

# An update on Genomic CDS, a complex ontology for pharmacogenomics and clinical decision support

José Antonio Minarro-Giménez<sup>1</sup>, Matthias Samwald<sup>1</sup>

<sup>1</sup>Section for Medical Expert and Knowledge-Based Systems; Center for Medical Statistics, Informatics, and Intelligent Systems; Medical University of Vienna; Vienna, Austria  
jose.minarrojimenez@meduniwien.ac.at  
matthias.samwald@meduniwien.ac.at

**Abstract.** Genetic data can be used to optimize drug treatment based on the genetic profiles of individual patients, thereby reducing adverse drug events and improving the efficacy of pharmacotherapy. The Genomic Clinical Decision Support (Genomic CDS) ontology utilizes Web Ontology Language 2 (OWL 2) reasoning for this task. The ontology serves a clear-cut medical use case that requires challenging OWL 2 DL reasoning. We present an update of the Genomic CDS ontology which covers a significantly larger number of clinical decision support rules and where inconsistencies presented in previous versions of the ontology have been removed.

**Keywords:** OWL, pharmacogenomics, clinical decision support

## 1 Motivation

Different patients can react drastically different to the same type of medication. The goal of personalized medicine and pharmacogenomics is to predict an individual patient's response by analyzing genetic markers that influence how medications are metabolized or able to bind to their targets.

To produce clinically valid and trustworthy predictions, no errors or ambiguities should arise in the process of inferring a patient's likely response from raw genetic data. Current formalisms, data infrastructures and software applications leave many opportunities for introducing such errors and ambiguities. Ontologies formalized with the Web Ontology Language 2 (OWL 2) could be an excellent choice for tackling this problem, but the complexity and potentially large scale of ontologies in this domain also pose formidable challenges to currently available OWL 2 reasoners.

## 2 The Genomic CDS ontology

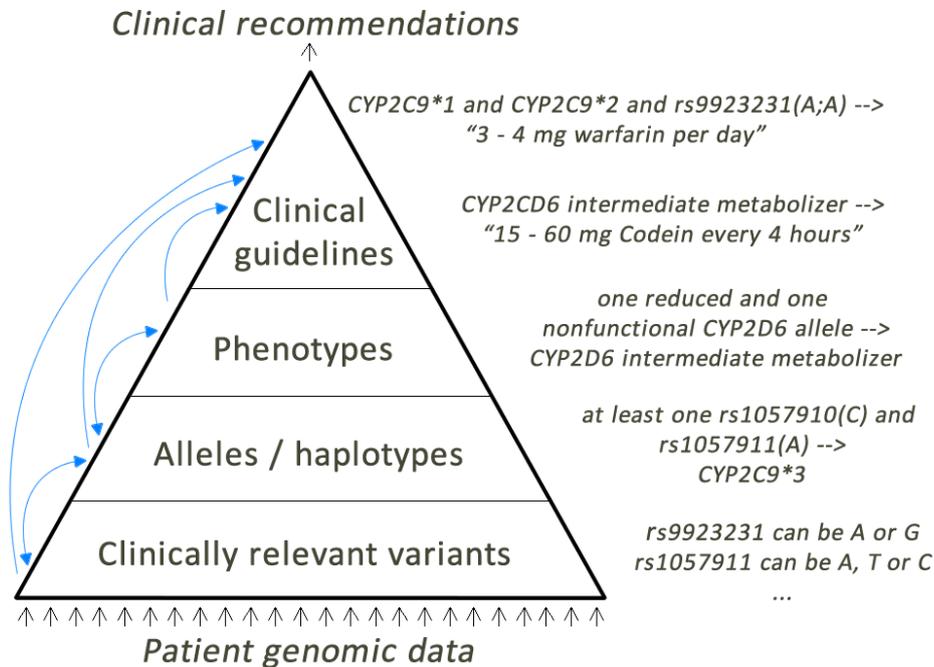
The Genomic Clinical Decision Support (Genomic CDS) ontology is an OWL 2 ontology aimed at representing pharmacogenomic knowledge and providing clinical decision support based on genetic patient data. The Genomic CDS ontology has been integrated into the Medicine Safety Code (MSC) system [1] in order to provide phar-

macogenomic decision support at the point-of-care. The different versions of the Genomic CDS ontologies can be downloaded from <http://www.genomic-cds.org/ont/snapshot-april-2014>

The goals of developing the ontology are:

- Providing a simple and concise formalism for representing pharmacogenomic knowledge
- Finding errors and lacking definitions in pharmacogenomic knowledge bases
- Automatically assigning alleles and phenotypes to patients
- Matching patients to clinically appropriate pharmacogenomic guidelines and clinical decision support messages
- Being able to detect inconsistencies between pharmacogenomics treatment guidelines from different sources.

