Analysis of pharmacogenetic traits in two distinct South African populations

Ogechi Ikediobi,[1*] Bradley Aouizerat,[2,3] Yuanyuan Xiao,[4] Monica Gandhi,[5] Stefan Gebhardt[6] and Louise Warnich[7]

[1]University of California, San Francisco, School of Pharmacy, Department of Clinical Pharmacy, San Francisco, CA 94143, USA
[2]University of California, San Francisco, School of Nursing, Department of Physiological Nursing, San Francisco, CA 94143, USA
[3]University of California, San Francisco Institute for Human Genetics, San Francisco, CA 94143, USA
[4]University of California, San Francisco, Center for Bioinformatics and Molecular Biostatistics, Department of Epidemiology and Biostatistics, San Francisco, CA 94143, USA
[5]University of California, San Francisco, School of Medicine, Department of Medicine, San Francisco, CA 94143, USA
[6]Stellenbosch University, Department of Obstetrics and Gynaecology, Tygerberg, South Africa
[7]Stellenbosch University, Department of Genetics, Stellenbosch, South Africa
*Correspondence to: Tel: +1 415 476 1230; E-mail: ikediobio@pharmacy.ucsf.edu

Abstract

Our knowledge of pharmacogenetic variability in diverse populations is scarce, especially in sub-Saharan Africa. To bridge this gap in knowledge, we characterised population frequencies of clinically relevant pharmacogenetic traits in two distinct South African population groups. We genotyped 211 tagging single nucleotide polymorphisms (tagSNPs) in 12 genes that influence antiretroviral drug disposition, in 176 South African individuals belonging to two distinct population groups residing in the Western Cape: the Xhosa ($n = 109$) and Cape Mixed Ancestry (CMA) ($n = 67$) groups. The minor allele frequencies (MAFs) of eight tagSNPs in six genes (those encoding the ATP binding cassette sub-family B, member 1 [ABCB1], four members of the cytochrome P450 family [CYP2A7P1, CYP2C18, CYP3A4, CYP3A5] and UDP-glucuronosyltransferase 1 [UGT1A1]) were significantly different between the Xhosa and CMA populations (Bonferroni $p < 0.05$). Twenty-seven haplotypes were inferred in four genes (CYP2C18, CYP3A4, the gene encoding solute carrier family 22 member 6 [SLC22A6] and UGT1A1) between the two South African populations. Characterising the Xhosa and CMA population frequencies of variant alleles important for drug transport and metabolism can help to establish the clinical relevance of pharmacogenetic testing in these populations.

Keywords: pharmacogenetics, South Africa, Xhosa, Cape Mixed Ancestry, HIV, antiretrovirals

Introduction

The field of pharmacogenomics aims to utilise the genetic composition of an individual to personalise therapeutic regimens and improve treatment out-comes. Most of the initial examples of the clinical utility of pharmacogenomics were elucidated for cancer treatments. Currently, however, there are more than 15 drugs used in the treatment of a variety of chronic diseases, such as cardiovascular disease, HIV/AIDS and seizures, for which the US Food and Drug Administration (FDA) recommends or requires pharmacogenomic testing to prevent drug-related toxicity or improve drug efficacy.[1] The increase in the number and breadth of drugs for which pharmacogenetic tests are recommended or required by the FDA is an indication of the impor-tant role that genetics plays in predicting treatment outcomes.

In order for pharmacogenetic testing to have the most impact in as many people possible, it is important to understand which genetic variants are

predictive of treatment outcomes in diverse populations. Most pharmacogenetics studies to date have been conducted in a limited number of population groups, most frequently in Western European and North American Caucasians. As a result of these limitations, genotype-to-phenotype correlates of drug response or toxicity for a number of drugs are clinically applicable in relatively few treated individuals. Furthermore, pharmacogenetic profiles characterised in Caucasians are often extrapolated for use and interpretation in other populations, in spite of at least two major problems with this method. First, it is clear that the population frequency of variants can differ markedly between populations, such as Caucasians. The differences in population frequencies of variant alleles has an impact on the clinical utility of pharmacogenetic testing, being more utilised in populations with a higher frequency of the variant allele than in populations in which the variant allele is rare. Secondly, ethnically specific variants exist in non-Caucasian populations which are more predictive of treatment outcomes than those identified in Caucasians. For example, although the UGT1A1*28 polymorphism is predictive of toxicity to the anticancer drug, irinotecan, in Caucasians, the UGT1A1*6 polymorphism is more predictive of irinotecan toxicity in Asians.[2] Such ethnically specific variation is not currently taken into account in most commercially available pharmacogenetic tests or on FDA drug labels.

African populations are among the most genetically diverse in the world.[3] In spite of this diversity, very few pharmacogenetics studies have been conducted in African populations. In fact, it is documented that there is inter-ethnic variability in pharmacogenetic traits between African populations.[4] Although the International HapMap Project has included three African populations, the Yoruba of Nigeria and the Maasai and the Luhya of Kenya, the population genetics of these three groups cannot represent the remaining populations in West and East Africa, or other populations living in Southern Africa. In order to achieve the goal of personalised medicine and individualisation of therapy in Africa, it is important carefully and

systematically to study pharmacogenetic traits in as many distinct African population groups as possible.

To bridge gaps in pharmacogenetic mapping in African populations, especially those residing in Southern Africa, this study prioritises the genotyping of 211 single nucleotide polymorphisms (SNPs) in 12 genes known to affect drug absorption, transport and metabolism in the Xhosa and Cape Mixed Ancestry (CMA) populations living in the Western Cape, South Africa. These genes are relevant for the pharmacokinetic disposition of a number of medications, including those used for the treatment of HIV infection, which is having a devastating impact in the region. The Xhosa population is indigenous to the Eastern Cape of South Africa, is the second largest ethnic group in South Africa and comprises approximately 17.6 per cent (~8 million) of the South African population.[5] The CMA population is known to have the highest rate of admixture worldwide, including mixes of European, African, South Asian and Indonesian ancestry, and comprises 8.9 per cent (~4 million) of the South African population.[5]

In this study, we characterised and examined differences in the minor allele frequency (MAF) estimates of pharmacogenetic alleles between the Xhosa and CMA populations. Secondly, we characterised haplotypes of the pharmacogenetic genes in both the Xhosa and CMA groups, and, finally, the MAF estimates we obtained for the Xhosa and CMA were compared with the HapMap estimates for other African, US and Asian populations. Taken together, these data should help lay the foundation for future pharmacogenetics studies in other South African populations, as well as the eventual use of pharmacogenetic testing, where clinically relevant, for the South African population.

Materials and methods

Study design
Written informed consent for the collection, storage and extraction of genomic DNA was obtained in English, Afrikaans or Xhosa from 176 unrelated HIV-positive South Africans.[6] The DNA

belonged to 109 Xhosa and 67 CMA individuals.[7] Ethnicity was determined by self-report. This study was approved by the individual Committees on Human Research at Stellenbosch University, South Africa and the University of California, San Francisco, USA.

