# Validation of methodology for efficient genotyping of *CYP2D6* and *CYP2C19*

**Beatriz Carvalho Henriques**
University of Alberta
**Amanda Buchner**
University of Alberta https://orcid.org/0000-0001-9574-4763
**Xiuying Hu**
University of Alberta https://orcid.org/0000-0001-8158-4412
**Yabing Wang**
University of Alberta
**Vasyl Yavorskyy**
University of Alberta https://orcid.org/0000-0001-5131-4374
**Rachael Dong**
University of Alberta https://orcid.org/0000-0003-3015-701X
**Kristina Martens**
University of Calgary https://orcid.org/0000-0002-3957-0647
**Michael Carr**
University of Alberta
**Bahareh Behroozi Asl**
University of Alberta https://orcid.org/0000-0003-0682-4417
**Keanna Wallace**
University of Alberta https://orcid.org/0000-0002-0864-6362
**Joshua Hague**
University of Alberta https://orcid.org/0000-0001-9993-6674
**Sudhakar Sivapalan**
University of Alberta https://orcid.org/0000-0001-8721-5896
**Wolfgang Maier**
University of Bonn
**Mojca Dernovsek**
University of Ljubljana https://orcid.org/0000-0001-5344-7007
**Neven Henigsberg**
University of Zagreb School of Medicine https://orcid.org/0000-0002-5303-1834
**Joanna Hauser**
Poznan University of Medical Sciences
**Daniel Souery**
Université Libre de Bruxelles and Psy Pluriel, Centre Européen de Psychologie Médicale
**Annamaria Cattaneo**
IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli https://orcid.org/0000-0002-9963-848X
**Ole Mors**
Aarhus University Hospital, Risskov
**Marcella Rietschel**
University of Mannheim https://orcid.org/0000-0002-5236-6149
**Gerald Pfeffer**
University of Calgary https://orcid.org/0000-0002-7657-7098
**Chad Bousman**
University of Calgary https://orcid.org/0000-0001-6303-8696
**Katherine Aitchison** ( aitchadm@ualberta.ca )
University of Alberta https://orcid.org/0000-0002-1107-3024

**Article**

**Keywords:** CYP2D6, CYP2C19, genotyping methodology

**Posted Date:** August 17th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-123677/v1



**Version of Record:** A version of this preprint was published at Translational Psychiatry on November 22nd, 2021. See the published version at https://doi.org/10.1038/s41398-021-01717-9.

## Abstract

There are some data associating variants in the *CYP2D6* and/or *CYP2C19* genes with concentration-to-dose ratios, efficacy, and retention in treatments. However, much of the above arises from relatively small studies or large datasets with limited genotyping methodologies. Our aim was to develop and validate comprehensive and accurate genotyping methodology for these two genes to facilitate regenotyping in large datasets and hence the generation of more accurate clinical associations. TaqMan copy number variant (CNV) assays for *CYP2D6* were used to identify samples from a relevant large dataset (GENDEP study, N = 853) with particularly challenging genotypes to call. These and those representing as broad a range of *CYP2D6* and *CYP2C19* genotypes as possible by prior available data (AmpliChip CYP450 and TaqMan *CYP2C19*17*) were chosen for further analysis (N = 96). Genotyping techniques employed were: Luminex *CYP2D6* xTAGv3 and Luminex *CYP2C19* xTAGv3, PharmacoScan, the Ion S5 AmpliSeq Pharmacogenomics Panel, TaqMan single nucleotide variant (SNV) assays, and, for the *CYP2D6* hybrids, long-range polymerase chain reactions (L-PCRs) with Sanger sequencing. Agena was also used for *CYP2C19*. The TaqMan SNV assays were able to assist with identifying which gene was duplicated or in tandem for multiple copy variants. A multiplex assay was adaptable for analysis of *CYP2D6* hybrid genes, with Sanger sequencing data being consistent with the data arising; we provide these data for efficient genotyping of such *CYP2D6* hybrid genes with adaptable multiplex methods. Consensus genotypes generated to date resulted in revision of assigned enzyme activity score for 28/96(29%) and 2/93 samples (2.2%) for CYP2D6 and CYP2C19, respectively.

## Introduction

Many antidepressants, atomoxetine, and several antipsychotics are metabolized by CYP2D6 and CYP2C19[1–7]. The gene (*CYP2D6*) encoding the enzyme CYP2D6 is on chromosome 22q13.2[8] adjacent to two pseudogenes, *CYP2D7* and *CYP2D8*[9]. The high homology between *CYP2D6* and these pseudogenes and the presence of flanking transposable genetic elements makes the region vulnerable to the generation of variable copy numbers of the *CYP2D6* gene and hybrid genes made up of sequence derived in part from *CYP2D7* and in part from *CYP2D6*[10,11]. Such variants are challenging to characterize for many technologies.

The *CYP2C19* gene encoding the CYP2C19 enzyme is located at chromosome 10q23.33, also together with other similar genes[12–14]. While structural variants of *CYP2C19* have recently been identified[15], the more commonly studied haplotypes result from single nucleotide variants (SNVs)[16]. Haplotypes are referred to as “star alleles” i.e. **2*, **3*, etc. as defined by PharmVar,[10] a consortium which maintains a rigorously curated catalogue of allelic variation in genes impacting drug metabolism, disposition and response.

Different *CYP2D6* or *CYP2C19* haplotypes may be associated with different levels of enzyme activity, ranging from haplotypes of loss-of-function (which give rise to no functional enzyme), to decreased function (which are associated with an enzyme with reduced metabolic activity), to gain-of-function (associated with increased activity). Haplotype frequencies vary between and within ethnic groups[1,4,17,18–20]. The study of clinical associations between variants in these genes and response to relevant medications has been to date limited by the challenging nature of the genotyping, particularly in the case of *CYP2D6*[21]. This gene is extremely polymorphic, with single or short sequence variants including indels (insertions/deletions), sequence derived from *CYP2D7* (described as “conversions” such as an exon 9 conversion), and copy number variants (deletions of the entire *CYP2D6* gene, gene duplications/multiplications, and hybrids[22] as above described). Our aim was to develop and validate efficient methodology, with a suggested algorithm (Fig. 1), for genotyping the range of common functional variants in these two genes.

## Materials And Methods

Ninety-six DNA samples (originating from venous blood) were selected from those previously genotyped for *CYP2D6* and *CYP2C19* using the AmpliChip CYP450 (Roche Molecular Systems) supplemented by the TaqMan assay C_469857_10 for *CYP2C19*17* as part of the Genome-based therapeutic drugs for depression (GENDEP) study[23]. Participants were all of self-reported White European ancestry. GENDEP was originally approved by ethics boards at all participating centres and approval for the work described herein was also provided by the University of Alberta Health Research Ethics Board - Biomedical Panel. Written informed consent was provided by all participants. The AmpliChip identified 32 *CYP2D6* variant haplotypes (**2, *3, *4, *5, *6, *7, *8, *9, *10, *11, *14, *15, *17, *19, *20, *25, *26, *29, *30, *31, *35, *36, *40, *41, *114 (*reported as **14A) *1xN (xN* referring to more than one copy*), *2xN, *4xN, *10xN, *17xN, *35xN, *41xN*). In addition, it covered *CYP2C19* haplotypes **2* and **3*. Sample DNA concentrations were ascertained using fluorimetry-based methods (Qubit or Quantifluor).

