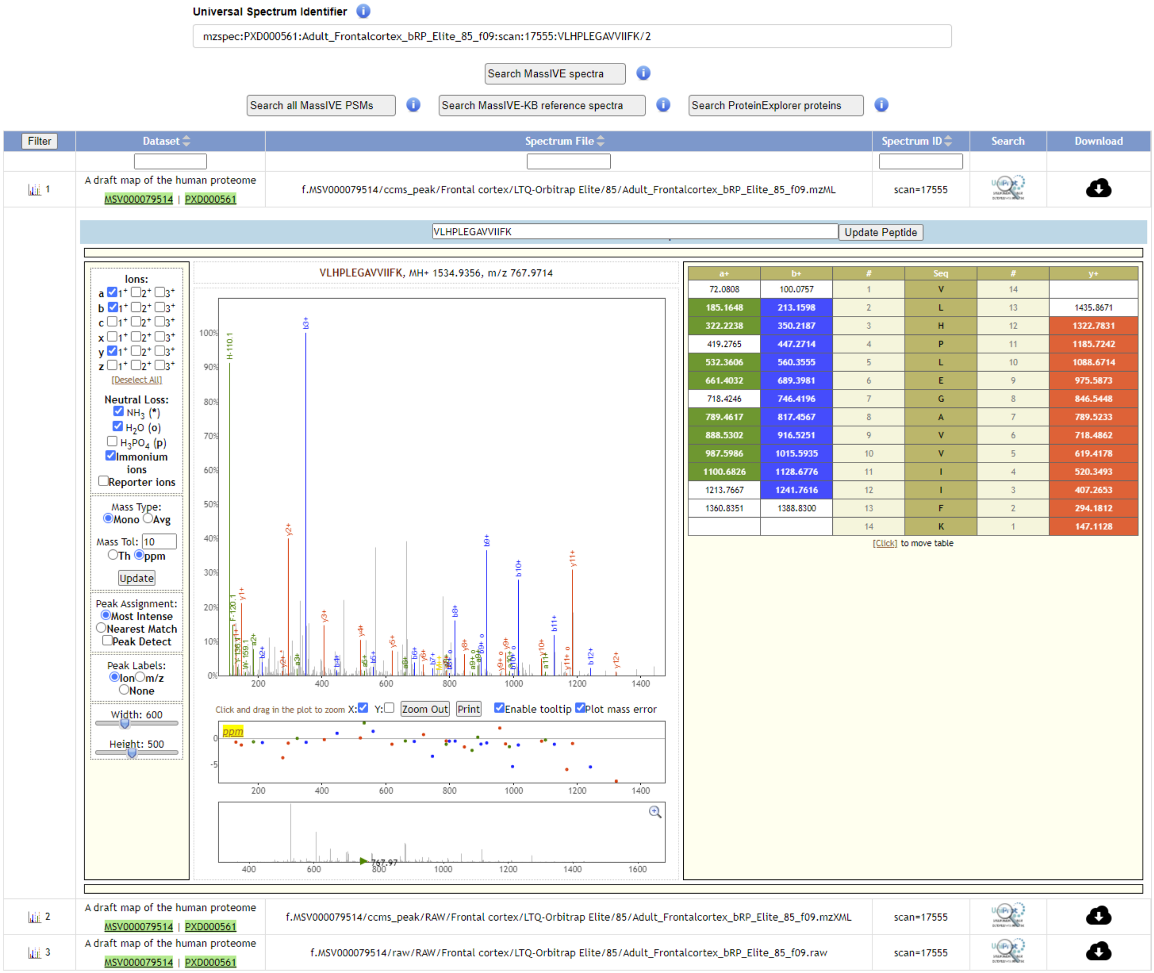
Supplementary Information

Universal Spectrum Identifier for mass spectra

Eric W. Deutsch[1,*], Yasset Perez-Riverol[2], Jeremy Carver[3], Shin Kawano[4], Luis Mendoza[1], Tim Van Den Bossche[5,6], Ralf Gabriels[5,6], Pierre-Alain Binz[7], Benjamin Pullman[3], Zhi Sun[1], Jim Shofstahl[8], Wout Bittremieux[9,10], Tytus D. Mak[11], Joshua Klein[12], Yunping Zhu[13], Henry Lam[14], Juan Antonio Vizcaíno[2], and Nuno Bandeira[3,9,*]



