

Name:	Kyu-pyo Kim
Affiliation:	Department of Oncology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea
Email:	kkp1122@amc.seoul.kr
Presentation type (select one):	Poster

CDM extension for loading and utilization of clinical genomic data

Yu Rang Park^{1,2}, Deokhoon Kim³, Kyu-pyo Kim⁴

¹Health Innovation BigData center, Asan Medical Center, Seoul, South Korea; ²Department of Biomedical Informatics Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; ³Department of Pathology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; ⁴Department of Oncology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea

Abstract

Due to the limited number of patients with specific genotype-phenotype, the importance of observational studies is pivotal in the era of precision medicine. Key issues in precision medicine are limited due to the lack of effective methods for extracting data from electronic health records (EHRs) and no data specification for clinical genomic data. We adopted OMOP Common Data Model (CDM) for extracting clinical data from EHR and extending the CDM for representing clinical genomic data to overcome these issues. Since March 2017, the Korean government has provided insurance for NGS-based cancer panel. To increase utilization of accumulated clinical genomic data, we have applied various processes to save the result of NGS cancer panel test to CDM including extension of CDM to create a genomic_data table based on the Mutation Annotation Format (MAF) file and genomic data API to visualize both clinical and genomic data. We propose and developed three new modules for extracting and loading clinical and genomic data: the Patient module, to represent patient information such as demographic and specimen; the ClinicalData module to represent clinical events such as surgery, chemotherapy, radiation therapy and radiological response evaluation; and the GenomicData module.

Introduction

Due to the rapid advancement of sequencing technologies, human genome has started to be adopted in the clinical settings to realize precision medicine. The real value of genomic data will be realized only when they are linked to high-quality, longitudinal, computationally amenable clinical information, allowing researchers to identify precise genotype-phenotype associations [1]. In addition, due to the limited number of patients with specific genotype-phenotype, the importance of observational studies is pivotal. One of main technical barriers is the lack of effective methods for extracting data from electronic health records (EHRs) which has made it difficult to obtain relevant clinical information for data amalgamation. Another barrier is there is no data specification for clinical genomic data from real practice. To overcome these barriers, we adopted OMOP Common Data Model (CDM) for extracting clinical data from EHR and extending the CDM for representing clinical genomic data.

Materials and Methods

In June 2017, Asan Medical Center completed the conversion of clinical data of 4.3 million patients into CDM. Since March 2017, the Korean government has provided insurance for NGS-based cancer panel. Since then, Asan Medical Center has conducted NGS Cancer panel in various cancer patients. To increase utilization of accumulated clinical genomic data, we extended the CDM by applying the following processes to save the result of NGS cancer panel test to CDM. First, we extended the CDM to create a genomic_data table, and the structure of this genomic table was referenced to MAF file, including the annotation information for the mutation among the various genomic files. The genes in AMC NGS cancer panel are loaded as organization defened concept in Concept table of CDM. All genes registered in the Concept table were named consistent with the HGNC nomenclature. Second, we loaded the genomic table of the extended CDM with the genomic data accumulated in EMR of Asan Medical Center. The order of the NGS Cancer panel test was also loaded in the Procedure table. Finally, we developed genomic data API to load clinical and genomic data from CDM into cBioPortal for integrated visualization and analysis of clinical and genomic data.

