

Single nucleotide and copy number polymorphisms of the *SULT1A1* gene in a South African Tswana population group

Authors:

Hlengiwe P. Mbongwa^{1,2}
 Petrus J. Pretorius^{1,2}
 Annemarie Kruger³
 Gerhard Koekemoer⁴
 Carolus J. Reinecke^{1,2}

Affiliations:

¹Centre for Human Metabonomics, North-West University, Potchefstroom, South Africa

²School of Physical and Chemical Sciences, North-West University, Potchefstroom, South Africa

³Africa Unit for Transdisciplinary Health Research, North-West University, Potchefstroom, South Africa

⁴Statistical Consultation Service, North-West University, Potchefstroom, South Africa

Correspondence to:
 Hlengiwe Mbongwa

Email:
 21046379@nwu.ac.za,
 hpmongwa@live.co.za

Postal address:
 Private Bag X6001,
 Potchefstroom 2520,
 South Africa

Dates:
 Received: 07 Aug. 2010
 Accepted: 28 June 2011
 Published: 04 Nov. 2011

How to cite this article:
 Mbongwa HP, Pretorius PJ, Kruger A, Koekemoer G, Reinecke CJ. Single nucleotide and copy number polymorphisms of the *SULT1A1* gene in a South African Tswana population group. *S Afr J Sci.* 2011;107(11/12), Art. #395, 6 pages. <http://dx.doi.org/10.4102/sajs.v107i11/12.395>

© 2011. The Authors.
 Licensee: AOSIS OpenJournals. This work is licensed under the Creative Commons Attribution License.

Previous studies on gene mapping have firmly established variation in segments of the DNA structure amongst different population groups. *SULT1A1* is one of four *SULT1A* genes that maps to the short arm of chromosome 16, and this area has been shown to contain many repetitive sequences and to be highly duplicated. Using the polymerase chain reaction (PCR)-based restriction fragment length polymorphism method, we set out to determine the *SULT1A1* genotype and allele frequency distributions in the largest sample studied to date: a homogeneous South African Tswana population of 1867 individuals from the Prospective Urban and Rural Epidemiological (PURE) study, and found the *SULT1A1*1* and *SULT1A1*2* alleles present at a frequency of 0.68 and 0.32, respectively. This finding corresponded with those obtained for the Black, Caucasian, and mixed-race South African groups reported in previous studies. Next, using a quantitative multiplex PCR method to estimate the *SULT1A1* gene number of copies in 459 subjects of our population, we discovered between one and five copies: 0.65% of the subjects had a single copy (allele deletion) and 60.14% of the subjects had three or more copies. Our findings correspond with an earlier study on a small African-American group, but differ from those based on two Caucasian groups. Whereas the genotype distribution was comparable to the Caucasian groups, there was a significant difference in the number of copies, which indicated a genetic link between Tswana and African-American populations despite differences in cultural lifestyle associated with their geographical location.

Introduction

Biochemical sulphation is part of the phase II detoxification metabolism in humans. Human cytosolic sulphotransferase (SULT) enzymes catalyse the sulphate conjugation of many exogenous substrates, such as drugs and some xenobiotics, but also endogenous substrates, such as thyroid hormones¹ and neurotransmitters.² This biotransformation process uses co-substrate 3'-phosphoadenosine-5'-phosphosulphate to transform the substrates to water-soluble compounds through catalysis by sulphotransferase enzymes,^{3,4} giving rise to the sulphonated product and the 3'-phosphoadenosine 5'-phosphate.⁵ This reaction usually is effective in rendering the xenobiotic compound less biologically active and stable enough to be excreted from the body.² However, the reaction sometimes plays a role in the bioactivation of xenobiotics.⁶ Population studies on *SULT1A1* polymorphisms have been extensive, as it has been shown that compromised sulphonation of xenobiotics by the less active *SULT1A1*2* enzyme may predispose an individual to cancer, especially in combination with other risk factors such as advanced age, smoking and a poor diet.^{2,6,7,8,9,10,11,12,13}

The genotype distribution covers only a partial aspect of genetic diversity; the copy number variation should also be considered. The study of human genetic variation at the DNA level constitutes a major challenge,^{14,15} ranging from large, microscopically visible chromosome anomalies to single nucleotide changes.¹⁶ Deletions, insertions, duplications and complex multisite variants,^{16,17,18} collectively termed copy number variations (CNVs) or copy number polymorphisms (CNP), are found in all humans.¹⁹ Often, CNPs and CNVs are associated with disease, and also with the evolution of the genome itself and have been observed for a number of genes encoding drug metabolising enzymes.²⁰ The extent to which large duplications and deletions contribute to human genetic variation and diversity is largely unknown.¹⁹ The dominating type of variation explored so far in the genome has been single nucleotide polymorphisms (SNPs), overshadowing the issue of CNPs (gains and deletions).¹⁵ The importance of genotype distribution of the *SULT1A1* gene is well documented. However, studies carried out previously on South African subjects contained individuals from different population groups (Black and mixed-race groups)²¹ and only a limited number of cases were included in a study on Nigerian and African-American groups.^{7,11} We set out to observe the *SULT1A1* genotypes and CNPs in a



