Waters™

應用手冊

Biological Interpretation of Breast Cancer Using Rapid Multi-Omic Profiling Methods

Adam King, Christopher J. Hughes, Giorgis Isaac, Lee A. Gethings, Robert S. Plumb

Waters Corporation

For research use only. Not for use in diagnostic procedures.

Abstract

In this application note, we demonstrate the utility of a high throughput, multi-omic workflow (proteomics and lipidomics) for a plasma-based sample set, consisting of healthy controls and breast cancer individuals. Resulting lipid and protein DIA datasets have been integrated and interrogated to provide insights into the biological pathway and associated networks connected to the pathogenesis of the disease.

Benefits

- Flexible, multi-omic data independent acquisition (DIA) workflows using the SYNAPT XS provide highly confident protein and lipid identifications.
- High throughput strategy that shows utility for large scale population studies.
- Combined DIA, high throughput approach has shown utility for the analysis of plasma from a breast cancer cohort.
- Integration of Progenesis processed data with Metacore, a Cortellis solution, allows for comprehensive pathway analysis to be conducted for multi-omic datasets.

Introduction

Breast cancer is one of the most common cancers diagnosed in women and the second leading cause of death after lung cancer. Breast cancer types consist of two main categories, termed in situ and invasive, with invasive being the most common (81% of cases).[1] Omic-based strategies have previously been shown to provide insight into the biological mechanisms of various cancer types, including breast-focused studies.[2,3]

Rapid profiling methods for discovery-based analyses have previously been demonstrated for omic-based studies,[4,5] whereby microbore chromatography is configured with data independent acquisition (DIA) schemas to allow for increased sample throughput, while retaining highly confident identifications.

Here, we demonstrate the utility of a high throughput, multi-omic workflow (proteomics and lipidomics) for a plasma-based sample set, consisting of healthy controls and breast cancer individuals. Resulting lipid and protein DIA datasets have been integrated and interrogated to provide insights into the biological pathway and associated networks connected to the pathogenesis of the disease.

Experimental

Samples

Plasma samples originating from 26 individuals were used for the study. These comprised of healthy controls (n=6) and breast cancer (n=20) diagnosed individuals(Figure 1).

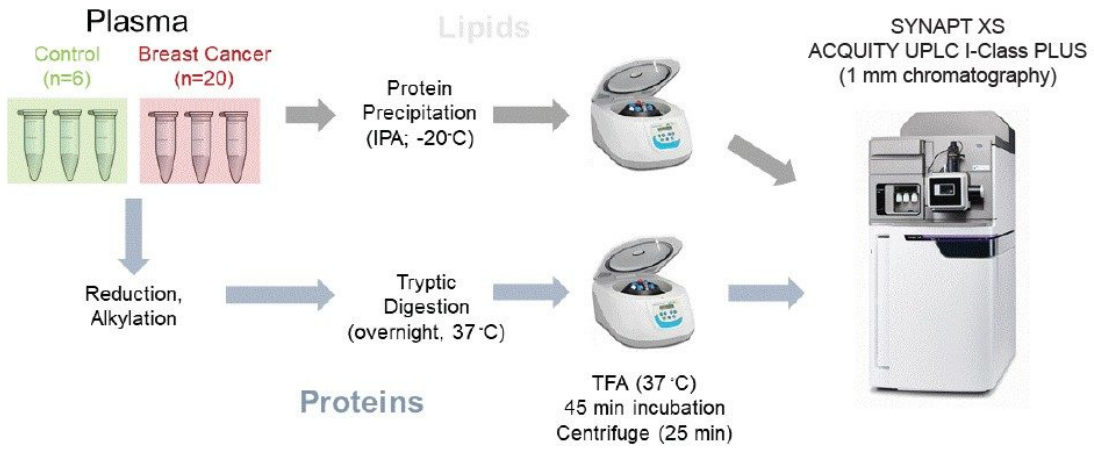


Figure 1. Experimental workflow detailing the preparation of plasma samples relating to control and breast cancer subjects.