Universal Spectrum Identifier

mzspec:PXD000561:Adult_Frontalcortex_bRP_Elite_85_f09:scan:17555:VLHPLEGAVVIIFK/2

Filter	Dataset	Spectrum File	Spectrum ID	Search	Download
1	A draft map of the human proteome MSV000079514 | PXD000561	f.MSV000079514/ccms_peak/Frontal cortex/LTQ-Orbitrap Elite/85/Adult_Frontalcortex_bRP_Elite_85_f09.mzML	scan=17555		
2	A draft map of the human proteome MSV000079514 | PXD000561	f.MSV000079514/ccms_peak/RAW/Frontal cortex/LTQ-Orbitrap Elite/85/Adult_Frontalcortex_bRP_Elite_85_f09.mzXML	scan=17555		
3	A draft map of the human proteome MSV000079514 | PXD000561	f.MSV000079514/raw/RAW/Frontal cortex/LTQ-Orbitrap Elite/85/Adult_Frontalcortex_bRP_Elite_85_f09.raw	scan=17555		

VLHPLEGAVVIIFK

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: 10 ○Th ●ppm

Update

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

Peak Labels: ●Ion ○m/z ○None

Width: 600

Height: 500

VLHPLEGAVVIIFK, MH+ 1534.9356, m/z 767.9714

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

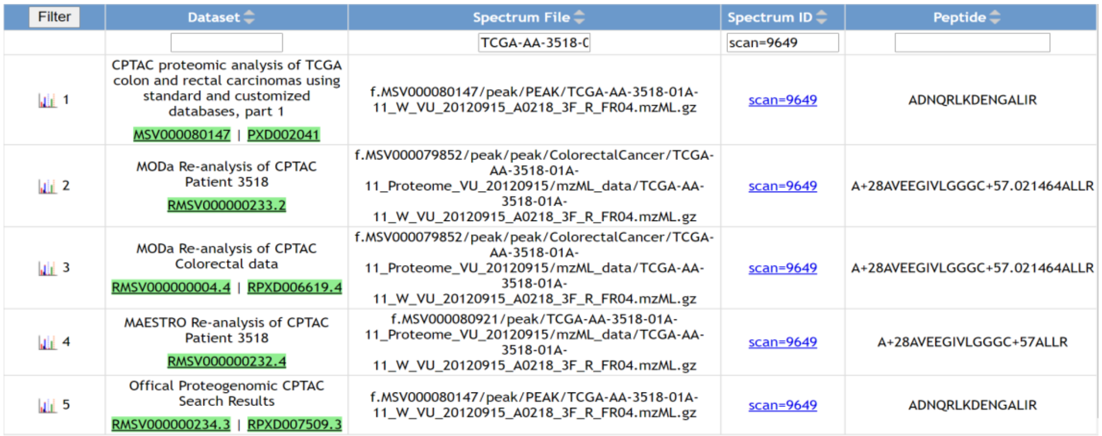
a+	b+	#	Seq	#	y+
72.0808	100.0757	1	V	14	
185.1648	213.1598	2	L	13	1435.8671
322.2238	350.2187	3	H	12	1322.7831
419.2765	447.2714	4	P	11	1185.7242
532.3606	560.3555	5	L	10	1088.6714
661.4032	689.3981	6	E	9	975.5873
718.4246	746.4196	7	G	8	846.5448
789.4617	817.4567	8	A	7	789.5233
888.5302	916.5251	9	V	6	718.4862
987.5986	1015.5935	10	V	5	619.4178
1100.6826	1128.6776	11	I	4	520.3493
1213.7667	1241.7616	12	I	3	407.2653
1360.8351	1388.8300	13	F	2	294.1812
		14	K	1	147.1128

[Click] to move table

Figure S1. Main MassIVE functionality for resolution of Universal Spectrum Identifiers.

