Cross-validation of technologies for genotyping *CYP2D6* and *CYP2C19*

Beatriz Carvalho Henriques[1], Vasyl Yavorskyy[1], Yabing Wang[1], Kristina Martens[2], Bahareh Behroozi Asl[1], Michael Carr[1], Joshua Hague[1], Xiuying Hu[1], Wolfgang Maier[3], Mojca Z. Dernovsek[4], Neven Henigsberg[5], Daniel Souery[6], Annamaria Cattaneo[7], Ole Mors[8], Marcella Rietschel[9], Cathryn Lewis[7], Gerald Pfeffer[2, 10], Chad Bousman[11], Katherine J. Aitchison[1*]

[1]Departments of Psychiatry and Medical Genetics, and Neuroscience and Mental Health Institute, University of Alberta, Edmonton, Canada
[2]Hotchkiss Brain Institute, Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, Canada
[3]University of Bonn, Bonn, Germany
[4]Institute Karakter, Ljubljana, Slovenia
[5]Croatian Institute for Brain Research, Zagreb, Croatia
[6]Université Libre de Bruxelles, Brussels, Belgium
[7]Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK
[8]Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark
[9]Central Institute of Mental Health, University of Heidelberg, Mannheim, Germany
[10]Alberta Child Health Research Institute, Department of Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, Canada
[11] Departments of Medical Genetics, Psychiatry, Physiology & Pharmacology, University of Calgary, Calgary, Canada

*Corresponding author:

Dr. Katherine J. Aitchison, Departments of Psychiatry and Medical Genetics, University

of Alberta, Edmonton, AB T6G 2E1, Canada

Telephone: 1-780-492-4018; Email: kaitchis@ualberta.ca

OCRID: orcid.org/0000-0002-1107-3024

Abstract

Background: CYP2D6 and CYP2C19 are cytochrome P450 enzymes involved in the metabolism of many medications from multiple therapeutic classes. The genes encoding them, *CYP2D6* and *CYP2C19*, have multiple functional variants that have been associated with different levels of enzyme activity. There is evidence that methodology for identifying such variants ("genotyping") effectively and efficiently could lead to substantial health care costs savings. There are various genotyping technologies available; we cross-validated several against each other.

Methods: Ninety-six samples with a variety of *CYP2D6* and *CYP2C19* genotypes according to prior AmpliChip CYP450 and TaqMan *CYP2C19*17* data were selected from the Genome-based therapeutic drugs for depression (GENDEP) study. Genotyping was performed with TaqMan copy number variant (CNV) analysis using probes specific for three different regions (*CYP2D6* only), the next generation sequencing-based Ion S5 AmpliSeq Pharmacogenomics Panel, PharmacoScan, and long-range polymerase chain reaction (XL-PCR) followed by TaqMan single nucleotide variant (SNV) analysis or the Agena MassARRAY. Variant to "star allele" translation was automated.

Results: The inter-platform concordance for *CYP2C19* was high (94-98%). For *CYP2D6*, the IonS5-PharmacoScan concordance was 94% for a range of variants tested apart from those with at least one extra copy of a *CYP2D6* gene (occurring at a frequency of 3.8%, 33/853), or a pseudogene-gene conversion (2%, 18/853).

Conclusions: Inter-platform concordance for *CYP2C19* was high, and, moreover, the Ion S5 and PharmacoScan data were 100% concordant with that from a TaqMan *CYP2C19*2* assay. The level of concordance was also good for *CYP2D6*, apart from those with at least one extra copy of a *CYP2D6* gene or a pseudogene-gene conversion. We have also demonstrated feasibility of using a NGS platform for genotyping *CYP2D6* and *CYP2C19*, with automated data interpretation methodology. *CYP2D6* and *CYP2C19* genotyping may thereby become much more readily accessible, with potential substantial savings in health care costs and public benefit.

Introduction

CYP2D6 and CYP2C19 are cytochrome P450 enzymes involved in hepatic phase I reactions that alter the hydrophobicity and reactivity of substrates. Substrates for CYP2D6 include at least 25% of commonly prescribed medications from a variety of classes, including antidepressants (e.g., venlafaxine, citalopram, fluoxetine, paroxetine, amitriptyline, nortriptyline, imipramine, desipramine, clomipramine, trimipramine,) antipsychotics (e.g., aripiprazole, risperidone, haloperidol, zuclopenthixol, and pimozide), antiarrhythmics (e.g., flecainide and propafenone), beta-blockers (e.g., metoprolol and propranolol), opioids (e.g., codeine), antiemetics (e.g., metoclopramide and ondansetron), antitussives (e.g., dextromethorphan) and anticancer agents such as the hormonal modulator tamoxifen.[2-5] Substrates for CYP2C19 include proton pump inhibitors (omeprazole and pantoprazole), the antithrombotic clopidogrel, antidepressants (fluoxetine, sertraline, paroxetine, citalopram, escitalopram, moclobemide, imipramine, amitriptyline, clomipramine), antipsychotics (e.g., clozapine and olanzapine), anxiolytics (diazepam and propranolol), the anticonvulsant phenytoin, and immunologically active agents (e.g., cyclophosphamide, nelfinavir, proguanil, and voriconazole).[5]

