Missing Signals: Copy Number and Rare Loss of Function Variants in *CYP2C19*

Daniel De Sloover, Hannah Williams, MPH, Cynthia L. Neben, PhD, Scott Topper, PhD, FACMG, Annette Leon, PhD, FACMG, Alicia Y. Zhou, PhD

Color Genomics, Burlingame, CA

Introduction

Clinical pharmacogenomic (PGx) implementation originated with genotyping technologies. As a result, this technology is unable to take novel variants into account. This legacy continues today with many standard PGx analyses unable to analyze novel genetic variation. Instead, genotyping assay focus on identifying only previously described haplotypes, called "star-alleles." Star-1 (*1) is commonly considered the default state and indicates that none of the interrogated alternative alleles are present. However, *1 does not exclude the possibility that a novel loss of function (LOF) or gain of function (GOF) variant is present.[1]

Due to ascertainment bias in the original design and implementation of many genotyping arrays, these methods may miss potentially impactful variants in individuals of non-European ancestry.[2] In addition, novel variants are often rare, which may also interfere with assay performance due to primer and probe binding inhibition resulting in allele drop-out. To standardize assay design and reporting of *CYP2C19* via genotyping, the Pharmacogenomics Working Group of the Association for Molecular Pathology (AMP) recently published a recommendation for a minimum set of alleles (*2,*3,*17) and a secondary set of tier 2 alleles.[3]

Color derives *CYP2C19* diplotypes from next-generation sequencing (NGS) data and reports on the established variants from PharmVar (*1, *2, *3, *4A, *4B, *10, *17).[4,5] In this study, we explored the data beyond those targets, to characterize the additional variation that is present. Here, we present novel predicted LOF (pLOF) CYP2C19 variants observed in 56,851 de-identified, research-consented individuals.

Methods

All individuals were ordered a Color test by a healthcare provider and provided informed consent to have their de-identified information and sample used in anonymized studies. Laboratory procedures were performed at the Color laboratory. Briefly, DNA was extracted, enriched for select regions using SureSelect XT probes, and then sequenced using NextSeq 500/550 or NovaSeq 6000 instrument. Sequence reads were aligned against the human genome reference GRCh37.p12, and variants were identified using a suite of bioinformatic tools.

Diplotype calls were computed using an implementation of Aldy[6] and Diplo, an internally developed tool, described previously.[7] Novel variants that are not included in the PharmVar allele tables were queried with the following quality filters in place: exonic calls depth >50X, GATK quality score >300, and allele fraction >30%.

Of the 56,851 individuals, genetic ancestry was calculated on 24,793 using fastNGSadmix[8] and the 1000 Genomes Project reference panel. Individuals were assigned genetic ancestry based on the seven geographical groups (Central/South Asian, SAS; East Asian, EAS; European, EUR; Near Eastern, NEA; Oceanian, OCE; Sub-Saharan African, SSA; and American, AME) and two admixed groups (African American/Afro-Caribbean, AAC and Latino, LAT) as described by Pharmacogenomics Knowledgebase (PharmGKB).[9] A third admixed group, Other Admixed, was included to account for admixed individuals who could not be classified as African American/Afro-Caribbean or Latino.

Conclusions

- This work indicates that NGS is a commensurable tool for clinically reporting PGx diplotypes and can help reduce the ethnic disparities in PGx testing.
- We have identified a long tail of recurrent, non-canonical variants are expected to have clear functional consequences, yet would likely be reported as "normal" or equivocal in existing clinical PGx genotyping assays.
- Consistent with recent population sequencing analyses[13], CNVs are an especially important LOF signal in *CYP2C19*. We observed a 0.27% global population frequency of CNVs in *CYP2C19*. Exon 1-5 and 2-5 deletions are the primary variants observed in ~4% of Caucasians reported as *1/*1.
- Further functional characterization of these novel variants is necessary to fully understand their impact on downstream drug metabolism.
- Genotyping assays will continue to miss a large number of important functional and haplotype-resolving variants that NGS is able to reveal. It is necessary to quickly interrogate and understand the functional impact of these novel variants and how they contribute to drug response.

