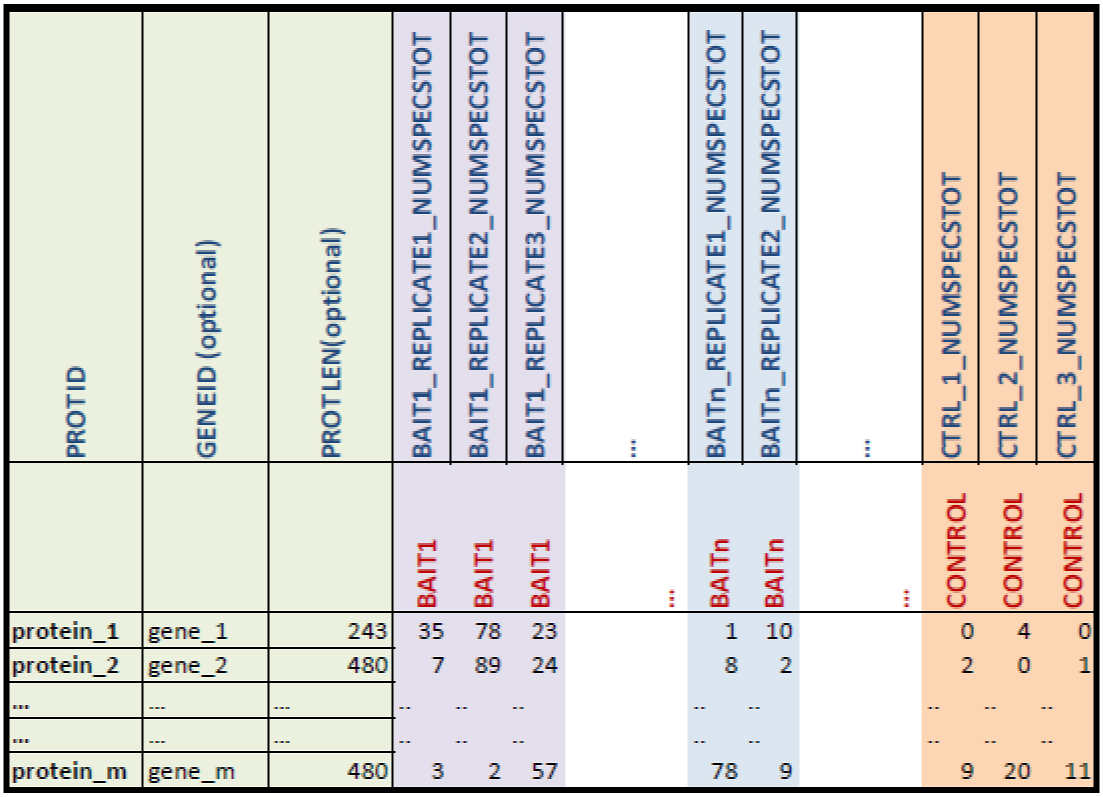
Case Study 3: HDAC5 interactome

Data from “Nuclear import of histone deacetylase 5 by requisite nuclear localization signal phosphorylation”, by Greco *et al.*, Mol Cell Proteomics 10:M110.004317 *(*2011), was used in this Case Study. The dataset used here (a fraction of the entire data used in the original manuscript) consisted of AP-MS experiments with a wild type HDAC5 as the bait protein (two biological replicates) and two negative control experiments. Briefly, eGFP tagged HDAC5 was stably expressed in a HEK293 cell line. The cells were cryogenically lysed and homogenized in a lysis buffer. Affinity purification was carried out by incubating the cell lysate with magnetic beads conjugated with anti-GFP antibodies. Cell lines (HEK293) expressing EGFP-FLAG were used as negative controls. After affinity purification, the samples were separated using SDS-PAGE and cut into X bands. The proteins extracted from each band were digested using trypsin. Peptides samples were then analyzed using an LTQ-Orbitrap XL mass spectrometer over a 90 minute LC gradient.

RAW mass spectrometry files were converted to mzXML format using ProteoWizard (*http://proteowizard.sourceforge.net/project.shtml*). The mzXML files were searched using X! Tandem against the human subset of the UniProt protein sequence database. An equal number of decoy (reverse) sequences and common contaminant proteins were appended to the database. The search results were further processed using the Trans-Proteomic Pipeline (TPP). ABACUS was used to generate the spectral count matrix from the TPP results. The ABACUS output was then manually edited to create a CRAPome input file. The data was subsequently analyzed using the CRAPome interface.

Preparation of the input file

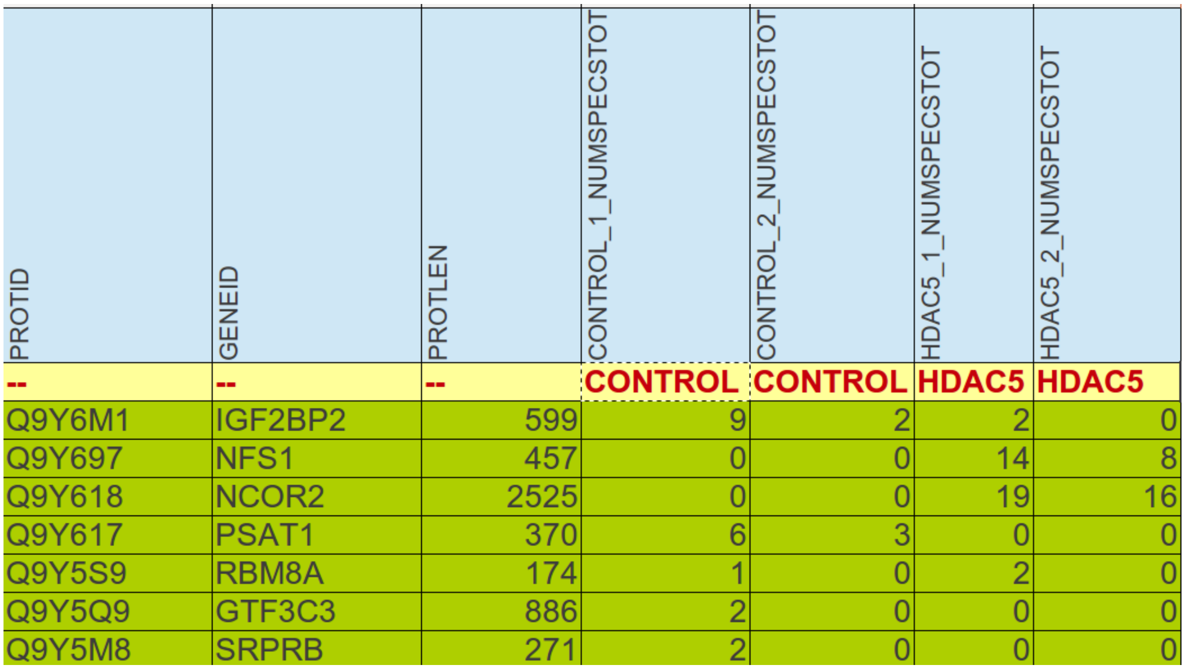
CRAPome supports input data formatted in two ways: 1) list format and 2) matrix format. The list format is a general format and is described in the CRAPome manuscript. The matrix format, now also supported by the CRAPome, provides an alternative option for users processing their data using the TPP/ABACUS. The matrix format is illustrated in Figure 1. Appendix I at the end of the file provides detailed instructions for running the TPP/ABACUS pipeline.