The gene (*CYP2D6*) encoding the enzyme CYP2D6 is on chromosome 22q13.2 [6] adjacent to two pseudogenes, *CYP2D7* and *CYP2D8* .[7] C*YP2D7* is highly homologous with *CYP2D6.* [7] This arrangement is susceptible to unequal crossover events in meiosis. For example, *CYP2D6* pairs with *CYP2D7* or part thereof, resulting in a deletion of the *CYP2D6* gene on one chromatid and an extra copy of *CYP2D6* on the other, or hybrid genes containing both *CYP2D6* and *CYP2D7* sequences. These structural variants have been historically challenging to detect by many different types of genotyping techniques. In addition, the region is one of the most variable areas of the human genome, with over 100 different haplotypes or "star alleles" (defined by particular genetic variants or combinations thereof on the same chromosome) having been identified and catalogued.[8,9] Moreover, new variants continue to be discovered. It is therefore hardly surprising that genotyping of *CYP2D6* for treatment optimization is not being routinely provided in clinical settings, despite the presence of guidelines indicating potential clinical utility.[10-12]

The *CYP2C19* gene encoding the CYP2C19 enzyme is likewise located at chromosome 10q23.33 together with other similar genes, namely *CYP2C18* (not expressed at the protein level), *CYP2C9*, and *CYP2C8.* [13-15] While structural variants of *CYP2C19* have recently been identified,[16] the more commonly studied "star allele" variants result from single base changes in the DNA sequence (single nucleotide variants or SNVs); more than 30 different variants in *CYP2C19* have been catalogued to date.[17]

Each *CYP2D6* or *CYP2C19* haplotype or "star allele" is associated with different levels of enzyme activity, ranging from poor metabolizer (PM) alleles, which give rise to no functional enzyme; to intermediate metabolizer (IM) alleles, which are associated with an enzyme with reduced metabolic activity; to ultrarapid metabolizer alleles (UMs), associated with increased activity.[5] In between the IM and

UM activities are alleles of normal metabolizer (NM) activity. The CYP2D6 and CYP2C19 enzyme activity groups vary markedly by ethnic group. CYP2D6 UMs have been found at a frequency of 0.9 - 4% in Whites and at higher frequency in Sub-Saharan Africa; CYP2D6 IMs are relatively common in East Asia and in Blacks; and CYP2D6 PMs are relatively common in Caucasians at 7% and rare in East Asians at 1%.[2] The frequency of CYP2C19 PMs is 2-5% in Caucasians, 2% in Saudi Arabians, 4% in Zimbabweans, 5% in Ethiopians, 13% in Koreans, 15-17% in Chinese, 21% in Indians, and 18-23% in Japanese.[2,5]

There is evidence that genotyping for CYP2D6 and CYP2C19 would reduce the number of adverse drug reactions (ADRs) to medications metabolized by these enzymes. For example, Pirmohamed et al. (2004) [18] found that antidepressants were the leading mental health class of medication and fifth leading medication class overall of ADR-related general hospital admissions. Methodology for genotyping these genes effectively and efficiently could therefore lead to substantial health care costs savings. The first step in introducing a new genetic test to a clinical genetics laboratory is establishing analytical validity. Our aim was to contribute to this step by cross-validating multiple technologies for genotyping *CYP2D6* and *CYP2C19* against each other and providing automated variant to "star allele" translation.

Methods

Ninety-six DNA samples originating from venous blood samples were selected from those previously genotyped for *CYP2D6* and *CYP2C19* using the AmpliChip CYP450 supplemented by a TaqMan assay C_469857_10 for *CYP2C19*17* as part of the Genome-based therapeutic drugs for depression (GENDEP) study[1]. Participants were all of self-reported White European ancestry. The AmpliChip identified 32 *CYP2D6* variant alleles (**2, *3, *4, *5, *6, *7, *8, *9, *10, *11, *14, *15, *17, *19, *20, *25, *26, *29, *30, *31, *35, *36, *40, *41, *1XN, *2XN, *10XN, *17XN, *35XN, *41XN*). In addition, it covered *CYP2C19* alleles **2* and **3*. The non-variant (known as wild-type) allele (**1*) for each gene was identified by default as an absence of the mutations defining the specific variants tested.

Samples were quantified using fluorimetry-based methods (Qubit or Quantifluor). TaqMan copy number 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, Aitchison laboratory), with a minor modification (TaqMan Universal Mastermix II no UNG was used instead of Genotyping MasterMix). The assays are specific for intron 2 and exon 9 in *CYP2D6*, with DNA not being amplified if the relevant region is *CYP2D7* sequence. Assays were run in quadruplicate with internal calibrators of known *CYP2D6* copy number. Data were analysed using Copy-Caller 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 95% and z-score value of less than 1.75).

Samples for which the TaqMan CNV call across the two probes were not equal were analysed with a third probe (assay ID Hs04502391_cn for *CYP2D6* intron 6). These samples and those with a unique combination *CYP2D6* and *CYP2C19* genotypes by the AmpliChip CYP450 and TaqMan *CYP2C19*17* assays, or no reference sample for that genotypic combination available at the time of selection (June 2019), or no call on the AmpliChip CYP450 were then taken forward for further analysis. The following genotyping techniques were employed: PharmacoScan (Thermo Fisher Scientific), Ion S5 AmpliSeq Pharmacogenomics Panel (Thermo Fisher Scientific), and TaqMan Drug Metabolism Genotyping Assays (Thermo Fisher Scientific). Samples were selected on the basis of the above data for the generation of specific long amplicons by polymerase chain reaction (known as XL-PCR) for the detection of specific gene *CYP2D6* deletion, duplication and conversion events.[19-21] Products of XL-PCRs for conversion events were In addition, a TaqMan assay for *CYP2C19*2* (assay ID C_25986767_70) was used as a further validation test for this SNV.

