

# → Advancing cancer research by improving access to NCI's proteogenomic data



The National Cancer Institute (NCI) estimates there were over 1.8 million new cancer cases in 2020 and 600,000 deaths. In the search for cures, NCI initiated an extensive analysis of the proteins expressed in cancer cells to understand how the disease starts and progresses, and to help in diagnoses and treatment.

To fuel their analysis, NCI receives large volumes of mass spectrometry data from research groups in the Clinical Proteomic Tumor Analysis Consortium (CPTAC). But they needed a better way to store and leverage the data: one central location that was accessible to all cancer researchers interested in the tumor proteome and maintains results for future research after the conclusion of each CPTAC cancer program.

ICF worked with NCI to develop a research data management solution and interactive portal that ensures speed and quality when handling large volumes of mass spectrometry data in the CPTAC—improving access to proteogenomic data.

The Proteomic Data Commons (PDC) portal is part of the NCI Cancer Research Data Commons (CRDC). It ensures proteomic data from each tumor is integrated with genomic sequence and patient clinical characteristics. And researchers across 140 countries have searched for data by cancer type, publication, or research program. In addition, PDC runs bioinformatic computational pipelines to identify peptides and proteins expressed in each tumor.

## Bridging the gap between health and technology

The CPTAC Data Coordinating Center and the PDC provide information about the cancer proteome to researchers around the world. The site provides secure and private areas for researchers to access and exchange data—as well as a public portal for the distribution of data from the CPTAC program and collaborators in the International Cancer Proteome Consortium (ICPC).

**“[ICF] merges a fundamental understanding of biology with expertise in data quality control and data security. Integrating all of these factors is key to delivering a secure and fast data portal for the scientific community.”** —Program manager, NCI

To date, 54K+ scientists have accessed data and downloaded 320K+ files, highlighting the breadth of researchers using the PDC to inform cancer research. The program has been featured in 18 landmark proteogenomic cancer publications—advancing our understanding of cancers, including ovarian, breast, colon, lung, pediatric, and adult brain cancer.

**“We pride ourselves on quality of data and attention to biological details because it informs real outcomes.”**

—Karen Ketchum, ICF, vice president, data analytics

**We're proud to support NCI in advancing our understanding of the molecular basis of cancer.**

[Click here](#) to learn more about how we bridge the gap between health and technology so organizations can focus on scientific discovery, patient care, and population health.

