Guideline on running Citrullination_Diagnostic_Ion_Analysis

Requirements

Python requirements

- Python version 3.6 or greater
- Dependencies: pandas, numpy, pyteomics, itertools, collections, statistics, re, io

File requirements

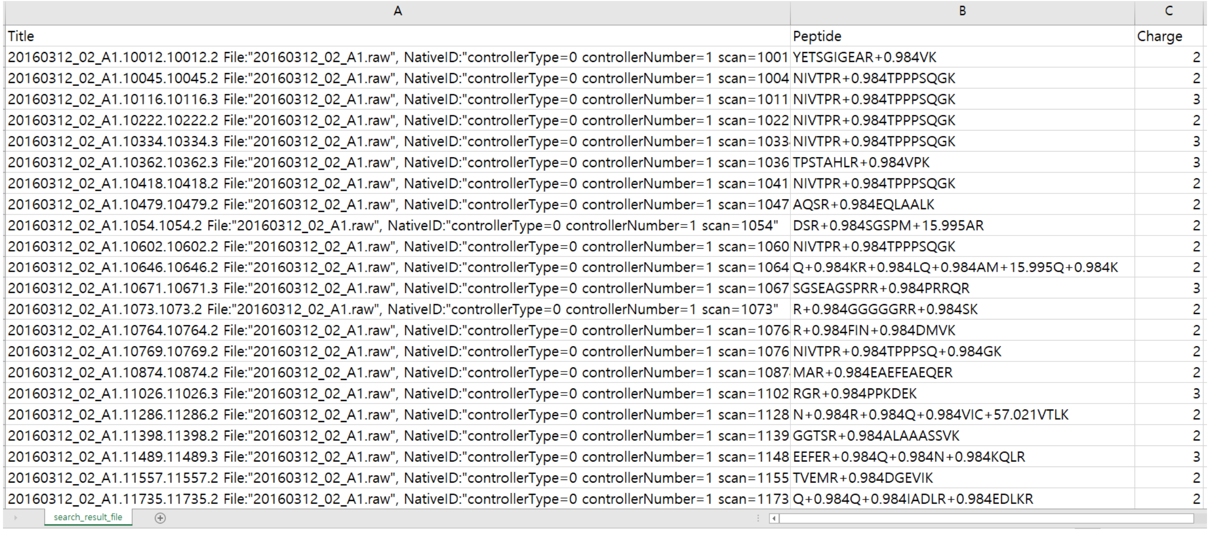
- Input search result file: The input file must be a .csv file(s) containing the following three columns (see next page for the details):

‘Title’	MS2 spectrum title as written in MGF file
‘Peptide’	Peptide sequence with modification delta mass rounded up to third decimal places
‘Charge’	Charge state of the peptide

- Spectrum file: The spectrum file must be a Mascot Generic Format (MGF) file(s) containing MS2 spectra corresponding to those matched to the PSMs in the input search result file. If MS2 spectra in the input search file and spectrum file are not equivalent, only the common MS2 spectra will be retained and subsequently processed.

Input search result file

Example input search result file:



A	B	C
Title	Peptide	Charge
20160312_02_A1.10012.10012.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1001	YETSGIGEAR+0.984VK	2
20160312_02_A1.10045.10045.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1004	NIVTPR+0.984TPPPSQGK	2
20160312_02_A1.10116.10116.3 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1011	NIVTPR+0.984TPPPSQGK	3
20160312_02_A1.10222.10222.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1022	NIVTPR+0.984TPPPSQGK	2
20160312_02_A1.10334.10334.3 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1033	NIVTPR+0.984TPPPSQGK	3
20160312_02_A1.10362.10362.3 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1036	TPSTAHLR+0.984VPK	3
20160312_02_A1.10418.10418.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1041	NIVTPR+0.984TPPPSQGK	2
20160312_02_A1.10479.10479.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1047	AQSR+0.984EQLAALK	2
20160312_02_A1.1054.1054.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1054"	DSR+0.984SGSPM+15.995AR	2
20160312_02_A1.10602.10602.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1060	NIVTPR+0.984TPPPSQGK	2
20160312_02_A1.10646.10646.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1064	Q+0.984KR+0.984LQ+0.984AM+15.995Q+0.984K	2
20160312_02_A1.10671.10671.3 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1067	SGSEAGSPRR+0.984PRRQR	3
20160312_02_A1.1073.1073.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1073"	R+0.984GGGGGRR+0.984SK	2
20160312_02_A1.10764.10764.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1076	R+0.984FIN+0.984DMVK	2
20160312_02_A1.10769.10769.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1076	NIVTPR+0.984TPPPSQ+0.984GK	2
20160312_02_A1.10874.10874.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1087	MAR+0.984EAEFEAEQER	2
20160312_02_A1.11026.11026.3 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1102	RGR+0.984PPKDEK	3
20160312_02_A1.11286.11286.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1128	N+0.984R+0.984Q+0.984VIC+57.021VTLK	2
20160312_02_A1.11398.11398.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1139	GGTSR+0.984ALAAASSVK	2
20160312_02_A1.11489.11489.3 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1148	EEFER+0.984Q+0.984N+0.984KQLR	3
20160312_02_A1.11557.11557.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1155	TVEMR+0.984DGEVIK	2
20160312_02_A1.11735.11735.2 File:"20160312_02_A1.raw", NativeID:"controllerType=0 controllerNumber=1 scan=1173	Q+0.984Q+0.984IADLR+0.984EDLKR	2

