BIOTECH SUPPORT GROUP

APPLICATION REPORT
APRIL 23, 2015

AlbuVoid™ LC-MS On-Bead
Differential Expression of Lung & Breast Cancer Sera Proteins Using Quantitative (iTRAQ) Proteomics

Introduction

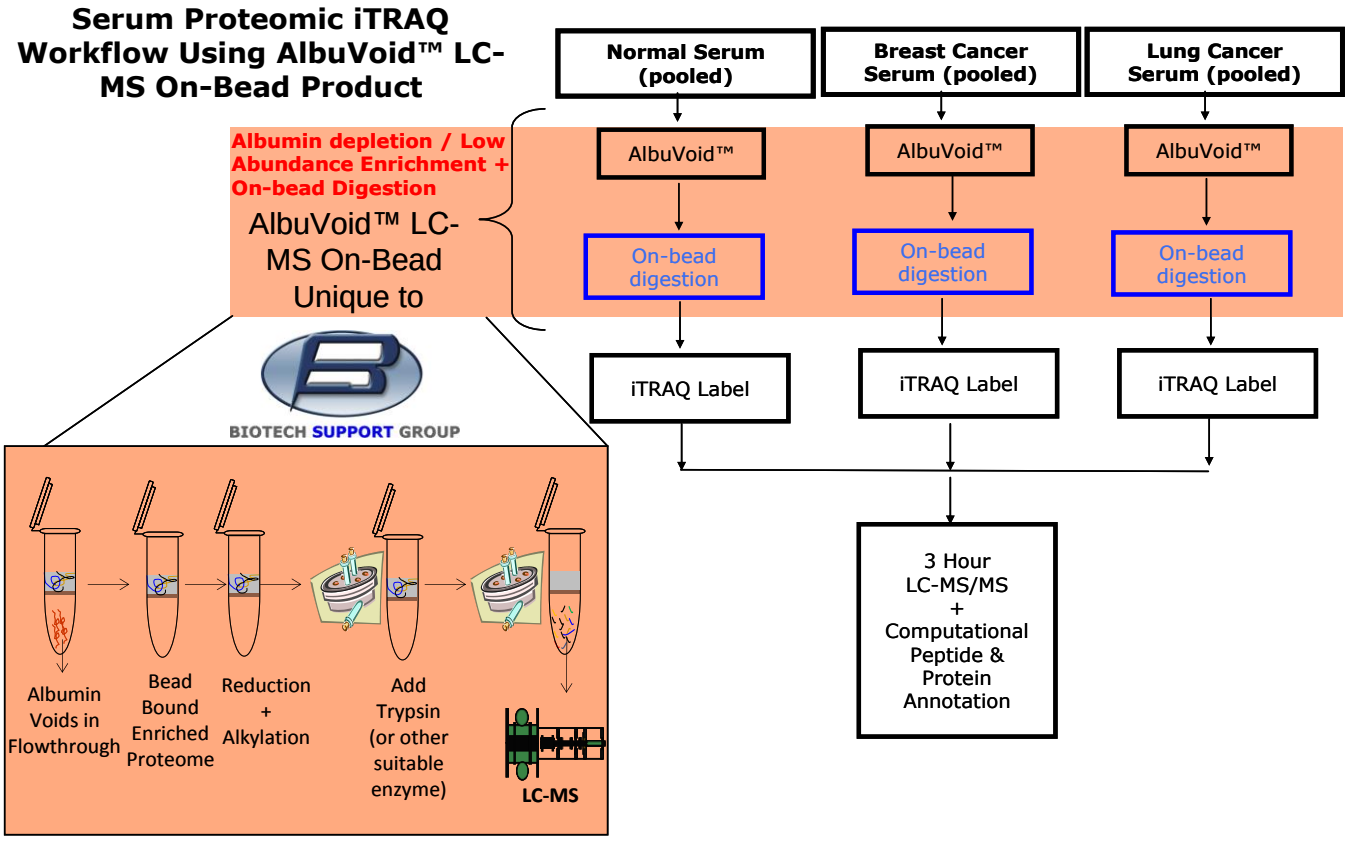
The discovery of cancer biomarkers that can personalize a treatment process has become an important research area in the proteomics field. For this, many proteomics approaches including two dimensional polyacrylamide gel electrophoresis (2DPAGE & 2D-DIGE), surface enhanced laser desorption/ionization time of flight (SELDI-ToF), protein arrays, and multidimensional protein identification technology (MudPIT) are being implemented in cancer research[1].

