Availability and access restrictions of cancer patient sequencing profiles for developing precision oncology methods

bachelor thesis exposé

Anna Göing

3rd October 2022

1 Introduction

Future medicine is most likely connected with individual treatment methods that are based on the analysis of a single persons physiology, biology and environmental settings, which require enormous amounts of filtering and careful selection of relevant features, to support specified medical personnel in making a decision. With the growing availability of all types of medical data, for instance from smart devices or improved diagnostic methods, the interpretation becomes very time consuming and hereby allows details to be unintentionally not being taken into account albeit being potentially relevant [Zug+16]. Especially when it comes to data that is not easily interpreted by humans, such as DNA sequencing profiles, a more automated approach that compares selected information with a reference database to identify similar cases of the past could be one way to support future medical decisions [Per+18].
In the field of precision oncology, a select number of late-stage cancer patient receive a treatment suggestion based on a number of factors, including of molecular data obtained from some form of DNA sequencing with the hope of improving the outcome of cancer patients [Le +12]. a database of such information could be used to train and test a machine learning algorithm that would predict a certain drug for an individual cancer patient. As such details are scattered around specific papers concerning all sorts of cancer variants, a large proportion of research time even in specialised institutions is taken up by merely comparing and acquiring published cancer treatment resources [Per+18].

2 Goals

This thesis will find out whether there are any cancer patient data freely available or accessible under restrictions that include relevant information to develop precision oncology methods. The result will present the accessibility of data, as well as all relevant key details regarding the corresponding papers.
The main focus hereby lies on listing variant profiles and the sequencing technology used to retrieve them, as well as the type of tumour identified. The resulting suggestion of therapy should be brought up by a Molecular Tumour Board or a similar designated unit to ensure a harmonised decision making process along all data found.

3 Related Work

Most of cancer patients that are cared for in clinical institutions receive treatment based on standardised procedures that check common biomarkers for general target therapy [22]. One important aspect of precision oncology however are a group of specialised people of a clinical institution, that come together to form a Molecular Tumour Board (MTB) and contribute their respective expertise. They discusses presented late-stage cancer patient cases based on annotated and analysed DNA variants [Lag22] to find an individual cancer treatment drug outside of standardised approaches that might prolong life [Le +12]. Whilst it has been shown, that across different MTBs, the same patient would not receive a similar treatment proposal [Rie+18], other studies have concluded that this might be due to different workflows and have suggested an optimised setup for better data-sharing [Rao+20]. Next to supporting the MTB itself, the actual prioritisation in analysis of gene-variants is also an issue to deal with, as different MTBs treat variants with varying importance [Lei+19]

During recent years, numerous Knowledge Bases (KB) of anonymous patient data and applied treatment procedures with ranging degrees of detail have been generated and compared to rate their usability, finding that albeit including different specialisations of certain cancer-types, redundant overlap is present [Pal+19]. The resulting attempt to harmonise such KBs in order to condense information and create a meta-knowledge-base for better retrieval has struggled to even transform variant names due to different tumour systems [Wag+20]. However it has been shown, that not only algorithms can be developed to match a patients genomics sequence to a known treatment method [Per+18], but also that if a meta-KB is used, the result is likely to improve [Wag+20].

4 Methodology

To answer the question of availability of such variant profiles a search through curated online resources, such as *PubMed* will be performed. Whilst starting with the above mentioned papers and following any references and citations, the search may lead to the actual, anonymous patient data used. By the help of *GoogleScholar*, cross-links can be successfully identified and depending on the retrieved papers might result in a quickly expanding set of potential papers, that could add to the pool of available patient data.
In order to keep track of any papers visited and store the potential information for future use, a table will be set up, containing one paper reference per row and all relevant information. Next to the *DOI* for quick access to any paper listed, the date of analysis will be important to compare for potential content updates, as well as the actual information of the study, stating whether the used data set is freely accessible, can be retrieved by fulfilling certain requirements or is not available at all. Potentially useful values for future use-cases will be the number of patients present in the study and type of cancer present, whilst making sure to agree on specific abbreviations or determining cancer categories beforehand, as well as the type of sequencing technology used. In case the paper possibly includes data on the treatment selected by an MTB, this information will be included, as well as potential details whether the drug was actually prescribed and how well it worked.
Multiple papers might relate to the same set of data and describe the ongoing process of treatment prediction, application and evaluation, but the table will be set up to only contain one paper per row. In order to quickly find all corresponding papers to one case, a special index will be used to link those different papers together, whilst not having to condense all information in one row and omitting paper credit at the same time. In this case, blanks will be filled with a

special symbol to inform that the missing information is completed by another paper.

