PANOPLY is a cloud-based platform for automated and reproducible proteogenomic data analysis, enabling the use of state-of-the-art statistical and machine learning algorithms to transform multi-omic data from cancer samples into biologically meaningful and interpretable results. Salient features of PANOPLY include:

- Comprehensive collection of algorithms from CPTAC landmark proteogenomic studies\(^{1,2}\) and more;
- Is easy to use;
- Integrates Genomic, Proteomic, and PTM data analysis;
- Automates flexible and reproducible workflows.

PANOPLY has been applied to routine proteogenomic analysis of a range of CPTAC datasets including breast cancer (BRCA), uterine cancer (UCEC)\(^{3}\), lung adenocarcinoma (LUAD)\(^{4}\), lung squamous cell carcinoma (LSCC), glioblastoma (GBM), pancreatic ductal adenocarcinoma (PDAC) and pediatric brain tumors (PBT).

**Introduction**

Recent technological advances in NGS and MS-based proteomics have enabled the rapidly advancing field of proteogenomics—the integrative analysis of genomic, transcriptomic, proteomic, and post-translational modification (PTM) data. Many landmark studies\(^{1,5-7}\) by the Clinical Proteomic Tumor Analysis Consortium (CPTAC, proteomics.cancer.gov) have highlighted its impact in promoting deeper insights in cancer biology and potential drug target identification. PANOPLY is a collection of state-of-the-art algorithms for proteogenomic and multi-omic data analysis, packaged in a simple and easy to use interface with the goal of producing biologically meaningful and interpretable results.

**PANOPLY User Interface**

- **Overview**
  - PANOPLY is a cloud-based platform for automated and reproducible proteogenomic data analysis, enabling the use of state-of-the-art statistical and machine learning algorithms to transform multi-omic data from cancer samples into biologically meaningful and interpretable results. Salient features of PANOPLY include:
  - Comprehensive collection of algorithms from CPTAC landmark proteogenomic studies\(^{1,2}\) and more;
  - Is easy to use;
  - Integrates Genomic, Proteomic, and PTM data analysis;
  - Automates flexible and reproducible workflows.

- **Features & Functionality**
  - **Flexible**
    - Easy to combine and customize new pipelines using docker images and Workflow Definition Language (WDL) specifications.
  - **Automated**
    - Preprogrammed executions.
    - Reuse previous computations (job abidance) to improve scalability and reduce costs.
  - **Reproducible**
    - Export and share entire pipelines and associated data, with version-control and associated digital object identifiers (DOI).
  - **Scalable and Secure**
    - Inherently scalable cloud-based architecture.
    - Appropriate access control to enforce data privacy requirements.

- **PANOPLY Architecture Overview**
  - PANOPLY implements a wide array of algorithms, applicable to all cancer types. In addition, disease-specific customizations can be easily added.

- **PANOPLY Application to CPTAC Data**
  - **Samples and Data**
    - 108 TCGA samples profiled with 37 iTRAQ 4-plexes.
    - 105 tumors – 25 basal-like, 29 luminal A, 33 luminal B, 18 HER2-enriched
    - 3 normal breast
    - 77 tumors used in study after QC
    - Global proteome and phosphoproteome
  - **Genome–WXS, CNV, RNAseq**
    - **Cost**
      - $10 per sample for compute and storage (including MS search, and proteogenomic analysis)

- **PTM-SEA (PTMseqDB)**
  - **K-means clustering performed on pathways derived from single sample GSEA analysis of phosphoproteome data identifies four distinct clusters.**
    - These clusters are unique to the phosphoproteome.
  - **Histogram of protein-mRNA correlation for 77 CPTAC breast cancer tumors.**
    - Statistically significant (FDR < 0.05) correlations are shown.
    - Median correlation = 0.38.

- **References**


**Availability**

This work was supported by grants U24CA210979 and U24CA210866 from the National Cancer Institute as part of Clinical Proteomic Tumor Analysis Consortium (CPTAC, http://proteomics.cancer.gov) initiative.