South African Tswana population group. In this study, we aimed to provide complementary details to the recently published comprehensive study on the genome of a single individual, Archbishop Emeritus Desmond Tutu, from the Sotho–Tswana population group.²²

Materials and methods

Subjects

This investigation forms part of an international study, the Prospective Urban and Rural Epidemiological (PURE) programme, which is a prospective cohort study that tracks changing lifestyles, risk factors and chronic disease using periodic standardised data collected in urban and rural areas of 16 countries in transition (developing countries). Full details of the PURE study were published on behalf of all participating scientific groups.²³ Ethical approval to participate in the South African component of the PURE (PURE-SA) study was obtained from the North-West University Ethics Committee (reference approval number 04M10). A total of 2010 apparently healthy African volunteers of Tswana ancestry gave informed consent to participate in the baseline study of PURE-SA in 2005 (1002 were from an urban setting around Potchefstroom and 1008 were from three rural villages ~450 km from Vryburg en route to Botswana). The South African population is quite heterogeneous in terms of ethnic and societal characteristics. The experimental group may, however, be regarded as relatively homogeneous, as all participants were of the same ethnic group and complied with the PURE definition of a community as ‘a group of people who have common characteristics and reside in a defined geographical area’²³.

DNA isolation from blood

Blood was collected intravenously into vacutainer tubes containing ethylenediaminetetraacetic acid (EDTA). Sufficient blood was available for the successful DNA analysis of 1867 of the original 2010 experimental subjects of the PURE study. The FlexiGene DNA kit (Qiagen, Hilden, Germany) was used to isolate DNA from blood. DNA was quantified using a Nano-Drop ND 1000 UV/Vis Spectrophotometer (Nano-Drop Technologies, Wilmington, DE, USA) and was kept at –20 °C until further use.

SULT1A1 genotyping

Genotypes for the *SULT1A1**2 polymorphism (Arg213His, rs9282861) were carried out by using a set of primers, 5'-GTT GGC TCT GCA GGG TTT CTA GGA-3' (forward) and 5'-CCC AAA CCC CCT GCT GGC CAG CAC CC-3' (reverse), which were designed to amplify a region in an intron flanking exon VII as described by Coughtrie et al.⁷ The primers did not contain a restriction enzyme recognition site. Amplification of DNA was performed in a 50- μ L incubation mixture comprising 10 pmol of each primer, 200 μ M dNTPs, 2 mM MgCl₂, 1 unit GoTaq (Promega, Madison, WI, USA), DNA polymerase, ~100 ng of genomic DNA and 1X Flexi

buffer (Promega, Madison, WI, USA). Initial denaturation was at 94 °C for 5 min. This initial denaturation was followed by 39 cycles of denaturation (at 94 °C for 30 s), annealing (at 63.4 °C for 30 s) and extension (at 72 °C for 30 s). The final extension was at 72 °C for 10 min. PCR performed under these conditions results in the specific amplification of a 333-bp fragment of DNA which includes exon VII of the *SULT1A1* gene. PCR reaction products (5 μ L) were digested with 2 units of HaeII (New England BioLabs, Ipswich, MA, USA) restriction enzyme for 90 min at 37 °C in a 20 μ L reaction mixture containing 1X NEBuffer 4 (pH 7.9; New England BioLabs, Ipswich, MA, USA) and 100 μ g/mL bovine serum albumin. After digestion, 9 μ L of digestion product was resolved on 3% (w/v) agarose gel for 1 h in a 1X TAE buffer (made in-house) (pH 8.5). The 333-bp fragment yielded two fragments of 168 bp and 165 bp, after digestion with HaeII.

The wildtype allele frequency distribution was calculated according to Snyder et al.²⁴ Allele frequency of AA = [(2 x AA) + AH]/2 x N and HH = 1 – AA. Where AA (Arg/Arg) is the total cases of *SULT1A1**1 form, AH (Arg/His) is the total cases of the *SULT1A1**1/*2 form, HH (His/His) is the total cases of *SULT1A1**2 form and N is the total number of cases.