PROTID	GENEID (optional)	PROTLEN(optional)	BAIT1_REPLICATE1_NUMSPECSTOT	BAIT1_REPLICATE2_NUMSPECSTOT	BAIT1_REPLICATE3_NUMSPECSTOT	...	BAITn_REPLICATE1_NUMSPECSTOT	BAITn_REPLICATE2_NUMSPECSTOT	...	CTRL_1_NUMSPECSTOT	CTRL_2_NUMSPECSTOT	CTRL_3_NUMSPECSTOT
			BAIT1	BAIT1	BAIT1	...	BAITn	BAITn	...	CONTROL	CONTROL	CONTROL
protein_1	gene_1	243	35	78	23		1	10		0	4	0
protein_2	gene_2	480	7	89	24		8	2		2	0	1
...	...	...	..	..	..		..	..		..	..	..
...	...	...	..	..	..		..	..		..	..	..
protein_m	gene_m	480	3	2	57		78	9		9	20	11

Figure 1: Matrix format of CRAPome input file. The second row should contain the bait name(s) (gene names). Control APs should be specified as 'CONTROL' or 'C'. The file should be in a tab delimited format.

The structure of the actual input file for the HDAC5 dataset used here is shown in Figure 2. The spectrum count matrix generated by the TPP/ABACUS (see Appendix I for detail) was manually edited. All decoy proteins and common contaminants added to the UniProt database were removed. Prior to that, decoys protein counts were used to estimate the protein identification False Discovery Rate (FDR) to ensure it was below 1%. Also were removed all keratin proteins (optional). One extra row (row 2 in Figure 2) was added to the file to label the columns. The file was saved in a tab delimited format, and then uploaded to CRAPome for subsequent interaction scoring analysis. The HDAC5 input file can be downloaded from the CRAPome website ('Supplementary Data' section).



PROTID	GENEID	PROTLEN	CONTROL_1_NUMSPECSTOT	CONTROL_2_NUMSPECSTOT	HDAC5_1_NUMSPECSTOT	HDAC5_2_NUMSPECSTOT
--	--	--	CONTROL	CONTROL	HDAC5	HDAC5
Q9Y6M1	IGF2BP2	599	9	2	2	0
Q9Y697	NFS1	457	0	0	14	8
Q9Y618	NCOR2	2525	0	0	19	16
Q9Y617	PSAT1	370	6	3	0	0
Q9Y5S9	RBM8A	174	1	0	2	0
Q9Y5Q9	GTF3C3	886	2	0	0	0
Q9Y5M8	SRPRB	271	2	0	0	0

Figure 2: The spectral count matrix generated by ABACUS for HDAC5 dataset. The output from ABACUS was manually edited to include the bait information. 'HDAC5_1_NUMSPECSTOT' and 'HDAC5_2_NUMSPECSTOT' are spectral counts in the AP-MS experiments with HDAC5 as the bait (two biological replicates). Hence "HDAC5" is specified in the second row. 'CONTROL_1_NUMSPECSTOT' and "CONTROL_2_NUMSPECSTOT" are the spectral counts in the negative controls in this dataset, and the columns are labeled as 'CONTROL' accordingly. When generating the matrix using tools other than ABACUS, the first three columns should be renamed PROTID, GENEID, and PROTLEN (note that the last two columns are optional).

Scoring interactions using CRAPome

Using this HDAC5 dataset, we discuss some of our observations that might be helpful for the analysis of similar datasets using '*Workflow 3: Analyze Your Data*'. Please refer to the user manual for a detailed description on how to use this workflow.

1. When only a few user generated controls are available (in this User Case, two), these controls may not capture the complete set of non-specific interactions. While it is possible to perform SAINT analysis with just two negative controls, the results of such an analysis need to be carefully checked. Also, as illustrated in Figure 3, the more conservative

empirical score (FC-B) is too close to the primary score (FC-A) to provide much additional useful information (Figure 3).

