Revolutionizing Precision Oncology through Collaborative Proteogenomics and Data Sharing

Henry Rodriguez[1,*] and **Stephen R. Pennington**[2]

[1]Office of Cancer Clinical Proteomics Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA [2]School of Medicine, Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield Dublin 4, Ireland

Abstract

The integration of proteomics into precision oncology presents opportunities that may transform the molecular analysis of cancer and accelerate basic and clinical cancer research. This Commentary discusses the importance of international collaboration and data sharing inspired by the Cancer Moonshot to accelerate the progress of multi-omic precision medicine–an approach that addresses the global diversity of people and of cancers.

Introduction

There is arguably no diagnosis a patient can receive from their doctor more frightening than "you have cancer." Cancer is the leading cause of death worldwide. An estimated 14.1 million new cancer cases and 8.2 million cancer deaths occurred in 2012 worldwide. These numbers are expected to increase due to a growing aging population, and adoption of behavioral and lifestyle factors known to contribute to cancer development. In the United States alone, approximately 38.5 percent of men and women will be diagnosed with cancer at some point during their lifetime, based on 2012–2014 data (SEER Data Base, https:// seer.cancer.gov).

Precision medicine in oncology aims to improve the diagnosis and treatment of cancer. Although genomics has shaped what precision medicine presently entails, science is approaching another inflection point. It is becoming increasingly clear that molecular phenotype measurements, particularly the constantly changing repertoire of proteins that make up an individual, are essential to our understanding of tumor biology and our ability to bring precision medicine to patients. Proteomics—the genome-wide measurement of proteins—and proteogenomics—a multi-omic approach of converging genomics, transcriptomics, and proteomics—are poised to transform oncology science and medicine by giving researchers and doctors new tools and data to more precisely target a patient's tumor based on their molecular profile (proteogenomics video: https://youtu.be/7qpKPiOZrks).

*Correspondence: henry.rodriguez@nih.gov.

This Commentary details how the successes of NCI-driven collaborative consortiums in cancer genomics and proteogenomics brought us to this turning point in precision medicine.

Evolution of Precision Medicine

For patients today, a cancer diagnosis typically begins with a clinical examination, followed by a blood test and medical imaging, and often ends with a histopathological analysis of tissue from a tumor biopsy. This screening and testing enables physicians to establish the cancer type and develop a therapeutic strategy with the patient.

Precision medicine is a concept recently promoted as a new paradigm for healthcare delivery and for selecting appropriate pharmaceutical interventions. It represents an unparalleled value proposition and is central to transitioning healthcare from a 20th century "one size fits all" approach to a 21st century personalized and precise strategy based on molecular characteristics. The underlying motivation is the belief that knowledge about a patient's tumor biology can help determine what treatment will work best for that patient. In addition to achieving more durable treatment responses, this approach may also help patients avoid undue costs, lost time, stress, and side effects associated with ineffective therapies. New omics-driven data may also guide the use of combination therapies and/or the development of new forms of treatment/pharmaceuticals. This is good news for patients because it heralds the opportunity to improve the patients' quality of life, and ultimately, find cures for cancer and other life-threatening diseases.

Currently, genomic approaches constitute most of precision-based medicine screening and testing, in large part due to the successes of genomic data in guiding diagnosis, informing prognosis, and supporting treatment decisions in a variety of cancers. Examples include the pioneering work that led to the development of the drug imatinib (Gleevec), designed to inhibit an altered enzyme (BCR-ABL) produced by the fusion of two genes in chronic myelogenous leukemia (Druker et al., 2006). Another example is the breast cancer drug trastuzumab (Herceptin), which works only for women whose tumors have a HER-2 positive genetic profile (Pegram and Slamon, 2000). Studies have also found lung cancer patients whose tumors have specific EGFR mutations respond to the drugs gefitinib (Iressa) and erlotinib (Tarceva), which target these mutations (Blackhall et.al., 2006). These therapies have not only transformed the lives of many patients, but also provide a powerful validation of the precision oncology approach.

Motivated by these early successes, it seemed that genomic-based approaches (for simplification in this Commentary, genome will encompass DNA and RNA) might be a key to delivering the promise of precision oncology. However, achieving that vision has proven somewhat elusive given our limited ability to wholly connect genotypes (genomic data) to disease properties or phenotypes.