Measurements

Two hundred and eleven tagging SNPs (tagSNPs) in 12 genes (Supplementary Table S1) that are important for drug absorption, transport and metabolism of antiretrovirals and other medications were selected using Snagger software.[8] This software takes into account different population frequencies of SNPs, as reported in HapMap, to generate a representative list of SNPs.[8] Because little is known about South African population genetic substructure and polymorphisms, HapMap Phase I, build 36 was used to select tagSNPs informative across all four population samples (ie Caucasian, Yoruba, Japanese and Han Chinese). In this manner, the likelihood of selecting for markers that may be informative in the Xhosa and the CMA populations is increased. Other SNPs were force-included based on their clinical pharmacogenetic relevance, as reported in the literature. All SNPs were included on a custom SNP genotyping array and DNA samples genotyped using the Illumina GoldenGate Assay kit (Illumina, Inc., San Diego, CA, USA).

Statistical analyses

Call rates, MAF and Hardy–Weinberg disequilibrium test p values were calculated using the R package. Chi-squared tests were used to test for Hardy–Weinberg disequilibrium. When small observed numbers were present for one or more genotype groups, Fisher's exact test was applied. Association analyses were performed using the co-dominant genetic model to report on SNPs with significantly different frequencies between the Xhosa and CMA groups. The significance criterion was set at a Bonferroni-corrected p value ≤ 0.05. In order to improve the quality of the genotype data, the SNP call rate was required to meet or exceed 90 per cent. The MAF of SNPs retained for association tests was required to meet or exceed 5 per cent in the Xhosa.

The R package, haplo.stats, was used to infer haplotypes. A sliding window haplotype association test was performed for each SNP represented in a given gene. This tests for association between haplotypes and an outcome. Given an ordered (by chromosomal locations) set of markers (1, 2, 3,...,*n*), sliding windows of overlapping haplotypes are tested in sequence (ie for window size = 3, markers 1-2-3 are treated as a single haplotype, then markers 2-3-4 are treated as a single haplotype, then markers 3-4-5, etc.). Haplotypes of varying sizes (2-, 3-, 4-SNP haplotypes) are assessed within each gene for this dataset. This haplotype test also assessed the association between identified haplotypes and outcome (in our case, ethnicity), as previously described by another group.[9]

Results

Study population

Our study population consisted of 176 HIV-positive, unrelated South African individuals, aged 21 and older. There were 109 Xhosa individuals, consisting of 79 females and 30 males, and 67 CMA individuals, consisting of 35 females and 32 males.

Comparison of SNPs between the Xhosa and CMA populations

Six of the 12 genes studied (those encoding the ATP binding cassette (*ABC*) sub-family B, member 1 [*ABCB1*], cytochrome P450 [*CYP*] *3A45*, UDP-glucuronosyltransferase 1 [*UGT1A1*], *CYP2C18*, *CYP3A4* and *CYP2A7P1*), in descending order of significance, contained at least one tagSNP that differed statistically between the Xhosa and the CMA populations (Table 1). The tagSNP results are presented in descending order of statistical significance (Table 1). Among the six genes, the MAF of eight of the 211 genotyped SNPs were statistically different between the Xhosa and the CMA, with the greatest difference found for

Table 1. Significantly different TagSNPs in the Xhosa and CMA populations

Gene	rsID	Minor allele	Xhosa MAF	CMA MAF	Bonferroni p value
ABCB1	rs13233308	T	0.01	0.18	1.77E-05
CYP3A5	rs4646450	C	0.03	0.17	0.004
UGT1A1	rs7572563	G	0.14	0.32	0.011
CYP2C18	rs2860840	T	0	0.09	0.015
CYP3A4	rs2738258	G	0.35	0.17	0.026
CYP2A7P1	rs11666982	G	0.22	0.43	0.028
UGT1A1	rs4148329	T	0.21	0.41	0.045
ABCB1	rs1202184	A	0.1	0.26	0.046

CMA, Cape Mixed Ancestry; MAF, minor allele frequency; rsID, ref SNP number ID.

the *ABCB1* SNP (rs13233308) ($p = 1.77$E-05; Table 1) and least difference found for *ABCB1* SNP (rs1202184) ($p = 0.0459$).

CYP3A5 SNP (rs4546450) occurred at a frequency of 0.03 in the Xhosa and 0.17 in the CMA ($p = 0.00393$). Two SNPs in *UGT1A1* were found to be statistically significantly different between the Xhosa and the CMA. The first SNP in *UGT1A1* (rs7572563) occurred at a frequency of 0.14 in the Xhosa and 0.32 in the CMA ($p = 0.0108$) and the second SNP in *UGT1A1* (rs4148329) occurred at a frequency of 0.21 in the Xhosa and 0.41 in the CMA population ($p = 0.0445$). One SNP in *CYP2C18* (rs2860840) is undetected in the Xhosa but occurred at a frequency of 0.09 in the CMA ($p = 0.0148$). A single SNP in *CYP3A4* (rs2738258) occurred at a frequency of 0.35 in the Xhosa and 0.17 in the CMA ($p = 0.0263$). A SNP in *CYP2A7* (rs11666982) occurred at a frequency of 0.22 in the Xhosa and 0.43 in the CMA ($p = 0.0279$).

Haplotype analysis

Based on the genotyped tagSNPs, haplotypes were constructed for each of the 12 genes. There were no identifiable haplotypes in the Xhosa and the CMA in the following genes: *ABCB1, ABCC2, CYP2A7P1, CYP2B6, CYP2C19, CYP2D6, CYP3A5* or *CYP3A7*; however, haplotypes were

identified in the gene encoding solute carrier family 22 member 6 (*SLC22A6*), *CYP2C18, CYP3A4* and *UGT1A1* (Table 2).

A total of four haplotypes were identified in the *SLC22A6* gene. The four-SNP TAGG haplotype of *SCL22A6* was found to occur at a significantly different frequency in the Xhosa (0.19) and in the CMA (0.05) ($p = 2.7$E-04; Table 2). In *CYP2C18*, a total of three haplotypes were identified. The four-SNP AAGC haplotype of *CYP2C18* occurred at a frequency of 0.40 in the Xhosa and 0.25 in the CMA ($p = 3.0$E-03). A total of seven haplotypes were identified in *CYP3A4*, which included the *1B SNP (rs2740574; Table 2). Six of the seven *CYP3A4* haplotypes were significantly different in terms of the population frequency between the Xhosa and the CMA. The four-SNP haplotype of *CYP3A4* which differed the most between the groups was the GCAG haplotype, which occurred at a frequency of 0.04 in the Xhosa, compared with 0.22 in the CMA population ($p = 3.3$E-05; Table 2).

A total of ten haplotypes in *UGT1A1* were identified. Unlike the other genes, two different haplotype blocks were identified in *UGT1A1*. One of the *UGT1A1* haplotype blocks consisted of four SNPs, composed of five haplotypes. Two of the four SNP haplotypes were significantly different in frequency between the Xhosa and CMA South African populations (Table 2).

