give the parameter setting for the baselines employed in the experiments for tagging performance comparison.

### 2.1 InsPecT

- 1) Tool download: <http://proteomics.ucsd.edu/Software/Inspect/>
- 2) Dataset format conversion (we use .mgf as spectrum file):  
[http://tools.proteomecenter.org/wiki/index.php?title=Main\\_Page](http://tools.proteomecenter.org/wiki/index.php?title=Main_Page)
- 3) Database Setup: run CMD as belows  

```
> python PrepDB.py FASTA [myDB.fasta]
```
- 4) Parameter InputFile:  

```
speactra, [FILENAME.mgf]
protease, Trypsin
mod, 57, C, fix
TagCount, 100
TagLength, 5
```
- 5) Run the InsPect:  

```
> InsPecT.exe -i InputFile.text -o OutputFile.txt
```

( In order to output the extracted tags, we can adopt the “DEBUG” module, more detail can be refer to document: <http://proteomics.ucsd.edu/Software/Inspect/InspectDocs/> )

### 2.2 PepNovo+

- 1) Tool download: <https://github.com/jmchilton/pepnovo>
- 2) Run the pepnovo+ with the below command:  

```
>PepNovo.exe -file [mgfPath] -model CID_IT_TRYP
C+57:M+16 -digest NON_SPECIFIC -tag_length 5 -
num_solutions 100 -fragment_tolerance 0.01 >
[OutputFilePath]
```

(Note that we utilize the high-resolution data version for tag extraction. In practice, If the MS/MS spectra come from high-resolution instruments, the sequencing performance can be improved by manipulating the tolerances. For instance if the spectra have fragment tolerances of 0.01, this can be set with the flag: `-fragment_tolerance 0.01`.)

### 2.3 SVM:

- 1) Package download: <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- 2) Parameter setting:

Input features	Peak intensity Edge mass error Node relevance degree
Feature Normalization	Yes
Kernel	RBF
C	512
Gamma	0.03125
Validation	5-fold cross-validation

(For convenience, the tag features produced by tag generator are used directly.)

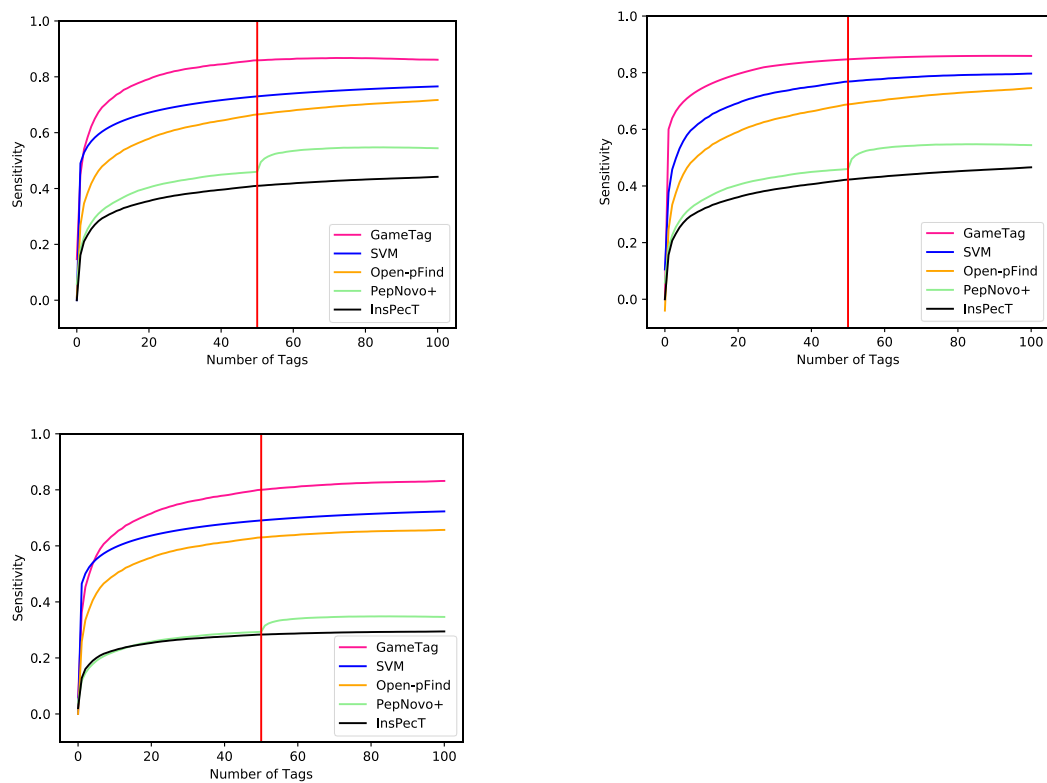
- 3) Train the SVM model as:  

```
>svm-train.exe -t 2 -g 0.03125 -c 512 -v 5 -s 1  
[DatasetName] [ModelName]
```
- 4) Predict the results using the trained SVM model:  

```
>svm-predict.exe [DatasetName] [ModelName] [OutputFile]
```

### 3. Additional Experimental Results

Due to the limitation of paper space, we will present some addition results here.



**Figure S1:** Sensitivity of different methods as a function of the number of tags considered for each spectrum on Gygi-Human-QE, Dong-Ecoli-QE, and Xu-Yeast-QEHF, respectively.