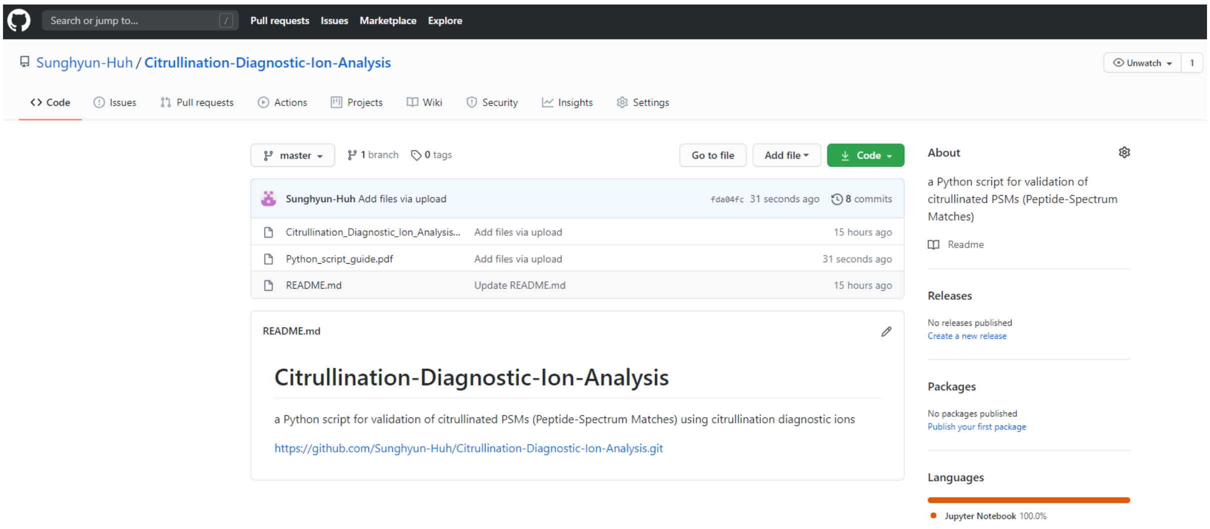
Format for peptide sequences should follow that of MS-GF+ search result. Specifically, modification delta masses should be rounded up to third decimal places. Currently allowed modifications are as follows:

Modification	**Mod on peptide**
Carbamidomethyl Cys	C+57.021
Oxidation Met	M+15.995
Deamidated Asn	N+0.984
Deamidated Gln	Q+0.984
Citrullinated Arg	R+0.984
Pyro-Glu from Glu	E-17.027
Pyro-Glu from Gln	Q-18.011
iTRAQ 4plex Lys	K+144.102
iTRAQ 8plex Lys	K+304.205
TMT Lys	K+229.163
iTRAQ 4plex N-term	+144.102
iTRAQ 8plex N-term	+304.205
TMT N-term	+229.163
Acetyl N-term	+42.011

Downloading the Python script

The Python script can be downloaded via following GitHub page:

https://github.com/Sunghyun-Huh/Citrullination-Diagnostic-Ion-Analysis



Sunghyun-Huh / Citrullination-Diagnostic-Ion-Analysis

Sunghyun-Huh Add files via upload fda04fc 31 seconds ago 8 commits

Citrullination_Diagnostic_Ion_Analysis...	Add files via upload	15 hours ago
Python_script_guide.pdf	Add files via upload	31 seconds ago
README.md	Update README.md	15 hours ago

README.md

Citrullination-Diagnostic-Ion-Analysis

a Python script for validation of citrullinated PSMs (Peptide-Spectrum Matches) using citrullination diagnostic ions

https://github.com/Sunghyun-Huh/Citrullination-Diagnostic-Ion-Analysis.git

About

a Python script for validation of citrullinated PSMs (Peptide-Spectrum Matches)

