Version 1.0.1

ProteomicsBrowser User Guide

Table of Contents

ProteomicsBrowser Downloading and Installation

For Microsoft Windows users, please download the Windows version of ProteomicsBrowser from https://medicine.yale.edu/keck/nida/proteomicsbrowser.aspx. After downloading, first unzip the compressed file and then double click the file "ProteomicsBrowser.exe" to run ProteomicsBrowser.

For Mac Operating System or Linux users, please download ProteomicsBrowser from https://medicine.yale.edu/keck/nida/proteomicsbrowser.aspx. After downloading, first unzip the compressed file and then double click the file "ProteomicsBrowser.jar" to run ProteomicsBrowser. If it does not work, open a command line terminal and enter the ProteomicsBrowser directory. Then use the following command in the command terminal:

java -jar ProteomicsBrowser.jar

Note: ProteomicsBrowser is built on Java 8. Be sure that the latest version of Java 8 is installed on the computer before running ProteomicsBrowser. All of the illustrations shown below and the example data that are discussed below are from data in the test project (test.pbproj) file that is stored in the testData folder after running the ProteomicsBrowser program.

Project Management

Create Project

Since the data in ProteomicsBrowser is organized in individual projects, before importing and processing data, the user must first create a project by clicking "File/New Project", giving the new project a name, and then saving it on the same computer (i.e., not a remote server) on which the ProteomicsBrowser code is stored. Alternatively, if the user chooses to first try the ProteomicsBrowser using the Test sample data set, then after clicking "Open Project" the user can open the "test.pbproj" project that is in the testData folder.

Save Project

After importing the data files (e.g., the Peptide and Sampleinfo files that are in the testData folder), the project can be saved from the menu "File/Save Project". After a project has been saved it can be re-opened again so that additional analyses of the same data can be carried out without having to again import the data.

Open Project

If a ProteomicsBrowser project has been created previously (e.g.,the test.pbproj file that is in the testData folder), it can be opened by clicking "File/Open Project".

Data Import and Data Format

Data Import

After creating a ProteomicsBrowser project, the proteomics data can be imported by using the menu item “File/Import Data”.

As shown in **Figure 1**, ProteomicsBrowser accepts two kinds of data for import: Peptide Data and Sample Data. Peptide Data is mandatory while Sample Data is optional when the peptide data is from only one sample. If the peptide data are from multiple samples, then a Sample Data file is also mandatory for data import. In addition, the organism database that corresponds to the peptide data must also be selected before clicking the “Import” button for data import. Finally, there is a check box at the bottom of the dialogue box: “Include Peptides Mapped to Multiple Proteins”. If this check box is selected, then peptides that map to multiple proteins will be included in the ProteomicsBrowser analyses. Otherwise, all of these peptides will be removed. Please see Peptides Mapped to Multiple Proteins for additional details. For detailed descriptions of the required formats for Peptide Data, Sample Information, and Protein Sequence Database files, please refer to the Data Format section. As an example, users can use the data in the “testData” folder and select “RAT” in “Organisms” to import data.

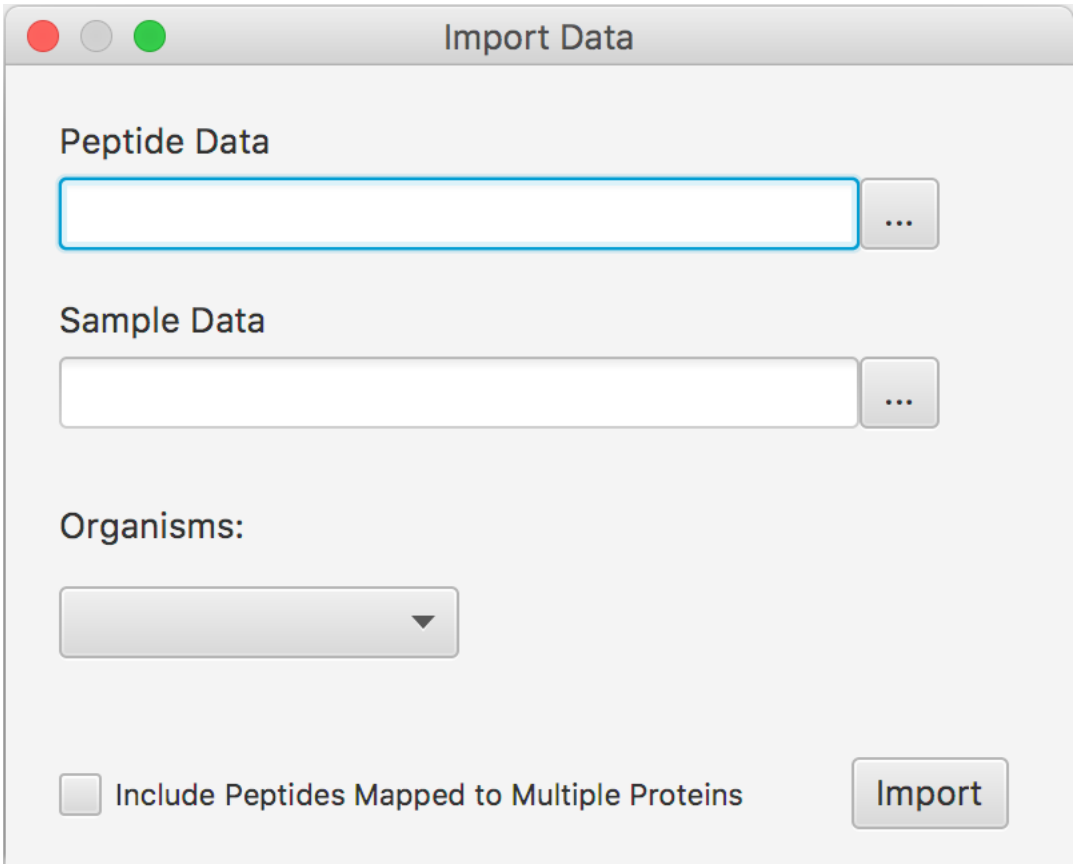


Figure 1

Data Format

Peptide Data File

The peptide data file is in a comma separated values (csv) format. The first row of the file contains the header information for each column. Each row after the header contains the corresponding information for each peptide. If there is only one sample the file must include the following columns:

id	Peptide id, unique for each peptide (character string)
sequence	Amino acid sequence of peptide (character string)
charge	Charge of peptide (integer)
protein	The name of the protein including the peptide (character string)
modification	Modification information for the peptide (character string)
abundance	Peptide abundance/intensity (numerical value)

The format of each column is described below:

id: id for each peptide, numbers, characters or a combination of numbers and characters such as 1, 2, 3 ,…; a, b, c, …; pep1, pep2, pep3, … . With the exception of the **few scenarios described below, the id should be unique for each peptide.**

For some modifications like phosphorylation, the exact modification position may not be known. In this case, there are two options to input the modification information. The first option is to provide a detailed text description of the modification in the modification section. Please check the **modification** section below for details. The second option is to record the modification information in two or more peptide records with the *same* id. In the latter case, it is important to note that the **ProteomicsBrowser combines together all peptides that share the same id**. For example, peptide 37384 has three records in the sample test data. All three of these records have the same information except for the modification column. The three modification records are [3] Phospho (ST), [8] Phospho (ST), and [10] Phospho (ST). Hence, although it has been determined that peptide 37384 contains one phosphorylation site, the exact position of this phosphorylation is unknown. Since the phosphorylation could be at either position 3, 8, or 10; the modification information is displayed as: “[3] Phospho (ST,NA);[8] Phospho (ST,NA); [10] Phospho (ST,NA)”.

Another scenario that could result in one id having multiple records is when one peptide can be mapped to multiple proteins. In this case, there are two options in ProteomicsBrowser depending upon the selection that was made during Data Import. If “Include Peptides Mapped to Multiple Proteins” was selected, then these peptides will be **INCLUDED** in the analysis. Otherwise, they will be **REMOVED** from the analysis. Please refer to Peptides Mapped to Multiple Proteins for more details.

sequence: the amino acid sequence of the peptide

charge: the ion charge state of the peptide. The charge should be a non-negative integer.

protein: the name of the protein to which the peptide has been assigned. The Browser accepts two formats: ProteinName and ProteinName_Organism. For example: DYHC1 and DYHC1_RAT.

modification: modification information of the peptide. It has the following format: **[modified position] modification type (amino acid at that is being modified)**. For example, [4] Oxidation (M), oxidation of Methionine (M) at position 4. If there are multiple modifications on an individual peptide, a vertical "|" or ";" can be used to separate them, such as "[4] Oxidation (M)|[5] Oxidation (M)" or "[4] Oxidation (M);[5] Oxidation (M)", which indicates two oxidations on M at positions 4 and 5. For modifications like phosphorylation the exact position of the modification may be unknown, but the probabilities of its occurring at several different locations may be known. For example, in the case of a phosphorylation with 90% probability of being on position 3 and 10% probability of being on position 7 in a peptide, the following format can be used to show this information: [3] Phospho (ST,90);[7] Phospho (ST,10). If the probability of occurrence at each possible position is unknown, the modification can be displayed as [3] Phospho (ST,NA);[7] Phospho (ST,NA).

abundance: calculated abundance/intensity of the peptide.

If additional peptide information is available, such as m/z, mass, etc; then an additional column with a corresponding header can be added for each in the first row.

When there are multiple samples, there will be more than one abundance value for each peptide. In this case, multiple columns of abundance data are added with each containing the corresponding sample id as its header. As noted above, when there are multiple samples, a sample information file must be imported together with the Peptide Data file. In this case the sample id in the peptide data file must be **exactly the same** as the sample id in the sample information file (see Sample Information File for more details). Otherwise, ProteomicsBrowser will not be able to match the two files together. Example files for peptide data and sample information are shown below. The sample ids are shown in red and they are exactly the same in both files.

Peptide Data File

id	Charge	m/z	Sequence	Modification	Protein	Normal_1	Normal_2	Normal_3	Normal_4	Normal_5	Normal_6	Disease_1	Disease_2	Disease_3	Disease_4	Disease_5	Disease_6
6040	3	778.753817	LVPLLLEDGGDAPAALEAAL		DYHC1_RAT	2737662.42	2036890.03	1836054.76	2136971.45	2280041.64	1763497.95	2915105.29	2342419.51	1782989.97	2089443.16	2128800.62	1911894.76
50536	2	1167.6258	LVPLLLEDGGDAPAALEAAL		DYHC1_RAT	204424.983	260016.481	181582.207	184257.803	234780.162	178508.101	357126.465	192633.983	154202.472	188488.744	168804.185	212180.749
86208	4	584.316716	LVPLLLEDGGDAPAALEAAL		DYHC1_RAT	28742.3292	17452.1603	19109.7413	22180.9062	27492.3144	12627.1351	21274.7899	33386.6947	21756.633	24064.0223	21180.5931	18144.945
101598	3	778.754025	LVPLLLEDGGDAPAALEAAL		DYHC1_RAT	35821.7087	21006.315	18248.2578	32019.0592	22057.9954	12533.3916	35769.0981	28681.926	21998.5161	17954.3028	26652.2148	13249.2543
7814	3	708.412752	GIFEALRPLETLPVEGLIR		DYHC1_RAT	1327826.39	909041.129	1064806.82	774652.033	1709936.64	1169128.97	994426.185	1055175.05	788114.072	495790.948	1139796.69	1311802.41
7886	3	447.938441	LLLIQAFRPDR		DYHC1_RAT	559175.115	697556.186	603634.59	477437.313	391544.656	441990.847	804610.78	542213.335	532041.262	469416.526	467530.424	520530.827

Sample Information File

SampleID	Group	Weight	Gender
Normal_1	Normal	134	F
Normal_2	Normal	155	M
Normal_3	Normal	178	F
Normal_4	Normal	142	F
Normal_5	Normal	187	M
Normal_6	Normal	167	M
Disease_1	Disease	199	M
Disease_2	Disease	203	M
Disease_3	Disease	179	F
Disease_4	Disease	213	F
Disease_5	Disease	188	F
Disease_6	Disease	197	M

Sample Information File

If there are multiple samples with each having its own set of peptide data, then a sample information file must be imported together with the peptide data file. Otherwise, a sample information file is optional. The sample information file is also in CSV format and a “SampleID” column must be included in this file. Each sample id in this column must be the same as the id in the header of the peptide data file (see Peptide Data File for more details). The other columns in the sample information file can contain any additional information that is needed for the particular analysis.

Protein Sequence Database File (Organisms)

The peptide sequences will be aligned to the protein sequences within the selected database. The protein sequences are stored in FASTA formatted files in the “db” folder in the case of MacOS and Linux or in the “app/db” folder for Windows. There are already three FASTA files for Human, Mouse and Rat respectively in the folder. Since the sample data in the testData folder is from rat, if this data set is being used then during data import Rat should be selected under the organism option. If other organisms are being studied the corresponding protein sequence files can be downloaded from the UniProt website: http://www.uniprot.org/uniprot/. If there are some synthetic proteins included in the samples or if the proteins are from multiple organisms, users can edit the FASTA file using the FASTA file format (https://zhanglab.ccmb.med.umich.edu/FASTA/).

Error/Warning Messages during Data Import

Protein Not Found in Database

ProteomicsBrowser searches protein sequences from FASTA files in the db folder. If there is no match for a specific protein within the data set then an error message, “The following proteins that were not found in the database have been removed from the analysis”, will be displayed as shown below in **Figure 2**. In this case, some proteins may have been removed from UniProt or their names may have been changed resulting in the proteins not found in the “database” error. For example, protein CALM_RAT was demerged into P0DP29, P0DP30 and P0DP31 on 2017-05-10. Since the FASTA files in the db folder were downloaded from UniProt on November 12-13, 2017, the CALM_RAT entry could not be found in the database. This challenge can be solved by downloading the newer FASTA file or by adding the needed protein records to the FASTA file that is in the db folder. The proteins that cannot be found in the database file that is in the db folder can be exported to a text file by clicking “Export”.

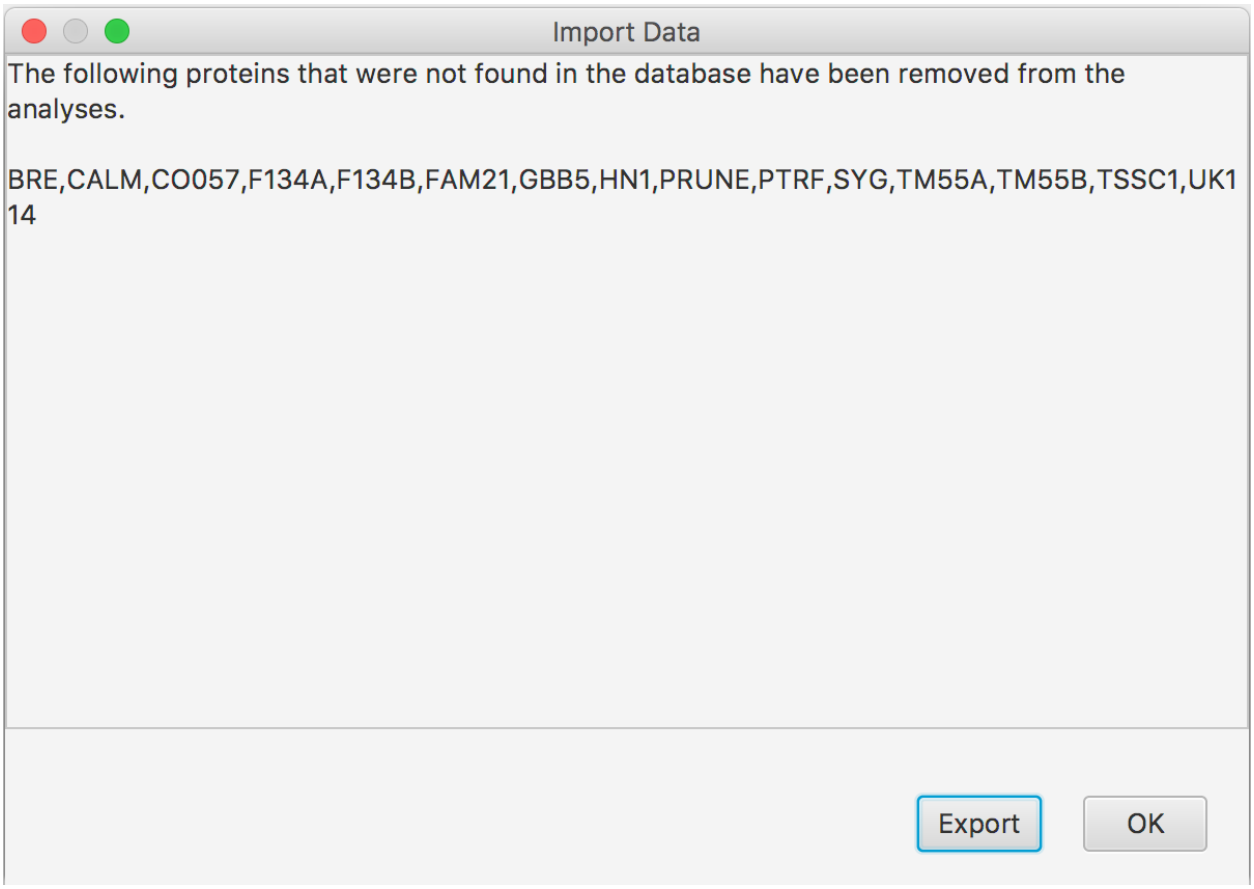


Figure 2

Peptide Sequence Does Not Match

ProteomicsBrowser searches the protein database to locate the start and end point of each peptide within each protein identified in the sample. If a peptide sequence cannot be exactly matched to a sequence within the specified protein, there will be an error message as shown below in **Figure 3**. This error might be caused by a UniProt Update. The unmatched peptide sequence and parent protein name can be exported to a csv file by clicking “Export”.

