Opening the Personal Gate between Technology and Health Care
M. Giacomini et al. (Eds.)

doi:10.3233/SHTI260700

Findings from the PM4Onco Workshop on Determining Patient Similarity Search Strategies for Personalized Therapy in Molecular Tumor Boards

Iryna MANUILOVA[a,1], Annemarie B. WEISE [a,] Julia FOLWARK [a], Sonja HIEMER[b], Susann SCHULZE[b], Manuela BENARY[c], Arsenij USTJANZEW[d], Cosima STRANTZ[e], Philipp UNBERATH[f], Silke D. WERLE[g], Benedikt Brors[h], Jan CHRISTOPH [a,e]

[a] *Junior Research Group (Bio-)Medical Data Science, Faculty of Medicine, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany*
[b] *Krukenberg Cancer Center Halle (Saale), Halle (Saale), Germany*
[c] *Charité Comprehensive Cancer Center, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany*
[d] *Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany*
[e] *Friedrich-Alexander-Universität Erlangen-Nürnberg, Institute for Medical Informatics, Biometrics and Epidemiology, Medical Informatics, Erlangen, Germany*
[f] *Medical Center for Information and Communication Technology, Universitätsklinikum Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*
[g] *Institute of Medical Systems Biology, Ulm University, Ulm, Germany*
[h] *Division Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany*

ORCiD ID: Iryna Manuilova 0009-0005-8821-1471, Annemarie B.Weise 0009-0003-6925-5941, Julia Folwark 0009-0006-2092-7964, Sonja Hiemer 0009-0004-2886-9485, Susann Schulze 0000-0002-3856-7956, Manuela Benary 0000-0002-3067-2030, Arsenij Ustjanzew 0000-0002-1014-4521, Cosima Strantz 0009-0007-3980-8233, Phillip Unberath 0000-0002-1269-9360, Silke D. Werle 0000-0002-5153-0269, Benedikt Brors 0000-0001-5940-3101, Jan Christoph 0000-0003-4369-3591

Abstract. The PM4Onco project advances precision oncology by developing analytical and visual methods to support data-driven therapy decisions in Molecular Tumor Boards (MTBs). An interdisciplinary workshop at Martin Luther University Halle-Wittenberg in March 2025 explored strategies for defining and integrating patient similarity into MTB workflows. Using a synthetic oncology dataset, participants with clinical and data science backgrounds defined similarity criteria and structured search strategies for patient similarity. The workshop outcomes provide structured decision frameworks that operationalize patient similarity reasoning in a

[1] Corresponding Author: Iryna.Manuilova@uk-halle.de

reproducible manner. Their implementation supports transparent and evidence-based analysis, fostering integration between algorithmic modeling and clinical expertise.

Keywords. Patient similarity, molecular tumor board, personalized oncology

1. Introduction

The PM4Onco project advances precision oncology by developing computational methods that support data-driven therapy decisions. Within this framework, the Patient Similarity (PatSim) task develops analytical and visual approaches to identify clinically comparable patients and to integrate similarity reasoning into the Molecular Tumor Board (MTB) workflow. While MTBs translate molecular and clinical data into personalized therapy recommendations, discussions often remain case-specific [1]. PatSim addresses this limitation by linking prior therapeutic experiences to current cases, bridging clinical expertise and algorithmic reasoning. An interdisciplinary workshop, building on the frameworks proposed in [2] and [3], explored these concepts. This manuscript summarizes the workshop outcomes and outlines foundations for integrating PatSim into MTBs.

2. Methods

An interdisciplinary workshop involving 17 clinicians, bioinformaticians, and data scientists from 13 German institutions to define domain-specific reasoning patterns for PatSim and translate them into methodological strategies for algorithmic development. The workshop comprised two phases. In Phase 1 (Exploratory Similarity Assessment), participants independently identified the five most similar patients to a given index case from 20 synthetic profiles without predefined criteria followed by group discussions to compare disciplinary reasoning approaches. As supportive background material, participants were provided with previously defined example groupings based on common similarity metrics (e.g., Euclidean, Cosine, Jaccard, and Gower), which served as contextual reference but were not applied as decision rules. In Phase 2 (Conceptual Strategy Development), mixed-discipline teams developed structured search strategies and flowcharts outlining hierarchical decision pathways for PatSim determination. They distinguished core determinants from contextual modifiers and formalized logical dependencies between parameters. Each workflow was applied to a cBioPortal-generated dataset [4].