SULT1A1 gene copy number

Because of the high costs of copy number analysis, not all 1867 PURE samples could be analysed for this part of the study. Hence, a proportional stratified sample of 459 was selected to investigate *SULT1A1* gene CNP. The stratification was based on gender, area and genotype. The *SULT1A1* copy number was estimated by calculating the height and area ratio of the 212 bp amplicon of *SULT1A1* to the reference 208 bp amplicon of *SULT1A2* as described by Hebring et al.²⁰ The set of PCR primers used were designed, as per Hebring et al.²⁰, to co-amplify a 212-bp fragment within exons 2 and 3 of *SULT1A1* and a 208-bp fragment within exons 3 and 4 of *SULT1A2*. In this case, *SULT1A2* was used as an internal control for copy number. Amplification of the required DNA fragments from ~25 ng genomic DNA was performed in a 25- μ L incubation mixture comprising 0.4 μ M each of *SULT1/2* forward (5'-6FAM/TCA CCG AGC TCC CAT CTT-3'), *SULT1/2* reverse (5'-GTTT GGG GCA GGT GTG TCT TCAG-3'), *Factor V* forward (5'-6FAM/ATG GAC TTC CAC ATT AGG GAC-3') and *Factor V* reverse (5'-GTTT GAA GGT AGT GGA TTC TCC ATCA-3') primers; 200 μ M dNTPs; 2 mM MgCl₂; 1 unit GoTaq DNA polymerase and 1X Flexi buffer (pH 9.0). Initial denaturation was at 95 °C for 10 min. This was followed by 29 cycles of denaturation (at 95 °C for 30 s), annealing (at 58 °C for 30 s), and extension (at 72 °C for 30 s). The final extension was at 72 °C for 10 min.

PCR product (1 μ L) was diluted a 100X with high-performance liquid chromatography-grade purified water, and 1 μ L of this dilution was added to a 9- μ L mixture of Hi-Di formamide and GeneScan 500 LIZ size standard (15 μ L LIZ size standard for every 1 mL of Hi-Di formamide; both Applied Biosystems, Foster City, CA, USA) in a PCR

microtitre well plate. After loading samples, the plate was incubated at 94 °C for 5 min (denaturation). Thereafter the microtitre well plate was centrifuged at 1500 rpm for 30 s at room temperature, and the samples were analysed further in a Hitachi 3130xl genetic analyser (Applied Biosystems, Foster City, CA, USA) according to the conditions recommended by the manufacturer. The peak heights between 100 and 7500 (arbitrary units) were accepted as the minimum and maximum (samples whose peak heights were higher than 7500 were re-analysed but diluted 150X, and samples whose peak heights were lower than 100 were re-analysed but diluted 75X). The samples were analysed in duplicate.

Statistical analyses

If X_1 , X_2 and X_3 indicate the number of times each of these genotypes are observed over the n trials, then $X = (X_1, X_2, X_3)$ follows a multinomial distribution with parameters n and $\pi = (\pi_1, \pi_2, \pi_3)$, where π_1, π_2 and π_3 are the true (population) occurrence probability of the three genotypes and $\pi_1 + \pi_2 + \pi_3 = 1$. The estimates of these probabilities were calculated and the method proposed by Goodman²⁵ was used to construct the 95% simultaneous confidence intervals for $\pi = (\pi_1, \pi_2, \pi_3)$. Chi-square dependency tests were also performed to determine if the distribution of the copy number differed between Caucasian and African populations. Histograms and bar charts are used for graphical presentation. SPSS version 18,²⁶ Statistica version 9,²⁷ R²⁸ and Excel²⁹ statistical software programs were used for the statistical analyses.

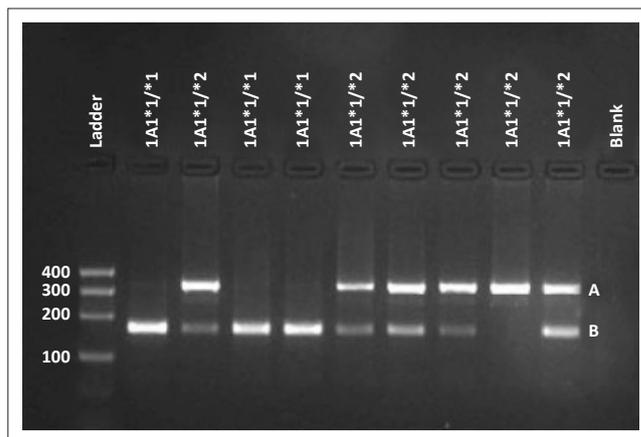
Results

SULT1A1 genotyping

The *SULT1A1* gene is one of the three closely related sulphotransferase genes located on chromosome 16p11.2-12.1.^{8,30} Digestion does not take place with DNA fragments amplified from *SULT1A1* alleles harbouring the CGC→CAC polymorphism at codon 213, as this alters the restriction site recognition sequence for HaeII (Figure 1). The wildtype, heterozygous and homozygous genotype groups were found to be 43.76%, 47.13% and 9.11% with 95% simultaneous confidence intervals given as [41.03; 46.52], [44.38; 49.90] and [7.63; 10.83], respectively. The wildtype allele frequency distribution was found to be 0.68 for the Tswana group.

Copy number estimation

The distribution of the calculated *SULT1A1/1A2* ratio values, which was used to estimate the number of *SULT1A1* gene copies, is shown in Figure 2. The copy number estimation in 459 subjects was as follows: 0.65% demonstrated a deletion within the *SULT1A1* gene (which had one copy), and 60.14% had three or more copies, with 95% simultaneous confidence intervals given as [0.17; 2.55] and [48.83; 74.00], respectively (Table 1). It has been shown that CNV is highly replicable and that biallelic CNVs may present in various classes, ranging from simple to complex profiles.



