The Commonality of the Cancer Serum Proteome Phenotype as analyzed by LC-MS/MS, and Its Application to Monitor Dysregulated Wellness

Haiyan Zheng\(^1\); Caifeng Zhao\(^1\); Swapan Roy\(^2\); Devjit Roy MD; Amenah Soherwardy\(^2\); Ravish Amin\(^2\); Matthew Kuruc\(^2\)

\(^1\)Rutgers Center for Integrative Proteomics, Piscataway, NJ; \(^2\)Biotech Support Group LLC, Monmouth Junction, NJ

Introduction and Objectives

For many diseases, pancreatic cancer for example, long term survival is critically dependent upon early detection. So various strategies for early detectable markers are being investigated. One such strategy is to identify a singular biomarker, derived from genomic analysis, and then determine its derivative protein concentration in blood. This has been challenging as many differentially regulated genes do not generate a differentially regulated protein. An alternative biomarker strategy is to consider panels of proteins as up and/or down regulated biomarkers, which differ in diseased and normal states. Following this strategy, many serum proteomic investigations have analyzed one cancer type or another, but our goal was to consider whether patterns of multiple proteins dysregulated in cancer could be observed regardless of the primary tumor, stage of progression or tumor burden. In this study we adopted a relatively new method, which combines Albumin depletion and on-bead digestion of the depleted serum in a seamless process, called AlbuVoid™ LC-MS On-Bead. From this method, we were able to compare isobaric labeled quantification of proteins, using LC-MS/MS from normal and disease state sera – for this case, breast, lung and pancreatic cancer. The methods spectrally quantified over 200 total proteins, and most notably enriched for sub-populations of the protease inhibitor Alpha-1-Antitrypsin, which as a ratio of bead-bound fractions to unbound fractions, is severely dysregulated in the cancer samples. From this data, we solicit that biomarkers from categorical panels can distinguish a cancer serum phenotype at very early stages, possibly even before presentation of any clinical evidence. Such ability to monitor dysregulation from wellness can now be refined by future more detailed cancer sera investigations.

For such a strategy to be viable, most proteins from normal/healthy sera should be within a small and consistent variance taking into account both technical and biological variance for the proteins measured. This has been established previously with the methods reported here\(^1\). With confidence that variance of normal is limited, a panel of proteins can be selected that are observed to be dysregulated (either up, down, or cyclical up/down) within the disease state; defining in our case a cancer phenotype. We cite one report supporting such a strategy using multiple protein markers in a risk prediction
model for lung cancer in lieu of stand-alone tests\(^2\). Yet in almost all previous studies, only one primary tumor site (lung or pancreas for example) was considered in the study. Notably, we found only one report which considers the common proteomic patterns of three cancer primary tumor cites - breast, colon and lung, revealing distinctive expression of acute-phase proteins\(^3\). Similarly, others report on a hyper-coagulation state with many cancers\(^4\). This suggests the feasibility of our goal, revealing a measureable cancer phenotype in serum, without regard to primary tumor of origin, clinical stage of progression, or tumor burden.

In our prior report, 3 cancer types – pancreatic (Pancr), breast (Brst) and lung cancer, with sera taken from clinically characterized stages I – IV, were compared against a pooled sera from matched 5 normal/healthy individuals of similar age and sex, in this case females, ages 40-60. In like manner, we also considered the variance within these same normal/healthy individuals to account for any combined technical and biological variance with our methods. A discussion of how these observations compare to previous observations by others and areas for future research has been reported previously\(^1\).

One particular observation stood out. The protein identified as SERPINA1 and more commonly known as Alpha-1-Antitrypsin (AAT) was down-regulated in all cancer samples tested, and in most cases, such down-regulation was severe – twofold or more relative to the normal reference. This contrasts to other reports that cite this protein as being up-regulated in cancer and inflammation, except when gene mutation drives functional inactivation. As we suspected what we were observing was a distinctive sub-population, a second series of tests herein reported, support our important conclusion that different sub-populations of Alpha-1-Antitrypsin are dysregulated in the cancer sera.

