Characterization of the genetic profile of *CYP2C19* in two South African populations

Aims: This study was aimed at elucidating the common sequence variation present in the *CYP2C19* gene within the South African Xhosa population and comparing it with the Cape Mixed Ancestry (CMA) population for possible future pharmacogenetic applications. **Materials & methods:** Common sequence variation was identified through the resequencing of 15 Xhosa individuals. The detected variants were prioritized for genotyping in an additional 85 Xhosa and 75 CMA individuals, while 5′-upstream variants were analyzed using dual luciferase reporter assays. **Results:** Resequencing of the Xhosa population revealed 30 variants, including the novel *CYP2C19*27* and *CYP2C19*28* alleles. *CYP2C19*27*, characterized by -1041G>A, caused a twofold decrease in luciferase activity, while *CYP2C19*28* is characterized by the nonsynonymous V374I variant. In addition, the previously characterized variants, *CYP2C19*2*, *CYP2C19*9* and *CYP2C19*17*, were present in both populations, while *CYP2C19*3* was only observed in the CMA population. **Conclusion:** Our data demonstrate that both the Xhosa and CMA populations exhibit unique genetic profiles that could influence the outcome of drug therapy in these populations.

KEYWORDS: Cape Mixed Ancestry ■ *CYP2C19* ■ pharmacogenetics ■ South Africa ■ Xhosa

Britt I Drögemöller[1], Galen EB Wright[1], Dana JH Niehaus[2], Liezl Koen[2], Stefanie Malan[1], Danielle M Da Silva[1], Renate Hillermann–Rebello[1], Anthony M La Grange[1], Mauritz Venter[1] & Louise Warnich[†1]

[1]Department of Genetics, Stellenbosch University, Private Bag XI, Matieland, 7602, South Africa
[2]Department of Psychiatry, Stikland Hospital, Stellenbosch University, South Africa
[†]Author for correspondence:
Tel.: +27 218 085 888
Fax: +27 218 085 833
lw@sun.ac.za

The frequent occurrence of adverse drug reactions (ADRs) and treatment failure contribute significantly to the economic and healthcare burdens worldwide, with fatal ADRs falling in the top ten causes of death [1,2]. It has, however, been argued that through the implementation of pharmacogenetics, the rate of ADRs could be reduced by 10–20%, and the efficiency of drugs could be increased by 10–15% [3]. *CYP* genes, which encode Phase I drug-metabolizing enzymes, have been demonstrated to be good candidates for pharmacogenetic studies [4], and may provide information to aid in the attainment of optimal treatment plans. Members of the *CYP2C* subfamily are collectively involved in the metabolism of approximately 20% of prescribed drugs [5] and have been categorized as known 'valid' pharmacogenomic biomarkers as they express measurable genetic polymorphisms proven to be associated with a variable drug response [6]. More specifically, the *CYP2C19* gene has been shown to be of particular value in the context of drugs such as antidepressants, antiulcer, antimalarial, anti-HIV and antiplatelet agents. All of these therapeutic agents are metabolized by the CYP2C19 enzyme and have been associated with ADR and/or treatment failure [7,8]. The examination of the *CYP2C19* gene is thus of value to the field of pharmacogenetics.

Several polymorphisms affecting the functioning of the CYP2C19 enzyme have been identified and individuals can be categorized into poor metabolizers (PMs), intermediate metabolizers (IMs), extensive metabolizers (EMs) and ultrarapid metabolizers (UMs), according to the level of enzyme functionality coded by the *CYP2C19* gene. Several populations have been genotyped for the nonfunctional *CYP2C19*2* and *CYP2C19*3* alleles and significant differences in the frequencies of these variants have been widely documented [9]. Unfortunately, thorough characterization of populations to identify novel variants using moderate-to-large sample sizes remains limited, especially in the genetically diverse African populations. A recent article highlighting the importance of *CYP2C19*17* [7] emphasized the need for thorough genotyping panels, with special attention to the 5′-upstream areas. Furthermore, the *CYP2C19* gene should be characterized for copy number variants (CNVs) that have not been reported for *CYP2C19* to date, but which are present in the *CYP2D6* and *CYP2A6* genes in African populations [10,11].

African populations are of special interest as they are the most ancient of populations and, consequently, their genomes have been exposed to greater diversification pressures than non-African populations, as reiterated in a recent paper by Tishkoff *et al.* [12]. The large number of ethnolinguistic groups present in Africa [101], although not necessarily dividing the groups

future medicine part of fsg

into isolated populations, may play a role in further subdividing the already genetically diverse African populations. More specifically, the 11 official languages spoken in South Africa hint at the number of diverse populations residing within the country. Although there remains a vast amount of information to be obtained from these populations, to our knowledge, only one South African population, the Venda population, which comprises only 2.3% of the South African population [102], has been examined for *CYP2C19* variation to date [13,14].

