

# **TIDD: Tool-Independent and Data-Dependent PSM rescoring method**

Copyright (C) 2022 BIS Labs, Hanyang Univ. Korea

## **Outline:**

|                                                    |    |
|----------------------------------------------------|----|
| Introduction                                       | 2  |
| Install and run                                    | 3  |
| File format                                        | 9  |
| - TIDD input file format: PSM search result [*tsv] | 9  |
| - TIDD input file format: MS/MS spectra [*mgf]     | 10 |
| - Test data: download link                         | 10 |
| Run: GUI                                           | 11 |
| - File upload & download                           | 11 |
| - Data management                                  | 12 |
| - Model Fitting                                    | 14 |
| Run: command line                                  | 16 |
| - PSM feature extraction                           | 16 |
| - Calculate XCorr score                            | 17 |
| - Fitting models                                   | 17 |

## Introduction

A Tool-Independent and Data-Dependent PSM rescoring method.

TIDD is a universal post-processing tool which supports confident peptide identifications regardless of the search engine adopted. TIDD can work for any (including newly developed search engines) because it calculates universal features that assess peptide-spectrum match quality while it allows additional features provided by search engines (or users) as well.

Here, we support two types of TIDD version -- a simple GUI based TIDD and command version of TIDD.

GUI

1. Data Management: extract features and check feature densities"),
2. Model fitting: fit SVM models"),

\*Computational resources supported by free version of "R studio cloud" is limited, so it takes a little time for feature extraction."),

\*Sample file description:

1. test\_required.tsv: contain 6 required features for feature extraction
2. \*\_feature.stv: contain common features except calculated Xcorr
3. \*\_xcorrColAdded.tsv: add calculated Xcorr feature
4. \*\_ModelFitting\_Results.tsv: add rescoring score feature

Required:

1. R (v4.1 or above) packages: shiny, shinythemes, shinydashboard, e1071, ROCR
2. java jre/jdk v1.8 or above

Download:

1. Download code from github: <https://github.com/HanyangBISLab/TIDD.git>

2.Download/Test code from R studio cloud:

<https://rstudio.cloud/spaces/178915/project/2994889>

3.Test GUI from shinyapps.io: <https://honglan-li.shinyapps.io/project/>

## Install and run

It is based on R and java programming.

So it required:

- Install java jdk (17 or above) and R(4.1.2 or above)
- For R, the following packages are needed to install.  
`install.packages("shiny")` `install.packages("shinythemes")`  
`install.packages("shinyFiles")` `install.packages("shinydashboard")`  
`install.packages("ROCR")` `install.packages("e1071")`
- Run shinyApp

Detail: R and Java based program

1. Install R version 4.1.2 (or above)

link: <https://cran.r-project.org/>

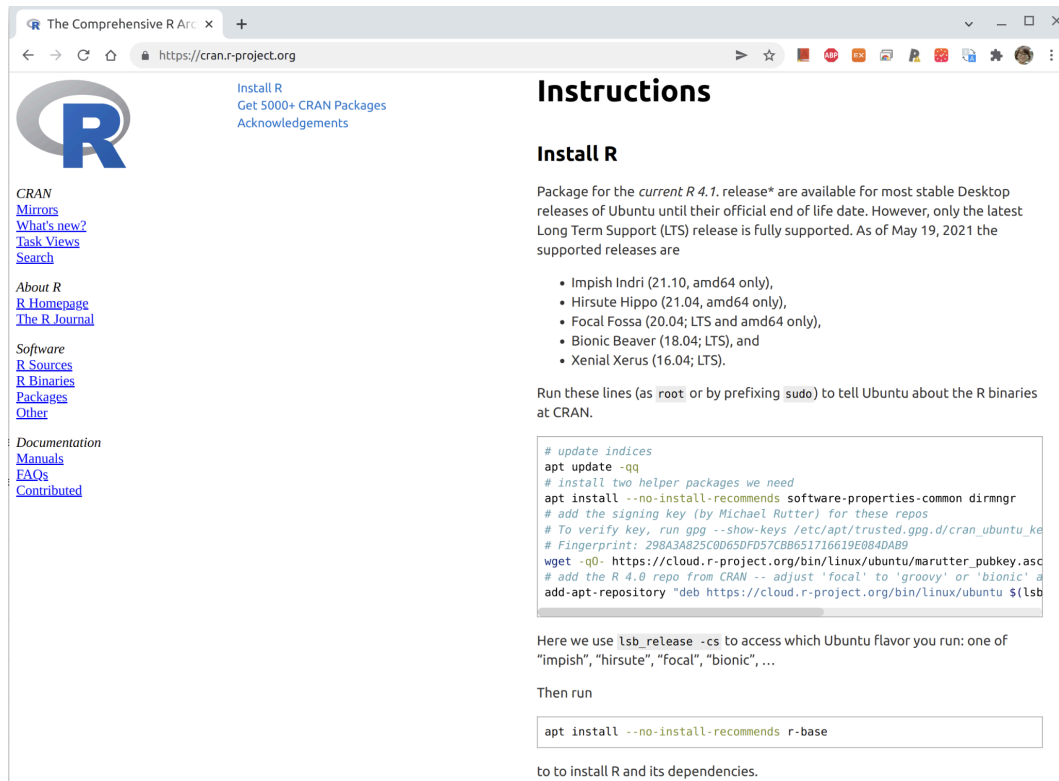
E.g. download R-4.1.2\* ver

The screenshot shows the CRAN website with the title 'The Comprehensive R Archive Network'. On the left, there is a sidebar with links: CRAN, Mirrors, What's new?, Task Views, Search, About R, R Homepage, The R Journal, Software, R Sources, R Binaries, Packages, Other, Documentation, Manuals, FAQs, and Contributed. The main content area is titled 'Download and Install R' and contains the following text: 'Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:'. Below this, there are three links: 'Download R for Linux (Debian, Fedora/Redhat, Ubuntu)', 'Download R for macOS', and 'Download R for Windows'. A blue box with the text 'Latest version (Now R 4.1.2)' is positioned over the 'Download R for Windows' link. Below the links, there is a section titled 'Source Code for all Platforms' with the following text: 'Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!'. Below this, there are four bullet points: 'The latest release (2021-11-01, Bird Hippie) [R-4.1.2.tar.gz](#), read [what's new](#) in the latest version.', 'Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).', 'Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.', and 'Source code of older versions of R is [available here](#)'.