Results

Table 1. Summary statistics of phenotypes mapped to PharmVar alleles by genetic ancestry.

Haplotypes were determined using the complete set of variants in the current PharmVar definition (*1-*19, *22-*26, *28-*35: Version 4.0.1). Phenotypes were determined from diplotypes via the most recent CPIC guideline for *CYP2C19* and Voriconazole therapy.[10] We observed phenotype frequencies across ancestry groups that are consistent with previous reports. Genetic ancestry was not calculated for 32,058 individuals, and as such those individuals were not included in this summary.

	Individuals	Ultra Rapid %	Rapid %	"Normal" %	Intermediate %	Poor %	Likely Intermediate, Likely Poor, and Indeterminate
Total	24,793	1,134 (4.2%)	6,594 (24.7%)	9,841 (39.7%)	6,796 (27.4%)	765 (3.1%)	211 (0.9%)
AAC	337	5.6%	24.9%	32.3%	31.5%	3.3%	2.4%
SAS	999	1.9%	8.7%	32.5%	45.6%	9.3%	1.9%
EAS	581	0.9%	7.6%	37.7%	39.9%	13.8%	0.2%
EUR	20,037	4.6%	26.7%	39.5%	26.0%	2.5%	0.7%
LAT	1244	2.3%	19.2%	52.5%	23.8%	1.8%	0.4%
AME	22	4.6%	-	86.4%	9.1%	-	-
NEA	79	6.3%	25.3%	41.8%	22.8%	2.5%	1.3%
SSA	529	4.4%	24.4%	30.6%	31.2%	3.2%	6.2%
OCE	17	-	11.76%	35.3%	47.1%	5.9%	-
Other	948	2.7%	19.1%	41.7%	31.5%	4.5%	0.4%

Figure 1. pLOF variants in CYP2C19 by type.

The majority of novel exonic variants in *CYP2C19* were pLOF missense (52.8%) and CNVs (33.2%). pLOF missense variants were classified using the variant effect prediction utility in Ensembl and the majority consensus pathogenicity calls between REVEL, SIFT, PolyPhen 2, DANN, MutationAssessor, MutationTaster, dbNSFP, FATHMM, MetaLR, and PROVEAN.[11] n = 56,581

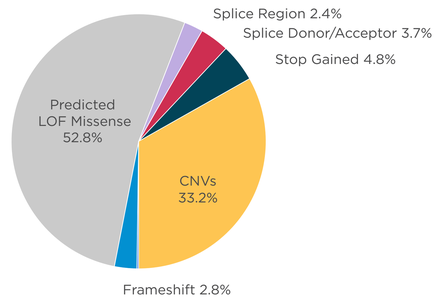


Figure 2. Schematic of canonical and pLOF variants in *CYP2C19*.

Top: pLOF variants by type and position within *CYP2C19*. Size of the bubble is proportional to the number of observations.[12] Bottom: Genomic structural variants. Yellow are copy losses. Blue are copy gains. n = 56,581.