So for cancer research, serum and plasma are especially attractive sample types as collection of blood is common, simple and only minimally invasive. Yet these samples can offer unique challenges in proteomic analyses. These challenges are being met by developments surrounding the AlbuVoid™ depletion product, coupled to on-bead digestion. A full report for this method entitled "AlbuVoid™ & "On-Bead Digestion - Tackling the challenges of serum proteomics (LC-MS)", can be downloaded from our website[2].

For quantitative proteomics, having workflows that can greatly reduce one or more highly abundant proteins, along with short, efficient and consistent digestions will be highly desirable. Indeed several reports show that quantitative precision is strongly influenced by variations in enzymatic digestion efficiency[3-5]. For these reasons, we continue to evaluate and report these efficiencies in our methods development for on-bead digestion using the new **AlbuVoid™ LC-MS On-Bead** product for Albumin depletion and consequent low abundance enrichment.

Isobaric tags for relative and absolute quantitation (iTRAQ) is a labeling method for tandem mass spectrometry to analyze protein amounts from multiple pooled samples at once. After Trypsin digestion, the reporter groups split from the peptide and generate small fragments unique and correspondent to the original sample. The intensity of each of these reporting peaks represents the relative quantity of a peptide. Peaks in the MS2 spectrum are then used to identify peptide sequences and by inference, the protein sequences from which they were derived. By comparing the amounts of peptides labeled with each iTRAQ reporter, quantitative differences from all proteins identified can be readily measured and differentially assessed.

In this report we demonstrate the new **AlbuVoid™ LC-MS On-Bead** product for comparing labeled (iTRAQ) quantification of proteins from normal and two disease state sera – breast and lung cancer. We consider the workflow advantages and the variety of options available to investigators using this new product and method.



Material & Methods

AlbuVoid™ Protocol for Albumin Depletion and Low Abundance Serum Protein Enrichment

Processes 50-100 µl serum per prep. In bold are the **AlbuVoid™ kit** components.

1. Weigh out 25 mg of **AlbuVoid™** matrix in a spin-tube (0.45µ SpinX centrifuge tube filter from Corning).

2.Add 125 µl of **Binding Buffer AVBB.** Vortex for 5 minutes at room temperature followed by centrifugation at 3000 rpm. Discard the supernatant.

3.Repeat step-2

4.Condition by adding 100 µl of **AVBB** and 50-100 µl of the **Serum.** Vortex for 10 min. and then centrifuge for 4 min. at 10,000 rpm.

5. Remove the albumin enriched supernatant (Flow-Through) **FT**.

6. To the beads, add 250 µl of **Wash Buffer AVWB.** Vortex for 5 min and centrifuge for 4 minutes at 10,000 rpm. Discard the **Wash.**

7. Repeat Step-6.

The AlbuVoid™ bead is now enriched with albumin depleted low abundance proteins. For LC-MS sample preparation, an on-bead digestion protocol can be applied.

8. After the final wash steps from Step 7 from the enrichment, add 10 µL 100mM DTT + 90 µL 10mM HEPES, pH 7, vortex 10 min, incubate ½ hr at 60 °C.

9.After cooling, add 20µl 200mM Iodoacetamide, and 80 µL 10mM HEPES, pH 7, incubate in dark for 45 min at room temp.

10.Centrifuge at 10,000 rpm (microfuge max setting) for 5 minutes, and discard supernatant.

11.Add 40 µL Sequencing-grade trypsin (0.4µg/µl, in 50mM acetic acid) + 60 µL 10mM HEPES, pH 7 to the beads. Digest overnight at 37°C or other optimized time period.

12. Centrifuge at 10,000 rpm (microfuge max setting) for 5 minutes, and retain peptide filtrate.

13.To further extract remaining peptides, add 150 µL 10% formic acid, vortex 10 min, centrifuge at 10,000 rpm (microfuge max setting) for 5 mins., and add this volume to the first volume.

14.Total is about 250µl. Prepare to desired final concentration. Store at -80 °C until LC-MS/MS.