# TaqMan copy number variant (CNV) assays

TaqMan copy number variant (CNV) assays for *CYP2D6* (assay IDs: Hs04083572_cn and Hs00010001_cn for intron 2 and exon 9 respectively; Thermo Fisher Scientific) were run according to the manufacturer’s protocol on a ViiA7 Real-Time PCR System (Thermo Fisher Scientific). Assays were run in quadruplicate[24]. Data were analyzed using CopyCaller software version 2.1 (Thermo Fisher Scientific) with internal calibrators of known *CYP2D6* copy number according to the manufacturer’s instructions (using a confidence level of at least 95%, most being above 99%).

Samples for which the TaqMan CNV call across the two probes were not equal were analyzed with a third probe (assay ID Hs04502391_cn for *CYP2D6* intron 6). These samples, and those representing as broad a range of *CYP2D6* and *CYP2C19* genotypes as possible by the AmpliChip CYP450 and TaqMan *CYP2C19*17* assays, or “no call” for *CYP2D6* on the AmpliChip CYP450 were then taken forward for further analysis (N = 96). The following

genotyping techniques were employed: Luminex *CYP2D6* xTAG v3 and Luminex *CYP2C19* xTAG v 3, PharmacoScan (Thermo Fisher Scientific, Waltham, MA, United States), Ion Torrent S5 AmpliSeq Pharmacogenomics Panel (Thermo Fisher Scientific), and TaqMan Drug Metabolism Genotyping Assays (Thermo Fisher Scientific). Data arising from these were then used to select samples for the generation of amplicons by long range polymerase chain reaction (known as L-PCR)[25,26].

## TaqMan SNV assays

Haplotype phasing for samples with three copies of the *CYP2D6* gene according to the TaqMan, Pharmacoscan, and/or AmpliSeq CNV probe data, and heterozygous SNV data was conducted by the following methodologies: TaqMan assays for the relevant *CYP2D6* SNVs on genomic DNA, and/or L-PCR specific for *CYP2D6* duplicated genes followed by genotyping of the L-PCR product using relevant TaqMan SNV assays. The TaqMan SNV assays used were for *CYP2D6*2*, **3*, **4*, **6*, **35* and **41* (with assay IDs C_27102425_10, C_32407232_50, C_27102431_D0, C_32407243_20, C_27102444_F0, and C_34816116_20 respectively). The TaqMan SNV assays for *CYP2D6*2, *4*, **10* (assay ID: C__11484460_40), and **35* were used on genomic DNA to conduct haplotype phasing for samples with CNV data consistent with three *CYP2D6* genes including a hybrid gene. In addition, a TaqMan assay (assay ID C__25986767_70) was used to cross-validate a *CYP2C19*2*-defining SNV, rs4244285. Samples were run in duplicate on a ViiA7 Real-Time PCR System (Thermo Fisher Scientific), with genotype calling after visual inspection, outlier exclusion, and manual adjustment of $C_T$ threshold settings as necessary. Data arising from duplicates were compared with each other using an automated method available from the authors at request.

## Luminex

The Luminex xTAG *CYP2D6* and *CYP2C19* Kits v3 (research use only versions) were run according to the manufacturer's instructions using on a Luminex 200 system (Luminex Molecular Diagnostics, Inc., Toronto, ON, Canada). The assays use multiplex allele specific primer extension (ASPE) with a bead-based assay system. Haplotypes covered for *CYP2D6* are: *CYP2D6*2-*12* (including the **5* gene deletion), **14, *15, *17, *29, *35, *41*, and gene duplication. Our own *CYP2D6* haplotype translator (available at request) also permits derivation of **4.013*, **70*, **107* and **114*. The RUO software for *CYP2C19* reports *CYP2C19*2-*10*, and **17*.

## Ion S5 AmpliSeq Pharmacogenomics Panel

Genotyping using the Ion Torrent S5 AmpliSeq Pharmacogenomics Panel (Thermo Fisher Scientific) was conducted according to the manufacturer's instructions using an Ion Chef instrument (Thermo Fisher Scientific, Waltham, MA, USA). Short stretches of genomic DNA were sequenced, including regions of *CYP2D6* designed to detect gene deletion, duplication and conversion (*CYP2D6-2D7* hybrid) events. Following sequencing, data were analyzed using the GeneStudio Data Analysis software (Thermo Fisher Scientific). Sequencing generated an average of 109,454 reads per sample (mean read length 142.5 bp), with two samples failing quality control (in a manner indicating likely insufficient template: mapped read numbers of 18 and 51). Variant calling by the Ion Torrent Variant Caller version 5.10.1.19 (Thermo Fisher Scientific) generated three text files: one with the genotype at each SNV (including 20 *CYP2D6* variants and 11 *CYP2C19* variants), one for the *CYP2D6* exon 9 CNV output, and one for the *CYP2D6* gene level CNV data (based on sequence across nine regions in *CYP2D6*). Haplotype translation files were created to derive *CYP2D6* and *CYP2C19* haplotypes including various hybrid configurations in conjunction with the AlleleTyper software (Thermo Fisher Scientific)[27].

## PharmacoScan

The PharmacoScan array-based technology was run at Neogen Genomics (Lincoln, Nebraska, USA). The resultant data including more than 100 variants in *CYP2D6* and 60 variants in *CYP2C19* were analyzed using the Axiom Analysis Suite 4.0.3.3 (Thermo Fisher Scientific). The latest version (v8.2) of the manufacturer's *CYP2D6* haplotype translation file was used. CNV calls were provided by probes for exon 9 of *CYP2D6* as well as for the 5' and 3' flanking regions as described[28].

## Long-range PCR assays with characterization of resultant amplicons

L-PCR as described with minor modifications to generate an amplicon specific for the duplicated *CYP2D6* gene[29]. In brief, for the L-PCR assay that generates the D amplicon (specific for duplicated *CYP2D6* genes), we used primers as described (Gaedigk et al. 2007)[29], i.e., forward and reverse 5'-CCAGAAGGCTTTGCAGGCTTCAG-3' and 5'-CGGCAGTGGTCAGCTAATGAC-3' respectively, with minor modifications to the PCR conditions. Amplicons were purified by gel extraction (GeneJET Gel Extraction Kit, Thermo Fisher Scientific, Waltham, MA, USA), and genotyped using the TaqMan SNV assays described above.

Samples with unequal calls across the TaqMan, PharmacoScan or AmpliSeq CNV probes were subjected to L-PCR assays to generate amplicons specific for *CYP2D6-2D7* or *CYP2D7-2D6* hybrids (E, G[26], or H[30]), with minor modifications. Amplicons were purified by gel extraction and subjected to Sanger sequencing (10 µl at 3.5 ng/µl per reaction) using BigDye Terminator version 3.1 chemistry, the Axygen CleanSEQ magnetic beads-based post-reaction clean up protocol (automated on a Biomek 3000 workstation), and a capillary 3130xl Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA, USA).

