ProteomicsBrowser: MS/Proteomics Data Visualization and Investigation

Gang Peng

Department of Biostatistics
Yale School of Public Health

May 1, 2019

Proteomic Data Analysis

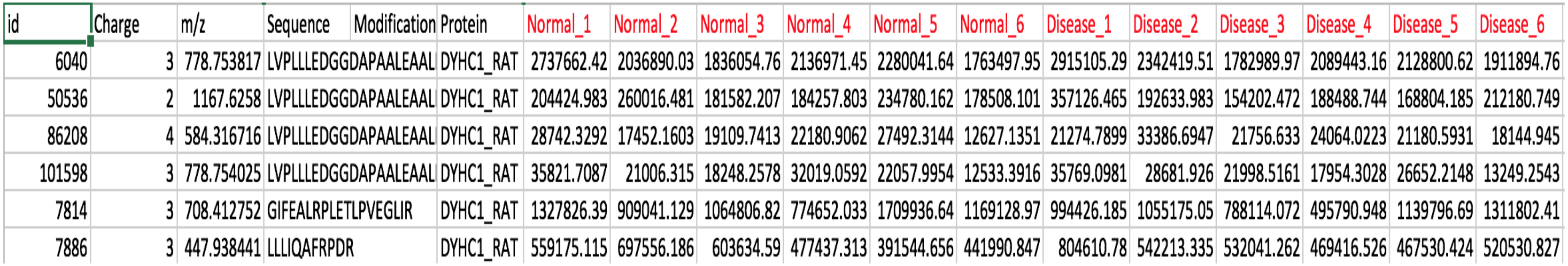
- Challenges
  - Large-scale
  - Complexity
  - Increasing numbers of post-translational modifications
- Advantages of Data Visualization
  - Show the information quick and clear
  - Identify the relationships and patterns
  - Show the details of difference

ProteomicsBrowser

- Java
  - Windows
  - macOS
  - Linux
- Update Database from UniProt

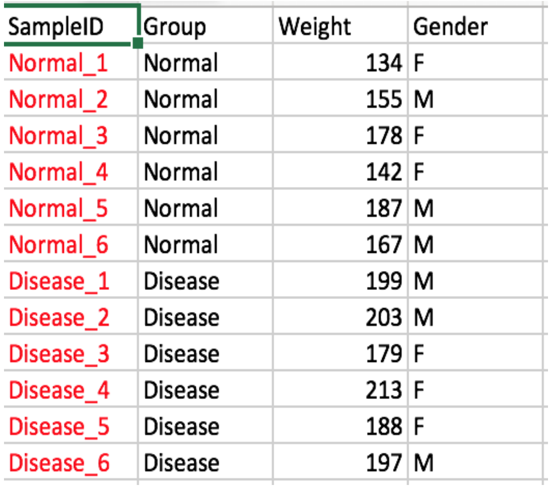
Input Data

- Peptide Data File



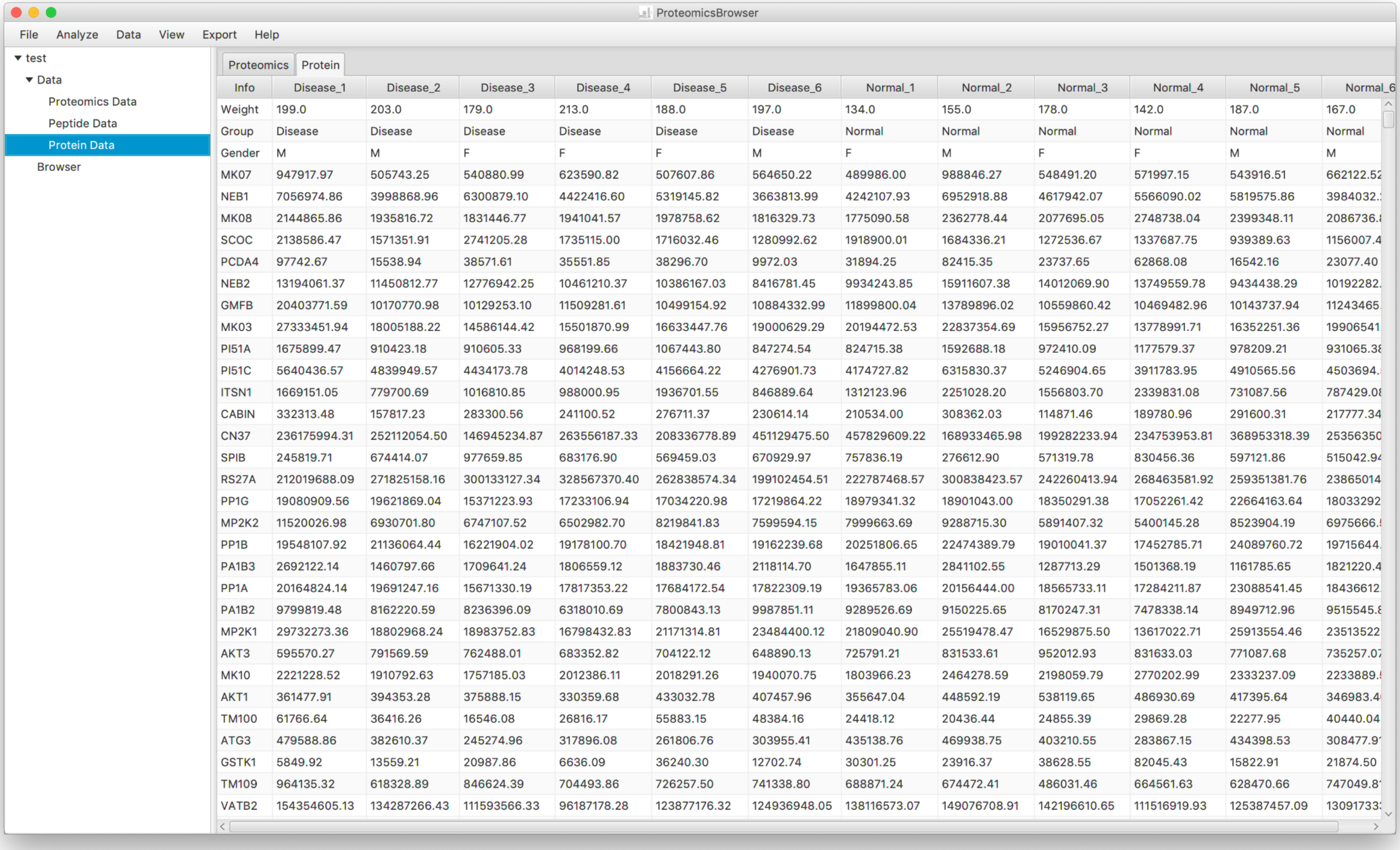
id	Charge	m/z	Sequence	Modification	Protein	Normal_1	Normal_2	Normal_3	Normal_4	Normal_5	Normal_6	Disease_1	Disease_2	Disease_3	Disease_4	Disease_5	Disease_6
6040	3	778.753817	LVPLLLEDGGDAPAALEAAL		DYHC1_RAT	2737662.42	2036890.03	1836054.76	2136971.45	2280041.64	1763497.95	2915105.29	2342419.51	1782989.97	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 Data 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

Table View



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
MK07	947917.97	505743.25	540880.99	623590.82	507607.86	564650.22	489986.00	988846.27	548491.20	571997.15	543916.51	662122.52
NEB1	7056974.86	3998868.96	6300879.10	4422416.60	5319145.82	3663813.99	4242107.93	6952918.88	4617942.07	5566090.02	5819575.86	3984032.
MK08	2144865.86	1935816.72	1831446.77	1941041.57	1978758.62	1816329.73	1775090.58	2362778.44	2077695.05	2748738.04	2399348.11	2086736.
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	13194061.37	11450812.77	12776942.25	10461210.37	10386167.03	8416781.45	9934243.85	15911607.38	14012069.90	13749559.78	9434438.29	10192282.
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	27333451.94	18005188.22	14586144.42	15501870.99	16633447.76	19000629.29	20194472.53	22837354.69	15956752.27	13778991.71	16352251.36	19906541
PI51A	1675899.47	910423.18	910605.33	968199.66	1067443.80	847274.54	824715.38	1592688.18	972410.09	1177579.37	978209.21	931065.38
PI51C	5640436.57	4839949.57	4434173.78	4014248.53	4156664.22	4276901.73	4174727.82	6315830.37	5246904.65	3911783.95	4910565.56	4503694.
ITSN1	1669151.05	779700.69	1016810.85	988000.95	1936701.55	846889.64	1312123.96	2251028.20	1556803.70	2339831.08	731087.56	787429.0
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	236175994.31	252112054.50	146945234.87	263556187.33	208336778.89	451129475.50	457829609.22	168933465.98	199282233.94	234753953.81	368953318.39	25356350
SPIB	245819.71	674414.07	977659.85	683176.90	569459.03	670929.97	757836.19	276612.90	571319.78	830456.36	597121.86	515042.9
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	19080909.56	19621869.04	15371223.93	17233106.94	17034220.98	17219864.22	18979341.32	18901043.00	18350291.38	17052261.42	22664163.64	18033292
MP2K2	11520026.98	6930701.80	6747107.52	6502982.70	8219841.83	7599594.15	7999663.69	9288715.30	5891407.32	5400145.28	8523904.19	6975666.
PP1B	19548107.92	21136064.44	16221904.02	19178100.70	18421948.81	19162239.68	20251806.65	22474389.79	19010041.37	17452785.71	24089760.72	19715644.
PA1B3	2692122.14	1460797.66	1709641.24	1806559.12	1883730.46	2118114.70	1647855.11	2841102.55	1287713.29	1501368.19	1161785.65	1821220.4
PP1A	20164824.14	19691247.16	15671330.19	17817353.22	17684172.54	17822309.19	19365783.06	20156444.00	18565733.11	17284211.87	23088541.45	18436612.
PA1B2	9799819.48	8162220.59	8236396.09	6318010.69	7800843.13	9987851.11	9289526.69	9150225.65	8170247.31	7478338.14	8949712.96	9515545.
MP2K1	29732273.36	18802968.24	18983752.83	16798432.83	21171314.81	23484400.12	21809040.90	25519478.47	16529875.50	13617022.71	25913554.46	23513522
AKT3	595570.27	791569.59	762488.01	683352.82	704122.12	648890.13	725791.21	831533.61	952012.93	831633.03	771087.68	735257.0
MK10	2221228.52	1910792.63	1757185.03	2012386.11	2018291.26	1940070.75	1803966.23	2464278.59	2198059.79	2770202.99	2333237.09	2233889.
AKT1	361477.91	394353.28	375888.15	330359.68	433032.78	407457.96	355647.04	448592.19	538119.65	486930.69	417395.64	346983.4
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	964135.32	618328.89	846624.39	704493.86	726257.50	741338.80	688871.24	674472.41	486031.46	664561.63	628470.66	747049.8
VATB2	154354605.13	134287266.43	111593566.33	96187178.28	123877176.32	124936948.05	138116573.07	149076708.91	142196610.65	111516919.93	125387457.09	13091733

