

### Overview

Purpose: To develop a software tool able to analyze and quantitate single amino variations on a large-scale.

Methods: A unique peptide mapping tool was proposed which is able to map all LC-MS signals eluted from the mass spectrometer for each identified protein. With PEAKS' de novo capabilities, the signals display not only identified peptide LC-MS data, but as well as de novo only LC-MS data.

Results: A new method for large-scale quantitative analysis of single amino acid variations

#### Introduction

Global quantification of the single amino acid variations (SAVs) is essential to investigate their roles in disease progression. Most proteomic software tools are only able to interpret peptide fragments through database searches and map database hits to the protein sequences. As a result, mutated peptides are neither identified nor mapped to the protein sequences, thus failing to provide direct quantitative analysis. This study demonstrates the benefits of using PEAKS, a de novo assisted search engine, to maximize the performance of peptide mapping. Subsequent quantification of these peptide variants also provides large-scale quantitative analysis of SAVs.

# Methods

A thorough analysis of the LC-MS/MS data using a complete software package PEAKS, was proposed to ensure a feasible method for the analysis of single amino acid variants on a large scale data set. An initial database search was performed to identify all proteins in the sample. These proteins then became the candidates for searching SAVs. For the remaining unidentified spectra, de novo sequencing of these scans were conducted. Resulting high confident de novo sequence tags were then mapped to the corresponding protein candidates (figure 1). To further analyze these tags, homology search of the spectra were performed within the peptide maps to find single amino acid variants (figure 2 and 3). Finally, relative quantification of mutant and normal forms of the peptide could ultimately be determined by feature intensities of their extracted ion chromatograms (figures 3, 4 and 5).



Figure 1. Peptide mapping of PEAKS DB and de novo only results









Accession	-10kp	Coverage			#Peptides	*	#Utique PTM		Aug. P	fass	Description		
🛛 💭 🕫 (Q15758) AAAT JALPIAN				13%									
- spigetd19/WEX9_HLMAW	121.56	II I I	11.1	175	8		7	2	1071	68	Serine, threanine-pro	tein kinase No	49 OS+Hana si
-O sp(P82650)R122_HUNW	121.55		<b>III</b>	30%	9		9		412	90	285 ribosomal protei	n S22, mitoche	ndial OS+Han
	121.52			45%	5		5		226	56 )	Partner of 114 and r	nago OS+Hom	o sapiens GN=4
Import P18858-1	121.50			1 8%	6		6	2	1017	36	CNA ligese 1 isoforr	n Iso 1	
🕀 🌑 spiq960h6 MS22H_HUMAN	121.48		100	23%	6		6	00	351	97 )	WA-binding protein	Museshi hono	log 2 OS+Homo
-O sp[000582[P0004_HUMAN	121.47			22%	9		9	20	586	35 )	Podocalyvin OS=Hor	io sapiens GN	«PODIL PE+1 S
—O spip55957[830_HUMAN	121.46			35%	5		5	8	219	95 )	9K3-interacting dom	sin death agu	nist OS+Homa s
—O splQ96AQ6/P8IP1_HLMAN	121.44	I II		14%	8		8	2	805	43 )	Pre-B-cell leukernia b	anscription fa	ctor-interacting
—O spiQ9UKHB/RALY_HUMAN	121.38	11 II I			12	12 9		2	32463		RNA-binding protein Raly OS+Homo sapiens GN=R		
Converge Repúblic Demons Unity Tage													
Peptide 🕇	Utique	-18pP	Ness	Length	ppm	n(z	87	Area	Scan	#Spec	Start	Erd	PTH
1 K.C(+57.02)/EEM/G/A/CH		12.93	1118.5026	10		560.2592	9.72		F15:1914		363	372	2
2 K.EVLDSPLDLAR.N	•	44,69	1276.6663	11	1.5	639.3414	135.25	1.466	F10:31378	9	179	189	
3 P.GELLER		19.12	699.4279	6	2.8	350.7222	35.70	6.0526	F13:7329	1	93	98	
<ul> <li>R.GPAGDATVASEILE</li> </ul>	•	47.54	1101.5302	12	1.8	551.7733	11.79	2.366	F13:2321	30	526	537	
5 2 R.NEPSNLVSAAFR.S	•	43.91	1434.7620	13	17	718.3895	138.51	3.5266	F7:38326	8	190	202	
6 V.SELPLOPLPL(sub V)PREEGNPLIK.H	•	34.45	2171.1724	20	1.5	1006.5951	150.11	6.385	F1:32599	7	503	522	2
7 V.SELPLDPUPVPTEEGNPLUCH	•	58.57	2157.1567	20	0.5	1079.5862	141.30	2.2466	F8:31518	3	503	522	
8 R.SISTIVEER.M	•	22.44	1134.4829	9	-2.4	568.2474	20.57		F3:4530	2	203	211	

Figure 4. PEAKS displays corresponding peptide intensities for quantitative analysis

#### Katherine Tran, Cheryl Lichti, Baozhen Shan, Carol Nilsson Bioinformatics Solutions Inc, Waterloo, ON University of Texas Medical Branch, Galveston, TX





# Experimental

Glioma stem cell (GSC) lines, which were isolated and prepared by The University of Texas M.D. Anderson Cancer Center, were reduced, alkylated and analyzed by LC-MS/MS on a Thermo Orbitrap Elite mass spectrometer (Lichti, et. al., 2015). The LC-MS/MS data was analyzed using PEAKS with 10 ppm parent mass error tolerance and 0.025 Da fragment mass tolerance. A maximum of 2 missed cleavages were allowed, in addition to one non-specific cleavage. Carbamidomethyl cysteine was set as a fixed modification, whereas variable modifications included methionine oxidation as well as phosphorylation on STY. Database search was performed using the custom database with FDR estimation enabled. Highly confident de novo tags were then further analyzed using PEAKS SPIDER. Peptides with a -10logP score >30, which contained SAVs or post-translational modifications (PTMs), were selected for further validation.

As many as 312 unique variation sites were quantified. 19 SAV-containing peptides have been verified. Those peptides represent 19 SAVs in 17 chromosome 19 proteins. The peptide maps generated by PEAKS indicate that this approach could not only serve to maximize peptide maps of any protein, but to also provide quantifiable information regarding these variants presented as peptide profiles (figure 3).

## Conclusions

Large-scale quantitative analysis of single amino acid variants is made feasible using PEAKS' automated peptide mapping tool.

### References

1. Lichti, C.F., Mostovenko, E., Wadsworth, P.A., Lynch, G.C., Pettitt, B.M., Sulman, E.P., Wang, Q., Lang, F.F., Rezeli, M., Marko-Varga, G., Vegvari, A., Nilsson, C.L.: Systematic identification of single amino acid variants in glioma stem-cell derived chromosome 19 proteins. J. Proteome Res. 14, 778-786 (2015).

# Annotated Chromatogram of Protein sp|015382|BCAT2\_HUMAN