This study places its focus on the Xhosa population, while also examining the Cape Mixed Ancestry (CMA) or Cape Coloured population. The Xhosa population comprises 17.6% of the South African population, making it the second largest unique South African population, while the CMA comprises 8.9% of the country's population [102]. The Xhosa, like the Venda, are a Bantu-speaking population belonging to the Niger–Kordofanian African macrofamily [15]; however, in a focused study by Lane *et al.* [16], comparing South African populations, the Venda and Xhosa were shown to cluster apart from one another. In addition, the CMA population is of special interest to population-based studies as this population has been shown to exhibit the highest level of intercontinental admixture observed worldwide [12], with genetic contributions from European, South Asian, Indonesian and Xhosa populations [17].

Therefore, this study performed a detailed analysis of the entire *CYP2C19* gene in the Xhosa population, creating a comprehensive *CYP2C19* genetic profile for these individuals. The information was subsequently used to genotype a cohort of CMA samples, in order to compare the frequencies of *CYP2C19* variants in the Xhosa to another large South African population group. Together, these data could aid in the development of successful pharmacogenetic profiles for other South African populations.

Materials & methods

Patient samples

Institutional approval was granted by the Committee for Human Research of Stellenbosch University (Stellenbosch, South Africa). For DNA collection purposes; written, informed consent was obtained from 100 Xhosa and 75 CMA healthy individuals from the Western Cape region, whose ethnicity was determined via self-report. Genomic DNA was extracted from the venous blood using the Miller *et al.* protocol [18]. A total of 15 of the Xhosa individuals were randomly selected for bidirectional sequencing of *CYP2C19*, in order to detect variants occurring at a frequency of more than 10% in the population with 95.8% certainty (from the binomial distribution). The remaining Xhosa and CMA individuals were genotyped for selected SNPs using PCR-RFLP.

PCR amplification & sequencing reactions

Primers were designed according to the *CYP2C19* DNA reference sequence (Ensembl Gene ID ENSG00000165841) [103] to amplify all nine exons, including at least 30 bp of the adjacent introns, 346 bp of the 3′-UTR and 2095 bp of the 5′-upstream region (refer to Online supplementary Table A for primer sequences; www.futuremedicine.com/doi/suppl/10.2217/pgs.10.90). PCR amplification reactions were prepared to a final volume of 25 µl (refer to Online supplementary Table B for PCR specifications; www.futuremedicine.com/doi/suppl/10.2217/pgs.10.90) and were purified using SureClean (Bioline, London, UK) for subsequent bidirectional semi-automated sequencing, using BigDye® Terminator v3.1 Cycle Sequencing chemistry (Applied Biosystems, CA, USA), Half Dye mix (Bioline) and the addition of 0.2% sodium dodecyl sulfate, at a purification cycle of 98°C for 5 min and 25°C for 10 min to remove unincorporated dye terminators. Capillary electrophoresis was performed by the Central Analytical Facility of Stellenbosch University on a 3130Xl Genetic Analyzer (Applied Biosystems) and the obtained sequencing data were compared with the reference sequence for the identification of variants.

Prioritization & genotyping of variants

The identified variants were prioritized for genotyping in an additional 85 healthy Xhosa and 75 CMA individuals according to the following criteria:

- Nonsynonymous mutations [19,20,104];
- Data obtained from the Human *CYP* Allele Nomenclature website [105];
- *In silico* splice site [21,22], transcription factor binding site [106,23,24] and nucleic acid folding and hybridization prediction (mFold) analyses [25];
- Linkage disequilibrium data from Haploview v3.31 [26].

In addition, the nonfunctional *CYP2C19*3* variant (rs4986893) and upstream *CYP2C19*17* variant (rs11188072) were genotyped in the entire Xhosa and CMA cohorts. All *CYP2C19*2* (rs4244285) variants identified in the Xhosa cohort were genotyped for the adjacent *CYP2C19*10* (rs6413438) variants to confirm which of these two variants were present. The prioritized variants were subsequently genotyped using PCR-RFLP analysis, through the preparation of a 20 µl restriction enzyme digestion reaction mix according to the manufacturer's recommendations (refer to Online supplementary Table C for restriction enzyme digest specifications; www.futuremedicine.com/doi/suppl/10.2217/pgs.10.90).

TaqMan® copy number assays

To detect *CYP2C19* CNVs in the Xhosa cohort, a TaqMan® copy number assay (Applied Biosystems) was performed, using probes Hs02932336_cn (located in intron 6 of *CYP2C19*) and the TaqMan Reference Copy Number Assay (RNase PH1), in a duplex real-time reaction. The assays were prepared to a final volume of 10 µl according to the manufacturer's instructions in a 384-well plate, using an EpMotion® pipetting robot (Eppendorf, Hamburg, Germany), with four replicates for each sample, and loaded onto a 7900HT Fast Real-Time PCR system (Applied Biosystems) at the specified reaction cycle. The results were then analyzed with CopyCaller™ Software (Applied Biosystems), with the 'Most Frequent Sample Copy Number' set as 2 owing to the fact that no deletion or duplication controls are available and the manual cycle threshold (C_T) set to 0.2. All samples used were diluted to 5 ng/µl, with A260:A280 ratios of greater than 1.7, as recommended by the manufacturers for consistency.