(a) Searching for all reported identifications of the USI spectrum

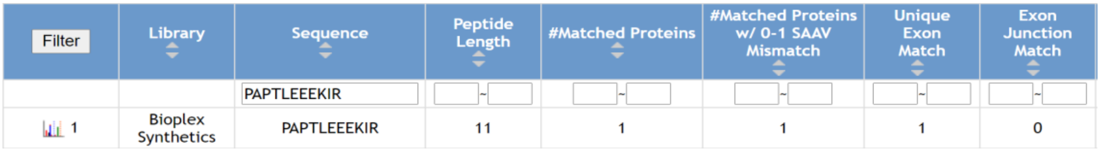
mzspec:PXD000561:Adult_Frontalcortex_bRP_Elite_85_f09:scan:17555:VLHPLEGAVVIIFK/2



Filter	Dataset	Spectrum File	Spectrum ID	Peptide
		TCGA-AA-3518-C	scan=9649	
1	CPTAC proteomic analysis of TCGA colon and rectal carcinomas using standard and customized databases, part 1 MSV000080147 | PXD002041	f.MSV000080147/peak/PEAK/TCGA-AA-3518-01A-11_W_VU_20120915_A0218_3F_R_FR04.mzML.gz	scan=9649	ADNQRLKDENGALIR
2	MODa Re-analysis of CPTAC Patient 3518 RMSV000000233.2	f.MSV000079852/peak/peak/ColorectalCancer/TCGA-AA-3518-01A-11_Proteome_VU_20120915/mzML_data/TCGA-AA-3518-01A-11_W_VU_20120915_A0218_3F_R_FR04.mzML.gz	scan=9649	A+28AVEEGIVLGGGC+57.021464ALLR
3	MODa Re-analysis of CPTAC Colorectal data RMSV000000004.4 | RPXD006619.4	f.MSV000079852/peak/peak/ColorectalCancer/TCGA-AA-3518-01A-11_Proteome_VU_20120915/mzML_data/TCGA-AA-3518-01A-11_W_VU_20120915_A0218_3F_R_FR04.mzML.gz	scan=9649	A+28AVEEGIVLGGGC+57.021464ALLR
4	MAESTRO Re-analysis of CPTAC Patient 3518 RMSV000000232.4	f.MSV000080921/peak/TCGA-AA-3518-01A-11_Proteome_VU_20120915/mzML_data/TCGA-AA-3518-01A-11_W_VU_20120915_A0218_3F_R_FR04.mzML.gz	scan=9649	A+28AVEEGIVLGGGC+57ALLR
5	Offical Proteogenomic CPTAC Search Results RMSV000000234.3 | RPXD007509.3	f.MSV000080147/peak/PEAK/TCGA-AA-3518-01A-11_W_VU_20120915_A0218_3F_R_FR04.mzML.gz	scan=9649	ADNQRLKDENGALIR

(b) Searching MassIVE-KB for reference spectra for the same peptide as in the USI

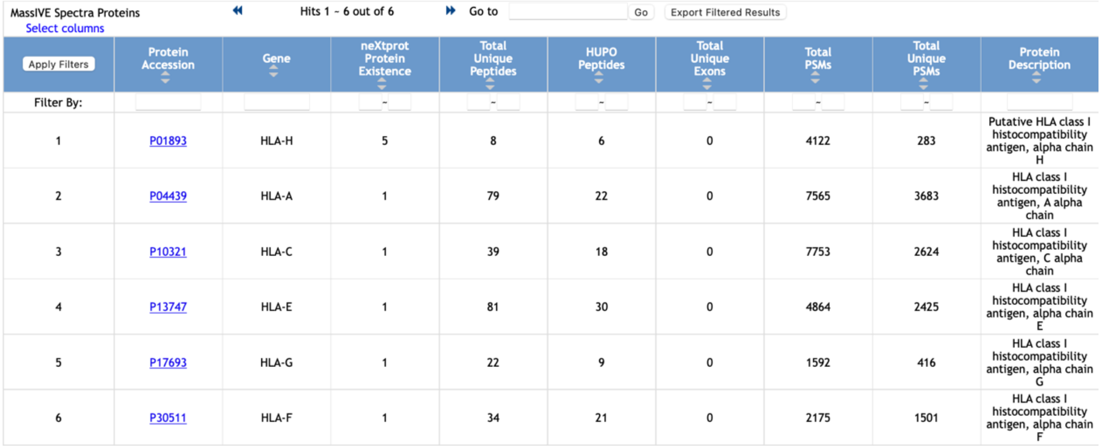
mzspec:PXD002041:TCGA-AA-3518-01A-11_W_VU_20120915_A0218_3F_R_FR04:scan:9649:ADNQRLKDENGALIR/2