Table 2. Inferred haplotypes in the Xhosa and CMA populations

Gene	Haplotypes				MAF		
					Xhosa	CMA	*p* Value
SLC22A6	rs6591722	rs4149170	rs12293966	rs955434			
	T	G	A	G	0.467	0.514	0.370
	T	A	A	A	0.233	0.199	0.460
	T	A	G	G	0.192	0.051	0.000
	A	G	A	G	0.082	0.147	0.091
CYP2C18	rs7085563	rs11188067	rs2860840	rs1326830			
	T	A	G	C	0.448	0.496	0.310
	A	A	G	C	0.403	0.258	0.003
	T	G	G	C	0.143	0.087	0.095
CYP3A4	rs2738258	rs7801671	rs2740574	rs7811025			
	G	A	G	G	0.127	0.290	0.003
	G	A	G	A	0.227	0.092	0.008
	A	C	G	G	0.180	0.084	0.014
	G	A	A	G	0.144	0.150	0.019
	G	C	A	G	0.049	0.224	3.30E-05
	A	A	G	G	0.100	0.079	0.140
	A	A	G	A	0.076	0.007	0.003
UGT1A1	rs8330	rs4148329	rs1500482	rs4663972			
	G	A	A	A	0.206	0.417	3.00E-05
	G	G	A	G	0.247	0.082	5.40E-05
	G	G	A	A	0.151	0.223	0.091
	C	G	G	G	0.183	0.156	0.520
	C	G	A	A	0.201	0.119	0.052
	rs2003569	rs887829	rs28900396				
	G	A	A		0.353	0.365	0.820
	G	G	A		0.129	0.330	5.20E-06
	A	G	A		0.177	0.154	0.640
	A	G	G		0.189	0.054	0.000
	G	G	G		0.149	0.094	0.110

CMA, Cape Mixed Ancestry; MAF, minor allele frequency

Table 3. Minor allele frequencies (MAFs) of clinically associated single nucleotide polymorphisms in African, US and Asian populations. MAF data for non-South African populations were obtained from the National Center for Biotechnology Information (NCBI) dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP)

Gene	rsID	Minor allele	Polymorphism (NTpos; AA change)	MAF								
				Xhosa (n = 109)	CMA (n = 67)	Yoruba (n = 60)	Luhya (n = 89)	Maasai (n = 143)	Caucasian (n = 60)	Chinese (n = 45)	Hispanic (n = 23)	AAm (n = 24)
ABC transporters												
ABCB1	rs2229109	A	1199G > A;S400N	0.0	0.0	0.0	NA	0.0	0.03	0.0	0.0	0.0
ABCB1	rs9282564	G	61A > G; N21D	0.0	0.0	0.0	NA	NA	0.1	0.0	0.07	0.0
ABCB1	rs1045642	C	3435C > T	0.89	0.8	0.88	NA	0.84	0.43	0.58	0.45	0.84
ABCB1	rs4148740	G	12386379A > G	0.14	0.17	0.13	NA	NA	0.1	0.02	NA	0.15
ABCB1	rs7787082	A	12391327G > A	0.69	0.61	0.68	0.71	0.62	0.18	0.46	NA	0.56
ABCB1	rs2032583	G	12394837A > G	0.14	0.18	0.25	0.25	0.23	0.16	0.03	0.04	0.15
ABCB1	rs3789243	G	12455162A > G	0.63	0.61	0.41	0.58	0.44	0.43	0.72	NA	0.6
ABCB1	rs1128503	T	1236C > T	0.13	0.23	0.12	0.11	0.14	0.39	0.68	0.54	0.06
ABCB1	rs10248420	G	12399262A > G	0.59	0.56	0.59	0.64	0.56	0.17	0.48	0.12	0.53
ABCB1	rs2235041	A	1240010G > A	0.12	0.07	0.12	0.1	0.1	0.0	0.0	0.0	0.14
ABCC2	rs8187707	T	4488C > T;H1496H	0.08	0.08	0.0	0.06	0.09	0.04	0.0	0.0	0.02
ABCC2	rs2273697	A	1249G > A;V417I	0.11	0.18	0.21	0.2	0.23	0.23	0.07	0.15	0.23
ABCC2	rs717620	T	-24C > T	0.03	0.04	0.04	0.02	0.04	0.18	0.22	0.23	0.06
Phase I												
CYP2B6	rs3745274	T	516G > T;Q172H,*6	0.2	0.23	0.45	0.31	0.37	0.25	0.15	0.27	0.46
CYP2B6	rs28399499	C	983T > C;I328T,*18	0.17	0.09	0.04	0.07	0.02	0.0	0.0	0.0	0.03
CYP2B6	rs36079186	C	593T > C; M198T,*27	0.0	0.0	NA	NA	NA	NA	NA	NA	NA

Continued

Table 3. Continued

Gene	rsID	Minor allele	Polymorphism (NTpos; AA change)	MAF								
				Xhosa (n = 109)	CMA (n = 67)	Yoruba (n = 60)	Luhya (n = 89)	Maasai (n = 143)	Caucasian (n = 60)	Chinese (n = 45)	Hispanic (n = 23)	AAm (n = 24)
CYP2C18	rs7896133	A	15213256G > A	0.2	0.2	0.12	0.11	0.16	0.07	0.08	NA	0.1
CYP2C18	rs2860840	T	15243758C > T	0.0	0.09	0	0.01	0.06	0.35	0.23	0.28	0.09
CYP2C18	rs10509675	A	15237414G > A	0.15	0.09	0.27	0.18	0.17	0.21	0.02	NA	0.13
CYP2C19	rs4917623	C	15358094T > C	0.11	0.2	0.18	0.16	0.29	0.49	0.6	NA	0.1
CYP2C19	rs12248560	T	-806C > T, *17	0.15	0.09	0.27	NA	NA	0.21	0.02	NA	0.15
CYP2C19	rs4244285	A	681G > A, *2	0.22	0.2	0.16	NA	NA	0.15	0.26	0.16	0.17
CYP2D6	rs3892097	A	1846G > A, *4	0.0	0.01	NA	NA	NA	0.5	NA	NA	NA
CYP3A4	rs4986910	C	1334T > C;M445T, *3	0.0	0.0	0.0	NA	NA	0.0	0.0	0.0	0.03
CYP3A4	rs4986909	T	1247C > T; P416L, *13	0.0	0.0	0.0	NA	NA	0.0	0.0	NA	NA
CYP3A4	rs28371759	C	878T > C; L293P, *18	0.0	0.0	NA	NA	NA	NA	NA	NA	NA
CYP3A4	rs4986907	A	485G > A; R162Q, *15A	0.01	0.0	0.0	NA	NA	0.0	0.0	0.0	0.03
CYP3A4	rs4987161	C	566T > C; F189S, *17	0.0	0.0	0.0	NA	NA	0.0	0.0	NA	NA
CYP3A4	rs2740574	G	-392A > G, *1B	0.77	0.58	0.75	NA	NA	0.03	0.0	0.11	0.68
CYP3A5	rs4646450	T	-1630 T > C	0.97	0.83	1.0	0.98	0.85	0.13	0.34	NA	NA
CYP3A5	rs10264272	A	711G > A;K208K, *6	0.2	0.17	0.15	0.25	0.14	0.0	0.0	0.02	0.11
CYP3A7	rs10211	G	129 A > G	0.35	0.52	0.75	0.72	0.44	0.06	0.34	0.19	0.58
Phase II												
UGT1A1	rs4148323	A	211G > A;G71R, *6	0.0	0.0	0.0	NA	NA	0.0	0.2	0.03	0.0
UGT1A1	rs887829	T	601324 C > T	0.35	0.36	0.55	NA	NA	0.28	0.08	0.35	0.57
UGT1A1	rs10929302	A	598536 G > A, *93	0.25	0.29	0.36	NA	NA	0.26	0.08	NA	NA