Downloaded file: window→ R-4.1.2-win.exe → “double click” to install R

ubuntu → click “ubuntu” → You can see the guideline for installing

<captures below is from <https://cran.r-project.org/>>



The screenshot shows the CRAN website with the 'Instructions' page for installing R. The page includes a sidebar with navigation links and a main content area with the title 'Instructions' and a sub-header 'Install R'. The text explains that packages for the current R 4.1 release are available for most stable Desktop releases of Ubuntu until their official end of life date. It lists supported releases: Impish Indri (21.10, amd64 only), Hirsute Hippo (21.04, amd64 only), Focal Fossa (20.04; LTS and amd64 only), Bionic Beaver (18.04; LTS), and Xenial Xerus (16.04; LTS). It provides a list of terminal commands to update indices, install helper packages, add the signing key, and add the R 4.0 repo from CRAN. It also mentions using `lsb_release -cs` to access the Ubuntu flavor and provides a command to install R and its dependencies.

**Instructions**

### Install R

Package for the *current R 4.1* release\* are available for most stable Desktop releases of Ubuntu until their official end of life date. However, only the latest Long Term Support (LTS) release is fully supported. As of May 19, 2021 the supported releases are

- Impish Indri (21.10, amd64 only),
- Hirsute Hippo (21.04, amd64 only),
- Focal Fossa (20.04; LTS and amd64 only),
- Bionic Beaver (18.04; LTS), and
- Xenial Xerus (16.04; LTS).

Run these lines (as `root` or by prefixing `sudo`) to tell Ubuntu about the R binaries at CRAN.

```
# update indices
apt update -qq
# install two helper packages we need
apt install --no-install-recommends software-properties-common dirmngr
# add the signing key (by Michael Rutter) for these repos
# To verify key, run gpg --show-keys /etc/apt/trusted.gpg.d/cran_ubuntu_key.asc
# Fingerprint: 298A3A825C0D65DFD57CBB651716619E084DAB9
wget -qO- https://cloud.r-project.org/bin/linux/ubuntu/marutter_pubkey.asc
# add the R 4.0 repo from CRAN -- adjust 'focal' to 'groovy' or 'bionic' as needed
add-apt-repository "deb https://cloud.r-project.org/bin/linux/ubuntu $(lsb_release -cs)-r4.0"
apt update
```

Here we use `lsb_release -cs` to access which Ubuntu flavor you run: one of "impish", "hirsute", "focal", "bionic", ...

Then run

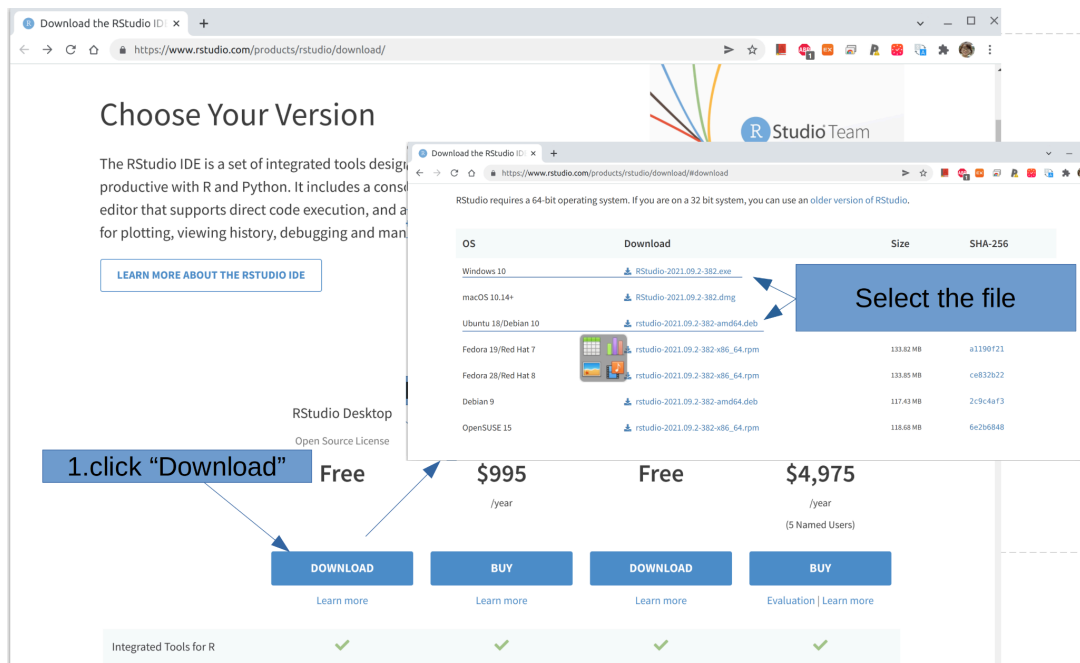
```
apt install --no-install-recommends r-base
```

to to install R and its dependencies.