Filter	Library	Sequence	Peptide Length	#Matched Proteins	#Matched Proteins w/ 0-1 SAAV Mismatch	Unique Exon Match	Exon Junction Match
		PAPTLEEEKIR	~	~	~	~	~
1	Bioplex Synthetics	PAPTLEEEKIR	11	1	1	1	0

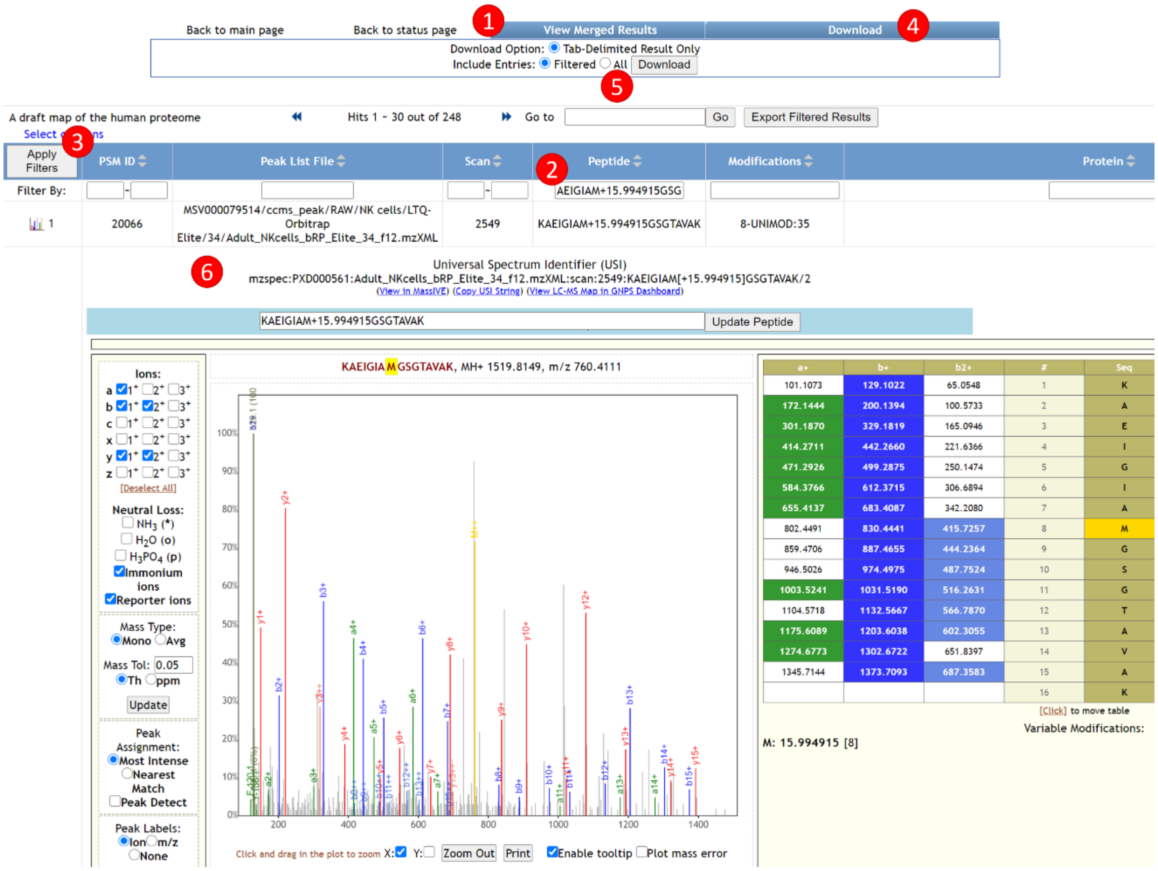
(c) Searching Protein Explorer for proteins containing USI peptide

