

https://dspace.org/

Pharmacogenomic varia.on in drug transporter encoding genes across African popula.ons

Patel, Yumna Dadi

2025
University of the Witwatersrand, Johannesburg
https://hdl.handle.net/10539/48275

Patel, Yumna Dadi . (2025). Pharmacogenomic varia.on in drug transporter encoding genes across African popula.ons [Master`s dissertation, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/48275

Downloaded from a DSpace repository. https://dspace.org/

Pharmacogenomic variation in drug transporter encoding genes across African populations

Yumna Dadi Patel

Student number: 2303732

Supervisor: Dr David Twesigomwe

Co-supervisor: Dr Houcemeddine Othman

A research report submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in partial fulfillment of the requirements for the Degree of Master of Science in Medicine (Genomic Medicine) by Research and Coursework

February, 2025





UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG



NATIONAL HEALTH
LABORATORY SERVICE

NATIONAL HEALTH LABORATORY SERVICE
School of Pathology, University of the Witwatersrand
DIVISION OF HUMAN GENETICS
Hospital Street, Johannesburg, 2001 I PO Box 1038, Johannesburg, 2000
[T] +27 11 489 9223 I [M] +27 78 080 8841 I [F] +27 11 489 9226 I [E] human.genetics@nhls.ac.za

30 November 2024

Re: Research report submission to the Faculty of Health Sciences, University of the Witwatersrand, by Ms Yumna Dadi Patel in partial fulfillment of the requirements for the degree Master of Science in Medicine (Genomic Medicine)

Dear examiner,

Thank you for agreeing to examine the research report submitted by Ms Yumna Dadi Patel (student number: 2303732) entitled *"Pharmacogenomic variation in drug-transporter encoding genes across African populations"*. This research report is being submitted by the candidate in the format of a "submissible" paper as per the Style Guide for Thesis, Dissertation and Research Reports (updated March 2016) by the University of the Witwatersrand Faculty of Health Sciences. This research report contributes 33% towards the final mark of the MSc (Med) Genomic Medicine degree for which the candidate is enrolled.

The candidate and her supervisors are considering the journal *Pharmacogenomics* as one of the preferred journal options for publication. This international journal is peer-reviewed and is published by Taylor & Francis. The submissible paper attached is in a free-format version which is permissible by the journal for purposes of review.

The research protocol has also been attached as an appendix for the purpose of providing an extended literature review and further contextualizing the study. The protocol was assessed and approved by the Faculty of Health Sciences Post-Graduate Assessors committee and is therefore not for examination.

Please refer to the Table of Contents for a complete list of the contents provided in this research report.

Yours sincerely,

Yumna Dadi Patel (Candidate)

Dr David Twesigomwe (Supervisor)

Dr Houcemeddine Othman (Co-supervisor)

This report is intended solely to record the observations and/or opinion of the writer. It does not constitute a medico-legal report

Declaration

I, Yumna Dadi Patel, declare that this Research Report is my own, unaided work. It is being submitted for the Degree of Master of Science in Medicine (Genomic Medicine) by Research and Coursework at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.



Yumna Dadi Patel

Student number: 2303732

Date: 18/02/2025

Presentations arising from this research project

1. Yumna Dadi Patel, Houcemeddine Othman, and David Twesigomwe. Pharmacogenomic variation in drug transporter encoding genes across African populations. Young and Early Career Geneticists Group (YEGG) Symposium , 26th October 2024 (poster presentation)
2. Yumna Dadi Patel, Houcemeddine Othman, and David Twesigomwe. Pharmacogenomic variation in drug transporter encoding genes across African populations. South African Society for Human Genetics (SASHG) Conference, 27th- 29th October 2024 (poster presentation)

Abstract

Variations in genes encoding enzymes involved in the absorption, distribution, metabolism, and excretion (ADME) of drugs significantly influence drug response across populations. While pharmacogenetic studies in African populations have increased, they predominantly focus on drug metabolism genes, leaving drug transporter genes, such as *SLCO1B1* (encoding gene for OATP1B1), underexplored. This gap is particularly critical as variations in *SLCO1B1* influence the response to drugs such as statins (cholesterol-lowering agents) and methotrexate (used in chemotherapy). This study aimed to characterize pharmacogenetic variation in *SLCO1B1* across African populations. High-depth whole genome sequencing data (n=1080) from sub-Saharan African populations were analysed, drawing from datasets generated by the Human Heredity and Health in Africa (H3Africa) Consortium, the 1000 Genomes Project and other collaborations. The StellarPGx pipeline was used for calling known *SLCO1B1* star alleles (i.e. haplotypes and structural variants) and identification of potentially novel haplotypes. Variants were annotated using the Ensembl Variant Effect Predictor (VEP), and missense variants were further evaluated using structural bioinformatics (DDGun) and visualised on OATP1B1 using PyMOL to assess their potential functional and structural impacts. The study identified over 17 unique *SLCO1B1* star alleles and five potentially novel haplotypes in sub-Saharan African populations. Star alleles such as **1* and **37*, associated with normal function phenotypes, exhibited high frequencies in sub-Saharan African populations. Haplotypes containing missense variants were further analysed and nine of these haplotypes (**9, *15, *19, *20, *27, *28, *30, *31, *44*), had negative Gibbs free energy scores, suggesting potential protein destabilisation. These findings underscore the importance of assessing African genomic diversity in pharmacogenomics research to refine drug-response predictions and develop tailored genotyping panels. Characterizing *SLCO1B1* genetic diversity in African populations lays a foundation for experimental and clinical functional studies to assay the impact on response to statins and other medications with a view towards supporting the broader application of precision medicine within African populations.

Acknowledgements

I would like to express my gratitude to my supervisor, Dr David Twesigomwe. Working on this project has been a valuable learning experience, and I am grateful for all the knowledge I have gained. Equally, I would like to express my heartful gratitude to my supervisor, Dr Houcemeddine Othman, for his invaluable guidance and encouragement. You embody the qualities of an exceptional educator and a compassionate guide.

I am also deeply grateful for the supportive and collaborative environment fostered by everyone at the Division of Human Genetics, and the scientific community at the Sydney Brennar Institute of Molecular Bioscience. Their shared expertise, encouragement, and readiness to assist have greatly enriched my academic journey and personal development.

Thank you to the Wits Postgraduate Merit award and the Wits Postgraduate Scholarship, which made this year financially attainable for me.

To my incredible friends and classmates, thank you for making this journey so much more meaningful and memorable. The moments we spent learning together, exchanging ideas, and cheering each other on are ones I will always treasure.

To my mum, sisters, and bhais, thank you for your unwavering love and constant encouragement. Your strength and belief in me have been my foundation through every step of this journey. Lastly, to my beloved late father, your love and guidance continue to inspire me every day. This achievement is as much yours as it is mine.

Table of Contents

Declaration ..... *iii*

Presentations arising from this research project ..... *iv*

Abstract ..... *v*

Acknowledgements vi

List of Figures ..... *ix*

List of Tables ..... *ix*

List of Abbreviations ..... *x*

Literature Review ..... *xi*

Research report in the format of a "submissible" paper ..... *xvi*

Abstract for Journal Submission ..... *xvii*

Introduction *1*

Materials and Methods *3*

Ethics approval 3

Population data 4

SLCO1B1 star allele analysis and phenotype prediction 5

Variant annotation and functional analysis 6

Variant annotation 6

Functional analysis on OATP1B1 7

Statistical Analysis 7

Results *8*

Star allele frequency distributions for *SLCO1B1* 8

SLCO1B1 phenotype distributions 13

Inferred novel *SLCO1B1* haplotypes 15

Functional impact of variants and star alleles 15

Discussion *19*

Conclusion *22*

Future research *22*

Data Availability Statement *22*

Code Availability Statement *23*

Author Contributions *23*

References *23*

Appendices *33*

Appendix A: Supplementary Tables and Figures *34*

Appendix B: Ethics Clearance Certificate *39*

Appendix C: Approved Research Protocol *42*

Appendix D: Plagiarism Declaration and Turnitin Report *43*

List of Figures

Figure 1: *SLCO1B1* gene locus ……2

Figure 2: Overview of the bioinformatic analysis performed in this study……6

Figure 3: Breakdown of *SLCO1B1* variants obtained from Ensemble VEP in SSA populations for this study……8

Figure 4: Phenotype distribution across African ethnolinguistic groups and global populations……14

Figure 5: Predicted effect of variation on OATP1B1 based on DDGun scores……17

Figure 6: Predicted stability scores of potential novel variants on OATP1B1……18

Figure 7: Structural representation of the OATP1B1 protein with key variants mapped……18

Supplementary Figure 1: Total depth of coverage for selected genes across various star allele diplotypes in different African populations……34

List of Tables

Table 1: sub-Saharan population groups included in each dataset used in the study……4

Table 2: *SLCO1B1* star allele (haplotype) frequencies across global populations……10

Table 3: *SLCO1B1* star allele (haplotype) frequencies across sub-Saharan African populations in this study……12

Table 4: Inferred novel variants for *SLCO1B1*……15

Supplementary Table 1: Assignment of predicted *SLCO1B1* likely phenotype based on genotype……35

Supplementary Table 2: *SLCO1B1* diplotype frequencies across global populations……36

Supplementary Table 3: In silico scores of variants for *SLCO1B1* (according to the ADME optimised framework)……38

List of Abbreviations

ACB	Afro-Caribbean
ADME	Absorption, Distribution, Metabolism, Excretion
AF	Allele frequency
AMR	Admixed-American
ASW	Americans with African-ancestry in South West USA
BAM	Binary Alignment Map
BEN	Participants from Benin
BF	Participants from Burkina Faso
BOT	Participants from Botswana
BRN	Berom in Nigeria
CMRN	Participants from Cameroon
CNVs	Copy Number Variants
CPIC	Clinical Pharmacogenomics Implementation Consortium
dbSNP	Single Nucleotide Polymorphism Database
EAS	East Asian participants
EM	Electron-microscopy
ESN	Esan in Nigeria
EUR	European participants
GHAN	Participants in Ghana
GWD	Madinka in western divisions of The Gambia
LWK	Luhya in Webuye, Kenya
MSL	Mende in Sierra Leone
OATP1B1	Organic Anion Transporting Polypeptide 1B1
PDB	Protein Data Bank
PGx	Pharmacogenomics
SAS	South Asian participants
SEB	south eastern Bantu in South Africa
SGDP	Simons Genome Diversity Project
SLCO1B1	Solute Carrier Organic Anion Transporter 1B1
SNVs	Single Nucleotide Variants
SSA	sub-Saharan Africa
VCF	Variant Call Format
VEP	Variant Effect Predictor
WGS	Whole Genome Sequencing
YRI	Yoruba in Ibadan, Nigeria
ZM	Participants from Zambia

Literature Review

Pharmacogenomics is a branch of precision medicine that explores how an individual's genomic profile impacts their response to drugs (Ahmed et al., 2016; Klein et al., 2019; Roden et al., 2019). Variations within ADME genes play an important role in drug responses (da Rocha et al., 2021). ADME proteins are typically classified as either phase I drug metabolisers (e.g. Cytochrome P450 enzymes), phase II drug metabolisers (e.g. N-acetyltransferases and Glutathione Transferases), drug transporters (*SLC* and *ABC* transporters), and modifiers (e.g. cytochrome P450 oxidoreductase), each playing a role in the pharmacokinetics of various drugs (da Rocha et al., 2021; Klein et al., 2019). Genetic variation in ADME genes can impact drug bioavailability, which is a function of the rate of drug metabolism and transport across cells/tissues (Schärfe et al., 2017). It is therefore understandable to project that functionally impactful variation in these genes, collectively, shapes drug efficacy and safety across populations.

Over the last two decades, participants involved in pharmacogenomics studies have predominantly been of European or Asian descent (Li et al., 2023). Furthermore, pharmacogenetic variation has been well characterised in drug metabolism genes such as cytochrome P450 genes. However, there is a lack of data on drug transporter-encoding genes, and the distribution of clinically relevant star alleles (haplotypes and structural variations) therein, particularly in African populations (Rajman et al., 2020). Star alleles are typically defined by various combination (i.e. haplotypes) of small nucleotide variations (SNVs; especially missense, frameshift, stop gain/loss, splice site variants) and/or structural variants; and are represented by star (*) allele nomenclature (Gaedigk et al., 2021). For polymorphic pharmacogenes like *CYP* and *SLC* genes, two star alleles (one on each chromosome) make up an individual's diplotype (e.g. *SLCO1B1*1/*2*).

Given that allele frequencies of ADME gene variants tend to vary across different populations (Ahmed et al., 2016; Zhou et al., 2017) and African populations are known to have a higher genetic diversity compared to other populations (Choudhury et al., 2017; da Rocha et al., 2021) it is important to obtain an in-depth characterisation of the pharmacogenomic landscape in Africa. The solute carrier (*SLC*) transporters and the ATP-binding cassette (*ABC*) transporters, comprise the two main drug transporter superfamilies (Lin et al., 2015). Understanding the distribution of common, rare, and novel star alleles and other forms of

genetic variation in drug transporter genes is crucial for informing pharmacokinetic studies in Africa and the design of genotyping platforms appropriate in the African context. Moreover, this information can be included in pharmacogenomics databases (for further meta-analyses) that combine data across multiple sources to identify patterns in drug response, especially in underrepresented African-ancestry populations (Rajman et al., 2017).

***SLC* transport genes**

SLC transporters are primarily involved in the uptake of small molecules into the cell (Lin et al., 2015). The *SLC* superfamily consists of 458 transporter proteins within 65 families. Their proteins are pivotal in drug absorption across cell membranes, distribution, and excretion (Pizzagalli et al., 2021). Key *SLC* genes, where variants potentially impacting drug transport have been reported globally, include *SLCO1B1*, *SLC10A1*, *SLC39A8*, *SLC22A1*, and *SLC22A6*.

In the African context, where genetic diversity is high and pharmacogenomic data is scanty, exploring variations within *SLC* genes can establish novel African-ancestry variants that can impact drug transport, thereby contributing to a more comprehensive understanding of pharmacogenomics in understudied populations.

SLCO1B1

The *SLCO1B1* gene is located on chromosome 12p12.1 and encodes the organic anion transporting polypeptide 1B1 (OATP1B1) protein (Ramsey et al., 2023). OATP1B1 is a major hepatic influx transport protein that facilitates the sodium-independent transport of organic ions across the cell membrane (Ciută et al., 2023). Genetic variations within *SLCO1B1* may result in decreased activity or abundance of the OATP1B1 protein, which can limit the hepatic clearance of drugs, thereby increasing systemic exposure to drug substrates and causing adverse effects (Ramsey et al., 2023). Moreover, some genetic variants have been found to increase gene expression and impact transport function in a substrate-dependent manner (Neuvonen et al., 2021). The distribution of *SLCO1B1* star alleles varies across different populations (Rajman et al., 2020).