**Methods**

**Test 1 Methods**

The basic workflow considered:

- Albumin Removal and low abundance bead bound enrichment of remaining sub-proteome (without Albumin)
- On-Bead Digestion, 4 hours to minimize proteolytic background
- Single 3 hour LC-MS, no peptide level fractionation
- TMT labels, ratio cancer/normal thresholds >1.5 up-regulated, <0.7 down-regulated
- >200 Proteins observed

A detailed analysis of the methods and the biomarker proteins has been previously documented as a poster report first presented at the 12th Annual US HUPO Conference, March 13-16, 2016. A reprint is available from the company website\(^1\).

**Test 1 Results and Discussion**

We note two reports that a collection of blood based biomarkers measuring inflammatory and acute phase response proteins, measured within a panel rather than singular tests, can model early detection of cancer 2 to 3 years prior to clinical evidence\(^2,5\). Our initial study supports others that describe common dysregulated functions within the cancer phenotype, regardless of the primary tumor or tumor burden. We similarly found **inflammatory**, **coagulation**, and **tissue remodeling** biomarkers which
are common to the 3 primary tumors tested and possibly common to the majority of primary tumors\(^1\). With the exception of *glycolysis* which is likely derived from tumor burden, such functional dysregulation within blood may serve a supportive premalignancy microenvironment necessary for cancer progression. We also observed proteins not previously reported with cancer associations, along with several with little or no previous annotation to serum/plasma at all; 2 of which were theoretical only based on gene sequence, and not previously reported at the protein level in the public domain. We categorize these proteins as *unknown* function.

**AlbuVoid™ Methods Reveal Altered Cancer Associate Protein Levels Categorized into 5 Separate Biomarker Panels:**
- Inflammation
- Coagulation
- Tissue Remodeling
- Glycolysis
- Unknown

A representative list of some of these categorical biomarkers are shown in Table 1.

**Table 1. Representative List of Categorical Cancer Serum Biomarkers**

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Protein Description</th>
<th>Category</th>
<th># Individuals Threshold /All</th>
<th>Cancer Types</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRP</td>
<td>C-reactive protein</td>
<td>Inflammation</td>
<td>7/11 Up-regulated</td>
<td>Pancr, Breast, Lung</td>
<td>Acute phase response protein, may scavenge nuclear material released from damaged circulating cells. Others report up-regulated in cancer (6) (7)</td>
</tr>
<tr>
<td>PIGR</td>
<td>Polymeric IgG</td>
<td>Inflammation</td>
<td>3/5 Up-regulated</td>
<td>Pancr, All Stages</td>
<td>Mediates transcellular transport of polymeric immunoglobulins. Others report up-regulated in cancer (8) (9)</td>
</tr>
<tr>
<td>SEMA3D</td>
<td>Sema domain</td>
<td>Inflammation</td>
<td>3/3 Down-regulated</td>
<td>Lung</td>
<td>Neural development, may have role in immune function. No known reports of cancer association.</td>
</tr>
<tr>
<td>IGLV3-27</td>
<td>Ig-like</td>
<td>Inflammation</td>
<td>9/10 5/5 Stg 1 Down-regulated</td>
<td>Pancr, Breast, Lung</td>
<td>Immunoglobulin. No known previous reports of cancer association.</td>
</tr>
<tr>
<td>PF4</td>
<td>Platelet Factor 4</td>
<td>Inflammation</td>
<td>13/13 9/13 severe Up-regulated</td>
<td>Pancr, Breast, Lung</td>
<td>Promotes blood coagulation, may play role in wound repair and inflammation. Others report up-regulated in cancer (10) (11) (12)</td>
</tr>
<tr>
<td>PPBP</td>
<td>Beta thromboglobulin aka CTAPIII/NAP-2</td>
<td>Coagulation</td>
<td>13/13 9/13 severe Up-regulated</td>
<td>Pancr, Breast, Lung</td>
<td>Platelet activation. Elevated levels predated lung cancer diagnosis by 29 months, with cyclical levels upon surgical resection and recurrent disease (2)</td>
</tr>
<tr>
<td>CLU</td>
<td>Clusterin</td>
<td>Tissue Remodeling</td>
<td>12/13 Up-regulated</td>
<td>Pancr, Breast, Lung</td>
<td>Multiple isoforms, associated with clearance of cellular debris and tissue remodeling. Up &amp; down regulation have been described for many cancers, may be isoform specific (3) (13)</td>
</tr>
</tbody>
</table>
Table 1 (Continued). Representative List of Categorical Cancer Serum Biomarkers

