



HAL
open science

High-level Biomedical Data Integration in a Semantic Knowledge Graph with OncodashKB for finding Personalized Actionable Drugs in Ovarian Cancer

Johann Dreo, Sebastian Lobentanzer, Ekaterina Gaydukova, Marko Baric, Altti Ilari Maarala, Taru Muranen, Jaana Oikkonen, Federico Bolelli, Vittorio Pipoli, Veli-Matti Isoviita, et al.

► To cite this version:

Johann Dreo, Sebastian Lobentanzer, Ekaterina Gaydukova, Marko Baric, Altti Ilari Maarala, et al.. High-level Biomedical Data Integration in a Semantic Knowledge Graph with OncodashKB for finding Personalized Actionable Drugs in Ovarian Cancer. *Cancer Genomics, Multiomics and Computational Biology*, European Association for Cancer Research, Apr 2024, Bergame, Italy. hal-04509599

HAL Id: hal-04509599

<https://hal.science/hal-04509599v1>

Submitted on 18 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Title: High-level Biomedical Data Integration in a Semantic Knowledge Graph with OncodashKB for finding Personalized Actionable Drugs in Ovarian Cancer

Johann Dreo^{1,2}, Sebastian Lobentanzer³, Ekaterina Gaydukova¹, Marko Baric¹, Altti Ilari Maarala⁵, Taru Muranen⁵, Jaana Oikkonen⁵, Federico Bolelli⁷, Vittorio Pipoli^{7,8}, Veli-Matti Isoviita⁵, Johanna Hynninen⁶, Benno Schwikowski¹

Author details: ¹Computational Systems Biomedicine Lab, Department of Computational Biology, Institut Pasteur, Université Paris Cité, Paris, France, ²Bioinformatics and Biostatistics Hub, Institut Pasteur, Université Paris Cité, Paris, France, ³Heidelberg University, Faculty of Medicine, and Heidelberg University Hospital, Institute for Computational Biomedicine, Bioquant, Heidelberg, Germany, ⁵Research Program in Systems Oncology, Research Programs Unit, Faculty of Medicine, University of Helsinki, Helsinki, Finland, ⁶Department of Obstetrics and Gynecology, University of Turku and Turku University Hospital, Turku, Finland, ⁷Univeristy of Modena and Reggio Emilia, Modena, Italy, ⁸University of Pisa, Pisa, Italy.

Background: The growing amount of biomedical knowledge about cancer in combination with genome-scale patient profiling data offers unprecedented opportunities for personalized oncology. However, the large amounts of knowledge and data require scalable approaches to providing actionable information to support clinicians in decision-making [1].

Objective: To develop software and methods that integrate all relevant clinical and genomic data about patients and that enables the discovery of optimal personalized treatment options, together with the supporting literature knowledge and data.

Methods: We exploit a Semantic Knowledge Graph (SKG), a type of database that represents medical data in the form of objects and relationships, linking previously unconnected information across several cancer databases. To build up this SKG (*OncodashKB*), we use the BioCypher library [2]. We then integrate clinical data from patients with high-grade serous ovarian cancer, including information on genome changes collected as part of the DECIDER project (<https://deciderproject.eu>). The SKG can then be queried to gather evidence paths linking patient-specific alterations to actionable drugs.

Results: Our approach provides a fully automated, systematic, and reproducible data integration workflow, along with the use of existing expert-made ontologies to provide interoperability and semantic descriptions. The integrated data is assessed by experts on molecular tumor boards and allows for the exploration of relevant clinical and genomic patient data in a visually accessible format, designed for ease of interpretation by clinicians. Importantly, we expect the system to reveal compelling findings/indications from multiple lines of converging evidence to a strong case from multi-omics patient data and a broad and automated review of the latest clinical and experimental knowledge.

Conclusion: Decision support systems using graph databases emerge as valuable tools by revealing new connections between various patient data and treatment options shown in an easy-to-understand format.

References:

- [1] Reisle, C. *et al.* A platform for oncogenomic reporting and interpretation. *Nat Commun* 13, 756 (2022). <https://doi.org/10.1038/s41467-022-28348-y>
- [2] Lobentanzer, S., Aloy, P., Baumbach, J. *et al.* Democratizing knowledge representation with BioCypher. *Nat Biotechnol* 41, 1056–1059 (2023). <https://doi.org/10.1038/s41587-023-01848-y>