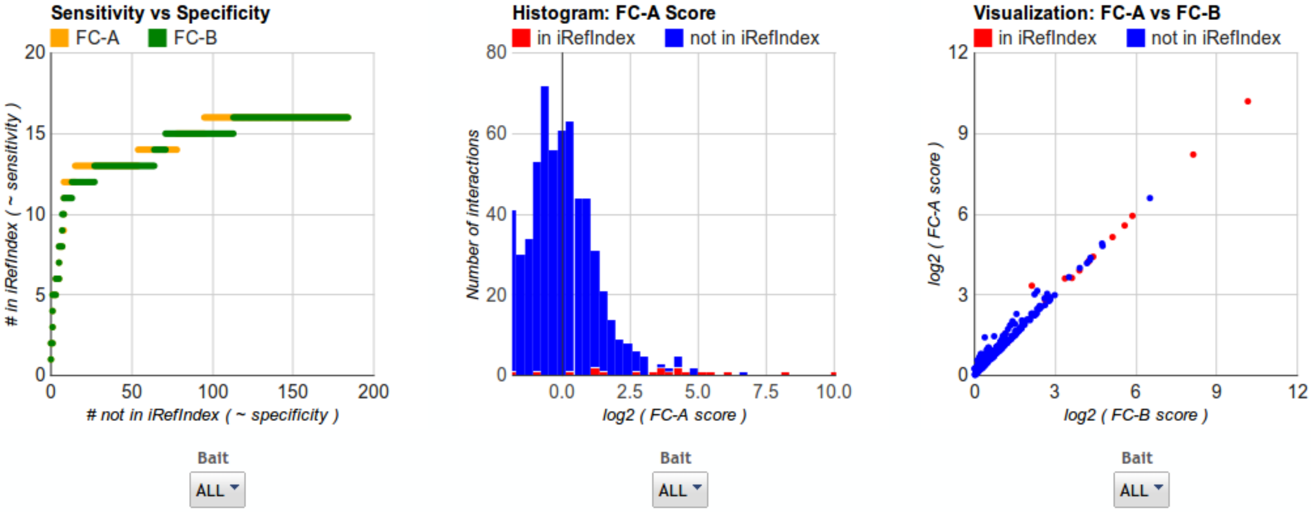


Figure 3: Analysis of HDAC5 data. No CRAPome controls were included in the analysis. The visualization plot (FC-A vs. FC-B) indicates that the more conservative FC-B score provides little additional information on top of the primary FC-A score.

2. When the user chooses to run SAINT, it is important to note that the performance of SAINT depends on the choice of options such as LowMode, MinFold and Normalize (see “Analyzing protein-protein interactions from affinity purification-mass spectrometry data with SAINT”, Curr Protoc Bioinformatics Chapter 8:Unit8.15 (2012) for a detailed discussion; the manuscript can be downloaded from the CRAPome website, “Supplementary Data” section). The “Sensitivity vs. Specificity” plots (ROC-like curves) on the “view results” page can be helpful when choosing the best analysis options for a given dataset. The score that has greater sensitivity for a given specificity can generally be considered a better score. In this case study, SAINT performs slightly better with 0 1 1 options (LowMode = 0; MinFold = 1 and Normalize = 1, which are the default options) compared to other options (e.g. 0 0 1, see Figure 4). Note, however, that the ROC plots may not be sufficiently informative when there are only a few known interactions for the bait as annotated in the iRefIndex database used to generate the ROC plots. Thus, while alternative SAINT options can be used in place of the default options (or the simple FC-A

score may be used instead of the SAINT score), deviation from the standard options is recommended only the ROC curves show a clear benefit of doing so.

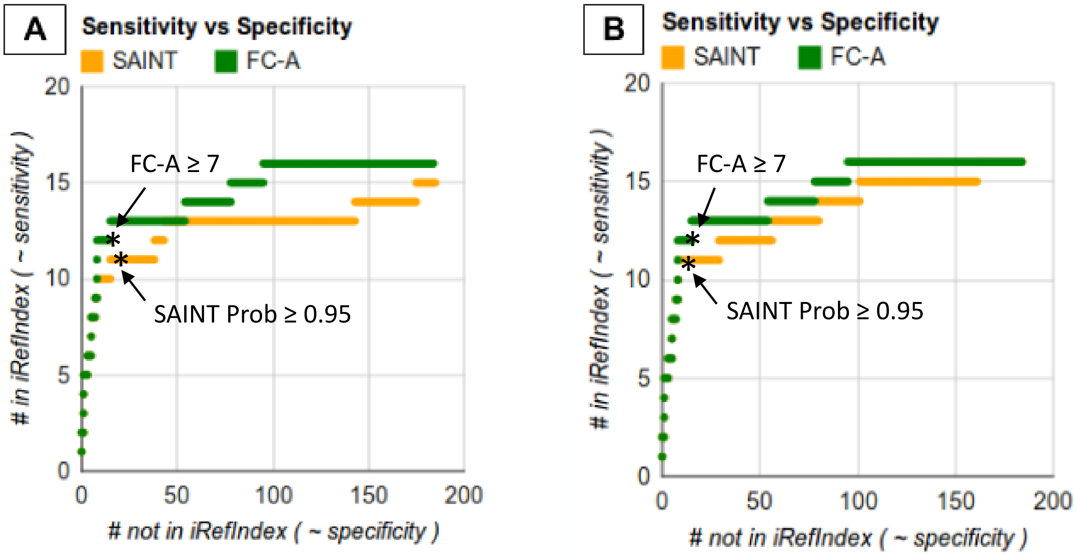
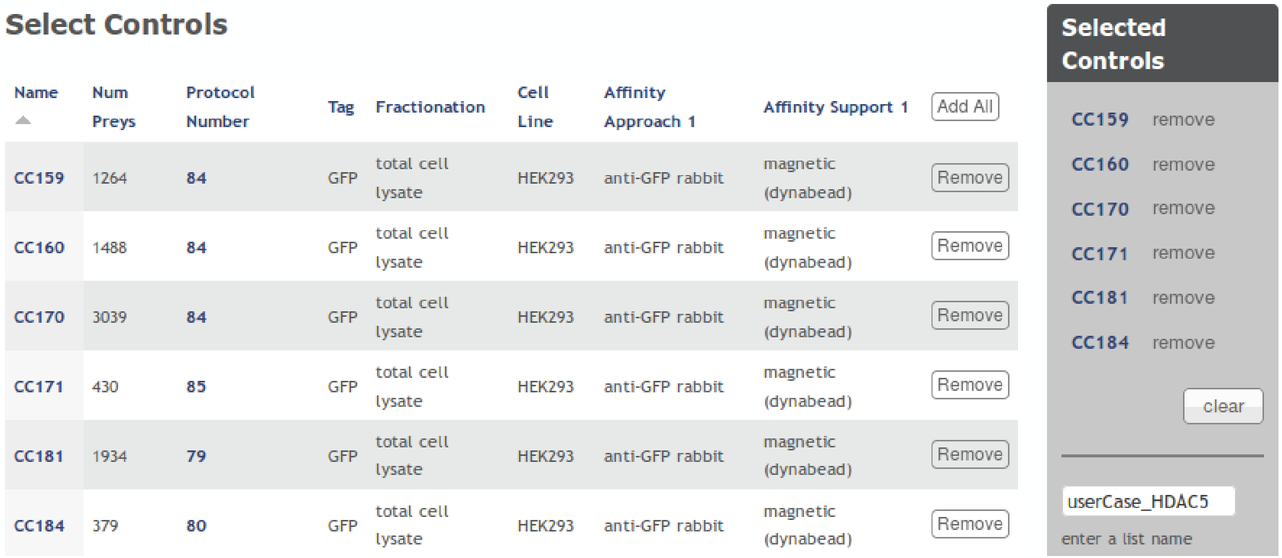


