JAMA Insights

Molecular Phenotyping With Proteomics

Devin K. Schweppe, PhD; Brian J. Beliveau, PhD; Andrew N. Hoofnagle, MD, PhD

A proteome is the entire set of proteins produced in a biological system and proteomics is the area of biochemistry dedicated to characterizing proteomes. With the goal of better understanding the intricate biochemical machinery that operates within cells, current proteomic methods can identify and quantify up to 10 000 proteins from human tissue samples[1] and have been used to explore disease pathophysiology, therapeutic responses, and the manifestations of genetic variation.

CME at jamacmelookup.com

Clinical Proteomics: A Protein at a Time

Since the 1950s, affinity reagents (eg, antibodies) have been leveraged in immunohistochemistry or immunoassays to quantify or determine the location of 1 or a few proteins in patient samples.[2] More modern methods using mass spectrometry have been developed to quantify clinically relevant biomarkers, which overcome the poor reproducibility and lack of standardization of antibody-based methods.[3] Over the last 2 decades, mass spectrometry–based protein assays have moved from concept to US Food and Drug Administration (FDA)–approved diagnostic tests, including a test for ovarian malignancies that quantifies the plasma concentrations of proteins cancer antigen 125 (CA 125), prealbumin, apolipoprotein AI, β2 microglobulin, and transferrin. Yet, in biomedical research, the term proteomics is more often associated with methods designed to more exhaustively study a proteome.

How It Works

Proteomics generates protein abundance phenotypes for cells, tissues, and organisms and mass spectrometry is the most common technique used to measure many proteins in complex biological samples. In this approach, proteins are digested with proteases into peptides (eg, trypsin). Peptides are then ionized and fragmented to break the peptide bonds between amino acids (**Figure**). This process generates a spectrum of peptide-specific fragment ions of varying mass to charge ratios that can be used to sequence the peptide of interest. Since the peptide and fragment ions are charged, a quantitative measurement of relative protein amounts can be determined by comparing the ion current–based intensity of the representative peptide precursor or fragment ions. Modern instrumentation can collect thousands of spectra per minute and requires high-performance computational infrastructure to process the large amount of data generated. As a result, proteomics can measure the abundance of a single protein or, starting from as little as 25 µg of protein input (~10% of the protein from a core needle biopsy), more than 10 000 proteins from human tissue samples.[1] Modern alternatives for mass spectrometry–based proteomics now exist that use DNA-barcoded affinity reagents (eg, antibodies or aptamers [nucleic acid molecules that recognize proteins]) that can then be read out with DNA microarray or DNA sequencing technologies for protein identification and quantification.[4]

Clinical Proteomics: Many Proteins at Once

Beyond the use of mass spectrometry–based proteomics for diagnostic testing, one of the largest and most successful research-focused campaigns to catalog proteome variation in clinical samples comes from the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC), which has analyzed the proteomes of more than 3000 tumor and normal tissue samples from 19 different cancer types, including ovarian serous cystadenocarcinomas (486 cases) and lung adenocarcinomas (316 cases).[5] By moving beyond the measurement of a single protein, these studies demonstrate how the integrated analysis of thousands of proteins can be leveraged to better understand molecular drivers of disease (eg, how genetic variation in the lung adenocarcinoma–associated kinase kelch-like ECH-associated protein 1 [KEAP1; UniProt: Q9Z2X8] reduces protein abundance with no effect on transcript abundance). This spurred efforts to quantify proteins in formal-fixed, paraffin-embedded tissues (ie, those commonly preserved during surgery) at a rate of 100 samples per day per instrument, suggesting that mass spectrometry–based proteomics assays can now be performed at a scale and speed compatible with phase 3 clinical trials.[6]

