

CSE 291: Biomolecular big data systems

Lecture 2: Introduction to MassIVE and ProteoSAFe

Spring 2019



*Center for
Computational
Mass
Spectrometry*



UCSDCSE
Computer Science and Engineering





CREATE AN ACCOUNT

Computer Science and Engineering
University of California, San Diego

Center for Computational Mass Spectrometry

User: Pass: Sign in

[Don't have an account? Register!](#)

[MassIVE Datasets](#) | [General Info](#) | [UCSD Proteomics](#) | [Future Tools](#) | [Demo](#) | [Contact](#)

Please Login to Use Workflows

MassIVE
Mass Spectrometry
Interactive Virtual Environment

MassIVE Repository Statistics

Public Datasets:	8,840	Proteins:	20,116
Number of Files:	3,592,101	Peptides:	5,803,803
Total Size:	162.77 TB	Peptide Variants:	11,110,562
Spectra:	1,907,166,887	PSMs:	452,647,478
Dataset Subscriptions:	2,660	Modifications:	505

Full member of the **Proteome Xchange** consortium

Search Dataset Identifiers or Metadata:

<https://massive.ucsd.edu/>

- Click “Register” to create a new account



CREATE AN ACCOUNT



Username

Name

Organization

Email

Password (confirmation)



SEARCH DATASETS

A screenshot of the Center for Computational Mass Spectrometry website. The header includes the UCSD CCMS logo, the text "Computer Science and Engineering University of California, San Diego", and the "Center for Computational Mass Spectrometry". Below the header are links for "MassIVE Datasets", "General Info", "UCSD Proteomics", "Future Tools", "Demo", and "Contact". On the right side of the header are fields for "User" and "Pass" with a "Sign in" button, and a link "Don't have an account? Register!".



Full member of the **Proteome Xchange** consortium

Please Login to Use Workflows

MassIVE Repository Statistics

Public Datasets:	8,840	Proteins:	20,116
Number of Files:	3,592,101	Peptides:	5,803,803
Total Size:	162.77 TB	Peptide Variants:	11,110,562
Spectra:	1,907,166,887	PSMs:	452,647,478
Dataset Subscriptions:	2,660	Modifications:	505

Search Dataset Identifiers or Metadata:

- Enter “nci-60” into the search box



SEARCH DATASETS

Search Datasets: nci-60

Submitted MassIVE Datasets ◀ Hits 1 ~ 4 out of 4 ▶ Go to Go Export Filtered Results

Select columns

Filter	Title	Submission Type	Description	Keywords	Principal Investigator	Species	Instrument	PTMs	Uploaded By
1	Global Proteome Analysis of the NCI-60 Cell Line Panel, part 3 MSV000082205 PXD005946	Partial	Show	+ LC-MS/MS	Bernhard Kuster	Homo sapiens	LTQ Orbitrap	MOD:00397 - "A protein modification that is produced by reaction with iodoacetamide, usually replacement of a reactive hydrogen with a methylcarboxamido group."	CCMS System (ccms)
2	Global Proteome Analysis of the NCI-60 Cell Line Panel MSV000082204 PXD005940	Partial	Show	+ LC-MS/MS	Bernhard Kuster	Homo sapiens	LTQ Orbitrap Elite	MOD:00397 - "A protein modification that is produced by reaction with iodoacetamide, usually replacement of a reactive hydrogen with a methylcarboxamido group."	CCMS System (ccms)



SEARCH DATASETS

MassIVE MSV000082204

Partial Public PXD005940

Global Proteome Analysis of the NCI-60 Cell Line Panel

[Subscribe](#) [Comment](#) [Reanalyze Spectra](#) [Add Reanalysis](#)

Description

The NCI-60 cell line collection is a very widely used panel for the study of cellular mechanisms of cancer in general and *in vitro* drug action in particular. It is a model system for the tissue types and genetic diversity of human cancers and has been extensively molecularly characterized. Here, we present a quantitative proteome and kinome profile of the NCI-60 panel covering, in total, 10,350 proteins (including 375 protein kinases) and including a core cancer proteome of 5,578 proteins that were consistently quantified across all tissue types. Bioinformatic analysis revealed strong cell line clusters according to tissue type and disclosed hundreds of differentially regulated proteins representing potential biomarkers for numerous tumor properties. Integration with public transcriptome data showed considerable similarity between mRNA and protein expression. Modeling of proteome and drug-response profiles for 108 FDA-approved drugs identified known and potential protein markers for drug sensitivity and resistance. To enable community access to this unique resource, we incorporated it into a public database for comparative and integrative analysis (<http://wzw.tum.de/proteomics/nci60>).

Keywords: LC-MS/MS ; NCI60 ; DTP

Contact

Principal Investigator: Bernhard Kuster, Chair of Proteomics and Bioanalytics Technical University of Munich Emil-Erlenmeyer-Forum 5 85354 Freising Germany, N/A
Submitting User: ccmss

Number of Files: 434
Total Size: 58.87 GB
Spectra: 2,552,251
Subscribers: 0

	Owner	Reanalyses
Proteins (reported):	0	34,782
Peptides:	0	74,837
Variant Peptides:	0	103,543
PSMs:	0	280,907

[FTP Download](#)

FTP Download Link (click to copy):
<ftp://massive.ucsd.edu/MSV000082204>

Species	Instrument	Modifications
Homo sapiens	LTQ Orbitrap Elite	MOD:00397 - 'A protein modification that is produced by reaction with iodoacetamide, usually replacement of a reactive hydrogen with a methylcarboxamido group.'

<https://massive.ucsd.edu/ProteoSAFe/QueryMSV?id=MSV000082204>



MASSIVE: COMMUNITY REANALYSIS

MassIVE MSV000079852

Complete Public

CPTAC, TCGA Cancer Proteome Study of Colorectal Tissue / Global Proteome Analysis of Normal Colon Epithelium Samples

Subscribe

Comment

Reanalyze Spectra

Compare Results

Add Reanalysis

Description

Ninety-five TCGA tumor samples were used in this study from 90 patients, with 5 samples representing different portions from the same tumor. The samples were generated from two TCGA cohorts: 64 are from Colon Adenocarcinoma (COAD) samples and 31 are from the Rectum.

Number of Files: 11,759
Total Size: 1.22 TB
Subscribers: 2

1. MSGF+ – Database search engine
2. MSPLIT – Spectral Library Search Engine
3. ENOSI – Proteogenomic Search Engine
4. MODa – Multi-blind modification database search engine
5. Raven – Spectral networks-based analysis and propagation of identifications
6. Maestro – MSPLIT, MSGF+, MODa cascade Search
7. MSPLIT-DIA – Spectral Library Search for SWATH
8. Upload your own! (*mzIdentML*, *mzTab*, *TSV*)

7	2016-08-18 12:29:55.0	ADD-MASSIVE-REANALYSIS	MSPLIT Reanalysis Results
8	2016-09-14 14:41:40.0	ADD-MASSIVE-REANALYSIS	MSPLIT MS/MS Mixture Re-analysis of CPTAC Colorectal data

<http://proteomics.ucsd.edu/ProteoSAFe>



DATASETS EVOLVE WITH COMMUNITY INPUT

MassIVE MSV000079852

 *Complete*  *Public*

CPTAC, TCGA Cancer Proteome Study of Colorectal Tissue / Global Proteome Analysis of Normal Colon Epithelium Samples

[Subscribe](#)

Comment

Reanalyze Spectra

Compare Results

Add Reanalysis

Description

Ninety-five TCGA tumor samples were used in this study from 90 patients, with 5 samples representing different portions from the same tumor. The samples originated from two TCGA cohorts: 64 are from Colon Adenocarcinoma (COAD) samples and 31 are from the Rectum Adenocarcinoma (READ) collection. The primary data from the liquid chromatography-tandem mass spectrometry (LC-MS/MS) global proteomic profiling of each tumor sample is associated with a data set in the table below. This work was accomplished by the Proteome Characterization Center (PCC) at Vanderbilt University led by Dr. Daniel C. Liebler. Clinical data files contain both the human readable TCGA bar codes and the UUIDs for each sample. Colon tissue samples (ascending and descending) were obtained from 30 patients. Each sample was analyzed with label free global proteomic profiling. These colon samples, while derived from colon cancer subjects, did not contain tumor. Samples were obtained from the Jim Ayers Institute for Precancer Detection and Diagnosis. Data sets below are labeled with the patient identification number and contain data for both ascending and descending tissue samples.

