Searching of Clinical Trials Made Easier in cBioPortal Using Patients’ Genetic and Clinical Profiles

Philipp Unberath¹ Lukas Mahlmeister¹ Niklas Reimer² Hauke Busch² Melanie Boerries³,4,* Jan Christoph¹,5,*

³Friedrich-Alexander University Erlangen-Nuremberg, Chair of Medical Informatics, Erlangen, Bayern, Germany
⁴University of Freiburg Faculty of Medicine, Institute of Medical Bioinformatics and Systems Medicine, University Medical Center Freiburg, Freiburg, Baden-Württemberg, Germany
⁵Martin-Luther-University Halle-Wittenberg, Faculty of Medicine, Junior Research Group (Bio-)Medical Data Science, Halle, Sachsen-Anhalt, Germany

Address for correspondence Philipp Unberath, MSc, Friedrich-Alexander University Erlangen-Nuremberg, Chair of Medical Informatics, Wetterkreuz 15, Erlangen, Bayern 91058, Germany (e-mail: philipp.unberath@fau.de).

Abstract

Background Molecular tumor boards (MTBs) cope with the complexity of an increased usage of genome sequencing data in cancer treatment. As for most of these patients, guideline-based therapy options are exhausted, finding matching clinical trials is crucial. This search process is often performed manually and therefore time consuming and complex due to the heterogeneous and challenging dataset.

Objectives In this study, a prototype for a search tool was developed to demonstrate how cBioPortal as a clinical and genomic patient data source can be integrated with ClinicalTrials.gov, a database of clinical studies to simplify the search for trials based on genetic and clinical data of a patient. The design of this tool should rest on the specific needs of MTB participants and the architecture of the integration should be as lightweight as possible and should not require manual curation of trial data in advance with the goal of quickly and easily finding a matching study.

Methods Based on a requirements analysis, interviewing MTB experts, a prototype was developed. It was further refined using a user-centered development process with multiple feedback loops. Finally, the usability of the application was evaluated with user interviews including the thinking-aloud protocol and the system usability scale (SUS) questionnaire.

Results The integration of ClinicalTrials.gov in cBioPortal is achieved by a new tab in the patient view where the genomic profile for the search is prefilled and additional parameters can be adjusted. These parameters are then used to query the application programming interface (API) of ClinicalTrials.gov. The returned search results

Keywords ► clinical decision support ► genomics ► clinical trial ► molecular tumor board ► trial matching ► genetic alteration

© 2022. The Author(s).
This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (https://creativecommons.org/licenses/by-nc-nd/4.0/)

THIEME Research Article 363
Published online: 2022-03-30
Background and Significance

Genome sequencing has become more cost-effective and feasible due to technical developments such as next-generation sequencing (NGS) methods.\(^1,2\) Oncology, in particular, benefits from these developments, as sequencing of the cancer genome can provide extended molecular tumor profiling, as well as valuable information on potential personalized treatment strategies.\(^3\) This consideration has led to the development of a variety of targeted therapy approaches.\(^4-6\)

Since the same targetable mutations can occur in different tumor entities, more and more clinical trials are being conducted as so-called basket studies which include tumor entities based on its specific mutation or biomarker.\(^7,8\) This leads to a much wider spread of tumor characterizations and therefore also to a reduction of matching patients and corresponding case numbers.\(^9\)

To cope with the complexity of molecular-based cancer treatment, many institutions have implemented molecular tumor boards (MTBs) complementing the already existing organ-based tumor boards.\(^10\) These MTBs are composed of interdisciplinary experts reviewing and discussing the complex personalized therapy options based on clinical and advanced molecular diagnostics. For treatment recommendations, it is crucial to find accessible clinical trials that match the genetic profile of the patient’s tumor to provide patient the opportunity to participate in these trials, or to include findings from completed trials. Especially in the last few years, it has been possible to expand the therapy guidelines based on genetic tumor profiling, as shown for example for lung carcinoma by Lindeman et al.\(^11\) However, mainly due to the increase in advanced diagnostics using next generation sequencing, there are still other unknown genetic findings where no guided line therapy is currently available.