Figure 4: ROC-like curves can be used to determine the most suitable score and/or scoring options for a given dataset. A) SAINT options LowMode=0; MinFold=0; Normalize=1 B) SAINT options LowMode=0; MinFold=1; Normalize =1 (default options). The score that consistently shows a more favorable balance between the number of interactions that are annotated in iRefIndex vs. those that are not, in the relevant range of the score cut-offs, can be considered to be a better score. Here, comparing SAINT results obtained using different options shows comparable performance, and thus the use of default scoring options is suggested. Also note that while the FC-A score appears to slightly outperform the SAINT score, the difference is negligible in the most relevant range of scores. Points corresponding to SAINT probability ≥ 0.95 or FC_A score ≥ 7 thresholds are labeled on the Figures. These points on the ROC curve represent a reasonable set of thresholds for these data because they capture the majority of previously known interactions for HDAC5 bait without admitting a large number of previously annotated interactions. Filtering the data using these thresholds (using either SAINT or FC-A score) results in approximately 30 interactions in each case (of which about slightly less than half were previously known).

3. When none or only a few user controls are available, inclusion of additional controls from the CRAPome should improve the scoring. Here, six additional controls were selected from the CRAPome repository based on the experimental protocol annotation (Figure 5).



Select Controls

Name	Num Preys	Protocol Number	Tag	Fractionation	Cell Line	Affinity Approach 1	Affinity Support 1	Add All
CC159	1264	84	GFP	total cell lysate	HEK293	anti-GFP rabbit	magnetic (dynabead)	Remove
CC160	1488	84	GFP	total cell lysate	HEK293	anti-GFP rabbit	magnetic (dynabead)	Remove
CC170	3039	84	GFP	total cell lysate	HEK293	anti-GFP rabbit	magnetic (dynabead)	Remove
CC171	430	85	GFP	total cell lysate	HEK293	anti-GFP rabbit	magnetic (dynabead)	Remove
CC181	1934	79	GFP	total cell lysate	HEK293	anti-GFP rabbit	magnetic (dynabead)	Remove
CC184	379	80	GFP	total cell lysate	HEK293	anti-GFP rabbit	magnetic (dynabead)	Remove

Selected Controls

CC159 remove
CC160 remove
CC170 remove
CC171 remove
CC181 remove
CC184 remove

userCase_HDAC5
enter a list name

Figure 5: Matching controls are selected from the CRAPome database using the filters on the "Select Controls" page. The following filters were applied in this case: Epitope Tag='GFP'; Fractionation='total cell lysate'; Cell Line='HEK293'; Affinity Approach 1='anti-GFP rabbit'; Affinity Support 1='magnetic (dynabead)'.

4. Inclusion of additional controls slightly improves the SAINT scoring (again measured using the ROC-like curve, data not shown). Furthermore, it improves the utility of the FC-B score as a secondary score in addition to the primary SAINT score. Proteins that have high SAINT score but very low FC-B score are more likely to be false interactions and require additional scrutiny (e.g., they can be exuded from the filtered dataset). In contrast, proteins with SAINT score that are reasonably high but fall below the score threshold (e.g. in 0.7-0.95 range in these data) and high FC-B score are worth additional consideration. Such proteins can be validated using orthogonal methods, or included in the higher level computational analysis.

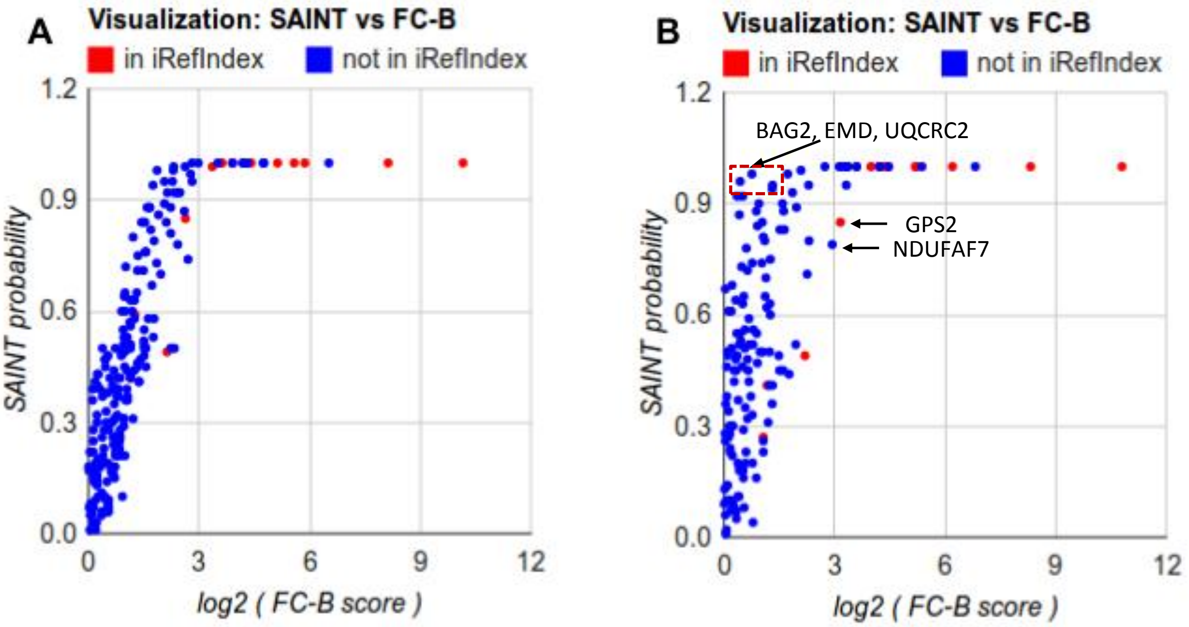
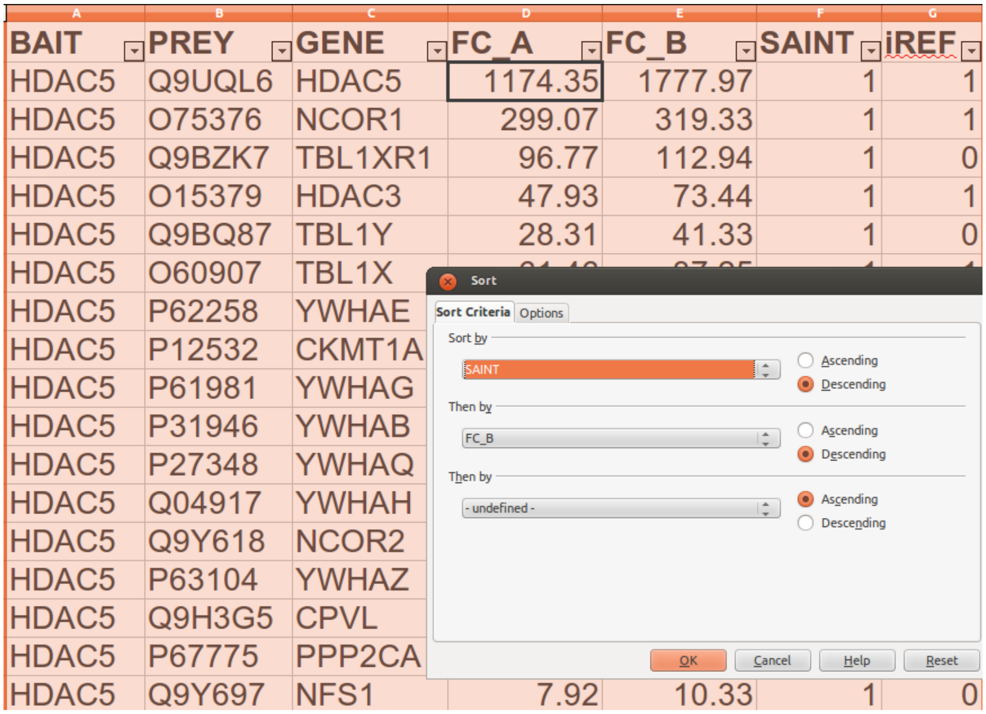