mzspec:PXD000865:00644_H11_P004899_B0P_A00_R1:scan:3727:PAGDGTFQK/2



Apply Filters	Protein Accession	Gene	neXtprot Protein Existence	Total Unique Peptides	HUPO Peptides	Total Unique Exons	Total PSMs	Total Unique PSMs	Protein Description
Filter By:			~	~	~	~	~	~	
1	P01893	HLA-H	5	8	6	0	4122	283	Putative HLA class I histocompatibility antigen, alpha chain H
2	P04439	HLA-A	1	79	22	0	7565	3683	HLA class I histocompatibility antigen, A alpha chain
3	P10321	HLA-C	1	39	18	0	7753	2624	HLA class I histocompatibility antigen, C alpha chain
4	P13747	HLA-E	1	81	30	0	4864	2425	HLA class I histocompatibility antigen, alpha chain E
5	P17693	HLA-G	1	22	9	0	1592	416	HLA class I histocompatibility antigen, alpha chain G
6	P30511	HLA-F	1	34	21	0	2175	1501	HLA class I histocompatibility antigen, alpha chain F

Figure S2. Advanced MassIVE functionality for resolution of Universal Spectrum Identifiers.



Download Option: ● Tab-Delimited Result Only
Include Entries: ● Filtered ○ All Download

A draft map of the human proteome Hits 1 ~ 30 out of 248 Go to

	PSM ID	Peak List File	Scan	Peptide	Modifications	Protein
Filter By:				AEIGIAM+15.994915GSG		
1	20066	MSV000079514/ccms_peak/RAW/NK cells/LTQ-Orbitrap Elite/34/Adult_NKcells_bRP_Elite_34_f12.mzXML	2549	KAEIGIAM+15.994915GSGTAVAK	8-UNIMOD:35	

Universal Spectrum Identifier (USI)
mzspec:PXD000561:Adult_NKcells_bRP_Elite_34_f12.mzXML:scan:2549:KAEIGIAM[+15.994915]GSGTAVAK/2

KAEIGIAM+15.994915GSGTAVAK

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+

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

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

Peak Labels: ● Ion ○ m/z ○ None

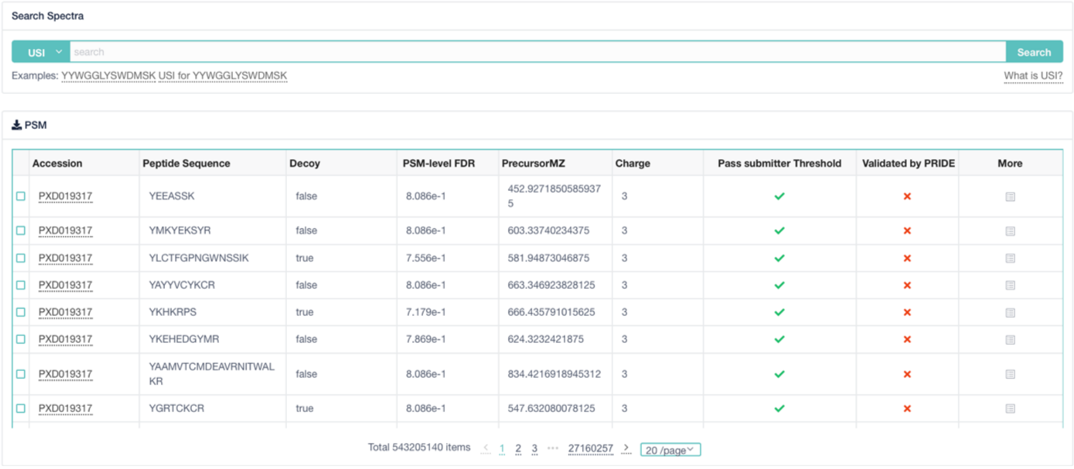
KAEIGIAMGSGTAVAK, MH+ 1519.8149, m/z 760.4111

