

# Challenges and Experiences Extending the cBioPortal for Cancer Genomics to a Molecular Tumor Board Platform

Niklas REIMER<sup>a,1,\*</sup>, Philipp UNBERATH<sup>b,\*</sup>, Hauke BUSCH<sup>a</sup>, Melanie BÖRRIES<sup>c,d</sup>, Patrick METZGER<sup>c</sup>, Arsenij USTJANZEW<sup>c</sup>, Christopher RENNER<sup>b</sup>, Hans-Ulrich PROKOSCH<sup>b</sup> and Jan CHRISTOPH<sup>b,f</sup>

<sup>a</sup>Group for Medical Systems Biology, Lübeck Institute of Experimental Dermatology, Universität zu Lübeck, Germany

<sup>b</sup>Department of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

<sup>c</sup>Institute of Medical Bioinformatics and Systems Medicine, Faculty of Medicine and Medical Center – University of Freiburg, Freiburg, Germany

<sup>d</sup>German Cancer Consortium (DKTK) Partner Site Freiburg and Cancer Research Center (DKFZ), Heidelberg, Germany

<sup>e</sup>Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), Johannes Gutenberg-University School of Medicine, Mainz, Germany

<sup>f</sup>Junior Research Group (Bio-)Medical Data Science, Martin-Luther-University Halle-Wittenberg, Faculty of Medicine, Halle, Germany

**Abstract.** In Molecular Tumor Boards (MTBs), therapy recommendations for cancer patients are discussed. To aid decision-making based on the patient's molecular profile, the research platform cBioPortal was extended based on users' requirements. Additionally, a comprehensive dockerized workflow was developed to support the deployment of cBioPortal and connected services. In this work, we present the challenges and experiences of nearly two years of implementing and deploying an MTB platform based on cBioPortal and compare those to findings of a previous study.

**Keywords.** cBioPortal, Molecular Tumor Board, MIRACUM, Precision Medicine, Genomics, Docker

## 1. Introduction

Next-generation sequencing (NGS) techniques are becoming widespread in personalized cancer treatment [1]. These data are fundamentally important in the context of Molecular Tumor Boards (MTBs), where experts from different fields, like oncology, bioinformatics, and systems medicine, jointly discuss therapy options for cancer patients

---

<sup>1</sup> Corresponding Author, Niklas Reimer, Group for Medical Systems Biology, Lübeck Institute of Experimental Dermatology, Universität zu Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany; E-mail: n.reimer@uni-luebeck.de.

\*These authors contributed equally to this work.

based on molecular data. While these advancements have demonstrated their potential to improve patient outcomes already [2,3], the management, analysis, and interpretation of these data poses a challenge to traditional healthcare systems. In research, however, several software tools supporting data processing and interpretation exist, one of which is the cBioPortal for Cancer Genomics [4,5] developed by the Memorial Sloan Kettering Cancer Center (MSKCC). In order to evaluate the impact of bringing such tools to MTBs and personalized cancer treatment, the MIRACUM Use Case 3 [6] is developing a comprehensive workflow and tool architecture from the sequencer to the clinician providing care to cancer patients. One crucial part of this goal is to deploy an MTB software platform based on cBioPortal and extended with various additional functionalities identified from extensive requirements analysis with real-world users in the clinics [7].

## **2. Objectives**

In this work, we share our experience in extending cBioPortal for use in a clinical setting. This includes the challenges and possible solutions encountered and the cBioPortal extensions deployed together with their architectural details, surrounding tools, and milestone releases' rollout. These developments are compared and evaluated concerning the findings of a previous study from 2018 that described extending cBioPortal in a research setting [8] regarding the heterogeneity of system environments and different levels of integrated data available. The deployment processes are monitored through feedback forms to collect issues encountered during the setup.

