Preemptive PGx testing using NGS reveals novel predicted LOF variants in *TPMT* and *NUDT15*

Hannah E. Williams, MPH, Daniel De Sloover, Cynthia L. Neben, PhD, Julian R. Homburger, PhD, Scott Topper, PhD, FACMG, Annette Leon, PhD, FACMG, Alicia Y. Zhou, PhD
Color Genomics, Burlingame, CA

Introduction

Thiopurines, including mercaptopurine, thioguanine, and azathioprine, are frequently used in the treatment of Acute Lymphoblastic Leukemia (ALL) and Irritable Bowel Disease (IBD).[1] Genetic variations resulting in the loss of function (LOF) or decreased enzyme activity of thiopurine S-methyltransferase (*TPMT*) and nudix hydrolase 15 (*NUDT15*) have been shown to be associated with hematopoietic toxicity, a potentially life-threatening complication associated with thiopurine treatment. As such, the Clinical Pharmacogenetics Implementation Consortium (CPIC) has published thiopurine dosing guidelines for preemptive pharmacogenetic (PGx) testing of *TPMT* and *NUDT15*, in order to mitigate the risk of drug toxicity without compromising overall treatment outcomes.[2]

Genotyping technologies are a common approach to PGx testing, however, this approach typically involves the interrogation of known variants leaving rare and novel variants undetected. The frequency of decreased and LOF variants in *TPMT* and *NUDT15* has been shown to vary across ancestry groups.[3,4] As a result, it is important to identify both known and novel variants in *TPMT* and *NUDT15* across broad populations, to further characterize the variation found in these pharmacogenes. Color, a population health company, derives diplotypes from next-generation sequencing (NGS) data and reports on established variants from PharmVar for clinical analysis. In this study, we explored the data beyond those established variants, in an effort to further characterize the variation found in *TPMT* and *NUDT15* across different ancestries. Here we present novel predicted LOF (pLOF) variants in *TPMT* and *NUDT15* by genetic ancestry using NGS technologies in 24,779 de-identified, research-consented individuals.

Methods

All individuals were ordered a Color test by a healthcare provider and consented 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 identified using a suite of bioinformatic tools. Diplotype calls were computed using an implementation of Aldy3 and Diplo, an internally developed tool, as described previously.[5] Novel pLOF variants were defined as any copy number variant, start lost, stop gained, frameshift, splice donor and splice acceptor variants (± 2), or missense variants where the majority of in silico functional predictors agreed on a deleterious consequence and are not established variants from PharmVar. Variants were queried with the following quality filters in place: exonic calls depth >50X, GATK quality score >300, and allele fraction >30%.

Genetic ancestry was calculated using fastNGSadmix[6] using 1000 Genomes Project as the 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).[7] A third admixed group (Other) was included to account for admixed individuals who could not be classified as African American/Afro-Caribbean or Latino.

Results

Table 1. Cohort Characteristics

The majority of individuals in our cohort were female (77.5%), with an average age of 47.5 (± 14.3 SD) years. SD, standard deviation. IQR, interquartile range.

	Characteristic	Individuals, N (%)
Total		24,779 (100.0)
Gender	Female	19,201 (77.5)
	Male	5,578 (22.5)
Age (Years)	Mean (SD)	47.5 (14.3)
	Median (IQR)	47.1 (36.0, 58.2)
Genetic Ancestry	EUR	20,022 (80.8)
	LAT	1,245 (5.0)
	SAS	999 (4.0)
	Other	948 (3.8)
	EAS	581 (2.4)
	SSA	529 (2.1)
	AAC	337 (1.4)
	NEA	79 (0.3)
	AME	22 (0.1)
	OCE	17 (0.1)
***TPMT* Phenotype**	Normal Metabolizer	22,665 (91.5)
	Intermediate Metabolizer	1,927 (7.7)
	Possible Intermediate Metabolizer	9 (0.1)
	Poor Metabolizer	45 (0.2)
	Indeterminate	133 (0.5)
***NUDT15* Phenotype**	Normal Metabolizer	23,953 (96.6)
	Intermediate Metabolizer	791 (3.2)
	Poor Metabolizer	23 (0.1)
	Indeterminate	12 (0.1)

Results

Figure 1. Observed combined genotype of *TPMT* and *NUDT15*.