Keywords: CPTAC

Contact

- Dataset Reanalyses

Dataset Reanalyses

◀ Hits 1 ~ 6 out of 6

▶ Go to

Go

Select columns

Filter	Container ▲	Datasets ▲	Reanalysis ▲	Title ▲	Deactivate
<input type="checkbox"/> checked only					
1	RMSV000000004	1	RMSV000000004.3	MSPLIT Reanalysis Results	
2	RMSV000000004	1	RMSV000000004.4	MODA Re-analysis of CPTAC Colorectal data	
3	RMSV000000004	1	RMSV000000004.5	MSPLIT MS/MS Mixture Re-analysis of CPTAC Colorectal data	
4	RMSV000000004	1	RMSV000000004.7	MSGF+ Re-analysis of CPTAC Colorectal data	
5	RMSV000000004	1	RMSV000000004.8	ENOSI Re-analysis of CPTAC Colorectal data	



BROWSE REANALYSES

- Dataset Reanalyses

Dataset Reanalyses					
Select columns					
Filter	Container	Datasets	Reanalysis	Title	Deactivate
<input type="checkbox"/> checked only	<input type="text"/>	<input type="checkbox"/> ~ <input type="checkbox"/>	<input type="text"/>	<input type="text"/>	
1	RMSV000000242	1	RMSV000000242.1	Example MSGF+ reanalyses for CSE190 Fall 2018	

- Click the reanalysis link to view its details



BROWSE REANALYSES

MassIVE Reanalysis - RMSV000000242.1

RPX011188.1

Example MSGF+ reanalyses for CSE190 Fall 2018

Description

Example reanalyses for CSE190 Fall 2018: MSGF+ searches of deep analysis of NCI60 Leukemia and Ovarian cancer cell lines

[See [results attachment job](#) for details]

Reanalyzed Datasets

- [MSV000082204](#) : Global Proteome Analysis of the NCI-60 Cell Line Panel

- Click “Browse Results” to explore the peptide and protein identification results for this reanalysis

Number of Files: 6
Total Size: 657.76 MB

Proteins (reported): 34,782
Peptides: 74,837
Variant Peptides: 103,543
PSMs: 280,907

[FTP Download](#) [Browse Results](#)

FTP Download Link (click to copy):
ftp://massive.ucsd.edu/RMSV000000242/2018-09-25_nuno

Species	Instrument	Modifications
Homo sapiens	LTQ Orbitrap Elite	UNIMOD:4 - "Iodoacetamide derivative."
		UNIMOD:7 - "Deamidation."
		UNIMOD:1 - "Acetylation."
		More...

<https://massive.ucsd.edu/ProteoSAFe/QueryMSV?id=RMSV000000242.1>



BROWSE REANALYSES

Example MSGF+ reanalyses for CSE190 Fall 2018								
Select columns								
	Result File	Total PSMs	Invalid PSMs	Peptides (Derived)	Proteins	Proteins (Derived)	"One-Hit Wonders"	PTMs
1	2018-09-25_nuno_6cb7a21b/ccms_result /MSV000082204_Ovarian_cancer_dee_c28129353f4f479c9864fd810e2b18d0.mzTab	162083 (1.0000% FDR)	825	83395 (0.9994% FDR)	31161	31161	View	8
2	2018-09-25_nuno_6cb7a21b/ccms_result /MSV000082204_Leukemia_deep_Uni_f587a70211c44286b668bce9d70313e3.mzTab	118824 (0.9993% FDR)	545	64613 (0.9993% FDR)	25916	25916	View	8

- “PSMs” = Peptide-Spectrum Matches (spectra identified as peptides by this search)
- “Peptides (Derived)” = Modified peptide sequences taken from all PSMs
- “Proteins (Derived)” = Proteins identified by all PSMs
- “One-Hit Wonders” = Proteins identified by only one peptide in this search
- “PTMs” = Post-Translational Modifications found in this search



RUN A WORKFLOW: LOGIN

The screenshot shows the homepage of the Center for Computational Mass Spectrometry at UCSD. At the top right is a login form with fields for "User" and "Pass" and a "Sign in" button. Below it is a link to "Register!". The main content area has two sections: "Workflow Selection" and "Basic Options".

Workflow Selection:

- Workflow: MS-GF+ (selected)
- Title: (empty input field)

Basic Options:

- Spectrum Files: Select Input Files (with a link to learn more about MS-GF+)
- Instrument: ESI-ION-TRAP
- Cysteine Protecting Group: Carbamidomethylation (+57)
- Number of Allowed ¹³C: 1
- Parent Mass Tolerance: 30 ppm
- Fragmentation Method: Specified in spectrum file
- Protease: Trypsin
- Number of Allowed Non-Enzymatic Termini: 1

<https://proteomics.ucsd.edu/ProteoSAFe/>



RUN A WORKFLOW: ENTER TITLE

Workflow Selection

Workflow: MS-GF+ ▾ Search Protocol: None ▾ Reset Form Save as Protocol

Title:

- Enter an appropriate title for your search job
- Example for the following analysis:
“MSV000082204, Leukemia deep, UniProt reference”



RUN A WORKFLOW: IMPORT DATASET FILES

Basic Options

Spectrum Files:

See [here](#) to learn more about MS-GF+.

Instrument: ESI-ION-TRAP

Fragmentation Method: Specified in spectrum file

Cysteine Protecting Group: Carbamidomethylation (+57)

Protease: Trypsin

Number of Allowed ¹³C : 1

Number of Allowed Non-Enzymatic Termini: 1

Parent Mass Tolerance: 30 ppm

- Click “Select Input Files”



RUN A WORKFLOW: IMPORT DATASET FILES

A screenshot of the CCMS software interface. At the top, there are three buttons: "Select Input Files", "Upload Files", and "Share Files". The "Share Files" button is highlighted with a blue rectangular border. Below this is a toolbar with icons for "Select Input Files" (blue folder), "Upload Files" (document with pencil), and "Share Files" (document with red X). The main area is divided into two sections: "Select Input Files" on the left and "Selected Files" on the right. The "Select Input Files" section shows a tree view of local drives and folders: CCMS_ProteomeDatabases, CCMS_School_2017, CCMS_SpectralLibraries, ccms_test, and speclibs. The "Selected Files" section shows two items under "Selected Spectrum Files": "Selected Spectrum Files" and "Selected Sequence Files". At the bottom are "Clear Selection" and "Finish Selection" buttons.

Select Input Files Upload Files Share Files

Select Input Files i

Spectrum Files Sequence Files

Selected Files

Selected Spectrum Files
Selected Sequence Files

Clear Selection Finish Selection

- A pop-up window will appear to select files
- You may need to enable pop-ups in your web browser
- Click “Share Files”



RUN A WORKFLOW: IMPORT DATASET FILES

A screenshot of a software interface for managing dataset files. At the top, there are three buttons: "Select Input Files", "Upload Files", and "Share Files". Below these are two main sections: "Share Files With User" and "Import Data Share". The "Import Data Share" section contains a text input field with the value "MSV000082204" and a blue-bordered "Import" button. To the right, under "Imported Data Shares", there is a list of four items, each preceded by a red circular icon with a white "X": "CCMS_ProteomeDatabases", "CCMS_School_2017", "CCMS_SpectralLibraries", and "speclibs".

