Deep learning to predict the impact of rare variation in drug metabolism genes

Russ B Altman, MD, PhD
Bioengineering, Genetics, Medicine, Biomedical Data Science
Stanford University
March 2021
PharmGKB & Helix Groups

Greg McInness
Adam Lavertu

Rachel Dalton & Erica Woodahl, U. Montana
Pharmacogenetics = variation in drug response due to genetic differences

• Drug should work as expected
• Change the dose of the drug
• Increased chance of toxicity for drug
• Use another drug
<table>
<thead>
<tr>
<th>Level 1A</th>
<th>CYP2D6</th>
<th>Drug</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CYP2D6</td>
<td>fluvoxamine</td>
<td>Efficacy/Toxicity, Depressive Mental Disorder</td>
</tr>
<tr>
<td></td>
<td>CYP2D6</td>
<td>tropicetron</td>
<td>Efficacy, Vomiting</td>
</tr>
<tr>
<td></td>
<td>CYP2D6</td>
<td>codeine</td>
<td>Efficacy/Toxicity, Pain</td>
</tr>
<tr>
<td></td>
<td>CYP2D6</td>
<td>amitriptyline, antidepressants, clomipramine, desipramine, doxepin, imipramine, nortriptyline, trimipramine</td>
<td>Dosage/Toxicity, Depression</td>
</tr>
<tr>
<td></td>
<td>CYP2D6</td>
<td>doxepin</td>
<td>Efficacy</td>
</tr>
</tbody>
</table>
As much as 23% of people in the US have a compromised ability to metabolize opioids.

1-5% are poor metabolizers => CODEINE DOES NOT WORK
1-21% are ultra metabolizer => MORPHINE SPIKES IN BLOOD
“Star” Alleles = Haplotypes of pharmacogenes

*1 = Wildtype (Reference Sequence)
*2 = some combination of SNP alleles
*3 = another combination
*4 = etc...

CYP2D6 has 161+ observed haplotypes (many are common)

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Variants (variant = variants with dbSNP rsID)</th>
<th>Impact</th>
<th>Function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2D6*1A</td>
<td></td>
<td>normal function</td>
<td>Kimura et al. 1989</td>
<td></td>
</tr>
<tr>
<td>CYP2D6*1B</td>
<td>3829G&gt;A</td>
<td>normal function</td>
<td>Marez et al. 1997</td>
<td></td>
</tr>
<tr>
<td>CYP2D6*1C</td>
<td>1979C&gt;T</td>
<td>normal function</td>
<td>Marez et al. 1997</td>
<td></td>
</tr>
<tr>
<td>CYP2D6*1D</td>
<td>2576C&gt;A</td>
<td>normal function</td>
<td>Marez et al. 1997</td>
<td></td>
</tr>
<tr>
<td>CYP2D6*1E</td>
<td>1870T&gt;C</td>
<td>normal function</td>
<td>Sachse et al. 1997</td>
<td></td>
</tr>
<tr>
<td>CYP2D6*2B</td>
<td>1038C&gt;T, 1662G&gt;C, 2851C&gt;T, 4181G&gt;C</td>
<td>normal function</td>
<td>Marez et al. 1997</td>
<td></td>
</tr>
</tbody>
</table>
### Some drugs metabolized by CYP2D6

<table>
<thead>
<tr>
<th>Antidepressants</th>
<th>Beta Blockers</th>
<th>Anti-cancer</th>
<th>Antipsychotics</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amitriptyline</td>
<td>Alprenolol</td>
<td>Tamoxifen</td>
<td>Haloperidol</td>
<td>Mexiletine</td>
</tr>
<tr>
<td>Clomipramine</td>
<td>Carvedilol</td>
<td></td>
<td>Perphenazine</td>
<td>Minaprine</td>
</tr>
<tr>
<td>Desipramine</td>
<td>Propafenone</td>
<td></td>
<td>Risperidone</td>
<td>Nebivolol</td>
</tr>
<tr>
<td>Imipramine</td>
<td>Bupranolol</td>
<td></td>
<td>Thioridazine</td>
<td>Nortriptyline</td>
</tr>
<tr>
<td>Fluoxetine</td>
<td>Clonidine</td>
<td></td>
<td>Zuclopenthixol</td>
<td>Nebivolol</td>
</tr>
<tr>
<td>Paroxetine</td>
<td>Debrisoquine</td>
<td></td>
<td>Ondansetron</td>
<td>Chlorpromazine</td>
</tr>
<tr>
<td>Tamoxetine</td>
<td>Metoprolol</td>
<td></td>
<td>Oxycodone</td>
<td>Codeine</td>
</tr>
<tr>
<td>Trimipramine</td>
<td>Propranolol</td>
<td></td>
<td>Perhexilene</td>
<td>Debrisoquine</td>
</tr>
<tr>
<td>Venlafaxine</td>
<td>Timolol</td>
<td></td>
<td>Phenacetin</td>
<td>Dextromethorphan</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Aripiprazole</td>
<td>Phenformin</td>
</tr>
</tbody>
</table>