Browser

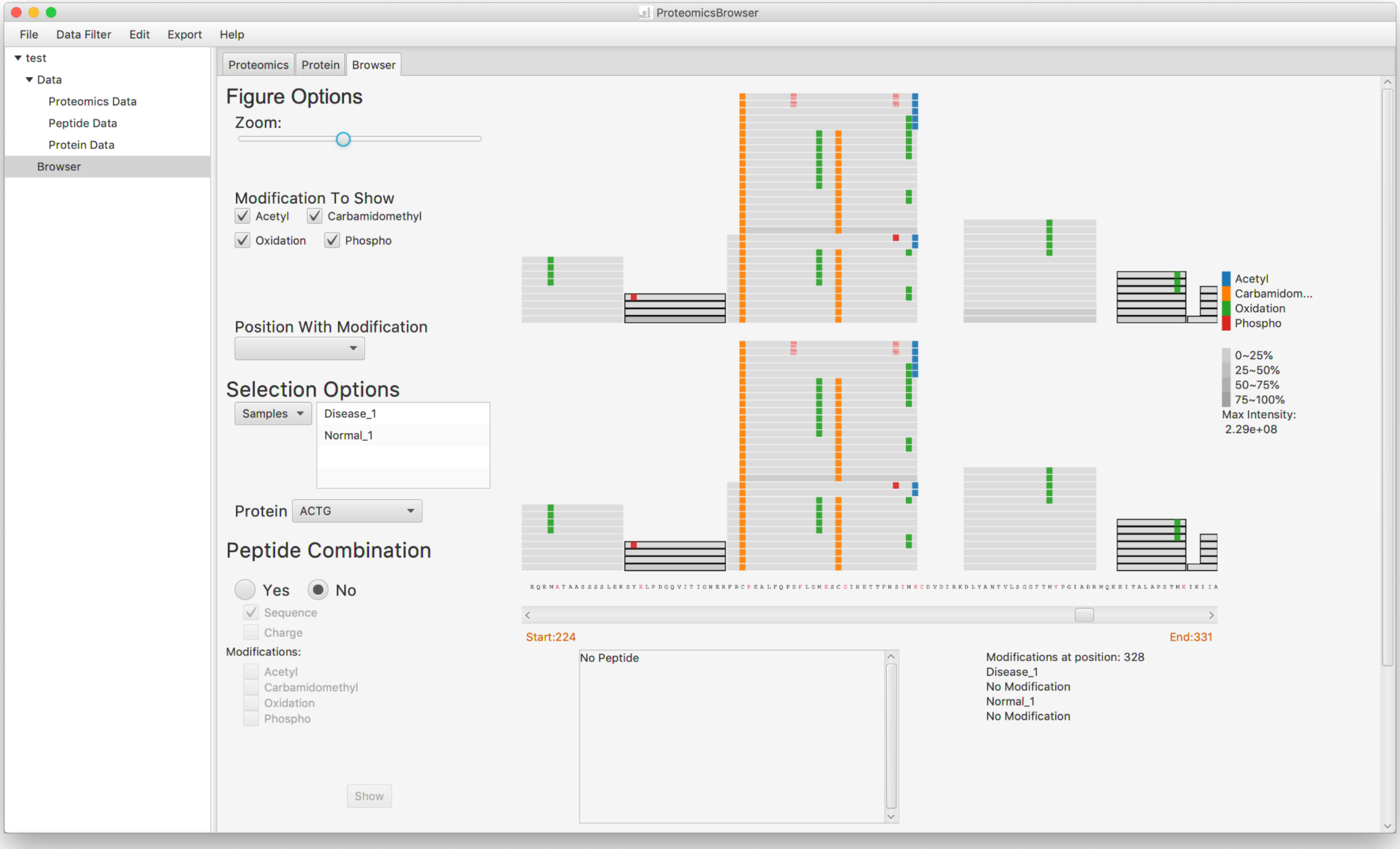


Figure Options

Zoom:

