

https://dspace.org/

Evaluation of iPSC-derived hepatocyte model to predict functional effects of the CYP2D6*17 genetic variant in activation of tamoxifen to endoxifen

Kariithi, Brian Timothy

2021

https://hdl.handle.net/10539/32535

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

Evaluation of iPSC-derived hepatocyte model to predict functional effects of the *CYP2D6*17* genetic variant in activation of tamoxifen to endoxifen



UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Candidate Name: Brian Timothy Kariithi

Student No.: 608875

Supervisor: Professor Michèle Ramsay

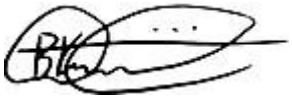
Co-Supervisors: Professor Collen Masimirembwa & Dr Janine Scholefield

Date: 25 March 2021

A dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Medicine in Human Genetics.

DECLARATION

I, Brian Timothy Kariithi, student number 608875, declare that this dissertation is my own work. It is being submitted for the Degree of Master of Science in Medicine in Human Genetics at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

Signature of Student: 

Date: 25th day of March 2021 signed at the University of the Witwatersrand, Johannesburg

PRESENTATIONS FROM THIS RESEARCH REPORT

1. Poster Presentation at the Southern African Society for Human Genetics Biennial Congress, 3-6 August 2019, Cape Town, South Africa.

   Poster title: Evaluation of iPSC-derived hepatocyte model to predict functional effects of the *CYP2D6*17* genetic variant in activation of tamoxifen to endoxifen.

   Authors: Brian Kariithi, Michèle Ramsay, Janine Scholefield and Collen Masimirembwa.

ABSTRACT

BACKGROUND: Drug efficacy can be affected by numerous factors including gene-drug and drug-drug interactions. CYP2D6 is an important enzyme for drug metabolism that catalyses the production of endoxifen from tamoxifen. The African-specific *CYP2D6*17* allele causes a reduced function effect on the CYP2D6 enzyme by decreasing substrate affinity. Various methods have been employed to study enzyme function in a physiologically relevant *in vitro* model. Recently, iPSCs have been used to develop cellular models for liver-like tissue and to study individual differences in drug metabolism and toxicity. **AIM:** The study aimed to assess the frequency of the *CYP2D6*17* variant in African populations and to evaluate its enzyme activity in activation of tamoxifen to endoxifen in iPSC-derived hepatocyte-like cells (SC-HLCs). **METHODS:** *CYP2D6* data was downloaded from the 1000 Genomes Project (KGP), African Genome Variation Project (AGVP) and PharmVar databases to establish haplotype and diplotype frequencies of reduced function variants, including *CYP2D6*17*, in African populations. SNP rs28371706, as a proxy for *CYP2D6*17*, was genotyped and sequenced to identify homozygotes. Skin samples were collected from homozygotes to isolate dermal fibroblasts for the development of induced pluripotent stem cell (iPSC) lines. iPSC lines were used for the creation of stem-cell-derived hepatocyte-like cell (SC-HLC) 3D spheroid models. HepG2 cells were cultured in 3D as spheroids and *CYP3A4* induction was performed using rifampicin incubations for 24-, 48- and 72-hour time periods. CYP3A4 enzyme activity was measured using a CYP3A4 luciferase assay. **RESULTS:** *CYP2D6*17* average haplotype frequencies from KGP, AGVP and PharmVar data for African populations were 20.70%, 19.95% and 20.0% respectively, and *CYP2D6*17* average homozygous diplotype frequencies for KGP, AGVP and PharmVar were 4.28%, 3.98% and 7.81% respectively. Sequencing identified five individuals with a C/C genotype, two were T/T, two were T/C heterozygote, and one was A/C at the rs28371706 SNP locus. HepG2 3D spheroids showed a 7.70% (1.08-fold) (Rifampicin2.5 µM) and a 2.87% (1.03-fold) (Rifampicin 25 µM) increase for the 24-hour normalised enzyme activity readings for *CYP3A4* rifampicin induction. The 72-hour *CYP3A4* rifampicin induction normalised enzyme activity readings showed a 12.62% (0.13-fold) (Rifampicin 2.5 µM) and a 92.64% (0.93-fold) (Rifampicin 25 µM) decrease in CYP3A4 activity. **DISCUSSION:** The distribution of the *CYP2D6*17* variant is likely due to early African population migratory patterns. This may have implications for tamoxifen treatment in patients with breast cancer and other drugs that are metabolised by CYP2D6 in African populations. This research has established the first step in generating iPSC lines that have the *CYP2D6*17* variant. The next

steps for this research are to establish cellular models for other *CYP* gene variants common in Africans to assess their effects on drug metabolism.

ACKNOWLEDGEMENTS

Firstly, I would like to thank and acknowledge my supervisors Prof. Michèle Ramsay, Prof. Collen Masimirembwa and Dr Janine Scholefield, for all their guidance and assistance throughout the entire process of this masters project. Without you all none of this would have been possible, so I thank you.

Dr Tracey Hurrell – I truly appreciate all the teaching and assistance you gave to me with all the enzyme activity assays that were necessary for this project.

Dr Dimakatso Gumede – I truly appreciate all the teaching and assistance you gave to me with all the tissue culturing techniques that were necessary for this project.

Dr Ananyo Choudhury – I truly appreciate all your assistance with access to the AGVP data needed for this project.

David Twesigomwe – I truly appreciate all your assistance with acquiring haplotype and diplotype predictions for my data.

Rajeshree Mahabeer – I truly appreciate all your assistance with facilitating and optimising the DNA extraction via use of buccal swabs.

Monica Araujo – I truly appreciate all your assistance with providing counselling for all the study participants that were recruited for skin biopsies.

Dr Laeeka Moosa – I truly appreciate all your assistance with performing the skin biopsies on the study participants at CMJAH.

Mahtaab Hayat – I truly appreciate all your assistance with preparing and formatting my proposal and complete dissertation.

Jocelyn Gayenga – I truly appreciate all your assistance with all administrative and financial issues throughout the masters. Without you it would have been impossible to stay afloat.

Dr Ezio Fok, Dr Jerolen Naidoo and Dr Shivani Goolab – I appreciate all the assistance and words of encouragement throughout my stay in the B.I.G lab at the CSIR.

All SBIMB staff and students – Thank you for all your support and encouragement.

All Department of Human Genetics staff and students – Thank you for all your support and encouragement.

All my friends and family – Thank you for the support and encouragement. It would have been impossible to do it without all your love and chats around hot meals.

Taahirah Daniels, my significant other – Thank you for all the late-night tea and chocochinos, and for believing in me the whole way through. I could not have done it without you.

FINANCIAL ACKNOWLEDGEMENTS

DST/NRF South African Research Chair Initiative (grant holder Michèle Ramsay)

DST/CSIR Inter-Bursary Support Programme

DEDICATION

I would like to dedicate this thesis to my family, friends and significant other. Thank you for believing in me and being my best fans and supporters.

TABLE OF CONTENTS

DECLARATION ..... **ii**

PRESENTATIONS FROM THIS RESEARCH REPORT ..... **iii**

ABSTRACT ..... **iv**

ACKNOWLEDGEMENTS ..... **vi**

DEDICATION ..... **vii**

LIST OF FIGURES ..... **xii**

LIST OF TABLES ..... **xiv**

ABBREVIATIONS LIST ..... **xv**

1. CHAPTER 1: INTRODUCTION & LITERATURE REVIEW **2**

1.1 INTRODUCTION **2**

1.1.1 Precision Medicine 2

1.1.2 Precision Medicine & Drug Discovery 2

1.1.3 Pharmacogenetics 3

1.1.4 Role of Drug Metabolism in Drug Discovery 4

1.1.5 Liver Physiology 6

1.2 CYP2D6: UNDERSTANDING OF THE ENZYME **8**

1.2.1 Cytochrome P450 summary 8

1.2.2 CYP2D6 Enzyme 10

1.2.3 *CYP2D6* Gene Description 11

1.2.4 *CYP2D6* Alleles 13

1.2.5 African-Specific Allele *CYP2D6*17* 14

1.2.6 Protein Structure of CYP2D6 & CYP2D6*17 16

1.2.7 Drugs Metabolised by CYP2D6*17 17

1.2.7.1 Tamoxifen 17

1.2.7.2 Codeine 19

1.3 CURRENT MODELS FOR DRUG METABOLISM **21**

1.3.1 Human Liver Microsomes 22

1.3.2 Other Subcellular Fractions: Cytosol & S9 Fractions 22

1.3.3 PXR Reporter Gene Assays 23

1.3.4 Liver Slices 23

1.3.5 Primary Human Hepatocyte Cellular Models 24

1.3.6 HepG2 Cellular Models 25

1.3.7 iPSC Cellular Models 26

1.4 CONCLUSION 29

1.5 STUDY RATIONALE 30

1.6 ETHICS 31

2. CHAPTER 2: CYP2D6*17 ALLELE FREQUENCIES IN DIFFERENT AFRICAN POPULATIONS AND IDENTIFICATION OF CYP2D6*17 HOMOZYGOTES 33

2.1 BRIEF INTRODUCTION 33

2.2 METHODS 35

2.2.1 Prevalence of *CYP2D6* Reduced Function Variants in African Populations 35

2.2.1.1 1000 Genomes Project Data 35
2.2.1.2 PharmVar Data 36
2.2.1.3 African Genome Variation Project Data 37

2.2.2 Identifying *CYP2D6*17* Homozygote Individuals to Establish Cell Lines 37

2.2.2.1 Study Participants 38
2.2.2.2 Participant Sampling: Buccal Swab & DNA Extraction 39
2.2.2.3 Primer Design & Synthesis 40
2.2.2.4 TaqMan Genotyping: Assay Design 40
2.2.2.5 Confirmatory PCR & Sequencing 42

2.3 RESULTS 45

2.3.1 Prevalence of *CYP2D6* Reduced Function Variants in African Populations 45

2.3.1.1 CYP2D6*17 Allele Frequency Data Extraction 45
2.3.1.2 CYP2D6*17 Haplotype Frequency Data Extraction 46
2.3.1.3 CYP2D6*17 Homozygous Diplotype Frequency Data Extraction 48

2.3.2 Identifying *CYP2D6*17* Homozygote Individuals to Establish Cell Lines 49

2.3.2.1 Participant Recruitment 49
2.3.2.2 Sampling & DNA Extraction for CYP2D6*17 Genotyping 51
2.3.2.3 TaqMan Assay Optimisation & CYP2D6 Genotyping 52
2.3.2.4 Confirmation of Genotypes by Sequencing of PCR products 55

2.4 DISCUSSION 58

2.4.1 Prevalence of Reduced Function *CYP2D6* Variants in African Populations & Potential Impact on Drug Metabolism 58

2.4.1.1 CYP2D6*17 SNP-associated Allele Frequencies in African Populations 58
2.4.1.2 CYP2D6*17 Haplotype & Diplotype Frequencies in African Populations 59

2.4.1.3 Other Reduced Function CYP2D6 Haplotype & Diplotypes frequencies 61

2.4.2 Identification of Individuals to Act as Cell Donors to Establish an SC-HLC *CYP2D6*17* Drug Metabolism Model 66

2.4.2.1 Recruitment of Study Participants & Buccal Swabs for DNA extraction 66
2.4.2.2 Genotyping & Sequencing of rs28371706 as a Proxy for CYP2D6*17 66

2.4.3 Strengths & Limitations Related to this Chapter's Work 68

2.5 CONCLUSION 70

3. CHAPTER 3: CELLULAR ASSAYS & EXPERIMENTS 72

3.1 BRIEF INTRODUCTION 72

3.2 METHODS 75

3.2.1 Skin Biopsy & Dermal Fibroblast Culturing 75

3.2.1.1 Preparations for Skin Biopsy 75
3.2.1.2 Skin Biopsy Procedure 75
3.2.1.3 Dermal Fibroblast Culturing 76
3.2.1.4 Haemocytometer Cell Counting 76
3.2.1.5 Cryopreservation & Thawing of Cells 77

3.2.2 Fibroblast Reprogramming 77

3.2.2.1 iMEF seeding & Conditioned Media 77
3.2.2.2 Fibroblast Reprogramming Using Sendai Virus 78

3.2.3 HepG2 Cell Line Culture & Spheroid Formation 80

3.2.4 Functionality of Spheroids *In-Vitro* 80

3.2.4.1 CYP3A4 Luciferase Assay 80
3.2.4.2 Protein Quantitation 81
3.2.4.3 Viability Staining 83

3.3 RESULTS 84

3.3.1 Skin Biopsy & Dermal Fibroblast Culturing from homozygote rs28371706 T & C Allele Individuals 84

3.3.1.1 Skin Biopsy Procedure 84
3.3.1.2 Dermal Fibroblast Isolation & Culturing 85

3.3.2 Fibroblast Reprogramming Using Sendai Virus & iPSC Culturing for SC-HLC Drug Metabolism Model 85

3.3.2.1 Attempts at Fibroblast Reprogramming 85

3.3.3 HepG2 Cell Line 3D Culture & In-Vitro Spheroid Functionality Assays 86

3.3.3.1 HepG2 Spheroid Culturing 86
3.3.3.2 Rifampicin Induction of CYP3A4 Expression in HepG2 Spheroids 86
3.3.3.3 Viability Staining 92

3.4 DISCUSSION 93

3.4.1 Dermal Fibroblasts derived iPSCs for SC-HLC Differentiation. 93

3.4.2 Use of Sendai Virus for iPSC Reprogramming 94

3.4.3 Comparison of HepG2 Spheroid with Other Hepatic Cell Culture Methods for Drug Metabolism Functionality Assays. 95
3.4.3.1 Rifampicin Induction of CYP3A4 Expression in hepatic cells. 95
3.4.4 Strengths & Limitations Related to this Chapter's Work 96

3.5 CONCLUSION 99

4. CHAPTER 4: FUTURE RESEARCH & CONCLUSIONS 101

4.1 FUTURE RESEARCH ON THE PREVALENCE OF *CYP2D6* VARIANTS IN AFRICAN POPULATIONS 102
4.2 COMPARISON OF CYP2D6 MODELLING IN SC-HLCs & OTHER CELLULAR MODELLING METHODS 104
4.2.1 RNA & Protein Expression Levels 104
4.3 CYP ENZYME ACTIVITY IN SC-HLCs, HEPG2 CELLS & PHHs 106
4.3.1 Luciferase Assay to Investigate Enzyme Activity 106
4.3.2 LC/MS-MS & Drug Compound Metabolism 106
4.4 POTENTIAL SOURCES OF CELLS TO DEVELOP IPSC MODELS FOR DRUG METABOLISM 110
4.4.1 Peripheral Blood & Urine-Derived Cells 110
4.5 OTHER DRUG METABOLISM MODELLING SYSTEMS FOR USE WITH SC-HLCs 112
4.5.1 Molecular Dynamics 112
4.5.2 *In Silico* Modelling 113
4.5.3 Microfluidic Systems 113
4.6 CONCLUSION 115

5. REFERENCES & APPENDICES 117

5.1 REFERENCES 117
5.2 APPENDICES 125
5.2.1 Appendix A: Ethics Clearance Certificates 125
5.2.2 Appendix B: Study Recruitment & Participation Forms 127
5.2.3 Appendix C: Bioinformatic Coding Script & Supplementary Data 137
5.2.4 Appendix D: ATCC MEF culture and inactivation with mitomycin C 139

LIST OF FIGURES

CHAPTER 1

Fig 1.1: Impact of introducing DMPK evaluation in the drug discovery and development value chain……5

Fig 1.2: Gross anatomical structure of the liver and the liver lobules……..……..7

Fig 1.3: Ratio of the CYP proteins found in the liver microsomes……..9

Fig 1.4: Participation, by percentages, of the CYPs in the metabolism of clinical drugs……..…..9

Fig 1.5: Annotation of the gene structure for *CYP2D7*……..……..……..12

Fig 1.6: *CYP2D6* gene annotation from UCSC Genome Browser…….13

Fig 1.7: Schematic diagram of *CYP2D6* and genetic arrangement of the *CYP2D6*17* variant……..…..15

Fig 1.8: Tertiary structure of CYP2D6 with the positions of some of the most common SNPs associated with the enzyme……..……..……..…….....15

Fig 1.9: Conformational changes in CYP2D6, and their effects on the protein structure and access to the active site……..……..……..……...17

Fig 1.10: Human metabolic pathways of tamoxifen……..……..……..……..18

Fig 1.11: Human metabolic pathway of codeine……..……..……..……..20

Fig 1.12: Liver lobule zonation and the metabolism the function of the hepatocytes in each zone……..……..……24

CHAPTER 2

Fig 2.1: Workflow for the bioinformatic analysis in the study…..……..……..…….....34

Fig 2.2: Workflow from sample collection……..……..……....37

Fig 2.3: Calculations for expected *CYP2D6*17* allele frequencies in the Zimbabwean Shona population……..……..……37

Fig 2.4: Polymerisation and detection process of the TaqMan Genotyping Assay for the SNP rs28371796……..……..……..……40

Fig 2.5A: Amplification curve plot illustrating the amplification of the fluorescent signals from the probes at each of the KAPA™ Taq Polymerase PCR cycles……..……..……..52

Fig 2.5B: Allelic discrimination plot illustrating the rs28371706 allele results from the TaqMan® Custom SNP Genotyping Assay Mix……..……..……52

Fig 2.6A: Amplification curve plot illustrating the amplification of the fluorescent signals from the probes at each of the Q5® Polymerase PCR cycles……..……..……..53

Fig 2.6B: Allelic discrimination plot illustrating the rs28371706 allele results from the TaqMan® Custom SNP Genotyping Assay Mix……..……..……..53

Fig 2.7A: Amplification curve plot illustrating the amplification of the fluorescent signals from the probes at each of the TaqPath™ PCR cycles……..……..……..54

Fig 2.7B: Allelic discrimination plot illustrating the rs28371706 allele results from the TaqMan® Custom SNP Genotyping Assay Mix……..……..……54

Fig 2.8A: Amplification curve plot illustrating the amplification of the fluorescent signals from the probes at each of the PCR cycles……54

Fig 2.8B: Allelic discrimination plot illustrating the rs28371706 allele results from the TaqMan® Custom SNP Genotyping Assay Mix……54

Fig 2.9: Gel electrophoresis of the PCR products during the rs28371706 PCR optimisation……55

Fig 2.10: Gel electrophoresis of the PCR products during the rs28371706 PCR optimisation……56

Fig 2.11: Haplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* overlaid with the migratory patterns of the first and second bantu expansions……64

Fig 2.12: Diplotype frequencies of the CYP2D6*10, CYP2D6*17, CYP2D6*29 and CYP2D6*41 overlaid with the migratory patterns of the first and second bantu expansions……65

CHAPTER 3

Fig 3.1: Workflow of experiments and assays utilised to achieve objectives 2 and 3……73

Fig 3.2: Calculation for cell concentration using a haemocytometer and trypan blue dye……75

Fig 3.3: Experimental timeline for the reprogramming of the dermal fibroblasts using the CytoTune™-iPS 2.0 Sendai Reprogramming Kit……77

Fig 3.4: Calculations for viral vector volumes for transduction of dermal fibroblasts during iPSC reprogramming……78

Fig 3.5: Conversion of P450-Glo™ substrate to luciferin by cytochrome P450 and detection mechanism……79

Fig 3.6: Phase-contrast images dermal fibroblast isolation process from a skin biopsy and cultured until ready for SeV reprogramming……83

Fig 3.7: Phase-contrast images showing iPSC clones from the reprogramming of dermal fibroblast line 3B4014 from the February Reprogramming……85

Fig 3.8: Phase-contrast images of HepG2 cells in 2D and 3D culturing spanning a 10-day period at 20x magnification……86

Fig 3.9: *Experiment 1* Luciferase assay results for 72Hr incubation with Rifampicin at concentrations of 2 µM and 2000 µM for *CYP3A4* induction……87

Fig 3.10: *Experiment 2* Luciferase assay results for 48Hr incubation with Rifampicin at concentrations of 20 µM for *CYP3A4* induction……88

Fig 3.11: BSA standard curve for total protein concentrations ranging from 25–2000 µg/ml at an Absorbance of 450 nm……88

Fig 3.12: *Experiment 3* luciferase assay *CYP3A4* activity after 24-hour with rifampicin induction……89

Fig 3.13: *Experiment 3* luciferase assay *CYP3A4* activity after 72-hour with rifampicin induction……90

Fig 3.14: *Experiment 3* luciferase assay activity over 24- to 72-hour incubation with rifampicin for *CYP3A4* induction……90

LIST OF TABLES

CHAPTER 1

Table 1.1: Major substrates of CYP2D6 and the concerns for particular metaboliser groups……..20

CHAPTER 2

Table 2.1: Genetic arrangement of *CYP2D6*2* & *CYP2D6*17* allele variants……..……..32

Table 2.2: SNP rs28371706 primer and probe information……..……...39

Table 2.3: Reagents for rs28371706 genotyping and volume for each PCR reaction……....42

Table 2.4: PCR cycling conditions for each PCR assay……43

Table 2.5: Allele frequencies of the SNPs that make up the *CYP2D6*17* haplotype, extracted from the KGP African population data……..…….45

Table 2.6: Allele frequencies of the SNPs that make up the *CYP2D6*17* haplotype, extracted from the AGVP African populations……..……..…..45

Table 2.7: Haplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* alleles from the KGP African populations……..……..……..…….46

Table 2.8: Haplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* alleles from the PharmVar populations……..……46

Table 2.9 Haplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* alleles from the AGVP populations……..……..……..……46

Table 2.10: Homozygous diplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* alleles from KGP population data estimates……..……..……48

Table 2.11: Homozygous diplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* alleles from PharmVar population data estimates……...48

Table 2.12: Homozygous diplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* alleles from AGVP population data estimates……..……..…..49

Table 2.13: Participant questionnaire data: parental ethnicity and drug use and response..……50

Table 2.14: DNA concentrations (ng/µl) and quality (260/280 ratio) for the study participant DNA sample extractions performed at two different laboratories……..…..…..51

Table 2.15: Comparison of rs28371706 SNP locus genotype calling using the TaqMan PCR assay and DNA sequencing of PCR product……..……57

CHAPTER 3

Table 3.1: Reagents and measurements for the RIPA buffer for cell lysis……..…..…..……81

Table 3.2: Dilutions and concentrations for BSA standards……..……..……...82

Table 3.3: Absorbance at 450 nm and total protein concentration of HepG2 spheroids in the Experiment 3……..……..……..89

CHAPTER 4

Table 4.1: Summary of Literature on CYP enzyme drug substates used in drug metabolism studies……...89

ABBREVIATIONS LIST

ADME – Absorption, Distribution, Metabolism & Excretion

ADR – Adverse Drug Reaction

AGVP – African Genome Variation Project

Amp-B – Amphotericin-B

BCA – Bicinchoninic Acid

BSA – Bovine Serum Albumin

BSL- 2 – Biosafety Level II

CMJAH – Charlotte Maxeke Johannesburg Academic Hospital

CNS – Central Nervous System

CNV – Copy Number Variant

CPIC – Clinical Pharmacogenetics Implementation Consortium

CSIR – Council for Industrial and Scientific Research

CYP – Cytochrome P450

DMEM – Dulbecco's Modified Eagle's Medium

DMPK – Drug Metabolism and Pharmacokinetics

ECM – Extracellular Matrix

ESC – Embryonic Stem Cell

ESN – Esan in Nigeria

ER – Endoplasmic Reticulum

FBS – Foetal Bovine Serum

FDA- Fluorescein Diacetate

GIT – Gastrointestinal Tract

GST – Glutathione S-Transferase

GWAS – Genome Wide Association Study

GWD – Gambian in Western Division, Gambia

H+/ER+ - Hormone/Oestrogen Receptor-positive

HREC – Human Research Ethics Committee

HSC – Hepatic Stellate Cell

huESC – Human Embryonic Stem Cell

HUC – Urine-derived Stem Cell

HWE – Hardy-Weinberg Equilibrium

iMEF – Inactivated Mouse Embryonic Fibroblast

IM – Intermediate Metaboliser

iPSC – Induced Pluripotent Stem Cell

KC – Kupffer Cell

KGP – 1000 Genomes Project

LSEC – Liver Sinusoidal Endothelial Cell

LWK – Luhya in Webuye, Kenya

MAO-A – Monoamine Oxidase A

MET – Mesenchymal-To-Epithelial Transition

MGB – Minor Groove Binder

MOI – Mode of Infection

MSL – Mende in Sierra Leone

NADP – Nicotinamide Adenine Dinucleotide Phosphate

NAT – N-Acetyl Transferase

NFQ – Non-Fluorescent Quencher

NHLS – National Health Laboratory Services

NK – Natural Killer Cells

NM – Normal Metaboliser

NR – Nuclear Receptor

PBC – Peripheral Blood Cell

PBS – Phosphate Buffer Saline

PCR – Polymerase Chain Reaction

PD – Pharmacodynamics

PI – Propidium Iodide

PenStrep – Penicillin/Streptomycin

PharmVar – Pharmacogene Variation Consortium

PHH – Primary Human Hepatocyte

PK - Pharmacokinetics

PM – Poor Metaboliser

PXR – Pregnane X Receptor

QC – Quality Control

REC – Research Ethics Committee

RFU – Relative Fluorescence Unit

RIPA – Radio-Immunoprecipitation Assay

RT-PCR – Reverse Transcriptase Polymerase Chain Reaction

SBIMB – Sydney Brenner Institute for Molecular Bioscience

SC-HLC – Stem-Cell-Derived Hepatocyte-Like Cell

SeV – Sendai Virus

SNP – Single Nucleotide Polymorphism

ST – Sulfotransferase

UM – Ultrarapid Metaboliser

UGT – Uridine Diphosphoglucuronosyl Transferase

WGS – Whole Genome Sequences

YRI – Yoruba in Ibadan, Nigeria

CHAPTER 1

INTRODUCTION & LITERATURE REVIEW

1. CHAPTER 1: INTRODUCTION & LITERATURE REVIEW

1.1 INTRODUCTION

1.1.1 Precision Medicine

In the discovery, development and clinical use of medicines, pharmacokinetics (PK) (process of uptake of drugs by the body and biotransformation they undergo) and pharmacodynamics (process of interaction of pharmacologically active substances with target sites in living systems) are important for safe and efficacious treatment outcomes (Duffus, Nordberg and Templeton, 2007; Zhang and Tang, 2018). Preclinical drug metabolism and pharmacokinetic studies (DMPK) require that we use *in silico*, *in vitro* and non-human animal models that will enable us to predict outcomes in humans (Jones and Rowland-Yeo, 2013; Jones *et al.*, 2015). Efforts for improving current models and innovating new models for this purpose are ongoing. This has unsurprisingly led to an unspoken race to create the best model for investigating interaction between drugs and the human body. The complexities of drug modelling are made more apparent with the variation seen in drug metabolism and the enzymes involved in this process (Kumar and Surapaneni, 2001). Understanding the interaction between genetics and drug metabolism is a crucial endeavour for the development of effective and safe treatments. Given the well-established interindividual and interpopulation variability in drug metabolism, the subject of DMPK has found itself central to the concept and practice of precision medicine (Jones *et al.*, 2015). This is because precision medicine aims to determine patient treatment based on an individual's genomic and environmental factors. The effect of genetic/genomic variation on drug response, referred to as pharmacogenetics/genomics, has therefore become a priority for many involved in pharmacological research (Dugger, Platt and Goldstein, 2018).

1.1.2 Precision Medicine & Drug Discovery

The impact that big data has had on medical science research, particularly genomic data, is enormous, and has caused an exponential increase in the rate of knowledge. It has seen better and more substantial comprehension of biological mechanisms and traits that contribute to disease and disorders (Denny *et al.*, 2018). The increased understanding of genetic influences on disease and drug responses has also kickstarted numerous new genetic-tailored therapies aimed at common and rare diseases. One of the great obstacles to these endeavours has been the lack of sufficient, sizeable and suitable cohorts with robust and credible data needed to study these biological influences on health (Denny *et al.*, 2018). In recent years, there has been a great amount of growth in large national and international cohorts and networks committed to host large repositories of data. The focus of this storage of data has been to

merge phenotypic data with genetic data from DNA biobanks. These cohorts have been used to fast-track new genomic information and drug discovery (Denny *et al.*, 2018).

One research approach that has contributed vastly to the increase in genetic information is the use of genome-wide association studies (GWAS). This is a genotyping approach that uses linkage disequilibrium in populations to distinguish between particular disease-associated variants and common single nucleotide polymorphisms (SNPs) (Dugger, Platt and Goldstein, 2018). In drug discovery, GWAS is able to provide researchers with insight into the genetic mechanisms that lead to disease. This led to several new candidate genes and pathways for disease targeting for common and complex diseases (Dugger, Platt and Goldstein, 2018).. However, GWAS has largely been unsuccessful in uncovering new targets needed for drug discovery. This is due to GWAS data limitations in revealing causal variants and specific genes involved in disease pathology (Dugger, Platt and Goldstein, 2018). Another problem is that majority of effects seen in GWAS studies lean towards an additive effect of thousands of SNPs that are linked in complex interconnected cellular networks. This has made it difficult to target a single or several SNPs, which may have a miniscule effect individually regarding drug-tissue interactions, for better drug development outcomes. This has left GWAS as a promising approach for variant discovery but rather ineffective for drug development (Dugger, Platt and Goldstein, 2018). From the birth of pharmacogenetics in the 1960s, the field has continued to grow through technological advances in genomics and bioinformatics (Roden *et al.*, 2011). When looking at the implementation of precision medicine policies and interventions, pharmacogenetics has been shown to be extremely valuable. It has delivered medical solutions through its many applications for better clinical outcomes and patient benefit (Ashley, 2016).

1.1.3 Pharmacogenetics

Pharmacogenetics is defined as “the study of how genetic differences influence the variability in patients response to drugs” (Roses, 2000). It is a field of study critical to the establishment of precision medicine and is the cornerstone of its implementation in healthcare. The strength of pharmacogenetics is its ability to profile individual patients and predict tailored treatment interventions for them. From a policy perspective, it could be able to provide healthcare practitioners, private institutions and governments with treatment interventions that are cost-effective with reduced risks of injury or adverse drug reactions (ADRs) (Roses, 2000). In Africa particularly, there is substantial variation due to the continent being the site of origin for modern humans, and having had longer timespans for these mutations, genetic recombination and drift to occur (Rajman *et al.*, 2017). There is also the heavy burden of

disease felt by Africa with an estimated 25% share in the global burden. This highlights the disproportionate effects of disease on Africa as it carries 15.5% of the global population (Murray *et al.*, 2012; Rajman *et al.*, 2017). This has also been met with the lack of a robust pharmaceutical industry indigenous to Africa to develop medicines for the continent resulting in 90% of medicines being imported (Chaudhuri and West, 2015). Additionally, of great concern are the high levels of ADRs in Africa due to numerous factors ranging from poor medicine efficacy to genetic individual variability, as well as medications not optimized for African populations (Chaudhuri and West, 2015; Rajman *et al.*, 2017). Hence, the need for the implementation of any and all systems that can alleviate the problem of a poor health burden, including pharmacogenetics (Rajman *et al.*, 2017).

Resolving the problem of ADRs is crucial in establishing precision medicine and drug development in Africa. Genetics is one of the major contributors to ADRs experienced, and 20-30% of the drug response variability is associated to genetic variation (Goh *et al.*, 2017). The cytochrome P450 (CYP) superfamily is highly associated with this genetic-based variation in drug response. Genes such as *CYP2D6* being highly polymorphic and metabolising 25% of all clinically used drugs (Goh *et al.*, 2017). Today, most pharmacogenetic studies are undertaken in Caucasian and Asian populations with African Americans being a proxy for all Africans. This highlights the need to generate African-based genetic data to better formulate associations between genetics and drugs for African populations (Rajman *et al.*, 2017). Increases in research on the African continent have been motivated by observed differences in clinically relevant genetic variants of drug metabolising enzymes (Rajman *et al.*, 2017). However, this genotyping and pharmacogenetic profiling of individuals will need to be cost-effective and robust in its implementation, and this reaffirms the need for suitable modelling systems for industrial, research and diagnostic purposes (Goh *et al.*, 2017).

1.1.4 Role of Drug Metabolism in Drug Discovery

Drug metabolism is the process of biotransformation of therapeutic compounds through structural modification by enzymes to different compounds called metabolites (Zhang and Tang, 2018). Primarily, metabolism of drugs or a xenobiotic substance is utilised for elimination of that compound from the body. Thus, the majority of drug metabolism results in the inactivation of the drugs. However, some of these metabolites can cause clinical responses that are either linked to the desired therapeutic response or exhibit off-target effects (Kumar and Surapaneni, 2001). There are three major drug-metabolising pathways that are utilised in the human body and the organ responsible for the majority of drug metabolism is the liver

(Almazroo, Miah and Venkataramanan, 2017). The different pathways are referred as: phase I (functionalisation reactions aimed at modification via the addition of a functional group), phase II (derivatization/conjugation reactions aimed at elimination from host) and phase III (drug transportation for elimination via bile production and efflux) (Kumar and Surapaneni, 2001; Almazroo, Miah and Venkataramanan, 2017).

In the initial stages of drug discovery and development, drug metabolism has become a key factor and input for selecting candidate drug compounds (Yengi, Leung and Kao, 2007). Drug metabolism is an important aspect of optimising promising lead therapeutic compounds to ensure drug efficacy and safety. It also has relevance in understanding the active metabolite compounds and their activity in the body pertaining to efficacy and safety when using the drug (Zhang and Tang, 2018). The growing knowledge of the relationship between clinically prescribed drugs and drug metabolism led to DMPK and safety profiles, in conjunction with clinical data, being used to develop better and safer drugs and treatment regimens. Over the 1990 to 2000 period integration of DMPK into the drug discovery process reduced attrition rates due to inadequate DMPK properties (Fig 1.1) (Prentis, Lis and Walker, 1988; Kola and Landis, 2004; Yengi, Leung and Kao, 2007).

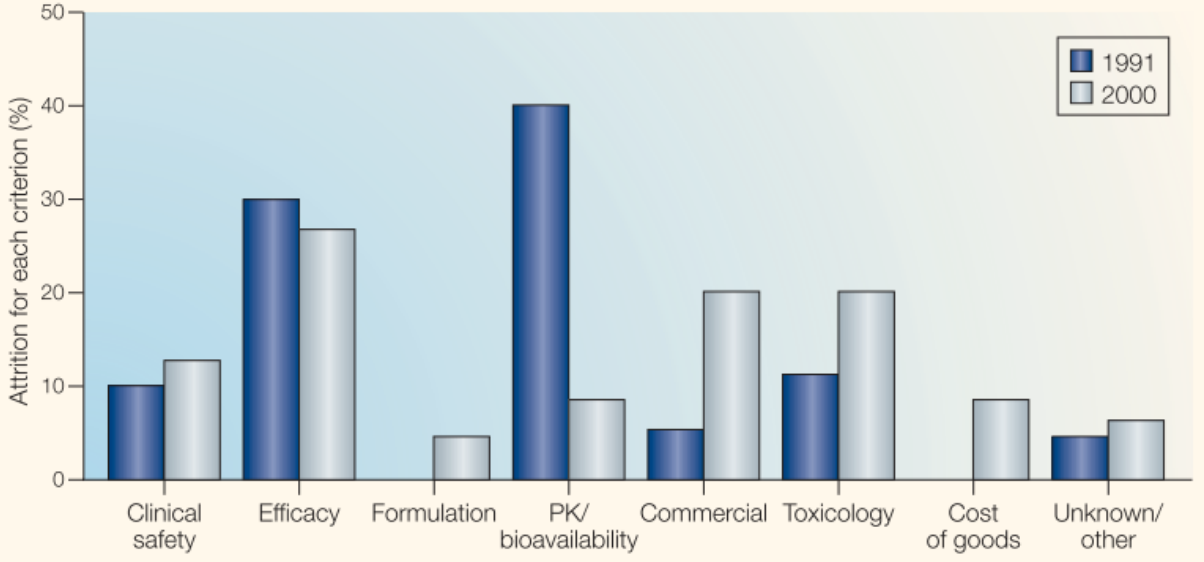


Fig 1.1: Impact of introducing DMPK evaluation in the drug discovery and development value chain (Kola and Landis, 2004). *Image from: Kola, I. and Landis, J. (2004).*

This prompted researchers to employ various approaches to decipher the underlying pharmacogenetics in drug metabolism to increase efficiency in drug discovery and development (Dugger, Platt and Goldstein, 2018). Another aspect of metabolism that is critical for drug development is understanding and evaluating the drug-drug interactions that can affect the enzymes involved in the process, by either induction or inhibition (Kumar and Surapaneni, 2001). When looking at the effects of drug metabolism on the efficacy and safety

of drugs, numerous factors must be considered. It is important to not only look at the chemistry of the metabolic process of interest, but the enzymatic, genetic, environmental, mechanistic and interactive factors that influence the desired outcome (Pearson and Wienkers, 2019). Some of the most important and most widely studied enzymes are those from the cytochrome P450 (CYP450) superfamily of enzymes, which have a critical role in processing of over 75% of the drugs on the market (Harvey *et al.*, 2001; Dong *et al.*, 2018).

1.1.5 Liver Physiology

The liver is the most important organ regarding metabolism in the body and can be compared to a large biochemical factory (Sherwood, 2016). It is also the largest organ and gland in the human body with a weight of approximately 1 500 g which is approximately 2.5% of the weight of the average adult body (Pawlina and Ross, 2016). Fig. 1.2 shows the gross and detailed anatomy of the liver where the hepatic portal vein accounts for approximately 75% of the liver blood supply and carries blood from the digestive tract, spleen, pancreas, and other abdominal organs. This blood is predominantly deoxygenated and carries nutrients, toxic materials, blood products, and endocrine secretions. The hepatic artery is a branch of the celiac trunk making up the outstanding 25% of the blood supply and brings oxygenated blood to the liver (Pawlina and Ross, 2016). Within the liver parenchyma, the branches of the hepatic portal vein, hepatic artery and common bile duct flow together in a grouping referred to as the portal triad (Pawlina and Ross, 2016).

The liver is functionally organised into units called lobules with three main methods to describing their structure in the liver, which are the classic lobule, the portal lobule, and the liver acinus (Pawlina and Ross, 2016). The classic lobule unit is hexagonal in shape and has the central vein anchored in the middle of the lobule unit with the portal triad at the corner of lobule. Blood flows from the branches of the portal triad through the sinusoids and sinusoidal capillaries towards the central vein (Pawlina and Ross, 2016; Sherwood, 2016). The liver also exhibits zonation which describes the patterns of degeneration, regeneration and toxicity effects on the liver parenchyma in relation to the central vein (Pawlina and Ross, 2016). This zonation is facilitated by morphogen signalling pathways such as the Wnt/B-catenin, hepatocyte growth factor and hedgehog signalling pathways. The Wnt/B-catenin signalling pathway is most active and highest in the area surrounding the central vein. In the cells around this area, the pathway is involved in regulating numerous functions including drug metabolism (Muriel, 2017).

