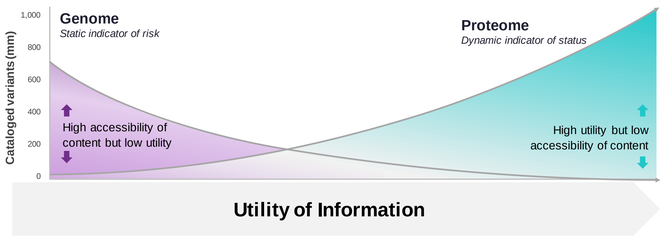
Application of the Proteograph™ Product Suite to the Identification of Differential Protein Isoform Plasma Abundance in Early Lung Cancer vs. Healthy Controls

Asim Siddiqui*, John E. Blume, Margaret K. R. Donovan, Marwin Ko, Ryan W. Benz, Theodore L. Platt, Juan C. Cuevas, Serafim Batzoglou and Omid C. Farokhzad

Proteograph Platform Delivers Unbiased, Deep and Rapid Proteomics at Scale

The ~20,000 genes in the human genome encode over one million protein variants, because of alternative splice forms, allelic variation and protein modifications. Though large-scale genomics studies have expanded our understanding of cancer biology through analysis of both tissue and biofluids, similarly-scaled unbiased, deep proteomics studies of biofluids have remained impractical due to complexity of workflows. We have previously described the Proteograph Product Suite, a novel platform that leverages the nano-bio interactions of nanoparticles for deep and unbiased proteomic sampling at scale. We have shown the utility of the Proteograph solution for unbiased and deep interrogation of plasma from 141 subjects: 80 pre-classified healthy controls and 61 samples from early-stage NSCLC patients to create a plasma biomarker classifier for the detection of NSCLC versus healthy controls with AUC of 0.91[1]. Here we present a further analysis of this data to dissect differences between patients and controls in plasma abundance of protein isoforms arising from alternative gene splicing.



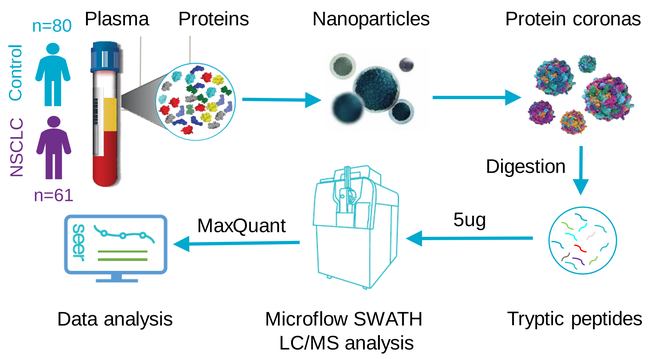
10M+ human exomes
1M+ genomes

~695M genetic variants catalogued

< 0.2% of genetic variants fully characterized

Core Technology

Proprietary Engineered Nanoparticles



Proteograph Product Suite

Sample is ready to be analyzed on most LC/MS instruments

Utilizing Proteograph Platform to Interrogate Protein Isoforms in a Non-small Cell Lung Cancer (NSCLC) Plasma Proteome Study