In the most common scenario, genetic patient data in OWL format are combined with the axioms of the Genomic CDS ontology, and an OWL reasoner is used to infer matching pharmacogenomic treatment recommendations. Several inference steps are needed to derive matching treatment recommendations from raw data about genetic markers (**Fig. 1.**). The raw data consist of small variants in the genetic code, which in most cases are single nucleotide polymorphisms (SNPs), such as an 'A' instead of a 'G' or a deletion/insertion of a nucleotide. *Alleles* are variants of a gene that are defined by containing sets of such small variants. *Phenotypes* are referring to the specific effects that certain small variants and alleles can have on the organism, e.g., how quickly a patient metabolizes a specific drug. Clinical guidelines can use small variants, alleles and/or phenotypes to match patients to treatment recommendations. The Genomic CDS ontology classifies patients according to four types of inference rules that are represented as subclasses of the *human* class. The class *human\_with\_genotype\_marker* represents the first inference step which gathers the raw genetic data and recognizes particular SNP variants. The class *human\_with\_genetic\_polimorphism* is related to the second inference step where the obtained SNP variants are matched to obtain the related alleles of the patient. The third inference step is associated to the class *human\_triggering\_phenotype\_inference\_rule* which represents patient's phenotype rules based on the SNP variants and alleles obtained in the previous steps. Finally, the class *human\_triggering\_CDS\_rule* represents the rules that conceptualize the clinical guidelines based on the combination of the inferred SNP variants, alleles and phenotypes.



**Fig. 1.** Through a series of inference steps, matching pharmacogenetic treatment guidelines are inferred from raw genetic patient data.

The human genome usually contains two copies of each gene (one from the father, one from the mother), with each copy potentially bearing multiple genetic variants. Because of this, the ontologies rely heavily on qualified cardinality restrictions with cardinalities of two, which seems to cause performance issues with most current OWL reasoners.

There are two version of the ontology and the corresponding ‘demo’ versions that include an example of the genetic data of a patient. The full versions of the ontology (genomic-cds\_rules\_full.owl and genomic-cds\_rules\_full\_demo.owl) contain the axioms that can be used to link SNPs variants such as “rs267607275(G;G)”, and alleles variants such as “TPMT \*1/\*2”, to a patient, whereas the light version of the ontology (genomic-cds\_rules.owl and genomic-cds\_rules\_demo.owl) only provide axioms related to allele variants. The light version of the ontology can reduce the complexity of the model which is useful when running reasoners with limited computing resources without losing expressiveness. Both versions of the ontology have ALCQ expressivity. They are characterized by extensive use of qualified cardinality restrictions.

Compared to the 2013 version of the ontology [2], the decision support rules encoded in the ontology were increased from 49 drug dosing recommendation rules to 298 drug dosing recommendation rules and 18 phenotype inference rules. This increase in the number of rules demands more computational resources to obtain the

inferred model. In order to reduce the complexity of the ontology and facilitate the reasoning process, we optimized the ontology by removing the axioms from the previous version of the ontology which had no actual effects to the decision support rules. Therefore, the number of genes and SNP variants that we currently cover is lower than in the 2013 version of the ontology. We also removed classes of genes, such as “ABCG2” or “NAT1”, that were not reflected in any rule and, consequently, the total number of genes represented in the ontology decreased from 72 to 38. Besides, the number of SNPs variants decreased from 822 to 674. Despite these removals, the number of defined alleles and drugs included in the 2014 version of the Genomic CDS ontology is larger. The statistics about the ontology are summarized in **Table 1**.

**Table 1.** Comparison of 2013 and 2014 versions of the Genomic CDS ontology. In the 2014 version, the number of defined drug treatment recommendation rules has increased but the number of genes and SNP variants has reduced due to removal of unnecessary axioms for triggering rules.