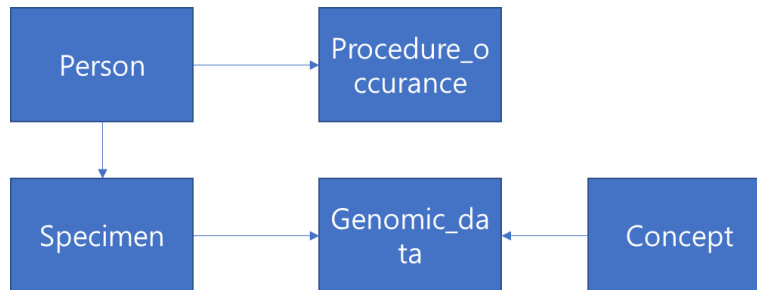


Figure 1 Subset of extended CDM structure

Result

As a result, 432 NGS cancer panel test orders occurred in 432 patients in Asan Medical Center since March 2017. Approximately 60% of these orders were in colon cancer, 20% in lung cancer, and the remaining 30% included breast cancer, stomach cancer, and ovarian cancer. Figure 1 shows the subset of extended CDM. We added one table named genomic data. The genomic_data table consisted of the following 14 items; ReferenceSeq, ObserveSeq, SourceSample, Chromosome, StarPosition, EndPosition, DNasequenceVariation, GeneId, VariantTranscriptReferenceSequenceId, DNARegionName, ProteinReference SequenceID, AminoAcideChange, and VariationID. The SouceSample and GeneId is a foreign key of Specimen and Concept table, respectively.

We loaded 432 specimen and procedure records in specimen and procedure_occurrence table, respectively. We used the Current Procedural Terminology (CPT) Version 4 code 81425 sequencing a genome to store the result of NGS cancer panel in procedure_occurrence table. 314 genes are loaded in concept table. And these concept ids of genes are linked to mutations from specific specimen of patients in the genomic_data table. A total of 2,902 clinically significant somatic mutations were obtained from bioinformatical analysis of 432 NGS cancer panel results, which were loaded on the genomic_data table.

For utilizing genomic data in extended CDM, we propose and developed three new modules for extracting and loading clinical and genomic data: The Patient module, to represent patient information such as demographic and specimen; the ClinicalData module to represent clinical events such as surgery, chemotherapy, radiation therapy and radiological response evaluation; and the GenomicData module to represent data from genomic_data table. To validate our genomic data API, we extract and loading the clinical and genomic data from extended CDM to cBioPortal (Figure 2).

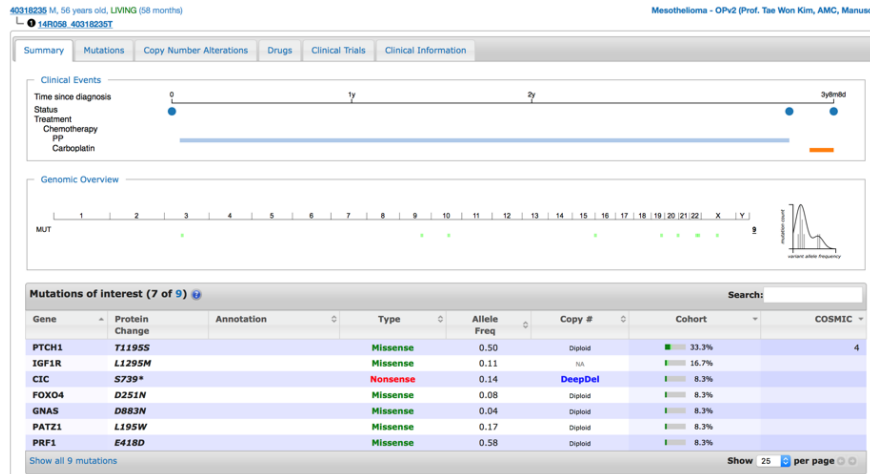


Figure 2 Example screenshot of cBioPortal interfaced with extended CDM

Conclusion

CDM is a *de facto* clinical contents standard suitable for multi-institutional studies and loading of various observational clinical data including hospitals around the world. Compared to the importance of clinical genomic data, current CDM is focused on structure for observational clinical data. In this poster, we expanded the CDM structure to add the clinical genome data in the existing CDM for concept validation for representing clinical genome data, developed the API for utilization, and loaded the genomic data accumulated in our medical center into the extended CDM. As more patients receive NGS in clinical practice in the era of precision medicine, the necessity of CDM-based clinical genomic data increases. The results of this poster can be an example of CDM-based clinical genomic data for the various organization participating in OHDSI.

References

1. Lawler, M., et al., *Sharing Clinical and Genomic Data on Cancer - The Need for Global Solutions*. New England Journal of Medicine, 2017. **376**(21): p. 2006-2009.