Abstract

The cBioPortal for Cancer Genomics is an open-source software platform that enables interactive, exploratory analysis of large-scale cancer genomics data sets with a user-friendly interface. It integrates genomic and clinical data, and provides a suite of visualization and analysis options, including OncoPrint, mutation diagram, variant interpretation, survival analysis, expression correlation analysis, alteration enrichment analysis, cohort and patient-level visualization, among others.

The public site ([https://www.cbioportal.org](https://www.cbioportal.org)) hosts data from almost 300 studies spanning individual labs and large consortia. Data is also available in the cBioPortal Datahub ([https://github.com/cBioPortal/datahub/](https://github.com/cBioPortal/datahub/)). In 2020 we added data from 21 studies, totaling almost 30,000 samples. In addition, we added data to existing TCGA PanCancer Atlas studies, including MSI status, mRNA-seq z-scores relative to normal tissue, microbiome data, and RPPA-based protein expression. The cBioPortal also supports AACR Project GENIE with a dedicated instance hosting the GENIE cohort of 112,000 clinically sequenced samples from 19 institutions worldwide ([https://genie.cbioportal.org](https://genie.cbioportal.org)).

The site is accessed by over 30,000 unique visitors per month. To support these users, we hosted a five-part instructional webinar series. Recordings of these webinars are available.
on our website and have already been viewed thousands of times.

In addition, more than 50 instances are installed at academic institutions and pharmaceutical/biotechnology companies. In support of these local instances, we continue to simplify the installation process: we now provide a docker compose solution which includes all microservices to run the web app as well as data validation, import and migration.

We continue to enhance and expand the functionality of cBioPortal. This year we significantly enhanced the group comparison feature; it is now integrated into gene-specific queries and supports comparison of more data types including DNA methylation, microbiome, and any outcome measure. We also expanded support of longitudinal data: the existing patient timeline has been refactored and now supports a wider range of data and visualizations; a new “Genomic Evolution” tab highlights changes in mutation allele frequencies across multiple samples from a patient; and samples can now be selected based on pre- or post-treatment status. Other features released this year include: allowing users to add gene-level plots for continuous molecular profiles in study view, enabling users to select the desired transcript on the Mutations tab, and integration of PathwayMapper.

The cBioPortal is fully open source (https://github.com/cBioPortal/) under a GNU Affero GPL license. Development is a collaborative effort among groups at Memorial Sloan Kettering Cancer Center, Dana-Farber Cancer Institute, Children's Hospital of Philadelphia, Princess Margaret Cancer Centre, Bilkent University and The Hyve.