Annals of human genetics

doi: 10.1111/j.1469-1809.2010.00585.x

Elucidation of *CYP2D6* Genetic Diversity in a Unique African Population: Implications for the Future Application of Pharmacogenetics in the Xhosa Population

Galen E. B. Wright[1], Dana J. H. Niehaus[2], Britt I. Drögemöller[1], Liezl Koen[2], Andrea Gaedigk[3] and Louise Warnich[1]*

[1]*Department of Genetics, Stellenbosch University, Private bag X1, Matieland 7602, South Africa*
[2]*Department of Psychiatry, Stikland Hospital, Stellenbosch University, PO Box 19090, Tygerberg 7505, South Africa*
[3]*Section of Developmental Pharmacology and Experimental Therapeutics, Children's Mercy Hospital & Clinics, 2401 Gillham Rd, Kansas City, Missouri 4108, USA*

Summary

Genetic variation of the *CYP2D6* gene has been associated with altered drug metabolism; however, limited studies have investigated *CYP2D6* sequence diversity in African populations. We devised a *CYP2D6* genotyping strategy to analyse the South African Xhosa population and genotype a Xhosa schizophrenia cohort, as CYP2D6 metabolises many antipsychotics and antidepressants.

The entire *CYP2D6* gene locus was sequenced in 15 Xhosa control individuals and the data generated were used to design a comprehensive genotyping strategy. Over 25 *CYP2D6* alleles were genotyped in Xhosa controls and Xhosa schizophrenia patients using long-range PCR, DNA sequencing and single nucleotide primer extension analysis. Bioinformatic algorithms were used to predict the functional consequences of relevant mutations and samples were assigned CYP2D6 activity scores.

A unique allele distribution was revealed and two rare novel alleles, *CYP2D6*73* and *CYP2D6*74*, were identified. No significant differences in allele frequencies were detected between Xhosa controls and schizophrenia patients. This study provides i) comprehensive data on a poorly characterised population, ii) a valuable *CYP2D6* genotyping strategy and iii) due to their unique genetic profile, provides the basis for pharmacogenetic intervention for Xhosa individuals.

Keywords: *CYP2D6*, pharmacogenetics, schizophrenia, South African population, Xhosa

Introduction

Elucidating the genetic components responsible for heterogeneity in drug response offers the potential to improve drug efficacy and reduce drug toxicity. This would be particularly beneficial to the field of psychiatry where the treatment of psychotic disorders such as schizophrenia is often inadequate (Lieberman et al., 2005; Arranz & De Leon, 2007). Accurate prediction of an individual's drug metabolism phenotype based on their genetic makeup and other factors (e.g. environmental and clinical) would allow for individualised treatment in the clinical setting and reduce the occurrence of adverse drug reactions and treatment non-response. The full catalogue of pharmacogenetic polymorphisms, however, remains to be determined and is further complicated by the fact that the frequency of genetic variants differs substantially between populations and ethnic groups (Zanger et al., 2004; Sistonen et al., 2007; Jakobsson et al., 2008). Therefore, before appropriate genotyping panels can be established, the extent of functional genetic variation in candidate genes needs to be determined, especially in understudied populations.

CYP2D6 is a drug metabolising enzyme that is of great importance to pharmacogenetic research, metabolising approximately 25% of all clinically prescribed drugs, including many anti-psychotics and anti-depressants (e.g. haloperidol and fluoxetine – for an extensive list see Zanger et al., 2004).

*Corresponding author: L. Warnich, Department of Genetics, Stellenbosch University, Private bag X1, Matieland 7602 South Africa. Tel: +27-21-808-5888; Fax: +21-21-808-5833; E-mail: lw@sun.ac.za

The polymorphic *CYP2D6* gene resides on chromosome 22q13, in close proximity to two homologous pseudogenes (Zanger et al., 2004). Genetic variation of this gene has been shown to have a strong effect on substrate metabolism and whole *CYP2D6* gene duplications and deletions, as well as hybrid genes, have been observed at significant frequencies in various populations (Kirchheiner et al., 2004; Zanger et al., 2004; Sistonen et al., 2007; Gaedigk et al., 2008; Gaedigk & Coetsee, 2008). Genotyping strategies are further complicated by the wide range of reduced-function and non-functional alleles, causing a variety of CYP2D6 phenotypes (Gaedigk et al., 2008). Approaches to inferring phenotype from *CYP2D6* genotype information, such as the CYP2D6 activity score (AS) approach described by Gaedigk et al. (2008) or the "semiquantitative gene dose" concept introduced by Steimer et al. (2004), facilitate the predictions of metaboliser status. In order to implement such tools in unique populations, novel *CYP2D6* alleles first need to be characterised and their respective capacity to metabolise CYP2D6 substrates assessed. *CYP2D6* genotype is believed to have relevance to predict antipsychotic response (Kobylecki et al., 2009), although more studies that measure clinical utility of genotyping of *CYP2D6* with regards to schizophrenia treatment are still required (Fleeman et al., 2010).

The implementation of pharmacogenetic technologies in African populations is particularly challenging. Not only have these groups been under-represented in past research, but it has been shown that African populations have the highest nucleotide diversity and lowest linkage disequilibrium in their genomes (Yu et al., 2002; Sistonen et al., 2007). This is due to a longer evolutionary history of African populations and their larger effective population size (Yu et al., 2002). Numerous studies have shown that the frequencies of the majority of *CYP2D6* alleles differ between population groups (Zanger et al., 2004; Sistonen et al., 2007; Gaedigk et al., 2008; Gaedigk & Coetsee, 2008). The *CYP2D6*17* and *CYP2D6*29* alleles are two examples of reduced function variants that occur at high frequencies in African cohorts and those of recent African origin, but are virtually absent from other population groups (Masimirembwa et al., 1996; Wennerholm et al., 2001; Sistonen et al., 2007). Despite the identification of many African-specific alleles, *CYP2D6* genotype-phenotype discordance remains a problem, indicating the need for more comprehensive genotyping strategies.

The Xhosa people represent a unique South African population and form part of the Nguni clan (Nurse et al., 1985; Seedat et al., 2004). According to the 2001 South African Population Census, 17.6% of the South African population speak isiXhosa as their first language, representing roughly 8 million individuals (http://www.statsonline.gov.za/census01/html/RSAPrimary.pdf). The Xhosa people, therefore, form an integral part of South African society, but to date no comprehensive study has been undertaken to determine the frequency of novel and known *CYP2D6* variants in this population or any other population of the Nguni group. Recently, novel *CYP2D6* alleles and a unique allele distribution were detected in a South African Mixed Ancestry cohort (Gaedigk & Coetsee, 2008), emphasising the need for similar studies to be undertaken in other South African populations. This study was aimed at setting up a unique strategy that could determine novel and known *CYP2D6* variants in the Xhosa population in a reasonably priced and time efficient manner when compared to other genotyping methods. Once established, the strategy was used to determine *CYP2D6* allele frequencies and AS in healthy Xhosa control individuals as well as Xhosa schizophrenia patients, since individuals suffering from this disorder are likely to receive a CYP2D6 substrate medication during pharmaceutical intervention.