Primers for sequencing (Supplementary Table 3) included those used to generate the L-PCR amplicons, those supplied by Dr. Gaedigk[31] for *CYP2D6* sequencing, one from Hosono et al[32], primers described by Broly et al.[33], and Gaedigk's universal fragment A reverse primer[31]. Sequence traces were aligned (to sequences available via the PharmVar[10] or archived[11] *CYP2D6* pages) and analyzed using SnapGene software version 5.1.4.1 (GSL Biotech LLC, Chicago, IL, USA).

# The Agena MassARRAY

The Agena MassARRAY (Agena Bioscience, San Diego, CA, USA) uses matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry technology for resolving oligonucleotides. We ran 8 *CYP2C19* variants to enable calling of 9 haplotypes. Genomic DNA was subjected to PCR followed by single-base extension with the extension products then being dispensed onto a SpectroCHIP Array and detected via mass spectrometry as described (Gaedigk et al., 2019)[28]. Haplotypes were assigned using Typer Analyzer software version v4.1.83 (Agena Bioscience).

## Results

*CYP2C19*

Percentage concordance for Luminex, AmpliSeq, PharmacoScan, and Agena with consensus *CYP2C19* genotype are shown in Table 1. All technologies apart from AmpliChip were able to detect *CYP2C19*6* and **8*. For *CYP2C19*2* and *CYP2C19*17*, data from all technologies were concordant.

*CYP2D6*

Comparative genotypic and CNV data across the technologies for samples with one and three copies of the *CYP2D6* gene are shown in Supplementary Tables 4 and 5 respectively. Owing to the "no calls" in the AmpliSeq CNV data, we revised these to manual calls, where possible, after reviewing the vcf files. This did result in an improvement in the degree of concordance with consensus genotypes for the AmpliSeq. Data for samples with a CNV call of two are shown in Supplementary Table 6.

While the AmpliChip provided haplotype phasing of *CYP2D6xNs* (i.e., assignment of a haplotype to one or other of the two chromosomes), the other technologies included herein do not offer that. We used TaqMan assays on genomic DNA to identify which haplotype was duplicated or multiplicated based on relative magnitude of signals arising from TaqMan wild-type and mutant probes for each assay (an example is shown in Figure 2). All of our *CYP2D6xNs* were duplications with the exception of one sample, which had a *CY2D6*41x3*. Consistent with this, our TaqMan CNV data were 4, 4, 4, and the raw PharmacoScan copy number probe calls were 4, 4, 4.

Data including consensus genotypes for the samples with hybrids genotyped to date are shown in Supplementary Table 7. Twenty samples had an unequal call across at least two out of three CNV probesets (TaqMan or PharmacoScan or AmpliSeq); for ten of these, the CNV pattern was consistent with a *CYP2D7-2D6* hybrid, and for ten with a *CYP2D6-2D7* hybrid. For all of the *CYP2D6-2D7* hybrids, the pattern was consistent with an extra *CYP2D6* gene, either on the same haplotype as the hybrid gene (in "cis"), or on the other chromosome 22 (in "trans"). This was also the case for seven of the *CYP2D7-2D6* hybrids. Amplicons consistent with hybrids[34] were generated for all 20 samples. Six samples had an unequal call across CNV probes for only one platform; four of these were genotyped as *CYP2D6* duplications (*CYP2D6*1x2/*4*, **1/*1x2*, **2x2/*1* and **2x2/*35*) and two as heterozygotes (**1/*2* and **1/*3.001*). For three of these, amplicon G was generated; however, it should be noted that the primer pair for this amplicon will also amplify up *CYP2D7* (Figure 2 in Black et al., 2012[26], observed where the *CYP2D6* downstream gene was **1*, **4*, or **41*; these three all had genotypes including the **1* and/or **4*, specifically **1/*2*, **1x2/*4*, **4x2/*1*).

L-PCR amplicons specific for *CYP2D7-2D6* hybrids aligned well to *CYP2D6*13* sequences (Supplementary Figures 1A and 1B). Specifically, three samples aligned to the GQ162807 sequence (previously *CYP2D6*77*) deposited by Gaedigk et al. (2010)[30], where *CYP2D6*13* is found in tandem arrangement with *CYP2D6*2*. The consensus genotype (from data including TaqMan assays) for these samples is *CYP2D6*13+*2/*1*. Three other amplicons aligned well to the EU093102 sequence for another *CYP2D6*13* variant, which is found as a single gene on one chromosome (Daly et al., 1996; Gaedigk et al., 2008)[35,36], consistent with the consensus genotypes for these samples: *CYP2D6*13/*4.013*, *CYP2D6*13/*1* and *CYP2D6*13/*1* (Supplementary Table 7). At this point, we have two remaining *CYP2D6*13s* for which the exact *CYP2D6*13* has not been identified; however, we have sufficient data to make a consensus genotype call for both (*CYP2D6*13+*4/*5*, and *CYP2D6*13+*2/*41*) and all *CYP2D6*13* haplotypes have the same CYP2D6 enzyme activity score (zero, i.e., do not encode any functional CYP2D6 protein) owing to a T insertion in exon 1 that is a frameshift mutation resulting in premature chain termination[22]. L-PCR amplicons for *CYP2D6-2D7* hybrids aligned to EU530605 (*CYP2D6*4-like*[34], Supplementary Figure 2), with one aligning to EU530606 (partial sequence for *CYP2D6*68*[34], Supplementary Figure 3).

*CYP2D6*13+*4/*5* represents a novel haplotype. The CNV data from TaqMan and AmpliSeq were 1, 2, 2 and 1, 2 consistent with the presence of a *CYP2D7-2D6* hybrid (a *CYP2D6*13*) with a switch region between intron 2 and intron 6. On alignment of the Sanger sequencing data, the sequence appeared as if the sample was heterozygous from a region consistent with this inferred switch region onwards, with the region prior to this aligning well to *CYP2D7*. This would be consistent with our long-PCR having amplified up both a *CYP2D7[REP6]* (a version of *CYP2D7* that has a *CYP2D6* version of the repeat element at its 3' region, consistent with the 3' primer used for the long-PCR) and a *CYP2D6*13* (legacy designations for the *CYP2D6*13* haplotypes *CYP2D6*67*, **78*, and **80*[22,37] all have switch regions in the relevant area). Comparative data from other technologies for this sample showed genotypes of *CYP2D6*4/*4* (AmpliChip), and no call with alternative calls of *CYP2D6*4/UNK* or **4.009/UNK* (Pharmacoscan). These indicate that a *CYP2D6*4* haplotype is also present. Haplotype phasing with the *CYP2D6*4* TaqMan assay indicated a deletion on one allele and the *CYP2D6*4* on the other. Therefore we deduced a configuration of *CYP2D6*13+CYP2D6*4/CYP2D6*5*, where the *CYP2D6*5* deletion has a

*CYP2D7[REP6]* followed by a deletion of the *CYP2D6* gene. We are proceeding to seek to amplify specific regions of the *CYP2D6*13* on their own without the *CYP2D7[REP6]*.