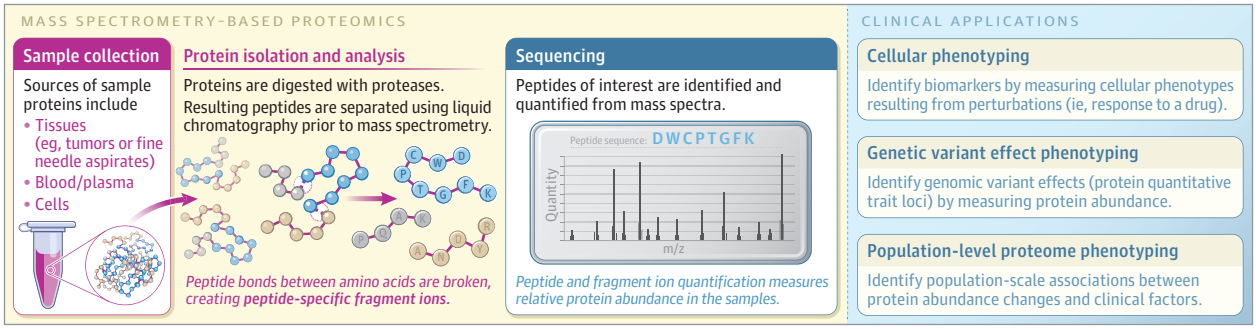
Proteogenomics

Proteomic workflows can be coupled with analysis of the entire spectrum of genes or transcripts in a biological system (genomics) to derive the functional consequences of genetic variation. Early experiments using outbred mouse strains[7] identified protein quantitative trait loci (pQTL), which are genetic variants within a protein-coding gene or in distal genomic regions that affect the levels and types of proteins expressed. For example, genetic variants in a single member (*CCT6A*; UniProt: P80317) of the chaperonin-containing TCP-1 (CCT) complex can alter the protein abundance of all 8 members of the complex. Proteomics methods that are not mass spectrometry based (eg, the antibody- and aptamer-based methods noted above) have recently identified associations between rare protein-coding genomic variants and protein abundance in plasma.[4] One study reported that more than 2000 pQTLs were associated with changes in the plasma concentration of proteins for which protein abundance has been linked to a disease or phenotypic trait, including a pQTL associated with lower abundance of chordin-like protein 2 (*CHRDL2*; UniProt: Q6WN34) and lower risk of colorectal cancer.[4] Another study found that variation in a DNA-modifying methylcytosine dioxygenase (TET2; UniProt: Q6N021), which resides in the nucleus of human cells, was associated with the plasma concentration of the acute myeloid leukemia–associated FMS-related tyrosine kinase 3 (FLT3; UniProt: P36888). In that study, the quantitative assessment of 2923 plasma proteins in 49 736 individuals highlighted the scale achievable with non–mass spectrometry, clinically facing proteomics. However, the correlation of results between alternative non–mass spectrometry platforms was modest.[4]

Clinical Considerations

The hope that proteomics would identify novel, clinically actionable biomarkers of disease, prognosis, and therapeutic efficacy has yet to be widely realized. Most clinical proteomics assays targeting more than a few proteins have been developed and implemented in a single

Figure. Proteomics Methods and Applications



laboratory,[3] require costly instrumentation,[3] and lack the spatial resolution of imaging-based approaches, such as immunohistochemistry. However, creative technological[8,9] and bioinformatic[10] advances have been transformative, leading to the ability of mass spectrometers to be operated by surgeons in the operating room to directly characterize patient tissues in just 1.5 seconds[8] or, coupled with laser-capture microdissection of formalin-fixed paraffin-embedded tissue, to identify that Janus kinase inhibitor treatment can resolve toxic epidermal necrolysis.[9] These advances may lead to more widespread application of proteomics in the armamentarium of clinical diagnostics.

Evidence Base

Several organizations have proposed a role for multiprotein analyte panels in decision-making to perform diagnostic procedures (eg, imaging or biopsy) or refer patients to specialists. However, the evidence to support recommendations for their use was often weak. For example, serum biomarker panels were recommended as an alternative to CA 125 when evaluating referrals to a gynecologic oncologist for adnexal masses that may require surgery (level C). As another example, a blood test that combines 3 biomarkers into a risk-based score for liver fibrosis was recommended for all people incidentally diagnosed with metabolic dysfunction-associated steatotic liver disease (formerly nonalcoholic fatty liver disease [NAFLD]) (level C). Although the US Centers for Medicare & Medicaid Services has adopted Proprietary Laboratory Analyses or Current Procedural Terminology codes to help with reimbursement of proteomics assays, the data needed to recommend their widespread use or strongly recommend their use in particular circumstances are lacking.

