Editorial

For reprint orders, please contact: reprints@futuremedicine.com

Pharmacogenomics

Accurately genotyping *CYP2D6*: not for the faint of heart

Charity Nofziger*,[1] & Markus Paulmichl[2,3]

[1]PharmGenetix Gmbh, Niederalm, Austria

[2]Center for Health & Bioresources, Austrian Institute of Technology, Vienna, Austria

[3]NESMOS Department, University of Rome Sapienza, Rome, Italy

*Author for correspondence: Tel.: +43 501015 300; charity.nofziger@me.com

> "achievement of a 100% accurate *CYP2D6* genotype may constitute only half of the proverbial uphill battle."

First draft submitted: 20 June 2018; Accepted for publication: 5 July 2018; Published online: 18 July 2018

Keywords: allele drop out • *CYP2D7* • personalized medicine • pharmacogenetics • sequencing

Genotyping is arguably one of the mostimportant pillars of personalized medicine, especially within the context of pharmacogenetics, where phenotype is often predicted from genotype in order to adjust therapy. As such, it is not difficult to realize the necessity of accurately determining the genotype. Multiple factors contribute to the relative ease and/or difficulty for accurate genotyping, ranging from gene structure to genotyping technique employed, and the gene encoding for the Phase I metabolizing enzyme, *CYP2D6*, undoubtedly falls on the arduous end of the spectrum. This is underscored by multiple studies that have evidenced discordant results between sequencing and single variant genotyping techniques[1,3].Although relatively small in size (~4400 nucleotides from starting ATG to stop codon) compared with other Phase I metabolizing enzyme encoding genes (~90,300 nucleotides for *CYP2C19*, for example), the polymorphic nature of *CYP2D6*, as well as its surrounding locus add to the complexity of being able to comprehensively (and thus correctly) genotype it.

The first challenge to face is copy number variation (CNV), as *CYP2D6* can be duplicated, multiplicated or deleted. Between zero all the way up to 12 copies of *CYP2D6* on a single allele have been described to date [4,5]. A simple assessment of the number of *CYP2D6* genes present in a particular sample is seemingly straightforward; complications lie in the chimeras and/or hybrids that exist with its nearest neighbor and psuedogene, *CYP2D7*. *CYP2D6* and *CYP2D7* can form two main types of hybrids based on their structure – either *CYP2D6*/*CYP2D7* or *CYP2D7*/*CYP2D6* hybrids. *CYP2D6*/*CYP2D7* hybrids usually have a 5′-derived *CYP2D6* region and a 3′-derived *CYP2D7* region, while *CYP2D7*/*CYP2D6* hybrids on the other hand, have a 5′-derived *CYP2D7* region and a 3′-derived *CYP2D6* region. The two types of hybrids also differ in the amount of sequence derived from each respective gene. For example, at least ten different *CYP2D7*/*CYP2D6* hybrids have been reported thus far, each differing in the amount of *CYP2D7*- and *CYP2D6*- derived regions contained within them [6,7]. Therefore, whether or not CNV assays will detect hybrid structures depends on: the region of *CYP2D6* targeted by the CNV assay and how much of and which part of the *CYP2D6* sequence is present in the respective hybrid. The use of at least two separate CNV assays targeting the extreme ends of the gene – for example, one targeting intron 2 (5′ end) and the other targeting exon 9 (3′ end) – is usually standard procedure, as mismatches between the result for both assays can indicate that a hybrid is present. While neither type of hybrid usually produces a functional enzyme, genetic variations present in the structures, especially within the *CYP2D6*-derived regions, can erroneously contribute to and/or interfere with genotyping assays [8–10]. Since there is no 'diagnostic' sequence variation that is indicative of either type of hybrid, and use of multiple CNV assays are not guaranteed to flag their presence, inclusion of other techniques, such as end point PCRs designed to specifically amplify such entities probably represents the best approach.

Having correctly calculated the number of *CYP2D6* genes and hybrids present in a particular sample, the next goal should be assessment of sequence variations occurring within them. Numerous sequence variations in *CYP2D6*,

encompassing point mutations, insertions, deletions and the like have been and continue to be discovered. One particularly controversial challenge associated with *CYP2D6* sequence variants is deciding which ones should be interrogated. Should the testing of sequence variations be based on known functional implication, frequency within a certain population or ethnicity, a combination of both, or something in addition? This may seem like an easy question to answer since many *CYP2D6* gene variants are more common in certain populations than others [11,12]. However, we must realize that some of this population-specific frequency data may be flawed, especially when considering: population/ethnicity categorization is often self proclaimed; full gene sequencing is often not performed in such studies and human migration is on the rise. While a plethora of commercially available *CYP2D6* genotyping panels are currently available, with some being more extensive in terms of variant coverage than others [13], an apparent drawback of genotyping panels designed to detect single sequence variants is the possibility of known and unknown mutations within the remaining, non-interrogated sequence of the gene. Why not just sequence the entire gene, then? Next generation sequencing techniques, along with software tools for analyzing and calling the *CYP2D6* genotype, are accordingly flooding the *CYP2D6* variant detection field [14-16]. The main concerns with next generation sequencing methods are the misalignment of sequence reads to the highly homologous *CYP2D7* and nondetection of the structural variants (hybrids and/or duplications and multiplications), although the technology is becoming increasingly refined to avoid these blunders. Modified methods using Pacific Bioscience's single-molecule real-time (SMRT) sequencing platform have generated concordant results in previous, comprehensively *CYP2D6* genotyped samples [17,18], but other snags inherent to single-molecule real-time technology should not be overlooked [19].

