Deriving Rules and Assertions From Pharmacogenomic Knowledge Resources In Support Of Patient Drug Metabolism Efficacy Predictions

Casey L. Overby\textsuperscript{1,2}, Beth Devine\textsuperscript{1}, Peter Tarczy-Hornoch\textsuperscript{1}, Ira Kalet\textsuperscript{1}

1 University of Washington, Seattle, WA
2 Columbia University, New York, NY (current affiliation)

JAMIA Journal Club
September 6, 2012
Pharmacogenomic evaluations of variability in drug exposure may facilitate personalized medicine.

- **Cancer patient taking tamoxifen**
- **The Doctor determines action based on individual characteristics**
- **The Doctor modifies current therapy**
- **The Doctor performs appropriate monitoring**
Pharmacogenomic evaluations of variability in drug exposure may facilitate personalized medicine

Cancer patient taking tamoxifen

The Doctor determines action based on individual characteristics

PGx profile data

The Doctor modifies current therapy

The Doctor performs appropriate monitoring

Motivation
Pharmacogenomic evaluations of variability in drug exposure may facilitate personalized medicine.

Cancer patient taking tamoxifen

The Doctor determines action based on individual characteristics

PGx profile data

The Doctor modifies current therapy

The Doctor performs appropriate monitoring

Motivation

TAM

CYP3A5

CYP2D6

CYP2C9

CYP3A5

N-Desmethyl TAM

4-OH TAM

Endoxifen
Phenotype scores based on pharmacogenetic evaluations can be predictive of endoxifen levels.

- Several examples of assigning gene or enzyme activity scores to predict drug or metabolite response/level [Borges et al 2010, Gaedigk et al 2008, Zineh et al 2004]
- Use a comparable scoring system
- Use scoring system to predict drug/metabolite levels
- How our scoring system and approach differs:
  - Incorporate contribution of several genes
  - Computational inference
Computational inference based on pharmacological knowledge

- The Drug Interaction Knowledge-base [Boyce et al 2009]

- Predicting metabolic inhibition and induction interactions

- Components
  - Evidence base
  - Knowledge base
  - The DIKB Evidence taxonomy

### DIKB Evidence Taxonomy

#### Evidence Types

**Clinical trial types**
- A pharmacokinetic clinical trial
- A genotyped pharmacokinetic clinical trial
- A phenotyped pharmacokinetic clinical trial

**Retrospective study types**
- A retrospective population PK study

**In vitro experiment types**
- A drug metabolism identification experiment
  - A CYP450, recombinant, drug metabolism identification experiment with possibly NO probe enzyme inhibitor(s)
  - A CYP450, human microsome, drug metabolism identification experiment using chemical inhibitors

**Observation based report**
- An observation-based ADE report (e.g. FDA Adverse Event Reporting System)
- A published observation-based ADE report

**Non-traceable statement types**
- A non-traceable, but possibly authoritative, statement
- A non-traceable drug-label statement
We use a computational approach to represent pharmacogenomic evaluations as a phenotype score.

1. Identify a clinical data source
2. Identify pharmacogenomic evidence source(s)
3. Derive assertions from sources to include in evidence base (EB) and knowledge base (KB)
4. Define rules to reason over EB and KB assertions and calculate phenotype scores using various approaches
   a. Initial literature base selection
   b. Weighting enzymes
   c. Scoring systems
5. Evaluate various approaches
Identify clinical data source

- Consortium on Breast Cancer Pharmacogenomics (COBRA)
- 30 breast cancer patients
- 20 mg/day tamoxifen
- genotype information: CYP3A5, -2D6, -2C9, and -2C19
- phenotype information: endoxifen ng/ml and NDM ng/ml at 4 months
Identify pharmacogenomics evidence sources

- Evidence Sources
  - PharmGKB (pharmgkb.org): drug metabolic knowledge (tamoxifen PK pathway)
  - SuperCYP (bioinformatics.charite.de/SuperCYP): gene variant - enzyme activity relationships
  - Primary literature review of gene variant - enzyme activity relationships (PharmGKB cited)
Methods