Some drugs metabolized by CYP2D6 include:
- Antidepressants: Amitriptyline, Clomipramine, Desipramine, Imipramine, Fluoxetine, Paroxetine, Tamoxetine, Trimipramine, Venlafaxine
- Beta Blockers: Alprenolol, Carvedilol, Propafenone, Bupranolol, Clonidine, Debrisoquine, Metoprolol, Propranolol, Timolol
- Anti-cancer: Tamoxifen
- Antipsychotics: Haloperidol, Perphenazine, Risperidone, Thioridazine, Zuclopenthixol, Ondansetron, Oxycodone, Perhexilene, Aripiprazole
- Other: Mexiletine, Minaprine, Nebivolol, Nortriptyline, Nebivolol, Chlorpromazine, Phenacetin, Phenformin, Methamphetamine, Bufuralol, Chlorpheniramine, Chlorburalol.


Codeine is bioactivated to morphine, a strong opioid agonist, by the hepatic cytochrome P450 2D6 (CYP2D6); hence, the efficacy and safety of codeine are governed by CYP2D6 activity. Polymorphisms are a major cause of CYP2D6 variability. We summarize evidence from the literature supporting this association and provide therapeutic recommendations for codeine based on CYP2D6 genotype. This document is an update to the 2012 Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for CYP2D6 genotype and codeine therapy.

cypalleles.ki.se. Clinical phenotype data are available for common alleles (Supplementary Tables S1–S5 online). However, many alleles have not been evaluated in clinical trials, and their clinical phenotypes are predicted based on the expected functional impact of their defining genetic variation or are extrapolated based on in vitro functional studies using different substrates.

Genetic test interpretation
Most clinical laboratories report CYP2D6 genotype using the star (*) allele nomenclature and may provide interpretation of genotypes.
Rare variants in UK Biobank Exomes

• Evaluated: variation in 8 key pharmacogenes (metabolizing enzymes, transporters) including CYP2D6
• 478 predicted-deleterious variants across all 8
• 244 of these not in gnomAD (resource for population variation)
• 6.1% of individuals carry one novel deleterious variant
• Each individual has an average of 12 drugs for which unusual response might expected
• Novel variants enriched in non-European populations
We need methods to assess the impact of novel or rare variations!
How can we predict the function of the novel haplotypes observed in population surveys?

==

How can we bring clinical pharmacogenetics to patients with rare variants?
Deep Learning

- Deep Learning is based on an analogy to neural processing = neural networks
- cf. processing of light in the retina.

http://www.arn.org/docs/glicksman/eyw_041101.htm
Deep Learning

Patterns of Local Contrast

Face Features

Input Layer

Hidden Layer 1

Hidden Layer 2

Output Layer

Pre-training

convolutional layers

64% tabby

33% Siamese

0.1% wooden spoon

dense layers

Transfer learning

convolutional layers (frozen)

95% beagle

4% basset hound

new dense layers
CYP2D6 “Activity Score”

Method for predicting metabolic phenotype from genotype (* allele)

Assigns a score to each haplotype based on known functional variants = sum of the haplotype scores

<table>
<thead>
<tr>
<th>Value assigned</th>
<th>Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>*1xN, *2xN, *35xN</td>
</tr>
</tbody>
</table>

Adapted from Gaedigk et al, 2007

Adapted from Gaedigk et al, 2017
IDEA for CYP2D6 Transfer Learning

• Generate 50,000 sequences on a natural gnoMAD background with known CYP2D6 variations embedded/spiked into these sequences
• Estimate the Activity Score of these sequences
• Train a model to learn how to assign Activity Scores
  • (This should force CNN to learn key sequence features)
• Use SPARSE experimental (360 samples) & database data (~60 * alleles with known function) to refine final layers
• Predict function of haplotypes & assess

Transfer learning used to train network

Activity score classification

Pretrain on simulated data

Measured activity regression

Tune on real data via semi-supervised learning

Star allele Classifier

Fine tune on star alleles sequences
CYP2D6: 14,407 base pairs in 9 exons

Experimental data (Erica Woodahl & Rachel Dalton)

(360 liver samples, sequenced CYP2D6, 2 activity measurements/sample)

161 variant sites

60 intronic, 56 exonic, 45 upstream/downstream
Gold standard data available from databases

A. CYP2D6 Star Allele Data

Star allele sequences and functions from PharmVar. Divided into training and validation sets.