Sample preparation

Lipidomics

A simple sample preparation procedure was adopted using protein precipitation with pre-cooled isopropanol (1:5 plasma:IPA). Samples were mixed for one minute and placed at -20 °C overnight. The extracted samples were centrifuged at a maximum of 10,300 g for 10 minutes at 4 °C before transferring the supernatant to Waters Total Recovery UPLC Vials (p/n: 186005669CV) for LC-MS analysis. Prepared samples were analyzed in triplicate for both ionization modes.

Proteomics

Undepleted plasma were prepared with 1% RapiGest SF Surfactant prior to reduction, alkylation, and overnight digestion with trypsin.

LC conditions (lipidomics)

LC system:	ACQUITY UPLC I-Class PLUS
Column(s):	ACQUITY UPLC BEH C_8 1.0 ×

	50 mm, 1.7 µm
Column temp.:	55 °C
Flow rate:	0.25 mL/min
Mobile phase:	Water:isopropanol:acetonitrile (50:25:25)/5 mM ammonium acetate/0.05% acetic acid (A) and isopropanol:cetonitrile (50:50)/5 mM ammonium acetate/0.05% acetic acid (B)
Gradient:	1% to 90% B
Run time:	3 min
Injection volume:	0.2 µL

LC conditions (proteomics)

LC system:	ACQUITY UPLC I-Class PLUS
Column(s):	ACQUITY UPLC CSH 1.0 × 100 mm, 1.7 µm
Column temp.:	45 °C
Flow rate:	50 µL/min
Mobile phase:	Water + 0.1% formic acid (A) and acetonitrile + 0.1% formic

	acid (B)
Gradient:	1% to 40% B
Run time:	15 min
Injection volume:	2 µL

MS conditions (lipidomics)

MS system:	SYNAPT XS
Ionization mode:	ESI (+/-)
Capillary voltage:	2.8 kV (+) 1.9 kV (-)
Acquisition mode:	$HDMS^E$
Acquisition rate:	Low and elevated energy functions at 0.1 seconds
Collision energy:	5 eV (low energy function); 20–45 eV linear collision energy ramp (elevated energy function)
Source temp.:	120 °C
Desolvation temp.:	500 °C
Cone gas flow:	50 L/hr
Desolvation flow:	800 L/hr

MS conditions (proteomics)

MS system:	SYNAPT XS
Ionization mode:	ESI (+)
Capillary voltage:	2.1 kV
Acquisition mode:	SONAR
Acquisition rate:	Low and elevated energy functions at 0.5 seconds
Collision energy:	5 eV (low energy function); 20–45 eV linear collision energy ramp (elevated energy function)
Source temp.:	120 °C
Desolvation temp.:	500 °C
Cone gas flow:	50 L/hr
Desolvation flow:	600 L/hr

Informatics

Progenesis QI and Progenesis QI for Proteomics were used for processing the lipidomic and proteomic datasets respectively. In both cases, data were aligned and normalized to provide label-free quantification. Lipid identifications were provided by searching against the LIPID MAPS database. Resulting identifications were further curated using the Lipid Reporter Tool.[6] Protein/peptide identifications were returned when searched against the UniProt human database, containing only curated entries. Carbamidomethyl (C) and oxidation (M) were considered as fixed and variable modifications respectively, with a 1% FDR applied. Statistical analysis in all cases was conducted using a combination of EZinfo (Umetrics, Umeå, Sweden) and MetaboAnalyst (University

of Alberta).[7] The curated lipid and protein results were directly outputted from Progenesis for pathway and network mapping using Metacore, a Cortellis solution (Clarivate Analytics, London, UK) (Figure 2).

