

# Novel variants of major drug-metabolising enzyme genes in diverse African populations and their predicted functional effects

Alice Matimba,<sup>1,2\*</sup> Jurgen Del-Favero,<sup>3,5</sup> Christine Van Broeckhoven<sup>4,5</sup> and Collen Masimirembwa<sup>1</sup>

<sup>1</sup>African Institute of Biomedical Science & Technology (AiBST), P.O. Box 2294, Harare, Zimbabwe

<sup>2</sup>Division of Human Genetics, IIDMM, University of Cape Town, Observatory 7925, South Africa

<sup>3</sup>Applied Molecular Genomics Group, Department of Molecular Genetics, VIB, Universiteitsplein 1, Wilrijk, B-2610, Belgium

<sup>4</sup>Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, VIB, Universiteitsplein 1, Wilrijk, B-2610, Belgium

<sup>5</sup>University of Antwerp (UA), Campus Drie Eiken, Universiteitsplein 1, Wilrijk, B-2610, Belgium

\*Correspondence to: Tel: +27 21 4066501; Fax: +27 21 4066826; E-mail: Alice.Matimba@uct.ac.za

Date received (in revised form): 15th August 2008

## Abstract

Pharmacogenetics enables personalised therapy based on genetic profiling and is increasingly applied in drug discovery. Medicines are developed and used together with pharmacodiagnostic tools to achieve desired drug efficacy and safety margins. Genetic polymorphism of drug-metabolising enzymes such as cytochrome P450s (CYPs) and N-acetyltransferases (NATs) has been widely studied in Caucasian and Asian populations, yet studies on African variants have been less extensive. The aim of the present study was to search for novel variants of *CYP2C9*, *CYP2C19*, *CYP2D6* and *NAT2* genes in Africans, with a particular focus on their prevalence in different populations, their relevance to enzyme functionality and their potential for personalised therapy. Blood samples from various ethnic groups were obtained from the AiBST Biobank of African Populations. The nine exons and exon–intron junctions of the *CYP* genes and exon 2 of *NAT2* were analysed by direct DNA sequencing. Computational tools were used for the identification, haplotype analysis and prediction of functional effects of novel single nucleotide polymorphisms (SNPs). Novel SNPs were discovered in all four genes, grouped to existing haplotypes or assigned new allele names, if possible. The functional effects of non-synonymous SNPs were predicted and known African-specific variants were confirmed, but no significant differences were found in the frequencies of SNPs between African ethnicities. The low prevalence of our novel variants and most known functional alleles is consistent with the generally high level of diversity in gene loci of African populations. This indicates that profiles of rare variants reflecting interindividual variability might become the most relevant pharmacodiagnostic tools explaining Africans' diversity in drug response.

**Keywords:** pharmacogenetics, cytochrome P450, N-acetyltransferase, single nucleotide polymorphisms, African populations

## Introduction

Pharmacogenetics describes patients' variation in response to therapy due to genetic factors.

Pharmacogenetics-based therapy is of special interest for drugs with narrow therapeutic indices, where impairment in metabolic activity might cause

difficulties in dose adjustment, resulting in increased susceptibility to adverse drug reactions (ADRs). The cytochrome P450 enzymes (CYPs) metabolise more than 80 per cent of clinically used drugs and most of them exhibit functionally significant genetic polymorphisms. The genes encoding CYP2C9, CYP2C19 and CYP2D6, as well as N-acetyltransferase 2 (NAT2), have been most extensively studied across various populations.<sup>1,2</sup> The presence of novel variants remains to be ascertained in African populations, however, particularly rare (<1 per cent frequency) single nucleotide polymorphisms (SNPs), which may contribute to a better understanding of interindividual variation in the metabolism of drugs.

The human CYP2C subfamily contains four highly homologous genes — *2C8*, *2C9*, *2C18* and *2C19* — which are located in a cluster on chromosome 10.<sup>3</sup> CYP2C9 is the main CYP2C enzyme, constituting 20 per cent of total human liver microsomal P450 content.<sup>4</sup>

*CYP2C9* and *CYP2C19* genes each contain nine exons and encode proteins of 490 amino acids in length. Although these genes are highly homologous (92 per cent), the enzymes differ in terms of substrate specificities.<sup>5</sup> Major variations in the occurrence of polymorphisms in both *CYP2C9* and *CYP2C19* genes have been reported in various populations. *CYP2C9* variants *CYP2C9\*2* and *CYP2C9\*3* are the most common and occur at frequencies of 0.11 and 0.08, respectively, in Caucasians.<sup>6</sup> Population-based pharmacokinetics–pharmacodynamics modelling of their effects has been explored for revising labels of CYP2C9 substrate drugs.<sup>7</sup>

Testing for *CYP2C9* genotypes can be used to predict the starting dose of the anticoagulant drug warfarin to avoid excessive bleeding episodes.<sup>8</sup> Other drugs affected by *CYP2C9* polymorphism are the antidiabetic agents glipizide and tolbutamide, the anti-epileptic agent phenytoin, the antihypertensive drug losartan and non-steroidal anti-inflammatory drugs (NSAIDs) such as ibuprofen and diclofenac.<sup>9</sup>

CYP2C19 metabolises omeprazole, diazepam and proguanil to a major extent. The common allelic variants, such as *CYP2C19\*2* and *CYP2C19\*3*, cause reduced enzyme activity and contribute to the poor metabolism of substrate drugs.<sup>10</sup> A polymorphism in

the promoter region has, however, been associated with increased enzyme activity.<sup>11</sup> Individuals carrying this variant may therefore require a higher dosage in order to achieve the therapeutic effect.

CYP2D6 metabolises a wide range of drugs, such as antiarrhythmic agents, tricyclic antidepressants, neuroleptics and anti-cancer agents.<sup>12</sup> *CYP2D6* is the most polymorphic CYP, with alleles causing a spectrum of phenotypic responses. The presence of multiple copies of the gene results in individuals described as ultra-rapid metabolisers. For example, individuals carrying duplicated or multi-duplicated active *CYP2D6* genes are very common among Ethiopians, compared with Caucasian, Oriental and other Black populations.<sup>13</sup> By contrast, whole gene deletions causing poor metaboliser phenotypes, have been observed across all populations. The African-specific alleles *CYP2D6\*17* and *CYP2D6\*29* cause reduced enzyme activity; individuals homozygous for these alleles are classified as intermediate metabolisers. Overall, Africans metabolise CYP2D6 substrates at a slower rate than Caucasians owing to the higher prevalence of these reduced-function alleles.<sup>14</sup>

So far, NAT2 has been found to comprise 19 major known haplotypes. Important drugs metabolised by this enzyme include the anti-tuberculosis drug isoniazid and the antibiotic co-trimoxazole. Some polymorphisms of NAT2 have been shown to affect the acetylation of these drugs and this may result in toxic side effects.<sup>15,16</sup> The most commonly known alleles are *NAT2\*5*, *NAT2\*6*, *NAT2\*7* and the African-specific *NAT2\*14*. In addition, other SNPs have been discovered and are awaiting characterisation of their phenotypic effects.

In clinical pharmacogenetics, we aim to optimise therapeutic outcome by prescribing drugs to patients at doses that are predicted to be efficacious and safe. Knowledge of the types of genetic variants of major drug-metabolising enzymes and their frequency in the population is therefore important for the design and deployment of pharmacodiagnostic tools to guide drug prescription. Only a few studies on genotype–phenotype relationships of drug effects have been carried out in African populations.<sup>16–20</sup> Therefore, limited knowledge of polymorphisms and their impact in Africans may underestimate the importance

of clinical applications of pharmacogenetics. Here, we report novel variants of the *CYP2C9*, *CYP2C19*, *CYP2D6* and *NAT2* genes found in African populations and their predicted functional effects.

## Materials and methods

### DNA samples

The study was carried out according to the Declaration of Helsinki (2000) of the World Medical Association and was approved by the Ethical Review Boards of Kenya, Nigeria, Tanzania and Zimbabwe. Informed consent was obtained from volunteers of the following ethnic groups: Hausa (20), Ibo (20), Luo (30), Maasai (13), San (40), Shona (23), Venda (9), Yoruba (20) and Tanzanian Mixed Bantu (12). Ethnicity was assigned based on the submission that parents and grandparents of the volunteers were of the same self-identified ethnic group. The exact numbers of samples analysed per gene are shown in Tables S1–S4 in the Appendix. DNA was extracted from whole blood samples stored in the AiBST Biobank of African Populations<sup>21</sup> using the QIAamp DNA Blood Mini Kit (Qiagen, KJ Venlo, The Netherlands).

### PCR and sequencing

Primers were designed using SNPBox and Primer3 software.<sup>22,23</sup> Their specificity for each gene studied was confirmed by a BLAST analysis search and comparison of genomic sequences in the National Center for Biotechnology Information (NCBI) databases (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Identical primers were used for the polymerase chain reaction (PCR) and sequencing, except where otherwise stated (Table S5). First-step exon amplification mixtures (20  $\mu$ l) contained 1x TiTaq buffer, 0.25 mM deoxyribonucleotide triphosphates, 0.5  $\mu$ M of each primer and 0.25 units of TiTaq and DNA template (5 ng/ $\mu$ l). For PCR, an initial denaturation at 94°C for two minutes was followed by 35 cycles at 94°C for 30 seconds, 59°C for 30 seconds and 72°C for 60 seconds. Sequencing reactions were started at 96°C for one minute followed by 25 cycles at 96°C for ten seconds, 50°C for five seconds and 60°C for four minutes, and resolved on an ABI 3730 DNA Analyzer (Applied Biosystems, Brussels, Belgium).

### Data analysis

Identification of SNPs was carried out using the novoSNP v2.1.9 software package.<sup>24</sup> Reference sequences were NC\_000010.9 for *CYP2C9*, NC\_000010.9 for *CYP2C19*, M33388 for *CYP2D6* and NC\_000008.9 for *NAT2*. All identified SNPs were compared with the NCBI Single Nucleotide Polymorphism database (dbSNP).<sup>25</sup> As SNPs can cause the introduction of pre-microRNA (miRNA) sites, this was included as part of the annotation in the novoSNP analysis procedure. Frequencies of SNPs were calculated using Genepop.<sup>26</sup>

HaploView v3.31<sup>27</sup> was used to determine haplotypes from sequence genotype data. Linkage disequilibrium plots were generated to assess the extent to which SNPs were likely to be linked and hence likely to occur on the same haplotype with logarithm of odds (LOD) score >3. The HaploView Tagger tool was used to estimate which SNPs were likely to be tagged by a single SNP in a predicted haplotype (threshold  $r^2 > 0.8$ ).

