Open Source Development Success through collaboration: Contributions to cBioPortal

the hyve

We empower scientists by building on open source software

List of Authors: **Sjoerd van Hagen[1], Pieter Lukasse[1], Sander de Ridder[1], Fedde Schaeffer[1], Priti Kumari[2], James Lindsay[2], JianJiong Gao[3], Benjamin Gross[3], Zachary Heins[3], Adam Abeshouse[3], Hongxin Zhang[3], Yichao Sun[3], Robert Sheridan[3], Onur Sumer[3], Stuart Watt[4], Chris Sander[2], Nikolaus Schultz[3], Ethan Cerami[2]**

[1] The Hyve, Weg der Verenigde Naties 1, 3527 KT Utrecht, The Netherlands
The Hyve, Cambridge Innovation Center, 1 Broadway, 14th Floor, Cambridge, MA 02142, USA
[2] Dana Farber Cancer Institute
[3] Memorial Sloan Kettering Cancer Center
[4] Princess Margaret Cancer Centre

Abstract

Through collaboration on open source software, pharma, commercial software development companies and cancer research institutions have proven successful in enhancing the cBioPortal platform by optimizing and extending it with new features.

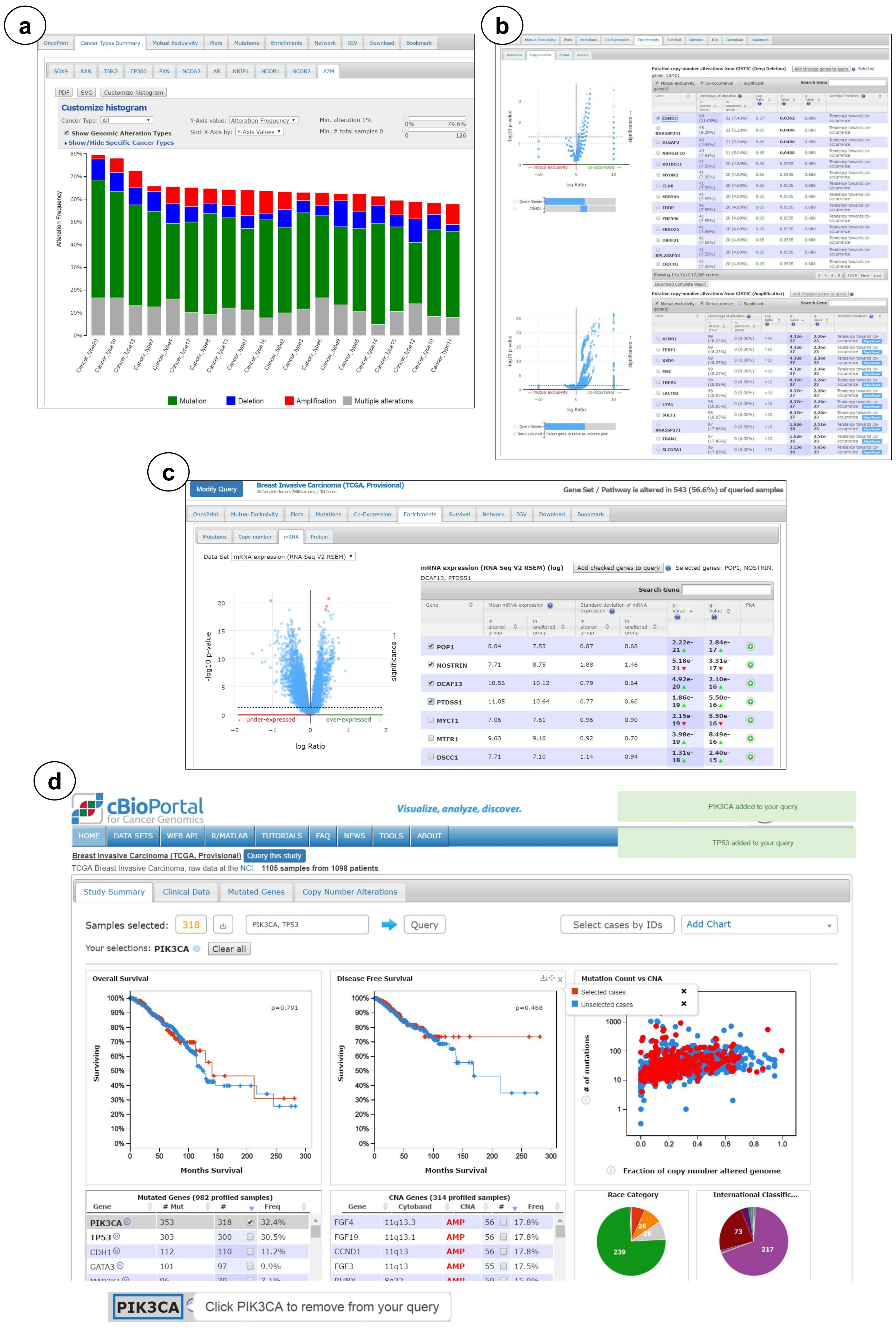


Figure 1: In the front end we added **(a)** a whole new pan-cancer view for studies comprising multiple cancer types, **(b,c)** new visualizations to the query results page to support better enrichment analysis of expression (mRNA, Proteins) and co-occurrence (copy number, mutations) and **(d)** new query options in the Study overview page. We have also implemented integration of documentation from the Wiki or Git and made the portal more customizable (logo, headers, news and FAQ), which is very important for open source software.

References