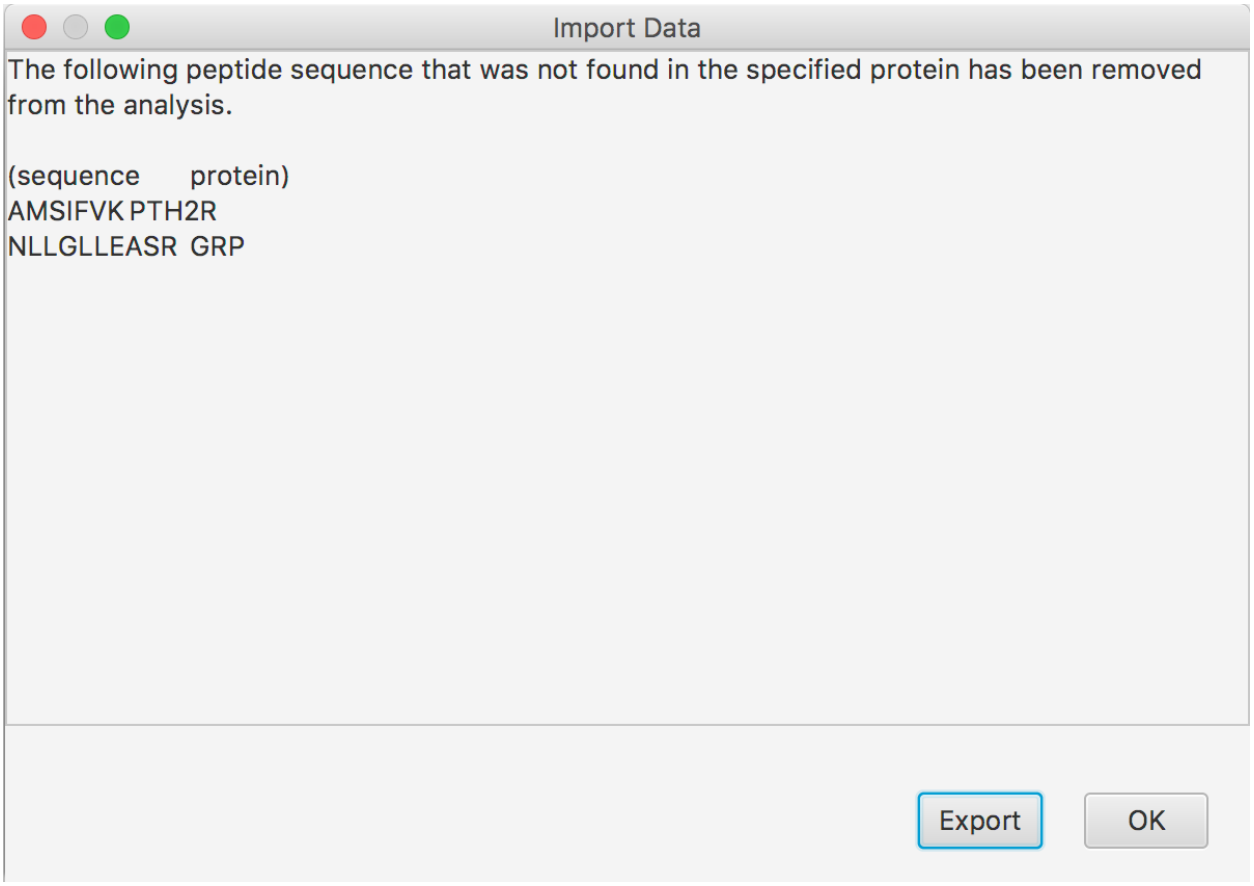


Figure 3

Peptides Mapped to Multiple Proteins

Some peptides might be aligned to more than one protein. For example, as shown below in **Figure 4**, peptide 10003 in the sample test data can be aligned to protein HNRH1_RAT and also to HNRH2_RAT. In this case, whether this peptide is removed from or kept in ProteomicsBrowser depends on the selection that was made during Data Import at “Include Peptides Mapped to Multiple Proteins”. If these peptides are removed, they are **NOT** included in the protein abundance calculations. If these peptides are retained, they are included in each matching protein abundance calculation and they will be displayed differently than other peptides. All of the peptide information can be exported to a csv file by clicking “Export”.

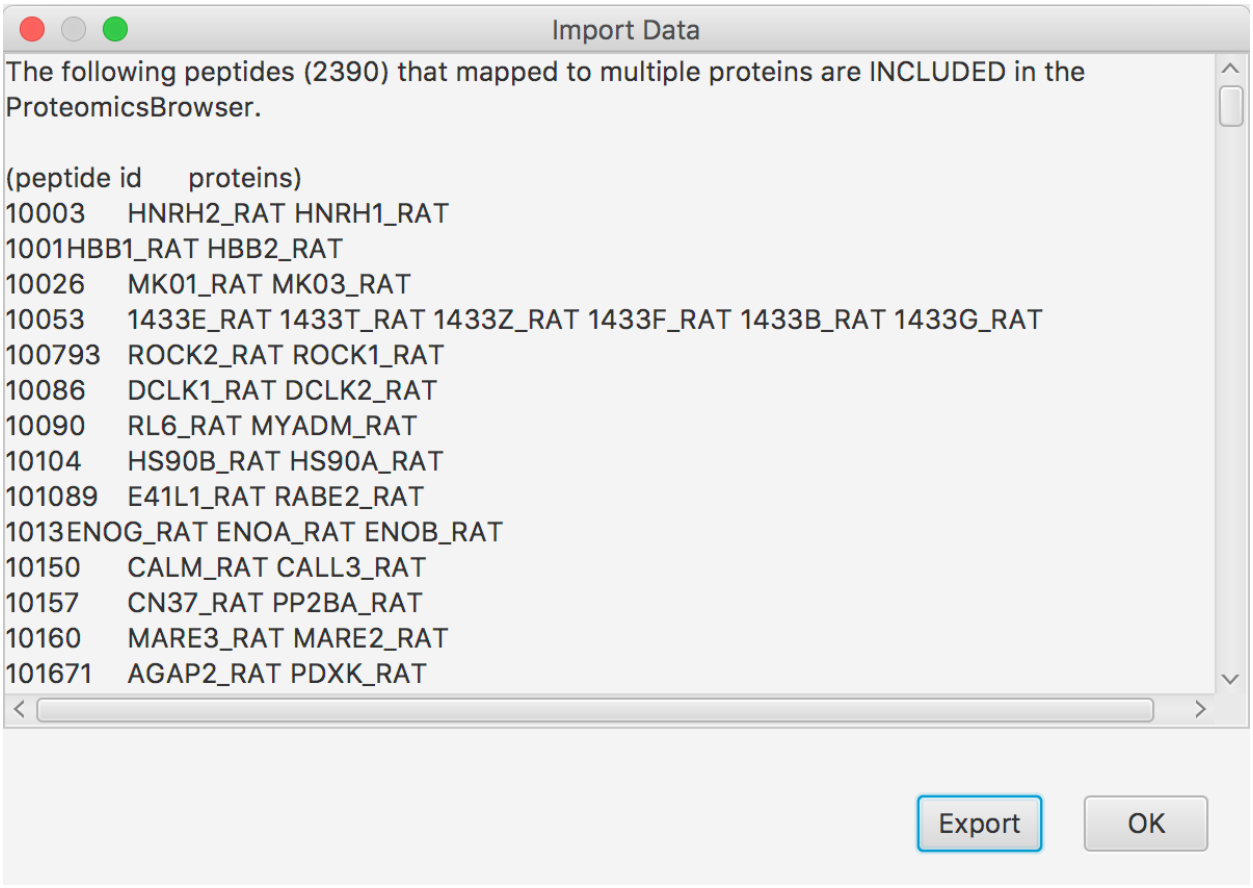
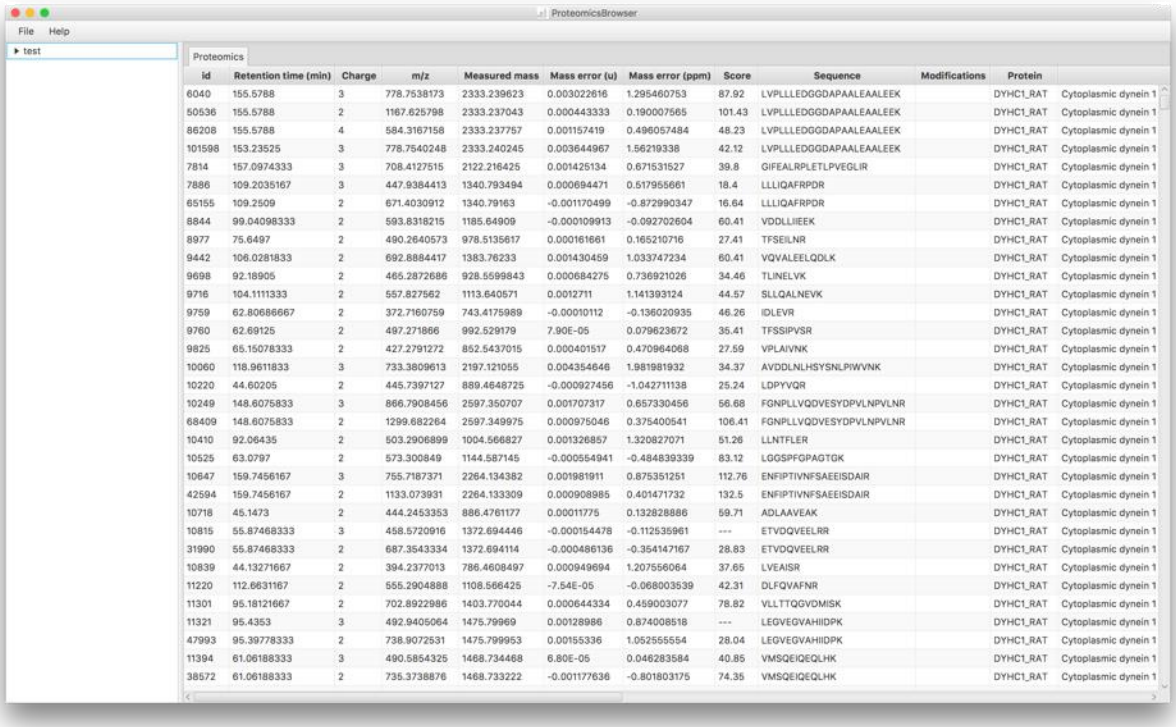

Import Data
The following peptides (2390) that mapped to multiple proteins are INCLUDED in the ProteomicsBrowser.
(peptide id proteins)
10003 HNRH2_RAT HNRH1_RAT
1001HBB1_RAT HBB2_RAT
10026 MK01_RAT MK03_RAT
10053 1433E_RAT 1433T_RAT 1433Z_RAT 1433F_RAT 1433B_RAT 1433G_RAT
100793 ROCK2_RAT ROCK1_RAT
10086 DCLK1_RAT DCLK2_RAT
10090 RL6_RAT MYADM_RAT
10104 HS90B_RAT HS90A_RAT
101089 E41L1_RAT RABE2_RAT
1013ENOG_RAT ENOA_RAT ENOB_RAT
10150 CALM_RAT CALL3_RAT
10157 CN37_RAT PP2BA_RAT
10160 MARE3_RAT MARE2_RAT
101671 AGAP2_RAT PDXK_RAT
Export
OK

Figure 4

General View of ProteomicsBrowser

After importing the peptide data, a window will be displayed that is similar to the one below in **Figure 5**:



id	Retention time (min)	Charge	m/z	Measured mass	Mass error (u)	Mass error (ppm)	Score	Sequence	Modifications	Protein	
6040	155.5788	3	778.7538173	2333.239623	0.003022616	1.295460753	87.92	LVPLLLEDGGDAPAALEAALEEK		DYHC1_RAT	Cytoplasmic dynein 1
50536	155.5788	2	1167.625798	2333.237043	0.000443333	0.190007565	101.43	LVPLLLEDGGDAPAALEAALEEK		DYHC1_RAT	Cytoplasmic dynein 1
86208	155.5788	4	584.3167158	2333.237757	0.001157419	0.496057484	48.23	LVPLLLEDGGDAPAALEAALEEK		DYHC1_RAT	Cytoplasmic dynein 1
101598	153.23525	3	778.7540248	2333.240245	0.003644967	1.56219338	42.12	LVPLLLEDGGDAPAALEAALEEK		DYHC1_RAT	Cytoplasmic dynein 1
7814	157.0974333	3	708.4127515	2122.216425	0.001425134	0.671531527	39.8	GIFEALRPLETLPVEGLIR		DYHC1_RAT	Cytoplasmic dynein 1
7886	109.2035167	3	447.9384413	1340.793494	0.000694471	0.517955661	18.4	LLLIQAFRPDR		DYHC1_RAT	Cytoplasmic dynein 1
65155	109.2509	2	671.4030912	1340.79163	-0.001170499	-0.872990347	16.64	LLLIQAFRPDR		DYHC1_RAT	Cytoplasmic dynein 1
8844	99.04098333	2	593.8318215	1185.64909	-0.000109913	-0.092702604	60.41	VDDLLIIEEK		DYHC1_RAT	Cytoplasmic dynein 1
8977	75.6497	2	490.2640573	978.5135617	0.000161661	0.165210716	27.41	TFSEILNR		DYHC1_RAT	Cytoplasmic dynein 1
9442	106.0281833	2	692.8884417	1383.76233	0.001430459	1.033747234	60.41	VQVALEELQDLK		DYHC1_RAT	Cytoplasmic dynein 1
9698	92.18905	2	465.2872686	928.5599843	0.000684275	0.736921026	34.46	TLINELVK		DYHC1_RAT	Cytoplasmic dynein 1
9716	104.1111333	2	557.827562	1113.640571	0.0012711	1.141393124	44.57	SLLQALNEVK		DYHC1_RAT	Cytoplasmic dynein 1
9759	62.80686667	2	372.7160759	743.4175989	-0.00010112	-0.136020935	46.26	IDLEVR		DYHC1_RAT	Cytoplasmic dynein 1
9760	62.69125	2	497.271866	992.529179	7.90E-05	0.079623672	35.41	TFSSIPVSR		DYHC1_RAT	Cytoplasmic dynein 1
9825	65.15078333	2	427.2791272	852.5437015	0.000401517	0.470964068	27.59	VPLAIVNK		DYHC1_RAT	Cytoplasmic dynein 1
10060	118.9611833	3	733.3809613	2197.121055	0.004354646	1.981981932	34.37	AVDDLNLHSYSNLPIWVNK		DYHC1_RAT	Cytoplasmic dynein 1
10220	44.60205	2	445.7397127	889.4648725	-0.000927456	-1.042711138	25.24	LDPYVQR		DYHC1_RAT	Cytoplasmic dynein 1
10249	148.6075833	3	866.7908456	2597.350707	0.001707317	0.657330456	56.68	FGNPLLVQDVESYDPVLNPVLNR		DYHC1_RAT	Cytoplasmic dynein 1
68409	148.6075833	2	1299.682264	2597.349975	0.000975046	0.375400541	106.41	FGNPLLVQDVESYDPVLNPVLNR		DYHC1_RAT	Cytoplasmic dynein 1
10410	92.06435	2	503.2906899	1004.566827	0.001326857	1.320827071	51.26	LLNTFLER		DYHC1_RAT	Cytoplasmic dynein 1
10525	63.0797	2	573.300849	1144.587145	-0.000554941	-0.484839339	83.12	LGGSPFGPAGTGK		DYHC1_RAT	Cytoplasmic dynein 1
10647	159.7456167	3	755.7187371	2264.134382	0.001981911	0.875351251	112.76	ENFIPTIVNFSAEEISDAIR		DYHC1_RAT	Cytoplasmic dynein 1
42594	159.7456167	2	1133.073931	2264.133309	0.000908985	0.401471732	132.5	ENFIPTIVNFSAEEISDAIR		DYHC1_RAT	Cytoplasmic dynein 1
10718	45.1473	2	444.2453353	886.4761177	0.00011775	0.132828886	59.71	ADLAAVEAK		DYHC1_RAT	Cytoplasmic dynein 1
10815	55.87468333	3	458.5720916	1372.694446	-0.000154478	-0.112535961	---	ETVDQVEELRR		DYHC1_RAT	Cytoplasmic dynein 1
31990	55.87468333	2	687.3543334	1372.694114	-0.000486136	-0.354147167	28.83	ETVDQVEELRR		DYHC1_RAT	Cytoplasmic dynein 1
10839	44.13271667	2	394.2377013	786.4608497	0.000949694	1.207556064	37.65	LVEAISR		DYHC1_RAT	Cytoplasmic dynein 1
11220	112.6631167	2	555.2904888	1108.566425	-7.54E-05	-0.068003539	42.31	DLFQVAFNR		DYHC1_RAT	Cytoplasmic dynein 1
11301	95.18121667	2	702.8922986	1403.770044	0.000644334	0.459003077	78.82	VLLTTQGVDMISK		DYHC1_RAT	Cytoplasmic dynein 1
11321	95.4353	3	492.9405064	1475.79969	0.00128986	0.874008518	---	LEGVEGVAHIIDPK		DYHC1_RAT	Cytoplasmic dynein 1
47993	95.39778333	2	738.9072531	1475.799953	0.00155336	1.052555554	28.04	LEGVEGVAHIIDPK		DYHC1_RAT	Cytoplasmic dynein 1
11394	61.06188333	3	490.5854325	1468.734468	6.80E-05	0.046283584	40.85	VMSQEIQEQLHK		DYHC1_RAT	Cytoplasmic dynein 1
38572	61.06188333	2	735.3738876	1468.733222	-0.001177636	-0.801803175	74.35	VMSQEIQEQLHK		DYHC1_RAT	Cytoplasmic dynein 1

Figure 5

At this point the user can navigate the Browser through the treeview on the left side of the window that is shown in **Figure 6**.