Individuals were grouped by total number of known risk alleles in *TPMT* and *NUDT15*: 0 risk alleles; 1 risk allele; ≥ 2 risk alleles. Approximately 11% of individuals in the cohort had 1 or more risk alleles. A total of 3 individuals had a combined genotype of 3 risk alleles. n=24,779.

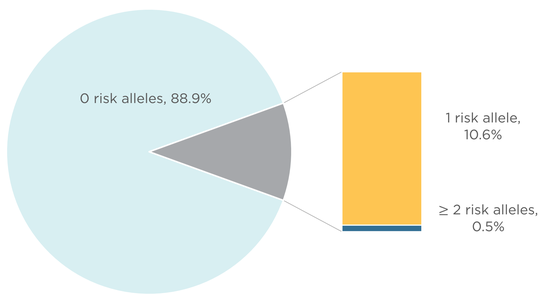


Figure 2. Observed combined genotype of *TPMT* and *NUDT15* by ancestry group.

Individuals were grouped by total number of known risk alleles in *TPMT* and *NUDT15*: 0 risk alleles; 1 risk allele; ≥ 2 risk alleles. Individuals of American or East Asian ancestry had the highest proportion of risk alleles, 22.7% and 19.6%, respectively.

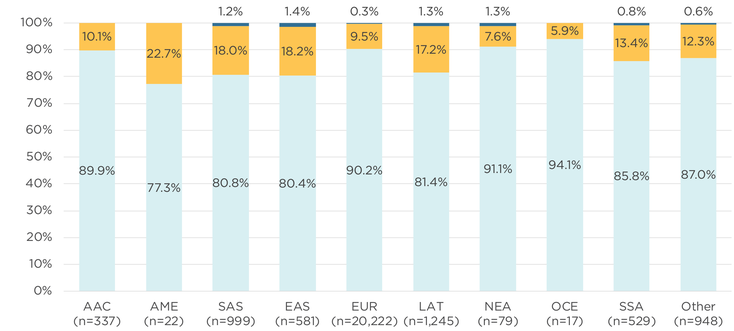


Figure 3. Novel exonic variants observed in *TPMT*.

We observed 31 novel exonic variants in *TPMT*, of which 18 (54.5%) were pLOF. The majority of novel variants were missense (n=20; pLOF=7; teal dots), followed by splice acceptor/donor (n=6; orange dots), stop gained (n=3; yellow dots), and frameshift (n=2; purple dots). All splice acceptor/donor, stop gained, and frameshift variants were classified as pLOF.

Figure 4. Novel exonic variants observed in *NUDT15*.

We observed 43 novel exonic variants in *NUDT15*, of which 22 (51.2%) were pLOF. The majority of novel variants were missense (n=34; pLOF=13; teal dots), followed by splice acceptor/donor (n=2; orange dots), inframe (n=1; gray dot), and frameshift (n=6; purple dots). All splice acceptor/donor, inframe, and frameshift variants were classified as pLOF.