## 2. Install R studio

link: <https://www.rstudio.com/products/rstudio/download/>

E.g. download Rstudio-2021.09.1\*



Install: window → double click "RStudio 202109\*.exe"

ubuntu → follow the guideline provided by the link below:

<https://linuxconfig.org/how-to-install-rstudio-on-ubuntu-20-04-focal-fossa-linux>

### 3. Install java 8

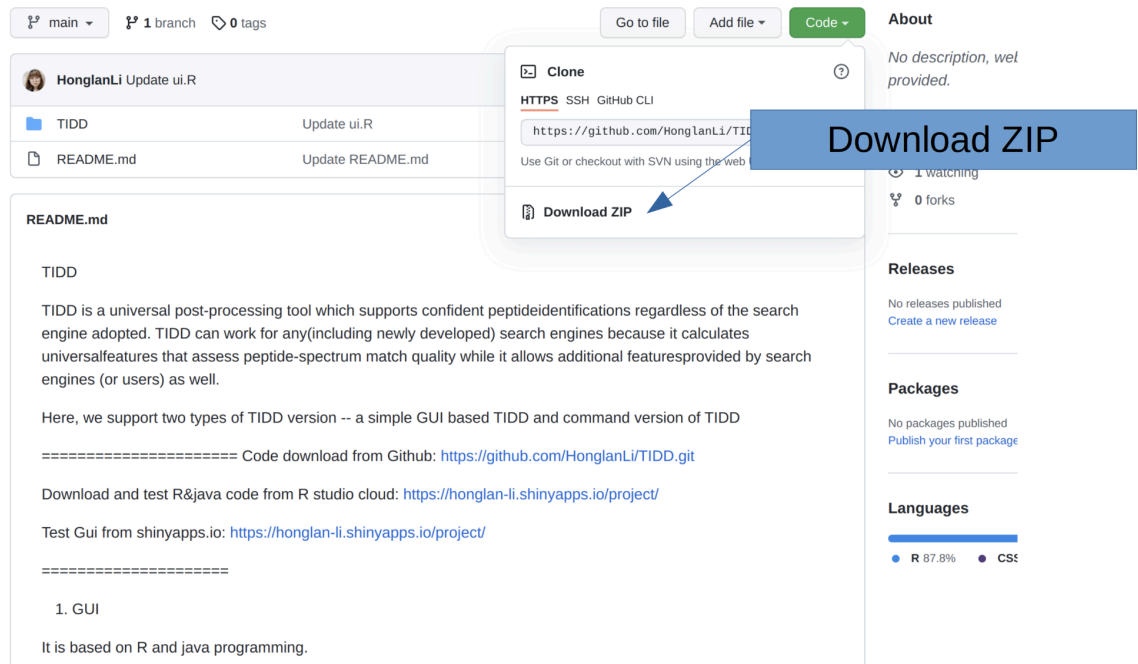
download: <https://www.oracle.com/java/technologies/downloads/>

Installation guide:

<https://docs.oracle.com/javase/10/install/installation-jdk-and-jre-microsoft-windows-platforms.htm#JSJIG-GUID-29333CFD-E7A6-498B-9317-97700C81D928>