Figure 6: Inclusion of additional CRAPome controls in the analysis improves the interaction scoring. The plot in panel is generated using A) the two user controls only (i.e. no CRAPome controls added), or B) with inclusion of six selected CRAPome controls. With more controls (panel B), FC-B becomes a useful secondary score effective at identifying protein interactions requiring additional scrutiny. For example, interactions having high (above 0.95) SAINT probabilities but very low FC-B scores (EMD, UQCRC2, and DNAJA3, indicated on the Figure) are likely to be non-specific interactions. On the other hand, interactions with SAINT score just below the 0.95 threshold but having high a FC-B score are more likely to be true interactions, and can potentially be included in the subsequent analysis (e.g. GPS2 and NDUFAF7, labeled in the Figure; note that GPS2 is a previously known interaction according to iRefIndex). SAINT was run with the default LowMode=0, MinFold=1, and Normalize=1 scoring options.

5. The results can be downloaded as a tab delimited file and sorted using the SAINT score (or FC-A score if SAINT was not run or if SAINT results were deemed suboptimal based on the analysis described above), see Figure 7.



BAIT	PREY	GENE	FC_A	FC_B	SAINT	iREF
HDAC5	Q9UQL6	HDAC5	1174.35	1777.97	1	1
HDAC5	O75376	NCOR1	299.07	319.33	1	1
HDAC5	Q9BZK7	TBL1XR1	96.77	112.94	1	0
HDAC5	O15379	HDAC3	47.93	73.44	1	1
HDAC5	Q9BQ87	TBL1Y	28.31	41.33	1	0
HDAC5	O60907	TBL1X				
HDAC5	P62258	YWHAE				
HDAC5	P12532	CKMT1A				
HDAC5	P61981	YWHAG				
HDAC5	P31946	YWHAB				
HDAC5	P27348	YWHAQ				
HDAC5	Q04917	YWHAH				
HDAC5	Q9Y618	NCOR2				
HDAC5	P63104	YWHAZ				
HDAC5	Q9H3G5	CPVL				
HDAC5	P67775	PPP2CA				
HDAC5	Q9Y697	NFS1	7.92	10.33	1	0