Proteograph Platform Enables Deep and Unbiased Plasma Proteomics

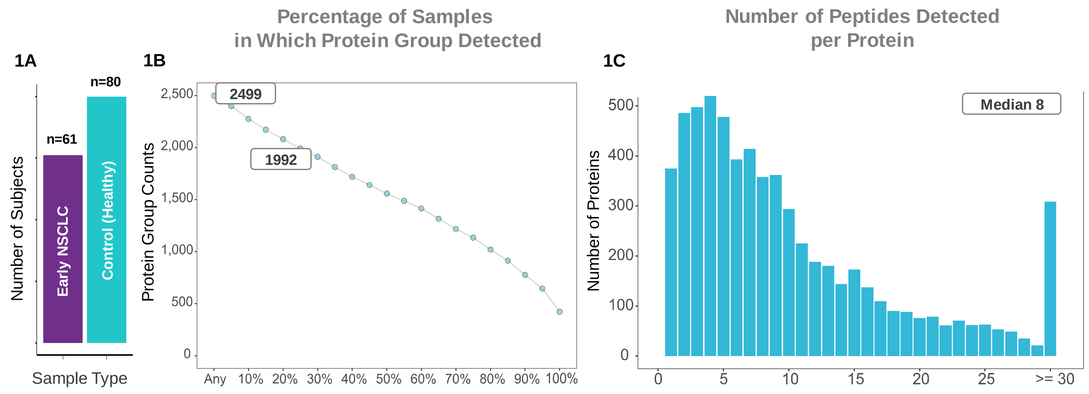


Figure 1. NSCLC study using the Proteograph Product Suite. Number of subjects in NSCLC study including healthy controls and early NSCLC samples (1A). Protein groups detected across 141 subjects in control vs. early NSCLC plasma samples (1B), across a percentage of the subjects in the NSCLC study. 2499 protein groups are found across all subjects and 1,992 in 25% of the subjects. Number of peptides for each protein group in the NSCLC study (1C) shows, 21,959 peptides are detected in total with a median 8 peptides per protein across the NSCLC plasma study using the Proteograph Product Suite.

Identification of Putative Protein Isoforms Using Peptide Abundance

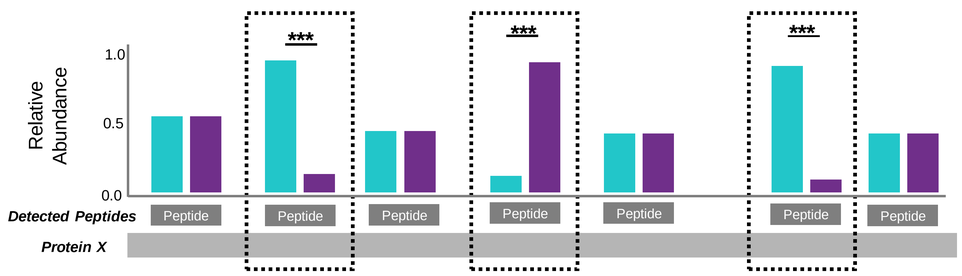


Figure 2. Overview of the putative protein isoform identification strategy. From 1,992 proteins we filtered to proteins present in at least 50% of subjects from either heathy or early cases and searched for peptides that had differential abundance between controls and cancer ($p < 0.05$; Benjamini-Hochberg corrected). Next, we filtered for proteins comprising sets of peptides where at least one peptide had significantly higher and another significantly lower plasma abundance in healthy controls vs. early NSCLC.

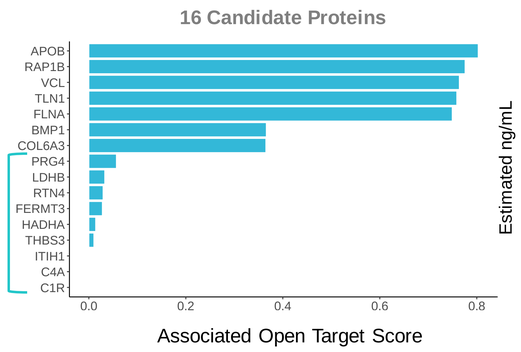


Figure 3. Identification of 16 putative protein isoforms and corresponding open targets score. Associated open targets score for lung carcinoma targets with putative protein isoforms. Nine novel lung carcinoma targets with little or non-existing information are highlighted (teal bracket).

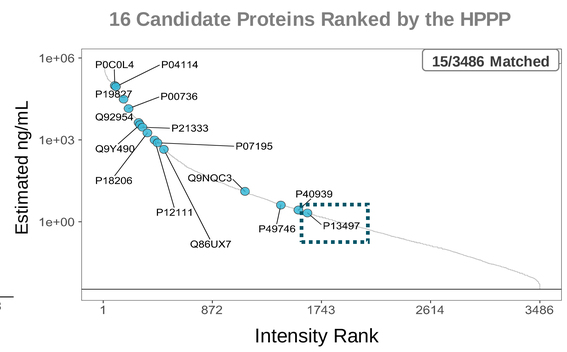


Figure 4. Putative protein isoforms matched to the Human Plasma Proteome Project (HPPP)[2]. 15 out 16 proteins were found in the HPPP. These were ranked by estimated concentration in plasma. P13497 (BMP1) is the least abundant. P61224 (RAP1B) is not present in the HPPP.