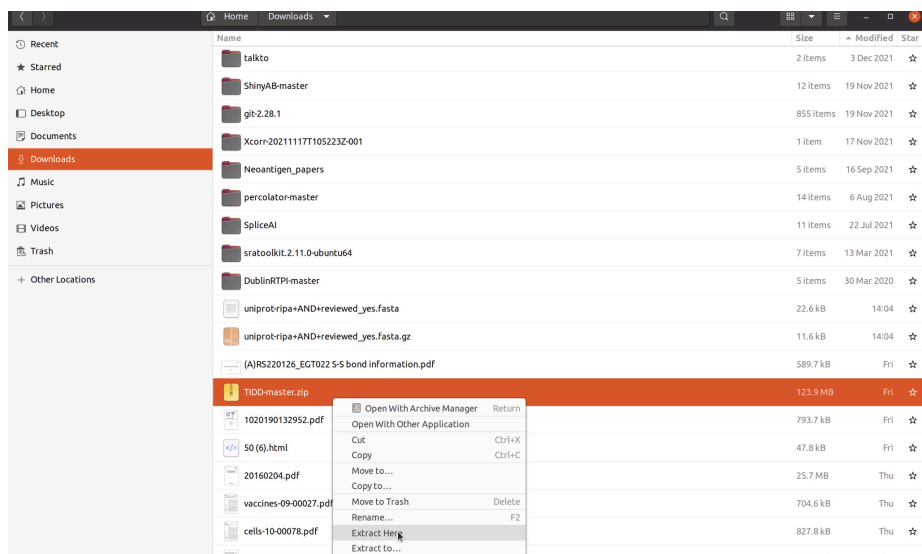
E.g. jdk-17.01\*

### 4. Download TIDD code from git: <https://github.com/HonglanLi/TIDD.git>



## 5. Unzip/Uncompress the TIDD\*.zip

Default location: ~/Download

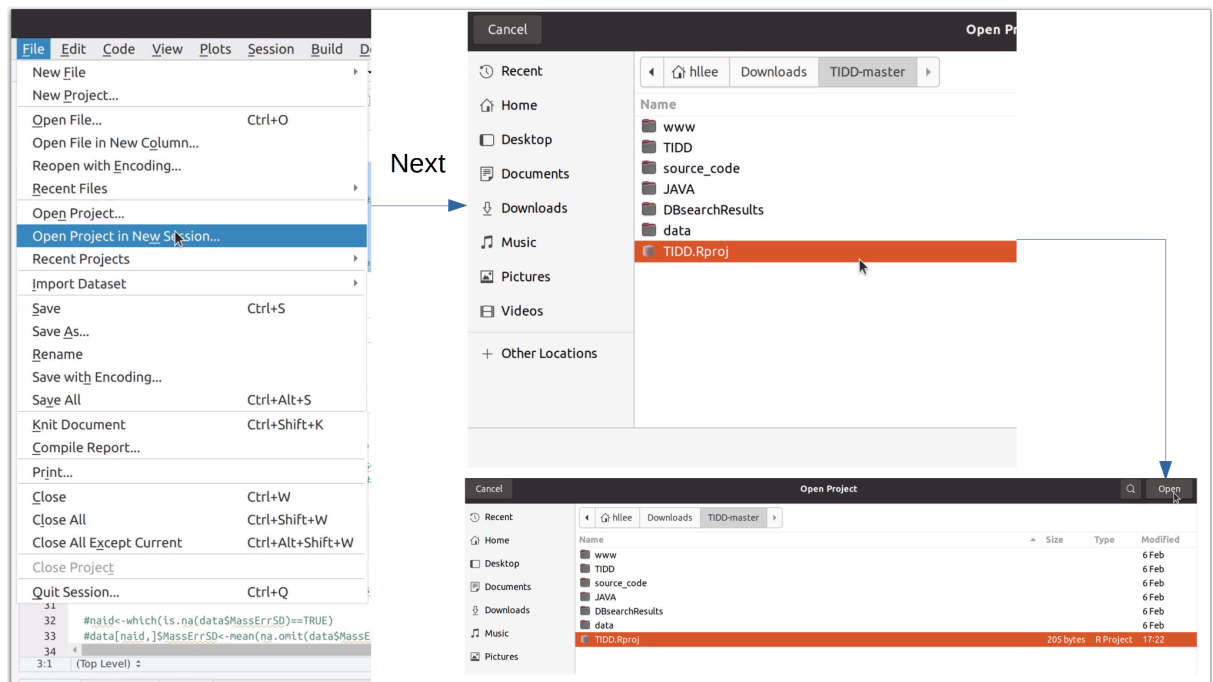


## 6. Open the R studio

Import TIDD project to R studio

File --> Open project in New Session --> select TIDD.rproj

for example: "~/Download/TIDD-master/TIDD/TIDD.rproj



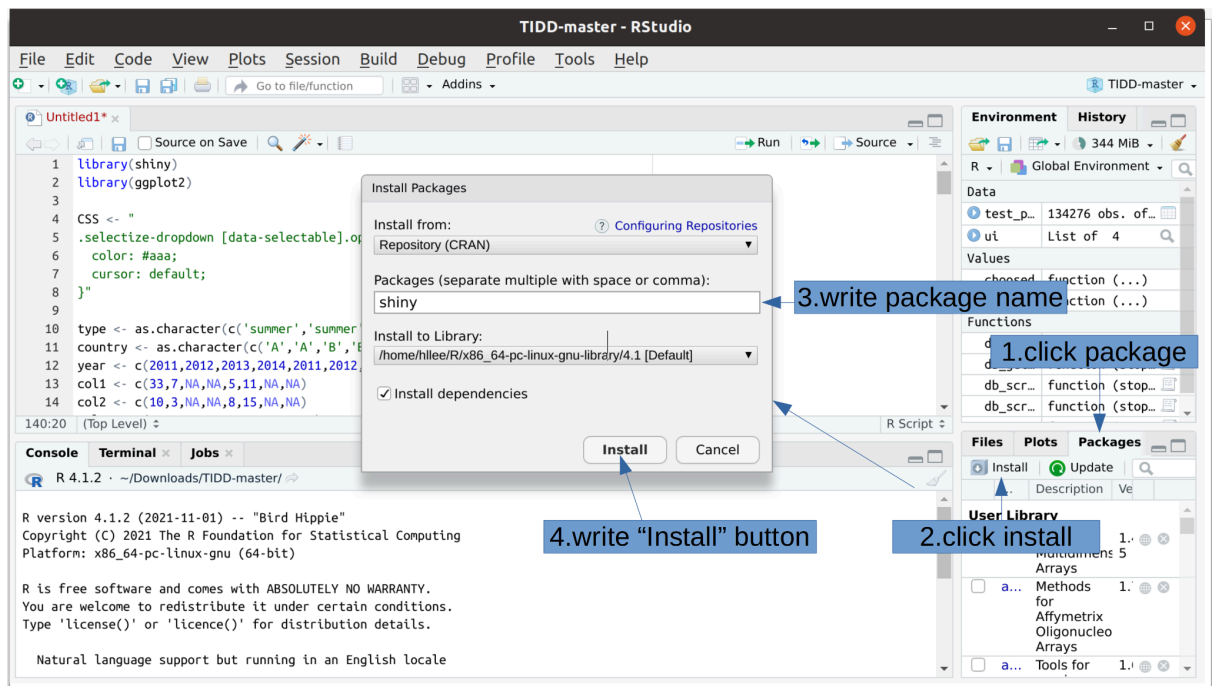
## 7. Install R packages

required packages: shiny, shinythemes, shinydashboard, shinyFiles, e1071, ROCR

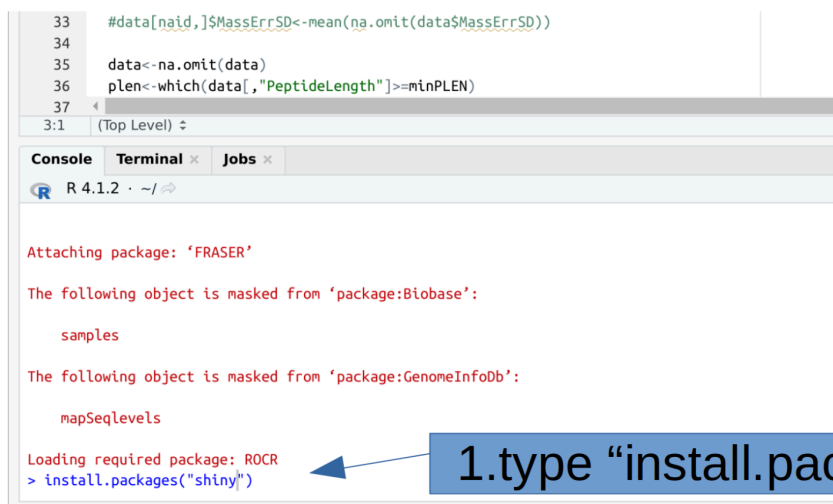
E.g. install shiny package

UI: click Packages --> install --> type "shiny" --> install

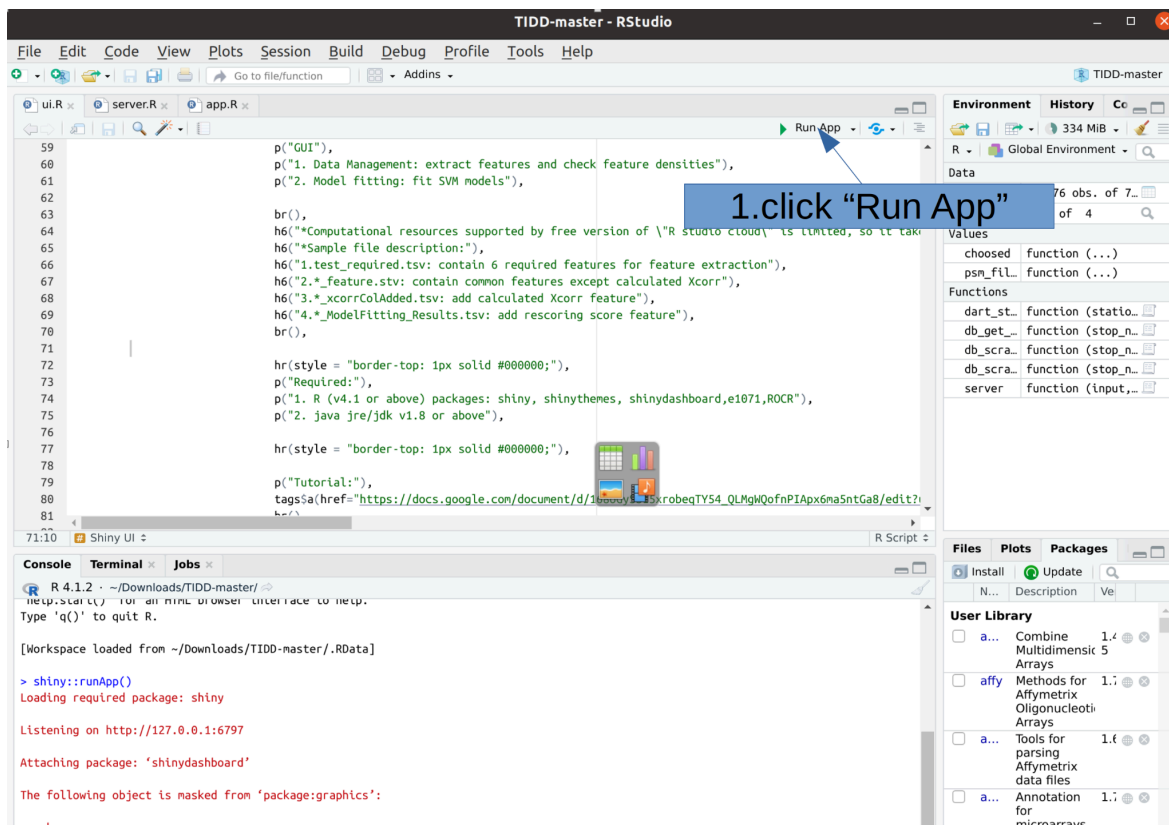




console: install.packages("shiny")



8. Run



## File format

### - TIDD input file format: PSM search result [\*tsv]

This file requires 6 features:

1. Filename
2. Scan
3. Charge
4. precursor mass
5. Peptide
6. Protein

To calculate NTT value, pre and next amino acids must be included in the protein information or peptide information as below:

Example1:

| Filename | Scan  | Charge | PrecursorM  | Peptide                      | Protein                                                                          |