Over 49 *SLCO1B1* star alleles have been catalogued by the Pharmacogene Variation Consortium (PharmVar) to date. This includes decreased function star alleles such as **5, *15, *23, *31*, and **46-*49,* and increased function alleles, *SLCO1B1*14* and *20. Detecting star

alleles/variants within these genes is essential as they encode transporter proteins involved in drug uptake and disposition, of drugs involved in the treatment of diseases such as diabetes and cardiovascular disorders (Ciută et al., 2023) that are prevalent in African populations (CDC, 2023). For instance, variations in *SLCO1B1* are linked to statin-induced myopathy, a critical consideration in cardiovascular treatment (Voora et al., 2022)

Other key *SLC* transport genes

Generally, the frequency data across Africa for many *SLCO1B1* star alleles are scarce, and recorded frequencies for alleles in African populations may be inaccurate due to the limited sample sizes There is evidence that variations in *SLC* genes besides *SLCO1B1* can influence drug transport. For instance, *SLC10A1* variants affect the transport of various drugs, including methotrexate and statins, potentially altering their efficacy and toxicity (Lin et al., 2015). *SLC39A8* variants can influence the uptake of essential minerals and metals (Park et al., 2015), which has implications for diseases such as hypertension, especially as magnesium and zinc contribute to regulating blood pressure (Chiu et al., 2021). Similarly, *SLC22A1* and *SLC22A6* variants are associated with drug transport, affecting medications like metformin and antiretrovirals used for HIV treatment (Lozano et al., 2018).

***ABC* transport genes**

Over 48 *ABC* genes have been identified and are divided into seven subfamilies. *ABC* are efflux transporters and regulate the movement of drugs across the cell membrane (Bruckmueller and Cascorbi, 2021).

ABCB1

ABCB1 is a drug transporter gene located on chromosome 7q21.12 and spans 210 kilobases. This gene consists of 29 exons, and the first two exons form part of the promoter region, therefore regulating the expression of *ABCB1* (Bruckmueller and Cascorbi, 2021). *ABCB1* encodes P-glycoprotein, also known as multi-drug resistance protein 1 (Sakurai et al., 2005). P-glycoprotein is an efflux transporter that plays an integral role in the bioavailability and clearance of drugs, such as chemotherapeutic agents and rifampicin, an antibiotic used to treat bacterial infections (Beloor Suresh et al., 2024).

A few studies in African populations have reported that variants in *ABCB1* can contribute to altering drug responses to rifampicin and antiretroviral medications, through influencing P-glycoprotein activity (Swart et al., 2012). Variants with these existing clinical annotations include *ABCB1* rs2032582 c.2677G>T/A, rs1045642 3435C>T, and rs1128503 1236C>T, to name a few (Sissung et al., 2014). However, reports on the clinical relevance of *ABCB1* variants regarding drug bioavailability largely remained unclear or conflicting as most variants are listed as level 3 or level 4 in the Pharmacogenomics Knowledge Base (PharmGKB; https://www.pharmgkb.org/gene). Moreover, adverse drug reactions mediated by P-glycoprotein are usually because of drug-drug interactions owing to its broad substrate specificity (Aszalos, 2007). Nonetheless, determining allele frequencies of rs2032582 and other variants in *ABCB1* in African populations is important for understanding the genetic landscape in drug transporter genes. These frequencies and novel variant discoveries could inform future genotyping efforts seeking to characterise the impact on African-ancestry *ABCB1* variants on either drug disposition or disease susceptibility.

ABCG2

The *ABCG2* gene is found on chromosome 14q22 and consists of 16 exons and 15 introns, spanning 66 kilobases (Fohner et al., 2017). *ABCG2* encodes Breast Cancer Resistance Protein (BCRP), which functions as an efflux transporter that has an influence on the pharmacokinetics on various drugs such as statins and antibiotics (Bruckmueller and Cascorbi, 2021). Therefore, variants in the BCRP can significantly alter protein function and impact drug response. However, only rs2231142 c.421C>A has a PharmGKB level 1A annotation, owing to its effect on the bioavailability of statins. The rs2231142 c.421C>A variant can reduce BCRP protein expression by 30 to 40%, leading to decreased activity in efflux from tissues (Cooper-DeHoff et al., 2022). Consequently, this increases drug exposure in tissues and plasma, impacting both drug efficacy and toxicity.

Allele frequencies for rs2231142 and other *ABCG2* variants vary among populations, and they have been characterised predominantly in European-ancestry and Asian-ancestry populations (Bruckmueller and Cascorbi, 2021). However sufficient data on African populations have not been obtained (Radouani et al., 2020). Due to their high genetic diversity, it is plausible that African populations may harbour novel functionally relevant *ABCG2* variants. The paucity of *ABCG2* allele distribution data for African populations warrants the need to analyse more

African genomes to obtain a better insight into whether Africans might respond differently to statins and other medications considering the diversity in the pharmacogenomic variation landscape on the continent.

Genotyping drug transporter genes is essential to better infer the potential pharmacokinetic profiles of several widely used drugs, and thus plays an important role in precision medicine and reducing adverse reactions. Despite the genetic diversity in African populations, there is a lack of data on the distribution of star alleles and other variant forms in drug transporter genes, *SLC* genes*, ABCB1* and *ABCG2* (Bruckmueller and Cascorbi, 2021; Lozano et al., 2018; Rajman et al., 2020). This causes difficulties in predicting phenotypes based on genotype and adjusting drug dosage for optimal therapy.

Study rationale

The increasing availability of high coverage whole genome sequence data through studies/consortia such as the Human, Heredity and Health in Africa (H3Africa) Consortium Projects (Choudhury et al., 2020), has provided an opportunity for explorative pharmacogenomics research (da Rocha et al., 2021). Moreover, this is also facilitated by current reproducible pharmacogenetic bioinformatic pipelines, such as StellarPGx (Twesigomwe et al., 2021). Characterising pharmacogenomic variation in drug transporter encoding genes across Africa is an important component towards informing strategies such as the design of appropriate pharmacogenetic testing panels and the design of ADME *in vitro* and *in vivo* functional studies pertinent to African populations. Combining the use of high-throughput sequence datasets from African populations and scalable bioinformatics tools/pipelines can catalyse the characterisation of known and potential novel star alleles in key drug transporter paving the way for more accurate pharmacogenetic testing and informing functional studies assessing the impact of novel African ancestry variation on drug response.

Study focus

Considering the timeframe of this MSc research project (3 months), the study focused on the *SLCO1B1* gene to ensure a detailed and thorough analysis within the available period. This focused approach allowed for a comprehensive investigation into the gene's pharmacogenomic variations across African populations, providing meaningful insights while adhering to the project’s time constraints.

Research report in the format of a "submissible" paper

Journal option: *Pharmacogenomics*

Pharmacogenomic variation in *SLCO1B1* across sub-Saharan African populations

Yumna Dadi Patel[1,2], Houcemeddine Othman[2,3], and David Twesigomwe[2]

[1]Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[2]Sydney Brenner Institute of Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[3]Laboratory of Cytogenetics, Molecular Genetics and Reproductive Biology (LR03SP02), Farhat Hached University Hospital, Ibn El Jazzar St.Sousse, 4000, Tunisia.

Conflict of interest:

The authors declare no conflict of interest for this work.

Abstract for Journal Submission

Introduction: Variations in genes encoding drug absorption, distribution, metabolism, and excretion (ADME) enzymes and transporters significantly influence drug response across populations. While pharmacogenetic research in African populations has grown, most studies focus on drug metabolism genes, with limited exploration of drug transporter genes like *SLCO1B1***,** which encodes the OATP1B1 hepatic transporter. Variations in *SLCO1B1* affect responses to widely used drugs, including statins (cholesterol-lowering agent) and methotrexate (anti-rheumatic and chemotherapy). This underscores the need for comprehensive investigation of variation in *SLCO1B1* especially in sub-Saharan African (SSA) populations, where there is extensive genetic diversity and the burden of hypercholesterolemia and other non-communicable diseases is rising.

Methods: This study analysed high-depth whole-genome sequencing data (n = 1080) from SSA populations, incorporating datasets from the Human Heredity and Health in Africa (H3Africa) Consortium, the 1000 Genomes Project, and other collaborations. The StellarPGx pipeline was used for calling *SLCO1B1* star alleles (haplotypes and structural variants). Variants defining these star alleles were annotated using the Ensembl Variant Effect Predictor (VEP), and the structural and functional impacts of missense variants were assessed using tools such as DDGun and PyMOL.

Results: 17 known *SLCO1B1* star alleles and eight potentially novel ones were identified in SSA. *SLCO1B1*1* and **37*, associated with OATP1B1 normal function, were observed at high frequencies. Of the haplotypes analysed, nine (**9, *15, *19, *20, *27, *28, *30, *31, *44*), had negative Gibbs free energy scores, suggesting potential protein destabilisation.

Conclusion: These findings underscore the importance of assessing African genomic diversity in pharmacogenomics research to refine drug-response predictions and develop tailored genotyping panels. Characterizing *SLCO1B1* genetic diversity in African populations lays a foundation for experimental and clinical functional studies to assay the impact on response to statins and other medications with a view towards supporting the broader application of precision medicine within African populations.

Keywords: drug transporter genes, *SLCO1B1,* sub Saharan Africa

Introduction

Pharmacogenomics (PGx) examines how genetic variation affects drug response. The majority of PGx research over the years has focused on genes encoding proteins involved in drug absorption, distribution, metabolism, and excretion (collectively known as ADME) (Ahmed et al., 2016; Klein et al., 2019; Roden et al., 2019). Genetic variants in ADME genes can impact drug bioavailability, which is determined by the rate of drug metabolism and transport across cells and issues (Schärfe et al., 2017).

While drug metabolism genes, such as those in the cytochrome P450 family, have been extensively studied, there remains a paucity of data regarding drug transporter genes. The main drug transporter genes belong to the solute carrier organic anion transporter (*SLC*) family and the ATP-binding cassette (*ABC*) superfamily. Genetic variation in these genes is collectively responsible for variability in response to medications such as statins (cholesterol-lowering agent), methotrexate (anti-rheumatic and chemotherapy), and warfarin (blood thinner), among other medications. Herein, the sub-study presented is focused on the solute carrier organic anion member 1B1 (*SLCO1B1*).

The *SLCO1B1* gene (Figure 1), located on chromosome 12p12.1, spans 109 kilobases and consists of 15 exons (Ramsey et al., 2023). The ATG start codon is situated in exon 2, while exon 1 forms part of the 5' untranslated region. This gene encodes the organic anion transporting polypeptide 1B1 (OATP1B1) protein, a crucial hepatic influx transporter responsible for the sodium-independent transport of organic ions across cell membrane (Ciută et al., 2023). Deleterious genetic variants within *SLCO1B1* can diminish the activity or abundance of OATP1B1, leading to impaired hepatic drug clearance, increasing systemic drug exposure, and a heightened risk of adverse effects (Ramsey et al., 2023). Conversely, certain variants may enhance gene expression and modify transport function in a substrate-dependent manner (Neuvonen et al., 2021). These variants may affect the structure and/or function of OATP1B1, which mediates the uptake and disposition of drugs commonly prescribed for widespread conditions such as diabetes and cardiovascular disease (Ciută et al., 2023) which are becoming increasingly common within African populations (CDC, 2023). For example, certain *SLCO1B1* variants are associated with statin-induced myopathy, an important consideration in cardiovascular treatment (Voora et al., 2022).

Currently, the Pharmacogene Variation Consortium (PharmVar) has catalogued over 49 star alleles (Ramsey et al., 2023). *SLCO1B1* star alleles (Figure 1b) are typically defined by combinations (i.e. haplotypes) of small nucleotide variants (SNVs), such as missense, frameshift, stop-gain/loss, and splice-site variants (Gaedigk et al., 2021; Ramsey et al., 2023). Furthermore, some star alleles are defined by structural variants, including copy number variants (CNVs i.e. deletions and duplications). An individuals diplotype consists of two star alleles, one inherited from each parent (e.g., *SLCO1B1*1/*2*).

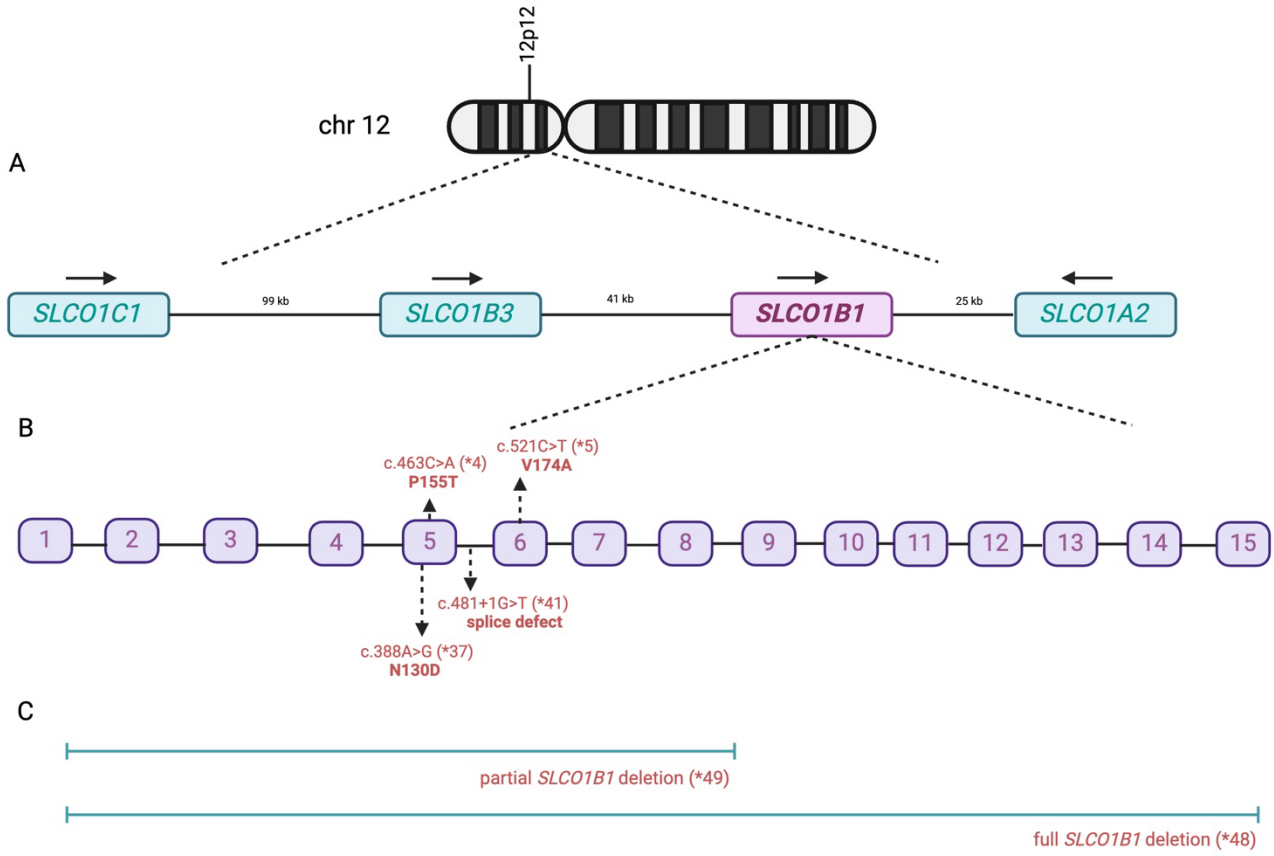


Figure 1: *SLCO1B1* gene locus. (A) The *SLCO1B1* locus on chromosome 12 includes *SLCO1B1, SLCO1B3, SLCO1B7*, and *SLCO1A2.* All genes except *SLCO1A2* are transcribed from the positive strand (indicated by black arrows). (B) *SLCO1B1* consists of 15 exons (purple boxes). Core variants defining **4, *5, *37*, and **41* are indicated at their respective positions. (C) Structural variants (**48* and **49*) are shown as blue lines representing major genomic rearrangements. (Created in BioRender. https://BioRender.com/y60m086)

Despite the extensive genetic diversity within African populations, there remains a significant gap in the distribution of star alleles and other variants in *SLCO1B1,* particularly in sub-Saharan Africa (SSA) (Li et al., 2023). The growing availability of high-coverage whole genome sequence (WGS) data through initiatives such as the Human, Heredity and Health in Africa (H3Africa) Consortium Projects (Choudhury et al., 2020) is opening doors to advancing pharmacogenomic research (da Rocha et al., 2021). This is further supported by reproducible bioinformatic pipelines such as StellarPGx (Twesigomwe et al.,2021), which facilitates identification of known and potentially novel star alleles across WGS datasets.

