Identification of Novel *CYP2D6* Haplotypes that Interfere with TaqMan Copy Number Analysis

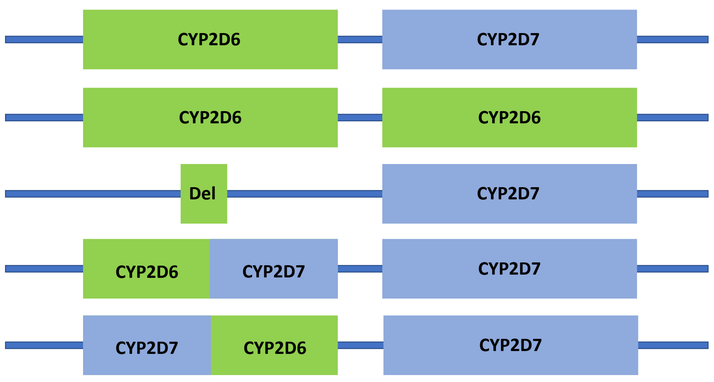
Amy Turner[1,2], Praful Aggarwal[1,2], Erin C. Boone[3], Cyrine-E. Haidar[4], Mary V. Relling[4], Ulrich Broeckel[1,2] and Andrea Gaedigk[3]

[1]Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI. [2]RPRD Diagnostics, Wauwatosa, WI. [3] Clinical Pharmacology, Toxicology & Therapeutic Innovation, Children's Mercy Kansas City, Kansas City, MO. [4] Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN

Background

Cytochrome P450 2D6 (*CYP2D6*) is a highly polymorphic gene encoding an enzyme critical in the metabolism of up to 25% of commonly prescribed drugs[1]. There are over 100 defined star allele (*) haplotypes, including gene conversions, dup/dels and gene-fusions (*D6/D7* hybrid), many of which have altered enzyme activity **(Figure 1**). Accurate copy number (CN) calling is critical in determining patient drug response[2-4].

Figure 1. *CYP2D6* and *CYP2D7* Structural Arrangement. *CYP2D6* is located next to the highly homologous pseudogene *CYP2D7*. This has led to the generation of stable duplications, deletions and *D6/D7* hybrid alleles.



TaqMan CNV assays (Thermo Fisher Scientific) utilize specific primers and labeled probes to evaluate the CNV state of a genomic region of interest. Polymorphisms within the primer or probe target sequence can generate false-positive CN calls. Three *CYP2D6* intra-gene regions can be tested using commercially available TaqMan CNV assays: intron 2, intron 6 and exon 9 (**Table 1 and Figure 2**). The *D6/D7* hybrids or conversions can be detected by discrepant CN calls between the three probe locations. The identification of haplotypes that cause such CN calls is critical in translating genotype into phenotype to predict drug response.

Table 1. Assay Information. Probe ID, gene location and chromosomal position for TaqMan CNV Assays used.

CYP2D6 CN Assays	location	GRCh38 Location
Hs00010001_cn	Exon 9	Chr22:42126649
Hs04502391_cn	Intron 6	Chr22:42127680
Hs04083572_cn	Intron 2	Chr22:42129681

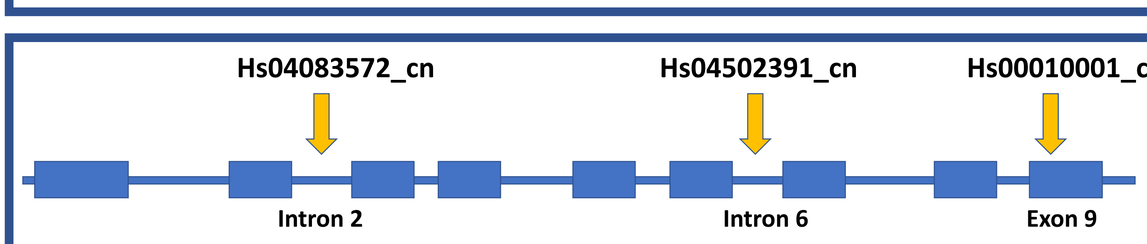


Figure 2. Representation of *CYP2D6* Gene Layout and Probe Position.