LC-MS/MS Analysis, Peptide & Protein Identifications

The following results were obtained by LC-MSMS at the Mass Spectrometry facility at Rutgers Proteomics Center. A nanoLC-MS/MS coupled a RSLC system (Dionex, Sunnyvale CA) interfaced with a LTQ Orbitrap Velos (ThermoFisher, San Jose, CA). Samples were loaded onto a self-packed 100µm x 2cm trap packed with Magic C18AQ, 5µm 200 A (Michrom Bioresources Inc, Aubum, CA) and washed with Buffer A(0.2% formic acid) for 5 min with flow rate of 10ul/min. The trap was brought in-line with the homemade analytical column (Magic C18AQ, 3µm 200 A, 75 µm x 50cm) and peptides fractionated at 300 nL/min with a multi-stepped gradient (4 to 15% Buffer B (0.16% formic acid 80% acetonitrile) in 35 min and 15-25%B in 65 min and 25-50%B in 50 min). Mass spectrometry data was acquired using a data-dependent acquisition procedure with a cyclic series of a full scan acquired in Orbitrap with resolution of 60,000 followed by MSMS scans (CID 35% of collision energy) of 20 most intense ions with a repeat count of two and the dynamic exclusion duration of 60 sec. The LC-MS/MS data was searched against the Human & Rat respective Ensembl databases using X!tandem with carbamidomethylation on cysteine as fixed modification and oxidation of methionine and deamidation on Asparagine as variable modifications using a 10 ppm precursor ion tolerance and a 0.4 Da fragment ion tolerance. We included only the peptides with loge<=-1.5 and charge state <=4. For spectra with multiple top hit peptides, only one was selected.