Scoring system

<table>
<thead>
<tr>
<th>Metabolizer activity</th>
<th>Gene</th>
<th>Allele activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultrapid metabolizer</td>
<td>CYP2C9</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>CYP2C19</td>
<td>Increased/increased</td>
</tr>
<tr>
<td></td>
<td>CYP2D6</td>
<td>Wild-type/increased; increased/increased</td>
</tr>
<tr>
<td></td>
<td>CYP3A5</td>
<td>Unknown</td>
</tr>
<tr>
<td>Extensive metabolizer</td>
<td>CYP2C9</td>
<td>Wild-type/wild-type</td>
</tr>
<tr>
<td></td>
<td>CYP2C19</td>
<td>Wild-type/wild-type</td>
</tr>
<tr>
<td></td>
<td>CYP2D6</td>
<td>Wild-type/wild-type; wild-type/decreased; wild-type/non-functional</td>
</tr>
<tr>
<td></td>
<td>CYP3A5</td>
<td>Unknown</td>
</tr>
<tr>
<td>Intermediate metabolizer</td>
<td>CYP2C9</td>
<td>Wild-type/non-functional</td>
</tr>
<tr>
<td></td>
<td>CYP2C19</td>
<td>Wild-type/non-functional</td>
</tr>
<tr>
<td></td>
<td>CYP2D6</td>
<td>Decreased/decreased; decreased/non-functional</td>
</tr>
<tr>
<td></td>
<td>CYP3A5</td>
<td>Unknown</td>
</tr>
<tr>
<td>Poor metabolizer</td>
<td>CYP2C9</td>
<td>Non-functional/non-functional</td>
</tr>
<tr>
<td></td>
<td>CYP2C19</td>
<td>Non-functional/non-functional</td>
</tr>
<tr>
<td></td>
<td>CYP2D6</td>
<td>Non-functional/non-functional</td>
</tr>
<tr>
<td></td>
<td>CYP3A5</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

Derived assertions from evidence and clinical data sources

- Clinical data source assertions
- Tamoxifen PK pathway assertions
- Enzyme contribution assertions
- ‘allelic variant’ – ‘enzyme activity’ assertions
- ‘genotype’ – ‘metabolizer activity’ assertions

Methods

Patient clinical data source facts

\[
\begin{align*}
\text{(patient} & \text{ (patient-MRN 001)} \\
\text{(population white)} & \\
\text{(medications tamoxifen)} & \\
\text{(enz-tested cyp2c9 cyp2c19 cyp2d6 cyp3a5))} \\
\text{(patient-enz-genotype} & \text{ (patient-MRN 001)} \\
\text{(enzyme cyp2c9)} & \\
\text{(allelic-variants cyp2c9_1 cyp2c9_1))} \\
\end{align*}
\]

PK pathway facts (PharmGKB)

\[
\begin{align*}
\text{(med-metabolite} & \text{ (name tamoxifen)} \\
\text{(downstream-metabolite N-desmethyeltamoxifen)} & \\
\text{(metabolizing-enzymes cyp3a4 cyp3a5))} \\
\text{(med-metabolite} & \text{ (name tamoxifen)} \\
\text{(downstream-metabolite 4-OHtamoxifen)} & \\
\text{(metabolizing-enzymes cyp2d6 cyp2b6 cyp2c9 cyp2c19 cyp3a4 cyp3a5))} \\
\end{align*}
\]

Allelic variant – enzyme activity facts (SuperCYP & PharmGKB)

\[
\begin{align*}
\text{(enz-allele-activity-publication} & \text{ (pubmed-id 17761971)} \\
\text{(enz-allele cyp2d6_10a)} & \\
\text{(enz-activity decreased)} & \\
\text{(drug tamoxifen)} & \\
\text{(evidence-type in vivo))} \\
\end{align*}
\]