|----------|-------|--------|-------------|------------------------------|----------------------------------------------------------------------------------|
| test.mgf | 19608 | 3      | 1449.831259 | IETLDPALIRPGR                | sp P62191 PRS4_HUMAN(pre=R,next=I);sp P62191-2 PRS4_HUMAN                        |
| test.mgf | 37707 | 2      | 1685.96616  | LPDGSEIPLPILLGR              | sp Q96KP4 CNDP2_HUMAN(pre=K,next=L)                                              |
| test.mgf | 32201 | 2      | 2001.002903 | IPEISIQDMTAQVTSPSGK          | sp P21333 FLNA_HUMAN(pre=K,next=T);sp P21333-2 FLNA_HUMAN                        |
| test.mgf | 11229 | 2      | 965.506199  | EVDIYTVK                     | sp Q9NR22 ANM8_HUMAN(pre=K,next=T);sp Q9NR22-2 ANM8_HUMAN;sp Q9NR22-3 ANM8_HUMAN |
| test.mgf | 36713 | 2      | 2752.472752 | VPAATSAIITNDGIGINPAQTAGNVFLK | sp P61960 UFM1_HUMAN(pre=K,next=H);sp P61960-2 UFM1_HUMAN                        |
| test.mgf | 26973 | 2      | 1268.721165 | SGGLQPVPVIFR                 | sp P11177 ODPB_HUMAN(pre=M,next=G);sp P11177-2 ODPB_HUMAN                        |
| test.mgf | 563   | 3      | 927.434836  | AHGNSGMVR                    | sp P18077 RL35A_HUMAN(pre=R,next=A)                                              |
| test.mgf | 9816  | 3      | 1293.700064 | NVNQSLLELHK                  | sp P02794 FRIH_HUMAN(pre=K,next=L)                                               |