### Prediction of functional effects of non-synonymous SNPs

Functional effects of non-synonymous SNPs were predicted using the Polyphen prediction programme<sup>28</sup> based on position-specific independent counts (PSIC) scores of multiple sequence alignments, as well as structural information, if available. The programme predicts the functional effects of SNPs based on occurrence in active/binding sites or in transmembrane regions, interference with disulphide or other bonds, compatibility with homologous sequences at that position, as well as mapping to known three-dimensional protein structures or validated homology models. Protein sequence accession numbers were obtained from Swiss-Prot<sup>29</sup> as *CYP2C9*: P11712; *CYP2C19*: P33261; *CYP2D6*: P10635; *NAT2*: P11245.

### Allele nomenclature

Allele nomenclature is assigned according to the Human Cytochrome P450 (CYP) Allele Nomenclature Committee<sup>30</sup> and the Arylamine N-acetyltransferase Gene Nomenclature Committee.<sup>31</sup>

Alleles, as borne by specific SNPs, are assigned numbers — for example, *CYP2C9*\*9 to define the 10535A > G mutation, the presence of which results in the amino acid change H251R.

## Results

In our African populations, novel and known SNPs were found in all drug-metabolising enzyme genes studied (Tables S1–S4). Novel SNPs for *CYP2C9*, *CYP2C19* and *CYP2D6* were grouped and assigned to haplotypes or groups of other known mutations if possible (Table 1). We mainly looked at the non-synonymous SNPs, since these are used to determine the eventual assignment of new functional alleles. *CYP2C9* 42519T > C (I327T) and 50341G > T (V490F) were assigned the new allele names *CYP2C9*\*31 and *CYP2C9*\*32, respectively. For *CYP2C19*, the 17869G > C/80161A > G (R186P/I331V) combination was assigned the new allele name *CYP2C19*\*22. It was not possible to assign new haplotypes/alleles for *CYP2C9* 50294A > G (N474S) and *CYP2C19* 12690G > A (V113I) because their linkage with other alleles such as *CYP2C9*\*9 (10535A > G; H251R) and *CYP2C19*\*2 (19154G > A; P227P), respectively, could not be excluded. Known mutations also had some synonymous SNPs, as well as non-coding SNPs, grouped to them (eg *CYP2C9*\*9, *CYP2C19*\*12 and *CYP2C19*\*13). It appears that the novel *CYP2D6*1608 G > A (V119M) SNP is found on the known *CYP2D6*\*29 allele, which is defined by 1659G > A (V136M) and 3183G > A (V338M). This haplotype group was therefore assigned the new name *CYP2D6*\*70. New alleles were not assigned for *CYP2D6* 1621G > T (R123L) and 4057G > A (G445E), since further work is required fully to establish the haplotypes. The novel *NAT2* SNPs did not appear to be linked to any other SNPs. HaploView-determined tag SNPs for *NAT2* were used to determine the major haplotype frequencies (Figure 1).

In *CYP2C9* (Table S1), three out of six non-synonymous SNPs — 42519T > C (I327T), 50294A > G (N474S) and 50341G > T (V490F) — were novel. Of these, I327T and V490F changes are

predicted to have a functional effect (Table 1); however, further inference of these amino acid changes with crystal structure information<sup>32</sup> and Gotoh's sequence alignments<sup>33</sup> indicates that they may not influence substrate recognition and binding. The most common non-synonymous *CYP2C9* allele in this study was *CYP2C9*\*9 (10535A > G; H251R), which is predicted to be damaging to enzyme function, although phenotypic studies in African individuals have shown no effect on the metabolism of the antiepileptic drug phenytoin. By contrast, the other known non-synonymous SNPs, such as *CYP2C9*\*5 (42619C > G; D360E) and *CYP2C9*\*6 (10601delA; K273fs) (Table S1), did cause reduced enzyme activity.<sup>17</sup>

The two novel non-synonymous SNPs discovered in *CYP2C19* (Table S1), 12690G > A (V113I) in exon 3 and 17869G > C (R186P) in exon 4, seem to cause very different effects on enzyme function, according to the physicochemical character of their amino acid changes (Table 1). Whereas the effect of V113I may be negligible, the change from the basic arginine to proline at position 186 seems to be functionally damaging, as predicted (PSIC score = 3.159).

Three novel non-synonymous SNPs were found in *CYP2D6*: 1608G > A (V119M), 1621G > T (R123L) and 4057G > A (G445E) (Table S3). Whereas the V119M and the R123L changes were predicted to have no effect on enzyme function (Table 1), they are located in the substrate recognition site SRS1. The G445E substitution may be functionally important (PSIC score = 3.063) owing to its close proximity to the 443 site, which is critical for the heme ligand binding in this enzyme according to the crystal structure.<sup>34</sup> Consistent with other African data,<sup>35,36</sup> the most common *CYP2D6* haplotypes contributing to the variability of drug response were *CYP2D6*\*2 (2850C > T; R296C and 4180G > C; S486T), *CYP2D6*\*17 (1023C > T; T107I) and *CYP2D6*\*29 (1659G > A; V136M and 3183G > A; V338M) (Table S3).

Four novel amino acid-changing SNPs were detected in *NAT2* (Table S4). The 641C > T (T214I) was predicted to have an effect on enzyme function (Table 1) because the amino acid at this

**Table I.** Grouping of novel SNPs and functional effect prediction

Gene	SNP grouping <sup>†</sup>	cDNA position	Amino acid change	Functional effect prediction (PSIC score)
<b>CYP2C9</b>	<b>42519T &gt; C (*31)</b>	<b>980</b>	<b>I327T</b>	<b>Possible functional damage (2.761)</b>
	<b>50294A &gt; G</b>	<b>1421</b>	<b>N474S</b>	<b>No functional damage (0.162)</b>
	47545A > T			
	50298A > T			
	<b>50341G &gt; T (*32)</b>	<b>1468</b>	<b>V490F</b>	<b>Possible functional damage (1.806)</b>
	10535A > G (*9)	752	H251R	Possible functional damage (2.239)
	50196C > T	1323	A441A	
<b>CYP2C19</b>	12122G > A			
	<b>12690G &gt; A</b>	<b>337</b>	<b>V113I</b>	<b>No functional damage (0.198)</b>
	<b>57453G &gt; C</b>			
	<b>90533C &gt; T</b>			
	<b>57575T &gt; C</b>			
	87290C > T (*13)	1228	R410C	
	90209A > C (*12)	1473	X491C	26 extra amino acids
	<b>90302C &gt; T</b>			
<b>CYP2D6</b>	<b>17869G &gt; C (*22)</b>	<b>557</b>	<b>R186P</b>	<b>Possible functional damage (3.159)</b>
	80161G > A	991	I331V	
	-175G > A			
	310G > T			
	843T > G			
	<b>1608G &gt; A (*70)</b>		<b>V119M</b>	<b>No functional damage (0.054)</b>
	1659G > A		V136M	Functional damage – reduced enzyme activity
	1661G > C			
	3183G > A		V338M	Functional damage – reduced enzyme activity
	3384A > C			
	4180G > C		S486T	No functional damage (0.267)
	4722T > G			
	214G > C			
223C > G				
227T > C				
843T > G				
<b>1621G &gt; T</b>		<b>R123L</b>	<b>No functional damage (1.236)</b>	

Continued

Table 1. Continued

Gene	SNP grouping <sup>†</sup>	cDNA position	Amino acid change	Functional effect prediction (PSIC score)
	1661G > C			
	2850C > T		R296C	No functional damage (0.254)
	3384A > C			
	3584G > A			
	3790C > T			
	4180G > C		S486T	No functional damage (0.267)
	843T > G			
	1661G > C			
	2850C > T		R296C	No functional damage (0.254)
	3384A > C			
	<b>4057G &gt; A</b>		<b>G445E</b>	<b>Possible functional damage, contact with functional site (3.063)</b>
	4180G > C		S486T	No functional damage (0.267)
<b>NAT2</b>	<b>10542A &gt; C</b>	<b>472</b>	<b>I158L</b>	<b>No functional damage (0.615)</b>
	<b>10659C &gt; T</b>	<b>589</b>	<b>R197X</b>	<b>No protein expressed</b>
	<b>10711C &gt; T</b>	<b>641</b>	<b>T214I</b>	<b>Possible functional damage, involved in ligand binding (1.257)</b>
	<b>10879T &gt; C</b>	<b>809</b>	<b>I270T</b>	<b>No functional damage (0.526)</b>

Abbreviations: PSIC, position-specific independent counts; SNP, single nucleotide polymorphism.

<sup>†</sup>Bold: novel non-synonymous SNPs; italic bold: novel intronic SNPs; italics: known non-synonymous SNPs; (\*) = described alleles carrying that particular mutation; SNP positions are according to reference sequences (Tables S2–S5).

