COVID19 Classification Project

Chao Wang, Steven Wu, Brian Wu

Introduction

Objective: Identify potential biomarkers and therapeutic targets for COVID-19 through the analysis of proteomic and metabolomic data from patient serum samples.

Input: Proteomic and metabolomic data obtained from serum samples of COVID-19 patients.

- Proteins: 402
- Peptides: 101,461
- Metabolites: 941 biochemicals

Output: A list of potential biomarkers and therapeutic targets for COVID-19.

Classification categories:

- COVID-19 symptoms: Non-severe-COVID-19 and Severe-COVID-19 samples.
- No COVID-19 symptoms: Healthy and Symptomatic-non-COVID-19 samples.

Data Preprocessing

Data Selection: Select columns with intensity for peptide variant

Missing Value Interpretation: Zero values in the processed matrix, which represent peptide intensities, are replaced with N/As.

	rowid	ccms_row_id	Algorithm	Filename	Cluster_index	Peptide	Unmodified_sequence	Charge
0	1	1	.MODA.	specs_ms.mgf	960991	K.[304.207]GARLIPEMDQIFTEVEMTTLE(K,304.207).V	.GARLIPEMDQIFTEVEMTTLEK.	4
1	2	2	.MODA.	specs_ms.mgf	763982	I.[304.207]FTEVEMTTLE(K,304.207).V	.FTEVEMTTLEK.	3
2	3	3	.MSGFPLUS.	specs_ms.mgf	902201	K.[304.207]LYQPEYQEVSTEEQR.E	.LYQPEYQEVSTEEQR.	3
3	4	4	.MSGFPLUS.	specs_ms.mgf	935503	K.[304.207]AANSLEAFIFETQD(K,304.207).L	.AANSLEAFIFETQDK.	3
4	5	5	.MODA.	specs_ms.mgf	297961	R.[304.207]YSHDF(N,-56.985)FH.I	.YSHDFNFH.	3
...	...	...	...	...	...	...	...	...
101456	101457	101457	.MODA.	specs_ms.mgf	480358	K.[304.207]YLGE(E,-68.078)YV(K,304.207).A	.YLGEEYVK.	3
101457	101458	101458	.MODA.	specs_ms.mgf	237950	K.[304.207]YL(G,55.921)EEYV(K,304.207).A	.YLGEEYVK.	4
101458	101459	101459	.MODA.	specs_ms.mgf	1037953	K.{187.018}[304.207]YLGEEYV(K,304.207).A	.YLGEEYVK.	2

Data Preprocessing(cont.)

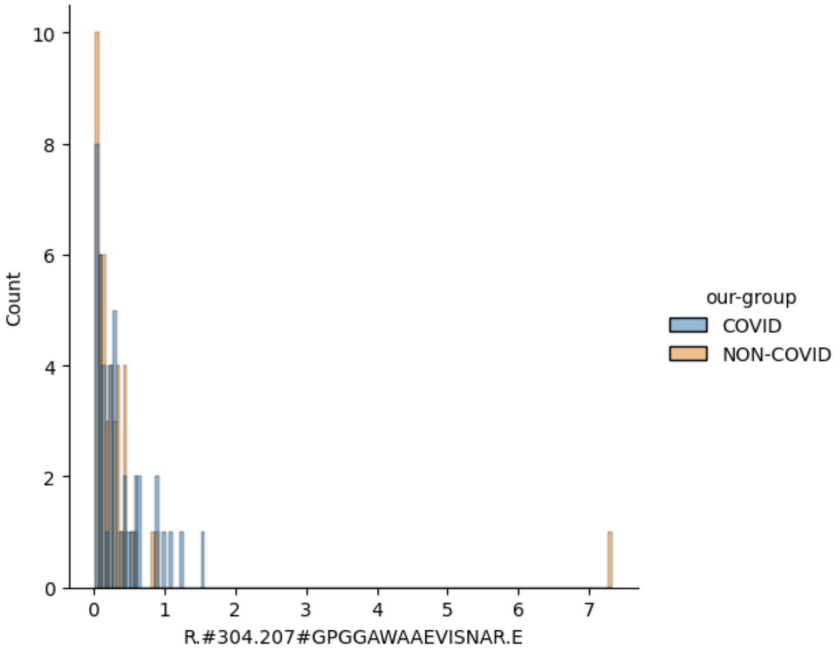
Data Re-organization: The matrix is reorganized so that each peptide is a column, and each row represents a patient sample.

Grouping and Labeling:

- Patient samples are grouped into two classes: COVID and NON-COVID.
  - 'COVID' includes non-severe-COVID-19 and severe-COVID-19 patients.
  - 'NON-COVID' includes healthy and symptomatic-non-COVID-19 patients.
- These groups are used as labels for classification.

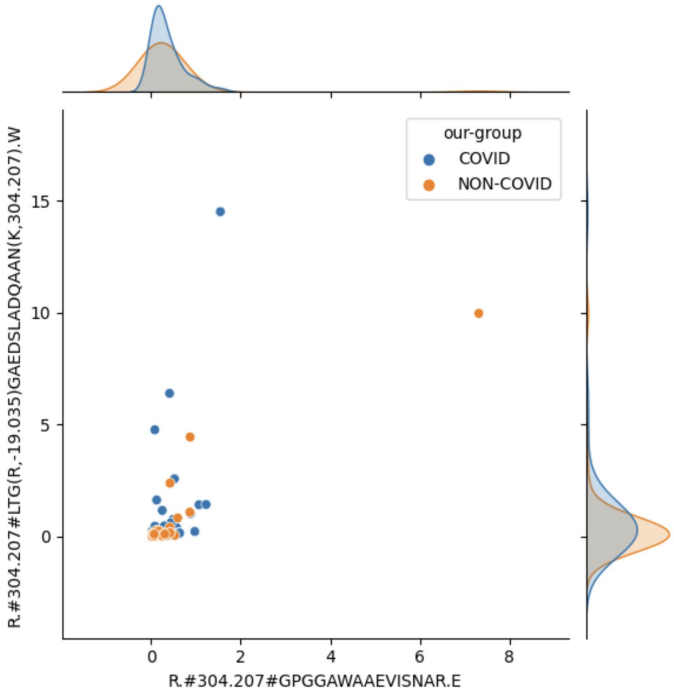
Data Preprocessing - histogram of peptide values

A histogram of peptide values was plotted to visually examine the distribution of peptide "R.#304.207#GPGGAWAAEVISNAR.E", grouped by 'our-group'.