Readme

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

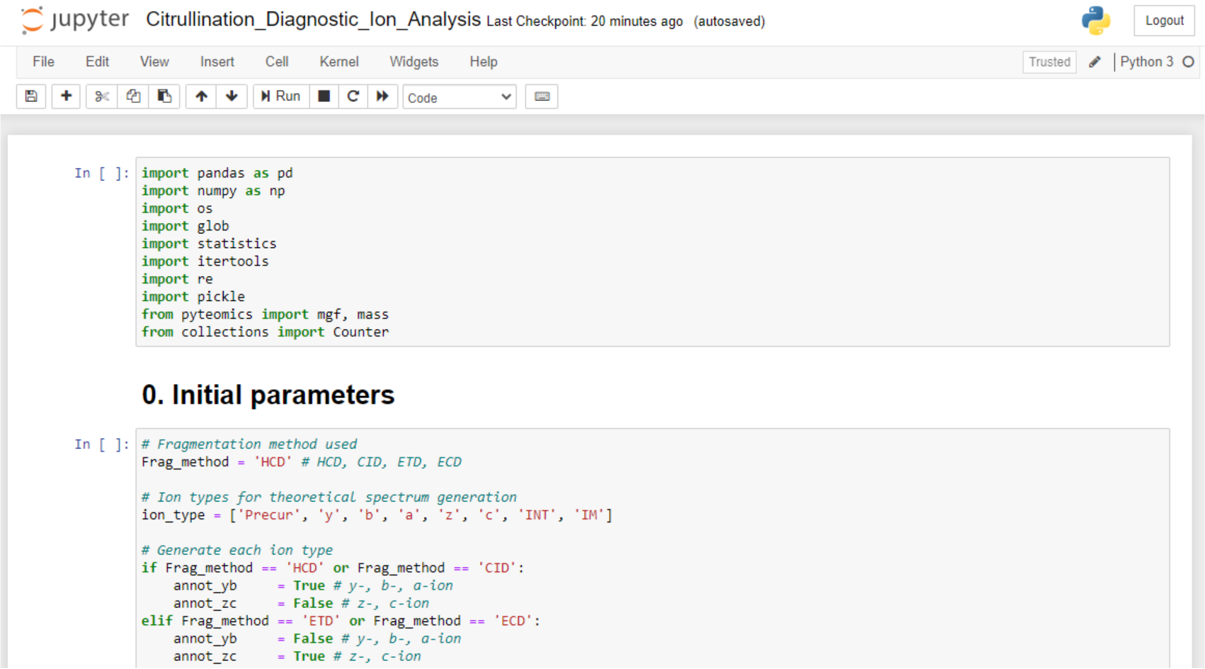
● Jupyter Notebook 100.0%

Users can download the Jupyter notebook via the following steps:

- Click on the ‘Citrullination_Diagnostic_Ion_Analysis.ipynb’.
- Click on the ‘Raw’.
- Press ctrl+s and manually type ‘ipynb’ after the filename to download as a .ipynb file.

Running the Python script

A snapshot of the Jupyter notebook:



```
import pandas as pd
import numpy as np
import os
import glob
import statistics
import itertools
import re
import pickle
from pyteomics import mgf, mass
from collections import Counter
```

0. Initial parameters

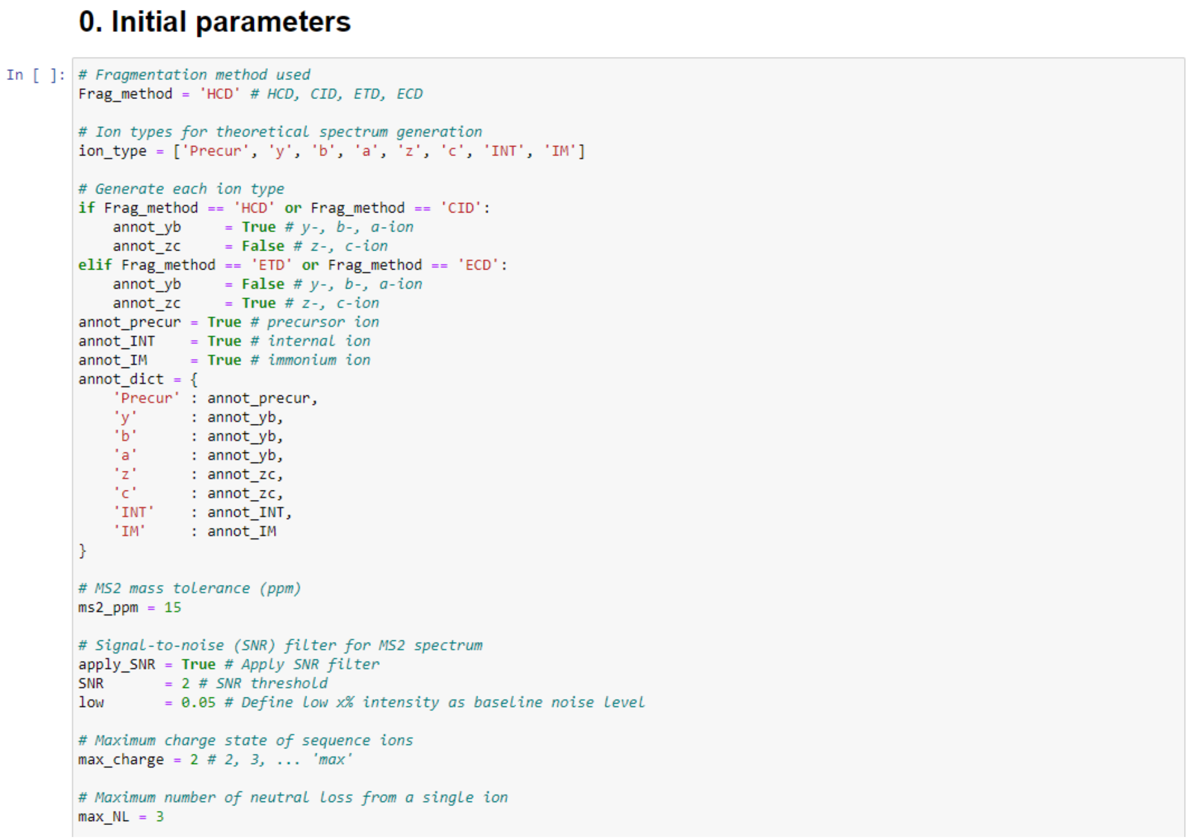
```
# Fragmentation method used
Frag_method = 'HCD' # HCD, CID, ETD, ECD

# Ion types for theoretical spectrum generation
ion_type = ['Precur', 'y', 'b', 'a', 'z', 'c', 'INT', 'IM']

# Generate each ion type
if Frag_method == 'HCD' or Frag_method == 'CID':
    annot_yb     = True # y-, b-, a-ion
    annot_zc     = False # z-, c-ion
elif Frag_method == 'ETD' or Frag_method == 'ECD':
    annot_yb     = False # y-, b-, a-ion
    annot_zc     = True # z-, c-ion
```