We hypothesize that the use of statins and other drugs, transported by OATP1B1, is likely to increase over the next decade in SSA due to the increasing burden of hypercholesterolemia and other non-communicable diseases. Therefore, enriching pharmacogenomic databases with information on *SLCO1B1* variation in SSA is important for precision medicine in this context. This study aimed to characterize genetic variation in *SLCO1B1* across diverse populations with a view towards informing future *SLCO1B1*-related pharmacogenomics research in the region. This was established by analysing the distribution of known pharmacogenetic variants and star alleles in *SLCO1B1*, identifying novel functional variants, and annotating and predicting the functional impact of rare and novel African-ancestry star alleles within *SLCO1B1*.

Materials and Methods

Ethics approval

Ethics approval for this study was granted by the Human Research Ethics Committee (HREC Medical) of the University of the Witwatersrand (protocol number: M240744). The data from all datasets was already deidentified. All analysis involving raw data was done on the ZA-Wits-Core computing cluster (PI: Prof Scott Hazelhurst). This sub-study falls under the parent study: Characterisation of pharmacogene allelic variation in multi-ancestry populations and development of a novel diplotype calling algorithm (ethics clearance certificate number: M200993).

Population data

High-coverage (30X) WGS data from 1080 SSA individuals was analysed from six data sources, namely the 1000 Genomes project (Byrska-Bishop et al., 2022), the Africa Wits-INDEPTH partnership for Genomics (AWI-Gen) dataset (Ramsay et al., 2016), Simons Genome Diversity Project (SGDP) (Mallick et al., 2016), the Southern African Human Genome Project (SAHGP) (Choudhury et al., 2017), the H3Africa-Baylor datasets (Choudhury et al., 2020) and the Cell Biology Research Lab (CBRL), which is associated with the National Institute of Communicable Diseases (NICD) and University of the Witwatersrand. The African ethnolinguistic groups included in each dataset are summarised in Table 1.

Whole-genome sequencing datasets (n=2000) from six global populations represented in the 1000 Genomes Project were included in the study to compare *SLCO1B1* star allele distributions in SSA with global populations. This comprised 347 admixed American (AMR) participants, 61 Americans with African ancestry in South West, USA (ASW), 96 African Caribbean (ACB) participants in Barbados, 503 European (EUR) participants, 504 East Asian (EAS) participants, and 489 South Asian (SAS) participants.

Table 1: sub-Saharan population groups included in each dataset used in the study.

Country/Population	Primary project	n
Public datasets		
Gambian in Western Division, Madinka (GWD)	1000 Genome Project	113
Luhya in Webuye, Kenya (LWK)	1000 Genome Project	99
Esan in Nigeria (ESN)	1000 Genome Project	99
Yoruba in Ibadan (YRI)	1000 Genome Project	108
Mende in Sierra Leone (MSL)	1000 Genome Project	85
Various African ethnolinguistic groups	SGDP	31
African genomes from African-based research projects/collaborations		
South-eastern Bantu (SEB) in South Africa	AWI-Gen	100
Burkina Faso	AWI-Gen; H3Africa-Baylor	33
Ghana	AWI-Gen; H3Africa-Baylor	26
South Africa	SAHGP	15
South Africa	CBRL (NICD, WITS)	85
Cameroon	H3Africa-Baylor	50
Benin	H3Africa-Baylor	50
Nigeria	H3Africa-Baylor	49

Zambia	H3Africa-Baylor	41
Botswana	H3Africa-Baylor	47
Mali	H3Africa-Baylor	50

SGDP: Simon Genome Diversity Project; AWI-Gen: Africa Wits-INDEPTH partnership for Genomics; SAHGP: Southern African Human Genome Project; CBRL (NICD, WITS): Cell Biology Research Lab, National Institute of Communicable Diseases, University of the Witwatersrand

SLCO1B1 **star allele analysis and phenotype prediction**

The StellarPGx pipeline (Twesigomwe et al., 2021) implemented in Nextflow (Di Tommaso et al., 2017), was used to call *SLCO1B1* star alleles. Briefly, StellarPGx uses GraphTyper (Eggertsson et al., 2017) to call SNVs from Binary Alignment Map (BAM) files. Before variant calling, Graphtyper realigns reads in a target region to a custom pangenome reference. A pangenome is created by augmenting a standard linear reference (such as GRCh38) with known variants. Using this "variant-aware" reference, reduces common biases from traditional linear reference genomes, resulting in more accurate variant detection (Eggertsson et al., 2017). Variant genotypes and allele depth data based on the realignment are stored in Variant Call Format (VCF) files. For star allele assignment, StellarPGx matches the detected SNVs in VCF files to known star allele combinations in its database, which is primarily based on PharmVar-curated definitions (https://www.pharmvar.org/gene/SLCO1B1).

To identify structural variants like gene duplications or deletions, StellarPGx relies on read-depth obtained using SAMtools depth (Li et al., 2009) to calculate a ratio between the target gene and the control genes (*VDR* and *EGFR*) to ascertain copy number changes. *VDR* and *EGFR* are used because they are relatively large and tend to have a stable copy number of two (MacDonald et al., 2014), probably owing to their biological importance. The ratio is then used to detect potential CNVs: a value significantly higher than 1 suggests gene duplication, while a value significantly lower than 1 suggests a deletion. These read-depth ratios are compared against thresholds based on expected values for stable genes, allowing StellarPGx to identify copy number changes in *SLCO1B1*. The structural variant data are then combined with SNV allele depth information from the VCF files to determine structural changes such as allele duplications and/or deletions. For samples with novel star alleles or suballeles StellarPGx includes a flag in the output; and suggests the closest known matching diplotype to facilitate manual variant inspection. An overview is shown on Figure 2.

SLCO1B1 drug transporter phenotype predictions were made according to the diplotype-phenotype translation information from the Clinical Pharmacogenomics Implementations Consortium (CPIC) guidelines (PharmGKB) (Cooper-DeHoff et al., 2022), as shown on Supplementary Table 2.

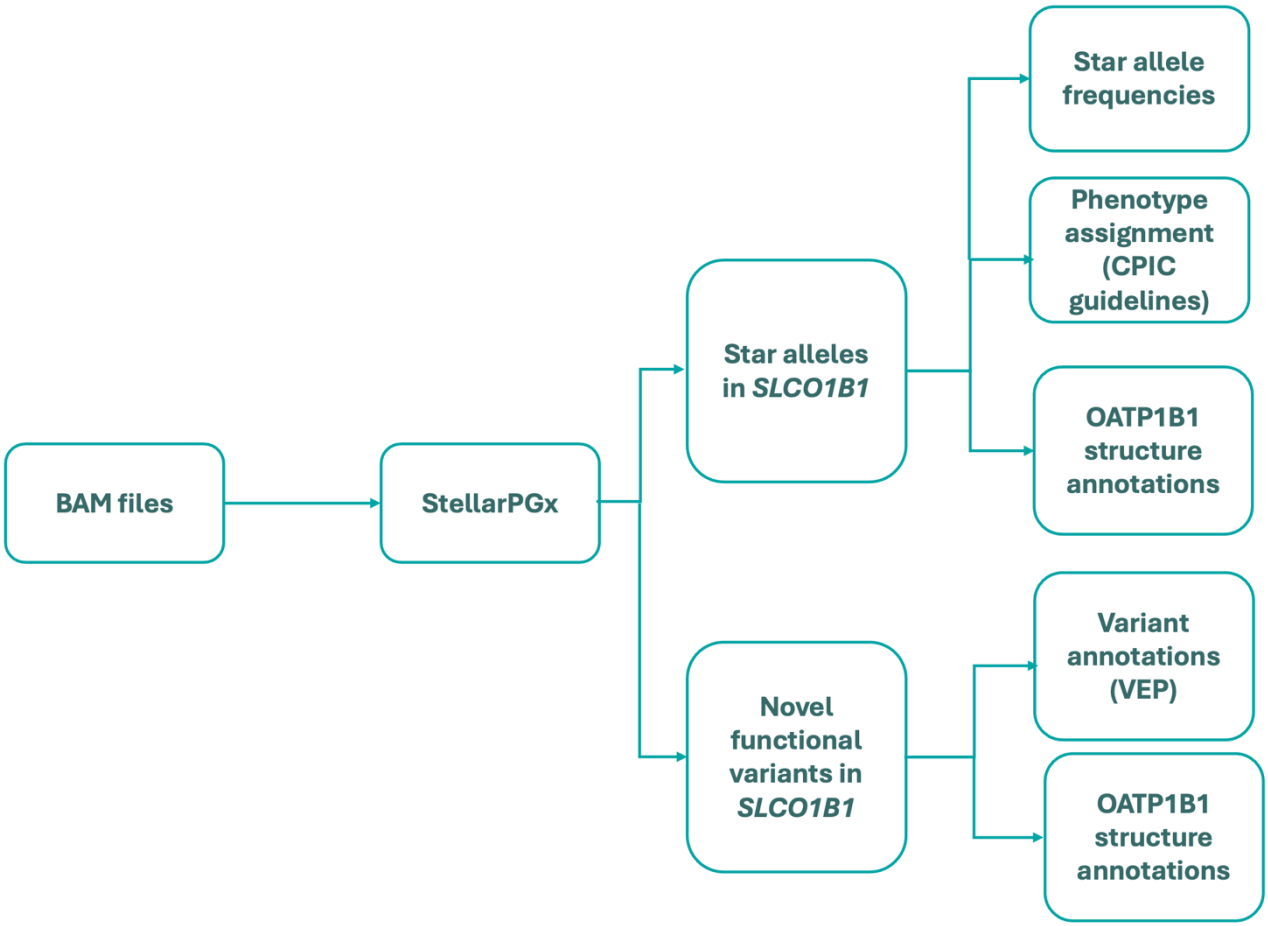


Figure 2: Overview of the bioinformatic analysis performed in this study. BAM: Binary Alignment Map; CPIC: Clinical Pharmacogenomics Implementation Consortium; VEP: Ensembl Variant Effect Predictor

Variant annotation and functional analysis

Variant annotation

All variants identified in this study were annotated using Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016). *In-silico* prediction tools including SIFT (Ng and Henikoff, 2003), Polyphen-2 (Adzhubei et al., 2013), LRT , CADD (Rentzsch et al., 2019), MutationAssessor (Reva

et al., 2011), PROVEAN (Y. Choi et al., 2012), and AlphaMissense (Cheng et al., 2023) were used to determine the functional impact of SNVs and/or indels. A consensus functional prediction based on the scores from these tools and recommended parameters for ADME genes (Zhou et al., 2024) were obtained. This consensus was reached, when at least half of the *in silico* scores predicted a variant to be deleterious.

Functional analysis on OATP1B1

Structural bioinformatics tools were used to evaluate how *SLCO1B1* missense variants impact the variation of the Gibbs free energy compared to a reference variant of the OATP1B1 protein. This analysis is predicated on the hypothesis that alterations in protein structure due to these variants may provide insights into their potential effects on protein function.

PDB files for the AlphaFold and electron microscopy (EM) structures were obtained from the AlphaFold Protein Structure Database (Varadi et al., 2024) and the Protein Data Bank (https://www.rcsb.org/structure/8HND), respectively (Identifiers: AF-Q9Y6L6-F1 and 8HND). The *SLCO1B1* missense variants were analysed using DDGun (Montanucci et al., 2019), which facilitates the prediction of the functional impact of amino acid substitutions by calculating changes in Gibbs free energy.

Following the Gibbs free energy calculations, a linear regression analysis was performed for both the AlphaFold and EM scores to determine the relationship between the structural alterations and the energy changes associated with the variants. Finally, the variants were visualized and mapped using the PyMOL Molecular Graphics System (DeLano, 2002).

Statistical Analysis

The summary output generated by StellarPGx was analysed to determine star allele and diplotype frequencies. To identify significant differences in star allele frequencies among populations, Fisher's Exact Test was conducted. For the correction of multiple comparisons, the false discovery rate (FDR) p-value adjustment for dependent observations was applied, following the Benjamini-Yekutieli method. Differences with p-values < 0.05 were considered statistically significant. Furthermore, deviations from Hardy-Weinberg equilibrium were assessed using version 1.2.8.1.3 of the Genetics package https://cran.r-project.org/web/packages/genetics/index.html within R version 4.1.0. A p-value >0.05

indicated that the allele distributions were consistent with Hardy-Weinberg equilibrium expectations.

Results

In SSA, a total of 3,562 variants were identified in the *SLCO1B1* gene, with 161 (4.5%) being novel variants. The distribution of these variants by type is shown in Figure 3.

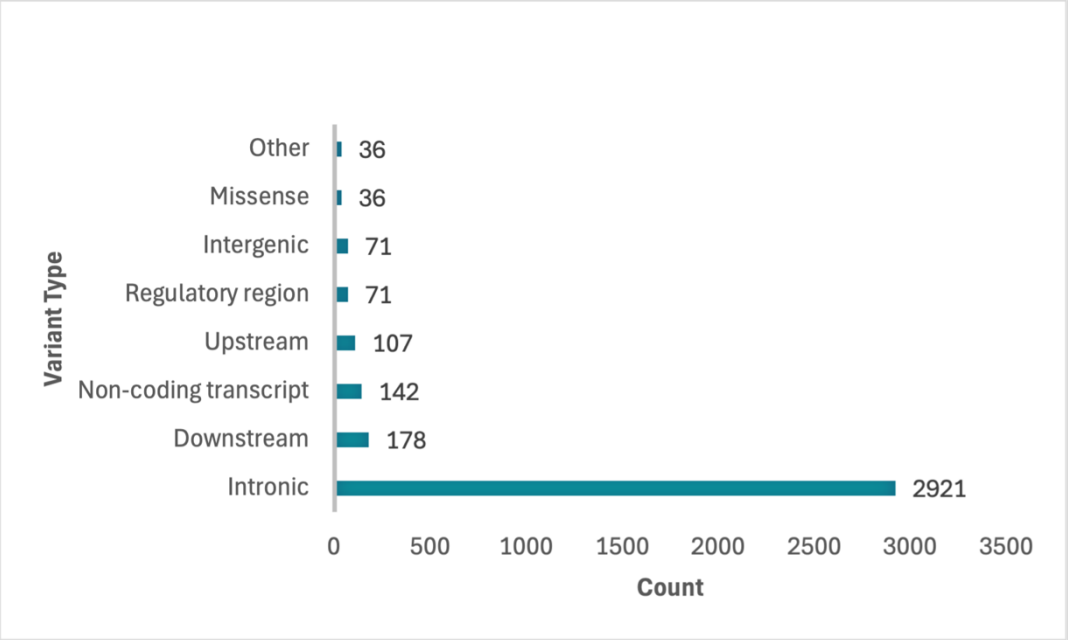


Figure 3: Breakdown of *SLCO1B1* variants obtained from Ensemble VEP in SSA populations for this study.