Version	Genotype rules	Drugs	Genes	Alleles variants	SNP variants
2014	298 (+ 18 phenotype rules)	62	38	664	674
2013	49	6	72	301	822

A simplified example of a rule for inferring an allele (CYP2C9 \*3) based on its single nucleotide polymorphisms, which also include a SNP insertion (rs72558188\_AGAAATGGAA), looks like this in Manchester syntax:

```
Class: 'human with CYP2C9 *3'
EquivalentTo:
  (has some rs1057910_C) and (has some rs1057911_A)
  and (has some rs1799853_C) and (has some rs2256871_A)
  and (has some rs28371685_C) and (has some rs28371686_C)
  and (has some rs56165452_T) and (has some rs57505750_T)
  and (has some rs67807361_C) and (has some rs72558184_G)
  and (has some rs72558187_T) and (has some rs72558189_G)
  and (has some rs72558190_C) and (has some rs72558192_A)
  and (has some rs72558193_A) and (has some rs7900194_G)
  and (has some rs9332130_A) and (has some rs9332131_A)
  and (has some rs9332239_C) and
  (has some rs72558188_AGAAATGGAA)
SubClassOf:
  has some CYP2C9_star_3
  human_with_genetic_polimorphism
```

An example of an axiom for inferring an adequate clinical decision support message for the anticoagulant drug warfarin (based on a combination of alleles and SNPs according to an official recommendation in the drug label):

```
Class: 'human triggering CDS rule 7'  
EquivalentTo:  
  (has some 'CYP2C9*1') and (has some 'CYP2C9*3')  
  and (has exactly 2 rs9923231_C)  
Annotations:  
  label "human triggering CDS rule 7",  
  CDS_message "3-4 mg warfarin per day should  
  be considered as a starting dose range for  
  a patient with this genotype according to  
  the Warfarin drug label (Bristol-Myers  
  Squibb)."
```

From the previous version of the ontology [2] we found that TrOWL<sup>1</sup> [3] is significantly more performant than the Hermit<sup>2</sup> reasoner and other OWL 2 DL reasoners when classifying and realizing the Genomic CDS ontology. Consequently, in this paper we evaluated versions 1.3 and 1.4 of TrOWL reasoners for classifying the full demo (MSC\_classes\_demo.owl) and the light demo (genomic-cds\_demo.owl) of the Genomic-CDS ontology. We compared the performance of the two version of the reasoner on a 64-bit Windows 7 machine with 4GB of memory and an Intel i5-2430 at 2.4GHz. The reasoner use version 3.0 of OWLAPI and JRE 6 update 29 to run each demo files. The results of this evaluation are shown in **Table 2**. As expected, the reasoners take more time to classify the full version of our demo ontology than the simplified one. Surprisingly, the latest version of TrOWL (1.4) takes slightly longer than the previous version (1.3) to classify the demo ontologies. The 1.4 version of TrOWL seems to be a minor revision of the 1.3 version due to the fact that the developers only highlight some fixed bugs on the ontological patterns. Our hypothesis is that such changes in the updated reasoner have increase the complexity of the inference process and, consequently, its performance is lower.

---

<sup>1</sup> <http://trowl.eu>

<sup>2</sup> <http://www.hermit-reasoner.com/>

**Table 2.** Reasoning performance using full and light versions of the Genomic CDS ontology with the 1.3 and 1.4 versions of the TrOWL reasoner.

<b>Ontology version</b>	<b>TrOWL 1.3</b>	<b>TrOWL 1.4</b>
MSC_classes_demo.owl (full ontology)	156 seconds	178 seconds
genomic-cds_demo.owl (light ontology)	111 seconds	122 seconds

### **3 Conclusions and outlook**

The Genomic CDS ontology is an example of an OWL 2 ontology for clinical genetics and decision support. The updated version of this ontology has covered an increased number of drug treatment recommendation rules and has improved some pharmacogenetic markers. As a consequence, the total number of axioms has increased, further increasing the demand for OWL reasoners that could deal with this type of ontologies in a reasonable time and with limited resources use.

### **4 Acknowledgements**

The research leading to these results has received funding from the Austrian Science Fund (FWF): [PP 25608-N15].

### **References**

1. Miñarro-Gimenez JA, Blagec K, Boyce R, Adlassnig K-P, Samwald M. An Ontology-Based, Mobile-Optimized System for Pharmacogenomic Decision Support at the Point-of-Care. *Plos One*. 9(5):e93769.
2. Samwald M. Genomic CDS: an Example of a Complex Ontology for Pharmacogenetics and Clinical Decision Support. Ulm, Germany: CEUR Workshop Proceedings; 2014. p. 128-33.
3. Thomas E, Pan JZ, Ren Y. TrOWL: Tractable OWL 2 Reasoning Infrastructure. *The Semantic Web: Research and Applications*. 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 – June 3, 2010, Proceedings, Part II. 2010. p. 431-5.