The liver comprises of various cell types each of which aid in facilitating and performing its numerous functions. The cell types found here are hepatocytes, cholangiocytes, natural killer

cells (NKs), liver sinusoidal endothelial cells (LSECs), Kupffer cells (KCs), and hepatic stellate cells (HSCs) (Muriel, 2017). The hepatocytes are large polygonal cells that account for approximately 80% of liver cell population and are involved in the metabolic and secretory functions of the liver (Pawlina and Ross, 2016; Muriel, 2017). The metabolic functions that hepatocytes are involved in are categorised into phase I and II metabolism. Phase I metabolism is involved in the oxidation of xenobiotics by hydroxylation and carboxylation, and this is predominantly performed by the cytochrome P450 (CYP) enzymes. Phase II metabolism is involved in conjugation of phase I metabolised products to make them more water-soluble for excretion via the kidneys (Pawlina and Ross, 2016). The CYP enzymes are critical for the elimination of xenobiotics from the body and the liver is at the centre of this function (Almazroo, Miah and Venkataramanan, 2017).

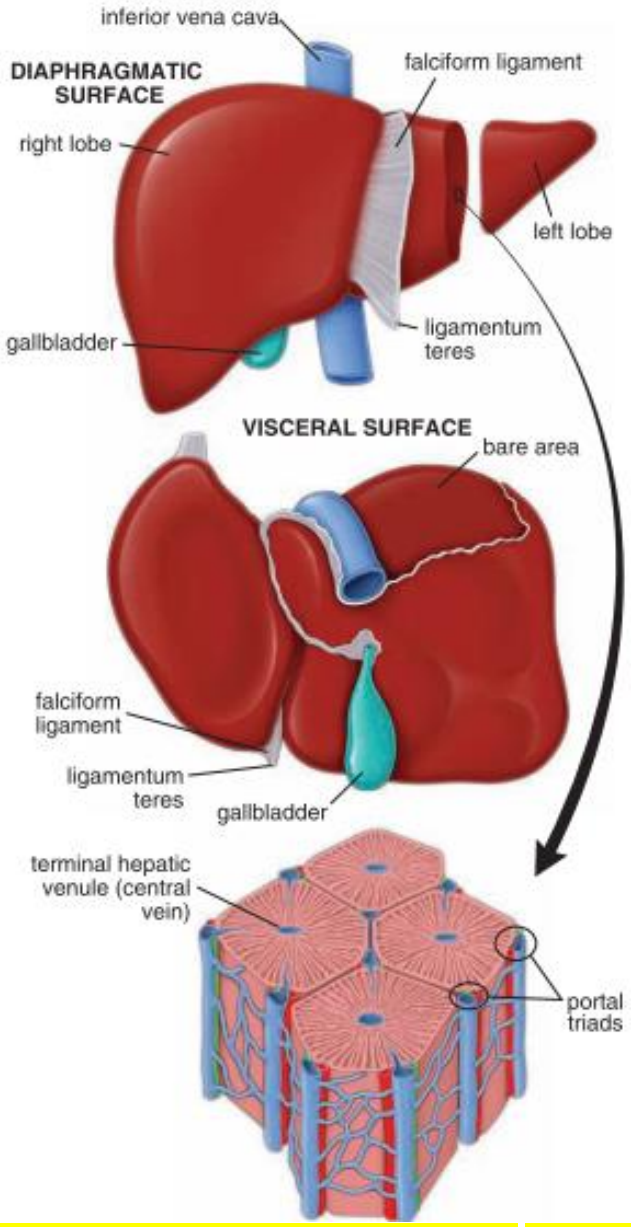


Fig 1.2: Gross anatomical structure of the liver and the liver lobules. *Image from: Pawlina, W. and Ross, M. H. (2016).*

1.2 CYP2D6: UNDERSTANDING OF THE ENZYME

1.2.1 Cytochrome P450 summary

The Cytochrome P450 (CYP) superfamily is made up of 18 families and 43 subfamilies (Fig 1.3) in humans, and is involved in the metabolism of various pharmacological and toxic agents (Dong *et al.*, 2018). In total, there are 57 enzymes that make up this CYP superfamily, some of which have not been completely characterised (Pearson and Wienkers, 2019). These CYP enzymes are membrane-bound proteins with 50 localised in the endoplasmic reticulum and seven in the mitochondria (Guengerich, Waterman and Egli, 2016). The history of these CYP enzymes dates to one of the most ancient microorganisms, archaebacteria. An example of a bacterium that houses CYP enzymes is *Sulfolobus solfataricus*, which uses CYP enzymes due to their tremendously stable structures that were necessary for adequate function in the treacherous environment of early earth (Zuber, Anzenbacherová and Anzenbacher, 2002). This bacteria's *CYP* gene, *CYP119*, shares approximately 18% homology with the human counterpart *CYP3A4* (Wright *et al.*, 1996).

In humans, there are more than 115 *CYP450* genes and pseudogenes ranging from *CYP1A1* to *CYP51P3* (Almazroo, Miah and Venkataramanan, 2017). From these, only 57 genes result in functional proteins of which 12 are responsible for approximately 75% of phase I drug metabolism (Zhou, Ingelman-Sundberg and Lauschke, 2017). Phase I drug-metabolising enzymes are involved in pharmaceutical drug modification and transformation that allows for easy excretion of these drug compounds as water-soluble products from the kidney and liver (Almazroo, Miah and Venkataramanan, 2017). The enzymes CYP1A2, CYP2B6, CYP2C9, CYP2C19, CYP2D6, CYP3A4 and CYP3A5 account for approximately 80% of the drugs metabolised by the CYP superfamily (Rendic and Guengerich, 2015). The enzymes CYP3A4 and CYP3A5 are responsible for 27% and 6% respectively, and CYP2C9 and CYP2D6 account for 10% and 13% correspondingly (Pearson and Wienkers, 2019) (Fig 1.4). Genes encoding CYP450 enzymes have numerous polymorphisms that give rise to different isozymes. CYP2D6, involved in tamoxifen and codeine metabolism, has been reported to have more than a 100 genetic variants and CYP2B6, responsible for efavirenz metabolism, has between 50 to 60 genetic variants (Tracy *et al.*, 2016).

The majority of the reactions catalysed by the CYP enzymes are carbon hydroxylation, heteroatom oxidation, bond oxidation, hydrocarbon desaturation, and halocarbon dehalogenation. These enzymes work in a systematic metabolism of compounds referred to as the P450 catalytic cycle, which entails a compound or drug binding to the enzyme active site (Zhang and Tang, 2018). The active site is often large and has preference for lipophilic

substrates, and hydrophobic and steric interactions are the main factors that affect the binding and orientation of the substrate to the active site amino acids (Zhang and Tang, 2018). The CYP enzymes will then bind two oxygen atoms to the cysteine bound iron of the heme on the active site and undergo oxygen activation, which splits the dioxygen molecule. One of the oxygens will be removed to form water, and the other will be introduced to the substrate as a hydroxyl and results in a hydroxylated product (Zuber, Anzenbacherová and Anzenbacher, 2002).

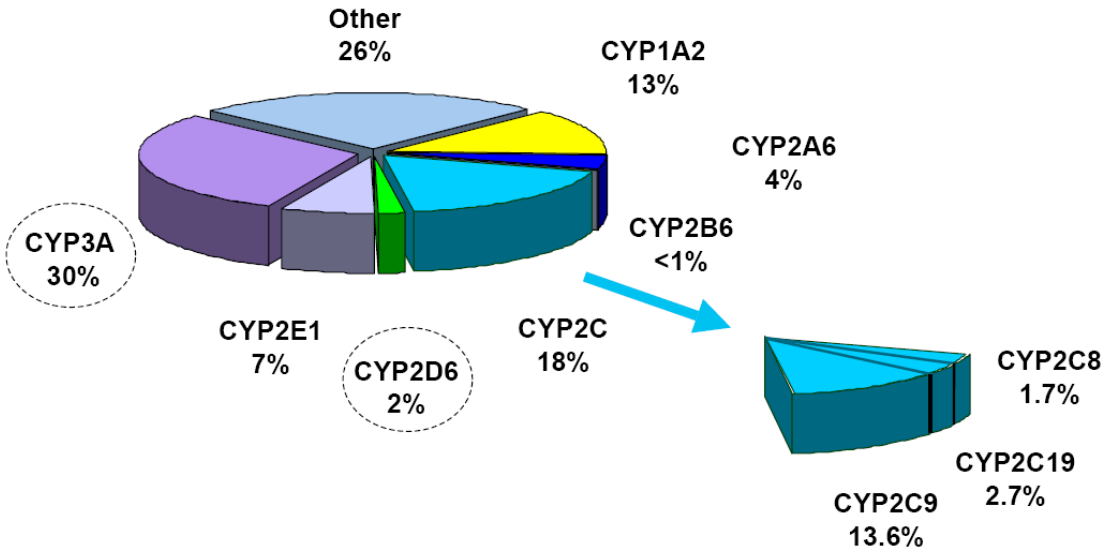


Fig 1.3: Ratio of the CYP proteins found in the liver microsomes. *Image from: Shimada et al., (1994), (Left) and Jasker et al., (1998), (Right).*

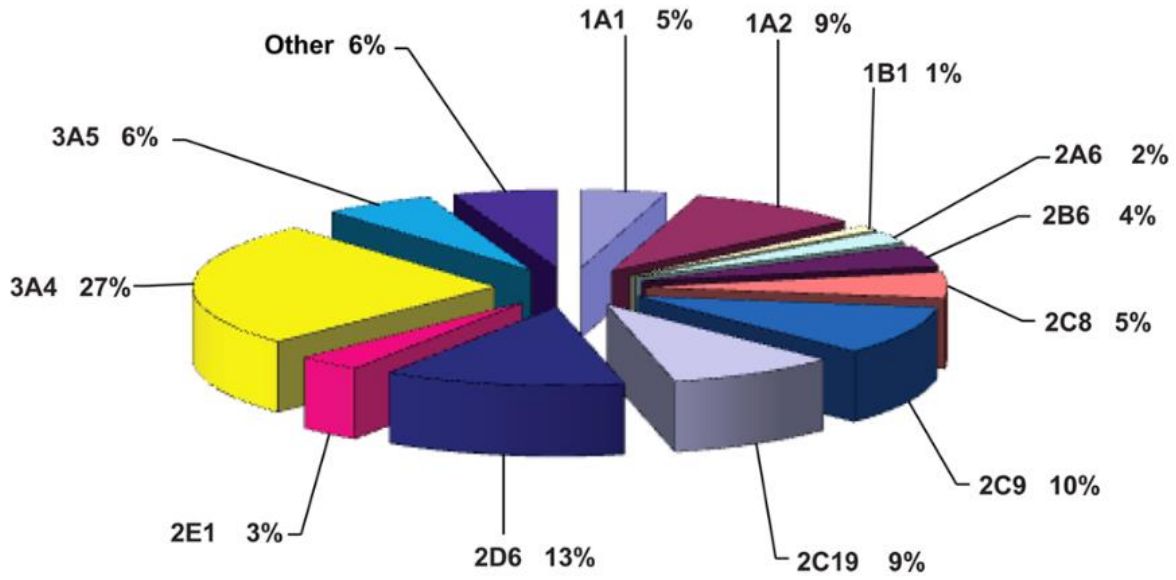


Fig 1.4: Participation, by percentages, of the CYPs in the metabolism of clinical drugs. *Image from: Rendic, S. and Guengerich, F.P., (2015).*

Genetic variants found within the *CYP* genes can have a large effect on the functioning of the enzymes. For example, one in 15 persons of African American or Caucasian ancestry will experience an increased beta-blocker response (Lynch and Price, 2007). These variants often

result in the expression of CYP proteins that exhibit reduced enzyme activity. Irrespective of the causes of these interindividual variations in enzyme activity, it is critical to establish better methods to study these differences as well as test them in a clinical setting (Tracy *et al.*, 2016). With this in mind, CYP2D6 is an enzyme that has a vast amount of variation at the genome level and has garnered noteworthy interest from researchers for its pharmacological and clinical implications.

1.2.2 CYP2D6 Enzyme

CYP2D6 is an enzyme from the CYP2 family which is the most important CYP family in the context of polymorphic enzymes for the metabolism of cancer drugs (Bertholee, Maring and van Kuilenburg, 2017). CYP2D6 metabolises approximately 20% of all clinically used drugs and thus, plays a critical role in drug elimination and efficacy (Tracy *et al.*, 2016). The known clinical drugs metabolised by CYP2D6 range from antidepressants, antipsychotics, antineoplastics, adrenergic antagonists and analgesics. CYP2D6 enzyme activity has a high degree of functional variability and can range from enzymes with no to increased function or activity. This functional variability has been associated with clinically significant differences in treatment outcomes for some substrate drugs (Nofziger *et al.*, 2019). This has led to CYP2D6 being the most well characterised enzyme of the CYP family (Nofziger *et al.*, 2019).

The discovery of CYP2D6 occurred in 1970 when English and German research groups independently discovered a series of unanticipated drug reactions experienced by participants from pharmacokinetic studies of the antihypertensive drug debrisoquine and antiarrhythmic drug sparteine (Mahgoub *et al.*, 1977; Eichelbaum *et al.*, 1979). In both studies, the consensus was that the participants affected were unable to oxidise the drugs, and that this inability appeared to be monogenic with an autosomal recessive mode of inheritance. By 1980, more findings showed that poor metabolisers (PMs) of debrisoquine were also PMs of sparteine, and this fuelled the search for more drugs that had a similar effect due to this genetic deficiency (Zanger, Raimundo and Eichelbaum, 2004). Additional drugs were found to exhibit altered rates of metabolism including metoprolol and bufuralol, with these findings both happening in the early 1980s (Lennard *et al.*, 1982; Dayer *et al.*, 1984). It was then theorised that one of the CYP enzymes was the culprit of this poor metabolism exhibited in particular individuals (Wolff *et al.*, 1985). This was made possible by steps taken to purify CYP2D6 from solubilised human liver microsomes (Distlerath *et al.*, 1985). By 1988, demonstrations of the PM phenotype were carried out in human liver microsomes showing poor bufuralol metabolism in the absence of a 50kDa protein from an enzyme (Zanger *et al.*, 1988).

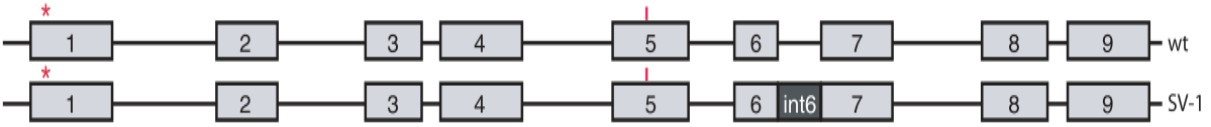
In 1987, the polymorphism for this debrisoquine/sparteine deficiency was located in the region of chromosome 22q13.1 and it was found that a *CYP2D* gene was involved (Eichelbaum *et al.*, 1987). In 1988, the novel and distinct full-length human cDNA from this enzyme was isolated, and a new member of the CYP2D family was named as CYP2D6 (Skoda *et al.*, 1988). This was followed by the elucidation of the entire sequence of the gene from the cDNA as well as identification of many alleles affecting activity by restriction fragment length polymorphisms (Skoda *et al.*, 1988; Kimura *et al.*, 1989). These alleles were isolated and then sequenced to investigate their DNA sequence and molecular effects from a genomic perspective (Kagimoto *et al.*, 1990). The advent of the polymerase chain reaction (PCR) technique and Sanger sequencing resulted in the addition of numerous genetic variants to be added to the list of no function or decreased functioning CYP2D6 alleles throughout the 1990s (Nofziger *et al.*, 2019). This led to the creation of the star allele nomenclature system to keep up with the boom in variants found with in *CYP2D6* gene and to track them systematically. The star allele nomenclature system was then applied to not only the *CYP2D6* gene but all the genes of the CYP family by researchers. (Daly *et al.*, 1996). In 2001, a website dedicated to this nomenclature was created called The Human Cytochrome P450 webpage and in 2017 this was changed to the website PharmVar (Nofziger *et al.*, 2019).

1.2.3 *CYP2D6* Gene Description

The *CYP2D6* gene region is complex and shares gene homology with pseudogenes combined with a complex molecular architecture. This has caused numerous technical problems in terms of sequencing and genotyping (Yang *et al.*, 2017). However, it has become one of the most important genes to study due to its clinical and pharmacological implications for drug metabolism and toxicity. This important gene has been estimated to have arisen 361.2 million years ago and the duplication event of the *CYP2D* gene most likely occurred in a common hominin ancestor between humans and great apes (Yang *et al.*, 2017). In humans, the *CYP2D* locus is found on chromosome 22q13.2 within a 45kb area, and contains three homologous genes which are *CYP2D6*, *CYP2D7* and *CYP2D8* (Zanger, Raimundo and Eichelbaum, 2004; Yang *et al.*, 2017). The *CYP2D8* gene is typical of a classic pseudogene containing numerous deletions and insertions and no open reading frame (Zanger, Raimundo and Eichelbaum, 2004). The *CYP2D7* gene is more similar to the functional gene *CYP2D6* and is classically marked with a T-insertion in exon 1 that has resulted in a lack of function in the gene (Zanger, Raimundo and Eichelbaum, 2004; Nofziger *et al.*, 2019). Additionally, CYP2D7 has a frameshift mutation, comprised of a deletion (138delT) in exon 1, and a splice variant, causing a 57 bp insertion after exon 6, that has been proposed to be involved in codeine to morphine demethylation (Pai *et al.*, 2004) (Fig 1.5). This supposedly functional variant has

only been described in East Asian and Hispanic Latino populations at frequencies of 1.1% and 0.7% respectively (Bhathena *et al.*, 2007). The variant also has two isoforms, one being a 1734 nt isoform which is found in the liver and brain and a 1655 nt isoform that is only expressed in the liver (Gaedigk, Gaedigk and Leeder, 2005). Though, it has still not been confirmed if this variant is truly functional, as studies on its metabolic activities have shown little or no function with CYP2D6 substrates (Gaedigk, Gaedigk and Leeder, 2005; Jukic *et al.*, 2018).

Fig 1.5: Annotation of the gene structure for *CYP2D7*. The asterisk (*) represents the 138delT frameshift mutation and the vertical line (|) represents the end of the frameshift and first alternative stop codon. The wildtype (wt) CYP2D7 gene is above and the splice variant (SV-1) is below. *Image from: Gaedigk et al., (2005).*

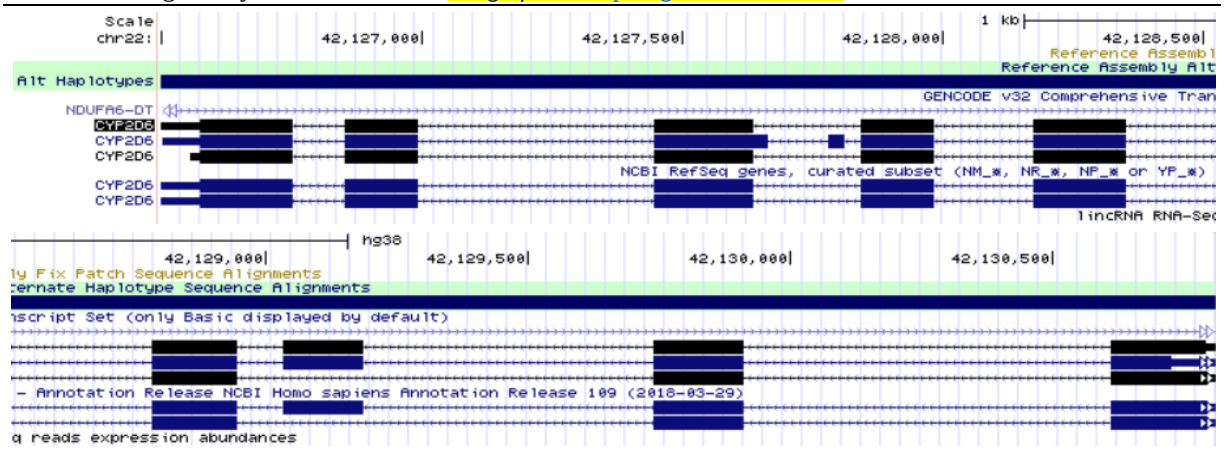


The *CYP2D6* gene that has been inherited by modern humans is 4382 nt in length according to the canonical *CYP2D6* RefSeq reference gene, and only has 9 exons and 8 introns (Fig 1.6) which is similar to all *CYP2D* genes (Zanger, Raimundo and Eichelbaum, 2004; Nofziger *et al.*, 2019). The regulatory mechanisms of *CYP2D6* are complex and not fully understood. One regulatory mechanism is via loss-of-function polymorphisms, which generally affect splicing and expression. The other is via gain-of-function variants, such as CNVs and multiple gene copies (Zanger and Schwab, 2013). An enhancer region is also present 115kb downstream from the gene, and this is associated with a two-fold increase in transcription of *CYP2D6*. The single nucleotide polymorphisms (SNPs) rs5758550 and rs133333 are located in this region, and are associated with increased activity of *CYP2D6* (Wang *et al.*, 2014). Using reverse transcriptase PCR (RT-PCR) it is possible to quantify the amount of mRNA expression from the *CYP2D6* gene. *CYP2D6* expression is low in humans even though it is one of the most important xenobiotic metabolizing enzymes (Zanger, Raimundo and Eichelbaum, 2004). CYPs, in general, have variability in their expression amongst individuals, however, CYP2D6 is one of the enzymes involved in many xenobiotic metabolism pathways and thus is expressed at a similar level according to population group (Guengerich, 1994).

The low levels of expression may be due to an efficient metabolism and faster turnover because of high substrate affinity to the CYP2D6 enzyme (Pan, Ning and Jeong, 2017; Pearson and Wienkers, 2019). *CYP2D6* is also expressed in numerous extrahepatic tissues, and in the gastrointestinal tract (GIT), there is expression of CYP enzymes found throughout the intestinal segments (Zanger, Raimundo and Eichelbaum, 2004; Fritz *et al.*, 2019). CYP2D6 is found mainly in the small intestine and in particular in the duodenum and

jejunum, with lower levels found in the rest of the colon (Fritz *et al.*, 2019). In the brain, *CYP2D6* is expressed and is involved in breakdown of drugs that participate in the dopamine pathways and neurotoxin metabolism (Kirchheiner *et al.*, 2011). In the lungs, there is a low expression of CYP2D6 and the CYP enzyme that is more highly expressed is CYP2E1 (Bernauer *et al.*, 2006).

Fig 1.6: *CYP2D6* gene annotation from UCSC Genome Browser. The *CYP2D6* gene is located on the reverse strand and in the annotation is shown from the view of the forward strand. The picture fragment above illustrates exon 9 – 5 (left to right) and the below exon 4 – 1 (left to right.) Exonic regions are specified by a block shape and intronic regions by line with arrows. *Image from: https://genome.ucsc.edu*



1.2.4 *CYP2D6* Alleles

Most of the *CYP2D6* alleles have been classified due to one or more SNPs or indels that are found in linkage disequilibrium (LD) affecting the transcription and translation of the CYP2D6 enzyme. Some of the other alleles have been characterised by rearrangements in the gene which cause large gene deletions, gene duplications/multiplications or copy number variants (CNVs), and gene hybrids with *CYP2D7* (Gaedigk, 2013). In the 1000 Genomes Project, there were 149 exonic single nucleotide variants found and seven exonic indels identified (Yang *et al.*, 2017). From this, more than 100 variants/alleles and sub-variants/sub-alleles have been classified according to the Human CYP Allele Nomenclature committee, and are being studied extensively to understand their effects on CYP2D6 enzyme activity (Muroi *et al.*, 2014). All the variation and differences are associated with changes in CYP2D6 activity depending on the variants, combinations of variants, genetic structural changes, copy and tandem repeats, or a lack of a functional gene (Muroi *et al.*, 2014). The dysfunctional alleles for the *CYP2D6* gene can be divided into three groups in terms of their effects on enzyme activity (Zanger, Raimundo and Eichelbaum, 2004).

The first group is null alleles, which are alleles that result in a lack of a functional protein product, and these alleles result in the PM categorisation when in homozygous or compound heterozygous states. In this group there are numerous different mechanisms that resulted in a

complete loss of a functioning enzyme (Zanger, Raimundo and Eichelbaum, 2004). Some alleles (**3, *4, *6, *8, *11, *15, *19, *20, *38, *40, *42*, and **44*) are involved in reading frame disruptions or splice site interference by either single base pair mutations or small indels resulting in truncated proteins from premature stop codons. A few alleles *(*7, *12, *14*, and **18*) have variants that allow for the translation of a full-length non-functional protein. There are also three alleles that contain deletions of the whole *CYP2D6* gene (**5*) or the *CYP2D6/2D7* hybrid genes (**13, *16*) (Zanger, Raimundo and Eichelbaum, 2004). The second group is the decreased enzyme activity associated alleles, made up of numerous alleles. The first allele to be classified as a decreased activity allele, and part of the intermediate metaboliser (IM) phenotype, was *CYP2D6*9*. Some of the most common as well as the most studied alleles of this group are **9, *10, *17* and **41* (Zanger, Raimundo and Eichelbaum, 2004). The last group is the increased enzyme activity associated alleles, which show an ultrarapid metabolism (UM) phenotype. This is due to numerous genetic phenomena that cause an increased activity in CYP2D6 or overexpression of *CYP2D6* gene. This is due to numerous genetic phenomena that cause an increased activity in CYP2D6 or overexpression of *CYP2D6* gene and is caused mainly in two ways. The first is by CNVs or tandem repeats, where large sections of the gene or the entire gene are duplicated and cause increased expression of more enzyme transcripts. The second is by variants that cause the overexpression of the gene such as the -1584 C>G promoter region SNP. (Zanger, Raimundo and Eichelbaum, 2004). When looking at the CNVs present in *CYP2D6*, there is still much to uncover, but they do play a significant role in enzyme activity diversity with *CYP2D6*1xN* or **2xN* seen at frequencies of 12% worldwide.(Dalton *et al.*, 2019).

1.2.5 African-Specific Allele *CYP2D6*17*

The *CYP2D6*17* allele is part of the decreased enzyme activity group and homozygous individuals are categorised as IMs (Wennerholm *et al.*, 2002; Zanger, Raimundo and Eichelbaum, 2004). It consists of a haplotype of four exonic single nucleotide variants (rs28371706, rs1058164, rs16947, rs1135840), three of which are missense variants and a single silent mutation (rs1058164) (Masimirembwa *et al.*, 1996; Oscarson *et al.*, 1997). The three missense variants (rs28371706, rs16947, rs1135840) cause substitution mutations (1022C>T, 2851C>T, 4181G>C) (Fig 1.7) in the gene that results in amino acid changes (T107I, R296C, S486T) in the polypeptide chain (Fig 1.8) (Masimirembwa *et al.*, 1996). The substitution mutations 2851T (rs16947) and 4181C (rs1135840) are also part of the *CYP2D6*2* allele, which is one of the most common alleles for *CYP2D6* and has a normal effect on enzyme activity. The 1022T variant (rs28371706) is the defining variant of the

*CYP2D6*17* allele and haplotype, and when found with the two previously mentioned missense variants causes the reduced function effect (Oscarson *et al.*, 1997).

Allele Variants	rs28371706	rs1058164	rs16947	rs1135840
Chromosome Position (GRCh38.p12)	22:42129770	22:42129130	22:42127941	22:42126611
SNP Transcript Position	1022bp	1661bp	2851bp	4181bp
CYP2D6 Reference	C	G	C	G
*CYP2D6*2*	C	C	T	C
*CYP2D6*17*	T	C	T	C

Fig 1.7: Schematic diagram of *CYP2D6* and genetic arrangement of the *CYP2D6*17* variant. Lines show the positions of the four SNP variants in the exons of the gene.

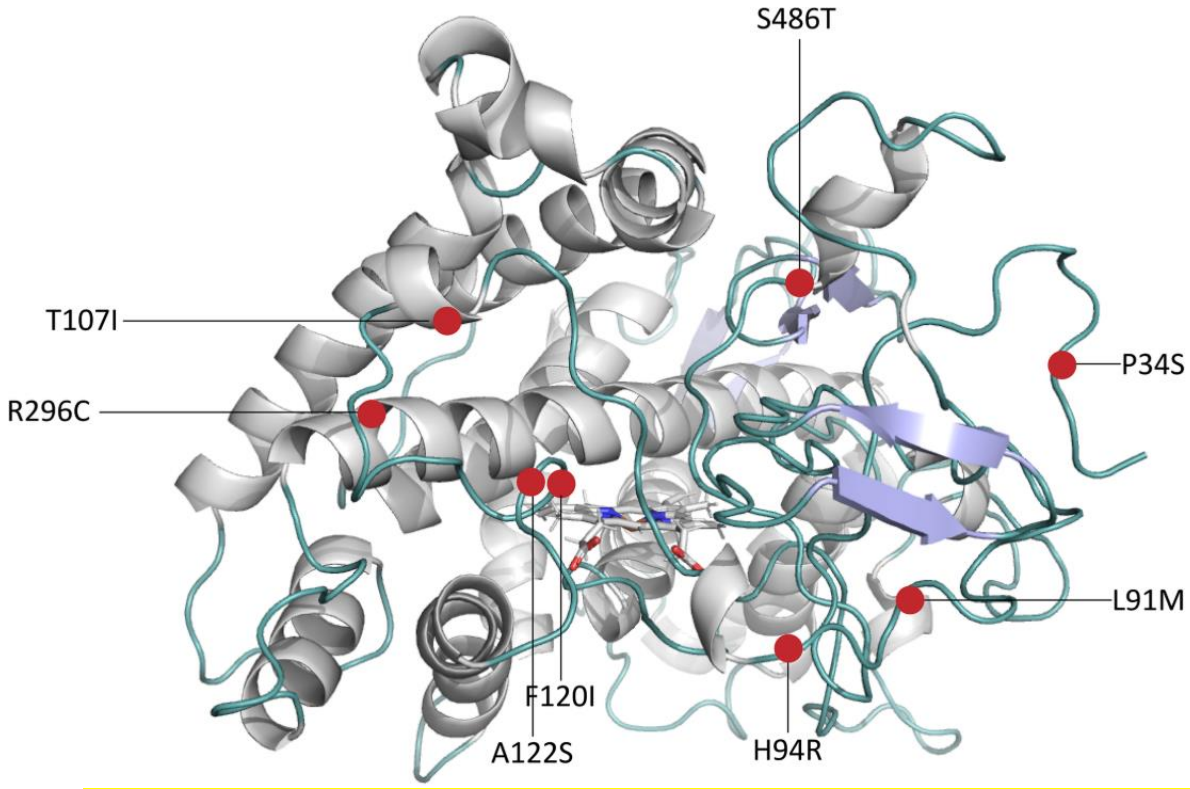


Fig 1.8: Tertiary structure of CYP2D6 with the positions of some of the most common SNPs associated with the enzyme. The red circles highlighted in the SNPs and their positions in the protein structure. The SNPs are associated with the following star alleles: *CYP2D6*2* (R296C and S486T), *CYP2D6*10* (P34S and S486T), *CYP2D6*17* (T107I, R296C and S486T), *CYP2D6*4* (P34S, L91M, H94R and S486T), and *CYP2D6*53* (F120I and A122S). *Image from Don et al., (2018).*

In recent years, the *CYP2D6*17* allele has garnered significant interest from African researchers, as it is essentially specific to African population and is not found in significant frequencies in other populations (Zhou, Ingelman-Sundberg and Lauschke, 2017). The frequencies of the *CYP2D6*17* allele vary across populations worldwide with approximately 20% in African populations and less than 1% in European and Asian populations (Zhou, Ingelman-Sundberg and Lauschke, 2017). When looking at the public genomic data sites and

databases, such as Ensembl (1000 Genomes Project (KGP) Data) and PharmVar (Clinical Pharmacogenetics Implementation Consortium (CPIC) data), this variant has frequencies of 20.7% and 19.9% respectively (Auton *et al.*, 2015; Nofziger *et al.*, 2019). Interestingly, there are similar frequencies in individuals of African-American descent in the PharmVar database (18.7%) and the KGP data (13.9%), hinting at conservation of this allele, being observed also following admixture of this population over time (Auton *et al.*, 2015; Nofziger *et al.*, 2019). However, Zhou et. al. suggests a different frequency in the Admixed American populations (0.7%) (Zhou, Ingelman-Sundberg and Lauschke, 2017). Nonetheless, this shows the importance of studying *CYP2D6*17* in the African context and also understanding the molecular implications it has on drug metabolism and clinical outcomes (Bapiro *et al.*, 2002; Zhou, Ingelman-Sundberg and Lauschke, 2017).

1.2.6 Protein Structure of CYP2D6 & CYP2D6*17

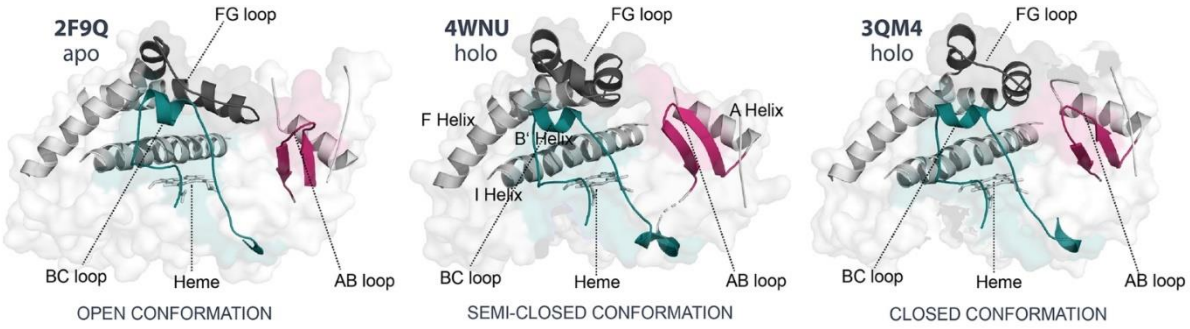
The 3D structure of drug-metabolising enzymes has always been a key component of drug development and design(Guengerich, Waterman and Egli, 2016). One of the important goals set out by early researchers into CYP2D6 was to acquire accurate knowledge of CYP2D6 3D structure to create templates to improve the *in-silico* drug design process. This also necessitated detailing, with great precision, the protein interactions that occurred between CYP2D6 and its drug substrates (Guengerich, Waterman and Egli, 2016). CYP enzymes generally all share a similar structure and there is agreement that the chemical properties of CYP enzymes are mostly invariant. The positioning of substrates in the active site, which houses the FeO^{+} that catalyses the oxidative reactions, is the reason for catalytic specificity across the CYP enzyme superfamily. Hence, the B-C loop conformation and position is responsible for the differences seen in CYPs and catalytic selectivity (Guengerich, Waterman and Egli, 2016).

The active site cavity has a “right foot” shape and is well defined with a volume of approximately 540 $Å^3$. The borders of the site are made up of the heme group, B-C Loop, F-helix, G-helix, I-helix, K-helix and β-sheet 1 loop, and the β-sheet 4 loop residues (Figure 6). Access to this “foot” cavity of the active site is done through the normal solvent channel that passes through the second substrate recognition site (SRS) 2 and the SRS6, found in the F-helix and β-sheet 4 respectively (Rowland *et al.*, 2006). The structure of CYP2D6 suggests that it is a relatively elastic protein that allows for substrates to pass through its ‘tunnels’ and induces a closed conformation around these substrates depending on their characteristics (Fig 1.9) (Guengerich, Waterman and Egli, 2016). The loops are critical in allowing this elasticity and flexibility, and if the loops become more rigid or flexible due to a mutation, this will

result in dissimilar thermodynamics and a changed enzyme activity. This inducible fit in the active site around substrates is responsible for the numerous drugs that are metabolised by CYP2D6 (Guengerich, Waterman and Egli, 2016).

Fig 1.9: Conformational changes in CYP2D6, and their effects on the protein structure and access to the active site. *Image from Don et al., (2018).*

Hydrogen bonding between amino acids is also important for the structural fold integrity of the protein. They can have similar effects on the enzyme activity by altering the



conformational flexibility, and this could affect enzyme thermodynamics (Don and Smieško, 2018). For the CYP2D6*17 enzyme variant, the interaction between the amino acid residues Thr^{107} and Asn^{255}, found in the G-helix of the FG-loop, is changed. This is hypothesised to cause a more flexible FG-loop which is associated with a closed conformation and the inactive state of the enzyme (Don and Smieško, 2018). Also, the mutated Ile^{107}, from the BC loop, interacts more with the Phe^{112} residue and less with the Leu^{110} amino acid. This is argued to result in an extended hydrophobic residue interaction that reduces the flexibility of the BC-loop. These two phenomena together are thought to hamper the access to tunnels 2c and 2e to the active site, and hence decrease enzyme activity (Don and Smieško, 2018).

1.2.7 Drugs Metabolised by CYP2D6*17

1.2.7.1 *Tamoxifen*

Tamoxifen is a selective oestrogen modulator that has been prescribed frequently for approximately 40 years in the treatment of hormone/oestrogen receptor-positive ($H^{+/}ER^{+)}$ breast cancer (Kiyotani *et al.*, 2012). It is also used in preventative treatment strategies for breast cancer recurrence as it reduces the risk of developing $H^{+/}ER^{+}$ and non-invasive breast cancer (Fisher *et al.*, 2005). It has been seen that tamoxifen administration in adjuvant therapy for five years shows a reduction in breast cancer recurrence risk by up to 50%, and has long-term survival benefits after 10 years of administration (Davies *et al.*, 2013). Tamoxifen has also been recommended as an adjuvant endocrine therapy for premenopausal women with $H^{+/}ER^{+}$ breast cancer, and also as an aromatase inhibitor treatment substitute for postmenopausal women with $H^{+/}ER^{+}$ breast cancer (Ahern *et al.*, 2017).

Tamoxifen is a prodrug, and its pharmacological relevant metabolite compound is endoxifen (4-hydroxyl-*N*-desmethyltamoxifen) which is the primary active metabolite. Endoxifen works by suppressing the $H^{+/}ER^{+}$ breast cancer by binding the receptors, ERα and ERβ, on breast cells (Goetz *et al.*, 2007). This antiestrogenic effect of endoxifen inhibits the expression of oestrogen-related genes that are involved in cell growth and angiogenesis via the autocrine and paracrine systems (Osborne, 1998). Endoxifen has a 100-fold higher affinity compared to tamoxifen for these receptors, and 30- to 100-fold higher suppression of cell proliferation (de Vries Schultink *et al.*, 2015). The entire process and reactions involved are catalysed by various CYP enzymes, with the primary pathway involving CYP3A4/5 and CYP2D6. There are two steps to this catalytic conversion and the first is performed mainly by CYP3A4/5 which metabolises tamoxifen to its metabolite *N*-desmethyltamoxifen (Jin *et al.*, 2005). CYP2D6 is involved in the second step in which *N*-desmethyltamoxifen is metabolised to endoxifen (Fig 1.10). The rate of reaction of tamoxifen to endoxifen metabolism has been seen to be less than 0.05 $pmol.min^{-1}.pmol^{-1}$ in CYP2D6*17 enzymes, while the wildtype CYP2D6*1 has a max velocity of 0.45 $pmol.min^{-1}.pmol^{-1}$ (Muroi *et al.*, 2014).

