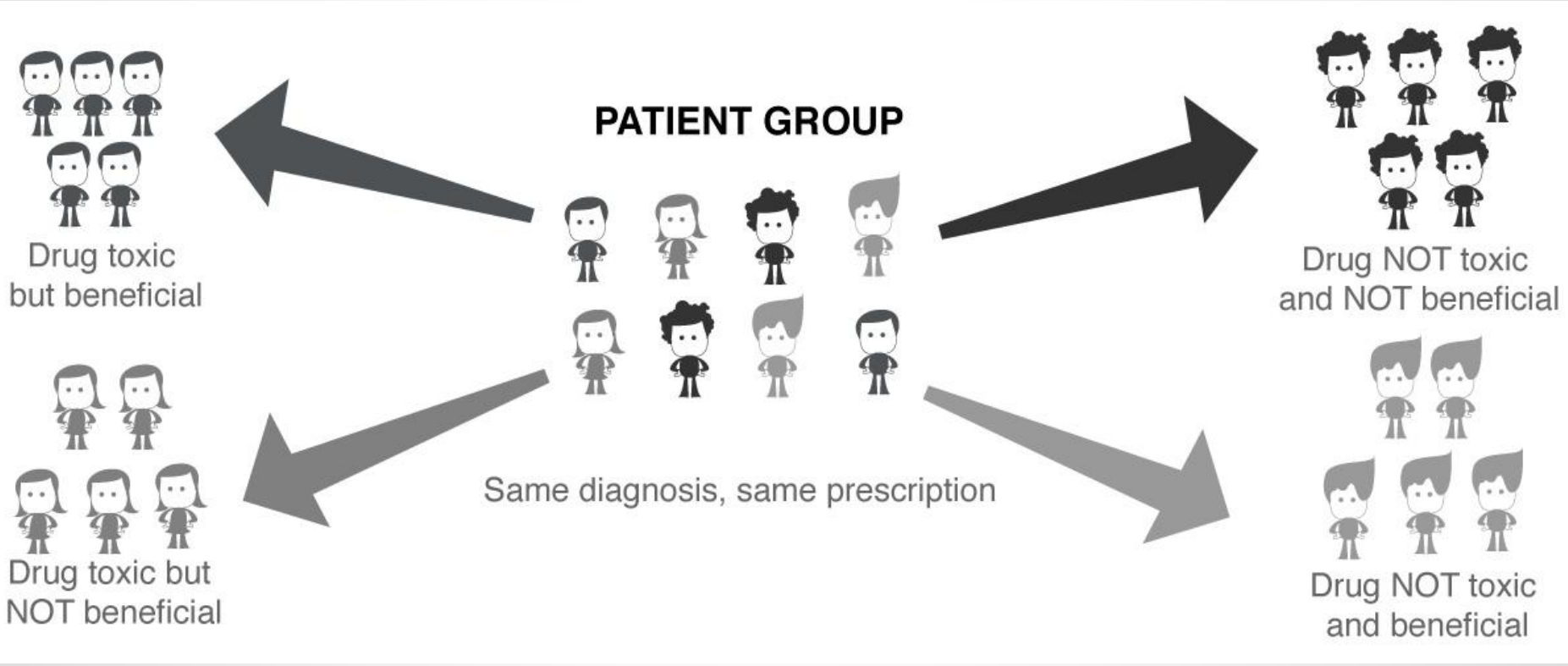
Semantically Enabling Genetic Medicine

Facilitating Patient - Guideline Matching and Pharmacogenetic Clinical Decision Support