Star allele frequency distributions for *SLCO1B1*

Seventeen previously known *SLCO1B1* star alleles were identified across SSA, with *SLCO1B1*1* and **37* being the most frequent (Tables 2 and 3). Both star alleles are associated with normal function. *SLCO1B1*1* was most prevalent among participants from Ghana (n=26), with an allele frequency of 22.2%, and least common in participants from Mali, at 7% (p-value = 0.009408). In contrast, *SLCO1B1*37* showed the highest allele frequency in participants from Mali (77%) and the lowest allele frequency in participants from Burkina Faso (50%, p-value = 0.008932). Other star alleles observed in SSA generally had low frequencies, with some appearing only as singletons, including potentially novel haplotypes (Tables 3 and 4). However,

no significant differences in star allele frequencies were observed between any of the SSA ethnolinguistic groups in this study, as assessed using Fisher's exact test. Additionally, all-star alleles tested showed no significant deviations from Hardy-Weinberg Equilibrium, suggesting the absence of underlying selection.

Potential copy number variants (CNVs) were detected in samples from South Africa (**20/*37*x2*, *37/*37*x2*, *41/*37*x2) and Ghana (**37/*37*x2). (Supplementary Figure 1). However, due to the limited data available, these findings remain inconclusive. Further analysis and additional data are needed to confirm the presence and impact of these potential CNVs as well as their contribution to pharmacogenomic variability in these populations.

The frequency of *SLCO1B1*1* in SSA was significantly lower compared to global populations (Table 3), such as in European participants (56.7%, p-value = 1.08×10^{-15}) and Admixed-American participants (50.3%, p-value = 1.62×10^{-15}). In contrast, *SLCO1B1*37* showed similar frequencies in SSA and East Asian populations (62.4%, p-value = 2.94×10^{-16}). A relatively high allele frequency for *SLCO1B1*37* was also observed in African-American/Afro-Caribbean populations (46.2%). Although other star alleles generally had low frequencies in SSA, they were still comparatively higher than those observed in global populations, albeit not significantly so.

Table 2: *SLCO1B1* star allele frequencies in SSA and across global populations

Haplotype	rsID of defining variant(s)	Variant alleles (NG_011745.1)	Protein change	Consequence	CPIC Function	Allele Frequencies (%)					
						SSA n=1080	EUR n=503	EAS n=504	SAS n=489	ASW/ ACB n=157	AMR n=347
*1					Normal	14.4	56.7	22.9	42.2	19.1	50.3
*4	rs11045819	g.50686C>A	P155T	Missense	Indeterminate	0.3	0.1	0	0	0	0.1
*5	rs4149056	g.52422T>C	V174A	Missense	No function	0	2.8	0	0.5	0	1.9
*9	rs59502379	g.79806G>C	G488A	Missense	No function	<0.1	0	0	0	0	0
*14	rs2306283 rs11045819	g.50611A>G g.50686C>A	N130D P155T	Missense Missense	Increased function	2.8	14.2	0.3	2.8	3.8	9.7
*15	rs2306283 rs4149056	g.50611A>G g.52422T>C	N130D V174A	Missense Missense	No function	0.8	12.9	11.9	3.7	5.1	11
*16	rs2306282	g.50675A>G	N151S	Missense	Indeterminate	0	0	0.4	0	0	0
*19	rs34671512	g.112849A>C	L643F	Missense	Indeterminate	0.2	0.3	0	2.5	0	0
*20	rs2306283 rs34671512	g.50611A>G g.112849A>C	N130D L643F	Missense Missense	Increased function	7.9	4.7	0.5	2	7	4.3
*24	rs2306283 rs11045852	g.50611A>G g.70758A>G	N130D I245V	Missense Missense	Indeterminate	0	0	0	3.9	0.3	0
*27	rs2306283 rs59113707	g.50611A>G g.76362C>G	N130D F400L	Missense Missense	Indeterminate	5.3	0	0	0	4.5	0.6
*28	rs2306283 rs11045852 rs11045853	g.50611A>G g.70758A>G g.70783G>A	N130D I245V R253Q	Missense Missense Missense	Indeterminate	0.1	0	0	0	0.3	0
*29	rs2306283 rs140790673	g.21329738A>G g.21392092C>T	N130D S682F	Missense Missense	Indeterminate	0	0	0.2	0	0	0
*30	rs2306283 rs79135870	g.50611A>G g.21331891A>G	N130D I222V	Missense Missense	Indeterminate	0	0	0	0	1.9	0

*31	rs2306283 rs59502379	g.50611A>G g.79806G>C	N130D G488A	Missense Missense	No function	4.2	0	0	0	3.8	0.3
*32	rs2306283 rs11045819 rs11045852	g.50611A>G g.50686C>A g.70758A>G	N130D P155T I245V	Missense Missense Missense	Indeterminate	0	0	0	0.2	0	0
*37	rs2306283	g.50611A>G	N130D	Missense	Normal	63.2	8	62.4	42	47.1	21.2
*38		g.50678G>A	R152L	Missense	Indeterminate	<0.1	0	0	0	0	0
*39	rs2306283 rs183501729	g.50611A>G g.70782C>T	N130D R253X	Missense Stop gain	Indeterminate	0	0	0.2	0	0	0
*40	rs4149056 rs34671512	g.52422T>C g.112849A>C	V174A L643F	Missense Missense	Indeterminate	0	0.1	0	0	0	0.1
*41	rs77271279	g.50705G>T		Splice defect	Indeterminate	3.4	0	0	0	2.2	0.1
*43	rs2306283 rs11045819 rs11045852 rs74064213	g.50611A>G g.50686C>A g.70758A>G g.79838A>G	N130D P155T I245V I499V	Missense Missense Missense Missense	Indeterminate	2.6	0	0	0	3.8	0.1
*44	rs2306283 rs11045852 rs74064213	g.50611A>G g.70758A>G g.79838A>G	N130D I245V I499V	Missense Missense Missense	Indeterminate	0.2	0	0	0	0	0
*45	rs71581941	g.96162C>T	R580X	Stop gain	Indeterminate	0	0.2	0.6	0	0	0
*46	rs2306283 rs4149056 rs71581941	g.50611A>G g.52422T>C g.96162C>T	N130D V174A R580X	Missense Missense Stop gain	No function	0	0	0	0	0	0.3
*47	rs2306283 rs4149056 rs72559747	g.50611A>G g.52422T>C g.74351C>G	N130D V174A P336R	Missense Missense Missense	No function	0	0	0.4	0	0	0

CPIC: Clinical Pharmacogenomics Implementation consortiums; SSA: Sub-Saharan Africa; EUR: European; EAS: East Asian; SAS: South Asian; ASW: African American; ACB: Afro-Caribbean; AMR: admixed American

Table 3: Star allele frequencies across SSA in this study

			Allele Frequencies (%)													
Haplotype	rsID of defining variant(s)	CPIC Function	YRI	LWK	GWD	MSL	ESN	SEB	BF	GHAN	BEN	BRN	CMRN	ZM	BOT	MALI
*1		Normal	13	13.1	13.7	11.2	9.1	13.3	21.2	22.2	12	10.2	12	19.5	17	7
*4	rs11045819	Indeterminate	0	0.5	0.9	0	0.5	0	0	0	0	0	0	0	1.1	0
*9	rs59502379	No function	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0
*14	rs2306283 rs11045819	Increased function	0.9	2	4.9	4.7	3	1.8	1.5	0	2	5.1	1	2.4	5.3	2
*15	rs2306283 rs4149056	No function	0.9	2	0	0	0	1.5	0	0	0	2	1	0	0	0
*19	rs34671512	Indeterminate	0	0	0	0	0	0	1.5	0	0	0	1	1.2	1.1	0
*20	rs2306283 rs34671512	Increased function	6.5	7.1	4.9	5.3	8.1	11.8	12.1	1.9	10	3.1	6	7.3	6.4	5
*27	rs2306283 rs59113707	Indeterminate	6.9	2.5	6.2	5.3	7.6	3.3	9.1	5.6	3	7.1	5	0	7.4	1
*28	rs2306283 rs11045852 rs11045853	Indeterminate	0	0	0.4	0.6	0	0	0	0	0	0	0	0	0	0
*30	rs2306283 rs79135870	Indeterminate	3.2	0.5	0.4	0.6	1.5	0.5	1.5	3.7	3	0	0	0	1.1	3
*31	rs2306283 rs59502379	No function	4.6	5.1	1.8	2.4	6.6	2.3	1.5	9.3	8	4.1	5	3.7	1.1	3
*37	rs2306283	Normal	53.2	65.2	62.8	60	57.6	59	50	51.9	56	59.2	58	57.3	55.3	77
*38		Indeterminate	0	0	0	0	0	0	0	0	0	0	1	0	0	0
*41	rs77271279	Indeterminate	5.1	2	4	5.9	2	2	1.5	3.7	1	6.1	5	3.7	0	2
*43	rs2306283 rs11045819 rs11045852 rs74064213	Indeterminate	4.2	0	0	0.6	3.5	4.8	0	9.3	4	2	5	4.9	1.1	0
*44	rs2306283 rs11045852 rs74064213	Indeterminate	0.9	0	0	0	0.5	0	0	0	0	1	0	0	0	0

CPIC: Clinical Pharmacogenomics Implementation consortium; YRI: Yoruba in Ibadan; LWK: Luhya in Webuye, Kenya; GWD: Gambian in the Western Division, The Gambia; MSL: Mende in Sierra Leone; ESN: Esan in Nigeria; SEB: Southeastern Bantu-speaking; BF: Burkina Faso; GHAN: Ghana; BEN: Benin; BRN: Berom, Nigeria; CMRN: Cameroon; ZM: Zambia; BOT: Botswana

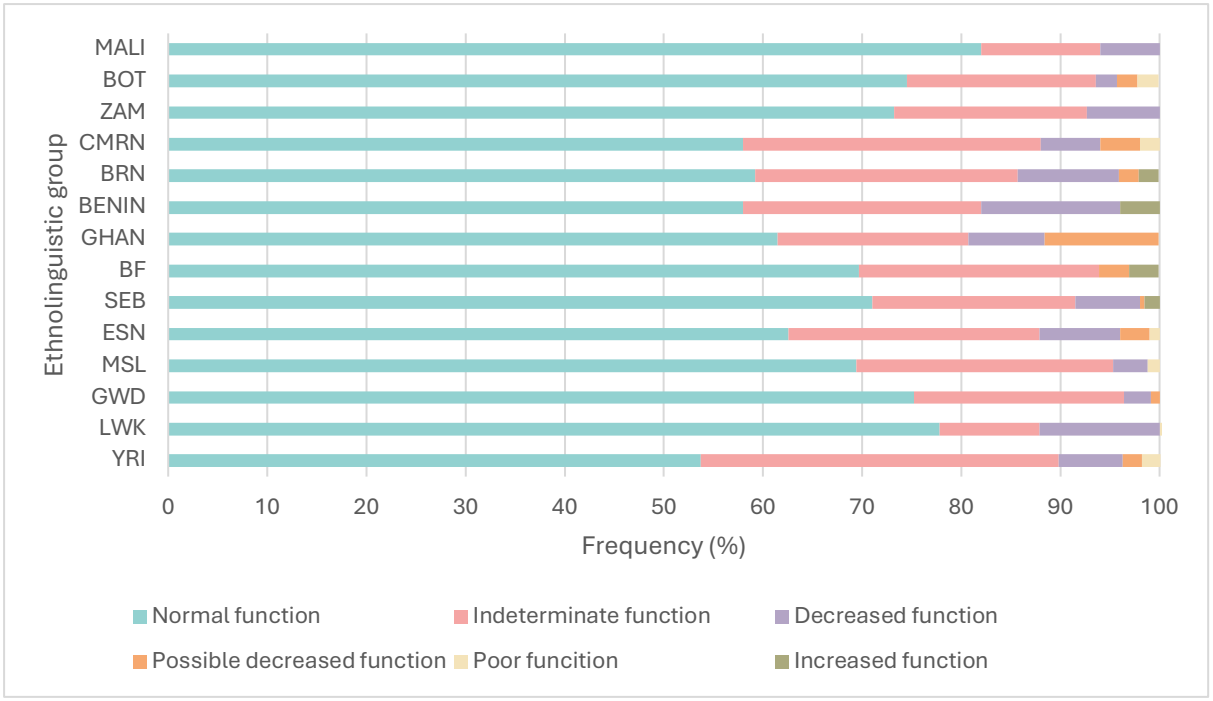
SLCO1B1 phenotype distributions

The majority of individuals in SSA had a normal function phenotype (67.7%), primarily due to the prevalence of the **1/*37* and **37/*37* diplotypes (Supplementary Table 2). This frequency aligns with the overall normal function phenotype observed in global populations (72.7%), as shown in Figure 4. However, the **37/*37* diplotype was less common in most global population groups, where there was a higher prevalence of the **1/*1* diplotype, particularly in European and Admixed American populations (Supplementary Table 2), however no statistical significance was observed between these two populations and SSA for the normal function phenotype.

Notably, 22.7% of the SSA population exhibited an indeterminate phenotype, which was significantly higher compared to global populations such as in, European (1.2%, p-value = $2.2x10^{-16}$) and Admixed American (2%, p-value = $2.2x10^{-16}$) groups. In contrast, the decreased function phenotype was significantly different from European ($p = 2.51 \times 10^{-16}$), Admixed American ($p = 9.25 \times 10^{-15}$), and East Asian ($p = 1.08 \times 10^{-14}$) populations when compared to SSA.