We also subjected the L-PCR G amplicons to genotyping using the Luminex *CYP2D6* assay, with a protocol modification. The resultant genotypes at 14 SNVs were consistent with the sequence data (Table 2). For example, the 4181G>C (rs1135840) variant was found in the Luminex and sequence data for samples aligned to GQ162807 or EU093102, while the 2851C>T (rs16947) variant was wild-type for samples aligned to EU093102 and variant for samples aligned to GQ162807. For the *CYP2D6-2D7* hybrids, our *CYP2D6* haplotype translator was able to identify some hybrids (Supplementary Table 7). An additional 3 (of the 96) samples have a CNV data pattern consistent with a hybrid and are still being characterized.

## Discussion

Consensus genotypes generated in 96 samples for *CYP2D6* and 93 samples for *CYP2C19* to date resulted in revision of assigned enzyme activity score for 28/96 (29%) and 2/93 samples (2.2%) for CYP2D6 and CYP2C19, respectively (sample selection enriched for structural variants in *CYP2D6*). These changes in assigned activity score were due to both changed genotype assignments and new genotype assignments for samples that were “no calls” on AmpliChip (Fig. 3). For *CYP2C19*, the highest concordance with consensus genotype was in the Luminex and PharmacoScan data (100%). Data from Luminex, Agena, TaqMan, AmpliSeq, PharmacoScan, and AmpliChip were 100% concordant for the *CYP2C19*2* and *CYP2C19*17*, the most common loss-of-function and gain-of-function haplotypes, respectively, in individuals of European ancestry. No adjustments in the prior AmpliChip data were therefore required for either of these haplotypes; prior clinical association analyses conducted on the basis of these *CYP2C19* haplotypes are therefore valid (Huezo-Diaz et al., 2012[23]; Fabbri et al., 2018[38]).

For *CYP2D6*, all technologies other than the AmpliChip were able to reliably detect the *CYP2D6*5*. Haplotype phasing of *CYP2D6xNs* was achieved by using relevant TaqMan assays on genomic DNA (Fig. 2), or by genotyping an amplicon specific for the *XN*. Although using allelic ratios to cluster TaqMan genotype data leaves a degree of uncertainty around genotypes (e.g., if only one probe amplifies, it may not be possible to distinguish between C/C, CC/C, CC/- and C/-), this technique can be used effectively to distinguish different heterozygote groups (Fig. 2). A strength of the sample set was the availability of prior AmpliChip data including haplotype phasing of *CYP2D6xNs*. The haplotype phasing thus achieved with our methods was consistent with the prior data, where available. One sample was genotyped as having a multiplication (i.e., more copies than 2), specifically of *CYP2D6*41*, which has been previously described[29,39]. The majority of the revisions in assigned enzyme activity score were due to the inability of AmpliChip to detect hybrids (Supplementary Table 7) and the inconsistency of *CYP2D6*5* detection by AmpliChip.

A focus of recent research on *CYP2D6* is the hybrid haplotypes[26,40–42]. Samples with *CYP2D6-2D7* or *CYP2D7-2D6* hybrid genes were identified through multiple methods including genotyping the L-PCR amplicons specific for *CYP2D7-2D6* hybrids (*CYP2D6*13* variants) using the Luminex *CYP2D6* assay. The resultant data were consistent with the specific *CYP2D6*13* variant sequences to which these samples had been aligned through Sanger sequencing. It is therefore possible that it might not be necessary to conduct Sanger sequencing of amplicons to identify such hybrid variants: screening for hybrids using CNV probes and multiplex SNV detection by methods such as Luminex or AmpliSeq may be sufficient. Our *CYP2D6* haplotype translator (Supplementary Table 1) was able to identify some hybrid tandems.

Our cross-technology comparisons suggest the following approach for efficient genotyping of *CYP2D6*: a multiplex SNV and CNV assay, haplotype phasing, and L-PCRs with sequencing where necessary (Fig. 1). Appropriate positive controls (e.g., from the Genetic Testing Reference Material Program (GeT-RM)[43], especially for the haplotypes that we did not see in this European sample set and which might be found in other ethnic groups, should be run with the assays. All of these assays other than the downstream processing of amplicons could be run in parallel, and the downstream processing of amplicons using a multiplex *CYP2D6* assay efficiently conducted. Different technologies have their strengths and weaknesses, particularly in regard to coverage of *CYP2D6* CNVs. A strategy that would be even more comprehensive, to include novel variants in non-coding regions, would include sequencing of the *CYP2D6-CYP2D7* intergenic region and the *CYP2D6* downstream region.

Limitations of this work include the following. Firstly, we have not covered *de novo* variants. Secondly, the work was conducted in a set of samples from European individuals being treated for depression, with samples being selected as being representative for genotypes available in the whole set and with enrichment for *CYP2D6* structural variants. As such, we did not find *CYP2D6* haplotypes that would be more commonly found in other ethnic groups, such as **29*, and therefore although the technologies were able to identify this haplotype, as none were detected in our data, we were not able to validate the detection thereof. However, of note, there are reference samples available with this haplotype from the Genetic Testing Reference Material Program (GeT-RM)[43]. Thirdly, theoretically it is possible that our CNV detection methods resulted in false positive calls for copy number loss in introns 2 and 6, owing to sequence variation in the relevant regions[44]. However, as we used three different technologies (AmpliSeq, Pharmacoscan, and TaqMan), covering probes in multiple regions of *CYP2D6* in addition to introns 2 and 6, and subjected any putative hybrid haplotypes to L-PCR and Sanger sequencing, we do not think this is a significant concern.

We suggest supplementing the *CYP2D6* SNV coverage described herein with TaqMan assays as follows a) for haplotyping *XNs* (13 assays including a custom assay for an exon 9 conversion SNV for *CYP2D6*36*, 6 of which we have already used to demonstrate methodology); b) to extend coverage to haplotypes of > 1% in any ethnic group (6 assays)[17,18]; c) another CNV assay, for the *CYP2D6* 5’ region; d) an assay (C__29692254_10) for rs5758550, which tags a region 100 kb downstream from *CYP2D6* proposed to act as an enhancer [45]. We have previously validated the C__29692254_10 assay against PharmacoScan data (concordance 100%)[46–48]. However, the functional consequence of this SNV on a background of a broad range of

different haplotypes is at present unknown[49]. The *CYP2C19* haplotypes included in the Luminex assay (*CYP2C19*2-*10*, and **17*) cover the *CYP2C19* change-of-function haplotypes currently identified in all ethnic groups at a frequency > 1%[19,20] apart from *CYP2C19*13, *15, *35*, and **18*. There are TaqMan assays available for the first three, while for the *CYP2C19*18* g.80156G > A SNV, a custom assay is required. For PharmacoScan, the four haplotypes are already covered. We also suggest an additional TaqMan assay for the c.463G > T variant (rs374036992) that may be found on the *CYP2C19*17* haplotype and introduces a premature stop codon[50], and an assay to enable *CYP2C19*17* haplotype phasing[50]. These suggestions for *CYP2C19* cover a greater range of haplotypes than suggested by previous authors[51], although we have not yet included coverage of the recently discovered *CYP2C19* structural variants[15]. TaqMan assays for both *CYP2D6* and *CYP2C19* can be efficiently multiplexed.