A, 333 bp.
B, 168 bp and 165 bp.

FIGURE 1: An electrogram showing representative results of a polymerase chain reaction-based restriction fragment length polymorphism assay to indicate the mutation in exons VII of *SULT1A1*. The genotypes represented are 1A1*1/*1 (both alleles are wildtype), 1A1*1/*2 (one allele is wildtype and the other is a mutant), 1A1*2/*2 (both alleles are mutants), a 400-bp ladder and a blank (no template).

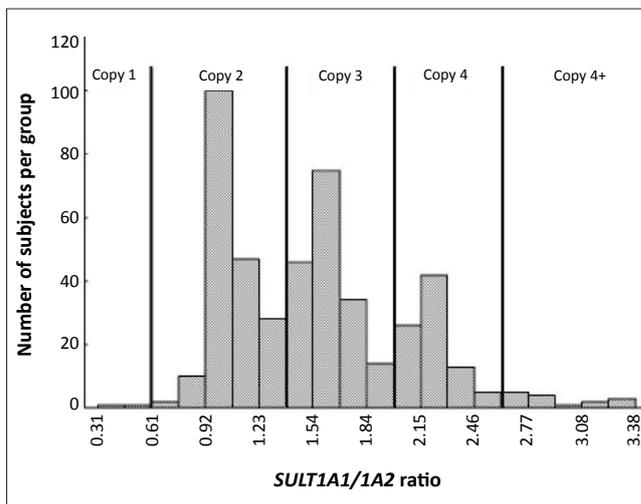


FIGURE 2: A histogram showing the copy number estimation for the 1867 subjects from the Tswana group studied. The samples from individuals with an A1/A2 ratio ≤ 0.6 ; >0.6 and ≤ 1.4 ; >1.4 and ≤ 2.0 ; >2.0 and ≤ 2.6 ; and >2.6 were defined as having a *SULT1A1* copy number of 1, 2, 3, 4, and 4+, respectively.

Discussion

The *SULT1A1*2* allele contains a non-synonymous polymorphism and has been extensively reported to occur in individuals in various populations (Table 2).^{7,9,10,11,21,31,32,33} We used PCR-based restriction fragment length polymorphism, a method known to observe the *SULT1A1* genotype distribution in a much bigger (1867), homogeneous South African population group. Availability of nearly 2000 samples from a well-defined profile for each participating individual brought with it uniqueness to this study. From the results presented in Table 2 there seems to be a distinct pattern that populations follow, depending on ancestry. Asian populations show a much higher wildtype genotype frequency when compared to the rest, therefore showing a much lower mutant allele frequency compared to Caucasian populations and African populations. A previous study by

TABLE 1: The number and percentage of subjects with one or more *SULT1A1* copies and the 95% simultaneous confidence interval distribution of the *SULT1A1* copy number.

<i>SULT1A1</i> copy number	Number of subjects ^a	Percentage of total	95% Simultaneous confidence interval
1	3	0.65	[0.17; 2.55]
2	182	39.65	[33.96; 45.64]
3	174	37.91	[32.29; 43.87]
4	85	18.52	[14.31; 23.63]
4+	15	3.27	[1.71; 6.16]

^a, Per copy number variation group.

Cann et al.³⁴ on mitochondrial DNA concluded that genetic lineage for human origins is in Africa, although the argument is complex, as noted in the study on the genome of the southern African Khoisan and Bantu.²²

A study by Dandara et al.²¹ focused on mixed-race and Black South African populations. Their Black subjects were mainly Xhosa-speaking, and came originally from either the Transkei area of the Eastern Cape or from the Western Cape. The mixed-ancestry individuals are a mix of nationalities resulting from gene flow amongst Black South Africans, Western Europeans, the San, the Khoikhoi, Indonesians and Malaysians who settled in the Cape from the 17th century onwards. The homozygous genotype (Table 2) in mixed-race populations compares relatively closely to the Black South African population but differs to that obtained for the Tswana population. *SULT1A1**2 polymorphism has consequences for the sulphonation of drugs and carcinogens.

We calculated 95% simultaneous confidence intervals for this study, as well as for the findings from previous studies (Table 2). This calculation gives lower and upper limits that can be used as a guide to how close the findings are to the true distribution of the population. In general, larger sample sizes produce lower and upper limits that are closer together.

Copy number polymorphisms are found in all humans¹⁹; results from previous studies are summarised in Table 3. From the results shown in Table 3, we can conclude that the

two Caucasian groups and the two African groups do not differ in distribution. This lack of difference was verified by using a Chi-square dependency test, which produced *p*-values of 0.651 and 0.72 for the Caucasian and African groups, respectively. We subsequently pooled the Caucasian groups and the African groups to test for a difference in distribution between the Caucasian and African groups, which revealed a *p* < 0.001. This pooled distribution is presented graphically in Figure 3.