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Protein Description</th>
<th>Category</th>
<th># Individuals Threshold /All</th>
<th>Cancer Types</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPN1</td>
<td>Carboxypeptidase N, serum</td>
<td>Tissue Remodeling</td>
<td>10/13 Up-regulated</td>
<td>Pancr, Breast, Lung All Stages</td>
<td>metallo-protease No known previous reports of cancer association</td>
</tr>
<tr>
<td>SERPIN G1</td>
<td>serpin peptidase inhibitor, clade G (C1 inhibitor), member 1</td>
<td>Tissue Remodeling Inflammation</td>
<td>12/13 Up-regulated</td>
<td>Pancr, Breast, Lung All Stages</td>
<td>protease inhibitor, levels rise ~2-fold during inflammation Others report Up-regulation in lung cancer (7)</td>
</tr>
<tr>
<td>SERPIN A1</td>
<td>Alpha-1-antitrypsin (AAT)</td>
<td>Tissue Remodeling Inflammation</td>
<td>13/13 9/10 severe Pancreatic, Breast Down-regulated</td>
<td>Pancr, Breast, Lung All Stages</td>
<td>Concentration can rise manyfold upon acute inflammation. Reports of up-regulation in cancer (14), and mutation driven down-regulation in emphysema and cancer (15). We observe strong down-regulation, severe in Breast and Pancreatic; suspect measuring sub-populations and not total populations</td>
</tr>
<tr>
<td>PFKM</td>
<td>Phosphofructokinase 1</td>
<td>Glycolysis</td>
<td>5/5 Up-regulated</td>
<td>Pancr, Breast, Stage 1</td>
<td>Key regulatory and rate limiting step of glycolysis (Warburg effect), unexpected cytosolic protein in serum/plasma. Glycolysis pathway observed in cancer sera by 2DE/LC-MS &amp; functional proteomics (16) (17)</td>
</tr>
<tr>
<td>C18orf63</td>
<td>No annotation, simply called Uncharacterized</td>
<td>Unknown Function</td>
<td>Up-regulated</td>
<td>Pancr, Breast, Lung All Stages</td>
<td>Not consistently observed with low spectral counts, but outliers are highly up-regulated</td>
</tr>
<tr>
<td>CFAP61</td>
<td>Cilia- and flagella-associated protein 61</td>
<td>Unknown Function</td>
<td>7/8 Down-regulated</td>
<td>Pancr, Breast, All Stages</td>
<td>not observed in Peptide Atlas in Human Plasma</td>
</tr>
<tr>
<td>CTD-2007N20 .1</td>
<td>No annotation</td>
<td>Unknown Function</td>
<td>6/8 Down-regulated</td>
<td>Pancr, Breast, All Stages</td>
<td>Not observed at the protein level in public data. Ensembl gene annotation: Immunoglobulin (Ig) and T-cell receptor (TcR)</td>
</tr>
</tbody>
</table>
Considerations for Test 2 – Testing for Sub-populations of Alpha-1-Antitrypsin

The interesting results that many cancer associated proteins fall into categories led us to consider the impact of those categories on our goal – to establish a cancer serum phenotype protein panel that could be monitored over time; threshold values to which would signal a dysregulation from wellness and trigger a deeper clinical evaluation. One particular protein – SERPINA1, also known as Alpha-1-Antitrypsin (AAT) in the category of tissue remodeling stood out. For several reason AAT was particularly interesting to us; its consistency in all 13 cancer individuals tested, the severity of the dysregulation and most importantly, the contradiction of our observations with those of others. As we were in essence performing a protein level separation by our methods, we suspected that what we might be observing was not the total AAT population, but a fractionated sub-population. So additional tests considered where these different sub-populations might fractionate and how they might be reported.