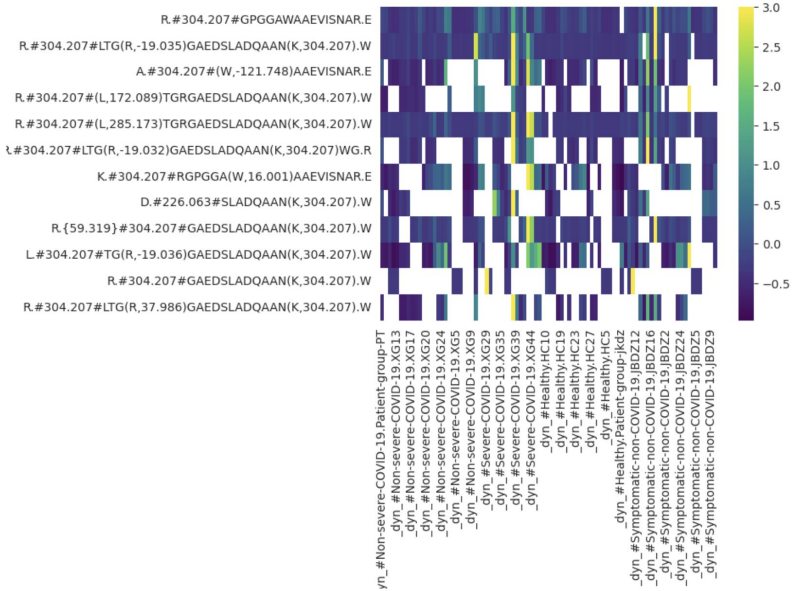


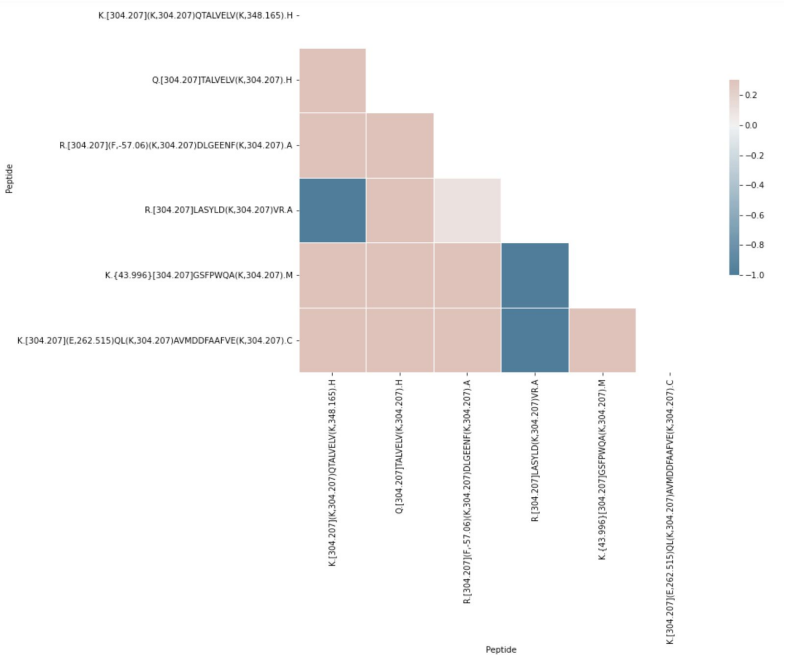
Data Preprocessing - relationship between peptide

2D plot to examine the relationship between the peptide "R.#304.207#GPGGAWAAEVISNAR.E" and "R.#304.207#LTG(R,-19.035)GAEDSLADQAAN(K,304.207).W", separated by 'our-group':



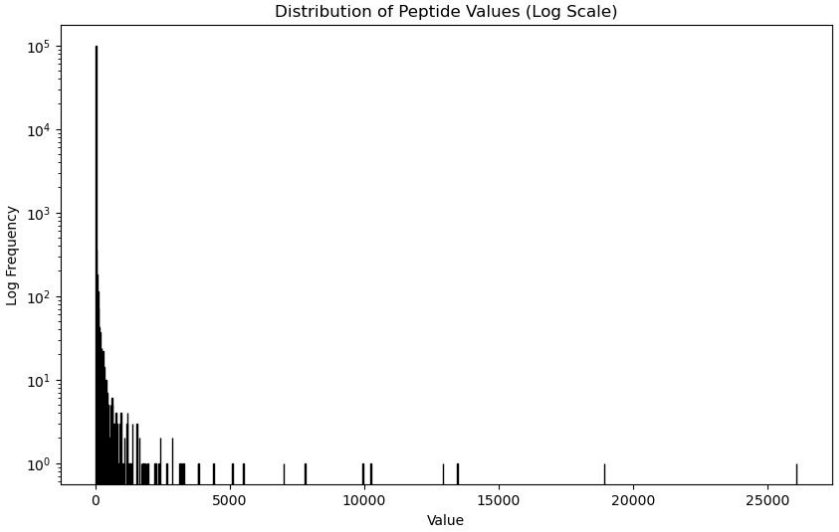
Data Preprocessing - heatmap





Feature Selection

- Variance and correlation of each peptide was calculated to identify potential discriminative features.
- The 99.99th percentile of the variants was calculated and used as the threshold for filtering high variance features.



Feature Selection (cont.)

- The top 100 peptides with the highest variance were selected as input features.
- Final dataset resulted in 90 samples and 104 features, including the label and group features.

```
Our variants threshold is  510.2783044466491
Index(['K.[304.207]Q(C,57.021)S(K,304.207)EDGGGWWYNR.C',
       'R.[304.207]MGPTELLIEMEDW(K,304.207)GD(K,304.207)V(K,304.207).A',
       'R.{303.216}[304.207]TP(C,57.021)TVS(C,57.021)NIPVVSG(K,304.207)E(C,57.021)EEIIR.K',
       'K.[304.207]NLNE(K,247.148)DYELL(C,57.021)LDGTR.K',
       'R.[304.207]SPSQADIN(K,304.207).I',
       'P.[304.207]SVSGSPGQSVTIS(C,57.021)TGT(S,-2.011)SDVGSYNR.V',
       'K.[304.207]GFYPSDIAVEWE(S,78.09)NGQPENNY(K,304.207).T',
       'Q.[304.207]SEDEADYY(C,57.021)A(I,8.99)WYSS.T',
       'R.[304.207](T,-13.936)(K,304.207)NDFTWF(K,304.207).L',
       'K.[304.207](C,57.021)HAGHLNGVY(Y,4.879)QGGTYS.K',
       ...
       '-.[304.207](V,42.011)SFLSALEEYT(K,304.207).K',
       '-.[304.207](V,43.006)QPYLDDFQ(K,304.207).K',
       'K.[304.207]SDVVYTDW(K,248.146).K',
       'K.{43.993}[304.207]SDVVYTDW(K,304.207).K',
       'K.[304.207]SDVVYTDW(K,347.2).K', 'I.[304.207](K,317.837)EAGDAESR.V',
       'K.[304.207]SNEEGSEE(K,304.207)GPEVR.E', '-.AIMDKKANIR.-',
       'R.[304.207]GSGGGSSGGSIGGRGSSSGGV(K,304.207).S',
       '-.SRAQLGGPEAAKSDETAAK.-'],
      dtype='object', name='Peptide', length=101)
```

Dataset Construction

Data Randomization:

To ensure unbiased model evaluation, the dataset was randomized. This process disrupts any inherent order in the data that may artificially inflate the model's performance.