### Example2:

| 1 SpecId | ScanNr                                                | ExpMass | Peptide                                   | Proteins | Charge |             |                                        |                          |
|----------|-------------------------------------------------------|---------|-------------------------------------------|----------|--------|-------------|----------------------------------------|--------------------------|
| 2        | 20150911_QE7_UPLC11_DBJ_SA_HELA_3MG_50FRAC_Ti02_1.mgf |         | GLT16_HUMAN;sp Q8N428-2 GLT16_HUMAN       | 3        | 578    | 1166.504765 | K.KM+15.9949NC+57.021464KS+79.9663FR.W | sp Q8N428 -              |
| 3        | 20150911_QE7_UPLC11_DBJ_SA_HELA_3MG_50FRAC_Ti02_1.mgf |         | 229552;CONTAMN_gi 113574                  | 2        | 579    | 1166.498119 | R.C+57.021464C+57.021464TKPESER.M      | CONTAMN_gi -             |
| 4        | 20150911_QE7_UPLC11_DBJ_SA_HELA_3MG_50FRAC_Ti02_1.mgf |         | RUVB2_HUMAN;DECOY_sp Q9Y230-2 RUVB2_HUMAN | 3        | 580    | 980.16234   | K.T+79.9663DMT+79.9663EGS+79.9663.-    | DECOY_sp Q9Y230 -        |
| 5        | 20150911_QE7_UPLC11_DBJ_SA_HELA_3MG_50FRAC_Ti02_1.mgf |         |                                           |          | 581    | 765.288091  | R.YGGSS+79.9663GR.P                    | tr K7ESH3 K7ESH3_HUMAN 2 |

## - TIDD input file format: MS/MS spectra [\*mgf]

MS/MS spectra [\*mgf] must contain title information.

The title format: filename+”. ”+start scan+”. ”+charge

Example:

```

BEGIN IONS
TITLE=test.2.2.2
RTINSECONDS=0.7896
PEPMASS=344.720764160156 64676.6015625
SCANS=2
CHARGE=2+
100.4279022 65.6441345215
101.0709763 204.4091339111
101.1071472 238.9379425049
102.0914917 183.1679229736
110.0713196 604.6396484375
111.0912781 213.9605712891
112.0756607 279.3070983887
113.1076431 147.8057403564
115.0876236 88.8695907593
120.0803528 154.9431304932
124.2364426 76.4476547241

```

Filename: test.mgf  
Scan: 2  
Charge: 2+  
  
TITLE=test.2.2.2

## - Test data: download link

Link: <https://github.com/HanyangBISLab/TIDD/tree/master/data>

PSM result file: ~/TIDD/data/PSM/

- Feature extraction input file: test\_required.tsv
- Model fitting input file: test\_required\_xcorrColAdded.tsv
- Link: <https://github.com/HanyangBISLab/TIDD/tree/master/data/PSM>

MGF file: ~/TIDD/data/test/test.mgf

- Link: <https://github.com/HanyangBISLab/TIDD/tree/master/data/MGF/test>

## Run: GUI

TIDD-GUI has 4 menus:

1. “About”: describes a brief introduction of TIDD, requirements for running TIDD, and a download link.
2. “File upload & download”: upload/download files to/from TIDD-GUI.
3. “Data management”: extract PSM features
4. “Model fitting”: fit model to rescore PSMs

## - File upload & download

Upload files:

1. Select file types <PSM file or MS/MS spectra file> for uploading
2. Click “Browse...” and select files for uploading
3. Done
  - a. PSM file: located unnder the “data/PSM/” folder
  - b. MGF file: located under the “data/MGF/Local/”

Download files:

1. Select one file for downloading
  - a. PSM file: the files under the ‘data/PSM/’ folder
  - b. MGF file: the files under the ‘data/PSM/test/’ folder
2. Click download button

TIDD v1.1   About   **File upload & download**   Data Management   Model Fitting

---

Upload your own files to web

\*Tab-delimited PSM file must contains:  
1.file name col.: must same as the corresponding mgf file name  
2.scan number col.: scan number in the \*.mgf file  
3.charge

...

TITLE=test.5169.5169.2

Select one file type to upload (Max size 2G per one time)

☐ PSM file (tab-delimited file)

☒ MS/MS spectra file (MGF)

Upload

Browse...   No file selected

Uploaded PSM or MGF files will be stored in "PSM or Local" directory respectively.

Download files from web

Select files to download:

PSM file:  
test\_required\_Feature\_xcorrColAdded.tsv

Download \*.tsv

MGF file:  
test.mgf

Download \*.mgf

Upload

Download

## - Data management

Here, a set of PSM features can be extracted.

Video tutorial: [link](#)

Feature extraction:

1. Select files and click the

TIDD v1.1   About   File upload & download   **Data Management**   Model Fitting

---

File Loading

Peptide-spectrum matching key info:  
1. filename  
2. scan number  
3. charge

Select one PSM file (tab-delimited):  
test\_required.tsv

Select one directory contained spectrum files (\*.mgf):  
test

First Load Files

PSMs   Preview (table)

Step1.  
Select PSM file  
E.g. test\_required.tsv

Step2.  
Select directory contained mgf files  
E.g. test

6 required features for PSM feature extraction:

Write the corresponding column indices in the same order as below:  
(separator between indices: ";")

1. filename  
2. scan number  
3. charge  
4. precursor mass  
5. peptide  
6. protein (separator ";")

required column indices:  
1,2,3,4,5,6

Decoy prefix:  
XXX\_

Fragment tol.(Da):  
0.025

Second Extract!

Extraction log   Summary&Plot

2. Fill in the 6 required information column indexes

TIDD v1.1 About File upload & download Data Management Model Fitting

File Loading

Peptide-spectrum matching key info:

1. filename
2. scan number
3. charge

Select one PSM file (tab-delimited):

test\_required.tsv

Select one directory contained spectrum files (\*.mgf):

test

First. Load Files

PSM Preview (table)

test\_required.tsv

Features in the psm result file.