These copy number findings confirm results from a much smaller African-American study, and suggests a possible genetic link between the African Tswana and the heritage of the African-Americans. The number of copies demonstrated by each subject does not seem to be conditioned by geography, but could be hereditary.

Copy number variations influence gene expression, phenotypic variation and adaptation by disrupting genes and altering gene dosage, and can cause disease, as in microdeletion or microduplication disorders, or confer risk to complex disease traits. CNVs often represent an appreciable minority of causative alleles at genes at which other types of mutation are strongly associated with specific diseases. Furthermore, CNVs can influence gene expression indirectly through position effects, predispose to deleterious genetic changes, or provide substrates for chromosomal change in evolution.^{16,35} Because many CNVs include genes that result in differential levels of gene expression, CNVs may account for a significant proportion of normal phenotypic variation.^{36,37}

Information on SNPs and CNPs clearly has a clinical significance. Several studies have established that interindividual and interethnic polymorphic differences are important in responses to factors like chemical exposure and drug treatment, and often relate to the detoxification ability of individuals.^{38,39} In this regard, it has been shown that the *SULT1A1* polymorphism has been associated with

TABLE 2: *SULT1A1* genotype (mean [95% confidence interval]) and allele frequencies in previously studied populations.

Population	Genotype						<i>n</i>	Allele frequency		
	<i>SULT1A1</i> *1		<i>SULT1A1</i> *1/*2		<i>SULT1A1</i> *2			<i>SULT1A1</i> *1	<i>SULT1A1</i> *2	Number of chromosomes tested
	Mean	95% CI	Mean	95% CI	Mean	95% CI				
Caucasian ⁷	46.7	[39.9; 53.7]	42.3	[35.6; 49.3]	10.9	[7.3; 16.1]	293	0.679	0.321	586
Caucasian ¹¹	35.2	[29.0; 41.9]	46.8	[40.1; 53.6]	18.1	[13.4; 23.9]	310	0.585	0.415	620
Caucasian ¹⁰	39.9	[35.5; 44.6]	45.5	[40.9; 50.1]	14.6	[11.6; 18.1]	666	0.627	0.373	1332
Caucasian ³¹	47.7	[38.2; 57.3]	37.1	[28.3; 46.8]	15.2	[9.5; 23.5]	151	0.662	0.338	302
Japanese ⁹	69.2	[59.4; 77.6]	28	[20.0; 37.7]	2.8	[0.9; 8.3]	143	0.832	0.168	286
Chinese ³²	83.3	[78.6; 87.2]	14.3	[10.7; 18.9]	2.4	[1.1; 4.8]	426	0.905	0.095	852
Japanese ³³	70.9	[59.3; 80.3]	25.2	[16.5; 36.6]	3.9	[1.3; 11.4]	103	0.835	0.165	206
Nigerian ⁷	53.9	[37.8; 69.2]	38.5	[24.2; 55.0]	7.7	[2.5; 21.3]	52	0.731	0.269	104
African-American ¹¹	49.5	[37.5; 61.5]	44.1	[32.5; 56.4]	6.5	[2.5; 15.4]	93	0.715	0.285	186
Black South African ²¹	46.5	[37.7; 55.6]	33.1	[25.2; 42.2]	20.4	[14.0; 28.6]	172	0.631	0.369	344
Mixed-ancestry South African ²¹	55.3	[43.1; 66.9]	30.9	[20.8; 43.1]	13.8	[7.4; 24.4]	94	0.707	0.293	188
Tswana (present study)	43.8	[41.0; 46.5]	47.1	[44.4; 49.9]	9.1	[7.6; 10.8]	1867	0.673	0.327	3734

Note: Please see the full reference list of the article, Mbongwa HP, Pretorius PJ, Kruger A, Koekemoer G, Reinecke CJ. Single nucleotide and copy number polymorphisms of the *SULT1A1* gene in a South African Tswana population group. *S Afr J Sci.* 2011;107(11/12), Art. #395, 6 pages. <http://dx.doi.org/10.4102/sajs.v107i11/12.395>, for more information. CI, confidence interval.

TABLE 3: Frequency of the *SULT1A1* copy number differences (mean [95% confidence interval]) in four population study groups.