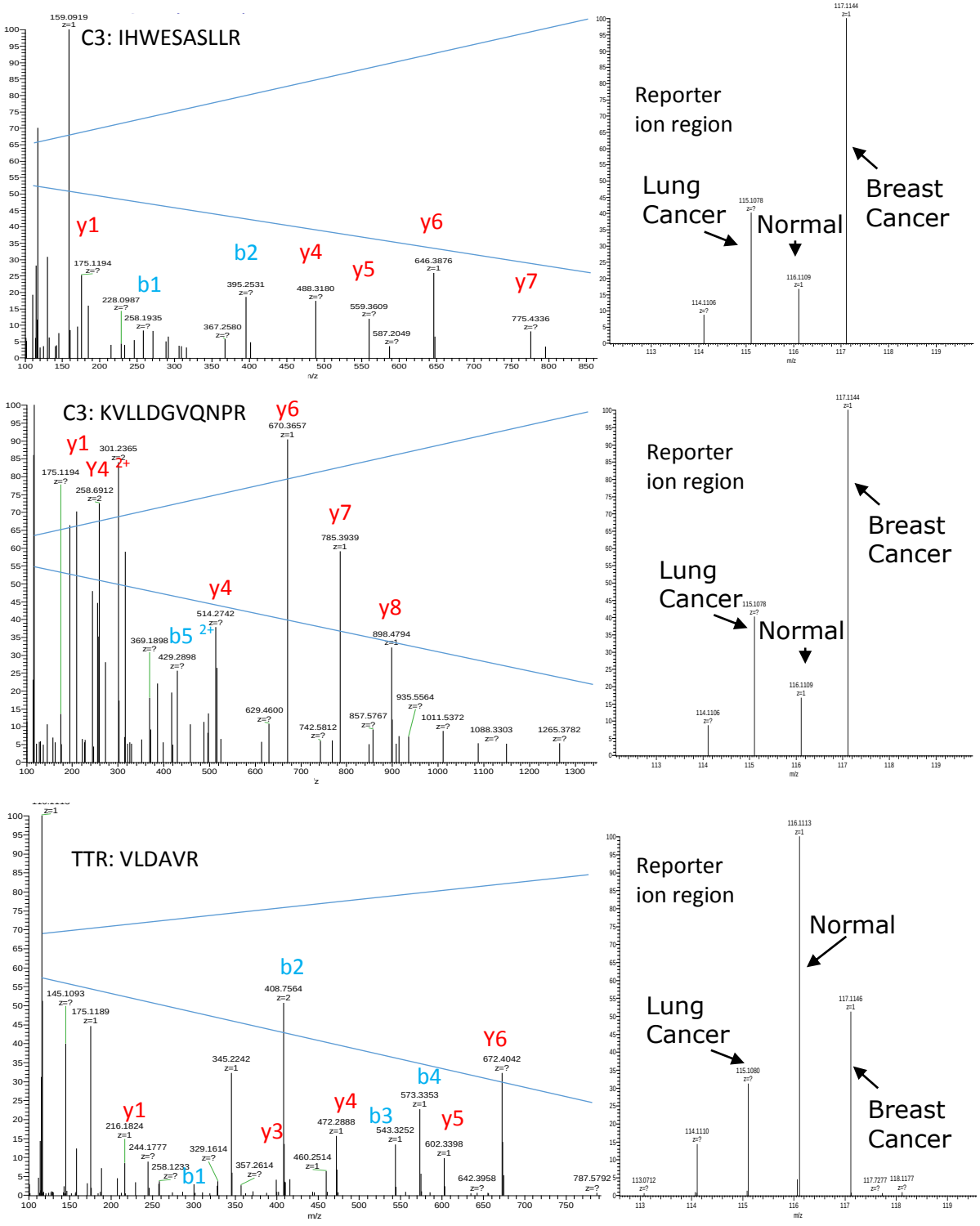


Figure 1 – iTRAQ labeled peptides from two representative proteins observed to be differentially quantified. On the right, the peptide features from the MS2 spectral profile, are magnified to illustrate the differences in reporter intensities between the three sera enriched with AlbuVoid™ and on-bead digested. Top: Complement Component 3 (C3), Middle: Complement Component 3 (C3), Bottom: Transthyretin (TTR).

Results and Discussion

Proteolytic efficiency affects all proteomic applications as computational methods rely on the corresponding predicted tryptic digestion in which every theoretical tryptic site is cleaved *in silico* and compared to the spectral profile from the LC-MS/MS instrument for sequence identification. For labeled quantification, the pooling requirement, requisite peptide data quality and signal to background noise, limits the number of quantifiable proteins to a fraction of the total identified without labels. In this experiment using a singular 3 hour LC-MS gradient, we observed 201 total proteins, of which 21 were differentially quantified between the 3 serum samples, shown in Tables 1 & 2 below.

In Figure 1 above, we highlight a few representative spectral profiles of iTRAQ reporter intensities that are associated with quantitative differences between peptides from normal and cancer sera. Complement Component 3 and Transthyretin serve as examples of proteins that are respectively observed over-expressed and under-expressed in the disease state. In Figure 2 to follow, the iTRAQ reporter signal intensities are plotted for individual proteins, illustrating a consistent ratio for most proteins from each of the three samples, along with some clear outliers.

Table 1. Proteins Observed to be Over-expressed

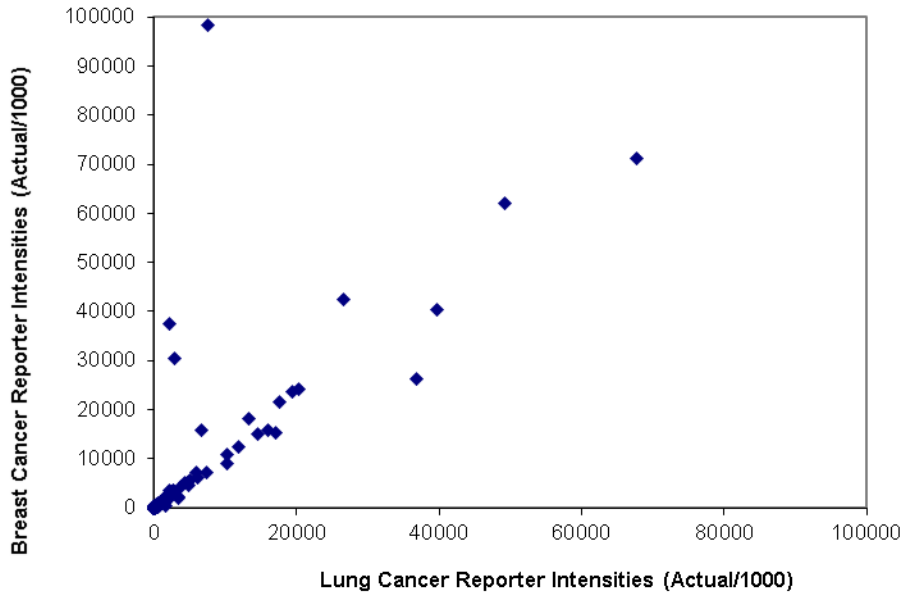
Over-expressed Protein Identification [gene name]	Total Reporter Intensity Ratio Lung Cancer / Normal	Total Reporter Intensity Ratio Breast Cancer / Normal	Total Spectral Count
Immunoglobulin kappa variable 1-17 [IGKV1-17]	**1.2**	**5.5**	**4**
Complement component 3 [C3]	**3.3**	**3.2**	**325**
Wolfram syndrome 1 (wolframin) [WFS1]	**4.2**	**3.1**	**1**
S100 calcium binding protein A9 [S100A9]	**2.3**	**2.4**	**2**
serum amyloid A1 [SAA1]	**3.8**	**2.0**	**13**
serum amyloid A2 [SAA2]	**3.8**	**2.0**	**43**
CD109 molecule [CD109]	**6.4**	**1.2**	**2**
unc-45 homolog A (C. elegans) [UNC45A]	**3.6**	**1.1**	**1**
Armadillo repeat containing 5 [ARMC5]	**3.3**	**1.0**	**1**
coagulation factor II (thrombin) [F2]	**4.3**	**1.0**	**4**
pro-platelet basic protein (chemokine (C-X-C motif) ligand 7) [PPBP]	**2.7**	**0.9**	**12**