Select Input Files Upload Files Share Files

Share Files With User i

Shared Users

Import Data Share i

MSV000082204 Import

Imported Data Shares

- ✖ CCMS_ProteomeDatabases
- ✖ CCMS_School_2017
- ✖ CCMS_SpectralLibraries
- ✖ speclibs

- Import dataset MSV000082204



RUN A WORKFLOW: IMPORT DATASET FILES

A screenshot of the CCMS software interface. At the top, there are three buttons: "Select Input Files" (highlighted with a blue border), "Upload Files", and "Share Files". Below these are two main sections: "Share Files With User" and "Shared Users". The "Share Files With User" section contains a text input field and a "Share" button. The "Shared Users" section is currently empty. In the bottom right corner, there is a "Import Data Shares" section. This section lists several datasets: "[Dataset MSV000082204] - 'Global Proteome Analysis of the NCI-60 Cell Line Panel'" (with a red error icon), "CCMS_ProteomeDatasets" (green checkmark), "CCMS_School_2017" (green checkmark), "CCMS_SpectralLibraries" (green checkmark), and "speclibs" (green checkmark).

Select Input Files Upload Files Share Files

Share Files With User i Shared Users

Import Data Share i

Import Data Shares

- [Dataset MSV000082204] - "Global Proteome Analysis of the NCI-60 Cell Line Panel"
- CCMS_ProteomeDatasets
- CCMS_School_2017
- CCMS_SpectralLibraries
- speclibs

- After importing, the dataset now appears in your list of imported data shares
- Click “Select Input Files” to return to file selection



RUN A WORKFLOW: IMPORT DATASET FILES

A screenshot of the CCMS software interface. At the top, there are three tabs: "Select Input Files" (highlighted in blue), "Upload Files", and "Share Files". Below the tabs, there are two main sections: "Select Input Files" on the left and "Selected Files" on the right. The "Select Input Files" section contains icons for adding files, a search bar, and a tree view of local directories. The "Selected Files" section shows a list of selected spectrum and sequence files. At the bottom are "Clear Selection" and "Finish Selection" buttons.

Select Input Files Upload Files Share Files

Select Input Files

Spectrum Files Sequence Files

Selected Files

Selected Spectrum Files
Selected Sequence Files

Clear Selection Finish Selection

CCMS_ProteomeDatabases
CCMS_School_2017
CCMS_SpectralLibraries

[Dataset MSV000082204] - "Global Proteome Analysis of the NCI-60 Cell Line Panel"

ccms_test
speclibs

- Expand newly imported dataset MSV000082204



RUN A WORKFLOW: IMPORT DATASET FILES

Select Input Files Upload Files Share Files

Select Input Files i

Spectrum Files Sequence Files

Selected Files

Selected Spectrum Files
Selected Sequence Files

Clear Selection Finish Selection

- Assign files for Leukemia cell line “CCRFCEM” to the “Spectrum Files” collection



RUN A WORKFLOW: IMPORT DATASET FILES

A screenshot of the CCMS software interface. At the top, there are three buttons: "Select Input Files", "Upload Files", and "Share Files". Below these are two main panes. The left pane, titled "Select Input Files", shows a file tree with several folders expanded, including "CCMS_ProteomeDatabases", "CCMS_School_2017", "CCMS_SpectralLibraries", and a dataset folder "[Dataset MSV000082204] - "Global Proteome Analysis of the NCI-60 Cell Line Panel"". This dataset folder contains sub-folders like "ccms_peak" which further contains "RAW" and other cancer type folders. The right pane, titled "Selected Files", shows a list of selected files under "Selected Spectrum Files", specifically "MSV000082204/ccms_peak/RAW//Leukemia/CCRFCEM". There are also buttons for "Clear Selection" and "Finish Selection".

Select Input Files Upload Files Share Files

Select Input Files Selected Files

Spectrum Files Sequence Files

Selected Spectrum Files
MSV000082204/ccms_peak/RAW//Leukemia/CCRFCEM

Selected Sequence Files

Clear Selection Finish Selection

CCMS_ProteomeDatabases
CCMS_School_2017
CCMS_SpectralLibraries
[Dataset MSV000082204] - "Global Proteome Analysis of the NCI-60 Cell Line Panel"
ccms_peak
ccms_parameters
ccms_peak
RAW
Brain cancer
Breast cancer
Colon cancer
Leukemia
CCRFCEM
Melanoma
NSCLC
Ovarian cancer
Prostate cancer
Renal cancer
ccms_statistics
raw
search
ccms_test
speclibs

- After assigning files, they should appear under the proper collection on the right



RUN A WORKFLOW: SELECT PROTEIN DATABASE

The screenshot shows a user interface for selecting input files for a proteomics workflow. At the top, there are three buttons: "Select Input Files" (highlighted in blue), "Upload Files", and "Share Files". Below these are two main sections: "Select Input Files" and "Selected Files".

Select Input Files: This section contains icons for file operations: a folder with a green plus sign, a pencil, and a red X. A list of directories is shown, with "CCMS_ProteomeDatabases" highlighted with a blue border. Other listed directories include "CCMS_School_2017", "CCMS_SpectralLibraries", "[Dataset MSV000082204] - Global Proteome Analysis of the NCI-60 Cell Line Panel", "ccms_test", and "speclibs".

Selected Files: This section shows a list of selected files. It includes a folder icon with a red X, a checked checkbox, and a folder icon labeled "Selected Spectrum Files". Below this is another folder icon with a red X and a folder icon labeled "Selected Sequence Files".

At the bottom right of the "Selected Files" section are two buttons: "Clear Selection" and "Finish Selection".

- Expand “CCMS_ProteomeDatabases” to select protein sequence database



RUN A WORKFLOW: SELECT PROTEIN DATABASE

Select Input Files Upload Files Share Files

Select Input Files i

Spectrum Files Sequence Files

Selected Files

Selected Spectrum Files
MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM

Selected Sequence Files

Clear Selection Finish Selection

CCMS_ProteomeDatabases

- uniprot
 - homo_sapiens
 - uniprot-human-swissprot-1_2_2017.fasta
 - uniprot-human-swissprot-5_5_2016.fasta
 - uniprot-human-swissprot-9_19_2018.fasta
 - uniprot-human_proteome_UP000005640_with_isoforms_2017-09-19.fasta**
 - uniprot-human_proteome_with_isotoms_UP000005640_2014-10-23.CC.fasta
 - uniprot-proteome_homo_sapiens_reference_proteome_11_8_2016.fasta
 - uniprot-proteome_homo_sapiens_reference_proteome_5_23_2016.fasta
 - mus_musculus
 - s_cerevisiae
 - CCMS_School_2017
 - CCMS_SpectralLibraries
 - [Dataset MSV000082204]
 - ccms_test
 - speclibs
 - Select UniProt human reference proteome database
“uniprot-human_proteome_UP000005640_with_isoforms_2017-09-19.fasta”



RUN A WORKFLOW: SELECT PROTEIN DATABASE

Select Input Files Upload Files Share Files

Select Input Files i

Spectrum Files Sequence Files

Selected Files

Selected Spectrum Files
MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM

Selected Sequence Files
CCMS_ProteomeDatabases/uniprot/homo_sapiens/uniprot-human_proteome_UP000005640_with_isoforms_2017-09-19.fasta

CCMS_ProteomeDatabases uniprot homo_sapiens mus_musculus s_cerevisiae CCMS_School_2017 CCMS_SpectralLibraries [Dataset MSV000082204] - "Global Proteome Analysis of the NCI-60 Cell Line Panel" ccms_test speclibs

Clear Selection Finish Selection

- After selecting all files, click “Finish Selection”



RUN A WORKFLOW: INPUT PARAMETERS

Basic Options

See [here](#) to learn more about MS-GF+.