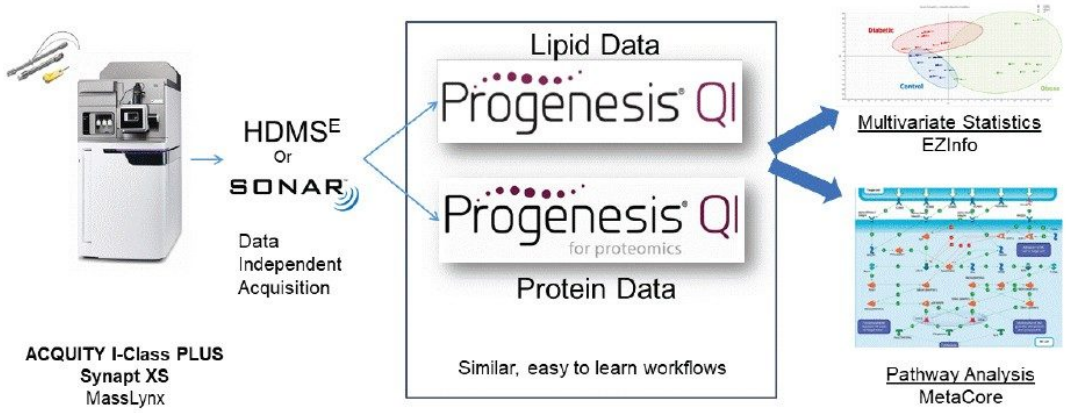


Figure 2. Lipidomic and proteomic LC-MS analyses and associated data processing workflow.

Results and Discussion

Extracts resulting from control and breast cancer samples were analyzed using a high throughput approach combined with DIA strategies. Ion mobility (IM)-based DIA was utilized for lipidomic analyses, while SONAR was employed for proteomics. Rapid profiling of the lipidome combined with multivariate statistical approaches highlighted key differences between the control and breast cancer groups (Figure 3). Orthogonal partial least squares (OPLS-DA) indicated phospholipids and neutral lipid species (i.e., triglycerides) as being the dominant drivers for separation between the two cohorts (Table 1).