Our suggested strategy for *CYP2D6* and *CYP2C19* genotyping is based on the data available to us. There might be other strategies (such as single molecule long read sequencing[52]) or modified versions of various technologies that could work just as well. Considerations for choice of methodology include expected batch sizes of samples, cost and ability to add novel haplotypes. We have derived haplotype translation files for both *CYP2D6* and *CYP2C19* to enable efficient genotyping, adaptable for use with multiple technologies[27]. Our suggested genotyping strategy resulted in more accurate phenotype assignment for CYP2D6 and CYP2C19 in a subset of the GENDEP pharmacogenomic clinical trial, justifying extension of this work to the remainder of this dataset. In addition, the strategy with the additional assays suggested above provides a method for comprehensive detection of haplotypes above a frequency of 1% in any ethnic group for clinical implementation.

## Declarations


### Acknowledgments

We thank Kelly Homb for technical support for the Luminex assays, Lin Chen for support with interpretation of data arising from the AmpliSeq Pharmacogenomics Panel, and Carsten A. Bruckner for support for PharmacoScan data interpretation including supplying the most recent manufacturer's version of the *CYP2D6* translation file. The work reported herein was funded by: a Canada Foundation for Innovation (CFI), John R. Evans Leaders Fund (JELF) grant (32147 - Pharmacogenetic translational biomarker discovery), an Alberta Innovates Strategic Research Project (SRP51_PRIME - Pharmacogenomics for the Prevention of Adverse Drug Reactions in mental health; G2018000868 to KJA and CB), an Alberta Centennial Addiction and Mental Health Research Chair (to KJA), Alberta Innovation and Advanced Education Small Equipment Grants Program (to KJA), the Neuroscience and Mental Health Institute, Department of Psychiatry, and the Faculty of Medicine and Dentistry at the University of Alberta. Infrastructure used in this project from GP's lab was supported by a Hotchkiss Brain Institute Dementia Equipment Fund grant (to CB and GP for the Ion S5) and a Canada Foundation for Innovation John R Evans Leaders Fund Grant (CFI-JELF) (36624 - Neuromuscular genetics program, to GP). GENDEP was funded by a European Commission Framework 6 grant, LSHB-CT-2003-503428. Roche Molecular Systems supplied the AmpliChip CYP450 Test arrays and some associated support. GlaxoSmithKline and the Medical Research Council (UK) contributed by funding add-on projects in the London centre. This paper represents independent research part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.


### Conflict of Interest

N.H. has participated in research supported by CSF project No. IP-09-2014-2979. D.S. has received grant/research support from Janssen and Lundbeck, and has served as a consultant or on advisory boards for Janssen and Lundbeck. C.A.B. reports a grant from Alberta Innovates Strategic Research Project G2018000868, during the conduct of the study; and he has received in-kind testing kits from Myriad Neuroscience, CNSDose, Genomind, and AB-Biotics for research purposes but has not received payments or received any equity, stocks, or options in these companies or any other pharmacogenetic companies. K.J.A. is a member of the Clinical Pharmacogenetics Implementation Consortium and the Pharmacogene Variation Consortium, has received two research grants in the last two years from Janssen Inc., Canada (fellowship grants for trainees) and provided consultancy services (unpaid) for HLS Therapeutics. All other authors have nothing to disclose.

## References


1. Aitchison K.J., Jordan B.D., Sharma T. The relevance of ethnic influences on pharmacogenetics to the treatment of psychosis. *Drug Metabol and Drug Interact.* **16**(1),15–38 (2000).
2. Lapetina D.L., Yang E.H., Henriques B.C., Aitchison K.J. Pharmacogenomics and psychopharmacology. In: Peter M. Haddad, David J. Nutt, eds. *Seminars in Clinical Psychopharmacology*. London: Cambridge University Press; 2020:153–204.
3. Jukic M.M., Haslemo T., Molden E., Ingelman-Sundberg M. Impact of *CYP2C19* Genotype on Escitalopram Exposure and Therapeutic Failure: A Retrospective Study Based on 2,087 Patients. *Am J Psychiatry.* **175**(5),463–470 (2018).
4. Carvalho Henriques B., et al. How Can Drug Metabolism and Transporter Genetics Inform Psychotropic Prescribing? *Front Genet.* **11**,491895 (2020).