- Training data:
  - 4 Normal
  - 8 Decreased
  - 19 No function

- Validation data:
  - 7 Normal
  - 6 Decreased
  - 12 No function

- Uncurated:
  - 71 Uncurated
Representation of (phased) sequence data

B. Data formatting

**Input:** *CYP2D6* star allele sequence  
**Output:** One-hot encoded sequence and annotation data

C. Functional prediction

**Input:** One-hot encoded sequence and annotation data  
**Output:** Functional probabilities: normal function score, no function score.
Binary annotations for variants

• In coding region?
• Allele freq < 0.05?
• Deleterious per vote of CADD, DANN, FATHMM, LOFTEE?
• Indel?

• In methylation mark?
• DNA hypersensitivity site?
• TF binding site?
• Known eQTL site?
• Known active site amino acid?
Transfer learning used to train network

Activity score classification

Pretrain on simulated data

Measured activity regression

Tune on real data via semi-supervised learning

Star allele Classifier

Fine tune on star alleles sequences
D. Conversion of ordinal scores to functional classes

**Input:** Functional probabilities

**Output:** CYP2D6 functional prediction
Validate predictions using *in vitro* data from large study.

71% variance explained by functional labels

Star allele function measured

---

Functional Characterization of Wild-type and 49 CYP2D6 Allelic Variants for N-Desmethyltamoxifen 4-Hydroxylation Activity

Yuka Muroi¹, Takahiro Saito¹, Masamitsu Takahashi¹, Kanako Sakuyama², Yui Niinuma¹, Miyabi Ito¹, Chihiro Tsukada¹, Kiminori Ohta², Yasuyuki Endo², Akifumi Oda³, Noriyasu Hirawawa¹ and Masahiro Hiratsuka¹.*
Evaluation of model components and annotations

Component evaluation

Annotation evaluation

Accuracy

Full model, No anno, No transfer, No transfer & No anno, Rare variant, Deleterious, Active site, TFBS, INDEL, Coding variant, eQTL, DNase, Methyl

Test accuracy, Train accuracy
Performance on novel uncurated star alleles

- Patients with these haplotypes would currently be told: "no information available"
Importance scores for core variants in star allele sequences

Legend
Functional prediction
- Normal
- Decreased
- No function

Variant annotation
- Benign
- Deleterious

Variant importance
- Less important
- More important

Importance using DeepLift
Shrikumar, Greenside & Kundaje
https://arxiv.org/abs/1704.02685

Conclusions

• Pharmacogenomics is entering clinical care and is useful chiefly in the context of common variants

• UK Biobank analysis indicates large numbers of people with variations in pharmacogenes that are not currently characterized, thus limiting impact.

• Deep learning methods (in this case with transfer learning) hold promise for predicting clinically useful pharmacogenomic phenotypes for novel (chiefly rare) variations in important genes.
Thanks!

russ.altman@stanford.edu
www.pharmgkb.org
Table 2. Drug-gene side effect relationship results. Associations are presented in three groups: drug-gene pairs with CPIC guidelines, pairs with no guidelines but evidence in PharmGKB, and novel associations. Phenotype is the gene phenotype (IM: Intermediate Metabolizer, PM: Poor Metabolizer, RM: Rapid Metabolizer, UM: Ultrarapid Metabolizer, IF: Increased Function, PF: Poor Function). Odds ratio is the odds ratio relative to normal metabolizer or normal function alleles. * indicates significance with Bonferroni adjusted p-value threshold of 1.0 x 10⁻⁵. Only results with a standard error less than 0.2 are included.