Running the Python script

A snapshot of initial parameters settings:



0. Initial parameters

In [ ]:

```
# Fragmentation method used
Frag_method = 'HCD' # HCD, CID, ETD, ECD

# Ion types for theoretical spectrum generation
ion_type = ['Precur', 'y', 'b', 'a', 'z', 'c', 'INT', 'IM']

# Generate each ion type
if Frag_method == 'HCD' or Frag_method == 'CID':
    annot_yb      = True # y-, b-, a-ion
    annot_zc      = False # z-, c-ion
elif Frag_method == 'ETD' or Frag_method == 'ECD':
    annot_yb      = False # y-, b-, a-ion
    annot_zc      = True # z-, c-ion
annot_precur = True # precursor ion
annot_INT    = True # internal ion
annot_IM     = True # immonium ion
annot_dict = {
    'Precur' : annot_precur,
    'y'      : annot_yb,
    'b'      : annot_yb,
    'a'      : annot_yb,
    'z'      : annot_zc,
    'c'      : annot_zc,
    'INT'    : annot_INT,
    'IM'     : annot_IM
}

# MS2 mass tolerance (ppm)
ms2_ppm = 15

# Signal-to-noise (SNR) filter for MS2 spectrum
apply_SNR = True # Apply SNR filter
SNR       = 2 # SNR threshold
low       = 0.05 # Define low x% intensity as baseline noise level

# Maximum charge state of sequence ions
max_charge = 2 # 2, 3, ... 'max'

# Maximum number of neutral loss from a single ion
max_NL = 3
```

Explanations of initial parameters are as follows:

Frag_method Fragmentation method used in the input data (value = ‘HCD’, ‘CID’, ‘ETD’, or ‘ECD’). If set as ‘HCD’ or ‘CID’, *y*-ion, *b*-ion, and *a*-ion will be generated for theoretical spectrum. If set as ‘ETD’ or ‘ECD’, *z*-ion, *c*-ion will be generated for theoretical spectrum. Commonly, precursor, internal, and immonium ion will be generated for all fragmentation method used.

ms2_ppm MS2 level mass tolerance in ppm (default = 15 ppm)

apply_SNR Determine whether to apply signal-to-noise filter to remove noise peaks (value = True or False; default = True)

SNR Signal threshold level. The average intensity of noise peaks (as defined in ‘low’) multiplied by this signal threshold level will be the final signal-to-noise filter. If ‘apply_SNR’ = True, all peaks below the signal-to-noise filter will be removed (default = 2)

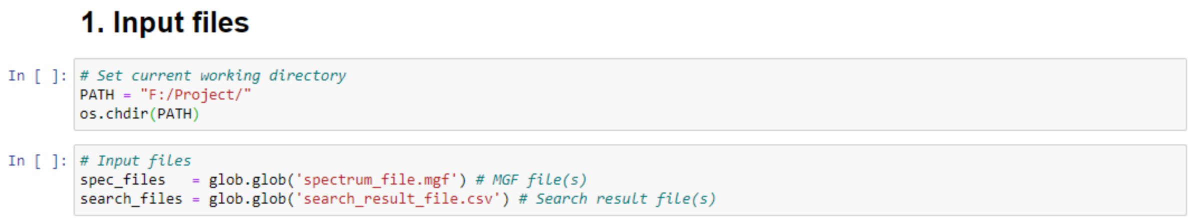
low Proportion of MS2 peaks regarded as noise. If ‘apply_SNR’ = True, all peaks below this noise level will be treated as noise (default = 0.05)

max_charge Maximum charge state of fragment ions (default = 2)

max_NL Maximum number of neutral losses from a single ion (default = 3)

Running the Python script

A snapshot of codes for loading input files:



1. Input files

```
# Set current working directory
PATH = "F:/Project/"
os.chdir(PATH)
```

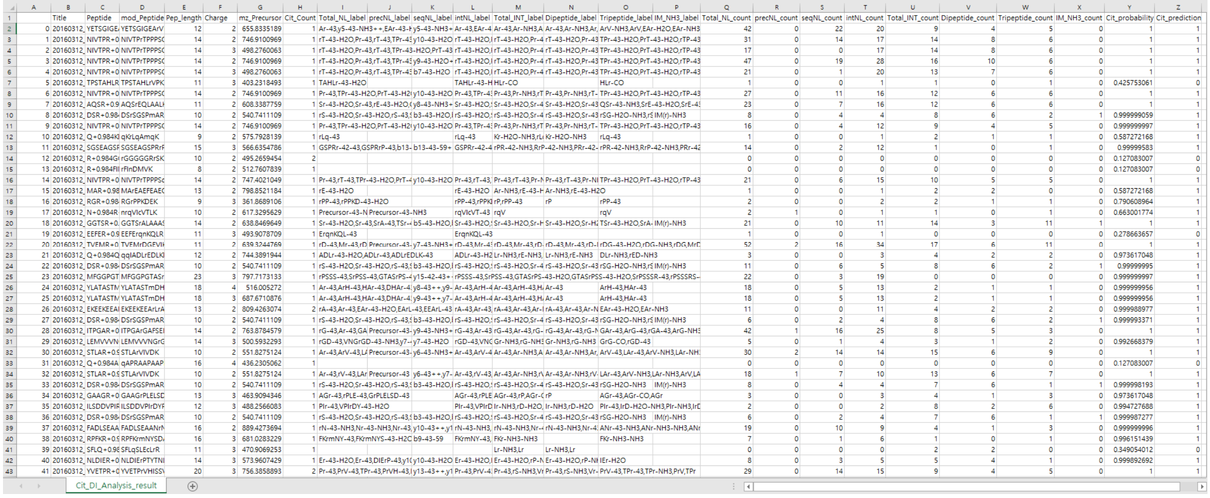
```
# Input files
spec_files   = glob.glob('spectrum_file.mgf') # MGF file(s)
search_files = glob.glob('search_result_file.csv') # Search result file(s)
```

Users can upload local input files via the following steps:

- Set the directory in which the input files are located.
- Copy and paste the input filenames. In case of multiple MGF or search result files in the same directory, type in ‘*.mgf’ or ‘*.csv’.

Output result file

Example output file:



	A	B	C	D	E	F	G	H	I	J	K
1		Title	Peptide	mod_Peptide	Pep_length	Charge	mz_Precursor	Cit_Count	Total_NL_label	precNL_label	seqNL_label
2	0	20160312_	YETSGIGE/	YETSGIGEArV	12	2	655.8335189	1	Ar-43,y5-43-NH3++,EAr-43-H		y5-43-NH3+
3	1	20160312_	NIVTPR+0	NIVTPrTPPPS(	14	2	746.9100969	1	rT-43-H2O,Pr-43,rT-43,TPr-4		y10-43-H2O
4	2	20160312_	NIVTPR+0	NIVTPrTPPPS(	14	3	498.2760063	1	rT-43-H2O,Pr-43,rT-43,TPr-43-H2O,PrT-43		
5	3	20160312_	NIVTPR+0	NIVTPrTPPPS(	14	2	746.9100969	1	rT-43-H2O,Pr-43,rT-43,TPr-4		y9-43-H2O+
6	4	20160312_	NIVTPR+0	NIVTPrTPPPS(	14	3	498.2760063	1	rT-43-H2O,Pr-43,rT-43,TPr-4		b7-43-H2O
7	5	20160312_	TPSTAHLR	TPSTAHLrVPK	11	3	403.2318493	1	TAHLr-43-H2O		
8	6	20160312_	NIVTPR+0	NIVTPrTPPPS(	14	2	746.9100969	1	Pr-43,TPr-43-H2O,PrT-43-H2(		y10-43-H2O
9	7	20160312_	AQSR+0.9	AQSrEQLAALI	11	2	608.3387759	1	Sr-43-H2O,Sr-43,rE-43-H2O,(		y8-43-NH3+
10	8	20160312_	DSR+0.98	DSrSGSPmAR	10	2	540.7411109	1	rS-43-H2O,Sr-43-H2O,rS-43,S		b3-43-H2O,I
11	9	20160312_	NIVTPR+0	NIVTPrTPPPS(	14	2	746.9100969	1	Pr-43,TPr-43-H2O,PrT-43-H2(		y10-43-H2O
12	10	20160312_	Q+0.984KI	qKrLqAmqK	9	2	575.7928139	1	rLq-43		
13	11	20160312_	SGSEAGSF	SGSEAGSPRrF	15	3	566.6354786	1	GSPRr-42-43,GSPRrP-43,b13-		b13-43-59+
14	12	20160312_	R+0.984G	rGGGGGRrSK	10	2	495.2659454	2			
15	13	20160312_	R+0.984FI	rFInDMVK	8	2	512.7607839	1			
16	14	20160312_	NIVTPR+0	NIVTPrTPPPSc	14	2	747.4021049	1	Pr-43,rT-43,TPr-43-H2O,PrT-4		y10-43-H2O
17	15	20160312_	MAR+0.98	MArEAEFEAE(	13	2	798.8521184	1	rE-43-H2O		
18	16	20160312_	RGR+0.98	RGrPPKDEK	9	3	361.8689106	1	rPP-43,rPPKD-43-H2O		
19	17	20160312_	N+0.984R	nrqVIcVTLK	10	2	617.3295629	1	Precursor-43-N	Precursor-43-NH3	
20	18	20160312_	GGTSR+0.	GGTSrALAAA!	14	2	638.8469649	1	Sr-43-H2O,Sr-43,SrA-43,TSr-4		b5-43-H2O,I
21	19	20160312_	EEFER+0.9	EEFErqnKQLR	11	3	493.9078709	1	ErqnKQL-43		
22	20	20160312_	TVEMR+0.	TVEMrDGEVII	11	2	639.3244769	1	rD-43,Mr-43,rD	Precursor-43-	y7-43-NH3+
23	21	20160312_	Q+0.984Q	qqIADLrEDLKI	12	2	744.3891944	1	ADLr-43-H2O,ADLr-43,ADLrEDLK-43		
24	22	20160312_	DSR+0.98	DSrSGSPmAR	10	2	540.7411109	1	rS-43-H2O,Sr-43-H2O,rS-43,S		b3-43-H2O,I
25	23	20160312_	MFGGPGT	MFGGPGTASr	23	3	797.7173133	1	rPSSS-43,SrPSS-43,GTASrPS-4		y15-42-43+
26	24	20160312_	YLATASTM	YLATASTmDH	18	4	516.005272	1	Ar-43,ArH-43,HAr-43,DHAr-4		y8-43++,y9-
27	25	20160312_	YLATASTM	YLATASTmDH	18	3	687.6710876	1	Ar-43,ArH-43,HAr-43,DHAr-4		y9-43++,y7-
28	26	20160312_	EKEEKEEAI	EKEEKEEArLrA	13	2	809.4263074	2	rA-43,Ar-43,EAr-43-H2O,EArL-43,EEArL-43		
29	27	20160312_	DSR+0.98	DSrSGSPmAR	10	2	540.7411109	1	rS-43-H2O,Sr-43-H2O,rS-43,S		b3-43-H2O,I
30	28	20160312_	ITPGAR+0	ITPGArGAFSEI	14	2	763.8784579	1	rG-43,Ar-43,GA	Precursor-43-	y9-43-NH3+
31	29	20160312_	LEMVVVN	LEMVVVNGrG	14	3	500.5932293	1	rGD-43,VNGrGD-43-NH3,y7-		y7-43-H2O
32	30	20160312_	STLAR+0.9	STLArVIVDK	10	2	551.8275124	1	Ar-43,ArV-43,LA	Precursor-43-	y6-43-NH3+
33	31	20160312_	Q+0.984A	qAPRAAPAAP	16	4	436.2305062	1			
34	32	20160312_	STLAR+0.9	STLArVIVDK	10	2	551.8275124	1	Ar-43,rV-43,LAr	Precursor-43	y6-43++,y7-
35	33	20160312_	DSR+0.98	DSrSGSPmAR	10	2	540.7411109	1	rS-43-H2O,Sr-43-H2O,rS-43,S		b3-43-H2O,I
36	34	20160312_	GAAGR+0	GAAGrPLELSD	13	3	463.9094346	1	AGr-43,rPLE-43,GrPLELSD-43		
37	35	20160312_	ILSDDVPIR	ILSDDVPIrDYF	12	3	488.2566083	1	PIr-43,VPIrDY-43-H2O		
38	36	20160312_	DSR+0.98	DSrSGSPmAR	10	2	540.7411109	1	rS-43-H2O,Sr-43-H2O,rS-43,S		b3-43-H2O,I
39	37	20160312_	FADLSEAA	FADLSEAANrN	16	2	889.4273694	1	rN-43-NH3,Nr-43-NH3,Nr-43,		y10-43++,y1
40	38	20160312_	RPFKR+0.9	RPFKrmNYSD/	16	3	681.0283229	1	FKrmNY-43,FKrmNYS-43-H2C		b9-43-59
41	39	20160312_	SFLQ+0.98	SFLqSLEcLrR	11	3	470.9069253	1			
42	40	20160312_	NLDIER+0	NLDIErPTYTNI	14	3	573.9607429	1	Er-43-H2O,Er-43,DIErP-43,y1(		y10-43-H2O
43	41	20160312_	YVETPR+0	YVETPrVHISS\	20	3	756.3858893	2	Pr-43,PrV-43,TPr-43,PrVH-43,I		y13-43++,y1