Materials and Methods

Study Population, Clinical Assessment and DNA Isolation

A cohort of 53 healthy unrelated Xhosa controls was recruited from the Cape Metropole, South Africa. A group of 51 Xhosa schizophrenia patients was recruited from in- and outpatient psychiatric services in the same capture area as the controls. The inclusion criterion for the controls and schizophrenia patients was Xhosa ethnicity (4/4 grandparents reported as of Xhosa origin), while patients additionally required a diagnosis of schizophrenia according to the diagnostic and statistical manual of mental disorders (DSM-IV) criteria (American Psychiatric Association, 1994). Written informed consent was obtained from all individuals participating in this study, or their caretakers, and institutional approval was granted by the Committee for Human Research of Stellenbosch University. Genomic DNA (gDNA) was extracted from participants' whole blood using the protocol described by Miller et al. (1988) and gDNA integrity was then confirmed by means of agarose gel electrophoresis. Additional subjects ($n = 323$ Xhosa controls; $n = 443$ Xhosa schizophrenia patients) were recruited in the same manner as the original cohort, with identical inclusion criteria as those described above. The gDNA, however, was of inferior quality due to a period of imperfect storage and hence was only used for limited genotype analysis (Supplementary Table S1).

Sequence Analysis of the *CYP2D6* Gene

The spectrum of common DNA sequence variation in the *CYP2D6* gene locus, which we defined to be polymorphisms of $\geq$10%, was determined by sequencing 15 of the Xhosa controls that had tested negative for the *CYP2D6*5* deletion allele (see genotyping methods below). If the actual prevalence of a *CYP2D6* allele in the Xhosa population is 10%, then the probability that at least one such allele will be detected in these samples

is 95.8% (from the binomial distribution). This was accomplished by amplifying a 6.6-kb *CYP2D6* specific product, containing the entire gene, upstream sequence and 3′-UTR, using primers 6.6 kb-F and 6.6 kb-R (Supplementary Table S1). The entire 6.6-kb template was subsequently purified with use of MSB® Spin PCRapace columns (Invitek GmbH, Berlin, Germany). Bi-directional DNA sequencing was then performed on the entire template, bar 432 bp, using appropriately spaced primers (Supplementary Table S1) and Big Dye v3.1 sequencing chemistry (Applied Biosystems, Foster City, CA, USA). If insufficient amplification was achieved, the 6.6-kb template was diluted 1000 times and the sequencing primers were used to perform nested PCRs. These amplicons were subsequently used as templates for sequencing reactions. Numbering of variants was done according to the Genbank reference sequence, M33388, in accordance with the *CYP* allele nomenclature committee criteria. Haploview v4.0 was then used to examine the pattern of linkage disequilibrium observed between the variants (Barret et al., 2005). All *CYP2D6* polymorphisms detected in the Xhosa population were analysed for splicing changes using the program, NetGene2 (Brunak et al., 1991).

Genotype Analysis

CYP2D6 polymorphisms were selected for genotype analysis in the entire cohort using the following criteria: (i) polymorphisms detected in the sequencing data that have been shown to affect CYP2D6 function or expression, (ii) any non-synonymous mutations that were identified in the sequencing data and (iii) *CYP2D6* alleles that have been found to have pharmacogenetic importance in other populations, especially those of African origin. The genotyping panel should consist of a sufficient number of *CYP2D6* alleles to allow genotype designation to be both specific and sensitive, while remaining both reasonably priced and time efficient.

The panel of alleles genotyped in this study comprised *CYP2D6*2*, **2A*, **3*, **4*, **5*, **6*, **8*, **10*, **14*, **17*, **28*, **29*, **30*, **39*, **40*, **41*, **43*, **45*, **46*, **49*, **50*, **53*, **58*, **64*, **65* and **70*, as well as gene duplications and hybrid genes. The *CYP2D6*2A* allele was genotyped separately from *CYP2D6*2*, as the G allele of the -1584C>G SNP has been associated with increased activity of CYP2D6 *in vivo* as well as fast metabolism (Raimundo et al., 2000; Løvlie et al., 2001). The genotyping strategy combined long-range PCR, DNA sequencing and multiplex single nucleotide primer extension analysis. The region that was sequenced contained numerous important *CYP2D6* polymorphisms, as identified by our gene sequencing data and the literature, making this approach feasible. Single nucleotide primer extension analysis allowed for efficient genotyping of polymorphisms spread across the *CYP2D6* gene, while long-range PCR was employed to detect major genomic rearrangements at the *CYP2D6* locus and to prevent amplification of the *CYP2D7* and *CYP2D8P* pseudogenes. The genotyping strategy was validated using the sequenced Xhosa samples as positive controls or samples of known genotype obtained in other studies (Gaedigk et al., 2008; Gaedigk & Coetsee, 2008).

Genotyping by sequencing

The first step in the genotyping process involved amplifying the 6.6-kb *CYP2D6* fragment described above. Following amplification, an aliquot of the template was diluted 1000 times and used for the nested amplification of a 1.5-kb fragment, encompassing exons 2 to 4, using the primers 2D6 Ex2 F and 2D6 Ex3–4 R (Supplementary Table S1). The nested PCR product was subsequently purified as described above and then sequenced using the two Ex2–4 sequencing primers (Supplementary Table S1). Sequencing reaction cycling consisted of an initial denaturation step of 94°C for 5 minutes, followed by 30 cycles of 94°C for 10 seconds, 50°C for 10 seconds and 60°C for 4 minutes.

Genotyping by SNaPshot

The remaining 6.6-kb product was purified using the Nucleofast PCR kit (Macherey-Nagel, Düren, Germany) on a vacuum manifold, then interrogated for eight single nucleotide polymorphisms (SNPs) employing the single nucleotide primer extension reaction, SNaPshot (Applied Biosystems) as reported by Sistonen et al. (2005). A membrane based PCR cleanup kit was the most effective purification method for this particular amplicon, as the residual primers were refractive to digestion by *Exo* I. Genotyping primers used in this reaction were either modified from those described by Sistonen et al. (2005) or designed to genotype polymorphisms that were not analysed in that study. Genotyping primer sequences and their concentrations in the SNaPshot reaction can be found in Supplementary Table S1. A schematic summary of the polymorphisms genotyped by the SNaPshot and DNA sequencing are given in Figure 1.