X: ☑ Y: ☐ ☑ Enable tooltip ☐ Plot mass error

a+	b+	b2+	#	Seq
101.1073	129.1022	65.0548	1	K
172.1444	200.1394	100.5733	2	A
301.1870	329.1819	165.0946	3	E
414.2711	442.2660	221.6366	4	I
471.2926	499.2875	250.1474	5	G
584.3766	612.3715	306.6894	6	I
655.4137	683.4087	342.2080	7	A
802.4491	830.4441	415.7257	8	M
859.4706	887.4655	444.2364	9	G
946.5026	974.4975	487.7524	10	S
1003.5241	1031.5190	516.2631	11	G
1104.5718	1132.5667	566.7870	12	T
1175.6089	1203.6038	602.3055	13	A
1274.6773	1302.6722	651.8397	14	V
1345.7144	1373.7093	687.3583	15	A
			16	K

Variable Modifications:
M: 15.994915 [8]

Figure S3. MassIVE functionality for exporting Universal Spectrum Identifiers for subsets of identified spectra, as is commonly important for identifications of modified peptides. This example illustrates how to export USIs for all identifications with an oxidized methionine on sequence AEIGIAMGSGTAVAK in one of the original human draft proteome datasets. Starting with "Browse Results" in the MassIVE dataset page for PXD000561, the figure illustrates that **(1)** clicking View Merged Results shows all identifications in the dataset, regardless of the number of separate searches; **(2)** entering a filter string in the Peptide field and **(3)** clicking Apply Filters selects all spectrum identifications matching the filter text; **(4)** clicking the Download tab and **(5)** selecting to download Filtered results produces a TSV file with all USIs for all filtered results that can be submitted with a manuscript to support the review process for the manuscript associated with a dataset. In addition, **(6)** USIs are also show for every row in the filtered results when clicking on the spectrum icon (leftmost column) to visualize the spectrum identification. The same results can also be communicated and accessed directly using the same URL that was used to generate this figure.