[1] The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Ethan Cerami, Jianjiong Gao, Ugur Dogrusoz, Benjamin E. Gross, Selcuk Onur Sumer, Bülent Arman Aksoy, Anders Jacobsen, Caitlin J. Byrne, Michael L. Heuer, Erik Larsson, Yevgeniy Antipin, Boris Reva, Arthur P. Goldberg, Chris Sander and Nikolaus Schultz. Cancer Discovery. May 1, 2012 2; 401. doi: 10.1158/2159-8290.CD-12-0095

[2] Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Gao J1, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Sci Signal. 2013 Apr 2;6(269):pl1. doi: 10.1126/scisignal.2004088.

[3] cBioPortal goes Open Source. Ward Weistra, http://thehyve.nl/cbioportal-goes-open-source/

[4] First cBioPortal hackathon. Ethan Cerami. http://biobits.org/cbio-portal-hackathon.html

[5] The Hyve joins first cBioPortal hackathon. Pieter Lukasse. http://thehyve.nl/first-cbioportal-hackathon/

DANA-FARBER CANCER INSTITUTE

Memorial Sloan-Kettering Cancer Center

UHN Princess Margaret Cancer Centre

cBioPortal for Cancer Genomics

Introduction

Approximately one year ago the popular cBioPortal for Cancer Genomics[1,2] was made open source[3]. In this last year its development community has grown and the platform has been extended with many new features. Here we detail some of the contributions The Hyve (Utrecht) has made to the platform, in collaboration with Dana Farber Cancer Institute (Boston), Memorial Sloan Kettering Cancer Center (New York) and Boehringer Ingelheim (BI RCV). The contributions can roughly be divided into three categories: (I) new data analysis features (Figure 1), (II) improvement of the data loading pipeline (Figure 2), and (III) performance optimizations of the front end to be able to host larger studies[4,5].