	L	M	N	O	P
1	intNL_label	Total_INT_label	Dipeptide_label	Tripeptide_label	IM_NH3_label
2	Ar-43,EAr-4	Ar-43,Ar-NH3,A	Ar-43,Ar-NH3,Ar,	ArV-NH3,ArV,EAr-H2O,EAr-NH3	
3	rT-43-H2O,I	rT-43-H2O,Pr-4	rT-43-H2O,Pr-43	TPr-43-H2O,PrT-43-H2O,rTP-43	
4	rT-43-H2O,I	rT-43-H2O,Pr-4	rT-43-H2O,Pr-43	TPr-43-H2O,PrT-43-H2O,rTP-43	
5	rT-43-H2O,I	rT-43-H2O,Pr-4	rT-43-H2O,Pr-43	TPr-43-H2O,PrT-43-H2O,rTP-43	
6	rT-43-H2O,I	rT-43-H2O,Pr-4	rT-43-H2O,Pr-43	TPr-43-H2O,PrT-43-H2O,rTP-43	
7	TAHLr-43-H	HLr-CO		HLr-CO	
8	Pr-43,TPr-43	Pr-43,Pr-NH3,rT	Pr-43,Pr-NH3,rT-	TPr-43-H2O,PrT-43-H2O,rTP-43	
9	Sr-43-H2O,S	Sr-43-H2O,Sr-4	Sr-43-H2O,Sr-43	QSr-43-NH3,SrE-43-H2O,SrE-43	
10	rS-43-H2O,S	rS-43-H2O,Sr-4	rS-43-H2O,Sr-43	rSG-H2O-NH3,rS	IM(r)-NH3
11	Pr-43,TPr-43	Pr-43,Pr-NH3,rT	Pr-43,Pr-NH3,rT-	TPr-43-H2O,PrT-43-H2O,rTP-43	
12	rLq-43	Kr-H2O-NH3,rL	Kr-H2O-NH3	rLq-43	
13	GSPRr-42-4	rPR-42-NH3,RrP-42-NH3,PRr-42-		rPR-42-NH3,RrP-42-NH3,PRr-42	
14					
15					
16	Pr-43,rT-43,T	Pr-43,rT-43,Pr-N	Pr-43,rT-43,Pr-NI	TPr-43-H2O,PrT-43-H2O,rTP-43	
17	rE-43-H2O	Ar-NH3,rE-43-H	Ar-NH3,rE-43-H2O		
18	rPP-43,rPPKI	rP,rPP-43	rP	rPP-43	
19	rqVIcVT-43	rqV		rqV	
20	Sr-43-H2O,S	Sr-43-H2O,Sr-H	Sr-43-H2O,Sr-H2	TSr-43-H2O,SrA-	IM(r)-NH3
21	ErqnKQL-43				
22	rD-43,Mr-43	rD-43,Mr-43,rD-	rD-43,Mr-43,rD-I	rDG-43-H2O,rDG-NH3,rDG,MrD	
23	ADLr-43-H2	Lr-NH3,rE-NH3,	Lr-NH3,rE-NH3	DLr-NH3,rED-NH3	
24	rS-43-H2O,S	rS-43-H2O,Sr-4	rS-43-H2O,Sr-43	rSG-H2O-NH3,rS	IM(r)-NH3
25	rPSSS-43,SrPSS-43,GTASrPS-43-H2O,GTASrPSS-43-H2O,SrPSSSR-43,rPSSSRS-				
26	Ar-43,ArH-4	Ar-43,ArH-43,H	Ar-43	ArH-43,HAr-43	
27	Ar-43,ArH-4	Ar-43,ArH-43,H	Ar-43	ArH-43,HAr-43	
28	rA-43,Ar-43,	rA-43,Ar-43,Ar-I	rA-43,Ar-43,Ar-N	EAr-43-H2O,EAr-NH3	
29	rS-43-H2O,S	rS-43-H2O,Sr-4	rS-43-H2O,Sr-43	rSG-H2O-NH3,rS	IM(r)-NH3
30	rG-43,Ar-43	rG-43,Ar-43,rG-	rG-43,Ar-43,rG-N	GAr-43,ArG-43,rGA-43,ArG-NH3	
31	rGD-43,VNC	Gr-NH3,rG-NH3	Gr-NH3,rG-NH3	GrG-CO,rGD-43	
32	Ar-43,ArV-4	Ar-43,Ar-NH3,A	Ar-43,Ar-NH3,Ar,	ArV-43,LAr-43,ArV-NH3,LAr-NH3	
33					
34	Ar-43,rV-43,	Ar-43,Ar-NH3,rV	Ar-43,Ar-NH3,rV-	LAr-43,ArV-NH3,LAr-NH3,ArV,LA	
35	rS-43-H2O,S	rS-43-H2O,Sr-4	rS-43-H2O,Sr-43	rSG-H2O-NH3	IM(r)-NH3
36	AGr-43,rPLE	AGr-43,rP,AGr-C	rP	AGr-43,AGr-CO,AGr	
37	PIr-43,VPIrD	Ir-NH3,rD-H2O,	Ir-NH3,rD-H2O	PIr-43,IrD-H2O-NH3,PIr-NH3,IrD	
38	rS-43-H2O,S	rS-43-H2O,Sr-4	rS-43-H2O,Sr-43	rSG-H2O-NH3	IM(r)-NH3
39	rN-43-NH3,	rN-43-NH3,Nr-4	rN-43-NH3,Nr-43	ANr-43-NH3,ANr-NH3-NH3,ANr	
40	FKrmNY-43,	FKr-NH3-NH3		FKr-NH3-NH3	
41		Lr-NH3,Lr	Lr-NH3,Lr		
42	Er-43-H2O,I	Er-43-H2O,rP-N	Er-43-H2O,rP-NI	IEr-H2O	
43	Pr-43,PrV-43	Pr-43,rS-NH3,Vr	Pr-43,rS-NH3,Vr-	PrV-43,TPr-43,TPr-NH3,PrV,TPr	