Genotyping by long-range PCR

Detection of the *CYP2D6*5* deletion allele, *CYP2D6* duplications, as well as *CYP2D6/CYP2D7* hybrid genes was performed in three individual multiplex long-range PCR reactions, followed by electrophoresis on 0.8% agarose gels. PCR reaction and cycling conditions were modified from those described by Gaedigk et al. (2007), Gaedigk & Coetsee, (2008) and Steen et al. (1995) (Supplementary Table S1). Hybrid gene testing was performed in the final step of the genotyping protocol on samples that had homozygous genotypes. Characterisation of the nature of gene duplication was performed by amplifying fragments described by Gaedigk et al. (2007) (Supplementary Table S1), then genotyping those for allele-defining SNPs using the protocols described above.

Detection and Characterisation of *CYP2D6*73* and *CYP2D6*74*

The *CYP2D6*73* allele occurred in a sample that was heterozygous for the *CYP2D6*5* deletion. The 6.6-kb *CYP2D6* product was amplified from this sample and completely sequenced according to the protocol described above. The *CYP2D6*74* allele was detected in a sample that was heterozygous for *CYP2D6*2* and the 6.6-kb product was therefore cloned using the

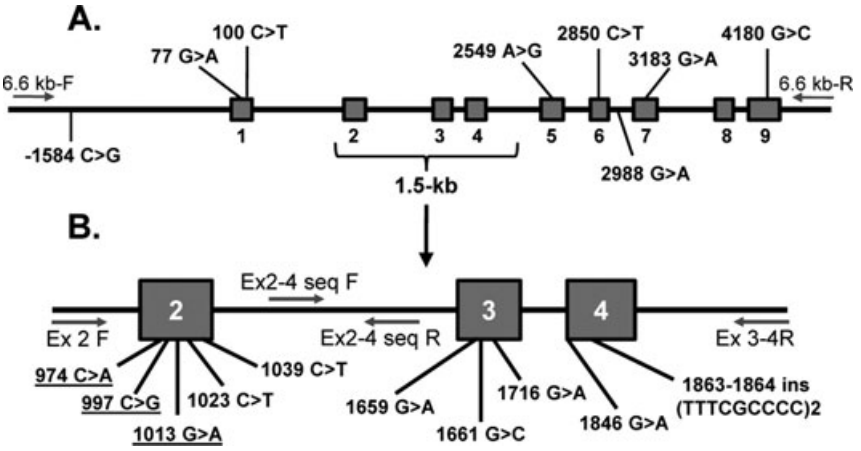

Figure 1 Schematic representation of the described *CYP2D6* genotyping strategy. (A) Depicts the 8 polymorphisms that were genotyped in the 6.6-kb fragment using SNaPshot as well as their locations in the gene. (B) The 1.5-kb nested PCR fragment was sequenced with the primers displayed in intron 2; the relative locations of significant polymorphisms found in this area are shown. *CYP2D6* polymorphisms not detected in this area in the 15 sequenced Xhosa controls, but found in the larger cohort, are underlined. Sequence variations not found in the Xhosa cohort, are not indicated in the figure.

pGEM®-T Easy Vector System (Promega, Madison, WI, USA), as per the manufacturer's protocol. Clones were genotyped for the 1013G>A SNP by means of PCR-restriction fragment length polymorphism (RFLP) (Supplementary Figure S1). The allele was then entirely sequenced and the presence of variants was confirmed by resequencing those regions in the original 6.6-kb product. Both alleles were submitted to the nomenclature committee for allelic designation (http://www.cypalleles.ki.se/). The functional effect of relevant non-synonymous sequence variations was predicted using the sorting intolerant from tolerant (SIFT) and PolyPhen algorithms (Ramensky et al., 2002; Ng & Henikoff, 2003). SIFT scores of <0.05 and normalised position-specific independent count (PSIC) scores of ≥ 1.5 were regarded as deleterious.

Statistical Analysis

Allele and genotype frequencies were calculated and deviations from Hardy-Weinberg Equilibrium (HWE) were determined using a Pearson χ^2 analysis in Tools For Population Genetic Analysis (TFPGA) Software v1.3. (Miller, 1997) If less than five individuals were observed for a particular genotype, a Fisher's exact analysis was performed instead. CYP2D6 activity scores (ASs) were calculated for all genotyped individuals using model A as described in Gaedigk et al. (2008). Novel *CYP2D6* alleles were classified as fully functional (i.e. AS = 1), as no phenotypic data were available. Differences between observed and expected allele and AS distribution between patients and controls were tested for using an analog to Fisher's exact test by way of appropriate contingency tables. *P*-values ≤ 0.05 were regarded as significant.

Results

Sequence Analysis of the *CYP2D6* Gene

Sequence analysis was performed on the *CYP2D6* gene locus for 15 Xhosa control individuals. High quality reads were obtained for the entire region of the 6.6-kb *CYP2D6* fragment. A total of 56 different polymorphisms, including the *CYP2D6*1*, **2*, **2AxN*, **4*, **10*, **17*, **29*, **40*, **43* and **45B* alleles, were detected in the 30 chromosomes. Of the detected polymorphisms 15 were located in the promoter, 14 in exons (8 non-synonymous SNPs, 5 synonymous SNPs and one 18 bp insertion), 23 in introns and 4 downstream of the 3′UTR (Supplementary Table S2). One of the exonic variations was a novel synonymous SNP in exon 5, 2452T>C, that was found in a sample initially genotyped as *CYP2D6*2AxN/*17*. Analysis of the promoter polymorphisms revealed that 13 of the 15 variants were either found in *Alu* or low complexity sequences. The -43insG and -176G>A promoter polymorphisms were not located in these regions and were each linked to the *CYP2D6*1* (D' 1.0, LOD 3.91, r^2 1.0) and *CYP2D6*29* (D' 1.0, LOD 5.32, r^2 1.0) alleles respectively. The -43insG polymorphism (haplotype undetermined) was originally detected by Solus et al. (2004) and appeared on half the *CYP2D6*1* alleles in the sequenced controls. The -176G>A polymorphism was detected in all samples with a *CYP2D6*29* allele and was one of five promoter variants that were linked to this allele.