Genotype – metabolizer activity facts (Review article)

\[
\begin{align*}
\text{(phenotype-activity-publication} & \text{ (pubmed-id 17761971)} \\
\text{(enzyme cyp2d6)} & \\
\text{(enz-allele1 decreased)} & \\
\text{(enz-allele2 nonfunctional)} & \\
\text{(phenotype-activity IM)} & \\
\text{(drug tamoxifen)} & \\
\text{(evidence-type review))} \\
\end{align*}
\]

Enzyme contribution facts (PharmGKB)

\[
\begin{align*}
\text{(enz-contribution} & \text{ (pubmed-id 15159443)} \\
\text{(drug tamoxifen)} & \\
\text{(enzyme cyp2d6)} & \\
\text{(major-minor major)} & \\
\text{(percent-contribution unknown))} \\
\end{align*}
\]
Derived assertions from evidence and clinical data sources

- Clinical data source assertions
- Tamoxifen PK pathway assertions
- Enzyme contribution assertions
- ‘allelic variant’ – ‘enzyme activity’ assertions
- ‘genotype’ – ‘metabolizer activity’ assertions

Patient clinical data source facts

(patient
(patient-MRN 001)
(population white)
(medications tamoxifen)
(enz-tested cyp2c9 cyp2c19 cyp2d6 cyp3a5))

(patient-enz-genotype
(patient-MRN 001)
(enzyme cyp2c9)
(allelic-variants cyp2c9_1 cyp2c9_1))

PK pathway facts (PharmGKB)

(med-metabolite
(name tamoxifen)
(downstream-metabolite N-desmethyltamoxifen)
(metabolizing-enzymes cyp3a4 cyp3a5))

(med-metabolite
(name tamoxifen)
(downstream-metabolite 4-OHtamoxifen)
(metabolizing-enzymes cyp2d6 cyp2b6 cyp2c9 cyp2c19 cyp3a4 cyp3a5))

Allelic variant – enzyme activity facts (SuperCYP & PharmGKB)

(enz-allele-activity-publication
(pubmed-id 17761971)
(enz-allele cyp2d6_10a)
(enz-activity decreased)
(drug tamoxifen)
(evidence-type in_vivo))

Genotype – metabolizer activity facts (Review article)

(phenotype-activity-publication
(pubmed-id 17761971)
(enzyme cyp2d6)
(enz-allele1 decreased)
(enz-allele2 nonfunctional)
(phenotype-activity IM)
(drug tamoxifen)
(evidence-type review))

Enzyme contribution facts (PharmGKB)

(enz-contribution
(pubmed-id 15159443)
(drug tamoxifen)
(enzyme cyp2d6)
(major-minor major)
(percent-contribution unknown))
Derived assertions from evidence and clinical data sources

- Clinical data source assertions
- Tamoxifen PK pathway assertions
- Enzyme contribution assertions
- ‘allelic variant’ – ‘enzyme activity’ assertions
- ‘genotype’ – ‘metabolizer activity’ assertions

Patient clinical data source facts

\begin{verbatim}
(patient
  (patient-MRN 001)
  (population white)
  (medications tamoxifen)
  (enz-tested cyp2c9 cyp2c19 cyp2d6 cyp3a5))

(patient-enz-genotype
  (patient-MRN 001)
  (enzyme cyp2c9)
  (allelic-variants cyp2c9_1 cyp2c9_1))
\end{verbatim}

PK pathway facts (PharmGKB)

\begin{verbatim}
(med-metabolite
  (name tamoxifen)
  (downstream-metabolite N-desmethyltamoxifen)
  (metabolizing-enzymes cyp3a4 cyp3a5))

(med-metabolite
  (name tamoxifen)
  (downstream-metabolite 4-OHtamoxifen)
  (metabolizing-enzymes cyp2d6 cyp2b6 cyp2c9 cyp2c19 cyp3a4 cyp3a5))
\end{verbatim}

Allelic variant – enzyme activity facts (SuperCYP & PharmGKB)

\begin{verbatim}
(enz-allele-activity-publication
  (pubmed-id 17761971)
  (enz-allele cyp2d6_10a)
  (enz-activity decreased)
  (drug tamoxifen)
  (evidence-type in_vivo))
\end{verbatim}