Train/Validation/Test Split:

- The randomized data was then divided into training, validation, and testing datasets.
- 80% of the data was used for training to allow the model to learn the intricate patterns and relationships within the data.
- 10% was used for validation to fine-tune model parameters and prevent overfitting during the training process.
- The remaining 10% was used for testing to evaluate the model's performance on unseen data.

Model Selection

Extreme Gradient Boosting (XGBoost): A decision-tree-based ensemble machine learning algorithm known for its speed, performance, and ability to handle a large number of features.

Random Forest: An ensemble learning method that constructs multiple decision trees, preventing overfitting and handling categorical and numerical data without scaling.

Logistic Regression: A simple yet effective model for binary and multiclass classification, serving as our baseline due to its computational efficiency and interpretability.

Model Fine-tuning

```
# fine-tuning, using f1 score: https://scikit-learn.org/stable/modules/model_evaluation.html
print(xgb.__version__)
params = { 'max_depth': [5, 7, 10, 15],
           'learning_rate': [0.01, 0.1, 0.2, 0.3],
           'subsample': np.arange(0.5, 1.0, 0.1),
           'colsample_bytree': np.arange(0.4, 1.0, 0.1),
           'colsample_bylevel': np.arange(0.4, 1.0, 0.1),
           'n_estimators': [50, 100, 150]}
xgbr = xgb.XGBClassifier(seed = xgb_random_seed)
clf = RandomizedSearchCV(estimator=xgbr,
                         param_distributions=params,
                         scoring='accuracy',
                         n_iter=10, # number of samples in random selection
                         cv = 2, # 2 fold cross validation
                         n_jobs=-1,   # use all processor
                         random_state= xgb_random_seed, # set the seed
                         verbose=4)
clf.fit(X_train+X_val, pd.concat([Y_train, Y_val]))
print("Best parameters:", clf.best_params_)
print("Highest accuracy: ", clf.best_score_)
```

```
1.7.4
Fitting 2 folds for each of 10 candidates, totalling 20 fits
Best parameters: {'subsample': 0.7, 'n_estimators': 50, 'max_depth': 10, 'learning_rate': 0.2, 'colsample_bytree': 0.7, 'colsam
ple_bylevel': 0.7}
```

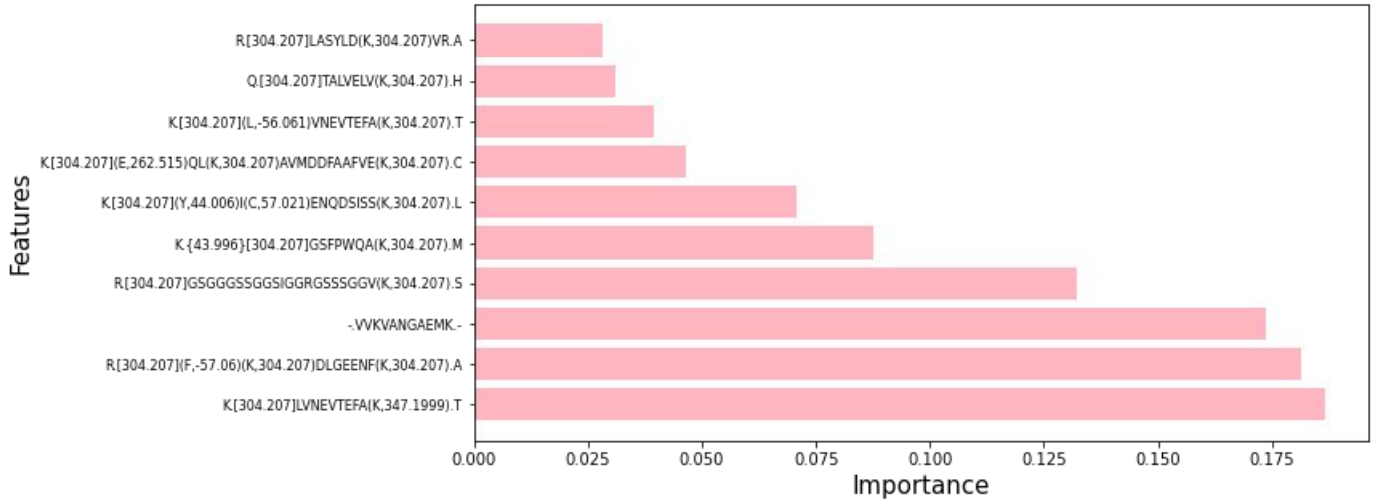
Model results

The model results for 100 features is shown below:

Algorithm	**Best Accuracy**	Precision	Recall	Accuracy	Balanced Accuracy	F1 Score
Logistic Regression	**0.78**	1.0	0.6	0.7778	0.8	0.75
Random Forest	**0.56**	0.6	0.6	0.5556	0.55	0.6
XGBoost	**0.89**	-	-	-	-	-