CYP2D6 Genotyping Strategy

A unique genotyping strategy was successfully designed and used to genotype an additional 38 healthy Xhosa controls and 51 Xhosa schizophrenia patients (Supplementary Figure S1). Due to the size of this cohort ($n = 104$), comparisons were made between the groups to determine whether a *CYP2D6* genotyping strategy devised in a cohort of Xhosa controls could be utilised in schizophrenia patients of the same ethnicity. However, these comparisons were not made to elucidate the role *CYP2D6* genotype plays on schizophrenia-susceptibility as this was not the aim of the study. The *CYP2D6* allele frequencies detected in these samples are displayed in Table 1. *CYP2D6*2A*, **5*, **41*, **1xN* and **4xN* alleles were found in this larger cohort in addition to those found in the 15 sequenced controls, as well as a further ten sequence variants (Supplementary Table S2). No novel splicing mutations were predicted by NetGene2. A total of four non-functional alleles, i.e. *CYP2D6*4*, **5*, **40* and **4xN*, were observed in the Xhosa. Notably, no hybrid genes were detected in the original study cohort ($n = 104$). All alleles were in HWE for the two groups. *CYP2D6*45* was the only allele to show significant differences in allele frequencies between the two groups ($P = 0.004$). The *CYP2D6*5* deletion allele was the major non-functional allele in the Xhosa. Over 40% of the gene duplications in this ethnic group were non-functional, i.e. *CYP2D6*4xN*.

Table 1 Frequencies of *CYP2D6* alleles observed in the Xhosa controls and schizophrenia patients

Allele	CONTROL ($n = 53$)		SCHIZOPHRENIA ($n = 51$)	
	Count	Freq (%)	Count	Freq (%)
**1*	25	23.6	25	24.5
**2*	13	12.3	15	14.7
**2A*	0	0.0	1	1.0
**4*	2	1.9	1	1.0
**5*	15	14.2	19	18.6
**10*	2	1.9	2	2.0
**17*	14	13.2	17	16.7
**29*	14	13.2	7	6.9
**40*	2	1.9	3	2.9
**41*	2	1.9	1	1.0
**43*	1	0.9	1	1.0
**45*	11	10.4	1	1.0
**73*	0	0.0	1	1.0
**74*	0	0.0	1	1.0
**1xN*	0	0.0	1	1.0
**2AxN*	3	2.8	3	2.9
**4xN*	2	1.9	3	2.9
Total Chr	106	/	102	/
NF	21	20.0	26	25.0

n, individuals; Freq, frequency; Chr, chromosomes; NF, non-functional alleles

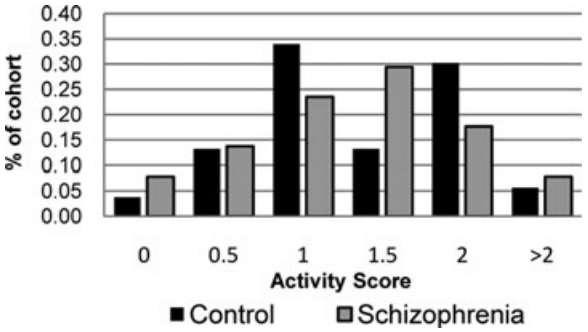

Figure 2 CYP2D6 activity score (AS)-Model A distribution (Gaedigk et al., 2008) in Xhosa schizophrenia patients and healthy Xhosa controls. PM subjects (i.e. no functional copies of *CYP2D6*) receive an AS of 0, while samples containing a functional allele as well as a functional gene duplication receive an AS > 2. The relatively high frequency of *CYP2D6*45* in the Xhosa control cohort is partially responsible for 30% of these subjects grouping in the AS 2 region. *CYP2D6*45* was originally classified as fully functional (Gaedigk et al., 2005) and was therefore assigned an AS of 1 in this study.

CYP2D6 Genotypes: Predicted Activity Scores and Phenotypes

The predicted AS for the two Xhosa cohorts is shown in Figure 2. No significant differences in AS distribution were detected between the cohorts ($P = 0.233$). A total of six poor metabolisers (PMs) were identified in this study, four in the schizophrenia group and two in the controls, all of which were accounted for by the *CYP2D6*5/*5, *5/*40* and **4xN/*40* genotypes. Individuals carrying more than two functional *CYP2D6* gene copies were defined as ultrarapid metabolisers (UMs) and occurred at a frequency of 3.8% in the entire study, with two samples detected in each group. Noteworthy, 12.5% of individuals were found at the extreme ends of the phenotype distribution by exhibiting genotypes assigned to the AS groups 0 (PM) or >2 (UM). These individuals (including eight schizophrenia patients) are most likely to respond in a non-ideal manner to CYP2D6 substrates.

Detection and Characterisation of *CYP2D6*73* and *CYP2D6*74*

The two novel alleles, *CYP2D6*73* and *CYP2D6*74*, were identified in the Xhosa schizophrenia cohort. The results of genotyping clones for the 974C>A *CYP2D6*74* SNP as well

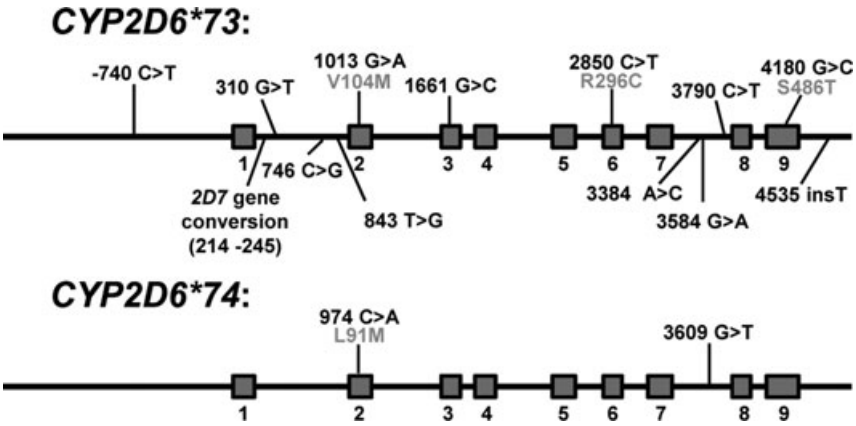

Figure 3 Schematic representation of the novel *CYP2D6*73* and **74* alleles. The key mutation of *CYP2D6*73* is 1013G>A (V104M), while that of *CYP2D6*74* is 974C>T (L91M).

as chromatograms of the key SNPs for both of the novel alleles can be found in Supplementary Figure S1. The *CYP2D6*73* allele-defining SNP, 1013G>A, occurs on the "*CYP2D6*2*-backbone" along with the 4535insT variation downstream of the 3′UTR. This novel SNP causes a V104M amino acid change in exon 2 of the CYP2D6 protein. Neither SIFT or PolyPhen algorithms predicted this mutation to be deleterious (SIFT score = 0.23, PSIC score = 0.767). *CYP2D6*74* is characterised by two SNPs, the 974C>A SNP in exon 2 that causes a L91M amino acid change and an intron 7 SNP, 3609G>T. 974C>A has previously been described to occur on a number of *CYP2D6*4* sub-variants, however, all *CYP2D6*4* detected in the Xhosa were negative for this SNP. *In silico* analysis of this variant with the SIFT algorithm predicted that it would affect protein function (SIFT score = 0.01), but PolyPhen predicted that this SNP would be benign (PSIC score = 0.512). During the genotyping procedure, a sample was identified as containing the synonymous 997C>G SNP, which has also been found to be linked to this subgroup of *CYP2D6*4* alleles as well as to *CYP2D6*2E*. This allele was also completely sequenced, but only this single sequence variation was detected. A schematic representation of the polymorphisms detected in the novel alleles can be found in Figure 3.