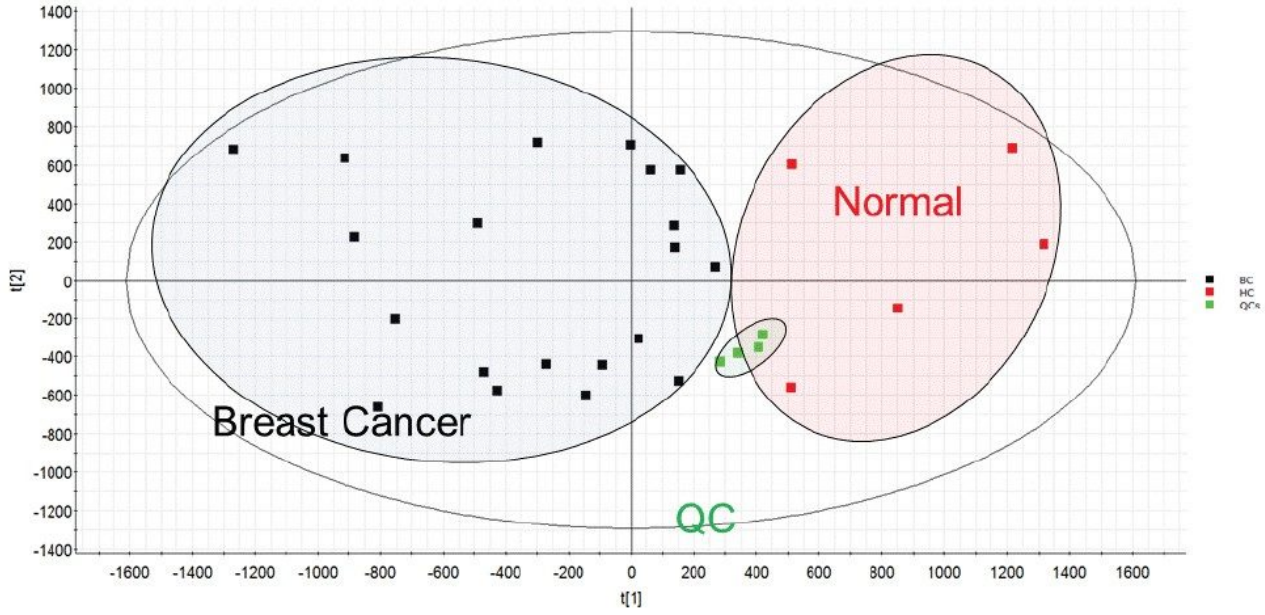


Figure 3. Multivariate statistical analysis representing the lipidomic data. Principal component analysis (PCA) clearly shows differentiation between the control (red) and breast cancer (black) subjects. The pooled QCs (green) are shown to cluster tightly, indicating high technical reproducibility.

Lipid identification	Neutral mass (Da)	*m/z*	Retention time (min)	CCS (Å²)	ΔCCS (Å²)	Chromatographic peak width (min)	Anova (p)	q Value	Max fold change	Minimum CV%
TG(52:3)	856.75	874.79	3.14	334.1	–	0.20	5.5E-05	0.00166	1.5	6.03
TG(52:4)	–	872.77	3.06	331.4	–	0.23	1.0E-04	0.00201	1.8	4.85
TG(54:5)	880.75	898.78	3.07	337.6	–	0.18	1.5E-04	0.00237	2.6	5.53
PS (40:4)	839.57	822.56	1.62	304.4	–	0.30	2.1E-04	0.00274	2.0	4.38
TG(54:3)	–	902.82	3.22	341.1	–	0.24	3.4E-04	0.00386	2.0	5.54
DG(34:1)	–	577.52	3.2	267.1	–	0.18	3.7E-04	0.00404	1.8	2.97
TG(50:1)	832.75	850.79	3.19	332.5	–	0.16	3.9E-04	0.00408	1.7	5.15
PS(O-36:2)	773.56	774.56	1.55	295.8	–	0.31	5.0E-04	0.00452	1.9	3.16
PS(O-38:3)	799.57	782.57	1.62	299.7	–	0.28	6.5E-04	0.00529	1.8	3.09
PS(36:1)	789.55	790.56	1.54	298.7	–	0.41	2.3E-03	0.01119	2.9	2.21
PC(36:4)	781.56	782.57	2.01	306.6	–	0.28	6.6E-03	0.02300	1.5	5.15
PC (38:4)	809.59	810.60	2.19	312.2	7.2	0.37	8.3E-03	0.02592	1.6	4.06
PC (36:2)	785.60	786.60	2.19	304.6	4.6	0.20	1.9E-02	0.04109	1.3	4.69
PS (38:2)	815.57	816.57	1.63	306.3	–	0.41	2.2E-02	0.04428	1.5	4.56
PC(34:2)	757.57	758.57	2.02	296.9	–	0.24	2.7E-02	0.04926	1.3	4.84

Table 1. Curated lipid identifications resulting from the statistical analysis. Identifications highlighted in red signify down regulation with respect to controls, while up regulated lipids with respect to controls are shown in green.

Complementary proteomic analysis of the plasma extracts identified 170 curated protein identifications and demonstrated a dynamic range >3 orders. Furthermore, when comparing the proteome using multivariate statistical approaches, significant differences between the control and breast cancer cohorts were revealed. Unsupervised principal component analysis (PCA) clearly shows this separation between groups (Figure 4A), while the most significant proteins contributing towards the variance are presented as a heatmap (Figure 4B). ApoA2 is under-expressed in those subjects with breast cancer. ApoA2 is involved in the lipid transportation and, more specifically, phospholipid binding, corresponding with the increased levels of PS lipids observed from the lipidomic analyses. Previous studies have also shown correlation between the role of ApoA2 in other cancer types, such as pancreatic.[8] LRG1, a protein involved in signalling within the TGF pathway, is shown to be over-expressed for breast cancer and has previously been reported as being involved in ovarian cancer (Figure 4C).[9]