Genotype – metabolizer activity facts (Review article)

\begin{verbatim}
(phenotype-activity-publication
  (pubmed-id 17761971)
  (enzyme cyp2d6)
  (en-z-allele1 decreased)
  (en-z-allele2 nonfunctional)
  (phenotype-activity IM)
  (drug tamoxifen)
  (evidence-type review))
\end{verbatim}

Enzyme contribution facts (PharmGKB)

\begin{verbatim}
(enz-contribution
  (pubmed-id 15159443)
  (drug tamoxifen)
  (enzyme cyp2d6)
  (major-minor major)
  (percent-contribution unknown))
\end{verbatim}
Derived assertions from evidence and clinical data sources

- Clinical data source assertions
- Tamoxifen PK pathway assertions
- Enzyme contribution assertions
- ‘allelic variant’ – ‘enzyme activity’ assertions
- ‘genotype’ – ‘metabolizer activity’ assertions

Patient clinical data source facts

```prolog
(patient
  (patient-MRN 001)
  (population white)
  (medications tamoxifen)
  (enz-tested cyp2c9 cyp2c19 cyp2d6 cyp3a5))

(patient-enz-genotype
  (patient-MRN 001)
  (enzyme cyp2c9)
  (allelic-variants cyp2c9_1 cyp2c9_1))

PK pathway facts (PharmGKB)

(med-metabolite
  (name tamoxifen)
  (downstream-metabolite N-desmethyltamoxifen)
  (metabolizing-enzymes cyp3a4 cyp3a5))

(med-metabolite
  (name tamoxifen)
  (downstream-metabolite 4-OHtamoxifen)
  (metabolizing-enzymes cyp2d6 cyp2b6 cyp2c9 cyp2c19 cyp3a4 cyp3a5))

Allelic variant – enzyme activity facts (SuperCYP & PharmGKB)

(enz-allele-activity-publication
  (pubmed-id 17761971)
  (enz-allele cyp2d6_10a)
  (enz-activity decreased)
  (drug tamoxifen)
  (evidence-type in_vivo))

Genotype – metabolizer activity facts (Review article)

(phenotype-activity-publication
  (pubmed-id 17761971)
  (enzyme cyp2d6)
  (enz-allele1 decreased)
  (enz-allele2 nonfunctional)
  (phenotype-activity IM)
  (drug tamoxifen)
  (evidence-type review))

Enzyme contribution facts (PharmGKB)

(enz-contribution
  (pubmed-id 15159443)
  (drug tamoxifen)
  (enzyme cyp2d6)
  (major-minor major)
  (percent-contribution unknown))
```
Derived assertions from evidence and clinical data sources

- Clinical data source assertions
- Tamoxifen PK pathway assertions
- Enzyme contribution assertions
- ‘allelic variant’ – ‘enzyme activity’ assertions
- ‘genotype’ – ‘metabolizer activity’ assertions

Methods

Patient clinical data source facts

(patient
(patient-MRN 001)
(population white)
(medications tamoxifen)
(enz-tested cyp2c9 cyp2c19 cyp2d6 cyp3a5))

(patient-enz-genotype
(patient-MRN 001)
(enzyme cyp2c9)
(allelic-variants cyp2c9_1 cyp2c9_1))

PK pathway facts (PharmGKB)

(med-metabolite
(name tamoxifen)
(downstream-metabolite N-desmethyltamoxifen)
(metabolizing-enzymes cyp3a4 cyp3a5))

(enz-contribution
(pubmed-id 15159443)
(drug tamoxifen)
(enzyme cyp2d6)
(major-minor major)
(percent-contribution unknown))

Allelic variant – enzyme activity facts (SuperCYP & PharmGKB)

(enz-allele-activity-publication
(pubmed-id 17761971)
(enz-allele cyp2d6_10a)
(enz-activity decreased)
(drug tamoxifen)
(evidence-type in_vivo))

Genotype – metabolizer activity facts (Review article)