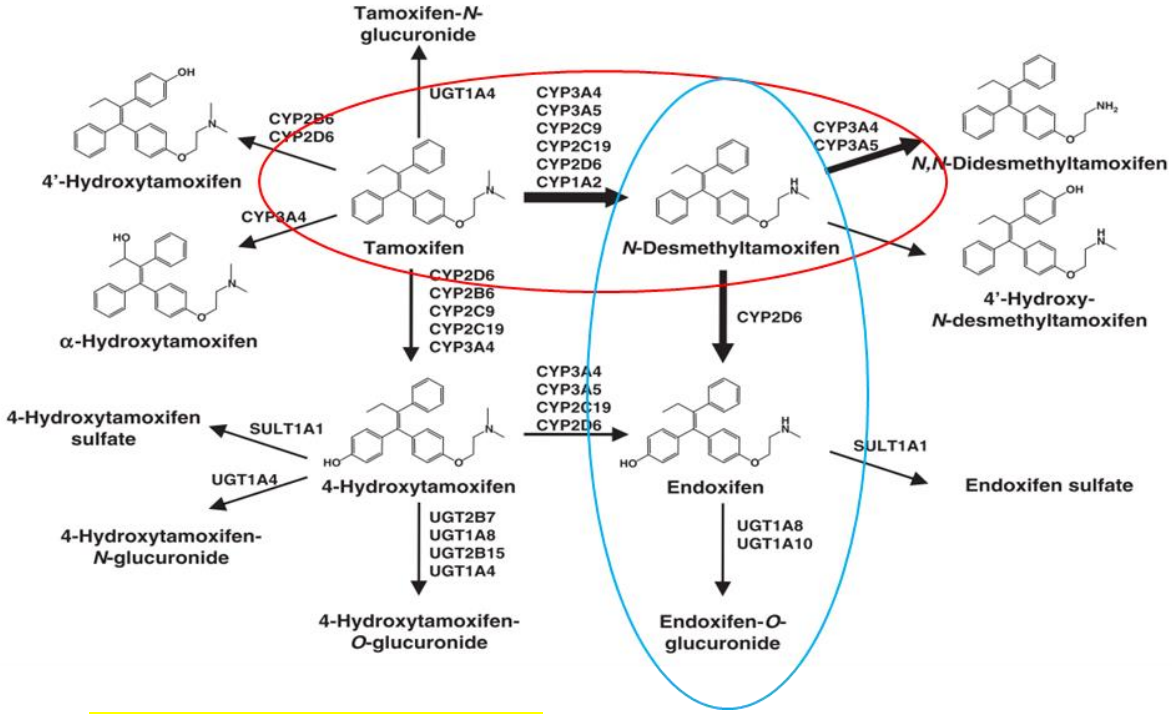


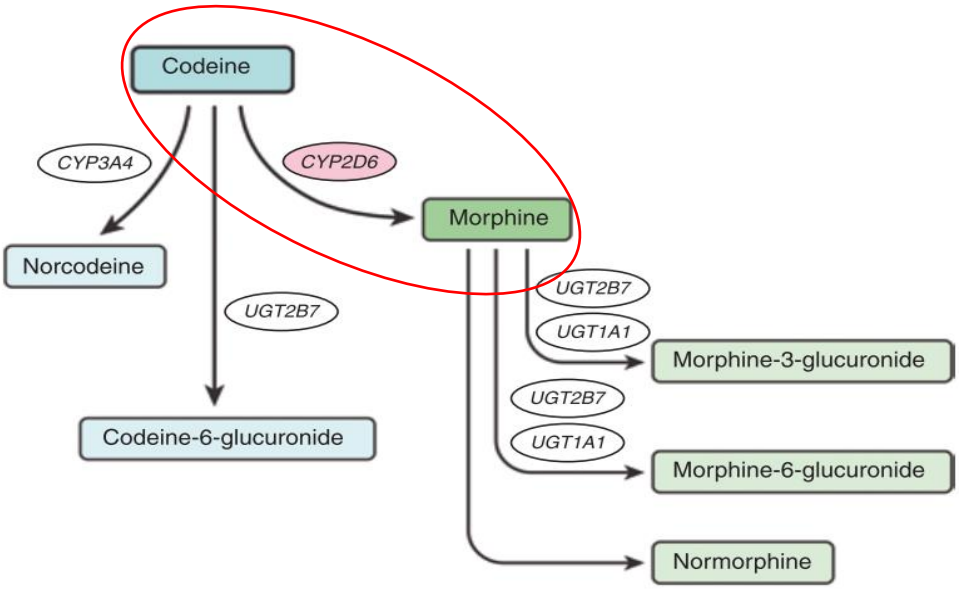
Fig 1.10: Human metabolic pathways of tamoxifen. In the red circle is the tamoxifen initial drug metabolism step performed by CYP3A4/5 enzyme. In the blue circle is the drug metabolism step performed by the CYP2D6 to endoxifen. *Image from: Muroi et al., (2014).*

This shows that for individuals who are homozygous for the reduced function CYP2D6*17 enzyme, there is a marked decrease in the amount of endoxifen produced from the metabolism of tamoxifen by the enzyme, rendering the treatment ineffective (Jin *et al.*, 2005).

1.2.7.2 *Codeine*

Codeine is an opioid analgesic drug that is used for mild to moderately severe pain relief and antitussive effects (Crews *et al.*, 2012). It interacts with the opioid receptors that are found throughout the body, such as the GIT and central nervous system (CNS) for example (Pratt *et al.*, 2018). However, codeine is only a prodrug and has a weak affinity for these μ-opioid receptors, and is converted to is active metabolite morphine, which has a 200-fold stronger affinity to μ opioid receptors (Crews *et al.*, 2012; Pratt *et al.*, 2018). Codeine is converted into morphine as a result of *O*-Demethylation by CYP2D6 where is converted into morphine-6-glucuronide through glucuronidation (Fig 1.11). This pathway is responsible for only 5-10% of codeine metabolism into the active metabolite morphine, and 80% is metabolised by UGT2B7 and CYP3A4 into the inactive metabolites' codeine-6-glucuronide and norcodeine respectively (Fig 1.11). However, this pathway is essentially responsible for opioid activity in the body, and the rate and quantity of converted codeine to morphine is heavily dependent on *CYP2D6* polymorphism (Crews *et al.*, 2012).

The main adverse effects of codeine are drowsiness, light-headedness, dizziness, sedation, shortness of breath, nausea, vomiting, and sweating. The more serious adverse effects are respiratory depression, circulatory depression, respiratory arrest, shock, and cardiac arrest (though the last four are seldomly experienced by patients and mainly seen in UM



individuals) (Crews *et al.*, 2012). *CYP2D6* polymorphisms that are linked to PM phenotypes result in decreased levels of morphine as well as a decrease in some adverse effects (i.e. constipation) and no effect on others (i.e. sedation and nausea). In the UM phenotype, there can be an increase in morphine levels to toxic systemic concentrations, even at very low dosages of codeine administration (Crews *et al.*, 2012). The CYP2D6*17 enzyme has been estimated to have average conversion rates of 80% to morphine as compared to CYP2D6*1 (Shen *et al.*, 2007). Though this may seem like an insignificant change, even the smallest change in rate can have significant effects on the therapeutic outcomes of codeine administration. CYP2D6*17 individuals may also experience poor response to the analgesic effects of codeine due to reduced morphine conversion (Pratt *et al.*, 2018). For PM and UM, as well as poor-responding IM, patients, tramadol is seen as an alternative drug for pain relief as it is responsible for the opioid receptor interaction that causes the analgesic effects. However, PM studies have shown reduced response to tramadol and so clinicians must be aware of the individual patient response to the medication (Crews *et al.*, 2012). To avoid complications with opioid treatment, clinicians should either choose alternative non-opioid drugs, or pair up an opioid with a non-opioid to mitigate the adverse effects, and to do this on a case-to-case basis (Crews *et al.*, 2012).

Fig 1.11: Human metabolic pathway of codeine. The red circle represents the codeine to morphine metabolic conversion by CYP2D6 *O*-demethylation. *Image from: Crews et al, (2012).*

Table 1.1 has a summary of some of the major substrates of CYP2D6 and the effects of each metaboliser group on the outcome of the drug metabolised.

Table 1.1: Major substrates of CYP2D6 and the concerns for particular metaboliser groups.

1.3 CURRENT MODELS FOR DRUG METABOLISM

The number of approaches to study drug metabolism has increased substantially as the technologies and techniques to study biological systems has improved. Some of the most widely available modelling systems are animal models and 2D human cellular models. These advancements have brought forth several questions of the pros and cons of all these models, sparking debate on which is best to use. Animal models for the study of drug metabolism include mice (Perlman, 2016), rabbits (Zuber, Anzenbacherová and Anzenbacher, 2002), monkey (Zuber, Anzenbacherová and Anzenbacher, 2002), pigs (Achour, Barber and Rostami-Hodjegan, 2012) and Zebrafish (MacRae and Peterson, 2015). In their natural state, animal models all face drawbacks from the interspecies difference between animals and humans which limit the extrapolation of animal DMPK to human DMPK (Montagutelli, 2015). There is, however, some relatively useful application of transgenic animal models where genes of human drug metabolising enzymes have been expressed and studied (Bissig *et al.*, 2018). These drawbacks have partly been met by the advances in human cell culture techniques in both 2D and 3D approaches, resulting in less use of animal models (Montagutelli, 2015). Currently most of the industry has turned their focus to vitro models as

Major substrates of CYP2D6

Substrate	Metabolite	Effects on Metaboliser Groups	Alternative Drugs for use
Codeine	Morphine	PM – decreased morphine levels and decrease in analgesic effects. IM – slightly decreased morphine levels and slight decrease in analgesic effects. UM – Increase in morphine levels to toxic concentrations. Risk of respiratory or circulatory depression, respiratory or cardiac arrest, and shock	Tramadol (may have a reduced response in PM individuals)
Primaquine	5-hydroxylprimaquine	PM & IM – Decreased levels of metabolite and decreased effects on malaria infection	None
Risperidone	Paliperidone	PM & IM – Decreased levels of metabolite for clearance and increase risk of side effects UM – Increased levels of metabolite formation for clearance and decrease effectiveness of the treatment	Haloperidol, Aripiprazole, Paliperidone
Tamoxifen	Endoxifen	PM & IM – Decreased levels of endoxifen and decrease in cancer treatment response. Increase levels of relapse in remission patients.	Aromatase inhibitors

being more predictive of human DMPK (Montagutelli, 2015). In this section, the *in vitro* systems that are being applied in drug discovery will be briefly reviewed.

1.3.1 Human Liver Microsomes

Liver microsomes have been commonly utilised in numerous studies to assess and investigate drug metabolism *in vitro* (Strathern, 2014). They are obtained from the endoplasmic reticulum of hepatocytes and are prepared by the use of liver homogenization and differential centrifugation (Pelkonen *et al.*, 1974; Li, 2005). They are mostly used for studying and evaluating the phase I oxidative metabolism of xenobiotic compounds with the supplementation of cofactors such as nicotinamide adenine dinucleotide phosphate (NADP) (Asha and Vidyavathi, 2010). They were discovered, in 1955, as a potential area for biotransformation and metabolism of drugs in rabbit liver (Brodie *et al.*, 1955; Palade and Siekevitz, 1956). By 1958, the enzymes, now known as the CYP enzymes, were identified and named in rat liver microsomes. The name P450 coming from the maximum absorbance of 450 nm leading the name Pigment-450 nm (Klingenberg, 1958). This led to new uses of liver microsomes to establish tools and techniques to study the metabolism of xenobiotic compounds for pharmacological interventions. In recent times, understanding the metabolic rate of different test compounds and investigating their metabolic half-life *in vitro*, often used liver microsomes as an adequate testing system, due to the amenability to automation and formats for high-throughput assays (Strathern, 2014). The expression of CYP enzymes in liver microsomes varies rather significantly due to interindividual differences between liver sample donors. However, this is overcome by pooling the liver microsomes to have an overall adequate drug metabolism response. This makes the CYP expression comparable to in-vivo human expression once microsome data is pooled. CYP expression and activity in microsomes is also affected by gender, and pooling aids in investigating gender-specific influences on drug metabolism (Parkinson *et al.*, 2004; Asha and Vidyavathi, 2010).

1.3.2 Other Subcellular Fractions: Cytosol & S9 Fractions

The Cytosol is not a commonly used model for phase I drug metabolism but is often used for phase II metabolism research. The phase II enzymes that are found within the cytosol are glutathione S-transferase (GST), N-acetyl transferase (NAT) and sulfotransferase (ST). However, due to the location of the phase II enzyme uridine diphosphoglucuronosyl transferase (UGT) being in the endoplasmic reticulum, cannot be studied using cytosol (Asha and Vidyavathi, 2010). S9 subcellular fractions are also extensively used in DMPK studies, however, they are not as commonly used as liver microsomes. They were first discovered in the 1970s and contain the same liver enzymes found in both cytosol and microsomes. The S9

fractions are often supplemented with NADPH, which is needed as an energy supply for the CYP enzymes. Due to their composition of both phase I and II enzymes, S9 fraction are useful in investigating metabolic activity and pathways where both are needed. However, this is met with the low levels of enzyme activity and can result in overlooked metabolism of xenobiotic compounds (Asha and Vidyavathi, 2010).

1.3.3 PXR Reporter Gene Assays

The pregnane X receptor (PXR) is a protein receptor that is part of the nuclear receptor (NR) family as it shares significant sequence homology with other NR proteins. The PXR is found in the liver and intestines, and is responsible for the regulation of numerous drug metabolism genes such as those that encode for the CYP and GST enzymes (Shukla *et al.*, 2011). PXR has also been identified as an important regulator of *CYP3A4*, and facilitates the drug-induced expression of CYP3A4 (Luo *et al.*, 2002; Shukla *et al.*, 2011). The ability of PXR to bind to CYP3A4 substrates have been shown to have a correlation with *CYP3A4* expression, and this PXR-induced expression is involved in approximately 60% of drug-drug interaction (Shukla *et al.*, 2011). This led to PXR reporter gene assays being used as a method to investigate *CYP3A4* drug inducers, and look at the drug metabolism pathways that *CYP3A4* is involved in (Luo *et al.*, 2002). PXR reporter gene assays have become attractive models for drug metabolism due to the lowered cost by assay compatibility with low-cost cells lines, it is more consistent in the results due to the lack of large variation between cell donor sources and reproducibility of the assay (Luo *et al.*, 2002; Raucy *et al.*, 2002). However, this system is hampered by its inability to observe drug interactions that are not linked to CYP3A4 (Luo *et al.*, 2002.

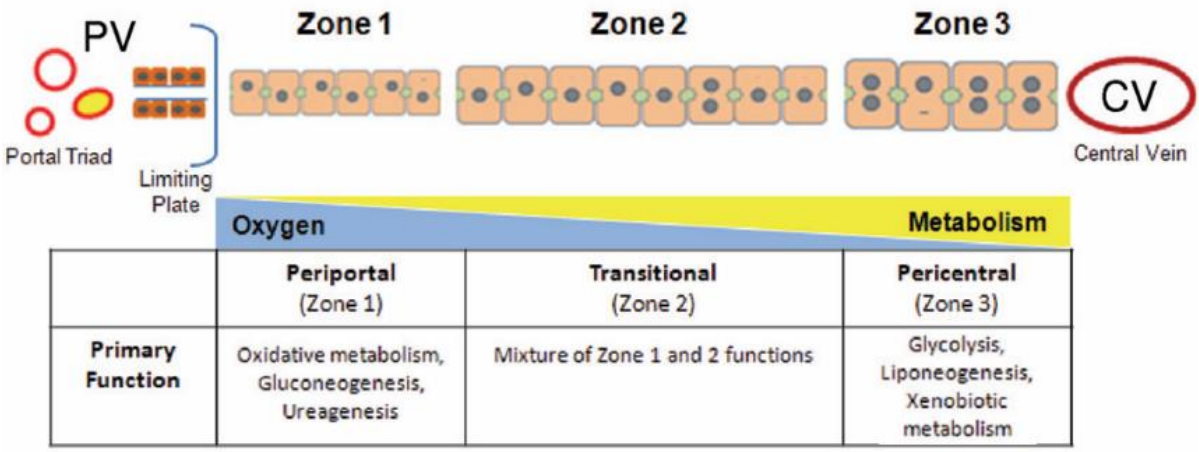
1.3.4 Liver Slices

Liver slices are portions of liver that have numerous liver cell types retained in their original structure and ECM *in vivo*. This allows for the study of how different cell types interact with each other and the extracellular environment. They also allow for research into the selective nature of drug transporters and drug-carrier conjugates to each cell type, which is critical for better understanding into drug targeting technologies (Olinga and Groothuis, 2001). Some of the attributes of liver slices that makes this an attractive model for drug metabolism is the low cost, maintenance of hepatocyte phenotype, and reproducibility of results (Palma, Doornebal and Chokshi, 2019). However, some of the drawbacks of this as a drug metabolism model are the short viability timespan, and there is an insufficient penetration of culture medium into the liver slices affecting the long-term viability of the slices (Asha and Vidyavathi, 2010).

1.3.5 Primary Human Hepatocyte Cellular Models

Human cell cultures have been frequently used to study liver biology and human diseases, as well as develop drugs and treatment regimens viable for clinical use. With the advent of new culturing techniques and technologies, the question about which human cell model should be used is still largely unanswered (LeCluyse *et al.*, 2012). With growing needed for more economic, regulated and practical systems to better understand and predict drug response, the need for more relevant and complex *in vitro* models has grown (LeCluyse *et al.*, 2012). For this to be achieved, aspects of *in vivo* liver functioning for modelling need to be understood (LeCluyse *et al.*, 2012). The liver functions in many ways to serve a vast array of critical roles in detoxification in the body. The structure of the liver aids in its function and this is seen in the zonation of the liver lobules and zonal expression of genes according to each zone (Fig 1.12). The best example of this is seen in the expression of CYP3A4, which is in the pericentral region, and is also the same area that is most affected by liver toxicity when exposed to particular compounds such as acetaminophen (LeCluyse *et al.*, 2012). This zone is made up of mature hepatocytes which are responsible for xenobiotic metabolism, and are the primary cell used for culturing in studies of drug metabolism (LeCluyse *et al.*, 2012).

The harvesting and isolation of primary human hepatocytes (PHHs) are the gold standard for accurate *in vitro* model for studying the metabolic functions of the liver (Bell *et al.*, 2016). These cells are mature liver cells that are fully differentiated and can maintain their *in vivo*



	Periportal (Zone 1)	Transitional (Zone 2)	Pericentral (Zone 3)
Primary Function	Oxidative metabolism, Gluconeogenesis, Ureagenesis	Mixture of Zone 1 and 2 functions	Glycolysis, Liponeogenesis, Xenobiotic metabolism

Fig 1.12: Liver lobule zonation and the metabolism the function of the hepatocytes in each zone. The abbreviations in the image stand for the following: PV = Portal Vein, CV = Central Vein. *Image from: LeCluyse, E.L. et al., (2012).*

expression and response when exposed to physiopathological factors. PHHs are capable of metabolising drugs for a maximum of two to three days (Gomez-Lechon *et al.*, 2004).

However, once they are plated in 2D culture they engage in a rapid and spontaneous dedifferentiation losing CYP metabolism functioning and overall protein expression within

the first 24 hours (Andersson *et al.*, 2018). This has been a serious limitation for their viability in long-term metabolism studies as their utility is limited to a short timespan (Bell *et al.*, 2016). Other problems PHH monolayers exhibit are hepatotoxin sensitivity and an inability to create *in vivo* zone-specific differentiation (LeCluyse *et al.*, 2012). This being said, PHHs remain the gold standard for drug metabolism and toxicity studies as they are the closest cells to liver for *in vitro* modelling (Gomez-Lechon *et al.*, 2004; Guo *et al.*, 2011). Some of the methods to overcome this problem are found in the culturing of PHHs in a 3D spheroid as this increases their viability and the production of the essential liver molecules (Bell *et al.*, 2016).

PHH spheroids are cellular 3D structures that are cultured on non-adhesive surfaces under specific conditions and create a deposit of ECM that forms a capsule around the spheroid (LeCluyse *et al.*, 2012). Several of the benefits of spheroids are that have a 3D cytoarchitecture important for *in vivo* like functioning and the ECM allows for cell-cell interactions critical for cell function. There is also cell organisation that takes on a histotypic manner similar to that found in the liver, which further aids with more accurate hepatocyte functioning (LeCluyse *et al.*, 2012). Regarding CYP enzyme activities, in 3D-spheroid culture systems the activities had minimal changes for five weeks in a serum-free culture. Overall, the spheroids' proteome levels were comparable to liver tissue proteome levels *in vivo* (Bell *et al.*, 2016). However, their viability in high-throughput drug metabolism studies are limited due the high cost associated with PHH culturing, cell scarcity and the heterogenous nature of donations and isolations of PHHs, and thus alternative cellular models have been pursued (Sullivan *et al.*, 2010).

1.3.6 HepG2 Cellular Models

HepG2 is a human liver cell line derived from liver tissue of a 15 year old Caucasian American male with differentiated hepatocellular carcinoma, and is the most well studied and frequently used liver cell line (LeCluyse *et al.*, 2012; Luckert *et al.*, 2017). HepG2 cells are epithelial-like in morphology and adherent in small aggregates when cultured in 2D monolayers. They possess 55 chromosomes as a result of their cancerous characteristics; however, they do not show any tumorigenic characteristics associated with cancer cells. They act in ways that are typical to that of hepatocytes *in vivo* and perform normal hepatocyte functions such as biotransformation of xenobiotic compounds (LeCluyse *et al.*, 2012). The attractiveness of HepG2 as a model for drug metabolism stems from their easy maintenance in culture, their wide availability, and cost effectiveness (Guo *et al.*, 2011). Though in comparison with PHHs, there are significant differences seen in the basal gene expression, such as an increase in cell cycle gene expression and decreased *CYP* gene expression

(LeCluyse *et al.*, 2012; Choi *et al.*, 2015). These shortcomings make them a good candidate for modelling simple hepatic functions, but they may not be appropriate for modelling more complex hepatic functions such as CYP enzyme drug metabolism critical for drug development studies (LeCluyse *et al.*, 2012).

In the cases of CYP1A2 and CYP3A4, it has been shown that HepG2 cell lines possess expression levels comparable to PHHs (Choi *et al.*, 2015). HepG2 cell lines have been shown to induce CYP1A2, CYP2B6, and CYP3A4 at a reproducible rate under specific conditions and these show the potential of their use in early drug discovery and metabolism studies (Choi *et al.*, 2015). The use of 3D structure culturing in a Matrigel matrix demonstrates that HepG2 cells can recover some of the hepatic functioning that is not seen in 2D monolayer growth. There was also an increase in xenobiotic metabolism, especially with CYP enzyme activity in comparison to 2D culturing of HepG2s (Choi *et al.*, 2015). In regards to CYP enzyme level production, there was an increase in the 3D culture and detection of CYP3A4, CYP2E1, CYP2C9, CYP2C19, and CYP2D6 even after 28 days compared to 3 days in 2D culture (Choi *et al.*, 2015). These increases in CYP and other hepatic cell functions in 3D spheroids displays the influence of long-term cultivation and partial differentiation of cells due to aging. Nonetheless, the success of increased cell function in HepG2 cell may not be feasible and substantial enough in relation to the high cost of 3D culturing (Choi *et al.*, 2015).

1.3.7 iPSC Cellular Models

Induced pluripotent stem cells (iPSCs) are derived from somatic cells and are able to undergo differentiation into a number of cell types (Takahashi *et al.*, 2007). This is achievable through the introduction of four transcription factors (Oct3/4, Sox2, c-Myc and Klf4), which are involved in early embryogenesis and pluripotency maintenance, via viral-mediated transfection (Takahashi and Yamanaka, 2006). This is frequently performed using dermal fibroblasts which were the first human cells that successfully undergo cellular reprograming into iPSCs (Takahashi and Yamanaka, 2006; Takahashi *et al.*, 2007). Dermal fibroblasts are still the most common choice for reprogramming of cells into iPSCs, and this is due to them being inexpensive to acquire and relative easily to work with in-vitro. Acquiring dermal fibroblasts is normally done via skin biopsy by a qualified dermatologist or by a commercially available cell line. Though, it is not easy to find fibroblasts from donors with rare pharmacogenetic variants or uncommon diseases (Raab *et al.*, 2014). Commercially derived fibroblast lines are available, however, it is difficult to identify donors with precise pharmacogenetic variants. This is particularly the case with those of African descent harbouring rare pharmacogenetic SNPs (Raab *et al.*, 2014). This is because almost all

commercially available lines are derived from donors in the US and Europe, thus minimizing the chance of securing donors with rare African variants. Thus, many studies secure the expertise of a dermatologist to perform skin biopsies on individuals post confirmation of genetic status to enable the study of rare genetic variants (Roses, 2000; Raab *et al.*, 2014). One of the most commonly used viral vectors for reprogramming of dermal fibroblasts and other cells is Sendai virus (SeV). SeV has become a very popular choice for viral vector delivery approaches due to its episomal expression of the reprogramming genes negating genome integration (Nakanishi and Otsu, 2012).

Reprogramming of iPSCs results in the cells acquiring the ability and capacity to differentiate into all germ layers, and subsequently into any cell type in the body (Nakanishi and Otsu, 2012). The ability for iPSCs to potentially differentiate into any cell type in the body has made them an extremely attractive liver-like cellular model for drug metabolism and toxicity studies. They also have important utility in studying the interindividual differences and genetic polymorphisms that affect drug metabolism in the liver (Sullivan *et al.*, 2010). Another use of iPSCs is to better understand the polygenetic mechanisms and cell environment factors that affect drug response and metabolism in a reproducible manner. This has resulted in iPSC differentiated liver-like cells as a viable contender for modelling disease, drug response, and accurate deciphering of the gene interactions within the liver (Sullivan *et al.*, 2010). The hepatocytes that are differentiated from iPSC cells are referred to as stem-cell-derived hepatocyte-like cells (SC-HLCs) which can be used to study liver functioning *in vitro* (Goldring *et al.*, 2017).

The process of making SC-HLCs is achieved by first guiding stem cells down a path to definitive endoderm by growth factor usage, such as activin A and Wnt3a. The next step is specification of hepatic progenitor cells and then the maturation of the derived hepatocytes (Meseguer-Ripolles *et al.*, 2018). These cell models have the potential to be used in numerous areas of study pertaining to drug metabolism in mature hepatocytes. This is due to SC-HLCs being capable of retaining their CYP expression profiles, especially with CYP2D6 (95% comparable with PHHs) and CYP3A4 (90% comparable with PHHs) (Goldring *et al.*, 2017). When established within a 3D structure, there is an increase in hepatic cell functioning as the 3D spheroid provides a cell environment that is more similar to the *in vivo* environment, and aids in their functioning (Meier *et al.*, 2017). The 3D spheroid CYP functioning and expression is still lower than that in PHHs but has a marked increase when comparing them to 2D cultures. The culture system also makes a vast difference when utilising 3D culture

techniques such as bioreactors which display increases in CYP expression and activity by increasing the developmental state through environmental specifications (Meier *et al.*, 2017).

The utilisation of SC-HLCs in modelling drug metabolism and toxicity carries promise despite having many challenging factors in creating hepatocytes with a mature cell profile. The other challenge is to achieve the mimicking of liver development *in vivo*, with regard to the zone-specific sensitivity including hepatocellular damage (Goldring *et al.*, 2017). It is also of immense importance to create iPSCs panels and lines that are representative of all the polymorphisms found in different populations. This is to ensure better outcomes when studying these effects in iPSC models, and to also ensure that each population has appropriate and effective drug compounds and dosages for treatment of disease (Takayama *et al.*, 2014). IPSC technology pertaining to these endeavours has not yet been perfected and is still in early stages of development and implementation, but there is more evidence mounting of their utility in drug metabolism (Scott, Peters and Dragan, 2013). As we advance in the field of drug metabolism, it appears that iPSC models could be very powerful and innovative tools to aid in drug development. Also, as we embark on achieving precision and personalised medicine, these technologies may be key to the implementation of some of the new technologies in healthcare (Takayama *et al.*, 2014).

1.4 CONCLUSION

There is a need for a better understanding of African pharmacogenetics and drug metabolism profiles, as well as better models to investigate pharmacokinetics and pharmacogenetics. Therefore, this study aimed to assess the frequency of the *CYP2D6*17* variant in African populations and to evaluate its enzyme activity on the activation of tamoxifen to endoxifen in iPSC-derived hepatocyte-like cells (SC-HLCs).

1.5 STUDY RATIONALE

The problem of insufficient knowledge on the prevalence of genetic variants, such as *CYP2D6*17*, and their effects on drug metabolism is a barrier to the advent of precision medicine platforms. Furthermore, the impact of incorrect drug prescription and inaccurate drug dosage allocation to individuals in specific populations has contributed extensively to ADR-related hospitalisation and death and needs tenable solutions to resolve this matter. The ever-growing need for more African-based research has prompted an opportunity for interventions using different approaches, particularly a cost-effective and powered approach. Thus, the study aimed *to assess the frequency of the CYP2D6*17 variant in African populations and to evaluate its enzyme activity on the activation of tamoxifen to endoxifen in patient derived iPSC-derived hepatocyte-like cells (SC-HLCs).*

The following objectives were proposed to achieve the aim of the study:

Objective One: To determine the allele and genotype frequencies of the reduced function variant *CYP2D6*17* in African populations:

a) Establish frequencies from publicly available data
b) Screen for *CYP2D6*17* homozygote(s) in black South Africans (to identify healthy donors of dermal fibroblasts for the creation of iPSCs for SC-HLCs model)

Objective Two: To develop the SC-HLC model:

a) From a newly established iPSC line from skin biopsies of an individual who is homozygous for *CYP2D6*17*
b) From a previously established iPSC line (homozygous for *CYP2D6*1 reference allele* – to be confirmed by sequencing)

Objective Three: To measure enzyme activity of CYP2D6 and compare the effects of *CYP2D6*17* to the reference *CYP2D6*1* variant in the conversion of *N*-desmethyltamoxifen to endoxifen using the SC-HLCs.

Due to time constraints on this study, major drug metaboliser CYP3A4 (not CYP2D6) was the enzyme used to test a proof of concept of the model. Rifampicin was used as the major drug metabolising enzymes, CYP3A4 (not CYP2D6) was used to test the viability of the in vitro system in HepG2 cell spheroids. The results from CYP3A4 activity can therefore not be extrapolated to how CYP2D6 would function but just serve to demonstrate functionality of the model system for studying CYP2D6 variants.

1.6 ETHICS

This was a prospective study undertaken in partnership between researchers from two research institutions, the University of the Witwatersrand (Sydney Brenner Institute for Molecular Bioscience (SBIMB) and the Department of Human Genetics, Faculty of Health Sciences) and the Council for Scientific and Industrial Research (CSIR) (Bioengineering and Integrative Genomics Group in the NextGen-Health Cluster).

Ethical clearance for the study was obtained through the Human Research Ethics Committee (HREC) (Medical) at the University of the Witwatersrand for the recruitment of study participants for saliva samples for DNA analysis and skin biopsies for creating iPSC lines (M190221) (Appendix A1). Ethical clearance had previously been obtained through the CSIR Research Ethics Committee (REC) for the "Development of a cellular model of infectious disease investigating drug toxicity" (Ref: 77/2013) (Appendix A2). A request for modification of this CSIR approved protocol was made and was approved for the inclusion of this study under the same approved clearance by the CSIR REC.

The HREC (Medical) and CSIR REC approved the information to participants leaflets and informed consent templates. Witten informed consent was obtained from each individual after they agreed to participated and once they had been informed of the purpose, implications, risks and benefits of this study. Information sheets were given to each individual and a thorough assessment of participant understanding of the information was undertaken (Appendix B1). It was also important for each individual to consent separately to a buccal swab for DNA extraction and analysis, as well as a punch skin biopsy for the development of iPSC lines (Appendix B2-3). Participation was voluntary and participants could withdraw at any time, without being disadvantaged in any way.

CHAPTER 2

*CYP2D6*17* ALLELE FREQUENCIES IN DIFFERENT AFRICAN POPULATIONS & IDENTIFICATION OF *CYP2D6*17* HOMOZYGOTES

2. CHAPTER 2: CYP2D6*17 ALLELE FREQUENCIES IN DIFFERENT AFRICAN POPULATIONS AND IDENTIFICATION OF CYP2D6*17 HOMOZYGOTES

2.1 BRIEF INTRODUCTION

It is important to understand drug response for a particular treatment, in order to assess its safety and efficacy. The metabolism of drugs during treatment is understood through the interactions between the host physiology and the chemical make-up of the drugs (Roses, 2000). The genetic variants that affect responses to drug metabolism are estimated to account for 40-70% of all adverse drug reactions (ADRs) (Zhou, Ingelman-Sundberg and Lauschke, 2017). This has created a need to identify the genetic variants responsible for differential metabolism and to determine their prevalence in African populations (Rajman *et al.*, 2017). It is also evident that approximately 15-30% of drug response variability has underlying contributions from genetic variants, which are predominantly genetic polymorphisms in human drug metabolism genes such as the cytochrome (CYP) P450 superfamily of enzymes (Zhou, Ingelman-Sundberg and Lauschke, 2017). CYP2D6 is an important enzyme for drug metabolism which is accountable for oxidation in over 70 prescribed drugs (Dong *et al.*, 2018). The *CYP2D6*17* allele causes a reduced function effect on the CYP2D6 enzyme and is associated with the intermediate metaboliser (IM) status in individuals who are homozygous for this variant (Wennerholm *et al.*, 2002; Zanger, Raimundo and Eichelbaum, 2004). As explained in Chapter 1, the *CYP2D6*17* variant protein is characterised by a haplotype made up of four single nucleotide variant (SNV) loci (rs28371706 [1022C>**T**], rs1058164 [1661G>**C**], rs16947 [2851C>**T**], rs1135840 [4181G>**C**]). Three of these are missense variants and one (rs1058164) is a silent mutation (Masimirembwa *et al.*, 1996; Oscarson *et al.*, 1997).

Table 2.1: Genetic arrangement of *CYP2D6*2* & *CYP2D6*17* allele variants.

Allele Variants	rs28371706	rs1058164	rs16947	rs1135840
CYP2D6 Reference	C	G	C	G
*CYP2D6*2*	C	C	T	C
*CYP2D6*17*	T	C	T	C

When studying and investigating drug metabolism and pharmacogenetics, it is critical to understand the genetic makeup of the target population/s. The impact of the genetic variations in genes involved drug metabolism and their population frequencies have a large effect on drug efficacy and effectiveness at a population level. This makes it important to establish

better tools for investigating and understanding these differences. Bioinformatic analysis was performed on genomic data from African populations in publicly available datasets from the 1000 Genomes Project (KGP), PharmVar and the African Genome Variation Project (AGVP) (which contained South African data on Zulu and Sotho populations). These data are stored at the SBIMB and were analysed to understand population genetic aspects of *CYP2D6* reduced function allele polymorphisms, and more specifically of *CYP2D6*17*.

To identify individuals who are *CYP2D6*17* homozygotes and could potentially donate dermal fibroblasts to develop iPSC cell lines, medical students from the University of the Witwatersrand were recruited to provide buccal swab samples in order to extract DNA and to genotype for *CYP2D6*17*. The SNP rs28371706, was genotyped as a proxy for the *CYP2D6*17* haplotype and was analysed using a TaqMan assay. The intention was to perform DNA sequencing on a limited number of samples to confirm the TaqMan genotypes, and this would be performed by a commercial laboratory.

Laboratory experiments were conducted at the National Health Laboratory Service (NHLS) in Braamfontein (DNA extraction) and CSIR in Pretoria (TaqMan assay), and the bioinformatic analyses were performed at the SBIMB at the University of the Witwatersrand.

This chapter outlines the approaches used in this study. Firstly to mine public genomic data to determine allele, genotype and haplotype frequencies for reduced function *CYP2D6* variants (including *CYP2D6*17*) in African populations, and secondly to identify individuals who are homozygous for *CYP2D6*17* and who could be healthy donors of dermal fibroblasts for the creation of iPSCs for SC-HLCs model.

2.2 METHODS

2.2.1 Prevalence of *CYP2D6* Reduced Function Variants in African Populations

This section describes the bioinformatic analysis of the public data on the *CYP2D6* gene, investigating the haplotype and star allele frequencies linked to reduced CYP2D6 enzyme function in the African populations from the KGP, PharmVar and AGVP datasets (Fig 2.1).

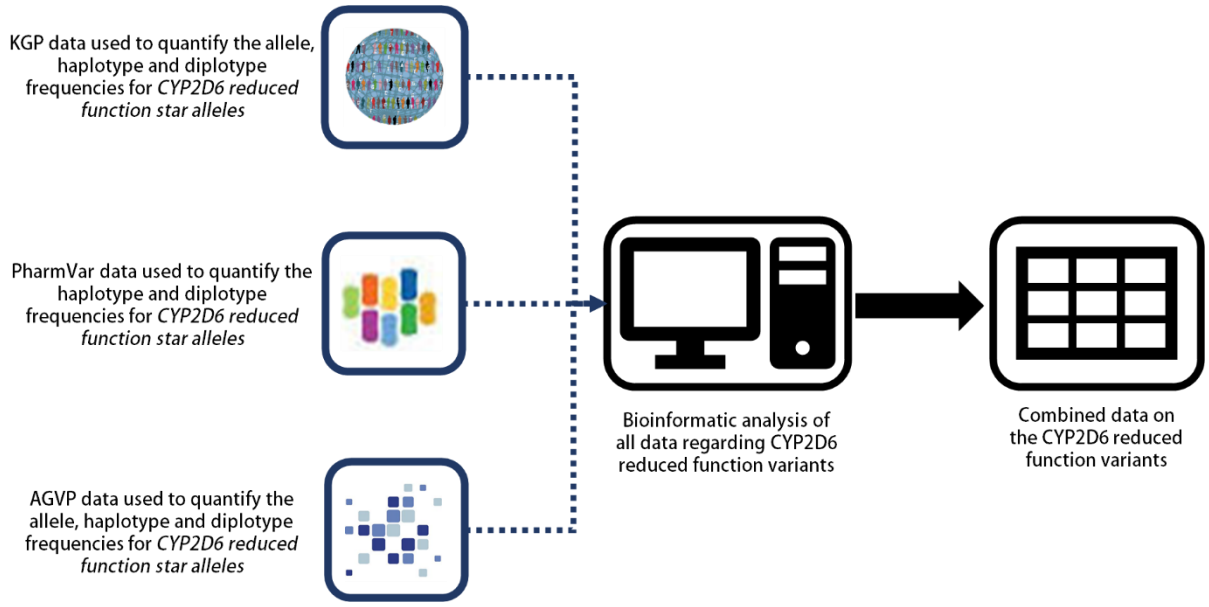


Fig 2.1 Workflow for the bioinformatic analysis in the study. The publicly available data came from three sources KGP, PharmVar and AGVP.

2.2.1.1 *1000 Genomes Project Data*

1000 Genomes Project (KGP) data is publicly available for download and can be accessed via the Ensembl Genome Browser. It was used to obtain information on the *CYP2D6* reduced function star alleles. KGP was established in 2008 and completed in 2015. It included three project phases and its aim was to provide a deep characterisation of the sequence variation seen in the human genome across different populations. This was done to create a resource that could be used to further investigate the relationships and interaction between genotype and phenotype (Altshuler *et al.*, 2010). The genomes of 2 504 individuals from 26 different population groups were investigated using low-coverage whole genome sequencing, deep exome sequencing and dense microarray genotyping. More than 88 million variants were phased onto high-quality haplotypes for further study and investigation (Auton *et al.*, 2015).

