COMMENTARY

Decoding the Pan-Cancer regulatory landscape: Insights from proteogenomics and post-translational modifications

Wenjun Deng[1,*] **and Shanye Yin**[2,3,*]

[1]Clinical Proteomics Research Center, Massachusetts General Hospital, Harvard Medical School, Boston 02155, USA

[2]Department of Pathology, Albert Einstein College of Medicine, Bronx 10461, USA

[3]Einstein Pathology Single-cell & Bioinformatics laboratory, Bronx 10461, USA

*Correspondence: wdeng@mgh.harvard.edu (W.D.); shanye.yin@einsteinmed.edu (S.Y.)

Received: August 22, 2023; Accepted: September 6, 2023; Published Online: October 30, 2023; https://doi.org/10.59717/j.xinn-life.2023.100032

Citation: Deng W. and Yin S. (2023). Decoding the Pan-Cancer regulatory landscape: Insights from proteogenomics and post-translational modifications. The Innovation Life **1(3)**, 100032.

Genomic studies have provided valuable insights into the diverse genetic alterations driving the development and progression of various cancers.[1,2] However, the intricate interplay between genomics and proteomics, particularly post-translational modifications (PTMs), remains largely unexplored. In a remarkable feat of scientific collaboration and data integration, two concurrent studies published on August 14th in *Cell* by the Clinical Proteomic Tumor Analysis Consortium (CPTAC) have reported the proteogenomics[3] and PTM landscape[4] of over a thousand tumors across ten cancer types (Figure 1). These findings underscore the potential of precision medicine driven by proteogenomic research, illustrating shared and distinct regulatory patterns across cancer types and identifying potential new therapeutic approaches.[5]

PAN-CANCER PROTEOGENOMIC LANDSCAPE

How genomic alterations impact protein expression and PTMs across different cancer types remains largely unknown. Mapping these changes will allow us to uncover commonalities and differences in the proteogenomic landscape, shedding light on potential oncogenic drivers and therapeutic vulnerabilities. Proteomic profiling of tumor cells also has the potential to uncover molecular insights that might be otherwise missed by genomics- and transcriptomics-driven cancer research. In the first study,[3] titled "Pan-cancer proteogenomics connects oncogenic drivers to functional states", the researchers examined proteogenomic data from 1,064 cases across ten cancer types, encompassing breast carcinoma (BRCA), colorectal adenocarcinoma (COAD), clear cell renal cell carcinoma (ccRCC), glioblastoma (GBM), head and neck squamous cell carcinoma (HNSCC), lung squamous cell carcinoma (LSCC), lung adenocarcinoma (LUAD), ovarian high-grade serous carcinoma (HGSC), pancreatic ductal adenocarcinoma (PDAC), and uterine corpus endometrial carcinoma (UCEC). The study systematically analyzed genetic alterations, DNA methylation, transcriptome, global proteomics, and phospho-proteomics, and correlated them with clinical data, pathological histology, and treatment outcomes to create multi-omics molecular profiles. To reveal specific molecular pathways in different cancer types, the researchers conducted clustering analysis of tumor samples, resulting in four distinct clusters, and further explored the representative pathway changes within these clusters.

