**OVERVIEW**

**PANOPLY**—A cloud-based platform for automated and reproducible proteogenomic data analysis

D. R. Mani, Myranda Maynard, Karsten Krugi, Ramani Kothadia, Karen Christianison, David Heiman, Karl R. Clauser, Gad Getz and Steven A. Carr

*Proteomics Platform, *Cancer Genome Computational Analysis, Broad Institute of MIT and Harvard, Cambridge, MA

---

**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. Key features of **PANOPLY** include:

- **Comprehensive collection of algorithms from CPTAC and PANOPLY** (latest version)
- **Is easy to use**
- **Integrates Genomic, Proteomic, and PTM data analysis**
- **Flexible and reproducible workflows**

It has been applied to routine proteomic analysis of a range of CPTAC datasets including breast cancer (BRCA), uterine cancer (UCS), lung adenocarcinoma (LUAD), 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 collection of proteogenomic datasets—the integrative analysis of genomic, transcriptomic, proteomic, and post-translational modification (PTM) data. Many landmark studies [1-7] by the Clinical Proteomic Tumor Analysis Consortium (CPTAC) have highlighted its impact in promoting deeper insights in cancer biology and in 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.

---

**Features & Functionality**

**PANOPLY** leverages Terra (*.app.terra.bio) to include proteogenomic workflows, it is a **Google cloud-based platform** developed at the Broad Institute for genome-scale analysis and data sharing. It is designed to be:

- **Flexible**—Easily combine and customize new pipelines using Docker images and Workflow Description Language (WDL) specifications.
- **Automated**—Preprogrammed executions. Reuse previous jobs (computations) 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**

---

**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**

[Illustrations show TCGA Breast cancer analysis]

---

**PANOPLY User Interface**

---

**PANOPLY Startup Notebook**

Provides simple step-by-step instructions for uploading data, identifying data types, specifying parameters, and setting up the **PANOPLY** workspace.

---

**Add your module to PANOPLY**

- **Docker**—Encapsulate your code files in a docker.
- **Docker Hub**—Build and push it to a docker registry of your choice.
- **WDL**—Write a WDL such that it:
  - Takes correct input from Terra Data Model
  - Calls the correct script, passing matched arguments
  - Sources the right container image
  - Terra Workflow—Save this WDL as a workflow on Terra
- **Plug in this method into PANOPLY using outputs from PANOPLY as input to your method.**

---

**Summary**

**PANOPLY** implements a high-throughput cloud-based platform for transforming genomic, proteomic, and other post-translational modification (PTM) data into biologically meaningful information easily accessible to scientists worldwide. The flexible, robust, reproducible, and automated platform has been applied to a range of CPTAC and other cancer datasets.

**Availability**

[https://app.terra.bio/](https://app.terra.bio/)

---

**References**

2. Gehron*, et al. (2013)
5. Petret*, et al. (2020)