The frequencies observed for the increased function, possible decreased function, and poor function phenotypes were comparatively lower than that observed for normal function and decreased function phenotypes. The highest frequency for increased function was observed in Admixed-American populations (4%, p-value = 1.04×10^{-3}), and European populations (2.6%, p-value = 0.04), compared to SSA (Figure 5). Moreover, the possible decreased function phenotype was only observed in SSA (1.4%), African-American/Afro-Caribbean (2.5%), and East Asian (0.2%) populations. Finally, no statistical significance was obeserved for the poor function phenotype across global populations and SSA.

a



b

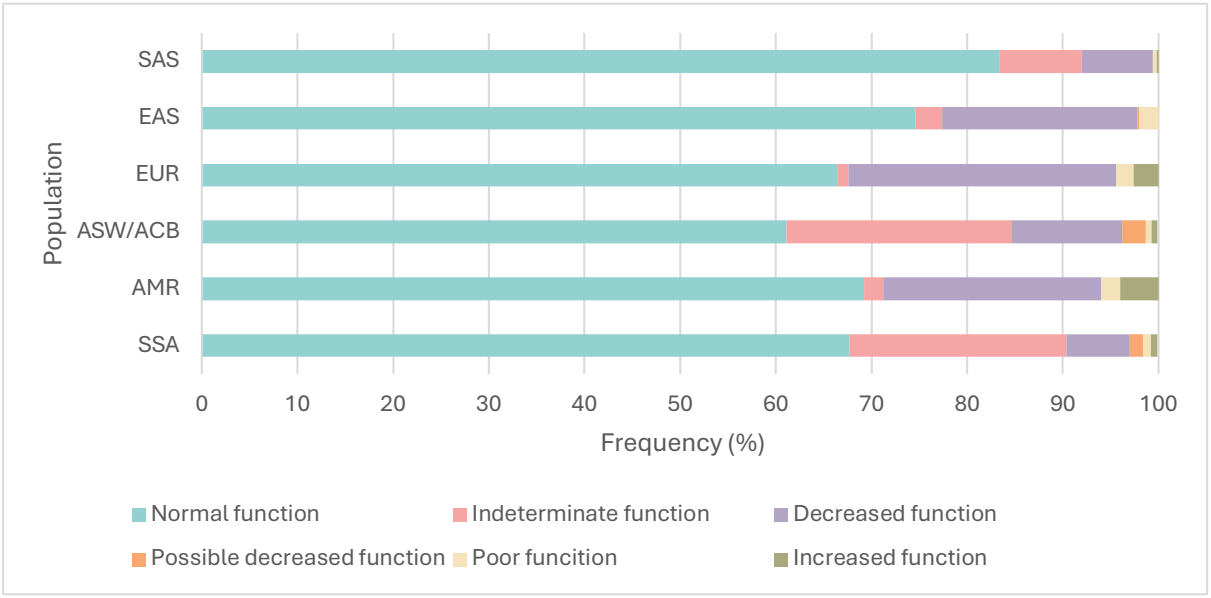


Figure 4: Phenotype distribution across (a) African ethnolinguistic groups and global populations (b). YRI: Yoruba in Ibadan; LWK: Luhya in Webuye, Kenya; GWD: Gambian in the Western Division, The Gambia; MSL: Mende in Sierra Leone; ESN: Esan in Nigeria; SEB: Southeastern Bantu-speaking; BF: Burkina Faso; GHAN: Ghana; BRN: Berom, Nigeria; CMRN: Cameroon; ZM: Zambia; BOT: Botswana; SSA: Sub-Saharan Africa; EUR: European; EAS: East Asian; SAS: South Asian; ASW: African American; ACB: Afro-Caribbean; AMR: admixed American

Inferred novel *SLCO1B1* haplotypes

In the SSA population, 13 individuals were predicted to carry novel haplotypes (Table 4), whereas a smaller proportion of individuals from other global populations exhibited predicted novel haplotypes. The variants defining these novel haplotypes are catalogued in the Single Nucleotide Polymorphism Database (dbSNP), however these were observed in combinations not yet catalogued by PharmVar, therefore classifying them as potentially novel star alleles that require further investigation.

Table 4: Inferred novel variants for *SLCO1B1*

#	Background allele	Additional core variant	Consequence	Population (s)	Frequency (%)
1	*1	rs59113707~76362C>G(F400L)	Missense (F400L)	MSL	2.4
2	*14	rs59113707~76362C>G(F400L)	Missense (F400L)	MSL	1.2
				SEB	0.5
				BEN	1
				CMRN	1
				ACB	0.5
3	*14	rs34671512~112849A>C(L643F)	Missense (L643F)	SGDP	1.6
4	*14	rs74064213~79838A>G(I499V)	Missense (I499V)	SEB	0.3
5	*15	rs34671512~112849A>C(L643F)	Missense (L643F)	SEB	0.3
				EUR	0.1
6	*20	rs11045852~70758A>G(I245V)	Missense (I245V)	SAS	0.2
7	*27	rs79135870~52764A>G(I222V)	Missense (I222V)	ACB	0.5
8	*37	rs71581941~96162C>T(R580X)	Stop-gain	EAS	0.2

MSL: Mende in Sierra Leone; SGDP: Simons Genome Diversity Project; SEB: Southeastern Bantu-speaking; BEN: Benin; CMRN: Cameroon; EUR: European; EAS: East Asian; SAS: South Asian; ACB: Afro-Caribbean

Functional impact of variants and star alleles

Of the 16 variants characterised in this study, only three (rs4149056, rs11045853, rs59502379) had *in silico* scores predicting them to have a deleterious effect (Supplementary Table 3); the remaining 13 variants were either classified as benign or had uncertain effects.

The DDGun analysis predicts the significance of impact on the protein structure by calculating the variation of the Gibbs free energy using electron microscopy (EM) and AlphaFold structural models, accounting for regions missing in the EM structure (Figure 5). The linear regression between these two datasets showed a strong correlation (R^2 = 0.9), indicating consistent predictions of the mutations' impact on protein stability.

All DDGun scores for SSA and other global populations were either negative or neutral. Negative scores point to potential destabilizing effects on the OATP1B1 protein. Of the variants (Figure 5a), rs72559747, which is one of the defining variants of **47* and located in a transmembrane domain (Figure 7), exhibited destabilizing scores of -2.3 (EM) and -2.5 (AlphaFold). The most pronounced destabilization for haplotypes (Figure 5b) was observed for the **5* allele, located in a transmembrane domain (Figure 7) with a score of -1.3, which aligns with its no-function phenotype. For scores closer to zero, which were closely associated with alleles exhibiting indeterminate phenotypes, the impact on stability appears minimal or neutral.

Among the inferred novel haplotypes observed in both SSA and global populations, all associated variants were predicted to be benign based on *in silico* functional prediction scores (Supplementary Table 2). This suggests that these variants may not significantly affect protein function under typical conditions. However, a subset of these variants showed evidence of a destabilizing effect on the OATP1B1 protein structure, as indicated by DDGun stability scores.

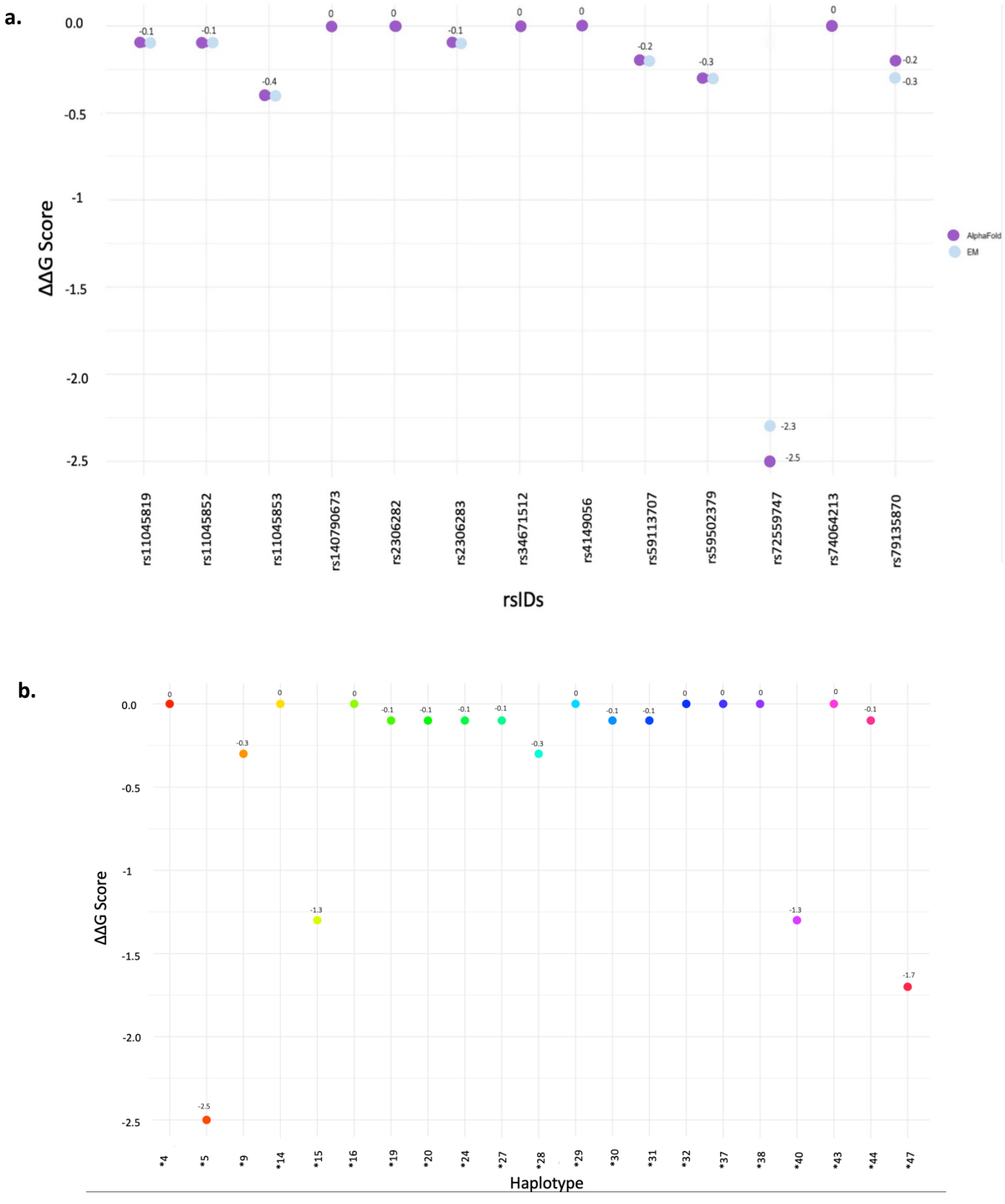


Figure 5: Predicted effect of variation on OATP1B1 based on DDGun scores. a. DDGun scores for missense variants (x-axis) on the electron microscopy (EM) and AlphaFold models. **b.** DDGun scores for star alleles/haplotypes (x-axis). For both a and b, the y-axis displays the DDGun stability scores. Variants with scores below zero are predicted to be destabilizing, whereas variants with a score of 0, are predicted to have a neutral effect.

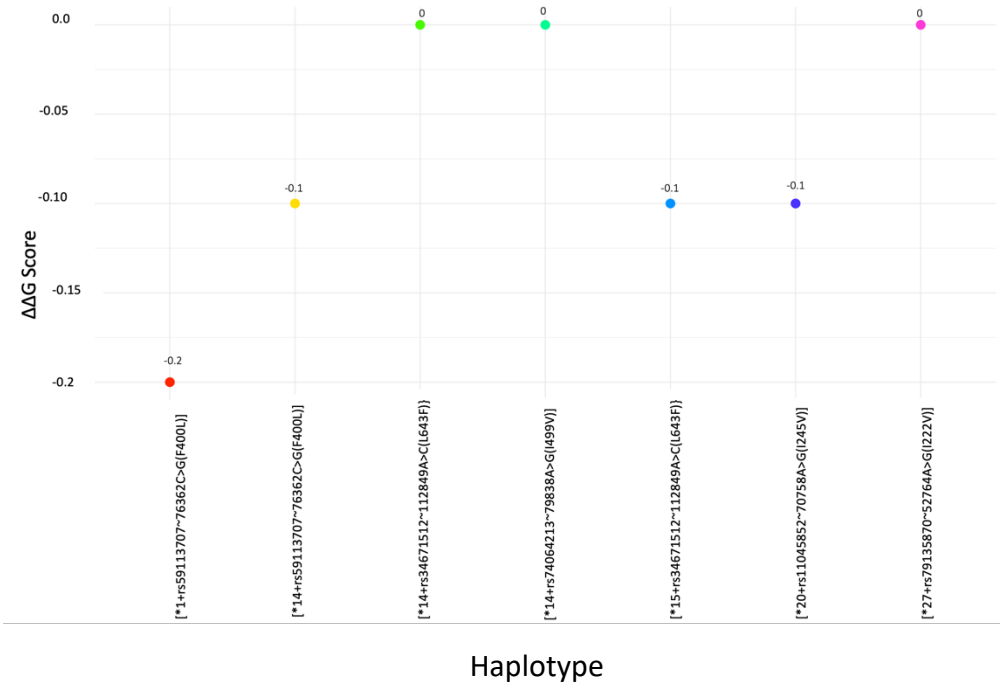


Figure 6: Predicted stability scores of potential novel haplotypes on OATP1B1.

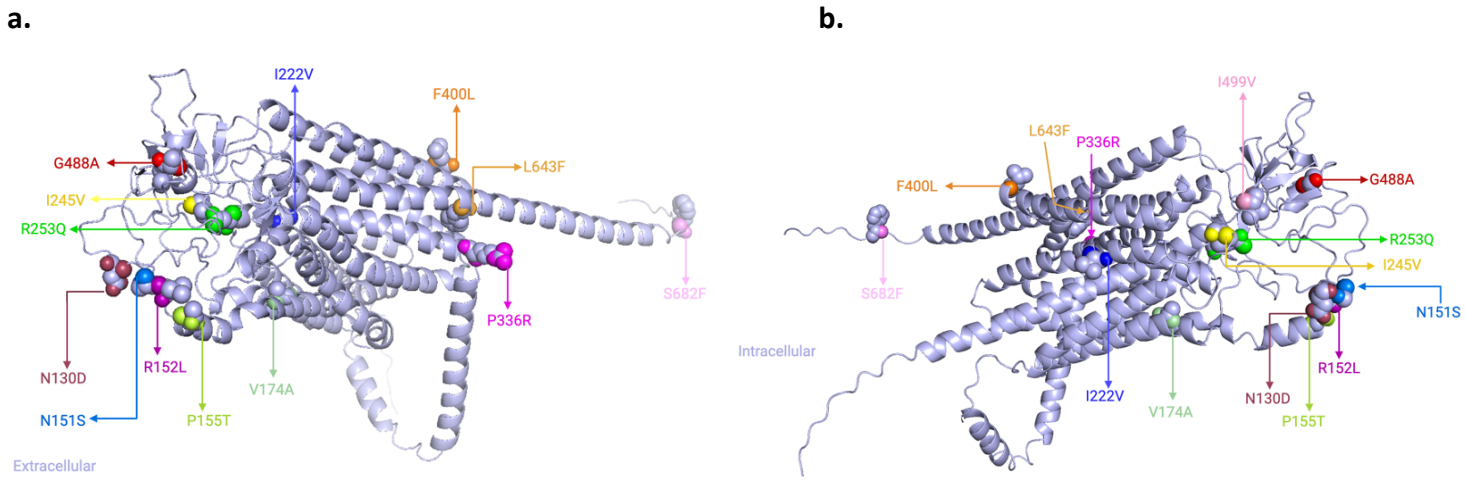


Figure 7: Structural representation of the OATP1B1 protein with key variants mapped. a. Extracellular View: This view shows the structure of OATP1B1 as seen from the extracellular side, with key amino acid substitutions highlighted. Important variants include **F400L** (orange), **S682F** (pink), **L643F** (orange), **P336R** (purple), **I499V** (pink), **G488A** (red), **R253Q** (green), **I245V** (yellow), **N151S** (blue), **R152L** (brown), **N130D** (maroon), **P155T** (purple), **V174A** (green), and **I222V** (blue). Each variant is labelled with its respective amino acid change and coloured for easy identification. **b. Intracellular View:** The structure of OATP1B1 from the intracellular side, showing the same set of variants with corresponding labels and colours. The intracellular perspective provides insight into the spatial orientation of these variants relative to the intracellular environment. Each colour-coded arrow corresponds to a specific variant, aiding in the visualization of their locations within the 3D structure of OATP1B1. These variants may influence the function or structural stability of the transporter, contributing to interindividual differences in drug transport activity.