*CYP2D6*45* 1716G>A Genotyping

Since *CYP2D6*45* was the only allele to show significant differences between patients and controls after the genotyping strategy was employed, the allele-defining SNP (i.e. non-synonymous 1716G>A), was therefore genotyped in additional samples to rule out Type I error. The *CYP2D6*46* allele also contains the 1716G>A SNP, yet additionally carries the non-synonymous 77G>A substitution. Since the 77G>A polymorphism is not believed to influence CYPD26 functionality (Gaedigk et al., 2008) and because the *CYP2D6*46* allele was not detected in the original Xhosa cohort, this polymorphism was therefore not genotyped. Since the gDNA integrity of the majority of these samples was inadequate for long-range PCR, preventing genotyping of the *CYP2D6*5* deletion allele, all samples that carried the respective allele/polymorphism (i.e. heterozygotes and homozygotes) were therefore grouped together during analyses. In total, 1716G>A genotypes were successfully obtained for 454 Xhosa schizophrenia patients and 334 Xhosa control individuals for this polymorphism (results include the original cohort of 104 individuals). This pooled analysis of 1716G>A was no longer significant ($P = 0.941$), with 13.7% of the schizophrenia patients and 13.5% of the control individuals being carriers for the 1716A allele.

Discussion

CYP2D6 Allele Distribution in the Xhosa

To our knowledge, this is the most comprehensive study of variation at the *CYP2D6* locus in an indigenous black African population. A total of 17 *CYP2D6* alleles were identified in the Xhosa cohort, displaying a unique and complex allele distribution. This wide range of *CYP2D6* variants reiterates the need for innovative genotyping methods to accurately predict CYP2D6 phenotype in African populations. Studies of other pharmacogenetically relevant CYP enzymes have shown that the identification of novel alleles can lead to the need to reassess predicted phenotypes, even in extensively studied populations (Ragia et al., 2009). A detailed comparison of

selected *CYP2D6* allele frequencies to other populations is presented in Supplementary Table S3. The allele frequencies in the Xhosa vary from those described in the south-eastern/south-western and north-eastern Bantu populations by Sistonen et al. (2007), even when results from this study are adjusted to account for differences in genotyping methods, emphasising the need to separate Bantu speaking Africans into subgroups. The Xhosa belong to the southernmost branch of this linguistic group and despite some admixture with the Khoisan population, their gene pool has remained relatively isolated throughout the generations (Nurse et al., 1985; Seedat et al., 2004). A recent study resequenced the *CYP2D6* gene in samples from eight African populations, but included only one South African population, the Venda (Matimba et al., 2009) and *CYP2D6* allele frequency data in South African Bantu populations remains limited. The Venda do not speak an Nguni language and allele frequencies reported for this group differed from the Xhosa, corresponding to autosomal DNA and Y-chromosome data previously obtained in these populations (Lane et al., 2002). The *CYP2D6*10* allele was present at an unusually high frequency for an African population ($n = 9$, 19%) and *CYP2D6*4* was absent from this particular cohort (Matimba et al., 2009). However, the small sample sizes of the majority of these African population studies make it difficult to make accurate *CYP2D6* allele frequency comparisons, especially for the rarer alleles.

The *CYP2D6*5* deletion allele was found at a relatively high frequency in the Xhosa, especially in the schizophrenia group (i.e. 18.6%). A similar frequency of this allele was also previously observed in a South African Mixed Ancestry population ($n = 99$, 17.2%) (Gaedigk & Coetsee, 2008). The Xhosa, however, remain different in allele distribution to the Mixed Ancestry group in many other aspects, with the lack of the novel *CYP2D6*64*, *CYP2D6*65* and hybrid alleles representing the most apparent examples. The non-functional *CYP2D6*40* allele was observed at a relatively high frequency in the entire cohort (i.e. 2.4%) and the role of this allele in indigenous African populations may have been underestimated as it remains understudied.

Novel *CYP2D6* Alleles: *CYP2D6*73* and *CYP2D6*74*

The rare novel alleles identified in the schizophrenia cohort require further characterisation to determine the functional consequences of the mutations described in Figure 3. It has been suggested that low frequency variants such as these might explain, at least in part, the variability in drug response observed in Africans (Matimba et al., 2009). *CYP2D6*73* is distinguished from *CYP2D6*2* by the non-synonymous 1013G>A (V104M) SNP and a T-insertion at position 4535, however, *in silico* analysis did not predict V104M to be detrimental to protein function. SIFT and PolyPhen algorithms also did not predict a functional change for the key SNP of *CYP2D6*17*, T107I, which is located in close proximity on the same conserved β'-helix of the CYP2D6 protein. In contrast, previous research has demonstrated that 1023C>T (T107I) on *CYP2D6*17* is contributing to the diminished function of this allele (Oscarson et al., 1997). Given the close proximity of *CYP2D6*73* 1013G>A and *CYP2D6*17* 1023C>T and the fact that both alleles are otherwise encoding identical proteins, it is possible that *CYP2D6*73* could also confer a similar activity. Extensive functional characterisation of the *CYP2D6*73* allele is, however, required as it has been suggested that CYP2D6.17 may exhibit substrate-dependant activity (De Leon et al., 2009). With regards to *CYP2D6*74*, the key non-synonymous 974C>A (L91M) SNP is not novel *per se* as it is found on the majority of *CYP2D6*4* alleles; however the functional effect of this amino acid has never been further elucidated because CYP2D6.4 activity is completely abolished by the 1846G>A splicing defect mutation. Although the 974C>A variant was not predicted to alter splicing function, this SNP should be prioritised for functional studies as the SIFT algorithm predicted that it may affect protein function (SIFT score = 0.01). A detailed comparison of selected alleles observed in this study to other resequenced *CYP2D6* alleles (Gaedigk, unpublished) can be found in Figure 4. This figure illustrates that some known alleles have undocumented sub-variants.