Methods and Results

The TaqMan CNV assays listed in **Table 1** were ran on gDNA samples. A subset of 20 samples with CN calls inconsistent between the assays were identified. To further characterize the haplotypes causing these CN calls, long range PCR (XL-PCR), Next-Gen sequencing (NGS), digital droplet PCT (ddPCR) and/or allele-specific Sanger sequencing were performed. The entire *CYP2D6* region was amplified for each sample as described by Black et al[5]. NGS libraries were prepared from the XL-PCR products and sequenced on the Ion Proton System (Thermo Fisher Scientific). Sequence analysis[6,7] identified three SNPs in intron 6 and four in intron 2 that are located within the TaqMan probe/primer binding sites. These SNPs were only present in samples that showed a discordant TaqMan CN call at the respective genomic locations.

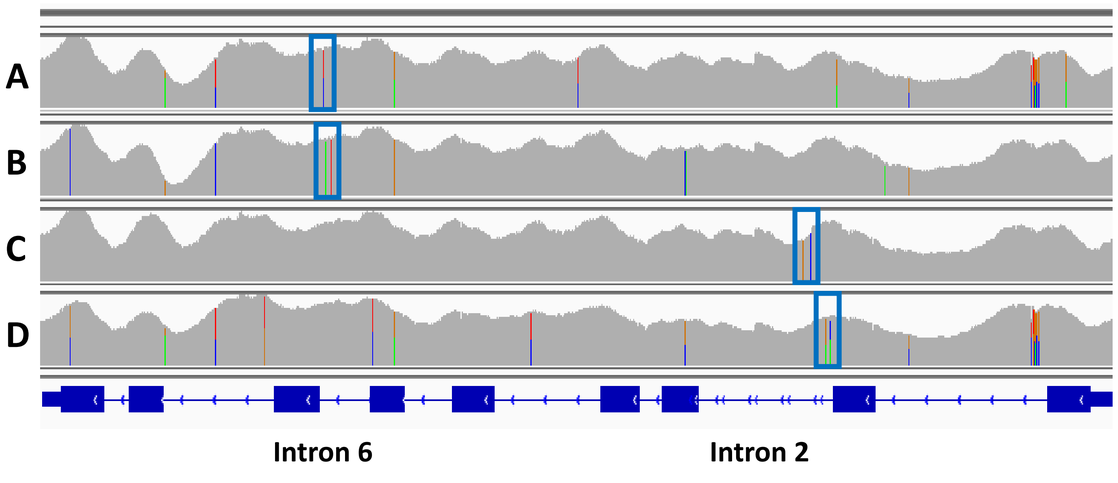


Figure 3. IGV Alignment of NGS Data to *CYP2D6*. Representative alignments of the seven SNPs in probe or primer sites. Samples **A** and **B** were found to have an intron 6 CN loss with assay Hs04502391. Samples **C** and **D** showed a intron 2 CN loss with assay hs04083572. The SNPs identified within the probe regions are indicated in blue boxes.