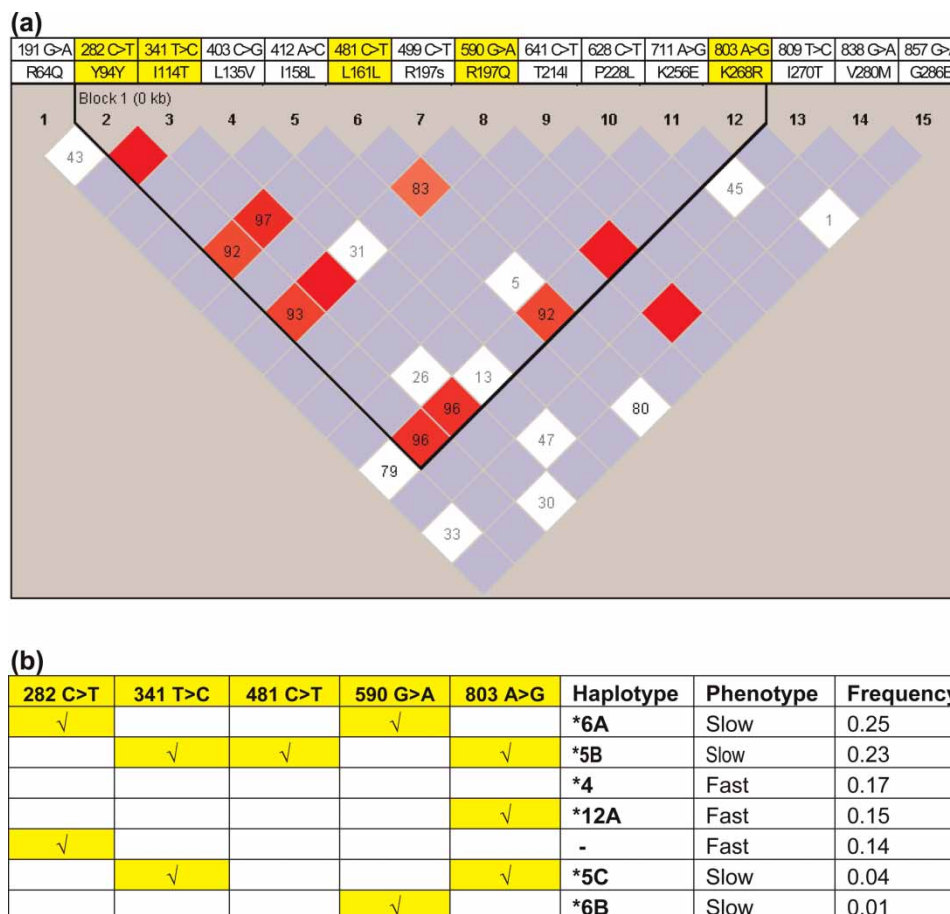
position was predicted to be involved in coenzyme A ligand binding as part of the acetylation process. The 589C > T (R197X) results in a stop codon being introduced and hence no protein is expressed. The most common alleles of *NAT2* in this study were *NAT2\*5* (341T > C; I114T) and *NAT2\*6* (590G > A; R197Q) (Table S4), which contribute largely to the slow acetylator phenotype in African populations. Figure 1 shows *NAT2* haplotypes and their frequencies in the total population studied. The most common sub-haplotypes were *NAT2\*6A* and *NAT2\*5B*, which affect enzyme function, followed by the wild type *NAT2\*4* and *NAT2\*12A*, which do not impair acetylation.

In addition to non-synonymous SNPs, numerous novel synonymous SNPs, SNPs in introns and at splice site junctions, were identified. SNPs at splice site junctions were investigated, but none of the

novel SNPs were located within the most critical -1 to -2 positions of the acceptor sites or the -2 to +4 positions of the donor sites.

## Discussion

Major genetic variability in drug-metabolising enzymes has been reported in Caucasian and Asian populations.<sup>37</sup> The aim of this study was to search for novel variants of the highly polymorphic cytochrome P450 (*CYP2C9*, *CYP2C19*, *CYP2D6*) and N-acetyltransferase 2 (*NAT2*) genes in Africans, using representative samples from our newly established Biobank.<sup>21</sup> This analysis was focused on the occurrence of alleles in African populations, their potential effects on enzyme function and the applicability of such data to personalised therapy.



**Figure 1.** N-acetyltransferase 2 haplotypes constructed from sequence and genotype data from the total population studied ( $n = 127$ ). (a) Linkage disequilibrium plot with the genomic positions indicated at the top. In yellow are the tag single nucleotide polymorphisms (SNPs) which define the major known haplotypes and are able to capture other SNPs within the same haplotype. The amino acid changes at the various positions are shown. (b) Haplotype frequencies. Haplotypes and phenotypes (acetylators) were assigned according to Consensus Arylamine N-Acetyltransferase (NAT) Gene Nomenclature.<sup>31</sup>

### African populations

Certain SNPs or haplotypes that have been reported as prevalent and functionally important in other populations are rare or have not yet been detected in African populations. For example, *CYP2C9*\*2 (R144C) and *CYP2C9*\*3 (I359L), while extensively studied in Asian and Caucasian populations and identified as rare alleles in African Americans (frequency ~1 per cent), were not found in Africans, neither in this study nor in a Beninese population.<sup>38</sup> By contrast, *CYP2C9*\*5 (D360E) and *CYP2C9*\*6 (K273fs) have been identified in African populations, although at low frequency (frequency = 0.01; Table S1). *CYP2C9*\*5 causes impaired enzyme

activity,<sup>6</sup> and *CYP2C9*\*6, first found in African Americans, is associated with phenytoin toxicity.<sup>39</sup> The importance of *CYP2C9*\*8 (R150H), *CYP2C9*\*9 (H251R) and *CYP2C9*\*11 (R335W), which were detected in limited studies in Africans<sup>17</sup> (partly including the present study), and the distribution of poor metabolisers in African populations remain unclear.

The US Food and Drug Administration (FDA) has recommended genotyping for *CYP2C9*\*2 and *CYP2C9*\*3 with the aim of better use of warfarin.<sup>40</sup> Since these variants are practically absent in populations of African origin, their use in current pharmacodiagnostic kits that identify individuals

carrying *CYP2C9\*2* and *CYP2C9\*3* may not be applicable in these populations. Test kits that detect *CYP2C9\*5*, *CYP2C9\*6*, *CYP2C9\*8*, *CYP2C9\*9* and *CYP2C9\*11*, as well as our novel SNPs, should be more predictive of the clinical response to *CYP2C9* substrate drugs in Africans. Before such tools can be developed and deployed for clinical use, however, further studies are required to establish the frequencies of these alleles in larger African populations, in addition to genotype–phenotype studies to establish their functional relevance.

*CYP2C19\*2* (splicing defect) and *CYP2C19\*3* (W212X) have been recommended as biomarkers for the administration of certain *CYP2C19* substrates.<sup>41</sup> The *CYP2C19* poor-metaboliser phenotype is detected in two to four per cent of Caucasians and in about 20 per cent of Asians, and these two variants account for 99 per cent of these poor metaboliser phenotypes.<sup>42,43</sup> Whereas *CYP2C19\*2* was the most frequent known defective variant in our study (frequency = 0.15; Table S2), we and various genotype–phenotype correlation studies have found *CYP2C19\*3* to be rare in most African populations (frequency = 0.01; Table S2<sup>44</sup>). We also identified one individual in the Maasai ethnic group who was heterozygous for this allele, and a few heterozygous individuals have previously been reported in a Tanzanian population.<sup>20</sup> Earlier data show that *CYP2C19\*2* accounts for over 70 per cent of slow metabolisers of S-mephenytoin.<sup>45</sup> The missing 30 per cent might be made up by *CYP2C19\*3* and other variants such as *CYP2C19\*12*, *CYP2C19\*13* and *CYP2C19\*15*, which would make these SNPs important contenders to include in genotyping panels for diagnostic purposes in Africans.

Our analysis of diverse African populations confirmed that *CYP2D6\*17* (T107I) and *CYP2D6\*29* (V136M, V338M) remain the most important functional SNPs in the metabolism of *CYP2D6* substrate drugs. Together with other less prevalent haplotypes, they explain why African populations generally have a larger number of intermediate metabolisers (~40 per cent) compared with Caucasian populations (~15 per cent).<sup>46</sup>

Based on the highly polymorphic *CYP2D6*, we used principal component analysis to investigate inter-ethnic variability. The fact that no significant differences were detected across ethnicities (data not shown) could be due to our small sample sizes; however, our data are consistent with a recent study illustrating that *CYP2D6* shows a high frequency of altered activity variants but no clear population structure.<sup>47</sup> It may also imply that the phenotype status of those populations is not significantly different either.

It has been speculated that the variation in acetylator (NAT2) status across major world populations reflects differences in dietary habits or the environment. There is a high prevalence of slow and intermediate acetylators in African populations, however, due to the common *NAT2\*5* (I114T), *NAT2\*6* (R197Q) and *NAT2\*14* (R64Q) alleles, which contribute largely to the slow acetylator phenotype. This is consistent with our data (Figure 1) and with a recent study of sub-Saharan populations which also indicates that the *NAT2\*5B* and *NAT2\*6A* haplotypes are more common than the wild-type haplotype *NAT2\*4*.<sup>1,48</sup>

### Enzyme function

Some mutations in coding regions cause amino acid changes that result in alterations of enzyme activity, substrate selectivity and, sometimes, protein stability. Ensuing functional differences cause different metaboliser phenotypes. So far, over 30 such variants have been reported for *CYP2C9*, approximately 20 for *CYP2C19* and over 60 for *CYP2D6*.<sup>30</sup>

We have predicted functional effects of novel non-synonymous SNPs discovered in this study (Table 1). These predictions were based on amino acid chemistry, conservation in the alignment of known sequences from the same protein families, and solved structures or homology modelling. Crystal structures of *CYP2C9* and *CYP2D6* have been reported<sup>32,34</sup> and structures of the other enzymes have been approximated by homology modelling.<sup>49,50</sup> It is assumed that such approximation is sufficiently accurate to predict functional effects in substrate recognition, binding and catalysis of reactions.<sup>51</sup>



Amino acid changes with a PSIC score of less than 1 are assumed not to be involved in any functional sites and are predicted not to affect enzyme function (eg N474S in *CYP2C9*, V113I in *CYP2C19*, V119M in *CYP2D6*, and I158L and I270T in *NAT2*). Some changes with PSIC scores slightly above 1 may still have modest effects on enzyme function — for example, R123L in *CYP2D6* (PSIC score = 1.236). As shown in a previous study, in which the *CYP2D6* sequence was aligned with Gotoh's sequence,<sup>33</sup> however, this residue is involved in the substrate recognition site SRS1.<sup>52</sup> The T214I change in *NAT2* (PSIC score = 1.257) seems to interfere with enzyme function because this residue is important for the interaction with the co-enzyme A ligand, according to homology model prediction.