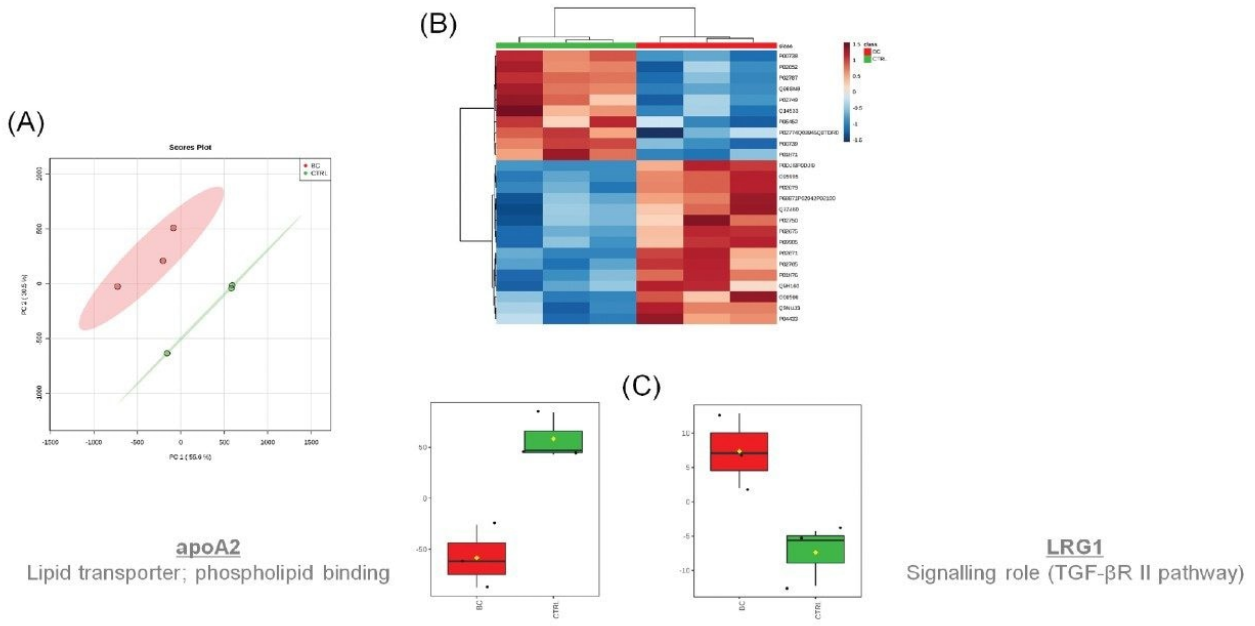
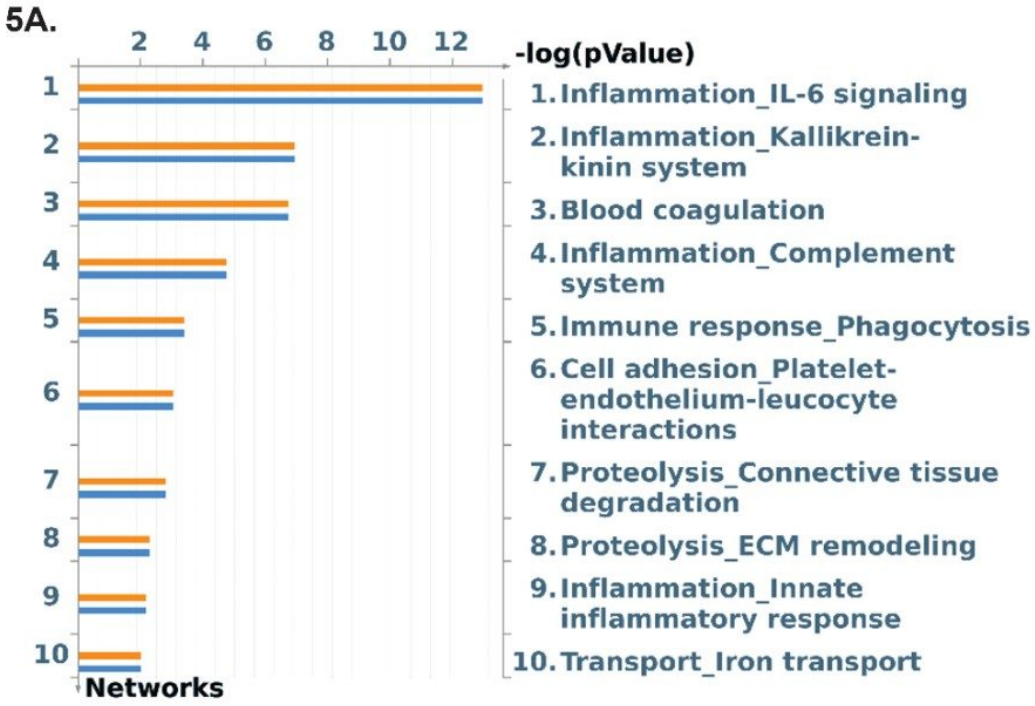