Modification To Show

☑ Acetyl ☑ Carbamidomethyl
☑ Oxidation ☑ Phospho

Position With Modification

Selection Options

Samples: Disease_1, Normal_1

Protein: ACTG

Peptide Combination

○ Yes ● No
☑ Sequence
☐ Charge

Modifications:
☐ Acetyl
☐ Carbamidomethyl
☐ Oxidation
☐ Phospho

Show

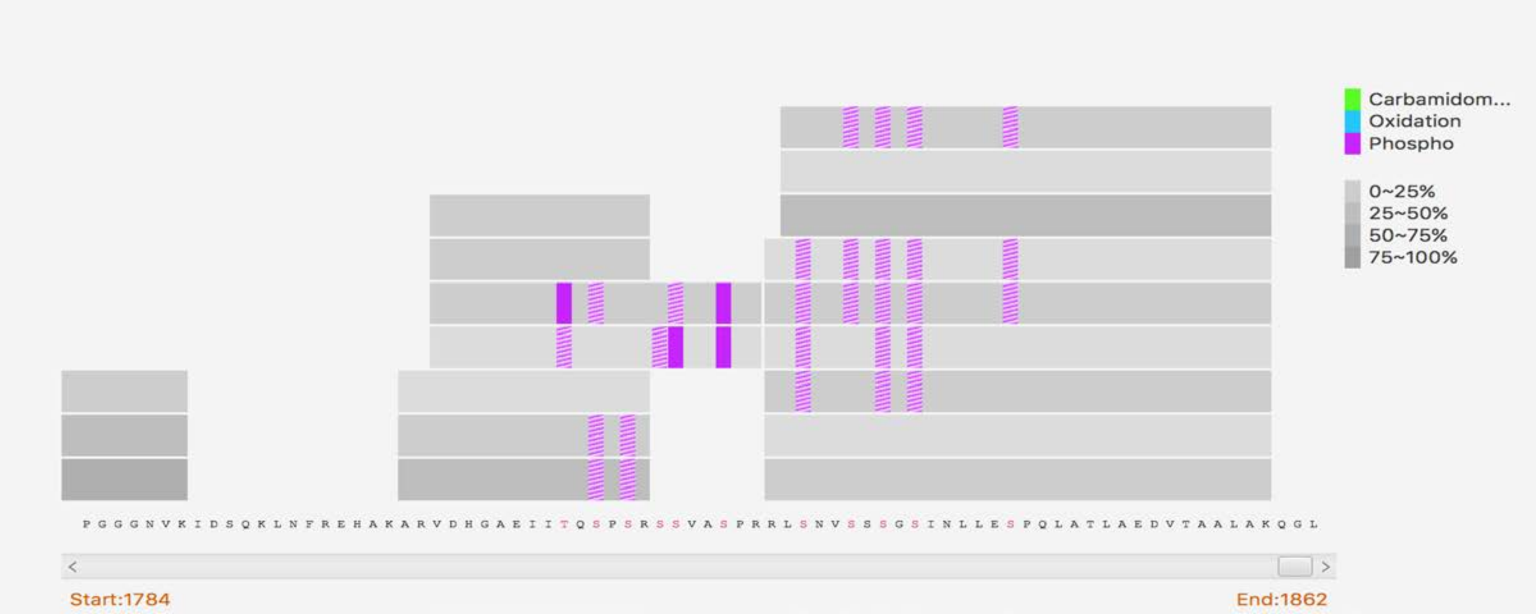
Start:224 End:331

Max Intensity: 2.29e+08

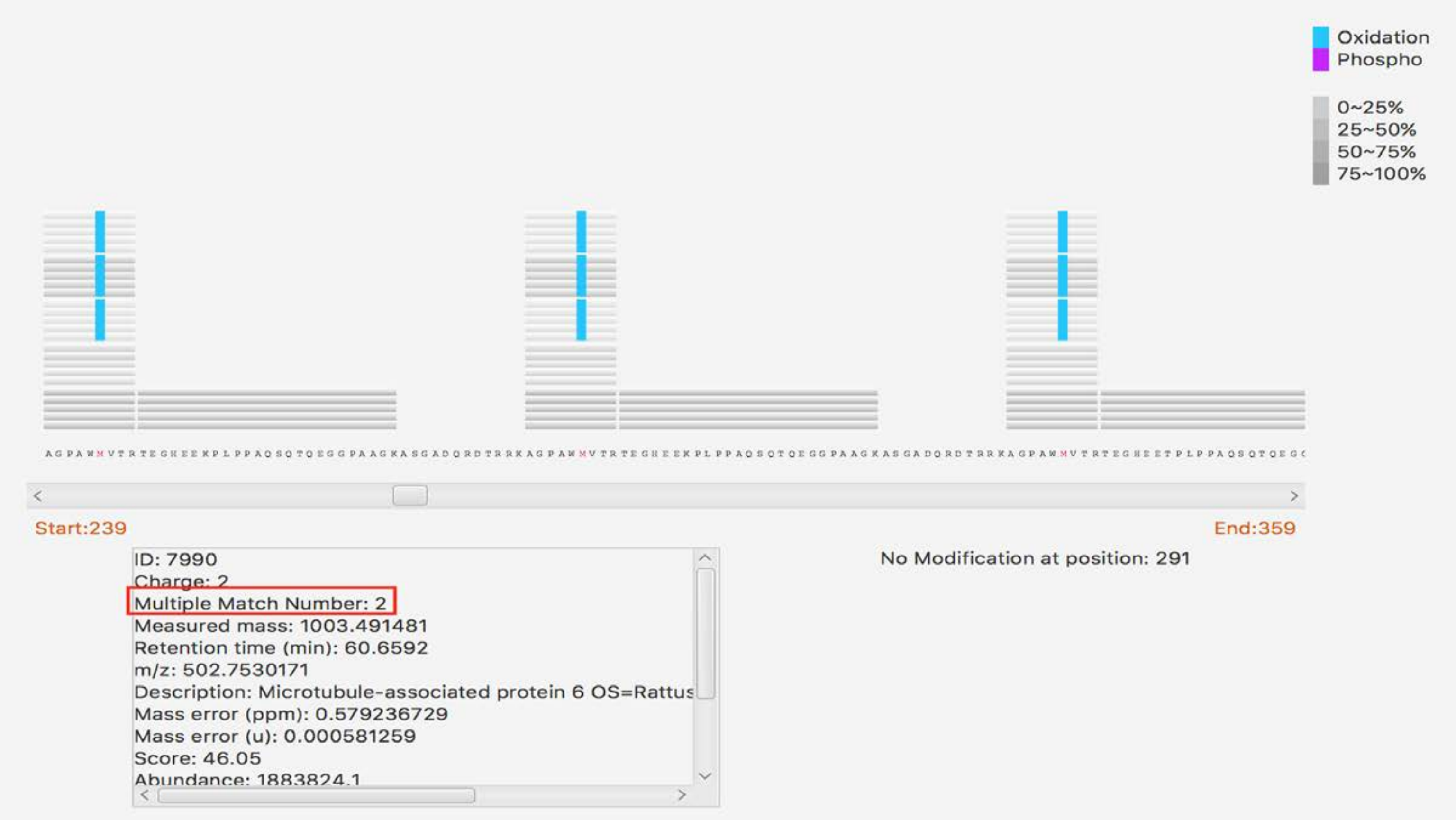
No Peptide

Modifications at position: 328
Disease_1
No Modification
Normal_1
No Modification

PTM with Unassigned Sites



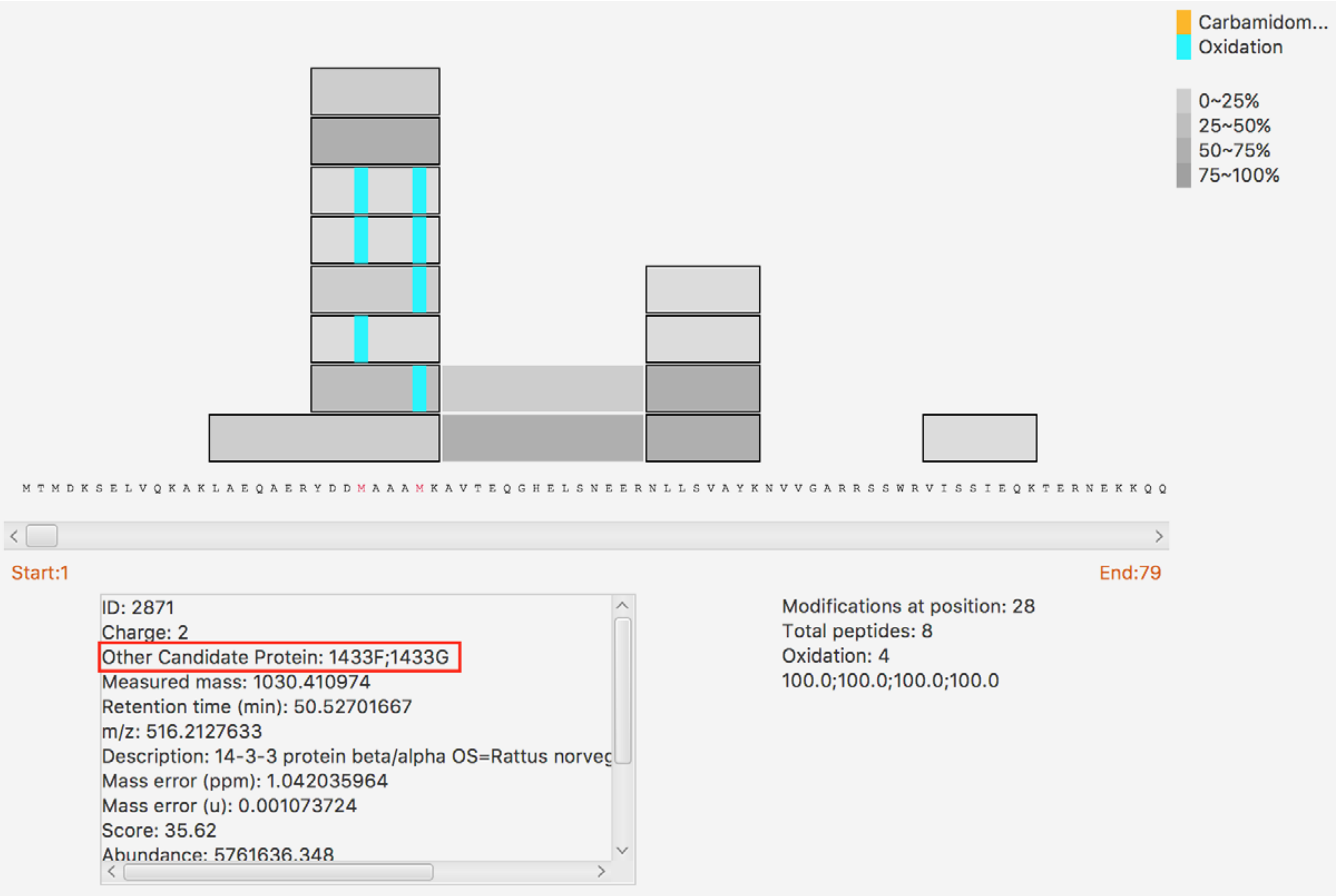
Peptides with Multiple Matches in A Protein



ID: 7990
Charge: 2
Multiple Match Number: 2
Measured mass: 1003.491481
Retention time (min): 60.6592
m/z: 502.7530171
Description: Microtubule-associated protein 6 OS=Rattus
Mass error (ppm): 0.579236729
Mass error (u): 0.000581259
Score: 46.05
Abundance: 1883824.1

No Modification at position: 291

Peptides Mapped to Multiple Proteins

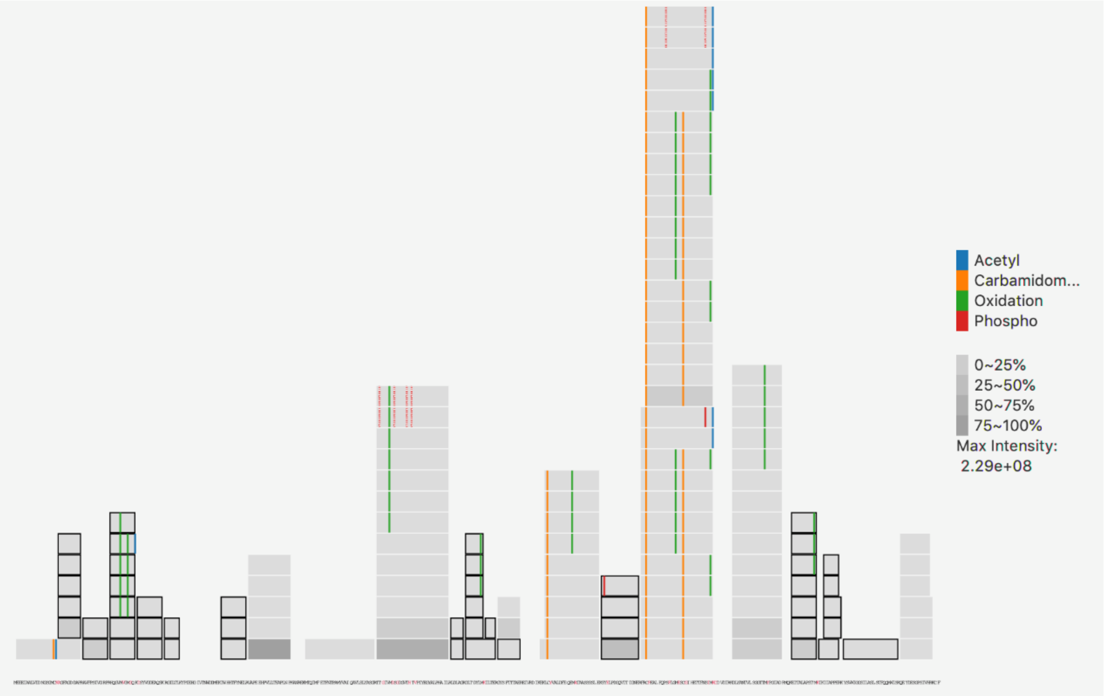


ID: 2871
Charge: 2
Other Candidate Protein: 1433F;1433G
Measured mass: 1030.410974
Retention time (min): 50.52701667
m/z: 516.2127633
Description: 14-3-3 protein beta/alpha OS=Rattus norveg
Mass error (ppm): 1.042035964
Mass error (u): 0.001073724
Score: 35.62
Abundance: 5761636.348