Column info (index, feature name):

|                                              |                                     |                                  |
|----------------------------------------------|-------------------------------------|----------------------------------|
| <input checked="" type="radio"/> 1. Filename | <input type="radio"/> 3. Charge     | <input type="radio"/> 5. Peptide |
| <input type="radio"/> 2. Scan                | <input type="radio"/> 4. PrecursorM | <input type="radio"/> 6. Protein |

Show column index and column name  
E.g. test\_required.tsv  
Column index 1: column name "Filename"  
Column index 2: column name "Scan"  
Column index 3: column name "Charge"  
Column index 4: column name "PrecursorM"  
Column index 5: column name "Peptide"  
Column index 6: column name "Protein"

6 required features for PSM feature extraction:

Write the corresponding column indices in the same order as below:  
(separator between indices: ",")

1. filename
2. scan number
3. charge
4. precursor mass
5. peptide
6. protein (separator ";")

required column indices:

1,2,3,4,5,6

Decoy prefix:

XXX\_

Fragment tol.(Da):

0.025

Second.Extract!

Fill in the column index

3. Click "Second Extract" button
- Show the state of running

File Loading

Peptide-spectrum matching key info:

1. filename
2. scan number
3. charge

Select one PSM file (tab-delimited):

test\_required.tsv

Select one directory contained spectrum files (\*.mgf):

test

First. Load Files

PSM Preview (table)

test\_required.tsv

Features in the psm result file.

Column info (index, feature name):

|                                              |                                     |                                  |
|----------------------------------------------|-------------------------------------|----------------------------------|
| <input checked="" type="radio"/> 1. Filename | <input type="radio"/> 3. Charge     | <input type="radio"/> 5. Peptide |
| <input type="radio"/> 2. Scan                | <input type="radio"/> 4. PrecursorM | <input type="radio"/> 6. Protein |

6 required features for PSM feature extraction:

Write the corresponding column indices in the same order as below:  
(separator between indices: ",")

1. filename
2. scan number
3. charge
4. precursor mass
5. peptide
6. protein (separator ";")

required column indices:

1,2,3,4,5,6

Decoy prefix:

XXX\_

Fragment tol.(Da):

0.025

Second.Extract!

Extracting program is running  
Do not click the any buttons until it finished

Extraction log Summary&Plot

Second.Extract!

Check the "Extraction log" tab whether it is finished or not

Extracting features now, it takes a few minites...

4. Done

test\_required.tsv

Select one directory contained spectrum files(\*.mgf):

test

First Load Files

Finished!

Extraction log Summary&Plot

Extracing features: finished [result file: "feature\_\*.tsv and \*\_Xcorr.tsv

DONE!

1.\_feature.tsv: contain annotated peaks - related features

2.\_xcorrColAdded.tsv: add calculated xcorr features

5. peptide

6. protein (separator ";")

required column indices:

1,2,3,4,5,6

Decoy prefix:

XXX\_

Fragment tol.(Da):

0.025

Second.Extract!

Check the "Extraction log" tab whether it is finished or not

## - Model Fitting

Video tutorial: [link](#)

Steps:

1. Select one PSM file → "LOAD <GO!>"

TIDD v1.1 About File upload & download Data Management Model Fitting

PSMs for model fitting

Tab-delimited file:

1. head: do not start with '#' symbol

2. one column for one feature

Select one PSM file:

test\_required\_Feature\_xcorrColAdded.tsv

LOAD <GO!>

Feature selection

Please select x, y and init score correctly.

x:

1. Select one PSM file

Default:

test\_required\_Feature\_xcorrColAdded.tsv

y (tar/dec): (Select TorD feature)

model:

Iterative SVM

init score: (Select xcorr or calXcorr feature)

k-CV:

3

Max train size:

1000

Max iteration (SVM):

2. Select features for modeling

TIDD v1.1   About   File upload & download   Data Management   **Model Fitting**

PSMs for model fitting

Tab-delimited file:

1. head: do not start with '#' symbol
2. one column for one feature

Select one PSM file :

test\_required\_Feature\_xcorrColAdded.tsv

LOAD <GO> !

Feature selection

Input file: test\_required\_Feature\_xcorrColAdded.tsv

Please select x, y and init score correctly.

x:

|                                                |                                                   |                                                     |
|------------------------------------------------|---------------------------------------------------|-----------------------------------------------------|
| <input type="checkbox"/> Filename              | <input checked="" type="checkbox"/> AbsDeltaMass  | <input checked="" type="checkbox"/> SeqCoverBion    |
| <input type="checkbox"/> Scan                  | <input checked="" type="checkbox"/> PeptideLength | <input checked="" type="checkbox"/> ConsecutiveYion |
| <input type="checkbox"/> Charge                | <input checked="" type="checkbox"/> TIC           | <input checked="" type="checkbox"/> ConsecutiveBion |
| <input checked="" type="checkbox"/> PrecursorM | <input checked="" type="checkbox"/> MaxIntALL     | <input checked="" type="checkbox"/> MassErrMean     |
| <input type="checkbox"/> Peptide               | <input checked="" type="checkbox"/> MaxYionInt    | <input checked="" type="checkbox"/> MassErrSD       |
| <input type="checkbox"/> Protein               | <input checked="" type="checkbox"/> MaxBionInt    | <input checked="" type="checkbox"/> NumofAnnoPeaks  |
| <input type="checkbox"/> Title                 | <input checked="" type="checkbox"/> SumYmatchInt  | <input checked="" type="checkbox"/> MissedCleavage  |
| <input type="checkbox"/> PrecursorM.1          | <input checked="" type="checkbox"/> SumBmatchInt  | <input checked="" type="checkbox"/> Triptic         |
| <input type="checkbox"/> PrecursorMZ           | <input checked="" type="checkbox"/> FracYmatchInt | <input checked="" type="checkbox"/> TorD            |
| <input type="checkbox"/> CalMW                 | <input checked="" type="checkbox"/> FracBmatchInt | <input checked="" type="checkbox"/> CalXcorr        |
| <input checked="" type="checkbox"/> DeltaMass  | <input checked="" type="checkbox"/> SeqCoverYion  |                                                     |

y (tar/dec): (Select TorD feature)

Charge

model:

Iterative SVM

init score: (Select xcorr or calXcorr feature)

Charge

k-CV:

3

Max train size:

1000

Max iteration (SVM):

1. Select features for modeling  
Default: TIDD features

3. Select parameters and click "Fitting <GO!>" button

2. one column for one feature

Select one PSM file :

test\_required\_Feature\_xcorrColAdded.tsv

LOAD <GO> !