5. Hicks J.K., et al. Clinical Pharmacogenetics Implementation Consortium guideline for *CYP2D6* and *CYP2C19* genotypes and dosing of tricyclic antidepressants. *Clin Pharmacol Ther.* **93**(5),402–408 (2013).
6. Hicks J.K., et al. Clinical pharmacogenetics implementation consortium guideline (CPIC) for *CYP2D6* and *CYP2C19* genotypes and dosing of tricyclic antidepressants: 2016 update. *Clin Pharmacol Ther.* **102**(1),37–44 (2017).
7. Pharmacogenetics. Dutch Pharmacogenetics Working Group. https://www.knmp.nl/patientenzorg/medicatiebewaking/farmacogenetica/pharmacogenetics-1. Published 2019. Accessed Dec. 5, 2019.
8. Gough A.C., et al. Localization of the CYP2D gene locus to human chromosome 22q13.1 by polymerase chain reaction, in situ hybridization, and linkage analysis. *Genomics.* **15**(2),430–432 (1993).
9. Kimura S., Umeno M., Skoda R.C., Meyer U.A., Gonzalez F.J. The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification of the polymorphic *CYP2D6* gene, a related gene, and a pseudogene. *Am J Hum Genet.* **45**(6),889–904 (1989).
10. Gaedigk A., 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).
11. PharmVar. *CYP2D6*. https://www.pharmvar.org/gene/CYP2D6. Accessed Oct 12, 2020.
12. Goldstein J.A., de Morais S.M. Biochemistry and molecular biology of the human CYP2C subfamily. *Pharmacogenetics.* **4**(6),285–299 (1994).
13. Chen Y., Goldstein J.A. The transcriptional regulation of the human CYP2C genes. *Curr Drug Metab.* **10**(6),567–578 (2009).
14. Nelson D.R., et al. Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics.* **14**(1),1–18 (2004).
15. Botton M.R., et al. Structural variation at the CYP2C locus: Characterization of deletion and duplication alleles. *Hum Mutat.* **40**(11),e37-e51 (2019).
16. PharmVar. *CYP2C19*. https://www.pharmvar.org/gene/CYP2C19. Accessed Oct 12, 2020.
17. *CYP2D6* Frequency Table. Gene-specific Information Tables for *CYP2D6* Web site. https://www.pharmgkb.org/page/cyp2d6RefMaterials. Accessed Oct 12, 2020.
18. *CYP2D6* Allele Functionality Table. Gene-specific Information Tables for *CYP2D6* Web site. https://www.pharmgkb.org/page/cyp2d6RefMaterials. Accessed Oct 12, 2020.
19. *CYP2C19* Frequency Table. Gene-specific Information Tables for *CYP2C19* Web site. https://www.pharmgkb.org/page/cyp2c19RefMaterials. Accessed Oct 12, 2020.
20. *CYP2C19* Allele Functionality Table. Gene-specific Information Tables for *CYP2C19* Web site. https://www.pharmgkb.org/page/cyp2c19RefMaterials. Accessed Oct 12, 2020.
21. Nofziger C., Paulmichl M. Accurately genotyping CYP2D6: not for the faint of heart. *Pharmacogenomics.* **19**(13),999–1002 (2018).
22. PharmVar. *CYP2D6* Structural Variation Document. https://a.storyblok.com/f/70677/x/7073c64512/cyp2d6_structural-variation_v2-1.pdf. Published 2021. Accessed July 2, 2021.
23. Huezo-Diaz P., et al. *CYP2C19* genotype predicts steady state escitalopram concentration in GENDEP. *J Psychopharmacol.* **26**(3),398–407 (2012).
24. Carvalho Henriques B., et al. Cross-validation of technologies for genotyping CYP2D6 and CYP2C19. *bioRxiv.*2019.2012.2024.870295 (2020).
25. Aitchison K.J., et al. Failure to respond to treatment with typical antipsychotics is not associated with *CYP2D6* ultrarapid hydroxylation. *Br J Clin Pharmacol.* **48**(3),388–394 (1999).
26. Black J.L., 3rd, Walker D.L., O'Kane D.J., Harmandayan M. Frequency of undetected *CYP2D6* hybrid genes in clinical samples: impact on phenotype prediction. *Drug Metab and Dispos.* **40**(1),111–119 (2012).
27. CYP2D6 & CYP2C19 Haplotype Translator Tools https://. Accessed July 2, 2021.
28. Gaedigk A., et al. Characterization of Reference Materials for Genetic Testing of *CYP2D6* Alleles: A GeT-RM Collaborative Project. *J Mol Diagn.* **21**(6),1034–1052 (2019).
29. Gaedigk A., et al. Cytochrome P4502D6 (*CYP2D6*) gene locus heterogeneity: characterization of gene duplication events. *Clin Pharmacol Ther.* **81**(2),242–251 (2007).
30. Gaedigk A., et al. CYP2D7-2D6 hybrid tandems: identification of novel CYP2D6 duplication arrangements and implications for phenotype prediction. *Pharmacogenomics.* **11**(1),43–53 (2010).
31. Gaedigk A., Riffel A.K., Leeder J.S. *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).
32. Hosono N., et al. *CYP2D6* Genotyping for Functional-Gene Dosage Analysis by Allele Copy Number Detection. *Clin Chem.* **55**(8),1546–1554 (2009).
33. Broly F., et al. An efficient strategy for detection of known and new mutations of the *CYP2D6* gene using single strand conformation polymorphism analysis. *Pharmacogenetics.* **5**(6),373–384 (1995).
34. Kramer W.E., et al. *CYP2D6*: novel genomic structures and alleles. *Pharmacogenet Genomics.* **19**(10),813–822 (2009).
35. Daly A.K., et al. Characterization and PCR-based detection of two different hybrid *CYP2D7P/ CYP2D6* alleles associated with the poor metabolizer phenotype. *Pharmacogenetics.* **6**(4),319–328 (1996).

36. Gaedigk A., Coetsee C. The *CYP2D6* gene locus in South African Coloureds: unique allele distributions, novel alleles and gene arrangements. *Eur J Clin Pharmacol.* **64**(5),465–475 (2008).
37. PharmVar. Archived CYP2D6 Allele Nomenclature. https://www.pharmvar.org/htdocs/archive/cyp2d6.htm. Published 2016. Accessed July 2, 2021.
38. Fabbri C., et al. Effect of cytochrome *CYP2C19* metabolizing activity on antidepressant response and side effects: Meta-analysis of data from genome-wide association studies. *Eur Neuropsychopharmacol.* **28**(8),945–954 (2018).
39. Gaedigk A., Twist G.P., Leeder J.S. *CYP2D6*, *SULT1A1* and *UGT2B17* copy number variation: quantitative detection by multiplex PCR. *Pharmacogenomics.* **13**(1),91–111 (2012).
40. Carvalho H.B., et al. CYP2D6: Detecting New Structures for Clinical Practice. Oral presentation at the 15th Annual Pharmacogenetics in Psychiatry Meeting; October, 2016; Maui.
41. Carvalho Henriques B., et al. *CYP2D6* Revisited in GENDEP: Inter-Platform Concordance. Poster presentation at the Society of Biological Psychiatry's 75th Annual Meeting; 2020; New York.
42. Carvalho Henriques B., et al. Cross-Validation of Technologies for Genotyping *CYP2D6* and *CYP2C19*. Poster presentation at the 28th Annual World Congress of Psychiatric Genetics; 2020.
43. Genetic Testing Reference Materials Coordination Program (GeT-RM). Centers for Disease Control and Prevention. https://www.cdc.gov/labquality/get-rm/index.html. Published 2019. Accessed Oct 12, 2020.
44. Turner A.J., et al. Identification of CYP2D6 Haplotypes that Interfere with Commonly Used Assays for Copy Number Variation Characterization. *J Mol Diagn.* **23**(5),577–588 (2021).
45. Wang D., Papp A.C., Sun X. Functional characterization of *CYP2D6* enhancer polymorphisms. *Hum Mol Genet.* **24**(6),1556–1562 (2015).
46. Slomp C.J., et al. Comparative Data for *CYP2D6* from Microarray and Taqman Technologies: A Pathway to Innovation. Poster presentation to Cold Spring Harbour Personal Genomes & Medical Genomics Meeting; Nov. 13–16, 2013; New York.
47. Aitchison K.J. Drug metabolizing enzyme associations with response to agents used in neuropsychopharmacology. Oral presentation in a Symposium (chaired) "Pharmacogenomic associations with response to neuropsychopharmacological agents" at The 41st Annual Meeting of the Canadian College of Neuropsychopharmacology; June. 27–30, 2018; Vancouver.
48. Wang Y., et al. In-depth analysis of CYP2D6, a gene highly relevant to neuropsychopharmacology. Poster presentation at the 42nd Annual Meeting of the Canadian College of Neuropsychopharmacology; June. 12–15, 2019; Montreal.
49. Boone E.C., et al. Long-Distance Phasing of a Tentative "Enhancer" Single-Nucleotide Polymorphism With CYP2D6 Star Allele Definitions. *Front Pharmacol.* **11**(486) (2020).
50. Skierka J.M., Black J.L. Analysis of compound heterozygous *CYP2C19* genotypes to determine cis and trans configurations. *Pharmacogenomics.* **15**(9),1197–1205 (2014).
51. Pratt V.M., et al. Recommendations for Clinical CYP2C19 Genotyping Allele Selection: A Report of the Association for Molecular Pathology. *J Mol Diagn.* **20**(3),269–276 (2018).
52. Buermans H.P.J., et al. Flexible and Scalable Full-Length *CYP2D6* Long Amplicon PacBio Sequencing. *Hum Mutat.* **38**(3),310–316 (2017).