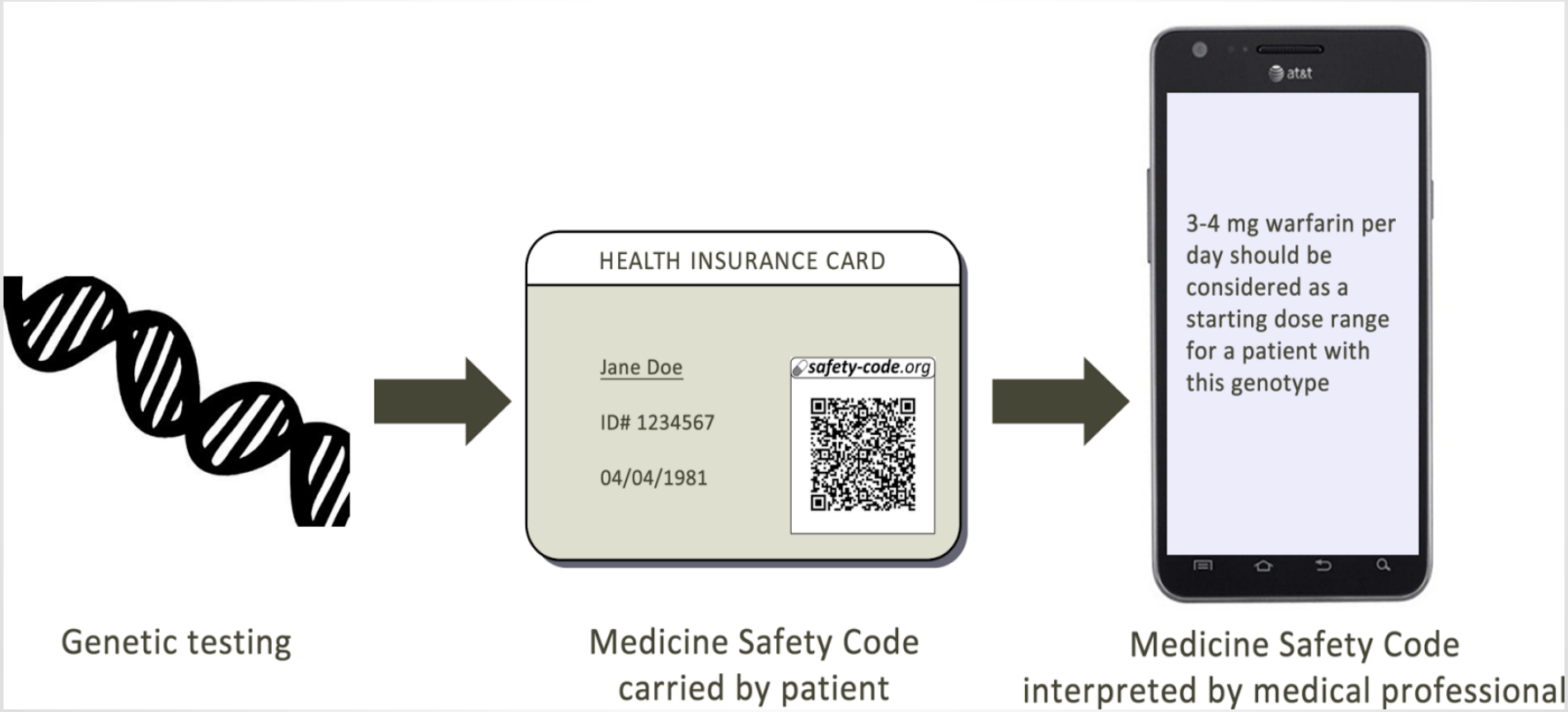
Matthias Samwald
Medical University of Vienna
W3C Semantic Web for Healthcare and Life Science Interest Group

Drug efficacy and toxicity can vary drastically between patients with different genetic profiles

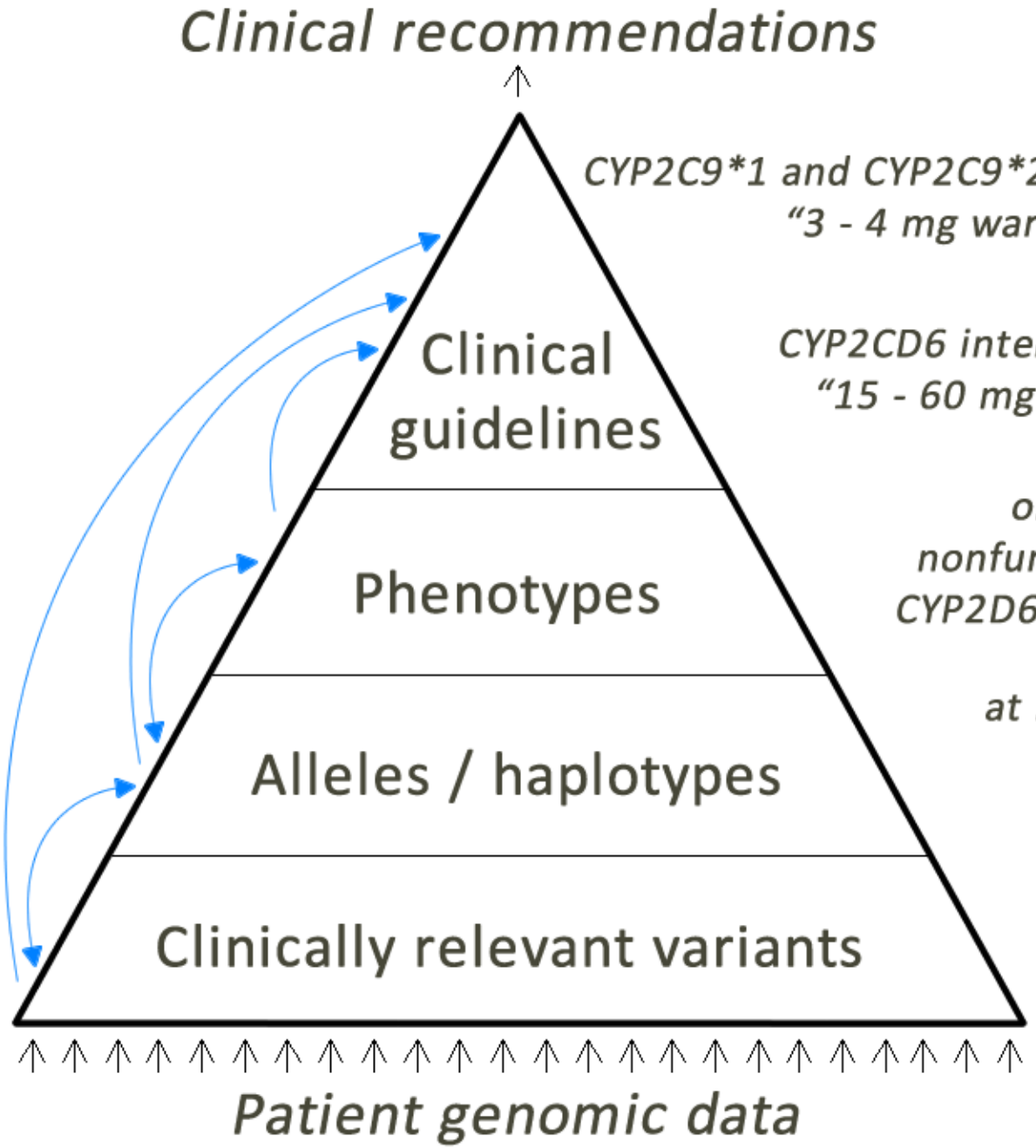


Up to 100,000 deaths and 2 million hospitalizations are caused by adverse drug reactions per year in the United States alone.