Ensembl is a genome browser website and database that creates a bioinformatic platform for use in genomic science. It is made up of numerous high-quality genome annotations, superimposed on the reference genome, and contains datasets of gene annotations, multiple alignments, gene homology and regulatory annotations of genomes (Flicek *et al.*, 2011). The

Haplosaurus tool in the Ensembl website was used to generate haplotypes for each of the *CYP2D6* star alleles of interest. Haplosaurus is a bioinformatic open-source software that is integrated into the Ensembl codebase and is used to compute haplotypes. It is made up of three main components which are the Ensembl website display of transcript haplotypes, Haplosaurus command line tool, and extensions to the Ensembl Perl application programming interface (API). For this study, the transcript display component was used to retrieve data on the *CYP2D6* reduced function alleles (Spooner *et al.*, 2018).

The KGP data retrieved for this study, pertaining to the *CYP2D6* reduced function alleles, were acquired only from the African-ancestry population data available. The African-ancestry data included five populations from Africa (ESN – Esan in Nigeria; GWD – Gambian in Western Division, Gambia; LWK – Luhya in Webuye, Kenya; MSL – Mende in Sierra Leone; YRI – Yoruba in Ibadan, Nigeria). The SNP allele frequencies and haplotype frequencies for each *CYP2D6* reduced function allele were retrieved and recorded.

2.2.1.2 *PharmVar Data*

PharmVar Data, which was publicly available for download, was used to further analyse the *CYP2D6* reduced function alleles. The Pharmacogene Variation Consortium (PharmVar) was established in September 2017 to create an innovative and interactive website for access to the data for the Human Cytochrome P450 (CYP) Nomenclature and the CYP genetic variants. PharmVar has also established a detailed database and expert panels made up of individuals with expertise on specific genes for the examining of pharmacogene interactions, providing information on genetic variants, and creating international standard operating procedures for investigating and reporting pharmacogenetic findings.

Data from the PharmVar database pertaining to frequency data for the haplotypes and diplotypes for the reduced function *CYP2D6* alleles was downloaded (Gaedigk, Sangkuhl, *et al.*, 2019). The population groups isolated for this study were African (2054 individuals), with the African population group being made up of individuals from Ethiopia (313 individuals), Ghana (193 individuals), South Africa (379 individuals), Tanzania (601 Individuals), Zimbabwe (191 individuals) and a mixture of other African countries (377 individuals). The star alleles *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* were chosen for haplotype and diplotype analyses. The allele/haplotype frequencies of each of the population groups were compared to assess the distribution of these genetic variants.

2.2.1.3 *African Genome Variation Project Data*

Data was extracted from the African Genome Variation Project (AGVP) Baganda (Uganda), Ethiopian and Zulu (South Africa) cohorts. The AGVP data used in this study was obtained from the WGS data acquired from each cohort. The cohorts consisted of 120 Ethiopian, 100 Baganda, and 100 Zulu individuals and was phased before analysis. The AGVP is a multinational collaboration established to assess the genetic diversity found within 18 ethno-linguistic groups from sub-Saharan Africa. A total of 1 801 individuals are part of the study providing the project with HumanOmni2.5M genotyping array (1 481 individuals) and whole genome sequences (WGS) (320 individuals) from various institutions (Gurdasani *et al.*, 2015).

The phased AGVP data was analysed using Plink (v1.904) (Purcell *et al.*, 2007) to obtain the minor allele frequencies for each of the SNPs that make up the *CYP2D6*17* haplotype in each individual population group and for the combined dataset. The phased AGVP data was then recoded into VCF files and organised by the sample ID. This was done to prepare the files for haplotype and diplotype analysis using the tool Astrolabe (v0.8.6.1) (Twist *et al.*, 2016) which needs VCF files as the input file format. Quality control (QC) before the analysis of the data using Astrolabe was not performed as there were no BAM files available for the WGS AGVP data which is needed as a QC reference. Astrolabe was used to make calls of each haplotype and diplotype and gave a text output file. Python was used to extract the Astrolabe data from the text file and organised them according to each of the sample IDs outputting them to a TSV file (Appendix C1).

The allele frequencies in the *CYP2D6* gene and the haplotype frequencies, using Haplosaurus, were compared to the allele and haplotype frequencies from data in PharmVar and in the AGVP data. The star alleles *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* were chosen for haplotype and diplotype analyses and comparisons if they had a haplotype frequency > 1% in all datasets and were reduced function alleles. The allele, genotype and haplotype data were summarized in tables.

2.2.2 Identifying *CYP2D6*17* Homozygote Individuals to Establish Cell Lines

This section describes the study participant recruitment, DNA extraction and genotyping to screen for *CYP2D6*17* homozygotes. In addition, dermal fibroblasts from two cell lines stored at the CSIR (BlkAfrCon and P7OxCon) also underwent DNA extraction and were genotyped to screen for their *CYP2D6*17* genotypes. Sequencing was performed to confirm the genotypes of the study participant DNA samples and the two CSIR dermal fibroblast cell

lines. Confirmed heterozygotes and homozygotes of *CYP2D6*17* from among the study participants were asked to donate dermal fibroblasts to establish iPSC lines (Fig 2.2).

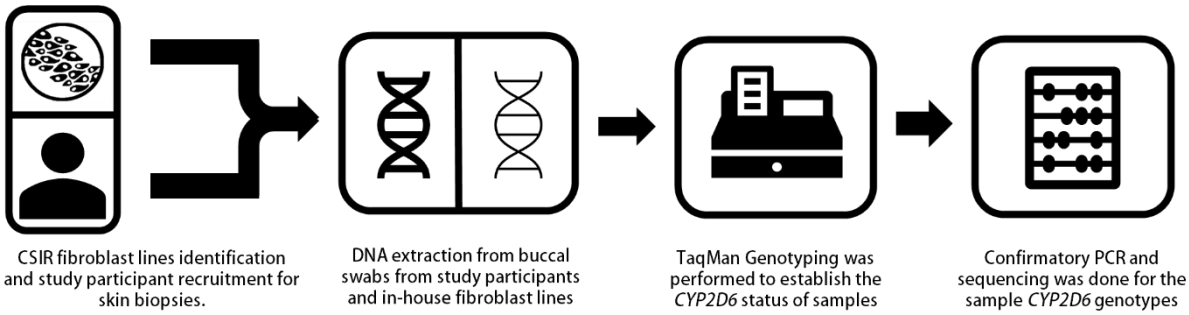


Fig 2.2: Workflow from sample collection (cell lines and study participants) to DNA extraction and genotyping, with confirmatory DNA sequencing, to identify cell lines and/or participants suitable for downstream construction of iPSC cell lines.

2.2.2.1 *Study Participants*

Study participants for the population *CYP2D*17* homozygote screening and skin biopsy were recruited from the 2nd and 3rd year medicine classes in the Faculty of Health Sciences at the University of the Witwatersrand. The classes were shown an informative recruitment video (*see link to view recruitment video* https://drive.google.com/file/d/1wZcZVXc-Ms2E1HPLssWdiI-J8Aiib5M7/view?usp=sharing) to present them with information on the project and ask for contact details to reach them (Appendix B4). The target was to recruit 50 black South African individuals. A power calculation was performed using allele frequencies based on the Zimbabwean allele frequency of 0.34 and assuming Hardy-Weinberg equilibrium (HWE) (Fig 2.3) to assess the likelihood of finding individuals homozygous for *CYP2D6*17* (Masimirembwa *et al.*, 1996). The calculation showed that approximately 11% (q^2 frequency = 0.11) of people would be *CYP2D6*17* homozygotes, hence we expected to find about five *CYP2D6*17* homozygous individuals for dermal fibroblast donation among the group of 50. Dermal fibroblasts cell lines, from the CSIR laboratory, also underwent DNA extraction to determine the genotypes for each line. There were two cell lines derived for individuals with African ancestry, BlkAfrCon and P7OxCon. Both cell lines were approved for use in this study by the CSIR REC.

p = C allele = 0.66
q = T allele = 0.34

$(p+q)^2 = p^2 + 2pq + q^2$
$= (0.66)^2 + 2(0.66)(0.34) + (0.34)^2$
$= 0.44 + 0.45 + 0.11$

$p^2 = 0.44$ *(~22 Individuals)*
$q^2 = 0.11$ *(~5 Individuals)*
$2pq = 0.45$ *(~23 Individuals)*

Fig. 2.3: Calculations for expected *CYP2D6*17* allele frequencies in the Zimbabwean Shona population. The major allele (p) and minor allele (q) frequencies are shown, as well as the frequencies for the homozygous wildtype genotype (p^2), the heterozygotes (2pq) individuals homozygous for the variant allele (q^2), under conditions of HWE.

Following an information session and informed consent, buccal swabs were collected for DNA extraction to determine each participant's *CYP2D6* status. A questionnaire was completed by each study participant during a brief interview to get more information and insight into the ethnicity of each individual (Appendix B5). Study participants who were heterozygous or homozygous for *CYP2D6*17* were then approached for a skin biopsy for the purpose of creating iPSC lines. A second consent form was completed before the skin biopsy was taken. Each study participant recruited was given a participant code that was used as an anonymous identifier for each sample in the study.

2.2.2.2 *Participant Sampling: Buccal Swab & DNA Extraction*

Buccal swabs were taken at the Department of Human Genetics (located at the NHLS) Braamfontein Offices) from study participants of black South African ancestry using the Qiagen Gentra Puregene® Buccal Cell Kit (Hilden, Germany). Two buccal swabs were taken per study participant by scraping the inner sides of the right and left cheeks 10 times per swab with a sterile nylon-bristle brush. Each brush was placed in 1.5 ml microcentrifuge tubes containing 600 µl Gentra Puregene® Cell Lysis Solution (Hilden, Germany) which contained an anionic detergent and DNA stabiliser.

The microcentrifuge tubes containing the brushes with the buccal cells were incubated at 65°C for 15 minutes and 3 µl of Gentra Puregene® Proteinase K (Hilden, Germany) at a concentration of 20 mg/ml was added. The tubes were further incubated overnight at 55°C and 3 µl of Gentra Puregene® RNase A Solution (Hilden, Germany) was added the next day to each tube. The tubes were mixed by inverting 25 times, and subsequently incubated for 30 minutes at 37°C. The tubes were vortexed for 20 seconds, cooled for 1 minute on ice and 20 µl of Gentra Puregene® Protein Precipitation Solution (Hilden, Germany) was added to each tube. The tubes were incubated on ice for 5 minutes and centrifuged for 3 minutes at 14 000 x g until the proteins precipitated into a white pellet. The supernatant from each of the tubes was poured into new 1.5 ml microcentrifuge tubes containing 600 µl of isopropanol and 1 µl of glycogen solution at a concentration of 20 mg/ml. The new tubes were mixed by inversion 50 times at room temperature and were centrifuged at 14 000 x g for 5 minutes until the DNA precipitated into a white pellet. The supernatant was removed, and 600 µl of 70% ethanol was added to the pellet as a wash step, followed by several inversions resuspending the pellet in the ethanol solution. The tubes were then centrifuged at 14 000 x g for 1 minute and the supernatant was discarded and the tubes air dried for 15 minutes. The DNA pellets were resuspended in 50 µl of Gentra Puregene® Hydration Solution (Hilden, Germany) and incubated for 1 hour at 65°C. All the tubes with newly extracted DNA were incubated

overnight at room temperature with gentle shaking. DNA concentration and quality were measured using the Implen NanoPhotometer® NP80 (Westlake Village, USA) at the NHLS Braamfontein Offices. To confirm the DNA concentration, the samples were also tested on the Thermo Fisher NanoDrop™ ND-1000 Spectrophotometer (Waltham, MA) at the CSIR.

2.2.2.3 *Primer Design & Synthesis*

The confirmatory PCR primers (Table 2.2) for the *CYP2D6*17* proxy SNP rs28371706 were designed using the Primer3 (v 0.4.0) primer generating software (http://bioinfo.ut.ee/primer3-0.4.0/), and were purchased from Inqaba Biotechnologies (Pretoria, South Africa) in 100 µM stocks. Primer sequences were synthesised using the Homo sapiens *CYP2D6* RefSeqGene sequence (Accession Number: NG_008376) with the sequences flanking SNP rs29371706 and NCBI Blast was used to evaluate the specificity (http:/www.ncbi.nih.gov/tools/primer-blast/). The primers were designed to have a PCR product with a size of 172 bp and the melting temperatures (T_m) of 58.3°C (Forward) and 62.7°C (Reverse) for annealing. The sequences of the TaqMan primers were unknown due to proprietary rights restrictions, however, the sequence of the probe that distinguishes the alleles was provided (Table 2.2). All the primers were reconstituted in dH_2O from the Merck Millipore Milli-Q® Direct Water Purification System (Darmstadt, Germany) to 10µM working stocks for PCR use. All the primers were kept cold at -20°C freezer. Table 2.1 summarises the relevant information for each primer.

Table 2.2: SNP rs28371706 primer and probe information

Primer/Probe	Sequence (5' – 3')	T_m (°C)	PCR Product Size (bp)
*CYP2D6*17* Forward	CGGTGGTCGTGCTCAATG	58.3	172
*CYP2D6*17* Reverse	GTCCCACGGAAATCTGTCTCTG	62.7	
TaqMan Genotyping VIC/FAM Probe	ACGCGGCCCGAAACCCAGGATCTGG**[G/A]**→ TGATGGGCACAGGCGGGCGGTCGGC	60.0 (Manual)	N/A

The red highlighted bases in the probe shows the base of interest that are associated with the rs28371706 SNP C/T alleles and would bind to them accordingly: G-C and A-T. The → represents the continuation of the sequence on the following line.

2.2.2.4 *TaqMan Genotyping: Assay Design*

A TaqMan assay was designed and performed to determine the *CYP2D6*17* genotypes by using rs28371706 as a proxy using the Thermo Fisher Scientific TaqMan® Custom SNP Genotyping Assay Mix (Waltham, MA) and Thermo Fisher Scientific TaqPath™ ProAmp™ Master Mix (Waltham, MA). The SNP genotyping assay mix contained the primers and TaqMan™ MGB probes for targeting the area of interest which was the rs28371706 T allele and primers for the reference C allele (Fig 2.4). This consisted of two probes with 5' end reporter dyes, VIC® dye for the C allele and FAM™ dye for the T allele, which were

complementary to the sequence surrounding rs28371706 (Table 2.1). A minor groove binder (MGB) was also present within the assay mix bound to the 5'end reporter dye and this functioned to increase the melting temperature (T_m) to allow for shorter probes and better sensitivity. A non-fluorescent quencher (NFQ) was also in the assay mix and bound to the 3'end of the probe. The commercial master mix consisted of Dual-Lock™ Taq DNA Polymerase, heat-labile uracil-DNA glycosylase (UNG), dNTPs with dUTP and ROX™ reference dye. The Dual-Lock™ Taq DNA Polymerase utilised a combination of two hot-start mechanisms for activation control which prevent polymerase activity at lower temperature ranges therefore decreasing non-specific binding and amplification. The heat-labile UNG functioned by inhibiting the reamplification of "carryover" PCR products in the samples by degrading dU-PCR products as well as misprimed and non-specific DNA. The dNTPs were used for the synthesis of the polynucleotides and dUTP was used to aid UNG activity in the reaction. The ROX™ Passive reference dye allowed for the normalisation, as an internal reference, of the reporter dye signal in the probes during data analysis. This is critical as fluctuations in fluorescence may occur due to changes in volume or concentration of samples.

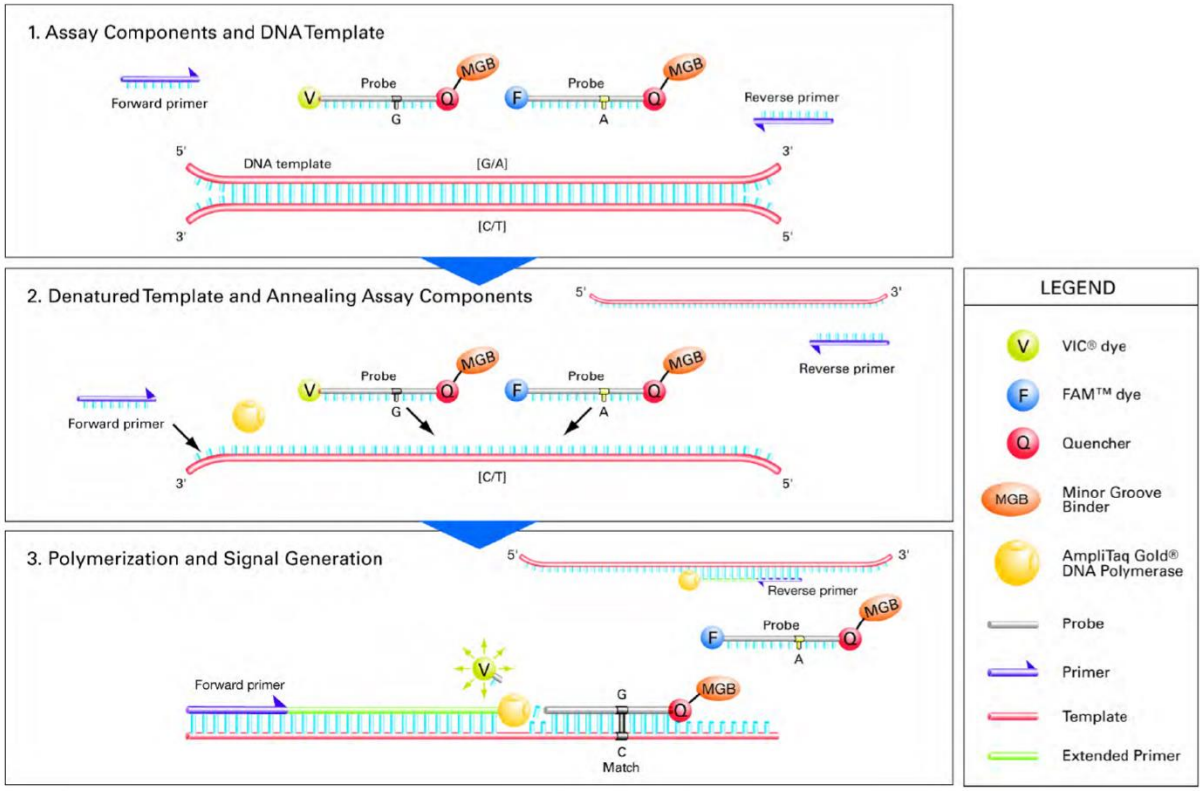


Fig 2.4: Polymerisation and detection process of the TaqMan Genotyping Assay for the SNP rs28371796, used as a proxy for the *CYP2D6*17* haplotype. *Image from: Applied Biosystems® TaqMan® SNP Genotyping Assays User Guide (2014).*

Reaction components for the TaqMan PCR for rs28371706 genotyping were 12.5 µl of 2x TaqPath™ ProAmp™ Master Mix (Thermo Fisher Scientific, MA), 1.25 µl of 20x TaqMan® Custom SNP Genotyping Assay Mix (Thermo Fisher Scientific, MA), 100ng/µl Genomic

DNA, and dH_2O (Table 2.3). The PCR was run on a Bio-Rad CFX96™ Real-Time System C1000™ Thermal Cycler (Hercules, USA) with the following steps: pre-read at 60°C for 30 sec; initial denaturation and enzyme activation at 95°C for 5 minutes; the cycling steps [denaturation at 95°C for 15 seconds and annealing/extension at 60°C for 60 seconds] for 40 cycles; and a post-read step at 60°C for 30 seconds (Table 2.4).

Two genotyping methods, using reagents from different suppliers, were also tested as alternatives, however, they were unable to bring forth results, but are none the less documented below. The TaqMan® Custom SNP Genotyping Assay Mix (Thermo Fisher Scientific, MA) was used together with Kapa Biosystems KAPA™ Taq ReadyMix PCR Kit (Wilmington, USA) (*Alternative Method 1*) and New England Biolabs Q5® High-Fidelity DNA Polymerase (Ipswich, USA) (*Alternative Method* 2).

Alternative Method 1 – For the KAPA™ Taq ReadyMix PCR Kit, the components for genotyping were 12.5 µl of 2X KAPA™ Taq ReadyMix (Wilmington, USA), 1.25 µl of 20X TaqMan® Custom SNP Genotyping Assay Mix (Thermo Fisher Scientific, MA), 1 µl of 100ng/µl genomic DNA, and 10.25 µl of dH_2O (Table 2.3). The PCR was run on a Bio-Rad CFX96™ Real-Time System C1000™ Thermal Cycler (Hercules, USA) in 200 µl PCR tubes, and the PCR steps and conditions were as follows: an initial denaturation at 95°C for 3 minutes, the cycling steps of denaturation at 95°C for 30 seconds, annealing at 60°C for 30 seconds and extension at 72°C for 1 minute for 35 cycles, and a final extension at 72°C for 1 minute (Table 2.4).

Alternative Method 2 – For the Q5® High-Fidelity DNA Polymerase (Ipswich, USA) reaction, the components for the genotyping were 5 µl of 5X Q5® Reaction Buffer (Ipswich, USA), 1.25 µl of each 10 µM forward and reverse primers, 0.25 µl of Q5® High-Fidelity DNA Polymerase (Ipswich, USA), 0.5 µl of 10 mM of dNTPs, 1 µl of 100ng/µl genomic DNA, and 15.75 µl of dH_2O (Table 2.3). Each PCR reaction was run on a Bio-Rad CFX96™ Real-Time System C1000™ Thermal Cycler (Hercules, USA) in 200 µl PCR tubes, and the PCR steps and conditions were as follows: initial denaturation at 98°C for 30 seconds, the cycling steps of denaturation at 98°C for 10 seconds, annealing at 60°C for 30 seconds and extension at 72°C for 30 seconds for 35 cycles, and a final extension at 72°C for 2 minutes (Table 2.4).

2.2.2.5 *Confirmatory PCR & Sequencing*

The TaqMan genotyping findings were confirmed by sequencing to assess the accuracy of the assay for the SNP rs28371706 (proxy for *CYP2D6*17*), using the SNP PCR products. The

Thermo Fisher Scientific DreamTaq™ Polymerase PCR Kit (Waltham, MA) was used to amplify the PCR products for sequencing. The components of the PCR were 0.2 µl of DreamTaq™ Polymerase (Thermo Fisher Scientific, MA), 2.5 µl of 10X DreamTaq™ Polymerase Buffer (Thermo Fisher Scientific, MA), 0.5 µl of 10 mM dNTPs, 0.5 µl of rs28371706 forward and reverse primers (Table 2.2), 1 µl of 100ng/µl genomic DNA, and 19.8 µl of dH_2O (Table 2.3). Each PCR reaction volume was 25 µl and put into a 200 µl PCR tube. The PCR procedure was run on a Bio-Rad T100™ Thermal Cycler (Hercules, USA) and the PCR steps and conditions were the following: initial denaturation at 95°C for 3 minutes, the cycling steps of denaturation at 95°C for 30 seconds, annealing at 62°C for 30 seconds and extension at 72°C for 1 minute for 35 cycles, and a final extension at 72°C for 5 minutes (Table 2.4). Approximately 5 µl of PCR products were then loaded on a 2% agarose gel and run at 100V for an hour in a Bio-Rad PowerPac™ Universal electrophoresis machine (Hercules, USA) to confirm a successful PCR without any contamination, and visualised on the Syngene G:BOX™ Chemi 16 Bio Imaging System (Cambridge, UK). The PCR products were extracted for sequencing using the Thermo Fisher Scientific GeneJET™ PCR Purification Kit (Waltham, MA). The samples were sent to Inqaba Biotechnologies (Pretoria, South Africa) for sequencing to confirm the genotype.

Table 2.3: Reagents for rs28371706 genotyping and volume for each PCR reaction.

Dream Taq (per 25 µl)	
Reagent	Volume (µl)
*CYP2D6*17* Forward Primer (10 µM)	0.5
*CYP2D6*17* Reverse Primer (10 mM)	0.5
Dream Taq Polymerase	0.2
Dream Taq Buffer (10x)	2.5
dNTPs (10 µM)	0.5
Genomic DNA (100 ng/µl)	1.0
dH_2O	19.8
TOTAL	25.0

KAPA Taq (per 25 µl)	
Reagent	Volume (µl)
TaqMan Genotyping Mix	1.25
KAPA Taq ReadyMix	12.5
Genomic DNA (100 ng/µl)	1.0
dH_2O	10.0
TOTAL	25

Q5 (per 25 µl)	
Reagent	Volume (µl)
TaqMan Genotyping Mix	1.25
Q5 Polymerase	0.25
Q5 Buffer (20x)	5.0
dNTPs (10 mM)	0.5
Genomic DNA (100 ng/µl)	1.0
dH_2O	17.0
TOTAL	25.0

TaqMan Genotyping (per 25 µl)	
Reagent	Volume (µl)
TaqPath ProAmp Master Mix	12.5
TaqMan Genotyping Mix	1.25
Genomic DNA (100 ng/µl)	1.0
dH_2O	1.25
TOTAL	25.0

Table 2.4: PCR cycling conditions for each PCR assay.

Dream Taq PCR Conditions (35 Cycles)				
Initial Denaturation	Denaturation	Annealing	Extension	Final Extension
95°C | 3 min	95°C | 30 sec	62°C | 30 sec	72°C | 1 min	72°C | 5 min
KAPA Taq PCR Conditions (35 Cycles)				
Initial Denaturation	Denaturation	Annealing	Extension	Final Extension
95°C | 3 min	95°C | 30 sec	60°C | 30 sec	72°C | 1 min	72°C | 1 min
Q5 PCR Conditions (35 cycles)				
Initial Denaturation	Denaturation	Annealing	Extension	Final Extension
98°C | 30 sec	98°C | 10 sec	60°C | 30 sec	72°C | 30 sec	72°C | 2 min
TaqMan Genotyping PCR Conditions (40 cycles)				
Pre- Read	Initial Denature	Denaturation	Extension	Post-Read
60°C | 30 sec	95°C | 5 min	95°C | 15 sec	60°C | 1 min	60°C | 30 sec

2.3 RESULTS

2.3.1 Prevalence of *CYP2D6* Reduced Function Variants in African Populations

Allele, haplotype and diplotype data from the *CYP2D6* reduced function alleles was extracted from KGP (using the Haplosaurus tool on Ensembl), PharmVar and the AGVP (using plink and Astrolabe tools). Please note that the "star allele" nomenclature usually refers to haplotypes, but these are often characterised by a single SNP that acts as a proxy for the specific haplotype. Therefore, this section refers to haplotypes that define the star alleles as alleles, with the understanding that in the strict genetic sense, they are not alleles.

Four reduced function alleles (*CYP2D6*10, CYP2D6*17, CYP2D6*29, CYP2D6*41*) were chosen for extraction and analysis of haplotype and diplotype data, due to their presence in all dataset predictions and higher than 1% frequency in all the AGVP, KGP and PharmVar datasets. All data was summarized in tables and compared across all African populations of interest to the study.

2.3.1.1 *CYP2D6*17 Allele Frequency Data Extraction*

From the KGP data, the allele frequencies for each of the four SNPs (rs16947, rs1058164, rs1135840 and rs28371706) that make up the *CYP2D6*17* haplotype (allele) were extracted using Ensembl Genome Browser, extracting data specifically from the African populations. The minor allele frequencies (MAFs) in the combined KGP African data are shown in Table 2.5. For the individual African populations from the KGP data, rs16047 T allele frequencies ranged from 35.4% (LWK) to 46.5% (GWD). The rs1058164 C allele frequencies ranged from 23.5% (MSL) to 35.0% (GWD). The rs1135840 C allele frequencies ranged from 23.5% (MSL) to 35.4% (GWD). The rs28371706 T allele frequencies ranged from 16.4% (GWD) to 28.2% (MSL) (Table 2.5).

From the AGVP data, the allele frequencies for each variant of *CYP2D6*17* was extracted from the AGVP WGS data from 320 African participants. When looking at the four SNPs (rs16947, rs1058164, rs1135840 and rs28371706) that make up the *CYP2D6*17* allele, the combined minor allele frequencies of each SNP was 38.4%, 10.6%, 30.6% and 20.2% respectively. For the individual African populations from the AGVP data, rs16047 T and rs1135840 C alleles frequencies were highest in the Zulu population, 47.0% and 36.0% respectively. The rs1058164 C and rs28371706 T allele frequencies were highest in the Bagandan population 12.5% and 23.5% respectively. With the CYP2D6*17 definitive SNP rs28371706 T, there was a combined allele frequency for all AGVP populations of 20.2% (Table 2.6).

Table 2.5: Allele frequencies of the SNPs that make up the *CYP2D6*17* haplotype, extracted from the KGP African population data.

KGP (1000 Genome Project) Allele Frequencies							
SNP Name	Minor Allele	Population % Combined (MAF)	ESN (%)	GWD (%)	LWK (%)	MSL (%)	YRI (%)
rs16947	T	44.6	43.4	46.5	35.4	38.8	44.0
rs1058164	C	32.8	34.8	35.0	24.7	23.5	33.3
rs1135840	C	32.4	33.8	35.4	25.3	23.5	32.4
rs28371706	T	21.8	26.3	16.4	19.2	28.2	25.5

Table 2.6: Allele frequencies of the SNPs that make up the *CYP2D6*17* haplotype, extracted from the AGVP African populations.

AGVP (African Genome Variation Project) SNP Frequencies					
SNP Name	Minor Allele	Population % Combined (MAF)	Zulu (%)	Ethiopian (%)	Bagandan (%)
rs16947	T	38.4	47.0	40.8	27.0
rs1058164	C	10.6	9.5	10.0	12.5
rs1135840	C	30.6	36.0	33.8	21.5
rs28371706	T	20.2	17.5	19.6	23.5

2.3.1.2 *CYP2D6*17 Haplotype Frequency Data Extraction*

For the KGP data, haplotypes for the star alleles *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* were generated using the Haplosaurus tool in Ensembl. These four star alleles (that are actually haplotypes) were chosen for analyses and comparisons due to having a haplotype frequency > 1% in all datasets, as well as being reduced function alleles. The overall African average haplotype frequency of *CYP2D6*17* was 20.7%, while *CYP2D6*10*, *CYP2D6*29* and *CYP2D6*41* were 8.93%, 8.85% and 18.2% respectively. The African population group frequencies for *CYP2D6*10* ranged from 2.02% (LWK) to 14.70% (MSL). For *CYP2D6*17*, frequencies ranged from 15.50% (GWD) to 25.90% (MSL). For *CYP2D6*29*, frequencies ranged from 7.07% (ESN) to 11.1% (GWD). For *CYP2D6*41*, frequencies ranged from 16.20% (YRI) to 26.8% (LWK) (Table 2.7).

For the PharmVar data, haplotypes for the star alleles *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* were extracted from the PharmVar database. The combined African haplotype frequency of *CYP2D6*17* was 19.95%, while *CYP2D6*10*, *CYP2D6*29* and *CYP2D6*41* were 6.61%, 8.95% and 9.71% respectively. For *CYP2D6*10*, the haplotype frequencies ranged from 3.10% (Ghanaian) to 9.11% (South African), with no data available for the Zimbabwean population. For *CYP2D6*17*, the haplotype frequencies ranged from 11.04% (Ethiopian) to 34.00% (Zimbabwean). For *CYP2D6*29*, the haplotype frequencies ranged from 6.14% (South African) to 19.80% (Tanzanian), with no data available for the Ethiopian, Ghanaian and Zimbabwean populations. For *CYP2D6*41*, the haplotype

frequencies ranged from 2.80% (Other African) to 22.94% (Ethiopian), with no data available for the Ghanaian, Tanzanian and Zimbabwean populations (Table 2.8).

Table 2.7: Haplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* alleles from the KGP African populations.

KGP (1000 Genome Project) Haplotype Frequencies						
Haplotype	AFR (%)	YRI (%)	ESN (%)	GWD (%)	MSL (%)	LWK (%)
*CYP2D6*10*	8.93	10.20	8.59	9.29	14.70	2.02
CYP2D6*17	**20.70**	**25.00**	**23.70**	**15.50**	**25.90**	**18.70**
*CYP2D6*29*	8.85	9.72	7.07	11.10	8.82	10.60
*CYP2D6*41*	18.20	16.20	16.70	17.70	18.80	26.80

*The bold highlighted row signifies the CYP2D6*17 data which was pertinent to the study and the establishment of an iPSC cellular model.*

Table 2.8: Haplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* alleles from the PharmVar populations.

PharmVar Haplotype Frequencies							
Haplotype	Combined African (%)	South African (%)	Ghanaian (%)	Tanzanian (%)	Ethiopian (%)	Zimbabwean (%)	Other African (%)
*CYP2D6*10*	6.61	9.11	3.10	3.80	8.47	-	5.23
CYP2D6*17	**19.95**	**17.30**	**27.70**	**13.99**	**11.04**	**34.00**	**17.84**
*CYP2D6*29*	8.95	6.14	-	19.80	-	-	6.70
*CYP2D6*41*	9.71	2.84	-	-	22.94	-	2.80

*The bold highlighted row signifies the CYP2D6*17 data which was pertinent to the study and the establishment of an iPSC cellular model. (-) represents no data being available for the particular population.*

Table 2.9: Haplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* alleles from the AGVP populations.

AGVP (African Genome Variation Project) Haplotype Frequencies				
Haplotype	Combined Data (%)	Zulu (%)	Ethiopian (%)	Bagandan (%)
*CYP2D6*10*	2.81	5.50	2.50	0.50
CYP2D6*17	**20.00**	**17.50**	**19.58**	**23.00**
*CYP2D6*29*	11.25	11.50	2.08	22.00
*CYP2D6*41*	5.78	2.50	7.92	6.50

*The bold highlighted row signifies the CYP2D6*17 data which was pertinent to the study and the establishment of an iPSC cellular model.*

For the AGVP data, haplotypes for the star alleles *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* were generated by using Astrolabe (v0.8.6.1). The combined African haplotype frequency for CYP2D6*17 was 20.0%, while CYP2D6*10, CYP2D6*29 and CYP2D6*41 were 2.81%, 11.25% and 5.78% respectively. The Zulu group had the highest haplotype frequency for *CYP2D6*10* of 5.50%. The Ethiopian group had the highest *CYP2D6*41* haplotype frequency of 7.92%. The Bagandan group had haplotype frequencies for *CYP2D6*17* and *CYP2D6*29* of 23.0% and 22.0% respectively (Table 2.9).

2.3.1.3 *CYP2D6*17 Homozygous Diplotype Frequency Data Extraction*

For the KGP data, homozygous diplotypes for the star alleles *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* were acquired by calculating for q^2 from the haplotype data generated by the Ensembl Haplosaurus tool. For each allele, an expected q^2 for homozygous diplotype frequency was estimated assuming Hardy-Weinberg equilibrium, as there was no diplotype data available from the Haplosaurus data. The overall African average homozygous diplotype frequency of *CYP2D6*17* was 4.28%, while *CYP2D6*10*, *CYP2D6*29* and *CYP2D6*41* were 0.80%, 0.78% and 3.31% respectively. For the individual African population groups, frequencies for *CYP2D6*17* ranged from 2.40% (GWD) to 6.71% (MSL). For *CYP2D6*10*, frequencies ranged from 0.04% (LWK) to 2.16% (MSL). For *CYP2D6*29*, frequencies ranged from 0.50% (ESN) to 1.23% (GWD). For *CYP2D6*41*, frequencies ranged from 2.62% (YRI) to 7.18% (LWK) (Table 2.10).

For the PharmVar data, homozygous diplotypes for the star alleles *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* were obtained from the PharmVar database. The African diplotype frequency of *CYP2D6*17* was 3.98%, while *CYP2D6*10*, *CYP2D6*29* and *CYP2D6*41* were 0.44%, 0.80% and 0.94% respectively. For the individual African populations, an expected q^2 for homozygous diplotype frequency was estimated assuming Hardy-Weinberg equilibrium, as there was no diplotype data available from PharmVar. For *CYP2D6*10*, the diplotype frequencies range from 0.09% (Ghanaian) to 0.83% (South African), with no Zimbabwean data available. For *CYP2D6*17*, the diplotype frequencies ranged from 1.22% (Ethiopian) to 11.56% (Zimbabwean). For *CYP2D6*29*, the diplotype frequencies ranged from 0.38% (South African) to 3.92% (Tanzanian), with no Ethiopian, Ghanaian or Zimbabwean data available. For *CYP2D6*41*, the diplotype frequencies were 0.08% (South African & Other African) and 5.26% (Ethiopian), with no Ghanaian, Tanzanian or Zimbabwean data available (Table 2.11).

For the AGVP data, homozygous diplotypes for the star alleles *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* were calculated using the data generated by Astrolabe (v0.8.6.1). An expected (exp) q^2 for homozygous diplotype frequency was also estimated assuming Hardy-Weinberg equilibrium. The combined African average diplotype frequencies for *CYP2D6*17* were 7.81% (observed) and 4.00% (expected), while *CYP2D6*10*, *CYP2D6*29* and *CYP2D6*41* were 0.63% (0.08% exp), 3.44% (1.27% exp) and 0.31% (0.33 exp) respectively. The Zulu cohort group had the highest *CYP2D6*10* diplotype frequencies of 2.00% (0.30% exp). The Ethiopian cohort group had the highest *CYP2D6*41* diplotype frequencies of 0.83% (0.63% exp). The Bagandan cohort group had the highest *CYP2D6*17*

and *CYP2D6*29* diplotype frequencies of 11.0% (5.29% exp) and 8.00% (4.84% exp) respectively (Table 2.12).

Table 2.10: Homozygous diplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* alleles from KGP population data estimates.

KGP (1000 Genome Project) Homozygous Diplotype Frequencies						
Diplotype	AFR (%)	YRI (%)	ESN (%)	GWD (%)	MSL (%)	LWK (%)
*CYP2D6*10*	0.80	1.04	0.74	0.86	2.16	0.04
CYP2D6*17	**4.28**	**6.25**	**5.62**	**2.40**	**6.71**	**3.50**
*CYP2D6*29*	0.78	0.94	0.50	1.23	0.78	1.12
*CYP2D6*41*	3.31	2.62	2.79	3.13	3.53	7.18

*The bold highlighted row signifies the CYP2D6*17 data which was pertinent to the study and the establishment of an iPSC cellular model.*

Table 2.11: Homozygous diplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* alleles from PharmVar population data estimates.

PharmVar Homozygous Diplotype Frequencies							
Diplotype	Combined African (%)	South African (%)	Ghanaian (%)	Tanzanian (%)	Ethiopian (%)	Zimbabwean (%)	Other African (%)
*CYP2D6*10*	0.44	0.83	0.09	0.14	0.72	-	0.27
CYP2D6*17	**3.98**	**2.99**	**7.67**	**1.96**	**1.22**	**11.56**	**3.18**
*CYP2D6*29*	0.80	0.38	-	3.92	-	-	0.45
*CYP2D6*41*	0.94	0.08	-	-	5.26	-	0.08

*The bold highlighted row signifies the CYP2D6*17 data which was pertinent to the study and the establishment of an iPSC cellular model.*

Table 2.12: Homozygous diplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* alleles from AGVP population data estimates.