Methods and Results

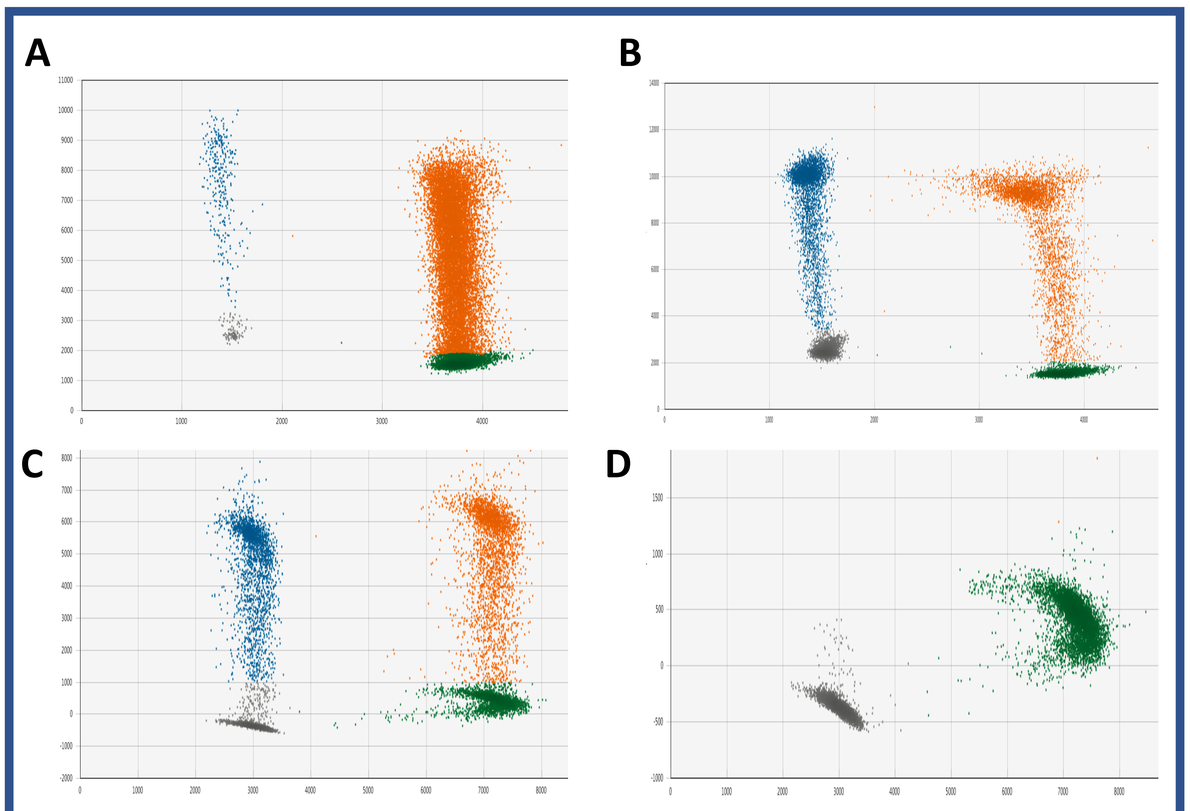


Figure 4. ddPCR Graphs. Samples show patterns reflective of containing SNPs under primer or probe regions. (**A)** and **(B)** show "tornado" and "rain" profiles for probe Hs04083572 with samples showing a CN loss in intron 2 (CN=1 for **(A)** CN=1 for **(B)**). **(C)** Tornado and **(D)** zero amplification for Hs04502391 with samples showing a CN loss in intron 6 (CN=1 for **(C)** and CN=0 for **(D)**). Y-Axis = channel 1 amplitude; X-Axis =Channel 2 amplitude.