The reason is that genetic mutations do not always result in the predicted change of the corresponding protein, and there are many other factors that contribute to tumor behavior, such as protein modifications, metabolism, and the microbiome. In addition, most tumors have many mutations, making it difficult to establish which are important drivers (and thus

potentially targetable or "druggable") and which are just "along for the ride." Even when cancer-driving mutation(s) are known, there may not be a drug that targets the particular gene(s). And sometimes, even if such drugs exist, it doesn't work to reduce the patients' tumors—sometimes because the cancer develops routes to bypass the impact of the targeted treatment, often via new mutations.

Integrating genomic and proteomic data through proteogenomics approaches may be one missing link to help make precision oncology more effective.

Genomics: A First Step Toward Precision Oncology

In February 2001, the Human Genome Project (HGP) published the results of its progress: a 90 percent complete sequence of the three billion base pairs in the human genome. This 15-year effort traces its ideological origins back to the mid-1980s, where a large part of the early work was devoted to developing improved sequencing technologies, at a cost near three billion U.S. dollars (U.S. National Research Council, https://www.nap.edu/catalog/1097/mapping-and-sequencing-the-human-genome).

In 2005, the National Cancer Advisory Board provided NCI with a framework for a collaborative program that utilized recent technological advances for detecting molecular aberrations (DNA, RNA, and proteins) in cancer patients to drive cancer research forward. Consequently, NCI launched The Cancer Genome Atlas (TCGA; https://cancergenome.nih.gov) in 2006 to understand the molecular basis of cancer. TCGA, a collaboration between NCI and the National Human Genome Research Institute, was initiated as a 3-year pilot project to determine the feasibility of cataloging the genomic alterations of three tumor types (glioblastoma multiforme, serous cystadenocarcinoma of the ovary, and squamous carcinoma of the lung). Achievement of the pilot project goals set the stage for the next phase of TCGA—an expansion in 2011 to 20 additional tumor types beyond those studied in the pilot. To date, TCGA has cataloged the genomes of 33 different tumor types including 10 rare cancers, based on paired tumor and normal tissue sets collected from 11,000 individuals.

TCGA was a gamble, a high-risk endeavor that perhaps only NCI was suited to take on. The hope was that a comprehensive understanding of the role of genomics in cancer biology would lead to improvements in how cancer is prevented, diagnosed, and treated. These hopes were met and exceeded, and we still have much to learn from the data produced by TCGA. In effect, TCGA jump-started the movement toward precision medicine in oncology and made this approach a real possibility for transforming cancer care. What's more, the collaborative model of TCGA united the cancer research world and demonstrated that global collaboration has the potential to greatly accelerate progress.

There is little doubt that TCGA has contributed significantly to the identification of markers for cancer prevention and diagnosis and novel targets for therapeutic drug development. It has also provided the foundation for a refined clinical understanding of patient stratification for therapy. Unequivocally, cancer researchers have made significant progress in identifying a new molecular taxonomy of cancer through genomics applications.

However, while genomics has formed the basis for current precision medicine approaches, genomics alone cannot explain all of the intricate connections between tumor biology and patient outcomes. Today, precision medicine is set to undergo a revolution—one that converges genomics with proteomics.

Moving Beyond the Genome

Proteomics has, in many quarters, been considered the next logical step in expanding our understanding of tumor biology because it provides information that complements genomic and transcriptomic data. In fact, at the same time that NCI was developing TCGA, the research community asked whether a similar atlas could be prepared for proteins encoded by genes in the cancer genome atlas. At the time, however, the National Academies determined that technologies used to detect protein changes were not as mature as their genome-based counterparts (U.S. National Research Council, https://www.nap.edu/catalog/10560/defining-the-mandate-of-proteomics-in-the-post-genomics-era).

NCI recognized that reliable proteomic technologies would be necessary to advance the molecular understanding of cancer and further the goals of precision medicine. With the abilities to drive scientific progress in focused areas, unite research groups with complementary skills, and take on high risk-high reward projects, NCI was uniquely suited to address this issue. Consequently, they launched the Clinical Proteomic Tumor Analysis Consortium (CPTAC; https://proteomics.cancer.gov) in 2006 to address rigor and reproducibility in proteomic technologies. CPTAC is a U.S. effort to accelerate the understanding of the molecular basis of cancer through the application of large-scale proteome and genome analysis. This collaborative consortium consists of institutions and investigators with expertise in proteomics, genomics, cancer biology, oncology, bio-informatics, and clinical chemistry.