Group of populations	Study	n	Percentage presence of the allelic form									
			1 allele		2 alleles		3 alleles		4 alleles		>4 alleles	
			Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
Caucasian	Caucasian–American ²⁰	362	4.7	[2.6; 8.5]	69.6	[63.1; 75.4]	21.0	[16.0; 27.0]	3.6	[1.8; 7.1]	1.1	[0.3; 3.6]
	Caucasian–Norwegian female breast cancer patients ²¹	151	5.3	[2.2; 12.2]	64.9	[54.5; 74.1]	21.9	[14.5; 31.6]	6.0	[2.6; 13.0]	2.0	[0.5; 7.5]
African	African–American ²⁰	99	0.0	[0.0; 6.3]	37.4	[26.0; 50.3]	37.4	[26.0; 50.3]	21.2	[12.6; 33.4]	4.0	[1.2; 12.6]
	South African Tswana (present study)	459	0.7	[0.2; 2.6]	39.2	[33.5; 45.2]	40.3	[34.6; 46.3]	16.3	[12.4; 21.3]	3.5	[1.9; 6.4]

Note: Please see the full reference list of the article, Mbongwa HP, Pretorius PJ, Kruger A, Koekemoer G, Reinecke CJ. Single nucleotide and copy number polymorphisms of the *SULT1A1* gene in a South African Tswana population group. *S Afr J Sci.* 2011;107(11/12), Art. #395, 6 pages. <http://dx.doi.org/10.4102/sajs.v107i11/12.395>, for more information. CI, confidence interval.

an increased risk of breast and oesophageal cancers.^{21,31,40} Hebbing et al.²⁰ have documented the presence of a common gene duplication or deletion event in *SULT1A1*. It was shown that individuals carrying additional copies of *SULT1A1* represented ‘rapid sulphators,’ whilst those with fewer copies were considered to be ‘slow sulphators’ when the *SULT1A1* enzyme activity was tested on liver and platelet homogenates. We are currently investigating the activity of *SULT1A1* enzyme activity on platelet homogenates. Preliminary results suggest that an increase in copy number benefits individuals of all *SULT1A1* genotypes in terms of sulphation abilities. The pharmacogenetic implications of differences in *SULT1A1* gene dosage might help explain individual differences in drug toxicity and/or efficacy in the clinical setting (Hebbing et al.²⁰).

Conclusion

Our results on the *SULT1A1* allele frequency distribution and copy number estimation per individual were from the largest and most well-defined population group studied thus far and provide data for a comparative view on the *SULT1A1* SNPs and CNPs between different population groups (i.e. Tswana, Black and mixed-race). Our findings, together with those by Dandara et al.²¹ on Black and mixed-race South Africans, bring new information on the *SULT1A1* allele frequency distribution regarding South African

populations. Identification of susceptible genes facilitates our understanding of the disease process and helps in the discovery of improved medicines and therapeutic regimens to match the target population.^{20,41} Also, our findings further correspond with those of an earlier study on a small African–American population in the USA, but differ from those on two Caucasian populations. Whereas the allele frequency distributions were comparable in the Caucasian population and our population, there was a significant difference in the number of copies, indicating a genetic link, despite differences in cultural lifestyle associated with geographical location.

Acknowledgements

We are grateful for financial assistance for this study from the National Research Foundation (Scarce Skills PhD Scholarship awarded to HPM), the North-West University Postgraduate Fund and the Biotechnology Partnership and Development (BioPAD) Project BPP007. Professor Scott Hebbing (Mayo Clinic College of Medicine, Rochester, MN, USA) provided the copy number estimation primer sequences, and his advice on the copy number estimation analysis is greatly appreciated. Finally, we thank the people who participated in the PURE study.

Competing interests

We declare that we have no financial or personal relationships which may have inappropriately influenced us in writing this article.

Authors’ contributions

H.P. Mbongwa performed the experiments and wrote the manuscript. C.J. Reinecke and P.J. Pretorius were project leaders and advised on the experimental and project design. G. Koekemoer performed the statistical analyses. A. Kruger approved our participation in the PURE study and gave access to the blood samples. As corresponding author, I confirm that each named author contributed to the final manuscript and approved the manuscript submitted for publication.

References

- Visser TJ, Kaptein E, Glatt H, Bartsch I, Hagen M, Coughtrie MWH. Characterization of thyroid hormone sulfotransferases. *Chem-Biol Interact.* 1998;109:279–291. [http://dx.doi.org/10.1016/S0009-2797\(97\)00139-7](http://dx.doi.org/10.1016/S0009-2797(97)00139-7)

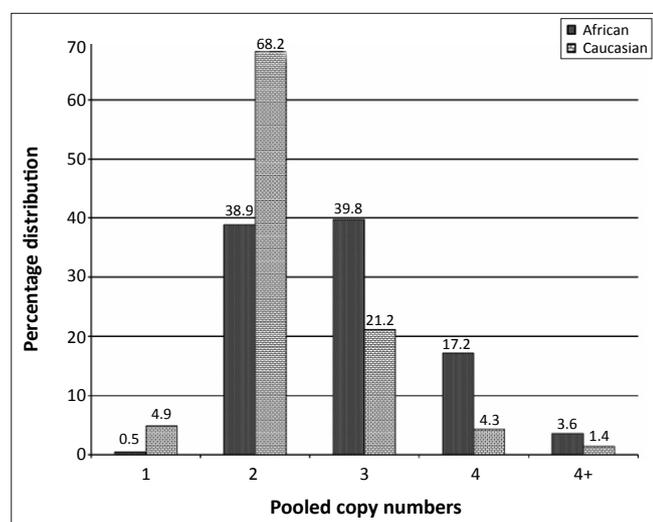


FIGURE 3: A bar graph showing the pooled distributions of the Caucasian and African populations for copy numbers 1 to 4+.