Abbreviations: AA, amino acid; AAm, African American; N = total number of samples per population; NA, no data available in the NCBI dbSNP database; NTpos, nucleotide position; Phase I, Phase I drug metabolising enzymes; Phase II, Phase II drug metabolising enzymes; CMA, Cape Mixed Ancestry; MAF, minor allele frequency
N.B. HapMap Caucasian minor alleles and MAFs are used as the reference in this table.

The second *UGT1A1* haplotype block consisted of three SNPs, composed of five haplotypes. Two of the three SNP haplotypes were found to be significantly different in frequency between the Xhosa and the CMA (Table 2). The most significant haplotype difference was the GGA haplotype of *UGT1A1*, which occurred at a frequency of 0.12 in the Xhosa and 0.33 in the CMA ($p = 5.2$E-06).

Comparison of SNPs between South African, the HapMap African, US and Asian populations

A comparison of the MAF of 35 pharmacogenetic SNPs with known functional or clinical associations in 10 genes (*ABCB1, ABCC2, CYP2B6, CYP2C18, CYP2C19, CYP2D6, CYP3A4, CYP3A5, CYP3A7* and *UGT1A1*) in the Xhosa and CMA populations is presented in Table 3. The MAF of the SNPs do not differ statistically between the Xhosa and CMA populations, except for *CYP3A5* rs4646450 ($p = 0.00393$) and *CYP2C18* rs2860804 ($p = 0.0148$).

Table 3 also shows a comparison between the allele frequencies obtained in the two distinct South African population groups in our study and available reports for other African populations, of which most data are known for the Yoruba from Nigeria and most recently the Maasai and the Luhya tribes of Kenya. In addition, the table displays a comparison of the allele frequencies in the African populations with other diverse populations in the USA and Asia.

Discussion

In this study, we analysed the allelic variation of 211 tag SNPs in 12 genes that are important in drug disposition and treatment outcome in two South African population groups: the Xhosa and the CMA. We identified both single SNPs and haplotypes which occurred at significantly different frequencies in the two populations.

In most sub-Saharan African countries, HIV/AIDS comprises one of the top socioeconomic and health burdens. It is estimated that 25 per cent of the adult population living in Southern Africa is infected with HIV, with an incidence of approximately 18 per cent in South Africa alone.[10] Given the high incidence of HIV/AIDS in South Africa, the greatest impact of pharmacogenetics may initially be made by improving treatment outcomes on antiretroviral therapy (ART).

In terms of the pharmacogenetic relevance of the *ABC* family of transporter genes, the evidence of their role in predicting HIV treatment-related toxicity is inconclusive. The presence of the *ABCB1* 3435T allele is associated with a decreased risk of hepatotoxicity in HIV patients treated with either efavirenz or nevirapine.[11,12] There is no conclusive evidence of the clinical significance of the *ABCB1* 1236C > T allele, however, although it appears minimally to affect the kinetics of the immunosuppressant drug cyclosporine.[13] Both efavirenz and nevirapine, which are non-nucleoside reverse transcriptase inhibitors (NNRTIs), are currently used in first-line treatment regimens of HIV-infected individuals in South Africa.[14] Therefore, it would be important to assess the importance and contribution of the *ABCB1* variant alleles to drug-related toxicity with these NNRTIs. Parathyras *et al.* studied the association between a number of variants of *ABCB1* and immune recovery in South Africans treated with ART and found no association between the well-known 3435T allele and immune recovery;[7] however, an association was found between the *ABCB1* G2677A SNP and immune recovery in this study.[7] Based on the results of the present study, it would be interesting to investigate whether there is an association between the two *ABCB1* tagSNPs (rs13233308 and rs1202184) found to be significantly different between the Xhosa and CMA populations and immune recovery.

According to the South African Department of Health, the current second-line ART regimen should include the anchoring agent lopinavir and ritonavir.[14] Therefore, pharmacogenetic traits of *CYP3A4* and *CYP3A5* may have an impact on the treatment outcome of second-line therapy. As protease inhibitors are both substrates and inhibitors of CYP3A, however, the influence of *CYP3A* gene

variation on ART treatment outcomes is difficult to discern[15] — although the *CYP3A4**1B variant allele is associated with variability in the pharmacokinetics of the protease inhibitor indinavir.[16] In fact, homozygotes for the *1B variant have a lower bioavailability of indinavir than heterozygotes and homozygotes for the *1A common allele.[16] Similarly, the common allele of *CYP3A5* A6986G is associated with increased clearance of indinavir.[17] Similar association studies should be carried out to assess the contribution of *CYP3A* variants to response or exposure to lopinavir in the South African population. It would be interesting to assess the influence of the *CYP3A5**6 variant that results in a loss-of-function of the CYP3A5 enzyme on lopinavir exposure and treatment outcome in South Africans, as this allele is more common in people of African descent than in Caucasians and Asians.[18,19] The variant occurred at a frequency of 0.2 in the Xhosa and 0.17 in the CMA populations in the present study. In addition, the influence of the *CYP3A4* SNP (rs2738258) and *CYP3A5* SNP (rs4646450), both found to be statistically significantly different between the Xhosa and CMA in the present study, on lopinavir exposure and treatment outcome should be investigated.

The single SNPs and haplotype structures inferred for the Xhosa and the CMA populations in the *UGT1A1* gene could be used more accurately to stratify the two populations in order to perform pharmacogenetic association studies. The South African HIV treatment guidelines changed in April 2010, and first-line ART. Now includes tenofovir in addition to either nevirapine or efavirenz and lamivudine.[14] Although both lamivudine and tenofovir are only nominally affected by CYP enzymes, they are glucuronidated in the liver and excreted unchanged through the kidneys.[15] Therefore, studies could be designed to assess the impact of *UGT1A1* SNPs (rs7572563 and rs4148329) and haplotypes on treatment outcomes of antiretroviral drugs that may undergo glucuronidation prior to excretion, such as tenofovir. It would make sense initially to genotype SNPs of the *UGT1A1* haplotype with known functional alleles such as the *UGT1A1**93 or the rs887829 SNP, both of which have been associated

with hyperbilirubinaemia,[20,21] and assess their impact on the response to tenofovir.