The first goal of CPTAC was to standardize proteomics methods so that data produced by different centers were reproducible and could be combined into a coherent dataset. This included standardization of comprehensive untargeted protein analyses by mass spectrometry (Tabb et al., 2010), standardization and best practices of targeted protein analyses by mass spectrometry (multiple reaction monitoring—MRM) (Addona et al., 2009; Kennedy et al., 2014; Carr et al., 2014; Whiteaker et al., 2014), adoption of a MRM assay (for thyroglobulin) by clinical reference laboratories, development of an open-source computational tool (Skyline) for designing MRM assays which is supported by instrument vendors (MacLean et al., 2010), development of Food and Drug Administration (FDA) mock 510(k) device clearance documents in targeted proteomic platforms done in coordination with the American Association for Clinical Chemistry (Regnier et al., 2010), and development of community-based proteomic data sharing policies (Kinsinger et al., 2011).

An Integrated Approach to Precision Medicine

To begin to incorporate these new proteomics technologies into precision oncology research, the second phase of CPTAC was initiated in 2011. The purposes of this pilot study were to comprehensively identify proteins derived from altered genes and related biological

processes, and to determine if this additional layer of molecular information would help us understand the molecular basis of cancers in ways that are not possible through genomics data alone. All data and resources (assays and antibodies) produced by CPTAC were to be made broadly available to the research community through public databases, to maximize utility and public benefit.

Because genomic and clinical datasets from TCGA are shared with the scientific community, CPTAC scientists were able to integrate new proteomic data with existing genomic data. Specifically, the investigators applied CPTAC's standardized proteomic workflows to three genomically-characterized tumor types from TCGA (colorectal, breast, and ovarian tumors).

These pilot studies demonstrated the ability of proteogenomics to reveal new insights into cancer biology and possibly therapeutic interventions for patients, while at the same time creating public resources that are widely used by the global cancer community (Zhang et al., 2014; Mertins et al., 2016; Zhang et al., 2016). One of the major insights gained from these studies is that genomic changes are not always present at the protein level. For example, although copy number alterations might show strong cis- and trans-effects on mRNA abundance, it does not always extend to the protein level. Thus, integrating genomic and proteomic data together can provide more information and insight into cancer development and growth.

Proteogenomic approaches enable a more complete characterization of the biological pathways associated with tumor development and metastasis, and have the potential to better match a patient's individual tumor to targeted therapies. A major challenge to cancer treatment is the ability to predict drug response/toxicity and the relatively rapid acquisition of resistance to such treatments. While there has been significant progress in stratifying patients according to their likelihood of responding to targeted therapies, the development of resistance/toxicity significantly limits their utility. The ability to predict drug response/ toxicity associated with targeted therapeutics is highly critical to the success of precision medicine, because it holds the promise to improve a patient's quality of life, and ultimately to find cures for cancer and other life-threatening diseases. Proteogenomic data provides this important information.

The concept of proteogenomics was in practice before the term was coined by Jaffe and colleagues (Jaffe et al., 2004) in 2004. However, it was not until the CPTAC pilot studies were carried out that the approach really began to gain traction. The integrative proteogenomic approach allowed CPTAC researchers to better connect cancer genotypes and phenotypes for the first time. The pilot project demonstrated the importance of analyzing different and complementary layers of molecular information within human tumors, thus providing the foundation for future proteogenomic studies.

An International Investment in Proteogenomics

The successes of CPTAC led NCI to strengthen its commitment to proteogenomic research. Inspired by the spirit of international collaboration and data sharing encouraged by the

Cancer Moonshot (a U.S. initiative to accelerate cancer research and to make more therapies available to more patients, while also improving our ability to prevent cancer and detect it at an early stage), NCI initiated the International Cancer Proteogenome Consortium (ICPC; https://icpc.cancer.gov) in 2016 (Figure 1). Based on the highly successful model of CPTAC, ICPC encourages international cooperation and investments in proteogenomic cancer research. Aligned with CPTAC, ICPC brings together some of the world's leading cancer and proteogenomic research centers that adopt standard operating procedures as appropriate for biospecimen collection and harmonized proteogenomic technologies/workflows to characterize commonly diagnosed cancers in their respective populations. The collected information will represent the global diversity—"a proteogenome atlas"—of people and commonly diagnosed cancers in their unique populations, hastening progress for patients.