## Tables

**e 1. Percentage concordance for Luminex, Ion S5, PharmacoScan, Agena, and AmpliChip and TaqMan *CYP2C19*17* prior data consensus *CYP2C19* genotype**

| sus Genotype | N | AmpliChip and TaqMan **17* (% concordance) | Luminex RUO (% concordance) | Ion S5 (% concordance) | PScan (% concordance) | Agena (% concordance) |
|---|---|---|---|---|---|---|
| **1/*1* | 44 | 100 | 100 | 97.7 (42/43)[1] | 100 (42/42)[2] | 95.2 (40/42)[3] |
| *1/*17* | 24 | 100 | 100 | 100 (22/22) | 100 (23/23) | 100 (22/22) |
| **1/*2* | 16 | 100 | 100 | 100 | 100 | 100 |
| *17/*17* | 1 | 100 | 100 | 0(0/1)[4] | NA[5] | 100 |
| *2/*17* | 4 | 100 | 100 | 100 | 100 | 100 (3/3) |
| **2/*2* | 2 | 100 | 100 (1/1) | 100 | 100 | 100 (1/1) |
| **1/*8* | 1 | 0 | 100 | 100 | 100 | 100 |
| **2/*6* | 1 | 0 | 100 | 100 | 100 | 100 |

One "no call" out of the 43 genotyped using this assay
A couple of *CYP2C19*27* haplotypes were found in this group; this is now classified as *CYP2C19*1.006*
Two "no calls" out of the 42 genotyped using this assay
One "no call"
Not assayed

enhanced validation, two more samples of *CYP2C19*17/*17* genotype by TaqMan were genotyped: one on IonS5, PharmacoScan Luminex (concordant on all three technologies), and one on Luminex (concordant). A TaqMan assay for *CYP2C19*2* cross-ated AmpliChip data 100% in the full GENDEP dataset.

**Table 2. Comparative Sanger sequencing and Luminex genotyping data for 14 *CYP2D6* variants in samples with *CYP2D7-2D6* hybrid haplotypes**

| ple | Method | 31G>A | 100C>T | 124G>A | 137_138insT | 882G>C | 1660G>A | 1708delT | 1847G>A | 2550delA | 2851C>T | 2936A>C | 2989G>A | 3184G>A | 4181G>C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 03102 | NA | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | WT | WT | Sequence not present | WT | MUT |
| 62807 | NA | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | MUT | WT | WT | WT | MUT |
| | Luminex | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | WT | WT | Low Signal | WT | MUT |
| | Sanger | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | WT | WT | Sequence not present | WT | MUT |
| | Luminex | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | WT | WT | Low Signal | WT | MUT |
| | Sanger | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | WT | WT | Sequence not present | WT | MUT |
| | Luminex | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | MUT | WT | WT | WT | MUT |
| | Sanger | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | MUT | WT | WT | WT | MUT |
| | Luminex | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | MUT | WT | WT | WT | MUT |
| | Sanger | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | MUT | WT | WT | WT | MUT |
| | Luminex | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | WT | WT | Low Signal | WT | MUT |
| | Sanger | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | WT | WT | Sequence not present | WT | MUT |
| | Luminex | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | MUT | WT | WT | WT | MUT |
| | Sanger | MUT | WT | WT | MUT | WT | WT | WT | WT | WT | MUT | WT | WT | WT | MUT |

Samples 1, 2, and 5 align to EU093102, sequence for the *CYP2D6*13* haplotype previously known as *CYP2D6*66*.
Samples 3, 4, and 6 align to GQ162807, sequence for the *CYP2D6*13* haplotype previously known as *CYP2D6*77* and found in a tandem arrangement with *CYP2D6*2*.