Ion S5 AmpliSeq Pharmacogenomics Panel

Genotyping using the Ion 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). Short stretches of genomic DNA were sequenced, including regions of *CYP2D6* designed to detect gene deletion, duplication and conversion (*CYP2D6-2D7* hybrid) events. In brief, library preparation was conducted using the Ion AmpliSeq Library Kit Plus (genomic DNA was subjected to a multiplex PCR using a pool of primer pairs). The resulting amplicons were then partially digested and ligated to barcode adaptors (Ion Express barcode adaptors 1-96 kit, Thermo Fisher Scientific), which were then purified using a magnetic bead technology (AMPure XP Reagent, Agencourt). Library quantitation was carried out by quantitative PCR using the Ion Library TaqMan Quantitation Kit (Thermo Fisher Scientific) and was followed by normalisation of library concentration to 40 pM. Two chips were then loaded, containing libraries 1-48 and 49-96 respectively. Following sequencing, data were accessed remotely using the GeneStudio Data Analysis software (Thermo Fisher Scientific). Ion S5 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 two regions in intron 2 to exon 4, one spanning exon4-intron3-exon3 and one in the exon-3-intron2 region) (genomic locations 42524932-42525095 and 42525131-42525245, GRCh38/hg38 assembly). Translator files were created to derive *CYP2D6* and *CYP2C19* alleles from the SNV and CNV data using the AlleleTyper software (Thermo Fisher Scientific). These translator files are available at https://flintbox.com/public/project/61820/.

PharmacoScan

PharmacoScan is an array-based technology in which the sample DNA is hybridized to short lengths of DNA pre-bound to a "gene chip;" this assay was run at Neogen Genomics (Nebraska, USA). Genomic DNA was amplified (including preferential amplification of *CYP2D6* and *CYP2C19*), with the amplicons then being fragmented, labelled, and hybridized to a PharmacoScan Array (Thermo Fisher Scientific). Arrays stained with a fluorescent antibody were scanned on a GeneTitan Multi-Channel Instrument (Thermo Fisher Scientific). The resultant data including more than 100 variants in *CYP2D6* and 60 variants in *CYP2C19* were analysed using the Axiom Analysis Suite 4.0.3.3 (Thermo Fisher Scientific). Updates were made to manufacturer's allele translation file for *CYP2D6* in order to improve the range of allele discrimination. CNV calls were provided by probes for exon 9 of *CYP2D6* as well as for the 5' and 3' flanking regions as described.[22]

Long-range PCR assays with characterisation of resultant amplicons

Haplotype assignment for samples with a call of three across the TaqMan or Pharmacoscan CNV probes was conducted by XL-PCR as described to generate an amplicon specific for the duplicated gene[20] followed by clean up of the resultant amplicons GenElute PCR Clean-Up Kit (Sigma Aldrich) and the use of TaqMan Drug Metabolising assays specific for the alleles found in samples of heterozygous *CYP2D6XN* genotype by at least one platform (**2*, **3*, **4*, **6*, **10, *35* and **41*, TaqMan assay IDs C_27102425_10, C_32407232_50, C_27102431_D0, C_32407243_20, C__11484460_40, C_27102444_F0, and C_34816116_20 respectively). Samples were run in duplicate on a ViiA7 Real-Time PCR System (Thermo Fisher Scientific), using automated genotype calling after visual inspection, outlier exclusion, and manual adjustment of C_T threshold settings. Data were exported using an automated method available from the authors at request.

Samples with unequal calls across the TaqMan, PharmacoScan or Ion S5 CNV probes were subjected to XL-PCR assays to generate products specific for *CYP2D6-2D7* or *CYP2D7-2D6* hybrids (amplicons E and G) as described, [21] with minor modifications. The resultant amplicons were then run on the Agena MassARRAYand the data compared with that arising from running the genomic DNA from the same samples through the Agena MassARRAY.

The Agena MassARRAY

The Agena MassARRAY (Agena Bioscience, San Diego, CA) uses matrix-assisted laser desorption/ionisation-time of flight (MALDI-TOF) mass spectrometry technology for resolving oligonucleotides. We used the Veridose Core plus *CYP2D6* copy number variant (CNV) panel for 18 *CYP2D6* variants plus gene deletion, duplication and *CYP2D6-2D7* conversion events and 9 *CYP2C19* variants. In addition, we ran five custom assays for *CYP2D6*: **25*, **35*, **40*, **42*, and **44*. 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) with star alleles being assigned using Typer Analyzer software version v4.1.83 (Agena Bioscience).

Results

CYP2C19

There was a mean of 1112.19 mapped reads per *CYP2C19* variant in the Ion S5 data. Data for *CYP2C19* are shown in Table 1. The *CYP2C19* Ion S5 data were concordant with the prior genotypic information from the AmpliChip CYP450 and TaqMan *CYP2C19*17* genotyping, apart from two samples, in each of which a poor metabolizer variant not included in AmpliChip CYP450 Test but included in the Ion S5 AmpliSeq Pharmacogenomics Panel was detected. These adjusted the prior genotypes of *CYP2C19*1/*1* and **1/*2* to **1/*8* and **2/*6* respectively, with corresponding changes in enzyme phenotypic assignment from NM and IM to IM and PM respectively. The genotypes for these two samples by PharmacoScan were also *CYP2C19*1/*8* and**2/*6*. There were two *CYP2C19*27* calls in the PharmacoScan data; these are, however, functionally indistinguishable from the *CYP2C19*1* allele (*CYP2C19*27* is now called *CYP2C19*1.006* on the PharmVar *CYP2C19 p*age). There was one functional discrepancy between the PharmacoScan data and other data: one of the samples that was positive for the *CYP2C19*27* by PharmacoScan was also called positive for a *CYP2C19*17* on that platform and was negative for *CYP2C19*17* on the prior TaqMan data and the IonS5. The IonS5 and PharmacoScan data were 100% concordant apart from those of *CYP2C19*1/*1* genotype by IonS5, for which the concordance was 94.5% (42/44, owing to one *CYP2C19*1/*27* by PharmacoScan and one *CYP2C19*17/*27* as above mentioned*)*. For the *CYP2C19*2*, the Ion S5, PharmascoScan and TaqMan data were 100% concordant.