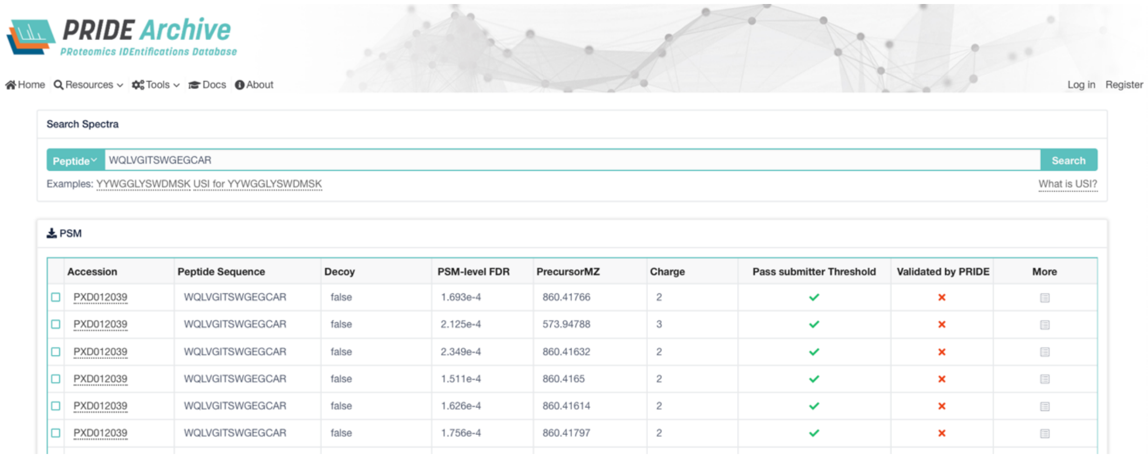


Examples: YYWGGLYSWDMSK USI for YYWGGLYSWDMSK

	Accession	Peptide Sequence	Decoy	PSM-level FDR	PrecursorMZ	Charge	Pass submitter Threshold	Validated by PRIDE	More
☐	PXD019317	YEEASSK	false	8.086e-1	452.9271850585937 5	3	✓	✗	▤
☐	PXD019317	YMKYEKSYR	false	8.086e-1	603.33740234375	3	✓	✗	▤
☐	PXD019317	YLCTFGPNGWNSSIK	true	7.556e-1	581.94873046875	3	✓	✗	▤
☐	PXD019317	YAYYVCYKCR	false	8.086e-1	663.346923828125	3	✓	✗	▤
☐	PXD019317	YKHKRPS	true	7.179e-1	666.435791015625	3	✓	✗	▤
☐	PXD019317	YKEHEDGYMR	false	7.869e-1	624.3232421875	3	✓	✗	▤
☐	PXD019317	YAAMVTCMDEAVRNITWAL KR	false	8.086e-1	834.4216918945312	3	✓	✗	▤
☐	PXD019317	YGRTCKCR	true	8.086e-1	547.632080078125	3	✓	✗	▤

Total 543205140 items < 1 2 3 ••• 27160257 > 20 /page

Figure S4. PRIDE Archive Spectra (https://www.ebi.ac.uk/pride/archive/spectra) provides access to millions of PSMs.