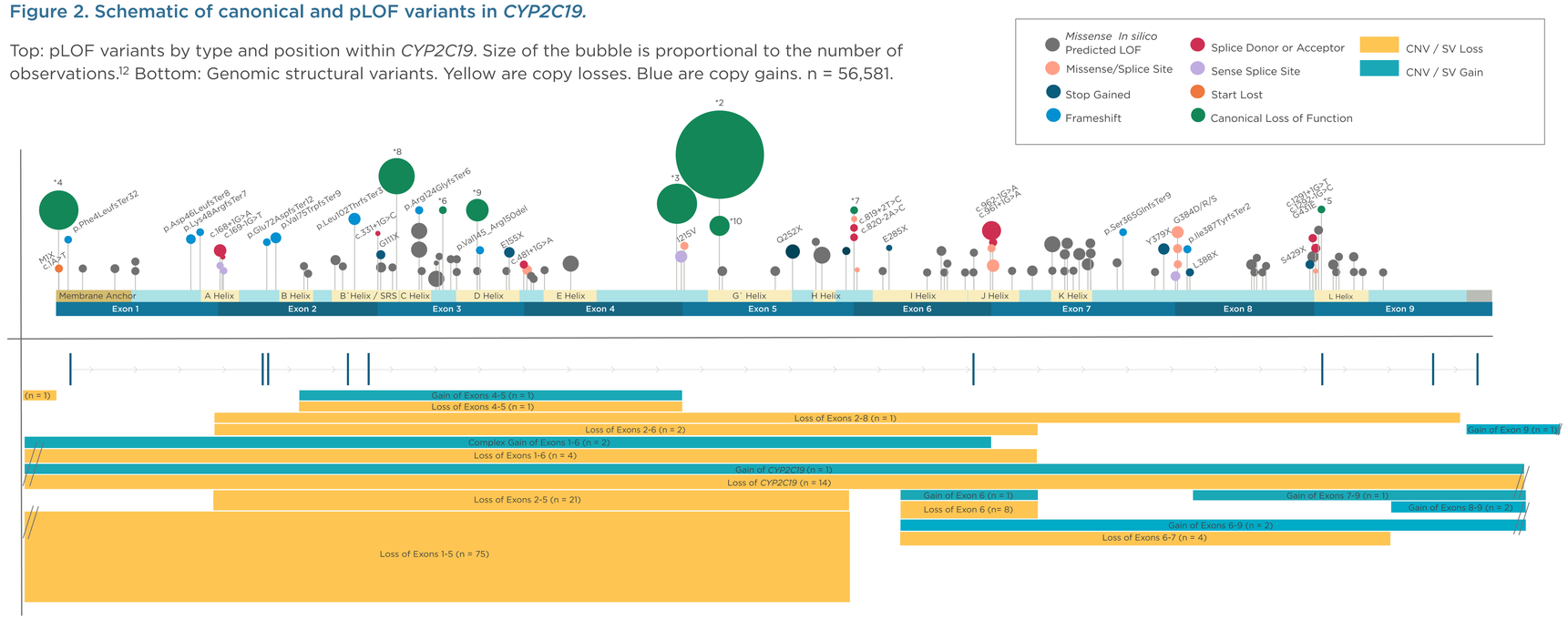


Figure 3. Copy number and variant allele frequency plots of recurrent CNVs in *CYP2C19* discovered through depth-based calling.

(A) Recurrent loss of exons 1-5 was observed in 75 individuals of primarily self-reported European descent. Interestingly, loss of exons 1-5 was recently reported in the Finnish pation at 0.4% - 0.8% frequency.[13] (B) Loss of exons 2-5 was observed in 21 self-reported European individuals. (C) Loss of exons 6 and 7 was observed in four Chinese individuals. (D) Gain of exons 8 and 9 was observed in two South Asian individuals. CNV calling algorithms include circular binary segmentation (cbs), fused lasso (flasso), and cumulative sums (cusum).