However, the search for suitable clinical trials is often performed manually, and therefore, it is a time-consuming process.\(^12\) One tool that can be employed to use bioinformatic and clinical resources to determine the clinical relevance of gene alterations for the MTB is the open-source platform cBioPortal developed by the Memorial Sloan Kettering Cancer Center.\(^13,14\) It allows analysis of clinical and molecular patient data of cancer patients and can be accessed as a public instance or deployed locally enabling customization of the portal and analysis of own patient data. Buechner et al.\(^15\) performed an extensive requirements analysis and specification of an MTB platform in which they found cBioPortal as a suitable basis for such an application. One of the identified still missing features for the use of cBioPortal in the clinical setting of an MTB was facilitating the search of suitable clinical studies for patients.

Utilizing cBioPortal as a local clinical and genomic data warehouse, there were already efforts to integrate clinical trials matching based on the genetic profile and clinical patient data. This is achieved by the integration of Match-Miner\(^16\) that matches the patient’s genetic profile to clinical trials that have previously been curated by domain experts and converted into a structured proprietary format. However, the curation constitutes a time-consuming process, therefore resulting in a relatively small number of available studies and a constant effort to keep the data up to date. Other solutions to find matching studies based on genetic characteristics are either proprietary, such as My Cancer Genome,\(^17\) or single-institution, such as TrialProspector,\(^18\) and therefore not publicly available.

Objectives

In this work, we developed an extension to cBioPortal to enable patient-centric clinical trials search. It is based on the integration of ClinicalTrials.gov and leverages cBioPortal’s structured molecular and clinical data which can be used to prefill the search fields. The main requirement was that the search for clinical trials should not require any manual curation or preprocessing of trial data and should be executed on the fly from the fronted of cBioPortal. It should aid MTB participants during the often laborious process of searching clinical trials for MTB patients, and therefore support an important one of their routine tasks. To account for the specific needs and requirements of the MTB participants, a prototype was implemented using a user-centered design approach including multiple feedback loops and evaluated in the context of a usability test.

Methods

Buechner et al.\(^15\) revealed the need of the MTB for the previously described extension in their requirements analysis and also already created a high-fidelity mockup for the design thereof. Adding on their findings, an in-depth requirements analysis for the specifics of the extension using open expert interviews was conducted. The interviews for this requirements analysis were performed in the course of a biweekly journ fixe about the advancement of processes and software in MTBs with over 20 regularly participating clinicians and medical informatics specialists of nine German partner sites. Based on the mockup (Fig. 1) and the resulting detailed requirements, a prototype was implemented using a user-driven development process with a total of three iterations over a period of 2 months. During this process, the
The high-fidelity mockup of the frontend integration of the ClinicalTrials.gov search as proposed by Buechner et al.15 It is derived from a screenshot of the patient view, adding the studies search tab using an image editing for the purpose of better demonstrating the feature. The mockup does not provide any functionality, but it served as a blueprint for the prototype.

The prototype was further refined iteratively, taking into account the feedback of the real-world users during the aforementioned jour fixes, particularly oncologists who provide MTB therapy recommendations for patients. This feedback of the prototype was collected by allowing users to try out the prototype and report any problems, ambiguities, or missing features during the jour fixes or afterward as written feedback. The prototype used real patient data from the cBioPortal public database and real study data from ClinicalTrials.gov.

To evaluate the extension, a formative usability test (the summative evaluation will be performed in a future study along with multiple other features extending cBioPortal for its use in MTBs) was conducted with real-world users. The call to participate in the usability test went to all potential probands, that is, 10 clinicians regularly attending MTBs and the aforementioned jour fixes. Five of these participants of the MTB were asked to perform a clinical trial search for a patient19 using the newly implemented extension as the scenario for the usability test. During the test process, the participants were instructed to review the patient’s genetic alterations and clinical data and select fitting data to feed into the search tool. In the first step of the search, the users were asked to prioritize studies the patient could realistically enroll in, most importantly meaning studies close to the patient’s place of residence. The second step should also include worldwide studies from which the user may only derive new therapy approaches, as enrollment of own patients would not be practical due to the place of residence. During the whole process, the user was asked to verbalize all thoughts (think-aloud protocol). Afterward, the participant was interviewed using open-style and Likert’s scale20 questions and finally asked to complete the system usability scale (SUS)21 questionnaire in German22 which is the native language of all participants. The usability tests were performed as web meetings using the teleconferencing software Zoom, taking approximately 30 to 60 minutes, with screen sharing enabled which allowed the recording of the users’ interactions with the user interface, as well as all feedback during the think-aloud protocol. These recordings were then systematically analyzed. All documented errors, difficulties, and hesitations from the usability perspective during the think-aloud protocol, as well as questions and suggestions for improvement from the open interview were grouped into categories and summarized to single change requests after collection.