AGVP (African Genome Variation Project) Homozygous Diplotype Frequencies								
Diplotype	Combined Data (%)		Zulu (%)		Ethiopian (%)		Bagandan (%)	
	Observed	*Expected*	*Observed*	*Expected*	*Observed*	*Expected*	*Observed*	*Expected*
*CYP2D6*10*	0.63	0.08	2.00	0.30	0.00	0.06	0.00	0.00
CYP2D6*17	**7.81**	**4.00**	**6.00**	**3.06**	**6.67**	**3.84**	**11.00**	**5.29**
*CYP2D6*29*	3.44	1.27	3.00	1.32	0.00	0.04	8.00	4.84
*CYP2D6*41*	0.31	0.33	0.00	0.06	0.83	0.63	0.00	0.42

*The bold highlighted row signifies the CYP2D6*17 data which was pertinent to the study and the establishment of an iPSC cellular model.*

2.3.2 Identifying *CYP2D6*17* Homozygote Individuals to Establish Cell Lines

The following results pertain to the recruitment of study participants for donation of dermal fibroblasts for the establishment of iPSC cell lines. It also includes the results of the genotyping of dermal fibroblasts used to establish the CSIR iPSC lines.

2.3.2.1 *Participant Recruitment*

Study participants were recruited for *CYP2D6*17* genotype screening to identify a homozygous individual for donation of dermal fibroblasts for the establishment of iPSC cell lines. Using allele frequencies from other African populations it was estimated that if 50

individuals were recruited, this sample size should yield on average five *CYP2D6*17* homozygous individuals. The recruitment was done in the second- and third-year medical classes from the Faculty of Health Sciences at the University of the Witwatersrand. Prospective participants were shown an informative recruitment video to present them with information on the project, and 22 individuals volunteered to participate in the study.

Of the initial 22 volunteers, only 10 agreed to undergo a buccal swab for DNA sampling after giving informed consent for the sampling and use of their DNA. A questionnaire was completed by each of the 10 study participants and information on their parental ethnic background, smoking activity, medical history and any ADR experience, were obtained. This was done to highlight potential information that could affect the DNA extraction and dermal fibroblast donation, as well as to assess ADR experiences of study participants (Table 2.13). A large mix of parental ethnicity and ancestry were observed and only three study participants stated that both parents were of the same ethnolinguistic backgrounds (1A2002 – Sesotho; 3A2008 - Sepedi; 3B4014 – Ndebele). Two participants opted not to reveal the ethnolinguistic background for their parents. Three participants indicated a history of ADRs triggered by chronic medication with drowsiness, severe stomach pain and vomiting experienced by these individuals. No participants stated being hospitalised due to ADRs and no participants were on any chronic medication at the time of the recruitment and DNA sampling. Only two study participants stated that they were smokers, with one stating that they smoke cigarettes daily (2A1004) and the other stating smoking only two to three times per week (3B1011).

Four students from the SBIMB (1, 2, 3, 4) were also recruited and consented to give buccal swab samples to test and optimize the TaqMan Genotyping assay. They consented for the sampling and use of their DNA after information sheets were given to each individual and a thorough assessment of participant understanding of the information was completed (Appendix B1).

Participant Code	Paternal Ancestry	Maternal Ancestry	Medication ADR History	Active Smoker
1A2002	Sesotho	Sesotho	YES - Drowsiness (Allergy & Pain Medication)	NO
1B1015	Tswana	Ndebele	NO	NO
1B2016	Shona	Zulu	NO	NO
2A1004	n/a	n/a	YES - Vomiting (Antibiotics & Injection - Stomach Infection	YES - Daily
2A2005	n/a	n/a	NO	NO
3A1007	Zulu	Zulu	NO	NO
3A2008	Sepedi	Sepedi	NO	NO

3B1011	Tsonga	Zulu	NO	YES - (2-3 Times per Week)
3B3013	Sesotho	Sepedi	YES - Severe Stomach Pain (Inconsistent Menstruation Medication)	NO
3B4014	Ndebele	Ndebele	NO	NO

Table 2.13: Participant questionnaire data: parental ethnicity and drug use and response.

2.3.2.2 *Sampling & DNA Extraction for CYP2D6*17 Genotyping*

Buccal cell collection and DNA extraction were performed using the Gentra Puregene® Buccal Cell Kit (Hilden, Germany). DNA was used for TaqMan genotyping for rs28371706, which was used as a proxy for *CYP2D6*17*. Two buccal swabs (one per cheek) were taken for DNA extraction and subsequent genotyping, to ensure that there was a "reserve" sample in case of contamination during or after the DNA extraction process. The DNA was extracted from both buccal swabs and kept in separate 1.5 ml microcentrifuge tubes that were labelled with the participant code and A or B for each swab. The DNA concentration measurements (absorbance at 250 nm) and 260/280 ratios (to identify protein contamination) were taken for each sample (Table 2.14). For the four test samples, the DNA was extracted and had 260/280 ratios ranging from 1.70 (3B) to 1.82 (1A) and concentrations ranging from 61.7 ng/µl (1A) to 383.5 ng/µl (2B). For TaqMan genotyping, confirmatory PCRs and sequencing, test sample 'A' aliquots were chosen with 'B' aliquots kept as reserves. For the study participant samples, the DNA was extracted and had 260/280 ratios ranging from 1.77 (3A1007/B) to 2.25 (1A2002/B) and concentrations ranging from 35.3 ng/µl (3B1011/B) to 332.9 ng/µl (3B4014/B). For TaqMan genotyping, confirmatory PCRs and sequencing, the study participants sample aliquots with a higher DNA concentration and a 260/280 ratio between 1.8 to 2.0 were chosen.

Table 2.14: DNA concentrations (ng/µl) and quality (260/280 ratio) for the study participant DNA sample extractions performed at two different laboratories.

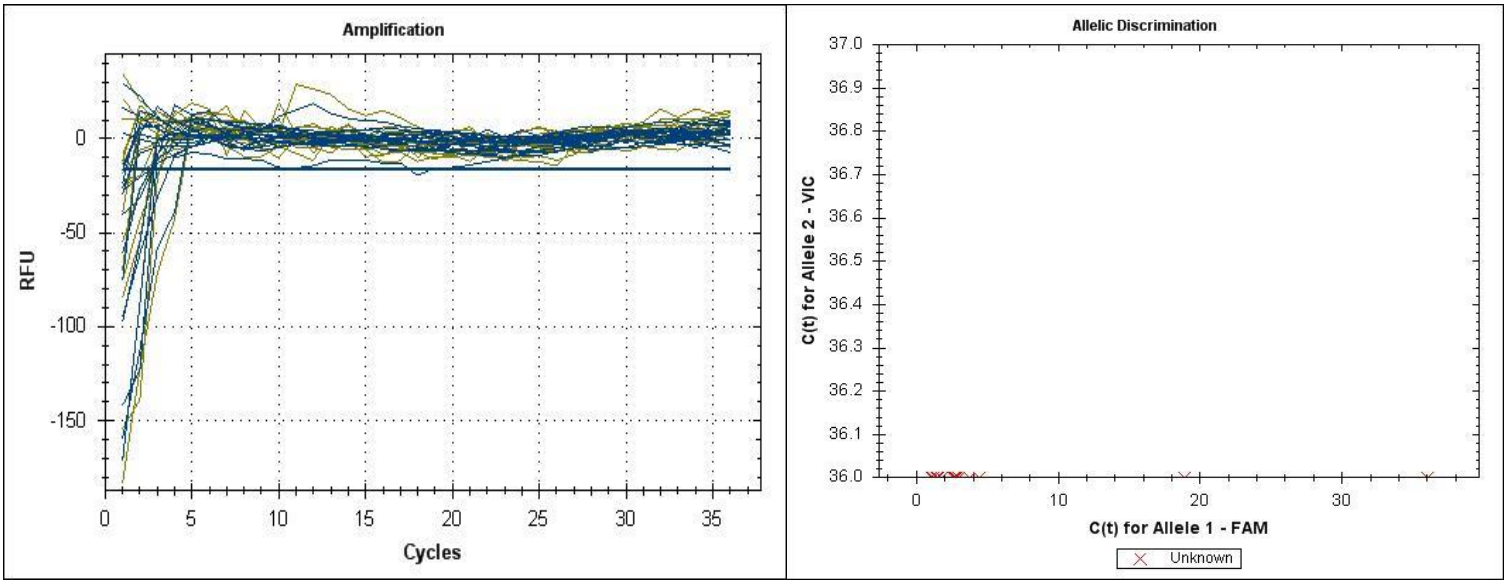
		NHLS Laboratory		CSIR Laboratory	
Sample		260/280 Ratio	Conc. (ng/µl)	260/280 Ratio	Conc. (ng/µl)
Test Samples					
1	A	2.30	44.6	1.82	61.7
	B	2.06	100.0	1.81	145.1
2	A	1.93	110.6	1.74	340.7
	B	1.86	136.8	1.73	383.5
3	A	2.13	71.4	1.79	94.3
	B	2.21	60.7	1.70	66.8
4	A	1.99	108.3	1.77	183.2
	B	1.91	177.9	1.72	224.7
Study Participant Samples					
1A2002	A	-	-	1.86	142.9

	B	-	-	2.25	58.1
1B1015	A	-	-	1.91	68.0
	B	-	-	1.85	68.4
1B2016	A	-	-	1.88	82.0
	B	-	-	2.04	52.8
2A1004	A	-	-	1.94	118.0
	B	-	-	1.84	111.1
2A2005	A	-	-	1.87	109.4
	B	-	-	1.98	56.8
3A1007	A	-	-	1.87	65.4
	B	-	-	1.77	92.5
3A2008	A	-	-	1.89	83.2
	B	-	-	2.11	55.5
3B1011	A	-	-	2.16	47.4
	B	-	-	1.88	35.3
3B3013	A	-	-	1.77	69.6
	B	-	-	1.89	74.4
3B4014	A	-	-	1.91	125.3
	B	-	-	1.84	332.9

2.3.2.3 *TaqMan Assay Optimisation & CYP2D6 Genotyping*

The TaqMan assay was initially optimized using two amplification platforms (*Alternative Method 1 and 2*) to establish which would be the most sensitive and cost-efficient assay for genotyping. The two platforms used were the KAPA™ Taq ReadyMix PCR Kit (Wilmington, USA) (*Alternative Method 1*) and Q5® High-Fidelity DNA Polymerase (Ipswich, USA) (*Alternative Method 2*) amplification platforms. The DNA used for the initial assay assessment was from the dermal fibroblast cell lines, BlkAfrCon and P7OxCon, that were used to create iPSC cell lines in the CSIR laboratory.

For the *Alternative Method 1* and *Alternative Method 2* reactions, a temperature gradient of 57-64°C and 60-70°C respectively were employed to determine the optimal T_m for binding of the TaqMan™ MGB probes and TaqMan™ Assay Mix primers to the DNA. However, the *Alternative Method 1* (Fig 2.5) and *Alternative Method* 2 (Fig 2.6) platforms were unable to



amplify the fluorescent signals from TaqMan™ MGB probes. The relative fluorescence units (RFU) figures for *Alternative Method 1* and *Alternative Method 2* started in at a very negative value and had a jagged shape to their curves. This could be due to master mix not being appropriate for the TaqMan™ MGB probes causing either buffer-nucleotide instability or weak binding of the probe. These were then judged as inadequate for this particular rs28371706 *CYP2D6*17* proxy genotyping assay.

Fig 2.5A (Left): Amplification curve plot illustrating the amplification of the fluorescent signals from the probes at each of the KAPA™ Taq Polymerase PCR cycles. Threshold was not met or passed, as signal did not amplify for both the VIC and FAM dyes. **Fig 2.5B (Right)**: Allelic discrimination plot illustrating the rs28371706 allele results from the TaqMan® Custom SNP Genotyping Assay Mix and the sequences of the PCR products for DNA samples: Oxford-derived Black African-descent (P7OxCon) and Black African-descent Fibroblasts (BlkAfrCon), and no-template control (NTC). No C(t) thresholds for FAM dye (T Allele) and VIC Dye as PCR failed, and fluorescent signal did not amplify.

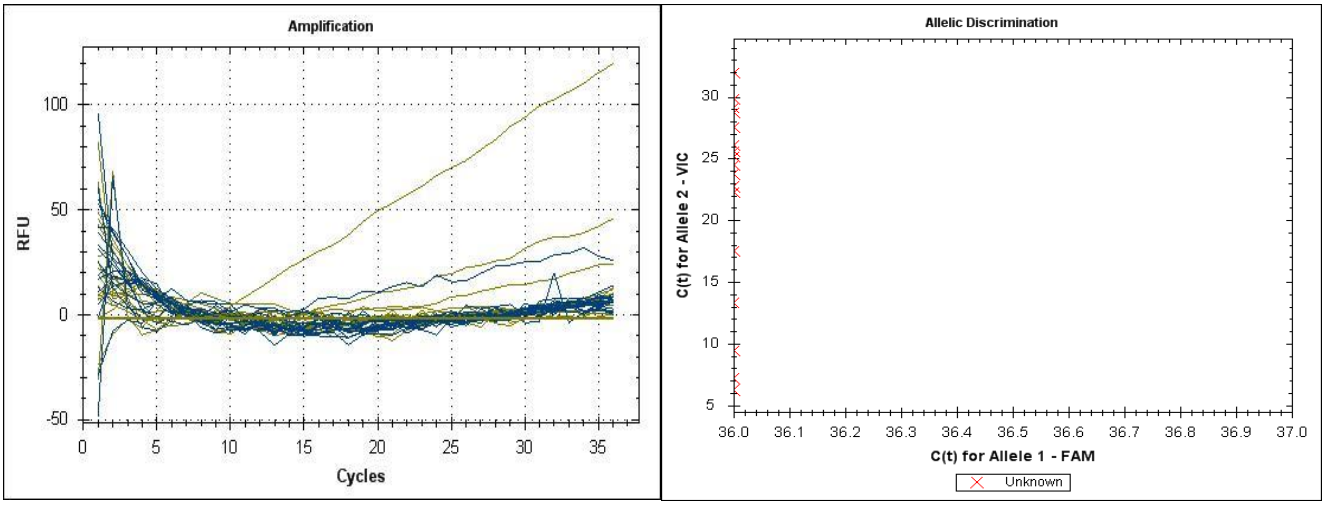


Fig 2.6A (Left): Amplification curve plot illustrating the amplification of the fluorescent signals from the probes at each of the Q5® Polymerase PCR cycles. Threshold was not met or passed, as signal did not amplify for both the VIC and FAM dyes. **Fig 2.6B (Right)**: Allelic discrimination plot illustrating the rs28371706 allele results from the TaqMan® Custom SNP Genotyping Assay Mix and the sequences of the PCR products for DNA samples: Oxford-derived Black African-descent (P7OxCon) and Black African-descent Fibroblasts (BlkAfrCon), and no-template control (NTC). No C(t) thresholds for FAM dye (T Allele) and VIC Dye as PCR failed, and fluorescent signal did not amplify.

A third method, using the TaqPath™ ProAmp™ Master Mix (Thermo Fisher Scientific, MA), was then used and was able to resolve the problem of failed amplification of the fluorescent signal. The DNA from test samples 1, 2, 3 and 4 (aliquot A), and the two CSIR iPSC cell lines described above (BlkAfrCon and P7OxCon), were used in optimisation the genotyping using the TaqPath™ Master Mix. Following a TaqMan assay run, the four test samples were found to be C/C homozygotes and the two iPSC cell lines were found to be C/T heterozygotes (Fig 2.7). The optimisation of the TaqPath™ method indicated that it was the most suitable amplification method for the genotyping the rs28371706 SNP. TaqMan genotyping was then performed on the 10 study participant DNA samples to determine their genotypes. The TaqMan™ MGB probes met the threshold of ~600 RFU for the florescent signal

amplification at approximately the 30th cycle of the PCR. In the allele discrimination, thresholds for a positive C allele and T allele signal was set at ~35 cycles and ~37 cycles respectively, with a positive signal being recorded before the cycle threshold. The results of the TaqMan genotyping using TaqPath™ platform indicated that four DNA samples (1A2002, 1B1015, 2A2005, 3A1007) had a C/C genotype at the rs28371704 SNP locus, five DNA samples (1B2016, 3A2008, 3B1011, 3B3013, 3B4014) had a T/T genotype, and one DNA sample (2A1004) was a T/C heterozygote (Fig 2.8).

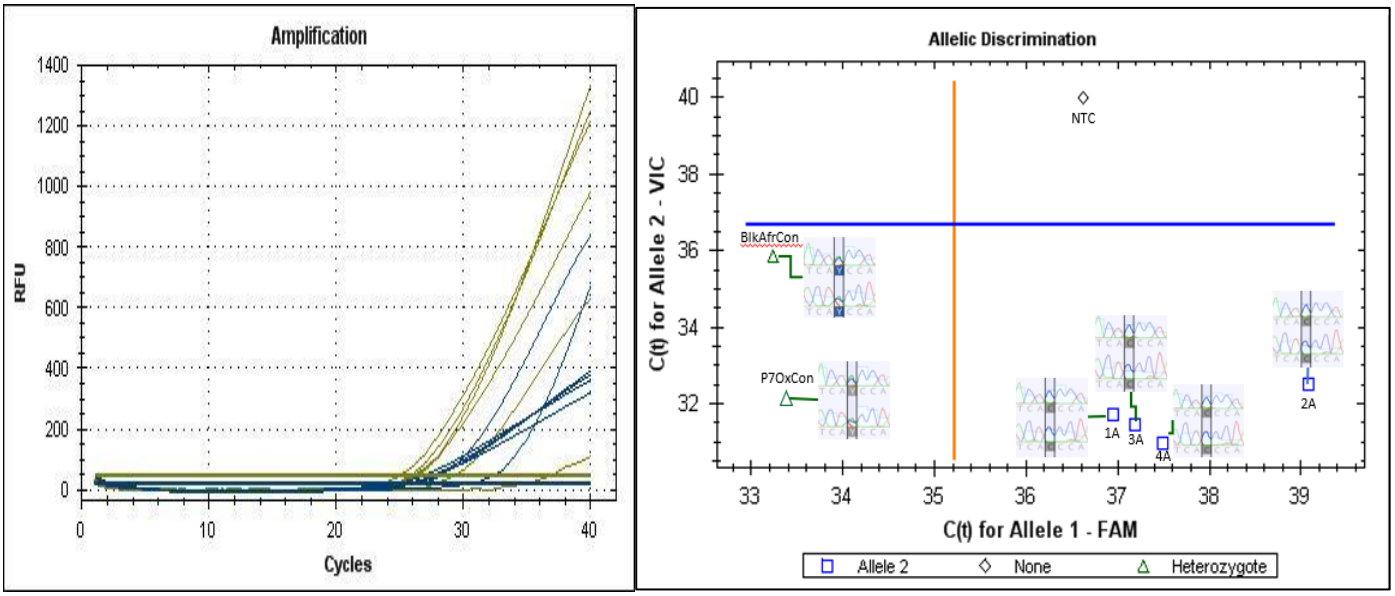
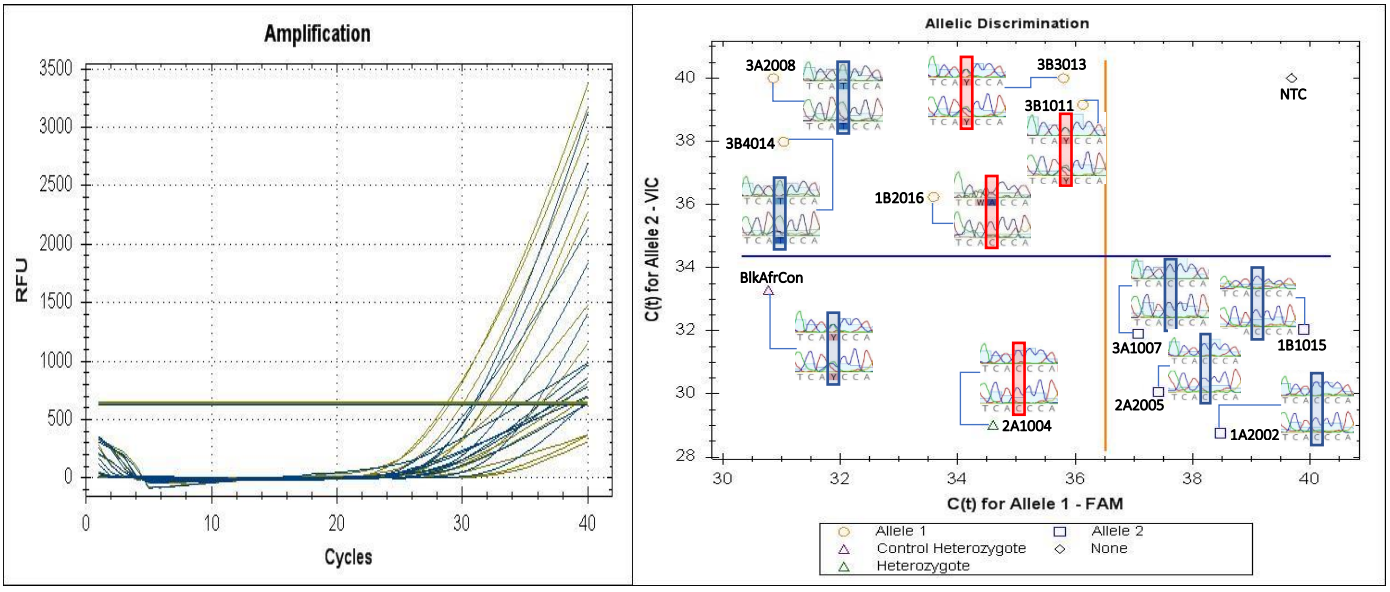


Fig 2.7A (Left): Amplification curve plot illustrating the amplification of the fluorescent signals from the probes at each of the TaqPath™ PCR cycles. Threshold was set at ~10 RFU for VIC dye and ~20 RFU for FAM dye signal amplification. **Fig 2.7B (Right)**: Allelic discrimination plot illustrating the rs28371706 allele results from the TaqMan® Custom SNP Genotyping Assay Mix and the sequences of the PCR products for DNA samples: 1A, 2A, 3A, 4A, Oxford-derived Black African-descent (P7OxCon), Black African-descent Fibroblasts



(BlkAfrCon), and Control Sample: Non-template control (NTC). C(t) thresholds for FAM dye (T Allele) = ~35 cycles; VIC Dye (C Allele) = ~37 cycles.

Fig 2.8A (Above): Amplification curve plot illustrating the amplification of the fluorescent signals from the probes at each of the PCR cycles. Threshold was set at ~600 RFU for both the VIC and FAM dye signal amplification. **Fig 2.8B (Right)**: Allelic discrimination plot illustrating the rs28371706 allele results from the TaqMan® Custom SNP Genotyping Assay Mix and the sequences of the PCR products for DNA samples: 1A2002, 1B1015, 1B2016, 2A1004, 2A2005, 3A1007, 3A2008, 3B1011, 3B3013, 3B4014, Black African-descent Fibroblasts (BlkAfrCon), and no-template control (NTC). C(t) thresholds for FAM dye (T Allele) = ~37 cycles; VIC Dye (C Allele) = ~35 cycles. Red rectangles on the sequencing image indicate mismatches between the TaqMan genotyping and sequencing, and blue rectangles indicate equivalent results between the TaqMan genotyping and sequencing.

2.3.2.4 *Confirmation of Genotypes by Sequencing of PCR products*

The primers for the confirmatory PCR were optimised with the DreamTaq™ Polymerase to establish the correct annealing temperature and DNA concentration. A PCR temperature gradient reaction was initially performed to determine the optimal T_m to use for the best PCR outcome. The PCR products were electrophoresed on a 2% agarose gel for 1 hr at a constant voltage of 100V. The first PCR showed that ~62°C was the ideal T_m due to it having the correct size fragment of 172 bp, however, contamination was found in the NTC lane on the gel (Fig 2.9). After numerous troubleshooting steps, including sterilising pipettes, plasticware and instrumentation, as well as testing reagents for contaminants and making new aliquots of reagents, the amplicon contamination was successfully resolved. The PCR temperature gradient was then repeated, and the results showed that an annealing temperature between 60-62°C was the best range for an efficient PCR, with 62°C chosen as the reaction T_m (Fig 2.10).

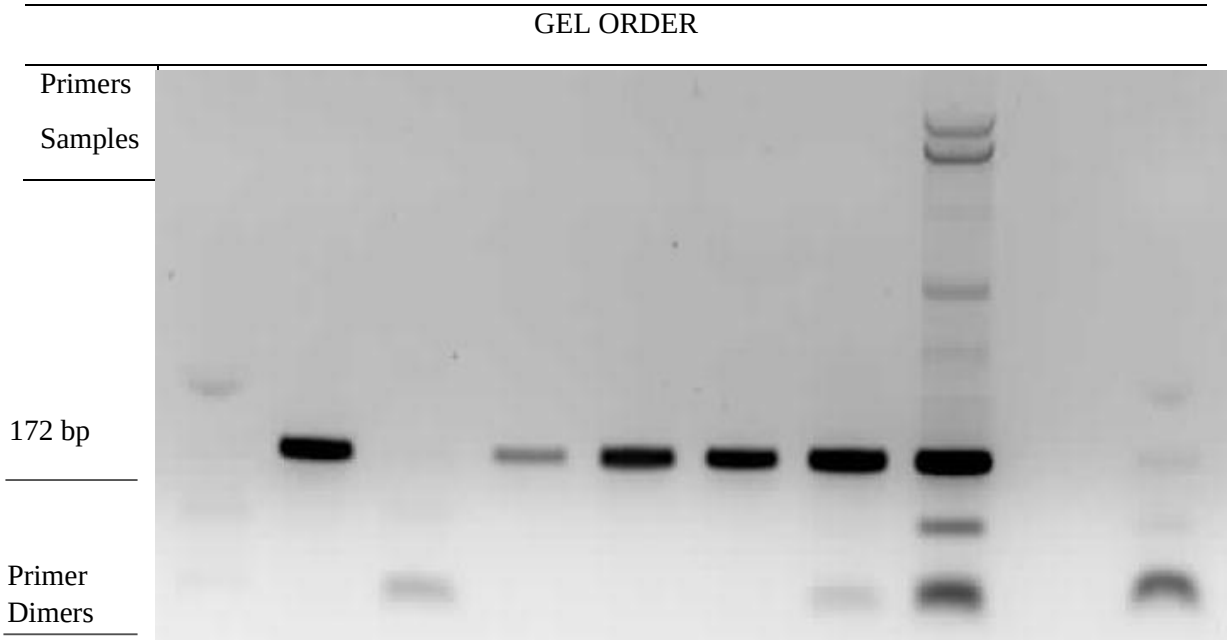


Fig 2.9: Gel electrophoresis of the PCR products during the rs28371706 PCR optimisation. PCR temperature gradient reaction was performed, and products were run on a 2% agarose gel. Contamination found from an amplicon in the No-template DNA negative control. *Lanes 1-8: DNA Samples; Lane 9: No-template DNA negative control (NTC).*

PCR using DreamTaq™ polymerase was then performed on the four test samples, two CSIR dermal fibroblast cell lines and all 10 participant DNA samples to extract PCR products for sequencing to confirm the genotypes. The sequencing results confirmed that all the test DNA samples had a C/C genotype and the BlkAfrCon and P7OxCon cell lines had a C/T genotype (Table 2.15). For the 10 study participant DNA samples, the PCR product sequencing results showed that five study participant DNA samples had the C/C genotype, two had the C/T genotype, another two had the T/T genotype at the rs28371706 locus, and one rare A/C genotype was found (Table 2.14). The rs28371706 locus is known to be triallelic (C/A/T), however, the A allele is rarely found in individuals. There were four discrepancies between

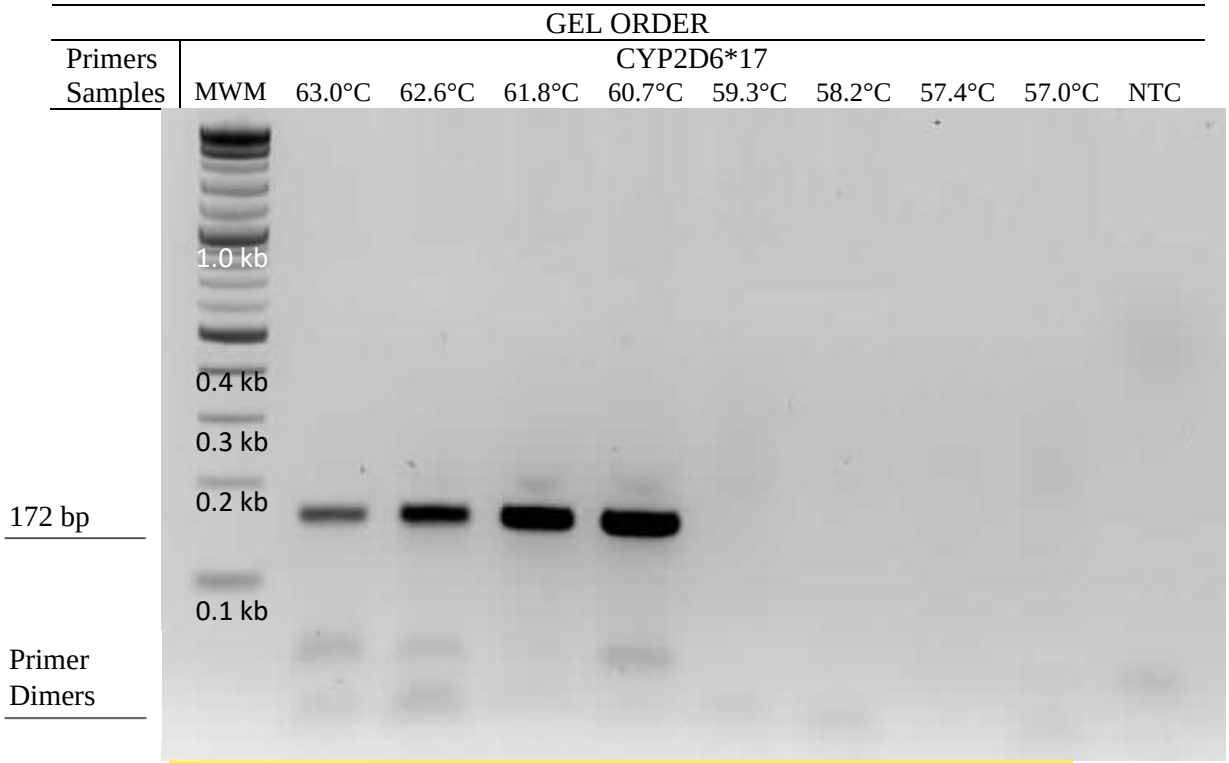


Fig 2.10 Gel electrophoresis of the PCR products during the rs28371706 PCR optimisation. PCR temperature gradient reaction was performed, and products were run on a 2% agarose gel. *Lane 1: Molecular Weight Marker (MWM; New England BioLabs® Quick-Load® Purple 2-Log DNA Ladder); Lanes 2-9: DNA Samples; Lane 9: No-template DNA negative control (NTC).*

the TaqMan™ genotyping (1B2016 T/T; 2A1004 C/T; 3B1011 T/T; 3B3013 T/T) and the sequence-based genotyping (1B2016 A/C; 2A1004 C/C; 3B1011 C/T; 3B3013 C/T). The sequencing is the gold standard and therefore the TaqMan genotyping may have presented false results. This could be due to the low number of samples used for the TaqMan which is not ideal in terms of distinguishing the genotypes based on different levels of fluorescence for the two alleles.

Two individuals were homozygous for the *CYP2D6*17* allele (based on rs28371706 as proxy) and were suitable as donors for developing iPSCs for the downstream assays on drug metabolism. They were selected, together with two control individuals who did not have the

*CYP2D6*17* allele to donate skin samples. They were (C/C: 1A2002 and 3A1007 and T/T: 3A2008 and 3B4014) and were asked to donate a skin biopsy to establish iPSC lines for the SC-HLC model. One participant (3A2008) declined to participate further after the call back for sample donation and the three other participants (1A2002, 3A1007, 3B4014) donated their skin samples after informed reconsent following a discussion with an HPCSA certified genetic counsellor from the NHLS in Braamfontein.

Table 2.15: Comparison of rs28371706 SNP locus genotype calling using the TaqMan PCR assay and DNA sequencing of PCR product

Test DNA Samples	TaqMan PCR Genotype	Sequencing Confirmed Genotype
1A	C/C	C/C
2A	C/C	C/C
3A	C/C	C/C
4A	C/C	C/C
CSIR Cell Lines		
BlkAfrCon	C/T	C/T
P7OxCon	C/T	C/T
Study Participant DNA Samples		
1A2002	C/C	C/C
1B1015	C/C	C/C
1B2016	T/T	A/C
2A1004	C/T	C/C
2A2005	C/C	C/C
3A1007	C/C	C/C
3A2008	T/T	T/T
3B1011	T/T	C/T
3B3013	T/T	C/T
3B4014	T/T	T/T

The text in red indicates DNA samples that showed discrepancies between the TaqMan genotyping and sequencing for rs28371706

2.4 DISCUSSION

The findings of this chapter focused on achieving two objectives for this study. Firstly, allele and haplotype frequencies of reduced function *CYP2D6* variants, with a focus on *CYP2D6*17,* were determined in several African populations using public genome data. Secondly, following genotype screening for rs28371706 (a proxy SNP for the *CYP2D6*17* allele) in black South Africans, a *CYP2D6*17* homozygote was identified and provided a skin biopsy for the establishment of a SC-HLC drug metabolism model. The results of this chapter present the outcomes of the data mining investigations on *CYP2D6* reduced function variants in African populations and discusses their potential impact on drug efficacy and potential adverse effects. It also outlines some of the challenges and limitations experienced in this study, and the impact on the other chapters of this study.

2.4.1 Prevalence of Reduced Function *CYP2D6* Variants in African Populations & Potential Impact on Drug Metabolism

The results provide information on the allele frequencies of SNPs associated with the *CYP2D6*17* allele (actually a haplotype). The haplotype and homozygous diplotype frequencies of four reduced function *CYP2D6* star alleles (*CYP2D6*10, CYP2D6*17, CYP2D6*29, CYP2D6*41)* were obtained by analysing data from the KGP, PharmVar and AGVP data.

2.4.1.1 *CYP2D6*17 SNP-associated Allele Frequencies in African Populations*

The *CYP2D6*17* allele is a haplotype defined by four SNPs (rs28371706, rs1058164, rs16947 and rs1135840) (haplotype TCTC), and the *CYP2D6*2* allele is a haplotype that differs from the *CYP2D6*17* allele by a single allele difference at the first SNP (haplotype CCTC), however it is defined by the last two SNPs of this haplotype.

For rs16947 and rs1135840, the KGP data showed African variant allele frequencies of 44.6% and 32.8% respectively, and in the AGVP data 38.4% and 30.6% respectively. These two SNPs (rs16947 and rs1135840) defining the *CYP2D6*2* allele, are found in all populations worldwide, and described as one of the most common alleles for *CYP2D6* (Zhou, Ingelman-Sundberg and Lauschke, 2017). CYP2D6*2 enzyme has a clinically normal function at the enzyme activity and metabolism of substrate levels (Marez *et al.*, 1997; Sakuyama *et al.*, 2008). When examining *CYP2D6*2* from KGP data, the allele frequencies ranged from 34.4% (South Asian) to 12.9% (East Asian), with the African allele frequency being 18.2% (Auton *et al.*, 2015). This is comparable to the AGVP data which had an African allele frequency of 16.6%, with the Bagandan group being 11.5%, the Ethiopian group being 22.9%, and Zulu group being 14.0%. According to Zhou et. Al. (2017), *CYP2D6* has global

allele frequencies ranging from 36.2% (South Asian) to 14.0% (East Asian). When looking at the African frequencies from this Zhou data, the allele frequency was 26.7% for *CYP2D6*2* (Zhou, Ingelman-Sundberg and Lauschke, 2017). This suggests that the data from this study agrees with the literature on the allele frequencies of *CYP2D6*2*. Also, knowing how important rs16947 and rs1135840 are for *CYP2D6*17* allele, their high frequencies appear to be associated to the relatively high frequencies of *CYP2D6*17* in African populations.

For rs1058164, the allele frequencies for the KGP data was 32.8% and AGVP data was 10.6%. The discrepancy seen between the two datasets is most likely due to differences in the populations included in the datasets. This may also be a result of this SNP being a silent mutation and evolutionary pressure on this SNP may not be as prominent in African populations as other SNPs in *CYP2D6*17*. This would be due to the silent mutation not creating a change in the amino acid sequence, as the amino acid it codes for is valine. The position of the rs1058164 SNP change 1661G>C is responsible for the last part of the codon sequence of valine, which does not result in an amino acid change (Medhasi *et al.*, 2016). However, this SNP is found in numerous *CYP2D6* variants and this commonality may be a result of linkage disequilibrium or some other evolutionary mechanism yet to be understood (Kim *et al.*, 2014).

The SNP rs28371706 is the definitive SNP of the *CYP2D6*17* allele with the T allele present in this haplogroup, found together with rs16947 T allele and rs1135840 C allele. For rs28371706, KGP and AGVP data showed an allele frequency of 21.8% and 20.2% respectively and shows agreement between what is publicly available in the KGP data and the AGVP analyses. The *CYP2D6*17* haplotype frequencies for the African data in each of the datasets were 20.70% (KGP), 19.95% (PharmVar), and 20.00% (AGVP), and shows agreement between the datasets and analyses. The regional variation shows high correlation with the frequency of *CYP2D6*17*, especially in the KGP ESN population and the AGVP data. This shows that the selective pressure for the rs28371706 T allele has resulted in the *CYP2D6*17* haplotype being more common in the African populations.

2.4.1.2 *CYP2D6*17 Haplotype & Diplotype Frequencies in African Populations*

The haplotype and diplotype frequencies of CYP2D6*17 are of importance when looking at this genetic variant's effects on CYP2D6 enzyme activity. The results here gave us insight into the prevalence of the variant in African populations after data was mined from AGVP, KGP and PharmVar.

The Bagandan group from the AGVP data had a *CYP2D6*17* haplotype frequency of 23.00% and this seemed to be more similar to the frequencies from the West African populations (KGP: ESN = 23.70%, YRI = 25.00%, MSL = 25.90%; PharmVar: Ghanaian = 27.70%) rather than the East African populations that generally have lower frequencies (KGP: LWK = 18.70%; PharmVar: Tanzanian = 13.99%, Ethiopian = 11.04%; PharmVar: Ethiopian = 19.58%). An explanation for this is found in the migratory patterns and origins of the Baganda people, which may have contributions from individuals that were a part of the Bantu migration of people from Cameroon and Congo to Shungwaya in Tanzania (Ssekamwa, 1986). There are similarities between the Bagandan *CYP2D6*17* haplotype frequency and data from Oluka et.al (2014), where Western Nilotes (Luo) from Kenya have a *CYP2D6*17* haplotype frequency of 22.5% (Oluka *et al.*, 2014). In this data, the Bantu (Kikuyu) group has a higher frequency (33.5%) than that in the Baganda group, highlighting potential origins or admixture with Nilote populations from Sudan and northern regions of Africa (Oluka *et al.*, 2014). However, to establish the contribution of these migratory patterns would require further research into genetic drift due to admixture during these migrations and how that could have affected the distribution of *CYP2D6* genotypes in those regions.

