Mind the Gaps: Novel Loss of Function *CYP2C19* Variants in 48,657 Individuals

Daniel De Sloover, Zheng Tan, PhD, FACMG, Zachary Langley, Hannah Williams, MPH, 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, it necessarily ignored novel variants. This legacy continues today with many standard PGx analyses ignoring most genetic variation and instead 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, previously uncharacterized, loss of function (LOF) or gain of function (GOF) variant is present[1].

Due to ascertainment bias in the single nucleotide polymorphisms (SNPs) included on many genotyping arrays and genotyping assays, these methods may miss potentially impactful variants in individuals of non-European ancestry[2]. Rare variants may also interfere with assay performance due to primer and probe binding inhibition resulting in allele drop-out. Conversely, due to the abundance of novel variants that next generation sequencing (NGS) will detect, it has been suggested that replacing such genotyping approaches is not technically feasible[3]. The field, therefore, finds itself at a crux where it is necessary to quickly discover, interrogate, and understand the breadth of novel variants to deepen our understanding in this field.

We derive diplotypes from NGS data by only reporting on the established variants from PharmVar for clinical analysis[4–6]. In this study, we explored the data beyond those targets, to characterize the additional variation that is present. Here, we present novel expected LOF *CYP2C19* variants observed in 48,657 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 human genome reference GRCh37.p12, and variants are identified using a suite of bioinformatic tools.

Diplotype calls were computed using an implementation of Aldy[3] and Diplo, an internally developed tool, as 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%.

Conclusions

- This work indicates that NGS is a commensurable tool for clinically reporting PGx diplotypes and can be used to reduce the ethnic disparities in PGx testing.
- 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. In this cohort, there were 3.7X more individuals with a rare (< 0.01% AF) expected or predicted LOF variant than those with a common expected or predicted LOF variant.
- Consistent with recent population sequencing analyses, CNVs are an especially important LOF signal in *CYP2C19*. We estimate that ~4% of Caucasians reported as **1/*1* carry a LOF CNV variant.
- Further functional characterization of these novel variants is necessary to more accurately represent the populations undergoing clinical PGx testing.

Results

Table 1. Summary statistics

Novel variants are defined as variants that are not present in the Pharmvar allele tables, even if they have an assigned rsID. Novel exonic expected LOF variants are defined as any copy number variant (CNV), start lost, stop gained, frameshift, splice donor and splice acceptor variants (+/- 2), or missense variant where a majority of *in silico* functional predictors agreed on a deleterious consequence. To note, several canonical LOF defined alleles do not meet this *in silico* threshold.

		Individuals n (%)	Individuals w/ Novel Variant n (%)	Individuals w/ Novel Exonic Variant n (%)	Individuals w/ Novel Exonic eLOF Variant n (%)
Total		48,657 (100%)	5,156 (10.6%)	1,824 (3.7%)	436 (0.9%)
Ethnicity*	African	653 (1.3%)	266 (40.7%)	113 (17.3%)	6 (0.9%)
	Asian	2,579 (5.3%)	503 (19.5%)	204 (7.9%)	38 (1.5%)
	Middle Eastern	171 (0.4%)	45 (26.3%)	17 (9.9%)	4 (2.3%)
	Unknown	967 (2.0%)	130 (13.4%)	44 (4.6%)	9 (0.9%)
	Indian	401 (0.8%)	83 (20.7%)	55 (13.7%)	7 (1.7%)
	Hispanic	2,207 (4.5%)	286 (13.0%)	110 (5.0%)	32 (1.4%)
	Native American	97 (0.2%)	8 (8.2%)	2 (2.1%)	0 (0.0%)
	Caucasian	38,530 (79.2%)	3,426 (8.9%)	1,138 (3.0%)	308 (0.8%)
	Multiple Ethnicities	3,052 (6.3%)	409 (13.4%)	141 (4.6%)	32 (1.0%)

*Ethnicity was reported by the individual; unknown includes information not provided.

Figure 1. Expected and predicted LOF variants in *CYP2C19* by type

The majority of novel exonic variants in *CYP2C19* were predicted LOF missense (61.1%) and CNVs (24.8%). Predicted LOF missense variants were classified through the variant effect prediction utility in Ensembl, using the majority consensus pathogenicity calls between REVEL, SIFT, PolyPhen 2, DANN, MutationAssessor, MutationTaster, dbNSFP, FATHMM, MetaLR, and PROVEAN[9].