Plasmid preparation, cell transfection & dual luciferase reporter assays

Dual luciferase reporter assays were used to determine the effect of selected 5′-upstream variants. After initial amplification of heterozygous *CYP2C19*28* and *CYP2C19*27* samples for the P+E1 fragment, a nested PCR was performed with Invitrogen (CA, USA) reagents (refer to Online supplementary Table A for primer sequences) to obtain a 2059 bp fragment ranging from directly upstream of the ATG start codon to 2095 bp upstream, to include the -2030C>T (ss230394406) and -2020C>A (ss230394407) variants. The promoterless pGL4.10 vector (Promega, WI, USA) [107] and purified fragments were digested with *Bgl*II and *Xho*I, and subsequently ligated, using T4 DNA ligase (Invitrogen). The resulting constructs were then transformed into *E.cloni*® chemically competent cells (Lucigen Corporation, WI, USA) and colonies were grown on LB agar plates (Lucigen Corporation) containing 50 µg/ml ampicilin. White recombinant colonies were selected, the orientation of the inserted fragments was verified using a vector specific primer and the constructs were genotyped for the relevant variants using PCR-RFLP analysis (refer to Online supplementary Table C for restriction enzyme digest specifications), after which all three of the desired fragments (Figure 1) were sequence verified. After extraction

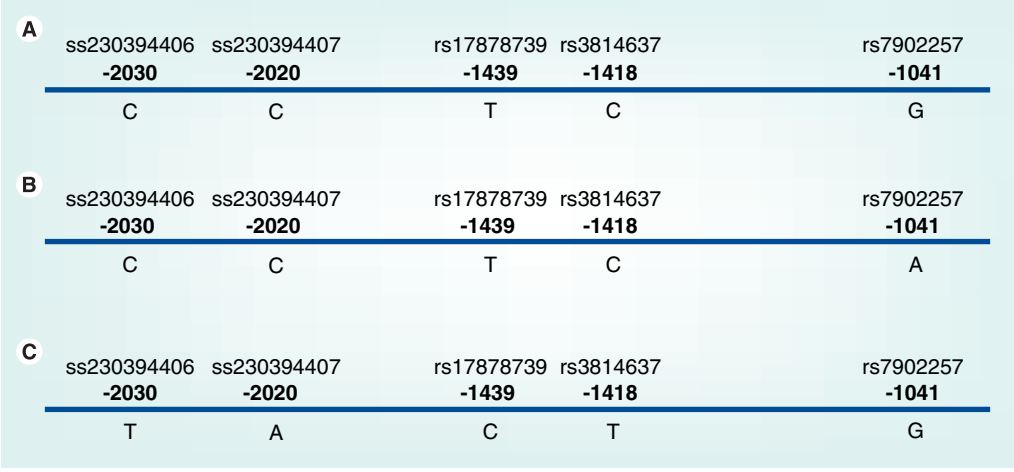

Figure 1. Fragments inserted into pGL4.10 vectors. Fragment A represents the 5′-upstream area of *CYP2C19*1*, fragment B represents the 5′-upstream area of *CYP2C19*27* and fragment C represents the 5′-upstream area of *CYP2C19*28*.

using a GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich [Pty] Ltd, South Africa) the constructs were transfected into HepG2 cells, which were used as representative of cells in liver metabolism *in vitro* owing to the fact that *CYP2C19* is predominantly expressed in the liver [27]. Last, the luminescence was measured on a GloMax™ 96 Plate Luminometer (Promega). All transfection experiments were performed independently, in triplicate, to validate the data obtained.

Statistical analysis

The genotyped SNPs were tested for Hardy–Weinberg equilibrium using an analog to Fisher's exact test in Tools For Population Genetic Analysis (TFPGA) Software v1.3 [108]. The data obtained from the luciferase assays were tested for normality using a Shapiro–Wilks test, and subsequent differences in luciferase activity were tested using a student's t-test. P-values of less than 0.05 were considered significant.

Results

Identification of variants

The resequencing data obtained from the 15 Xhosa individuals revealed 30 variants, all of which were in Hardy–Weinberg equilibrium. Among the detected variants were five novel variants, of which the novel nonsynonymous V374I variant (ss230394410) has been designated *CYP2C19*28* [105]. In addition, the -1041G>A (rs7902257) variant has been designated *CYP2C19*27* [105]. After identification of these variants in the 15 sequenced Xhosa individuals, the variants were prioritized as described in the 'Materials & methods' section and genotyped using PCR-RFLP in the remaining Xhosa and CMA individuals (refer to Table 1 and Online supplementary Figure 1 for the detected and genotyped variants; www.futuremedicine.com/doi/suppl/10.2217/pgs.10.90). *CYP2C19*3*, *CYP2C19*10* and CNVs were not detected in the Xhosa cohort. The calculated copy numbers of the examined samples ranged from 1.71 to 2.21, the ΔCT standard deviation was more than 0.21 and all samples showed confidence values greater than 95%.