Results

The detailed requirements analysis revealed that the search results should be displayed as a table within a new tab of the patient view of cBioPortal (~Fig. 2). The adjustable search parameters should consist of the following parameters: (1) variants and copy number variations (CNVs), (2) generic free-text keywords, (3) recruiting statuses of the study, (4) trial location by country, and (5) the age and sex of the patient. To be able to review the proposed trial results as fast as possible, the table should include (1) the recruiting status, (2) found keywords, (3) the title of the study, (4) conditions/diagnoses, (5) interventions, and (6) locations of partner sites. During the iterative feedback rounds with real-world users, prior requirements were adapted and refined and the prototype could be further developed. Multiple comments were related only to the user interface, for example, the way of entering study locations, how to select relevant alterations, or...
improving the explanation of fields using tooltips. But there were also suggestions requiring adaptations of the general requirements and therefore changes to multiple layers including the user interface and the logical handling of user input that affects the query of clinical trials. First, it should also be possible to differentiate between required and optional keywords and an option to specify whether they should be combined with "AND" or "OR". Second, the location search should be improved to not only enable a search by country but a proximity search using a city as input.

The aforementioned search parameters are then used without further modification or enrichment to build a request to query the application programming interface (API) of ClinicalTrials.gov. Scoring of the search results was achieved first filtering the results (if specified) by country, recruiting status, and selected alterations. Afterward, the remaining trials were ranked by a combined weighted query based on feedback of the users during the iterative feedback rounds which consisted of (importance in descending order) count of found keywords, matching tumor entity, distance to closest partner site, matching sex, and matching age.

Next, different methods of integrating the extension were considered, including integration in the cBioPortal frontend (running in the user’s browser), the cBioPortal backend (running on a dedicated server), and as a standalone service, running on the server in addition to the cBioPortal backend (for more information about the architecture of cBioPortal, refer to reference Unberath et al24). Since cBioPortal does not provide a plugin concept, the extension should be coupled as loosely as possible to the codebase to ensure better compatibility with future updates. This indicates an implementation in the frontend which also promotes the feature of retrieving search results on the fly. Since results are not cached, curated, or preprocessed in advance by the server, the display of results should be focused on ranking trials according to the number of matching criteria, listing results with less matches at the end rather than excluding them completely. This leaves the final assessment of the relevance of all found studies to the user. Another key feature of the extension stems from its usage in a clinical setting with real patients, the proximity search, which considers the patient’s place of residence and ranks the studies with matching geographical sites higher accordingly.

Finally, the concluding usability evaluation was performed in October 2020 with five clinicians with different backgrounds (oncology, systems medicine, or bioinformatics), who regularly participate in MTBs. Several additional shortcomings were identified that were not identified during the feedback rounds and needed to be addressed for finalizing the tool. Most importantly, it was not clear to the users that by design the tumor entity was invariably part of the ranking of the tool in the background, as it is always present in cBioPortal samples. All users therefore demanded that ranking parameters should be made selectable and customizable for the query. Other important feedback included improving the explanation of how search results were filtered and ranked and how to enter the optional and required keywords (each mentioned by three participants). The participants also mentioned additional nice to have features, such as the ability to search for the phase of the study, for interventions or alterations based on pathways (each mentioned by one participant). The latter can be considered as a suggestion for improvement for further developments.