ICPC was built on the initial success of a Memorandum of Understanding (MOU) with institutions in Australia. Soon after the first MOU was signed, six additional MOUs involving institutions spanning seven countries (Canada/Germany, China, Japan, South Korea, Switzerland, and Taiwan [Republic of China]) joined this global vision. Today, ICPC spans 12 counties, encompassing multiple institutions that have pledged to share their data to speed progress against common and rare cancers in diverse populations around the world.

To encourage additional institutions to join this global partnership, the Human Proteome Organization (HUPO; https://www.hupo.org, a leading voice on promoting the science of proteomics) held a Global Leadership Gala on September 16, 2017, in Dublin, Ireland with the theme, "International Cooperation in the Fight Against Cancer." The event, organized by the authors, in coordination with ICPC representatives—many of whom are members and leaders in HUPO—brought together leaders in the proteomic and proteogenomic research fields in the public and private sector to encourage greater cooperation and collaboration to accelerate the use of proteogenomics to fight cancer around the world. Videos from the HUPO 2017 Global Leadership Gala (http://ow.ly/eCpb30gg55M), have been made public to accelerate and raise awareness for cancer research (Figure 2).

Advancements in science and health-care are made possible through widespread access to results from cutting-edge research, enabling scientists to use and build on this knowledge. This principle was clearly demonstrated in the HGP, TCGA, and CPTAC, where researchers built upon the work of others to create an armamentarium of data resources for the community. These collective resources have paid dividends beyond what anyone could have conceived. Just as the HGP, TCGA, and CPTAC recognized the power of having community resources of high-quality data, a parallel opportunity exists today with ICPC. Each ICPC team agreed to make their data (genomic, proteomic, and imaging) available to the public through a "Data Sharing Pledge," accessible on NCI portals. A similar effort inspired by CPTAC and the Cancer Moonshot, is the Applied Proteogenomics OrganizationaL Learning and Outcomes network (APOLLO; https://apollo.cancer.gov), which is a U.S. effort between the NCI, Department of Defense, and Department of Veterans Affairs. Sharing Data (open data ecosystems) among researchers to generate new hypotheses and study cancer on large population scales, representative of the diversity of people around the world, is anticipated to advance our understanding of cancer biology and its translation to patient care, and is being met with enthusiasm, not resistance. In the spirit of the Cancer Moonshot, these partnerships

will continue to break down silos and allow scientists to share information with the long-term goal of ending cancer as we know it.

Concluding Remarks

Cancer is without doubt one of the biggest public health challenges facing humanity. Defining the future landscape of cancer research and care is challenging, in part, because cancer biology is very complicated, even more so when connecting molecular data to patient treatment and care. As researchers and physicians, we need to learn as much as we can about the entire biological narrative before we make life-changing decisions for and with our patients. From researchers to frontline medical practitioners, there is an eagerness to re-tool cancer care for tomorrow's challenges.

Precision medicine is a team sport: it is the integration of several layers of molecular information with clinical and family histories to better determine a person's predisposition to disease and to provide a more coherent and individualized approach to health care. Our understanding of cancer biology and our efforts to prevent and cure cancer with precision medicine can benefit from an integrated multi-omic (proteogenomic) approach. The potential to fully integrate DNA, RNA, and protein data via proteogenomics approaches can and will open new opportunities to better elucidate the interconnection of cancer signaling pathways and to establish a more accurate design, prediction and monitoring of therapeutic interventions for patients.

International collaborations such as TCGA, CPTAC, and ICGC have united experts across the cancer research world with a common goal. By working together to streamline efforts, troubleshoot issues, standardize protocols, and share progress, these collaborative initiatives are uniquely suited to address the challenges of precision medicine.

Let's imagine a future involving multi-omic precision medicine, bigger international collaborations, and wider data access. We have the audacious intent of shaping that future and we welcome others to join this and similar efforts and accelerate our fight against cancer.