Spectrum Files: 0 files and 1 folder are selected

Instrument: ESI-ION-TRAP ▾

Cysteine Protecting Group: Carbamidomethylation (+57) ▾

Number of Allowed ^{13}C : ppm ▾

Parent Mass Tolerance: ppm ▾

Fragmentation Method: ▾

Protease: Trypsin ▾

Number of Allowed Non-Enzymatic Termini: ▾

- Returning to the workflow input form, adjust remaining parameters:
- Set “Number of Allowed ^{13}C ” to 2
- Set “Parent Mass Tolerance” to 50 ppm
- Set “Fragmentation Method” to HCD



RUN A WORKFLOW: INPUT PARAMETERS

Allowed Post-Translational Modifications

Maximum Number of PTMs Permitted in a Single Peptide:

	Mass (Da)	Residues:	Type
<input checked="" type="checkbox"/> Oxidation	+15.994915	M	OPTIONAL
<input type="checkbox"/> Lysine Methylation	+14.015650	K	OPTIONAL
<input checked="" type="checkbox"/> Pyroglutamate Formation	-17.026549	Q	OPTIONAL, N-TERMINAL
<input type="checkbox"/> Phosphorylation	+79.966331	STY	OPTIONAL
<input checked="" type="checkbox"/> N-terminal Carbamylation	+43.005814	*	OPTIONAL, N-TERMINAL
<input checked="" type="checkbox"/> N-terminal Acetylation	+42.010565	*	OPTIONAL, N-TERMINAL
<input checked="" type="checkbox"/> Deamidation	+0.984016	NQ	OPTIONAL
<input type="checkbox"/> iTRAQ8plex:13C(6)15N(2)	+304.199040	K	FIXED
<input type="checkbox"/> iTRAQ8plex:13C(6)15N(2)	+304.199040	*	FIXED, N-TERMINAL
<input type="checkbox"/> TMT Tandem Mass Tag	+229.162932	K	FIXED
<input type="checkbox"/> TMT Tandem Mass Tag	+229.162932	*	FIXED, N-TERMINAL

- Check the boxes for “Oxidation”, “Pyroglutamate Formation”, “N-terminal Carbamylation”, “N-terminal Acetylation”, and “Deamidation”



RUN A WORKFLOW: INPUT PARAMETERS

More Options

Sequence Database:

Include Common Contaminants

Additional Sequences:

1 file and 0 folders are selected

Spectrum-Level FDR

0.01

Peptide-Level FDR

0.01

FPR

10e-9

- Check the box for “Include Common Contaminants”
- Select “Peptide-Level FDR” (leave value at 0.01)



RUN A WORKFLOW: SUBMIT

Workflow Submission

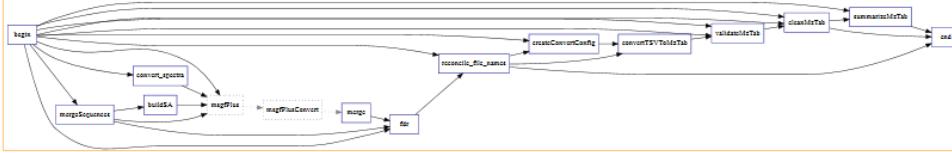
Email me at <your email address>

Submit

- Click “Submit” to launch your search job



RUN A WORKFLOW: SUBMIT

Job Status	
Workflow	MSGF_PLUS
Status	RUNNING [Clone]
User	ccms_test (ccms.web@gmail.com), CCMS
Title	MSV000082204, Leukemia deep, UniProt reference
Date Created	2018-09-26 17:05:14.0
Execution Time	35 seconds
Progress	
Spectrum Files	<p>MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_F11_P003814_B0N_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_A10_P003814_B0A_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_H12_P003814_B0X_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_F10_P003814_B0F_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_A11_P003814_B01_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_H10_P003814_B0H_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_F12_P003814_B0V_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_A12_P003814_B0Q_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_G12_P003814_B0W_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_G10_P003814_B0G_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_G11_P003814_B0O_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_C12_P003814_B0S_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_D12_P003814_B0T_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_C11_P003814_B0K_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_D10_P003814_B0D_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_C10_P003814_B0C_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_D11_P003814_B0L_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_E11_P003814_B0M_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_B10_P003814_B0B_A00_R1.mzML MSV000082204/ccms_peak/RAW/Leukemia/CCRFCEM/00522_E10_P003814_B0E_A00_R1.mzML</p> <p>More available</p>
Additional Sequences	CCMS_ProteomeDatabases/uniprot/homo_sapiens/uniprot-human_proteome_UP000005640_with_isofoms_2017-09-19.fasta

- This job is now running. It will take a while to finish.
- In the meantime, we can see results from a finished version of the same job.



BROWSE WORKFLOW RESULTS

Job Status

Workflow	MSGF_PLUS
Status	DONE [Clone] [Browse mzTab Result Files] Legacy Views [Group by Spectrum] [Group by Peptide] [Group by Protein]
User	nuno (bandeira@ucsd.edu), UCSD
Title	MSV000082204, Leukemia deep, UniProt reference
Re-Analyze Task Outputs	Import to Re-analyze Task Data Attach Reanalysis Results to Dataset
Date Created	2018-09-17 13:08:02.0
Execution Time	8 hours 55 minutes 45 seconds

<https://proteomics.ucsd.edu/ProteoSAFe/status.jsp?task=f587a70211c44286b668bce9d70313e3>

- Click “Browse mzTab Result Files” to explore the peptide and protein identification results for this search



BROWSE WORKFLOW RESULTS

MSV000082204, Leukemia deep, UniProt reference		◀	Hits 1 ~ 1 out of 1	▶	Go to	Go			
Select columns									
Filter	Result File	Total PSMs	Invalid PSMs	Peptides (Derived)	Proteins	Proteins (Derived)	"One-Hit Wonders"	PTMs	
1	MSV000082204__Leukemia_deep__Uni_f587a70211c44286b668bce9d70313e3.mzTab	118824 (0.9993% FDR)	545	64613 (0.9993% FDR)	25929	25929	View	8	

- Columns are the same as seen before in reanalysis results
- Click the “Total PSMs” link to view the Peptide-Spectrum Matches from this search



BROWSE WORKFLOW RESULTS

MSV000082204, Leukemia deep, UniProt reference

Hits 1 ~ 30 out of 118824 Go to [] Go Export Filtered Results

Select columns

	PSM ID	Add to Library	Peak List File	Scan	Peptide	Modifications	Protein
1	1	Link	MSV000082204/ccms_peak /RAW/Leukemia/CCRFCEM /00522_H10_P003814_B0H_A00_R1.mzML	12268	ILNIFGVIK	null	sp P02786 TFR1_HUMAN;tr G3V0E5 G3V0E5_HUMAN
2	2	Link	MSV000082204/ccms_peak /RAW/Leukemia/CCRFCEM /00522_C11_P003814_B0K_A00_R1.mzML	8614	EM+15.994915VLELIR	2-UNIMOD:35	Show
3	3	Link	MSV000082204/ccms_peak /RAW/Leukemia/CCRFCEM /00522_C11_P003814_B0K_A00_R1.mzML	8614	EM+15.994915VLEIIR	2-UNIMOD:35	sp Q96I24 FUBP3_HUMAN;sp Q96I24-2 FUBP3_HUMAN
4	4	Link	MSV000082204/ccms_peak /RAW/Leukemia/CCRFCEM /00522_B12_P003814_B0R_A00_R1.mzML	10775	DADVQNFSFISK	null	tr H3BS70 H3BS70_HUMAN

- Each row corresponds to one PSM, with columns to identify the spectrum, matched peptide sequence, modifications and all identified proteins



BROWSE WORKFLOW RESULTS

A screenshot of a PSM search results table from the MS-GF+ search engine. The table has 13 columns: Valid, QValue, sorting_score, Precursor, decoy, EValue, SpecEValue, MSGFScore, IsotopeError, DeNovoScore, PrecursorError(ppm), PepQValue, and PepQValue. The MSGFScore column is highlighted with a blue border and a red arrow points to it from the top right. The data shows four rows of PSMs, all marked as VALID. The MSGFScore values range from 4.65e-9 to 4.86e-9. The PepQValue column is empty for all rows.

