Genomic CDS

An RDF/OWL Knowledge Base for Query Answering and Decision Support in Clinical Pharmacogenetics

Matthias Samwald
Medical University of Vienna
W3C Semantic Web for Healthcare and Life Science Interest Group
Drug efficacy and toxicity can vary drastically between patients with different genetic profiles.

Significant cause of morbidity and mortality!
One reason why many promising therapeutics in development fail to reach patients!
Pharmacogenetic assays and treatment algorithms are becoming more and more numerous.
Pharmacogenetic assays and treatment algorithms are becoming more and more numerous.
Pharmacogenetic assays and treatment algorithms are becoming more and more numerous.

Sequencing-based:
- PGRNseq

Microarray-based:
- 23andMe
- Affymetrix DMET chip
- Florida/Stanford chip
We are creating an ontology-based (OWL 2 DL) framework for representing pharmacogenetic knowledge and providing pharmacogenetic clinical decision support.
Clinical recommendations

having CYP2C9*1 and CYP2C9*2 and rs9923231(A;A)
-> “3 - 4 mg warfarin per day”

being CYP2CD6 intermediate metabolizer
-> “15 - 60 mg Codeine every 4 hours”

having one reduced function and one nonfunctional CYP2D6 allele
-> CYP2D6 intermediate metabolizer

having at least one rs1057910(C) and one rs1057911(A)
-> CYP2C9*3

rs9923231 can be A or G
rs1057911 can be A, T or C
Clinical recommendations

- having CYP2C9*1 and CYP2C9*2 and rs9923231(A;A)
  -> “3 - 4 mg warfarin per day”

- being CYP2CD6 intermediate metabolizer
  -> “15 - 60 mg Codeine every 4 hours”

- having one reduced function and one nonfunctional CYP2D6 allele
  -> CYP2D6 intermediate metabolizer

Clinical guidelines

Phenotypes

Alleles / haplotypes

- having at least one rs1057910(C) and one rs1057911(A)
  -> CYP2C9*3

Clinically relevant variants

Patient genomic data

rs9923231 can be A or G
rs1057911 can be A, T or C
Clinical recommendations

- Having CYP2C9*1 and CYP2C9*2 and rs9923231(A;A) → “3 - 4 mg warfarin per day”
- Being CYP2CD6 intermediate metabolizer → “15 - 60 mg Codeine every 4 hours”

Phenotypes
- Having one reduced function and one nonfunctional CYP2D6 allele → CYP2D6 intermediate metabolizer
- Having at least one rs1057910(C) and one rs1057911(A) → CYP2C9*3

Alleles / haplotypes
- Rs9923231 can be A or G
- Rs1057911 can be A, T or C

Clinically relevant variants

Patient genomic data
Clinical recommendations

- having CYP2C9*1 and CYP2C9*2 and rs9923231(A;A)
  \[ \rightarrow \text{“3 - 4 mg warfarin per day”} \]

- being CYP2CD6 intermediate metabolizer
  \[ \rightarrow \text{“15 - 60 mg Codeine every 4 hours”} \]

Clinical guidelines

- having one reduced function and one nonfunctional CYP2D6 allele
  \[ \rightarrow \text{CYP2D6 intermediate metabolizer} \]

Phenotypes

- having at least one rs1057910(C) and one rs1057911(A)
  \[ \rightarrow \text{CYP2C9*3} \]

Alleles / haplotypes

- rs9923231 can be A or G
- rs1057911 can be A, T or C

Clinically relevant variants

Patient genomic data
Clinical recommendations

- having CYP2C9*1 and CYP2C9*2 and rs9923231(A;A)
  -> “3 - 4 mg warfarin per day”

- being CYP2CD6 intermediate metabolizer
  -> “15 - 60 mg Codeine every 4 hours”

- having one reduced function and one nonfunctional CYP2D6 allele
  -> CYP2D6 intermediate metabolizer

- having at least one rs1057910(C) and one rs1057911(A)
  -> CYP2C9*3

- rs9923231 can be A or G
  rs1057911 can be A, T or C
Small variations are scattered throughout genome, usually there are two copies of each gene

Patient’s GENE1 status

One copy from the mother: A C G A C A T A ...

One copy from the father: A C G A C A T C ...

rs1  rs2 rs3
Alleles / haplotypes (specific variants of genes) can be defined through logical axioms
Logical axioms can be used to infer alleles from partial, ambiguous test results

A) Allele definitions, B) simple case, C) ambiguous case requiring haplotype resolution
Of course, some of the genes in the ontology are far larger than that.
This is how it actually looks in the ontology

1 Class: 'human with CYP2C9*3'
2 EquivalentTo:
3   has some rs1057910_C
4 SubClassOf:
5   has some 'CYP2C9 *3',
6   (has some rs1057910_C) and
7   (has some rs1057911_A) and
8   (has some rs1799853_C) and
9   (has some rs2256871_A) and
10  (has some rs28371685_C) and
11  (has some rs72558188_AGAAATGGAA) and
12  (has some rs72558189_G) and
13  (has some rs9332239_C)
   ...