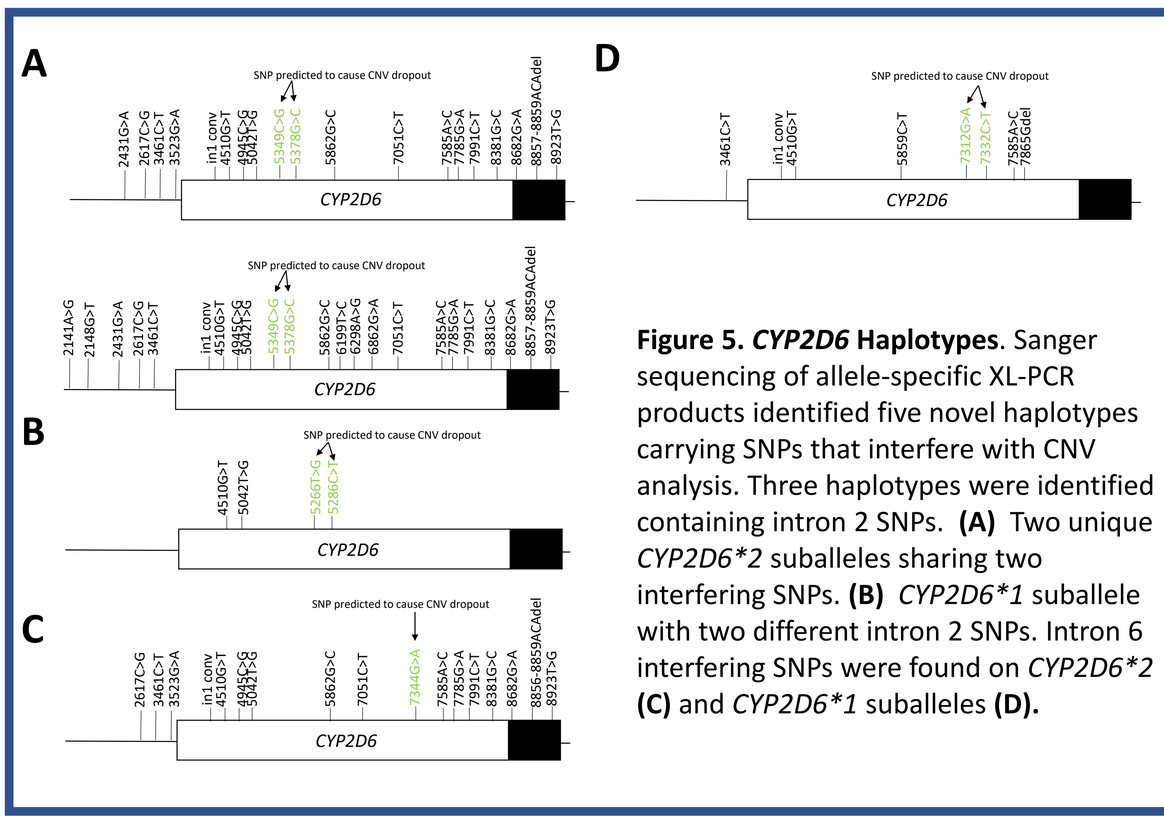


Figure 5. *CYP2D6* Haplotypes. Sanger sequencing of allele-specific XL-PCR products identified five novel haplotypes carrying SNPs that interfere with CNV analysis. Three haplotypes were identified containing intron 2 SNPs. **(A)** Two unique *CYP2D6*2* suballeles sharing two interfering SNPs. **(B)** *CYP2D6*1* suballele with two different intron 2 SNPs. Intron 6 interfering SNPs were found on *CYP2D6*2* **(C)** and *CYP2D6*1* suballeles **(D).**

Conclusions

In total, five novel haplotypes (suballeles) with SNPs within introns 2 and 6 that interfere with respective CNV assays were identified (**Figure 5**) .

In summary, performance of widely used TaqMan CNV assays is affected by the presence of sequence variation within the probe or primer target regions. The identification of these haplotypes and further characterization of additional variations will be important for clinical interpretation and could directly impact patient care.

References

1) Vasiliou et al. Acad Emerg Med. 2014 Aug; 21(8): 879–885.
2) https://cpicpgx.org
3) https://www.pharmgkb.org
4) https://www.pharmvar.org/
5) Black et al. Pharmacogenet Genomics. 2009 Oct; 19(10): 813–822.
6) James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. **Integrative Genomics Viewer**. Nature Biotechnology 29, 24–26 (2011) Dunnenberger et al. Annu Rev Pharmacol Toxicol. 2015;55:89-106
7) Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov**. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration**. Briefings in Bioinformatics 14, 178-192 (2013).