Goals: Solving this problem by creating a barrier-free system for storing and interpreting personal pharmacogenomic information



The backend of this systems:
OWL 2 for pharmacogenomic knowledge representation and clinical decision support



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

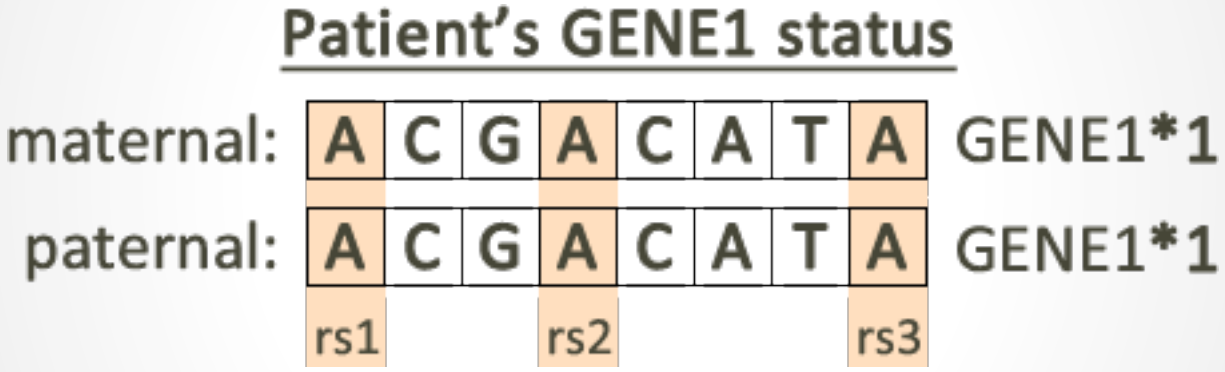
CYP2CD6 intermediate metabolizer -->
"15 - 60 mg Codein every 4 hours"

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

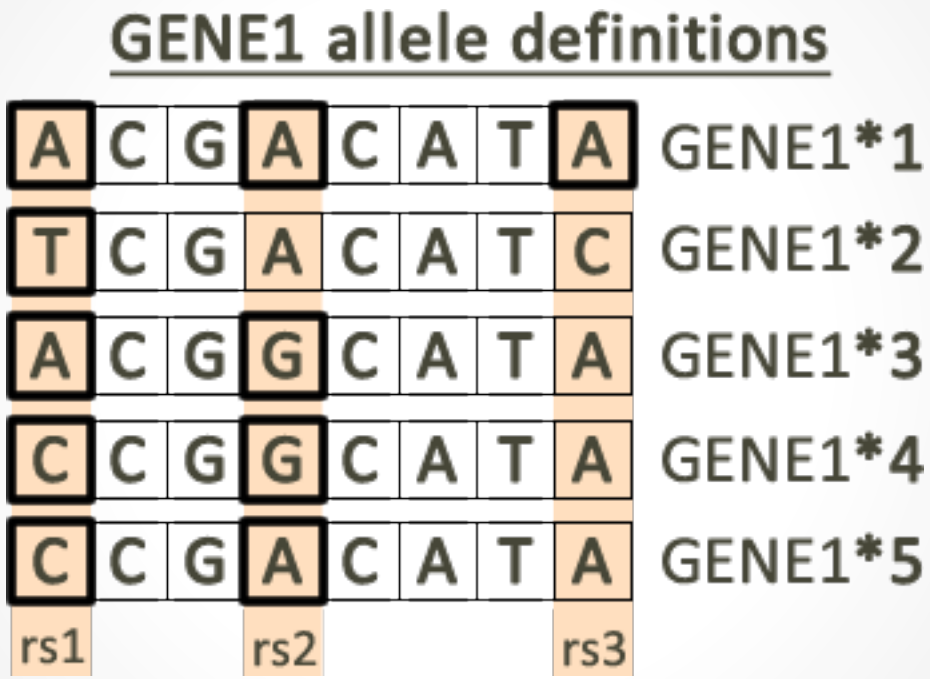
at least one rs1057910(C) and
rs1057911(A) -->
*CYP2C9*3*

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

Simplified example for a single gene



Some variants are necessary; some variants and variant combinations are necessary and sufficient for calling an allele.