References

Addona TA, Abbatiello SE, Schilling B, Skates SJ, Mani DR, Bunk DM, Spiegelman CH, Zimmerman LJ, Ham AJ, Keshishian H, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nat Biotechnol. 2009; 27:633–641. [PubMed: 19561596]

Blackhall F, Ranson M, Thatcher N. Where next for gefitinib in patients with lung cancer? Lancet Oncol. 2006; 7:499–507. [PubMed: 16750500]

Carr SA, Abbatiello SE, Ackermann BL, Borchers C, Domon B, Deutsch EW, Grant RP, Hoofnagle AN, Hüttenhain R, Koomen JM, et al. Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. Mol Cell Proteomics. 2014; 13:907–917. [PubMed: 24443746]

Druker BJ, Guilhot F, O'Brien SG, Gathmann I, Kantarjian H, Gattermann N, Deininger MW, Silver RT, Goldman JM, Stone RM, et al. IRIS Investigators. Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. N Engl J Med. 2006; 355:2408–2417. [PubMed: 17151364]

Jaffe JD, Berg HC, Church GM. Proteogenomic mapping as a complementary method to perform genome annotation. Proteomics. 2004; 4:59–77. [PubMed: 14730672]

Kennedy JJ, Abbatiello SE, Kim K, Yan P, Whiteaker JR, Lin C, Kim JS, Zhang Y, Wang X, Ivey RG, et al. Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins. Nat Methods. 2014; 11:149–155. [PubMed: 24317253]

Kinsinger C, Apffel J, Baker M, Bian X, Borchers CH, Bradshaw R, Brusniak Mi-Y, Chan DW, Deutsch E, Domon B, et al. Recommendations for Mass Spectrometry Data Quality Metrics for Open Access Data (Corollary to the Amsterdam Principles). Mol Cell Proteomics. 2011; 10(12):O111.015446.

MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010; 26:966–968. [PubMed: 20147306]

Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, Wang X, Qiao JW, Cao S, Petralia F, et al. NCI CPTAC. Proteogenomics connects somatic mutations to signalling in breast cancer. Nature. 2016; 534:55–62. [PubMed: 27251275]

Pegram M, Slamon D. Biological rationale for HER2/neu (c-erbB2) as a target for monoclonal antibody therapy. Semin Oncol. 2000; 27(5, Suppl 9):13–19.

Regnier FE, Skates SJ, Mesri M, Rodriguez H, Težak Z, Kondratovich MV, Alterman MA, Levin JD, Roscoe D, Reilly E, et al. Protein-based multiplex assays: mock pre-submissions to the US Food and Drug Administration. Clin Chem. 2010; 56:165–171. [PubMed: 20007858]

Tabb DL, Vega-Montoto L, Rudnick PA, Variyath AM, Ham AJL, Bunk DM, Kilpatrick LE, Billheimer DD, Blackman RK, Cardasis HL, et al. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. J Proteome Res. 2010; 9:761–776. [PubMed: 19921851]

Whiteaker JR, Halusa GN, Hoofnagle AN, Sharma V, MacLean B, Yan P, Wrobel JA, Kennedy J, Mani DR, Zimmerman LJ, et al. Clinical Proteomic Tumor Analysis Consortium (CPTAC). CPTAC Assay Portal: a repository of targeted proteomic assays. Nat Methods. 2014; 11:703–704. [PubMed: 24972168]

Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, Chambers MC, Zimmerman LJ, Shaddox KF, Kim S, et al. NCI CPTAC. Proteogenomic characterization of human colon and rectal cancer. Nature. 2014; 513:382–387. [PubMed: 25043054]

Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, Zhou JY, Petyuk VA, Chen L, Ray D, et al. CPTAC Investigators. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. Cell. 2016; 166:755–765. [PubMed: 27372738]

Figure 1. A Global Effort in Proteogenomics

Collaboration among nations and institutions to capture and share integrated multi-omic proteogenomics data, which represents the global diversity of people and commonly diagnosed cancers in their unique populations.

Figure 2. International Cooperation in the Fight Against Cancer
Group photo of several attendees at the HUPO 2017 Global Leadership Gala with former
Vice President Joseph R. Biden, Jr. Venue: Royal College of Physicians of Ireland