CYP2D6

While the Ion S5 SNV data were of high quality (mean 1,111.13 mapped reads per variant), the CNV data had a relatively high proportion of no calls (4.21 and 18.95 for the gene level and exon 9 CNV probes respectively). The CNV data from PharmacoScan together with the prior TaqMan CNV data were therefore

used as a basis for grouping samples for further analysis. Comparative genotypic and CNV data across the three platforms are shown in Table 2.

For samples with a CNV call of two across the TaqMan or Pharmacoscan CNV probes, the concordance between Ion S5 and PharmacoScan was 94.1% (32/34), with the two non-concordant samples having variants detected by PharmacoScan that are not currently in the Ion S5 panel (*CYP2D6*22* and *CYP2D6*25*) (Supplementary Table 1). For samples with a CNV call of one across the probes, there was 100% concordance between the two platforms (Table 2). For samples with a CNV call of three across the TaqMan or Pharmacoscan CNV probes, the concordance was 76.5% (13/17) (Table 3). The four non-concordant samples had an undetected call for the Ion S5 exon 9 CNV probe, with a resultant likely erroneous putative assignment of hybrid allele containing genotypes in three of those ([23]). Concordance was less for samples with unequal calls across the TaqMan or Pharmacoscan or Ion S5 CNV probes, consistent with hybrids (formed from part of the *CYP2D6* gene and part of its adjacent *CYP2D7* gene).

Comparing the above with the prior data from AmpliChip revealed the following: For the CNV call two samples (Supplementary Table S1), the PharmacoScan-AmpliChip concordance was 77.1%, the Ion S5-AmpliChip concordance 82.4%. For the CNV call one samples, the PharmacoScan and AmpliChip genotypes were consistent with the CNV probe data, which were also consistent across TaqMan, PharmacoScan and Ion S5 CNV probes. The AmpliChip genotypes were not consistent with the CNV probe data and therefore likely represent erroneous prior calls in 41.7% (5/12). For the CNV three call samples, the PharmacoScan-AmpliChip and IonS5-AmpliChip concordance values were 73.3% (11/15) and 60% (9/15) respectively.

Samples with a CNV call of three across the TaqMan or Pharmacoscan CNV probes were all genotyped as having a *CYP2D6XN* (i.e., duplication or multiplication of the *CYP2D6* gene) by at least one platform apart from one sample. This sample had no prior AmpliChip data, and was genotyped as *CYP2D6*1/*4* by PharmacoScan and *CYP2D6*1/*4* and *CYP2D6*1/*4* or **1.011/*4.002* (functionally equivalent genotypes) on Ion S5. While the CNV call was three for the TaqMan probes for intron 2 and exon 9, it was two for the Pharmacoscan and Ion S5 CNV probes. The Pharmacoscan, TaqMan and Ion S5 exon 9 CNV probes all overlap in *CYP2D6* location ([22], Figure 1 and Thermo Fisher, personal communication). The relative locations of the three TaqMan probes and the Pharmacoscan 5' flanking and 3' flanking probes have been previously described ([22], Figure 1). The remaining Ion S5 *CYP2D6* "gene level" *CYP2D6* CNV call is integrated from probes in the intron 2 to exon 4 region as described above. The most likely consensus genotype for this sample is therefore *CYP2D6*1/*4*.

While the AmpliChip provided haplotyping of the *CYP2D6XN* (i.e., assignment to one or other of the two chromosomes), neither PharmacoScan nor Ion S5 attempt to do that. *CYP2D6XN* positive samples comprised samples with the following alleles: *CYP2D6*1*, **2*, **3*, **4*, **6,* and **41*. Haplotype assignment

may be conducted by XL-PCR followed by product clean-up and the use of relevant TaqMan assays as above described.

Sixteen samples had an unequal call across at least two out of three CNV probesets (TaqMan or PharmacoScan or Ion S5); for eight of these, the CNV pattern was consistent with a *CYP2D7-2D6* hybrid, and for eight 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 allele as the hybrid gene (in "cis"), or on the other chromosome 22 (in "trans"). This was also the case for four of the *CYP2D7-2D6* hybrids. The expected XL-PCR amplicons were generated for all of these 16 samples. For an additional five samples with putative hybrids owing to unequal call across the Ion S5 CNV probes but equal call across the PharmacoScan and TaqMan CNV probes, XL-PCR products consistent with hybrid types predicted by the Ion S5 CNV pattern could not be reliably generated. As the Ion S5 CNV probes have at present a relatively high failure rate and would therefore appear to be less reliable than the CNV probes of the other platforms, we conclude that the Ion S5 data were misleading, and in fact these samples do not represent hybrids. One sample was genotyped as *CYP2D6*1/*2* by all platforms and copy number 2 on all CNV probes apart from the TaqMan intron 2 CNV probe, which we conclude was an outlier in the data. Genotypes for the hybrids were predicted using the combination of available data and the samples may be further analysed by subjecting the XL-PCR amplicons to genotyping using, for example, the Agena MassARRAY.

Discussion

For *CYP2C19*, the overall Ion S5-PharmacoScan functional genotype concordance was 97.8% (92/94). Moreover, the Ion S5 and PharmacoScan *CYP2C19*2* data were 100% concordant with that from a TaqMan *CYP2C19*2* assay (assay ID C_25986767_70).

