A comprehensive landscape of CYP2D6 variation across 30,000 individuals

Shishi Luo[1], Ruomu Jiang[1], Jasmine Dhaliwal[1], Sharoni Jacobs[1], William Lee[1]

[1] Helix LLC

Introduction

The cytochrome P450 family 2 subfamily D member 6 (CYP2D6) gene is involved in the metabolism of numerous important prescription drugs. CYP2D6 is highly polymorphic and harbors structural variation due to its homology with the pseudogene CYP2D7, making it challenging to genotype. This not only hampers efforts to apply CYP2D6 genotyping in a clinical setting but also limits the ability of researchers to uncover novel drug associations with CYP2D6 variants.

- We present a comprehensive assessment of CYP2D6 variation in a cohort of 30,000 individuals.
- Our **clinically validated** pipeline reports **106 star alleles**, including structural variants and amplifications.
- We find that less comprehensive tests are potentially miss-genotyping 17% of samples. This leads to the mischaracterization of 7.7% of ultrarapid metabolizers and 4.4% of poor metabolizers.

What current tests are missing

Commercial PGx tests typically report on the most common CYP2D6 star alleles, omitting a large fraction of the 100+ alleles documented by the Pharmacogene Variation Consortium (PharmVar.org). Using a panel of star alleles based on Del Tredici et al [1] that is representative of these commercial tests ('short panel' below), we quantify the consequences of using panels that are not comprehensive.

Samples with an incorrect star allele using short panel. (Out of 32,277)	16.8%
Ultrarapid metabolizers mischaracterized with short panel. (Out of 775)	7.7%
Poor metabolizers mischaracterized with short panel. (Out of 1997)	4.4%
Samples with allele of uncertain function. (Out of 32,277)	1.6%

Alleles in short panel: *1, *2, *3, *4, *4N, *5, *6, *9, *10, *17, *29, *35, *36, *41. (Amplifications also reported)