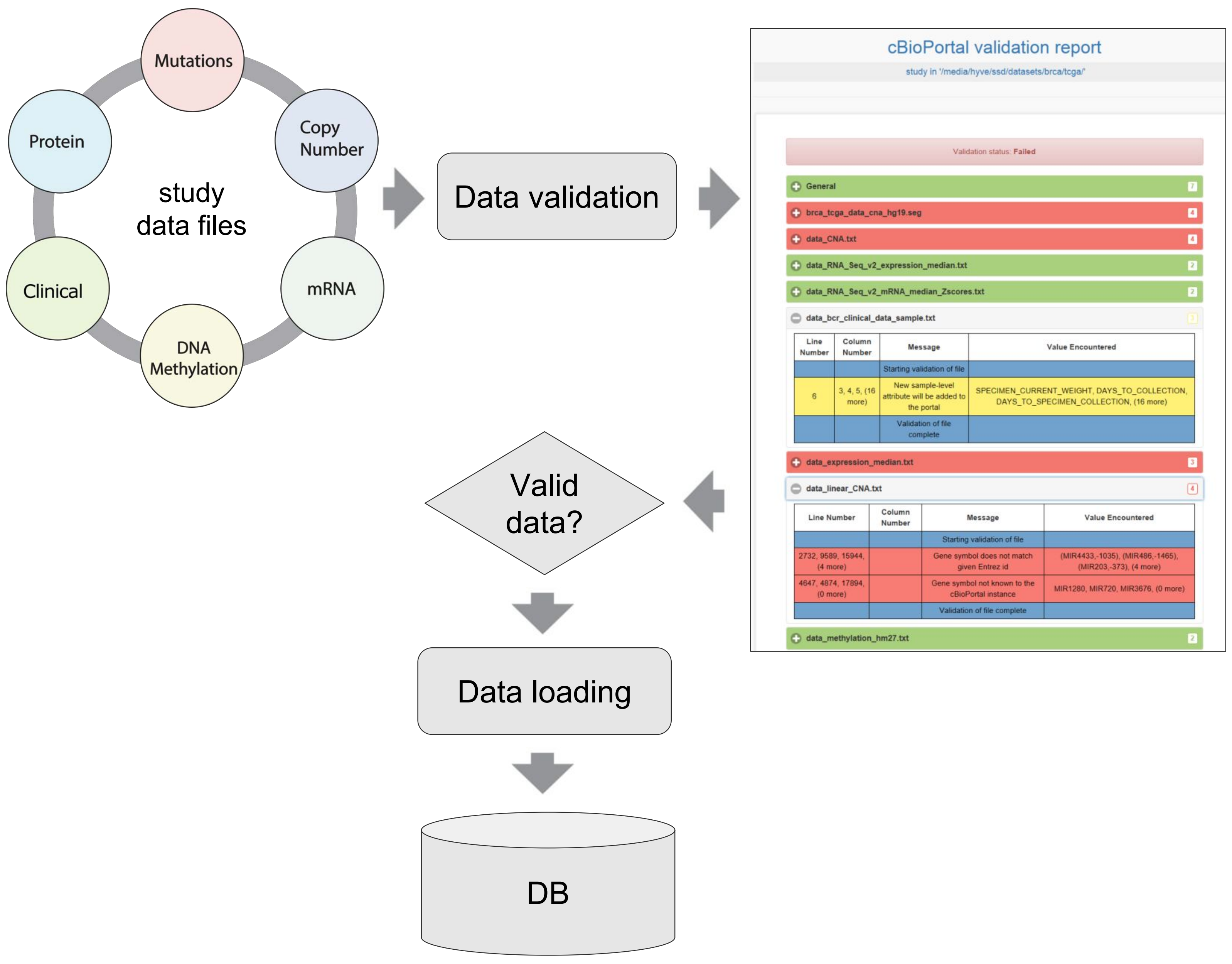
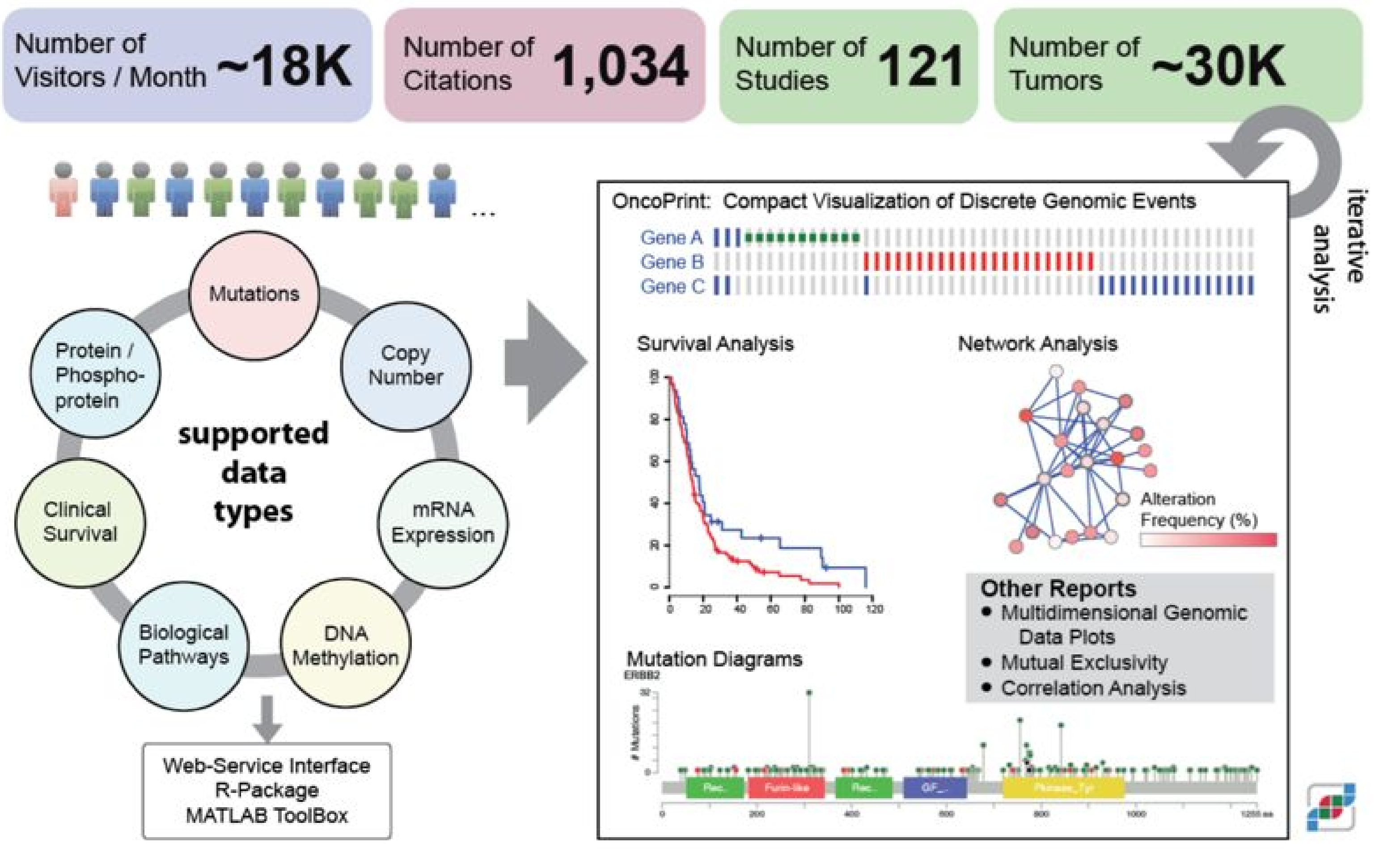


Figure 2: In the data loading pipeline we have introduced a strict separation between the validation step and the loading step. This "separation of concerns" design principle makes the code easier to understand and maintain and simplifies the process of adding new datasets to a local cBioPortal installation. Special effort was spent on making the validator easy to use, which is exemplified by clearer error messages and the generation of an HTML validation report.

About the cBioPortal community

Our community consists of a group of software engineers, bioinformaticians and cancer biologists building software solutions for precision medicine for cancer patients. Our overall goal is to build infrastructure to support clinical decisions for personalized cancer treatment by utilizing "big data" of cancer genomics and patient clinical profiles.



Our multi-institutional team currently has more than 30 active members, primarily from Memorial Sloan Kettering Cancer Center in New York, the Dana-Farber Cancer Institute in Boston, Princess Margaret Cancer Centre in Toronto, Children's Hospital of Philadelphia, and The Hyve, a bioinformatics company from the Netherlands.

Contact: office@thehyve.nl. US: 1 Broadway, 14th Floor Cambridge, MA 02142 United States. Tel: +1 617-869-9256
The Netherlands: Weg der Verenigde Naties 1, Room 1.10 3527 KT Utrecht The Netherlands. Tel: +31 30 7009713

http://thehyve.nl