Discussion

In this study, an extensive analysis of *SLCO1B1* genetic variation in multiple SSA ethnolinguistic groups was conducted, examining star alleles across these groups to uncover patterns unique to the African pharmacogenomic landscape. The focus on SSA ethnolinguistic groups stems from the underrepresentation of these groups in global genomic studies (Corpas et al., 2024), and the increase in burden of hypercholesterolemia and other non-communicable diseases (Cassambai et al., 2024) , where *SLCO1B1* gene-drug pairs are relevant .

*SLCO1B1*37*, an allele associated with normal transporter function, was observed at frequencies of 50% or higher in all African ethnolinguistic groups analysed in this study. This prevalence contrasts sharply with the lower frequencies observed in European populations, where *SLCO1B1*1*—a different normal-function allele—is more common. Additional alleles with reduced or no function were also observed at low frequencies in African populations. For instance*, SLCO1B1*5* and *SLCO1B1*15,* which are known to reduce OATP1B1 function and increase plasma statin levels (Choi et al., 2012; Kane, 2012; Voora et al., 2022) were rare in African populations compared to much higher frequencies in European populations (Tables 3 and 4) in this study. This aligns with findings from other studies (Chigutsa et al., 2011; Pasanen et al., 2008; Rajman et al., 2020), which have noted the rarity of **5* in Sub-Saharan Africa, further reinforcing our results. However, the accuracy of these frequency estimates may be limited due to small sample sizes and the potential for inaccurate haplotype assignments, particularly in admixed populations. Interestingly, while the variant, rs4149056 (c.521T>C; V174A), which defines *SLCO1B1*5,* is commonly observed in many haplotype combinations, **5* occurred less frequently. Moreover, in populations where genetic linkage is weaker or more recombination has occurred, such as in African populations, the frequency of *SLCO1B1*5* may be further diminished due to its reliance on specific variant combinations to form the haplotype (Santos et al., 2011). This observation underscores the complex interplay between genetic diversity and recombination events in shaping haplotype frequencies, particularly in ancestrally diverse populations.

In contrast, other African-specific alleles, such as *SLCO1B1*20*, were observed at higher frequencies and are linked to increased transporter function, adding a layer of complexity to drug response predictions within African populations. The higher activity associated with **20* could potentially result in faster hepatic clearance of certain drugs, which could warrant dose

adjustments to ensure therapeutic efficacy or prevent toxicity (Radouani et al., 2020). Such population-specific patterns underscore the importance of characterizing allele distributions in diverse groups to improve pharmacogenomic precision and optimize therapeutic dosing (Radouani et al., 2020)

We observed a relatively high prevalence of indeterminate phenotypes in African populations, some which were driven by the presence of novel star alleles. Examining both *in silico* scores and $\Delta\Delta$G scores helped to estimate functional impact of possibly novel haplotypes and existing alleles with unknown functions. DDGun utilizes a sequence-structure-based approach to assess how missense variants affect protein stability, enabling a detailed understanding of the structural consequences of these mutations. This analysis is predicated on the hypothesis that alterations in protein structure due to these variants may provide insights into their potential effects on protein function. For instance, if the $\Delta\Delta$G score is significantly decreased (<0), the mutation likely destabilises the protein, potentially reducing its functional efficiency (Montanucci et al., 2019). Conversely, a slight decrease, or minimal change in $\Delta\Delta$G suggest that the mutation may not critically impact stability, but could alter binding affinity or specificity (Montanucci et al., 2019).

For example, *SLCO1B1*5 and *15,* with a $\Delta\Delta$G score of -2.5 and -1.3 respectively, showed substantial destabilising effects on OATP1B1. This supports clinical observations that carriers of *SLCO1B1*5* and **15* experience no transporter function and are at an increased risk of drug toxicity (C.-I. Choi et al., 2012; Kane, 2012). The structural instability of these star alleles is likely due to the amino acid substitution at position 174, changing valine (V174) to alanine (Ala); as both star alleles contain this variant. Valine, a nonpolar and hydrophobic amino acid, is replaced by alanine, which is also nonpolar but smaller in size and slightly less hydrophobic. This alteration in side chain size can create subtle structural gaps in the protein, potentially affecting the stability of the transmembrane helix and decreasing its rigidity, which is important for the efficient uptake of drugs such as statins into hepatocytes (Sainz and Serra, 2023).

These structural alterations could have a significant effect on the pharmacokinetics of drugs. A destabilized OATP1B1 protein may exhibit reduced transport efficiency, leading to decreased hepatic uptake of statins and other substrates (Shan et al., 2023; Tirona et al., 2001). This

reduction in transporter function can result in higher plasma concentrations of these drugs, thereby increasing the risk of dose-dependent adverse effects such as statin-induced myopathy. Additionally, mutations that alter binding affinity or specificity without substantially affecting protein stability might change the substrate selectivity of OATP1B1 (Kane, 2012). This can lead to variable drug responses among individuals carrying these variant. For example, impaired transport of statins into hepatocytes can not only elevate systemic drug levels but also reduce the drug's therapeutic action within the liver, necessitating dosage adjustments to achieve optimal efficacy while minimizing adverse effects (Sainz and Serra, 2023).

To build on this discussion of novel alleles and their structural implications, we identified rare, potentially African-specific variants, such as F400L and L643F, that further highlight the unique genetic diversity in sub-Saharan populations and underscore the need to explore how these variants may influence transporter activity and drug response. Both these variants showed moderate destabilising effects that suggest potential alterations in transporter activity. Phenylalanine (F) to leucine (L) at position 400 represents a shift from a bulky aromatic residue to a smaller, nonpolar one (Chen et al., 2024). This change could reduce the structural integrity of the protein by altering hydrophobic packing within the transmembrane helices, potentially leading to mild destabilization Leucine at position 643, however, is replaced by phenylalanine, introducing an aromatic ring where it may create steric hindrance or change the local hydrophobic environment, which could influence the transporter's efficiency in binding certain substrates (Ramsey et al., 2012; Soko et al., 2018). While these variants are not yet characterized experimentally and clinically, *in silico* predictions suggest that these structural changes may affect transporter activity in ways that could alter drug response or bioavailability in African individuals carrying these variants.

While this study provides valuable insights into *SLCO1B1* variation in SSA, there were limitations. Firstly, the relatively small sample sizes for some populations may impact the robustness of allele frequency comparisons. Moreover, the WGS data used in this study was generated using short read technology, which presented a challenge in analysing and confirming the presence of structural variants. Lastly, the reliance on *in silico* tools and structural bioinformatics for functional assessment, while informative, cannot entirely replace experimental validation and clinical observations. It is also important to note that response to

statins and other medicines reliant on OATP1B1 in their pathway, as well as clinical outcomes, can be influenced by other factors such as age, environmental exposures, diet and lifestyle, medication adherence, and drug-drug interactions, all of which contribute to the complexity of pharmacogenomic studies (Banerjee et al., 2019; Drenth-van Maanen et al., 2020)

Future research should expand on African-specific cohorts, ideally incorporating functional assays, like protein expression and drug transport assays to validate the structural effects of identified *SLCO1B1* variants and star alleles.

Conclusion

This study underscores the unique pharmacogenomic landscape of SSA populations by highlighting the distinct allele distributions compared to other global populations, and potential functional implications, and clinical relevance of *SLCO1B1* variants. The high prevalence of normal function alleles such as *SLCO1B1*37* reflects African genetic diversity. Moreover, the identification of novel African-specific alleles emphasize the need the for more comprehensive and functional studies that include diverse populations to better understand the genetic factors influencing drug response. Characterizing *SLCO1B1* genetic diversity in African populations lays a foundation for experimental and clinical functional studies to assay the impact on response to statins and other medications with a view towards supporting the broader application of precision medicine within African populations.

Future research

Future research could expand upon these findings by examining additional transporter genes beyond *SLCO1B1,* as many other genes within the *SLC* and *ABC* transporter families also play significant roles in drug absorption and distribution. Investigating a broader range of transporter genes could offer a more comprehensive understanding of pharmacogenomic variability and further enhance precision medicine for African populations. However, due to the three-month time constraint of this master's research, focusing solely on *SLCO1B1* allowed for a more in-depth analysis within the limited timeframe.

Data Availability Statement

The 1000 Genomes Project (www.internationalgenome.org/) and SGDP datasets

(https://www.simonsfoundation.org/simons-genome-diversity-project/) are publicly available. The high coverage genomes from the SAHGP (n=15) (EGAD00001003791) and the Africa Wits-INDEPTH Partnership for Genomic Research study (n=159) (EGAD00001006418, EGAD00001004448) are available from the European Genome–Phenome Archive (https://ega-archive.org/).

Code Availability Statement

The StellarPGx code used in this study can be found at https://github.com/SBIMB/StellarPGx. Scripts used for statistical analysis are available at https://github.com/yumnadp/SLCO1B1.

Author Contributions

Y Dadi Patel, D Twesigomwe and H Othman designed the study. Y Dadi Patel analysed the data and drafted the manuscript. Y Dadi Patel, D Twesigomwe and H Othman contributed to interpretation of the study findings. Y Dadi Patel, D Twesigomwe and H Othman revised the manuscript for important intellectual content. D Twesigomwe and H Othman supervised the research. All authors approved the final version of the manuscript.

Funding

This research was supported by the Wits Postgraduate Merit Award, and the Wits Postgraduate Scholarship through the Kirsch Foundation.

References

Adzhubei, I., Jordan, D.M., Sunyaev, S.R., 2013. Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. Chapter 7, Unit7.20. https://doi.org/10.1002/0471142905.hg0720s76

Ahmed, S., Zhou, Z., Zhou, J., Chen, S.-Q., 2016. Pharmacogenomics of Drug Metabolizing Enzymes and Transporters: Relevance to Precision Medicine. Genomics Proteomics Bioinformatics 14, 298–313. https://doi.org/10.1016/j.gpb.2016.03.008

Aszalos, A., 2007. Drug-drug interactions affected by the transporter protein, P-glycoprotein (ABCB1, MDR1) II. Clinical aspects. Drug Discov. Today 12, 838–843. https://doi.org/10.1016/j.drudis.2007.07.021

Banerjee, B.D., Kumar, R., Thamineni, K.L., Shah, H., Thakur, G.K., Sharma, T., 2019. Effect of Environmental Exposure and Pharmacogenomics on Drug Metabolism. Curr. Drug Metab. 20, 1103–1113. https://doi.org/10.2174/1389200221666200110153304

Beloor Suresh, A., Rosani, A., Patel, P., Wadhwa, R., 2024. Rifampin, in: StatPearls. StatPearls Publishing, Treasure Island (FL).

Bruckmueller, H., Cascorbi, I., 2021. ABCB1, ABCG2, ABCC1, ABCC2, and ABCC3 drug transporter polymorphisms and their impact on drug bioavailability: what is our current understanding? Expert Opin. Drug Metab. Toxicol. 17, 369–396. https://doi.org/10.1080/17425255.2021.1876661

Byrska-Bishop, M., Evani, U.S., Zhao, X., Basile, A.O., Abel, H.J., Regier, A.A., et al., 2022. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell 185, 3426-3440.e19. https://doi.org/10.1016/j.cell.2022.08.004

Cassambai, S., Tetteh, J., Highton, P., Kunutsor, S.K., Darko, D.O., Jeffers, S., et al., 2024. Prevalence of Cardiometabolic Diseases in Sub-Saharan Africa: A Systematic Review and Meta-Analysis. https://doi.org/10.2139/ssrn.5017937

CDC, 2023. Heart Disease Facts | cdc.gov [WWW Document]. Cent. Dis. Control Prev. URL https://www.cdc.gov/heartdisease/facts.htm (accessed 2.28.24).

Chen, S., Jin, C., Ohgaki, R., Xu, M., Okanishi, H., Kanai, Y., 2024. Structure–activity characteristics of phenylalanine analogs selectively transported by L-type amino acid

transporter 1 (LAT1). Sci. Rep. 14, 4651. https://doi.org/10.1038/s41598-024-55252-w

Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., et al., 2023. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science 381, eadg7492. https://doi.org/10.1126/science.adg7492

Chigutsa, E., Visser, M.E., Swart, E.C., Denti, P., Pushpakom, S., Egan, D., et al., 2011. The SLCO1B1 rs4149032 Polymorphism Is Highly Prevalent in South Africans and Is Associated with Reduced Rifampin Concentrations: Dosing Implications ▿ . Antimicrob. Agents Chemother. 55, 4122–4127. https://doi.org/10.1128/AAC.01833-10

Chiu, H.-F., Venkatakrishnan, K., Golovinskaia, O., Wang, C.-K., 2021. Impact of Micronutrients on Hypertension: Evidence from Clinical Trials with a Special Focus on Meta-Analysis. Nutrients 13, 588. https://doi.org/10.3390/nu13020588

Choi, C.-I., Lee, Y.-J., Lee, H.-I., Kim, B.-H., Kim, M.-J., Jang, C.-G., et al., 2012. Effects of the SLCO1B1*15 allele on the pharmacokinetics of pitavastatin. Xenobiotica 42, 496–501. https://doi.org/10.3109/00498254.2011.632030

Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., Chan, A.P., 2012. Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE 7, e46688. https://doi.org/10.1371/journal.pone.0046688

Choudhury, A., Aron, S., Botigué, L.R., Sengupta, D., Botha, G., Bensellak, T., et al., 2020. High-depth African genomes inform human migration and health. Nature 586, 741–748. https://doi.org/10.1038/s41586-020-2859-7

Choudhury, A., Ramsay, M., Hazelhurst, S., Aron, S., Bardien, S., Botha, G., et al., 2017. Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans. Nat. Commun. 8, 2062. https://doi.org/10.1038/s41467-017-00663-9

Ciută, A.-D., Nosol, K., Kowal, J., Mukherjee, S., Ramírez, A.S., Stieger, B., et al., 2023. Structure of human drug transporters OATP1B1 and OATP1B3. Nat. Commun. 14, 5774. https://doi.org/10.1038/s41467-023-41552-8

Cooper-DeHoff, R.M., Niemi, M., Ramsey, L.B., Luzum, J.A., Tarkiainen, E.K., Straka, R.J., et al., 2022. The Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for SLCO1B1, ABCG2, and CYP2C9 and statin-associated musculoskeletal symptoms. Clin. Pharmacol. Ther. 111, 1007. https://doi.org/10.1002/cpt.2557

Corpas, M., Siddiqui, M.K., Soremekun, O., Mathur, R., Gill, D., Fatumo, S., 2024. Addressing Ancestry and Sex Bias in Pharmacogenomics. Annu. Rev. Pharmacol. Toxicol. 64, 53–64. https://doi.org/10.1146/annurev-pharmtox-030823-111731

da Rocha, J.E.B., Othman, H., Botha, G., Cottino, L., Twesigomwe, D., Ahmed, S., et al., 2021. The Extent and Impact of Variation in ADME Genes in Sub-Saharan African Populations. Front. Pharmacol. 12.

Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., Notredame, C., 2017. Nextflow enables reproducible computational workflows. Nat. Biotechnol. 35, 316–319. https://doi.org/10.1038/nbt.3820

Drenth-van Maanen, A.C., Wilting, I., Jansen, P.A.F., 2020. Prescribing medicines to older people—How to consider the impact of ageing on human organ and body functions. Br. J. Clin. Pharmacol. 86, 1921–1930. https://doi.org/10.1111/bcp.14094

Eggertsson, H.P., Jonsson, H., Kristmundsdottir, S., Hjartarson, E., Kehr, B., Masson, G., et al., 2017. Graphtyper enables population-scale genotyping using pangenome graphs. Nat. Genet. 49, 1654–1660. https://doi.org/10.1038/ng.3964

Fohner, A.E., Brackman, D.J., Giacomini, K.M., Altman, R.B., Klein, T.E., 2017. PharmGKB summary: very important pharmacogene information for ABCG2. Pharmacogenet. Genomics 27, 420–427. https://doi.org/10.1097/FPC.0000000000000305

Fujita, A., Kojima, K., Patriota, A.G., Sato, J.R., Severino, P., Miyano, S., 2010. A fast and robust statistical test based on likelihood ratio with Bartlett correction to identify Granger causality between gene sets. Bioinforma. Oxf. Engl. 26, 2349–2351. https://doi.org/10.1093/bioinformatics/btq427

Gaedigk, A., Casey, S.T., Whirl-Carrillo, M., Miller, N.A., Klein, T.E., 2021. Pharmacogene Variation Consortium: A Global Resource and Repository for Pharmacogene Variation. Clin. Pharmacol. Ther. 110, 542–545. https://doi.org/10.1002/cpt.2321

Kane, M., 2012. Simvastatin Therapy and SLCO1B1 Genotype, in: Pratt, V.M., Scott, S.A., Pirmohamed, M., Esquivel, B., Kattman, B.L., Malheiro, A.J. (Eds.), Medical Genetics Summaries. National Center for Biotechnology Information (US), Bethesda (MD).

Klein, K., Tremmel, R., Winter, S., Fehr, S., Battke, F., Scheurenbrand, T., et al., 2019. A New Panel-Based Next-Generation Sequencing Method for ADME Genes Reveals Novel Associations of Common and Rare Variants With Expression in a Human Liver Cohort. Front. Genet. 10.

Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–1081. https://doi.org/10.1038/nprot.2009.86

Li, B., Sangkuhl, K., Whaley, R., Woon, M., Keat, K., Whirl-Carrillo, M., et al., 2023. Frequencies of pharmacogenomic alleles across biogeographic groups in a large-scale biobank. Am. J. Hum. Genet. 110, 1628–1647. https://doi.org/10.1016/j.ajhg.2023.09.001

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lin, L., Yee, S.W., Kim, R.B., Giacomini, K.M., 2015. SLC Transporters as Therapeutic Targets: Emerging Opportunities. Nat. Rev. Drug Discov. 14, 543–560. https://doi.org/10.1038/nrd4626

Lozano, E., Briz, O., Macias, R.I.R., Serrano, M.A., Marin, J.J.G., Herraez, E., 2018. Genetic Heterogeneity of SLC22 Family of Transporters in Drug Disposition. J. Pers. Med. 8, 14. https://doi.org/10.3390/jpm8020014

MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L., Scherer, S.W., 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. 42, D986-992. https://doi.org/10.1093/nar/gkt958

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., et al., 2016. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538, 201–206. https://doi.org/10.1038/nature18964

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., et al., 2016. The Ensembl Variant Effect Predictor. Genome Biol. 17, 122. https://doi.org/10.1186/s13059-016-0974-4

Montanucci, L., Capriotti, E., Frank, Y., Ben-Tal, N., Fariselli, P., 2019. DDGun: an untrained method for the prediction of protein stability changes upon single and multiple point variations. BMC Bioinformatics 20, 335. https://doi.org/10.1186/s12859-019-2923-1

Neuvonen, M., Hirvensalo, P., Tornio, A., Rago, B., West, M., Lazzaro, S., et al., 2021. Identification of Glycochenodeoxycholate 3-O-Glucuronide and Glycodeoxycholate 3-

O-Glucuronide as Highly Sensitive and Specific OATP1B1 Biomarkers. Clin. Pharmacol. Ther. 109, 646–657. https://doi.org/10.1002/cpt.2053

Ng, P.C., Henikoff, S., 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 31, 3812–3814. https://doi.org/10.1093/nar/gkg509

Park, J.H., Hogrebe, M., Grüneberg, M., DuChesne, I., von der Heiden, A.L., Reunert, J., et al., 2015. SLC39A8 Deficiency: A Disorder of Manganese Transport and Glycosylation. Am. J. Hum. Genet. 97, 894–903. https://doi.org/10.1016/j.ajhg.2015.11.003

Pasanen, M.K., Neuvonen, P.J., Niemi, M., 2008. Global Analysis of Genetic Variation in SLCO1B1. Pharmacogenomics.

PharmVar [WWW Document], n.d. URL https://www.pharmvar.org/gene/SLCO1B1 (accessed 2.28.24).

Pizzagalli, M.D., Bensimon, A., Superti-Furga, G., 2021. A guide to plasma membrane solute carrier proteins. Febs J. 288, 2784–2835. https://doi.org/10.1111/febs.15531

Pymol: an open-source molecular graphics tool. – ScienceOpen [WWW Document], n.d. URL https://www.scienceopen.com/document?vid=4362f9a2-0b29-433f-aa65-51db01f4962f (accessed 11.5.24).

Radouani, F., Zass, L., Hamdi, Y., Rocha, J.D., Sallam, R., Abdelhak, S., et al., 2020. A Review of Clinical Pharmacogenetics Studies in African Populations. Pers. Med. 17, 155–170. https://doi.org/10.2217/pme-2019-0110

Rajman, I., Knapp, L., Hanna, I., 2020. Genetic Diversity in Drug Transporters: Impact in African Populations. Clin. Transl. Sci. 13, 848–860. https://doi.org/10.1111/cts.12769

Rajman, I., Knapp, L., Morgan, T., Masimirembwa, C., 2017. African Genetic Diversity: Implications for Cytochrome P450-mediated Drug Metabolism and Drug Development. EBioMedicine 17, 67–74. https://doi.org/10.1016/j.ebiom.2017.02.017

Ramsay, M., Crowther, N., Tambo, E., Agongo, G., Baloyi, V., Dikotope, S., et al., 2016. H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. Glob. Health Epidemiol. Genomics 1, e20. https://doi.org/10.1017/gheg.2016.17

Ramsey, L.B., Bruun, G.H., Yang, W., Treviño, L.R., Vattathil, S., Scheet, P., et al., 2012. Rare versus common variants in pharmacogenetics: *SLCO1B1* variation and methotrexate disposition. Genome Res. 22, 1–8. https://doi.org/10.1101/gr.129668.111

Ramsey, L.B., Gong, L., Lee, S.-B., Wagner, J.B., Zhou, X., Sangkuhl, K., et al., 2023. PharmVar GeneFocus: SLCO1B1. Clin. Pharmacol. Ther. 113, 782–793. https://doi.org/10.1002/cpt.2705

Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., Kircher, M., 2019. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 47, D886–D894. https://doi.org/10.1093/nar/gky1016

Reva, B., Antipin, Y., Sander, C., 2011. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 39, e118. https://doi.org/10.1093/nar/gkr407

Roden, D.M., McLeod, H.L., Relling, M.V., Williams, M.S., Mensah, G.A., Peterson, J.F., et al., 2019. Pharmacogenomics. Lancet Lond. Engl. 394, 521–532. https://doi.org/10.1016/S0140-6736(19)31276-0

Sainz, J.I.S. de M., Serra, M.B., 2023. Influence of pharmacogenetics on the diversity of response to statins associated with adverse drug reactions. Adv. Lab. Med. Av. En Med. Lab. 4, 341–352. https://doi.org/10.1515/almed-2023-0123

Sakurai, A., Tamura, A., Onishi, Y., Ishikawa, T., 2005. Genetic polymorphisms of ATP-binding cassette transporters ABCB1 and ABCG2: therapeutic implications. Expert Opin. Pharmacother. 6, 2455–2473. https://doi.org/10.1517/14656566.6.14.2455

Santos, P.C., Soares, R.A., Nascimento, R.M., Machado-Coelho, G.L., Mill, J.G., Krieger, J.E., et al., 2011. SLCO1B1 rs4149056 polymorphism associated with statin-induced myopathy is differently distributed according to ethnicity in the Brazilian general population: Amerindians as a high risk ethnic group. BMC Med. Genet. 12, 136. https://doi.org/10.1186/1471-2350-12-136

Schärfe, C.P.I., Tremmel, R., Schwab, M., Kohlbacher, O., Marks, D.S., 2017. Genetic variation in human drug-related genes. Genome Med. 9, 117. https://doi.org/10.1186/s13073-017-0502-5

Shan, Z., Yang, X., Liu, H., Yuan, Y., Xiao, Y., Nan, J., et al., 2023. Cryo-EM structures of human organic anion transporting polypeptide OATP1B1. Cell Res. 33, 940–951. https://doi.org/10.1038/s41422-023-00870-8

Sissung, T.M., Goey, A.K.L., Ley, A.M., Strope, J.D., Figg, W.D., 2014. Pharmacogenetics of Membrane Transporters: A Review of Current Approaches. Methods Mol. Biol. Clifton NJ 1175, 91–120. https://doi.org/10.1007/978-1-4939-0956-8_6

SLCO1B1 [WWW Document], n.d. . PharmGKB. URL https://www.pharmgkb.org/gene/PA134865839/prescribingInfo#guideline-annotations (accessed 11.5.24).

Soko, N.D., Masimirembwa, C., Dandara, C., 2018. A cost effective RFLP method to genotype Solute carrier organic anion 1B1 (SLCO1B1) c.1929A>C (p.Leu643Phe, rs34671512); a variant with potential effect on rosuvastatin pharmacokinetics. BMC Res. Notes 11, 384. https://doi.org/10.1186/s13104-018-3469-4

Swart, M., Ren, Y., Smith, P., Dandara, C., 2012. ABCB1 4036A>G and 1236C>T Polymorphisms Affect Plasma Efavirenz Levels in South African HIV/AIDS Patients. Front. Genet. 3, 236. https://doi.org/10.3389/fgene.2012.00236

Tirona, R.G., Leake, B.F., Merino, G., Kim, R.B., 2001. Polymorphisms in OATP-C. J. Biol. Chem. 276, 35669–35675. https://doi.org/10.1074/jbc.M103792200

Twesigomwe, D., Drögemöller, B.I., Wright, G.E.B., Siddiqui, A., da Rocha, J., Lombard, Z., et al., 2021. StellarPGx: A Nextflow Pipeline for Calling Star Alleles in Cytochrome P450 Genes. Clin. Pharmacol. Ther. 110, 741–749. https://doi.org/10.1002/cpt.2173

Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., et al., 2024. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. Nucleic Acids Res. 52, D368–D375. https://doi.org/10.1093/nar/gkad1011

Voora, D., Baye, J., McDermaid, A., Narayana Gowda, S., Wilke, R.A., Nicole Myrmoe, A., et al., 2022. SLCO1B1*5 Allele Is Associated With Atorvastatin Discontinuation and Adverse Muscle Symptoms in the Context of Routine Care. Clin. Pharmacol. Ther. 111, 1075–1083. https://doi.org/10.1002/cpt.2527

Zhou, Y., Ingelman-Sundberg, M., Lauschke, V.M., 2017. Worldwide Distribution of Cytochrome P450 Alleles: A Meta-analysis of Population-scale Sequencing Projects. Clin. Pharmacol. Ther. 102, 688–700. https://doi.org/10.1002/cpt.690

Zhou, Y., Pirmann, S., Lauschke, V.M., 2024. APF2: an improved ensemble method for pharmacogenomic variant effect prediction. Pharmacogenomics J. 24, 1–11. https://doi.org/10.1038/s41397-024-00338-x

Appendices

Appendix A: Supplementary Tables and Figures

Supplimentary Figure 1: Total depth of coverage for *SLCO1B1* and control genes (*EGFR* and *VDR* in individuals where StellarPGx identified occurrences of potential structural variants).

Supplementary Table 1: Assignment of predicted *SLCO1B1* phenotype based on genotype.

Supplementary Table 2: *SLCO1B1* diplotype frequencies across global populations

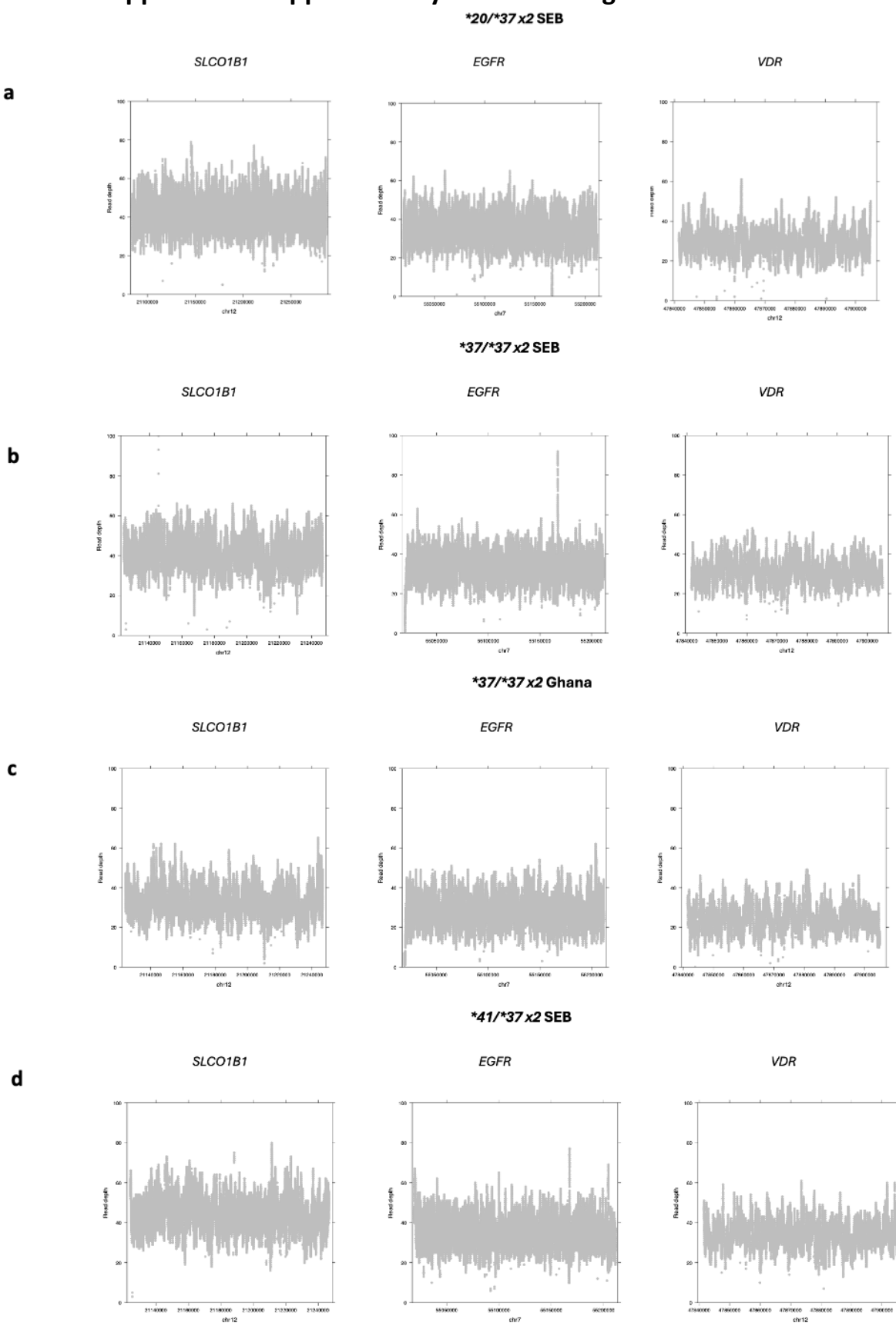
Supplementary Table 3: In silico scores of variants for *SLCO1B1* (according to the ADME optimised framework).