Table 2. Proteins Observed to be Under-expressed

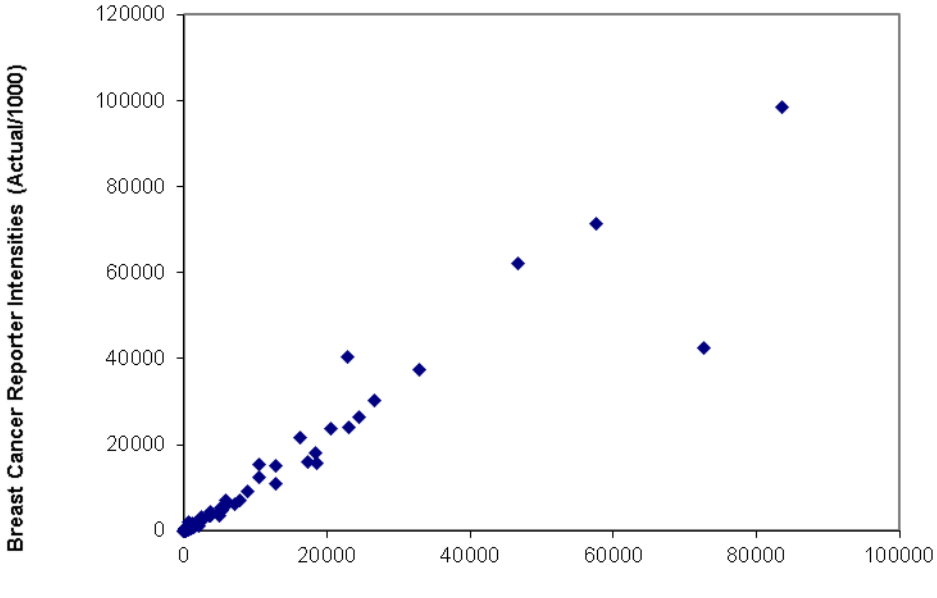
Under-expressed Protein Identification [gene name]	Total Reporter Intensity Ratio Lung Cancer / Normal	Total Reporter Intensity Ratio Breast Cancer / Normal	Total Spectral Count
BRF1, RNA polymerase III transcription initiation factor 90 kDa subunit [BRF1]	0.4	1.1	3
Fibrinogen alpha chain [FGA]	0.2	1.1	401
Fibrinogen beta chain [FGB]	0.2	1.1	141
Fibrinogen gamma chain [FGG]	0.2	1.1	126
sperm specific antigen 2 [SSFA2]	0.1	0.7	3
splicing factor 3a, subunit 1, 120kDa [SF3A1]	0.3	0.6	1
IQ motif containing E [IQCE]	0.3	0.6	1
Transthyretin [TTR]	0.4	0.5	7
amyloid P component, serum [APCS]	0.4	0.5	4
immunoglobulin heavy constant delta [IGHD]	0.3	0.4	7