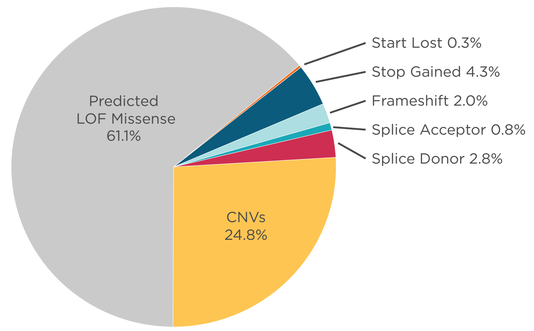
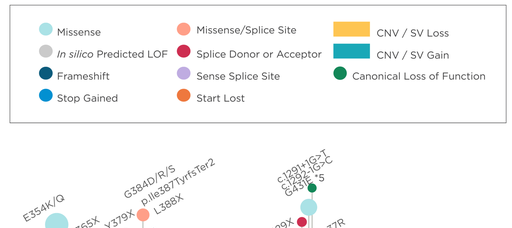


Figure 2. Schematic of novel exonic variants in *CYP2C19*

Of the 436 individuals who had an expected or predicted LOF variant, 399 individuals had a rare expected or predicted LOF variant. Rare variants are defined as variants with a population frequency <0.01 allele frequency. Top: Size of the bubble is approximately proportional to the number of observations. Bottom: Genomic structural variants. Yellow is copy losses. Blue is copy gains.



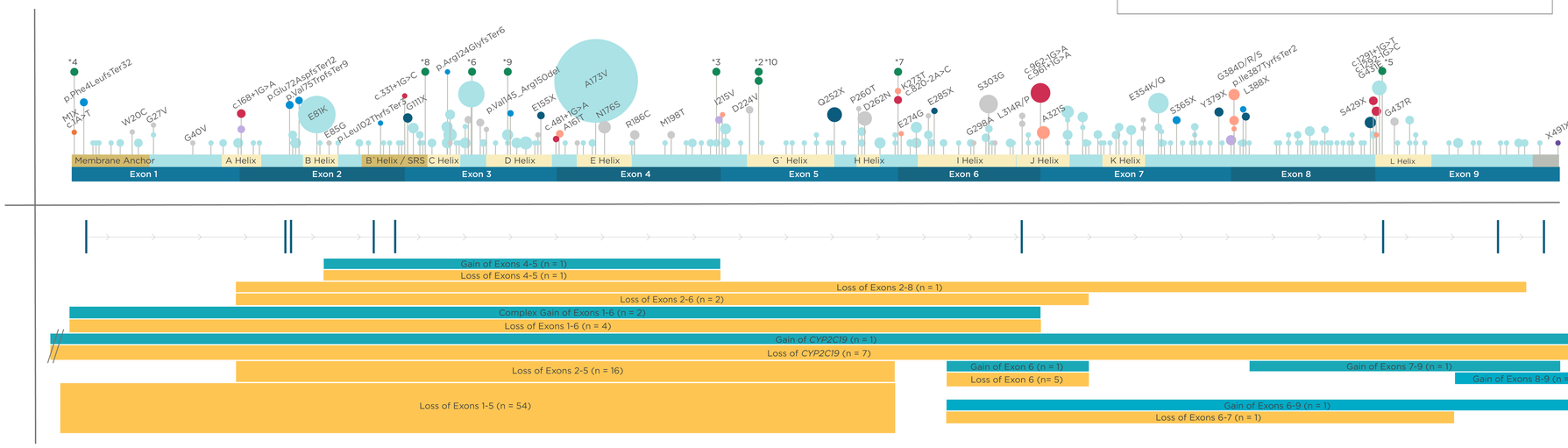


Figure 3. Copy number and variant allele frequency plots of recurrent CNVs in *CYP2C19* discovered through depth-based calling