For *CYP2D6*, the IonS5 data showed a high rate of concordance with PharmacoScan solution for *CYP2D6* variants over a range of alleles tested apart from the *XN* and hybrid variants. Despite greater allele coverage by the PharmacoScan relative to the Ion S5, for samples without complex structural variants of *CYP2D6* genes, there were no genotypes that were inconsistent between the two platforms. To date it has been thought that it was not possible to employ next generation sequencing technology on genomic DNA for genotyping *CYP2D6*; rather, that one would have to first amplify up the whole gene and its surrounding regions and then use such technology.[22] The Ion S5 AmpliSeq Pharmacogenomics Panel is to our knowledge the first product that uses a next generation platform to genotype *CYP2D6* variants from genomic DNA. Sixty-two different *CYP2D6* genotypes according to the AmpliChip CYP450 data were represented in the 96 samples analyzed herein, equating to 82.7% (65/75) of the range of AmpliChip genotypes in the GENDEP data. Alleles not represented in the prior AmpliChip data include the hybrid variants and others that are infrequent in Caucasians.[24] We have herein demonstrated the feasibility of using

this technology in this manner for a substantial proportion of the variety of *CYP2D6* genotypes and all of the common *CYP2C19* genotypes seen in Whites. We have also developed tools to facilitate the bioinformatic interpretation of such data into "star alleles" of known function, so that enzyme activity can be deduced. Moreover, these tools could be adapted for the interpretation of data from specific variants in these genes arising from any other platform.

Both platforms were also able to call *CYP2D6*5* (gene deletion) variants in a manner consistent with each other and more reliably than the AmpliChip. Given the failure to detect *CYP2D6*5* reliably by the AmpliChip, the rest of available GENDEP samples were screened for *CYP2D6*5* using the TaqMan intron 2 and exon 9 probes, which revealed seven samples (7/853, 0.8%) with a CNV one call consistent with the presence of a previously undetected *CYP2D6*5* allele.

For the CNV call three samples (consistent with a duplicated *CYP2D6* gene or *XN* call, specifically *X2*, on one chromosome), the genotype on at least one platform was consistent with an *XN* call, with phase assignment of the *XN* being possible by amplifying the duplicated gene and using the resultant amplicon as a template for TaqMan SNV genotyping. The frequency of *CYP2D6XNs* by TaqMan CNV analysis in GENDEP was 3.8% (33/853), consistent with prior data.[19]

A focus of recent research on *CYP2D6* is the hybrid alleles.[21,25] Using TaqMan *CYP2D6* CNV probes, we have identified data patterns consistent with hybrid alleles in 2%, 18/853 GENDEP samples, and presented such data.[26] Where CNV call was unequal across platforms, the particular pattern of CNV call was used to infer types of hybrid allele configurations, and hence which XL-PCR set of primers should be employed to generate templates for SNV genotyping. Our previous validation work found a 100% concordance between TaqMan and DMETPlus *CYP2D6*41* genotyping and between TaqMan and PharmacoScan rs5758550 genotyping. [23,27-29]

The Luminex *CYP2C19* xTAG v3 has been validated against the Motorola Life Sciences eSensor DNA detection system. [30] From 2006 onwards the AmpliChip *CYP2C19* genotypes were problematic owing to the identification of a relatively common variant conferring increased enzyme activity (*CYP2C19*17*) [31] that was not covered by the assay. As this test was approved for *in vitro* diagnostic use by both the U. S. Federal Drug Administration and the European Medicines Agency, it could not be modified without going through costly regulatory approval again, and it was hence removed from the market amidst growing concerns regarding its comprehensiveness and sensitivity, despite its strengths including the apparent ability to discriminate heterozygous *CYP2D6XN* calls.

Data arising from the Luminex CYP2D6 xTAG v3 has been compared with the AmpliChip CYP450 Test, as well as other genotyping platforms including AutoGenomics INFINITI, ParagonDx, and LDT SNaPShot. [32] *CYP2D6* TaqMan assays have been compared with data arising from mPCR-RETINA,

Sanger sequencing, long-PCR for *CYP2D6*5*, and next-generation sequencing data available via the 1000 Genome Project. [33] In a recent publication, [22] the range of *CYP2D6* alleles available in reference samples was increased by using multiple technologies including PharmacoScan, Agena iPLEX CYP2D6 v1.1, TaqMan assays, XL-PCR, digital droplet PCR, next-generation sequencing of *CYP2D6* amplicons, and long-read single-molecule real-time sequencing of the whole *CYP2D6* gene. However, the use of a next generation sequencing platform to genotype genomic DNA was not included in the above, and much of the data interpretation appears to have been done manually, without automated methodology. While software exists to call *CYP2D6* alleles from next generation full sequencing data,[34] the novel contribution described herein is the use of the IonS5 AmpliSeq Panel, which employs short sequence reads as a genotyping tool and our allele translation files for the interpretation of *CYP2D6* and *CYP2C19* data arising.

We propose the following methodology for genotyping for *CYP2D6*: use a high throughput platform (e.g., Ion S5, PharmacoScan or Agena) with automated allele translation plus TaqMan *CYP2D6* CNV analysis, followed by XL-PCR for gene duplications or hybrids and using the resultant amplicons as templates for genotyping (using the same high throughput platform as used for genotyping the genomic DNA), and compare the data from the amplicons with that from genomic DNA to derive genotypes for the relevant *CYP2D6* structural variants. The utilization of a greater number of *CYP2D6* CNV probes by the platforms where necessary and improvements in performance of such probes where indicated should enhance the feasibility of this suggested workflow.