C necessary

C necessary and sufficient

Of course, some of the genes in the ontology are far larger than that

Haplotype Name	rs1799853	rs1057910	rs56165452	rs28371686	rs9332131	rs67807361	rs7900194	rs2256871	rs9332130	rs28371685	rs9332239
*1	C	A	T	C	A	C	G	A	A	C	C
*2	T	A	T	C	A	C	G	A	A	C	C
*3	C	C	T	C	A	C	G	A	A	C	C
*4	C	A	C	C	A	C	G	A	A	C	C
*5	C	A	T	G	A	C	G	A	A	C	C
*6	C	A	T	C	del	C	G	A	A	C	C
*7	C	A	T	C	A	A	G	A	A	C	C
*8	C	A	T	C	A	C	A	A	A	C	C
*9	C	A	T	C	A	C	G	G	A	C	C
*10	C	A	T	C	A	C	G	A	G	C	C
*11	C	A	T	C	A	C	G	A	A	T	C
*12	C	A	T	C	A	C	G	A	A	C	T
*13	C	A	T	C	A	C	G	A	A	C	C
*14	C	A	T	C	A	C	G	A	A	C	C
*15	C	A	T	C	A	C	G	A	A	C	C
*16	C	A	T	C	A	C	G	A	A	C	C
*18	C	C	T	C	A	C	G	A	A	C	C

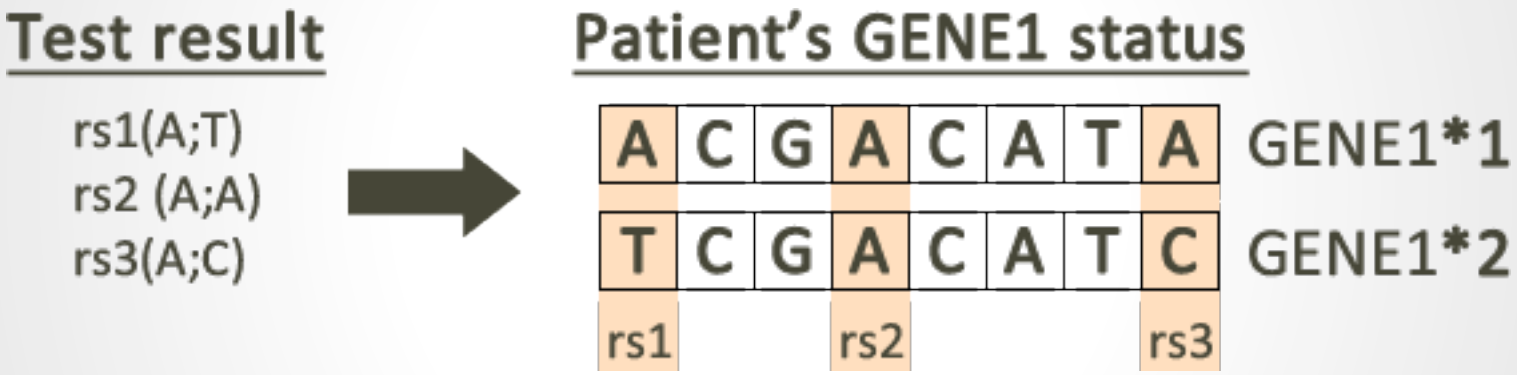
This is how it actually looks in the ontology