Fig. 2 The prototype of the newly developed tab in the cBioPortal patient view of the ClinicalTrials.gov search.
The results of the quantifiable questions of the usability evaluation can be found in Table 1. The evaluation of the SUS questionnaire resulted in an overall SUS score of 83.5 (number of participants was five, standard deviation = 1.158, individual scores were 92.5, 95, 90, 65, and 75, respectively). The SUS score of 83.5 is in the middle of the “acceptable” range and represents a “good” on the adjective rating scale.25

### Table 1 The results of the quantifiable questions of the usability evaluation

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>The function of each field of the search mask was comprehensible.</td>
<td>3.80</td>
<td>0.75</td>
</tr>
<tr>
<td>The information provided in the search results helped to identify suitable or unsuitable studies.</td>
<td>3.80</td>
<td>0.75</td>
</tr>
<tr>
<td>The order in which the studies are displayed was helpful.</td>
<td>3.25</td>
<td>1.01</td>
</tr>
<tr>
<td>The application overall facilitates the search for suitable studies.</td>
<td>4.00</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Abbreviation: SD, standard deviation. Note: Scores range from 1 (strongly disagree) to 5 (strongly agree).

The results of the quantifiable questions of the usability evaluation can be found in Table 1. The evaluation of the SUS questionnaire resulted in an overall SUS score of 83.5 (number of participants was five, standard deviation = 1.158, individual scores were 92.5, 95, 90, 65, and 75, respectively). The SUS score of 83.5 is in the middle of the “acceptable” range and represents a “good” on the adjective rating scale.25

### Discussion

Clinical trials are a major driver in the advancement of cancer treatment and while the willingness of patients to participate in trials is estimated as high as 70%, less than 5% of cancer patients enroll in clinical trials.26 Unger et al identified multiple barriers in different categories from which this work focuses on the structural barrier of identifying available clinical trials. Although being resource intensive, the process of searching and identifying appropriate trials is a crucial part of the preparation of patients for the MTB, as these patients usually already have exhausted all guideline-based treatment options or have rare tumors.30,31

While there are other works regarding software solutions for MTBs, these developments are still in early stages. Halfmann et al developed an MTB support tool also using a user-driven development process. Their tool focuses on the presentation part of an MTB and not on the preparation including the search for clinical trials for MTB patients and furthermore, the study did not include a usability evaluation. Fegerel et al are currently developing a solution focusing on the administrative part including communication methods for performing virtual tumor boards. Studies in the field of recruitment support often tackle different aspects than searching fitting trials for a specific patients, for example, optimize eligibility criteria or assess trial population representativeness, supporting data collection or helping to find patients for screening.

A key factor in enrollment of MTB patients in clinical trials is matching genomic alterations of the patient’s tumor and the study’s inclusion criteria, yet these alterations are not specified in a well-structured matter and often only found study’s descriptive text.39 Therefore, several attempts have been made to curate precision cancer studies from trial registries like ClinicalTrials.gov or institution-specific registries to generate a structured database to query. The drawback of these approaches generally is the high maintenance of curating trials and keeping the data up to date.40 Additionally, institution-specific systems, like MatchMiner or the Phase One Spot Tracker (POST), do not publish curated datasets. While those data would not directly benefit the enrollment at other institutions (depending on the proximity), they would constitute a valuable resource for benchmarking or validating the results of new matching tools. However, such a validation was not in the scope of this work and was therefore not performed.

The focus was on streamlining the process of the manual search by integrating ClinicalTrials.gov into cBioPortal and prefilling data already available for the patient.

In contrast to aforementioned tools, this work focuses on finding studies by directly using the necessary information for the search from the patient data already available in cBioPortal. It is obvious that such an approach can never achieve sensitivity and specificity levels of tools that require upfront efforts from domain experts. Therefore, this tool focuses on a high usability to enable physicians to query up to date and publicly available trial registries (currently solely ClinicalTrials.gov) on the fly as fast and straightforward as possible. The tool does have technical limitations, some of which originating from the API of ClinicalTrials.gov. For instance, while the advanced search of ClinicalTrials.gov web interface features a practical distance search, the API lacks this feature. Thus, the extension implements an own distance search using a set of predefined locations which needs to be further expanded. For the search of synonyms rather than exact string matches, the API provides the same functionality as the web interface, utilizing the Medical Subject Headings (MeSH) to search for all synonyms of the given term. Currently, the extension uses the HUGO Gene Symbol from cBioPortal for querying the API, but in the context of molecular profiling, it could be beneficial to also include more specialized vocabularies, such as Gene Ontology, for a broader synonym coverage, and using gene identifiers as HUGO symbols may change over time.44 Future developments could also consider pathways of the patient’s genes affected by mutations.