It is clear that there are differences in the MAF of key pharmacogenetic alleles in South African populations compared with other African populations (Table 3). Of particular interest, the loss of function *CYP2B6**18 variant allele is thought to occur most frequently in West African populations, with a reported MAF of 0.04.[22] In the present study, however, we find that it occurs at a frequency of 0.17 in the Xhosa and 0.09 in the CMA populations, compared with a reported frequency of 0.07 in the Luhya and 0.02 in the Maasai populations (Table 3). The *CYP2B6* gene plays an important role in the metabolism of two of the first-line ART drugs used in South Africa: efavirenz and nevirapine. The *CYP2B6**18 SNP is the only coding SNP in *CYP2B6*.[23] The variant is associated with elevated plasma concentrations of efavirenz and nevirapine and hepatotoxicity in HIV patients from Mozambique treated with either drug.[24–26] To our knowledge, the present study is the first report on the MAF of the *CYP2B6**18 variant in the Xhosa and the CMA populations. Given that efavirenz and nevirapine are both first-line treatment agents in this region, further investigation of the association between *CYP2B6* null variant alleles and adverse reactions in South African populations is warranted. Such findings have important implications for the incidence of adverse reactions to efavirenz and nevirapine in different African populations.

The current study is the first of its kind systematically to characterise tagging and clinically relevant pharmacogenetic SNPs in the two South African populations; however, there are inherent limitations to our analyses. First, as this was a purely descriptive study, there are no associations made with any disease (eg HIV) or treatment outcomes; however, this work lays the foundation for the future study of such associations. Secondly, whereas the sample size of the Xhosa sample is adequate, the sample size of the CMA population is modest and it is possible that lower frequency alleles could not be detected in this group. Thirdly, the frequency of SNPs typed in this study were previously characterised in other

populations and therefore we cannot rule out the presence of novel SNPs for which we did not test in the Xhosa and CMA populations, such as those recently reported in *CYP2C19* and *CYP2D6* in these populations.[5,27] Fourthly, the haplotypes inferred are limited by the sample size of our population and there may be others that remain to be identified. Lastly, 12 genes that are known to be associated with treatment outcomes in HIV infection were characterised but there are likely to be more genes that remain to be studied.

Conclusion

To our knowledge, this is the largest pharmacogenetics study of two distinct South African population groups. Our work shows that there are significant differences in the frequencies of variant alleles in several genes (*ABCB1, CYP2A7P1, CYP2C18, CYP3A4, CYP3A5* and *UGT1A1*) associated with treatment outcome in the Xhosa and the CMA populations of South Africa. It also shows that for the majority of SNPs analysed, there is great similarity in allele frequency between the two groups. Such work is of great importance for laying the foundation for ethnicity-specific genotype-to-phenotype correlates of treatment outcome for these various enzyme polymorphisms and their drug substrates. Importantly, we also identified novel haplotype structures in four genes (*CYP2C18, CYP3A4, SLC22A6* and *UGT1A1*) in the two distinct South African populations. The haplotypes could be used, in addition to single SNPs, to more accurately stratify patient groups according to ethnicity and to aid in identifying associations between causative variants and drug response.

It is clear from this work and that of others that not all African groups share the same allele frequencies of key pharmacogenetic genes.[4,5,7] Therefore, it is important that studies such as this is performed in as many populations as possible, to generate the most useful information on the clinical application of pharmacogenetics for these specific populations. Caution is advised in using a single African population in pharmacogenetics studies since it cannot be representative for all Africans.

Acknowledgments

References

1. Ikediobi, O.N., Shin, J., Nussbaum, R.L., Phillips, K.A. *et al.* (2009), 'Addressing the challenges of the clinical application of pharmacogenetic testing', *Clin. Pharmacol. Ther.* Vol. 86, pp. 28–31.

2. Marques, S.C. and Ikediobi, O.N. (2010), 'The clinical application of UGT1A1 pharmacogenetic testing: Gene-environment interactions', *Hum. Genomics* Vol. 4, pp. 238–249.

3. Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C. *et al.* (2009), 'The genetic structure and history of Africans and African Americans', *Science* Vol. 324, pp. 1035–1044.

4. Yen-Revollo, J.L., van Booven, D.J., Peters, E.J., Hoskins, J.M. *et al.* (2009), 'Influence of ethnicity on pharmacogenetic variation in the Ghanaian population', *Pharmacogenomics J.* Vol. 9, pp. 373–379.

5. Drögemöller, B.I., Wright, G.E.B., Niehaus, D.J.H., Koen, L. *et al.* (2010), 'Characterization of the genetic profile of CYP2C19 in two South African populations', *Pharmacogenomics* Vol. 11, pp. 1095–1103.

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

7. Parathyras, J., Gebhardt, S., Hillermann-Rebello, R., Grobbelaar, N. *et al.* (2009), 'A pharmacogenetic study of CD4 recovery in response to HIV antiretroviral therapy in two South African population groups', *J. Hum. Genet.* Vol. 54, pp. 261–265.

8. Edlund, C.K., Lee, W.H., Li, D., van den Berg, D.J. *et al.* (2008), 'Snagger: A user-friendly program for incorporating additional information for tagSNP selection', *BMC Bioinformatics* Vol. 9, p. 174.

9. Mathias, R.A., Gao, P., Goldstein, J.L., Wilson, A.F. *et al.* (2006), 'A graphical assessment of p-values from sliding window haplotype tests of association to identify asthma susceptibility loci on chromosome 11q', *BMC Genet.* Vol. 7, p. 38.

10. The Joint United Nations Programme on HIV/AIDS, http://www.unaids.org (last accessed 30th April, 2011).

11. Haas, D.W., Bartlett, J.A., Andersen, J.W., Sanne, I. *et al.* (2006), 'Pharmacogenetics of nevirapine-associated hepatotoxicity: An Adult AIDS Clinical Trials Group collaboration', *Clin. Infect. Dis.* Vol. 43, pp. 783–786.

12. Ritchie, M.D., Haas, D.W., Motsinger, A.A., Donahue, J.P. *et al.* (2006), 'Drug transporter and metabolizing enzyme gene variants and non-nucleoside reverse-transcriptase inhibitor hepatotoxicity', *Clin. Infect. Dis.* Vol. 43, pp. 779–782.

13. Zhang, Y.T., Yang, L.P., Shao, H., Li, K.X. *et al.* (2008), 'ABCB1 polymorphisms may have a minor effect on ciclosporin blood concentrations in myasthenia gravis patients', *Br. J. Clin. Pharmacol.* Vol. 66, pp. 240–246.