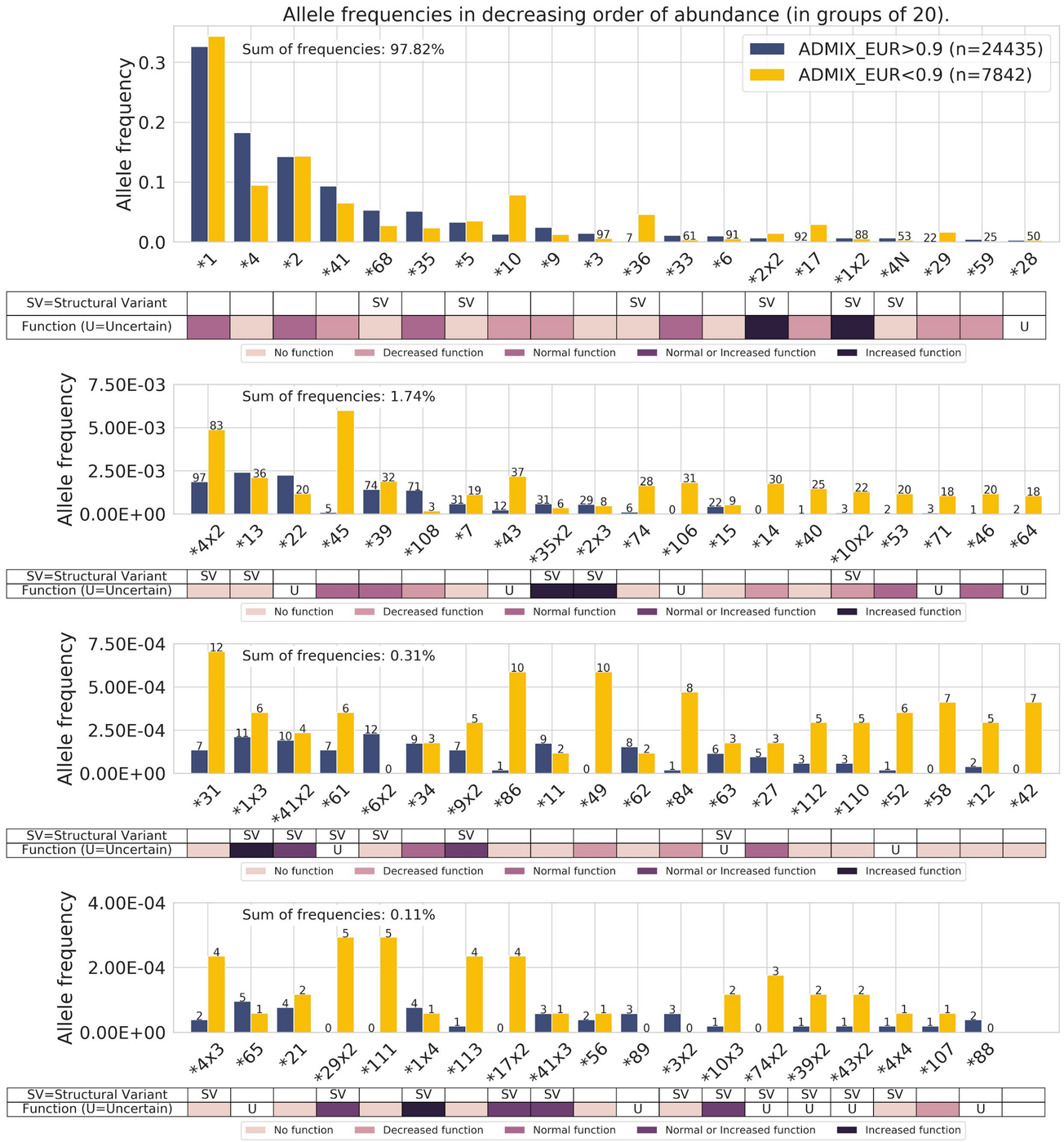
Methods

Our method builds upon existing NGS-based star allele callers Aldy [2] and Stargazer [3]:

1. Reads mapping to the CYP2D6-CYP2D7 locus are filtered and processed.
2. Allele depths at 97 variants and read counts per exon are determined.
3. The **most likely** allele combination that explains the data observed in step 2 is reported, along with a **quality score**.

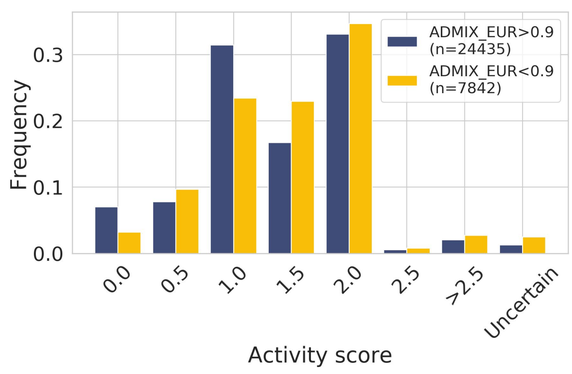
This method is clinically validated with average accuracy of **99.5%.**

Genotypic variation



- ADMIX_EUR is the ADMIXTURE [4] coefficient for the European reference population (range is 0.0-1.0).
- For alleles that appear fewer than 99 times in a group, the allele count is displayed on the bar.
- Not shown: 24 alleles that were each reported only once; 36 alleles that were not detected in the cohort; *82, because it is outside the pipeline's reportable range.

Phenotypic variation



Metabolizer status	Activity Score
Poor	0.0
Intermediate	0.5
Normal	1.0-2.0 inclusive*
Ultrarapid	>2.0

*Boundaries continue to be debated. Based on Table 1 in [5].

References:
[1] Del Tredici, Andria L., et al. "Frequency of CYP2D6 alleles including structural variants in the United States." Frontiers in Pharmacology 9 (2018): 305.
[2] Numanagić, Ibrahim, et al. "Allelic decomposition and exact genotyping of highly polymorphic and structurally variant genes." Nature Communications 9.1 (2018): 828.
[3] Lee, Seung-been, et al. "Stargazer: a software tool for calling star alleles from next-generation sequencing data using CYP2D6 as a model." Genetics in Medicine 21.2 (2019): 361.
[4] Alexander, David H., John Novembre, and Kenneth Lange. "Fast model-based estimation of ancestry in unrelated individuals." Genome Research 19.9 (2009): 1655-1664.
[5] Goetz, Matthew P., et al. "Clinical pharmacogenetics implementation consortium (CPIC) guideline for CYP2D6 and tamoxifen therapy." Clinical Pharmacology & Therapeutics 103.5 (2018): 770-777.

Visit Helix.com