Model results

Based on XGBoost result, we also calculated the important of features



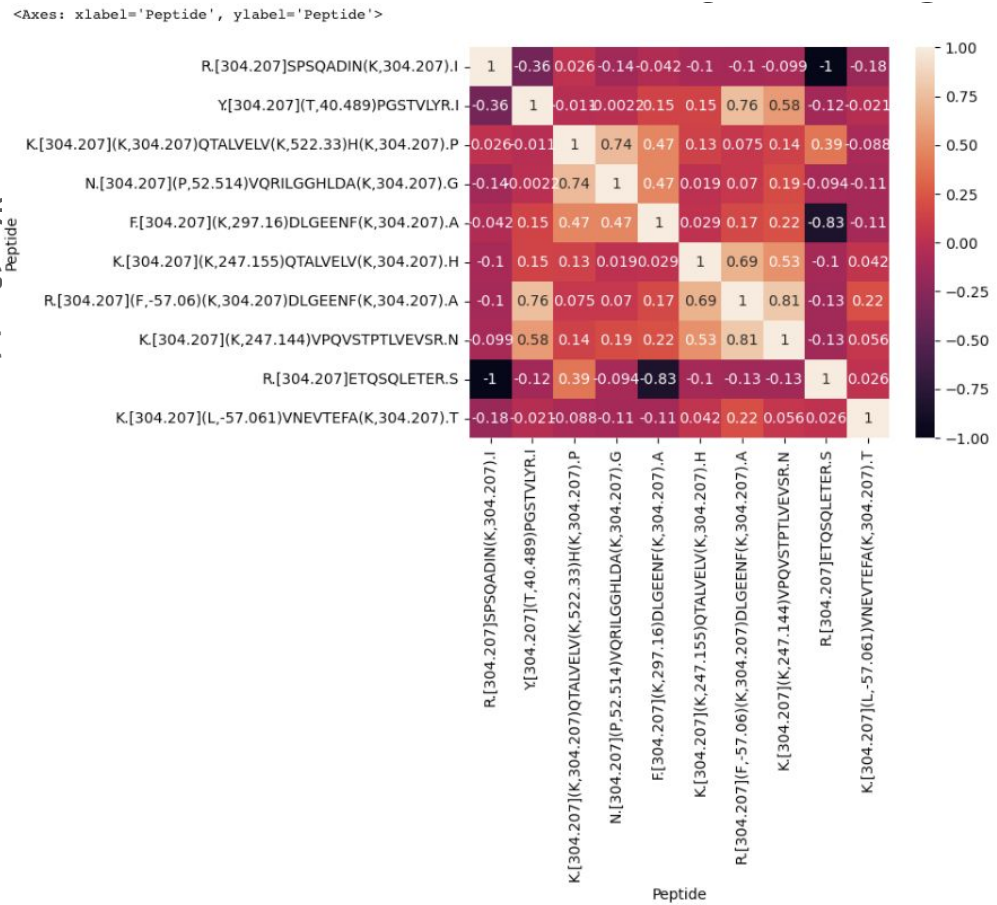
Assessment on Important Features for the Classification

We first identified the top features contributing to the classification task. This was achieved by utilizing the feature importance capability of the XGBoost model.

- 'K.[304.207](K,304.207)QTALVELV(K,522.33)H(K,304.207).P',
- 'R.[304.207]SPSQADIN(K,304.207).I',
- 'Y.[304.207](T,40.489)PGSTVLYR.I',
- 'N.[304.207](P,52.514)VQRILGGHLDA(K,304.207).G',
- 'F.[304.207](K,297.16)DLGEENF(K,304.207).A',
- 'K.[304.207](K,247.155)QTALVELV(K,304.207).H',
- 'R.[304.207](F,-57.06)(K,304.207)DLGEENF(K,304.207).A',
- 'K.[304.207](K,247.144)VPQVSTPTLVEVSR.N',
- 'R.[304.207]ETQSQLETER.S',
- 'K.[304.207](L,-57.061)VNEVTEFA(K,304.207).T'

Assessment (cont.)

To assess how we ssification es) correlate with the target classes hese top 10 features and the t



Assessment on Important Features for the Classification (cont.)

To
fur
va
co

	dyn#Empty	_dyn_#Healthy	_dyn_#Non-severe-COVID-19	_dyn_#Norm	_dyn_#Severe-COVID-19	_dyn_#Symptomatic-non-COVID-19
304.207]SPSQADIN(K,304.207).I	-0.054420	-0.000358	-0.150935	-0.059345	0.410520	-0.172184
[304.207](T,40.489)PGSTVLYR.I	-0.019837	-0.082234	0.256518	-0.020039	-0.083710	-0.110565
K.[304.207] LVELV(K,522.33)H(K,304.207).P	NaN	-0.435757	0.472856	-0.096241	0.256406	-0.251796
N.[304.207] 514)VQRILGGHLDA(K,304.207).G	-0.082981	-0.355232	0.220348	-0.081645	0.400751	-0.201455
K,297.16)DLGEENF(K,304.207).A	-0.095708	-0.389533	0.566488	-0.101019	0.134395	-0.236804
247.155)QTALVELV(K,304.207).H	-0.039478	-0.142940	0.392454	-0.038492	-0.042581	-0.191131
R.[304.207](F,-57.06) 304.207)DLGEENF(K,304.207).A	-0.031382	-0.137852	0.397662	-0.036093	-0.035166	-0.207785
(K,247.144)VPQVSTPTLVEVSR.N	-0.040875	-0.147100	0.396063	-0.038979	-0.025496	-0.203093
R.[304.207]ETQSQLETER.S	-0.095128	-0.111983	-0.177255	-0.093039	-0.156129	0.548287
-57.061)VNEVTEFA(K,304.207).T	-0.014741	-0.074742	0.212223	-0.016710	-0.036346	-0.099141

Assessment on peptide/variant identification

Introduction: This assessment focused on confirming the spectrum identification of the top features and understanding the impact of modifications on peptide spectra.

Top Features: The top five peptides identified in our model were:

- 'R.[304.207]SPSQADIN(K,304.207).I'
- 'Y.[304.207]PGSTVLYR.I'
- 'K.[304.207]QTALVELV(K,522.33)H(K,304.207).P'
- 'N.[304.207]VQRILGGHLDA(K,304.207).G'
- 'F.[304.207]DLGEENF(K,304.207).A'

Confirmation of Spectrum Identification: To validate the spectrum identification for these top peptides, we compared their spectral peaks with reference spectra from the MassIVE database. A high cosine similarity score provides strong evidence for correct identification.

Assessment on peptide/variant identification

Relation of Modified Peptides to Unmodified Spectra: Modifications can significantly alter the spectra of peptides. For the top five modified peptides, we evaluated the relationship between their spectra and the spectra of their unmodified counterparts. We focused on comparing the peak patterns and cosine similarity.