ITRAQ Labeled Peptides & Reporter Intensity Signals

A



B



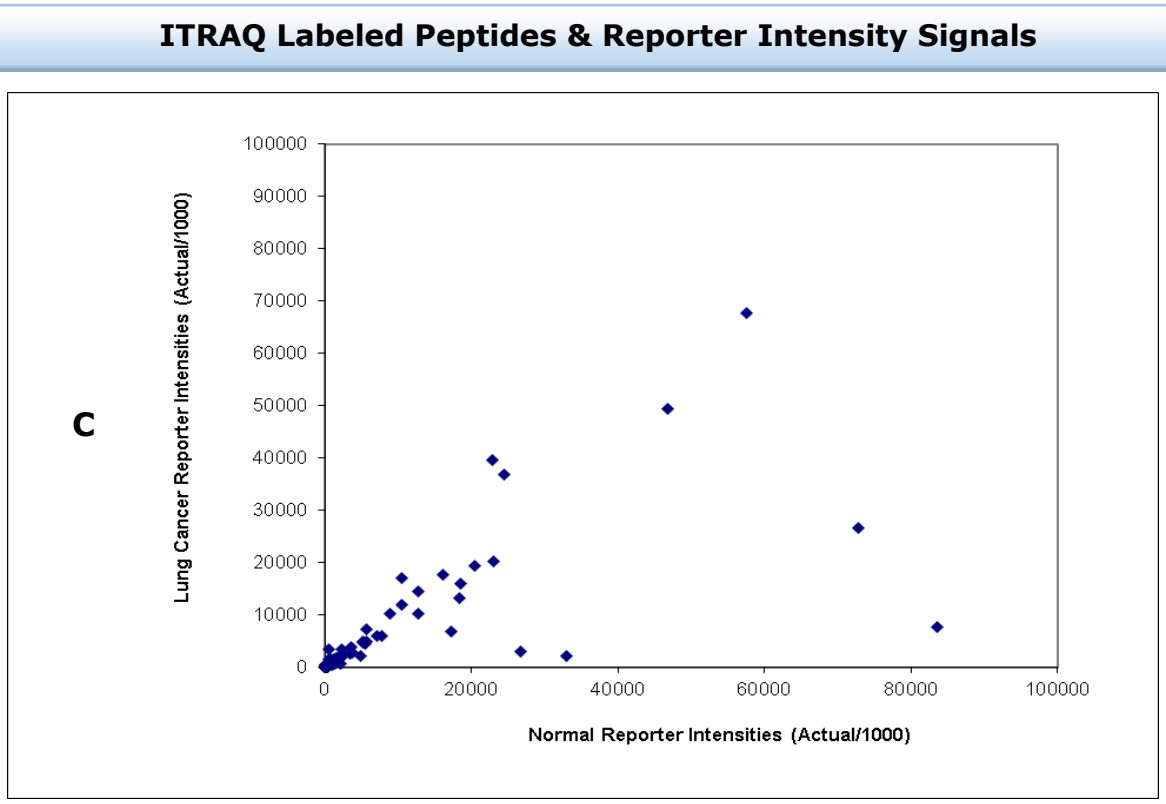


Figure 2 – The reporter intensity signals were added together for each of the iTRAQ labeled peptides supporting the associated protein identification. Each of the additive peptide reporter intensities were then plotted for each protein comparing the following sample pairs: A. Breast Cancer Serum Proteins vs. Lung Cancer Serum Proteins. B. Breast Cancer Serum Proteins vs. Normal Serum Proteins. C. Lung Cancer Serum Proteins vs. Normal Serum Proteins. The linearity between the great majority of proteins identified from each sample demonstrates that the AlbuVoid™ LC-MS On-Bead product and method is suitable for quantitative serum proteomics.

While a review of all potential serum cancer biomarkers is beyond the scope of this report, we nevertheless have observed several correlations with observations by others. Our data supports previous references that Complement Component 3, Serum Amyloid A, and CD109, are over-expressed in lung cancer sera[6-9]. Interestingly, Transthyretin has been reported to be under-expressed in both pancreatic and ovarian cancer sera[10,11]. Inexplicably, we observed under-expression of Fibrinogen in Lung Cancer serum, while Sheng et al., reports Fibrinogen as a prognostic indicator when over-expressed is lung cancer patient serum[12]. Nevertheless, such confluence with other strategies and methods for discovery, give confidence to our workflow as being suitable for proteomic biomarker discovery. We now consider that workflow in terms of efficiency and productivity compared to other LC-MS workflows.

Workflow & Productivity Considerations

The **AlbuVoid™** beads are based on a porous silica chemical platform (called NuGel™) coupled to mixed-mode polymeric ligands. The library was designed to facilitate weak binding of proteins, allowing for progressive enrichment of the low abundance proteome, with specialized voiding properties empirically derived through optimized buffer and binding conditions. Because of its specialized voiding properties, the **AlbuVoid™ LC-MS On-Bead** process depletes two of the three most abundant proteins in serum: Albumin and Transferrin. In this quantitative analysis, Albumin spectral counts are 6% of the total, with no evidence of Transferrin. The third most abundant protein is not a singular protein but rather a heterogeneous mix of Immunoglobulins that report as many gene products. These remain in relatively high abundance, about 18% of the total spectral counts.

