**Reproducible PDX Genomic Data Analysis**

Boris Jurič, Muhammad Usman, Zdenka Dudová, Radim Peša, Dalibor Stuchlík, and Aleš Křenek

Masaryk university

Acquisition, processing, and interpretation of next-generation sequencing (NGS) data became ubiquitous in cancer research because of bringing strong evidence on biological processes related to tumor growth etc. On the other hand, besides multiple experimental methods, wide variety of computational approaches exists, yielding reproducibility and comparison of such results extremely difficult.

Our work is focused on patient-derived xenograft (PDX) mice models – in vivo human tumor implants, a well-established technique in translational cancer research and treatment selection. Besides the generic issues of all NGS data and their processing, the specific problem of PDX based data is an intrinsic mixture of human tumor and mouse tissue – those cannot be fully separated during sample preparation, the problem must be addressed computationally. Therefore, specific computational pipeline setups are also used.

The [EurOPDX](http://europdx.eu/) consortium invests significant effort into harmonization of its members data [2], including the outputs of PDX-specific NGS pipelines. However, the NGS data processing itself has not been well coordinated yet.

We present an integrated solution of several such pipelines ([1], <https://github.com/jrderuiter>, …) implemented in Galaxy – all the required tools were wrapped properly and connected into smoothly running workflows. The results of the workflow can be directly propagated to a testing clone of [EurOPDX data portal](https://dataportal.europdx.eu/) for immediate comparison with other data already published in the Data Portal, visualization, etc.

The solution – bioinformatics pipelines and a clone of Data Portal including all visualization tools – is available as a service, where data can be uploaded and processed, or as a set of Docker containers which can be run by the user close to the data, allowing to handle non-disclosed, sensitive data etc., while still offering the full functionality.

[1] Woo, X. Y. *et al.* (2019). Genomic data analysis workflows for tumors from patient-derived xenografts (PDXs): challenges and guidelines. *BMC Medical Genomics*, 12.

[2] Dudová, Z., Conte, N., Mason, J. et al. The EurOPDX Data Portal: an open platform for patient-derived cancer xenograft data sharing and visualization. BMC Genomics 23, 156 (2022). https://doi.org/10.1186/s12864-022-08367-1

*Maximum abstract length is ½ of A4 page.*