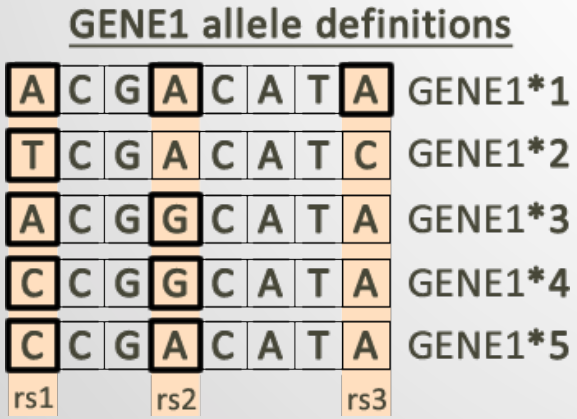
```
1  Class: 'human with CYP2C9*3'
2  EquivalentTo:
3   has some rs1057910_C
4  SubClassOf:
5   has some 'CYP2C9 *3',
6    (has some rs1057910_C) and
7    (has some rs1057911_A) and
8    (has some rs1799853_C) and
9    (has some rs2256871_A) and
10   (has some rs28371685_C) and
11   (has some rs72558188_AGAAATGGAA)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

This is how it actually looks in the ontology

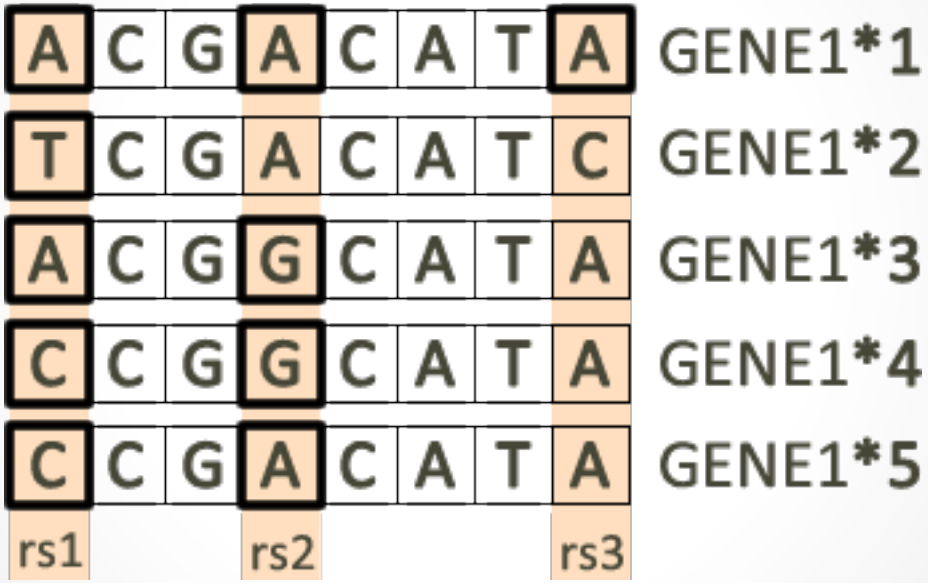
```
1  Class: 'human with CYP2C9*18'
2  EquivalentTo:
3   (has some rs1057910_C) and
4   (has some rs1057911_T) and
5   (has some rs72558193_C)
6  SubClassOf:
7   has some 'CYP2C9 *18',
8   (has some rs1057910_C) and
9   (has some rs1057911_T) and
10  (has some rs1799853_C) and
11  (has some rs2256871_A) and
12  (has some rs28371685_C) and