The next confounding variable with genotyping *CYP2D6,* assuming the CNV, number of hybrids and all possible sequence variations within them have been identified, is phasing, or assigning the detected variants to a particular allele. This task is quite easy with samples that, for example, are homozygous mutant for all sequence variations detected, or that have two copies and are heterozygous for only one sequence variation. Difficulty increases exponentially with samples containing more than two copies and/or more than one heterozygous sequence variation. There is an almost endless amount of software tools designed for computational phasing to choose from, but experimentally derived results, for example from Sanger sequencing of allele-specific amplicons remains the gold standard.

One aspect we have not discussed in depth is the vast array of techniques that can be used for detecting sequence variations. While the various methods, ranging from basic restriction fragment length polymorphism to more advanced massive parallel sequencing come with their own pros and cons, there is one issue that plagues them all – allele drop out. Drop out occurs when a gene and/or sequence variation on one or more alleles is not detected due to PCR interference, in one way or another, and these events are usually only detected when the same sequence variation is interrogated by two separate techniques (Sanger sequencing and end point genotyping using hydrolysis probes, for example). Multiple allele drop out events for popular *CYP2D6* sequence variation assays have been described thus far, and considering the highly polymorphic nature of this gene, more are likely to come [9,20–23].

Gross, but usually unintended genotyping errors associated with *CYP2D6* include the assignment of a homozygous wild-type (**1/*1*) call to samples in which none of the interrogated sequence variations within the genotyping panel were detected. For example, a sample tested only for *CYP2D6* g.1846G>A (a loss of function sequence variation defining the **4* allele, rs3892097) should not be described *a priori* as **1/*1* if found negative for this particular variation. Any organization performing *CYP2D6* genotyping should be especially careful in this regard and may consider recommendations for test result reporting specified in Kalman *et al.* [24]. Another area of confusion rests in the numerous haplotypes that exist, especially with those that harbor (possibly overlapping) combinations of function-altering sequence variations. Consider, for example, the g.2850C>T (rs16947) variation, which is often set as the **2* allele defining variant. The *CYP2D6*2* allele and all associated haplotypes (A through M) are considered to produce normal functioning enzyme. Not immediately apparent, however, is that this same variant occurs in at least 38 other *CYP2D6* star alleles, 13 and six of which are currently defined as loss of function or decrease of function alleles, respectively. The newly founded Pharmacogene Variation Consortium (PharmVar) is the official pharmacogene nomenclature forum, (www.pharmvar.org) and serves as a data repository for genetic variations in *CYP2D6* and other cytochrome P450 encoding genes [25]. It is complete with insightful tools and informative documents explaining the basics of the *CYP2D6* gene locus, star allele naming and haplotype versioning.

In conclusion, the likely best strategy to unequivocally assess *CYP2D6* genotype is to employ a battery of techniques to capture the: copy number; presence and nature of hybrids; sequence variations and allele specificity of any sequence variations detected. However, achievement of a 100% accurate *CYP2D6* genotype may constitute only

half of the proverbial uphill battle. How can this be? The reality is that the functional consequence of approximately 42% of the *CYP2D6* star alleles currently reported on PharmVar, for example, are unknown. This hurdle is far from being overcome, especially when one considers that novel sequence variations without accompanying functional (*in vitro* or *in vivo*) data are continually being discovered. Next consider the *CYP2D6* star alleles (10%) slated to decrease enzyme function. This presumably quantitative categorization is, unfortunately, analogous to a wolf in sheep's clothes, since there is little conclusivity as to exactly how much the activity is decreased. The good news is that the remaining approximately 48% of characterized *CYP2D6* star alleles are associated with either normal or complete loss of protein function.

Financial & competing interests disclosure

C Nofziger is employed by Pharmgenetics Gmbh. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