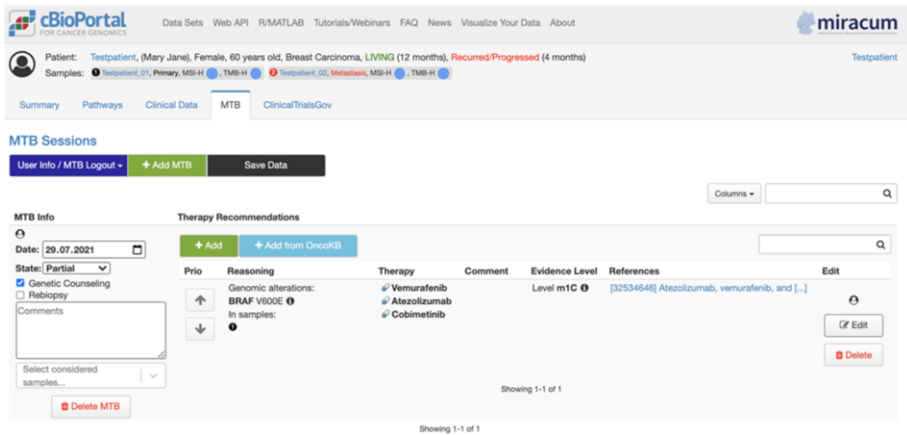
## **3. Development**

Having gained the first experience with extending cBioPortal (version 1.11.3) in a research setting in 2018 [8], the development of additional functionalities started in late 2019 with the cBioPortal version 3.1.2, derived from a detailed requirements analysis for the use of cBioPortal as an MTB platform [7]. Since it was expected that not all extensions were suitable for contribution to the main cBioPortal project, the development was carried out on a separate repository, forked from the original codebase but kept up to date alongside updates to the cBioPortal codebase [9]. Forking was only necessary for the cBioPortal frontend project as none of the implemented functionalities needed adaptations to the cBioPortal backend or database. During the 22 months of development, a total of 45 updates releases were also transferred to our forked version. All issues during such updates were caused by refactoring in the cBioPortal codebase or moving reusable parts of the codebase into separate packages. Apart from that, about 15 times, merge conflicts occurred but were easy to solve. This means that updates can usually be applied with little or no effort, and long-term support for features implemented in the fork would indeed be possible.

The most significant extensions were two new tabs in the patient view of cBioPortal. Firstly, one to allow for the search of clinical trials based on genomic and clinical data of the patient achieved with direct integration of ClinicalTrials.gov (publication currently under revision). Secondly, one to enable structured and standardized documentation of therapy recommendations of the MTB as shown in Figure 1. This also included a novel and more detailed authentication and authorization concept and the integration of the

service FhirSpark to provide a FHIR-compliant way to store therapy recommendations [10]. Both extensions were developed using a user-centered design process and evaluated through a usability test.

There were also several minor extensions and adaptations, e.g., the integration of approval status of drugs by the European Medicines Agency directly in the OncoKB annotation already available in cBioPortal, import and display of LoH-mutations, and handling of internal PDF documents by rendering them through the web browser instead of depending on external Google services.



**Figure 1.** The newly implemented MTB tab offers structured documentation of therapy recommendations by selecting clinical and mutation data for reasoning, drugs as a therapy, and the corresponding evidence level based on approval state and available references. Multiple therapy recommendations can be prioritized.

#### 4. Deployment

A major challenge in software deployment was enabling cBioPortal and its connected tools and services, like databases and annotation tools, to be distributed as simply as possible. The official cBioPortal GitHub project site initially provided a solution for the deployment of cBioPortal via Docker [11], requiring manual intervention at multiple points, including a manual setup of the containers. Therefore, we developed a custom workflow [12] for the standardized and simplified deployment of our extended version of cBioPortal, including a MySQL database, session service, an on-premise instance of Genome Nexus [13], and the FhirSpark service described in section 3. This solution also includes the option to deploy the standard cBioPortal in a research setting, which was later used by the team of MSKCC as the basis for the official cBioPortal Docker project [14].

Our Docker-based workflow successfully enabled the deployment of cBioPortal at all ten consortial sites of MIRACUM, along with other partner sites from the HIGHmed [15] and SMITH [16] consortia, as well as sites from the Bavarian Centre for Cancer Research (BZKF) and the German Cancer Consortium (DKTK). The deployment at the consortial sites was accompanied by structured feedback forms to collect comments and problems during the installation. The evaluation of the forms revealed that some sites delegated the deployment to the IT department, while at other sites, domain experts handle the deployment. The only notable difficulties were caused by the integration with

other tools, like the configuration of the identity provider Keycloak [17] or site-specific barriers like highly restrictive firewalls, incomplete proxy configurations, or issues integrating Transport Layer Security encryption.

As MIRACUM Use Case 3 intends to provide a comprehensive workflow for MTB case preparation, the deployment project builds upon MIRACUM-Pipe [18]. This sequencing pipeline includes advanced variant annotation and generates an interactive PDF report and files needed for importing mutation data in cBioPortal. Using this complete setup, the whole process from the sequencer to visualization in cBioPortal could be successfully tested at three sites with over 370 whole-exome and panel sequencing cases.

Although extensive documentation of the import data format of cBioPortal is readily available, the generation of the required files to import both clinical and molecular data is challenging. Especially when adding new patients to an existing study, file management becomes prone to error and hardly practicable by hand because IDs must be consistent across patient, sample, and case-list files. At the same time, newly added columns must also be populated in existing patients. To aid this process, *cbpManager* [19] was developed and integrated into the dockerized deployment workflow. The tool features a user interface that allows for easy import file generation and management.

## 5. Discussion and Conclusion

The major challenge on the development side is the long-term support of features like the therapy recommendation that are specific to the described use case as they require a regular merging with the upstream codebase. Even though the codebase structure is significantly improving, like refactoring code into separate reusable packages and separating the backend and frontend project [8], this step requires programming skills and thorough testing. However, a plugin concept in cBioPortal would help overcome this issue, depending on its customization capabilities.

Due to the varying delegation of the deployment mentioned in section 4 and complex domain-specific features, it will be necessary to provide detailed documentation. Especially edge cases like different types of proxy configurations were added to the Docker Compose solution, including test data sets for easier verification of proper functionality.