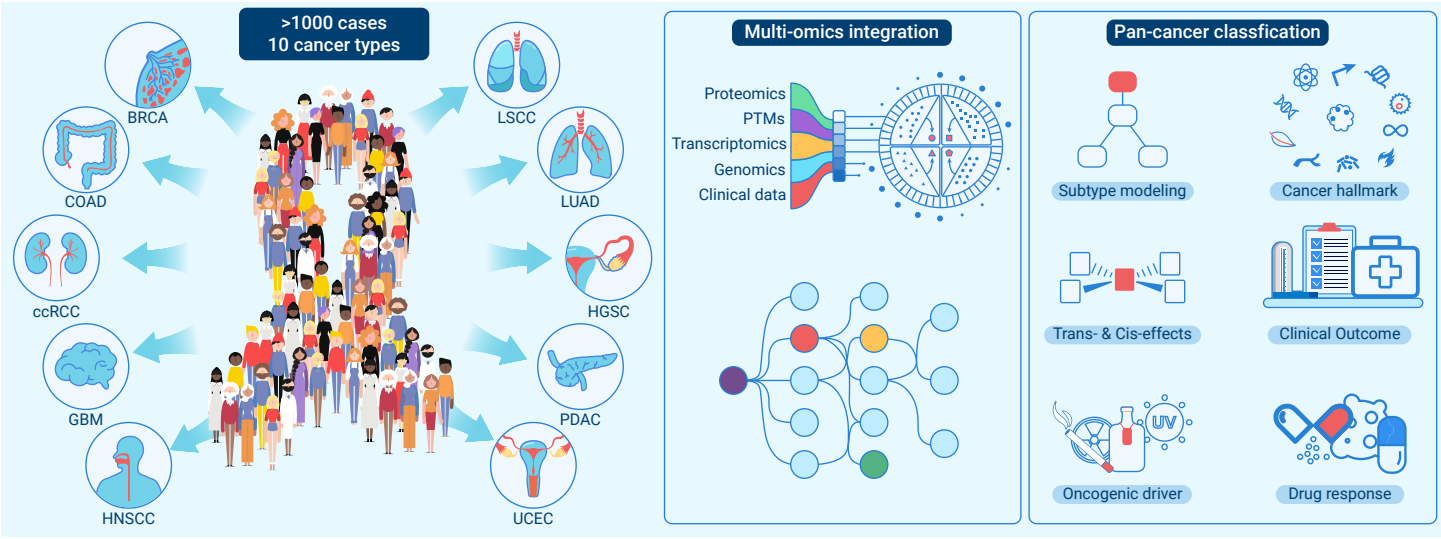

Figure 1. The panorama encompassing the CPTAC pan-cancer dataset The diverse molecular data types accessible within the CPTAC pan-cancer cohort consisting of >1000 cases of 10 cancer types. Genome, transcriptome, proteome, and PTMs and clinical data are available for all ten cancer types.

Importantly, the multi-omics pan-cancer analysis uncovered how gene mutations influence protein levels and interactions, while PTMs, especially phosphorylation, significantly impact protein interactions. Key findings encompass the connection between point mutations and copy-number changes with the changes at the protein level as well as restructuring of protein-protein interactions (PPIs). The study found that TP53 missense mutations were associated with higher protein abundance, whereas frameshift indel and nonsense mutations were associated with lower protein expression. Employing proteomic co-expression data to indirectly infer alterations in PPI networks, the study identified candidate PPIs exhibiting context-specific changes related to specific cancer types or driver mutations. Intriguingly, several driver alterations occurred at the interface of known interacting proteins, such as PIK3R1-PIK3CA, SMAD4-SMAD2, and PPP2R1A-PPP2R2A. Furthermore, investigating the trans-effects of driver events revealed a trend of molecular convergence, where cancer genes within the same pathway exhibited similar molecular patterns. This finding provided a rationale for the mutual exclusivity often observed among oncogenic drivers. Intriguingly, the study also identified instances of negative correlations in molecular fingerprints, often linked to mutually exclusive driver genes. This phenomenon suggested that a mutation in one gene renders the cellular environment incompatible with the oncogenic effects of a mutation in the second gene, potentially indicating opportunities for synthetic lethal vulnerabilities and innovative therapeutic approaches. Significantly, a considerable number of cancer genes tend to align with comparable molecular states, as indicated by sequence-based kinase activity profiles. Phosphoproteomic analyses

revealed that phosphorylation modifications also profoundly impact protein interactions. Notably, the study revealed that EGFR mutant samples exhibited kinase activity patterns opposite to those in STK11 and KRAS mutant samples, providing theoretical support for treatment strategies beyond genetic testing.

Further expanding on the pan-cancer exploration, the researchers identified key protein alterations driving cancer development. Distinct from other large-scale tumor characterization studies typically lacking normal controls, most of the CPTAC cohorts included matched or unmatched normal-adjacent tissues (NATs), allowing the investigation of tumor-specific signals compared to NATs. A total of 6,517 upregulated proteins and 7,030 downregulated proteins in tumors versus NATs were identified in different cancers. Among these, PLOD2 protein abundance displayed a negative correlation with overall survival in several cancers, suggesting its potential as a pan-cancer prognostic biomarker. Additionally, a consistent connection was found between kinase activity patterns and enriched pathways in different tumor types, reinforcing the importance of protein and PTM analysis in cancer research. In relation to new antigens and immunotherapies, the researchers systematically predicted neoantigens binding to patient-specific human leukocyte antigen class I alleles. The study revealed a positive correlation between neoantigen load and tumor mutation burden, as well as correlation with T-cell infiltration level, which is highly relevant to potential vulnerabilities for immunotherapeutic interventions. Importantly, the study also suggested that mutations that could potentially produce tumor neoantigens may not be effective targets for treatment if there is minimal or no expression of the abnormal protein product, which could potentially explain the lack of response to immunotherapy observed in some patients. Furthermore, the researchers developed the C3PO algorithm to study the impact of multi-gene mutations on tumor phenotypes. The results indicated that gene variations have complex effects on the proteome, with tumor-specific protein differences resulting from cumulative changes at multiple molecular levels including DNA, RNA, and protein.

PTMs AS FUNCTIONAL REGULATORS

PTMs encompass a diverse array of chemical modifications that can alter a protein's structure, stability, localization, and interaction partners. In cancer, aberrant PTMs contribute to the rewiring of signaling pathways, allowing cells to evade growth control mechanisms and promote uncontrolled proliferation. However, the shared regulatory patterns of PTMs among different cancer types, the interplay between PTMs, and the formation of regulatory networks by multiple PTMs are still poorly understood. In the parallel study,[4] "Pan-cancer analysis of post-translational modifications reveals shared patterns of protein regulation," the CPTAC team delved into the universe of PTMs, along with comprehensive genomic, transcriptomic, proteomic data from 1,110 patients of 11 tumor types, showcasing a unified pan-cancer analysis of PTMs' impact on protein regulation. Importantly, the researchers discerned patterns of PTM dysregulation that were unique to specific cancer subtypes, shedding light on novel therapeutic avenues. Furthermore, the researchers explored cross-talk between phosphorylation and acetylation, unraveling the complex interplay between these modifications and their potential role in driving cancer progression.