Under-representation of *CYP2D6*45*

After the comprehensive genotyping strategy was employed, *CYP2D6*45* was the only allele to occur at a significantly different frequency between the case and control groups. However, when the allele-defining SNP of *CYP2D6*45* (1716G>A) was genotyped in additional samples, to rule out Type I error, the results were no longer statistically significant ($P = 0.941$). These data collected in the large cohort, combined with the fact that the *CYP2D6*45* allele is probably fully functional (Gaedigk et al., 2005), indicate that this allele is not involved in schizophrenia risk reduction in the Xhosa. A few studies have detected an association between *CYP2D6* genotype and schizophrenia-susceptibility, although these findings remain controversial and more research is required in this regard (for review, see Dorado et al., 2007).

CYP2D6 AS Distribution

No significant differences in AS distribution were detected between the schizophrenia patients and control individuals. The AS grouping observed in the Xhosa individuals shows

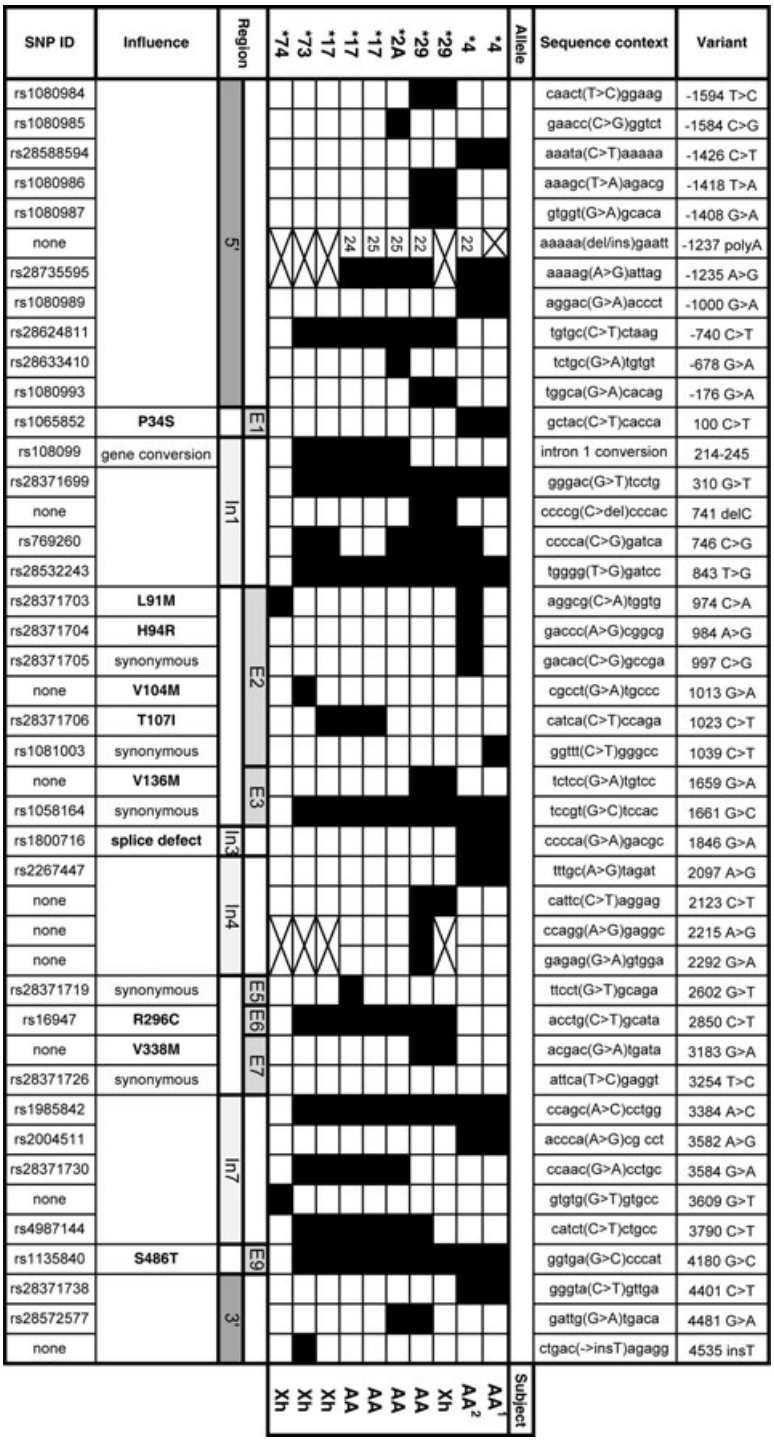

Figure 4 Comparison of selected *CYP2D6* alleles detected in this study to other resequenced alleles (Gaedigk, unpublished). Black boxes indicate the presence of a variant when compared to the reference sequence and "*X*" denotes a region for which no sequence was obtained. [1]All the *CYP2D6*4* alleles in the Xhosa (Xh) and 25% of those in African Americans (AA) possess the sub-variant with the 1039C>T SNP. [2]75% of *CYP2D6*4* alleles in African Americans possess this sub-variant with the 974C>A, 984A>G and 997C>G SNPs, which was not found in the Xhosa.

half of the samples ($n = 52$) in the 1–1.5 region. This is typical of African populations and is caused by the high number of reduced function alleles found in these groups (Gaedigk et al., 2008). The individuals most likely to benefit from pharmacogenetic testing before receiving CYP2D6 substrate medications are the 12.5% of the cohort that fall into AS groups 0 and >2 and represent subjects at the extreme end of the phenotype distribution. These results are important with regard to schizophrenia treatment with antipsychotics and antidepressants, as drug efficacy could be improved by adjusting medication dosage in these individuals prior to treatment (for review, see Kirchheiner et al., 2004). This is significant for Xhosa schizophrenia patients as the majority of individuals in this study were treated with the CYP2D6 substrate antipsychotic, haloperidol (Koen et al., 2008). However, since no *CYP2D6* genotype-phenotype studies have been undertaken in the Xhosa, it remains to be elucidated whether the AS system employed here accurately reflects CYP2D6 phenotype in this ethnic group (Gaedigk et al., 2008) and should be validated by future research. A total of 47 non-functional alleles were observed in the Xhosa controls and schizophrenia patients (23.0%). This non-functional allele frequency is similar to that found in Caucasians (25.8%) and the South African Mixed Ancestry population (26.8%) (Gaedigk et al., 2008; Gaedigk & Coetsee, 2008) and greater than that found in African Americans (15.6%) (Gaedigk et al., 2008). The number of UMs in the pooled study cohort (3.8%) is higher than the frequency found in South African Mixed Ancestry individuals (2.3%), while the PM frequency in these two populations was comparable (~6% in both populations) (Gaedigk & Coetsee, 2008).