R.[304.207]SPSQADIN(K,304.207).I Update Peptide

Ions:
a ☑1+ ☐2+ ☐3+
b ☑1+ ☑2+ ☑3+
c ☐1+ ☐2+ ☐3+
x ☐1+ ☐2+ ☐3+
y ☑1+ ☑2+ ☑3+
z ☐1+ ☐2+ ☐3+
[Deselect All]

Neutral Loss:
☐ NH_3 (*)
☐ H_2O (o)
☐ H_3PO_4 (p)
☑ Immonium ions
☑ Reporter ions

Mass Type:
◉ Mono ○ Avg

Mass Tol: 0.05
◉ Th ○ ppm

Update

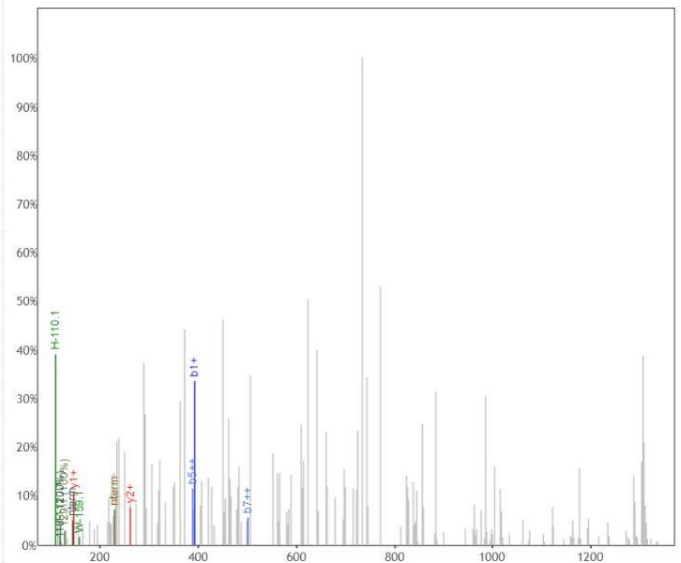
Peak Assignment:
◉ Most Intense
○ Nearest Match
☐ Peak Detect

Peak Labels:
◉ Ion ○ m/z
○ None

Width: 600

Height: 500

SPSQADINK
Spectrum m/z 732.9110, charge 2, MH+ 1464.8142
Identification m/z 421.9003, charge 3, MH+ 1263.6863



Click and drag in the plot to zoom X:☑ Y:☐ Zoom Out Print ☑Enable tooltip ☐Plot mass error

a+	b+	b2+	b3+	#	Seq	#	y+	y2+	y3+
364.2514	392.2463	196.6268	131.4203	1	S	9			
461.3042	489.2991	245.1532	163.7712	2	P	8	872.4472	436.7272	291.4873
548.3362	576.3311	288.6692	192.7819	3	S	7	775.3945	388.2009	259.1363
676.3948	704.3897	352.6985	235.4681	4	Q	6	688.3624	344.6849	230.1257
747.4319	775.4268	388.2170	259.1471	5	A	5	560.3039	280.6556	187.4395
862.4588	890.4537	445.7305	297.4894	6	D	4	489.2667	245.1370	163.7604
975.5429	1003.5378	502.2725	335.1841	7	I	3	374.2398	187.6235	125.4181
1089.5858	1117.5807	559.2940	373.1984	8	N	2	261.1557	131.0815	87.7234
				9	K	1	147.1128	74.0600	49.7091

Explained Intensity: 5.53% (5/8 backbone breaks)

[Click] to move table

Variable Modifications:

S: 304.207 [1]

Y.[304.207](T,40.489)PGSTVLYR.I Update Peptide

Ions:
a ☑1+ ☐2+ ☐3+
b ☑1+ ☑2+ ☑3+
c ☐1+ ☐2+ ☐3+
x ☐1+ ☐2+ ☐3+
y ☑1+ ☑2+ ☑3+
z ☐1+ ☐2+ ☐3+
[Deselect All]

Neutral Loss:
☐ NH_3 (*)
☐ H_2O (o)
☐ H_3PO_4 (p)
☑ Immonium ions
☑ Reporter ions

Mass Type:
◉ Mono ○ Avg

Mass Tol: 0.05
◉ Th ○ ppm
Update

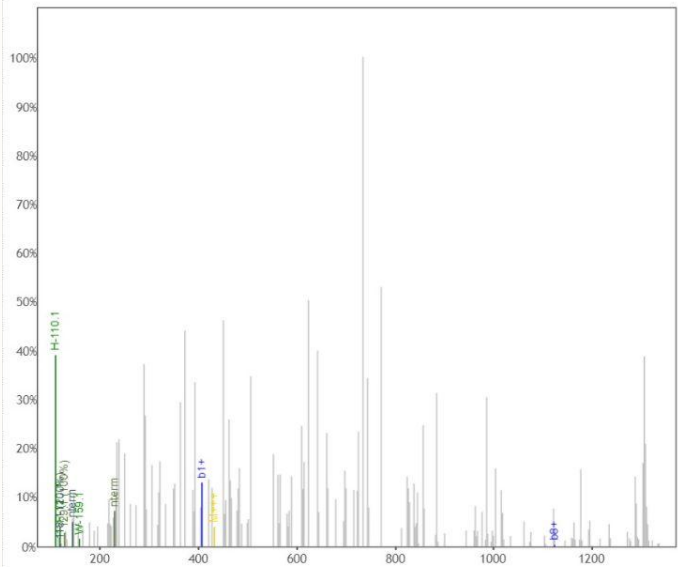
Peak Assignment:
◉ Most Intense
○ Nearest Match
☐ Peak Detect

Peak Labels:
◉ Ion ○ m/z
○ None

Width: 600

Height: 500

TPGSTVLYR
Spectrum m/z 732.9110, charge 2, MH+ 1464.8142
Identification m/z 433.2526, charge 3, MH+ 1297.7434



Click and drag in the plot to zoom X:☑ Y:☐ Zoom Out Print ☑Enable tooltip ☐Plot mass error

a+	b+	b2+	b3+	#	Seq	#	y+	y2+	y3+
378.2670	**406.2620**	203.6346	136.0922	1	T	9			
475.3198	503.3147	252.1610	168.4431	2	P	8	892.4887	446.7480	298.1677
532.3413	560.3362	280.6717	187.4502	3	G	7	795.4359	398.2216	265.8168
619.3733	647.3682	324.1877	216.4609	4	S	6	738.4145	369.7109	246.8097
720.4210	748.4159	374.7116	250.1435	5	T	5	651.3824	326.1949	217.7990
819.4894	847.4843	424.2458	283.1663	6	V	4	550.3348	275.6710	184.1164
932.5735	960.5684	480.7878	320.8610	7	L	3	451.2663	226.1368	151.0936
1095.6368	**1123.6317**	562.3195	375.2154	8	Y	2	338.1823	169.5948	113.3989
				9	R	1	175.1190	88.0631	59.0445

Explained Intensity: 3.32% (2/8 backbone breaks)

[Click] to move table

Variable Modifications:

T: 304.207 [1]