The Ethiopian group from the AGVP has a significantly different *CYP2D6*17* haplotype frequency (19.58%) when compared the Ethiopian data from PharmVar (11.04%). This is likely because the different studies included different ethnolinguistic groups from Ethiopia, and it is know that the origins of the populations native to Ethiopia can be very different (Pagani *et al.*, 2012). Even though the Ethiopian populations are very diverse, it is possible to broadly cluster them into two groups: the Semitic-Cushitic group, characterised by a strong non-African contribution; and Nilotic and Omotic group, which is characterised by strong West and East African contributions with almost no non-African contributions (Pagani *et al.*, 2012). Looking at the ancestry of the individuals in this group may give indications of contributions from other African and world populations that have influenced the allele frequencies in those groups. The AGVP Ethiopian group has allele frequencies more similar to populations of South Africa and the LWK group in Kenya, while the PharmVar Ethiopian group seems similar to the Tanzanian population. The AGVP Ethiopian data has contributions from individuals with Oromo and Amhara ethnolinguistic backgrounds (Pagani *et al.*, 2012). These ethnic populations are closely linked to the Semitic-Cushitic groups, which is an explanation with the closer link to the East African and South African populations (Pagani *et al.*, 2012; Choudhury *et al.*, 2018). The South African populations have known contributions with East African populations and this may be confirmed further with the similar frequencies with the LWK and AGVP Ethiopian groups (Choudhury *et al.*, 2018). The PharmVar

Ethiopian population has more similarities with the Tanzanian populations, and because it is a combination of numerous Ethiopian populations there may be more Middle Eastern and non-African contributions present in this population (Aklillu *et al.*, 1996; Pagani *et al.*, 2012).

For the homozygous *CYP2D6*17* diplotypes, the AGVP Zulu and Ethiopian groups showed similarities between the West African populations (KGP: ESN = 5.62%, YRI = 6.25%, MSL = 6.71%; PharmVar: Ghanaian = 7.67%), which may also be due to the migration patterns of the Bantu populations (Ssekamwa, 1986; Choudhury *et al.*, 2018). This is also seen in the AGVP Bagandan (11.00%) and PharmVar Zimbabwean (11.56%) groups, which could again be due to Bantu migration contributions in each of the populations (Ssekamwa, 1986; Choudhury *et al.*, 2018). However, discrepancies were seen between the PharmVar South African (2.99%) and AGVP Zulu (6.00%) groups, as well as the two Ethiopian groups from the AGVP (6.67%) and PharmVar (1.22%) data. It is also of interest to see the differences between the observed and expected diplotype frequencies. The *CYP2D6*17* frequencies had an approximately difference of 3% for the Zulu, Ethiopian and Combined data, with the Bagandan populations having a much larger difference of approximately 6%. This shows the extent of the prevalence of this IM variant in African populations especially in the Bagandan populations.

2.4.1.3 *Other Reduced Function CYP2D6 Haplotype & Diplotypes frequencies*

In addition to examining frequencies of *CYP2D6*17*, the data mining part of the study also included three other reduced function star alleles: *CYP2D6*10*, *CYP2D6*29* and *CYP2D6*41*. This was to examine the prevalence of the of these common "reduced enzymatic function" *CYP2D6* variants in African populations and highlight the importance and significance of *CYP2D6*17* in both a pharmaceutical research and clinical interventions for these populations.

For *CYP2D6*10* haplotype frequencies, there seemed to be a few differences between the AGVP data and the PharmVar and KGP data. The two most apparent differences are seen between the Ethiopian (2.50%) and Zulu (5.50%) AGVP data, and the PharmVar Ethiopian (8.47%) and South African (9.11%) data. For the *CYP2D6*10* diplotype data, again there were comparable differences in the frequencies of the PharmVar South African (0.83%) and Ethiopian (0.72%) populations, and the AGVP Zulu (2.00%) and Ethiopian (0.00%) populations.

With regards to the South African data, the discrepancy in the haplotype frequency stems mainly from a study in a Xhosa population that is part of the PharmVar data and found a

*CYP2D6*10* haplotype frequency of 19.23%. Other studies from the PharmVar data had haplotype frequencies (Coloured = 2.53%; Mixed populations = 5.10%) more similar to the Zulu group (Gaedigk and Coetsee, 2008; Dodgen *et al.*, 2016). This seems to be the case with the diplotype frequencies as well, as both datasets are based on the assumption that these populations are in HWE equilibrium. Using this assumption, the Xhosa population would have a diplotype frequency of 3.69%, showing similarities with the AGVP Zulu group (2.00%), the Coloured population would be at 0.06%, and the Mixed populations would be at 0.26%. The Xhosa and Zulu populations share a common Bantu ancestry and may be responsible for the similarity between the *CYP2D6*10* frequencies in the two groups (Choudhury *et al.*, 2018).

The Ethiopian data for *CYP2D6*10* had a large difference between the PharmVar (8.47%) and AGVP (2.50%) haplotype frequencies. This also shows the differences in the makeup of the two sets of data, with the AGVP Ethiopian data (2.50%) seem more similar to the Tanzanian (3.80%) and LWK (2.02%) groups, and PharmVar Ethiopian data (8.47%) very similar to the PharmVar South African and KGP West African (ESN: 8.59%; GWD: 9.29%; YRI: 10.20%) populations. This again resembles the migratory patterns of the different groups in the Ethiopian populations and contributions from West and East Africa, as well as non-African populations that may be able to explain these *CYP2D6*10* frequencies (Aklillu *et al.*, 1996; Pagani *et al.*, 2012; Choudhury *et al.*, 2018).

The Bagandan AGVP group has a low *CYP2D6*10* frequency (0.50%), in contrast to the higher frequencies in the AGVP Zulu (2.00%) and KGP MSL (2.16%) groups and this could also be indicative of Bantu migration contributions to the *CYP2D6* genetic distributions. This is also the case when comparing the rest of the KGP West African (ESN = 0.74%, GWD = 0.86%, YRI = 1.04%) and PharmVar South African (0.83%) populations, as well as the Ghanaian (0.09%) and LWK (0.04%) populations (Ssekamwa, 1986; Choudhury *et al.*, 2018).

For *CYP2D6*29*, the haplotype frequencies showed large differences and a few populations did not have any data for *CYP2D6*29* (PharmVar: Ghanaian, Ethiopian and Zimbabwean populations; AGVP: Ethiopian). The Bagandan AGVP group had a haplotype frequency similar to the PharmVar Tanzanian population, which again highlights the possible link between the two populations due to contributions of the Bantu migration patterns from West Africa to Tanzania (Ssekamwa, 1986). The Zulu AGVP group has similar allele frequencies to the West African populations (KGP: GWD = 11.10%, LWK = 10.60%; AGVP: Zulu = 11.50%), which may also be due to the Bantu expansion that resulted in the migration of Bantu-speaking populations from West African to Southern Africa (Choudhury *et al.*, 2018).

When looking at diplotypes, the AGVP Zulu (3.00%) and PharmVar Tanzanian (3.92%) data appear similar comparing frequencies. The other groups seem to vary between 0.38% (PharmVar South African) and 1.23% (GWD) with exception of the Baganda group with had a diplotype frequency significantly larger than all the other groups at 8.00%, and may be an anomaly of the potential inaccuracies of the Astrolabe prediction.

For *CYP2D6*41*, there were also a wide range of haplotype frequencies recorded in each of the datasets in this study and a few populations did not have any data for *CYP2D6*41* (PharmVar: Ghanaian, Tanzanian and Zimbabwean populations). The Zulu AGVP and PharmVar South African groups had similar haplotype frequencies (PharmVar: South African = 2.84%; AGVP: Zulu = 2.50%). The two Ethiopian groups have a large difference between them, with the AGVP Ethiopian population significantly lower than its counterpart group in the PharmVar data. This again can be linked to the differences in the composition of the ethnic groups in the AGVP and PharmVar datasets (Pagani *et al.*, 2012; Choudhury *et al.*, 2018). This was seen as the AGVP Ethiopian data had more in common with all South African and PharmVar Other African populations. The similarity between the PharmVar South and Other African was also indicative of the admixture that is present within the South African population as it often presents itself as an admixed group much like the Other African data. The Bagandan group had haplotype frequencies similar the AGVP Ethiopian group, which may again be from admixture with common East African populations (Ssekamwa, 1986). The results for diplotype frequencies highlighted similarities between the KGP West African populations (YRI = 2.62%, ESN = 2.79%, GWD = 3.13%, MSL = 3.53%) with the LWK population having a higher diplotype frequency (7.18%). The PharmVar Ethiopian diplotype frequency (5.26%) appeared to lie between the KGP West African and LWK populations, indicating again the possibility of contributions from migrations of people from each of the populations.

All these findings show the importance of pharmacogenetic research to establish the extent of genetic variation on pharmacological interventions and adverse drug reactions. The implications of this high *CYP2D6*17* prevalence in African populations could also result in high levels of adverse drug reactions and poor treatment efficacy, especially the Zimbabwean and Bagandan populations. This shows the need for better healthcare policies and clinical guidelines to address its effects on health intervention outcomes. There also needs to be further research into the effects of *CYP2D6*17* on all relevant drug substrates of CYP2D6, and more investigations into its prevalence in other African populations.

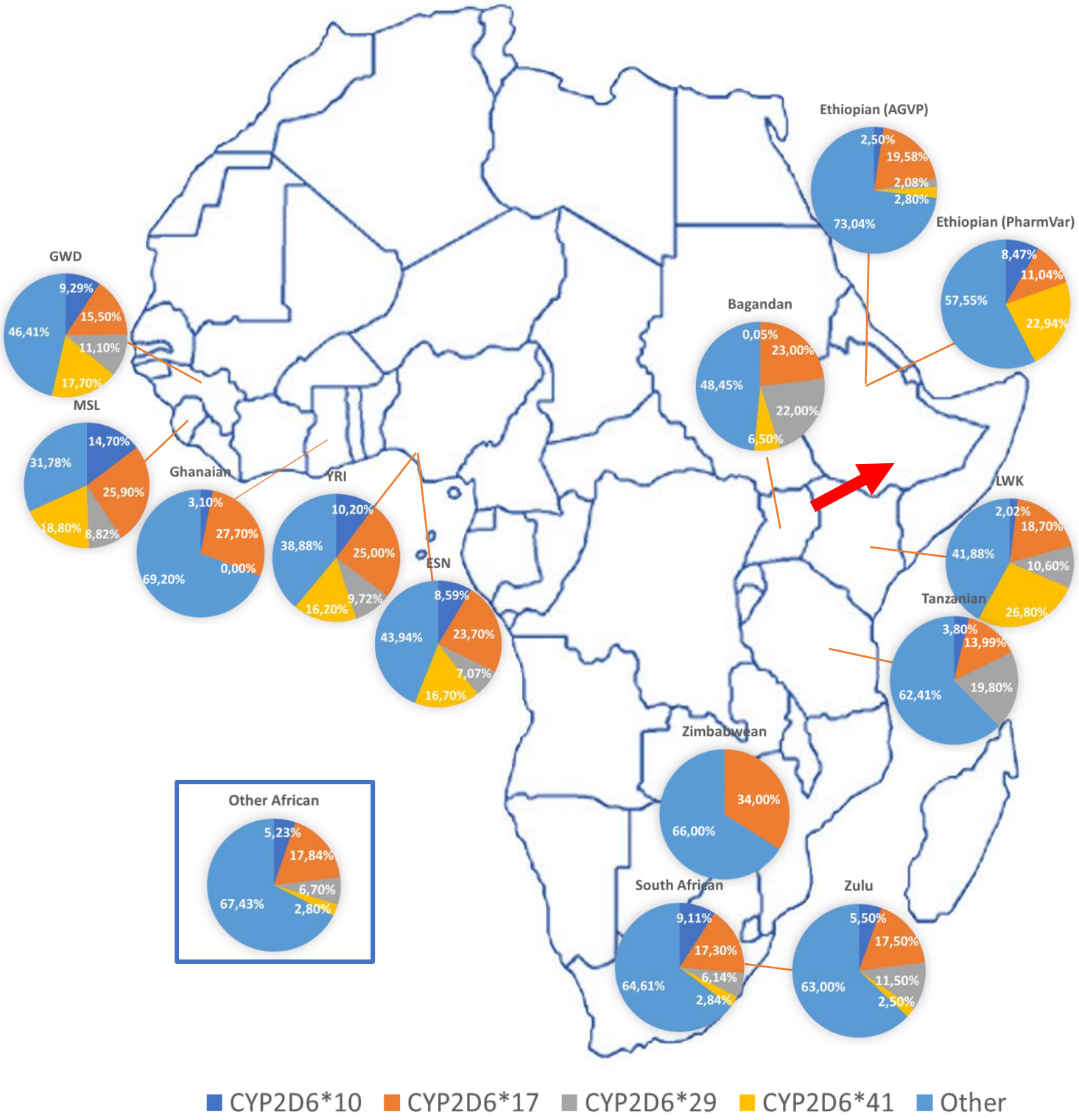


Fig 2.11: Haplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* overlaid with the migratory patterns of the first and second bantu expansions. Red arrows show the first bantu expansion from West Africa to East and South Africa. **Gold** line shows the second bantu expansion from East Africa down to South Africa.

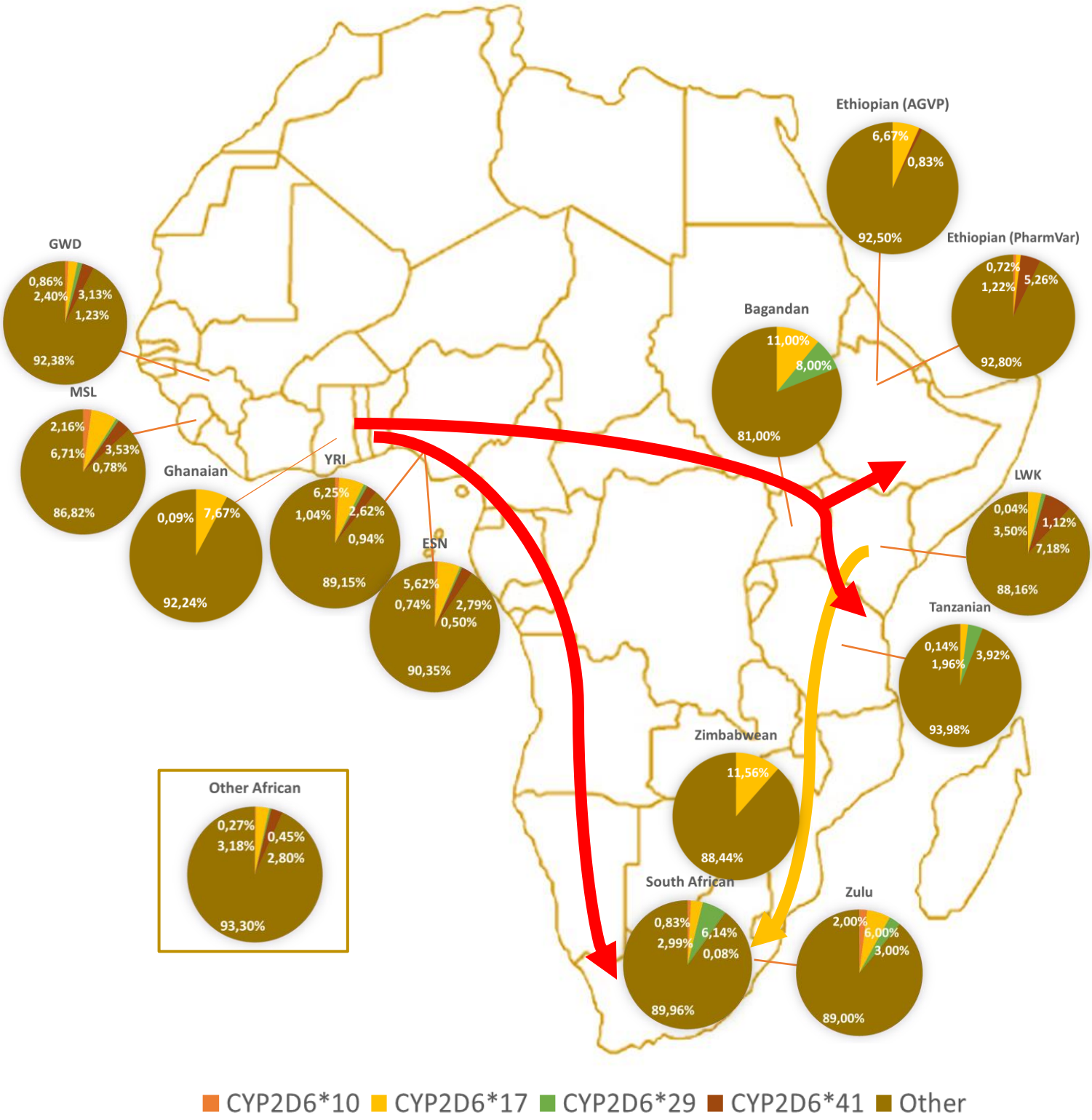


Fig 2.12: Diplotype frequencies of the *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29* and *CYP2D6*41* overlaid with the migratory patterns of the first and second bantu expansions. Red arrows show the first bantu expansion from West Africa to East and South Africa. Gold line shows the second bantu expansion from East Africa down to South Africa.

2.4.2 Identification of Individuals to Act as Cell Donors to Establish an SC-HLC *CYP2D6*17* Drug Metabolism Model

This section describes the process of identifying individuals that were *CYP2D6*17* homozygotes and therefore potential cell donors to develop the iPSC lines for pharmacokinetic drug metabolism studies. This was achieved by recruiting study participants, requesting a buccal swab for DNA sampling, performing a TaqMan genotyping on rs28371706 as a *CYP2D6*17* proxy, confirming the genotypes by sequencing of PCR products, and then identifying the *CYP2D6*17* homozygote individuals for a skin biopsy.

2.4.2.1 *Recruitment of Study Participants & Buccal Swabs for DNA extraction*

The initial target for recruitment was 50 black South African individuals, whom we estimated would potentially yield five *CYP2D6*17* homozygotes. However, during the recruitment process only 22 individuals provided contact details and an indication that they would be willing to participate in the study. However, once re-contacted only 10 individuals agreed to provide a buccal swab for DNA genotyping. The recruitment of individuals for research studies is a difficult task that often results in withdrawals and cancellations by the potential participants. Due to time limitations it was therefore decided to proceed with DNA sampling on the 10 participants and if no *CYP2D6*17* homozygous individuals were identified, another round of recruitment would be undertaken.

The buccal swab DNA sampling on the 10 participants took place at the NHLS in Braamfontein after an information session and provision of informed consent. This was followed by a brief interview and questionnaire for each participant. The brief interview and questionnaire revealed that admixture between ethnolinguistic groups was common in Johannesburg as numerous different groups had converged here for socio-economic reasons. Only a few participants stated experiences of ADRs when taking particular medications. This is likely due to the age of the participants, as students tend to be “young and healthy”. They are unlikely to have been exposed to regular use of chronic medications which are usually prescribed for mainly aged-related illnesses. The buccal swabs gave good yields of DNA for downstream analyses. The DNA concentrations varied in a range from sample 3B1001B (35.3ng/µl) to sample 3B (383.5 ng/µl) and the variation could reflect different numbers of cells obtained in the process of collecting the buccal samples.

2.4.2.2 *Genotyping & Sequencing of rs28371706 as a Proxy for CYP2D6*17*

The genotyping was performed using a commercially available TaqMan genotyping kit for rs28371706, as a proxy for *CYP2D6*17* since it is the defining SNP locus in the haplotype (Gaedigk *et al.*, 2015). The presence of the T allele defines the *CYP2D6*17* allele, and the C

is present in the reference sequence. The rs28371706 T allele is also part of the haplotypes for the *CYP2D6*40, CYP2D6*58, CYP2D6*64* and *CYP2D6*82* alleles, and possibilities of false positives were considered. However, these star alleles are found in much lower frequencies in African populations and it was concluded that they would be less relevant, and were therefore not included in the genotyping experiment (Gaedigk *et al.*, 2015; Dodgen *et al.*, 2016). The use of two different methods (*Alternative Method 1 and 2*) to optimise the TaqMan PCR illustrated the difficulties of optimising commercial kits using different reagents rather than using all the recommended reagents supplied with the kit. *Alternative Method 1 and 2* did not work optimally as the master mix components did not successfully bind to the TaqMan™ MGB probes. After extensive trouble shooting, the Thermo Fisher™ approved reagents were used, the assay was optimised and the TaqMan PCR assay was successful.

The TaqMan genotyping results showed that the standard Thermo Fisher™ based method was able to determine the rs28371706 genotypes in all the DNA samples. The four test DNA samples all had the C/C genotype while among the study participants who were genotyped, five were T/T, four were C/C and one was C/T. After confirmation using Sanger sequencing, four of the genotypes were not validated and the genotype calling was revised as follows: 1B2016: T/T→A/C; 2A1004: C/T→C/C; 3B1011: T/T→C/T; 3B3013: T/T→C/T. There was no apparent pattern characterised by the changes. The discrepancies may be due to the small number of samples that were included in the TaqMan PCR assay, which would have affected the ability of the software to accurately determine the genotypes. It is known that larger sample sizes improve the accuracy of the genotype clustering and calling. The addition of more DNA samples, approximating 50, and a rerun of the TaqMan PCR assay may have resolved this issue and resulted in better allelic discrimination.

Although the rs28371706 genotyping assay had been validated by the commercial company, there may also be other reasons for the errors in genotyping. The failure of the assay may have been related to interference from two pseudogenes, *CYP2D7* and *CYP2D8*, with similar sequences found in humans (Endrizzi *et al.*, 2002; Gaedigk *et al.*, 2015). This may have been resolved by sequencing the entire gene target areas of the *CYP2D6*, *CYP2D7* and *CYP2D8* genes specifically to confirm the sequence used for the PCR and TaqMan probes were specific for the *CYP2D6* gene, and would not bind to the sequences of the other genes (Endrizzi *et al.*, 2002; Gaedigk *et al.*, 2015; Gaedigk, Turner, *et al.*, 2019). It is, however, likely that the commercial company was aware of this and would have taken it into consideration when designing the assay.

Following sequencing, one sample (1B2016) was shown to have the A/C genotype, which was a rare find as it has a genotype frequency of 0.016% in African populations (Wang *et al.*, 2020). It would be prudent to investigate this further to confirm the validity of this sequencing result, as interference from pseudogenes during the PCR amplification process may have affected the result. This variant causes an amino acid change from threonine to asparagine (T107N) at this position (Wang *et al.*, 2020). Since Sanger sequencing is the gold standard for genotyping, the result from the sequencing was accepted as being more accurate and the selection of participants for the skin biopsy was performed using these results. The following participants were invited back (1A2002, 3A1007, 3A2008, 3B4014) with one declining due to unavailability (3A2008) and the others (1A2002, 3A1007, 3B4014) accepting to undergo a skin biopsy to donate skin samples.

2.4.3 Strengths & Limitations Related to this Chapter's Work

This chapter highlighted numerous methods on how to attain data on the prevalence of *CYP2D6*17* and other star alleles found in this gene. It also highlighted the process of identifying individuals that are homozygous for *CYP2D6*17*. However, there were some limitations for this study. These limitations are the following:

1. There were some concerns over the accuracy of the haplotypes and diplotypes that were generated using Astrolabe. This was due to the BAM file format required for sequenced data for analysis with Astrolabe, and the AGVP data provided did not have BAM files. The BAM files are used as a reference for the Astrolabe program to allow it to check the minimum coverage of the sequenced data (Twist *et al.*, 2016). However, Astrolabe was able to generate data that predicted haplotypes and diplotypes present in the AGVP data.

2. The study only looked at data available from the populations of interest in the AGVP, KGP and PharmVar datasets. The prevalence of *CYP2D6* reduced function variants in other African populations not a part of these datasets, cannot be obtained using data from these studies.

3. The use of rs28371706 genotyping as a proxy for *CYP2D6*17* status may have given false positives due to the SNP allele also defining the *CYP2D6*40*, *CYP2D6*58*, *CYP2D6*64* and *CYP2D6*82* haplotypes. However, due to their low frequencies in the African populations it was decided that they would be negligible for this study. An

alternative to mitigate this problem would have been to perform a long-range PCR and sequencing the PCR product to attain the entire *CYP2D6* sequence. This may result in more accurate genotyping of the study participants as *CYP2D6*17*.

4. The presence of *CYP2D7* and *CYP2D8* pseudogenes may have interfered with the genotyping of rs28371706 and resulted in false positives. This is due to these pseudogenes sharing regions of homology with the *CYP2D6* gene. This issue could have been investigated by performing sequencing or an RFLP assay for these genes to rule out any possible interference from those genes. Steps were however taken to ensure that the primers used in the assay were very specific the rs28371706 locus in *CYP2D6* (Gaedigk *et al.*, 2015).

The findings of this study showed many strengths as well and provided insight into the *CYP2D6*17* variant and its presence in African populations. These strengths were the following:

1. The findings of this study support the assumptions of the first and second Bantu migration events that took place on the African continent. The frequencies of the *CYP2D6* variants found in this study seem to correlate with the movements of people from West to East Africa, and from East Africa to southern Africa. Further research on the evolutionary pressures exerted on the *CYP2D6* gene may give more insight into the effect of migration and admixture on the prevalence of these *CYP2D6* variants.

2. The rs29371706 genotyping identified individuals homozygous for *CYP2D6*17*. The skin biopsies from these individuals would make it possible to establish *CYP2D6*17* SC-HLCs that could be used for modelling the effects of this allele on drug metabolism. This would make these cells a rather valuable asset in investigating these effects and might aid in affirming and revealing knowledge about this variant.

2.5 CONCLUSION

This chapter describes the use of public genomic data to understand the distribution of *CYP2D6* reduced function variants in African populations. It also describes the process of identifying an individual who was homozygous for *CYP2D6*17*. The overall aim of the study was to develop a *CYP2D6*17* homozygous SC-HLC model to test the function of the *CYP2D6 *17* isozyme on various compounds, and so successfully identified one *CYP2D6*17* homozygote was critical for the downstream experiments. The procedures following the skin biopsies and the establishment of a SC-HLC model from the dermal fibroblasts isolated from the biopsies will be described in Chapter 3.

CHAPTER 3

CELLULAR ASSAYS & EXPERIMENTS FOR DRUG METABOLISM MODEL

3. CHAPTER 3: CELLULAR ASSAYS & EXPERIMENTS

3.1 BRIEF INTRODUCTION

Induced pluripotent stem cells (iPSCs) are derived from somatic cells but are often derived from dermal fibroblasts and are able to differentiate into various cell types (Takahashi *et al.*, 2007). This makes iPSCs of great interest when looking to model disease and create liver-like tissue to investigate the mechanism involved in drug metabolism and response, as well as, understanding drug toxicity. Beyond creating liver-like tissue for drug metabolism studies, iPSCs also have the potential for studying polymorphisms effects in the CYP genes via gene editing. This also allows iPSCs to be utilised for studying gene-gene and gene-environment interactions. All of these factors make iPSCs a good candidate for modelling disease, drug development and gene effects on drug metabolism (Sullivan *et al.*, 2010). IPSCs differentiated into liver-like tissue are termed stem-cell-derived hepatocyte-like cells (SC-HLCs) (Goldring *et al.*, 2017). SC-HLC models can be used to perform enzymatic activity studies to investigate the effects of the *CYPD6*17* variants on the functionality of the enzyme.

This chapter outlines the cellular aspects of this study in achieving objectives two and three, which are:

- Objective Two (2): *To develop the SC-HLC model.*
- Objective Three (3): *To measure enzyme activity of CYP2D6 and compare the effects of CYP2D6*17 to the reference CYP2D6*1 variant in the conversion of N-desmethyltamoxifen to endoxifen using the SC-HLCs.*

In this study, a 3D SC-HLC model was to be formed using iPSCs derived from dermal fibroblasts acquired from study participants skin biopsies. This chapter reviews the skin biopsy procedure, isolation and culturing of donor dermal fibroblast, dermal fibroblast reprogramming into iPSCs, culturing of the iPSCs, HepG2 cell culturing, HepG2 and SC-HLC spheroid formation, and enzyme functionality assays on HepG2 and SC-HLC spheroids. Sections of this study concerned with iPSC characterisation, SC-HLC differentiation and spheroid formation, CYP2D6 enzyme activity assays, and drug metabolism assays were removed due to time constraints caused by the COVID-19 pandemic. Optimisations using HepG2 cells were utilised to demonstrate a proof of concept and technical ability. CYP3A4 (not CYP2D6) was the enzyme used to test a proof of concept of the model. Rifampicin was used as a test compound for CYP3A4 expressed in HepG2 cell spheroids to show the impact of 3D culture on CYP enzyme expression. These results cannot be used to imply any drug metabolism effects of CYP2D6*17 variants on the CYP3D6 enzyme (Fig 3.1).

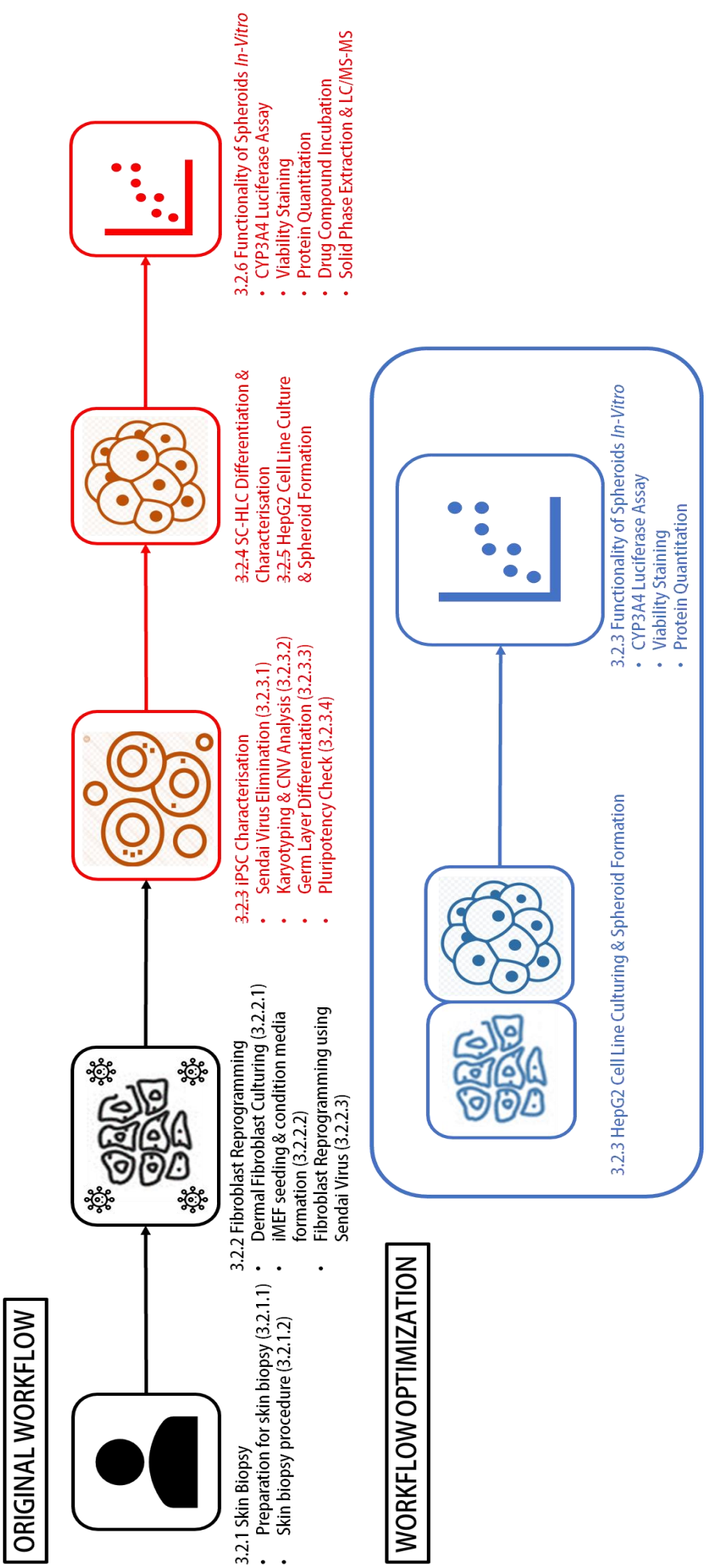


Fig 3.1: Workflow of experiments and assays utilised to achieve objectives 2 and 3. Steps in red indicate experiments/assays removed from the study to due shortened time in the CSIR laboratory. Steps circled in blue indicate experiments performed to generate results for a "proof of concept and technical ability".

3.2 METHODS

Experiments were conducted in a biosafety level II (BSL-2) tissue culture laboratory at the CSIR. Cells were cultured using sterile techniques and under class II, Labconco® type A2 biological safety cabinets (Kansas City, MO), and maintained within water-jacketed Thermo Fisher Scientific Forma™ Direct Heat CO2 incubators (Waltham, MA) at standard incubator conditions of 37°C and 5% CO_2 incubation.

3.2.1 Skin Biopsy & Dermal Fibroblast Culturing

3.2.1.1 *Preparations for Skin Biopsy*

An antibiotic and antimycotic culture solution for transportation of the samples were prepared in a 15 ml centrifuge tube containing Thermo Fisher Scientific Gibco® phenol red-free Dulbecco's Modified Eagle's Medium (DMEM; Waltham, MA) with 10% heat-inactivated Thermo Fisher Scientific Gibco® foetal bovine serum (FBS; Waltham, MA), 1 µg/ml of amphotericin-B (Amp-B), and 10 µg/ml penicillin/streptomycin (PenStrep) at for each of the three samples taken (Vangipuram *et al.*, 2013). Sterile 10 cm culture dishes were also used for culturing and isolation of dermal fibroblast from each sample skin biopsy.

3.2.1.2 *Skin Biopsy Procedure*

The skin biopsy procedure was performed at the Dermatology Outpatients Department (OPD) at Charlotte Maxeke Johannesburg Academic Hospital (CMJAH). A qualified dermatologist performed the skin biopsies with a 4 mm round punch instrument which were kept in the 15 ml centrifuge tubes containing culture media for transportation. Skin biopsies were placed into a 10 cm culture plate with culture media under the biological safety cabinets using sterile forceps. The skin biopsies were dissected into smaller pieces approximately 2 mm x 2 mm in size using sterile sharp scalpels. These skin biopsy pieces were washed in antibiotic and antimycotic culture media and placed on 100-200 µl droplets of antibiotic and antimycotic media in new sterile 10 cm culture plates. Each droplet containing a dissected skin sample piece was covered with a glass slip to limit movement, allow for cell attachment and reduced evaporation of liquid. Dermal fibroblasts were grown in culture and once fibroblasts were isolated from the skin biopsy samples, they were transferred to 6-well plates with 2 ml of antibiotic and antimycotic culture media. These were incubated at 37°C for a week and media was replaced every 48 hours adding 2 ml of antibiotic and antimycotic culture media. After one week, 2 ml of DMEM with 10% FBS, Amp-B and PenStrep was added to each well and replaced every 48-72 hours.

3.2.1.3 *Dermal Fibroblast Culturing*

Dermal fibroblasts were isolated from skin biopsies of donors (1A2002, 3A1007 and 3B4014) and cultured in TPP® T25, T75 and T150 (Trasadingen, Switzerland) flasks. Fibroblasts were cultured in DMEM media containing 10% FBS, and 2 mM GlutaMax (Thermo Fisher Scientific, MA). Fibroblasts were counted using a haemocytometer after staining with trypan blue dye. Cells from each well were then transferred to TPP® T75 (Trasadingen, Switzerland) flasks for fibroblast reprogramming into iPSCs for liver differentiation (Vangipuram *et al.*, 2013). Cells were passaged seven times after being cultured to 70-80% confluence. Cells were then washed in 1X Sigma-Aldrich Phosphate Buffer Saline (PBS) (St. Louis, MO) and dissociation reagent Thermo Fisher Scientific Gibco® TrypLE™ Express Enzyme (Waltham, MA) was added to detach the cells from the flask or well surface. Cells flasks or wells were incubated at 37°C for approximately 5-10 minutes followed by adding them to 15 ml centrifuge tubes and centrifuged at 300 x g for 4 minutes (Thermo Fisher Scientific, 2016a). Supernatant was discarded and cells were resuspended in DMEM culture media then added to culture flasks.

3.2.1.4 *Haemocytometer Cell Counting*

All cells were counted using a haemocytometer and trypan blue staining and observed under an Olympus CKX41 Inverted Fluorescence Microscope (Tokyo, Japan). Culture media was removed from culture plates or flasks and discarded. Cells were then washed in 1 x PBS (Sigma-Aldrich, MO) and dissociation reagent TrypLE™ Express Enzyme (Thermo Fisher Scientific, MA) was added to detach the cells from the flask or well surface. Supernatant was discarded and 2 ml of culture media was added to the tubes and resuspended in media solution for counting. Approximately 100 µl of media solution contain the cells was added to 1.5 ml microcentrifuge tubes containing 400 µl of trypan blue. Trypan blue and cell solution was mixed gently then approximately 10 µl of this solution was added to each chamber of a haemocytometer under a glass coverslip. The haemocytometer was placed under an inverted fluorescence microscope and an average cell count was taken using trypan blue (Fig 3.2) (Abcam, 2015).

$$\text{Cell Conc. (cell/ml)} = \text{Average Cell Count} \times 10^4 \times 5 \text{ (dilution factor)}$$

Fig 3.2: Calculation for cell concentration using a haemocytometer and trypan blue dye.

3.2.1.5 *Cryopreservation & Thawing of Cells*

Freezing medium was freshly prepared and consisted of DMEM, 20% FBS and 10% DMSO which was then stored at 4°C. Cells were then washed with 1 x PBS (Sigma-Aldrich, MO) and dissociation reagent TrypLE™ Express Enzyme (Thermo Fisher Scientific, MA) was added to detach the cells from the flask or well surface. Cells flasks or wells were incubated at 37°C for approximately 5-10 minutes followed by placing them to 15 ml centrifuge tubes and centrifuged at 300 x g for four minutes (Thermo Fisher Scientific, 2016a). Supernatant was discarded and 2 ml of culture media was added to the tubes and resuspended in the solution. Cells were counted and resuspended in freezing media and added to 2 ml cryovials (Kremsmünster, Austria) at a cell density of $2x10^6$ cells/ml. Cryovials were stored at -80°C for 24 hours and transferred to liquid nitrogen for more permanent storage (Thermo Fisher Scientific, 2016a).

Thawing of cells was performed by removing cryovials containing cells of interest from liquid nitrogen storage and immediately incubated at 37°C until a small piece of ice was left in the cryovial. Cells were then transferred dropwise to 15 ml centrifuge tubes containing prewarmed culture media containing DMEM, 10% FBS and 2 mM of GlutaMax (Thermo Fisher Scientific, MA). Cells were centrifuged at 300 x g for four minutes and the supernatant was discarded with the remaining cell pellet resuspended in culture media. Solution with cells was transferred to culture flasks and incubated at 37°C until needed (Thermo Fisher Scientific, 2016a).

