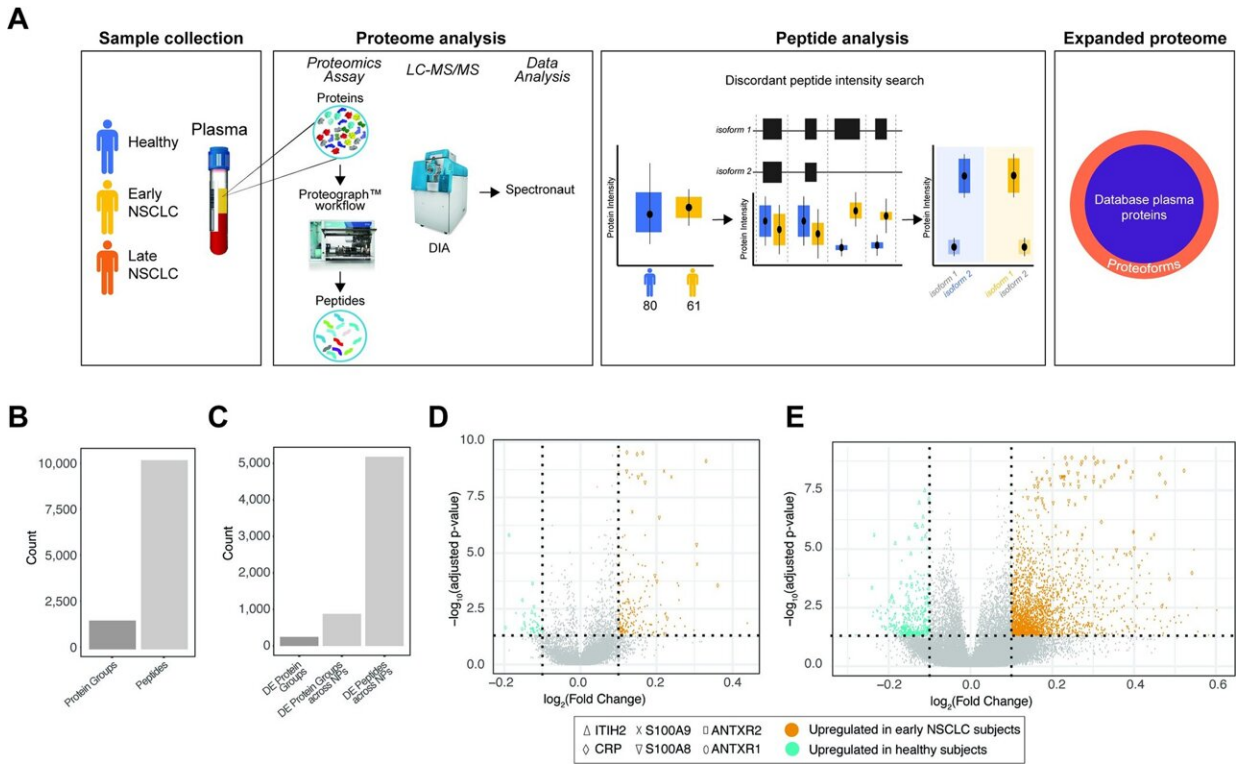
Protein isoforms could serve as novel non-small cell lung cancer biomarkers or therapeutic targets

March 29 2023



Proteome analysis of healthy and NSCLC subjects using a 5 NP plasma workflow. A. Overview of this proof-of-concept proteoform identification study. Plasma samples were collected from healthy (blue), early non-small cell lung cancer (NSCLC; yellow), late NSCLC (orange), and co-morbid (green) subjects (Sample Collection). The plasma proteomes were analyzed for each of these subjects, which included protein extraction, protein discovery using the NP-based Proteograph platform, then DIA protein/peptide identification and quantification using LC-MS/MS and search algorithms (Proteome Analysis).