Extensive *CYP2D6* Genotyping Strategy

The genotyping strategy utilised in this study was comprehensive, testing for DNA variants ranging from major genomic rearrangements at the *CYP2D6* locus to SNPs. Over 25 *CYP2D6* alleles and gene duplications were genotyped in the cohort and two novel *CYP2D6* alleles were identified. The combination of SNaPshot analysis and partial sequencing (i.e. 1/3 of the *CYP2D6* gene's exons) was also less labour-intensive and more thorough than most conventional approaches (e.g. RFLP analysis), while still remaining reasonably priced and time efficient. The genotyping strategy described here can be performed and results analysed within 72 hours. While complete sequencing of the *CYP2D6* gene is the most sensitive approach, it remains too costly for routine diagnostic use and is 5-fold more expensive than the methods described in this study. The novel genotyping strategy could be used in future pharmacogenetic studies in the Xhosa schizophrenia population as all of the alleles initially detected in the control cohort were also detected in the patient group. In developing nations such as South Africa, innovative approaches such as those described here, are required before the benefits of molecular medicine can be realised in the clinical setting.

Conclusion

Extensive pharmacogenetic studies in African populations such as the Xhosa remain sparse and the small sample sizes used in many other studies make it difficult to determine if the described frequencies are accurate. This study therefore meets the demand for resequencing data of polymorphic drug metabolising enzymes in unique populations, where treatment failure and adverse drug reactions are serious concerns (Mehta et al., 2008). The data presented here are not only valuable to the Xhosa schizophrenia population, but can also be used in pharmacogenetic studies of other diseases in this ethnic group. Our *CYP2D6* genotyping strategy can also be adapted to include polymorphisms representing other *CYP2D6* alleles if required and could easily be employed in other African populations and therefore aid in comprehensively characterising *CYP2D6* variation in these underrepresented groups in future research. The relationship between *CYP2D6* variants and the schizophrenia-susceptibility and pharmaceutical treatment response also requires further enquiry in a larger Xhosa cohort, as well as the functional characterisation of the novel alleles. The characterisation of the *CYP2D6* gene locus in the South African Xhosa population therefore supplies valuable information to local physicians and geneticists and provides the first step towards pharmacogenetic intervention in Xhosa patients.

Acknowledgements

Financial support was granted by the National Research Foundation (NRF) and the Harry Crossley Foundation. GW received bursaries from the Harry Crossley Foundation, the NRF and Stellenbosch University. The authors are greatly indebted to Dr M Venter for assistance with the molecular cloning and critical reading of the manuscript. We would also like to acknowledge Mr C van Heerden and Ms R Veikondis (Central Analytical Facility, Stellenbosch University) for assistance with DNA sequence and SNaPshot analyses.

References

American Psychiatric Association. (1994) *Diagnostic and Statistical Manual of Mental Disorders. 4th ed*. Washington, DC: APA.

Arranz, M. J. & de Leon, J. (2007) Pharmacogenetics and pharmacogenomics of schizophrenia: a review of the last decade of research. *Mol Psychiatry* **12**, 707–747.

Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265.

Brunak, S., Engelbrecht, J. & Knudsen, S. (1991) Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J Mol Biol* **220**, 49–65.

de Leon, J., Susce, M. T., Johnson, M., Hardin, M., Maw, L., Shao, A., Allen, A. C., Chiafari, F. A., Hillman, G. & Nikoloff, D. M. (2009) DNA microarray technology in the clinical environment: the AmpliChip CYP450 test for CYP2D6 and CYP2C19 genotyping. *CNS Spectr* **14**, 19–34.

Dorado, P., Peñas-LLedó, E. M. & LLerena, A. (2007) CYP2D6 polymorphism: implications for antipsychotic drug response, schizophrenia and personality traits. *Pharmacogenomics* **8**, 1597–1608.

Fleeman, N., McLeod, C., Bagust, A., Beale, S., Boland, A., Dundar, Y., Jorgensen, A., Payne, K., Pirmohamed, M., Pushpakom, S., Walley, T., de Warren-Penny, P. & Dickson, R. (2010) The clinical effectiveness and cost-effectiveness of testing for cytochrome P450 polymorphisms in patients with schizophrenia treated with antipsychotics: a systematic review and economic evaluation. *Health Technol Assess* **14**, 1–157.

Gaedigk, A., Bhathena, A., Ndjountché, L., Pearce, R. E., Abdel-Rahman, S. M., Alander, S. W., Bradford, L. D. & Leeder, J. S. (2005) Identification and characterization of novel sequence variations in the cytochrome P4502D6 (CYP2D6) gene in African Americans. *Pharmacogenomics J* **5**, 173–182.

Gaedigk, A., Ndjountché, L., Divakaran, K., Bradford, L. D., Zineh, I., Oberlander, T. F., Brousseau, D. C., McCarver, D. G., Johnson, J. A., Alander, S. W., Riggs, W. K. & Leeder, J. S. (2007) Cytochrome P4502D6 (CYP2D6) gene locus heterogeneity: characterization of gene duplication events. *Clin Pharmacol Ther* **81**, 242–251.

Gaedigk, A. & Coetsee, C. (2008) The CYP2D6 gene locus in South African Coloureds: unique allele distributions, novel alleles and gene arrangements. *Eur J Clin Pharmacol* **64**, 465–475.

Gaedigk, A., Simon, S. D., Pearce, R. E., Bradford, L. D., Kennedy, M. J. & Leeder, J. S. (2008) The CYP2D6 activity score: translating genotype information into a qualitative measure of phenotype. *Clin Pharmacol Ther* **83**, 234–242.

Jakobsson, M., Scholz, S. W., Scheet, P., Gibbs, J. P., VanLiere, J. M., Fung, H.-C. Szpiech, Z. A., Degnan, J. H., Wang, K., Guerreiro, R., Bras, J. M., Schymick, J. C., Hernandez, D. G., Traynor, B. J., Simon-Sanchez, J., Matarin, M., Britton, A., van de Leemput, J., Rafferty, I., Bucan, M., Cann, H. M., Hardy, J. A., Rosenberg, N. A. & Singleton, A. B. (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **45**, 998–1003.

Kirchheiner, J., Nickchen, K., Bauer, M., Wong, M.-L., Licinio, J., Roots, I. & Brockmöller, J. (2004) Pharmacogenetics of antidepressants and antipsychotics: the contribution of allelic variations to the phenotype of drug response. *Mol Psychiatry* **9**, 442–472.