Modifications at position: 28
Total peptides: 8
Oxidation: 4
100.0;100.0;100.0;100.0

Peptide Combination

- Overlapping peptides
  - Partial proteolytic cleavage
  - Multiple charge states
  - Post-translational modifications (PTMs)

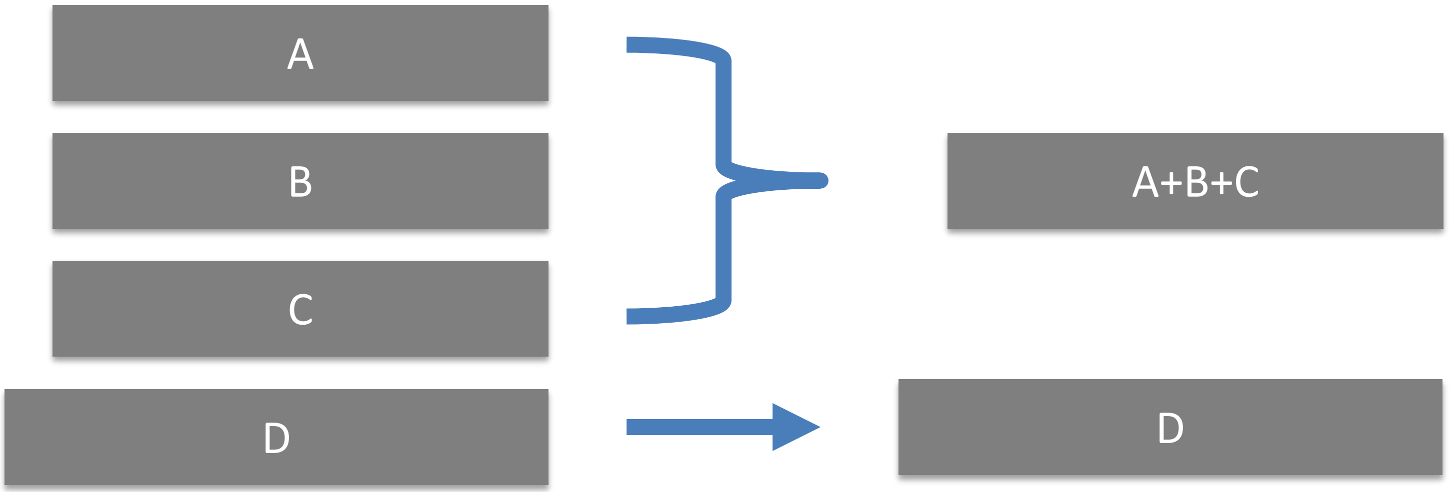


Peptide Combination

- Sequence Based Combination
  - Peptides with same sequence
  - Other criteria

- PTM Based Combination
  - Peptides with same PTM at selected position