14. South African Department of Health website. http://www.doh.gov.za (last accessed 30th April, 2011).

15. Mahungu, T.W., Johnson, M.A., Owen, A. and Back, D.J. (2009), 'The impact of pharmacogenetics on HIV therapy', *Int. J. STD AIDS* Vol. 20, pp. 145–151.

16. Bertrand, J., Treluyer, J.M., Panhard, X., Tran, A. *et al.* (2009), 'Influence of pharmacogenetics on indinavir disposition and short-term response in HIV patients initiating HAART', *Eur. J. Clin. Pharmacol.* Vol. 65, pp. 667–678.

17. Anderson, P.L., Lamba, J., Aquilante, C.L., Schuetz, E. *et al.* (2006), 'Pharmacogenetic characteristics of indinavir, zidovudine, and lamivudine

therapy in HIV-infected adults: A pilot study', *J. Acquir. Immune Defic. Syndr.* Vol. 42, pp. 441–449.

18. Kuehl, P., Zhang, J., Lin, Y., Lamba, J. *et al.* (2001), 'Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression', *Nat. Genet.* Vol. 27, pp. 383–391.

19. Balram, C., Zhou, Q., Cheung, Y.B. and Lee, E.J. (2003), 'CYP3A5*3 and *6 single nucleotide polymorphisms in three distinct Asian populations', *Eur. J. Clin. Pharmacol.* Vol. 59, pp. 123–126.

20. Sanna, S., Busonero, F., Maschio, A., McArdle, P.F. *et al.* (2009), 'Common variants in the SLCO1B3 locus are associated with bilirubin levels and unconjugated hyperbilirubinemia', *Hum. Mol. Genet.* Vol. 18, pp. 2711–2718.

21. Innocenti, F., Grimsley, C., Das, S., Ramírez, J. *et al.* (2002), 'Haplotype structure of the UDP-glucuronosyltransferase 1A1 promoter in different ethnic groups', *Pharmacogenetics* Vol.12, pp. 725–733.

22. Mehlotra, R.K., Bockarie, M.J. and Zimmerman, P.A. (2007) 'CYP2B6 983T>C polymorphism is prevalent in West Africa but absent in Papua New Guinea: Implications for HIV/AIDS treatment', *Br. J. Clin. Pharmacol.* Vol. 64, pp. 391–395.

23. Klein, T.E., Chang, J.T., Cho, M.K., Easton, K.L. *et al.* (2001), 'Integrating genotype and phenotype information: An overview of the PharmGKB Project', *Pharmacogenomics J.* Vol. 1, pp. 167–170.

24. Klein, K., Lang, T., Saussele, T., Barbosa-Sicard, E. *et al.* (2005), 'Genetic variability of CYP2B6 in populations of African and Asian origin: Allele frequencies, novel functional variants, and possible implications for anti-HIV therapy with efavirenz', *Pharmacogenet. Genomics* Vol. 15, pp. 861–873.

25. Wang, J., Sönnerborg, A., Rane, A., Josephson, F. *et al.* (2006), 'Identification of a novel specific CYP2B6 allele in Africans causing impaired metabolism of the HIV drug efavirenz', *Pharmacogenet. Genomics* Vol. 16, pp. 191–198.

26. Ciccacci, C., Borgiani, P., Ceffa, S., Sirianni, E. *et al.* (2010), 'Nevirapine-induced hepatotoxicity and pharmacogenetics: A retrospective study in a population from Mozambique', *Pharmacogenomics* Vol. 11, pp. 23–31.

27. Wright, G.E.B., Niehaus, D.J.H., Drögemöller, B.I., Koen, L. *et al.* (2010), 'Elucidation of *CYP2D6* genetic variants in a unique African population: Implications for the future application of pharmacogenetics in the Xhosa population', *Ann. Hum. Genet.* Vol. 74, pp. 340–350.

Table S1. Minor allele frequencies (MAF) of 211 tagging single nucleotide polymorphisms (SNPs) in the Xhosa and Cape Mixed Ancestry (CMA) populations

Gene	SNP	Minor allele	Call rate	MAF		HWE *p* value	
				CMA	Xhosa	CMA	Xhosa
ABCB1	rs2229109	A	1	0.0	0.0	NA	NA
ABCB1	rs1202185	G	1	0.187	0.216	0.789	0.242
ABCB1	rs10233247	G	1	0.075	0.101	2.22E-16	0.0
ABCB1	rs2235046	A	1	0.254	0.211	0.84	0.624
ABCB1	rs4148743	A	1	0.41	0.358	0.248	0.691
ABCB1	rs2373589	A	1	0.373	0.339	0.0551	0.297
ABCB1	rs6946119	G	1	0.067	0.041	0.556	0.653
ABCB1	rs10275831	A	1	0.037	0.05	0.751	0.579
ABCB1	rs1858923	G	1	0.239	0.128	0.428	0.863
ABCB1	rs2235033	G	1	0.463	0.472	0.25	0.898
ABCB1	rs2157926	T	1	0.537	0.413	0.0328	0.574
ABCB1	rs1922242	T	1	0.448	0.491	0.779	0.629
ABCB1	rs2032588	A	1	0.127	0.197	0.931	0.453
ABCB1	rs10267099	G	1	0.09	0.11	0.421	0.197
ABCB1	rs7787082	G	1	0.396	0.317	0.0048	0.0708
ABCB1	rs4148740	G	1	0.179	0.142	0.48	0.532
ABCB1	rs1211152	A	1	0.0	0.0	NA	NA
ABCB1	rs7796247	A	1	0.037	0.05	0.751	0.579
ABCB1	rs1202175	G	1	0.187	0.216	0.789	0.242
ABCB1	rs1922241	A	1	0.216	0.193	0.535	0.52
ABCB1	rs6465118	A	1	0.224	0.229	0.652	0.493
ABCB1	rs4728705	A	1	0.134	0.133	0.826	0.44
ABCB1	rs13229143	G	1	0.239	0.124	0.428	0.772
ABCB1	rs6950978	T	1	0.0	0.0	NA	NA
ABCB1	rs1202172	C	1	0.351	0.362	0.505	0.776
ABCB1	rs4148732	G	1	0.0	0.0	NA	NA
ABCB1	rs10274587	A	1	0.134	0.128	0.826	0.304
ABCB1	rs3789243	A	1	0.396	0.372	0.205	0.0968
ABCB1	rs4437575	G	1	0.418	0.353	0.881	0.557