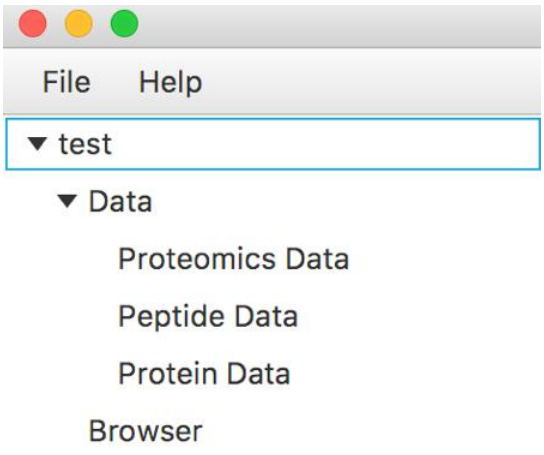


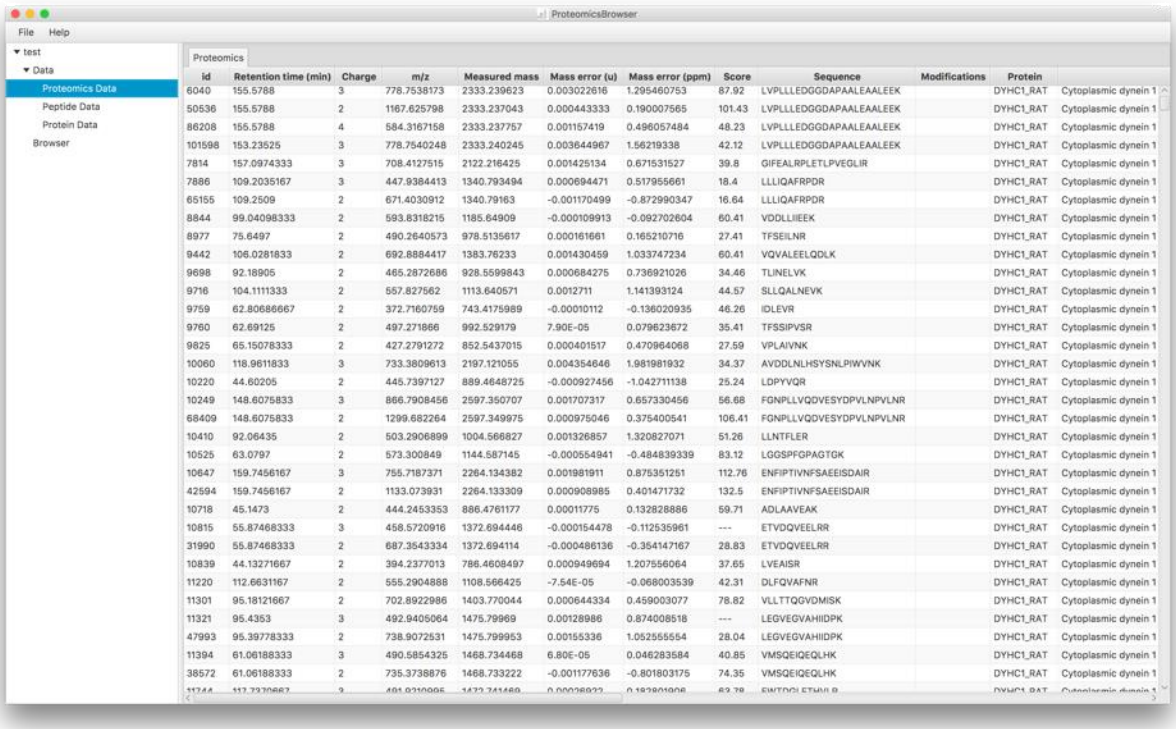
Figure 6

There are two main parts in the treeview, Data and Browser. The Data section contains three sub-parts: Proteomics Data, Peptide Data, and Protein Data. When the user clicks an item in the treeview, a different tab will be displayed on the right side of the window.

Data Tab

Proteomics Data Tab

After selecting “Data/Proteomics Data” from the treeview on the left of the Browser, a table view like that in **Figure 7** will be displayed that shows the imported peptide data.

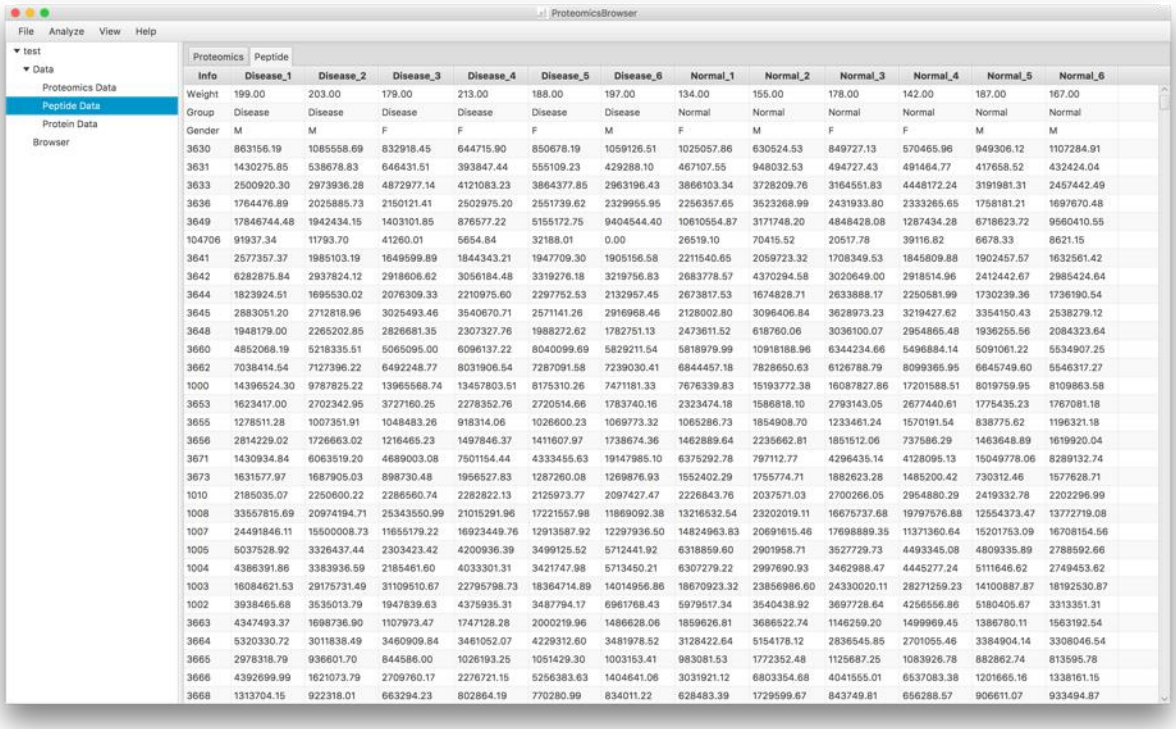


id	Retention time (min)	Charge	m/z	Measured mass	Mass error (u)	Mass error (ppm)	Score	Sequence	Modifications	Protein	
6040	155.5788	3	778.7538173	2333.239623	0.003022616	1.295460753	87.92	LVPLLLEDGGDAPAALEAALEEK		DYHC1_RAT	Cytoplasmic dynein 1
50536	155.5788	2	1167.625798	2333.237043	0.000443333	0.190007565	101.43	LVPLLLEDGGDAPAALEAALEEK		DYHC1_RAT	Cytoplasmic dynein 1
86208	155.5788	4	584.3167158	2333.237757	0.001157419	0.496057484	48.23	LVPLLLEDGGDAPAALEAALEEK		DYHC1_RAT	Cytoplasmic dynein 1
101598	153.23525	3	778.7540248	2333.240245	0.003644967	1.56219338	42.12	LVPLLLEDGGDAPAALEAALEEK		DYHC1_RAT	Cytoplasmic dynein 1
7814	157.0974333	3	708.4127515	2122.216425	0.001425134	0.671531527	39.8	GIFEALRPLETLPVEGLIR		DYHC1_RAT	Cytoplasmic dynein 1
7886	109.2035167	3	447.9384413	1340.793494	0.000694471	0.517955661	18.4	LLLIQAFRPDR		DYHC1_RAT	Cytoplasmic dynein 1
65155	109.2509	2	671.4030912	1340.79163	-0.001170499	-0.872990347	16.64	LLLIQAFRPDR		DYHC1_RAT	Cytoplasmic dynein 1
8844	99.04098333	2	593.8318215	1185.64909	-0.000109913	-0.092702604	60.41	VDDLLIIEEK		DYHC1_RAT	Cytoplasmic dynein 1
8977	75.6497	2	490.2640573	978.5135617	0.000161661	0.165210716	27.41	TFSEILNR		DYHC1_RAT	Cytoplasmic dynein 1
9442	106.0281833	2	692.8884417	1383.76233	0.001430459	1.033747234	60.41	VQVALEELQDLK		DYHC1_RAT	Cytoplasmic dynein 1
9698	92.18905	2	465.2872686	928.5599843	0.000684275	0.736921026	34.46	TLINELVK		DYHC1_RAT	Cytoplasmic dynein 1
9716	104.1111333	2	557.827562	1113.640571	0.0012711	1.141393124	44.57	SLLQALNEVK		DYHC1_RAT	Cytoplasmic dynein 1
9759	62.80686667	2	372.7160759	743.4175989	-0.00010112	-0.136020935	46.26	IDLEVR		DYHC1_RAT	Cytoplasmic dynein 1
9760	62.69125	2	497.271866	992.529179	7.90E-05	0.079623672	35.41	TFSSIPVSR		DYHC1_RAT	Cytoplasmic dynein 1
9825	65.15078333	2	427.2791272	852.5437015	0.000401517	0.470964068	27.59	VPLAIVNK		DYHC1_RAT	Cytoplasmic dynein 1
10060	118.9611833	3	733.3809613	2197.121055	0.004354646	1.981981932	34.37	AVDDLNLHSYSNLPIWVNK		DYHC1_RAT	Cytoplasmic dynein 1
10220	44.60205	2	445.7397127	889.4648725	-0.000927456	-1.042711138	25.24	LDPYVQR		DYHC1_RAT	Cytoplasmic dynein 1
10249	148.6075833	3	866.7908456	2597.350707	0.001707317	0.657330456	56.68	FGNPLLVQDVESYDPVLNPVLNR		DYHC1_RAT	Cytoplasmic dynein 1
68409	148.6075833	2	1299.682264	2597.349975	0.000975046	0.375400541	106.41	FGNPLLVQDVESYDPVLNPVLNR		DYHC1_RAT	Cytoplasmic dynein 1
10410	92.06435	2	503.2906899	1004.566827	0.001326857	1.320827071	51.26	LLNTFLER		DYHC1_RAT	Cytoplasmic dynein 1
10525	63.0797	2	573.300849	1144.587145	-0.000554941	-0.484839339	83.12	LGGSPFGPAGTGK		DYHC1_RAT	Cytoplasmic dynein 1
10647	159.7456167	3	755.7187371	2264.134382	0.001981911	0.875351251	112.76	ENFIPTIVNFSAEEISDAIR		DYHC1_RAT	Cytoplasmic dynein 1
42594	159.7456167	2	1133.073931	2264.133309	0.000908985	0.401471732	132.5	ENFIPTIVNFSAEEISDAIR		DYHC1_RAT	Cytoplasmic dynein 1
10718	45.1473	2	444.2453353	886.4761177	0.00011775	0.132828886	59.71	ADLAAVEAK		DYHC1_RAT	Cytoplasmic dynein 1
10815	55.87468333	3	458.5720916	1372.694446	-0.000154478	-0.112535961	---	ETVDQVEELRR		DYHC1_RAT	Cytoplasmic dynein 1
31990	55.87468333	2	687.3543334	1372.694114	-0.000486136	-0.354147167	28.83	ETVDQVEELRR		DYHC1_RAT	Cytoplasmic dynein 1
10839	44.13271667	2	394.2377013	786.4608497	0.000949694	1.207556064	37.65	LVEAISR		DYHC1_RAT	Cytoplasmic dynein 1
11220	112.6631167	2	555.2904888	1108.566425	-7.54E-05	-0.068003539	42.31	DLFQVAFNR		DYHC1_RAT	Cytoplasmic dynein 1
11301	95.18121667	2	702.8922986	1403.770044	0.000644334	0.459003077	78.82	VLLTTQGVDMISK		DYHC1_RAT	Cytoplasmic dynein 1
11321	95.4353	3	492.9405064	1475.79969	0.00128986	0.874008518	---	LEGVEGVAHIIDPK		DYHC1_RAT	Cytoplasmic dynein 1
47993	95.39778333	2	738.9072531	1475.799953	0.00155336	1.052555554	28.04	LEGVEGVAHIIDPK		DYHC1_RAT	Cytoplasmic dynein 1
11394	61.06188333	3	490.5854325	1468.734468	6.80E-05	0.046283584	40.85	VMSQEIQEQLHK		DYHC1_RAT	Cytoplasmic dynein 1
38572	61.06188333	2	735.3738876	1468.733222	-0.001177636	-0.801803175	74.35	VMSQEIQEQLHK		DYHC1_RAT	Cytoplasmic dynein 1

Figure 7

Peptide Data Tab

After selecting “Data/Peptide Data” from treeview on the left of the Browser, a table will be displayed like that in **Figure 8** that shows peptide abundance/intensity data. The cells in the first three rows in the first column of the table contain the information (Weight, Group, Gender) in the sample information file that is displayed in the corresponding rows for each of the six Disease and six Normal samples in the Test sample set. If additional information is included in the sample information file (e.g. BMI, Heart rate, Age, etc.), this information will be displayed similarly in additional top header rows in the “Data/Peptide Data” view. The next rows after these list the abundance/intensity for each peptide in each of the samples. The first column of these rows contains the peptide id. From the second column onward, each successive column contains the data from the sample whose identity is in the column header.