    ...
```

The trouble with some cheap genetic tests is that we don‘t know on which strand each specifc observed variant is located, we only know that they are one of the strand

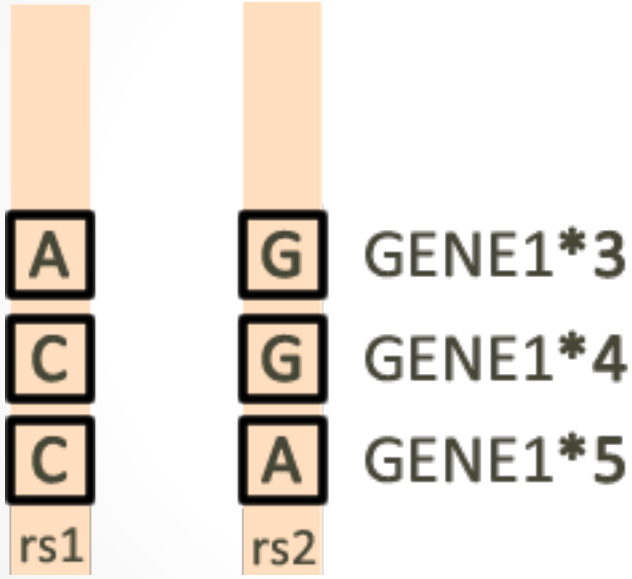




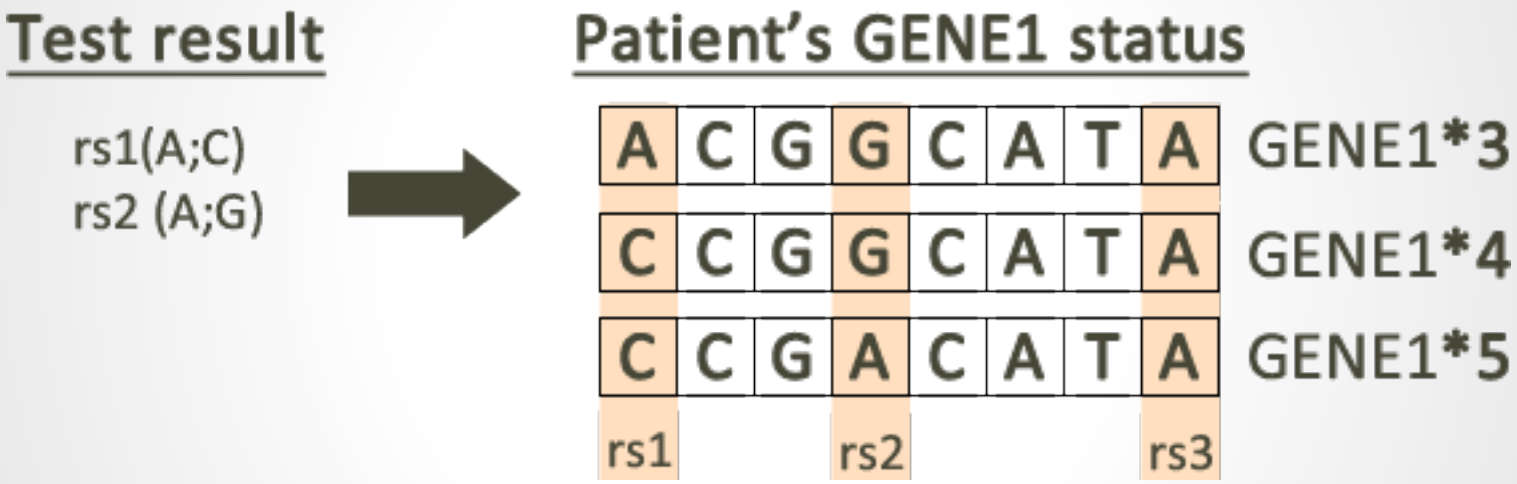
GENE1 allele definitions

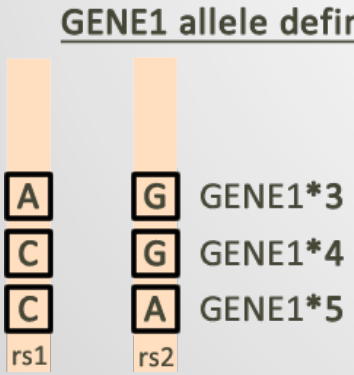


GENE1 allele definitions



In some cases, these test results are somewhat ambiguous, which can lead can lead to logical inconsistencies

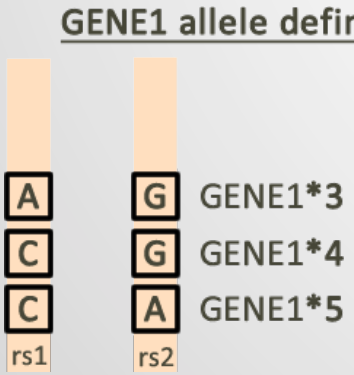




Three alleles match, but only two are allowed! The reasoner flags the ontology as inconsistent.

Allele definitions in OWL 2

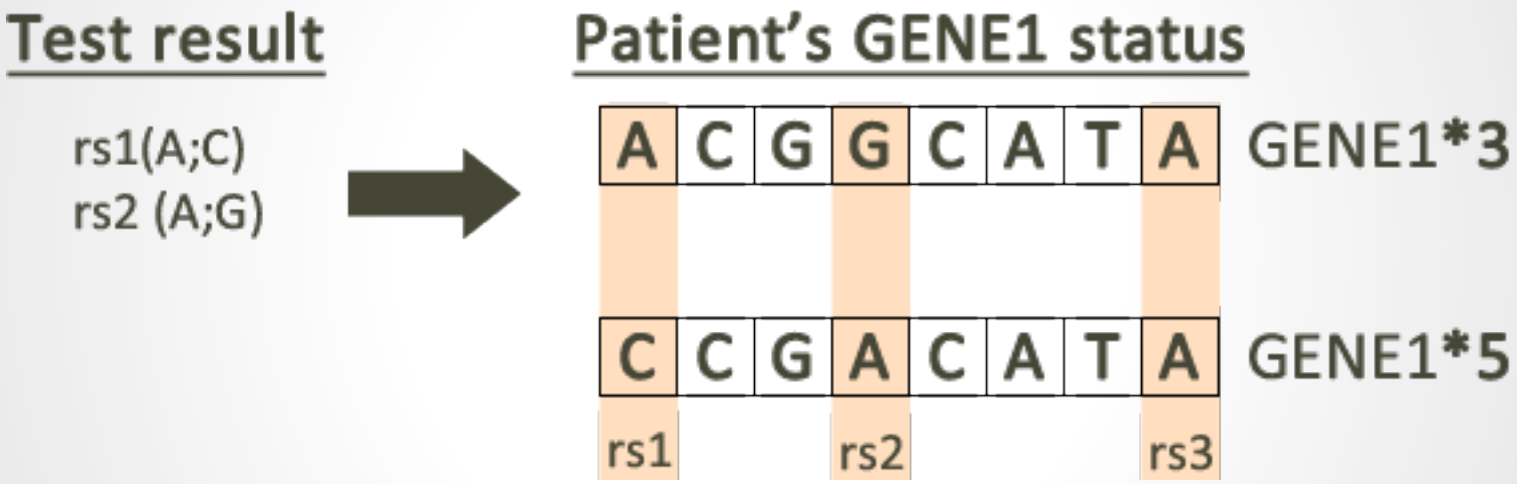
```
Class: 'human with GENE1*3'
    SubClassOf: human
    EquivalentTo:
            (has some rs1_A) and
            (has some rs2_G),
            has some GENE1*3