It should be noted, however, that during the actual progress of this thesis, the specific order or content of the table is subject to change to suit the challenges faced and keep a clean and easily accessible structure for future data usage.

5 Results

It cannot be precisely predicted what this literature search will produce and how the results could be evaluated. However depending on the amount of papers found, a semantic integration of data to harmonise any used IDs of genes or variants would enhance the future usage of the information, whilst an analysis with descriptive statistics could be done in case of a great quantity of data.
Regardless of how little rows of the finale table will include freely available patient data, the result of this study might prove potential for future research in this area. The mere absence of any free patient data would be a great concern and restriction for any possible development of new tools, that either pre-analyse or filter patient data for MTBs or even suggest a treatment based on curated data, which do not have access to such data from inside the system.

References

[22] 'S1-Leitlinie: Tumorgenetik – Diagnostik im Kontext maligner Erkrankungen'. In: *Medizinische Genetik* 34.1 (31st May 2022), pp. 53–68. ISSN: 1863-5490, 0936-5931. DOI: 10.1515/medgen-2022-2112.

[Lag22] Alessandro Laganà, ed. *Computational Methods for Precision Oncology*. Vol. 1361. Advances in Experimental Medicine and Biology. Cham: Springer International Publishing, 2022. ISBN: 978-3-030-91835-4 978-3-030-91836-1. DOI: 10.1007/978-3-030-91836-1.

[Le +12] Christophe Le Tourneau, Maud Kamal, Olivier Trédan, Jean-Pierre Delord, Mario Campone et al. 'Designs and challenges for personalized medicine studies in oncology: focus on the SHIVA trial'. In: *Targeted Oncology* 7.4 (1st Dec. 2012), pp. 253–265. ISSN: 1776-260X. DOI: 10.1007/s11523-012-0237-6.

[Lei+19] Jonas Leichsenring, Peter Horak, Simon Kreutzfeldt, Christoph Heining, Petros Christopoulos et al. 'Variant classification in precision oncology'. In: *International Journal of Cancer* 145.11 (2019). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijc.32358, pp. 2996–3010. ISSN: 1097-0215. DOI: 10.1002/ijc.32358.

[Pal+19] Steffen Pallarz, Manuela Benary, Mario Lamping, Damian Rieke, Johannes Starlinger et al. 'Comparative Analysis of Public Knowledge Bases for Precision Oncology'. In: *JCO Precision Oncology* 3 (Dec. 2019). Publisher: Wolters Kluwer, pp. 1–8. DOI: 10.1200/PO.18.00371.

[Per+18] Júlia Perera-Bel, Barbara Hutter, Christoph Heining, Annalen Bleckmann, Martina Fröhlich et al. 'From somatic variants towards precision oncology: Evidence-driven reporting of treatment options in molecular tumor boards'. In: *Genome Medicine* 10.1 (15th Mar. 2018), p. 18. ISSN: 1756-994X. DOI: 10.1186/s13073-018-0529-2.

[Rao+20] Shruti Rao, Beth Pitel, Alex H. Wagner, Simina M. Boca, Matthew McCoy et al. 'Collaborative, Multidisciplinary Evaluation of Cancer Variants Through Virtual Molecular Tumor Boards Informs Local Clinical Practices'. In: *JCO Clinical Cancer Informatics* 4 (Nov. 2020). Publisher: Wolters Kluwer, pp. 602–613. DOI: 10.1200/CCI.19.00169.

[Rie+18] Damian T. Rieke, Mario Lamping, Marissa Schuh, Christophe Le Tourneau, Neus Basté et al. 'Comparison of Treatment Recommendations by Molecular Tumor Boards Worldwide'. In: *JCO Precision Oncology* 2 (Nov. 2018). Publisher: Wolters Kluwer, pp. 1–14. DOI: 10.1200/PO.18.00098.

[Wag+20] Alex H. Wagner, Brian Walsh, Georgia Mayfield, David Tamborero, Dmitriy Sonkin et al. 'A harmonized meta-knowledgebase of clinical interpretations of somatic genomic variants in cancer'. In: *Nature Genetics* 52.4 (Apr. 2020). Number: 4 Publisher: Nature Publishing Group, pp. 448–457. ISSN: 1546-1718. DOI: 10.1038/s41588-020-0603-8.

[Zug+16] Jon Zugazagoitia, Cristiano Guedes, Santiago Ponce, Irene Ferrer, Sonia Molina-Pinelo et al. 'Current Challenges in Cancer Treatment'. In: *Clinical Therapeutics* 38.7 (1st July 2016). Publisher: Elsevier, pp. 1551–1566. ISSN: 0149-2918, 1879-114X. DOI: 10.1016/j.clinthera.2016.03.026.