Figure 4. Multivariate statistical analysis of the high throughput proteomic data. Unsupervised principal component analysis (A) shows clear separation between healthy controls and breast cancer subjects. The corresponding heatmap (B), highlights clear regions of differential protein expression between the two cohorts. The under- and over-expression of ApoA2 and LRG1, respectively, are provided as examples (C). The box-whisker plots indicate significant differences and tight conservation of the data within each group.

In order to provide biological context, both the proteomic and lipidomic datasets were subjected to pathway analysis using Metacore, a Cortellis solution. Streamlined integration of Progenesis with Metacore enables multi-omic datasets to be readily interrogated for both pathway and network analyses. Several pathways resulting from enrichment analysis were highlighted as being statistically significant based on the contribution from both lipids and proteins (Figure 5) with inflammation IL-6 signalling being the highest ranking. The relationship between complexes identified as a result of enrichment analysis can be highlighted through network analysis. Focusing on the IL-6 signalling pathway, the signal transducer and activator of transcription 3 (STAT3) is responsible for mediating the transcriptional programs downstream of several cytokine, growth factor, and oncogenic stimuli. Its expression and activity are consistently linked to cellular transformation, as well as tumor initiation and progression. All links/ associations between the identified proteins are shown to interact through STAT3 (Figure 5).