Valid	QValue	sorting_score	Precursor	decoy	EValue	SpecEValue	MSGFScore	IsotopeError	DeNovoScore	PrecursorError(ppm)	PepQValue
VALID	0.0099932721812901	0.920677558818075	509.82236	0	0.12003902	4.8644737e-9	57	2	65	-19.188227	0.000264283311710153
VALID	0.0099932721812901	0.920749163837409	509.78262	0	0.12001923	4.9348534e-9	101	0	102	-2.0952322	0.00283813066907013
VALID	0.0099932721812901	0.920749163837409	509.78262	0	0.12001923	4.9348534e-9	101	0	102	-2.0952322	0.00283813066907013
VALID	0.0099932721812901	0.92077543523829	490.5795	0	0.12001197	4.6532755e-9	8	0	27	-2.1150403	0

- More columns can be found by scrolling to the right, to help score and evaluate each PSM
- Sort by [descending](#) MSGFScore to see the highest-ranked PSMs, as determined by the MS-GF+ search engine



BROWSE WORKFLOW RESULTS

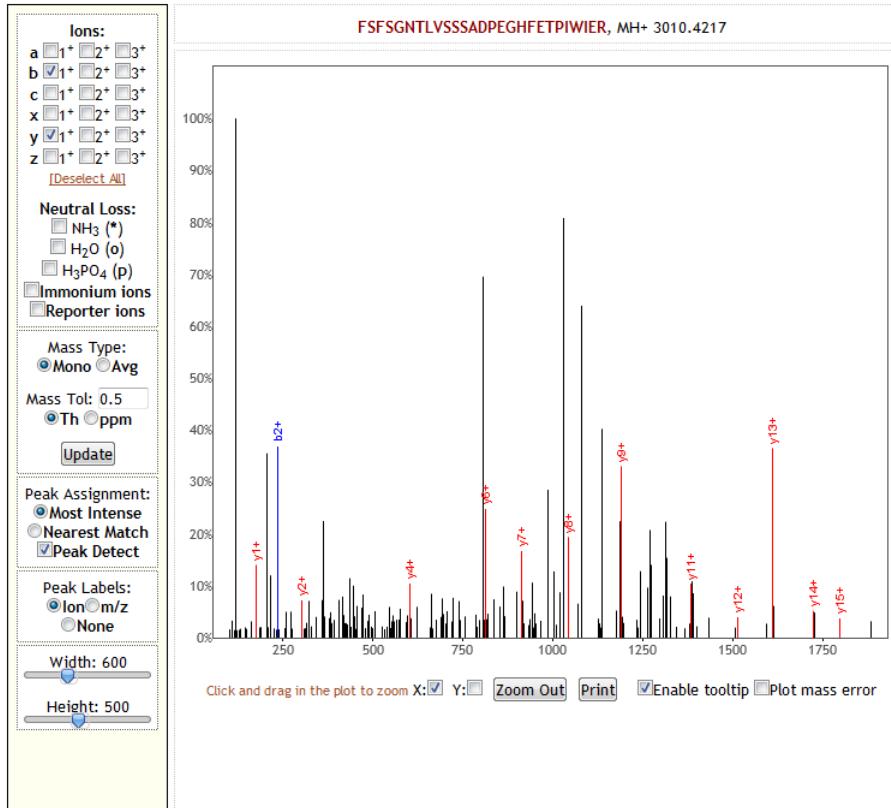
A screenshot of a web-based bioinformatics tool for browsing mass spectrometry (MS) workflow results. The interface includes a header with the title 'MSV000082204, Leukemia deer' and a search bar indicating 'Hits 1 ~ 30 out of 118824'. The main content is a table with columns for Peak List File, Scan, Peptide, Modifications, and Protein. The first row shows a peptide with sequence FSFGNTLVSSADPEGHFETPIWIER and modifications null. The second row shows a peptide with sequence EMQEVTNLQEVVFDSLHATAYQNTALGR and modifications null. The third row shows a peptide with sequence TIQENLNLAQNSAIGC+57.021464HVVAEDLK and modifications 18-UNIMOD:4. The fourth row shows a peptide with sequence +42.010565M+15.994915DEKVFTKELDQWIEQLNEC+57.021464KQLSESQVK and modifications 20-UNIMOD:4,0-UNIMOD:1,1-UNIMOD:35. Each row includes a 'Link' column with a colored icon (red, green, blue, orange) and a 'Show' button. A blue arrow points from the text below to the red icon in the first row's 'Link' column. A small inset window in the top-left corner displays a chromatogram with multiple colored peaks.

Peak List File	Scan	Peptide	Modifications	Protein
MSV000082204/ccms_peak /RAW/Leukemia/CCRFCEM /00522_G10_P003814_80G_A00_R1.mzML	10904	FSFGNTLVSSADPEGHFETPIWIER	null	Show
MSV000082204/ccms_peak /RAW/Leukemia/CCRFCEM /00522_G11_P003814_800_A00_R1.mzML	14442	EMQEVTNLQEVVFDSLHATAYQNTALGR	null	sp O75439 MPPB_HUMAN;tr G3V0E4 G3V0E4_HUMAN
MSV000082204/ccms_peak /RAW/Leukemia/CCRFCEM /00522_D11_P003814_80L_A00_R1.mzML	12926	TIQENLNLAQNSAIGC+57.021464HVVAEDLK	18-UNIMOD:4	sp P13796 PLSL_HUMAN;tr Q5TB3N Q5TB3N_HUMAN
MSV000082204/ccms_peak /RAW/Leukemia/CCRFCEM /00522_A12_P003814_80Q_A00_R1.mzML	12382	+42.010565M+15.994915DEKVFTKELDQWIEQLNEC+57.021464KQLSESQVK	20-UNIMOD:4,0-UNIMOD:1,1-UNIMOD:35	Show

- Click the colored icon on the left of the top scoring row to see the spectrum annotated with its matched peptide