Sequence Based Combination



A

B

C

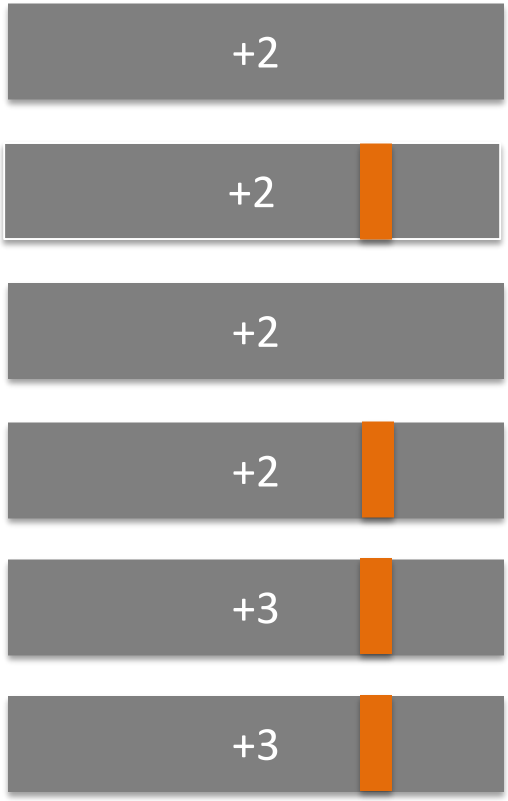
D

A+B+C

D

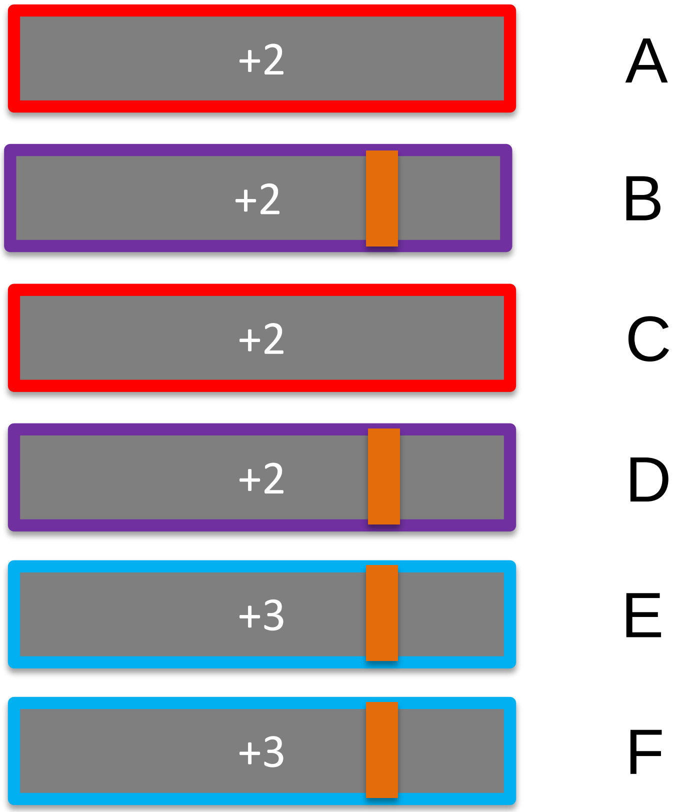
Sequence Based Combination

Same PTM and Same Charge



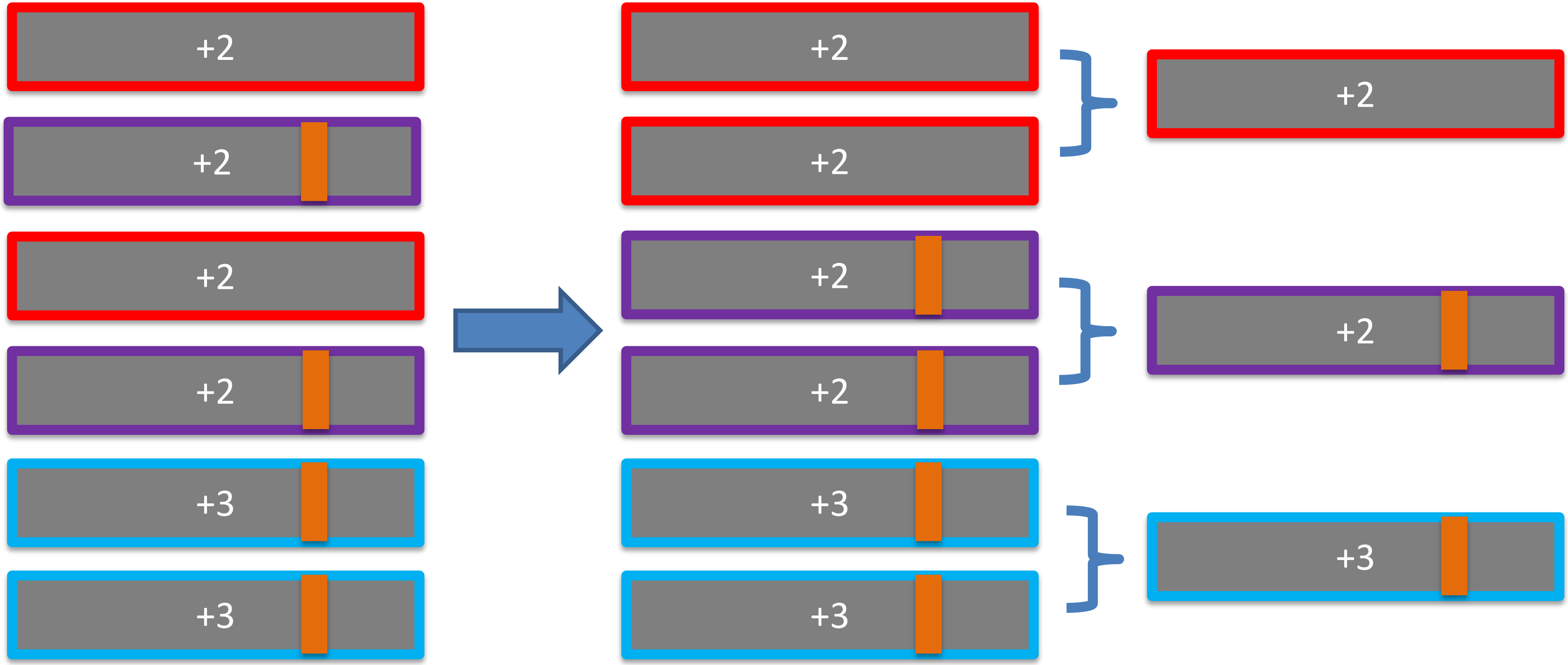
Sequence Based Combination

Same PTM and Same Charge



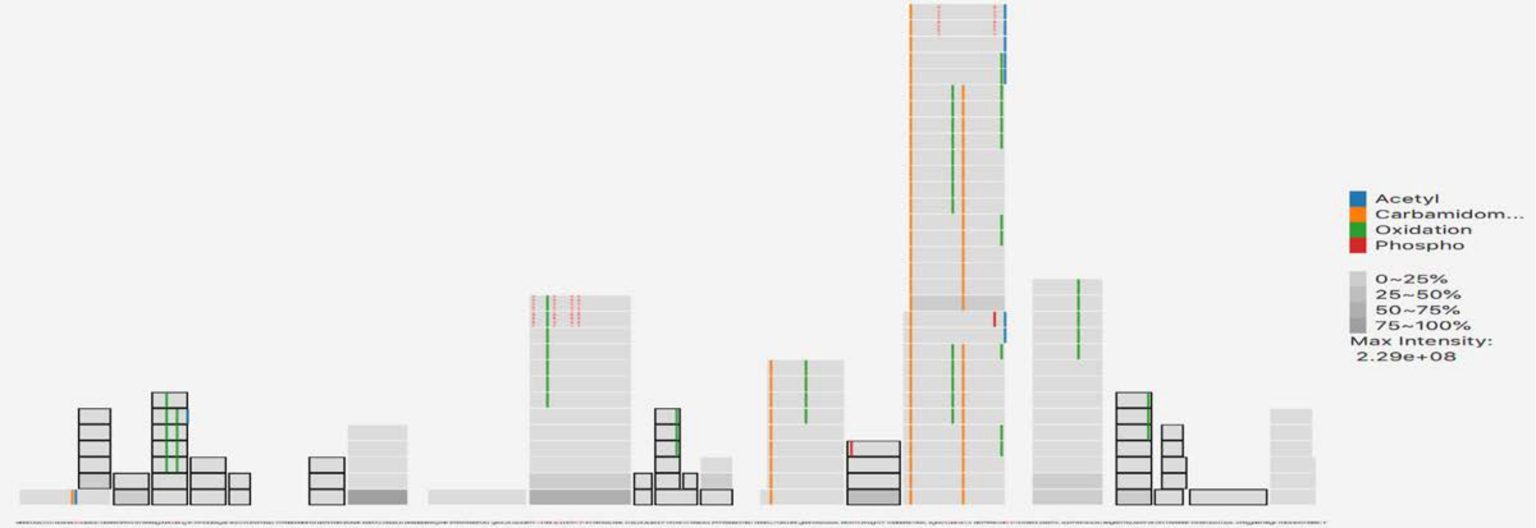
Sequence Based Combination

Same PTM and Same Charge

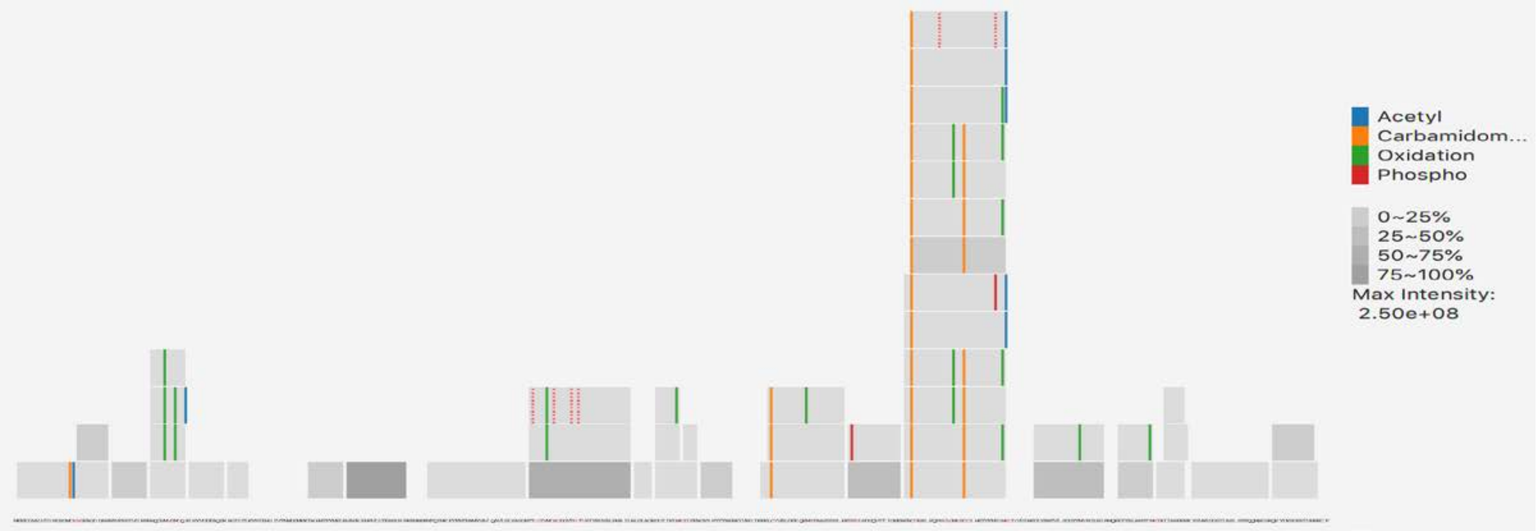