Sort

Sort Criteria | Options

Sort by: SAINT ○ Ascending ● Descending

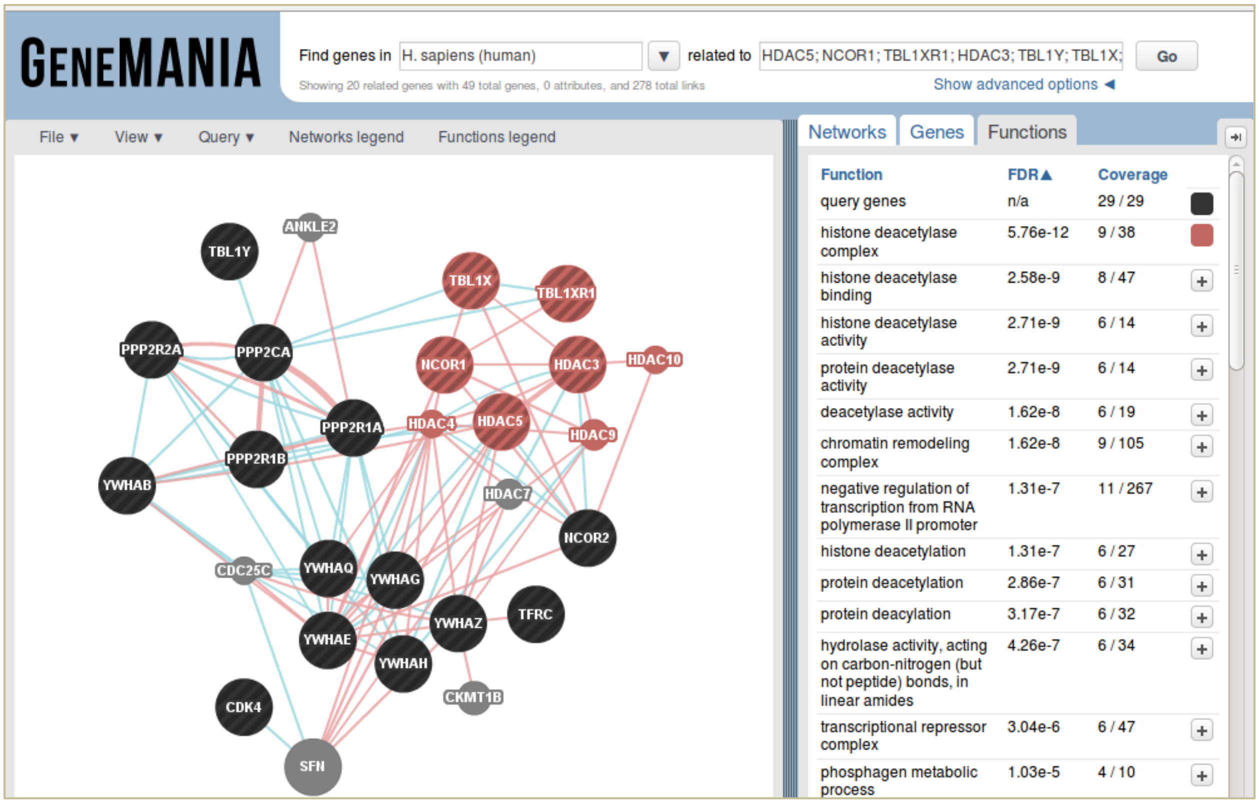
Then by: FC_B ○ Ascending ● Descending

Then by: - undefined - ● Ascending ○ Descending

OK | Cancel | Help | Reset

Figure 7: The results are downloaded in the table format and are sorted and filtered based on SAINT probabilities and/or empirical FC scores.

6. The filtered list of protein interactions can be used for network analysis and visualization. To provide one example, all interactions passing the SAINT probability threshold of 0.95 (29 interactions) were uploaded to GeneMANIA (Mostafavi *et al*. “Combining many interaction networks to predict gene function and analyze gene lists”, Proteomics 12:1687-96 (2012)). The network generated by GeneMANIA is based on prior knowledge and provides a biological context to the analysis. The presence of 'Histone deacetylase complex” as one of the top scoring concepts/functions in this network provided additional confidence for the overall quality of the data (Figure 8). Overall, the analysis recovered the main elements of the interaction network reported in the original publication. Further exploration of the data using GeneMANIA can be useful for understanding the functional role of novel HDAC5 interactions identified in this dataset.



Find genes in H. sapiens (human) ▼ related to HDAC5; NCOR1; TBL1XR1; HDAC3; TBL1Y; TBL1X;

Showing 20 related genes with 49 total genes, 0 attributes, and 278 total links

Function	FDR▲	Coverage	
query genes	n/a	29 / 29	■
histone deacetylase complex	5.76e-12	9 / 38	🟥
histone deacetylase binding	2.58e-9	8 / 47	+
histone deacetylase activity	2.71e-9	6 / 14	+
protein deacetylase activity	2.71e-9	6 / 14	+
deacetylase activity	1.62e-8	6 / 19	+
chromatin remodeling complex	1.62e-8	9 / 105	+
negative regulation of transcription from RNA polymerase II promoter	1.31e-7	11 / 267	+
histone deacetylation	1.31e-7	6 / 27	+
protein deacetylation	2.86e-7	6 / 31	+
protein deacylation	3.17e-7	6 / 32	+
hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides	4.26e-7	6 / 34	+
transcriptional repressor complex	3.04e-6	6 / 47	+
phosphagen metabolic process	1.03e-5	4 / 10	+

Figure 8: Analysis of top scoring (29) interactions using GeneMANIA. The panel on the right shows the top scoring biological functions/concepts in the network. 'Histone deacetylase complex' is among the top scoring concepts.

Appendix A

Tutorial on processing MS/MS data using X! Tandem, the Trans-Proteomic pipeline (TPP), and ABACUS.

PART A: Protein identification using X! Tandem database search and the TPP

This part of the tutorial was written for running the Trans-Proteomic Pipeline (TPP) on a local Windows computer. TPP tools are accessed through the Petunia, the Graphical user interface of TPP.

Information about the TPP can be found in the following manuscript:

Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, Tasman N, Sun Z, Nilsson E, Pratt B, Prazen B, Eng JK, Martin DB, Nesvizhskii AI, Aebersold R. "A guided tour of the Trans-Proteomic Pipeline," Proteomics 1150-9 (2010).

Additional reading:

1. Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. J Proteomics, 2010.
2. Nesvizhskii AI, Aebersold R. Interpretation of shotgun proteomic data: the protein inference problem. Mol Cell Proteomics. 2005.

Technical help for installing/running the TPP:

1. Wiki: http://tools.proteomecenter.org/wiki/index.php?title=Software:TPP
2. TPP support Google group: http://groups.google.com/group/spctools-discuss

1. Install Trans-Proteomic Pipeline (TPP)

Instructions for downloading and installing TPP are available here:
http://tools.proteomecenter.org/wiki/index.php?title=Windows_Installation_Guide

Notes:

a. Prior to installing the TPP, you need to install Active Perl and restart your computer. Active Perl for windows can be downloaded from http://www.activestate.com/activeperl/downloads. Install TPP after successfully installing Active Perl, and restart the computer again.

b. While installing Active Perl, make sure you select the correct installable based on whether your computer is running a 32 bit or 64 bit operating system. To know what version your computer is running, please the instructions on http://support.microsoft.com/kb/827218

2. *Copy input data to your local hard disk*

RAW files are first converted to mzXML format using ProteoWizard and copied to the local disk. For convenience, we recommend organizing the data as follows.

a. Place all the files belonging to a single experiment (i.e., a biological replicate or a control) in a one directory.

b. Place all the experiments in an analysis in a master folder.