BROWSE WORKFLOW RESULTS



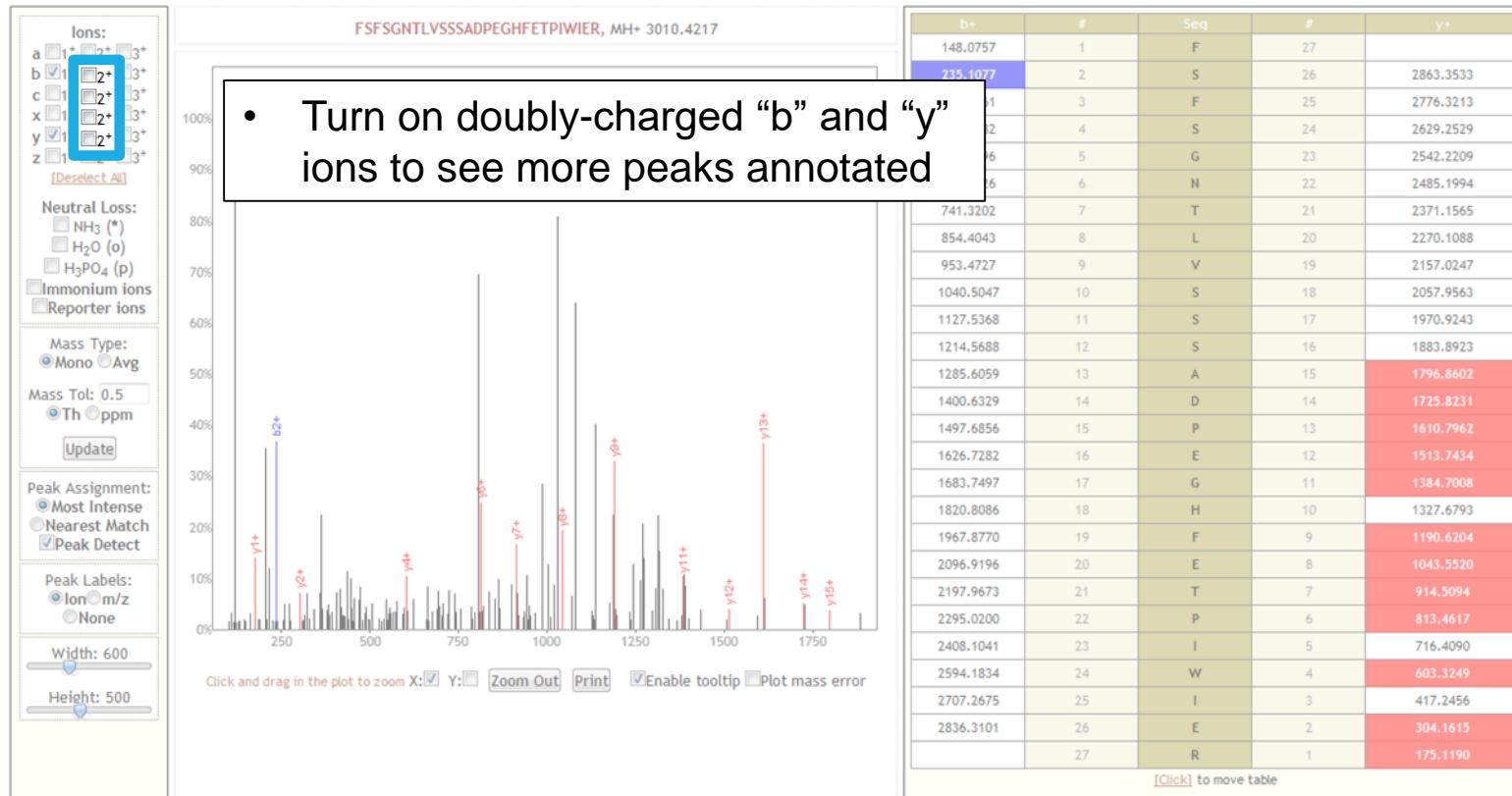
b+	#	Seq	#	y+
148.0757	1	F	27	
235.1077	2	S	26	2863.3533
382.1761	3	F	25	2776.3213
469.2082	4	S	24	2629.2529
526.2296	5	G	23	2542.2209
640.2726	6	N	22	2485.1994
741.3202	7	T	21	2371.1565
854.4043	8	L	20	2270.1088
953.4727	9	V	19	2157.0247
1040.5047	10	S	18	2057.9563
1127.5368	11	S	17	1970.9243
1214.5688	12	S	16	1883.8923
1285.6059	13	A	15	1796.8602
1400.6329	14	D	14	1725.8231
1497.6856	15	P	13	1610.7962
1626.7282	16	E	12	1513.7434
1683.7497	17	G	11	1384.7008
1820.8086	18	H	10	1327.6793
1967.8770	19	F	9	1190.6204
2096.9196	20	E	8	1043.5520
2197.9673	21	T	7	914.5094
2295.0200	22	P	6	813.4617
2408.1041	23	I	5	716.4090
2594.1834	24	W	4	603.3249
2707.2675	25	I	3	417.2456
2836.3101	26	E	2	304.1615
	27	R	1	175.1190

[\[Click\]](#) to move table

- Lots of annotated peaks, but many still unmatched (black), since we're not looking at doubly-charged fragments