K.[304.207](K,304.207)QTALVELV(K,522.33)H(K,304.207).P Update Peptide

Ions:
a ☑1+ ☐2+ ☐3+
b ☑1+ ☑2+ ☑3+
c ☐1+ ☐2+ ☐3+
x ☐1+ ☐2+ ☐3+
y ☑1+ ☑2+ ☑3+
z ☐1+ ☐2+ ☐3+
[Deselect All]

Neutral Loss:
☐ NH_3 (*)
☐ H_2O (o)
☐ H_3PO_4 (p)
☑Immonium ions
☑Reporter ions

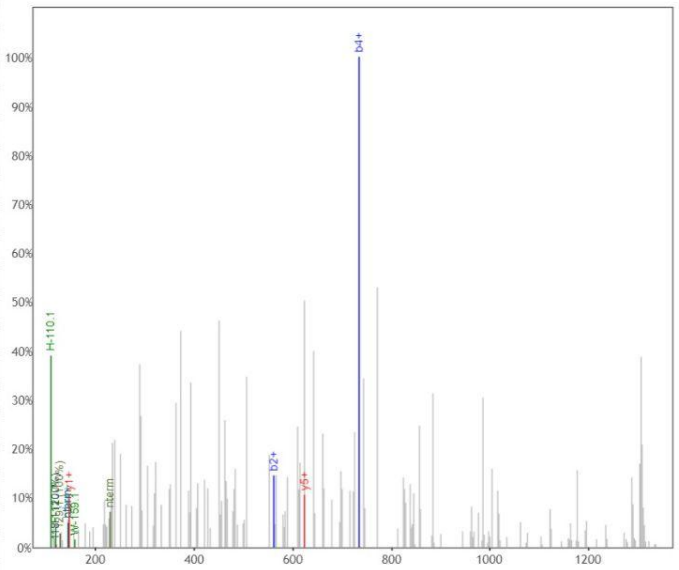
Mass Type: ◉Mono ○Avg
Mass Tol: 0.05
◉Th ○ppm
Update

Peak Assignment: ◉Most Intense ○Nearest Match
☐Peak Detect

Peak Labels: ◉Ion ○m/z ○None

Width: 600
Height: 500

KQTALVELVKHK
Spectrum m/z 732.9110, charge 2, MH+ 1464.8142
Identification m/z 566.6914, charge 3, MH+ 1698.0596



Click and drag in the plot to zoom X:☑ Y:☐ Zoom Out Print ☑Enable tooltip ☐Plot mass error

a+	b+	b2+	b3+	#	Seq	#	y+	y2+	y3+
405.3143	433.3092	217.1583	145.1079	1	K	12			
533.3729	561.3678	281.1875	187.7941	2	Q	11	1265.7576	633.3824	422.5907
634.4206	662.4155	331.7114	221.4767	3	T	10	1137.6990	569.3531	379.9045
705.4577	733.4526	367.2299	245.1557	4	A	9	1036.6513	518.8293	346.2220
818.5418	846.5367	423.7720	282.8504	5	L	8	965.6142	483.3108	322.5429
917.6102	945.6051	473.3062	315.8732	6	V	7	852.5302	426.7687	284.8482
1046.6528	1074.6477	537.8275	358.8874	7	E	6	753.4617	377.2345	251.8254
1159.7368	1187.7317	594.3695	396.5821	8	L	5	624.4192	312.7132	208.8112
1258.8052	1286.8002	643.9037	429.6049	9	V	4	511.3351	256.1712	171.1165
1386.9002	1414.8951	707.9512	472.3032	10	K	3	412.2667	206.6370	138.0937
1523.9591	1551.9540	776.4807	517.9895	11	H	2	284.1717	142.5895	95.3954
				12	K	1	147.1128	74.0600	49.7091

Explained Intensity: 8.40% (5/11 backbone breaks)

[Click] to move table

Variable Modifications:

K: 304.207 [1]

N.[304.207](P,52.514)VQRILGGHLDA(K,304.207).G Update Peptide

Ions:
a ☑1+ ☐2+ ☐3+
b ☑1+ ☑2+ ☑3+
c ☐1+ ☐2+ ☐3+
x ☐1+ ☐2+ ☐3+
y ☑1+ ☑2+ ☑3+
z ☐1+ ☐2+ ☐3+
[Deselect All]

Neutral Loss:
☐ NH_3 (*)
☐ H_2O (o)
☐ H_3PO_4 (p)
☑ Immonium ions
☑ Reporter ions

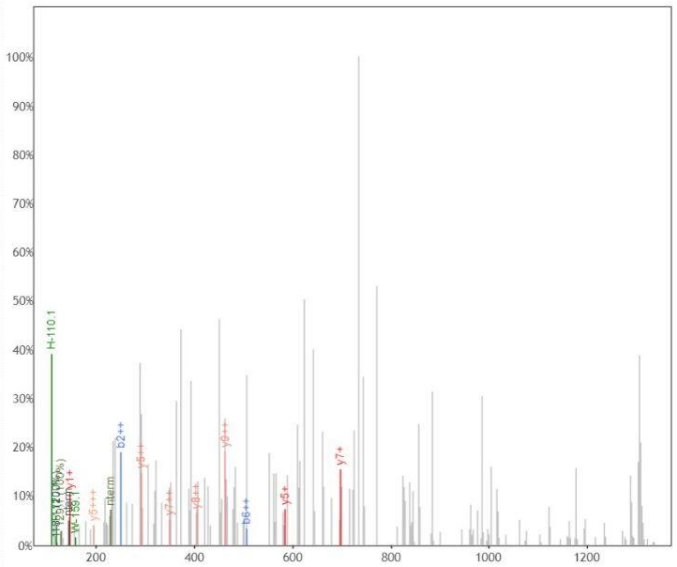
Mass Type: ◉ Mono ○ Avg
Mass Tol: 0.05
◉ Th ○ ppm
Update

Peak Assignment:
◉ Most Intense
○ Nearest Match
☐ Peak Detect

Peak Labels:
◉ Ion ○ m/z
○ None

Width: 600
Height: 500

PVQRILGGHLDAK
Spectrum m/z 732.9110, charge 2, MH+ 1464.8142
Identification m/z 570.0111, charge 3, MH+ 1708.0188