Interestingly, the *CYP2C19*17* -3402C>T (rs11188072) variant, which has been reported to occur in perfect linkage disequilibrium with the *CYP2C19*17* -806C>T (rs12248560) variant in all cases with one exception [28], was found to occur in five Xhosa individuals in the absence of the -806C>T variant. These individuals, along with the previously reported African–American individual [28], provide proof for all four haplotypic combinations in populations of African ancestry. After classification of the 100 Xhosa individuals according to metabolizer class, it was observed that PMs, IMs, EMs and UMs occurred at a frequency of 0.03, 0.49, 0.39 and 0.09, respectively, while the frequencies in the CMA population were 0.08, 0.40, 0.35 and 0.17, respectively. The metabolizer class genotype classification system is described in Table 2.

In silico analysis

Transcription factor binding site analysis with three distinct programs consistently predicted the removal of GATA-factor binding sites as a result of the -2030C>T (*CYP2C19*28*) variant. Two of the three programs utilized predicted the removal of a GATA-1 or GATA-2 site (Online supplementary Table D; www.futuremedicine.com/doi/suppl/10.2217/pgs.10.90). The addition of an octamer binding protein-1 (Oct-1) binding site as a result of the -1041G>A (*CYP2C19*27*) variant was only predicted by the MATCH transcription factor binding site predictive program; however, this was the only change that could explain the decrease in expression as a result of the -1041G>A variant (Online supplementary Table E; www.futuremedicine.com/doi/suppl/10.2217/pgs.10.90). Additional *in silico* analyses, such as splice site, mFold and nonsynonymous mutation analysis, did not reveal any significant results for any of the previously uncharacterized variants.

Dual luciferase reporter assays

Statistical analysis showed that there was a significant decrease in the fold induction of the construct containing the *CYP2C19*27* 5′-upstream region (-1041A) when compared with the construct containing the *CYP2C19*1* 5′-upstream region (-1041G; $p = 0.0077$) (Figure 2). Although the construct containing the *CYP2C19*28* 5′-upstream variants showed a trend towards decreased luciferase activity when compared with the *CYP2C19*1* construct, no significant difference in fold induction was observed ($p = 0.0928$) (Figure 2).

Discussion

This study focused on elucidating common *CYP2C19* sequence variation in the South African Xhosa population. Novel variants were identified, and it was revealed that the Xhosa exhibits a unique genetic profile, differing from other populations examined to date, including the South African CMA and Venda.

Table 1. *CYP2C19* variants identified in the Xhosa population and the corresponding frequency comparisons in other populations.

gDNA position	Region	Accession number	Allele	Effect on protein	Effect on enzyme activity	Frequencies						
						Xhosa (n = 15)	*Xhosa (n = 100)*	*CMA (n = 75)*	*Venda (n = 9)* [14]	*Venda (n = 75)* [13]	*Asian (n = 400)* [40]	*Caucasian (n = 283)* [41]
-3402C>T	5′	rs11188072	**17*				0.13	0.14				
- 2030C>T†	5′	ss230394406	**28*				0.09	0.08				
- 2020C>A†	5′	ss230394407	**28*				0.09	0.08				
-1439T>C	5′	rs17878739	**28*				0.09	0.08				
-1418C>T	5′	rs3814637				0.17						
-1041G>A	5′	rs7902257	**27*		Decrease *in vitro*		0.33	0.08			0.00	0.01 [107]
-806C>T	5′	rs12248560	**17*		Increase		0.10	0.14			0.01	0.20
-783C>T	5′	rs11568729				0.07						
-98T>C	5′	rs4986894	**2C*			0.17			0.11			
55A>C	Exon 1	rs17882687	**15/*28*	I>L	Unknown		0.09	0.08	0.00		0.01	
99C>T	Exon 1	rs17885098	**2C*	None	None	0.83			0.17			
183T>C	Intron 1	rs17882201				0.17			0.06			
188C>T	Intron 1	rs17881883				0.03			0.11			
12013T>C	Intron 1	rs17884832				0.10			0.00			
12122G>A	Intron 1	rs7916649	**2C*			0.77			0.28			
12126G>A†	Intron 1	ss230394408				0.03			0.00			
12637C>T†	Intron 2	ss230394409				0.03			0.06			
12662A>G	Intron 2	rs12769205	**2C*			0.20			0.22			
12784 G>A	Exon 3	rs17884712	**9*	R>H	Decrease *in vitro*		0.09	0.04	0.06		0.00	
17948G>A	Exon 4	rs4986893	**3*	Premature stop codon	Nonfunctional		0.00	0.07	0.00	0.00	0.03	0.00
19076T>C	Intron 4	rs57752480				0.07			0.00			
19154G>A	Exon 5	rs4244285	**2C*	Splicing defect	Nonfunctional		0.21	0.17	0.17	0.22	0.25	0.13
19520G>A	Intron 5	rs12571421	**2C*			0.17						
57740C>G	Intron 5	rs4417205	**2C*			0.20			0.22			
79936T>A	Intron 6	rs28399513	**2C*			0.17						
80160C>T	Exon 7	rs3758580	**2C*			0.13			0.17			
80161A>G	Exon 7	rs3758581	**2C/*27/*28*	I>V	None	1.00			1.00			
80290G>A†	Exon 7	ss230394410	**28*	V>I	Unknown		0.01	0.01	0.00		0.00	0.00
87106T>C	Intron 7	rs4917623				0.03			0.06			
87475G>C	Intron 8	rs17880188				0.07			0.06			
87522C>T	Intron 8	rs17885567							0.11			
89909C>T	Intron 8	rs12268020				0.17			0.17			
90011A>G	Intron 8	rs4451645				0.07			0.11			