In addition to the complexity of the locus, another barrier that has limited *CYP2D6* to date has been the cost of the genotyping technology. The cheapest way to genotype is in fact by sequencing. The most high throughput (and therefore potentially most cost-effective) sequencing methodology is NGS. And while we have developed translators specifically for use with the Ion S5, they could be adapted for other NGS platforms such as Illumina. Should an NGS provider develop a panel including not only a greater number of *CYP2D6* SNVs covered than those included in the Ion S5 AmpliSeq Panel but also a greater number of more reliable *CYP2D6* CNV probes and offer this at a competitive price, this could be a game-changer: moving *CYP2D6* from the position of being a locus that by virtue of its challenging complexity has been inconsistently genotyped to one for which genotyping becomes accessible and affordable.

Conclusion

The inter-platform concordance for *CYP2C19* was high (94-98%). For *CYP2D6*, the IonS5-PharmacoScan concordance was 94% for a range of variants tested apart from those with at least one extra copy of a *CYP2D6* gene or a pseudogene-gene conversion. We have also demonstrated feasibility of using a NGS platform for genotyping *CYP2D6* and *CYP2C19*, and, moreover, that this provides adequate coverage of

CYP2D6 for approximately 94% of Whites. Further, we have generated automated methods for translating NGS *CYP2D6* sequencing data into *CYP2D6* haplotypes (star alleles). We suggest a minimum set of assays for *CYP2D6* and *CYP2C19* as follows: either NGS or another relevant platform (e.g., PharmacoScan or Agena MassARRAY) in combination with specific TaqMan probes for detecting *CYP2D6* copy number, with follow through of a subset of such data by XL-PCR and *CYP2D6* amplicon genotyping. *CYP2D6* and *CYP2C19* genotyping may thereby become much more readily accessible, with potential substantial savings in health care costs and public benefit.

Acknowledgments

We thank Carsten A. Bruckner for supplying the manufacturer's version of the *CYP2D6* translation for PharmacoScan and for his helpful feedback regarding draft edits to the file. The work reported herein was funded by: an Alberta Innovates Strategic Research Project (PRIME - Pharmacogenomics for the Prevention of Adverse Drug Reactions in mental health;l PI KJA, Co-PI Chad Bousman), an Alberta Centennial Addiction and Mental Health Research Chair (to KJA), the Canada Foundation for Innovation (CFI), John R. Evans Leaders Fund (JELF) (Pharmacogenetic translational biomarker discovery to KJA), Alberta Innovation and Advanced Education Small Equipment Grants Program (to KJA), and 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 sequencer) and a Canada Foundation for Innovation John R Evans Leaders Fund Grant (CFI-JELF) (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.

References

1. Huezo-Diaz P, Perroud N, Spencer EP, et al. CYP2C19 genotype predicts steady state escitalopram concentration in GENDEP. *Journal of psychopharmacology.* Mar 2012;26(3):398-407.
2. Aitchison KJ, Jordan BD, Sharma T. The relevance of ethnic influences on pharmacogenetics to the treatment of psychosis. *Drug metabolism and drug interactions.* 2000;16(1):15-38.
3. Lapetina DL, Yang EH, Henriques BC, Aitchison KJ. Pharmacogenomics and psychopharmacology. In: Peter M. Haddad, David J. Nutt, eds. *Seminars in Clinical Psychopharmacology*. London: Cambridge University Press; 2020:153-204.
4. Jukic MM, 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. *The American journal of psychiatry.* May 1 2018;175(5):463-470.
5. Yang EH, Henriques BC, Lapetina DL, et al. Insights from drug metabolising enzyme and transporter pharmacogenomics for mental health and addictions. Frontiers in Genetics, Pharmacogenetics and Pharmacogenomics section, Research Topic "Precision Psychiatry from a Pharmacogenetics Perspective," submitted 2019.
6. Gough AC, Smith CA, Howell SM, Wolf CR, Bryant SP, Spurr NK. Localization of the CYP2D gene locus to human chromosome 22q13.1 by polymerase chain reaction, in situ hybridization, and linkage analysis. *Genomics.* Feb 1993;15(2):430-432.
7. Kimura S, Umeno M, Skoda RC, Meyer UA, Gonzalez FJ. The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene. *American journal of human genetics.* Dec 1989;45(6):889-904.
8. Gaedigk A, Ingelman-Sundberg M, Miller NA, et al. The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. *Clinical pharmacology and therapeutics.* Mar 2018;103(3):399-401.
9. PharmVar. CYP2D6. 2019; https://www.pharmvar.org/gene/CYP2D6. Accessed Dec. 5, 2019.
10. Hicks JK, Swen JJ, Thorn CF, et al. Clinical Pharmacogenetics Implementation Consortium guideline for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants. *Clinical pharmacology and therapeutics.* May 2013;93(5):402-408.
11. Hicks JK, Sangkuhl K, Swen JJ, et al. Clinical pharmacogenetics implementation consortium guideline (CPIC) for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants: 2016 update. *Clinical pharmacology and therapeutics.* Jul 2017;102(1):37-44.
12. KNMP. Pharmacogenetics 2019; https://www.knmp.nl/patientenzorg/medicatiebewaking/farmacogenetica/pharmacogenetics-1. Accessed Dec. 5, 2019.
13. Goldstein JA, de Morais SM. Biochemistry and molecular biology of the human CYP2C subfamily. *Pharmacogenetics.* Dec 1994;4(6):285-299.
14. Chen Y, Goldstein JA. The transcriptional regulation of the human CYP2C genes. *Current drug metabolism.* Jul 2009;10(6):567-578.
15. Nelson DR, Zeldin DC, Hoffman SM, Maltais LJ, Wain HM, Nebert DW. Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics.* Jan 2004;14(1):1-18.
16. Botton MR, Lu X, Zhao G, et al. Structural variation at the CYP2C locus: Characterization of deletion and duplication alleles. *Human mutation.* Nov 2019;40(11):e37-e51.
17. PharmVar. CYP2C19. 2019; https://www.pharmvar.org/gene/CYP2C19. Accessed Dec. 5, 2019.