(phenotype-activity-publication
(pubmed-id 17761971)
(enzyme cyp2d6)
(enz-allele1 decreased)
(enz-allele2 nonfunctional)
(drug tamoxifen)
(evidence-type review))
Derived assertions from evidence and clinical data sources

- Clinical data source assertions
- Tamoxifen PK pathway assertions
- Enzyme contribution assertions
- ‘allelic variant’ – ‘enzyme activity’ assertions
- ‘genotype’ – ‘metabolizer activity’ assertions

Patient clinical data source facts

(patient
  (patient-MRN 001)
  (population white)
  (medications tamoxifen)
  (enz-tested cyp2c9 cyp2c19 cyp2d6 cyp3a5))

(patient-enz-genotype
  (patient-MRN 001)
  (enzyme cyp2c9)
  (allelic-variants cyp2c9_1 cyp2c9_1))

PK pathway facts (PharmGKB)

(med-metabolite
  (name tamoxifen)
  (downstream-metabolite N-desmethyltamoxifen)
  (metabolizing-enzymes cyp3a4 cyp3a5))

(med-metabolite
  (name tamoxifen)
  (downstream-metabolite 4-OHtamoxifen)
  (metabolizing-enzymes cyp2d6 cyp2b6 cyp2c9 cyp2c19 cyp3a4 cyp3a5))

Allelic variant – enzyme activity facts (SuperCYP & PharmGKB)

(enz-allele-activity-publication
  (pubmed-id 17761971)
  (enz-allele cyp2d6_10a)
  (enz-activity decreased)
  (drug tamoxifen)
  (evidence-type in_vivo))

Genotype – metabolizer activity facts (Review article)

(phenotype-activity-publication
  (pubmed-id 17761971)
  (enzyme cyp2d6)
  (enz-allele1 decreased)
  (enz-allele2 nonfunctional)
  (drug tamoxifen)
  (evidence-type review))

Enzyme contribution facts (PharmGKB)

(enz-contribution
  (pubmed-id 15159443)
  (drug tamoxifen)
  (enzyme cyp2d6)
  (major-minor major)
  (percent-contribution unknown))
Defined rules to calculate phenotype scores:

evidence selection

- **Enzyme-oriented approach**
  - 85 ‘allelic variant’ – ‘enzyme activity’ facts (SuperCYP)

- **Drug-oriented approach**
  - 7 facts (SuperCYP) + facts from 9 additional publications cited in PharmGKB

**Enzyme allele activity facts (SuperCYP & PharmGKB)**

- (enz-allele-activity-publication (pubmed-id 17761971))
- (enz-allele cyp2d6_10a)
- (enz-activity decreased)
- (drug tamoxifen)
- (evidence-type in_vivo)
Defined rules to calculate phenotype scores: phenotype score calculation

- Enzyme activity scoring: Un-weighted approach
  - Increased allele activity: 1.5
  - Wild-type allele activity: 1.0
  - Decreased allele activity: 0.5
  - Non-functional allele activity: 0.0

- Enzyme activity scoring: Weighted approach
  - Gene encodes major metabolizing enzyme: 1.0
  - Gene encodes minor metabolizing enzyme: 0.5

- Metabolizer activity scoring approach
  - Ultrarapid metaboliser (UM): 1.5
  - Extensive metaboliser (EM): 1.0
  - Intermediate metaboliser (IM): 0.5
  - Poor metaboliser (PM): 0.0

Enzyme contribution facts (PharmGKB)
(enz-contribution (pubmed-id 15159443) (drug tamoxifen) (enzyme cyp2d6) (major-minor major) (percent-contribution unknown))
Defined rules to calculate phenotype scores: phenotype score calculation

- **Enzyme activity scoring: Un-weighted approach**
  - Increased allele activity: 1.5
  - Wild-type allele activity: 1.0
  - Decreased allele Activity: 0.5
  - Non-functional allele activity: 0.0

  \[
  p_{score} = \sum_{i=1}^{g} \left( \sum_{j=1}^{2} a_{ij} \right)
  \]