†*Novel SNPs.*
CMA: Cape Mixed Ancestry; gDNA: Genomic DNA; H: Histidine; I: Isoleucine; L: Leucine; R: Arginine; V: Valine.

Table 2. Metabolizer class genotype classification systems.

Category	Grouping	Score
Allele	*2	0
	*3	0
	*9	0.5
	*1, *15, 27*†, 28*†	1
	*17	1.5
Metabolizer class	Poor metabolizer	0
	Intermediate metabolizer	1
	Extensive metabolizer	2
	Ultrarapid metabolizer	2.5

The two CYP2C19 *gene copies present in each individual are each given a score based on the allele present. The sum of these scores is then calculated to determine the metabolizer class.*
†As the in vivo *functional effect of the novel alleles has not yet been determined, they were classified as functional.*

Where most studies have examined only the *CYP2C19*2* and *CYP2C19*3* variants, this analysis has revealed the presence of other relevant variants in both the Xhosa and CMA populations, which could influence the phenotype of patients. This correlates to data obtained for *CYP2D6* in the Xhosa and CMA populations, where the novel variants *CYP2D6*64*, **65*, **66*, **73* and **74* were detected [10,29]. These data emphasize the requirement for appropriate genotyping platforms in these and other South African populations.

The novel variant *CYP2C19*28*, V374I, does not appear to exert a major effect on the CYP2C19 enzyme, as none of the algorithms used predicted a significant impact on the protein structure. To our knowledge, there are no known data indicating that the substituted amino acid plays an essential role in the active site of the protein; however, these predictions should be verified *in vivo*. The *CYP2C19*27* allele was, however, shown to result in a decrease in luciferase activity. As the area in which this variant occurs has not been shown to be involved in *CYP2C19* expression [30,31], the addition of a repressor transcription factor binding site, namely Oct-1, was the only *in silico* prediction providing a possible explanation for the obtained results. Oct-1 has been shown to repress the expression of the rat *CYP1A1* and *CYP4A2* genes [32,33]; therefore, the role of Oct-1 on *CYP2C19* expression warrants examination in the future.

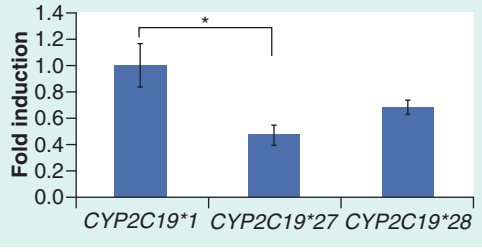

Figure 2. Fold induction ± standard error of the mean of constructs containing *CYP2C19*1*, *CYP2C19*27* and *CYP2C19*28* 5′-upstream variants.
*p = 0.0077.

Although preliminary bioinformatic analyses suggested that -2030C>T results in the removal of a GATA-factor binding site, more specifically GATA-2, which has recently been shown to stimulate *CYP2C19* [34], these results were not validated by the dual luciferase reporter assays. It is important to bear in mind that the other *CYP2C19*28* variants occurring alongside the -2030T variant could have influenced the results obtained for that construct. In addition, other factors, such as additional GATA factors within the 5′ region [34], may compensate for the disruption of this GATA transcription factor binding site.