While two Immunoglobulin gene products (IGKV1-17 & IGHD) were on the differentiated list of proteins for this investigation, the high level of Immunoglobulin abundance may mask other proteins of interest. For this reason, we will consider ways to minimize the influence of the Immunoglobulin derivative peptides with adjustments to LC-MS parameters in the future. Optionally, one may reduce the influence of the immunoglobulin fraction by additional protein level depletion using conventional biological-based adsorbents. This allows the **AlbuVoid™ LC-MS On-Bead** user great latitude in designing a workflow that meets the needs of any particular investigation.

As has been discussed by others in the field, there is always a tradeoff between effort, throughput and proteome coverage[14]. We consider how the new **AlbuVoid™ LC-MS On-Bead** product can be applied to increase the efficiency of the workflow, expand proteome coverage along with the productivity of LC-MS instrument time. Table 3 illustrates this point, with approximate total proteins identified from workflows described in the proteomic literature[7-11,13-15].

Table 3 – Workflow Comparison of Methods and Productivity

Method	Total Serum Proteins Identified by LC-MS, Label-free * With Label	Workflow Simplicity	LC-MS/MS Instrument Time (minutes)	Relative cost including labor & instrument service time (estimated)
Immuno-depletion	**150-200** ***130**	**++**	**180**	**$$$**
+ peptide level fractionation	**350** ***200-300**	**+**	**1200+**	**$$$$**
2-Dimensional Protein level fractionation, no peptide level fractionation	**100-200**	**+**	**varied**	**$$$$**
AlbuVoid™ LC-MS On-Bead, Singular LC-MS (no peptide level fractionation)	**250-350[2]** ***200**	**++++**	**180**	**$**
AlbuVoid™ LC-MS On-Bead, 2 On-Bead Digest Times	**400-500[2]**	**++**	**360**	**$$**

However it is accounted for, the utility cost of LC-MS instrument time is by far the largest singular expense within the overall workflow, Therefore, LC-MS productivity in terms of total proteins identified and quantified per allocated instrument service time, is an important metric to consider in the optimization of workflows. Because of the complications of integrating quantitative intensities from multiple peptide fractions, along with the associated costs and only marginal gains in proteome coverage, so far we have elected not to consider peptide level fractionation in our investigations using **AlbuVoid™ LC-MS On-Bead**. We recommend users of this product to consider peptide level fractionation and its relative benefits as appropriate or not for their particular investigation. The same is true for gains in productivity using more than one on-bead digestion time[2]. Prospectively, this may be more productive than peptide level fractionation, and is an area for future research.

Conclusions

With the complexity associated with LC-MS/MS analysis, many instrument parameters and computational stringencies will influence the total numbers of proteins that can be identified and quantified. So it is nearly impossible to make any definitive evaluations about workflow efficiencies and final data output. Nevertheless, we conclude that the **AlbuVoid™ LC-MS On-Bead** product and protocol supports efficient workflows as it greatly reduces the abundance of Albumin and Transferrin peptides, and enriches the low abundance proteome. It provides a suitable alternative or addition to other depletion strategies, and enrichment performance can be optimized for the application and goals of the investigation. For example, the product protocol specifies a starting volume range of 50-100 µls. Within this range, the influences of volume per prep can then be optimized for performance, with consideration for total protein amounts, low abundance identifications or quantitative precision. Thus, this new method of low abundance enrichment integrated with on-bead digestion offers many workflow advantages:

- Unique chemically derived beads, it is a consumable, one-time use product with no potential for cross-contamination or reduced performance upon regeneration.
- Species agnostic; human, rat, mouse, goat, sheep, porcine and bovine sera have been tested
- No in-gel digests, no solution digests, no C18 desalting, more consistent, reproducible results
- Compatibility with quantitative label (i.e., iTRAQ) and label-free LC-MS methods
- Versatile, cost-effective workflows

We suggest that it is practical and productive to evaluate different digestion times as this can impact both missed cleavages, and non-specific cleavages. When non-specific cleavages occur, higher proteolytic background can obscure sequence-rich features leaving many proteins unaccounted for. For quantitative applications, the speed and efficiency of **AlbuVoid™ LC-MS On-Bead** digestions can minimize many of the inconsistencies of proteolytic hydrolysis during the generation of serum or plasma peptides. This will prove advantageous for both quantitative discovery (shotgun) as well as targeted SRM/MRM applications.