Sequence Based Combination

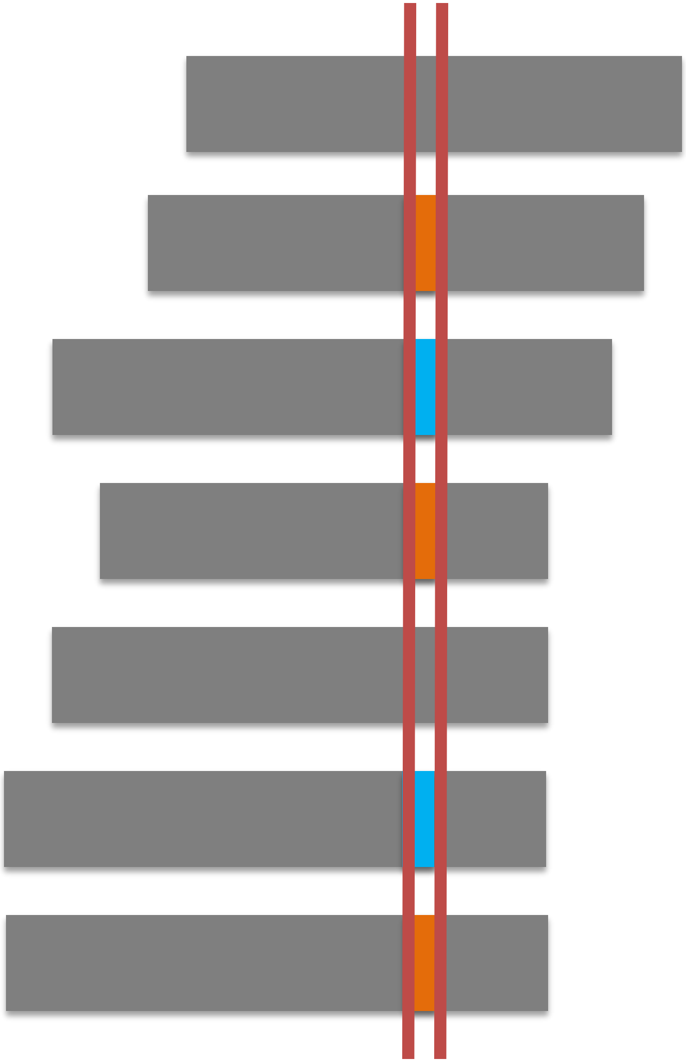
Before



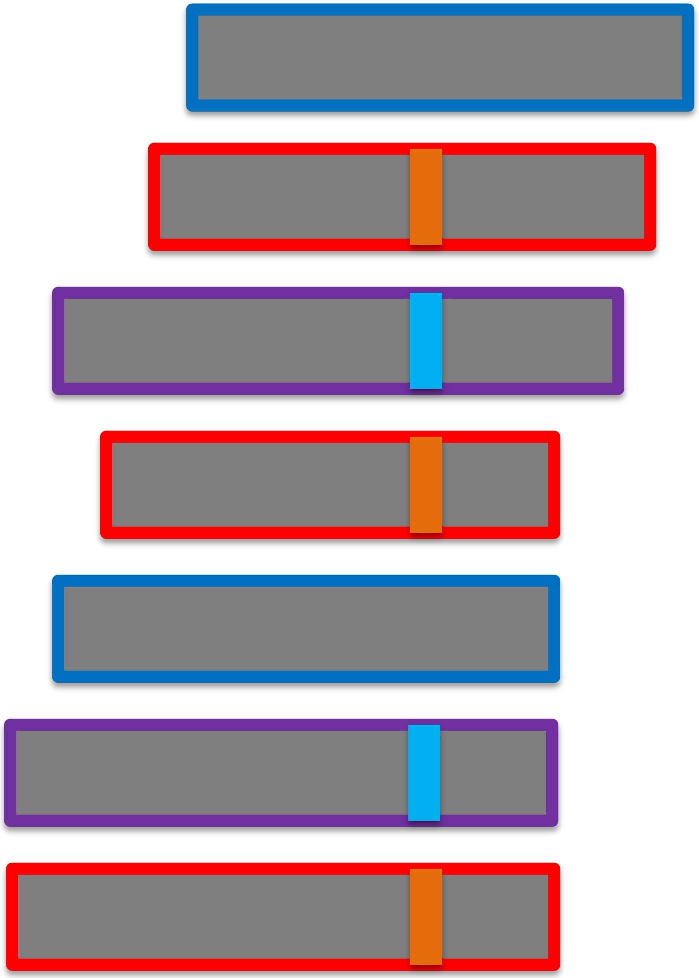
After



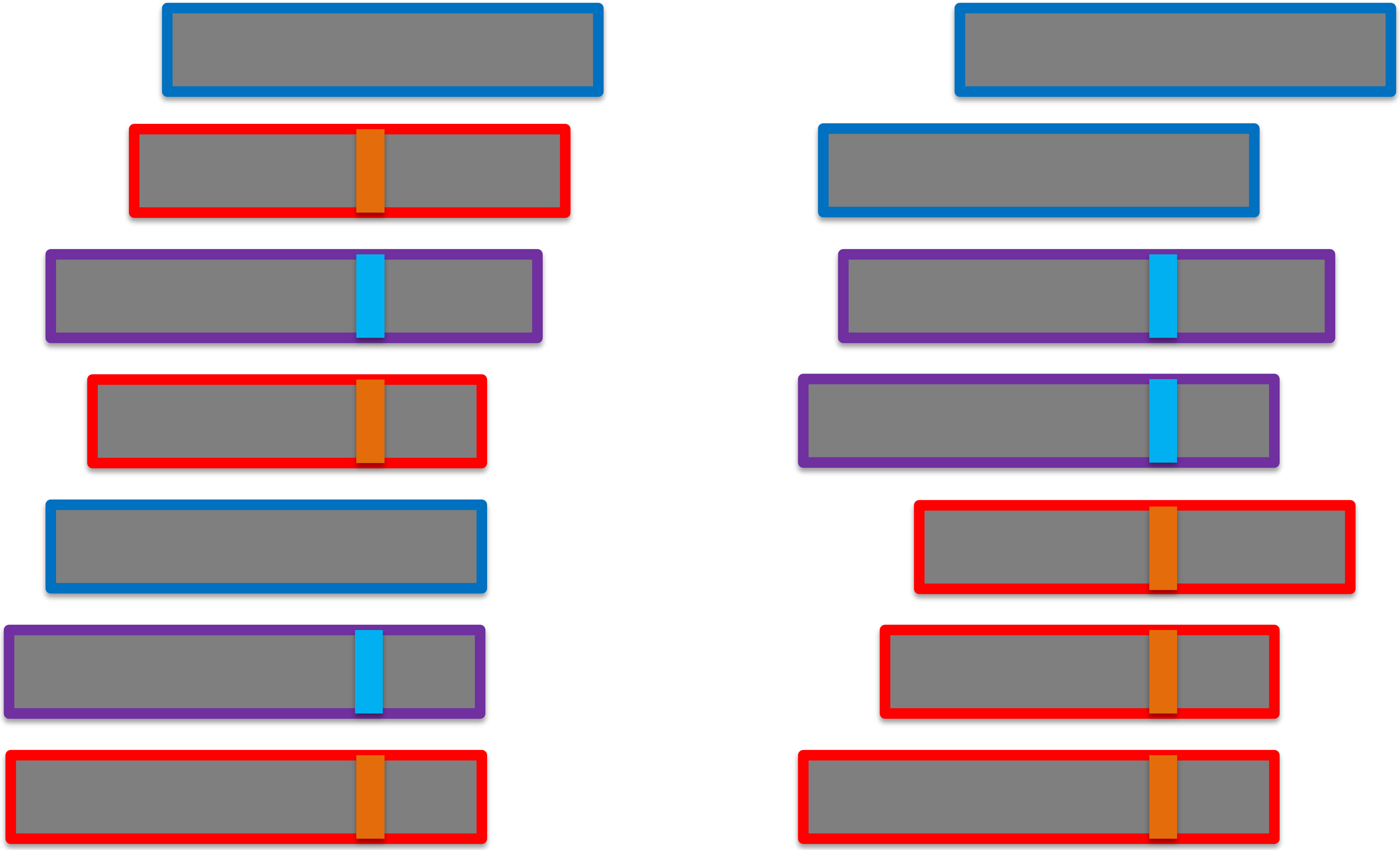
PTM Based Combination



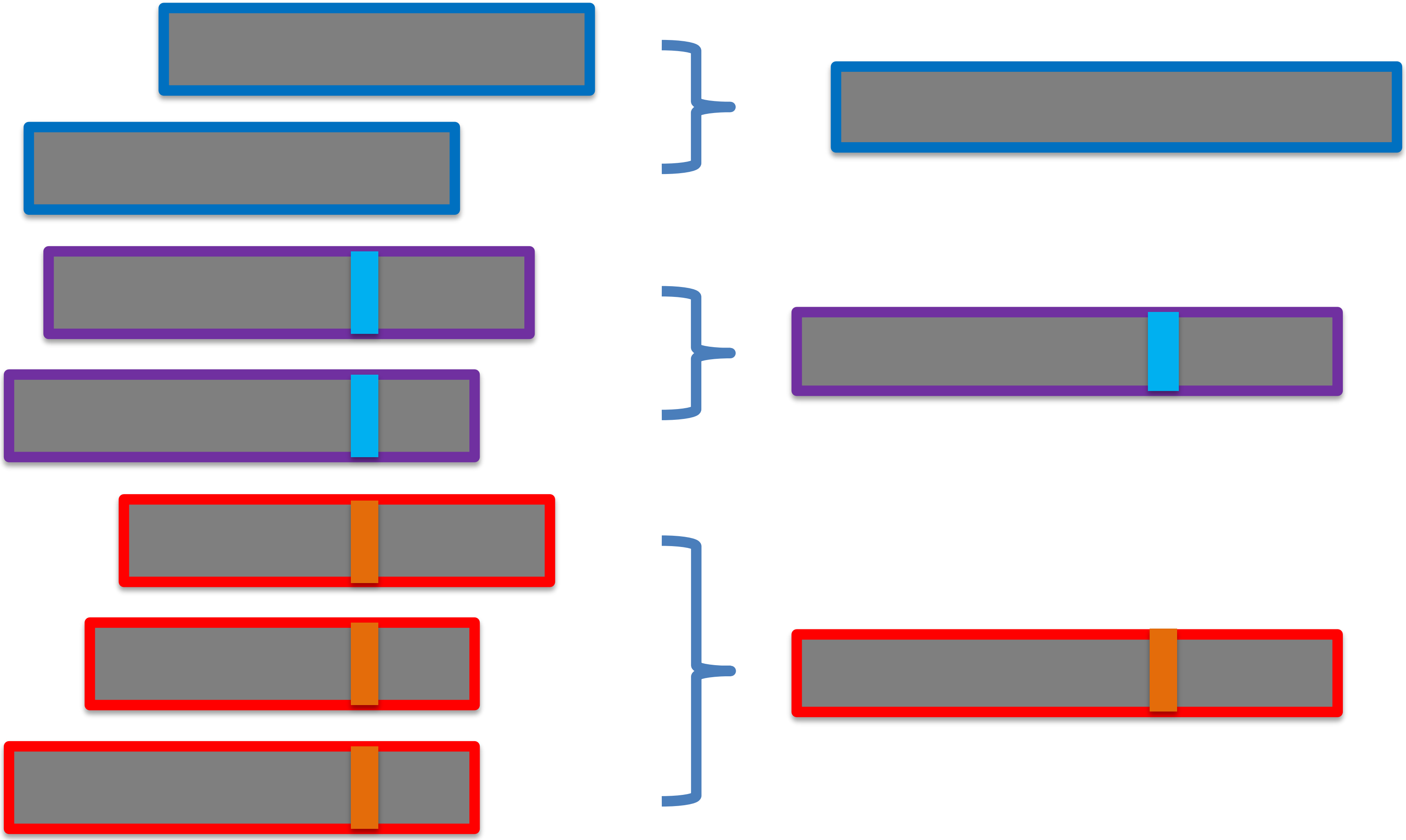
PTM Based Combination



PTM Based Combination



PTM Based Combination

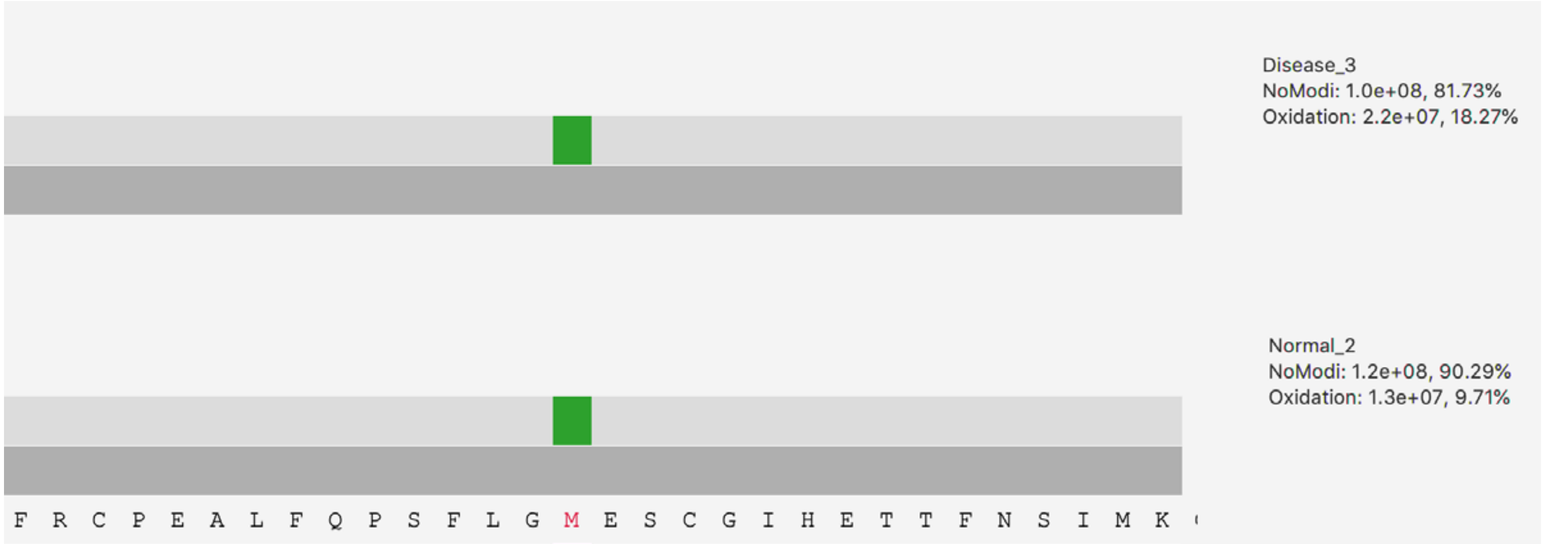


PTM Based Combination

Disease_3

Normal_2

PTM Based Combination

