Using Pathway Analysis to Develop and Verify Targeted Proteomics Studies

Christine Miller and Steven Fischer Agilent Technologies, Santa Clara, CA HUP0 2012

233

Agilent Technologies

Introduction

The discovery of new biomarkers may be driven by proteomics, metabolomics, or genomics experiments and can result in a list of proteins of interest for subsequent targeted analysis. Preclinical validation of these putative biomarkers requires assays that are both specific and quantitative for target proteins and such assays are typically multiplexed, multiple reaction monitoring (MRM) analyses to achieve the required high-throughput. New pathway analysis and data visualization software facilitates the development of such targeted proteomics experiments by allowing simultaneous, integrated, multi-omics analysis and pathway visualization. A list of proteins of interest can then be exported and analyzed. The quantitative results from this targeted analysis can then be imported into the same software for statistical analysis and pathway visualization thus facilitating both discovery and verification. This workflow will be illustrated using a model sample set.

Experimental

Samples: In order to model this workflow, treated and unstimulated HeLa cell lysates were purchased (Millipore). The unstimulated (US) cell lysate was used as the control. The treated lysates had been exposed to 1) heat shock and arsenite (HS-Ars), 2) interferon α (IFN), 3) tumor necrosis factor (TNF) and 4) epidermal growth factor (EGF).

Sample preparation: For each sample, 100 or 200 μ g of total protein was reduced, alkylated (carbamidomethyl) and digested with trypsin using a modified FASP protocol using 2,2,2-trifluoroethanol for solubilization and denaturation. After overnight digestion, the trypsinized lysate was acidified with formic acid and stored at -80°C until use.

LC/MS/MS Analysis: For each injection, 500 ng tryptic digest was loaded. Each cell lysate was analyzed multiple times (minimum of n=4) on the HPLC-Chip/MS system interfaced to a 6490 QQQ with iFunnel technology using a 90 minute gradient method. The HPLC-Chip used was the new Polaris chip (G4240-6230) which is comprised of a 360 nL enrichment column and a 150 mm x 75 μ m analytical column which were both packed with Polaris C18-A, 3 μ m material (Agilent). Data was in acquired with retention-time scheduling using dynamic MRM mode.

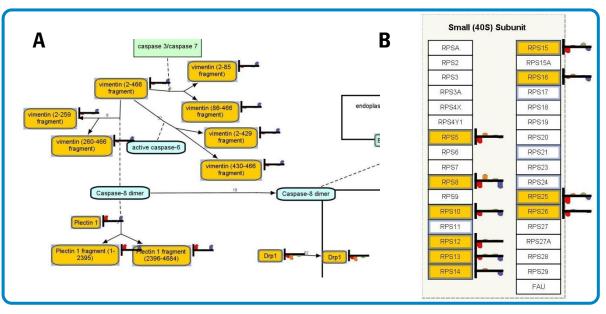
Data processing: The QQQ data was first processed by MassHunter Quantitative Analysis software to automatically extract the MRM chromatograms and calculate areas. The Quantitative software also have very convenient data review tools for assessing a large data set.

Area results for all targeted peptides were imported for into Mass Profiling Pro 12 (MPP), a data analysis and visualization package that includes an optional Pathways Analysis package.

Results and Discussion

Pathway Directed Target Analysis

Based on protein identification results from Spectrum Mill, differentially expressed proteins were mapped onto pathways in MPP (see companion poster 232). Two of the pathways (shown below) were selected for further targeted analysis by triple quadrupole mass spectrometry.



Q-TOF based label-free differential analysis displayed on portions of the A) Apoptotic execution phase WP1784_44958 pathway and B) Cytoplasmic Ribosomal Proteins WP477_41118 pathway. The heat strips to the right of each detected protein indicate the relative abundance in each sample treatment group.