<table>
<thead>
<tr>
<th>Group</th>
<th>Drug</th>
<th>Gene</th>
<th>Level of Evidence</th>
<th>Phenotype</th>
<th>ICD-10</th>
<th>Code definition</th>
<th>Odds ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPIC Guidance</td>
<td>citalopram</td>
<td>CYP2C19</td>
<td>1A</td>
<td>IM</td>
<td>B02</td>
<td>Herpes zoster</td>
<td>0.53</td>
<td>8.76E-05</td>
</tr>
<tr>
<td></td>
<td>simvastatin</td>
<td>SLCO1B1</td>
<td>1A</td>
<td>IF</td>
<td>M65</td>
<td>Synovitis and tenosynovitis</td>
<td>1.82</td>
<td>1.42E-04</td>
</tr>
<tr>
<td></td>
<td>amitriptyline</td>
<td>CYP2C19</td>
<td>1A</td>
<td>RM</td>
<td>R53</td>
<td>Malaise and fatigue</td>
<td>1.55</td>
<td>1.74E-04</td>
</tr>
<tr>
<td></td>
<td>amitriptyline</td>
<td>CYP2C19</td>
<td>1A</td>
<td>UM</td>
<td>J30</td>
<td>Vasomotor and allergic rhinitis</td>
<td>1.94</td>
<td>2.75E-04</td>
</tr>
<tr>
<td></td>
<td>codeine</td>
<td>CYP2D6</td>
<td>1A</td>
<td>PM</td>
<td>A52</td>
<td>Late syphilis</td>
<td>1.78</td>
<td>3.30E-04</td>
</tr>
<tr>
<td></td>
<td>ibuprofen</td>
<td>CYP2C9</td>
<td>1A</td>
<td>PM</td>
<td>E13</td>
<td>Other specified diabetes mellitus</td>
<td>2.00</td>
<td>4.90E-04</td>
</tr>
<tr>
<td></td>
<td>clobidogrel</td>
<td>CYP2C19</td>
<td>1A</td>
<td>RM</td>
<td>B08</td>
<td>Viral infections characterized by skin and mucous membrane lesions</td>
<td>0.59</td>
<td>5.17E-04</td>
</tr>
<tr>
<td></td>
<td>tamoxifen</td>
<td>CYP2D6</td>
<td>1A</td>
<td>IM</td>
<td>C50</td>
<td>Malignant neoplasm of breast</td>
<td>0.62</td>
<td>6.98E-04</td>
</tr>
<tr>
<td></td>
<td>simvastatin</td>
<td>SLCO1B1</td>
<td>1A</td>
<td>PF</td>
<td>M79</td>
<td>Unspecified soft tissue disorders</td>
<td>1.49</td>
<td>7.46E-04</td>
</tr>
<tr>
<td></td>
<td>simvastatin</td>
<td>SLCO1B1</td>
<td>1A</td>
<td>DF</td>
<td>M65</td>
<td>Synovitis and tenosynovitis</td>
<td>1.79</td>
<td>7.75E-04</td>
</tr>
<tr>
<td>No Guidance</td>
<td>citalopram</td>
<td>CYP2D6</td>
<td>3</td>
<td>IM</td>
<td>J45</td>
<td>Asthma</td>
<td>1.44</td>
<td>9.13E-05</td>
</tr>
<tr>
<td></td>
<td>citalopram</td>
<td>CYP2D6</td>
<td>3</td>
<td>IM</td>
<td>J50</td>
<td>Heart failure</td>
<td>1.56</td>
<td>1.12E-04</td>
</tr>
<tr>
<td></td>
<td>simvastatin</td>
<td>CYP2C9</td>
<td>3</td>
<td>PM</td>
<td>J01</td>
<td>Acute sinusitis</td>
<td>1.74</td>
<td>1.56E-04</td>
</tr>
<tr>
<td></td>
<td>citalopram</td>
<td>CYP2D6</td>
<td>3</td>
<td>IM</td>
<td>J64</td>
<td>Unspecified pneumoconiosis</td>
<td>1.56</td>
<td>5.74E-04</td>
</tr>
<tr>
<td></td>
<td>propranolol</td>
<td>CYP2D6</td>
<td>4</td>
<td>IM</td>
<td>O86</td>
<td>Other puerperal infections</td>
<td>1.85</td>
<td>6.38E-04</td>
</tr>
<tr>
<td>Novel associations</td>
<td>diazepam</td>
<td>CYP2C9</td>
<td>NA</td>
<td>PM</td>
<td>M19</td>
<td>Osteoarthritis</td>
<td>2.33</td>
<td>4.52E-06*</td>
</tr>
<tr>
<td></td>
<td>zopiclone</td>
<td>CYP2C9</td>
<td>NA</td>
<td>IM</td>
<td>H91</td>
<td>Unspecified hearing loss</td>
<td>2.20</td>
<td>1.73E-05</td>
</tr>
<tr>
<td></td>
<td>loratadine</td>
<td>CYP2D6</td>
<td>NA</td>
<td>IM</td>
<td>M16</td>
<td>Osteoarthritis of hip</td>
<td>1.98</td>
<td>1.20E-04</td>
</tr>
<tr>
<td></td>
<td>tramadol</td>
<td>CYP2B6</td>
<td>NA</td>
<td>PM</td>
<td>H61</td>
<td>Disorders of external ear</td>
<td>1.95</td>
<td>1.86E-04</td>
</tr>
<tr>
<td></td>
<td>quinine</td>
<td>SLCO1B1</td>
<td>NA</td>
<td>IF</td>
<td>N39</td>
<td>Disorders of urinary system</td>
<td>1.95</td>
<td>1.87E-04</td>
</tr>
</tbody>
</table>