To our knowledge, this is the first report of *CYP2C19*17* in a South African population. Although this variant has been shown to affect the plasma concentration of several substrates [7,8], it has remained largely neglected in most studies. *CYP2C19*10* (rs6413438) has been detected in the African–American population at a frequency of 0.03 [35] and may thus contribute to the erroneous genotyping of *CYP2C19*2* in other populations of African descent. Nonetheless, the re-examination of all individuals genotyped for *CYP2C19*2* in this study did not reveal the presence of the *CYP2C19*10* allele in the Xhosa population. As *CYP2C19*10* and *CYP2C19*2* are both nonfunctional [36], their misclassification will not impact pharmacogenetic applications. The absence of *CYP2C19*10* in the Xhosa population and the different linkage pattern of the *CYP2C19*17* variants serve as valuable reminders that data from studies involving African–American individuals cannot necessarily be inferred on to other African populations. Similarly, differences in the frequency and presence of variants were observed between the South African Xhosa and Venda populations [14,15], indicating that these two groups may be distinct from each other. Furthermore, all variants genotyped in the CMA population, including the novel and *CYP2C19*3* variants, were detected, emphasizing a high level of admixture and suggesting that additional variants, both previously identified and novel, may be present (refer to Table 1 for frequency comparisons). The comparisons of these populations highlight the need to consider the pharmacogenetic applications of

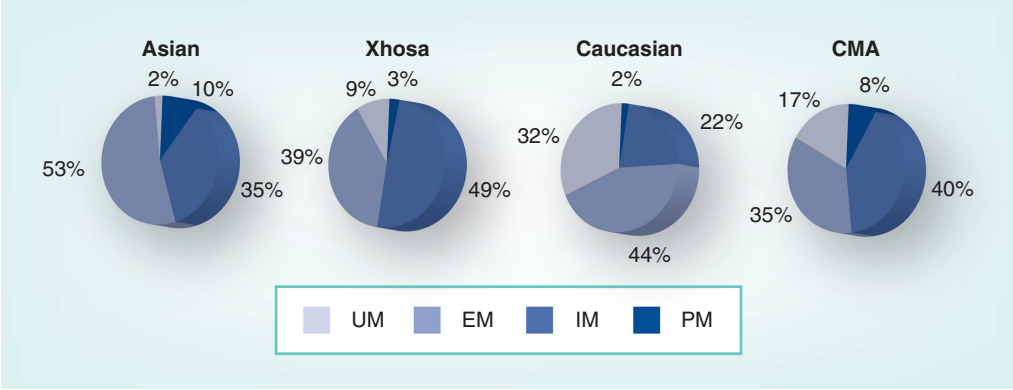

Figure 3. Differences in the frequencies of metabolizer classes observed in the Xhosa, Asian [40], Caucasian [41] and CMA populations (determined as described in Table 2 using data from Table 1). Of note, the Caucasian population was not genotyped for *CYP2C19*9,* which contributed to the number of IMs observed in both the Xhosa and CMA populations. However, this allele has been reported to occur predominantly in individuals of African descent [35]; thus, it is not likely to contribute significantly to the number of IMs observed in the Caucasian population.
CMA: Cape Mixed Ancestry; EM: Extensive metabolizer; IM: Immediate metabolizer; PM: Poor metabolizer; UM: Ultrarapid metabolizer.

these groups independently from one another. When comparing the genotype-inferred metabolizer classes of the Xhosa with the CMA and published data for the Caucasian and Asian populations, the Asians showed the highest frequency of PMs, the Xhosa showed the highest frequency of IMs and the Caucasians showed the highest frequency of UMs. The CMA population appears to form intermediate frequencies of the metabolizer groups (Figure 3), reflecting the high level of admixture observed in this population. The differences in frequencies of metabolizer classes observed between the Xhosa and CMA populations serve, once again, as a reminder that the different populations residing in South Africa may need to be considered independently.

The results obtained from this study may aid in the reduction of ADRs and treatment failure if applied in South Africa. At present, the application of *CYP2C19* pharmacogenetics shows most promise with regards to therapeutic agents such as clopidogrel [37], antimalarial agents [8], proton pump inhibitors and antidepressants [7], as these drugs are utilized in South Africa. The urgent requirement for pharmacogenetic applications in South Africa is highlighted by a study reporting that 14% of all patients admitted to a South African hospital exhibited ADRs, with a fatality rate five- to ten-fold higher than that reported in US and UK populations [38].

Conclusion

The data obtained from this study demonstrate that the *CYP2C19* genetic profile of the Xhosa population differs from that of the Caucasian, Asian, CMA and other African populations. This emphasizes the need to determine the genetic profiles of other South African populations in order to obtain comprehensive pharmacogenetic guidelines for the much needed implementation of optimal treatment plans within the South African context.

Future perspective

DNA sequencing technology is advancing at an exponential rate, allowing for quicker and cheaper analyses of genomes. The challenges inherent in analyses of the highly diverse African genomes will gradually lessen and allow for the identification of rare variants contributing to altered metabolic enzyme function. In future, applicable algorithms will need to be developed to allow for the calculation of the correct dosage and drug to be administered, based on, among other things, genotype, physiological/disease status, environmental factors and concomitant drugs. This may be useful in African countries with the high infection rate of HIV/AIDS, as it has been reported that the CYP enzyme activity of infected individuals is significantly altered [39]. In addition, non-traditional genetic variation such as CNVs, epigenetics and expression profiling, will need to be further considered.

Acknowledgements

We would like to acknowledge the Xhosa individuals for their participation in this study, Ms M Bosman for technical assistance, the Central Analytical Facility of Stellenbosch University for assistance with sequence analyses and Dr O Ikediobi for critical reading of the manuscript.