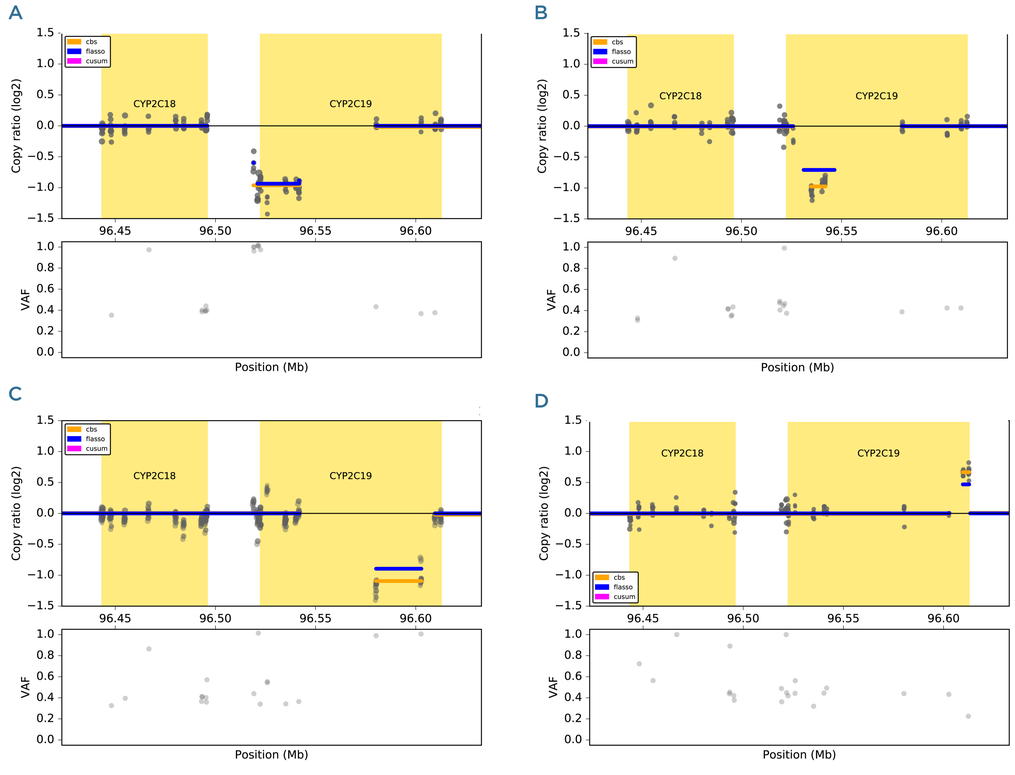
CNV calling algorithms include circular binary segmentation (cbs), fused lasso (flasso), and cumulative sums (cusum).

(A) Recurrent loss of exons 1-5 was observed in 54 individuals of primarily European descent. Of those, 10 (18.5%) individuals also had a loss of *CYP2C18*. Interestingly, loss of exons 1-5 was recently reported in the Finnish population at 0.4% - 0.8% frequency[9].

(B) Loss of exons 2-5 was observed in 16 Caucasian individuals.

(C) Loss of exons 6 and 7 was observed in three Chinese individuals.

(D) Gain of exons 8 and 9 was observed in two South Asian individuals.



References

1. Kalman LV, Agúndez JAG, Appell ML, et al. *Clin Pharmacol Ther.* 2016.
2. Lachance J, Tishkoff SA. *Bioessays.* 2013.
3. Ji Y, Si Y, McMillin GA, Lyon E. *Expert Rev Mol Diagn.* 2018.
4. Numanagić I, Malikić S, Ford M, et al. *Nat Commun.* 2018.
5. Gaedigk A, Ingelman-Sundberg M, Miller NA, et al. *Clin Pharmacol Ther.* 2018.
6. Klein TE, Chang JT, Cho MK, et al. *Pharmacogenomics J.* 2001.
7. Color Genomics. Color Medication Response Genetic Test. https://s3.amazonaws.com/color-static-prod/pdfs/whitepaper/Color_PGx_White_Paper.pdf. Published October 1, 2019. Accessed December 3, 2019.
8. Relling MV, Klein TE. *Clin Pharmacol Ther.* 2011.
9. McLaren W, Gil L, Hunt SE, et al. *Genome Biol.* 2016.
10. Santos M, Niemi M, Hiratsuka M, et al. *Genet Med.* 2018.