The SERPIN superfamily of proteins

Tissue remodeling is an essential feature of cancer, and proteases play a key role. Consequently, the balance and regulation of proteolytic activity is essential to biomarker discoveries and possibly to therapeutic intervention. Many of these regulating proteins fall under the SERPIN superfamily of suicidal protease inhibitors. One such inhibitor SERPINA1, known more commonly as Alpha-1-Antitrypsin (AAT), has several isoforms observed in plasma using 2-DE\textsuperscript{18}. Others report conformational properties of AAT having multiple effects on tumor cell viability and diverse roles in tumorigenesis, suggesting such isoforms may display a specific basis for diagnosis of cancer\textsuperscript{14,19,20}. Yet, most often in proteomics, all sub-populations of AAT are rolled into and counted as one homogeneous population, or as in the case of immuno-depletion, simply ignored as background noise. As a result, the regulation, balance and dynamism within these systems and its impact on disease progression cannot be properly investigated. As an example, AAT substrates can activate proMMP-2, a metalloproteinase involved in tumor invasion and angiogenesis\textsuperscript{21}. Key sub-populations of the SERPIN superfamily of protease inhibitors are represented here\textsuperscript{22}.
Current proteomic methods inadequately count the many differential sub-populations (also known as proteoforms) of key protease inhibitors in serum. From the data observed and reported in Table 1, on our first series of tests we suspected that our methods had enriched for a sub-population of AAT. To further validate this hypothesis, we report on a second series of tests; to determine if a second sub-population distinguishable by peptide reporting features could be observed in the unbound (or flow-through), distinct from the bead-bound to which we initially observed. For this second test, we only considered a pooled pancreatic cancer serum vs. a pooled normal control serum.

Methods for Test 2

TMT labels were applied to the tryptic peptides generated from paired samples: pooled pancreatic cancer sera (n=5), and normal/healthy, age & sex matched pooled individuals (n=5). The separation protocols were the same as in Test 1. Digest methods for the bead were the same as in Test 1. As we also analyzed the Flow-through from the beads (the unbound fraction) and Untreated, 2ul serum equivalent of AlbuVoid™ Flow-through and Load (Untreated) were run on SDS-PAGE as gel plugs and in-gel digested with trypsin using standard protocols.

**Basic AlbuVoid™ Workflow**

- Serum diluted in Binding Buffer added to tube and vortexed
- Diluted serum in bead
- Centrifugation

**Bead-bound proteome sub-population**
- Inflammation
- Coagulation
- Tissue Remodeling
- Glycolysis
- Unknown

**Flow-through proteome sub-population**
- Special focus on AAT
## AAT TMT Ratio: Pooled Pancreatic Cancer / Pooled Normal

<table>
<thead>
<tr>
<th>AAT Peptide Region</th>
<th>start</th>
<th>Amino acid sequence</th>
<th>end</th>
<th>Bead Bound TMT Ratio</th>
<th>Sp Ct Bead Bound</th>
<th>Flow-Through (Unbound) TMT Ratio</th>
<th>Sp Ct Flow-Through</th>
<th>Serum Untreated TMT Ratio</th>
<th>Sp Ct Serum Untreated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjacent RCL Tryptic</td>
<td>360</td>
<td>AVLTIDEK</td>
<td>367</td>
<td>0.35</td>
<td>9</td>
<td>1.78</td>
<td>21</td>
<td>1.53</td>
<td>14</td>
</tr>
<tr>
<td>RCL Cleaved</td>
<td>368</td>
<td>GTEAAGAMFLEAIM</td>
<td>382</td>
<td></td>
<td></td>
<td>1.05</td>
<td>7</td>
<td>1.16</td>
<td>23</td>
</tr>
<tr>
<td>RCL Intact</td>
<td>368</td>
<td>GTEAAGAMFLEAIMSIPPEVK</td>
<td>389</td>
<td>0.77</td>
<td>5</td>
<td>1.75</td>
<td>1</td>
<td>1.34</td>
<td>50</td>
</tr>
<tr>
<td>RCL Cleaved</td>
<td>383</td>
<td>SIPPEVK</td>
<td>389</td>
<td></td>
<td></td>
<td>1.45</td>
<td>27</td>
<td>1.44</td>
<td>460</td>
</tr>
<tr>
<td>Total all peptide features</td>
<td></td>
<td></td>
<td></td>
<td>0.54</td>
<td>132</td>
<td>1.57</td>
<td>372</td>
<td>1.44</td>
<td>460</td>
</tr>
</tbody>
</table>

We solicit that for the first time, our methods have unraveled sub-populations distinguishing an **Intact RCL region sub-population** from a **cleaved RCL region sub-population**. Remaining for future investigation are the peptide level features that report the covalently modified protease substrate complex.

