Canonical representation of variants to serve as standard concepts in the OMOP Standardized Vocabularies

Genomic Variant Harmonization in the OMOP Standardized Vocabularies

PRESENTER: **Denys Kaduk**

INTRO:

Clinical research in precision oncology requires concise, standardized and searchable interpretations of detected variants. Currently, there is no terminology available in the public domain that would collect variants relevant in oncology, which is necessary for standardized analytics in a research network. For OHDSI, representing the various knowledgebases in a standardized manner requires an open, interoperable sharing of variant interpretation and an automated methodology so a comprehensive approach to cancer precision medicine can be developed

METHODS

We developed a simple consolidation approach of single nucleotide polymorphism, insertions, deletions and duplications based on gene symbol, sequence type (g, c, p), versions and locations. We explicitly ignored refseqs, as they are highly variable, and the remaining attributes are unambiguous in most cases. (Figure 3)

RESULTS

We incorporated the variant collections of six source vocabularies: National Cancer Institute Thesaurus (NCIt), College of American Pathologists Cancer Checklists (CAP), Clinical Interpretation of Variants in Cancer (CIViC), Cancer Genome Interpreter (CGI), The Clinical Knowledgebase by The Jackson Laboratory (JAX) and ClinVar. Genomic, Transcript and Protein variants have each their standard concept using the OMOP Extension vocabulary in the MEASUREMENT domain (figure 1) and are hierarchically connected to each other and to the gene variants from the HGNC vocabulary (figure 4). More than 55,000 new concepts have been created, displaying variants and genes. Source variants map to these to standard concepts (figure 2).

Take a picture to download the full paper

Figure 1. Concept table representation Example

Concept_id	35978081
Concept_name	Variant in Chromosome 14 Substitution in position 105246551 of C replaced by T
Domain_id	Measurement
Vocabulary_id	OMOP Extension
Concept_class_id	Genomic Variant
Standard_concept	S
Concept_code	NC_000014:g.105246551C>T
Valid_start_date	9/24/2020
Valid_end_date	12/31/2099
Invalid_reason	

Figure 2. Source variant to Standard variant relationship

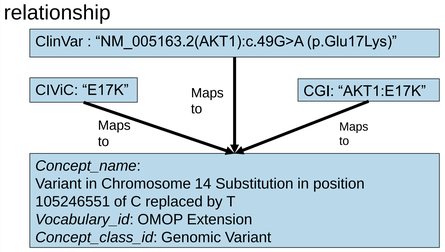


Figure 3. Consolidation of variants

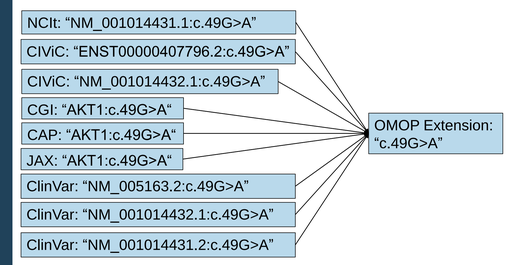
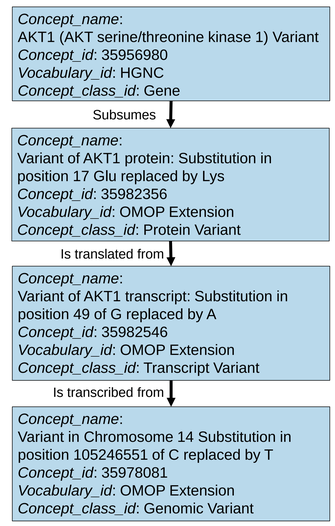


Figure 4. Hierarchy between Standard variants



Denys Kaduk, Violetta Komar, Asieh Golozar, Peter Robinson, Alex H. Wagner, Michael Gurley, Seng Chan You, Shin Seojeong, Shaadi Mehr, Andrew Williams, Rimma Belenkaya, Meera Patel, Shilpa Ratwani, Ron Miller, Christian Reich

ODYSSEUS DATA SERVICES INC

OHDSI