Acknowledgements
We thank Dr. Haiyan Zheng and her colleagues at the Rutgers Proteomics Center for help with experimental design and LC-MS data.

We thank Dr. Xing Wang at Array Bridge Inc. (St. Louis, MO) for providing the serum samples.

References

1. Maurya, P., Meleady, P., Dowling, P. and Clynes, M. Review: Proteomic Approaches for Serum Biomarker Discovery in Cancer. ANTICANCER RESEARCH 27: 1247-1256 (2007)

2. Biotech Support Group Application Report, entitled "AlbuVoid™ & On-Bead Digestion Tackling the challenges of serum proteomics (LC-MS)". April 16, 2015. http://biotechsupportgroup.com/sites/default/files/AlbuVoid%20On-Bead%20Application%20Report.Rev4%20042015_0.pdf

3. Addona, T.A., et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nat. Biotechnol. (2009) 27, 633-685.

4. Kuzyk, M.A., Smith, D., Yang, J.C., Cross T.J., Jackson, A.M., Hardie, D.B., Anderson, N.L., and Borchers, C.H. (2009) Multiple reaction monitoring-based, multiplexed, absolute quantification of 45 proteins in human plasma. Mol.Cell. Proteomics 8, 1860-1877.

5. Shuford, C.M., Sederoff, R.R., Chiang, V.L., and Muddiman, D.C. (2012) Peptide production and decay rates affect the quantitative accuracy of protein cleavage isotope dilution mass spectrometry (PC-IDMS). Mol. Cell. Proteomics 11, 814-823.

6. Verhaegen, H. et al. Increase in Serum Complement Levels In Cancer Patients With Progressing Tumors. Cancer 1976. 38: pp 1608-1613

7. Comparative Proteomic Analysis of Non-small-cell Lung Cancer and Normal Controls Using Serum Label-Free Quantitative Shotgun Technology Jun Pan Æ Hai-Quan Chen Æ Yi-Hua Sun Æ Jun-Hua Zhang Æ Xiao-Yang Luo. Lung (2008) 186:255–261. DOI 10.1007/s00408-008-9093-7

8. Zhang. F., et al. SWATH™- and iTRAQ-based quantitative proteomic analyses reveal an overexpression and biological relevance of CD109 in advanced NSCLC. Journal of Proteomics, Volume 102, 6 May 2014, Pages 125–136.

9. Sung, HJ., et al. Identification and Validation of SAA as a Potential Lung Cancer Biomarker and its involvement in Metastatic Pathogenesis of Lung Cancer J Proteome Res., 2011, 10(3), pp 1383-1395. DOI: 10.1021/pr101154j

10. Ehmann M., et al. Identification of Potential Markers for the Detection of Pancreatic Cancer Through Comparative Serum Protein Expression Profiling. Pancreas: March 2007 - Volume 34 - Issue 2 - pp 205-214. doi: 10.1097/01.mpa.0000250128.57026.b2

11. Goufman, E.I., et al. Two-dimensional electrophoretic proteome study of serum thermostable fraction from patients with various tumor conditions. Biochemistry (Moscow) 2006, Volume 71, Issue 4, pp 354-360.

12. Sheng, L., et al. Serum fibrinogen is an independent prognostic factor in operable nonsmall cell lung cancer. 1 December 2013. Volume 133, Issue 11. Pages 2511–2757.

13. Song, S.H., et al. Proteomic Profiling of Serum from Patients with Tuberculosis. Ann Lab Med. 2014 Sep;34(5):345-353. http://dx.doi.org/10.3343/alm.2014.34.5.345.

14. Dayon, L., and Kussmann, M., Proteomics of human plasma: A critical comparison of analytical workflows in terms of effort, throughput and outcome. EuPA Open Proteomics. Volume 1, 2013, Pages 8–16. doi:10.1016/j.euprot.2013.08.001

15. Fujii K, Nakano T, Kanazawa M, Akimoto S, Hirano T, Kato H and Nishimura T. Clinical-scale high-throughput human plasma proteome analysis: lung adenocarcinoma. Proteomics 5(4): 1150-1159, 2005.