3. Results

3.1. Conceptual and Hierarchical Structure of Similarity Criteria

Participants established a hierarchical conceptual framework for defining PatSim within the MTB context, integrating biological mechanisms, clinical context, and translational applicability according to therapeutic relevance. This led to the following concept: At its core, molecular alterations anchor the similarity assessment. Precise nucleotide- and amino acid-level annotation ensures mechanistic validity, as distinct variants can differently affect protein structure, signaling, and therapy response. This precision supports

variant-specific drug approvals and sequence-based models. Pathogenicity classification using databases such as ClinVar and COSMIC [5] further distinguishes actionable from uncertain variants. The tumor entity sets biological and regulatory boundaries. As the oncogenic effect of a mutation depends on cellular context, co-mutations, and the tumor microenvironment, evidence from the same histologic type is more reliable than cross-entity comparisons. Finally, the clinical context includes prior therapies, resistance mechanisms, and disease stage. These factors reshape the tumor's molecular profile through therapeutic pressure and clonal evolution, linking molecular data to dynamic patient trajectories and ensuring that computational similarity reflects therapeutic relevance.

3.2. Methodological Orientations in Similarity Assessment

Within this hierarchical framework, three methodological orientations emerged, representing a distinct analytical perspective and weightings of molecular, clinical, and systemic similarity (Figure 1).

1. Mutation-Centered Precision emphasizes variant-level specificity and mechanistic reproducibility, prioritizing exact nucleotide or amino acid substitutions and pathogenic classification aligned with regulatory logic linking targeted therapies to defined mutations. Computationally, it applies high-resolution similarity metrics based on sequence identity or mutational distance.

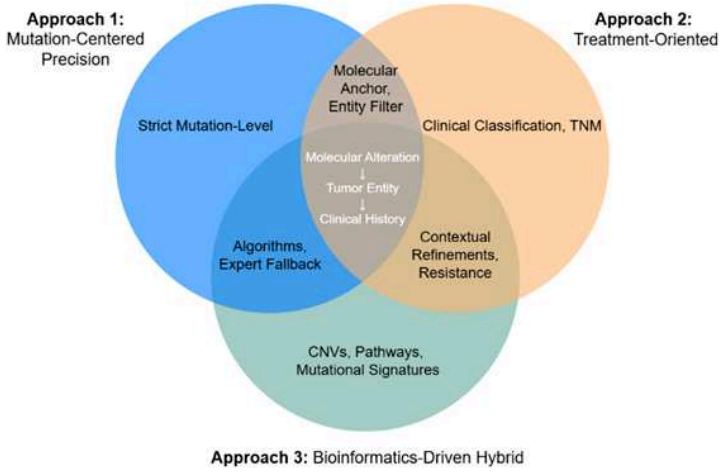

Figure 1. Comparative perspectives of methodological orientations in MTB similarity assessment.

2. Treatment-Oriented Approach extends similarity assessment beyond genomic identity toward therapeutic comparability. It defines similarity through shared treatment trajectories, disease stage, and acquired resistance mechanisms. By capturing therapy-driven tumor evolution, this approach emphasizes pragmatic comparability over molecular identity and embeds similarity modeling in real-world clinical decision-making.

3. Bioinformatics-Driven Hybrid Approach integrates molecular, clinical information within a unified computational framework. Moving beyond single variants, it includes copy-number changes, mutational signatures, and pathway-level dysregulations to cap-

ture network-level biological coherence. PatSim here emerges from functionally comparable molecular networks and signaling cascades, refined by comorbidities and physiological parameters. This model bridges molecular precision with systemic coherence.

3.3. Integration of Patient Similarity into MTB Decision Logic

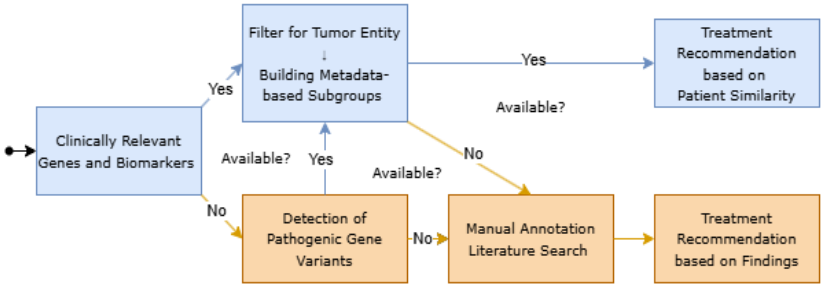

Figure 2. Conceptual decision workflow for therapy recommendations in the MTB use case, including molecular profiling, tumor-type filtering, and similarity analysis. In cases of incomplete data, pathogenicity assessment and literature review complement expert evaluation.

