Digging Deeper into Cancer with Proteogenomics

A growing collection of techniques for combining genomic and proteomic data reveals ever more about the molecular biology of cancer, and that knowledge will lead to even more advanced treatments

by Mike May

Data drives many modern advances in understanding and treating cancer. In particular, proteogenomics—the combination of data from the genomes and proteomes—promises ways to better understand the pathways that underlie cancer and how to treat them.[1] "Integrated proteogenomics has shown unequivocally that proteomics adds value over a genome-only approach to understanding tumor biology," says Amanda Paulovich, MD, PhD, professor and Aven Foundation endowed chair at the Fred Hutchinson Cancer Center in Seattle.

Amanda Paulovich, MD, PhD, professor and Aven Foundation endowed chair, Fred Hutchinson Cancer Center

More than the sequences of DNA and RNA are required to explore cancer more deeply. "Genomic sequences and copy-number analyses, as well as RNASeq data, are not reliable indicators of protein expression levels and protein activities or post-translational modifications," Paulovich explains. "Many post-transcriptional processes impact the proteome."

As Paulovich notes, an article by Henry Rodriquez, PhD—founding director of the office of cancer clinical proteomics research at the U.S. National Cancer Institute—and his colleagues depicts the transition in precision oncology from a genome-centric approach to the use of proteogenomics.[2]

The activity of proteins, as well as their post-translational modifications, really come into play in therapies. "Since most modern therapies target proteins, not nucleic acids, it is imperative that we be able to monitor proteins directly, rather than make unreliable inferences from nucleic-acid profiles," Paulovich says. "Fortunately, tremendous advances in mass spectrometry-based proteomics, both untargeted and targeted, over the past decade have enabled analytically robust analysis of the human proteome, in some cases in clinical settings."[3]

Advances in technology also make it easier to study more cancer patients when needed. "The ability to conduct large-scale studies, both in terms of sample throughput and target content, is enabling researchers to describe phenotypes or signatures of a particular disease state in a much greater level of detail than was previously possible and in much larger populations or cohorts for greater statistical power," says Fiona Kaper, PhD, vice president, advanced science assay research at Illumina in San Diego. "This is due to: the availability of large sample collections, such as population-scale biobanks; new technologies, such as the ones offered by Olink and SomaLogic, that combine high target plexity proteomics with high-throughput readouts, such as next-generation sequencing or microarrays; and the favorable economics of available technologies to fund studies at scale." As Kaper adds,

Fiona Kaper, PhD, vice president, advanced science assay research, Illumina

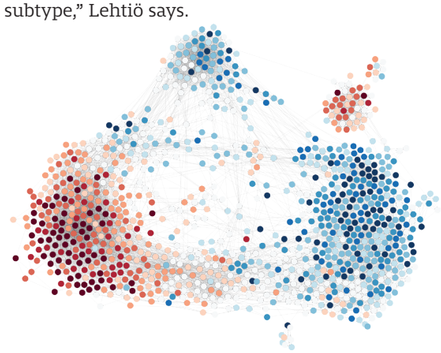
"Computer power and analytical tools and methods that can handle very large data sets and integrate different 'omic data types into interpretable results are key."

Overcoming poor outcomes

Most, maybe all, scientists who apply proteogenomics to cancer research agree that new technology is crucial to digging deeper into this disease. "Mass spectrometry-based proteomics has gotten better so we now work with smaller amounts of materials," says Janne Lehtiö, PhD, professor of medical proteomics at the Karolinska Institute in Stockholm. "The clinical application of proteogenomics is facing a major leap forward because of the technology developments."

Using a form of high-resolution liquid chromatography (LC)-mass spectrometry (MS), Lehtiö and his colleagues analyzed the proteomes of people with chronic lymphocytic leukemia (CLL).[4] Then, the scientists combined proteomic data with existing genomic, transcriptomic, and drug-perturbation data on these patients. From this work, Lehtiö says that they "found a new chronic leukemia subtype with a poor prognosis" for patients, and the researchers validated that subtype with an additional cohort. The new CLL subtype emerged from proteomic data. "If you just looked at genomics, you wouldn't be able to determine this subtype," Lehtiö says.