	Q	R	S	T	U	V	W	X	Y	Z
1	Total_NL_count	precNL_count	seqNL_count	intNL_count	Total_INT_count	Dipeptide_count	Tripeptide_count	IM_NH3_count	Cit_probability	Cit_prediction
2	42	0	22	20	9	4	5	0	1	1
3	31	0	14	17	14	8	6	0	1	1
4	17	0	0	17	14	8	6	0	1	1
5	47	0	19	28	16	10	6	0	1	1
6	21	0	1	20	13	7	6	0	1	1
7	1	0	0	1	1	0	1	0	0.425753061	0
8	27	0	11	16	12	6	6	0	1	1
9	23	0	7	16	12	6	6	0	1	1
10	8	0	4	4	8	6	2	1	0.999999059	1
11	16	0	4	12	9	4	5	0	0.999999997	1
12	1	0	0	1	2	1	1	0	0.587272168	1
13	14	0	2	12	1	0	1	0	0.99999583	1
14	0	0	0	0	0	0	0	0	0.127083007	0
15	0	0	0	0	0	0	0	0	0.127083007	0
16	21	0	6	15	10	5	5	0	1	1
17	1	0	0	1	2	2	0	0	0.587272168	1
18	2	0	0	2	2	1	1	0	0.790608964	1
19	2	1	0	1	1	0	1	0	0.663001774	1
20	21	0	10	11	14	3	11	1	1	1
21	1	0	0	1	0	0	0	0	0.278663657	0
22	52	2	16	34	17	6	11	0	1	1
23	3	0	0	3	4	2	2	0	0.973617048	1
24	11	0	6	5	8	6	2	1	0.99999995	1
25	22	0	3	19	0	0	0	0	0.999999997	1
26	18	0	5	13	2	1	1	0	0.999999956	1
27	18	0	5	13	2	1	1	0	0.999999956	1
28	11	0	0	11	4	2	2	0	0.999988977	1
29	6	0	2	4	8	6	2	1	0.999993371	1
30	42	1	16	25	8	5	3	0	1	1
31	5	0	1	4	3	1	2	0	0.992668379	1
32	30	2	14	14	15	6	9	0	1	1
33	0	0	0	0	0	0	0	0	0.127083007	0
34	18	1	7	10	13	6	7	0	1	1
35	8	0	4	4	7	6	1	1	0.999998193	1
36	3	0	0	3	4	1	3	0	0.973617048	1
37	2	0	0	2	8	2	6	0	0.994727688	1
38	6	0	2	4	7	6	1	1	0.999987277	1
39	19	0	10	9	4	1	3	0	0.99999996	1
40	7	0	1	6	1	0	1	0	0.996151439	1
41	0	0	0	0	2	2	0	0	0.349054012	0
42	8	0	3	5	5	4	1	0	0.999892692	1
43	29	0	14	15	9	4	5	0	1	1

Explanations of output columns are as follows:

Column	Explanation
mod_peptide	Simplified peptide with a predefined set of symbols for modifications
Pep_length	Peptide length
mz_Precursor	Theoretical precursor *m/z*
Cit_Count	Number of citrullinated sites
Total_NL_label	Annotations of all diagnostic neutral loss ions
precNL_label	Annotations of precursor neutral losses
seqNL_label	Annotations of sequence ion neutral losses
intNL_label	Annotations of internal ion neutral losses
Total_INT_label	Annotations of all diagnostic internal ions
Dipeptide_label	Annotations of diagnostic dipeptides
Tripeptide_label	Annotations of diagnostic tripeptides
IM_NH3_label	Annotation of IM(Cit)-NH_3
Total_NL_count	Number of all diagnostic neutral loss ions
precNL_count	Number of precursor neutral losses
seqNL_count	Number of sequence ion neutral losses
intNL_count	Number of internal ion neutral losses
Total_INT_count	Number of all diagnostic internal ions
Dipeptide_count	Number of diagnostic dipeptides
Tripeptide_count	Number of diagnostic tripeptides
IM_NH3_count	Number of IM(Cit)-NH_3
Cit_probability	Probability (P) of citrullination status calculated by the EN model (HCD data only)
Cit_prediction	Classification of citrullination status using a P cutoff >0.5 (HCD data only)