The effect of the R186P change in *CYP2C19* leads to a change in electrostatic charge and possibly geometry; hence, it is predicted to affect the protein dramatically, giving a high PSIC score. The high score observed for G445E in *CYP2D6* might be due to its interaction with position 443, which is important for heme-ligand binding,<sup>34</sup> and therefore has a high probability of affecting enzyme function.

Whereas some defective splice site variants are well understood — for example, *CYP2D6*\*4 (1846G > A), which occurs at the zero acceptor position of exon 4 — functional indications are less clear if mutations lie further away from splice site junctions. Rogan *et al.*<sup>53</sup> have used information theory analysis to show how other intronic and synonymous mutations may contribute to splice site effects in *CYP* genes.<sup>53</sup> For example, the defective allele *CYP2C19*\*2 (19154G > A) results in a synonymous mutation (P227P), yet it has been associated with reduced enzyme activity. Further investigations showed that this mutation introduces a cryptic splice site 40 nucleotides downstream, resulting in a truncated protein. We used information theory to analyse novel synonymous SNPs and intronic SNPs within the splice sites (−25 to +2 for exon acceptor sites and −3 to +6 for exon donor sites) of *CYP2C9*, *CYP2C19* and *CYP2D6* but did not find any significant effects on splice site recognition (data not shown).

Pre-miRNA sequences are involved in the regulation of protein expression. Mutations in these sequences, as well as insertions of new pre-miRNA sequences, could affect enzyme expression, yet *CYP1B1* is the only *CYP* that has been found to be miRNA regulated so far.<sup>54</sup> In the present study, we did not find any pre-miRNA sequences introduced in the 3' untranslated region (UTR) regions, yet in *CYP2C19*, 18818T > C in intron 4 and 19332G > A in intron 5 introduce miRNA binding sites for has-mir-139 and has-mir-448, respectively (Table S2). Since miRNA binding sites mostly act within the 3' UTR, however, these mutations would not be expected to have any effects.

In summary, our data, in conjunction with other studies of sub-Saharan Africans and African Americans,<sup>17,19,55,56</sup> indicate low heterogeneity in the frequency of functional mutations. In the genes studied, most functionally important SNPs have been found. What remains is to determine their prevalence across populations and to evaluate the functional effects of the novel SNPs. Expressing variant proteins and analysing their substrate turnover to show impaired enzymatic activity was beyond the scope of this study. We envisage that such analyses will strengthen our findings, however, and might become essential for the pharmacokinetic assessment of individual variants in order to meet regulatory requirements for diagnostic use.

### Personalised therapy

Our data indicate the importance of *CYP2C9*, *CYP2C19*, *CYP2D6* and *NAT2* for genotype assessment, including the identified novel SNPs, so that optimisation of drug use in African populations can be considered under appropriate clinical scenarios. This could enable correct dose adjustment for individuals who are likely to experience ADRs owing to poor metabolism or an inadequate therapeutic effect owing to ultra-rapid metabolism. It is noteworthy, however, that other factors, which are not related to the newly identified SNPs but affect the clinical pharmacology of prescribed medications, may play a role in clinical ADRs or therapeutic failure.

The incorporation of *CYP2C9* genotyping as part of pre-prescription diagnosis for individuals being treated with drugs metabolised by this enzyme<sup>57</sup> indicates the immediate utility of pharmacogenetics. Likewise, pre-prescription genotyping has been recommended for *CYP2D6*-metabolised drugs with a narrow therapeutic window, such as some antipsychotic agents.<sup>58</sup> *NAT2* genotype information can be used to predict the phenotypic status of individuals to enable dose adjustment of anti-tuberculosis drugs such as isoniazid.

## Conclusions

We have started to identify and catalogue novel variants (SNPs) of genes that are important in drug metabolism. We have confirmed African-specific variants but found modest variation between different African ethnicities, indicating similar metabolic profiles for most drugs, yet stressing inter-individual variability. The low frequency of our new *CYP2C9*, *CYP2C19*, *CYP2D6* and *NAT2* alleles seems to have reduced their impact at the population level. The generally high level of diversity in gene loci of African populations, however, indicates that rare variants (incidence of less than 1 per cent) and inter-individual variability might bear extra weight in explaining Africans' phenotypic diversity. As genome-wide association studies turn up new variants at high pace, the character of molecular diagnostics shifts from single genes to profiles, encompassing low frequency variants as their main constituents.

We have predicted the functional effects of non-synonymous SNPs and suggest genotype-phenotype studies to investigate the effects of these SNPs in individuals. Eventually, we recommend the genotyping of African populations to establish the prevalence of functionally important haplotypes towards the development of relevant pharmacodiagnostic tools for these populations.

## Acknowledgments

Alice Matimba was a recipient of the Flemish International Council scholarship for her PhD studies in collaborating between the Department of Molecular Genetics, University of

Antwerp (UA) and the African Institute of Biomedical Science and Technology (AiBST). The authors acknowledge the contribution of the personnel of the VIB Genetic Service Facility (<http://www.vibgeneticservicefacility.be>) for the genetic analysis. We also thank Anastasia Guantai, Margaret Oluka, Emmanuel Chigutsa, Oluseye Bolaji, Ben Ebeshi and the members of the Consortium for the Study of Pharmacogenetics of African populations (CoPhA) for the extensive sample collection in the AiBST Biobank.

## References

- Sabbagh, A., Langaney, A., Darlu, P., Gerard, N., Krishnamoorthy, R. and Poloni, E.S. (2008), 'Worldwide distribution of *NAT2* diversity: implications for *NAT2* evolutionary history', *BMC Genet.* Vol. 9, p. 21.
- Wijnen, P.A., Op den Buijsch, R.A., Drent, M. et al. (2007), 'Review article: The prevalence and clinical relevance of cytochrome P450 polymorphisms', *Aliment. Pharmacol. Ther.* Vol. 26 (Suppl. 2), pp. 211–219.
- Gray, I.C., Nobile, C., Muresu, R., Ford, S. and Spurr, N.K. (1995), 'A 2.4-megabase physical map spanning the *CYP2C* gene cluster on chromosome 10q24', *Genomics* Vol. 28, pp. 328–332.
- Shimada, T., Yamazaki, H., Mimura, M., Inui, Y. and Guengerich, F.P. (1994), 'Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: Studies with liver microsomes of 30 Japanese and 30 Caucasians', *J. Pharmacol. Exp. Ther.* Vol. 270, pp. 414–423.
- Tsao, C.C., Wester, M.R., Ghanayem, B. et al. (2001), 'Identification of human *CYP2C19* residues that confer S-mephenytoin 4'-hydroxylation activity to *CYP2C9*', *Biochemistry* Vol. 40, pp. 1937–1944.
- Yasar, U., Eliasson, E., Dahl, M.L., Johansson, L., Ingelman-Sundberg, M. and Sjoqvist, F. (1999), 'Validation of methods for *CYP2C9* genotyping: Frequencies of mutant alleles in a Swedish population', *Biochem. Biophys. Res. Commun.* Vol. 254, pp. 628–631.
- Dickinson, G.L., Lennard, M.S., Tucker, G.T. and Rostami-Hodjegan, A. (2007), 'The use of mechanistic DM-PK-PD modelling to assess the power of pharmacogenetic studies — *CYP2C9* and warfarin as an example', *Br. J. Clin. Pharmacol.* Vol. 64, pp. 14–26.
- Higashi, M.K., Veenstra, D.L., Kondo, L.M. et al. (2002), 'Association between *CYP2C9* genetic variants and anticoagulation-related outcomes during warfarin therapy', *JAMA* Vol. 287, pp. 1690–1698.
- Miners, J.O. and Birkett, D.J. (1998), 'Cytochrome P4502C9: An enzyme of major importance in human drug metabolism', *Br. J. Clin. Pharmacol.* Vol. 45, pp. 525–538.
- Goldstein, J.A. (2001), 'Clinical relevance of genetic polymorphisms in the human *CYP2C* subfamily', *Br. J. Clin. Pharmacol.* Vol. 52, pp. 349–355.
- Sim, S.C., Risinger, C., Dahl, M.L. et al. (2006), 'A common novel *CYP2C19* gene variant causes ultrarapid drug metabolism relevant for the drug response to proton pump inhibitors and antidepressants', *Clin. Pharmacol. Ther.* Vol. 79, pp. 103–113.
- Ingelman-Sundberg, M. (2005), 'Genetic polymorphisms of cytochrome P450 2D6 (*CYP2D6*): Clinical consequences, evolutionary aspects and functional diversity', *Pharmacogenomics J.* Vol. 5, pp. 6–13.
- Aklillu, E., Persson, I., Bertilsson, L. et al. (1996), 'Frequent distribution of ultrarapid metabolizers of debrisoquine in an Ethiopian population carrying duplicated and multiduplicated functional *CYP2D6* alleles', *J. Pharmacol. Exp. Ther.* Vol. 278, pp. 441–446.
- Zanger, U.M., Raimundo, S. and Eichelbaum, M. (2004), 'Cytochrome P450 2D6: Overview and update on pharmacology, genetics, biochemistry', *Naunyn Schmiedebergs Arch. Pharmacol.* Vol. 369, pp. 23–37.
- Donald, P.R., Siregel, F.A., Venter, A. et al. (2004), 'The influence of human N-acetyltransferase genotype on the early bactericidal activity of isoniazid', *Clin. Infect. Dis.* Vol. 39, pp. 1425–1430.