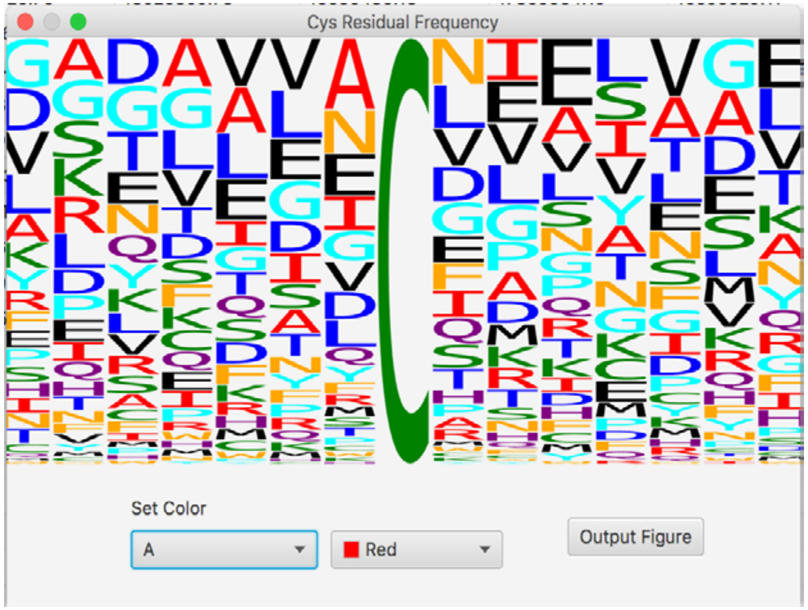


Disease_3
NoModi: 1.0e+08, 81.73%
Oxidation: 2.2e+07, 18.27%

Normal_2
NoModi: 1.2e+08, 90.29%
Oxidation: 1.3e+07, 9.71%

Other Functions

- Residue Frequency
- Correlation
- Statistic Tests

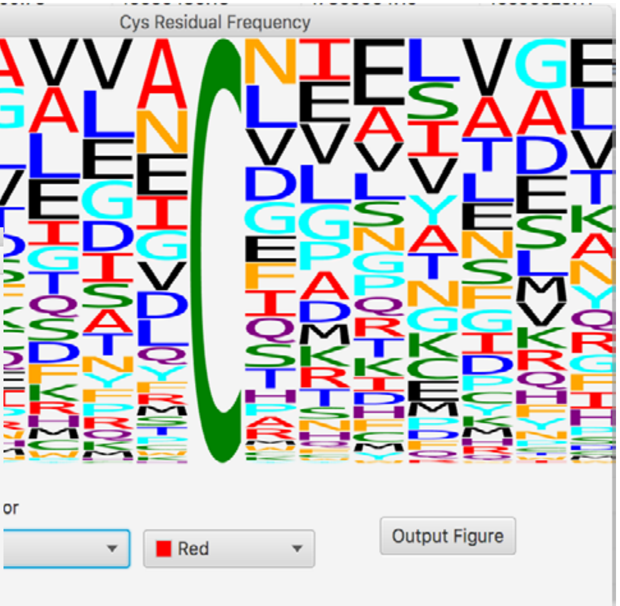


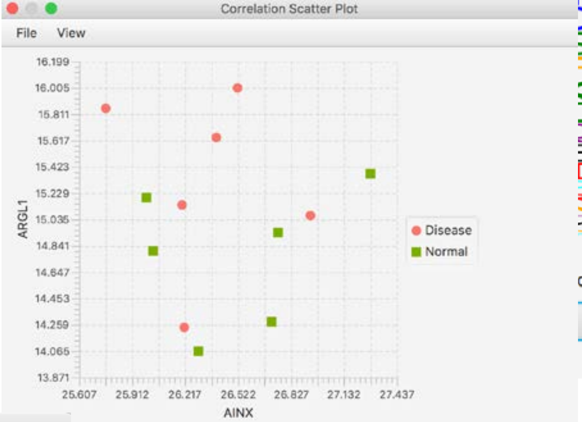
Other Functions

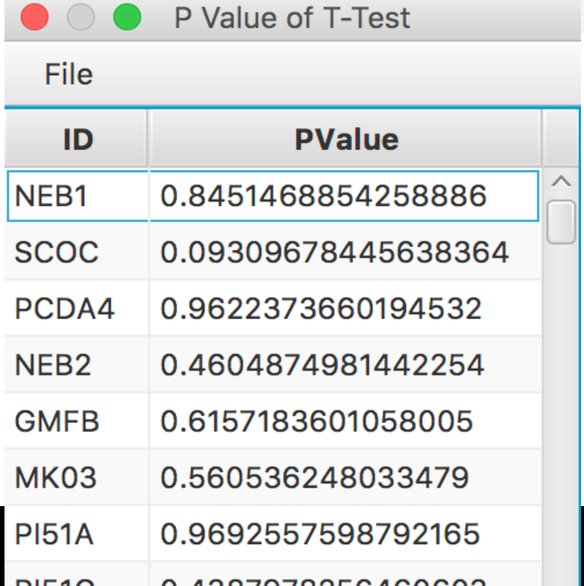
- Residue Frequency
- Correlation
- Statistic Tests