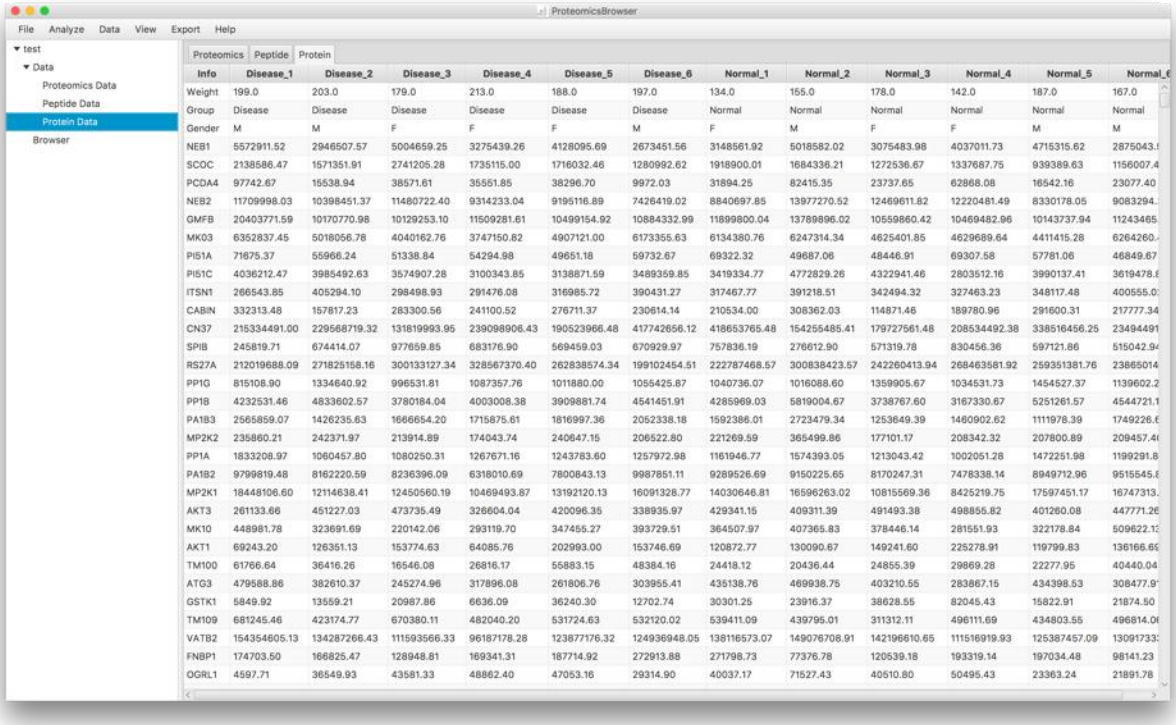


Info	Disease_1	Disease_2	Disease_3	Disease_4	Disease_5	Disease_6	Normal_1	Normal_2	Normal_3	Normal_4	Normal_5	Normal_6
Weight	199.00	203.00	179.00	213.00	188.00	197.00	134.00	155.00	178.00	142.00	187.00	167.00
Group	Disease	Disease	Disease	Disease	Disease	Disease	Normal	Normal	Normal	Normal	Normal	Normal
Gender	M	M	F	F	F	M	F	M	F	F	M	M
3630	863156.19	1085558.69	832918.45	644715.90	850678.19	1059126.51	1025057.86	630524.53	849727.13	570465.96	949306.12	1107284.91
3631	1430275.85	538678.83	646431.51	393847.44	555109.23	429288.10	467107.55	948032.53	494727.43	491464.77	417658.52	432424.04
3633	2500920.30	2973936.28	4872977.14	4121083.23	3864377.85	2963196.43	3866103.34	3728209.76	3164551.83	4448172.24	3191981.31	2457442.49
3636	1764476.89	2025885.73	2150121.41	2502975.20	2551739.62	2329955.95	2256357.65	3523268.99	2431933.80	2333265.65	1758181.21	1697670.48
3649	17846744.48	1942434.15	1403101.85	876577.22	5155172.75	9404544.40	10610554.87	3171748.20	4848428.08	1287434.28	6718623.72	9560410.55
104706	91937.34	11793.70	41260.01	5654.84	32188.01	0.00	26519.10	70415.52	20517.78	39116.82	6678.33	8621.15
3641	2577357.37	1985103.19	1649599.89	1844343.21	1947709.30	1905156.58	2211540.65	2059723.32	1708349.53	1845809.88	1902457.57	1632561.42
3642	6282875.84	2937824.12	2918606.62	3056184.48	3319276.18	3219756.83	2683778.57	4370294.58	3020649.00	2918514.96	2412442.67	2985424.64
3644	1823924.51	1695530.02	2076309.33	2210975.60	2297752.53	2132957.45	2673817.53	1674828.71	2633888.17	2250581.99	1730239.36	1736190.54
3645	2883051.20	2712818.96	3025493.46	3540670.71	2571141.26	2916968.46	2128002.80	3096406.84	3628973.23	3219427.62	3354150.43	2538279.12
3648	1948179.00	2265202.85	2826681.35	2307327.76	1988272.62	1782751.13	2473611.52	618760.06	3036100.07	2954865.48	1936255.56	2084323.64
3660	4852068.19	5218335.51	5065095.00	6096137.22	8040099.69	5829211.54	5818979.99	10918188.96	6344234.66	5496884.14	5091061.22	5534907.25
3662	7038414.54	7127396.22	6492248.77	8031906.54	7287091.58	7239030.41	6844457.18	7828650.63	6126788.79	8099365.95	6645749.60	5546317.27
1000	14396524.30	9787825.22	13965568.74	13457803.51	8175310.26	7471181.33	7676339.83	15193772.38	16087827.86	17201588.51	8019759.95	8109863.58
3653	1623417.00	2702342.95	3727160.25	2278352.76	2720514.66	1783740.16	2323474.18	1586818.10	2793143.05	2677440.61	1775435.23	1767081.18
3655	1278511.28	1007351.91	1048483.26	918314.06	1026600.23	1069773.32	1065286.73	1854908.70	1233461.24	1570191.54	838775.62	1196321.18
3656	2814229.02	1726663.02	1216465.23	1497846.37	1411607.97	1738674.36	1462889.64	2235662.81	1851512.06	737586.29	1463648.89	1619920.04
3671	1430934.84	6063519.20	4689003.08	7501154.44	4333455.63	19147985.10	6375292.78	797112.77	4296435.14	4128095.13	15049778.06	8289132.74
3673	1631577.97	1687905.03	898730.48	1956527.83	1287260.08	1269876.93	1552402.29	1755774.71	1882623.28	1485200.42	730312.46	1577628.71
1010	2185035.07	2250600.22	2286560.74	2282822.13	2125973.77	2097427.47	2226843.76	2037571.03	2700266.05	2954880.29	2419332.78	2202296.99
1008	33557815.69	20974194.71	25343550.99	21015291.96	17221557.98	11869092.38	13216532.54	23202019.11	16675737.68	19797576.88	12554373.47	13772719.08
1007	24491846.11	15500008.73	11655179.22	16923449.76	12913587.92	12297936.50	14824963.83	20691615.46	17698889.35	11371360.64	15201753.09	16708154.56
1005	5037528.92	3326437.44	2303423.42	4200936.39	3499125.52	5712441.92	6318859.60	2901958.71	3527729.73	4493345.08	4809335.89	2788592.66
1004	4386391.86	3383936.59	2185461.60	4033301.31	3421747.98	5713450.21	6307279.22	2997690.93	3462988.47	4445277.24	5111646.62	2749453.62
1003	16084621.53	29175731.49	31109510.67	22795798.73	18364714.89	14014956.86	18670923.32	23856986.60	24330020.11	28271259.23	14100887.87	18192530.87
1002	3938465.68	3535013.79	1947839.63	4375935.31	3487794.17	6961768.43	5979517.34	3540438.92	3697728.64	4256556.86	5180405.67	3313351.31
3663	4347493.37	1698736.90	1107973.47	1747128.28	2000219.96	1486628.06	1859626.81	3686522.74	1146259.20	1499969.45	1386780.11	1563192.54
3664	5320330.72	3011838.49	3460909.84	3461052.07	4229312.60	3481978.52	3128422.64	5154178.12	2836545.85	2701055.46	3384904.14	3308046.54
3665	2978318.79	936601.70	844586.00	1026193.25	1051429.30	1003153.41	983081.53	1772352.48	1125687.25	1083926.78	882862.74	813595.78
3666	4392699.99	1621073.79	2709760.17	2276721.15	5256383.63	1404641.06	3031921.12	6803354.68	4041555.01	6537083.38	1201665.16	1338161.15
3668	1313704.15	922318.01	663294.23	802864.19	770280.99	834011.22	628483.39	1729599.67	843749.81	656288.57	906611.07	933494.87

Figure 8

Protein Data Tab

Similar to the peptide data tab, this table in **Figure 9** shows the abundance of each protein as calculated by the ProteomicsBrowser. As described in more detail in the **Protein Abundance** section below, the Browser calculates protein abundances by using the method (Raw, iBAQ, Top3) that has been selected by the user from the menu “Data/Integration”.



Info	Disease_1	Disease_2	Disease_3	Disease_4	Disease_5	Disease_6	Normal_1	Normal_2	Normal_3	Normal_4	Normal_5	Normal_6
Weight	199.0	203.0	179.0	213.0	188.0	197.0	134.0	155.0	178.0	142.0	187.0	167.0
Group	Disease	Disease	Disease	Disease	Disease	Disease	Normal	Normal	Normal	Normal	Normal	Normal
Gender	M	M	F	F	F	M	F	M	F	F	M	M
NEB1	5572911.52	2946507.57	5004659.25	3275439.26	4128095.69	2673451.56	3148561.92	5018582.02	3075483.98	4037011.73	4715315.62	2875043.5
SCOC	2138586.47	1571351.91	2741205.28	1735115.00	1716032.46	1280992.62	1918900.01	1684336.21	1272536.67	1337687.75	939389.63	1156007.4
PCDA4	97742.67	15538.94	38571.61	35551.85	38296.70	9972.03	31894.25	82415.35	23737.65	62868.08	16542.16	23077.40
NEB2	11709998.03	10398451.37	11480722.40	9314233.04	9195116.89	7426419.02	8840697.85	13977270.52	12469611.82	12220481.49	8330178.05	9083294.1
GMFB	20403771.59	10170770.98	10129253.10	11509281.61	10499154.92	10884332.99	11899800.04	13789896.02	10559860.42	10469482.96	10143737.94	11243465.
MK03	6352837.45	5018056.78	4040162.76	3747150.82	4907121.00	6173355.63	6134380.76	6247314.34	4625401.85	4629689.64	4411415.28	6264260.4
PI51A	71675.37	55966.24	51338.84	54294.98	49651.18	59732.67	69322.32	49687.06	48446.91	69307.58	57781.06	46849.67
PI51C	4036212.47	3985492.63	3574907.28	3100343.85	3138871.59	3489359.85	3419334.77	4772829.26	4322941.46	2803512.16	3990137.41	3619478.8
ITSN1	266543.85	405294.10	298498.93	291476.08	316985.72	390431.27	317467.77	391218.51	342494.32	327463.23	348117.48	400555.0
CABIN	332313.48	157817.23	283300.56	241100.52	276711.37	230614.14	210534.00	308362.03	114871.46	189780.96	291600.31	217777.34
CN37	215334491.00	229568719.32	131819993.95	239098906.43	190523966.48	417742656.12	418653765.48	154255485.41	179727561.48	208534492.38	338516456.25	23494491
SPIB	245819.71	674414.07	977659.85	683176.90	569459.03	670929.97	757836.19	276612.90	571319.78	830456.36	597121.86	515042.94
RS27A	212019688.09	271825158.16	300133127.34	328567370.40	262838574.34	199102454.51	222787468.57	300838423.57	242260413.94	268463581.92	259351381.76	23865014
PP1G	815108.90	1334640.92	996531.81	1087357.76	1011880.00	1055425.87	1040736.07	1016088.60	1359905.67	1034531.73	1454527.37	1139602.2
PP1B	4232531.46	4833602.57	3780184.04	4003008.38	3909881.74	4541451.91	4285969.03	5819004.67	3738767.60	3167330.67	5251261.57	4544721.1
PA1B3	2565859.07	1426235.63	1666654.20	1715875.61	1816997.36	2052338.18	1592386.01	2723479.34	1253649.39	1460902.62	1111978.39	1749226.6
MP2K2	235860.21	242371.97	213914.89	174043.74	240647.15	206522.80	221269.59	365499.86	177101.17	208342.32	207800.89	209457.4
PP1A	1833208.97	1060457.80	1080250.31	1267671.16	1243783.60	1257972.98	1161946.77	1574393.05	1213043.42	1002051.28	1472251.98	1199291.8
PA1B2	9799819.48	8162220.59	8236396.09	6318010.69	7800843.13	9987851.11	9289526.69	9150225.65	8170247.31	7478338.14	8949712.96	9515545.8
MP2K1	18448106.60	12114638.41	12450560.19	10469493.87	13192120.13	16091328.77	14030646.81	16596263.02	10815569.36	8425219.75	17597451.17	16747313.
AKT3	261133.66	451227.03	473735.49	326604.04	420096.35	338935.97	429341.15	409311.39	491493.38	498855.82	401260.08	447771.26
MK10	448981.78	323691.69	220142.06	293119.70	347455.27	393729.51	364507.97	407365.83	378446.14	281551.93	322178.84	509622.13
AKT1	69243.20	126351.13	153774.63	64085.76	202993.00	153746.69	120872.77	130090.67	149241.60	225278.91	119799.83	136166.69
TM100	61766.64	36416.26	16546.08	26816.17	55883.15	48384.16	24418.12	20436.44	24855.39	29869.28	22277.95	40440.04
ATG3	479588.86	382610.37	245274.96	317896.08	261806.76	303955.41	435138.76	469938.75	403210.55	283867.15	434398.53	308477.9
GSTK1	5849.92	13559.21	20987.86	6636.09	36240.30	12702.74	30301.25	23916.37	38628.55	82045.43	15822.91	21874.50
TM109	681245.46	423174.77	670380.11	482040.20	531724.63	532120.02	539411.09	439795.01	311312.11	496111.69	434803.55	496814.06
VATB2	154354605.13	134287266.43	111593566.33	96187178.28	123677176.32	124936948.05	138116573.07	149076708.91	142196610.65	111516919.93	125387457.09	13091733
FNBP1	174703.50	166825.47	128948.81	169341.31	187714.92	272913.88	271798.73	77376.78	120539.18	193319.14	197034.48	98141.23
OGRL1	4597.71	36549.93	43581.33	48862.40	47053.16	29314.90	40037.17	71527.43	40510.80	50495.43	23363.24	21891.78

Figure 9

Browser Tab

When the Browser item is selected in the treeview, a panel will appear on the right of the window that is similar to that shown in **Figure 10** below.

Figure 10

The Browser can only show one protein from one sample at a time. After the user selects the Sample and Protein of interest under Selection Options in the middle of the left side of the screen, the Browser will then show all of the identified peptides in the selected protein.

Figure 11

There are several sections in the Browser as shown above in **Figure 11** for the MTAP2 protein from the Disease 1 sample in the Test Data Set. Each identified peptide is shown in Section 1. The Browser region that is shown can be exported as a figure by using the “Export/Browser Figure” menu option in the Header. Each gray bar in the figure corresponds to one peptide. The depth of the grayscale color of the bar indicates the relative abundance level of each of the peptides in the selected protein from a specified sample. If there are post-translational modifications (PTMs) within the peptide, they are shown in different colors which can be changed by the user from the menu “Edit/Modification Color”. The legend in the top of Section 2 indicates which color has been assigned to which PTM. The gray scale bars at the bottom of Section 2 indicate the relative percentage abundance of each peptide in the selected protein as compared to the most abundant peptide from this same protein in this same sample. The number below the gray scale bars indicates the intensity of the most intense peptide from the selected protein in the selected sample. Hence, if peptide “X” has an intensity of 289,000 while the most intense peptide isolated from the same protein in the same sample has an intensity of 794,000 then the relative % intensity of peptide X = 289,000/794,000 x 100% = 36.3%. The coloration of this peptide “X” would correspond

to the 25-50% gray scale as shown in the legend of Section 2. The slider in Section 3 can be used to quickly browse through different positions in the protein. The sequence of the selected protein is shown at the top of Section 3. The red residue numbers at the beginning (Start 621) and end (End 724) of the sequence being visualized are shown in red font at the bottom left and right of Section 3 respectively. When the mouse cursor is moved in Section 1 over the color-coded peptide bars, the respective peptide information and modifications that are shown in Section 4 will change depending upon the position of the mouse cursor. In Section 4, the peptide information is listed on the left, and the modification information is listed on the right. If the user clicks on a particular peptide, then a popup window will appear that shows the peptide information as listed below in **Figure 12** for that peptide. On the left of the window is the operation panel. The user can zoom in/out of the visible peptide region by using the control in Section 5. In Section 6, the user can choose the modifications that will be visualized in the Browser. In Section 7, the user can navigate to any modification position of interest that has been selected in Section 6. In Section 8, the user can select the sample and protein of interest that will be displayed in the large Browser window on the right. Section 9 shows the options for peptide combination, please see Peptide Combination for additional details.