Development of MRM-based Methods for Detected Targets

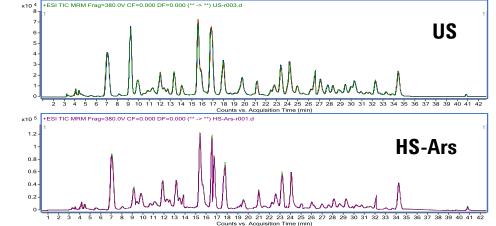
The SwissProt accession numbers for the detected differentially expressed proteins (yellow nodes in figures A and B above) were copied and then pasted into MRM Selector in Spectrum Mill software (below). MRM Selector automatically creates MRM transition lists based discovery data with a number of user-specified criteria. For this study, up to 5 transitions per peptide and 5 peptides per protein were exported from the data-dependent Q-TOF analyses. Because the same LC analysis was applied to the targeted study, retention time information could be exported for direct use in dynamic MRM methods.

	roll up proteins into groups ps ready for displaySorted by: P			to groups 0.24808 sec		Death	m/z	MH*	WH.	M						
*	Filename	z	Spectrum Intensity	Sequence	RT (min)	Peak Width (sec)					or					
EGF-r00	<u>34 10612 10612 2</u>	2	1.00e+005	(R) ILLIK (I)	30.42	47.64	294.2106	587.413	0.00	12	2.0					
P08670	_TLLIK_y3(+1)	294.2	373.3	0.79;;10.1												
P08670	_TLLIK_y4(+2)	294.2	243.7	0.79;;10.1												
P08670	TLLIK_y4(+1)	294.2	486.4	0.79;;10.1												
P08670	TLLIK y2(+1)	294.2	260.2	0.79;;10.1												
2 US-r001	16743 16743 3	3	6.35e+005	(R) FANYIDKVR(F)	40.47	49.12	375.8745	1125.605	0.00	38	3.4					
P08670_	FANYIDKVR_y8(+2)	375.9	489.8	0.82;;8.7												
P08670	FANYIDKVR y7(+2)	375.9	454.3	0.82;;8.7												
P08670	FANYIDKVR_y6(+1)	375.9	793.5	0.82::8.7												
P08670	FANYIDKVR y5(+1)	375.9	630.4	0.82;;8.7												
P08670	FANYIDKVR y6(+2)	375.9	397.2	0.82::8.7												
3 EGF-r00	04.23484.23484.4	4	2.16e+005	(R) LGDLYEEEMRELRR (Q)	65.73	47.52	452.9811	1808.896	0.00	66	3.6					
P08670	LGDLYEEEMRELRR y13(+3)	453.0	565.9	0.79;;11.5												
	1001VEEENDELOD 441.0		500 C	A 74 44 5												
Select Re	esults for MRM Selection			on and Sorting					RM Paran							
Select N	ARMs Save As Loa	d	Filter res	sults by:				13	Text file	export						
a constant	MassHunter QQQ DMRM +		valid		•			T	p n Tran	sitions:	5 2	80	% of	precur	sor 🔽 y-io	ns only
Format:			Protein	grouping method: 1 shared	l, expand	subgro	ups 🔻	1	rigger th	reshold	- 500	for ()	second	lary transi	tions
Data dire	ctories: Select		Sort pro	teins by: Score	-										e only •	
Chris/1	2Aug27-HeLa/EGF		Filter by	protein score: > • 0					an classes)			301/110	gine	in criary	le only -	
	2Aug27-HeLa/HS-Ars	10	i mer bj			1		D	well time	(ms):	10					
	2Aug27-HeLa/IFN	13	1 10 10 10 10 10	eptides for MRM: Limit to	• 5			1	Use pea	k width	of: 1	mir	1			
Chris/1	2Aug27-HeLa/TNF	-	Rank pe	eptides by: Score -					ollision E							
Search re	esult files:	Sort MR	M List by: Precursor m/z		E= slope											
*.spo	*		Filter pe	eptides by:					Use ac							
			Score	> • 0 % SPI: >	- 0			2	-state:	slope	interc	ept:				
					sallowed	AAe-	none		2	3.6	-4.8					
	*				All	nnia.			3	3.6	-4.8					
			Peptide		AII											
									4	3.6	-4.8					