18. Pirmohamed M, James S, Meakin S, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *Bmj.* Jul 3 2004;329(7456):15-19.
19. Aitchison KJ, Munro J, Wright P, et al. Failure to respond to treatment with typical antipsychotics is not associated with CYP2D6 ultrarapid hydroxylation. *British journal of clinical pharmacology.* Sep 1999;48(3):388-394.
20. Gaedigk A, Ndjountche L, Divakaran K, et al. Cytochrome P4502D6 (CYP2D6) gene locus heterogeneity: characterization of gene duplication events. *Clinical pharmacology and therapeutics.* Feb 2007;81(2):242-251.
21. Black JL, 3rd, Walker DL, O'Kane DJ, Harmandayan M. Frequency of undetected CYP2D6 hybrid genes in clinical samples: impact on phenotype prediction. *Drug metabolism and disposition: the biological fate of chemicals.* Jan 2012;40(1):111-119.
22. Gaedigk A, Turner A, Everts RE, et al. Characterization of Reference Materials for Genetic Testing of CYP2D6 Alleles: A GeT-RM Collaborative Project. *The Journal of molecular diagnostics : JMD.* Nov 2019;21(6):1034-1052.
23. Lee D, Rossolatos D, Curran S, et al. Genotyping CYP2C19: Comparison of Various Technologies. Personal Genomes & Medical Genomics; Nov. 13-16 2013; New York.
24. Del Tredici AL, Malhotra A, Dedek M, et al. Frequency of CYP2D6 Alleles Including Structural Variants in the United States. *Frontiers in pharmacology.* 2018;9:305.
25. Carvalho HB, Wang Y, Whitford R, et al. CYP2D6: Detecting New Structures for Clinical Practice. The 15th Annual Pharmacogenetics in Psychiatry Meeting; October 2016; Maui. .
26. Carvalho HB, Talarico F, Lodhi R, et al. CYP2D6 and CYP2C19: Toward better clinical associations with functional dipoltype. World Congress of Psychiatric Genetics; Oct. 26-31, 2019; Los Angeles.
27. Slomp CJ, Rossolatos D, Tsapakis EM, et al. Comparative Data for CYP2D6 from Microarray and Taqman Technologies: A Pathway to Innovation. . Personal Genomes & Medical Genomics; Nov. 13-16, 2013; New York
28. Aitchison KJ. Drug metabolizing enzyme associations with response to agents used in neuropsychopharmacology. The 41st Annual Meeting of the Canadian College of Neuropsychopharmacology; June. 27-30, 2018; Vancouver
29. Wang Y, Carvalho HB, Behroozi AB, et al. In-depth analysis of CYP2D6, a gene highly relevant to neuropsychopharmacology. The 42nd Annual Meeting of the Canadian College of Neuropsychopharmacology; June. 12-15, 2019; Montreal
30. Pickering JW, McMillin GA, Gedge F, Hill HR, Lyon E. Flow cytometric assay for genotyping cytochrome p450 2C9 and 2C19: comparison with a microelectronic DNA array. *American journal of pharmacogenomics : genomics-related research in drug development and clinical practice.* 2004;4(3):199-207.
31. Sim SC, Risinger C, Dahl ML, et al. A common novel CYP2C19 gene variant causes ultrarapid drug metabolism relevant for the drug response to proton pump inhibitors and antidepressants. *Clin Pharmacol Ther.* Jan 2006;79(1):103-113.
32. Pratt VM, Zehnbauer B, Wilson JA, et al. Characterization of 107 genomic DNA reference materials for CYP2D6, CYP2C19, CYP2C9, VKORC1, and UGT1A1: a GeT-RM and Association for Molecular Pathology collaborative project. *The Journal of molecular diagnostics : JMD.* Nov 2010;12(6):835-846.
33. Fang H, Liu X, Ramirez J, et al. Establishment of CYP2D6 reference samples by multiple validated genotyping platforms. *The pharmacogenomics journal.* Dec 2014;14(6):564-572.
34. Lee SB, Wheeler MM, Patterson K, McGee S, Dalton R, Woodahl EL, Gaedigk A, Thummel KE, Nickerson DA. Stargazer: a software tool for calling star alleles from next-generation sequencing data using CYP2D6 as a model. *Genet Med* 2019 Feb;21(2):361-372.

.