16. Donald, P.R., Parkin, D.P., Seifart, H.I. *et al.* (2007), 'The influence of dose and N-acetyltransferase-2 (NAT2) genotype and phenotype on the pharmacokinetics and pharmacodynamics of isoniazid', *Eur. J. Clin. Pharmacol.* Vol. 63, pp. 633–639.
17. Allabi, A.C., Gala, J.L., Horsmans, Y. *et al.* (2004), 'Functional impact of CYP2C95, CYP2C96, CYP2C98, and CYP2C911 in vivo among black Africans', *Clin. Pharmacol. Ther.* Vol. 76, pp. 113–118.
18. Bathum, L., Skjelbo, E., Mutabingwa, T.K., Madsen, H., Horder, M. and Brosen, K. (1999), 'Phenotypes and genotypes for CYP2D6 and CYP2C19 in a black Tanzanian population', *Br. J. Clin. Pharmacol.* Vol. 48, pp. 395–401.
19. Gaedigk, A., Bhatena, A., Ndjountche, L. *et al.* (2005), 'Identification and characterization of novel sequence variations in the cytochrome P4502D6 (CYP2D6) gene in African Americans', *Pharmacogenomics J.* Vol. 5, pp. 173–182.
20. Herrlin, K., Massele, A.Y., Jande, M. *et al.* (1998), 'Bantu Tanzanians have a decreased capacity to metabolize omeprazole and mephenytoin in relation to their CYP2C19 genotype', *Clin. Pharmacol. Ther.* Vol. 64, pp. 391–401.
21. Matimba, A., Oluka, M.N., Ebeshi, B.U. *et al.* (2008), 'Establishment of a biobank and pharmacogenetics database of African populations', *Eur. J. Hum. Genet.* Vol. 7, pp. 780–783.
22. Weckx, S., De Rijk, P., Van Broeckhoven, C. and Del Favero, J. (2004), 'SNPbox: Web-based high-throughput primer design from gene to genome', *Nucleic Acids Res.* Vol. 32, pp. W170–W172.
23. Rozen, S. and Skaletsky, H. (2000), 'Primer3 on the WWW for general users and for biologist programmers', *Methods Mol. Biol.* Vol. 132, pp. 365–386.
24. Weckx, S., Del Favero, J., Rademakers, R. *et al.* (2005), 'novoSNP, a novel computational tool for sequence variation discovery', *Genome Res.* Vol. 15, pp. 436–442.
25. NCBI Single Nucleotide Polymorphism Database (dbSNP). <http://www.ncbi.nlm.nih.gov/SNP>, The National Library of Medicine (NLM) [cited April 14 2008].
26. Raymond, M. and Rousset, F. (1995), 'GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism', <http://genepop.curtin.edu.au>, Laboratoire de Genetique et Environnement [cited August 21 2007].
27. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005), 'Haploview: Analysis and visualization of LD and haplotype maps', *Bioinformatics* Vol. 21, pp. 263–265.
28. Ramensky, V., Bork, P. and Sunyaev, S. (2002), 'Human non-synonymous SNPs: Server and survey', <http://genetic.bwh.harvard.edu/pph>, European Molecular Biology Laboratory [cited September 1 2008].
29. The Uniprot Consortium (2008), 'The universal protein resource (UniProt)', <http://www.expasy.uniprot.org>, The Uniprot Consortium [cited November 27 2007].
30. Home Page of the Human Cytochrome P450 (CYP) Allele Nomenclature Committee, <http://www.cypalleles.ki.se>, Karolinska Institute [cited February 21 2001].
31. Consensus Arylamine N-Acetyltransferase (NAT) Gene Nomenclature, <http://louisville.edu/medschool/pharmacology/NAT.html>, University of Louisville School of Medicine [cited October 14 2007].
32. Williams, P.A., Cosme, J., Ward, A., Angove, H.C., Matak, V.D. and Jhoti, H. (2003), 'Crystal structure of human cytochrome P450 2C9 with bound warfarin', *Nature* Vol. 424, pp. 464–468.
33. Gotoh, O. (1992), 'Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences', *J. Biol. Chem.* Vol. 267, pp. 83–90.
34. Rowland, P., Blaney, F.E., Smyth, M.G. *et al.* (2006), 'Crystal structure of human cytochrome P450 2D6', *J. Biol. Chem.* Vol. 281, pp. 7614–7622.
35. Masimirembwa, C., Persson, I., Bertilsson, L., Hasler, J. and 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.* Vol. 42, pp. 713–719.
36. Wennerholm, A., Johansson, I., Hidestrand, M., Bertilsson, L., Gustafsson, L.L. and Ingelman-Sundberg, M. (2001), 'Characterization of the CYP2D6\*29 allele commonly present in a black Tanzanian population causing reduced catalytic activity', *Pharmacogenetics* Vol. 11, pp. 417–427.
37. Mizutani, T. (2003), 'PM frequencies of major CYPs in Asians and Caucasians', *Drug Metab. Rev.* Vol. 35, pp. 99–106.
38. Allabi, A.C., Gala, J.L., Desager, J.P., Heusterspreute, M. and Horsmans, Y. (2003), 'Genetic polymorphisms of CYP2C9 and CYP2C19 in the Beninese and Belgian populations', *Br. J. Clin. Pharmacol.* Vol. 56, pp. 653–657.
39. Kidd, R.S., Curry, T.B., Gallagher, S., Edeki, T., Blaisdell, J. and Goldstein, J.A. (2001), 'Identification of a null allele of CYP2C9 in an African-American exhibiting toxicity to phenytoin', *Pharmacogenetics* Vol. 11, pp. 803–808.
40. Thompson, C.A. (2007), 'FDA encourages genetics-aided warfarin dosing', *Am. J. Health Syst. Pharm.* Vol. 64, pp. 1994–1996.
41. Furuta, T., Shirai, N., Kodaira, M.S. *et al.* (2007), 'Pharmacogenomics-based tailored versus standard therapeutic regimen for eradication of *H. pylori*', *Clin. Pharmacol. Ther.* Vol. 81, pp. 521–528.
42. Goldstein, J.A., Ishizaki, T., Chiba, K. *et al.* (1997), 'Frequencies of the defective CYP2C19 alleles responsible for the mephenytoin poor metabolizer phenotype in various Oriental, Caucasian, Saudi Arabian and American black populations', *Pharmacogenetics* Vol. 7, pp. 59–64.
43. Ibeanu, G.C., Goldstein, J.A., Meyer, U. *et al.* (1998), 'Identification of new human CYP2C19 alleles (CYP2C19\*6 and CYP2C19\*2B) in a Caucasian poor metabolizer of mephenytoin', *J. Pharmacol. Exp. Ther.* Vol. 286, pp. 1490–1495.
44. Xie, H.G., Kim, R.B., Stein, C.M., Wilkinson, G.R. and Wood, A.J. (1999), 'Genetic polymorphism of (S)-mephenytoin 4'-hydroxylation in populations of African descent', *Br. J. Clin. Pharmacol.* Vol. 48, pp. 402–408.
45. Masimirembwa, C., Bertilsson, L., Johansson, I., Hasler, J.A. and Ingelman-Sundberg, M. (1995), 'Phenotyping and genotyping of S-mephenytoin hydroxylase (cytochrome P450 2C19) in a Shona population of Zimbabwe', *Clin. Pharmacol. Ther.* Vol. 57, pp. 656–661.
46. Gaedigk, A., Simon, S.D., Pearce, R.E., Bradford, L.D., Kennedy, M.J. and Leeder, J.S. (2008), 'The CYP2D6 activity score: Translating genotype information into a qualitative measure of phenotype', *Clin. Pharmacol. Ther.* Vol. 83, pp. 234–242.
47. Sistonen, J., Sajantila, A., Lao, O., Corander, J., Barbujani, G. and Fuselli, S. (2007), 'CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure', *Pharmacogenet. Genomics* Vol. 17, pp. 93–101.
48. Sabbagh, A. and Darlu, P. (2006), 'SNP selection at the NAT2 locus for an accurate prediction of the acetylation phenotype', *Genet. Med.* Vol. 8, pp. 76–85.
49. Walraven, J.M., Trent, J.O. and Hein, D.W. (2007), 'Computational and experimental analyses of mammalian arylamine N-acetyltransferase structure and function', *Drug Metab. Dispos.* Vol. 35, pp. 1001–1007.
50. Wang, J.F., Wei, D.Q., Li, L., Zheng, S.Y., Li, Y.X. and Chou, K.C. (2007), '3D structure modeling of cytochrome P450 2C19 and its implication for personalized drug design', *Biochem. Biophys. Res. Commun.* Vol. 355, pp. 513–519.
51. de Graaf, C., Oostenbrink, C., Keizers, P.H. *et al.* (2007), 'Molecular modeling-guided site-directed mutagenesis of cytochrome P450 2D6', *Curr. Drug Metab.* Vol. 8, pp. 59–77.
52. Bapiro, T.E., Hasler, J.A., Ridderstrom, M. and Masimirembwa, C.M. (2002), 'The molecular and enzyme kinetic basis for the diminished activity of the cytochrome P450 2D6.17 (CYP2D6.17) variant. Potential implications for CYP2D6 phenotyping studies and the clinical use of CYP2D6 substrate drugs in some African populations', *Biochem. Pharmacol.* Vol. 64, pp. 1387–1398.
53. Rogan, P.K., Svojanovsky, S. and Leeder, J.S. (2003), 'Information theory-based analysis of CYP2C19, CYP2D6 and CYP3A5 splicing mutations', *Pharmacogenetics* Vol. 13, pp. 207–218.
54. Tsuchiya, Y., Nakajima, M., Takagi, S., Taniya, T. and Yokoi, T. (2006), 'MicroRNA regulates the expression of human cytochrome P450 1B1', *Cancer Res.* Vol. 66, pp. 9090–9098.

55. Blaisdell, J., Mohrenweiser, H., Jackson, J. *et al.* (2002), 'Identification and functional characterization of new potentially defective alleles of human CYP2C19', *Pharmacogenetics* Vol. 12, pp. 703–711.
56. Patin, E., Harmant, C., Kidd, K.K. *et al.* (2006), 'Sub-Saharan African coding sequence variation and haplotype diversity at the NAT2 gene', *Hum. Mutat.* Vol. 27, p. 720.
57. Gage, B.F and Lesko, L.J. (2008), 'Pharmacogenetics of warfarin: Regulatory, scientific, and clinical issues', *J. Thromb. Thrombolysis* Vol. 25, pp. 45–51.
58. Kirchheiner, J. (2008), 'CYP2D6 phenotype prediction from genotype: Which system is the best?', *Clin. Pharmacol. Ther.* Vol. 83, pp. 225–227.