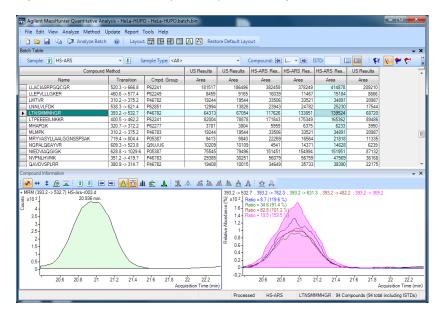
Results and Discussion

QQQ Reproducibility

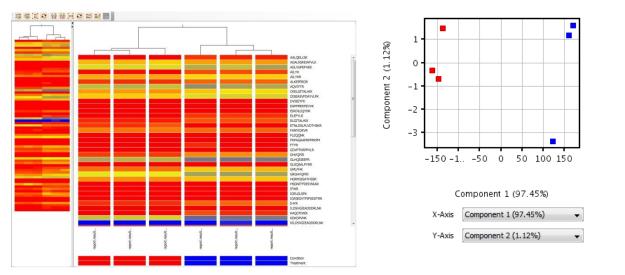
Dynamic MRM analysis was performed on the unstimulated and HS-Ars samples. The figure below shows overlaid TIC MRM chromatograms for the replicate analyses demonstrating excellent reproducibility.



MassHunter quantitative results (below) were exported to MPP.



These results were then subjected to statistical analysis (t-test) to determine the differential features. Again, PCA and hierarchical clustering can be used to visualize the quality of the data.



, , , , ,

Workflow



MassProfiler Pro

Pathway analysis of differentially expressed proteins
Export of protein accession numbers from pathways of interest



Creation of MRMs

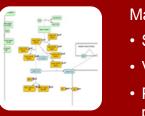
 MRM Selector in Spectrum Mill for export of DMRM method based on discovery data

 Peptide Selector or Skyline for *in silico* prediction of transitions



MRM analysis on target peptides
6490 iFunnel QQQ
Retention-time scheduled dynamic MRM mode based on discovery results

MassHunter Quantitative Analysis
Data browser to review quantitative results
Export of peak areas for all target compounds to MPP



MassProfiler ProStatistical analysis of quantitative resultsVisualization of differential peptides

 Pathway analysis at the protein level based on peptide targets

Development of MRM-based Methods for Non-Detected Targets

It is also possible to develop MRM assays for any or all proteins in an pathway regardless of detection in the protein discovery phase. In this case, the GAS2 protein from the apoptotic pathway was selected to illustrate that *in silico* MRM transitions can be predicted using either Skyline (A) or Peptide Selector in Spectrum Mill (B). This same approach is used when the pathway-directed experiment is based on genomics or metabolomic information

Α

📩 Skyline

<u>File Edit View Settings H</u>elp

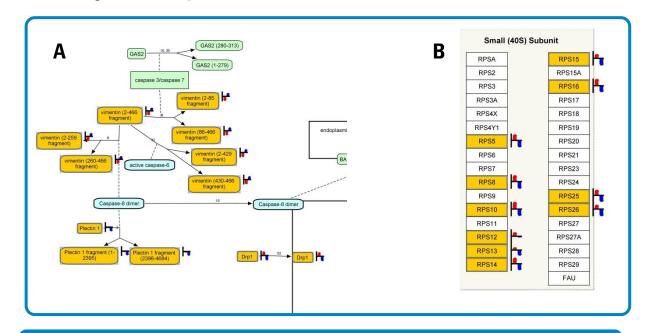
MRM transitions can be exported and applied to a pooled sample to determine which peptides can be detected.