Select target/decoy column  
Default: TorD  
<value 1 for target, 0 for decoy>

Select init score used for the first iteration  
Default: CalXcorr

K-cross validation

The number of training size

The number of iteration in iterative-SVM model

y (tar/dec): (Select TorD feature)

TorD

model:

Iterative SVM

init score: (Select xcorr or calXcorr feature)

CalXcorr

k-CV:

3

Max train size:

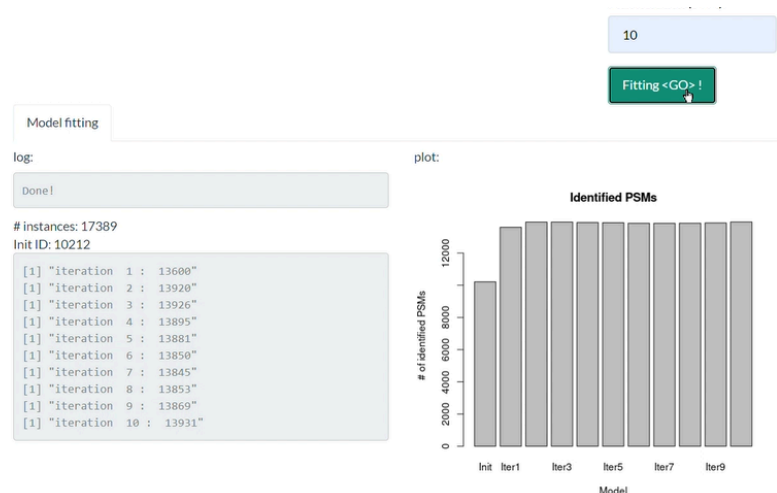
1000

Max iteration (SVM):

10

Fitting <GO> !

4. Show the identification results per iteration





## Run: command line

It contains 3 steps:

1. Extract PSM features
2. Calculated XCorr scores for PSMs
3. Fitting models and re-ranking PSMs

### - PSM feature extraction

Two types of codes are uploaded in the github (<https://github.com/HanyangBISLab/TIDD.git>)

The first one is a runnable "TIDD.jar" file, which located in the "TIDD/JAVA", and the other is "FeatureExtraction" source codes located in the "TIDD/source\_code"

Run TIDD.jar

-Help: "java -jar TIDD.jar -h"

-Run: java -jar <options> <attributes>

-Options:

-i <filename> : PSM search result file [Required]

: tab-delimited tsv file

-dir <filename> : MS/MS data file used for search [Required]

: Available formats [mgf]

-o <output\_filename>: Output file path for extracted features [Optional]

: If not specified, it's automatically generated

-id <list of index with separator "," >: list of column index in the tab-delimited PSM result file [required]

: SpectrumFile column index, ScanNum column index, charge column index, precursorMZ column index, peptide column index, Protein column index

-dec <String>: decoy prefix [required]

-ft <double>: fragment tolerance (Unit:Da) [required]

-phospho <boolean>: true for phospho data [optional]

: default values is false

- Example: java -jar TIDD.jar -i psm\_test.tsv -d C:/Test/ -id 1,2,3,4,5,11 -dec XXX\_ -ft 0.025

## - Calculate XCorr score

Two types of codes are uploaded in the github (<https://github.com/HanyangBISLab/TIDD.git>)

The first one is a runnable "Xcorr\_v2.jar" file, which located in the "TIDD/JAVA", and the other is "Xcorr" source codes located in the "TIDD/source\_code"

Run Xcorr\_v2.jar

-Help: "java -jar Xcorr\_v2.jar -h"

-Run: "java -jar xcorr.jar spectrumDir resultFile binWidth binOffset titleCol peptideCol precursorMZCol chargeCol mgffilenameCol header"

- "spectrumDir" should be a mgf directory

- "resultFile" should be a tsv file

- column number is a zero-based number

- binWidth is default is 1.0005079 for low resolution. 0.02 is recommended for high resolution

- binOffset default is 0.4 for low resolution. 0.0 is fine for high resolution

- header value can be either true (result file has a header) or false (no header in the result file)

- Example: java -jar Xcorr\_v2.jar /Test/ test.tsv 0.02 0.0 5,6,2,3,1 1

## - Fitting models

For model fitting, we uploaded one "TIDD\_ModelFit.R" file under the "TIDD/source\_code" (<https://github.com/HanyangBISLab/TIDD.git>)

Run TIDD\_ModelFit.R :

- Change the parameters inside the R code first.

<Change line 32 to 40>

TIDD\_ModelFit.R

```
32 inputfile<-"/home/hllee/Documents/TIDD/modif/phosph/AML/AML_phospho_comet/AML_phospho_format_Feature.tsv"
33 outputfile<-"/home/hllee/Documents/TIDD/modif/phosph/AML/AML_phospho_comet/AML_phospho_format_Feature_MF_impute.tsv"
34 MaxNum_iteration<-10
35 minPLEN<-6
36 trainsize<-10000
37 k_cross<-5
38 out_ycol<-"TorD"
39 init_score<-"Xcorr"
40 features=c("TorD", "Xcorr", "AbsDeltaMass", "Charge", "PeptideLength", "Triptic", "MissedCleavage", "PrecursorM", "DeltaMass", "CalMW", "TIC", "SumYmatchInt"
41 #features=c("TorD", "CalXcorr", "NumofAnnoPeaks", "ConsecutiveYion", "SeqCoverYion")
```

- Run "Rscript TIDD\_ModelFit.R"