Even though the development of *cbpManager* lowers the barrier when managing studies with cBioPortal, it still requires manual curation of data that will also be available in the local data integration centers. Therefore, long-term solutions should use these as primary data sources for automated ETL (extract, transform, load) workflows, requiring less manual interaction.

Currently, all provided tools are intended for research use only and are not approved for the diagnosis or treatment of individual patients. However, future efforts in the project will also cover the tools' compliance within the scope of the Medical and In Vitro Diagnostic Regulation (MDR and IVDR).

This work demonstrates how cBioPortal can be extended and integrated with other tools to a comprehensive and easily deployable MTB software solution. While the concurrent development and continuous updates of a forked cBioPortal are not trivial, refactoring the original project significantly impacted the maintainability. However, contributing as many features as possible to the main project should remain the primary goal.

## Acknowledgment

The project is funded by the German Federal Ministry of Education and Research (BMBF), Grand IDs 01ZZ1802Z (HiGHmed), 01ZZ1803N (SMITH), 01ZZ1801A, 01ZZ1801B, and 01ZZ1801F (all MIRACUM). We especially thank Nikolaus Schultz and Jianjiong Gao from the MSKCC, who provided us with feedback on the identified requirements and our planned changes to cBioPortal. The present work was performed in (partial) fulfillment of the requirements for obtaining the degree “Dr. rer. biol. hum.” from the Friedrich-Alexander-Universität Erlangen-Nürnberg (PU).

## References

- [1] Garraway LA, Verweij J, Ballman KV. Precision Oncology: An Overview. *JCO* 2013;31:1803–5.
- [2] Hoefflin R, Geißler A-L, Fritsch R, Claus R, Wehrle J, Metzger P, et al. Personalized Clinical Decision Making Through Implementation of a Molecular Tumor Board: A German Single-Center Experience. *JCO Precision Oncology* 2018;1–16.
- [3] Hoefflin R, Lazarou A, Hess ME, Reiser M, Wehrle J, Metzger P, et al. Transitioning the Molecular Tumor Board from Proof of Concept to Clinical Routine: A German Single-Center Analysis. *Cancers* 2021;13:1151.
- [4] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discovery* 2012;2:401–4.
- [5] Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science Signaling* 2013;6:p11–p11.
- [6] Prokosch H-U, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, et al. MIRACUM: Medical Informatics in Research and Care in University Medicine: A Large Data Sharing Network to Enhance Translational Research and Medical Care. *Methods Inf Med* 2018;57:e82–91.
- [7] Buechner P, Hinderer M, Unberath P, Metzger P, Boeker M, Acker T, et al. Requirements Analysis and Specification for a Molecular Tumor Board Platform Based on cBioPortal. *Diagnostics* 2020;10:93.
- [8] Unberath P, Knell C, Prokosch H-U, Christoph J. Developing New Analysis Functions for a Translational Research Platform: Extending the cBioPortal for Cancer Genomics. *Studies in Health Technology and Informatics* 2019;258:46–50.
- [9] nr23730. cbioportal-frontend - React Frontend of cBioPortal 2020. <https://github.com/nr23730/cbioportal-frontend> (accessed July 29, 2021).
- [10] Reimer N, Unberath P, Busch H, Ingenerf J, FhirSpark – Implementing a Mediation Layer to Bring FHIR to the cBioPortal for Cancer Genomics. In: Mantas J, Stoicu-Tivadar L, Chronaki C, Hasman A, Weber P, Gallos P, et al., editors. *Studies in Health Technology and Informatics*, IOS Press; 2021.
- [11] Merkel D. Docker: lightweight Linux containers for consistent development and deployment. *Linux J* 2014;2014:Article 2.
- [12] buschlab. MIRACUM-cbioportal 2020. <https://github.com/buschlab/MIRACUM-cbioportal> (accessed July 29, 2021).
- [13] Memorial Sloan Kettering Cancer Center. Genome Nexus - Annotation and Interpretation of Genetic Variants in Cancer n.d. <https://www.genomenexus.org/> (accessed July 28, 2021).
- [14] Memorial Sloan Kettering Cancer Center. Run cBioPortal using Docker Compose 2020. <https://github.com/cBioPortal/cbioportal-docker-compose> (accessed July 28, 2021).
- [15] Haarbrandt B, Schreiwies B, Rey S, Sax U, Scheithauer S, Rienhoff O, et al. HiGHmed – An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. *Methods Inf Med* 2018;57:e66–81.
- [16] Winter A, Stäubert S, Ammon D, Aiche S, Beyan O, Bischoff V, et al. Smart medical information technology for healthcare (SMITH). *Methods of Information in Medicine* 2018;57:e92–105.
- [17] WildFly. Keycloak - Open Source Identity and Access Management n.d. <https://www.keycloak.org/> (accessed July 28, 2021).
- [18] AG-Boerries. MIRACUM-Pipe 2019. <https://github.com/AG-Boerries/MIRACUM-Pipe> (accessed July 29, 2021).
- [19] Ustjanzew A, Marini F. cbpManager: Generate, manage, and edit data and metadata files suitable for the import in cBioPortal for Cancer Genomics. <https://arsenij-ust.github.io/cbpManager/index.html>; 2021.