Other Functions

- Residue Frequency
- Correlation
- Statistic Tests

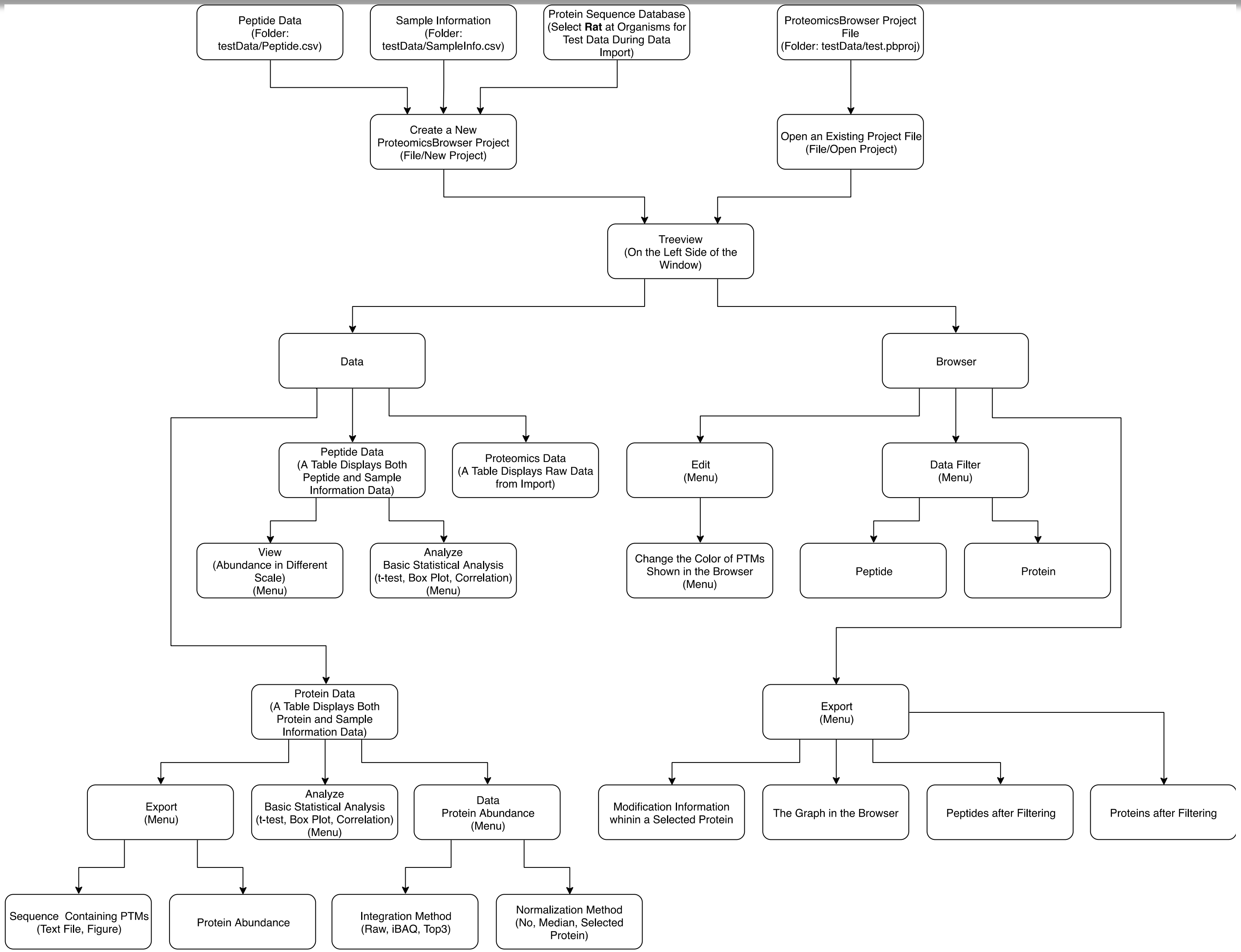






ID	PValue
NEB1	0.8451468854258886
SCOC	0.09309678445638364
PCDA4	0.9622373660194532
NEB2	0.4604874981442254
GMFB	0.6157183601058005
MK03	0.560536248033479
PI51A	0.9692557598792165

ProteomicsBrowser



Peptide Data
(Folder:
testData/Peptide.csv)

Sample Information
(Folder:
testData/SampleInfo.csv)

Protein Sequence Database
(Select **Rat** at Organisms for
Test Data During Data
Import)

ProteomicsBrowser Project
File
(Folder: testData/test.pbproj)

Create a New
ProteomicsBrowser Project
(File/New Project)

Open an Existing Project File
(File/Open Project)

Treeview
(On the Left Side of the
Window)

Data

Browser

Peptide Data
(A Table Displays Both
Peptide and Sample
Information Data)

Proteomics Data
(A Table Displays Raw Data
from Import)

Edit
(Menu)

Data Filter
(Menu)

View
(Abundance in Different
Scale)
(Menu)

Analyze
Basic Statistical Analysis
(t-test, Box Plot, Correlation)
(Menu)

Change the Color of PTMs
Shown in the Browser
(Menu)

Peptide

Protein

Protein Data
(A Table Displays Both
Protein and Sample
Information Data)

Export
(Menu)

Export
(Menu)

Analyze
Basic Statistical Analysis
(t-test, Box Plot, Correlation)
(Menu)

Data
Protein Abundance
(Menu)

Modification Information
whinin a Selected Protein

The Graph in the Browser

Peptides after Filtering

Proteins after Filtering

Sequence Containing PTMs
(Text File, Figure)

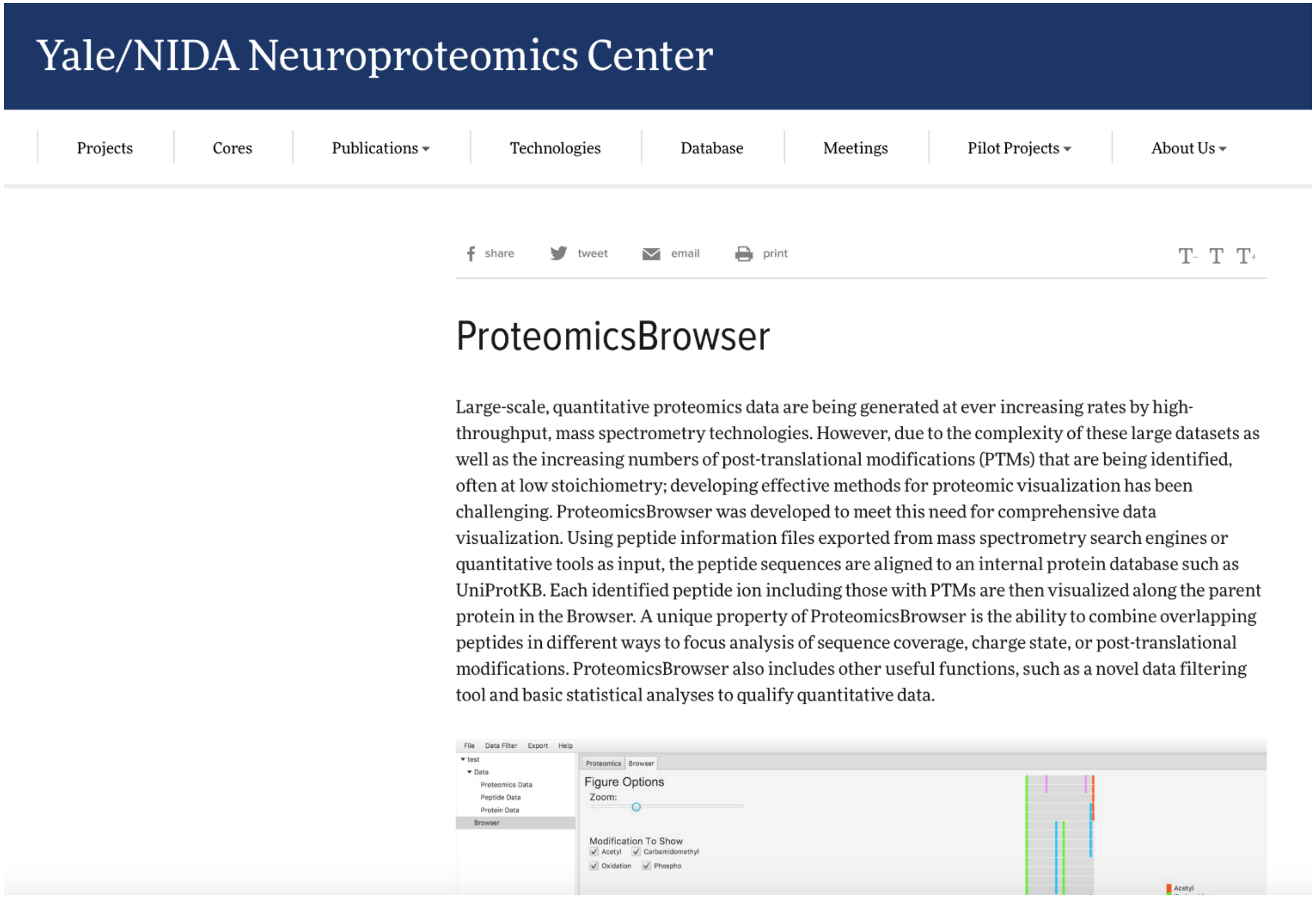
Protein Abundance

Integration Method
(Raw, iBAQ, Top3)

Normalization Method
(No, Median, Selected
Protein)

Availability

- https://medicine.yale.edu/keck/nida/proteomicsbrowser.aspx



Yale/NIDA Neuroproteomics Center

ProteomicsBrowser

Large-scale, quantitative proteomics data are being generated at ever increasing rates by high-throughput, mass spectrometry technologies. However, due to the complexity of these large datasets as well as the increasing numbers of post-translational modifications (PTMs) that are being identified, often at low stoichiometry; developing effective methods for proteomic visualization has been challenging. ProteomicsBrowser was developed to meet this need for comprehensive data visualization. Using peptide information files exported from mass spectrometry search engines or quantitative tools as input, the peptide sequences are aligned to an internal protein database such as UniProtKB. Each identified peptide ion including those with PTMs are then visualized along the parent protein in the Browser. A unique property of ProteomicsBrowser is the ability to combine overlapping peptides in different ways to focus analysis of sequence coverage, charge state, or post-translational modifications. ProteomicsBrowser also includes other useful functions, such as a novel data filtering tool and basic statistical analyses to qualify quantitative data.

Acknowledgement

- Department of Biostatistics
  - Hongyu Zhao
- Department of Psychiatry
  - Rashuan Wilson
  - Angus C. Nairn
  - Melissa Monsey
  - Jane Taylor
- Department of Genetics
  - Yishuo Tang
- Department of Molecular Biophysics and Biochemistry
  - Kenneth Williams
  - TuKiet T. Lam