Table 1. Percentage concordance for Ion S5 and PharmacoScan with AmpliChip CYP450 prior data by *CYP2C19* genotype

AmpliChip Genotype	N	Ion S5 (% concordance)	PharmacoScan (% concordance)
**1/*1*	45	97.8[1]	93.3[2]
**1/*17*	25	100[3]	100
**1/*2*	18	94.4[4]	94.4[4]
**17/*17*	1	100	0[5]
**2/*17*	4	100[6]	100
**2/*2*	2	100[7]	100

[1]All **1/*1* except one **1/*8*

[2]All **1/*1* except three (**1/*27*, **1/*8*, **17/*27*; note that *CYP2C19*27* has been reclassified in PharmVar as *CYP2C19*1.006*)

[3]100% of available data (N=24, with no call for one sample)

[4]All **1/*2* except one *2/*6

[5]**1/*17*

[6]All **2/*17* with one **2.002/*17*

[7]One **2/*2.002*

Table 2. Comparative *CYP2D6* data for samples with a CNV call of 1, with pairwise inter-platform concordance

TaqMan CNV data			PMS CNV data			IonS5	IonS5	PMS r8	AmpliChip	IonS5	consensus genotype	AmpliChip vs PMS	AmpliChip vs IonS5	PMS vs IonS5
I2	**I 6**	**E9**	**5'**	**E9**	**3'**	**gene**	**E9**							
1		1	1	1	1	1	1	**4/*5*	*4XN/*5*	**4/*5*	**4/*5*	Yes	Yes	Yes
1	1	1	1	1	1	1	1	**5/*35*	**5/*35*	**5/*35*	**5/*35*	Yes	Yes	Yes
1	1	1	1	1	1	1	1	**2/*5*	**2/*2*	**2.001/*5*	**2/*5*	No	No	Yes
1		1	1	1	1	1	1	**5/*10*	**5/*10*	**5/*10, *5/*10.003*	**5/*10*	Yes	Yes	Yes
1		1	1	1	1	1	1	**1/*5*	**1/*4*	**1/*5*	**1/*5*	No	No	Yes
1	1	1	1	1	1	1	1	**5/*41*	**5/*41*	**5/*41*	**5/*41*	Yes	Yes	Yes
1		1	1	1	1	1	1	**1/*5*	**1/*2XN*	**1/*5*	**1/*5*	No	No	Yes
1		1	1	1	1	1	1	**2/*5*	**2/*4*	**2.001/*5*	**2/*5*	No	No	Yes
1		1	1	1	1	1	1	**3A/*5*	**3/*5*	**3/*5*	**3/*5*	Yes	Yes	Yes
1		1	1	1	1	1	1	**5/*35*	**5/*35*	**5/*35*	**5/*35*	Yes	Yes	Yes
1		1	1	1	1	1	1	**3A/*5*	**4/*35*	**3/*5*	**3/*5*	No	No	Yes
1		1	1	1	1	1	1	**5/*41*	**5/*41*	**5/*41*	**5/*41*	Yes	Yes	Yes

PMS = PharmacoScan, I2 = intron 2, I6 = intron 6, E9 = exon 9, 5' = 5' flanking region, 3' = 3' flanking region; vs = versus

Table 3. Comparative *CYP2D6* data for samples with a CNV call of 3, with pairwise inter-platform concordance

TaqMan CNV data			PMS CNV data			Ion S5	IonS5	PMS r8	AmpliChip	IonS5	AmpliChip vs PMS	AmpliChip vs IonS5	PMS vs IonS5
I2	**I6**	**E9**	**5'**	**E9**	**3'**	**Gene**	**E9**						
3		3	3	3	3	3	3	*(*1/*4)XN*	**1XN/*4*	**1/*4XN, *1.011/*4.002XN, *1XN/*4*	Yes	Yes	Yes
3		3	3	3	3	3	UND	*(*1/*2)XN*	**1XN/*2*	*{*1/*2.001N, *1XN/*2.001, *2.001XN/*61}*	Yes	Yes	Yes
3		3	3	3	3	3	3	*(*1/*2)XN*	**1/*2XN*	**1/*2.001XN, *1XN/*2.001*	Yes	Yes	Yes
3		3	3	3	3	3	3	*(*1/*2)XN*	**1XN/*2*	**1/*2.001XN, *1XN/*2.001*	Yes	Yes	Yes
3		3	3[1]	3	3	3	3	*(*2/*4)XN*	**2XN/*4*	**2.001/*4XN, *2.001XN/*4*	Yes	Yes	Yes
3		3	3	3	3	3	3	*(*2/*4)XN*		**2.001/*4XN, *2.001XN/*4*			Yes
3		3	2	3	3	3	3	**2/*2*	**2/*2XN*	**2.001/*2.001XN*	No	Yes	Yes
3	3	3	2	3	3	3	3	**2/*2*	**2/*2XN*	**2.001/*2.001XN*	No	Yes	Yes
3	3	3	3	3	3	3	3	**2/*2XN*	**2XN/*41*	**2/*2.001XN, *2.001/*2XN*	No	No	Yes
3		3	3	3	3	3	3	*(*1/*4)XN*	**1/*4XN*	**1/*4XN, *1.011/*4.002XN, *1XN/*4*	Yes	Yes	Yes
3		3	2	3	3	2	UND	**2/*41*	**2/*41XN*	**2.001/*41*	No	No	Yes
3		3	2	2	2	2	2	**1/*4*		**1/*4, *1.011/*4.002*			Yes
3		3	3	3	3	UND	3	*(*1/*6)NX*	**1XN/*6*	**1XN/*6*	Yes	Yes	Yes
3		3	3	3	3	2	UND	**1/*1XN*	**1/*1XN*	*{*1/*1, *1/*61, *1XN/*5,...}*[2]	Yes	No	No
3		3	3	3	3	2	UND	**1/*1NX*	**1/*1XN*	*{*1/*1, *1/*61, *1XN/*5,...}*[2]	Yes	No	No
3		3	3	3	3	2	UND	*(*1/*2)XN*	**1XN/*2*	*{*1/*63, *61/*63}*[2]	Yes	No	No
3		3	3	3	3	2	UND	*(*2/*4)XN*	**2/*4XN*	**2.001/*4*	Yes	No	No

[1]Raw data consistent with a CNV call of 4 for the 5' flanking CNV probe; [2]Ion S5 tentative calls with **61* or **63* likely incorrect as associated with undetected exon 9 probe