Integration of Clinical and Genomic Data Mapped to 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 Speybroek[4], Martine Lewi[5], Valerie Vandeweerd[6]

[1] *Center for Human Genetics, University Hospitals Leuven, Belgium*
[2] *Department of Urology, University Hospitals Leuven, Belgium*
[3] *Illumina*
[4] *IT EMEA Departmental Janssen Pharmaceutica, Beerse, Belgium*
[5] *Global Data, Platforms & Partnerships Global Commercial Data Science, USA*
[6] *Clinical Innovation Research & Development Janssen Pharmaceutica, Beerse, Belgium*

Background:

Enriching clinical patient data with omics data can help understand and predict disease and treatment outcomes in bladder cancer.[1, 2, 3] However, omics and non-omics data are often siloed and inaccessible due to healthcare system fragmentation and privacy concerns.[4, 5] Federated data platforms enable data accessibility, usability, and security while complying with regulations. The OMOP Common Data Model allows privacy-preserving, large-scale genotype-phenotype research. ATHENA is a federated data network for multiple myeloma and bladder cancer in Belgium. This case focuses on non-muscle invasive bladder cancer (NMIBC) and the integration of clinical and genomics data.

Methods

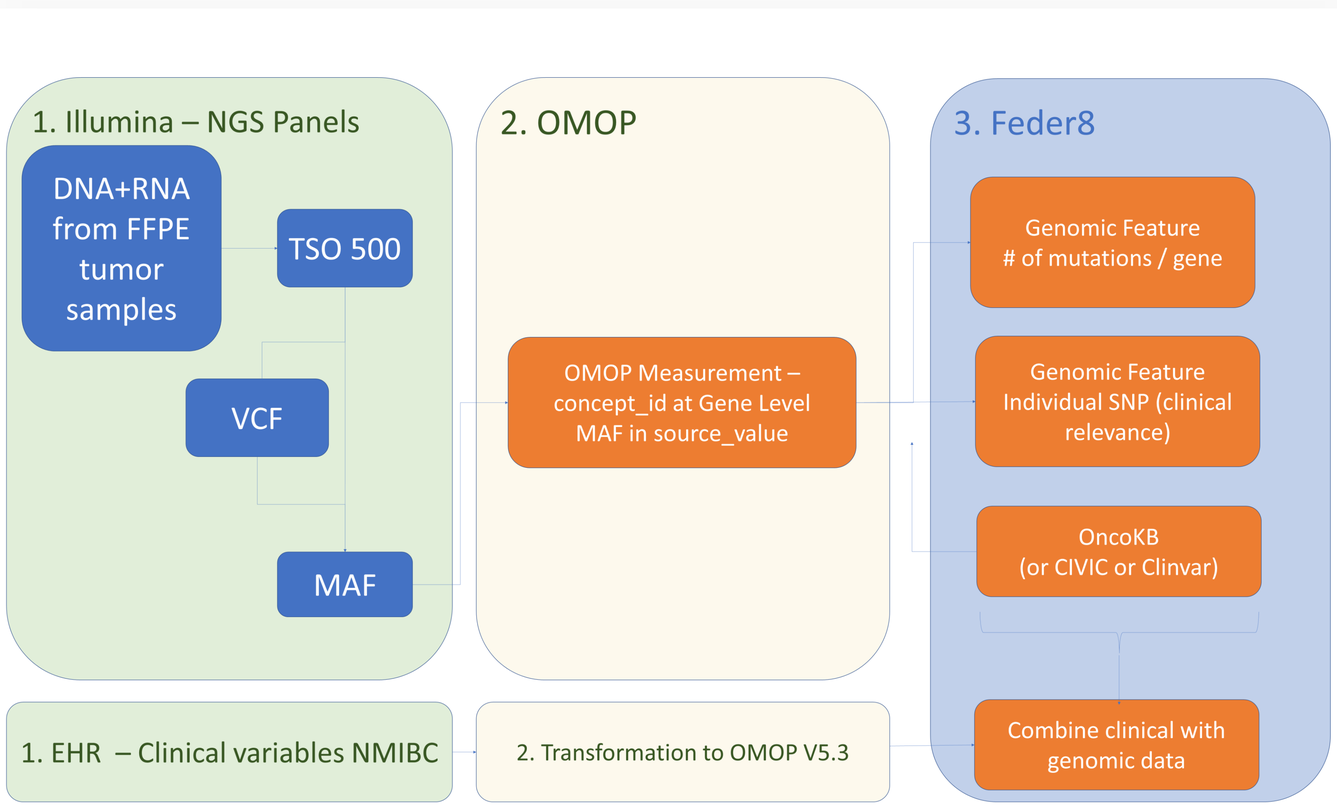

Figure 1: high level process genomics and clinical data integration

1. Extraction of de-identified clinical and genomic data from different databases of multiple healthcare institutions across Belgium. By date analyzed DNA and RNA from 102 FFPE tumor samples from UZ Leuven using TSO500 assay for genomic profiling.[6,7]
2. Collection of clinical data from EHR system for bladder cancer patients and mapped to OMOP V5.3. TSO500 data output are converted to measurements using OMOP genomic vocabulary and OncoKB precision oncology knowledge base vocabulary:
   - Microsatellite instability (MSI) and tumor mutational burden (TMB) as numerical measurements.
   - Variants with known clinical significance are filtered using OncoKB vocabulary and mapped back to standard concept id.
   - Fusion products mapped as standard OMOP genomic concepts.
   - Genomic variants that were impossible to map to a standard concept were only mapped to a custom concept (>2 b code)
   - Full mutation annotation (MAF) from Illumina assay is stored in measurement_source_value.
3. The feder8 platform profiles the clinical and genomic data from participating partners and accommodates centrally defined queries to query any available concept in any of the participating sources.

Results

- A federated data network that combines clinical and genomic data and maps it to the OMOP Common Data Model
- The federated approach keeps the data secure at each institution
- The common data model enables researchers and clinicians to do research on a larger scale.
- Future efforts will aim to grow the network with more institutions and more diverse datasets to improve the representativeness and generalizability of research findings.

Limitation: Loss of granularity, specifically for the alterations of unknown clinical significance, due to limited (standardized) set of genomic concepts. Remaining challenge of integrating the data into one single OMOP instance and matching it to the correct person_id.

VLAIO Vlaanderen is ondernemen
Project number ATHENA: HBC.2019.2528