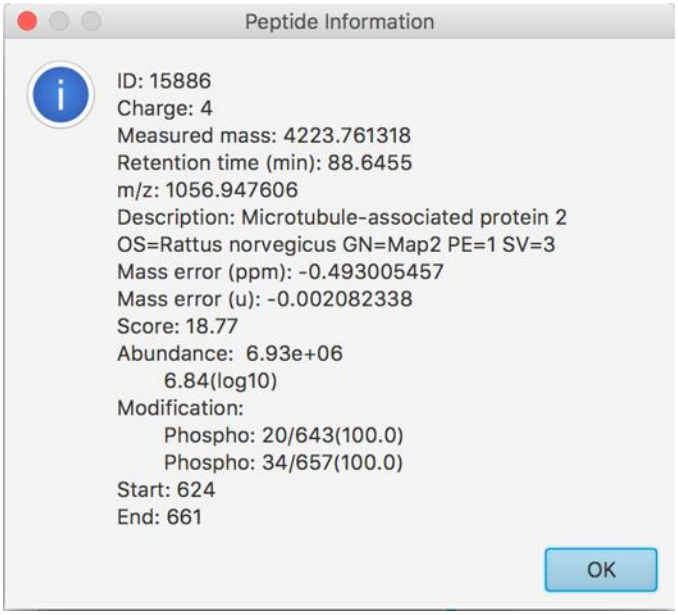


Figure 12

PTMs with Unassigned Sites

For some PTMs such as phosphorylation the exact modification site may be unknown. In this case the modification will be shown with a twill pattern instead of a solid bar as shown in **Figure 13** below for the MTAP2 protein from the Disease 1 sample:

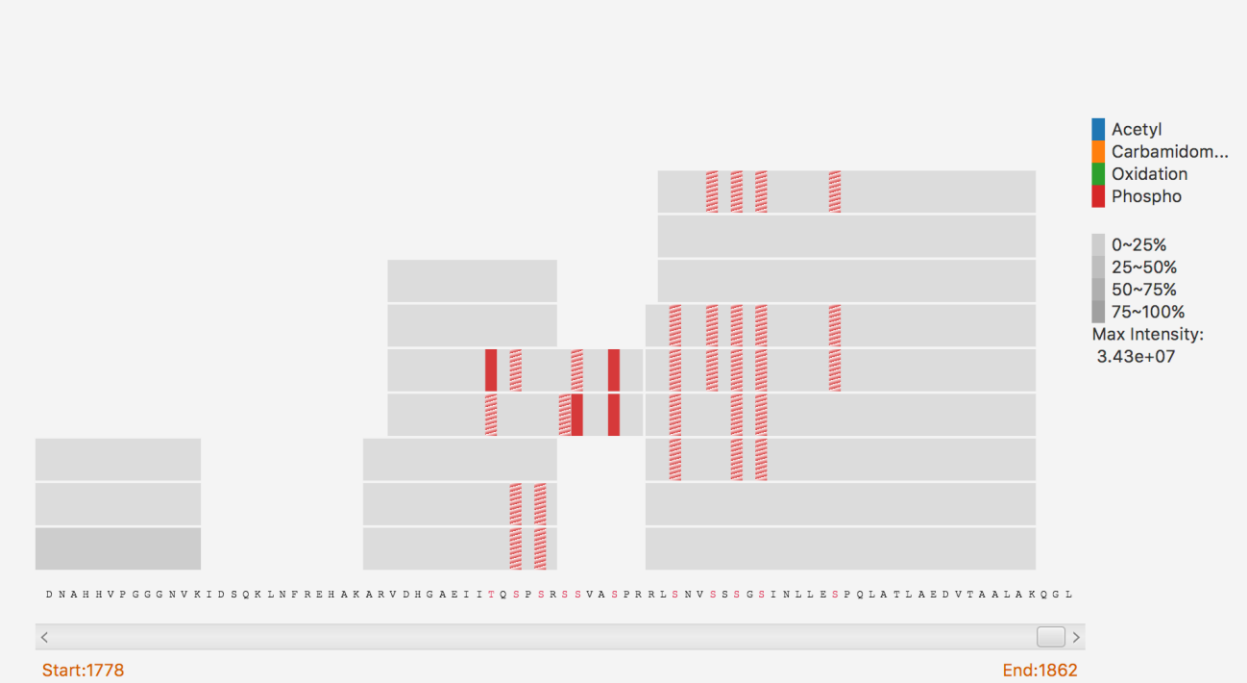


Figure 13

Peptides with Multiple Matches in a Protein

Some peptides may be mapped to multiple positons in a protein. In this case, the peptides are shown with horizontal stripes instead of a solid bar as shown in **Figure 14** below. When the mouse cursor is positioned over one of these peptides, a number is shown in the peptide information box that alerts the user of the number of times this peptide has been matched to previous sequences in the selected protein. Hence, the number 2 in the red Multiple Match Number box in **Figure 14** indicates that this is the second match to this point in the MAP6 protein sequence for this peptide in the Disease 1 test sample. Should this peptide also match to another position that occurs later in the sequence the Multiple Match Number will increase to 3 for the next match, and it will continue to increment by one for each successive match in the overall protein sequence.

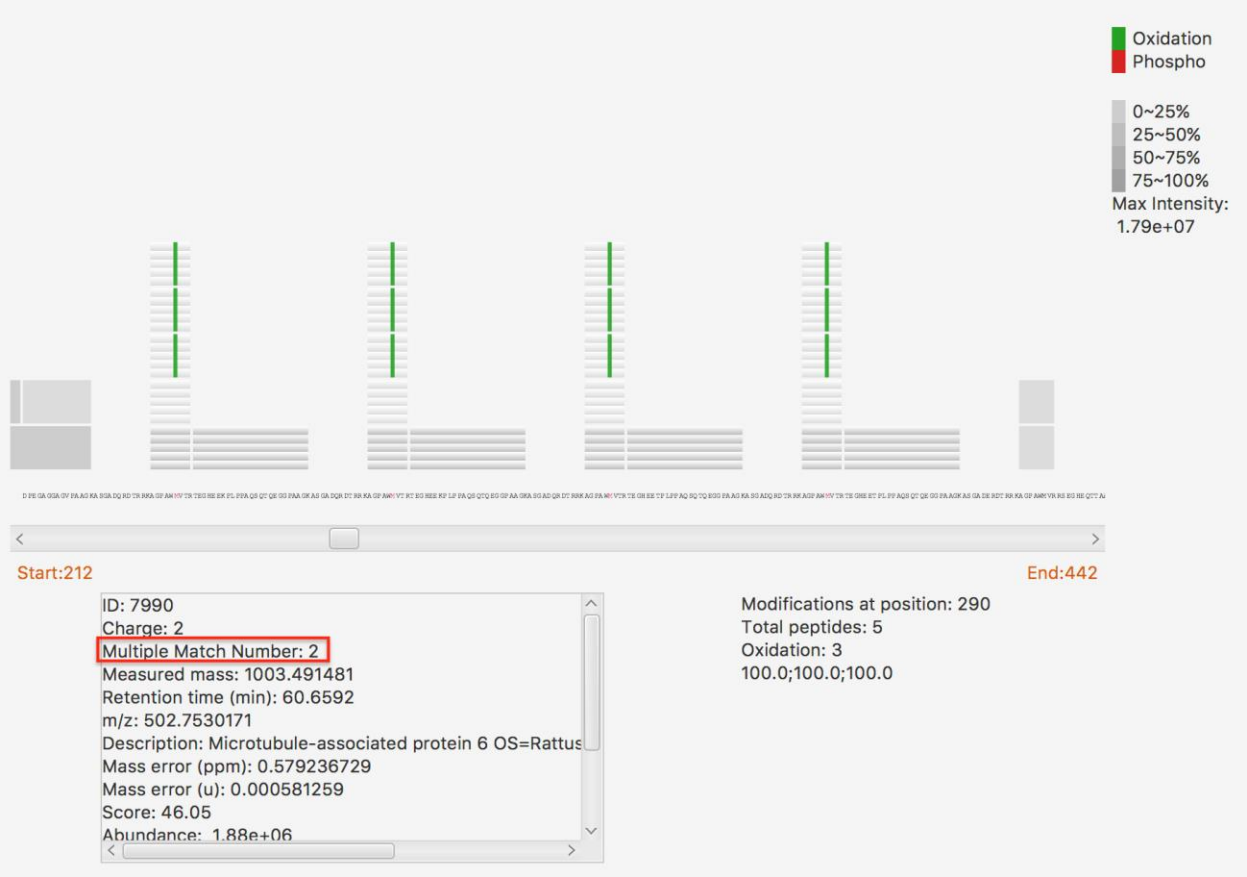


Figure 14

Peptides Mapped to Multiple Proteins

If the user chooses “Include Peptides Mapped to Multiple Proteins” during Data Import, then those peptides that are mapped to multiple proteins will be included in the ProteomicsBrowser window and calculations. These peptides can be easily spotted in the Browser window by the solid box borders around the edges of their bars. **Figure 15** below shows protein 1433B in the Test sample data set. It is evident from this figure that many peptides that are shown can be mapped to proteins other than 1433B. For example, the “Other Candidate Protein” entry that is highlighted by a red box in the peptide information field in **Figure 15** indicates that peptide 2871, which spans residues 21-29 and is represented by the second bar from the bottom, can be mapped to the 1433B, 1433F, and 1433G proteins.

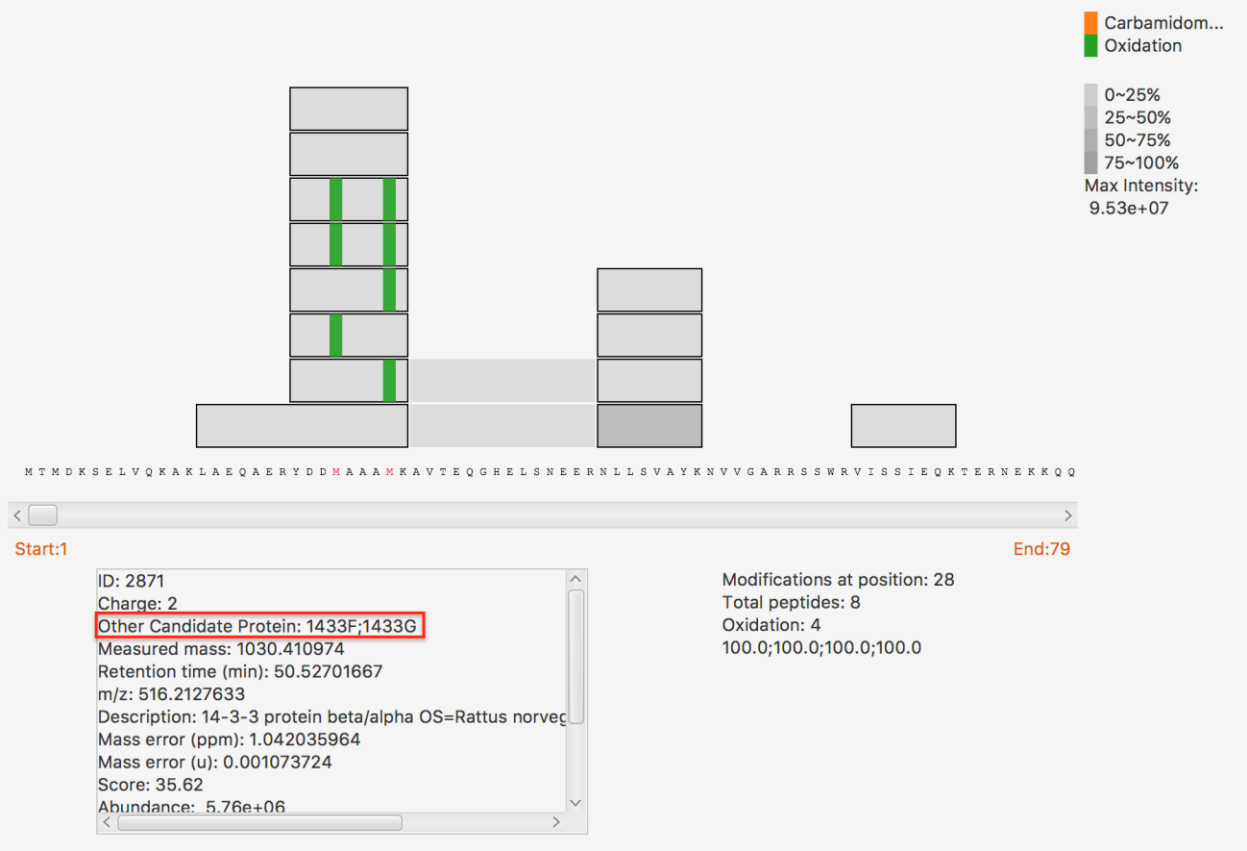


Figure 15

If the user chooses to not include peptides in the analysis that can be mapped to multiple proteins (i.e., when importing the data into the ProteomicsBrowser), these peptides will be removed from protein 1433B as shown below in **Figure 16**. In this case, all peptides that can be mapped to multiple proteins have been removed.

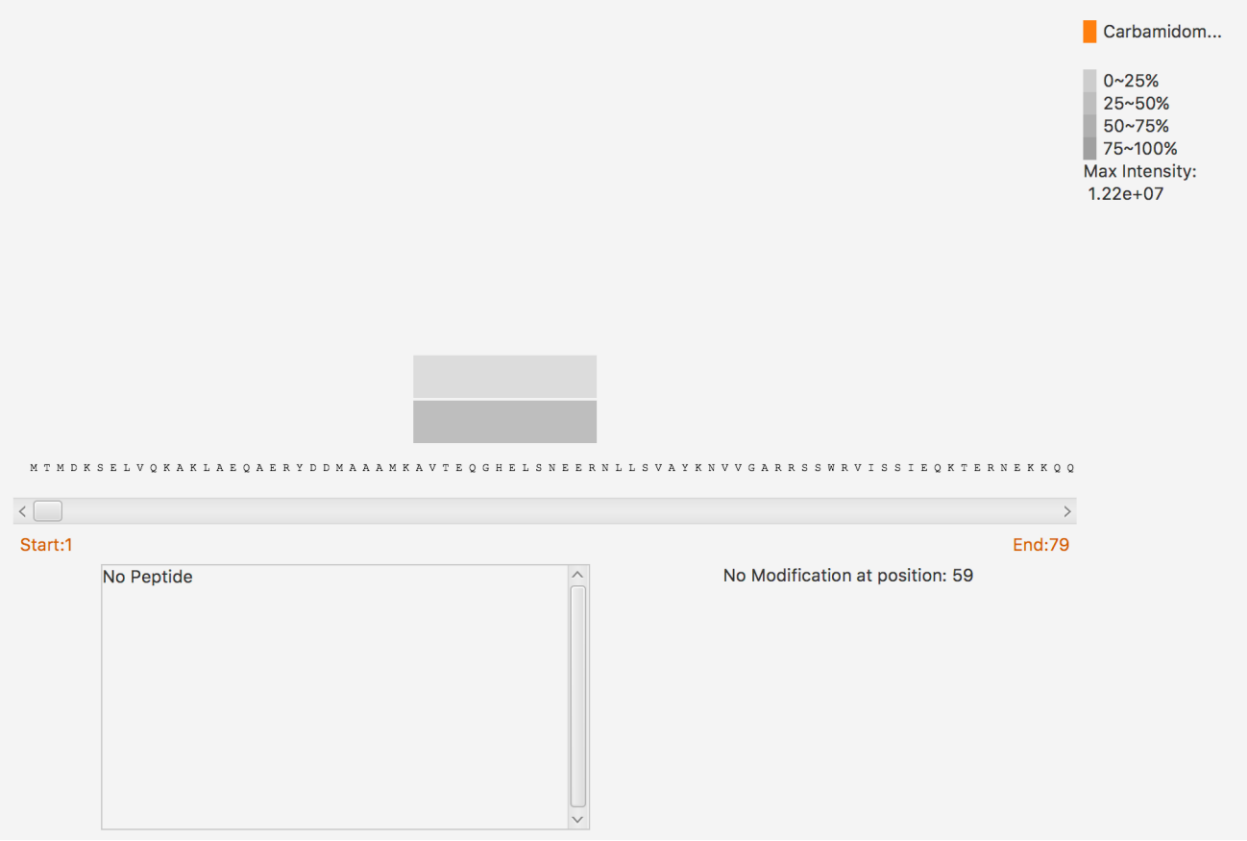


Figure 16

Functions in ProteomicsBrowser

Peptide Combination and Quantify PTM

Peptide Combination

Clicking “Yes” in the Peptide Combination Section on the bottom left of the ProteomicsBrowser window (i.e., **Figure 17** below and Section 9 in **Figure 11** above), enables several criteria that can be selected to combine peptide ions in the selected protein.

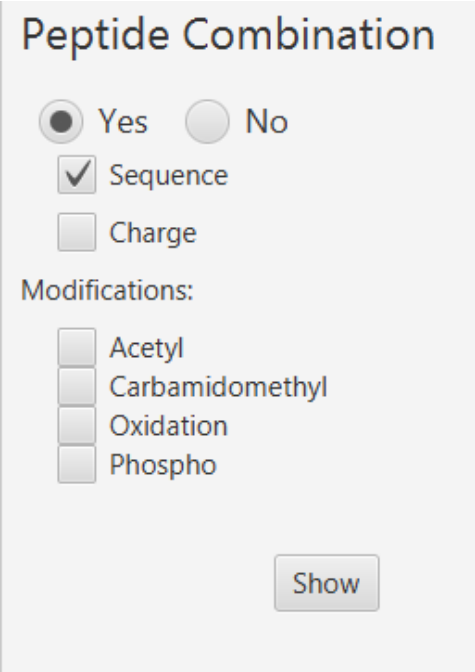


Figure 17

Since the criterion “Sequence” is always selected, only those peptides with the same sequence can be combined together. “Sequence” combines and sums all ions that have the same sequence *regardless* of their charge and/or post-translational modifications. If additional criteria are selected, then those peptides with the same sequence and the same additional selected criteria will be combined together. Thus, if Sequence, Charge, and Oxidation are selected then those peptides with the same sequence, charge, and oxidized residues will be combined together after clicking “Show”.