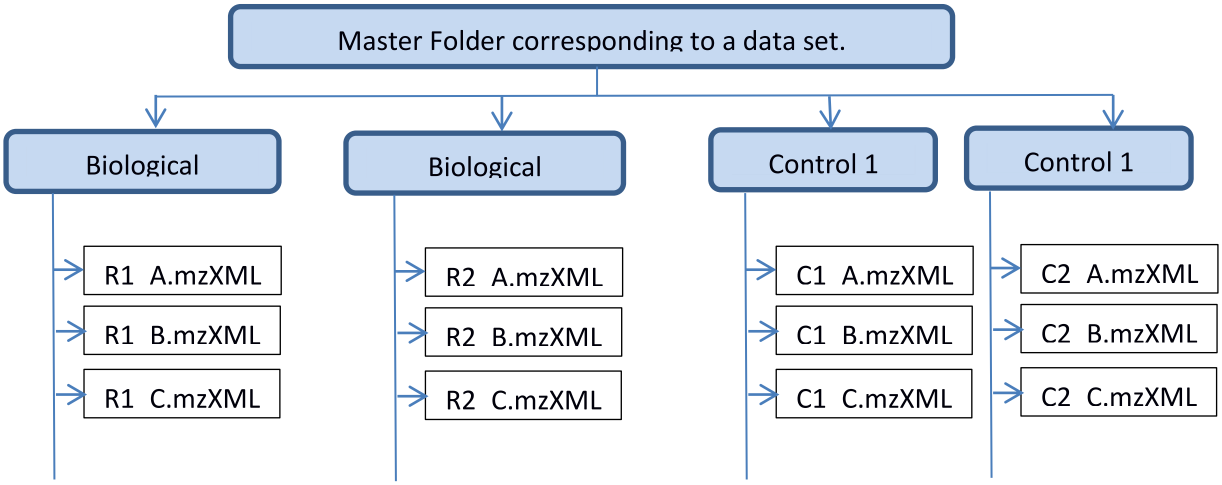


Figure 9. Experimental data, file structure

3. *Login to TPP server on your computer*

Once TPP is successfully installed, double click the “Trans-Proteomic Pipeline” icon on your desktop to launch Petunia, the graphical user interface of TPP. The GUI opens in a browser (e.g., Firefox). You can use the *guest* as user name **and** password to log in. If there is no icon on your desktop, use the following URL: http://localhost/tpp-bin/tpp_gui.pl?Action=display&page=home

Once you are on the **Home** page, please select **Tandem** as the analysis pipeline, which is just below the *Welcome* message. This option refers to the X! Tandem MS/MS database search tool that is provided as a default tool with the TPP.

4. Locate and view input data

To locate and view your input data using the TPP GUI, do the following:

a. Login to Petunia, the web interface of TPP.

b. Mouse-over on the Utilities portion in the navigation links near the top of the Petunia webpage; a pop-up menu should appear. Select the Browse files item in this menu. Alternatively, instead of mouse over, you can click on the larger Utilities portion of the navigation bars located below the navigation links

5. Search MS/MS data with X! Tandem database search tool

A custom version of the popular open-source search engine X! Tandem is bundled and installed with the TPP. To search your data using X! Tandem:

a. Click the **Database Search** menu under **Analysis Pipeline** to access the X! Tandem search interface. You can see the 'Database Search' menu when you mouse over the 'Analysis Pipeline' portion of navigation links at the top of the page (or use the navigation bars instead)

b. Under **Specify mzXML Input Files**, click on '**Add Files'** to add mzXML files from each of your experiments**.**

c. Similarly, under **Specify Tandem Parameters File,** choose the X! Tandem parameters file. A sample parameter file is available for download from the supplementary data section on CRAPome website. This file defines the database search parameters that override the full set of default settings referenced in the file isb_default_input file. The default file is present in the default directory: *C:/Inetpub/wwwroot/ISB/data/parameters/* . For more information, please go to *http://thegpm.org/tandem/api/index.html*

d. Select the sequence database to search against (for example, UniProt database). A database appended with an equal number of decoy (reverse) sequences will facilitate the calculation of false discovery rate (FDR) at a later stage.

e. Select **Convert to PepXML** option. Since each search engine provides results in different ways, the TPP requires that they be converted to a common format for downstream processing (PepXML format).

f. Start the search by clicking on **Run Tandem Search**. While you are waiting, you can periodically refresh the screen to see the output from X! Tandem printed on the screen.

g. To refresh the screen, click on **UPDATE THIS PAGE** link at the bottom left corner of the page. When the search is finished, the folder will contain more files. The main files are the X! Tandem output files (in PepXML format; file names end with *.tandem.pep.xml), one for each mzXML file, e.g. **XYZ.tandem.pep.xml**

6. TPP analysis (PeptideProphet/ProteinProphet)

PeptideProphet provides statistical validation of search engine results by assigning a probability to each peptide-spectrum match. *ProteinProphet* is a protein inference tool that takes as input the list if identified peptides (output from *PeptideProphet*), groups peptides into proteins, and computes a probability of correct identification at the protein level. To run these tools, do the following:

a. Click on the **Analyze Peptides** tab under the *Analysis Pipeline* section in *Petunia* to access the *Xinteract* interface. *Xinteract* is a general utility that is able to launch several components of the TPP, including *PeptideProphet*.

b. Select 'all' **tandem.pep.xml** corresponding to a single experiment (biological replicate or control). Make sure only ***.tandem.pep.xml** files are selected for the analysis; you can edit the selections using the checkboxes and *Remove* button on the right-hand side.

c. Under *PeptideProphet Options*, find and select appropriate options. For example, if your data was generated on a high mass accuracy instrument, you would select **'Use accurate mass**

binning' option. Please refer to the following wiki for details: http://tools.proteomecenter.org/wiki/index.php?title=Software:TPP

Note: We also recommend selecting **'Only use Expect Score as the discriminant'** if you are processing high mass accuracy data searched using narrow mass tolerance like 10ppm.

d. Select **'Run ProteinProphet afterwards'.** With this option, ProteinProphet will start automatically after PeptideProphet (you can also run it separately, see below).

e. Also, find **Enter additional options to pass directly to the command-line (expert use only!).** You need to specify the appropriate output file name here. For example, if you are processing the data for 'replicate 1' in a data set and wish to name your output as 'R1', you would enter **–NR1** in the text box. You can also specify additional options like –PPM if you are processing high mass accuracy data. This option does the modeling on PPM scale.

Note: If the process goes to "finished" state very quickly and does not produce pep.xml then it is very likely that there was an error in specifying the parameters.

f. Click on **Run XInteract** at the bottom of the page to run *PeptideProphet* (and *ProteinProphet* if you selected this option as suggested above).

g. Once the command finishes, you can click on the link that appears in the **Command Status** box to view and analyze the results. Alternatively, mouse over the **Utilities** navigation link**, s**elect the **Browse file s**ystem, and view the output files.

h. The protein summary file is called **interact-*.prot.shtml**. Click on **View** to open up this file in the browser to see the list of identified proteins. Note that the actual data is in the file with *.xml extension (same name).