1. Cohn I, Paton TA, Marshall CR *et al.* Genome sequencing as a platform for pharmacogenetic genotyping: a pediatric cohort study. *NPJ Genomic Med.* 2, 19 (2017).
2. Fang H, Liu X, Ramirez J *et al.* Establishment of *CYP2D6* reference samples by multiple validated genotyping platforms. *Pharmacogenomics J.* 14(6), 564–572 (2014).
3. Pratt VM, Everts RE, Aggarwal P *et al.* Characterization of 137 genomic DNA reference materials for 28 pharmacogenetic genes: a GeT-RM collaborative project. *J. Mol. Diagn.* 18(1), 109–123 (2016).
4. Gaedigk A, Blum M, Gaedigk R, Eichelbaum M, Meyer UA. Deletion of the entire cytochrome P450 *CYP2D6* gene as a cause of impaired drug metabolism in poor metabolizers of the debrisoquine/sparteine polymorphism. *Am. J. Hum. Genet.* 48(5), 943–950 (1991).
5. Johansson I, Lundqvist E, Bertilsson L, Dahl ML, Sjoqvist F, Ingelman-Sundberg M. Inherited amplification of an active gene in the cytochrome P450 *CYP2D* locus as a cause of ultrarapid metabolism of debrisoquine. *Proc. Natl Acad. Sci. USA* 90(24), 11825–11829 (1993).
6. Sim SC, Daly AK, Gaedigk A. CYP2D6 update: revised nomenclature for *CYP2D7/2D6* hybrid genes. *Pharmacogenet. Genomics* 22(9), 692–694 (2012).
7. Gaedigk A, Jaime LK, Bertino JS Jr *et al.* Identification of novel CYP2D7-2D6 hybrids: nonfunctional and functional variants. *Front. Pharmacol.* 1, 121 (2010).
8. Black JL 3rd, Walker DL, O'Kane DJ, Harmandayan M. Frequency of undetected *CYP2D6* hybrid genes in clinical samples: impact on phenotype prediction. *Drug Metab. Dispos.* 40(1), 111–119 (2012).
9. Riffel AK, Dehghani M, Hartshorne T *et al. CYP2D7* sequence variation interferes with TaqMan *CYP2D6 (*) 15* and *(*) 35* genotyping. *Front. Pharmacol.* 6, 312 (2015).
10. Gaedigk A. Complexities of *CYP2D6* gene analysis and interpretation. *Int. Rev. Psychiatry* 25(5), 534–553 (2013).
11. Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Klein T, Leeder JS. Prediction of CYP2D6 phenotype from genotype across world populations. *Genet. Med.* 19(1), 69–76 (2017).
12. Del Tredici AL, Malhotra A, Dedek M *et al.* Frequency of *CYP2D6* alleles including structural variants in the United States. *Front. Pharmacol.* 9, 305 (2018).
13. Yang Y, Botton MR, Scott ER, Scott SA. Sequencing the *CYP2D6* gene: from variant allele discovery to clinical pharmacogenetic testing. *Pharmacogenomics* 18(7), 673–685 (2017).
14. Lee SB, Wheeler MM, Patterson K *et al.* Stargazer: a software tool for calling star alleles from next-generation sequencing data using CYP2D6 as a model. *Genet. Med.* doi:10.1038/s41436-018-0054-0 (2018) (Epub ahead of print).
15. Numanagic I, Malikic S, Ford M *et al.* Allelic decomposition and exact genotyping of highly polymorphic and structurally variant genes. *Nat. Commun.* 9(1), 828 (2018).
16. Qiao W, Wang J, Pullman BS, Chen R, Yang Y, Scott SA. The CYP2D6 VCF translator. *Pharmacogenomics J.* 17(4), 301–303 (2017).
17. Buermans HP, Vossen RH, Anvar SY *et al.* Flexible and scalable full-length *CYP2D6* long amplicon PacBio sequencing. *Hum. Mutat.* 38(3), 310–316 (2017).
18. Qiao W, Yang Y, Sebra R *et al.* Long-read single molecule real-time full gene sequencing of cytochrome P450-2D6. *Hum. Mutat.* 37(3), 315–323 (2016).
19. Laver TW, Caswell RC, Moore KA *et al.* Pitfalls of haplotype phasing from amplicon-based long-read sequencing. *Sci. Rep.* 6, 21746 (2016).

20. Gaedigk A, Freeman N, Hartshorne T *et al.* SNP genotyping using TaqMan technology: the *CYP2D6*17* assay conundrum. *Sci. Rep.* 5, 9257 (2015).

21. Gaedigk A, Riffel AK, Leeder JS. CYP2D6 haplotype determination using long range allele-specific amplification: resolution of a complex genotype and a discordant genotype involving the *CYP2D6*59* allele. *J. Mol. Diagn.* 17(6), 740–748 (2015).

22. Rasmussen HB, Werge T. Novel variant of *CYP2D6*6* is undetected by a commonly used genotyping procedure. *Pharmacol. Rep.* 63(5), 1264–1266 (2011).

23. Scantamburlo G, Tziolia K, Zopf M *et al.* Allele drop out conferred by a frequent *CYP2D6* genetic variation for commonly used *CYP2D6*3* genotyping assays. *Cell. Physiol. Biochem.* 43(6), 2297–2309 (2017).

24. Kalman LV, Agundez J, Appell ML *et al.* Pharmacogenetic allele nomenclature: international workgroup recommendations for test result reporting. *Clin. Pharmacol. Ther.* 99(2), 172–185 (2016).

25. Gaedigk A, Ingelman-Sundberg M, Miller NA *et al.* The Pharmacogene Variation (PharmVar) consortium: incorporation of the human cytochrome P450 (CYP) allele nomenclature database. *Clin. Pharmacol. Ther.* 103(3), 399–401 (2018).