Appendix B: Ethics Clearance certificate

Appendix C: Approved Research Protocol

Appendix D: Plagiarism Declaration and Turnitin Report

Appendix A: Supplementary Tables and Figures



Supplementary Figure 1: Total depth of coverage for *SLCO1B1* and control genes (*EGFR* and *VDR* in individuals where StellarPGx identified occurrences of potential structural variants). (a) An individual from the South African dataset (SEB) had **20* as haplotype 1 and potentially **37*x2 as haplotype 2. (b) An individual from the SEB dataset had **37* as haplotype 1 and potentially **37*x2 as haplotype 2. (c) An individual from the Ghana dataset had **37* as haplotype 1 and potentially **37*x2 as haplotype 2. (d) An individual from the SEB dataset had **41* as haplotype 1 and potentially *37x2 as haplotype 2. These duplications were inferred because the ratio of the coverage across *SLCO1B1* to *EGFR,* and particularly *VDR,* was approximately 3:2.

Supplementary Table 1: Assignment of predicted *SLCO1B1* phenotype based on genotype. Retrieved from Cooper-DeHoff et al., (2022)

Gene	Phenotype	Genotype	Examples of diplotypes
SLCO1B1	Increased function	An individual carrying two increased function alleles	**14/*14*
	Normal function	An individual carrying two normal function alleles or one normal plus one increased function allele	**1/*1, *1/*14*
	Decreased function	An individual carrying one normal or increased function allele plus no function allele.	**1/*5, *1/*15*
	Possible decreased function	An individual carrying one no function allele plus one unknown/uncertain allele	**5/*6, *5/*10, *5/*43*
	Poor function	An individual carrying two no function alleles	**5/*5, *5/*15, *15/*15*
	Indeterminate	An individual carrying one normal function allele plus one uncertain or unknown function allele OR allele combinations with uncertain and/or unknown function alleles	**1/*7, *1/*10, *7/*10*

Supplementary Table 2: *SLCO1B1* diplotype frequencies across global populations

		Frequencies (%)					
Diplotype	CPIC Function	SSA	EUR	EAS	SAS	ASW/ ACB	AMR
*1/*1	Normal	2	32	5.4	21.5	3.2	25.9
*1/*4	Indeterminate	0	0.2	0	0	0	0
*1/*5	Decreased	0	2.4	0	0.6	0	2.9
*1/*9	Decreased	<0.1	0	0	0	0	8.6
*1/*14	Normal	1	16.9	0	2.2	3.2	11
*1/*15	Decreased	<0.1	16.1	5.6	2.5	0.6	0
*1/*16	Indeterminate	0	0	0.2	0	0	0
*1/*19	Indeterminate	0	0.2	0	0	0	0
*1/*20	Normal	2	5.4	0	0.2	1.3	4
*1/*24	Indeterminate	0	0	0	0.2	0	0
*1/*27	Indeterminate	2	0	0	0	3.8	0.6
*1/*28	Indeterminate	<0.1	0	0	0	0	0
*1/*30	Indeterminate	<0.1	0	0	0	0.6	0
*1/*31	Decreased	1	0	0	0	1.9	0
*1/*37	Normal	15	8.2	29	35.6	22.9	20.7
*1/*40	Indeterminate	0	0	0	0	0	0.3
*1/*41	Indeterminate	0	0	0	0	1.3	0.3
*1/*43	Indeterminate	<0.1	0	0	0	0.6	0.3
*1/*45	Indeterminate	0	0	0.4	0	0	0
*14/*5	Decreased	0	1.2	0	0	0	0
*14/*14	Increased	0	1.8	0	0.2	0	2
*14/*15	Decreased	0	4.6	0	0	0	1.7
*14/*19	Indeterminate	0	0.2	0	0	0	0
*14/*20	Increased	<0.1	0.6	0	0	0	1.7
*14/*27	Indeterminate	2	0	0	0	0.6	0
*14/*31	Decreased	<0.1	0	0	0	0.6	0.3
*14/*37	Normal	2	1.2	0.6	2.9	2.5	2.9
*14/*41	Indeterminate	<0.1	0	0	0	0.6	0
*14/*43	Indeterminate	<0.1	0	0	0	0	0
*14/*46	Decreased	0	0.2	0	0	0	0
*15/*5	Poor	0	0.6	0	0	0	0.3
*15/*15	Poor	0	1	1.6	0.4	0.6	1.2
*15/*20	Decreased	0	1.4	0.8	0.2	0.6	0.3
*15/*30	Possible decreased	0	0	0	0	0.6	0
*15/*31	Poor	<0.1	0	0	0	1.9	0.3
*15/*37	Decreased	1	0.8	13.7	3.3	3.2	5.8
*15/*41	Possible decreased	0	0	0	0	0.6	0
*15/*43	Possible decreased	0	0	0	0	0.6	0
*15/*45	Possible decreased	0	0	0.2	0	0	0
*15/*46	Possible decreased	0	0.2	0	0	0	0.3
*15/*47	Poor	0	0	0.4	0	0	0
*16/*37	Indeterminate	0	0	0.6	0	0	0
*19/*24	Indeterminate	0	0	0	4.9	0	0
*19/*37	Indeterminate	<0.1	0.2	0	0	0	0

*20/*5	Decreased	0	0.2	0	0	0	0
*20/*20	Increased	1	0.2	0	0	0.6	0.3
*20/*24	Indeterminate	0	0	0	2.2	0	0
*20/27	Indeterminate	1	0	0	0	0	0
*20/*31	Decreased	<0.1	0	0	0	0.6	0
*20/*32	Indeterminate	0	0	0	0.4	0	0
*20/*37	Normal	9	1.4	0.2	0.6	8.9	2
*20/*41	Indeterminate	<0.1	0	0	0	0.6	0
*20/*43	Indeterminate	1	0	0	0	0.6	0
*24/*37	Indeterminate	0	0	0	0	0	0
*24/*40	Indeterminate	0	0	0	0.4	0	0
*27/*4	Indeterminate	1	0	0	0	0	0.3
*27/*27	Indeterminate	<0.1	0	0	0	0	0
*27/*30	Indeterminate	<0.1	0	0	0	0.6	0
*27/*31	Possible decreased	1	0	0	0	0	0
*27/*37	Indeterminate	4	0	0	0	3.2	0.3
*27/*41	Indeterminate	<0.1	0	0	0	0	0
*27/*44	Indeterminate	<0.1	0	0	0	0	0
*28/*37	Indeterminate	<0.1	0	0	0	0.6	0
*29/*37	Indeterminate	0	0	0.4	0	0	0
*30/*31	Possible decreased	<0.1	0	0	0	0	0
*30/*37	Indeterminate	2	0	0	0	1.3	0
*30/*43	Indeterminate	0	0	0	0	0.6	0
*31/*31	Poor	<0.1	0	0	0	0	0
*31/*37	Decreased	0	0	0	0	3.8	0
*31/*41	Possible decreased	3	0	0	0	0	0
*31/*43	Possible decreased	<0.1	0	0	0	0.6	0
*37/*5	Decreased	0	1.2	0	0.2	0	0.6
*37/*37	Normal	39	1.4	39.4	20.4	20.4	4.9
*37/*38	Indeterminate	<0.1	0	0	0	0	0
*37/*39	Indeterminate	0	0	0.4	0	0	0
*37/*40	Indeterminate	0	0.2	0	0	0	0
*37/*41	Indeterminate	4	0	0	0	1.9	0
*37/*43	Indeterminate	3	0	0	0	4.5	0
*37/*44	Indeterminate	<0.1	0	0	0	0	0
*37/*45	Indeterminate	0	0	0.2	0	0	0
*37/*46	Decreased	0	0	0	0	0	0.3
*37/*47	Indeterminate	0	0	0.4	0	0	0
*41/*41	Indeterminate	<0.1	0	0	0	0	0
*41/*43	Indeterminate	<0.1	0	0	0	0	0
*43/*43	Indeterminate	<0.1	0	0	0	0	0
*45/*45	Indeterminate	0	0	0.2	0	0	0
*48/*20	Decreased	<0.1	0	0	0	0	0
*48/*27	Possible decreased	<0.1	0	0	0	0	0
*48/*31	Poor	<0.1	0	0	0	0	0
*48/*43	Possible decreased	<0.1	0	0	0	0	0

CPIC: Clinical Pharmacogenomics Implementation consortiums; SSA: Sub-Saharan Africa; EUR: European; EAS: East Asian; SAS: South Asian; ASW: African American; ACB: Afro-Caribbean; AMR: admixed American

Supplementary Table 3: In silico scores of variants for *SLCO1B1* (according to the ADME optimised framework)

Variant	Consequence	SIFT	PolyPhen2	CADD	LRT	PROVEAN	MutationAssessor	AlphaMissense	Consensus
rs11045819	missense				X				
rs4149056	missense		X	X	X	X		X	X
rs2306283	missense								
rs2306282	missense								
rs34671512	missense								
rs11045852	missense				X				
rs11045853	missense	X	X	X	X	X	X		X
rs59113707	missense								
rs140790673	missense								
rs79135870	missense								
rs59502379	missense	X	X	X		X	X	X	X
rs183501729	stop-gain			X	X				
rs74064213	missense								
rs71581941	stop-gain								
rs77271279	splice defect			X					
rs72559747	missense								

The presence of an 'X' indicates that the variant is predicted to be deleterious according to the in silico tool. The ADME-optimized thresholds for deleteriousness are as follows: CADD > 19.19, SIFT < 0.0376, PolyPhen-2 > 0.3841, LRT < 0.0025, PROVEAN <-3.286, and MutationAssessor >2.

In this study, we employed a consensus-based approach to classify the functional impact of genetic variants, utilizing multiple *in silico* prediction tools. A variant was deemed deleterious if at least half of these tools had scores indicating a damaging or harmful effect on protein structure and/or function. This majority-rule criterion was selected to balance sensitivity and specificity, acknowledging the diverse methodologies and potential biases inherent in each tool. This approach is particularly relevant in the context of pharmacogenes, which often exhibit high genetic variability, making their interpretation more challenging than typical Mendelian disease genes. This methodology aligns with the framework proposed by Zhou et al., (2024). The authors found that standard computational tools, often optimized for highly conserved disease-associated genes, underperformed when applied to pharmacogenes, which are subject to different selective pressures and functional constraints.

For variants that did not meet the consensus threshold for being classified as deleterious, meaning fewer than half of the in silico tools predicted them to be damaging, this was because the majority of scores instead indicated a benign effect. These variants were, therefore, classified as non-deleterious. As the binary classification of 'deleterious' and 'non-deleterious' is limiting with regard to characterising the impact of the variants on drug response, we leveraged structural bioinformatics-based analyses to gain more insights on their effect as described in the Results.

Appendix B: Ethics Clearance Certificate

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG	HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

Office of the Deputy Vice-Chancellor (Research and Innovation)

TO: Ms Y Dadi Patel
School of Pathology
Division of Human Genetics
Faculty of Health Sciences

E-mail: 2303732@students.wits.ac.za

CC: Supervisor: Drs D Twesigomwe and H Othman
David.Twesigomwe@wits.ac.za
and <HREC-Medical Research Office@wits.ac.za>

FROM: Mr Iain Burns
Human Research Ethics Committee (Medical)
Tel: 011 717 1252

E-mail: Iain.Burns@wits.ac.za

DATE: 1 August 2024

REF: R14/49

PROTOCOL NO: **M240744** (This is your ethics application reference number. Please quote it in all enquiries, oral or written, relating to this study.)

PROJECT TITLE: *Pharmacogenomic variation in drug transporter encoding genes across African populations*

Please find attached the Clearance Certificate for the above project. I hope it goes well and that an article in a recognized publication comes out of it. This will reflect well on your professional standing and contribute to Government funding of the University.



MSWorks2000/Iain0007/Clearscan.wps

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R49 Ms Y Dadi Patel

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M240744

NAME: (Principal Investigator) Ms Y Dadi Patel **DEGREE:** MSc (Med)

DEPARTMENT: School of Pathology
Division of Human Genetics
Faculty of Health Sciences

TITLE: *Pharmacogenomic variation in drug transporter encoding genes across African populations*

DATE CONSIDERED: 26 July 2024

DECISION: Approved unconditionally

CONDITIONS: Sub-study under M20/09/93

NOTE: If contact information regarding student study participants is required, please contact the Registrar's office <Nicoleen.Potgieter@wits.ac.za>

SUPERVISORS: Drs D Twesigomwe and H Othman

APPROVED BY:

Professor P Ruff, Chairperson, HREC (Medical)

DATE OF APPROVAL: 1 August 2024 **EXPIRY DATE:** 31 July 2029

This Clearance Certificate is valid for 5 years from the date of approval. An extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office secretariat on the 3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report** in the format available at https://wits.ac.za/research/researcher-support/research-ethics/ethics-committees/. This report is due annually, for the duration of the Clearance Certificate, beginning on the first anniversary of Date of Approval above. Unreported changes to the study may invalidate the clearance given by the HREC (Medical).

Signature of Principal Investigator Date

Appendix C: Approved Research Protocol

UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Private Bag 3 Wits, 2050
Fax: 027117172119
Tel: 02711 7172076

Reference: Mrs Sandra Benn
E-mail: sandra.benn@wits.ac.za

01 July 2024
Person No: 2303732
PAG

Miss Y Dadi Patel
25 Avonwold Road
Saxonwold
2132
South Africa

Dear Miss Yumna Dadi Patel

Master of Science in Medicine: Approval of Title

We have pleasure in advising that your proposal entitled *Pharmacogenomic variation in drug transporter encoding genes across African populations* has been approved. Please note that any amendments to this title have to be endorsed by the Faculty's higher degrees committee and formally approved.

Yours sincerely

Mrs Sandra Benn
Faculty Registrar
Faculty of Health Sciences

Appendix D: Plagiarism Declaration and Turnitin Report

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

FACULTY OF HEALTH SCIENCES

PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I Yumna Dadi Patel (Student number: 2303732) am a student registered for the degree of MSc (Med) Genomic Medicine in the academic year 2024.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included as an appendix a report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document.

Signature: _____ Date: 29/11/2024