## Appendix: Supplementary Tables

Table S1. CYP2C9 single nucleotide polymorphism (SNP) frequencies

NC_000010.9	mRNA position	SNP	mRNA feature	Effect	dbSNP	Hausa (13)	Luo (12)	Maasai (11)	San (13)	Shona (23)	Venda (9)	TZ Bantu (12)	Total (93)
96829291	-375	T > C	5' UTR		rs9332103	0.04	nd	0	0	0	nd	nd	0.01
96829916	251	T > C	Intron		rs9332104	0.08	0.10	0.27	0.12	0.22	0.17	0.11	0.16
96833076	3411	T > C	Intron		rs9332120	0	0.10	0	0.04	0.13	0.17	0.14	0.08
96833152	3487	A > G	Intron	splice site	rs12769205	0	0.04	0	0.17	0.11	0.06	0.09	0.07
96833165	3499	T > A	Intron	splice site	rs9332121	0	0	0	0	0.03	0	0	0.01
96838628	8963	T > C	Intron		nrs	0	0.05	0	0	0	0	0	0.01
96838697	9032	G > C	Intron		nrs	0	0.10	0.14	0.15	0.15	0.13	0.14	0.12
96838734	9069	G > A	Intron		Novel	0	0.05	0.05	0	0	0	0.05	0.02
96839116	9451	T > C	Intron		rs17443251	0	0.05	0	0	0	0	0.06	0.01
96839976	10311	A > G	Intron		rs9332129	0	0.17	0.15	0.15	0.14	0.13	0.19	0.13
96840012	10347	T > C	Intron		Novel	0	0	0	0	0.03	0	0	0.01
96840200	10535	A > G	Exon 5	H251R (*9)	rs2256871	0.18	0.17	0.05	0.15	0.11	0.06	0	0.11
96840266	10601	wt > delA	Exon 5	K273 fs (*6)	nrs	0.04	0	0	0	0	0.07	0	0.01
96863014	33349	A > G	Intron		rs9332172	0.17	0.23	0.23	0.15	0.28	0.25	0.50	0.26
96863323	33658	A > G	Intron		rs9332174	0.13	0.14	0.27	0.12	0.20	0.19	0.10	0.17
96872080	42415	C > T	Intron		Novel	0	0	0.06	0	0	0	0	0.01
96872134	42469	T > C	Intron		rs9332197	0	0	0.05	0	0	0	0	0.01
96872184	42519	T > C	Exon 7	I327T (*31)	Novel	0.04	0	0	0	0	0	0	0.01
96872284	42619	G > C	Exon 7	D360E (*5)	rs28371686	0	0	0	0	0.02	0.06	0	0.01
96877210	47545	A > T	Intron		rs9332230	0	0	0	0	0	0	0.05	0.01
96877258	47593	T > C	Intron		rs9332232	0.07	0.14	0.05	0.04	0.18	0.17	0.05	0.11
96877304	47639	C > T	Intron		rs2298037	0	0	0	0	0.03	0.06	0	0.01
96879721	50056	A > T	Intron		rs1934969	0.67	0.50	0.50	0.40	0.24	0.13	0.32	0.38

Continued

Table S1. Continued

NC_000010.9	mRNA position	SNP	mRNA feature	Effect	dbSNP	Hausa (13)	Luo (12)	Maasai (11)	San (13)	Shona (23)	Venda (9)	TZ Bantu (12)	Total (93)
96879790	50125	C > T	Intron		Novel	0	0	0	0	0	0	0.05	0.01
96879861	50196	C > T	Exon 9	A441A	rs2017319	0.04	0.14	0.05	0.04	0.20	0.17	0.05	0.10
96879959	50294	A > G	Exon 9	N474S	Novel	0	0.05	0	0	0	0	0.00	0.01
96879963	50298	A > T	Exon 9	G475G	rs1057911	0	0	0	0	0	0	0.05	0.01
96880006	50341	G > T	Exon 9	V490F (*32)	Novel	0	0	0	0	0	0	0.05	0.01
96880078	50413	C > T	3' UTR		rs9332240	0	0.05	0	0	0.03	0	0.00	0.01
96880099	50434	C > T	3' UTR		rs9332241	0.08	0.05	0	0	0.02	0.13	0.14	0.05
96880166	50501	C > T	3' UTR		rs9332243	0	0.05	0	0	0.03	0.00	0.00	0.01

mRNA position = relative to A of ATG start codon; wt = wild type; del = deletion; UTR = untranslated region; fs = frameshift; (\*) = described alleles carrying that particular mutation; nrs = rs number not yet assigned; nd = not determined. Number of individual samples studied per population is indicated in bold in parenthesis.

**Table S2.** *CYP2C19* single nucleotide polymorphism (SNP) frequencies

NC_000010.9	mRNA position	SNP	mRNA feature	Effect	dbSNP	Hausa (20)	Yoruba (20)	Ibo (20)	Luo (30)	Maasai (13)	Shona (15)	Venda (9)	TZ Bantu (10)	Total (137)
96653591	-97	T > C	5' UTR		rs4986894	0.13	0.18	0.33	0.07	0.08	0.2	0.11	0.15	0.15
96653743	55	A > C	Exon 1	I19L (*15)	rs17882687	0	0	0	0.05	0	0	0	0.05	0.02
96653787	99	T > C	Exon 1	P33P	rs17885098	0.05	0.08	0.15	0.18	0.15	0.17	0.17	0.30	0.15
96653871	183	T > C	Intron		rs17882201	0	0	0	0	0	0	0.06	0	<0.01
96653876	188	G > A	Intron		rs17881883	0	0	0.03	0	0	0.07	0.11	0.05	0.02
96653919	231	A > C	Intron		Novel	0	0	0	0.01	0	0	0	0	<0.01
96665810	12122	A > G	Intron		rs7916649	0.5	0.37	0.29	0.58	0.50	0.25	0.28	0.33	0.41
96665994	12306	G > A	Intron		rs17878649	0	0	0.03	0.08	0.04	0.10	0.06	0.10	0.05
96666148	12460	G > C	Exon 2	E92D	rs17878459	0	0	0	0	0	0.03	0	0	<0.01
96666295	12607	wt > insC	Intron		Novel	0	0	0	0.03	0.04	0	0	0	0.01
96666325	12637	C > T	Intron		Novel	0.03	0.1	0	0	0.04	0	0.06	0	0.03
96666350	12662	A > G	Intron	Splice site	rs12769205	0.16	0.2	0.33	0.09	0.12	0.27	0.22	0.2	0.18
96666378	12690	G > A	Exon 3	V113I	Novel	0	0	0	0	0	0	0.06	0	<0.01
96666472	12784	G > A	Exon 3	R144H (*9)	rs17884712	0	0	0	0	0	0	0.06	0	<0.01
96671557	17869	G > T	Exon 4	R186P (*22)	Novel	0	0	0	0	0	0	0	0.06	<0.01
96671636	17948	G > A	Exon 4	W212X (*3)	rs4986893	0	0	0	0	0.04	0	0	0	<0.01
96671895	18207	G > A	Intron		Novel	0	0	0	0.02	0	0	0	0.06	<0.01
96671917	18229	T > A	Intron		rs17884938	0.06	0.03	0.05	0.07	0.00	0.10	0	0	0.05
96671942	18254	T > C	Intron		Novel	0.03	0.03	0	0	0	0	0	0	0.01
96672506	18818	T > C	Intron	Pre-miRNA (has-mir-139)	Novel	0	0.03	0	0	0	0	0	0	<0.01

Continued

Table S2. Continued

NC_000010.9	mRNA position	SNP	mRNA feature	Effect	dbSNP	Hausa (20)	Yoruba (20)	Ibo (20)	Luo (30)	Maasai (13)	Shona (15)	Venda (9)	TZ Bantu (10)	Total (137)
96672599	18911	A > G	Intron		rs7088784	0.05	0.08	0.15	0.14	0.15	0.17	0.17	0.25	0.13
96672764	19076	T > C	Intron	Splice site	Novel	0	0	0	0	0	0.03	0	0	<0.01
96672842	19154	G > A	Exon 5	P227P (*2)	rs4244285	0.13	0.15	0.33	0.07	0.08	0.23	0.17	0.15	0.15
96673020	19332	G > A	Intron	Pre-miRNA (has-mir-448)	Novel	0	0	0	0	0.08	0	0	0	<0.01
96711141	57453	G > C	Intron		Novel	0	0	0.04	0.03	0	0.03	0	0	0.02
96711200	57512	A > G	Intron		Novel	0.03	0.05	0.08	0.03	0.04	0	0	0.05	0.04
96711255	57567	A > T	Intron		Novel	0	0	0.04	0.03	0	0.07	0.11	0.10	0.04
96711263	57575	T > C	Intron		Novel	0	0	0.04	0.03	0	0.03	0	0	0.02
96711325	57637	wt > delG	Intron		Novel	0.03	0.05	0.08	0.07	0.08	0.10	0.00	0.10	0.06
96711366	57678	T > G	Intron		rs28399511	0	0	0	0	0.04	0	0	0	<0.01
96711428	57740	G > C	Intron		rs4417205	0.14	0.15	0.21	0.09	0.13	0.23	0.22	0.20	0.15
96711677	57989	G > C	Intron		Novel	0	0	0	0.02	0	0.1	0	0	0.02
96733848	80160	C > T	Exon 7	V330V	rs3758580	0.12	0.18	0.25	0.07	0.08	0.17	0.17	0.15	0.13
96733849	80161	G > A	Exon 7	V331I	rs3758581	0.03	0.03	0	0	0.04	0	0	0	0.01
96734317	80629	T > A	Intron		Novel	0.03	0.05	0	0	0.05	0	0	0	0.01
96740794	87106	T > C	Intron		rs4917623	0.38	0.13	0.21	0.31	0.23	0.03	0.06	0.15	0.22
96740978	87290	T > C	Exon 8	R410C (*13)	rs17879685	0	0	0.04	0.03	0	0.03	0	0	0.02
96741001	87313	A > C	Exon 8	G417G	rs17886522	0.03	0.05	0.08	0.07	0.04	0.10	0	0	0.06
96741110	87422	A > G	Intron		Novel	0.03	0	0	0	0.08	0	0	0	0.02
96741163	87475	G > C	Intron		rs17880188	0.08	0.13	0	0.07	0.04	0.11	0.06	0	0.07