Compared to the genome and proteome, PTMs exhibited more discrete patterns across different tumor types, indicating a broader and more cancer-specific regulatory role of PTMs. Using tools like SignatureAnalyzer, the data were processed and analyzed, revealing multidimensional features spanning 33 subclusters. Most features extended across multiple tumor types, including significant enrichment of DNA damage response (DDR) and proliferation pathways (MYC and E2F), as well as pathways related to myogenesis and epithelial-mesenchymal transition (EMT). Additionally, the study identified 22 proteins in the pan-cancer cohort with significant phosphorylation site clustering features and one protein (SWI/SNF chromatin remodeling factor ARID1A protein) with significant acetylation site clustering.

Based on the pan-cancer PTM dataset, the study further explored the regulatory features of PTMs in five hallmark biological processes of cancer, including DNA repair, immune response, cellular metabolism, histone regulation, and kinase regulation. The study found that PTM dysregulation was associated with different DNA damage repair mechanisms. Phosphorylation-centric analysis revealed and characterized information patterns that couldn't be detected at the genome and transcriptome levels. For example, through analysis of the homologous recombination deficiency (HRD) cluster, the study found significant differences in DNA repair protein phosphorylation closely correlated with the severity of hypoxia. Interactions between cellular metabolism and immune response were previously established, and the study further described differences in PTM regulation between different tumor subtypes, revealing a connection between changes in acetylation of metabolic proteins and tumor immune status. The study identified four broad immune subtypes in various cancer types: immune-cold, immune-cool, immune-warm, and immune-hot subtypes, driven by acetylation-associated distinct metabolic phenotypes. Additionally, the study comprehensively analyzed the interplay between acetylation and phosphorylation using a kinase library, revealing that the phosphorylation of Thr/Ser kinases is influenced by proximal acetylation, allowing us to predict potentially responsible interfering kinases.

A RICH DATA RESOURCE

The comprehensive proteogenomic datasets meticulously curated by CPTAC have orchestrated the amalgamation of vast genomic, transcriptomic, proteomic, and clinical data, hailing from diverse cancer cohorts as well as normal tissues. This collective endeavor has culminated in the establishment of a unified and all-encompassing resource, tailored to the discerning needs of the cancer research community.[5] This multifaceted repository of data converges harmoniously to construct a panoramic representation of cellular states, bestowing researchers with an unparalleled opportunity to unravel the intricate regulatory dynamics that interlink DNA mutation events, acting as triggers for perturbed signaling networks, with the ultimate manifestation of cellular phenotypes.

With an aim to augment data reusability, the authors have introduced a computational framework for harmonizing data, coupled with a diverse array of dissemination strategies meticulously designed to impart accessibility to both the raw and processed datasets. The raw proteogenomic data engendered by the CPTAC initiative is made widely available through established platforms such as The Genomic Data Commons (GDC) and Proteomic Data Commons (PDC). To further facilitate access, CPTAC has ingeniously developed a Python application programming interface (API), which seamlessly streams finalized quantitative data tables, providing a gateway to the harmonized pan-cancer datasets. Furthermore, the authors have devised multiple user-friendly web portals dedicated to data visualization and analysis. Among these innovative platforms are PepQuery, LinkedOmicsKB, PTMcosmos, ProTrackPath, and NGlycositeAtlas portal, each serving as a virtual gateway for researchers to explore and glean insights from the intricate proteogenomic landscape. This holistic approach not only propels scientific inquiry but also bolsters collaborations and accelerates the translation of data-driven discoveries into transformative advancements in cancer research and treatment.

REFERENCES

1. Mani, D.R., Krug, K., Zhang, B., et al. (2022). Cancer proteogenomics: Current impact and future prospects. Nat. Rev. Cancer **22**: 298–313. DOI: 10.1038/s41568-022-00446-5.
2. Rodriguez, H., Zenklusen, J.C., Staudt, L.M., et al. (2021). The next horizon in precision oncology: Proteogenomics to inform cancer diagnosis and treatment. Cell **184**: 1661–1670. DOI: 10.1016/j.cell.2021.02.055.
3. Li, Y., Porta-Pardo, E., Tokheim, C., et al. (2023). Pan-cancer proteogenomics connects oncogenic drivers to functional states. Cell **186**: 3921-394410. DOI: 10.1016/j.cell.2023.07.014.
4. Geffen, Y., Anand, S., Akiyama, Y., et al. (2023). Pan-cancer analysis of post-translational modifications reveals shared patterns of protein regulation. Cell **186**: 3945-3967. DOI: 10.1016/j.cell.2023.07.013.
5. Li, Y., Dou, Y., Da Veiga Leprevost, F., et al. (2023). Proteogenomic data and resources for pan-cancer analysis. Cancer Cell **41**: 1397–1406. DOI: 10.1016/j.ccell.2023.06.009.

FUNDING AND ACKNOWLEDGMENTS

This work was in part supported by NIH R21 CA267527-01 and the Feldstein Medical Foundation grant. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.