Class: 'human with GENE1*4'
[…]
```

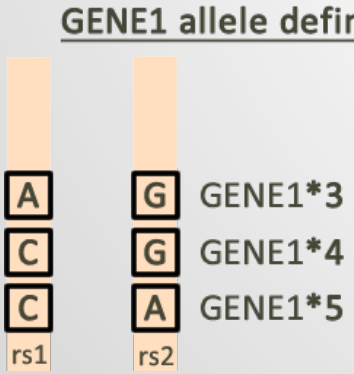


Experimental extension: telling the reasoner that each variant can only be used to call a single allele

```
Class: 'human with GENE1*3'
    SubClassOf: human
    EquivalentTo:
            (has some (rs1_A that (taken_by some GENE1*3)) and
            (has some (rs2_G that (taken_by some GENE1*3)),
            has some GENE1*3

Class: 'human with GENE1*4'
[…]

ObjectProperty: 'taken by'
    Characteristics: Functional

Class: GENE1
    EquivalentTo:
        GENE1_star_3
         or GENE1_star_4
         or GENE1_star_5
```

With these additions, the only possible combination is inferred, ontology becomes consistent





Inferred by HermiT, but not by TrOWL
See file *genomic-cds-haplotype-resolution-demo.owl*

Dosing guideline from an FDA drug label

Genotype		*CYP2C9*					
		**1/*1*	**1/*2*	**1/*3*	**2/*2*	**2/*3*	**3/*3*
VKORC1	*GG*	5-7 mg	5-7 mg	3-4 mg	3-4 mg	3-4 mg	0.5-2 mg
	AG	5-7 mg	3-4 mg	3-4 mg	3-4 mg	0.5-2 mg	0.5-2 mg
	AA	3-4 mg	3-4 mg	0.5-2 mg	0.5-2 mg	0.5-2 mg	0.5-2 mg

Dosing guideline from an FDA drug label