Click and drag in the plot to zoom X:☑ Y:☐ Zoom Out Print ☑Enable tooltip ☐Plot mass error

a+	b+	b2+	b3+	#	Seq	#	y+	y2+	y3+
374.2721	402.2670	201.6372	134.7605	1	P	13			
473.3405	501.3355	**251.1714**	167.7833	2	V	12	1306.7590	653.8831	436.2578
601.3991	629.3940	315.2007	210.4695	3	Q	11	1207.6906	604.3489	403.2350
757.5002	785.4951	393.2512	262.5032	4	R	10	1079.6320	540.3196	360.5489
870.5843	898.5792	449.7932	300.1979	5	I	9	923.5309	462.2691	308.5151
983.6684	1011.6633	**506.3353**	337.8926	6	L	8	810.4468	405.7271	270.8205
1040.6898	1068.6847	534.8460	356.8998	7	G	7	**697.3628**	349.1850	233.1258
1097.7113	1125.7062	563.3567	375.9069	8	G	6	640.3413	320.6743	214.1186
1234.7702	1262.7651	631.8862	421.5932	9	H	5	**583.3198**	292.1636	195.1115
1347.8543	1375.8492	688.4282	459.2879	10	L	4	446.2609	223.6341	149.4252
1462.8812	1490.8761	745.9417	497.6302	11	D	3	333.1769	167.0921	111.7305
1533.9183	1561.9132	781.4603	521.3093	12	A	2	218.1499	109.5786	73.3882
				13	K	1	**147.1128**	74.0600	49.7091

Explained Intensity: 7.10% (6/12 backbone breaks)

[Click] to move table

Variable Modifications:

P: 304.207 [1]

F.[304.207](K,297.16)DLGEENF(K,304.207).A Update Peptide

Ions:
a ☑1+ ☐2+ ☐3+
b ☑1+ ☑2+ ☑3+
c ☐1+ ☐2+ ☐3+
x ☐1+ ☐2+ ☐3+
y ☑1+ ☑2+ ☑3+
z ☐1+ ☐2+ ☐3+
[Deselect All]

Neutral Loss:
☐ NH_3 (*)
☐ H_2O (o)
☐ H_3PO_4 (p)
☑ Immonium ions
☑ Reporter ions

Mass Type:
◉ Mono ○ Avg

Mass Tol: 0.05
◉ Th ○ ppm
Update

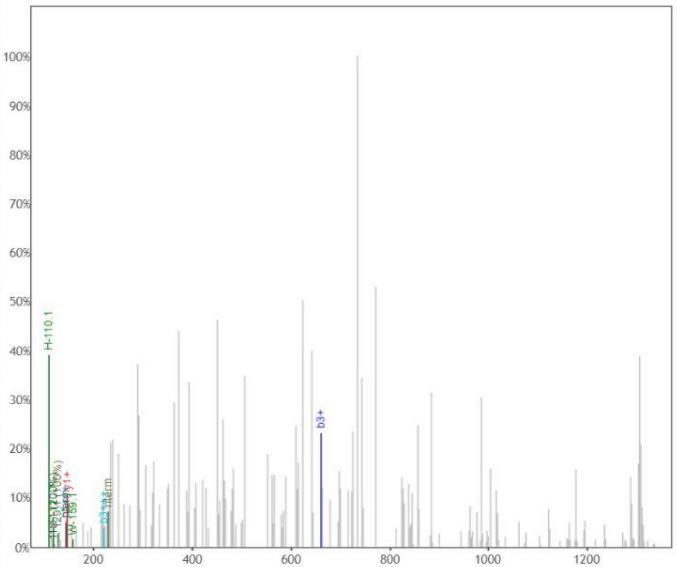
Peak Assignment:
◉ Most Intense
○ Nearest Match
☐ Peak Detect

Peak Labels:
◉ Ion ○ m/z
○ None

Width: 600

Height: 500

KDLGEENFK
Spectrum m/z 732.9110, charge 2, MH+ 1464.8142
Identification m/z 461.9194, charge 3, MH+ 1383.7438



Click and drag in the plot to zoom X:☑ Y:☐ Zoom Out Print ☑Enable tooltip ☐Plot mass error

a+	b+	b2+	b3+	#	Seq	#	y+	y2+	y3+
405.3143	433.3092	217.1583	145.1079	1	K	9			
520.3413	548.3362	274.6717	183.4502	2	D	8	951.4418	476.2245	317.8188
633.4253	661.4202	331.2138	221.1449	3	L	7	836.4149	418.7111	279.4765
690.4468	718.4417	359.7245	240.1521	4	G	6	723.3308	362.1690	241.7818
819.4894	847.4843	424.2458	283.1663	5	E	5	666.3093	333.6583	222.7746
948.5320	976.5269	488.7671	326.1805	6	E	4	537.2667	269.1370	179.7604
1062.5749	1090.5698	545.7885	364.1948	7	N	3	408.2241	204.6157	136.7462
1209.6433	1237.6382	619.3228	413.2176	8	F	2	294.1812	147.5942	98.7319
				9	K	1	147.1128	74.0600	49.7091

Explained Intensity: 4.19% (3/8 backbone breaks)

[Click] to move table

Variable Modifications:

K: 304.207 [1]