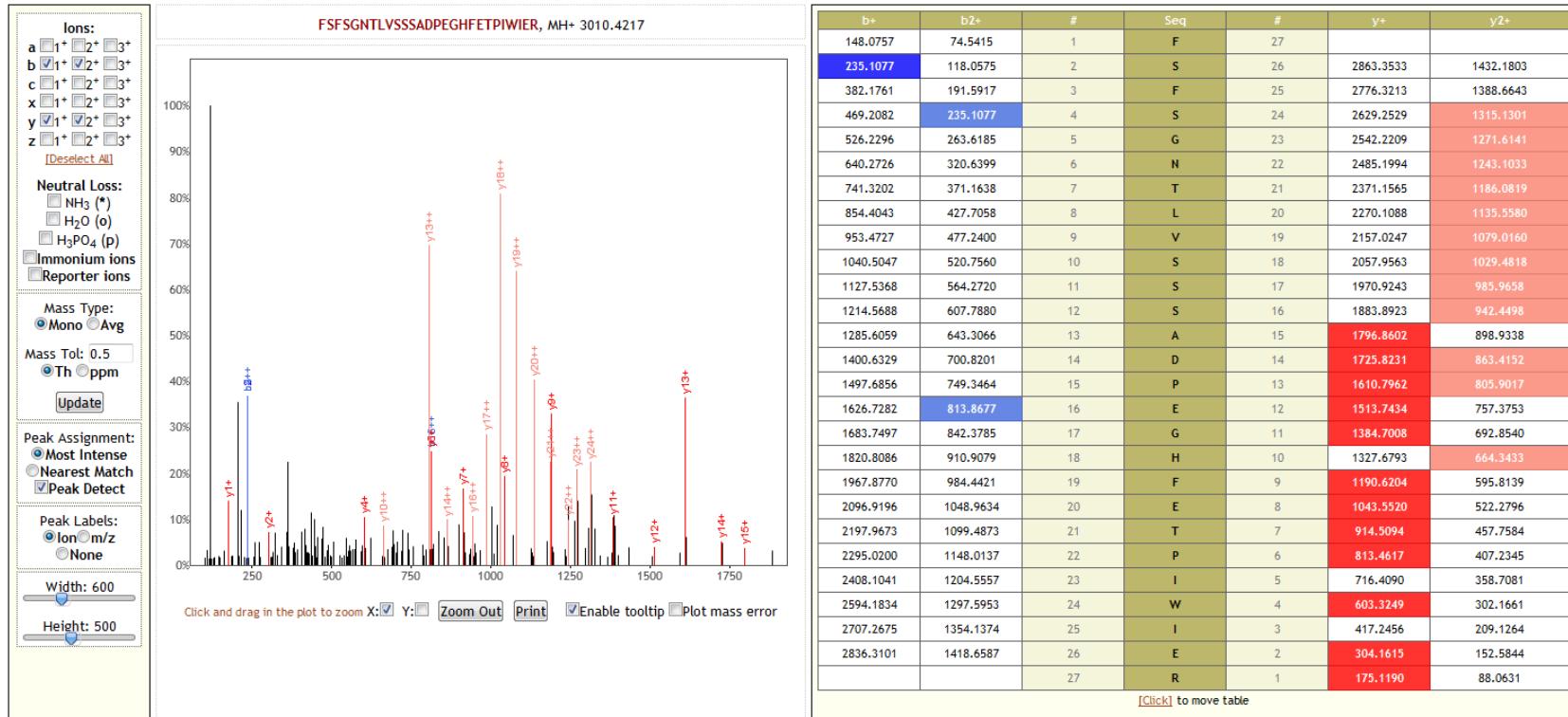


BROWSE WORKFLOW RESULTS





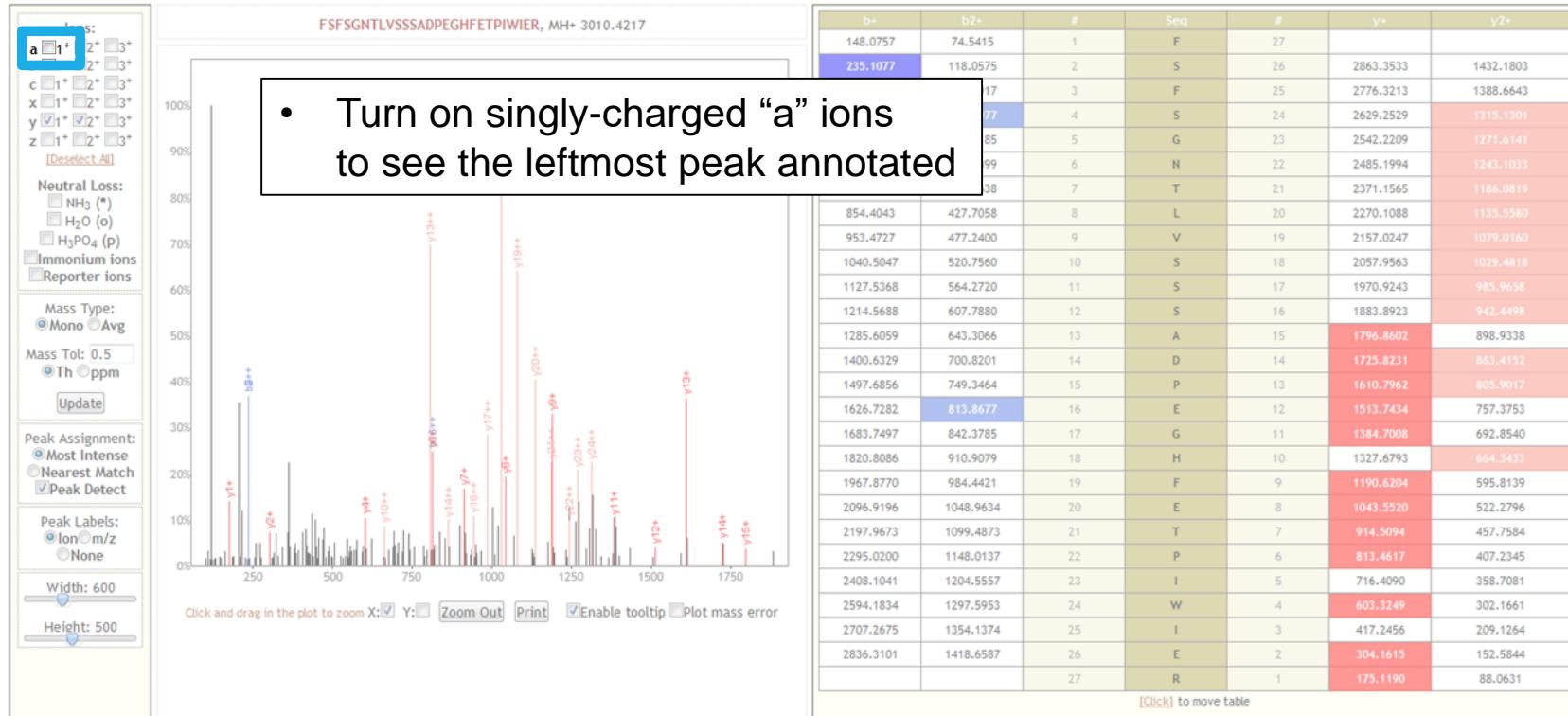
BROWSE WORKFLOW RESULTS



- Many more peaks annotated, but the highest (leftmost) peak is still unmatched

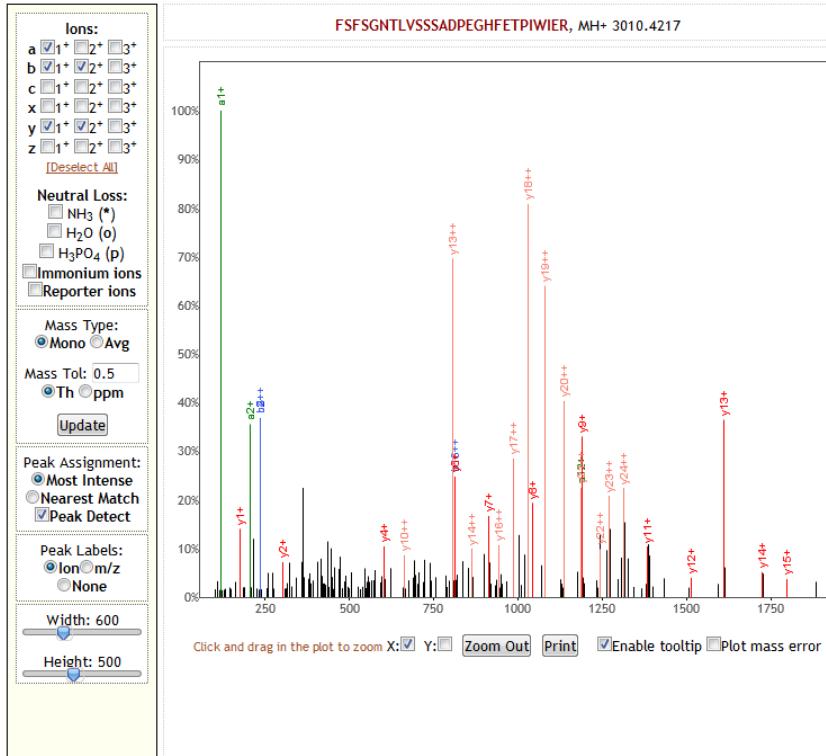


BROWSE WORKFLOW RESULTS





BROWSE WORKFLOW RESULTS



a+	b+	b2+	#	Seq	#	y+	y2+
120.0808	148.0757	74.5415	1	F	27		
207.1128	235.1077	118.0575	2	S	26	2863.3533	1432.1803
354.1812	382.1761	191.5917	3	F	25	2776.3213	1388.6643
441.2132	469.2082	235.1077	4	S	24	2629.2529	1315.1301
498.2347	526.2296	263.6185	5	G	23	2542.2209	1271.6141
612.2776	640.2726	320.6399	6	N	22	2485.1994	1243.1033
713.3253	741.3202	371.1638	7	T	21	2371.1565	1186.0819
826.4094	854.4043	427.7058	8	L	20	2270.1088	1135.5860
925.4778	953.4727	477.2400	9	V	19	2157.0247	1079.0160
1012.5098	1040.5047	520.7560	10	S	18	2057.9563	1029.4818
1099.5419	1127.5368	564.2720	11	S	17	1970.9243	985.9658
1186.5739	1214.5688	607.7880	12	S	16	1883.8923	942.4498
1257.6110	1285.6059	643.3066	13	A	15	1796.8602	898.9338
1372.6379	1400.6329	700.8201	14	D	14	1725.8231	863.4152
1469.6907	1497.6856	749.3464	15	P	13	1610.7962	805.9017
1598.7333	1626.7282	813.8677	16	E	12	1513.7434	757.3753
1655.7548	1683.7497	842.3785	17	G	11	1384.7008	692.8540
1792.8137	1820.8086	910.9079	18	H	10	1327.6793	664.3433
1939.8821	1967.8770	984.4421	19	F	9	1190.6204	595.8139
2068.9247	2096.9196	1048.9634	20	E	8	1043.5520	522.2796
2169.9724	2197.9673	1099.4873	21	T	7	914.5094	457.7584
2267.0251	2295.0200	1148.0137	22	P	6	813.4617	407.2345
2380.1092	2408.1041	1204.5557	23	I	5	716.4090	358.7081
2566.1885	2594.1834	1297.5953	24	W	4	603.3249	302.1661
2679.2726	2707.2675	1354.1374	25	I	3	417.2456	209.1264
2808.3152	2836.3101	1418.6587	26	E	2	304.1615	152.5844
			27	R	1	175.1190	88.0631

[Click] to move table



BROWSE WORKFLOW RESULTS

A screenshot of a search results table from the MS-GF+ search engine. The table has 12 columns: Valid, QValue, sorting_score, Precursor, decoy, EValue, SpecEValue, MSGFScore, IsotopeError, DeNovoScore, PrecursorError(ppm), and PepQValue. A blue callout box highlights the "MSGFScore" column, which contains numerical values such as 57, 101, 101, and 8. A red circle with a question mark is placed over the "MSGFScore" header. The table also includes a header row with sorting arrows for each column.