Proteograph Data Sheds Light on Biological Consequences of Protein Isoforms

BMP1 Shows Differential Isoform Abundance Pattern

The lowest abundant putative protein isoform, BMP1 from the list of candidate proteins identified in this study, comprises four protein coding isoforms. Two of these isoforms are substantially longer (~400-800 residues) than the other two isoforms covering additional exons. Peptides mapping to exons that cover all four protein isoforms (5A) have higher abundance in cancer relative to controls, whereas peptides mapping to exons that cover only the two longer isoforms have higher abundance in healthy controls. BMP1 is known to play a dual role in cancer, acting as both suppressor and activator[3] and this differential pattern of isoform abundance may shed further light into BMP1's role in cancer.

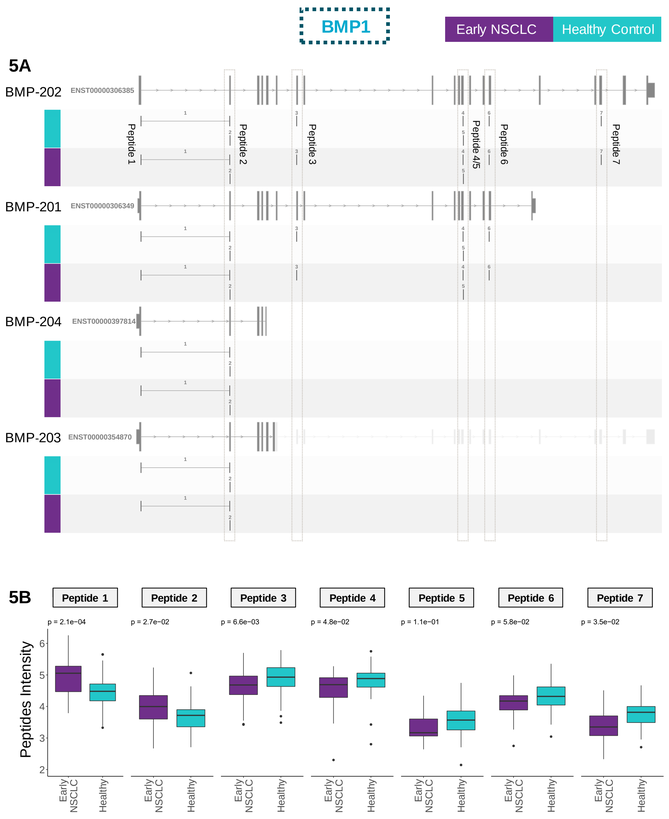


Figure 5. BMP1 exon map and peptide intensity. BMP1 peptides from NSCLC early and control subjects mapped to four BMP protein-coding transcripts (5A) and corresponding peptide intensities (5B) shows; two of the peptides (purple shading) are more abundant in NSCLC and five of the peptides (teal shading) are more abundant in healthy controls.

We demonstrated that measurements at the peptide level for the plasma proteome enable quantification of differential isoform abundance patterns, which were inaccessible to prior methods of lesser scale, depth, or coverage compared to the Proteograph platform. By extending our approach to include additional features such as protein amino acid variants and PTMs, we anticipate extending our knowledge to enable proteogenomics.

The Proteograph Product Suite has the throughput required for the identification of protein isoforms across the proteome in an unbiased manner, enabling cancer biology research and biomarker discoveries at scale.

References:

1. Blume et al. Nat. Comm. (2020)
2. Deutsch et al. J. Proteome Res.(2018)
3. Bach et al. Mol. Ther. Oncolytics (2018

Publications

Seer, Inc., Redwood City, CA - *asiddiqui@seer.bio