Proteoforms were then identified using a discordant peptide intensity search, which included examining peptide mappings to known protein coding isoforms and using differential abundance to discover protein isoforms. Together, these identified proteoforms represent an expanded plasma proteome database not captured in standard MS-based or targeted proteomic studies (Expanded proteome). B. Barplots showing the number of peptides and proteins groups retained after filtering to those present in at least 50% of subjects from either heathy or early NSCLC. C. Barplots showing the number of differentially abundant (DA): 1) protein groups, with collapsed abundances using MaxLFQ; 2) protein groups across NPs (i.e., DA independently across NPs); and 3) peptides across NPs. D. Volcano plot showing the significance (adjusted p-value; y-axis) and fold change (x-axis) from calculating the differential abundance of protein groups across NPs between healthy and early NSCLC subjects. Protein groups with a log2(Fold Change) greater or less than 1.0 and adjusted p-value PLOS ONE (2023). DOI: 10.1371/journal.pone.0282821

Proteomics is the study and analysis of proteins, including their structures, functions, interactions, and modifications within a biological system. Humans have approximately 20,000 genes that produce over a million different protein variants, including an estimated 100,000 protein isoforms. These protein variants comprise the vast complexity of the proteome that has remained largely unexplored.

There is a growing body of literature demonstrating that protein isoforms are of high biomedical importance, because isoforms of a protein that arise from the same gene can have distinct biological functions and play different roles in cellular processes. The study of the proteome at the level of protein variants across populations can inform our understanding of biology and provide future scientific advancements in how we identify and treat disease and manage health.

In a study published on March 29 in the journal *PLoS One*, Seer

scientists collaborating with Luis Diaz, M.D. from the Memorial Sloan Kettering Cancer Center found four structurally distinct protein isoforms that were differentially expressed in non-small cell lung cancer (NSCLC) patients. The identification of protein isoforms that are differentially expressed in NSCLC was made possible by comparing the proteomic profiles of people with cancer versus healthy subjects at peptide-level resolution using deep, unbiased proteomics methods.

"These findings demonstrate that distinct protein isoforms may differentially contribute to diverse biological mechanisms and to the pathogenesis of cancers, potentially paving the road to identify new diagnostic markers or new therapeutic targets," said Dr. Diaz, Head of the Division of Solid Tumor Oncology at Memorial Sloan Kettering Cancer Center. "What is especially encouraging here is that these protein isoforms were detected in plasma, a readily accessible sample type, enabling cancer detection and monitoring through liquid biopsy evaluation of patients."

The study analyzed 188 plasma proteomes from NSCLC patients and healthy controls to identify disease-associated protein isoforms. These protein variants play distinct roles in the body's biological pathways and can influence disease predisposition and progression. Seer's deep, unbiased proteomics platform revealed multiple isoforms of proteins, some of which showed significantly different behavior in NSCLC when compared to healthy controls, including bone morphogenetic protein 1 (BMP1), complement component 4a (C4a), complement component 1r (C1r) and lactate dehydrogenase B (LDHB).

One key isoform of the protein BMP1 had never been previously associated with NSCLC. The researchers found that healthy people had higher levels of a standard long form of the protein, while NSCLC patients had higher levels of a shortened isoform of the protein that lacks the ability to release collagen, which could impact the tumor

microenvironment. The findings suggest BMP1 may play a role in NSCLC disease pathogenesis.

"This study identified novel disease-associated protein isoforms in NSCLC, demonstrating the utility of an unbiased analysis of the proteome at peptide-level resolution, which provides a deeper, more nuanced assessment of the human proteome," said Asim Siddiqui, Senior Vice President of Research at Seer. "These findings underpin the importance of an unbiased approach to reveal novel biological content such as cancer-specific protein isoforms."

Future research into the discovered protein isoforms may validate their role in cancer, utility as novel biomarkers, or potential as therapeutic targets for NSCLC. By using a deep, unbiased approach to investigate the proteome at peptide-level resolution, novel protein isoforms that play a biologically significant role can be discovered for a variety of diseases, including cancer, that would otherwise not be discovered using targeted technologies.

More information: Margaret K. R. Donovan et al, Functionally distinct BMP1 isoforms show an opposite pattern of abundance in plasma from non-small cell lung cancer subjects and controls, *PLOS ONE* (2023). DOI: 10.1371/journal.pone.0282821

Provided by Seer