Janne Lehtiö, PhD, professor of medical proteomics, Karolinska Institute



This example of protein-correlation network analysis from a dataset generated with high-resolution isoelectric focusing liquid chromatography-mass spectrometry on a sample from a patient with chronic lymphocytic leukemia (CLL) shows protein abundance, which is color-coded and varies between patients with and without CLL. Credit: Janne Lehtiö

Beyond finding new molecular signatures of cancer, Lehtiö says, you can use a "smaller cohort for interesting results by moving to proteogenomics, since by layering the data you can get sample-specific, genotype-phenotype analysis." So, existing proteogenomic tools allow scientists to work with both smaller and larger cohorts.

This approach will work with a wide range of cancers. As another example, Lehtiö and his colleagues applied proteogenomics to non-small cell lung cancer (NSCLC).[5] This work revealed six proteome subtypes with interesting connections to oncogenic drivers, outcomes, as well as aberrant proteins, so called cancer neoantigens, that can be used for development of cancer vaccines. "We determined the tumor mutation burden on the DNA level, but also used proteomics data to study more complex aberrant proteins, caused by genomic aberrations the tumor harbor," Lehtiö explains. "To develop rational, targeted-therapy combinations and connect immunotherapies to these, we need proteogenomics as biomarker analysis to understand both targetable cancer-driving pathways as well as immune-evasion mechanisms at once."

Predicting treatment responses in ovarian cancer

Many recent research projects explore ways to predict the impact of cancer treatments. As one example, Paulovich and her colleagues studied resistance to platinum-based chemotherapy in women with high grade serous ovarian cancers (HGSOCs).[6,7]

Using LC-MS/MS to quantify proteins and machine-learning algorithms to analyze the data, Paulovich says that they "identified an ensemble prediction model of chemo-refractoriness based on 64 proteins, and it detects a subset of chemotherapy refractory tumors with very high specificity and is validated in two independent patient cohorts." In addition, the scientists identified five novel subtypes of HGSOC based on protein-pathway expression, which might suggest "different mechanisms of refractoriness and implicate potential subtype-specific treatment approaches, including immune therapies or metabolic inhibitors," Paulovich says.

The scientists also took other approaches to supporting the idea that protein-pathway expression in the five subtypes might predict therapeutic vulnerabilities. For example, the team studied patient-derived xenografts from patients who presented with chemo-refractory HGSOC. The researchers found that the effect of platinum-based chemotherapy could be improved with pharmacological inhibition or CRISPR knock out of a gene connected to fatty-acid oxidation.

The results from these studies show just some of the power of applying proteogenomics in oncology. "Despite over three decades of research on platinum responses in cancer, no predictive biomarker has been translated into clinical use," Paulovich says. "Predictors of refractory disease could spare these patients the unnecessary toxicity of a platinum-based regimen and provide a means to triage these patients in clinical trials to identify effective therapies for refractory disease."

points out that this platform can be used to "translate biological pathways to a list of candidate protein biomarkers, which enables improved experiment planning and setup, post-run data analysis, and analytical tools."

Olink's technology is already being applied in many ways. As one example, Lawley says that it "is enabling projects like the UK Biobank Pharma Proteomics Project where 13 pharma partners have come together to drive proteomics on around 60,000 UK Biobank samples."

Other companies also strive to expand datasets that can help scientists understand and treat cancer and other diseases. In many ways, that objective depends on enabling the readout of non-DNA based modalities, such as proteomes, on high-throughput analytical platforms, such as NGS. As an example, Kaper says, "Illumina's NGS technology roadmap has continually increased the data quality and output per sequencing run, driving down the cost per data point, while at the same time increasing the speed of data generation." Consequently, she says, "NGS therefore provides the fastest, highest throughput, most flexible, and most economical readout technology for different multi-omic modalities, including genomes, methylomes, transcriptomes, and proteomes."

Making the most of data analysis, however, often depends on collections of technology platforms. For instance, "combining Illumina's NGS readout with large target–content panels, such as SomaLogic's SomaScan platform or Olink's Explore platform, will enable researchers to analyze thousands of protein targets in hundreds of samples simultaneously," Kaper notes. As a result, scientists can explore protein abundance in more depth and at scale. Kaper says that such a capability "will drive wider adoption and incorporation of proteomics in multi-omic studies, deriving ever increasing value and understanding from each sample."