Based on the workshop findings, we defined a decision logic integrating PatSim reasoning into the MTB process (Figure 2). The workflow includes:
1. Extraction and annotation of molecular alterations define the core feature space.
2. Pathogenicity verification using reference databases to retain actionable variants.
3. Tumor-entity filtering to ensure biologically coherent comparisons.
4. Retrieval of precedent cases and literature evidence linking molecular similarity to therapeutic outcomes.
5. Evidence grading to quantify the strength of correspondence.
6. Clinical contextualization to align similarity with the individual patient trajectory.

4. Discussion and Conclusions

PatSim is often treated as a computational task in which features are empirically weighted and operationalized as numerical distance in a feature space**.** In oncology, algorithmic patient matching has mainly been discussed in terms of data availability and methodological challenges [6, 7]. By contrast, our framework embeds semantic and clinical logic into similarity assessment, prioritizing interpretability and clinician reasoning. Accordingly, participants applied a semantic hierarchy, recognizing molecular alterations as anchors. Biologically, this reflects the role of driver mutations as key determinants of disease behavior. Clinically, the tumor entity defines similarity. For instance, identical mutation in lung and breast cancer rarely implies comparable treatment relevance [8]. Therefore, computational PatSim framework must account for these nested dependencies to remain consistent with clinical reasoning. The workshop revealed three complementary orientations: mutation-centered precision, clinical-treatment focus, and bioinformatics-driven hybrid modeling. Together, they outline a multiscale perspective of precision oncology, spanning **v**ariant, pathway, and patient levels**.** Recognizing this complementarity is crucial for designing PatSim systems that are mechanistically faithful and clinically usable. PatSim cannot be reduced to algorithmic output alone. Computational similarity without interpretability risks irrelevance, while clinical intuition without

formalization lacks reproducibility. Hence effective PatSim requires bidirectional translation between clinical reasoning and algorithmic logic. The established six-step workflow captures the logic of translational oncology, from data extraction to clinical contextualization. Several limitations must be considered. The use of synthetic oncology data enabled controlled analysis but cannot reflect the heterogeneity of real-world MTB data. Synthetic datasets may also introduce implicit biases into similarity reasoning. In addition, the workshop-based consensus reflects expert judgment shaped by specific institutional and disciplinary practices. These factors may limit the generalizability of the proposed approach and indicate the need for validation using real-world patient data. Nevertheless, the framework makes similarity reasoning in MTB discussions explicit and reproducible, supporting transparency and consistency while preserving clinical judgment. In practical MTB settings, the workflow could serve as a basis for semi-automatic preselection of comparable patient cases. The six steps define explicit criteria and decision points for systematic filtering, while final interpretation and decision-making remain under clinician control. Validation could be performed using retrospective MTB cases to assess whether the workflow improves consistency and transparency of case selection. Overall, the framework provides a structured basis for integrating computational support into MTBs while complementing clinical expertise.

Acknowledgement

This research was funded by the German Federal Ministry of Research, Technology and Space through the PM4Onco project (01ZZ2322A-Y). The present work was performed in fulfillment of the requirements for obtaining the degree "Dr. rer. medic." at the Medical Faculty of the Martin Luther University Halle Wittenberg (IM).

References

[1] Pflugradt L, Metzger P, Gräßel L, Strantz C, Blaumeiser A, Busch H, et al. PM4Onco: personalisierte Medizin für die Onkologie – effektive Datennutzung für molekulare Tumorboards. Onkologie. 2024;30:898-904, doi: 10.1007/s00761-024-01593-0.

[2] Strantz C, Boehm D, Unberath P. Advancing usability and decision support for molecular tumor boards: insights from PM4Onco's visual analytics workshop. Stud Health Technol Inform. 2025;327:512-6, doi: 10.3233/SHTI250390.

[3] Seligson ND, et al. Recommendations for patient similarity classes: results of the AMIA 2019 workshop on defining patient similarity. J Am Med Inform Assoc. 2020;27(11):1808-12, doi: 10.1093/jamia/ocaa192.

[4] Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal for cancer genomics. Sci Signal. 2013;6(269):pl1, doi: 10.1126/scisignal.2004088.

[5] Park KJ. Clinical interpretation challenges of germline-shared somatic variants in cancer. Lab Med. 2022;53(1):24-9, doi: 10.1093/labmed/lmab020.

[6] Manuilova I, Bossenz J, Weise AB, Boehm D, Döbel M, Werle SD, et al. Uncovering the understanding of the concept of patient similarity in cancer research and treatment: scoping review. J Med Internet Res. 2025;27:e71906, doi: 10.2196/71906.

[7] Dai L, Zhu H, Liu D. Patient similarity: methods and applications. arXiv. 2020;arXiv:2012.01976, doi: 10.48550/arXiv.2012.01976.

[8] Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. The role of mutation catalogs in cancer genomics. Nat Rev Cancer. 2011;11(10):818-25, doi: 10.1038/nrc3133.