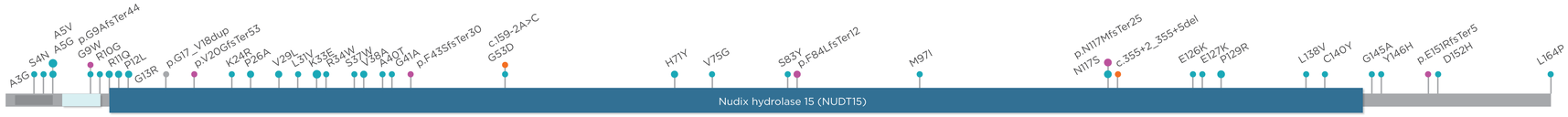


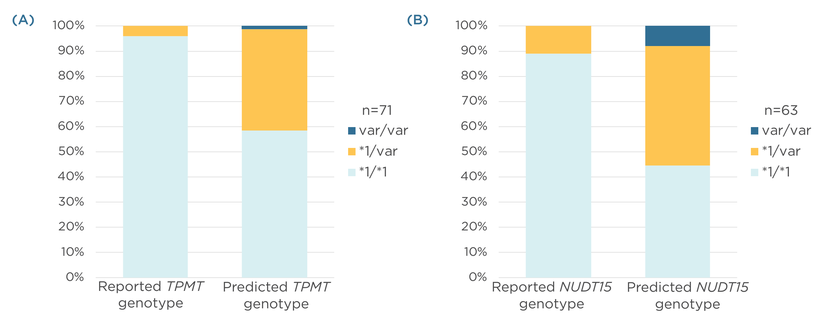
Table 2. Novel variants observed in *TPMT* and *NUDT15* by ancestry group.

Individuals of Oceania ancestry had the highest proportion of novel variants at 5.9%.

Genetic Ancestry	Proportion of novel variants (n=135)	
	Novel variants / Total	%
EUR	105 / 20,022	0.5
LAT	8 / 1,245	0.6
SAS	3 / 999	0.3
Other	8 / 948	0.8
EAS	3 / 581	0.5
SSA	5 / 529	0.9
AAC	2 / 337	0.6
OCE	1 / 17	5.9

Figure 5. Novel pLOF variants may result in change of reported *TPMT* and *NUDT15* phenotype.

(A) Of the 71 individuals with a novel *TPMT* variant, 28 (39.4%) had a pLOF variant that could result in a genotype with poorer function than reported. (B) Of the 63 individuals with a novel *NUDT15* variant, 33 (52.4%) had a pLOF variant that could result in a genotype with poorer function than reported. *1/*1 is wild-type; *1/var is one variant; var/var is two variants.



Conclusions

- The use of NGS technology allows for the detection of both known and novel variants in *NUDT15* and *TPMT*, which becomes increasingly important as more diverse populations are sequenced.
- Overall, 11.1% of our population had one or more risk alleles in *NUDT15* and/or *TPMT*. American ancestry had the highest proportion of risk alleles (22.7%), followed by East Asian ancestry (19.6%).
- We identified 43 novel exonic variants in *NUDT15* (21 benign, 22 pLOF) and 31 novel exonic variants in *TPMT* (13 benign, 18 pLOF) in a total of 135 individuals. Of those with a novel variant, 45.2% (n=61) were identified as having a pLOF variant that would result in a change from a *1/*1 or *1/var haplotype to a *1/var or var/var haplotypes in *TPMT* or *NUDT15*.
- Within our cohort, the ancestry groups with the highest proportion of novel variants were groups that have historically been underrepresented in genetics research, including Oceania (5.9%) and Sub-Saharan African (0.9%).
- The majority of our cohort consisted of healthy individuals with no reported history of thiopurine use. As a result, phenotypic data was not available. Further interrogation is needed to determine the functional consequences of these novel variants in order to better characterize and predict thiopurine response.

References

1. Schaeffeler E, Jaeger SU, Klumpp V, et al. *Genet Med*. 2019.
2. Relling MV, Schwab M, Whirl-Carrillo M, et al. *Clin Pharmacol Ther*. 2019.
3. Yang JJ, Landier W, Yang W, et al. *J Clin Oncol*. 2015.
4. Liu C, Yang W, Pei D, et al. *Clin Pharmacol Ther*. 2017.
5. Color Medication Response Genetic Test. Color Genomics.https://static.getcolor.com/pdfs/whitepaper/Color_PGx_White_Paper.pdf. Published October 1, 2019. Accessed June 3, 2019.
6. Jørsboe E, Hanghøj K, Albrechtsen A. *Bioinformatics*. 2017.
7. Huddart R, Fohner AE, Whirl-Carrillo M, et al. *Clin Pharmacol Ther*. 2019.