3.2.2 Fibroblast Reprogramming

3.2.2.1 *iMEF seeding & Conditioned Media*

Inactivated mouse embryonic fibroblasts (iMEFs) were cultured to harvest secreted growth factors needed to supplement the iPSC reprogramming of dermal fibroblasts. iMEFs were thawed and cultured in TPP® T75 flasks (Trasadingen, Switzerland) at a seeding density of 4-$5.33x10^4$ cells/cm^2. These were cultured with media consisting of DMEM containing 10% FBS, 2 mM GlutaMax (Thermo Fisher Scientific, MA) (Fig 3.3). Conditioned media was prepared to replace the iMEFs on day 14 of reprogramming as iMEFs stop producing growth factors necessary for stem cell reprogramming and clone growth. Conditioned media was prepared by harvesting 25 ml of human embryonic stem cell (huESC) media, which consisted of Thermo Fisher Scientific Gibco® DMEM/F12 (Waltham, MA), KnockOut™ Serum Replacement (Waltham, MA), 10 mM MEM Non-Essential Amino Acids Solution (Waltham, MA), 55 mM β-mercaptoethanol (Waltham, MA) and 10 µg/ml PenStrep. Conditioned media was collected every 24 hours for seven days consecutively and stored in a fridge at 4°C. Fresh

10 µg/ml bFGF (Thermo Fisher Scientific, MA) was added to the huESC media before supplementing the iPSCs in culture with iMEFs (Thermo Fisher Scientific, 2016b). For iMEF source and inactivation see Appendix D.

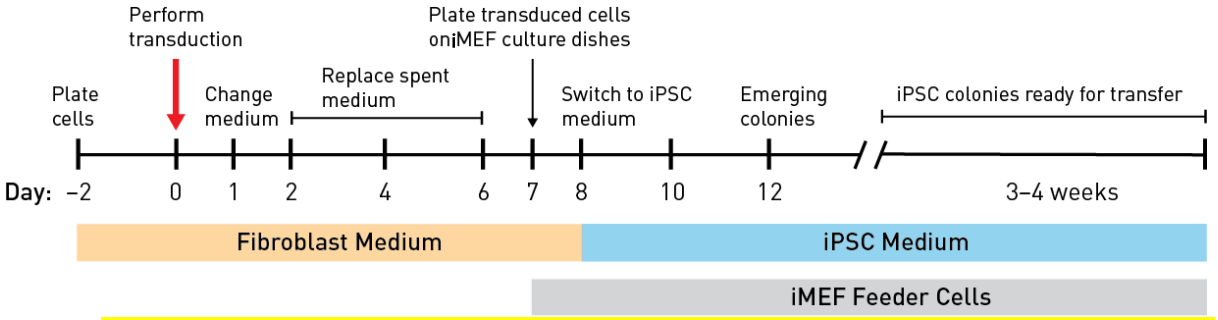


Fig 3.3: Experimental timeline for the reprogramming of the dermal fibroblasts using the CytoTune™-iPS 2.0 Sendai Reprogramming Kit. Red arrow indicated the day of transduction via a Sendai virus infection, and the black arrow indicates the day which the transduced fibroblast cells were transferred to an iMEF feeder for reprogramming. *Image from: Invitrogen CytoTune™-iPS 2.0 Sendai Reprogramming Kit User Guide (2016).*

3.2.2.2 *Fibroblast Reprogramming Using Sendai Virus*

Dermal fibroblasts underwent reprogramming to become iPSCs by transduction using the Thermo Fisher Scientific CytoTune™-iPS 2.0 Sendai Reprogramming Kit (Waltham, MA with a Sendai virus (SeV) vectors containing the four Yamanaka factors (Oct, Sox2, Klf4, and c-Myc). iPSC cells were derived from donor (1A2002, 3A1007 and 3B4014) dermal fibroblasts, which were reprogrammed by members of the CSIR laboratory using the CytoTune™-iPS 2.0 Sendai Reprogramming Kit (Thermo Fisher Scientific, MA). Cells were cultured in TPP® 6-well plate (Trasadingen, Switzerland), consisting of iMEF Feeder layers, at two seeding densities ($2x10^4$ and $5x10^4$ cells/cm^2). The iPSCs were cultured in huESC media as outlined above.

The CytoTune™-iPS 2.0 Sendai Reprogramming Kit consists of three SeV-based vectors containing the factors for reprogramming of somatic cells in human into iPSCs. The CytoTune™ 2.0 KOS Sendai vector (Thermo Fisher Scientific, MA comprises of human factors Klf4, Oct3/4 and Sox2 (Thermo Fisher Scientific, 2016b). The CytoTune™ 2.0 hc-Myc Sendai vector (Thermo Fisher Scientific, MA) contains human factors c-Myc. The CytoTune™ 2.0 hKlf4 Sendai vector (Thermo Fisher Scientific, MA) contains human factors Klf4. The viral vectors used in this kit are episomal and remain in the cytoplasm using the cell's RNA replicating machinery to multiply (Thermo Fisher Scientific, 2016b). An initial viral titre was calculated for first reprograming SeV transduction attempt using a multiplicity of infection (MOI) of 5:5:3 (KOS = 5; hc-Myc = 5; hKlf4 = 3) for an infection of $2x10^4$cells/well in three wells of a 6-well plate, with the other three being negative controls not undergoing an infection (Fig 3.4). Two more reprogramming attempts were tried using an MOI of 10:10:6 (KOS = 10; hc-Myc = 10; hKlf4 = 6) (Fig 3.4). These were added to the

fibroblast culture media which consisted of DMEM (Thermo Fisher Scientific, MA) with 10% FBS (Thermo Fisher Scientific, MA), 2 mM of GlutaMax (Thermo Fisher Scientific, MA and 10 µg/ml PenStrep.

The reprogramming took place over a period of approximately 5-6 weeks (Fig 3.3) with the initial step of plating of dermal fibroblasts (three SeV transduced wells and three negative control wells at seeding density of $2.13x10^3$ cells/cm^2 per well) in TPP® 6-well plate (Trasadingen, Switzerland) coated with iMEFs seeded at 4-$5.33x10^4$ cells/cm^2 on day -2. Cells had a passage number of four (Black African Con fibroblast line) and five (each donor skin biopsy sample). Cells were cultured in media

INITIAL TRANSDUCTION CALCULATIONS

Volume of KOS Viral Vector (µl) = $\frac{\text{MOI (CIU/cell) x Number of Cells}}{\text{Titre of Virus (CIU/ml) x } 10^{-3} \text{ (ml/µl)}}$

= $\frac{\text{5 CIU/cell x } 2x10^4 \text{ cells}}{1.1x10^8 \text{ CIU/ml x } 10^{-3} \text{ (ml/µl)}}$

= **0.9 µl**

Volume of hc-Myc Viral Vector (µl) = $\frac{\text{5 (CIU/cell) x Number of Cells}}{\text{Titre of Virus (CIU/ml) x } 10^{-3} \text{ (ml/µl)}}$

= $\frac{\text{5 CIU/cell x } 2x10^4 \text{ cells}}{1.0x10^8 \text{ CIU/ml x } 10^{-3} \text{ (ml/µl)}}$

= **1.0 µl**

Volume of hKlf4 Viral Vector (µl) = $\frac{\text{MOI (CIU/cell) x Number of Cells}}{\text{Titre of Virus (CIU/ml) x } 10^{-3} \text{ (ml/µl)}}$

= $\frac{\text{3 CIU/cell x } 2x10^4 \text{ cells}}{9.3x10^7 \text{ CIU/ml x } 10^{-3} \text{ (ml/µl)}}$

= **0.625 µl**

REVISED TRANSDUCTION CALCULATION

Volume of KOS Viral Vector (µl) = $\frac{\text{MOI (CIU/cell) x Number of Cells}}{\text{Titre of Virus (CIU/ml) x } 10^{-3} \text{ (ml/µl)}}$

= $\frac{\text{10 CIU/cell x } 2x10^4 \text{ cells}}{1.1x10^8 \text{ CIU/ml x } 10^{-3} \text{ (ml/µl)}}$

= **1.8 µl**

Volume of hc-Myc Viral Vector (µl) = $\frac{\text{MOI (CIU/cell) x Number of Cells}}{\text{Titre of Virus (CIU/ml) x } 10^{-3} \text{ (ml/µl)}}$

= $\frac{\text{10 CIU/cell x } 2x10^4 \text{ cells}}{1.0x10^8 \text{ CIU/ml x } 10^{-3} \text{ (ml/µl)}}$

= **2.0 µl**

Volume of hKlf4 Viral Vector (µl) = $\frac{\text{MOI (CIU/cell) x Number of Cells}}{\text{Titre of Virus (CIU/ml) x } 10^{-3} \text{ (ml/µl)}}$

= $\frac{\text{6 CIU/cell x } 2x10^4 \text{ cells}}{9.3x10^7 \text{ CIU/ml x } 10^{-3} \text{ (ml/µl)}}$

= **1.25 µl**

Fig 3.4: Calculations for viral vector volumes for transduction of dermal fibroblasts during iPSC reprogramming. *Image from: Invitrogen CytoTune™-iPS 2.0 Sendai Reprogramming Kit User Guide (2016).*

consisting of DMEM with 10% FBS, 2 mM GlutaMax (Thermo Fisher Scientific, MA), and 10 µg/ml PenStrep until day 0 (the day of transductions). On day 0, culture media containing the SeV vectors was added to each well according to the MOI for each cell concentration and incubated overnight at 37°C.

On day one, culture media containing SeV vectors was removed and fresh dermal fibroblast media without SeV vectors was added into each well, with media replaced every two days from day two to six. On day six, iMEFs were thawed and added into the wells of four 6-well plates (TPP®, Switzerland) containing dermal fibroblast media, and incubated overnight to allow iMEFs to attach to the plate. On day seven, transduced/infected SeV fibroblasts were added to iMEF plates at two seeding densities (3 wells at $2.13x10^3$ cells/cm^2 and 3 wells at $5.32x10^3$ cells/cm^2) and incubated overnight to allow for attachment of the fibroblasts. On day eight, the reprogramming fibroblasts were changed to huESC medium. From days nine to 13, reprogramming fibroblasts were culture on iMEF feeders with huESC media until the iMEFs stop secreting growth factors (day 13). From day 12 onwards, plates were observed for the emergence of iPSC colonies. Once the colonies were identified, they were cultured by changing media every day until they were ready for clonal transfer to new plates for expansion and characterisation.

3.2.3 HepG2 Cell Line Culture & Spheroid Formation

The HepG2 cell line was used for the optimisation of experiments and assays necessary for this study i.e. the optimisation of 3D spheroid cultures and CYP3A4 luciferase assay. HepG2 cells were thawed at 37°C and cultured for one week in a TPP® T75 flask (Trasadingen, Switzerland), containing Gibco® DMEM Media (Thermo Fisher Scientific, MA) with 10% FBS and Gibco® GlutaMax (Thermo Fisher Scientific, MA) and transferred to a TPP® 12-well plate (Trasadingen, Switzerland) for 18-21 days (*Experiments 1 & 2*) and 10-16 days (*Experiment 3*) with culture media to aggregate into spheroids. Sigma-Aldrich 3D Petri-Dish® Micromolds (St. Louis, MO) were used to cast molds for aggregating HepG2 cells into spheroids. Molds were cast using sterile 2% Sigma-Aldrich Agarose (St. Louis, MO) and 1 x PBS (Sigma-Aldrich, MO) gel, which were further sterilised for 15 minutes multiple times via a 254 nm UV light. HepG2 cells were then aggregated into spheroids in the micro-moulds at $2x10^3$ cell per spheroid. Media for each well was changed every 48 hours by performing a 50% change. On day 10 (*Experiment 3*) and day 18 (*Experiments 1 & 2*), all spheroids were used in different experiments testing for CYP3A4 enzyme expression and spheroid viability.

3.2.4 Functionality of Spheroids *In-Vitro*

3.2.4.1 *CYP3A4 Luciferase Assay*

A luciferase assay was utilised to investigating the expression of the CYP enzymes in our SC-HLC model. HepG2 spheroids were used to optimise this assay and were cultured for 10 days (*Experiment 3*) and 18 days (*Experiments 1 & 2*) in 12-well plate (TPP®, Switzerland). *Experiments 1* and *2* were part of the assay optimisation, while *Experiment 3* was performed

to illustrate the proof of concept and technical ability. CYP3A4 induction was performed by incubating the spheroids with rifampicin at 2 µM, 2.5 µM, 20 µM, 25 µM and no rifampicin for 24, 48 and 72 hours. The spheroids were harvested at each time point then centrifuged for 10 seconds at 300 x g in 1.5 ml microcentrifuge tubes. The supernatant was removed keeping spheroids within the tubes and were washed with 1 x PBS. DMEM media containing 10% FBS, 2 mM GlutaMax (Thermo Fisher Scientific, MA) and 10 µM Promega P450-Glo™ Luciferin-IPA Enzyme (Madison, USA) (Fig 3.5) was added to each tube. From each tube, 100 µl of the sample was added to a TPP® 96-well plate (Trasadingen, Switzerland) with one blank sample containing only media with luciferin enzyme and were incubated at 37°C for 24 hours.

The samples were transferred to a new white 96-well plate in duplicates of 30 µl each well, and 30 µl of Promega P450-Glo™ Luciferin Detection Buffer (Madison, WI) was added to the samples then incubated at room temperature for 20 min. Duplicates were defined as technical duplicates where sample solution measurements were the same in each well for each sample. The white 96-well plate was wrapped in aluminium foil to prevent light exposure to the light sensitive luciferin enzyme and substrates. Luminescence signals for CYP3A4 in both rifampicin induced and no-rifampicin control samples were recorded on a Tecan Infinite™ F500 Microplate Reader (Männedorf, Switzerland). After luminescence signals were recorded, the spheroids were harvested in 1.5 ml centrifuge tubes for protein quantitation.

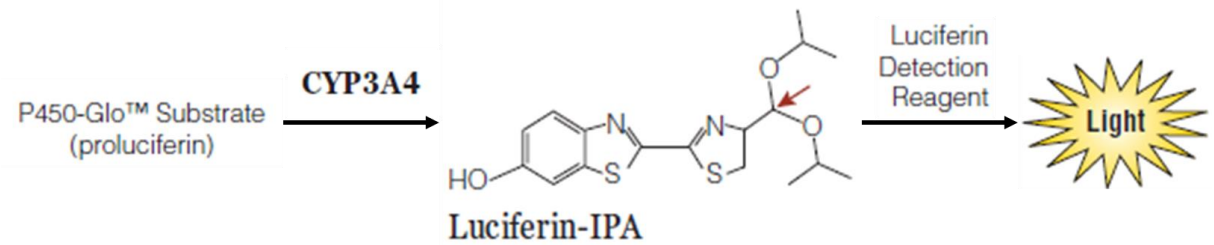


Fig 3.5: Conversion of P450-Glo™ substrate to luciferin by cytochrome P450 and detection mechanism. *Image from Promega (2016).*

3.2.4.2 *Protein Quantitation*

Protein quantitation is an important step for measuring the amount of protein in a particular sample to normalise luminescence data. A bicinchoninic acid (BCA) protein analysis assay was used to quantitate the total protein found in the HepG2 spheroids. BCA protein analysis works by utilising colorimetric detection using Cu^{2+} reduction and a BCA reagent. Cu^{2+} reduction is performed by the protein in an alkaline medium and is reduced to Cu^{+}, which when combined with BCA forms a selective and sensitive colorimetric detection assay (Smith *et al.*, 1985). A colour reaction is formed when these are combined, and a purple colour is observed when two BCA molecules chelate with the Cu^{+} ion. This colour reaction is due to

the protein macromolecular structure, number of peptide bonds and four amino acids (cysteine, cystine, tryptophan and tyrosine) (Walker and Walker, 2003).

HepG2 spheroids were lysed using radio-immunoprecipitation assay (RIPA) lysis buffer to extract the protein for analysis. The RIPA buffer consists of 25 mM Tris-HCL (pH 7.6), 150 mM NaCl, 1% Nonidet P-40, 1% sodium deoxycholate, 0.1% sodium dodecyl sulphate, (all Sigma-Aldrich; St. Louis, MO), made up to 100 ml with ddH_2O (Table 3.1). Before lysing spheroid cells, 0.08% protease inhibitor was added to the buffer mix and all reagents were stirred using a magnetic flea. 100 µl of cold RIPA buffer was added to each 1.5 ml microcentrifuge tube as it was estimated that $3.89x10^6$ – $4.86x10^6$ cells were harvested from the spheroids. Samples were stored on ice for 5 minutes, vortexed and centrifuged at 14 000 x g for 15 minutes at 4°C. The supernatant was then transferred to new 1.5 ml microcentrifuge tubes for further BCA analysis preparations.

Table 3.1: Reagents and measurements for the RIPA buffer for cell lysis.

Reagent	Final Volume	Final Concentration
Tris-HCL (1M, pH 7.6)	2.5 ml	25 mM
NaCl (sodium chloride)	0.876g or 3 ml of 5M solution	150 mM
Nonidet P-40	1 ml	1%
Sodium deoxycholate (10%)	10 ml	1 %
SDS (10%)	1 ml	0.1 %
ddH_2O	85.5 or 82.5 ml	–
Protease Inhibitors	80 µl	0.08%
TOTAL	100 ml	

Table 3.2: Dilutions and concentrations for BSA standards.

Sample	Volume of Diluent (µl)	Volume of BSA (µl)	Final BSA Concentration
A	80	20	2 000 µg/ml
B	85	15	1 500 µg /ml
C	90	10	1 000 µg /ml
D	92.5	7.5	750 µg /ml
E	95	5	500 µg /ml
F	97.5	2.5	250 µg /mg
G	98.75	1.25	125 µg /ml
H	99.75	0.25	25 µg /ml
Blank	-	-	-

For the BCA procedure, eight dilutions of bovine serum albumin (BSA) standard were prepared with 1 x PBS and a 10 mg/ml BSA stock solution (Table 3.2), and dilutions for the samples of interest/unknown samples were also prepared in 1 x PBS. A BCA working reagent was also prepared by preparing two reagent mixes (Reagent A and B) and adding them together at a ratio of 50:1 (Reagent A:Reagent B). Reagent A consisted of 1% sodium bicinchoninate (BCA), 10 mM sodium carbonate, 200 mM NaOH, 5 mM sodium bicarbonate,

and 12 ml with dH_2O, which was then adjusted to a pH of 11.25 using 10 M NaOH. Reagent B consisted of 4% cupric sulphate in 10 ml of dH_2O. The 10 µl of the BSA standard dilutions and 10 µl of the unknown samples were added together into a TPP® 96-well plate (Trasadingen, Switzerland) in duplicates. Two wells were selected to be blanks, one for the standard curve analysis which contained 10 µl of 1 x PBS in each well. In each well, 200 µl of BCA working reagent was added and mixed thoroughly, then the 96-well plate was incubated at 60°C for 30 minutes and placed to cool down to room temperature. Measurements for absorbance at 450 and 620 nm were recorded on a Tecan Infinite™ F500 Microplate Reader (Männedorf, Switzerland) for the total protein quantity.

3.2.4.3 *Viability Staining*

Spheroids were assessed for their viability after 16 days (*Experiment 3*) or 21 days (*Experiments 1 & 2*) days in culture using a fluorescein diacetate (FDA) and propidium iodide (PI) live/dead assay. FDA is used to measure the number of cells that are viable and live in culture. This is due to FDA (which is non-fluorescent) traversing the cell membrane and being converted to its metabolite fluorescein which fluoresces green. The PI is used to measure the number of dead cells in a culture. This is due to PI inability to be taken up by cell with an intact cell membrane. When there is damage to the integrity of the cell membrane, PI enters the cell and intercalates with the DNA double helix and fluoresces red (Jones and Senft, 1985; Ibidi, 2015).

Spheroids were cultured for 16 days (*Experiment 3*) or 21 days (*Experiments 1 & 2*) in a 12-well plate (TPP®, Switzerland) after which media from wells for staining was removed. Spheroids from 2 wells were harvested and transferred to 1.5 ml microcentrifuge tubes and washed in 1 ml of 1 x PBS per tube. A staining solution was prepared consisting of Gibco® DMEM (Thermo Fisher Scientific, MA), 2 mM of Gibco® GlutaMax (Thermo Fisher Scientific, MA), FDA in 1 x PBS solution at 5 mg/ml, and PI in 100% methanol at 2 mg/µl. The plate was covered with aluminium foil to prevent light interference and incubated at 37°C for five minutes. The staining solution was removed, and the cells were washed with 1 x PBS, then culture media was added to the wells of the plate. The spheroids were analysed for cell viability under a CKX41 Inverted Fluorescence Microscope (Olympus, Japan). This was a qualitative analysis of the spheroids and was performed on approximately three to five from two spheroid gel molds HepG2 spheroids.

3.3 RESULTS

3.3.1 Skin Biopsy & Dermal Fibroblast Culturing from homozygote rs28371706 T & C Allele Individuals

3.3.1.1 *Skin Biopsy Procedure*

Genotyping of CYP2D6*17 proxy rs28371706 identified two individuals that were homozygous for the T allele (3A2008 and 3B4014) and two individuals homozygous for the C allele (1A2002 and 3A1007) for skin biopsies. However, one individual (3A2008) was unable to undergo a skin biopsy and withdrew their participation in the study.

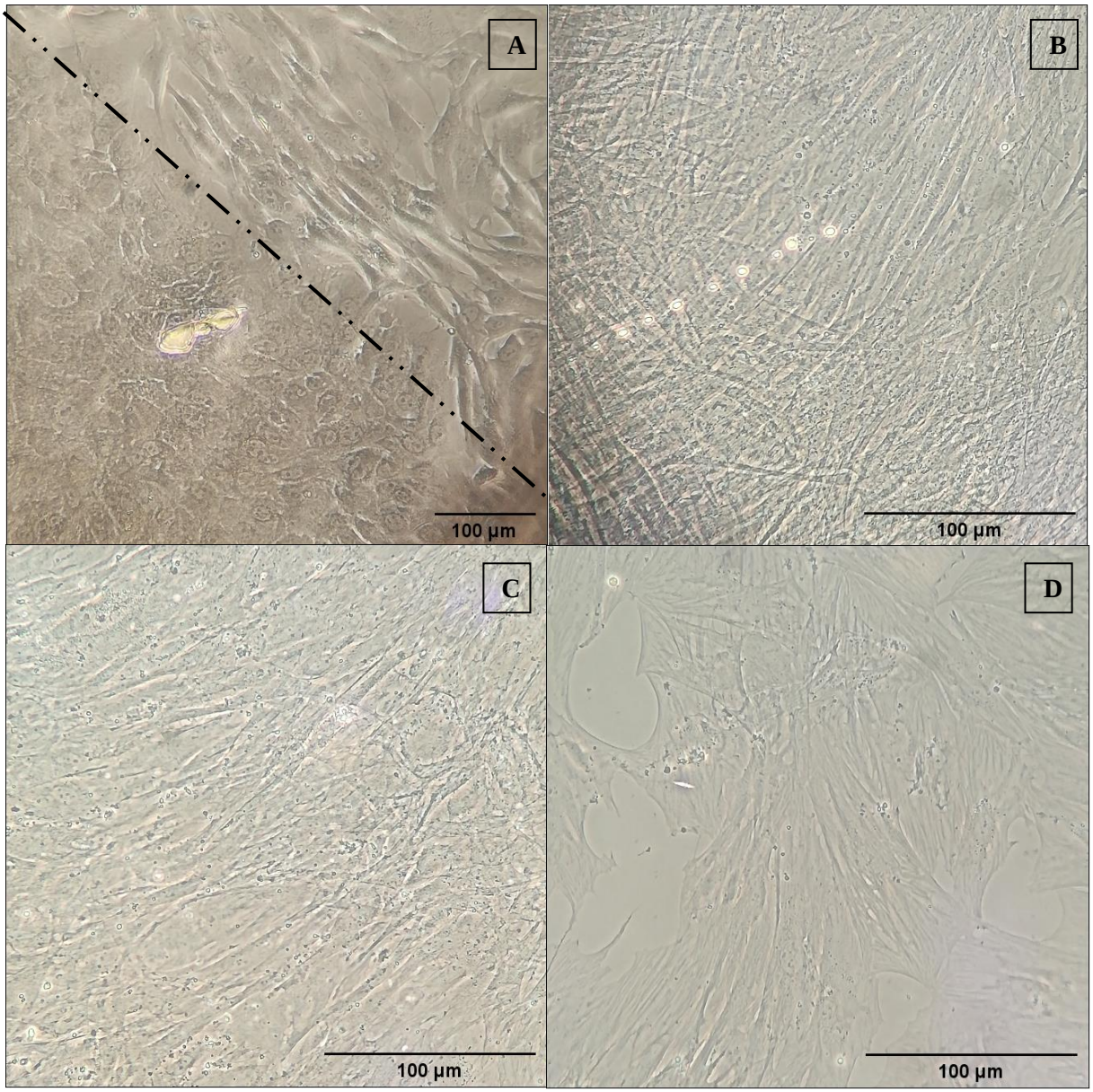


Fig 3.6: Phase-contrast images dermal fibroblast isolation process from a skin biopsy. Images were obtained at 20x objectives and cells were grown as a monolayer in a 37°C incubator with 5% CO_2 in air. **(A)** Transition barrier has formed between keratinocytes and dermal fibroblasts in our 3B4014 donor line which has a genotype of T/T at the rs28371706 for CYP2D6*17. Line separate the keratinocyte (bottom left) and dermal fibroblast (top right) cell masses from the 3B4014 donor line. Fibroblast morphology was elongated spindle shapes and keratinocyte morphology was polygonal shapes. Dermal fibroblasts were isolated from skin samples 1A2002 **(B)**, 3A1007 **(C)** and 3B4014 **(D)** after skin biopsy and cultured until ready for SeV reprogramming.

3.3.1.2 *Dermal Fibroblast Isolation & Culturing*

On day 12, dermal fibroblasts were observed to have emerged in the cultures and their isolation was more evident. By day 19 of culturing cells from skin biopsy samples, keratinocytes and other skin cell types were seen to be visibly separated from the dermal fibroblasts into two distinct epidermal and dermal layers (Fig 3.6A). After 28 days, all other skin cells had disappeared, and the dermal fibroblasts had proliferated substantially (Fig 3.6 B,C,D). These cells were transferred to T75 flasks (TPP®, Switzerland and culture in DMEM media (Thermo Fisher Scientific, MA). Dermal fibroblasts were then counted and transferred to 6-well plates (TPP®, Switzerland) at two different cell densities. The remainder of the dermal fibroblasts were added to Grenier Cryo.S™ 2 ml vials (Kremsmünster, Austria) at cell densities of $2x10^6$ cells/ml per vial and contained freezing medium.

3.3.2 Fibroblast Reprogramming Using Sendai Virus & iPSC Culturing for SC-HLC Drug Metabolism Model

3.3.2.1 *Attempts at Fibroblast Reprogramming*

Three different attempts (*Reprogramming 1*, *Reprogramming 2* and *Reprogramming 3*) at reprogramming were undertaken as the initial reprogramming attempts were unsuccessful and did not produce iPSC clones for the 1A2002 and 3A1007 dermal fibroblast lines. However, there was successful reprogramming of the 3B4014 dermal fibroblasts in initial attempts, and a few iPSC clones were observed. On day 0 of *Reprogramming 1*, dermal fibroblasts in TPP® 6-well plate (Trasadingen, Switzerland) were transduced with Sendai virus vectors to induce reprogramming into iPSCs at an MOI for each vector of 5:5:3 (KOS:hc-Myc:h-Klf4). *Reprogramming 2* was performed to repeat transduction of the 1A2002, 3A1007 and 3B4014 dermal fibroblast lines. The fibroblasts were again transferred into a TPP® 6-well plate (Trasadingen, Switzerland) at an increased MOI of 10:10:6 (KOS:hc-Myc:h-Klf4). However, there were no iPSC colonies formed in the 1A2002 and 3A1007 dermal fibroblast lines. The 3B4014 dermal fibroblast line produced three clones, and these were then cultured further to bulk up the iPSCs. *Reprogramming 3* was performed to repeat transduction for the dermal fibroblast lines 1A2002, 3A1007 and 3B4014 dermal fibroblast lines with an MOI of 10:10:6 (KOS:hc-Myc:h-Klf4). However, no iPSC colonies were generated in the 1A2002 dermal fibroblast line. The 3A1007 and 3B4014 dermal fibroblast line did produce a few clones, and these were then cultured further to bulk up the iPSCs.

Fibroblast reprogramming experiments were suspended and subsequently removed from the study due to time constraints. However, the iPSC clones were passaged, cryopreserved and stored in liquid nitrogen. After lockdown restrictions on laboratory use were lifted, iPSC

clones were thawed and reprogramming was resumed, allowing for phase contrast imaging of the iPSCs in both iMEF feeders and feeder-free cultures (Fig 3.7).

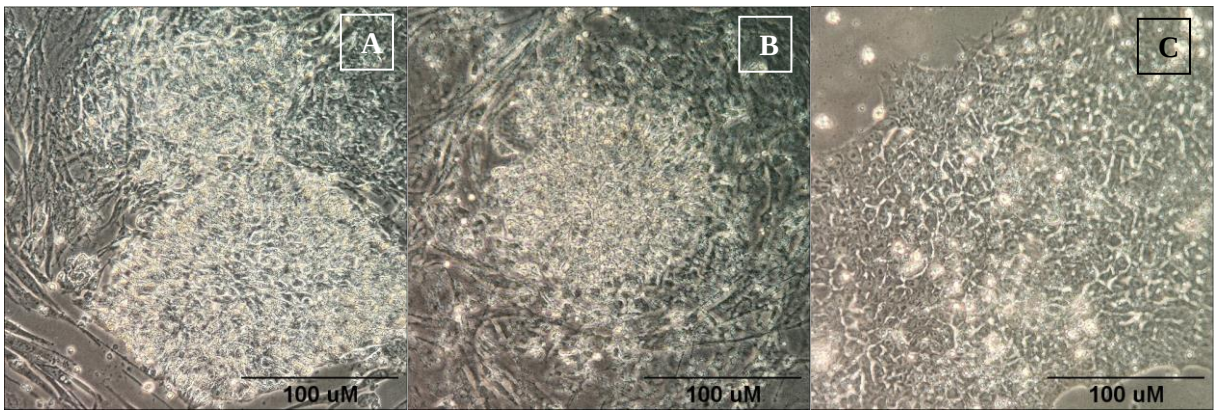


Fig 3.7: Phase-contrast images showing iPSC clones from the reprogramming of dermal fibroblast line 3B4014 from the *February* Reprogramming. Images were obtained at 20x objectives and cells were grown as a monolayer in a 37°C incubator with 5% CO_2 in air. iPSC clones were grown in both iMEF feeders **(A&B)** and feeder-free cultures **(C)**. Clone morphology was round compact cell colonies with well-defined and smooth edges, and bright white contrast due to the high density of the iPSC cells. iMEF morphology was an elongated spindle shape and keratinocyte morphology was polygonal shape.

3.3.3 HepG2 Cell Line 3D Culture & In-Vitro Spheroid Functionality Assays

3.3.3.1 *HepG2 Spheroid Culturing*

HepG2 cells were used in this study as a means for learning and to optimise the assays and experiments needed to establish the SC-HLC drug metabolism model. However, due the COVID-19 pandemic significantly affecting the time required to complete the SC-HLC model, it was then used as a substitute to illustrate proof of concept and technical ability. *Experiment 3* was undertaken and HepG2 cells used for this experiment had a cell density of $4.15x10^6$ cells/ml, and $2x10^3$ cells were seeded into each micromold in a 12-well plate. HepG2 cells were allowed to aggregate and cultured until day 10 then CYP3A4 enzyme activity studies were performed (Fig 3.8). Spheroids grew to an approximate size of 150 µM with the majority of the spheroids being similar in shape and size. On Day 2, cells had begun to aggregate into the 3D structure of the spheroids and a spheroid outer membrane had begun to emerge. By day 4, the outer membrane had fully formed as a distinct outer edge of the spheroid. On day 6, a necrotic centre had started to form which persisted until day 10, and the outer membrane had also persisted and was visibly more prominent around the spheroid edge.

3.3.3.2 *Rifampicin Induction of CYP3A4 Expression in HepG2 Spheroids*

Rifampicin inductions of *CYP3A4* were performed on the HepG2 spheroids to assess the expression of CYP3A4 enzymes in these cells. *CYP3A4* luciferase assays were used to assess the amount of CYP3A4 enzyme expressed after rifampicin induction in HepG2 spheroids from *Experiment 1, 2* and *3*. BCA protein quantitation was only performed in *Experiment 3*

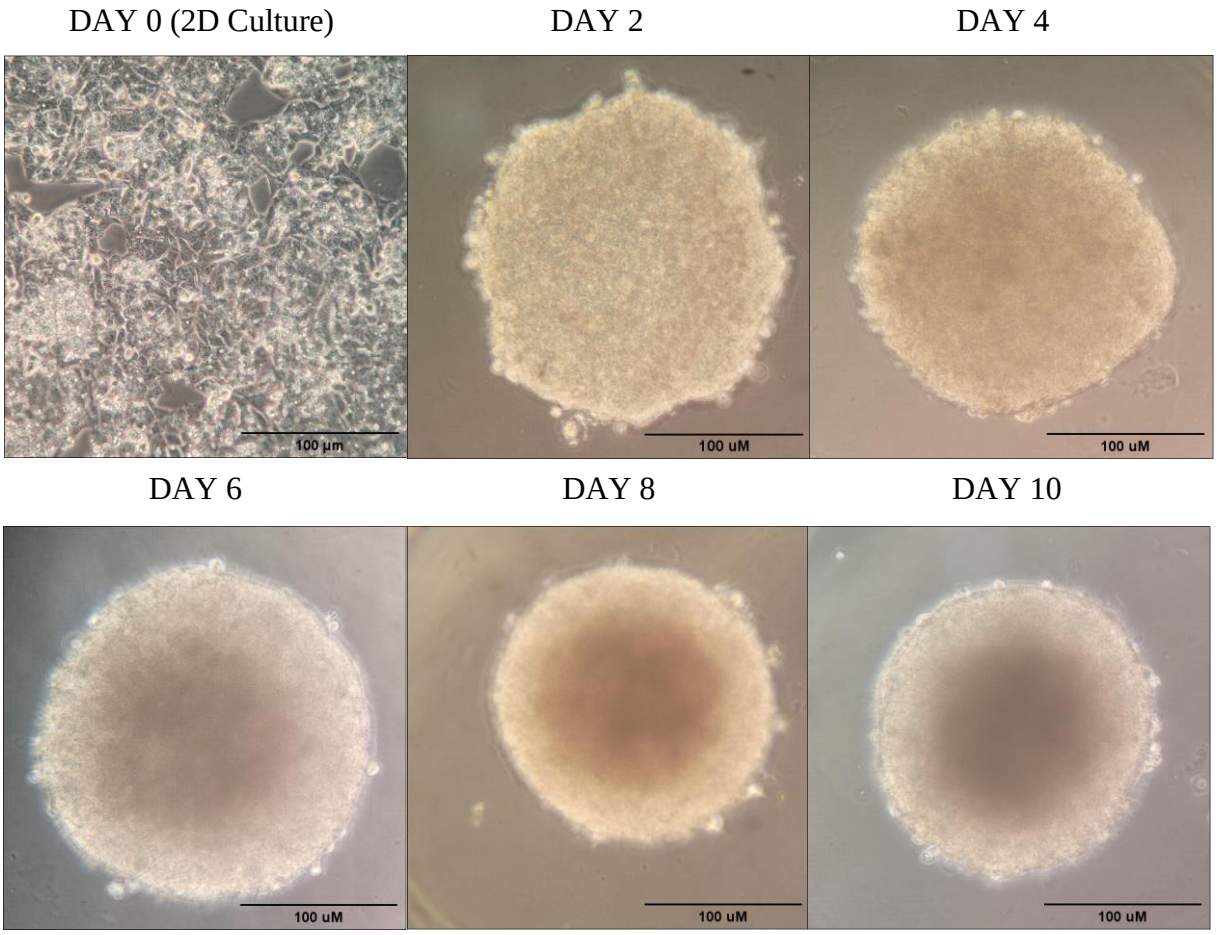


Fig 3.8: Phase-contrast images of HepG2 cells in 2D and 3D culturing spanning a 10-day period. Images were obtained at 20x objectives and cells were grown in 3D as spheroids in 2% agarose micro-molds in a 37°C incubator with 5% CO_2 in air. Spheroid morphology is characterised by a densely compacted 3D cell mass which is spherical in shape and has a distinct and well-defined outer edge. Dark inner circle indicates a growing necrotic cell mass over time as nutrient supply to the centre is compromised by the dense packing of cells

and not in *Experiments 1* and *2* as spheroid extraction in these experiments yielded unsuitable spheroids for BCA analysis.

For *Experiment 1*, *CYP3A4* induction was performed using two concentrations of rifampicin (2 µM and 2000 µM) and one without any rifampicin for a 72-hour incubation at 37°C. Luminescence readings at 1000 microseconds showed average measurements of 42527 (No Rifampicin), 79702 (Rifampicin at 2 µM) and 364 (Rifampicin at 2000 µM) (Fig 3.9). For *Experiment 2*, *CYP3A4* induction was also performed using one concentration of rifampicin (20 µM) and again one without any rifampicin for a 48-hour incubation at 37°C. Luminescence readings at 1000 microseconds showed average measurements of 68377 (No Rifampicin) and 83418 (Rifampicin at 20 µM) (Fig 3.10). For *Experiment 1* and *2*, values were not normalised to total protein by BCA due to poor quality of the spheroids after being harvested and stored.