- **Enzyme activity scoring: Weighted approach**
  - Gene encodes major metabolizing enzyme: 1.0
  - Gene encodes minor metabolizing enzyme: 0.5

  \[
  p_{score} = \sum_{i=1}^{g} \left( \sum_{j=1}^{2} a_{ij} w_i \right)
  \]

- **Metabolizer activity scoring approach**
  - Ultrarapid metaboliser (UM): 1.5
  - Extensive metaboliser (EM): 1.0
  - Intermediate metaboliser (IM): 0.5
  - Poor metaboliser (PM): 0.0

  \[
  p_{score} = \sum_{i=1}^{g} \text{pheno}_i \quad p_{score} = \sum_{i=1}^{g} \left( \text{pheno}_i w_i \right)
  \]
Tamoxifen case study evaluation: statistical methods

- Are phenotype scores predictive of endoxifen/NDM plasma ratio?
  - Quantile linear regression
    - IV: phenotype score
    - DV: endoxifen/NDM plasma ratio

- Are phenotype scores more predictive than simple methods for representing genotypes?
  - Quantile linear regression
    - IV: simple metric
    - DV: endoxifen/NDM plasma ratio

- Simple metrics:
  - Single gene
  - Multiple genes
  - Additive model (wt/wt=0, vt/wt=1, vt/vt=2)

![Distribution of genotypes](chart.png)

**Methods**

CYP3A5 Wt=*1, Vt=*3,*6
CYP2D6 Wt=*1, Vt=*4,*6
CYP2C9 Wt=*1, Vt=*2,*3
CYP2C19 Wt=*1, Vt=*2.
Are phenotype scores predictive of endoxifen/NDM plasma ratio?

Results

Enzyme Activity Scoring

Metabolizer Activity Scoring
Are phenotype scores predictive of endoxifen/NDM plasma ratio?

**Results**

**Enzyme Activity Scoring**

**Metabolizer Activity Scoring**

![Graph showing correlation between phenotype scores and endoxifen/NDM plasma ratio.](image)

- $r^2 = 0.10^a$
- $p = 0.04$

- $r^2 = 0.04$
- $p = 0.49$

- $r^2 = 0.03$
- $p = 0.65$
Are phenotype scores more predictive than simple metrics?

One scoring approach performed better than simple metrics to represent genotypes for each gene.
Are phenotype scores more predictive than simple metrics?

One scoring approach performed better than a simple metric to represent multiple genes.
Challenges and limitations

- Providing score in a useful way
  - Promotes appropriate use to make informed health decisions
  - Ensure the score is correct and current for the patient
- Validity of score to predict response to therapy should be tested
  - CYP2D6 enzyme activity $\rightarrow$ endoxifen concentrations $\rightarrow$ clinical outcomes
- Practical issues of implementing this model in an EHR should be investigated
  - Data and evidence source access, security, connection and coordination
Contributions and future directions

• Single score that characterizes combined activity of enzymes in a patient-specific way

• Download and data access capabilities to support automated extraction of evidence

• Other forms of clinical data and additional knowledge resources

• Validation of methods for calculating phenotype scores with a separate dataset

• More complicated polygenic models are needed to represent heterogeneity in clinical outcomes
Acknowledgements

- Anonymous reviewers
- Dr. Ken Thummel (UW Pharmaceutics)
- Dr. Houda Hachad (UW Pharmaceutics)
- Dr. Isabelle Ragueneau-Majlessi (UW Pharmaceutics)
- Dr. David Flockhart & COBRA group for sharing data
- UW Informatics research training grant (NIH NLM #T15 LM7442)
- UW Genome training grant (NIH NHGRI #T32 HG000053)
- Institute of Translational Health Sciences (NIH NCRR 1 UL1 RR 025014)
- Columbia Training in Biomedical Informatics (NIH NLM #T15 LM007079)
- Mentored Clinical Scientist Training Grant (AHRQ 5K08 HS014739, PI: Devine)
Questions?

Contact information:
Casey Lynnette Overby, PhD
casey@dbmi.columbia.edu