Continued

Table S1. Continued

Gene	SNP	Minor allele	Call rate	MAF		HWE *p* value	
				CMA	Xhosa	CMA	Xhosa
ABCB1	rs4728709	G	1	0.522	0.408	0.108	0.261
ABCB1	rs4728707	A	1	0.209	0.188	0.0306	0.928
ABCB1	rs2032583	G	1	0.187	0.147	0.591	0.618
ABCB1	rs17149866	G	1	0.03	0.028	2.22E-16	0.0
ABCB1	rs1882479	G	1	0.119	0.106	0.959	0.0698
ABCB1	rs1055302	A	1	0.231	0.33	0.687	0.211
ABCB1	rs12334183	G	1	0.246	0.317	0.202	0.395
ABCB1	rs10264990	G	1	0.299	0.275	0.25	0.902
ABCB1	rs17149699	A	1	0.284	0.321	0.713	0.917
ABCB1	rs1882478	G	1	0.425	0.399	0.289	0.0243
ABCB1	rs10248420	A	1	0.448	0.413	0.00595	0.339
ABCB1	rs10236274	G	1	0.254	0.289	0.136	0.327
ABCB1	rs3213619	G	1	0.224	0.193	0.652	0.977
ABCB1	rs10225473	G	1	0.127	0.119	0.931	0.186
ABCB1	rs17327442	A	1	0.478	0.459	2.22E-16	0.0
ABCB1	rs17149792	A	1	0.119	0.266	0.959	0.529
ABCB1	rs1202169	G	1	0.261	0.225	0.718	0.781
ABCB1	rs6949448	A	1	0.187	0.151	0.591	0.708
ABCB1	rs1202184	A	1	0.269	0.106	0.919	0.218
ABCB1	rs4148745	A	1	0.127	0.138	0.309	0.39
ABCB1	rs1002205	C	0.994	0.276	0.292	0.947	0.025
ABCB1	rs17327624	A	0.994	0.209	0.185	0.956	0.85
ABCB1	rs1202181	G	0.994	0.187	0.213	0.789	0.276
ABCB1	rs2235023	A	0.994	0.269	0.333	0.919	1.0
ABCB1	rs4148734	A	0.994	0.067	0.032	0.174	0.728
ABCB1	rs13233308	A	0.994	0.182	0.018	0.328	0.845
ABCB1	rs1045642	T	0.994	0.205	0.11	0.349	0.754
ABCB1	rs4148738	G	0.989	0.194	0.164	0.683	0.922
ABCB1	rs7779562	G	0.983	0.4	0.444	0.179	0.153

Continued

Table S1. Continued

Gene	SNP	Minor allele	Call rate	MAF		HWE *p* value	
				CMA	Xhosa	CMA	Xhosa
ABCB1	rs9282564	G	0.977	0.008	0.0	0.95	NA
ABCB1	rs10808071	G	0.977	0.164	0.194	0.0379	0.0732
ABCB1	rs2235041	A	0.977	0.077	0.126	0.282	0.256
ABCB1	rs2235074	NA	0.665	NA	NA	NA	NA
ABCB1	rs1002204	NA	0.119	NA	NA	NA	NA
ABCC2	rs2073337	G	1	0.403	0.491	0.34	0.387
ABCC2	rs7476245	A	1	0.097	0.092	0.606	0.0167
ABCC2	rs8187707	A	1	0.082	0.083	0.374	0.347
ABCC2	rs17216282	G	1	0.194	0.161	0.248	0.198
ABCC2	rs8187674	G	1	0.045	0.055	0.701	0.543
ABCC2	rs2180989	C	1	0.276	0.317	0.586	0.684
ABCC2	rs11595888	A	1	0.0	0.0	NA	NA
ABCC2	rs4148396	A	1	0.321	0.303	0.537	0.997
ABCC2	rs4148399	C	1	0.045	0.133	0.701	0.44
ABCC2	rs4148398	A	1	0.187	0.101	0.591	0.907
ABCC2	rs17112266	A	1	0.052	0.078	0.652	0.377
ABCC2	rs7067971	G	1	0.201	0.138	0.832	0.39
ABCC2	rs2804398	A	1	0.269	0.317	0.469	0.684
ABCC2	rs3740065	G	1	0.313	0.445	0.812	0.822
ABCC2	rs3740063	A	1	0.5	0.477	0.903	0.647
ABCC2	rs2273697	A	1	0.187	0.11	0.591	0.101
ABCC2	rs7898096	A	1	0.09	0.211	0.421	0.0702
ABCC2	rs2002042	A	0.994	0.201	0.185	0.832	0.408
ABCC2	rs2756109	A	0.983	0.47	0.448	0.691	0.779
ABCC2	rs4148391	T	0.977	0.0	0.0	NA	NA
ABCC2	rs717620	T	0.898	0.04	0.037	0.743	0.709
ABCC2	rs8187710	NA	0.341	NA	NA	NA	NA
ABCC2	rs2756112	NA	0	NA	NA	NA	NA
CYP2A7P1	rs7250597	A	1	0.082	0.165	0.464	0.499

Continued

Table S1. Continued

Gene	SNP	Minor allele	Call rate	MAF		HWE *p* value	
				CMA	Xhosa	CMA	Xhosa
CYP2A7P1	rs7249735	C	1	0.06	0.092	2.22E-16	0.0
CYP2A7P1	rs7255146	A	0.994	0.227	0.339	0.774	0.128
CYP2A7P1	rs11666982	A	0.994	0.433	0.227	0.471	0.161
CYP2B6	rs1042389	G	1	0.231	0.271	0.0138	0.633
CYP2B6	rs8192719	A	1	0.306	0.317	0.464	0.0294
CYP2B6	rs28399499	G	1	0.09	0.174	2.22E-16	0.0
CYP2B6	rs35773040	A	1	0.0	0.0	NA	NA
CYP2B6	rs35303484	G	1	0.0	0.0	NA	NA
CYP2B6	rs2279345	A	1	0.216	0.161	0.18	0.398
CYP2B6	rs7260329	A	1	0.209	0.106	0.956	0.218
CYP2B6	rs3211371	G	1	0.0	0.0	NA	NA
CYP2B6	rs16974799	A	1	0.306	0.335	0.464	0.0694
CYP2B6	rs7255374	A	1	0.366	0.399	0.614	0.18
CYP2B6	rs36060847	A	1	0.0	0.0	NA	NA
CYP2B6	rs8100458	G	1	0.336	0.252	0.394	0.136
CYP2B6	rs2113103	A	1	0.261	0.151	0.104	0.0626
CYP2B6	rs35979566	A	1	0.0	0.0	NA	NA
CYP2B6	rs36079186	G	1	0.0	0.0	NA	NA
CYP2B6	rs34223104	G	1	0.007	0.005	0.951	0.962
CYP2B6	rs2279342	A	1	0.015	0.018	0.901	0.845
CYP2B6	rs8192712	G	1	0.03	0.009	0.801	0.923
CYP2B6	rs2054675	G	1	0.306	0.339	0.464	0.0514
CYP2B6	rs3745274	C	1	0.231	0.202	0.0138	0.00829
CYP2B6	rs11882450	G	0.994	0.06	0.079	0.603	0.66
CYP2B6	rs4803417	C	0.994	0.159	0.115	0.0321	0.683
CYP2B6	rs6508963	A	0.994	0.007	0.023	0.951	0.805
CYP2B6	rs34097093	A	0.983	0.0	0.0	NA	NA
CYP2B6	rs2306606	G	0.983	0.201	0.184	0.0389	0.0203
CYP2B6	rs2279343	NA	0	NA	NA	NA	NA