IIE	; V	VI	псі	ιķ	JE	۶Þ١	liu	ie	3	L.	all be				Ľ	1 🦾					י רי	Gi .	Ŧ					
te N-ter s: CM thin 3 re ide inus en (H)	ent in Swissi rminal Gin t sidues Peptide C termin Free Acid (43903 txt)	to pyro(ilu Masses ar	e Uniques - Uniques - Count Sequen	1055	20238											ророророророророророророророророророро	3903 HEAI EIL/ EIL/ 	NLLPI ALWL ETFM GALL SGSH 1.2776 S [y7 G [y6 S [y7 G [y6 S [y7 G [y6 S [y7 C [y6 S [y7 C [y6 C [MK.E TNILLC EK.L QQLAI PTK.I FFAR. - 868 - 771 - 868 - 771 - 868 - 771 - 684 - 627 SW <u>C</u> F ELGR. LIK.L SAPS DAVK (.C [20 (FFAG ¹)	[28, 36] GK.E [50, 55] STMQ V (79, D [99, 44312 .3784 4.3464 7.3249 3.D [1 SEGL] SEGL] 1 [143 [158, PSPS R [19] 98, 21! YLLK.I	37, 43 9] EK.F 88] 109] + + + + 10, 12 VLHK , 153] 167] PSSK 6, 206 5] H [25]	[60, 7 20] .Q [12 1 5.S [17 6] 1, 264	21, 13 71, 19				
14 E24 I	9 G19 H6 H	1 K30 L3	18 M10 N12	P21 Q9 R15	5 S28 T	14 1/14 1	N4 Y6																					
Acc #	% Coverage	RP- RT	MH+	m/2	m/3	# DB peps	Start AA	End AA	Score	Prev.	Sequence	Next	bı	Ys - Yt C-side	of Asp, Gl of Pro	u												
43903		18.42	1127.5480	564.2776	5	1	100	110	1.00	(IPCK)	TSAPSOSFFAR	(DNT)	189.09	1026.50	939.47	868.43	771.38	684.35	627.32	540.29	393.22	246.16	175.12					
43903		23.50	1485.8312	743.4192	2	1	38	50	1.00	(LPMK)	EDLALWLTNLLGK	(EIT)	245.08	1356.79	1241.76	1128.68	1057.64	944.56	758.48	645.39	544.35	430.30	317.22	204.13	147.11			
43903				749.3905		-	252				VGGGWETFAGYLLK					1284.66							536.34	373.28	260.20	147.11		
43903				536.8055							YGVEPPGLIK					753.45							_		111			
043903	21.0		2099.9979	1050.5026	5	1	172	191	0.91	(KLEK)	EEGEETLSAPSPSPSPSSK	(SSG)	243.13	1970,96	1857.87	1728.83	1600.77	1471.73	1342,68	1241.64	1128.55	1041.52	970,48	873.43	786.40	689.35	02.31 50	5.26
10 YS	WLASRHE	ANLLP	KEDL AL	NUTRILLOK	EITA	ETFNER	LENG	DALLO	L AE	INGERER	E 80																	
T SAS	SCSFFAR	DNTAN	FLSWC BD	LOVDETCL	FESE	GLVLHY	-	EVCLOL	L ELO	PEIAABY	G 160																	
S APS	PSPSPSPSS	KSSGK	STON LL	DDAVKRIS	EDPP	CKCENK	FLV	ERLSOG	R YR	GENILF	I 240																	
	A's) of the pr		ANDOK TS	PIQSKSPT	LÄDIG	NPEWYL	. vvs	VRAKVŘ	K zij	ś	313																	

e Selector - Aglient Spectrum Mil Rev. 4.0.127

В

Because the accession number was included in the MRM table as the "Compound Group", pathway analysis could be performed on the exported results as well (below). Note that the QQQ results quite clearly show expression level changes in these proteins with HS-Ars treatment.



Conclusions

- Pathway analysis based on differential 'omics experiments can facilitate the development of targeted analyses for proteins of interested
- Targeted MRM-based assays can be easily developed based on proteomics discovery results or in silica prediction in combination with "scouting" analyses to confirm the best targets.
- Quantitative results from target analyses can be subjected to the same statistical evaluation, visualization and pathway analysis using MPP software
- This approach in combination with 'omics-based discovery facilitates biologically-driven workflows