Conclusions

The measurement of proteins has been a hallmark of clinical diagnosis and care for decades. With proteomics' ability to analyze thousands of proteins in patient samples at high throughput, this rapidly evolving technology will augment our understanding of and ability to manage a wide range of diseases.

ARTICLE INFORMATION

Author Affiliations: Department of Genome Sciences, University of Washington, Seattle (Schweppe, Beliveau); Brotman Baty Institute for Precision Medicine, Seattle, Washington (Schweppe, Beliveau); Institute for Stem Cell & Regenerative Medicine, University of Washington, Seattle (Schweppe, Beliveau); Department of Laboratory Medicine & Pathology, University of Washington, Seattle (Hoofnagle).

Corresponding Author: Devin K. Schweppe, PhD, University of Washington, 3720 15th Ave NE, William H. Foege Hall, Seattle, WA 98195 (dkschwep@uw.edu).

Published Online: January 30, 2025. doi:10.1001/jama.2024.28089

Conflict of Interest Disclosures: Dr Schweppe reported receiving grants from Pew Charitable Trusts, W.M. Keck Foundation, National Institutes of Health, Andy Hill CARE Fund, and Washington Research Foundation; financial support from AI Proteins; nonfinancial support from Thermo Fisher Scientific, Seer Inc, and Genentech through collaborative research agreements; consulting fees from Matchpoint Therapeutics outside the submitted work; and a patent pending cosubmitted with Dr Beliveau. Dr Beliveau reported receiving grants from National Institute of General Medical Sciences, National Human Genome Research Institute, National Institute of Mental Health, Washington Research Foundation, and W.M. Keck Foundation. Dr Hoofnagle reported receiving grants from National Institute of Diabetes and Digestive and Kidney Diseases, National Institute on Aging, and National Cancer Institute.

Note: Source references are available through embedded hyperlinks in the article text online.

REFERENCES

1. Satpathy S, Jaehnig EJ, Krug K, et al. Microscaled proteogenomic methods for precision oncology. *Nat Commun*. 2020;11(1):532. doi:10.1038/s41467-020-14381-2

2. Coons AH. Fluorescent antibodies as histochemical tools. *Fed Proc*. 1951;10(2):558-559.

3. Phipps WS, Kilgore MR, Kennedy JJ, et al. Clinical proteomics for solid organ tissues. *Mol Cell Proteomics*. 2023;22(11):100648. doi:10.1016/j.mcpro.2023.100648

4. Eldjarn GH, Ferkingstad E, Lund SH, et al. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature*. 2023;622(7982):348-358. doi:10.1038/s41586-023-06563-x

5. Li Y, Dou Y, Da Veiga Leprevost F, et al; Clinical Proteomic Tumor Analysis Consortium. Proteogenomic data and resources for pan-cancer analysis. *Cancer Cell*. 2023;41(8):1397-1406. doi:10.1016/j.ccell.2023.06.009

6. Tüshaus J, Eckert S, Schliemann M, et al. Towards routine proteome profiling of FFPE tissue. EMBO J. November 18, 2024. Accessed January 2, 2025. https://www.embopress.org/doi/full/10.1038/s44318-024-00289-w

7. Chick JM, Munger SC, Simecek P, et al. Defining the consequences of genetic variation on a proteome-wide scale. *Nature*. 2016;534(7608):500-505. doi:10.1038/nature18270

8. Zhang J, Sans M, DeHoog RJ, et al. Clinical translation and evaluation of a handheld and biocompatible mass spectrometry probe for surgical use. *Clin Chem*. 2021;67(9):1271-1280. doi:10.1093/clinchem/hvab098

9. Nordmann TM, Anderton H, Hasegawa A, et al. Spatial proteomics identifies JAKi as treatment for a lethal skin disease. *Nature*. 2024;635(8040):1001-1009. doi:10.1038/s41586-024-08061-0

10. Osipov A, Nikolic O, Gertych A, et al. The Molecular Twin artificial-intelligence platform integrates multi-omic data to predict outcomes for pancreatic adenocarcinoma patients. *Nat Cancer*. 2024;5(2):299-314. doi:10.1038/s43018-023-00697-7