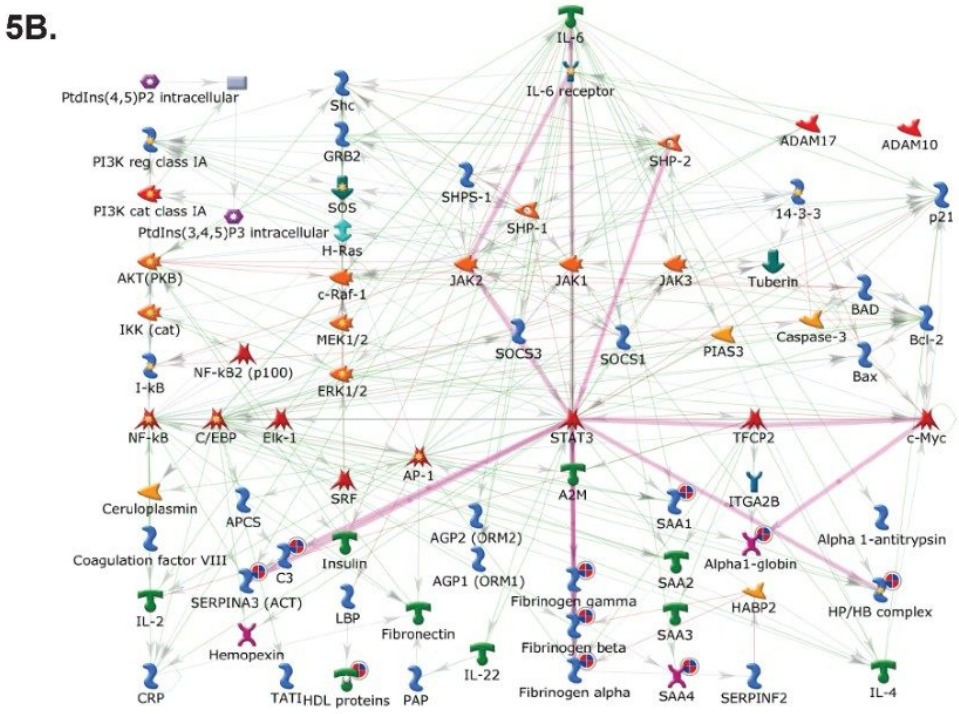


Figure 5. Metacore, a Cortellis solution, pathway analysis reveals signficant pathways of interests being

implicated in breast cancer mechanisms. The top 10 most statistically signficiant pathways (-logP) are listed (upper) following enrichment analysis with associated network analyses (lower) corresponding with the inflammation IL-6 signalling pathway. A number of key protein complexes are highlighted, all of which interact through STAT3. The pink lines indicate the association/connectivity between these complexes.

Conclusion

- A rapid workflow providing qualitative and quantitative multi-omic data for human plasma extracts is demonstrated using a breast cancer cohort.
- DIA-based acquisition schemas are shown to be complementary with high throughput strategies, while maintaining high coverage and highly confident identifications.
- Rapid lipidomic analysis utilized with HDMS^E, highlighted phospholipids (PCs and PSs) and neutral lipids (TGs and DGs) as differential markers between healthy controls and breast cancer subjects.
- Rapid proteomic analysis combined with SONAR provides a number of significantly relevant proteins, which included ApoA2 and LRG1.
- Subjecting the combined lipid and protein datasets through Metacore, a Cortellis solution, provided comprehensive pathway analysis. Several key pathways were identified, including IL-6 signalling.

References

1. American Cancer Society. Breast Cancer Facts and Figures 2019-2020. Atlanta: American Cancer Society, Inc. 2019.
2. Yan F. *et al.* Lipidomics: a promising cancer biomarker. *Clin Transl. Med*., 2018, 7:21.
3. Qin, X.J. *et al.* Proteomic studies in breast cancer (Review). *Oncol. Lett.,* 2012, 3(4): 735–743.

4. King, A. *et al.* Development of a rapid profiling method for the analysis of polar analytes in urine using HILIC-MS and ion mobility enabled HILICMS. *Metabolomics*, 2019, 15:17.

5. Hughes, C. and Gethings, L. Analysis of Plasma Proteins using 1 mm Scale Chromatography and the Xevo G2-XS QTof Operating in SONAR Acquisition Mode. Waters Tech Brief, 720006493EN, 2019.

6. Lipid Reporter. https://sourceforge.net/projects/ lipid-reporter/ (accessed March 2020).

7. Chong, J. *et al.* Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Current Protocols in Bioinformatics*, 2019, 68, e86.

8. Honda, K. *et al.* Potential usefulness of apolipoprotein A2 isoforms for screening and risk stratification of pancreatic cancer. *Biomark Med.,* 2016, 10(11): 1197–1207.

9. Anderson, J. *et al.* Leucine-rich alpha-2- glycoprotein-1 is upregulated in sera and tumors of ovarian cancer patients. *Journal of Ovarian Research*, 2010, 3:21.

Featured Products

ACQUITY UPLC I-Class PLUS System <https://www.waters.com/134613317>

SYNAPT XS High Resolution Mass Spectrometer <https://www.waters.com/135020928>

Progenesis QI <https://www.waters.com/134790652>

Progenesis QI for Proteomics <https://www.waters.com/134790665>

720006815, March 2020