Kobylecki, C. J., Jakobsen, K. D., Hansen, T., Jakobsen, I. V., Rasmussen, H. B. & Werge, T. (2009) CYP2D6 genotype predicts antipsychotic side effects in schizophrenia inpatients: a retrospective matched case-control study. *Neuropsychobiology* **59**, 222–226.

Koen, L., Magni, P., Niehaus, D. J. & Le Roux A. (2008) Antipsychotic prescription patterns in Xhosa patients with schizophrenia or schizoaffective disorder. *Afr J Psychiatry* **11**, 287–290.

Lane, A. B., Soodyall, H., Arndt, S., Ratshikhopha, M. E., Jonker, E., Freeman, C., Young, L., Morar, B. & Toffie, L. (2002) Genetic substructure in South African Bantu-speakers: evidence from autosomal DNA and Y-chromosome studies. *Am J Phys Anthropol* **119**, 175–185.

Lieberman, J. A., Stroup, T. S., McEvoy, J. P., Swartz, M. S., Rosenheck, R. A., Perkins, D. O., Keefe, R. S., Davis, S. M., Davis, C. E., Lebowitz, B. D., Severe, J., Hsiao J. K. & Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Investigators (2005) Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med* **353**, 1209–1223.

Løvlie, R., Daly, A. K., Matre, G. E., Molven, A. & Steen, V. M. (2001) Polymorphisms in CYP2D6 duplication-negative individuals with the ultrarapid metabolizer phenotype: a role for the CYP2D6*35 allele in ultrarapid metabolism? *Pharmacogenetics* **11**, 45–55.

Masimirembwa, C., Persson, I., Bertilsson, L., Hasler, J. & Ingelman-Sundberg, M. (1996) A novel mutant variant of the CYP2D6 gene (CYP2D6*17) common in a black African population: association with diminished debrisoquine hydroxylase activity. *Br J Clin Pharmacol* **42**, 713–719.

Matimba, A., Del-Favero, J., Van Broeckhoven, C. & Masimirembwa, C. (2009) Novel variants of major drug-metabolising enzyme genes in diverse African populations and their predicted functional effects. *Hum Genomics* **3**, 169–190.

Mehta, U., Durrheim, D. N., Blockman, M., Kredo, T., Gounden, R. & Barnes, K. I. (2008) Adverse drug reactions in adult medical inpatients in a South African hospital serving a community with a high HIV/AIDS prevalence: prospective observational study. *Br J Clin Pharmacol* **65**, 396–406.

Miller, S. A., Dykes, D. D. & Polesky, H. F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 1215.

Miller M. (1997) Tools for population genetic analyses (TFGA) 1.3: a Windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by the author.

Ng, P. C. & Henikoff, S. (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812–3814.

Nurse, G. T., Weiner, J. S. & Jenkins, T. (1985) The Khoi as migrants and nomads. In: *The peoples of Southern Africa and their Affinities*. (ed. G. A. Harrison), p81–83. Oxford: Clarendon Press.

Oscarson, M., Hidestrand, M., Johansson, I. & Ingelman-Sundberg, M. (1997) A combination of mutations in the CYP2D6*17 (CYP2D6Z) allele causes alterations in enzyme function. *Mol Pharmacol* **52**, 1034–1040.

Ragia, G., Arvanitidis, K. I., Tavridou, A. & Manolopoulos, V. G. (2009) Need for reassessment of reported CYP2C19 allele frequencies in various populations in view of CYP2C19*17 discovery: the case of Greece. *Pharmacogenomics* **10**, 43–49.

Raimundo, S., Fischer, J., Eichelbaum, M., Griese, E.-U., Schwab, M. & Zanger, U. M. (2000) Elucidation of the genetic basis of the common 'intermediate metabolizer' phenotype for drug oxidation by CYP2D6. *Pharmacogenetics* **10**, 577–581.

Ramensky, V., Bork, P. & Sunyaev, S. (2002) Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* **30**, 3894–3900.

Seedat, S., Emsley, R. A. & Stein, D. J. (2004) Land of promise: challenges and opportunities for research in South Africa. *Mol Psychiatry* **9**, 891–892.

Sistonen, J., Fuselli, S., Levo, A. & Sajantila, A. (2005) CYP2D6 genotyping by a multiplex primer extension reaction. *Clin Chem* **51**, 1291–1295.

Sistonen, J., Sajantila, A., Lao, O., Corander, J., Barbujani, G. & Fuselli, S. (2007) CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure. *Pharmacogenet Genomics* **17**, 93–101.

Solus, J. F., Arietta, B. J., Harris, J. R., Sexton, D. P., Steward, J. Q., McMunn, C., Ihrie, P., Mehall, J. M., Edwards, T. L. & Dawson, E. P. (2004) Genetic variation in eleven phase I drug metabolism genes in an ethnically diverse population. *Pharmacogenomics* **5**, 895–931.

Steen, V. M., Andreassen, O. A., Daly, A. K., Tefre, T., Børresen, A. L., Idle, J. R. & Gulbrandsen, A. K. (1995) Detection of the poor metabolizer-associated CYP2D6(D) gene deletion allele by long-PCR technology. *Pharmacogenetics* **5**, 215–223.

Steimer, W., Zöpf, K., von Amelunxen, S., Pfeiffer, H., Bachofer, J., Popp, J., Messner, B., Kissling, W. & Leucht, S. (2004) Allele-specific change of concentration and functional gene dose for the prediction of steady-state serum concentrations of amitriptyline and nortriptyline in CYP2C19 and CYP2D6 extensive and intermediate metabolizers. *Clin Chem* **50**, 1623–1633.

Wennerholm, A., Johansson, I., Hidestrand, M., Bertilsson, L., Gustafsson, L. L. & Ingelman-Sundberg, M. (2001) Characterization of the CYP2D6*29 allele commonly present in a black Tanzanian population causing reduced catalytic activity. *Pharmacogenetics* **11**, 417–427.

Yu, N., Chen, F. C., Ota, S., Jorde, L. B., Pamilo, P., Patthy, L., Ramsay, M., Jenkins, T., Shyue, S. K. & Li, W. H. (2002) Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* **161**, 269–274.

Zanger, U. M., Raimundo, S. & Eichelbaum M. (2004) Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn-Schmiedeberg's Arch Pharmacol* **369**, 23–37.

Supporting Information

Additional supporting information may be found in the online version of this article:

Figure S1 Examples of *CYP2D6* genotyping results
Table S1 *CYP2D6* primers and reaction conditions
Table S2 *CYP2D6* polymorphisms detected in the Xhosa cohort and the frequencies detected in each of the groups
Table S3 Selected *CYP2D6* allele frequencies in different populations

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organised for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received: 8 February 2010
Accepted: 4 April 2010