One approach to improve the fit of our tool’s search results would have been to preprocess or automatically curate trials from ClinicalTrials.gov in advance. Previous studies shown that even sophisticated curation pipelines require manual post-treatment.45 This fact also discouraged the implementation of an extract, transform, load (ETL)-pipeline from ClinicalTrials.gov to the MatchMiner trial format Clinical Trial Markup Language (CTML). Preprocessing trials in the form of building a custom registry would have also required a solution that is server-, for example either cBioPortal backend- order standalone-tool-based. As one of our main goals was to contribute our extension to the public cBioPortal open-source project, solutions requiring additional computing resources server-sided or manual data updates were not feasible. With the chosen frontend integration of the search tool, all computing is performed in the browser of the end user.

While developing our extension, we focused on a user-centered design with multiple feedback loops to be able to...
analyze and then support the workflow of the physicians as much as possible. This greatly improved the design and development process as most of the times, translating the requirements of the real-world users to technical features implemented in software is not trivial and should be reviewed and revised when necessary. In summary, the findings of the evaluation including the SUS score of 83.5 showed overall, the tool facilitates the search for matching trials (∼ Table 1).

The extension described in this work is only one of the needed functionalities identified by Buechner et al15 to use cBioPortal as an MTB platform. The search for clinical trials will be refined using the results of the usability test and gradually other key features will be additionally implemented to supply MTB participants with a comprehensive solution for their routine tasks. The completed prototype for the MTB platform will then be extensively evaluated in a future study.

Limitations
The final usability evaluation was intended to be the last, more detailed round of feedback in the development process, rather than a summative evaluation of a final product. However, the five participants match the often cited threshold to identify the most serious usability issues46 and provided valuable feedback from real-world users regularly participating in MTBs. The evaluation provides an indication of the usability of the application but needs to be confirmed after the tool has been finalized during the aforementioned extensive evaluation of the completed MTB platform. In particular, the large deviation of SUS scores indicates usability issues that may only apply to a subgroup of users and should be further investigated.

Conclusion
Using a user-centered design process, we developed an integration of the ClinicalTrials.gov registry with cBioPortal. The architecture of the integration is as lightweight as possible, only being coupled with the cBioPortal frontend and requiring no additional server-sided resources. The final evaluation showed a good overall usability and that the tool can assist physicians to find appropriate studies for individual patients in the preparation of MTBs with less manual effort.

Clinical Relevance Statement
This work proposes a tool for searching clinical trials based on molecular alterations in tumors of cancer patients. This facilitates the process of preparing patient data and discussing therapy recommendations in the clinical setting during Molecular Tumor Boards.

Multiple Choice Questions

1. How was the search tool integrated?
   a. In the backend of cBioPortal
   b. In the frontend of cBioPortal
   c. As a standalone service
   d. As a combination of frontend and backend integration

Correct Answer: The correct answer is option b. The integration of the tool in cBioPortal is frontend-based. As cBioPortal does not provide a plugin concept, implementing the tool in the frontend achieves a loose coupling to the codebase and better compatibility with future updates.

2. What was the main requirement of application?
   a. Transforming trials to a genomics-optimized data format
   b. Facilitate the curation of trial data by experts
   c. Searching trials without the need for manual curation in advance
   d. Enabling patients to search trials in which they could participate

Correct Answer: The correct answer is option c. The main requirement was that the search for clinical trials should not require any manual curation or preprocessing of trial data and should be executed on the fly. The application should be used by experts in or during the preparation of an MTB.

Protection of Human and Animal Subjects
The project was performed in compliance with the World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects. Ethical approval was not required.

Funding
This research has been conducted within the MIRACUM project. MIRACUM is funded by the German Federal Ministry of Education and Research (BMBF), grant IDs 01ZZ1801A (P.U., L.M., and J.C.) and 01ZZ1801B (M.B.). This work was supported by the BMBF funded HiGHmed project, grant ID 01ZZ18022 (H.B., N.R.) and by the German Research Foundation (DFG), grant IDs CRC 850 and CRC 1160 (M.B.).

Conflict of Interest
None declared.

Acknowledgments
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 Service RF. Gene sequencing. The race for the $1000 genome. Science 2006;311(5767):1544–1546