## Figures

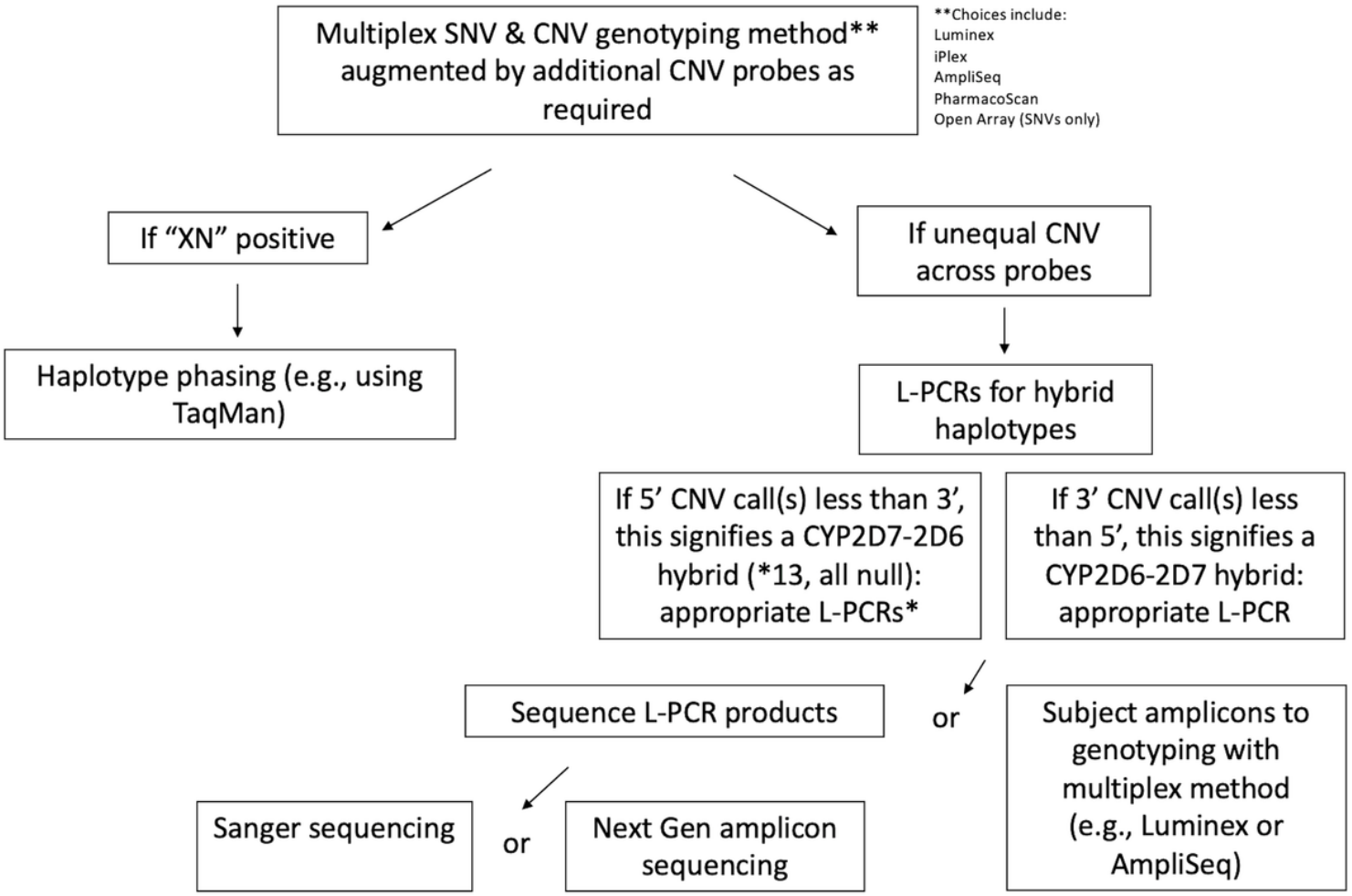


Figure 1

A suggested algorithm for CYP2D6 genotyping

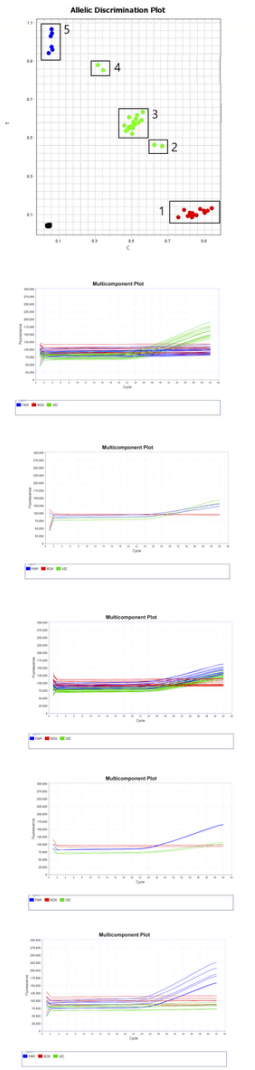


**Figure 2**

A-F. TaqMan CYP2D6*41 allelic discrimination plot (A), with multicomponent plots (B-F) showing different allelic ratios corresponding to the clusters marked 1-5. This is a C(Ref)>T(Var) SNV. (B) Amplification of only C. Sample genotypes are *1/*1, *1/*1, *1XN/*1, *1X2/*2, *2X2/*1, *1X2/*5, *1X2/*35. (C) Amplification of C and T, the former more than the latter, indicating a heterozygous *41 sample with more copies of C than T. Sample genotype is *2X2/*41. (D) Amplification of C and T to the same extent. Sample genotypes are *1/*41, *2/*41, *2/*41, *4/*41, *4/*41, *6/*41, *9/*41, *10/*41, *35/*41. (E) Amplification of C and T, the latter more than the former. Sample genotype is *41X3/*3. (F) Amplification of only T. Sample genotypes are *5/*41, *5/*41 *41/*41 (where *5 is a deletion of CYP2D6).

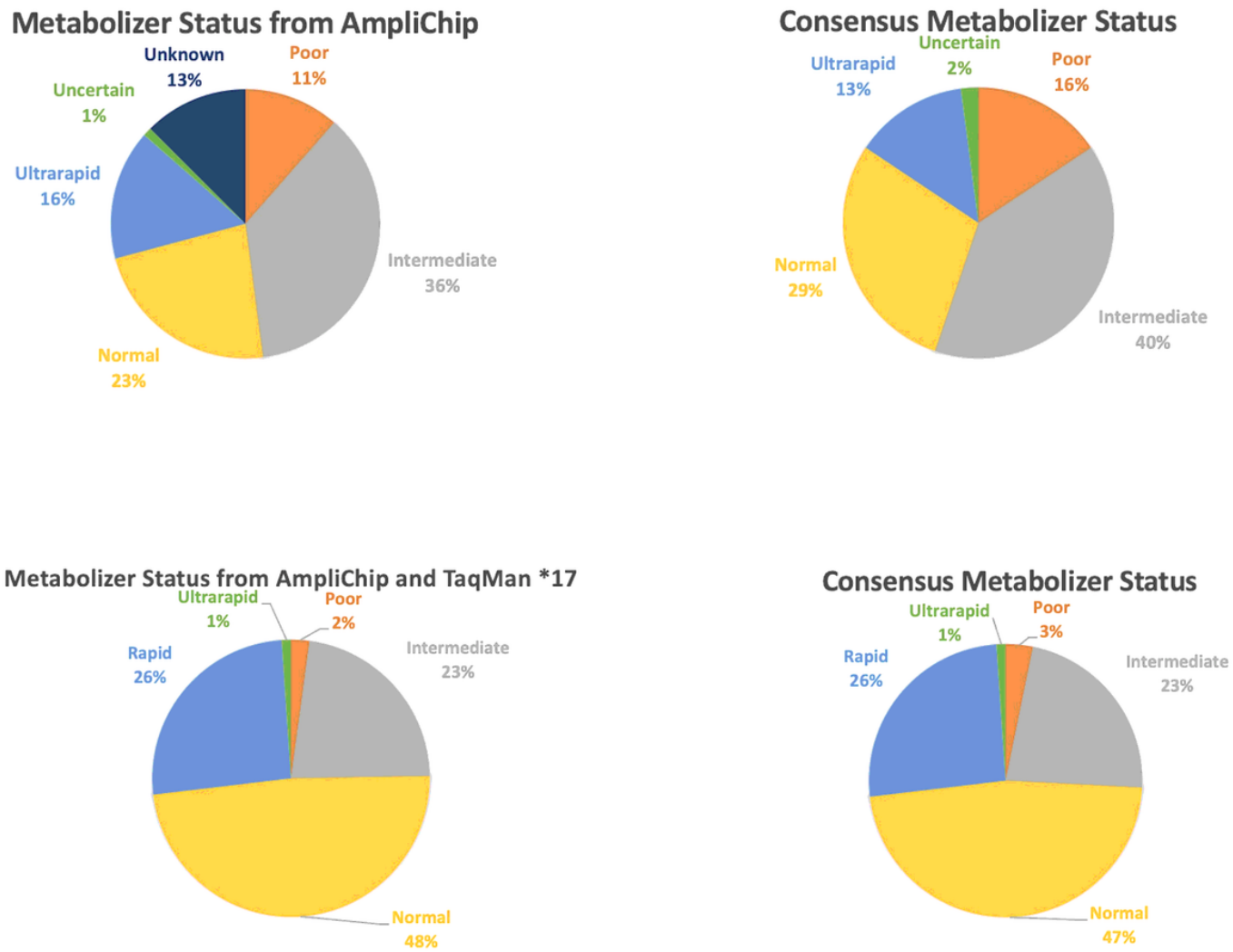


**Figure 3**

A-B. (A) Change in distribution of CYP2D6 metabolizer status deduced from genotype, from prior data to revised consensus data (B) Change in distribution of CYP2C19 metabolizer status deduced from genotype, from prior data to revised consensus data

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable120210702ed.xlsx
- SupplementaryTable220210711.xlsx
- SupplementaryTable320210715ed.xlsx
- SupplementaryTable420210702ed2.docx
- SupplementaryTable520210702ed2.xlsx
- SupplementaryTable620210702ed2.xlsx
- SupplementaryTable720210702ed2.xlsx
- SuppFig1ANew.tif
- SuppFig1BNew.tif
- SuppFig2.tif
- SuppFig3New.tif