2. Hempel N, Gamage N, Martin JL, McManus ME. Human cytosolic sulfotransferase *SULT1A1*. *Int J Biochem Cell Biol*. 2007;39:685–689. <http://dx.doi.org/10.1016/j.biocel.2006.10.002>
3. Negishi M, Pedersen LG, Petrotchenko E, et al. Structure and function of sulfotransferases. *Arch Biochem Biophys*. 2001;390:149–157. <http://dx.doi.org/10.1006/abbi.2001.2368>, PMID:11396917
4. Rath VL, Verdugo D, Hemmerich S. Sulfotransferase structural biology and inhibitor discovery. *Drug Discov Today*. 2004;9:1003–1011. [http://dx.doi.org/10.1016/S1359-6446\(04\)03273-8](http://dx.doi.org/10.1016/S1359-6446(04)03273-8)
5. Taskinen J, Ethell B, Lautala P, Hood AM, Burchell B, Coughtrie MWH. Conjugation of catechols by recombinant human sulfotransferases, UDP-glucuronosyltransferases, and soluble catechol O-methyltransferase: Structure-conjugation relationships and predictive models. *Drug Met Dispos*. 2003;31:1187–1197. <http://dx.doi.org/10.1124/dmd.31.9.1187>, PMID:12920175
6. Coughtrie MWH. Sulfation through the looking glass – recent advances in sulfotransferase research for the curious. *Pharmacogenomics*. 2002;2:297–308. <http://dx.doi.org/10.1038/sj.tpj.6500117>, PMID:12439736
7. Coughtrie WHC, Gilissen RAHJ, Shek B, et al. Phenol sulphotransferase *SULT1A1* polymorphism: Molecular diagnosis and allele frequencies in Caucasian and African populations. *Biochem J*. 1999;337:45–49. <http://dx.doi.org/10.1042/0264-6021:3370045>, PMID:9854023, PMCID:1219934
8. Weinshilboum R. Phenol sulfotransferase inheritance. *Cell Mol Neurobiol*. 1988;8(1):27–34. <http://dx.doi.org/10.1007/BF00712908>, PMID:3042142
9. Ozawa S, Shimizu M, Katoh T, et al. Sulfating-activity and stability of cDNA expressed allozymes of human phenol sulfotransferase, ST1A3*1 ((213)Arg) and ST1A3*2 ((213)His), both of which exist in Japanese as well as Caucasians. *Biochemistry*. 1999;126:271–277.
10. Sun XF, Ahmadi A, Arbman G, Wallin A, Askild D, Zhang H. Polymorphisms in sulfotransferase 1A1 and glutathione S-transferase P1 genes in relation to colorectal cancer risk and patients' survival. *World J Gastroenterol*. 2005;11:6875–6879. PMID:16425401
11. Nowell S, Ratnasingham DL, Ambrosone CB, et al. Association of *SULT1A1* phenotype and genotype with prostate cancer risk in African-Americans and Caucasians. *Cancer Epidemiol Biomarkers*. 2004;13:270–276. <http://dx.doi.org/10.1158/1055-9965.EPI-03-0047>
12. Glatt HR. Bioactivation of mutagens via sulfation. *FASEB J*. 1997;11:314–321. PMID:9141497
13. Nowell S, Sweeney C, Winters M, et al. Association between sulfotransferase 1A1 genotype and survival of breast cancer patients receiving tamoxifen therapy. *Nat Cancer Instit*. 2002;94:1635–1640.
14. Buckley PG, Mantripragada KK, Piotrowski A, De Stahl TD, Dumanski JP. Copy-number polymorphisms: Mining the tip of an iceberg. *Trends Genet*. 2005;21:315–317. <http://dx.doi.org/10.1016/j.tig.2005.04.007>, PMID:15922827
15. McCarroll SA, Hadnott TN, Perry GH, et al. Common deletion polymorphism in human genome. *Nature Genet*. 2006;38:86–92. <http://dx.doi.org/10.1038/ng1696>
16. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. 2006;444:444–454. <http://dx.doi.org/10.1038/nature05329>, PMID:17122850, PMCID:2669898
17. Fredman D, White SJ, Potter S, Eichler EE, Den Dunnen JT, Brookes AJ. Complex SNP-related sequence variation in segmental genome duplications. *Nature Genet*. 2004;36(8):861–866. <http://dx.doi.org/10.1038/ng1401>, PMID:15247918
18. Feuk L, Marshall CR, Wintle RF, Scherer SW. Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum Mol Gen*. 2006;15(1):R57–R66. <http://dx.doi.org/10.1093/hmg/ddl057>, PMID:16651370
19. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. *Science*. 2004;305:525–527. <http://dx.doi.org/10.1126/science.1098918>, PMID:15273396