**Variant Sub-populations of Alpha-1-Antitrypsin (AAT) severely distinguished In Cancer Serum Phenotype**

**Bead-bound sub-population severely down-regulated in cancer.**  
LC-MS reports intact RCL peptide region

**Flow-through sub-population up-regulated in cancer.**  
LC-MS reports cleaved RCL peptides in Flow-through

**AlbuVoid™ Protein Level Separation**
The amino acid region of the RCL is 368-392, so the adjacent RCL tryptic peptide at Lys367, highlighted in gray serves as a good comparison between the 3 observable serum populations:

- **Bead-Bound** – The sub-population of proteins that bind and are observed by the AlbuVoid™ on-bead methods
- **Flow-Through (Unbound)** – The sub-population of proteins that flow-through the AlbuVoid™ beads, unbound
- **Untreated** – The total population of proteins that are observable in serum without any sample enrichment, that is without the use of AlbuVoid™

Highlighted in orange is the RCL intact peptide. Highlighted in green are the two RCL peptides that are cleaved at Met382, during suicidal substrate interaction; note that these peptides were not observed in the Bead-Bound fraction. These data suggest that the overall AAT population is dominated by the sub-population up-regulated and collected in the Flow-through fraction of AlbuVoid™, and this same sub-population dominates the analysis when untreated sera is investigated. Such is the case in acute AAT up-regulation commonly observed with malignancies and inflammation. However, using our methods we distinguish a sub-population separated by the bead and reporting with the bound fraction, as being severely down-regulated with cancer! While this observation may have potential biological significance, no conclusion about the particular cancer-specific proteoform uncovered can be made at this time. Nevertheless, from a biomarker perspective, this serves an additional multiplier benefit; measured as a ratio, these two sub-populations may be the most distinguishable pattern of early dysregulation observable in the cancer serum phenotype to date. For this analysis, the ratio of the Adjacent RCL Tryptic peptide region would be 1.78/0.35=5. Additional tests report a similar pattern for all three primary tumors, lung, breast as well as pancreas, suggesting that this dysregulated AAT ratio phenotype is common to many if not most primary tumors. As isobaric label ratios in discovery methods can sometimes compress the reporting difference, this ratio may become much greater once more targeted quantitative methods are developed, a goal of future tests.

**Conclusions**

- Whole new fields of cancer biomarker discoveries rests in the data-rich features of the diverse variety of conformational and proteoform variants associated with classical high abundance proteins like the SERPIN superfamily (i.e., Alpha-1-Antitrypsin). Being nothing short of an interference biomolecule, AAT for example, is quite often immuno-depleted prior to LC-MS. More generally, mid to high abundance proteins are quantified as one homogeneous population, negating the disease specificity that can be obtained through the discreet quantification of the multiple sub-populations available to measure. Our methods begin to unravel and sort these variant sub-populations (also known as proteoforms) so that peptide reporting functional features can distinguish these with more detail, leading to characteristic disease profiles.

- Future investigations will consider the refinement of the reporting features of AAT proteoforms at peptide and functional levels, to distinguish sub-population signatures of Alpha-1-Antitrypsin in cancer. Such new reporting features may serve as a model for the SERPIN superfamily of suicidal protease inhibitors, many of which are involved in cancer and other disease pathologies.
Finally, we solicit that there is a measurable serum cancer phenotype that can be modeled with categorical proteins taken from inflammation, blood coagulation, tissue remodeling, glycolysis, as well new markers of unknown function. Within this framework, we suggest it feasible to baseline monitor normal/healthy individuals and determine a dysregulation pattern associated with cancer generally but not necessarily for a particular primary tumor. Under the guidance of a physician, such dysregulated patterns may serve as a ‘liquid biopsy’, forming an early indicator for cancer before clinical evidence. These same patterns may be prognostic and even offer therapy guidance. More refined algorithms for each of these purposes: early detection, prognosis and therapeutic options, can be made with a more detailed investigation in the future. We welcome inquiries in this regard.

For more information, contact:

Matt Kuruc
mkuruc@biotechsupportgroup.com

Tel: 732-274-2866

www.biotechsupportgroup.com
Acknowledgement: We thank Dr. Andrew Thompson at Proteome Sciences plc (Surrey UK) for donating the TMT-6 labels.

References