Executive summary

Identification of variants

- Resequencing of the Xhosa population revealed 30 variants, five of which were novel.

Detection of* CYP2C19 *alleles

- Two novel alleles, designated *CYP2C19*27* and *CYP2C19*28*, were detected in both the Xhosa and Cape Mixed Ancestry populations.
- The previously identified *CYP2C19*2*, *CYP2C19*9*, *CYP2C19*15* and *CYP2C19*17* variants were detected in both populations, while *CYP2C19*3* was only detected in the Cape Mixed Ancestry population.

Dual luciferase reporter assays

- *CYP2C19*27* was shown to cause a twofold decrease in luciferase activity.

Conclusion

- The South African populations examined displayed unique genetic profiles that are likely to affect drug metabolism.
- It remains to be determined whether the high frequency of intermediate metabolizers in these populations contributes to the high number of adverse drug reactions observed.

Financial & competing interests disclosure

We would like to acknowledge the Harry Crossley Foundation and South African National Research Foundation for financial assistance. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Bibliography

Papers of special note have been highlighted as:
- of interest

1 Wester K, Jonnson AK, Sigset O, Druid H, Hagg S: Incidence of fatal adverse drug reactions: a population based study. *Br. J. Clin. Pharmacol.* 65, 573–579 (2008).

2 Allison M: Is personalized medicine finally arriving? *Nat. Biotechnol.* 26, 509–517 (2008).

3 Ingelman-Sundberg M: Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, the present and future. *Trends Pharmacol. Sci.* 25, 193–200 (2004).

4 Oscarson M: Pharmacogenetics of drug metabolizing enzymes: importance for personalised medicine. *Clin. Chem. Lab. Med.* 41, 573–580 (2003).

5 Goldstein JA: Clinical relevance of genetic polymorphisms in the human *CYP2C* subfamily. *Br. J. Clin. Pharmacol.* 52, 349–355 (2001).

6 Andersson T, Flockhart DA, Goldstein DB *et al.*: Drug-metabolizing enzymes: evidence for clinical utility of pharmacogenomic tests. *Clin. Pharmacol. Ther.* 78, 559–581 (2005).

7 Li-Wan-Po A, Girard T, Farndon P, Cooley C, Lithgow J: Pharmacogenetics of *CYP2C19*: functional and clinical implications of a new variant *CYP2C19*17*. *Br. J. Clin. Pharmacol.* 69, 222–230 (2010).

8 Janha RE, Sisay-Joof F, Hamid-Adiamoh M *et al.*: Effects of genetic variation at the *CYP2C19*/*CYP2C9* locus on pharmacokinetics of chlorcycloguanil in adult Gambians. *Pharmacogenomics* 10(9), 1423–1431 (2009).

9 Sistonen J, Fuselli S, Palo JU, Chauhan N, Padh H, Sajantila A: Pharmacogenetic variation at *CYP2C9*, *CYP2C19* and *CYP2D6* at global and microgeographic scales. *Pharmacogenet. Genomics* 19, 170–179 (2009).

▪ **Good review summarizing the varying frequencies of *CYP2C19* alleles in individuals of difference ethnicities.**

10 Wright GEB, Niehaus DJH, Drögemöller BI, Koen L, Warnich L: Elucidation of *CYP2D6* genetic diversity in a unique African population: implications for the future application of pharmacogenetics in the Xhosa population. *Ann. Hum. Genet.* 74(4), 340–350 (2010).

11 Mwenifumbo JC, Zhou Q, Benowitz NL, Sellers EM, Tyndale RF: New *CYP2A6* gene deletion and conversion variants in a population of black African descent. *Pharmacogenomics* 11(2), 189–198 (2010).

12 Tishkoff SA, Reed FA, Friedlaender FR *et al.*: The genetic structure and history of Africans and African Americans. *Science* 324, 1035–1044 (2009).

▪ **Provides evidence for the high level of diversity in African genomes and a need for more thorough genetic analysis of these populations.**

13 Dandara C, Masimirembwa CM, Magimba A *et al.*: Genetic polymorphism of *CYP2D6* and *CYP2C19* in East- and Southern African populations including psychiatric patients. *Eur. J. Clin. Pharmacol.* 57, 11–17 (2001).

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

▪ **Indication of which *CYP2C19* genotypes could be expected to be present in African populations as well as proof that these populations differ from one another.**

15 Ehret C: Bantu expansions: re-envisioning a central problem of early African history. *Int. J. Afr. Hist. Stud.* 34, 5–41 (2001).

16 Lane AB, Soodyall H, Arndt S *et al.*: Genetic substructure in South African Bantu-speakers: evidence from autosomal DNA and Y-chromosome studies. *Am. J. Phys. Anthropol.* 119, 175–185 (2002).

17 Patterson N, Petersen DC, van der Ross RE *et al.*: Genetic structure of a unique admixed population: implications for medical research. *Hum. Mol. Genet.* 19, 411–419 (2010).