Improving precision

The increase in proteogenomic-driven knowledge about cancer's development should enhance treatment options—especially creating more therapies for specific patients. "Proteogenomics is increasing our understanding of cancer biology, but the ultimate goal is to use that knowledge to improve cancer outcomes, especially through personalized/precision oncology," Paulovich says. "For personalized/precision oncology to succeed, we need predictive biomarkers to match patients with efficacious therapies." Single genes or proteins are not enough to understand the complexity of drug responses. Instead, collections of information must be applied to cancer. As Paulovich says: "An unanswered question is: Can multi-analyte proteogenomic predictive biomarker signatures be translated into clinical labs and change clinical practice to improve outcomes while reducing healthcare costs?"

Making the most of proteomics, though, depends on expertise in various fields. "One of the challenges of proteogenomics in general is that you need a team of people—an expert in proteomics, an expert in genomics, and an expert in statistics—to really be able to figure out how these puzzle pieces fit together," Donovan explains. In some cases, a software solution could help. For example, an expert in proteomics could use software to bring in genomics when analyzing data.

Such software also helps deal with the volumes of data produced in proteogenomics. Just as an example, a genomics study on a large cohort—say, 10,000 people—could produce a table of 10,000 rows (patients) and 20,000 columns (genes), which makes 200 million data cells. Add the estimated more than one million proteoforms and that creates what Donovan describes as "a very high dimensional problem." Plus, that problem is destined to expand. "There are other 'omic' data types beyond genomics and proteomics," she says. "Epigenomics tells us about the regulation of the genome and then post-translational modifications of proteins tells us more about a protein's function." As Donovan adds: "There can be a lot of growth from putting together all these different molecular phenotypes, because that's systems biology."

Tomorrow's cancer studies will explore even more broadly based versions of 'omics, but that must await advances in technology. As Kaper says, "there are currently no technologies available that can convert additional modalities, such as metabolomics, into a format that is compatible with a DNA-based readout, such as NGS." As a result, she says, "these modalities therefore cannot—yet—benefit from a high-throughput technology that can interrogate many samples and analytes in parallel, limiting their inclusion into large-scale, multi-omic studies."

The speed of development in proteogenomic tools and techniques, as well as the ongoing advances in other areas of 'omics, promise to soon make it possible to explore cancer in even more ways. With that capability, scientists will learn much more about the molecular biology of cancer and find ways to treat it in more precise and personalized ways. ■

References

1. Mani, D.R., Krug, K., Zhang, B., et al. Cancer proteogenomics: current impact and future prospects. Nature Reviews Cancer 22(5):298–313 (2022).
2. Rodriguez, H., Zenklusen, J.C., Staudt, L.M., et al. The next horizon in precision oncology: Proteogenomics to inform cancer diagnosis and treatment. Cell 184(7):1661–1670. (2021).
3. Zhang, B., Whiteaker, J.R., Hoofnaagle, A.N., et al. Clinical potential of mass spectrometry-based proteogenomics. Nature Reviews Clinical Oncology 16(4):256–268. (2019).
4. Herbst, S.A., Verterlund, M., Helmbolt, A.J., et al. Proteogenomics refines the molecular classification of chronic lymphocytic leukemia. Nature Communications 13:6226 (2022).
5. Lehtiö, J., Arslan, T., Siavelis, I., et al. Proteogenomics of non-small cell lung cancer reveals molecular subtypes associated with specific therapeutic targets and immune-evasion mechanisms. Nature Cancer 2:1224–1242 (2021).
6. Huang, D., Savage, S.R., Calinawan, A.P., et al. A highly annotated database of genes associated with platinum resistance in cancer. Oncogene 40(46):6395–6405 (2021).
7. Huang, D., Chowdhury, S., Wang, H.I., et al. Multiomic analysis identifies CPT1A as a potential therapeutic target in platinum-refractory, high-grade serous ovarian cancer. Cell Reports Medicine 2(12):100471 (2021).
8. Donovan, M.K.R., Huang, Y., Blume, J.E., et al. Peptide-centric analyses of human plasma enable increased resolution of biological insights into non-small cell lung cancer relative to protein-centric analysis. BioRxiv (2022). https://doi.org/10.1101/2022.01.07.475393.

Mike May, is a freelance writer and editor with more than 30 years of experience. He earned an MS in biological engineering from the University of Connecticut and a PhD in neurobiology and behavior from Cornell University. He worked as an associate editor at *American Scientist*, and he is the author of more than 1,000 articles for clients that include *GEN*, *Nature*, *Science*, *Scientific American*, and many others. In addition, he served as the editorial director of many publications, including several *Nature Outlooks and Scientific American Worldview.*