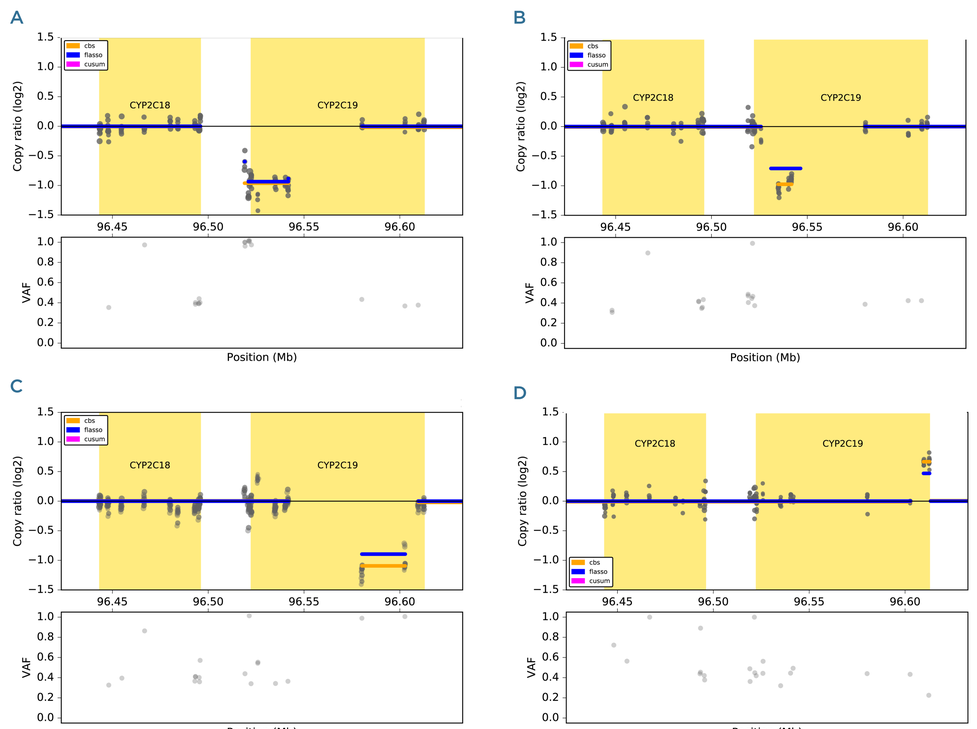
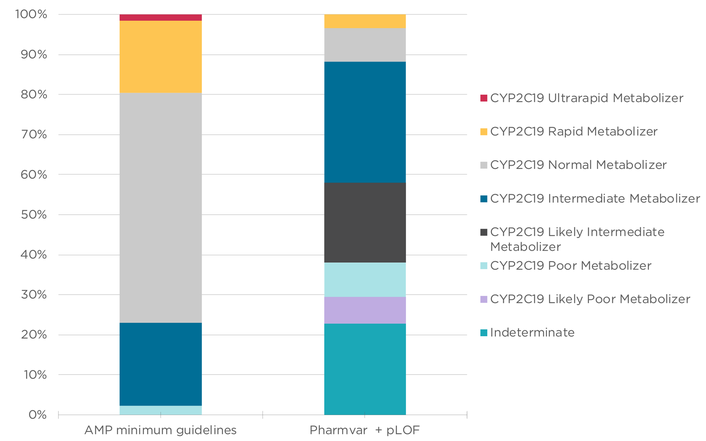


Figure 4. Potential change of reported *CYP2C19* phenotype.

Of the 24,793 individuals with genetic ancestry, 738 (2.98%) had a pLOF variant and/or allele other than *2, *3, or *17 that could result in a different phenotype than would have been reported.



References

1. Kalman LV, Agúndez J, Appell ML, et al. *Clin Pharmacol Ther.* 2016.
2. Lachance J, Tishkoff SA. *Bioessays.* 2013.
3. Pratt VM, Del Tredici AL, Hachad H, et al. *J Mol Diagn.* 2018.
4. Gaedigk A, Ingelman-Sundberg M, Miller NA, et al. Clin *Pharmacol Ther.* 2018.
5. Klein TE, Chang JT, Cho MK, et al. *Pharmacogenomics J.* 2001.
6. Numanagić I, Malikić S, Ford M, et al. *Nat Commun.* 2018.
7. Color Medication Response Genetic Test. http://www.color.com/pgxwhitepaper. Published 2018. Accessed May 4, 2018.
8. Jørsboe E, Hanghøj K, Albrechtsen A. *Bioinformatics.* 2017.
9. Huddart R, Fohner AE, Whirl-Carrillo M, et al. *Clin Pharmacol Ther.* 2019.
10. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for CYP2C19 and Voriconazole Therapy. *Clin Pharmacol Ther.* 2018.
11. McLaren W, Gil L, Hunt SE, et al. *Genome Biol.* 2016.
12. Jay JJ, Brouwer C. *PLoS One.* 2016.
13. Santos M, Niemi M, Hiratsuka M, et al. *Genet Med.* 2018.