```
1 Class: 'human triggering CDS rule 9'
2 Annotations:
3       CDS_message "0.5-2 mg warfarin per day should be considered
4       as a starting dose range for a patient with this genotype
5       according to the warfarin drug label."
6 EquivalentTo:
7       (has some 'CYP2C9 *1') and
8       (has some 'CYP2C9 *3') and
9       (has exactly 2 rs9923231_T)
```

Describing an individual patient in OWL

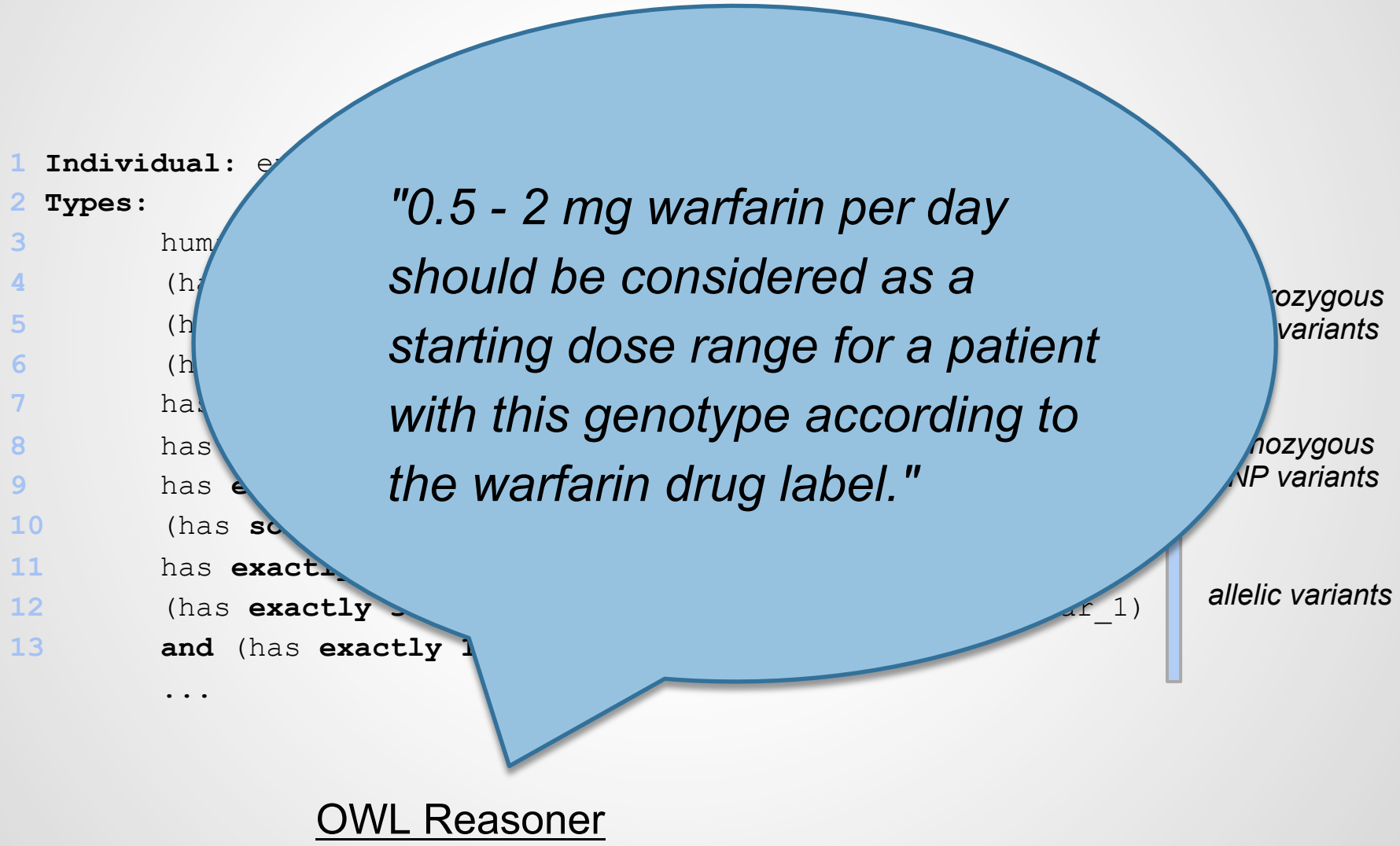
```
1  Individual: example_patient
2  Types:
3         human,
4         (has some rs1208_A)      and (has some rs1208_G),
5         (has some rs8192709_C)   and (has some rs8192709_T),
6         (has some rs9934438_A)   and (has some rs9934438_G),
7         has exactly 2 rs10264272_C,
8         has exactly 2 rs9923231_T,
9         has exactly 2 rs12720461_C,
10        (has some 'CYP2C9 *1') and (has some 'CYP2C9 *3'),
11        has exactly 2 'CYP2C19 *1',
12        (has exactly 3 CYP2D6) and (has exactly 2 'CYP2D6 *1')
13        and (has exactly 1 'CYP2D6 *2')
          ...
```

heterozygous SNP variants

homozygous SNP variants

allelic variants and CNV

Describing an individual patient in OWL



Phenotype (Genotype)	Therapeutic Dose Recommendation	Level of Evidence
PM (two inactive (*3-*8, *11-*16, *19-*21, *38, *40, *42) alleles)	Insufficient data to allow calculation of dose adjustment. Select alternative drug (e.g., citalopram, sertraline) or monitor amitriptyline and nortriptyline plasma concentration.	Published controlled studies of moderate quality* relating to phenotyped and/or genotyped patients or healthy volunteers, and having relevant pharmacokinetic or clinical endpoints.
IM (two decreased-activity (*9, *10, *17, *29, *36, *41) alleles or carrying one active (*1, *2, *33, *35) and one inactive (*3-*8, *11-*16, *19-*21, *38, *40, *42) allele, or carrying one decreased-activity (*9, *10, *17, *29, *36, *41) allele and one inactive (*3-*8, *11-*16, *19-*21, *38, *40, *42) allele)	Reduce dose by 25% and monitor plasma concentration or select alternative drug (e.g., citalopram, sertraline).	Published controlled studies of good quality* relating to phenotyped and/or genotyped patients or healthy volunteers, and having relevant pharmacokinetic or clinical endpoints.
UM (a gene duplication in absence of inactive (*3-*8, *11-*16, *19-*21, *38, *40, *42) or decreased-activity (*9, *10, *17, *29, *36, *41) alleles)	Insufficient data to allow calculation of dose adjustment. Select alternative drug (e.g., citalopram,sertraline) or monitor (E-10-hydroxy)amitriptyline plasma concentration.	Published controlled studies of moderate quality relating to phenotyped and/or genotyped patients or healthy volunteers, and having relevant

Definitions can become quite complex –
OWL reasoning helps identify inconsistencies and lacking definitions

Qualified cardinality restriction (exactly 2 alleles, all of them decreased-activity)

IM (two decreased-activity (*9, *10, *17, *29, *36, *41) alleles or carrying one active (*1, *2, *33, *35) and one inactive (*3-*8, *11-*16, *19-*21, *38, *40, *42) allele, or carrying one decreased-activity (*9, *10, *17, *29, *36, *41) allele and one inactive (*3-*8, *11-*16, *19-*21, *38, *40, *42) allele)	Reduce dose by 25% and monitor plasma concentration or select alternative drug (e.g., citalopram, sertraline).	Pu of ph ge he ha ph en
UM (a gene duplication in absence of inactive (*3-*8, *11-*16, *19-*21, *38, *40, *42) or decreased-activity (*9, *10, *17, *29, *36, *41) alleles)	Insufficient data to allow calculation of dose adjustment. Select alternative drug (e.g., citalopram,sertraline) or monitor (E-10-hydroxy)amitriptyline plasma concentration.	Pu of to ge he ha

Negation / 0 cardinality

Definitions can become quite complex –
OWL reasoning helps identify inconsistencies and lacking definitions

OWL 2 DL reasoning (especially realization / inferring types of patient individual) needs to be *fast!!*

TrOWL is massively more performant than the HermiT reasoner in classifying our demo ontology

	HermiT 1.3.8	TrOWL 1.1	MORe JFact 0.1.5	MORe Hermit 0.1.5
genomic-cds-light-demo.owl (2150 classes, 9500 axioms)	3 h 48 m	1.5 s	did not terminate within 1 h	did not terminate within 1 h
genomic-cds-demo.owl (2300 classes, 11000 axioms)	detected inconsistencies and terminated without making inferences	5.8 s	did not terminate within 1 h	did not terminate within 1 h

These ‘demo’ ontologies also include the genetic profile of a single patient.

The ‘light’ version contains only necessary&sufficient variants for alleles, full ontology also contains necessary variants for alleles

Ontologies have **ALCQ expressivity**.

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

Conclusions

(did not find an owl version, sorry)

Conclusions

- TrOWL performs best, but only provides partial results
- Need a better documentation of what inferences TrOWL can do / can not do. Could it be enhanced to reliably cover this use-case?
- Create reasoner optimized for *ALCQ* with qualified cardinality restrictions (with cardinalities >1)?
- There seems to be opportunity for (automated, semi-automated) modularization
- Reasoner should try to make inferences even when 'local' inconsistencies in a module are found (e.g., unresolvable ambiguity for a specific gene)…

Conclusions

- ‘Annoying’ limitation inherent to OWL 2D DL: no cardinality restrictions on transitive properties or property paths). Would make modelling much simpler. E.g.: *there are two baskets in the room, one basket contains exactly two eggs, the other contains exactly four eggs. Does the room contain at least three eggs?*

Thanks

W3C collaborators:

Michel Dumontier (Carleton University)

Robert R. Freimuth (Mayo Clinic)

Richard Boyce (University of Pittsburgh)

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]

http://www.genomic-cds.org/

http://safety-code.org/

Additional slides

You probably don‘t want to see them

Closing the world

```
Class: human
SubClassOf:
        has exactly 2 rs1,
        has exactly 2 rs2,
        has exactly 2 rs3,
        has exactly 2 GENE1

DisjointClasses:
rs1, rs2, rs3

DisjointClasses:
GENE1_star_1, GENE1_star_2,
   GENE1_star_3, …
```