18 Miller SA, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16, 1215 (1988).

19 Ng PC, Henikoff S: SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814 (2003).

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

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

22 Dogan RI, Getoor L, Wilbur WJ, Mount SM: SplicePort: an interactive splice-site analysis tool. *Nucleic Acids Res.* 35, 285–291 (2007).

23 Kel AE, Gössling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E: MATCH: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.* 31, 3576–3579 (2003).

24 Quandt K, Frech K, Karas H, Wingender E, Werner T: MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.* 23, 4878–4884 (1995).

25 Zuker M: Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415 (2003).

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

27 Guengerich PF: Human cytochrome P450 enzymes. In: *Cytochrome P450: Structure, Mechanism and Biochemistry. 3rd Edition.* Oritz de Montellano PR (Ed.). Plenum Publishers, NY, USA, 378 (2005).

28 Kearns GL, Leeder JS, Gaedigk A: Impact of the *CYP2C19*17* allele on the pharmacokinetics of omeprazole and pantoprazole in children: evidence for a differential effect. *Drug Metab. Dispos.* 38(6), 894–897 (2010).

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

30 Arefayene M, Skaar TC, Zhao X *et al.*: Sequence diversity and functional characterization of the 5′-regulatory region of human *CYP2C19*. *Pharmacogenetics* 13, 199–206 (2003).

31 Chen Y, Goldstein JA: The transcriptional regulation of the human *CYP2C* genes. *Curr. Drug Metab.* 10, 567–578 (2009).

32 Bhat R, Weaver JA, Sterling KM, Bresnick E: Nuclear transcription factor Oct-1 binds to the 5′ upstream region of *CYP1A1* and negatively regulates its expression. *Int. J. Biochem. Cell Biol.* 28, 217–227 (1996).

33 Fiala-Beer E, Lee AC, Murray M: Regulation of the rat *cyp2A4* gene promoter by c-Jun and octamer binding protein-1. *Int. J. Biochem. Cell. Biol.* 39, 1235–1247 (2007).

34 Mwinyi J, Hofmann Y, Pedersen RS *et al.*: The transcription factor GATA-4 regulates cytochrome P4502C19 gene expression. *Life Sci.* 86, 699–706 (2010).

35 Blaisdell J, Mohrenweiser H, Jackson J *et al.*: Identification and functional characterization of new potentially defective alleles of human *CYP2C19*. *Pharmacogenetics* 12, 703–711 (2002).

36 Rasmussen H, Werge T: Misclassification of allele *CYP2C19*10* as *CYP2C19*2* by a commonly used PCR-RFLP procedure. *Genet. Test.* 12, 57–58 (2008).

37 Ellis KJ, Stouffer GA, McLeod HL, Lee CR: Clopidogrel pharmacogenomics and risk of inadequate platelet inhibition: US FDA recommendations. *Pharmacogenomics* 10, 1799–1817 (2009).

38 Mehta U, Durrheim DN, Blockman M, Kredo T, Gounden R, Barnes KI: Adverse drug reactions in adult medical inpatients in a South African hospital serving a community with a high HIV/AIDS prevalence: prospective observational study. *Br. J. Clin. Pharmacol.* 65, 396–406 (2007).

▪ **Highlights the high prevalence of adverse drug reactions in South Africa, which is influenced among other things, by the HIV/AIDS epidemic. This article provides evidence that better treatment plans are required in South Africa.**

39 Jones AE, Brown KC, Werner RE *et al.*: Variability in drug metabolizing enzyme activity in HIV-infected patients. *Eur. J. Clin. Pharmacol.* 66(5), 475–485 (2010).

40 Chen L, Qin S, Xie J *et al.*: Genetic polymorphism analysis of *CYP2C19* in Chinese Han populations from different geographic areas of mainland China. *Pharmacogenomics* 9, 691–702 (2008).

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

▪ **Demonstrates the importance of genotyping *CYP2C19*17* in populations and, thus, the need for more thorough genotyping panels.**

Websites

101 Ethnologue
www.ethnologue.com

102 Statistics South Africa (Census 2001)
www.statssa.gov.za/census01/HTML/default.asp

103 Ensembl
http://apr2007.archive.ensembl.org/index.html

104 SNPs3D
www.snps3d.org

105 Human Cytochrome P450 (CYP) Allele Nomenclature Committee
www.cypalleles.ki.se

106 Pattern search for transcription factor binding sites
www.gene-regulation.com/cgi-bin/pub/programs/patch/bin/patch.cgi

107 Promega figure display
www.promega.com/figures/popup.asp?partno=e6651&product=pgl4.10%5b%3ci%3eluc2%3c%2fi%3e%5d+vector&fn=4721ma

108 Miller M: Tools for population genetic analyses (TFPGA) 1.3: a Windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by the author (1997)
www.marksgeneticsoftware.net/_vti_bin/shtml.exe/tfpga.htm