Valid	QValue	sorting_score	Precursor	decoy	EValue	SpecEValue	MSGFScore	IsotopeError	DeNovoScore	PrecursorError(ppm)	PepQValue
VALID	0.0099932721812901	0.920677558818075	509.82236	0	0.12003902	4.8644737e-9	57	2	65	-19.188227	0.000264283311710153
VALID	0.0099932721812901	0.920749163837409	509.78262	0	0.12001923	4.9348534e-9	101	0	102	-2.0952322	0.00283813066907013
VALID	0.0099932721812901	0.920749163837409	509.78262	0	0.12001923	4.9348534e-9	101	0	102	-2.0952322	0.00283813066907013
VALID	0.0099932721812901	0.92077543523829	490.5795	0	0.12001197	4.6532755e-9	8	0	27	-2.1150403	0

- More columns can be found by scrolling to the right, to help score and evaluate each PSM
- Sort by [ascending](#) MSGFScore to see the lowest-scored PSMs, as determined by the MS-GF+ search engine



BROWSE WORKFLOW RESULTS

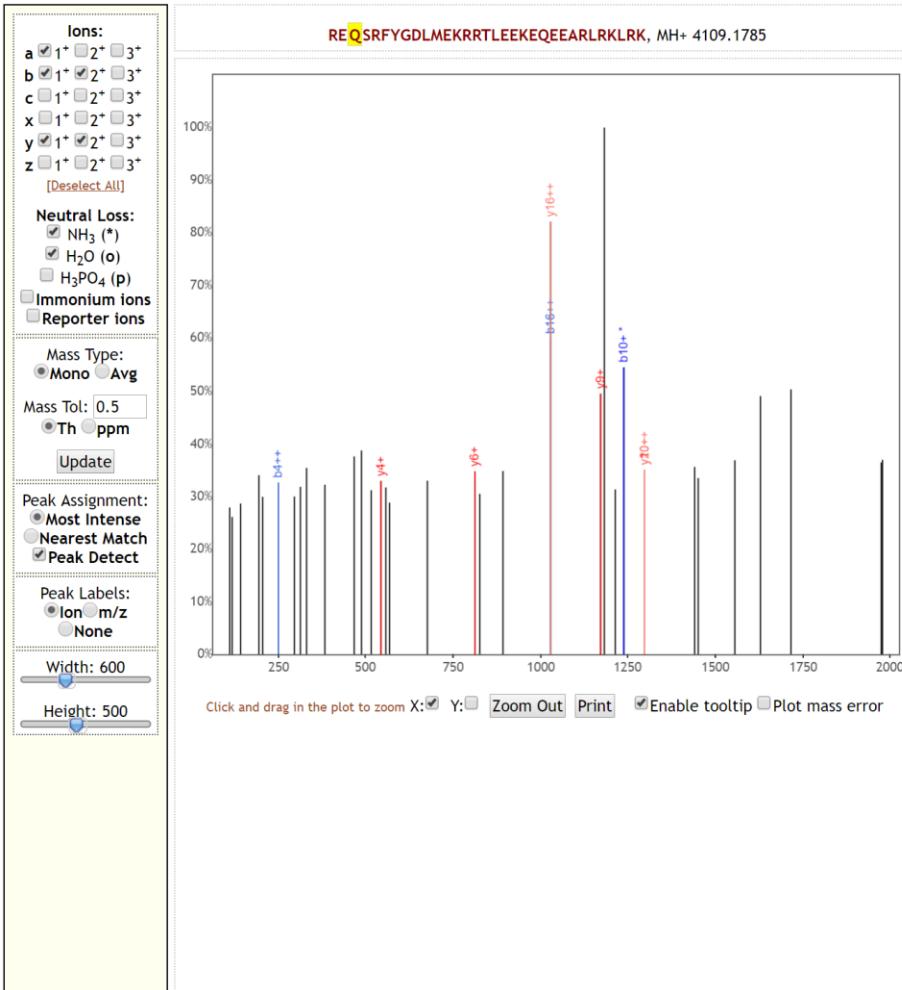
MSV000082204, Leukemia deep, UniProt reference						
Select columns						
Filter	PSM ID	Scan	Peptide	QValue	EValue	MSGFScore
1	5666	14133	TGKIPVTDEEQTNVPYIYAIQDILEDKVLTQVAIQAGR	0.00313570487483531	0.035181865	-24
2	5827	Link	REQ+0.984016SRFYGDLMEKRRTLEEKEQEEARLRKLK	0.00290257780766882	0.030906903	-24
3	5666	Link	HDSKWFKEPYFVHAVEWGSHVYFFFR	0.00295068643214305	0.031699926	-23
4	17517	Link	TDALQPYPHEYVPWVTVN+0.984016GKPLEDQTQLLTV+57.021464QLYQGK	0.000552649758215731	0.0050756456	-22
5	17518	Link	TDALQPYPHEYVPWVTVNGKPLEDQ+0.984016TQLLTV+57.021464QLYQGK	0.000552649758215731	0.0050756456	-22

- Select the second PSM on the list

REQ+0.984016SRFYGDLMEKRRTLEEKEQEEARLRKLK



BROWSE WORKFLOW RESULTS



a+	b+	b2+	#	Seq	#	y+	y2+
129.1135	157.1084	79.0578	1	R	32		
258.1561	286.1510	143.5791	2	E	31	3953.0774	1977.0423
387.1987	415.1936	208.1004	3	Q	30	3824.0348	1912.5210
474.2307	502.2256	251.6164	4	S	29	3694.9922	1847.9997
630.3318	658.3267	329.6670	5	R	28	3607.9602	1804.4837
777.4002	805.3951	403.2012	6	F	27	3451.8591	1726.4332
940.4635	968.4585	484.7329	7	Y	26	3304.7907	1652.8990
997.4850	1025.4799	513.2436	8	G	25	3141.7273	1571.3673
1112.5119	1140.5069	570.7571	9	D	24	3084.7059	1542.8566
1225.5960	1253.5909	627.2991	10	L	23	2969.6789	1485.3431
1356.6365	1384.6314	692.8193	11	M	22	2856.5949	1428.8011
1485.6791	1513.6740	757.3406	12	E	21	2725.5544	1363.2808
1613.7741	1641.7690	821.3881	13	K	20	2596.5118	1298.7595
1769.8752	1797.8701	899.4387	14	R	19	2468.4168	1234.7120
1925.9763	1953.9712	977.4892	15	R	18	2312.3157	1156.6615
2027.0240	2055.0189	1028.0131	16	T	17	2156.2146	1078.6109
2140.1080	2168.1029	1084.5551	17	L	16	2055.1669	1028.0871
2269.1506	2297.1455	1149.0764	18	E	15	1942.0828	971.5451
2398.1932	2426.1881	1213.5977	19	E	14	1813.0403	907.0238
2526.2882	2554.2831	1277.6452	20	K	13	1683.9977	842.5025
2655.3308	2683.3257	1342.1665	21	E	12	1555.9027	778.4550
2783.3893	2811.3843	1406.1958	22	Q	11	1426.8601	713.9337
2912.4319	2940.4268	1470.7171	23	E	10	1298.8015	649.9044
3041.4745	3069.4694	1535.2384	24	E	9	1169.7589	585.3831
3112.5116	3140.5066	1570.7569	25	A	8	1040.7163	520.8618
3268.6127	3296.6077	1648.8075	26	R	7	969.6792	485.3433
3381.6968	3409.6917	1705.3495	27	L	6	813.5781	407.2927
3537.7979	3565.7928	1783.4001	28	R	5	700.4941	350.7507
3665.8929	3693.8878	1847.4475	29	K	4	544.3929	272.7001
3778.9770	3806.9719	1903.9896	30	L	3	416.2980	208.6526
3935.0781	3963.0730	1982.0401	31	R	2	303.2139	152.1106
			32	K	1	147.1128	74.0600

[Click] to move table

Variable Modifications:

Q: 0.984016 [3]