7. Generate the 'combined' ProteinProphet file

All the experiments in a dataset are merged using ProteinProphet into a single 'combined' file, which helps ABACUS to create a master list of proteins across all the experiments in a dataset.

a. Mouse over on the **Analysis Pipeline** section of the navigation links at the top of the page and click on **Analyze Proteins**.

b. Select ‘all’ the PeptideProphet output files from ‘all’ replicates and controls in a dataset (i.e., all directories in a dataset).

c. Specify the output file name as “interact-COMBINED.prot.xml” and the desired output location.

d. Click on **Run ProteinProphet** at the bottom of the page to run ProteinProphet.

Part B: Generate spectral count matrix using ABACUS

1. Organize the data

Manually copy individual pep.xml and prot.xml files (from the corresponding directories) into a new directory. Make sure you copy*.pep .xml and *.prot.xml files and not files with other extensions. Also copy the interact-COMBINED.prot.xml file into the same folder.

2. Run ABACUS

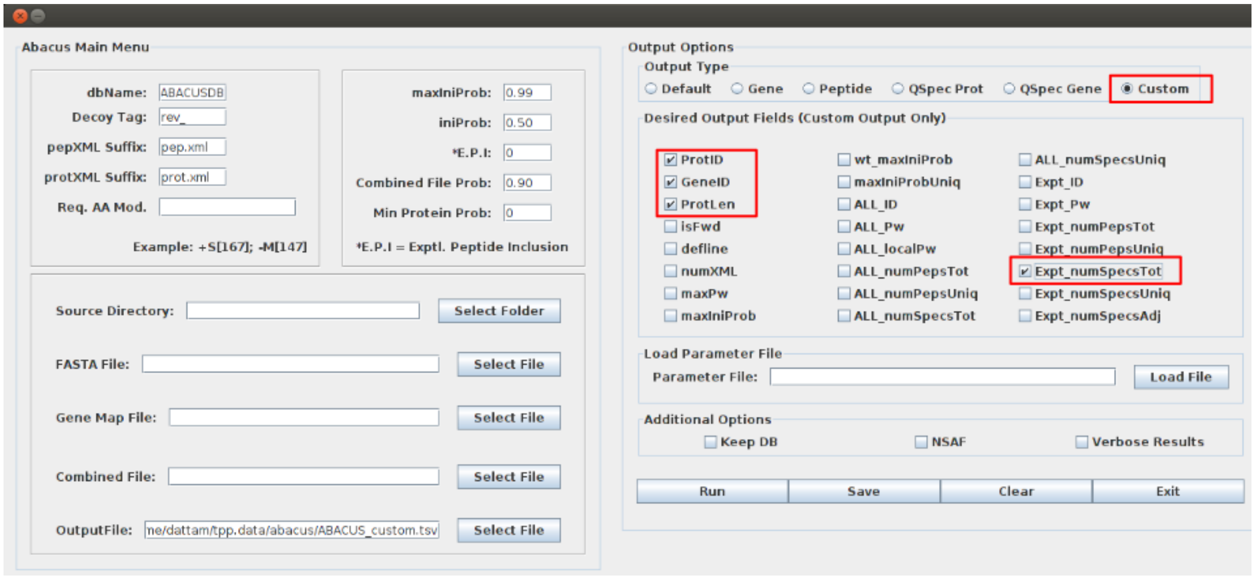
a. Download abacus.jar file from http://abacustpp.sourceforge.net/ .

b. Double click on abacus.jar file to launch the application.

c. Select the options shown in ‘Figure 10’ to generate the spectral count matrix. (Please refer to the user manual available for download at http://sourceforge.net/projects/abacustpp/files/ for details on how to set the parameters in the left panel).

d. Click on “Run” to generate the spectral count matrix.

   i. The user interface of ABACUS (Figure 10) consists of two panels. The left panel allows the user to set the parameters that are used to filter the data (the values shown in the Figure are the default values). The right panel is used to generate the output in the desired format. For generating CRAPome input files, select 'custom' output format as shown in Figure 10.



dbName: ABACUSDB
Decoy Tag: rev_
pepXML Suffix: pep.xml
protXML Suffix: prot.xml
Req. AA Mod.:
Example: +S[167]; -M[147]

maxIniProb: 0.99
iniProb: 0.50
*E.P.I: 0
Combined File Prob: 0.90
Min Protein Prob: 0
*E.P.I = Exptl. Peptide Inclusion

Source Directory:
FASTA File:
Gene Map File:
Combined File:
OutputFile: ne/dattam/tpp.data/abacus/ABACUS_custom.tsv

Output Type
○ Default ○ Gene ○ Peptide ○ QSpec Prot ○ QSpec Gene ● Custom

Desired Output Fields (Custom Output Only)

☑ ProtID
☑ GeneID
☑ ProtLen
□ isFwd
□ defline
□ numXML
□ maxPw
□ maxIniProb
□ wt_maxIniProb
□ maxIniProbUniq
□ ALL_ID
□ ALL_Pw
□ ALL_localPw
□ ALL_numPepsTot
□ ALL_numPepsUniq
□ ALL_numSpecsTot
□ ALL_numSpecsUniq
□ Expt_ID
□ Expt_Pw
□ Expt_numPepsTot
□ Expt_numPepsUniq
☑ Expt_numSpecsTot
□ Expt_numSpecsUniq
□ Expt_numSpecsAdj

Parameter File:

□ Keep DB □ NSAF □ Verbose Results

Figure 10: Abacus interface to generate the spectral count matrix. The highlighted options should be selected to generate a CRAPome compatible file.

ii. The ABACUS output file should be edited manually to specify the bait name for each experiment (see Figure 2). Controls should be specified as 'C' or 'CONTROL'. The first three columns do not need to be labeled in any specific way (e.g. it could be left blank or could contain'--' characters as shown in Figure 2), but the columns should be named be as specified, i.e. 'PROTID', 'GENEID' and 'PROTLEN' (GENEID and PROTLEN columns could be omitted altogether).

iii. Decoy proteins, common contaminants added to the searched database, and keratin proteins were removed.

iv. The file should be saved in a tab delimited format.