20. Hebringer SJ, Adjei AA, Baer JL, et al. Human *SULT1A1* gene: Copy number differences and functional implications. *Hum Mol Gen*. 2007;16:463–470. <http://dx.doi.org/10.1093/hmg/ddl468>, PMID:17189289
21. Dandara C, Li DP, Walther G, Parker MI. Gene-environment interaction: The role of *SULT1A1* and *CYP3A5* polymorphisms as risk modifiers for squamous cell carcinoma of the oesophagus. *Carcinogenesis*. 2006;27(4):791–797. <http://dx.doi.org/10.1093/carcin/bgi257>, PMID:16272171
22. Schuster SC, Miller W, Ratan A, et al. Complete Khoisan and Bantu genomes from southern Africa. *Nature Lett*. 2010;463:943–947. <http://dx.doi.org/10.1038/nature08795>, PMID:20164927
23. Teo K, Chow CK, Vaz M, Rangarajan S, Yusuf S. The Prospective Urban Rural Epidemiology (PURE) study: Examining the impact of societal influences on chronic noncommunicable diseases in low-, middle-, and high-income countries. *Am Heart J*. 2009;158:1–8. <http://dx.doi.org/10.1016/j.ahj.2009.04.019>, PMID:19540385
24. Population genetics. In: Snyder LA, Freifelder D, Hartl DL, editors. *General genetics*. Boston, MA: Jones and Bartlett, 1985: 494–528.
25. Goodman LA. On simultaneous confidence intervals for multinomial proportions. *Technometrics*. 1965;7:247–254. <http://dx.doi.org/10.2307/1266673>
26. Statistical Package for the Social Sciences (SPSS). Version 18. Armonk, NY: IBM; 2009.
27. Statistica. Version 9. Tulsa, OK: StatSoft; 2009.
28. R. Version 2.12.0. Vienna: R Development Core Team; 2010.
29. Windows Excel 7. Redmond, WA: Microsoft Corporation; 2009.
30. Hempel N, Masahiko N, McManus ME. Human *SULT1A* genes: Cloning and activity assays of the *SULT1A* promoters. *Methods Enzymol*. 2005;400:147–165. [http://dx.doi.org/10.1016/S0076-6879\(05\)00009-1](http://dx.doi.org/10.1016/S0076-6879(05)00009-1)
31. Gjerde J, Hauglid M, Breilid H, et al. Effects of *CYP2D6* and *SULT1A1* genotypes including *SULT1A1* gene copy number on tamoxifen metabolism. *Ann Oncol*. 2007;19(1):56–61. <http://dx.doi.org/10.1093/annonc/mdm434>, PMID:17947222
32. Han DF, Zhou X, Hu MB, et al. Sulfotransferase 1A1 (*SULT1A1*) polymorphism and breast cancer risk in Chinese women. *Toxicol Lett*. 2004;150:167–177. <http://dx.doi.org/10.1016/j.toxlet.2004.01.012>, PMID:15093672
33. Ohtake E, Kakihara F, Matsumoto N, et al. Frequency distribution of phenol sulfotransferase 1A1 activity in platelet cells from healthy Japanese subjects. *Eur J Pharm Sci*. 2006;28:272–277. <http://dx.doi.org/10.1016/j.ejps.2006.02.008>, PMID:16621480
34. Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. *Nature*. 1987;325:31–36. <http://dx.doi.org/10.1038/325031a0>, PMID:3025745
35. Korbel JO, Kim PM, Chen X, et al. The current excitement about copy-number variation: How it relates to gene duplications and protein families. *Curr Opin Struct Biol*. 2008;18:1–9. <http://dx.doi.org/10.1016/j.sbi.2008.02.005>, PMID:18511261, PMCID:2577873
36. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. *Nat Genet Suppl*. 2007;39:S37–S42.
37. Freeman JL, Perry GH, Feuk L, et al. Copy number variation: New insights in genome diversity. *Genome Res*. 2006;16:949–961.
38. Desta Z, Zhao X, Shin JG, Flockhart DA. Clinical significance of the cytochrome P450 2C19 genetic polymorphism. *Clin Pharmacokinet*. 2002;41:913–958.
39. Devi SS, Vinayagamoorthy N, Agrawal M, et al. Distribution of detoxifying genes polymorphism in Maharastrian population of central India. *Chemosphere*. 2008;70:1835–1839.
40. Dalhoff K, Buus JK, Enghusen PH. Cancer and molecular biomarkers of phase 2. *Methods Enzymol*. 2005;400:618–627. [http://dx.doi.org/10.1016/S0076-6879\(05\)00035-2](http://dx.doi.org/10.1016/S0076-6879(05)00035-2)
41. Raftogiannis RB, Wood TC, Otterness DM, et al. Phenol sulfotransferase pharmacogenetics in humans: Association of common *SULT1A1* alleles with TS PST phenotype. *Biochem Biophys Res Commun*. 1997;239:298–304. <http://dx.doi.org/10.1006/bbrc.1997.7466>, PMID:9345314