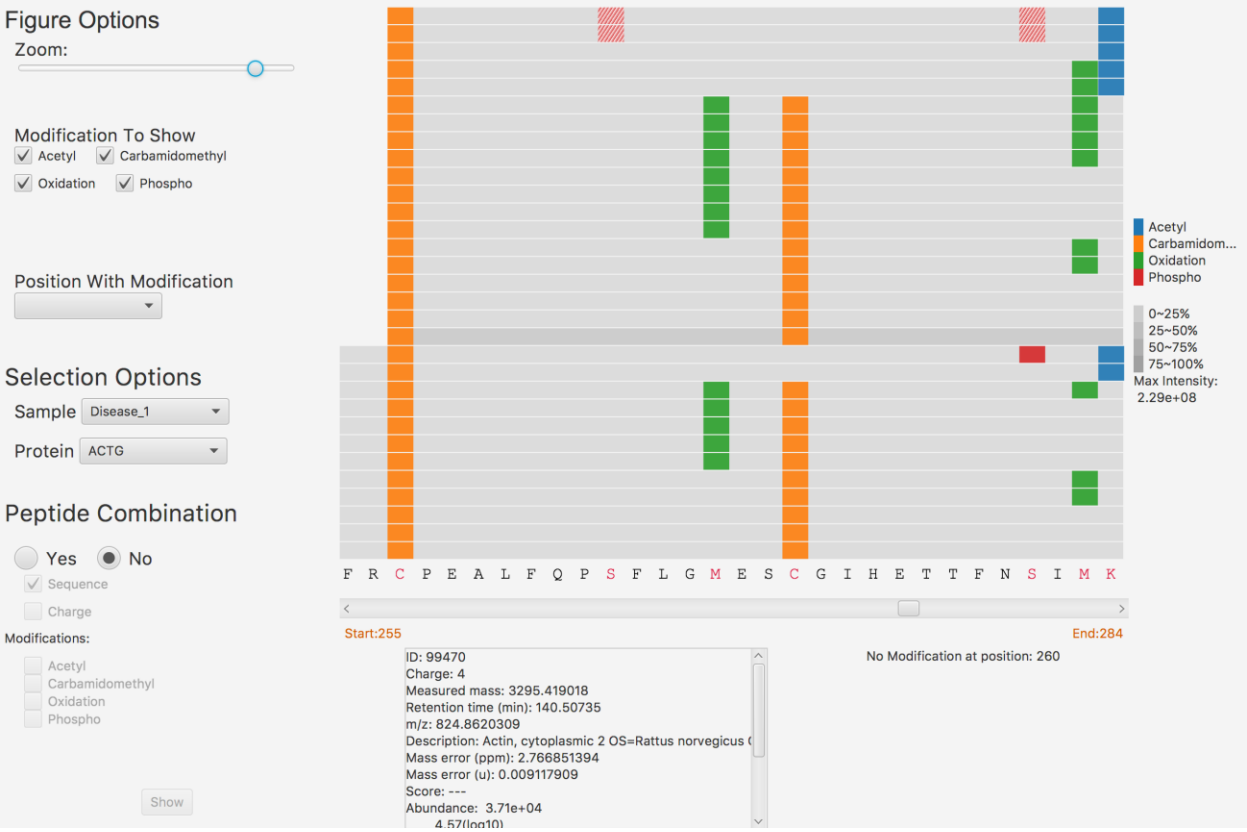


Figure 18

For example, **Figure 18** shows a close-up view of the 31 peptide ions that overlap residues 255 – 284 in the ACTG protein from the Disease 1 test sample. Clicking “Yes” in the Peptide Combination panel and then “Show” will combine together the 31 peptide ions into only the two groups of peptide ions shown in **Figure 19** that each have the same sequences.

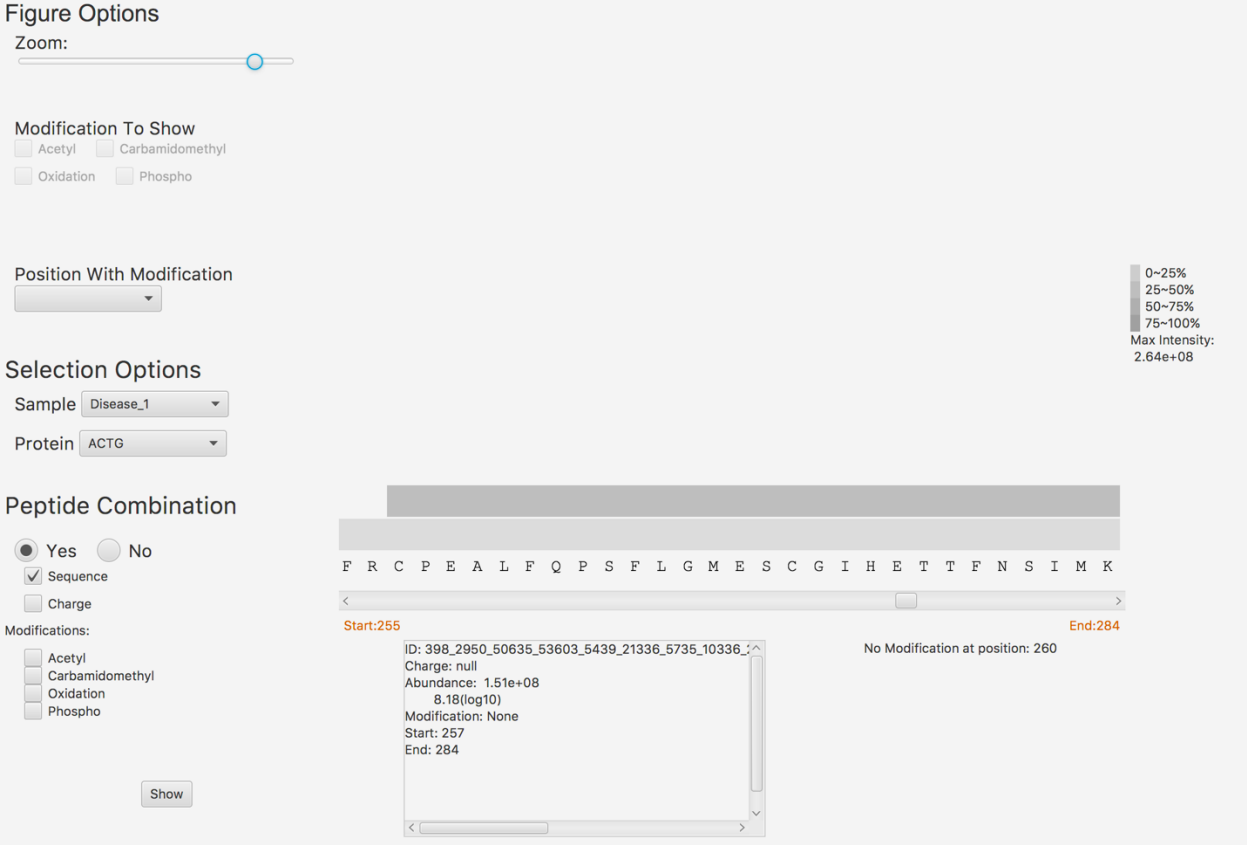


Figure 19

It is apparent from **Figure 19** that the overall yield of the combined 257-284 peptide ions that are in the upper, darker bar is greater than that of the combined 255-284 peptide ions that result from incomplete cleavage at Arg 256. Indeed, the ratio of the ProteomicsBrowser summed intensities (1.51e+08/1.79e+07) indicates that the relative yield of the 255-284/257-284 peptide ions is 8.4. Since the "Sequence" function combines together all peptides with the same sequence *regardless of their PTMs*, users analyzing site specific PTM quantification should *not* use this function. Selecting "Charge" and then clicking "Show" will combine together the 31 peptide ions into the five groups of peptide ions shown in **Figure 20** below that each have the same sequences *and* charges.

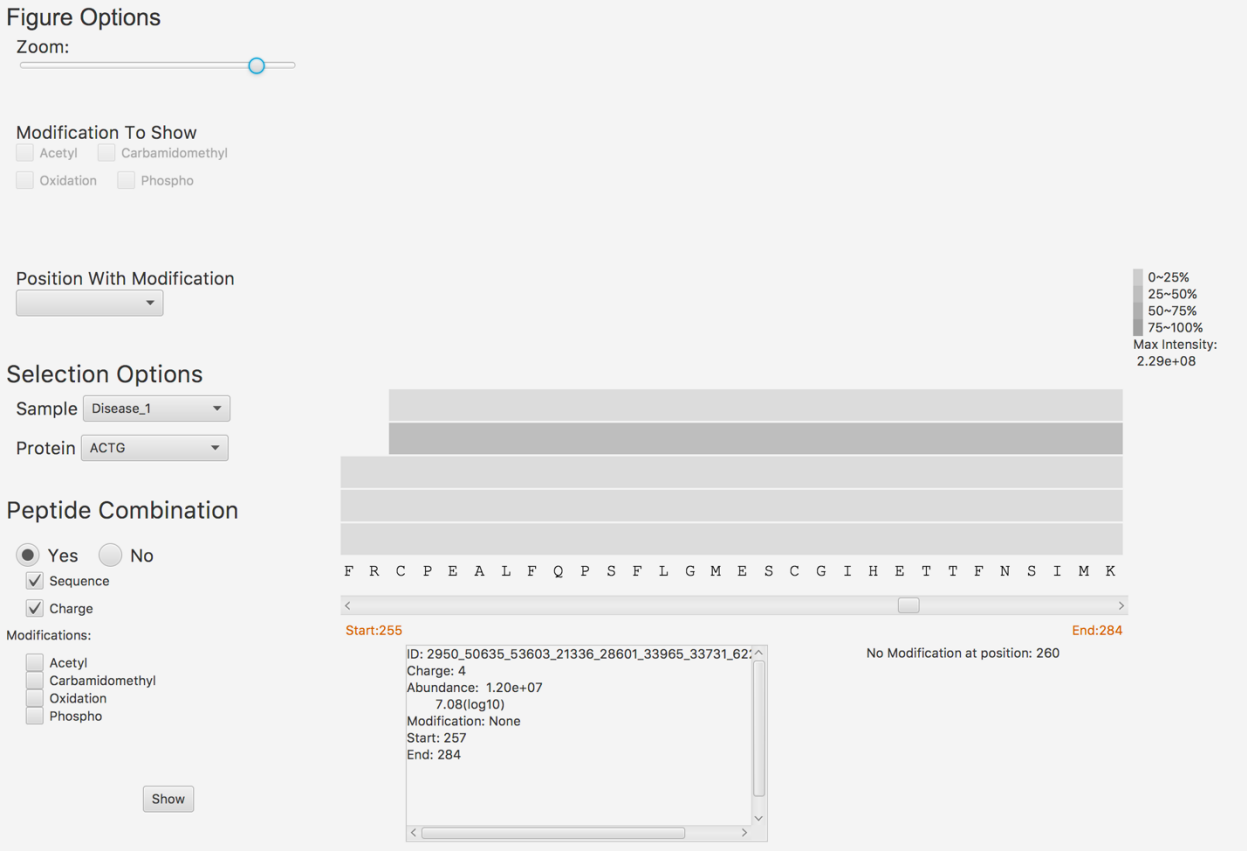


Figure 20

As an example, the top bar in **Figure 20** contains nine peptide ions that span residues 257-284 and that all have a charge of +4. After carrying out peptide combination, the peptide id will be a combination of all the id's of the combined peptides. If peptides A, B and C are combined together, the new peptide id would be A_B_C. Hence, in the above example, the peptide id of the top bar, 2950_50635_53603_21336_28601_33965_33731_62236_99470, is comprised of the peptide ids of the nine combined peptide ions represented by this bar. Finally, selecting "Oxidation" and clicking "Show" results in combining together the 31 peptide ions into the 16 groups of peptide ions depicted in **Figure 21** below that each have the same sequence, charge**,** and oxidized methionine residues.

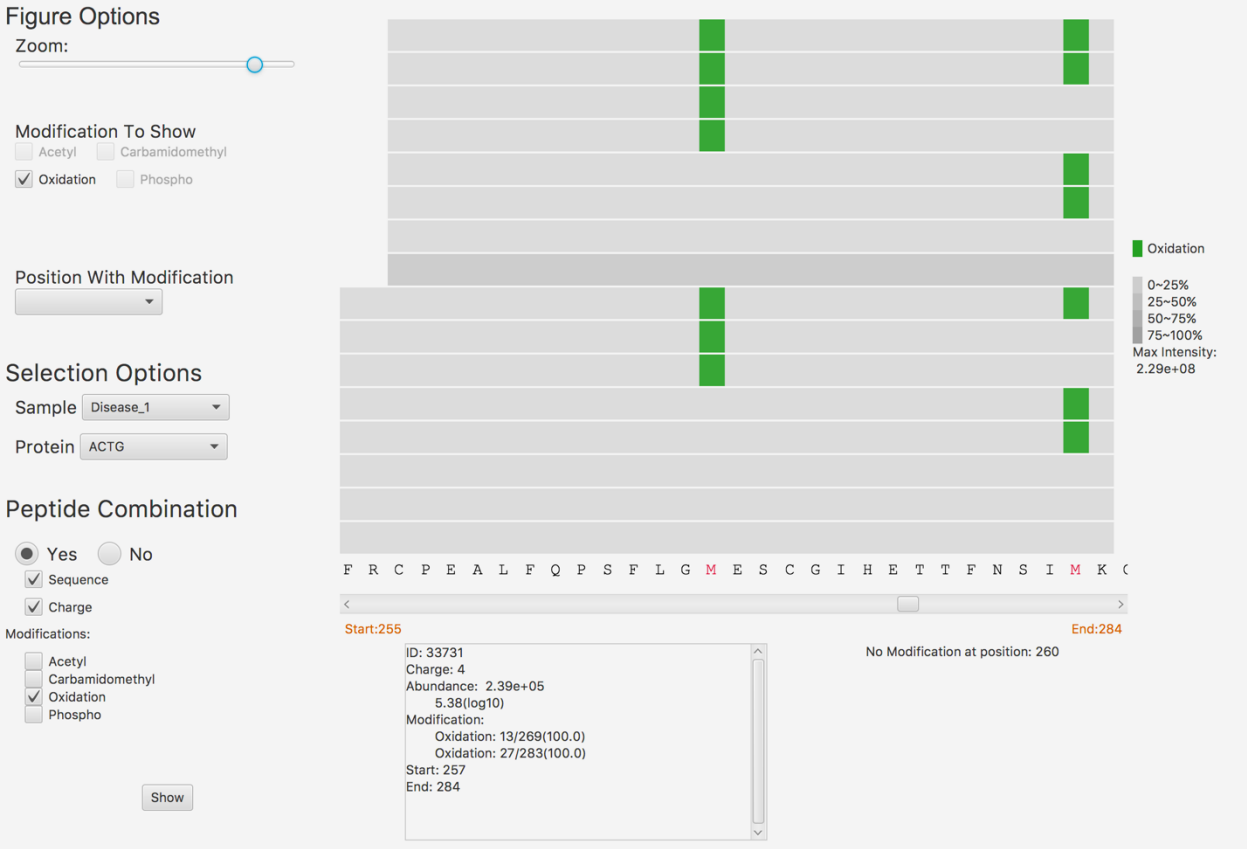


Figure 21

As an example, the top bar in **Figure 21** represents the only peptide ion (#33731) that spans residues 257-284, has a charge of +4, and that also has oxidized methionines at residues 269 and 283 in the ACTG protein from the Disease 1 test sample.

