Integration of Clinical and Genomic Data Mapped to the OMOP Common Data Model in a Federated Data Network in Belgium

<Tatjana Jatsenko>[1], <Murat Akand>[2], < Joris Robert Vermeesch>[1], <Dries Rombaut>[3] , Michel Van Speybroeck>[4] , <Martine Lewi>[5] , <Valerie Vandeweerd>[6]

[1] Center for Human Genetics, UZ Leuven, [2] Dept. of urology KUL, [3] Illumina, [4] Janssen EMEA Data Science, [5]Janssen Global Commercial Data Science, [6]Janssen R&D/Clinical Innovation

Background

There is a strong belief in the benefits of enriching clinical patient data with omics data to gain a deeper understanding of the complex multifactorial causes of a disease, as well as predicting treatment success.[1, 2, 3] However, due to the decentralized nature of healthcare systems and the sensitive nature of genomic and clinical information, data is often stored in siloed, inaccessible locations.[4,5] Federated data platforms are emerging as effective solutions to achieve data accessibility, usability, and security while maintaining compliance with governance and privacy regulations. ATHENA, a disease-specific federated data network in Belgium in multiple myeloma and bladder cancer, exemplifies the power of personalized medicine through multimodal data management and analysis. In this case, we focus on the effective combination of clinical data and research-grade genomics data in non-muscle invasive bladder cancer (NMIBC). Leveraging the Observational Medical Outcomes Partnership (OMOP) Common Data Model, this solution facilitates privacy-preserving, large-scale genotype-phenotype research in NMIBC. Through ATHENA, personalized medicine approaches can harness the potential of both clinical and genomic data.

Methods

To create the federated data network, we established collaborations with multiple healthcare institutions across Belgium, including hospitals, research institutes, and genomic sequencing centers. The participating institutions agreed to contribute de-identified clinical and genomic data from their respective databases. To ensure privacy and data security, we implemented robust data governance protocols, including data encryption, strict access controls, and compliance with relevant data protection regulations.

Clinical data collected from the EHR system from patients with transitional cell bladder cancer was mapped to OMOP V5.3. In addition, DNA and RNA extracted from FFPE tumor samples were analyzed using the Trusight Oncology 500 assay (TSO500) from Illumina.[6] The TSO500 assay is a clinical research solution for comprehensive genomic profiling of a range of solid tumors. Different genomic features such as micro-satellite instability (MSI), tumor mutational burden (TMB), single nucleotide variants (SNVs), copy number variants, indels, fusions across 523 genes for DNA and 55 for RNA are analyzed through this assay.[7]

The TSO500 data output was converted to a set of measurements:

1. MSI and TMB as numerical measurements.
2. SNVs recorded with the respective gene from the OMOP Genomic vocabulary as measurement_concept measurement_concept_id. In a second step, for mutations with known clinical impact, the specific mutation is retrieved, joined through OncoKB vocabulary and the corresponding concept from the OMOP genomic vocabulary is used a as

measurement_concept_id.

3. Fusion products will be mapped as standard OMOP genomic concepts.
4. The measurement_source_value contained the full annotation from the Illumina assay in Mutation Annotation format (MAF).

Results

The proof of concept illustrates the possibility to integrate clinical data with exhaustive NGS data using OMOP. In this case, a lightweight approach was chosen to map all genomics data to the measurement entity. This allows all clinical and genomics data to be immediately accessible through the conventional OHDSI tools and methods. Given the limited set of (standardized) genomic concepts in the current OMOP vocabularies, data transformation goes with a loss of granularity, especially for the alterations of unknown clinical significance.

In addition, the logistical challenge of integrating the data back into a single OMOP instance and attributing it to the right person_id when the genomics data is generated in another entity, remains a challenge.

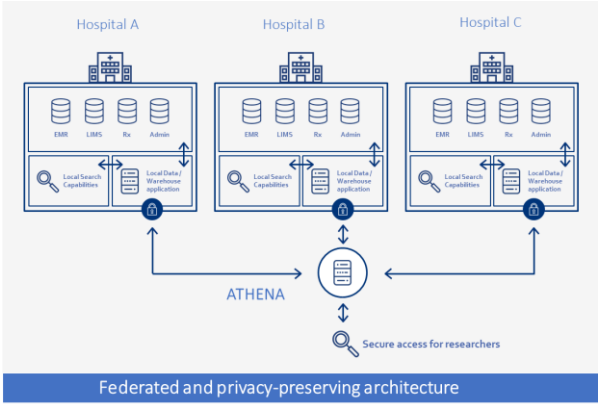

Figure 1.

Conclusion

The development of a federated data network incorporating clinical and genomic data and mapping it to the OMOP Common Data Model represents a significant milestone in developing a privacy preserving approach to enable precision medicine research. The federated approach ensures that data remains securely stored at each institution, mitigating privacy concerns and legal barriers associated with data sharing. By harmonizing the data to a common data model, researchers and clinicians can gain insights into disease mechanisms, treatment outcomes, and genotype-phenotype associations on a larger scale. This infrastructure lays the foundation for collaborative research projects and paves the way for personalized medicine approaches that utilize both clinical and genomic data. Future efforts will focus on expanding the network to include additional institutions and incorporating more diverse datasets to enhance the representativeness and generalizability of research findings.

In summary, the development of a federated data network in Belgium, incorporating clinical and genomic

data and mapping it to the OMOP Common Data Model, holds great promise for accelerating precision medicine research and improving patient care.

References

1. Letai. Functional precision cancer medicine—moving beyond pure genomics. Nature Medicine 23 (9): 1028-1035 (2017).
2. Brandão, Pondé and Piccart-Gebhart. MammaprintTM: A comprehensive review. Futur. Oncol., 15, 207–224 (2019).
3. Gerstung, Papaemmanuil, Martincorena, Bullinger, Gaidzik, Paschka, Heuser, Thol, Bolli, Ganly, Ganser, McDermott, Döhner, Schlenk, Döhner & Campbell. Precision oncology for acute myeloid leukemia using a knowledge bank approach. Nat Genet 49, 332–340 (2017).
4. Asiimwe, R., Lam, S., Leung, S., Wang, S., Wan, R., Tinker, A., et al. (2021). From biobank and data silos into a data commons: Convergence to support translational medicine. J. Transl. Med. 19, 493. doi:10.1186/s12967-021-03147-z
5. Garden, H. (2021). Building and sustaining collaborative platforms in genomics and biobanks for health innovation (OECD Science, Technology and Industry Policy Papers No. 102). OECD Sci. Technol. Industry Policy Pap 102. doi:10.1787/11d960b7-en
6. Illumina. TruSight Oncology 500 [Internet]. United States: Illumina, Inc; [cited 2023 June 8]. Available from: https://www.illumina.com/products/by-type/clinical-research-products/trusight-oncology-500.html
7. Illumina. Analysis of TMB and MSI Status with TruSight™ Oncology 500 [Internet]. United States: Illumina, Inc; [cited 2023 June 8]. Available from: https://science-docs.illumina.com/documents/LibraryPrep/trusight-oncology-500-tmb-analysis-html-1170-2018-009/Content/Source/Library-Prep/TruSight/Oncology/trusight-oncology-500-tmb-analysis-1170-2018-009/trusight-oncology-500-tmb-analysis-1170-2018-009.html