Continued

Table S1. Continued

Gene	SNP	Minor allele	Call rate	MAF		HWE *p* value	
				CMA	Xhosa	CMA	Xhosa
CYP2C18	rs11188059	A	1	0.03	0.0	0.801	NA
CYP2C18	rs932809	A	1	0.224	0.248	0.801	0.5
CYP2C18	rs12243416	A	1	0.09	0.151	0.421	0.711
CYP2C18	rs1326830	A	1	0.067	0.0	0.174	NA
CYP2C18	rs7896133	A	1	0.201	0.202	0.832	0.128
CYP2C18	rs2296680	A	1	0.209	0.234	0.956	0.277
CYP2C18	rs2296684	A	1	0.067	0.119	0.174	0.682
CYP2C18	rs2860840	A	1	0.09	0.005	0.421	0.962
CYP2C18	rs10509675	A	1	0.09	0.151	0.421	0.711
CYP2C18	rs1409656	G	1	0.112	0.151	0.302	0.708
CYP2C18	rs7917985	A	1	0.157	0.266	0.551	0.889
CYP2C18	rs7085563	A	0.994	0.333	0.404	0.0648	0.0581
CYP2C18	rs1010570	A	0.994	0.485	0.495	0.114	0.564
CYP2C18	rs1409655	A	0.994	0.098	0.064	0.375	0.38
CYP2C18	rs11188067	G	0.983	0.1	0.148	0.37	0.778
CYP2C19	rs10509677	G	1	0.022	0.014	0.851	0.884
CYP2C19	rs4917623	G	1	0.209	0.115	0.427	0.683
CYP2C19	rs1322179	A	1	0.201	0.211	0.832	0.509
CYP2C19	rs12248560	A	1	0.09	0.151	0.421	0.711
CYP2C19	rs4244285	A	1	0.201	0.22	0.832	0.339
CYP2C19	rs7101258	C	1	0.172	0.183	0.378	0.395
CYP2C19	rs7915414	A	0.994	0.418	0.44	0.0309	0.966
CYP2D6	rs5758589	A	1	0.321	0.335	0.537	0.339
CYP2D6	rs9623531	G	0.983	0.408	0.338	0.259	0.151
CYP3A4	rs4986910	G	1	0.0	0.0	NA	NA
CYP3A4	rs28371759	G	1	0.0	0.0	NA	NA
CYP3A4	rs17161829	A	1	0.09	0.22	0.421	0.0669
CYP3A4	rs7801671	C	1	0.328	0.284	6.29E-05	3.33E-05
CYP3A4	rs17161886	C	1	0.104	0.252	0.34	0.0455

Continued

Table S1. Continued

Gene	SNP	Minor allele	Call rate	MAF		HWE *p* value	
				CMA	Xhosa	CMA	Xhosa
CYP3A4	rs2687117	A	1	0.172	0.275	0.403	0.721
CYP3A4	rs12333983	A	1	0.433	0.271	0.471	0.328
CYP3A4	rs2740574	A	1	0.425	0.234	0.575	0.606
CYP3A4	rs4986909	A	1	0.0	0.0	NA	NA
CYP3A4	rs2738258	A	1	0.172	0.358	0.00932	0.691
CYP3A4	rs10267228	C	0.994	0.306	0.333	0.676	0.0833
CYP3A4	rs7811025	A	0.983	0.159	0.341	1.54E-11	0
CYP3A4	rs4987161	G	0.983	0.0	0.0	NA	NA
CYP3A5	rs4646450	C	1	0.172	0.032	0.0817	5.98E-09
CYP3A5	rs1419745	G	1	0.433	0.495	0.00671	0.773
CYP3A5	rs4646446	A	1	0.104	0.046	0.34	0.616
CYP3A5	rs10224569	A	0.994	0.194	0.329	0.248	0.309
CYP3A5	rs10264272	A	0.989	0.179	0.201	0.48	0.427
CYP3A7	rs10211	A	1	0.522	0.358	0.00101	0.663
CYP3A7	rs2687144	A	1	0.53	0.431	0.0024	0.774
CYP3A7	rs2687074	C	1	0.164	0.252	0.0508	0.295
CYP3A7	rs2687136	A	0.994	0.508	0.463	0.0139	0.191
SLC22A6	rs12293966	G	1	0.082	0.193	0.464	0.977
SLC22A6	rs6591722	A	1	0.149	0.083	0.0158	0.112
SLC22A6	rs2276300	A	1	0.0	0.0	NA	NA
SLC22A6	rs4149170	A	1	0.284	0.45	0.404	0.691
SLC22A6	rs955434	A	0.994	0.22	0.234	0.894	0.985
SLC22A6	rs10897310	NA	0.591	NA	NA	NA	NA
UGT1A1	rs4663972	G	1	0.239	0.431	0.143	0.621
UGT1A1	rs28900396	G	1	0.149	0.339	0.625	0.142
UGT1A1	rs4148323	A	1	0	0	NA	NA
UGT1A1	rs3755319	A	1	0.396	0.261	0.806	0.823
UGT1A1	rs10199882	G	1	0.164	0.156	0.288	0.326
UGT1A1	rs11563251	G	1	0.56	0.44	0.996	0.469

Continued

Table S1. Continued

Gene	SNP	Minor allele	Call rate	MAF		HWE *p* value	
				CMA	Xhosa	CMA	Xhosa
UGT1A1	rs1018124	G	1	0.09	0.087	0.421	0.836
UGT1A1	rs4148328	A	1	0.269	0.124	0.469	0.14
UGT1A1	rs9784064	A	1	0.119	0.197	0.267	0.0103
UGT1A1	rs10929303	A	1	0.299	0.381	0.0771	0.465
UGT1A1	rs887829	A	1	0.366	0.353	0.109	0.503
UGT1A1	rs4148329	A	1	0.418	0.216	0.393	0.242
UGT1A1	rs2003569	A	1	0.209	0.367	0.125	0.269
UGT1A1	rs7572563	G	1	0.328	0.147	0.0741	0.0725
UGT1A1	rs3771342	A	1	0.164	0.165	0.00446	0.475
UGT1A1	rs8330	C	1	0.276	0.385	0.586	0.0907
UGT1A1	rs28946889	A	0.994	0.0	0.0	NA	NA
UGT1A1	rs929596	G	0.994	0.333	0.312	0.14	0.0398
UGT1A1	rs1500482	G	0.994	0.144	0.193	0.527	0.208
UGT1A1	rs4663971	C	0.983	0.469	0.477	0.4	0.184
UGT1A1	rs6431630	NA	0	NA	NA	NA	NA

Abbreviations: Call rate, percentage of samples with a genotype; HWE, Hardy–Weinberg equilibrium; NA, no data available
*Grey boxes indicate SNPs with call rates <90 per cent