Quantify PTM

This feature is unique in that it only allows the user to select a single position that has a PTM. Hence, this function is only available when the user selects a position under “Position With Modification” as shown in **Figure 22**. The goal of this feature is to provide an easy way to visualize and to relatively quantify the extent of a selected PTM modification at any *single* residue of interest in any protein of interest. This feature thus combines and sums *all* peptides that contain the selected residue in the selected protein into one group of peptides that do and into a second group of peptides that do not contain the PTM of interest. Thus, this feature always will result in the Browser showing just two horizontal bars that extend to the most N-terminal and C-terminal positions of any of the peptides that overlap the modified residue of interest. To use this feature the user selects the residue of interest in the “Position with Modification” box, clicks “Yes” under “Peptide Combination”, de-selects the “Sequence” box, and then clicks “Show” in **Figure 22**. A new window will then open that shows the results. For this function to work properly the user must **unselect** the “Sequence” criterion before clicking the “Show” button. Otherwise, the peptides will be combined based on their having the same sequence as described in “Peptide Combination”. In the example shown in **Figures 22** and **23**, the selected residue is Lysine 284 in the ACTG protein from the Disease 1 sample. Unselecting “Sequence” and clicking “Show” opens the new window depicted below in **Figure 23.**

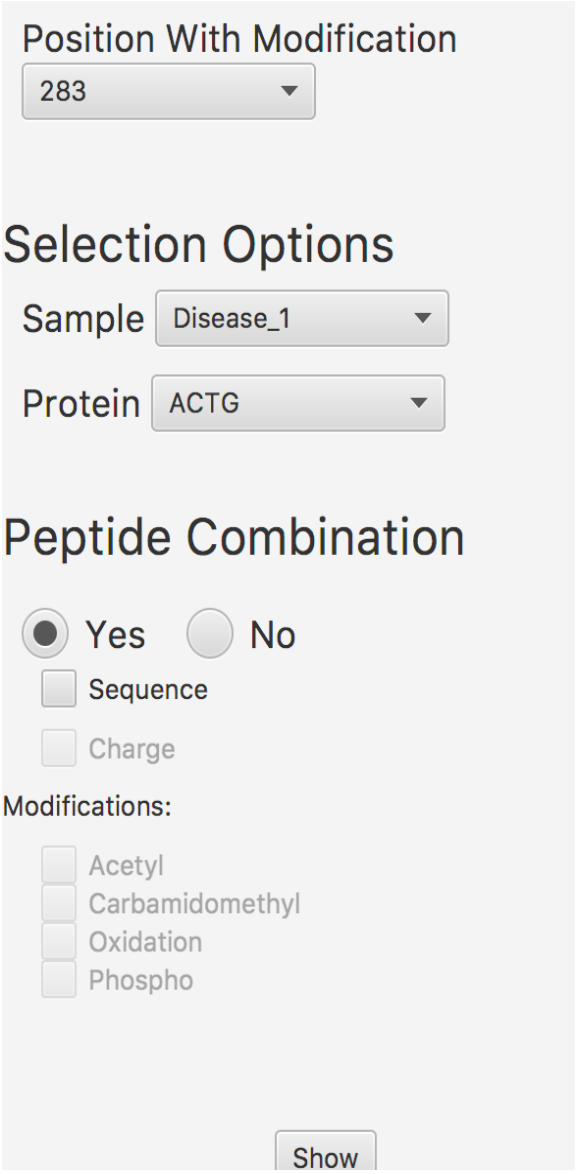


Figure 22

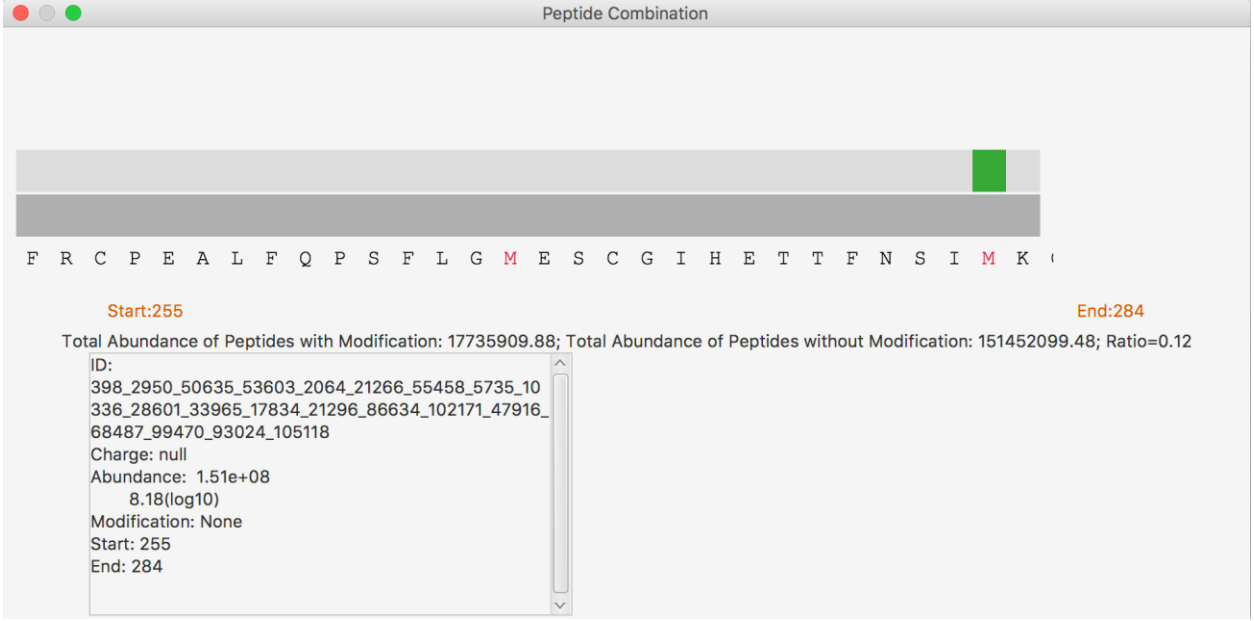


Figure 23

The bottom bar in **Figure 23,** which is where the cursor was resting when **Figure 23** was made, represents the 24 peptide ions that span residues 255-284 that do *not* have an acetylated Lysine at residue 284 in the ACTG protein from the Disease 1 test sample. The top bar in **Figure 23** represents the 7 peptide ions that span residues 255-284 that *do* have an acetylated Lysine at residue 284 in the ACTG protein from the Disease 1 test sample. As indicated in **Figure 23**, the combined abundance of the 7 acetylated peptide ions is 3,163,101 while the combined abundance of the 24 non-acetylated peptides is 166,024,908. Thus the ratio of the acetylated to the non-acetylated peptide ions is 0.02 which means that about 2% of Lysine 284 in the ACTG protein from the Disease 1 test sample is acetylated.

Export Sequences Containing PTMs

When the user is in the Data/Protein Data Tab, it is possible to export all peptide sequences that contain selected PTMs (Menu: Export/Export Sequences Containing PTMs). There are two available options, 1) export sequences to a text file or 2) show the same information in a figure.

Export Sequences Containing PTMs to a Text File

After clicking “Export/Export Sequences Containing PTMs/Export Text File”, the dialogue box shown below in **Figure 24** will appear:

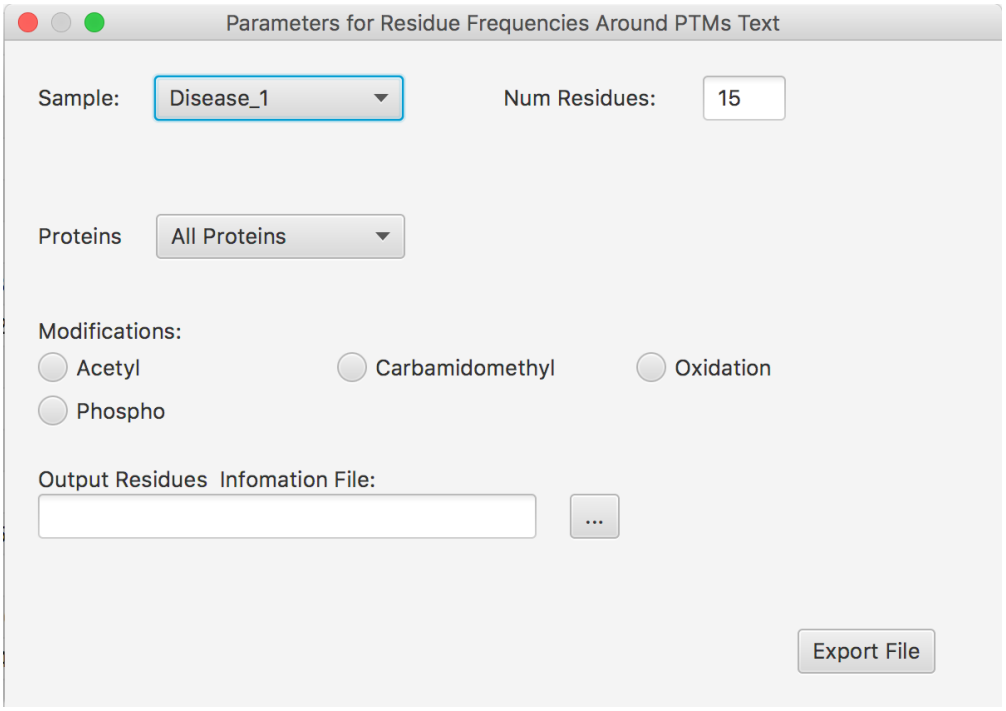


Figure 24

The “Sample” box allows the user to select the id of the sample from which the sequences will be output. At “Num Residues”, the user inputs a positive integer ***n*** that determines the number of amino acid residues that will be output on both sides of the selected modification. The default choice in the “Protein” box is “All Proteins” which results in exporting sequences around the selected PTMs from all proteins in the sample. If, instead, the “Select Proteins” option in the dropdown menu is selected then the user will have the option of importing a text file that lists the names of the proteins that are to be included in the analyses. In this case, one protein should be listed per line as in the following example:

CO3

CPSM

PYC

…

After selecting the modification type(s), the user then needs to select a file in which to save the output by clicking the "…" button. Then click the "Export File" button to export the modified sequences to the selected file. The following provides an example of the comma-separated value (csv) output file format:

KLRGEDGESECVINYVEK [11]Carbamidomethyl(C) AL1L1
QLLVRKLRGEDGESECVINYVEKAVKKLTLQ 404 389 419

KLRGEDGESECVINYVEK [11]Carbamidomethyl(C) AL1L1
QLLVRKLRGEDGESECVINYVEKAVKKLTLQ 404 389 419

GEDGESECVINYVEK [8]Carbamidomethyl(C) AL1L1 QLLVRKLRGEDGESECVINYVEKAVKKLTLQ
404 389 419

IAVIGQSLFGQEVYCQLRK [15]Carbamidomethyl(C) AL1L1
KIAVIGQSLFGQEVYCQLRKEGHEVVGVFTI 17 2 32

As indicated above, there are 7 columns. The first column contains the sequence of the peptide containing the selected modification. The second column contains the modification type and the position in the peptide. The third column contains the protein name. The fourth column contains the sequence from the -***n*** amino acid residue to +***n*** amino acid residue based on the modified position. The last three columns list the modification position, start and end position of the sequence in column 4 in the protein. In the example above, ***n***=15.

Export Sequences Containing PTMs to a Figure

After clicking "Export/Export Sequences Containing PTMs/Show Figure", a dialogue box will appear that is similar to the one below in **Figure 25**.

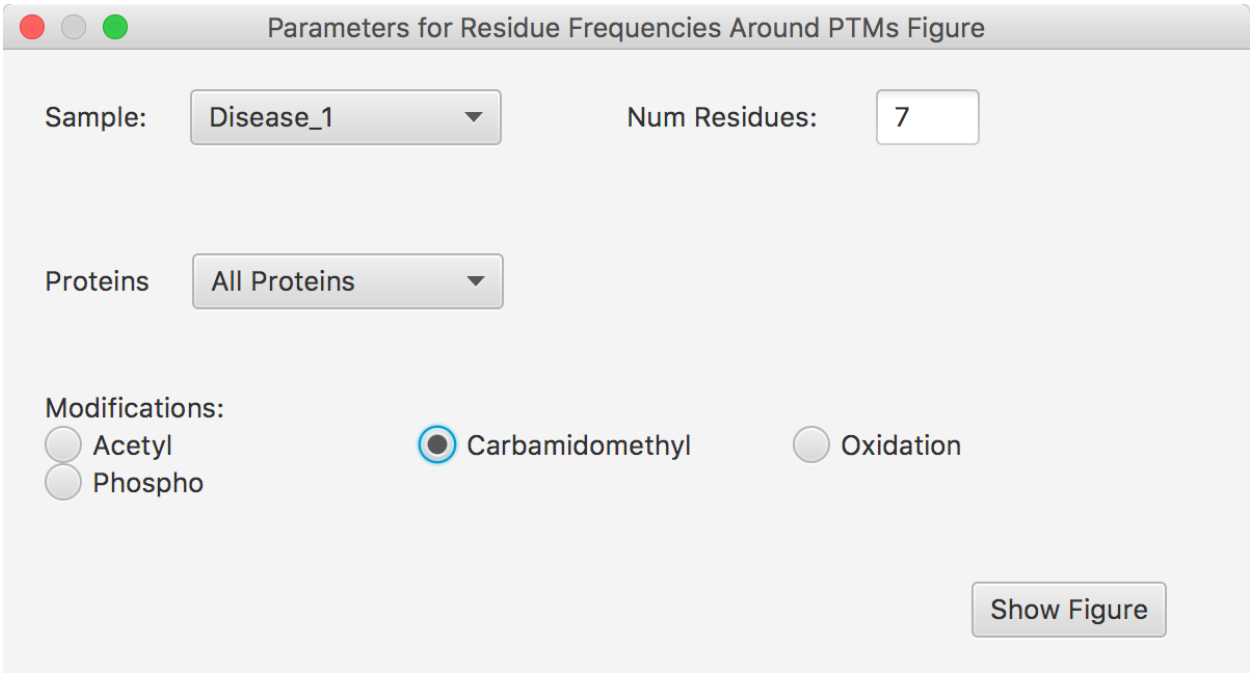


Figure 25

After selecting all of the parameters and clicking “Show Figure”, a window will appear that is similar to the one shown below in **Figure 26**:

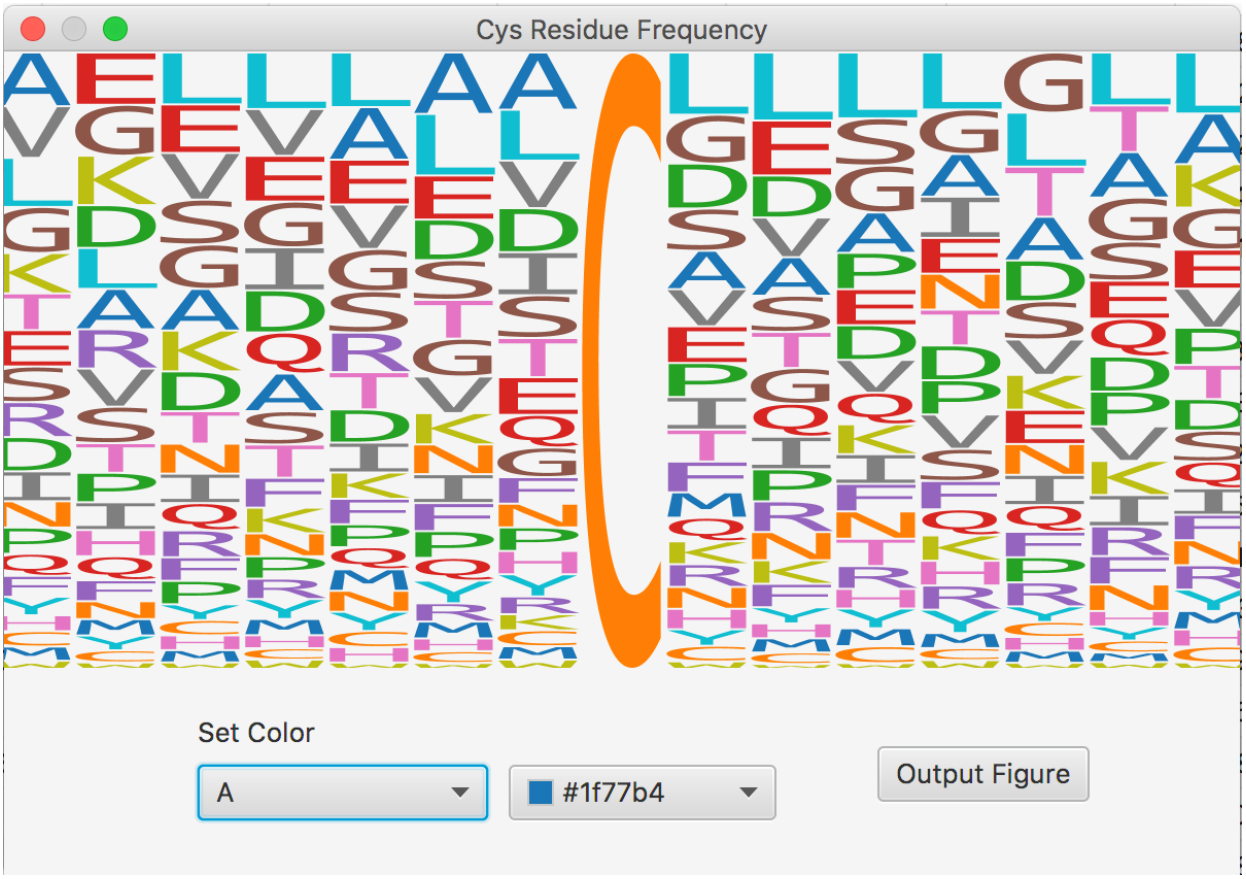


Figure 26

The height of each letter in **Figure 26** is proportional to the corresponding amino acid residue frequency in the selected sequences. The user can also choose the color of each amino acid and output the figure by clicking the “Output Figure” button. The algorithm that is used to generate figures like the above is based on the BlockLogo visualization scheme described by Olsen et al, (2013).

Export Modification Information of a Protein

In the Browser tab, the user can click “Export/Modification Info” in the menu to export all of the modification information for the protein selected in the Browser.