Continued



**Table S2.** Continued

NC_000010.9	mRNA position	SNP	mRNA feature	Effect	dbSNP	Hausa (20)	Yoruba (20)	Ibo (20)	Luo (30)	Maasai (13)	Shona (15)	Venda (9)	TZ Bantu (10)	Total (137)
96741210	87522	C > T	Intron		rs17885567	0.05	0.08	0.21	0.10	0.08	0.13	0.11	0.10	0.1
96743266	89578	T > A	Intron		rs12779363	0	0	0.06	0.04	0	0.03	0	0	0.01
96743597	89909	C > T	Intron		rs12268020	0.22	0.31	0.17	0.09	0.17	0.20	0.17	0.20	0.19
96741001	90011	A > G	Intron	Splice site	rs4451645	0.13	0.14	0	0.07	0.04	0.13	0.11	0	0.08
96743897	90209	A > C	Exon 9	X491C; 26 extra aa (*12)	nrs	0	0	0	0.04	0	0.03	0	0	0.01
96743989	90301	C > T	3' UTR		Novel	0.01	0	0	0	0	0	0	0	<0.01
96743990	90302	C > T	3' UTR		Novel	0	0	0	0.04	0	0.03	0	0	0.01
96744221	90533	C > T	3' UTR		Novel	0	0	0.05	0.04	0	0.04	0	0	0.02

mRNA position = relative to A of ATG start codon; wt = wild type; del = deletion; ins = insertion; UTR = untranslated region; pre-miRNA = introduction of a pre-miRNA sequence; X = stop codon; aa = amino acid; (\*) = described alleles carrying that particular mutation; nrs = rs number not yet assigned; nd = not determined. Number of individual samples studied per population is indicated in bold in parenthesis.

**Table S3.** CYP2D6 single nucleotide polymorphism (SNP) frequencies

M33388	mRNA position	SNP	mRNA feature	Effect	dbSNP	Hausa (20)	Yoruba (20)	Ibo (20)	Luo (29)	Maasai (13)	Shona (15)	Venda (9)	TZ Bantu (10)	Total (136)
1444	-175	G > A	5' UTR		rs1080993	0.05	0.12	0.31	0.22	0.05	0.11	0.06	0.30	0.17
1469	-150	C > T	5' UTR		nrs	0.09	0.03	0	0	0	0	0	0	0.01
1534	-85	T > C	5' UTR		nrs	0.04	0	0	0	0	0.03	0	0	0.01
1577	-42	wt > insG	5' UTR		rs28371695	0.19	0.35	0.13	0.21	0.05	0.20	0.13	0.10	0.19
1696	77	G > A	Exon 1	R26H (*43)	rs28371696	0.04	0	0.03	0	0	0.03	0	0.05	0.02
1701	82	C > T	Exon 1	R28C (*22)	nrs	0	0	0	0	0	0	0	0.05	<0.01
1719	100	C > T	Exon 1	P34S (*10)	rs1065852	0.15	0.12	0.10	0.09	0.05	0	0.19	0.10	0.10
1833	214	G > C	Intron		rs1080995	0.50	0.41	0.27	0.36	0.45	0.57	0.17	0.43	0.38
1840	221	C > A	Intron		rs1080996	0.50	0.41	0.27	0.38	0.45	0.64	0.30	0.43	0.40
1842	223	C > G	Intron		rs1080997	0.50	0.41	0.27	0.36	0.45	0.56	0.20	0.38	0.38
1846	227	T > C	Intron		rs1080998	0.50	0.41	0.27	0.36	0.45	0.50	0.25	0.50	0.38
1851	232	G > C	Intron		rs1080999	0.50	0.41	0.30	0.36	0.50	0.70	0.25	0.75	0.42
1852	233	A > C	Intron		rs1080999	0.50	0.41	0.27	0.38	0.45	0.67	0.25	0.50	0.40
1864	245	A > G	Intron		rs1081000	0.50	0.41	0.27	0.31	0.45	0.70	0.25	0.67	0.39
1929	310	G > T	Intron		rs28371699	0	0.27	0.07	0.25	0.31	0.25	0	nd	0.18
2273	654	C > T	Intron		Novel	nd	0	0	0.07	nd	0.08	0.07	0.07	0.07
2365	746	C > G	Intron		nrs	nd	nd	nd	0.36	nd	0.40	0.31	0.50	0.40
2462	843	T > G	Intron		rs28371702	0.14	0.40	0.20	0.33	0.38	0.33	0.13	0.20	0.29
2625	1006	C > T	Exon 2	R101R	Novel	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.01
2642	1023	C > T	Exon 2	T107I (*17)	rs28371706	0.20	0.20	0.22	0.22	0.17	0.20	0.19	0.15	0.20
2658	1039	C > T	Exon 2	F112F	rs1081003	0.00	0.13	0.13	0.04	0.00	0.00	0.13	0.05	0.06
2686	1067	T > G	Intron		Novel	0.13	0.13	0.19	0.04	0.08	0.07	0.19	0.10	0.11
3227	1608	G > A	Exon 3	V119M (*70)	Novel	0	0	0	0.02	0	0	0	0	<0.01
3240	1621	G > T	Exon 3	R123L	Novel	0	0	0	0	0	0	0	0.05	<0.01
3278	1659	G > A	Exon 3	V136M (*29)	rs1058164	0.11	0.10	0.28	0.24	0.04	0.17	0.06	0.25	0.17
3280	1661	G > C	Exon 3	V136V	rs28371708	0.29	0.43	0.35	0.32	0.46	0.37	0.33	0.30	0.35
3335	1716	G > A	Exon 3	E155K (*45)	rs28371710	0.08	0	0	0.09	0.04	0	0.17	0	0.05
3465	1846	G > A	Intron	Splicing defect (*4)	nrs	0.03	0.08	0.08	0.04	0.04	0.00	0.00	0.05	0.04
3483	1863_1864	ins (TTTCGC CCC)X2	Exon 4	174_175ins(FRP)X2	nrs	0	0	0	0.04	0.08	0	0	0	0.02

Continued

**Table S3.** Continued

M33388	mRNA position	SNP	mRNA feature	Effect	dbSNP	Hausa (20)	Yoruba (20)	Ibo (20)	Luo (29)	Maasai (13)	Shona (15)	Venda (9)	TZ Bantu (10)	Total (136)
3485	1866	C > T	Exon 4	N175N	nrs	0	0	0	0.02	0.00	0	0	0	<0.01
3488	1869	T > C	Exon 4	G176G	nrs	0	0.03	0	0.00	0.00	0	0	0	<0.01
3617	1998	T > C	Exon 4	F219F	novel	nd	nd	0	0.00	nd	0.03	0	0.05	0.02
4194	2575	C > A	Exon 5	P267P	nrs	nd	nd	nd	0.05	nd	0.03	0.22	0	0.07
4221	2602	G > T	Exon 5	L276L	novel	nd	nd	nd	0.05	nd	0	0.06	0	0.02
4280	2661	G > A	Intron		nrs	nd	nd	nd	0.05	nd	0.03	0.11	0.05	0.06
4379	2760	T > A	Intron		Novel	nd	nd	nd	0.00	nd	0.10	0.06	0	0.04
4469	2850	C > T	Exon 6	R296C (*2)	nrs	nd	nd	nd	0.55	nd	0.63	0.44	0.65	0.58
4607	2988	G > A	Intron		nrs	nd	nd	nd	0.00	nd	0.03	0	0	0.01
4802	3183	G > A	Exon 7	V338M (*29)	nrs	0.13	0.10	0.29	0.20	0.04	0.17	0.06	0.13	0.16
4873	3254	T > C	Exon 7	H361H	rs2743457	0.09	0.00	0.00	0.07	0.08	0	0.13	0	0.04
4880	3259_3260	wt > insTG	Exon 7	375 fs (*42)	nrs	0	0	0	0	0	0.03	0	0	<0.01
5003	3384	A > C	Intron		nrs	0.30	0.45	0.34	0.28	0.42	0.37	0.25	0.38	0.65
5016	3397	C > A	Intron		novel	0	0	0	0	0	0	0.06	0	<0.01
5180	3561	G > C	Intron		novel	0	0	0	0.02	0	0	0.06	0.06	0.01
5201	3582	A > G	Intron		nrs	0.08	0.11	0.11	0.09	0.04	0.00	0.13	0.00	0.08
5203	3584	G > A	Intron		nrs	0.54	0.34	0.26	0.43	0.46	0.47	0.44	0.44	0.41
5326	3707	G > A	Intron		nrs	0	0	0.03	0	0	0.03	0	0	0.01
5349	3721	wt > delGT	Intron		nrs	0	0.03	0	0	0	0	0	0	<0.01
5409	3790	C > T	Intron	Splice site	nrs	0.53	0.34	0.26	0.44	0.54	0.47	0.44	0.44	0.42
5472	3853	G > A	Exon 8	E410K (*27)	nrs	0	0	0	0.06	0.04	0	0	0.06	0.02
5652	4033	C > T	Intron	Splice site	Novel	0	0	0	0.02	0	0	0.06	0.06	0.01
5676	4057	G > A	Exon 9	G445E	Novel	0	0	0	0.04	0	0	0	0	0.01
5799	4180	G > C	Exon 9	S486T	rs1135850	0.68	0.55	0.66	0.72	0.63	0.63	0.75	0.67	0.66
6013	4394	wt > delAG	3' UTR		Novel	0	0	0	0	0	0.10	0.06	0	0.02
6020	4401	C > T	3' UTR		nrs	0.09	0.10	0.11	0.07	0.04	0	0.25	0.06	0.08
6100	4481	G > A	3' UTR		nrs	0.12	0.08	0.03	0.11	0.21	0.23	0.07	0.19	0.12
6275	4656	wt > delACA	3' UTR		nrs	0.44	0.09	0.03	0.23	0.31	0.08	0.33	0.33	0.20
6341	4722	T > G	3' UTR		nrs	0.63	0.57	0.73	0.57	nd	nd	nd	0.25	0.58

mRNA position = relative to A of ATG start codon; wt = wild type; del = deletion; ins = insertion; UTR = untranslated region; s = stop codon; (\*) = described alleles carrying that particular mutation; fs = frame shift; aa = amino acid; nrs = rs number not yet assigned; nd = not determined. Number of individual samples studied per population is indicated in bold in parenthesis.