This is how it actually looks in the ontology

1 Class: 'human with CYP2C9*18'
2 EquivalentTo:
3 (has some rs1057910_C) and
4 (has some rs1057911_T) and
5 (has some rs72558193_C)
6 SubClassOf:
7 has some 'CYP2C9 *18',
8 (has some rs1057910_C) and
9 (has some rs1057911_T) and
10 (has some rs1799853_C) and
11 (has some rs2256871_A) and
12 (has some rs28371685_C) and
...
Experimental extension for haplotype resolution for ambiguous test results

Class: 'human with GENE1*3'
  SubClassOf: human
  EquivalentTo:
    (has some (rs1_T that (taken_by some GENE1*3))) and
    (has some (rs2_G that (taken_by some GENE1*3)),
    has some GENE1*3
  SubClassOf:
    has some (rs3_A that (taken_by some GENE1*3))

Class: polymorphism
  SubClassOf:
    'taken by' exactly 1 allele

Class: GENE1
  EquivalentTo:
    GENE1_star_1 or GENE1_star_2 or GENE1_star_3 or
    GENE1_star_4 or GENE1_star_5
Dosing guideline from an FDA drug label

<table>
<thead>
<tr>
<th>Genotype</th>
<th>( ^{*1/!*1} )</th>
<th>( ^{*1/!*2} )</th>
<th>( ^{*1/!*3} )</th>
<th>( ^{*2/!*2} )</th>
<th>( ^{*2/!*3} )</th>
<th>( ^{*3/!*3} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>VKORC1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>5-7 mg</td>
<td>5-7 mg</td>
<td>3-4 mg</td>
<td>3-4 mg</td>
<td>3-4 mg</td>
<td>0.5-2 mg</td>
</tr>
<tr>
<td>AG</td>
<td>5-7 mg</td>
<td>3-4 mg</td>
<td>3-4 mg</td>
<td>3-4 mg</td>
<td>0.5-2 mg</td>
<td>0.5-2 mg</td>
</tr>
<tr>
<td>AA</td>
<td>3-4 mg</td>
<td>3-4 mg</td>
<td>0.5-2 mg</td>
<td>0.5-2 mg</td>
<td>0.5-2 mg</td>
<td>0.5-2 mg</td>
</tr>
</tbody>
</table>
Dosing guideline from an FDA drug label

Class: 'human triggering CDS rule 9'

Annotations:

CDS_message "0.5-2 mg warfarin per day should be considered as a starting dose range for a patient with this genotype according to the warfarin drug label."

EquivalentTo:

(has some 'CYP2C9 *1') and
(has some 'CYP2C9 *3') and
(has exactly 2 rs9923231_T)
Describing an individual patient in OWL

1 Individual: example_patient
2 Types:
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
...
Describing an individual patient in OWL

"0.5 - 2 mg warfarin per day should be considered as a starting dose range for a patient with this genotype according to the warfarin drug label."

OWL Reasoner
A virtual patient cohort

We are running our OWL 2 reasoner over hundreds of publicly available human genomes to analyse how pharmacogenetic decisions support would influence treatment in patient populations.
OWL 2 DL reasoning needs to be *fast!!*
TrOWL is significantly more performant than other openly available reasoners for our demo ontology

- **TrOWL 1.1:** 5.8 s
- HermiT 1.3.8: did not terminate within 1 h
- Fact++: did not terminate within 1 h
- Pellet: did not terminate within 1 h
- JFact: did not terminate within 1 h
- MORe JFact 0.1.5: did not terminate within 1 h
- MORe Hermit 0.1.5: did not terminate within 1 h

Demo ontology has ALCQ expressivity and approx. 2300 classes, 11000 axioms. The demo ontology also includes the genetic profile of a single patient.

Tested with reasoner plugins in Protégé 4.3, Running on an Amazon EC2 “High-Memory Extra Large Instance” virtual machine, Microsoft Windows Server 2008, 17.1 GB of memory, 64-bit platform, two virtual cores with 3.25 EC2 compute units each
We are creating a barrier-free system for storing and interpreting personal pharmacogenetic information (based on 2D barcodes and web-based decision support).
We are creating a barrier-free system for storing and interpreting personal pharmacogenetic information (based on 2D barcodes and web-based decision support)
Next steps

• Integrating with HL7 standards for clinical genomics and decision support (OpenCDS?)

• Forming collaborations with stakeholders from
  o clinical practice (early adopters of prospective pharmacogenetic testing)
  o pharmaceutical companies
  o health insurance providers
  o genetic testing service providers

• Testing in realistic clinical settings
Take-home messages

• Genomic CDS ontology + Medicine Safety Code system provide a comprehensive solution for clinical pharmacogenetics

• RDF/OWL 2 is capable of providing both decision support functionality as well as flexible knowledge bases for molecular, personalized medicine in an integrated model

• OWL 2 reasoner performance continues to improve significantly (what we are doing today was not possible 1-2 years ago)
Thanks!

Local team:
Jose Antonio Miñarro Gimenez
Klaus-Peter Adlassnig

W3C partners:
Richard Boyce (University of Pittsburgh)
Robert R. Freimuth (Mayo Clinic)
Michel Dumontier (Carleton University)
Simon Lin (Marshfield Clinic)
Robert L. Powers (Predictive Medicine, Inc.)
Joanne S. Luciano (Rensselaer Polytechnic Institute)
Eric Prud’hommeaux (W3C)
M. Scott Marshall (MAASTRO Clinic)

Funding:
Austrian Science Fund (FWF): [PP 25608-N15]

Links:
http://www.genomic-cds.org/
http://safety-code.org/
http://samwald.info/