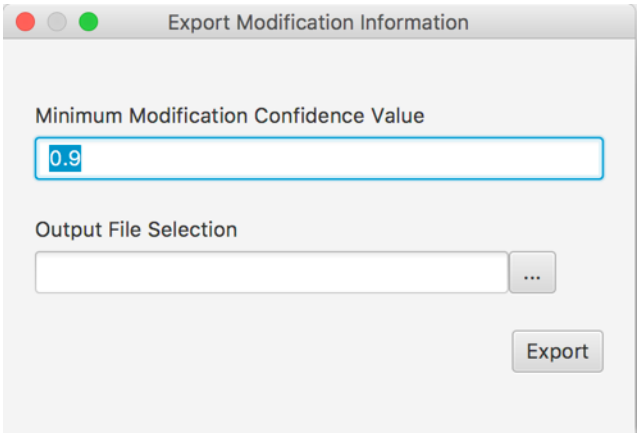


Figure 27

In the above dialogue box in **Figure 27**, the user needs to set two parameters. The first parameter is the “Minimum Modification Confidence Value”, which ranges from 0 to 1 (i.e., 100% confidence). Any modifications with confidence values less than this cutoff will not be shown in the output. **If the cutoff is set at 0**, then all modifications will be exported including “NAs”. Please see the modification format in the Peptide Data File for a detailed description. The second setting that needs to be made is to select the output file by clicking the “…” button. After clicking the “Export” button, a file will be generated that has the following format:

Position:650	TotalPep:13	Oxidation:1
Position:656	TotalPep:13	Oxidation:2
Position:679	TotalPep:13	Oxidation:2
Position:697	TotalPep:5	Carbamidomethyl:5
Position:708	TotalPep:5	Oxidation:2
Position:711	TotalPep:5	Carbamidomethyl:5

The first column shows the modification position. The second column gives the total number of peptides that span this position and the last column gives the modification type and the number of peptides that contain this modification.

Peptide Filter

In the “Browser” tab, the user can filter out some peptides in a selected protein by clicking “Data Filter/Peptides” in the menu.

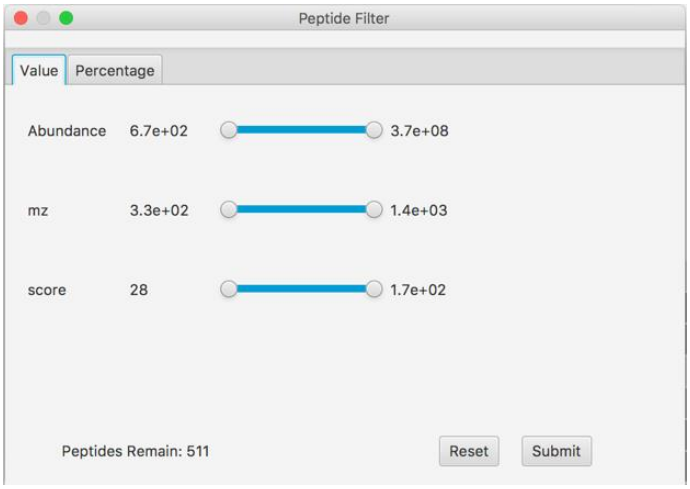


Figure 28

As an example, the above dialogue box in **Figure 28** provides options for filtering out peptides according to three different criteria (abundance, m/z, score) that were included in the input file. The cutoff values for these criteria can be chosen directly by value or by percentage if the user is in the “Percentage” tab.

Protein Filter

In the “Browser” tab, the user can also filter the proteins in a selected sample by clicking “Data Filter/Proteins” in the menu.

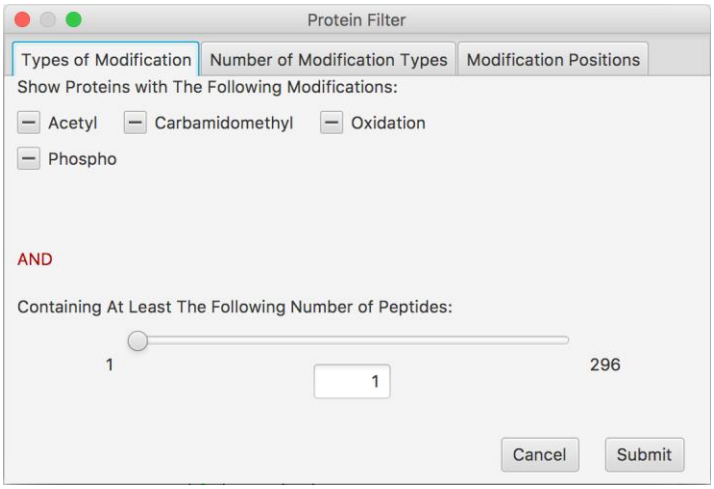


Figure 29

Two criteria are used for filtering as shown in the Protein Filter dialogue box above in Figure 29. On the top of the dialogue box, there are three options to select proteins with

modifications: Types of Modification, Number of Modification Types and Modification Positions. For “Types of Modification”, the user can select those proteins with or without the specified types of modifications. For each modification type, there are three possible states: checked(√), unchecked(□) and undefined(-). Checked indicates that the proteins must include the modification type; unchecked indicates that the proteins cannot include the modification type; and undefined indicates that the protein can either contain or not contain the modification type. For example, if the user checks Acetyl and Carbamidomethyl, unchecks Oxidation and undefines Phospho, the Browser will only show those proteins with both Acetyl and Carbamidomethyl groups that do not also have sites of Oxidation. These proteins can either contain Phospho sites or not since this modification is undefined. For “Number of Modification Types”, the user can select the number of modification types that the selected proteins must include. For example, if the user selects 2, the Browser will only show those proteins that contain two or more different *types* of modifications. **NOTE**: With this particular filter the Browser *only* considers modification types not numbers of modifications. Proteins with two or more sites of Oxidation will be filtered out if the user selects 2 since these proteins contain only one type of modification (i.e. Oxidation). For “Modification Position”, the user can select only those proteins that contain at least that number of modified positions. **NOTE**: Since only ONE of the three modification filter options can be considered at any given time by the Browser, in this example the Browser will ignore the “Number of Modification Types” and the “Modification Positions” filters because the user has already chosen to use the “Types of Modification” filter. At the bottom of the dialogue box, the user can filter out those proteins whose identifications are based on fewer than the selected number of identified peptides. After filtering, the selected proteins can be exported to a text file by clicking “Export/Proteins after Filtering” in the menu. To select the number that will be used in the filter, the user can either select a number with the slider or enter a number and press “ENTER” key in the text field that is under the slider. **NOTE:** The Browser only shows those proteins that meet BOTH of the two criteria, modification and number of peptides, selected by the user. For example, if the user selects at least 3 modification types and at least 10 peptides, the browser will only show those proteins with at least 3 types of modifications AND that have at least 10 identified peptides. The filtered proteins can be exported to a text file by using the menu item, “Export/Proteins after Filtering”.

Protein Abundance

ProteomicsBrowser calculates protein abundance from peptide abundance/intensity using whichever approach is chosen by the user in the Data/Integration Menu option on the Protein Data tab. The three different methods for calculating protein abundance are: Raw, iBAQ and Top3.

RAW

The RAW value is the sum of the peak area intensities for all peptides that have been aligned to the protein. Peptides that can be aligned to more than one protein are not included.

iBAQ

iBAQ = raw value/the number of theoretical tryptic peptides (6-30 amino acids) from that protein. The number of theoretical tryptic peptides (6-30 amino acids) from that protein is calculated based on the approach described by Fabre et al (2014) from an *in silico* protein digestion. For Trypsin, cleavage occurs at the C-terminal side of lysine(K) or arginine (R), except where these residues are directly followed by a proline (P).

Top3

Top3 is the summation of the top 3 peptides from each protein that have the largest intensities.

Basic Statistical Analysis

ProteomicsBrowser also includes some basic functions to analyze the abundance of peptides and proteins when the user is in the “Data/Peptide Data” or “Data/Protein Data” Tabs. **All of these analyses are based on the linear values without log2 or log10 transformation.**

T-Test

Clicking “Analyze/T-Test” in the menu brings up the dialogue box shown below in **Figure 30**.

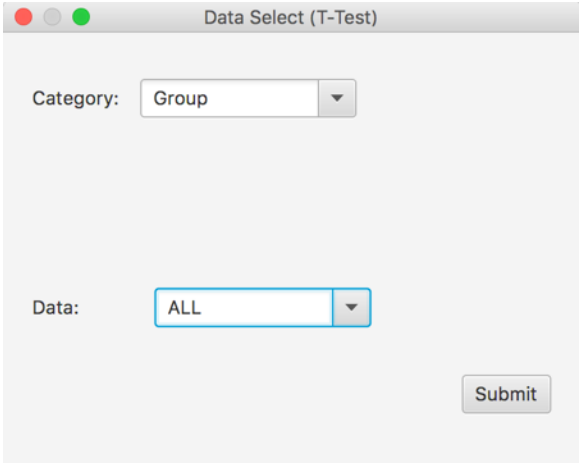


Figure 30

In “Category”, the user can select the feature that will be used to divide the samples into two groups. For the sample test data set, there are two categories in “Group”: Disease and Normal. After making a selection within Group, ProteomicsBrowser will perform a t-test between the Disease and Normal samples for the selected proteins/peptides in “Data”. Within “Data” the user can select one protein/peptide or all proteins/peptides for the analysis.

Selecting a numeric variable in “Category” like “Weight”, will bring up a slightly different dialogue box as shown below in **Figure 31**. By default, the Browser divides all of the samples into 3 groups according to the selected variable and then uses the first tertile and the last tertile as two groups. The cutoff can be adjusted with the slider.

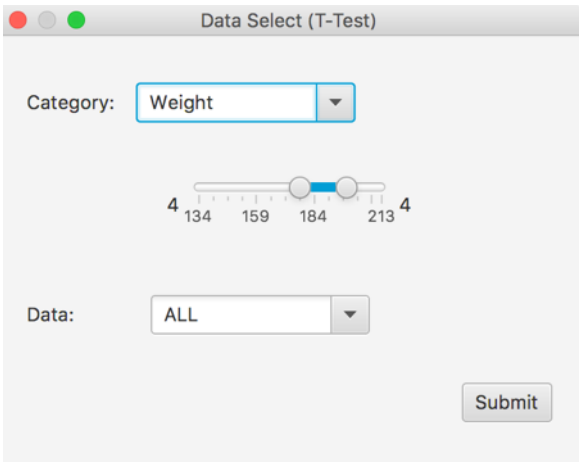
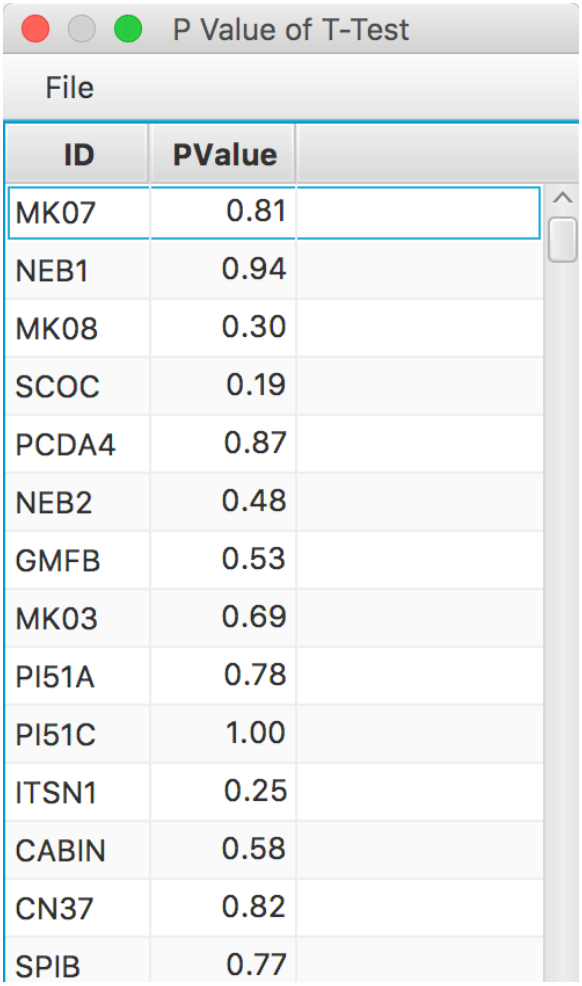


Figure 31

After clicking the “Submit” button, a table will be displayed with the p-values for the selected proteins/peptides as shown below in **Figure 32**.



ID	PValue
MK07	0.81
NEB1	0.94
MK08	0.30
SCOC	0.19
PCDA4	0.87
NEB2	0.48
GMFB	0.53
MK03	0.69
PI51A	0.78
PI51C	1.00
ITSN1	0.25
CABIN	0.58
CN37	0.82
SPIB	0.77

Figure 32

Box Plot

The Box plot function is similar to that of the T-Test. After selecting a feature for the “Category” and a “Protein/peptide” of interest, a box plot will be displayed as shown below in **Figure 33**. After the plot is generated, the View option can be used to select Regular, Log2, Log 10. In addition, the Jitter option can be used to display the individual data points.

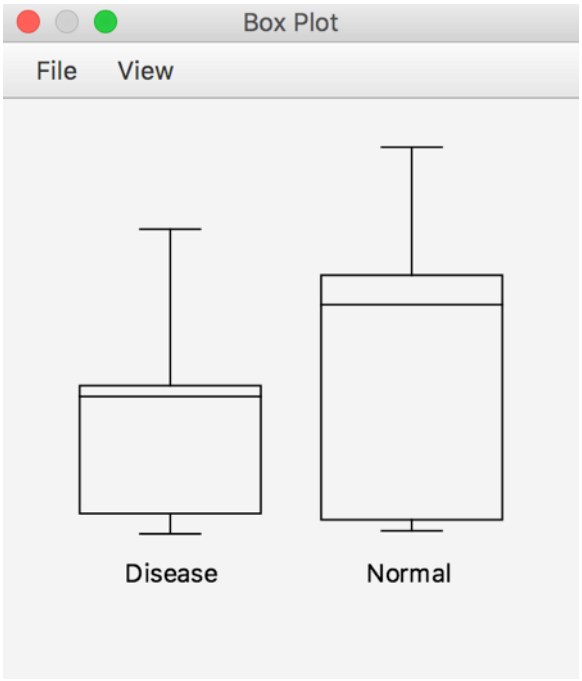


Figure 33

Correlation

The correlation function shows the correlation between two different proteins/peptides. After clicking “Analyze/Correlation” in the menu, a dialogue box will appear as shown below in **Figure 34**. Next, the user selects the category that the correlation should be performed on in the “Category” option.

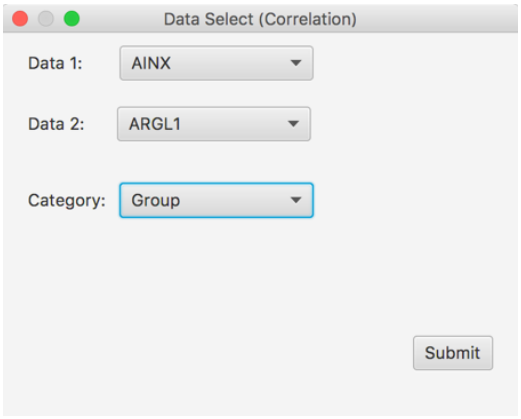


Figure 34

After selecting a pair of proteins/peptides and clicking the “Submit” button, a scatter plot of the selected proteins/peptides will be displayed as shown below in **Figure 35**.

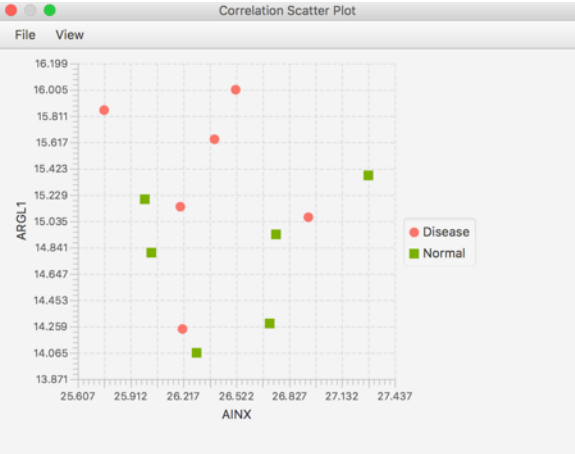


Figure 35

Abundance

Normalization

The protein abundances can be normalized between samples using either of two methods: median normalization and normalization based on a selected protein. Of course, users also can select no normalization. Any of the three options can be selected from the “Data/Normalization” menu.

View

For both peptide data and protein data, ProteomicsBrowser has three options to display the abundances: linear, $\log_2$ scale, and $\log_{10}$ scale. Users can select any one of these options using the “View/Scale” menu.

Literature Cited:

Fabre, B., Lambour, T., Bouyssie, D., Menneteau, T., Monserrat, B., Burlet-Schultz, O., Bousquet-Dubouch, M.,(2014) Comparison of label-free quantification methods for the determination of protein complexes subunits stoichiometry. EuPA Open Proteomics 4:82-86

Olsen, L.R., Kudahl, U.J., Simon, C., Sun, J., Schönbach, C., Reinherz, E.L., Zhang, G.L., Brusic V.et al. (2013) BlockLogo: visualization of peptide and sequence motif conservation. Journal Immunological Methods:400-401:37-44.