**Table S4.** N-acetyltransferase 2 single nucleotide polymorphism (SNP) frequencies

NC_000008.9	mRNA position	SNP	Effect	dbSNP	Hausa (20)	Yoruba (20)	Ibo (19)	Luo (16)	Maasai (12)	San (40)	Total (127)
8950	191	G > A	R64Q (*14)	rs1801279	0.03	0.08	0.13	0.20	0.08	0.09	0.10
9041	282	C > T	Y94Y	rs1041983	0.4	0.44	0.55	0.44	0.38	0.29	0.39
9100	341	T > C	I114T (*5)	rs1801280	0.33	0.14	0.34	0.27	0.5	0.2	0.27
9162	403	C > G	L135V	nrs	0	0.03	0	0.03	0	0	<0.01
9231	472	A > C	I158L	Novel	0	0	0	0.03	0	0	<0.01
9240	481	C > T	L161L	rs1799929	0.25	0.14	0.34	0.27	0.46	0.14	0.24
9348	589	C > T	R197X	Novel	0	0	0	0	0	0.01	<0.01
9349	590	G > A	R197Q (*6)	rs1799930	0.32	0.33	0.29	0.30	0.25	0.20	0.27
9400	641	C > T	T214I	Novel	0	0	0	0.03	0	0	<0.01
9442	683	C > T	P228L	nrs	0	0	0	0.03	0	0	<0.01
9525	766	A > G	K256E	nrs	0	0	0	0	0	0.03	0.01
9562	803	A > G	K268R	rs1208	0.37	0.39	0.40	0.44	0.54	0.43	0.42
9568	809	T > C	I270T	Novel	0	0	0	0	0	0.13	0.04
9597	838	G > A	V280M	nrs	0.06	0.03	0.05	0.03	0	0	0.02
9616	857	G > A	G286E (*7)	rs1799931	0.03	0.03	0.03	0.03	0.04	0.01	0.02

mRNA position = relative to A of ATG start codon; X = stop codon; (\*) = described alleles carrying that particular mutation; nrs = rs number not yet assigned. Number of individual samples studied per population is indicated in bold in parenthesis.

**Table S5.** Polymerase chain reaction and sequencing primers

Gene	Exon	First PCR primers	Sequencing primers
<b>CYP2C9</b>	5' UTR	<b>cyp2C9-5'FLF</b> ATCCTCAACTCAGTATGTCAGC <b>cyp2C9-5'FLR</b> ATCACCTAGGTCCACTATATGC	<b>cyp2C9-5'FLSF1</b> ATCCTCAACTCAGTATGTCAGC <b>cyp2C9-5'FLSR1</b> ACCTTTACCATTAAACCCCC <b>cyp2C9-5'FLSF2</b> CAATTCCTGCCTTCAGGA <b>cyp2C9-5'FLSR2</b> AAGGACTTTGACCCACTGAT
	1	<b>cyp2C9-1F</b> GGAATGTACAGAGTGGACAATGG <b>cyp2C9-1R</b> GATCCCACAATACCTTACCATTAC	
	2&3	<b>cyp2C9-2&amp;3F</b> GACCTGCTGAATATGTTGATGTG <b>cyp2C9-2&amp;3R</b> CCCGCTTCACATGAGCTAAC	<b>cyp2C9-2SF</b> TCTTGAACCTCCTGACCTTGT <b>cyp2C9-2SR</b> GGAGCTCTGTAAGTCTCTGT <b>cyp2C9-3SF</b> AGGAGTTTTCTGGAAGAGG <b>cyp2C9-3SR</b> GGAAAAACACTGCTCTTAACTC
	4*	<b>cyp2C9-4F</b> CAGCTAGGTTGTAATGGTCAACTC <b>cyp2C9-4R</b> GCTAATGGGCTTAGAAATCAGG	***
	5*	<b>cyp2C9-5F</b> TCATCTGGTTAGAATTGATCCTCTG <b>cyp2C9-5R</b> GCTATTAACCTACCGCCTCAACTTC	***

Continued

Table S5. Continued

Gene	Exon	First PCR primers	Sequencing primers
	6*	<b>cyp2C9-6F</b> GAGGAAATGGACCTAGAGACCTTC <b>cyp2C9-6R</b> CCCATTGTAATCACCATTAGTTTG	***
	7*	<b>cyp2C9-7F</b> GTGCATCTGTAACCATCCTCTCT <b>cyp2C9-7R</b> CAGACACTAGGACCTGTTACAAACC	***
	8*	<b>cyp2C9-8F</b> AGAAGGTTGCATCCAAGTATCC <b>cyp2C9-8R</b> GAGTTCTTGGGTACCTCACTGGT	***
	9	<b>cyp2c9-9F</b> CTCATCCATCCATTCATTCATG <b>cyp2c9-9R</b> CTCTAACACTCACCCAAAATAGC	<b>cyp2c9-9SF</b> CTCATCCATCCATTCATTCATG <b>cyp2c9-9SR</b> CGAATGTTCACTAGATCTTCAG <b>cyp2c9-9S2F</b> CTGCAGCTCTCTTTCTC <b>cyp2c9-9S2FR</b> CTCTAACACTCACCCAAAATAGC
<b>CYP2C19</b>			
	1	<b>cyp2C19-1F</b> CAATTATGACGGTGCATTGG <b>cyp2C19-1R</b> CACTTCCCTTACTGTTTACCCTCA	***
	2&3	<b>cyp2C19-2&amp;3F</b> GTTCTTGAAGCTGGGTATTTGTC <b>cyp2C19-2&amp;3R</b> AGCAAAGTTCAGGAGAACATAGG	<b>cyp2C19-2SF</b> AATTCAGAATATTTGAGCCTGTGTG <b>cyp2C19-2SR</b> GGTTTTCTCAACTCCTCCACAA <b>cyp2C19-3SF</b> GCCTGGGATCTCCCTCCTAGTTT <b>cyp2C19-2&amp;3R</b> AAGCAAAGTTCAGGAGAACATAGG
	4	<b>cyp2C19-4F</b> CAGCTAGGCTGTAATTGTTAATTG <b>cyp2C19-4R</b> GAGTAATGGAAGACTCCAAAGTGC	***
	5	<b>cyp2C19-5F</b> TTCAATTCAGAGGCTGCTTG <b>cyp2C19-5R</b> CTATGATGCTTACTGGATATTCATGC	***
	6	<b>cyp2C19-6F</b> CAGCATATAAACAGAGCCAAAGAC <b>cyp2C19-6R</b> ACACCATTAATGGGACAGATTAC	***
	7	<b>cyp2C19-7F</b> CCTAGCTTAAGGCACAGTTACACA <b>cyp2C19-7R</b> GAAAGACTCAAGGTGTCAAGATGTC	***
	8	<b>cyp2C19-8F</b> GCCTTAAGCTCATGCCTCTTATTAC <b>cyp2C19-8R</b> GGCAGAATTCAACCAACCTTACTT	***
	9	<b>cyp2C19-9F</b> TCATTGTTTAGTTGCCTATCCATC <b>cyp2C19-9R</b> CCATCTTCACCTTTGCCTTTC	***

Continued

Table S5. Continued

Gene	Exon	First PCR primers	Sequencing primers
<b>CYP2D6</b>			
	1&2	<i>cyp2D6ex1_2F</i> ACCAGGCCCTCCACCGG <i>cyp2D6ex1_2R</i> CTCTCTGCCAGCTCGG	<b>CYP2D6ex1_2F</b> ACCAGGCCCTCCACCGG <i>cyp2D6ex1SR</i> GTTTCACCCACCACCCATGTTT <i>cyp2D6ex2SF</i> CTCCACCTGCTCACTCCTGGTA <i>cyp2D6ex2SR</i> CCTCCCTAGTGCAGGTGGTTTCT
	3&4	<i>cyp2D6ex3_4F</i> ATTTCCCAGCTGGAATCC <i>cyp2D6ex3_4R</i> GAGACTCCTCGGTCTCTC	<i>cyp2D6ex3_4SF</i> GAGCATAGGGTTGGAGTGGGTG <i>cyp2D6ex3_4R</i> GAGACTCCTCGGTCTCTC
	5&6	<i>cyp2D6ex5_6F</i> GCCTGAGACTTGCCAGG <i>cyp2D6ex5_6R</i> CCGGCCCTGACACTCCTTCT	<i>cyp2D6ex5_6F</i> GCCTGAGACTTGCCAGG <i>cyp2D6ex5_6R</i> CCGGCCCTGACACTCCTTCT
	7,8,9	<i>cyp2D6ex7_9F</i> GGATCCTGTAAGCCTGACCTC <i>cyp2D6ex7_9R</i> ACTGAGCCCTGGGAGGTAGGTAG	<i>cyp2D6ex7_9F</i> GGATCCTGTAAGCCTGACCTC <i>cyp2D6ex7SR</i> GTGGTGGCATTGAGGACTAGGTG <i>cyp2D6ex8SF</i> GTCCAGAGTATAGGCAGGGCTGG <i>cyp2D6ex8SR</i> AGCACAAAGCTCATAGGGGATG <i>cyp2D6ex9SF</i> CTTCCTTCTTCACCTCCCTGC <i>cyp2D6ex9SR</i> AATATGGCCTCCAGGCTGAGT
<b>NAT2</b>	2	<b>NAT2ex2F</b> GAAGCATATTTTCAAAGAATTGG <b>NAT2ex2R</b> GCATTTTAAGGATGGCCTGT	<b>NAT2ex2F</b> GAAGCATATTTTCAAAGAATTGG <b>NAT2SF1</b> TGCCAAAGAAGAAACACCAA <b>NAT2SR2</b> ACCTCGAACAATTGAAGATTTTGA <b>NAT2ex2R</b> GCATTTTAAGGATGGCCTGT

\*\*\*same set of primers used for sequencing.