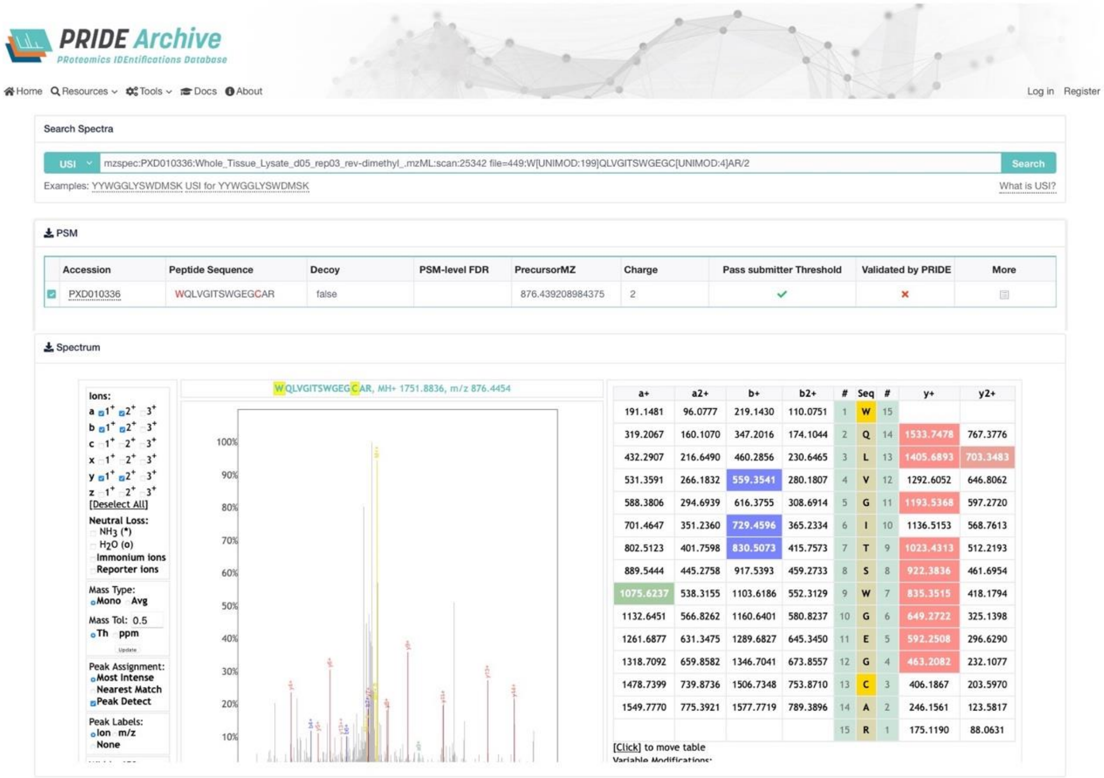


Peptide: WQLVGITSWGEGCAR

Examples: YYWGGLYSWDMSK USI for YYWGGLYSWDMSK

	Accession	Peptide Sequence	Decoy	PSM-level FDR	PrecursorMZ	Charge	Pass submitter Threshold	Validated by PRIDE	More
☐	PXD012039	WQLVGITSWGEGCAR	false	1.693e-4	860.41766	2	✓	✗	▤
☐	PXD012039	WQLVGITSWGEGCAR	false	2.125e-4	573.94788	3	✓	✗	▤
☐	PXD012039	WQLVGITSWGEGCAR	false	2.349e-4	860.41632	2	✓	✗	▤
☐	PXD012039	WQLVGITSWGEGCAR	false	1.511e-4	860.4165	2	✓	✗	▤
☐	PXD012039	WQLVGITSWGEGCAR	false	1.626e-4	860.41614	2	✓	✗	▤
☐	PXD012039	WQLVGITSWGEGCAR	false	1.756e-4	860.41797	2	✓	✗	▤

Figure S5. Screenshot showing the result from searching for the peptide sequence WQLVGITSWGEGCAR.



mzspec:PXD010336:Whole_Tissue_Lysate_d05_rep03_rev-dimethyl_.mzML:scan:25342 file=449:W[UNIMOD:199]QLVGITSWGEGC[UNIMOD:4]AR/2

Examples: YYWGGLYSWDMSK USI for YYWGGLYSWDMSK

	Accession	Peptide Sequence	Decoy	PSM-level FDR	PrecursorMZ	Charge	Pass submitter Threshold	Validated by PRIDE	More
☑	PXD010336	WQLVGITSWGEGCAR	false		876.439208984375	2	✓	✗	

WQLVGITSWGEGCAR, MH+ 1751.8836, m/z 876.4454

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^+

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

Mass Type:
● Mono ○ Avg

Mass Tol: 0.5
● Th ○ ppm

Peak Assignment:
● Most Intense
○ Nearest Match
☑ Peak Detect

Peak Labels:
● Ion ○ m/z
○ None

a+	a2+	b+	b2+	#	Seq	#	y+	y2+
191.1481	96.0777	219.1430	110.0751	1	W	15		
319.2067	160.1070	347.2016	174.1044	2	Q	14	1533.7478	767.3776
432.2907	216.6490	460.2856	230.6465	3	L	13	1405.6893	703.3483
531.3591	266.1832	559.3541	280.1807	4	V	12	1292.6052	646.8062
588.3806	294.6939	616.3755	308.6914	5	G	11	1193.5368	597.2720
701.4647	351.2360	729.4596	365.2334	6	I	10	1136.5153	568.7613
802.5123	401.7598	830.5073	415.7573	7	T	9	1023.4313	512.2193
889.5444	445.2758	917.5393	459.2733	8	S	8	922.3836	461.6954
1075.6237	538.3155	1103.6186	552.3129	9	W	7	835.3515	418.1794
1132.6451	566.8262	1160.6401	580.8237	10	G	6	649.2722	325.1398
1261.6877	631.3475	1289.6827	645.3450	11	E	5	592.2508	296.6290
1318.7092	659.8582	1346.7041	673.8557	12	G	4	463.2082	232.1077
1478.7399	739.8736	1506.7348	753.8710	13	C	3	406.1867	203.5970
1549.7770	775.3921	1577.7719	789.3896	14	A	2	246.1561	123.5817
				15	R	1	175.1190	88.0631

[Click] to move table

Figure S6. Screenshot showing the mass spectrum corresponding to the selected PSM (peptide sequence WQLVGITSWGEGCAR in dataset PXD010336).