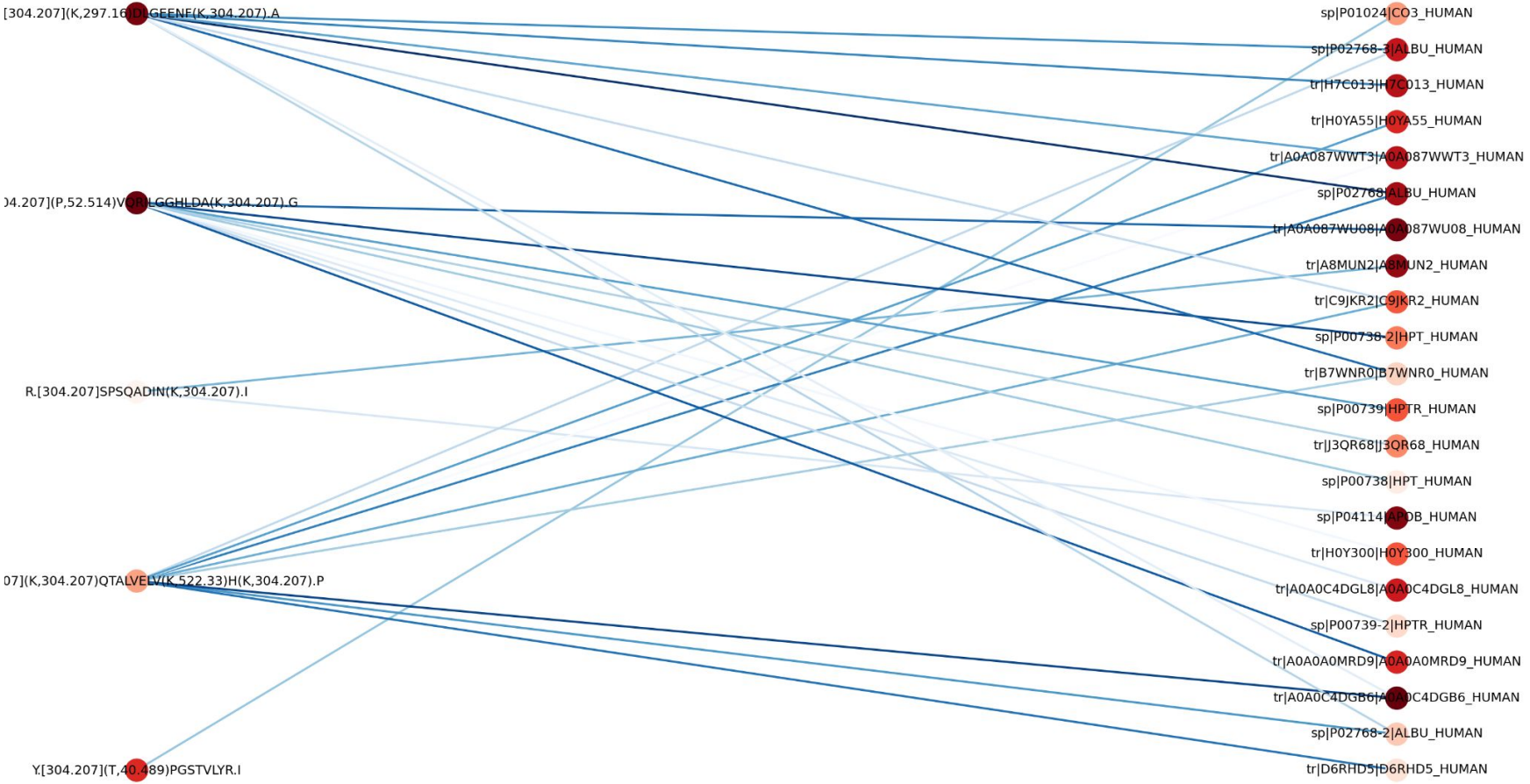
Assessment on Protein Identification

Objective: To examine the relationships between peptides identified in the peptide identification assessment phase and their related proteins using top 5 peptides from proteomics data.

Bipartite Graph: Visual representation of peptide-protein mapping. Nodes on the left represent peptides, on the right - proteins. Edges represent peptide's presence in a protein.

Findings:

- Unique peptide-protein match: 'Y.304.207PGSTVLYR.I' with 'sp|P01024|CO3_HUMAN'. High certainty in identification.
- Seven proteins identified with multiple peptides. Increases reliability of protein identification.


[304.207](K,297.16)DLGEENF(K,304.207).A
04.207](P,52.514)VQRLGGHLDA(K,304.207).G
R.[304.207]SPSQADIN(K,304.207).I
07](K,304.207)QTALVELV(K,522.33)H(K,304.207).P
Y.[304.207](T,40.489)PGSTVLYR.I
sp|P01024|CO3_HUMAN
sp|P02768-3|ALBU_HUMAN
tr|H7C013|H7C013_HUMAN
tr|H0YA55|H0YA55_HUMAN
tr|A0A087WWT3|A0A087WWT3_HUMAN
sp|P02768|ALBU_HUMAN
tr|A0A087WU08|A0A087WU08_HUMAN
tr|A8MUN2|A8MUN2_HUMAN
tr|C9JKR2|C9JKR2_HUMAN
sp|P00738-2|HPT_HUMAN
tr|B7WNR0|B7WNR0_HUMAN
sp|P00739|HPTR_HUMAN
tr|J3QR68|J3QR68_HUMAN
sp|P00738|HPT_HUMAN
sp|P04114|APOB_HUMAN
tr|H0Y300|H0Y300_HUMAN
tr|A0A0C4DGL8|A0A0C4DGL8_HUMAN
sp|P00739-2|HPTR_HUMAN
tr|A0A0A0MRD9|A0A0A0MRD9_HUMAN
tr|A0A0C4DGB6|A0A0C4DGB6_HUMAN
sp|P02768-2|ALBU_HUMAN
tr|D6RHD5|D6RHD5_HUMAN

Assessment on Protein Identification (cont.)

```
Number of unique peptides: 1
Number of proteins with multiple peptides identified: 7
----------------------
Peptide to Protein:
Y.[304.207](T,40.489)PGSTVLYR.I → sp|P01024|CO3_HUMAN
----------------------
Protein to Peptide:
sp|P04114|APOB_HUMAN → R.[304.207]SPSQADIN(K,304.207).I
tr|A8MUN2|A8MUN2_HUMAN → R.[304.207]SPSQADIN(K,304.207).I
sp|P01024|CO3_HUMAN → Y.[304.207](T,40.489)PGSTVLYR.I
tr|D6RHD5|D6RHD5_HUMAN → K.[304.207](K,304.207)QTALVELV(K,522.33)H(K,304.207).P
tr|H0YA55|H0YA55_HUMAN → K.[304.207](K,304.207)QTALVELV(K,522.33)H(K,304.207).P
sp|P00738-2|HPT_HUMAN → N.[304.207](P,52.514)VQRILGGHLDA(K,304.207).G
sp|P00738|HPT_HUMAN → N.[304.207](P,52.514)VQRILGGHLDA(K,304.207).G
sp|P00739-2|HPTR_HUMAN → N.[304.207](P,52.514)VQRILGGHLDA(K,304.207).G
sp|P00739|HPTR_HUMAN → N.[304.207](P,52.514)VQRILGGHLDA(K,304.207).G
tr|A0A087WU08|A0A087WU08_HUMAN → N.[304.207](P,52.514)VQRILGGHLDA(K,304.207).G
tr|A0A0A0MRD9|A0A0A0MRD9_HUMAN → N.[304.207](P,52.514)VQRILGGHLDA(K,304.207).G
tr|A0A0C4DGL8|A0A0C4DGL8_HUMAN → N.[304.207](P,52.514)VQRILGGHLDA(K,304.207).G
tr|H0Y300|H0Y300_HUMAN → N.[304.207](P,52.514)VQRILGGHLDA(K,304.207).G
tr|J3QR68|J3QR68_HUMAN → N.[304.207](P,52.514)VQRILGGHLDA(K,304.207).G
tr|H7C013|H7C013_HUMAN → F.[304.207](K,297.16)DLGEENF(K,304.207).A
```

Q&A