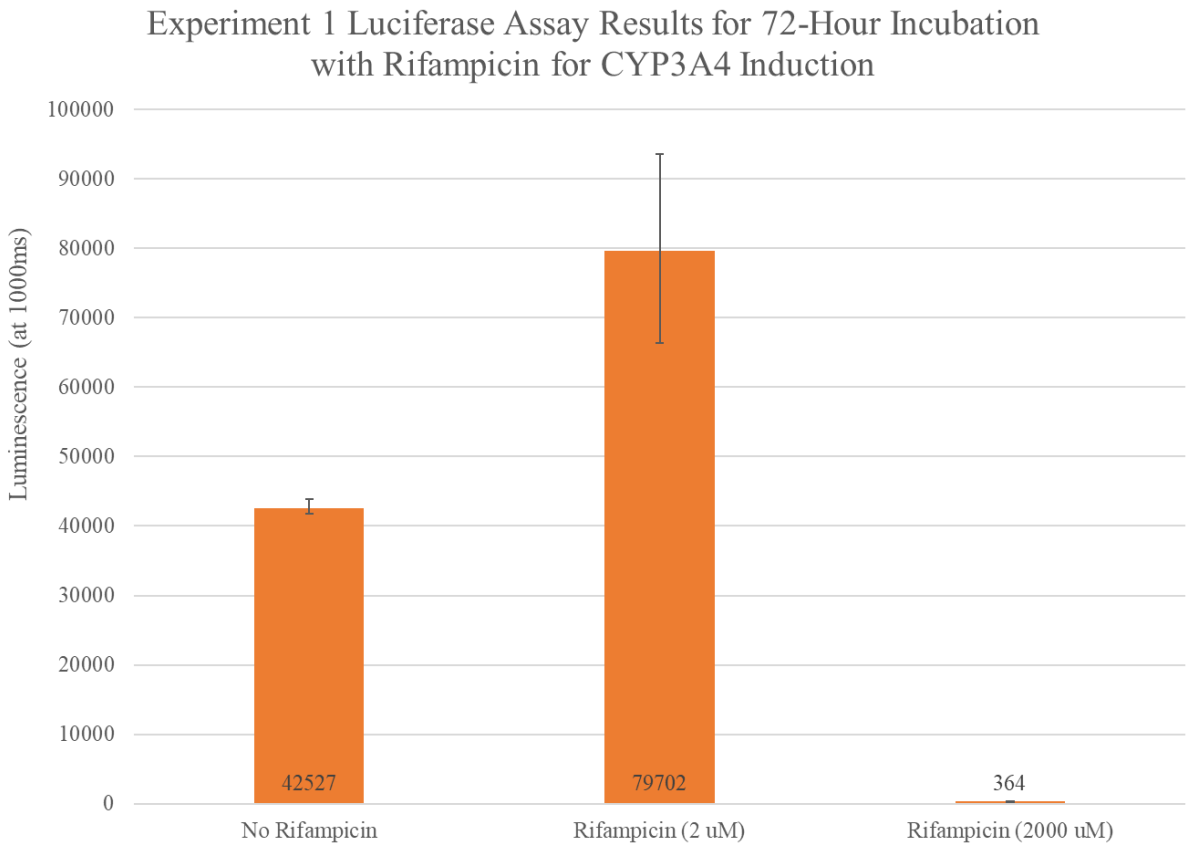


Fig 3.9: *Experiment 1* Luciferase assay results for 72-hour incubation with Rifampicin at concentrations of 2 µM and 2000 µM for *CYP3A4* induction. Data was not normalised to total protein concentration. Luminescence measurements detected CYP3A4 luciferin substrate at an exposure time of 1000 microseconds. All reactions were run in duplicate. Standard error estimates are based on best fit of the average.

For *Experiment 3*, *CYP3A4* induction was performed using two concentrations of rifampicin (2.5 µM and 25 µM) and one without any rifampicin for 24-hour, 48-hour and 72-hour incubations at 37°C. For the 24-hour incubation, luminescence readings at 1000 microseconds showed average measurements of 18413 (No Rifampicin), 17184 (Rifampicin at 2.5 µM) and 22591 (Rifampicin at 25 µM). For the 48-hour incubation, a problem occurred during the assay resulting no luminescence being detected by the reader. This likely could have been a problem with bubbles interfering with the luminescence reader or an unsuccessful incubation with luciferase. For the 72-hour incubation, luminescence readings at 1000 microseconds showed average measurements of 4677 (No Rifampicin), 4813 (Rifampicin at 2.5 µM) and 433 (Rifampicin at 25 µM).

BCA protein quantitation was performed to the spheroids in the *Experiment 3* to normalise the luciferase assay data to total protein concentrations of spheroids at each time-point incubation. A standard curve was generated using BSA standards with a protein concentration ranging from 25–2000 µg/ml (Fig 3.11) and was used to generate the *Experiment 3* total protein concentrations from the absorbance of each sample (Table 3.3). For both the 24-hour and 72-hour incubations, luminescence readings were normalised to indicate CYP3A4 activity, which

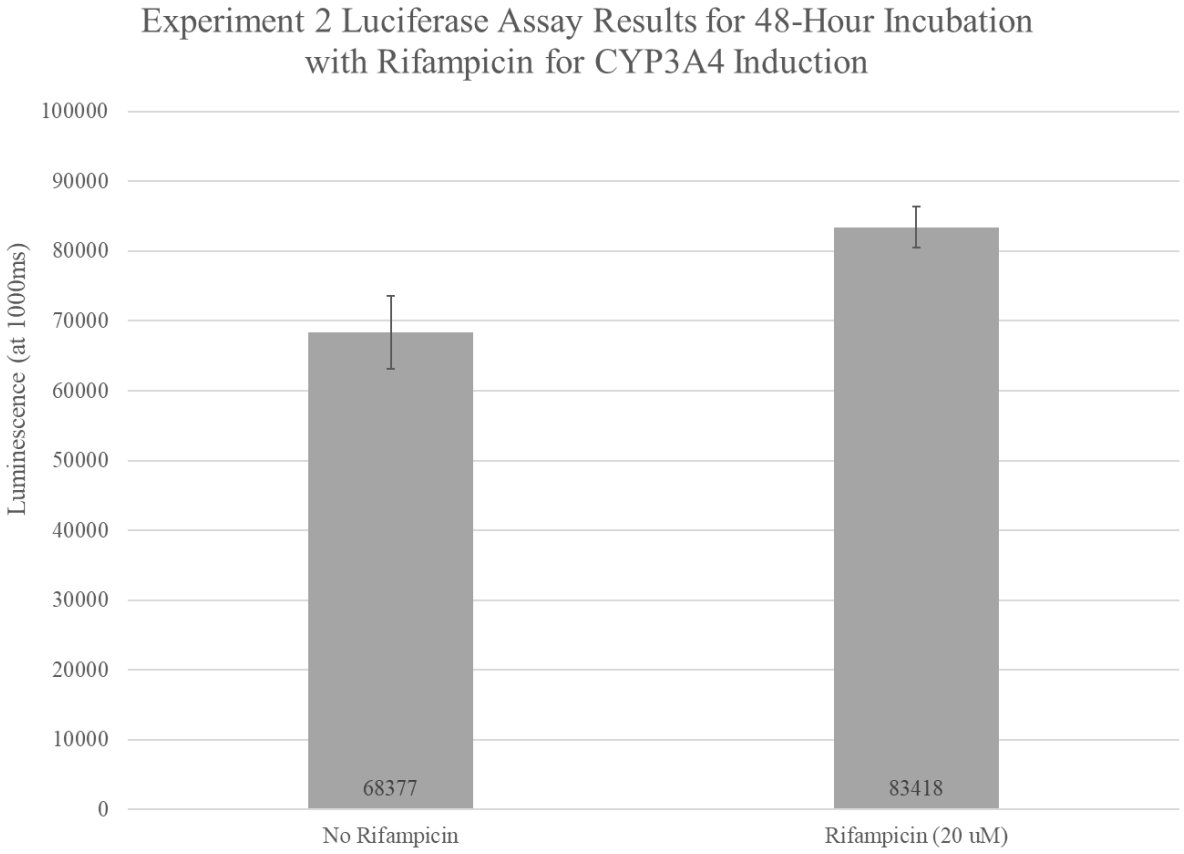


Fig 3.10: *Experiment 2* Luciferase assay results for 48-hour incubation with Rifampicin at concentrations of 20 µM for *CYP3A4* induction. Data was not normalised to total protein concentration. Luminescence measurements detected CYP3A4 luciferin substrate at an exposure time of 1000 microseconds. All reactions were run in duplicate. Standard error estimates are based on best fit of the average.

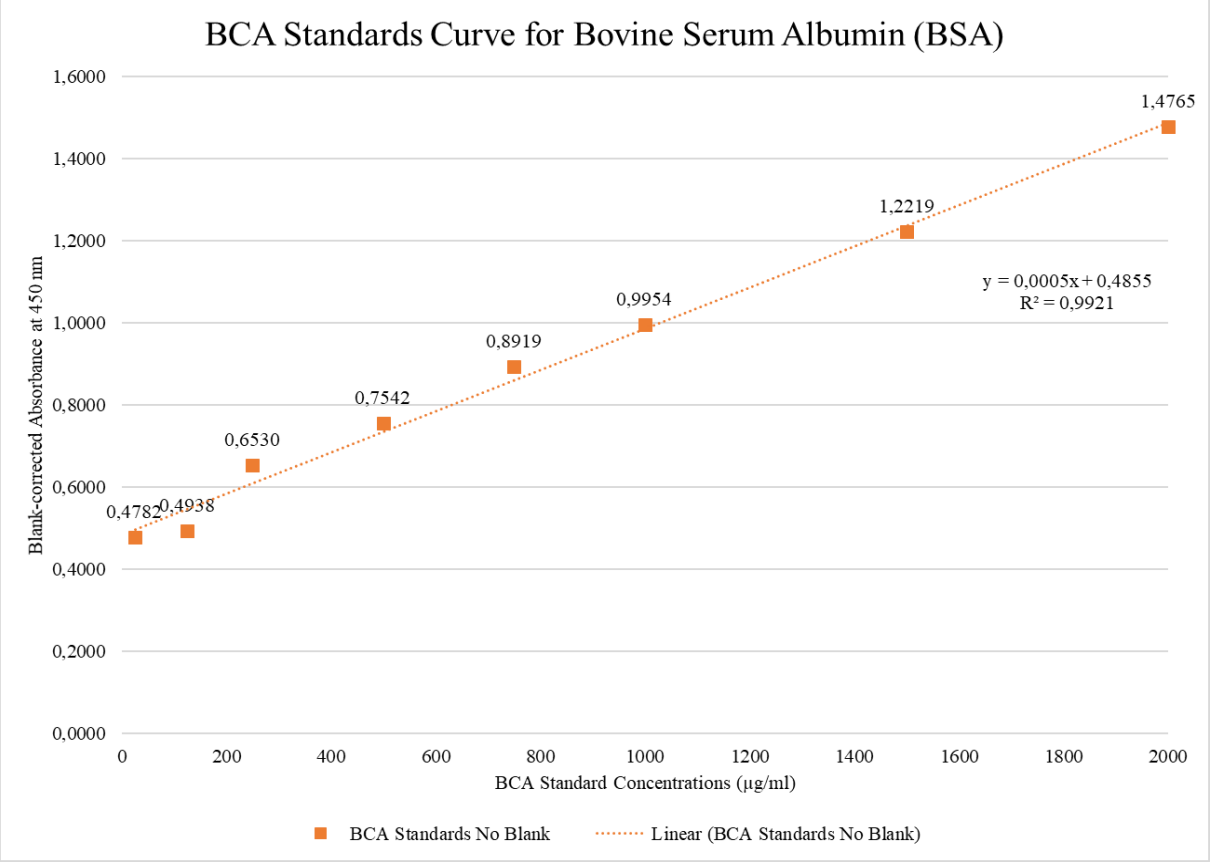


Fig 3.11: BSA standard curve for total protein concentrations ranging from 25–2000 µg/ml at an Absorbance of 450 nm. Orange line indicates the linear regression line (y = 0.0005x + 0.4855) for BSA standards. $R^2 = 0.9921$

Table 3.3: Absorbance at 450 nm and total protein concentration of HepG2 spheroids in the *Experiment 3*.

	Experiment 3 24 Hour Incubation		*Experiment 3 72* Hour Incubation	
Rifampicin Conc. (µM)	Absorbance (450 nm)	Total Protein Conc. (µg/ml)	Absorbance (450 nm)	Total Protein Conc. (µg/ml)
25	1.3846	1798.2	1.4284	1885.8
2.5	1.1387	1306.4	1.3550	1739.0
No Rifampicin	1.2393	1507.6	1.2226	1472.2

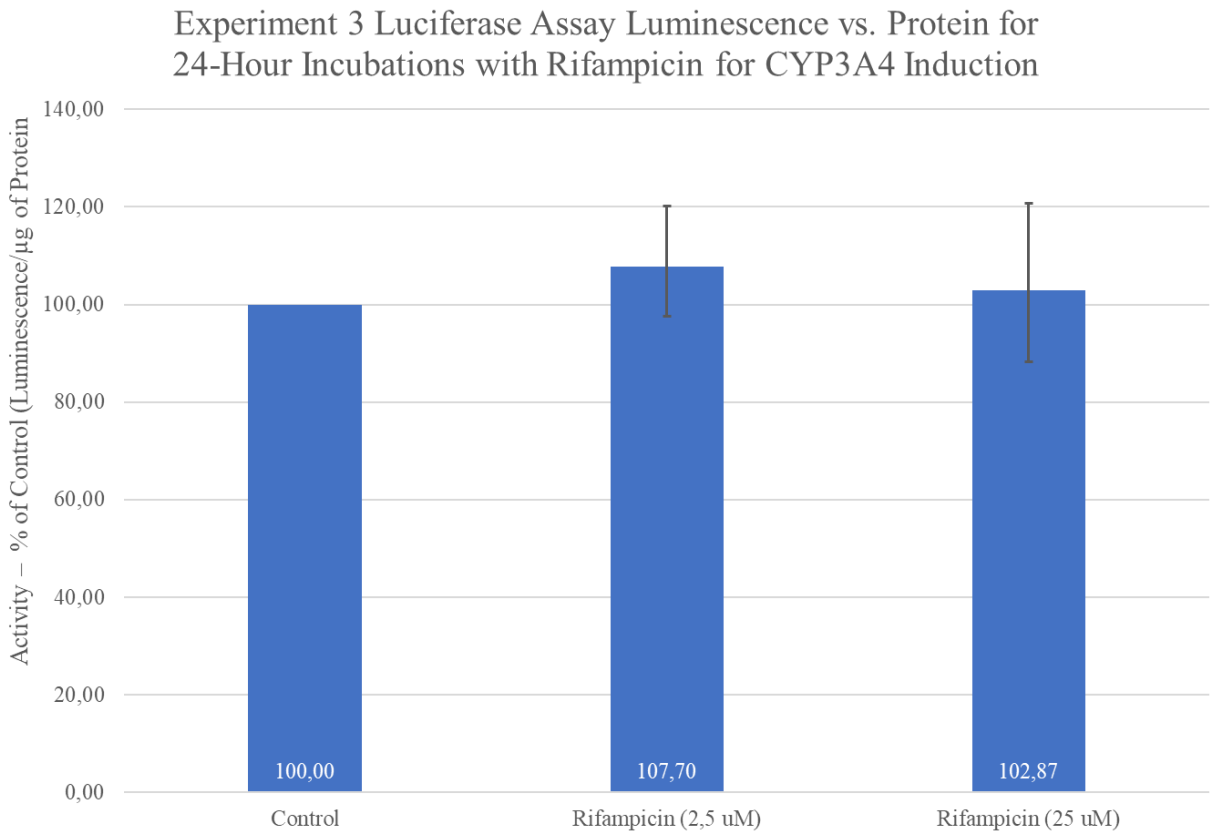


Fig 3.12: *Experiment 3* luciferase assay CYP3A4 activity after 24-hour with rifampicin induction. Luminescence values were normalised to BCA total protein concentrations as a measurement of luminescence versus protein concentration. All reactions were run in duplicate. Standard error estimates are based on best fit of the average.

was calculated from luminescence divided by total protein concentration of each sample. The 24-hour normalised enzyme activity readings showed a 7.70% (increase (Rifampicin2.5 µM)) and a 2.87% increase (Rifampicin 25 µM) in CYP3A4 induction and activity (Fig 3.12). The 72-hour normalised enzyme activity readings showed a 12.62% decrease (Rifampicin 2.5 µM) and a 92.64% decrease (Rifampicin 25 µM) in CYP3A4 induction and activity (Fig 3.13). Luminescence reading from the 48-hour incubations were not used to generate for BCA protein quantitation normalisation or CYP3A4 activity.

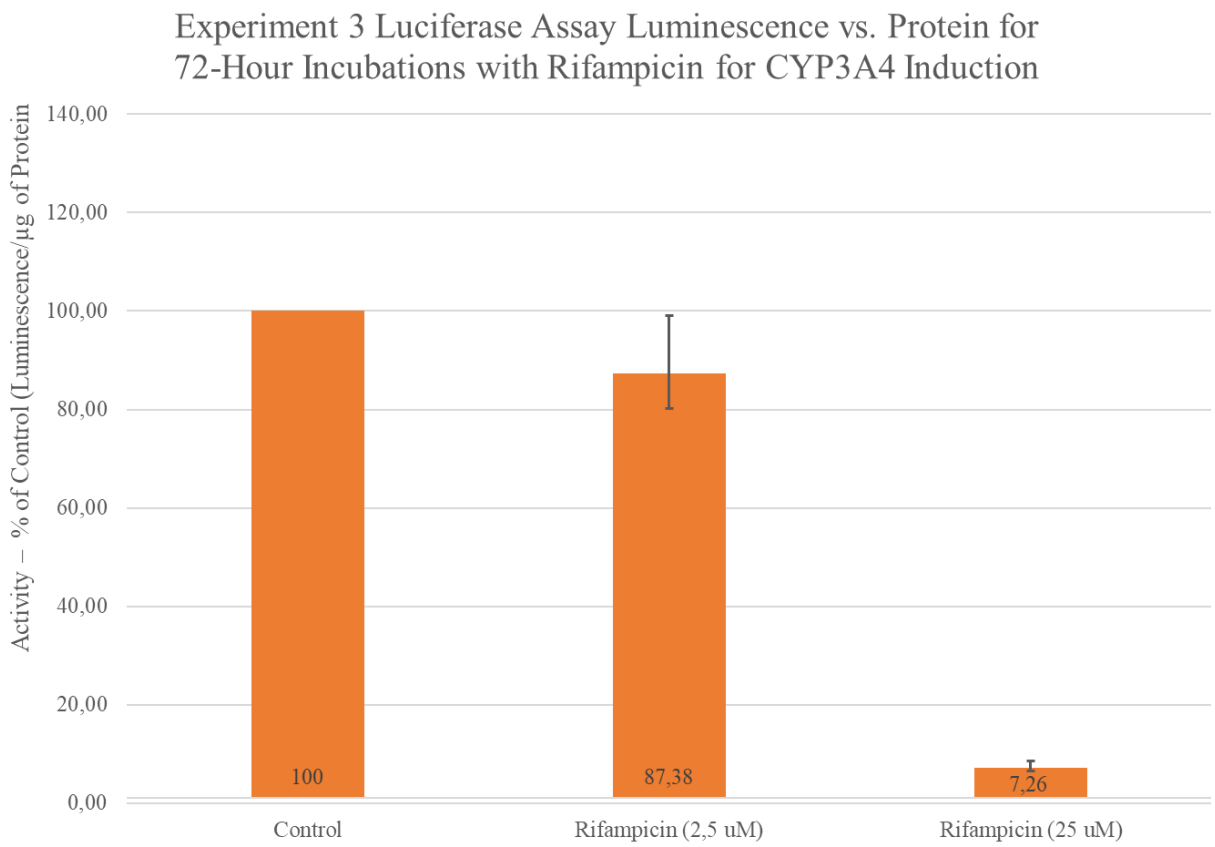


Fig 3.13: *Experiment 3* luciferase assay CYP3A4 activity after 72-hour with rifampicin induction. Luminescence values were normalised to BCA total protein concentrations as a measurement of luminescence versus protein concentration. All reactions were run in duplicate. Standard error estimates are based on best fit of the average.

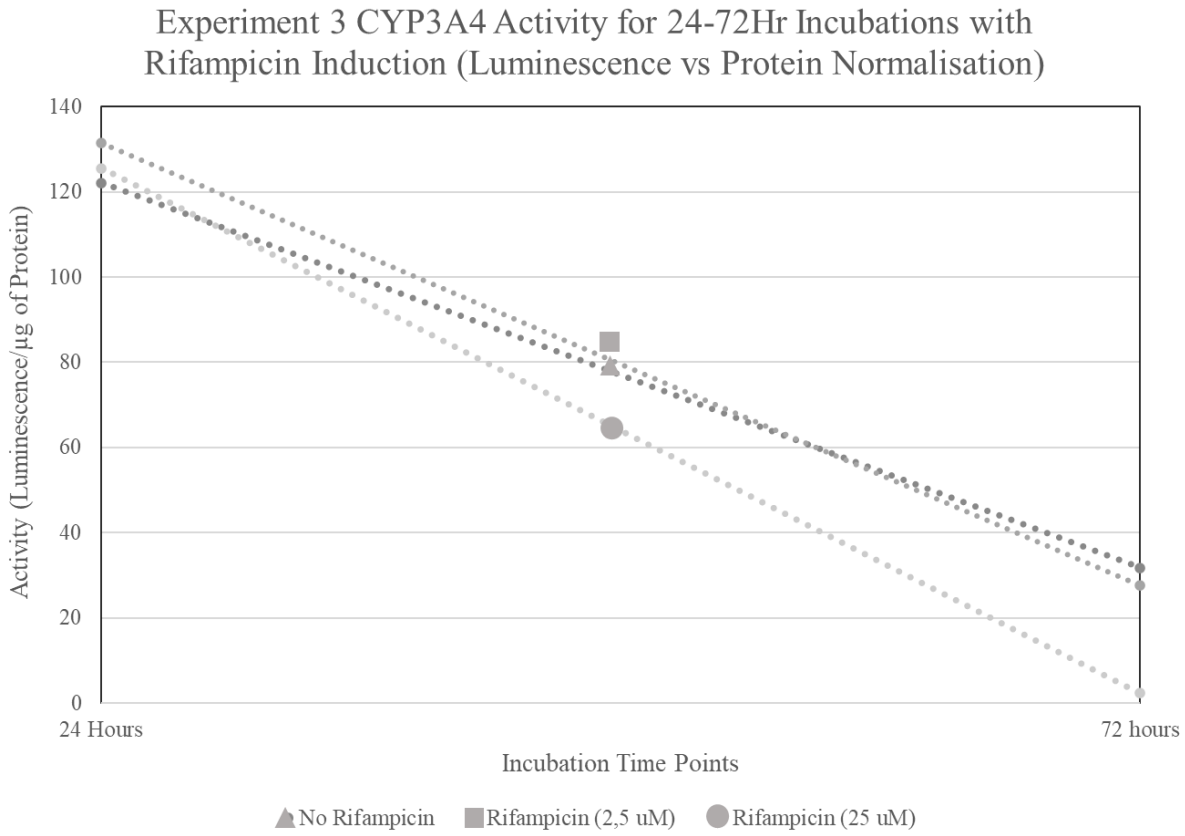


Fig 3.14: *Experiment 3* luciferase assay activity over 24- to 72-hour incubation with rifampicin for CYP3A4 induction. Activity was measured as luminescence divided by total protein concentration. Activity readings for 48-hour rifampicin incubation were not included as there were no reliable data points acquired.

The CYP3A4 activity data from *Experiment 3* was also analysed across the 24- and 72-hour incubations looking for trends of activity from the rifampicin induced *CYP3A4* expression (Fig 3.14). This showed a slight increase in activity at 24-hour incubation with 2.5 and 25 µM rifampicin over the No rifampicin samples. However, induction for 72-hours with rifampicin showed a large decrease in CYP3A4 activity for 25 µM rifampicin induction and a slight decrease for 2.5 µM rifampicin compared to no rifampicin inductions.

3.3.3.3 *Viability Staining*

The viability of the HepG2 spheroids were assessed in *Experiment 2* and *3* using fluorescein diacetate (FDA) and propidium iodide (PI) live/dead staining assay. This was performed on day 21 for *Experiment 2* and on day 16 for *Experiment 3*. The FDA/PI staining performed on *Experiment 2* spheroids, however, the resolution of the images from the staining were inadequate to determine the number of viable and dead cells in the spheroids. In *Experiment 3*, the FDA/PI staining was unsuccessful, which was likely caused by inadequate preparation of staining solution, over saturation of spheroids with staining solution or shortened incubation time to allow the staining of the spheroids.

3.4 DISCUSSION

3.4.1 Dermal Fibroblasts derived iPSCs for SC-HLC Differentiation.

In this study dermal fibroblasts were used as the cell source for iPSC reprogramming and for the establishment of a CYP2D6*17 drug metabolism model. Dermal fibroblasts are the most frequently used and was the first cell source used for iPSC reprogramming. This is due to their low cost, relative ease in culturing, vast knowledge of fibroblasts and their frequent use in different research fields (Raab *et al.*, 2014). They are normally acquired from skin biopsies on adult persons or from foreskin removed during circumcision. Donor variation and donor health status from these biopsies and samples have allowed for the study of genetic and cell variation effects on disease development and outcomes (Raab *et al.*, 2014). Dermal fibroblasts acquired from study participant skin biopsies are expected to be suitable for drug metabolism model for CYP2D6*17 after iPSC reprogramming. In addition, they have the capacity to be used in other cellular models, especially in all the other components of ADME drug interactions, and even used in the creation of numerous bioengineered tissues for disease modelling (Saha and Jaenisch, 2009; McGivern and Ebert, 2014). This would, however, require further investigations into whole genome sequencing, transcriptomics and epigenomics of these dermal fibroblasts. This is to ensure no cancerous mutations are found within these genomes, as well as account for the epigenetic influences of the dermal fibroblast on the re-engineered tissue (Saha and Jaenisch, 2009; Raab *et al.*, 2014).

However, there are a few shortcomings with using dermal fibroblasts as a cell source for iPSC reprogramming. One common problem encountered during reprogramming is the inefficient reprogramming percentage, which is found between 0.01-0.5% for dermal fibroblasts (Raab *et al.*, 2014). There are also evidently longer time periods of 3-5 weeks for stem cell colonies to form after the reprogramming. Alternatively, keratinocyte reprogramming takes approx. 2-3 weeks for stem cell colonies to start appearing in culture (Raab *et al.*, 2014). A commonly suggested reason for these differences in reprogramming efficiencies has been the mesenchymal-to-epithelial transition (MET). *In vivo*, dermal fibroblasts are found in the skin dermis layer and are of mesenchymal cell origin. They are mainly responsible for the production and generation of precursor molecules crucial in the formation of the ECM. During reprogramming, an initial MET step is undertaken by upregulating epithelial-associated genes such as ECPAM and CDH1. This may be responsible for the increases in time periods of reprogramming, due to the added time needed for MET. This may also account for decreased yields of positively reprogrammed iPSCs, due to unsuccessful MET steps (Raab *et al.*, 2014). With regards to the iPSC colonies in this study, there was poor

reprogramming efficacy evident from the numerous reprogramming procedures. The addition MET step may have resulted in slower times and unsuccessful yields with dermal fibroblast reprogramming as highlighted above.

3.4.2 Use of Sendai Virus for iPSC Reprogramming

Sendai virus (SeV) is classified as parainfluenza virus type I from the *Paramyxoviridae* family and is a respiratory virus found in mice and rats (Nakanishi and Otsu, 2012; Thermo Fisher Scientific, 2016b). SeV is a non-segmented single-stranded RNA virus with a viral envelope that is approx. has a diameter of 150-250 nm. The RNA genome of the virus contains six genes that codes for six proteins (nucleocapsid, phosphoprotein, matrix protein, fusion protein, haemagglutinin-neuraminidase, and large protein) for the viral structures of SeV (Nakanishi and Otsu, 2012; Thermo Fisher Scientific, 2016b). SeV is capable of producing large amounts of protein without entering the nucleus and integrating in the genomes of infected cells. The SeV vector instead remains in the cytoplasm using the replication machinery of the infected cell and is normally completely eliminated from the cell by the 10^{th} passage. SeV has a reprogramming efficiency between 0.5-1.4% and is able to infect the majority of somatic cell types (Rao and Malik, 2012).

This has resulted in the widespread use of SeV by researchers as an ideal vector for reprogramming of dermal fibroblasts and other cell types (Rao and Malik, 2012). However, as a reprogramming vector it does come with its difficulties and shortcomings, which are the low reprogramming efficiency compared to other reprogramming methods. Though it has the highest reprogramming efficiency of all virus methods and has commercially available kits making it one of the more used methods. There is also potential risk of viral genes being reactivated in the infected cells (Rao and Malik, 2012; Raab *et al.*, 2014). In this study, there were multiple steps of optimisation of dermal fibroblast reprogramming with SeV and most of the transductions were unsuccessful. However, *Experiment 3* showed the most success because of the change in MOI for the SeV transduction. The unsuccessful transductions may be due to reasons mentioned before with the use of dermal fibroblasts as a cell sources, or this may be a result of the known poor efficiency of SeV in cells in comparison to non-viral reprogramming methods. This could be relieved by using different vectors which are able to achieve a high reprogramming percentage and also do not integrate into the cell genome, such as lentiviral miRNA (efficiency = 10.4-11.6%) or miRNA (efficiency 0.6-4.4%) (Rao and Malik, 2012).

3.4.3 Comparison of HepG2 Spheroid with Other Hepatic Cell Culture Methods for Drug Metabolism Functionality Assays.

3.4.3.1 *Rifampicin Induction of CYP3A4 Expression in hepatic cells.*

In this study, objectives (2) and (3) were formulated to compare the expression of CYP enzymes in different cell lines cultured in 2D and 3D platforms, and subsequent *CYP2D6*17* expression and enzyme activity. However, endeavours to achieve this were impeded by time restraints due to the COVID-19 pandemic. Nonetheless, a HepG2 cell line was used to establish functionality of *CYP3A4* after incubation with inducer rifampicin for 24-, 48- and 72 hours. Rifampicin is known to be a strong inducer of *CYP3A4* and has shown to cause a 2.4-fold to 4.7-fold increase hepatic CYP3A4 protein (Niemi *et al.*, 2003). In PHHs, rifampicin has been seen to cause a 150-fold increase in the expression of *CYP3A4* at a concentration of 10 µM. In HepG2 cells, a concentration of rifampicin at 10 µM resulted in a 3.0-fold to 5.5-fold of CYP3A4 protein expression with additional induction of the *CYP3A4* gene with only PXR and HNFα transcription factors (Li and Chiang, 2006).

CYP3A4 induction using rifampicin incubation for 24 hours at a concentration of 2.5 µM and 25 µM was performed and showed a 1.1-fold and 1.07-fold increase in CYP3A4 activity, respectively. What also appears to occur is toxicity from rifampicin after 24 hours of incubation with the spheroids. This was confirmed as the 72-hour "No Rifampicin" data had a higher luminescence reading and activity than all the other concentration. A rifampicin incubation time of 72-hours is a commonly used timepoint in the literature (Niemi *et al.*, 2003; Li and Chiang, 2006). However, more investigations need to be performed to determine if the differences in this study are due to errors in the experiment execution or variation in the HepG2 cell line influencing the induction. Hence, for this HepG2 cell line 24-hours and rifampicin concertation at 2.5 µM would be the most suitable time point for inducing the expression of *CYP3A4.* From this an approximation of CYP3A4 can be read at also 2-fold and 1.2-fold increases. The expression of CYP3A4 was important as it is involved in the majority of phase I metabolism (Pearson and Wienkers, 2019). It was also used to infer functional expression of CYP2D6 enzymes in the HepG2 cells as it is part of majority of the metabolic pathways that CYP2D6 is involved in. In addition, when looking at the tamoxifen metabolism pathway, CYP3A4 is critically important for the conversion of tamoxifen to *N*-desmethyltamoxifen which is subsequently metabolised by CYP2D6 (Muroi *et al.*, 2014).

In this study, readings showed that there was functional expression of CYP3A4 in the HepG2 cells that was lower that what is observed in other studies highlighting expression of CYP3A4 in HepG2. The HepG2 spheroids showed slight increases in CYP3A4 enzyme activity though

this was at considerably lower concentrations than those seen in the literature. Further research into this HepG2 cell line may indicate variation in the line that may result in lower fold increases from rifampicin induction. Furthermore, repeating the rifampicin incubation experiments may yield different results that are closer to what is seen in the literature. The 3D culturing of these HepG2 cells is stated to increase CYP enzyme expression and HepG2 spheroids in this study highlighted expression of CYP3A4. However, more studies into the CYP3A4 expression in 2D culture as a comparison may be required to confirm this hypothesis. From the luciferase data functional expression of CYP3A4 expression indicated that the initial metabolism of tamoxifen in these HepG2 cells by CYP3A4 was properly functioning, but this would need confirmation via LC/MS-MS analysis.

3.4.4 Strengths & Limitations Related to this Chapter's Work

This chapter highlighted numerous methods on how to establish a cellular model to investigate the effects of CYP2D6*17 variants on tamoxifen to endoxifen metabolism. The objectives assigned to achieved this were objectives (2) and (3), however, there were changes to the study due to time constraints. Nonetheless, a HepG2 cell line was used to demonstrate a proof of concept and technical ability, and garner data for the functionality of HepG2 spheroids to study variant effects on drug metabolism. Though the data highlights CYP3A4 functioning *in-vitro,* the inability to generate SC-HLCs to establish a cellular drug metabolism model, and the inability to perform enzyme activity experiments has resulted in some limitations in this study. These limitations are the following:

1. Though iPSCs colonies were attained from dermal fibroblasts acquired form the 3A1007 and 3B4014 skin biopsies, the characterisation of these iPSC lines was not completed due time constraints. This resulted in no data being generated for these iPSCs and subsequent change to a HepG2 cell to attain the results needed for this study. There was no differentiation of the iPSCs acquired from the skin biopsies into SC-HLCs for analysis of the effects of *CYP2D6*17* variant on CYP2D6 drug metabolism. Even though this was the case, the iPSC lines will be characterised at a future date and subsequent studies maybe be able to produce data on the effects of the *CYP2D6*17* variant on CYP2D6 drug metabolism.

2. The whole *CYP2D6* sequence for the study participants was not completed and this may result in unexpected complications or discrepancies for the SC-HLCs derived from each study participant. Though this was the case, there was a significantly low chance of acquiring an individual who was not a *CYP2D6*17* homozygote while also

having the T/T genotype linked to the SNP rs28371706. However, *CYP2D6* status of the reference iPSC line (3A1007) acquired through skin biopsy is unknown and this may affect the data that is generated using it as a reference because of the numerous possibilities of other variants being present in this line.

3. The use of dermal fibroblasts as a cell source may affect the epigenetic makeup of the iPSC cells generated from this study. This may impact the expression levels of *CYP3A4* and *CYP2D6* resulting in inaccurate measurements of enzyme activity and drug metabolism in the cellular model created. This can be overcome by investigating the epigenomes of the dermal fibroblasts and determining which chromatin and methylation patterns may affect the expression of the CYP genes of interest.

4. The functional assay results in this chapter were performed on HepG2 cell lines and cannot be used to make inference on SC-HLC functionality. This requires future research using the same functional assays to determine the functionality of the SC-HLCs that will be produced from the iPSCs generated in this study.

5. Insufficient amount of statistical analysis throughout the chapter. This was due to the qualitative nature of the analysis done to the spheroids for viability and assessing growth. There were also problems with the viability staining assay and this hindered the ability for adequate assessment of spheroid cell viability. These were due to time constraints on the project and thus not enough time was available to optimise these assays for this model. Future steps should be taken to readdress these statistical shortcomings and incorporate better statistical analysis into the work.

The findings of this chapter also showed some strengths and may reveal insights into the *CYP2D6*17* variant and its presence in African populations. These strengths were the following:

1. At time of writing, the iPSC lines generated in this study were the first stem cell lines that included the African-specific CYP2D6*17 variant, and this can be used to study numerous CYP2D6*17 drug interactions. Future research should expand the knowledge on the enzyme activity effects of this variants and drug metabolism differences seen in all substrates of CYP2D6.

2. The HepG2 spheroid cultures showed slighted increased activity due to rifampicin induction of *CYP3A4*, which is expected from the literature. However, more time may

have been needed to optimise this induction to get increases in activity that were comparable to the literature. Further research into 3D culturing of iPSCs should be performed to affirm the effects of 3D culturing on liver function and rifampicin induction of *CYP3A4*.

3.5 CONCLUSION

This chapter describes the establishment of a SC-HLC model to investigate the effects of CYP2D6*17 on the metabolism of tamoxifen to endoxifen. Skin biopsies were performed, and dermal fibroblasts were isolated for reprogramming to generate iPSCs, Due to time constraints, changes were made to use a HepG2 cell line to demonstrate the concept outlined in this study. The HepG2 cell line cultured in 3D was able to show functional expression of enzyme CYP3A4 and thus it was inferred that CYP enzyme expression was apparent in these spheroids. Future studies should resume the establishment on an iPSC model and further investigate the functional effects of *CYP2D6*17* on drug metabolism.

CHAPTER 4

FUTURE RESEARCH & CONCLUSIONS

4. CHAPTER 4: FUTURE RESEARCH & CONCLUSIONS

This study aimed to *assess the frequency of the CYP2D6*17 variant in African populations and to evaluate its enzyme activity on the activation of tamoxifen to endoxifen in iPSC-derived hepatocyte-like cells (SC-HLCs)*. This was achieved by accomplishing three objectives. Objective (1) was *to determine the allele and genotype frequencies of the reduced function variant CYP2D6*17 in African populations.* This objective was successfully completed and highlighted the effect of African population migratory patterns on the distribution of *CYP2D6* variants in African individuals from different regions on the continent. Objective (2) was *to develop the SC-HLC model* and Objective (3) was *to measure enzyme activity of CYP2D6 and compare the effects of CYP2D6*17 to the reference CYP2D6*1 variant in the conversion of N-desmethyltamoxifen to endoxifen using the SC-HLCs*. These two objectives were only partially completed and due to restrictions to entering the laboratory during the COVID-19 pandemic. Future studies by others will continue pursuing the establishment of a SC-HLC model and other cellular models to assess drug metabolism in the background of specific genetic variants common in Africans. Moving forward from this study, there are numerous avenues for future research using the iPSCs generated by this study to investigate, more specifically, the effects of *CYP2D6*17* and other *CYP2D6* variants on drug metabolism. Below are some potential areas and techniques for further research into drug metabolism modelling.

4.1 FUTURE RESEARCH ON THE PREVALENCE OF *CYP2D6* VARIANTS IN AFRICAN POPULATIONS

The frequency of *CYP2D6* variants within African populations is an important part of ensuring that precision medicine for better pharmaceutical use and outcomes can be implemented on the African continent (Matimba *et al.*, 2009; Dugger, Platt and Goldstein, 2018). *CYP2D6* has over 100 different variants that have been reported and many of these have not yet been studied in African populations causing a gap of knowledge in these populations (Matimba *et al.*, 2009; Nofziger *et al.*, 2019). Future research aimed at characterising these variants is needed to ensure better drug development and implementation of appropriate treatments for African populations to ensure better health outcomes for Africans (H3Africa, 2011; Dugger, Platt and Goldstein, 2018). There is also a need to investigate some of the other *CYP* gene variants associated with differing responses to pharmaceutical interventions. *CYP* genes such as *CYP2B6*, *CYP2C9* and *CYP2C19* are also known to carry numerous genetic variants associated with differences in enzyme activity (Tracy *et al.*, 2016).

CYP2B6 is another locus that is also highly polymorphic has 38 star allele variants reported to date. The enzyme is responsible for approximately 5% of all phase I drug metabolism in the liver (Rendic and Guengerich, 2015; Desta *et al.*, 2019). Decrease function variants *CYP2B6*16* and *CYP2B6*18* are found at frequencies of 6.5% and 7.1%, respectively, in African populations (Zhou, Ingelman-Sundberg and Lauschke, 2017). CYP2B6 is also the major metaboliser of the drug efavirenz, which is used in the treatment of treatment-naïve HIV-positive patients. Efavirenz is also involved in increasing expression of *CYP2B6* via auto-induction which results in efavirenz aiding its own metabolism. The two variants, *CYP2B6*16* and *CYP2B6*18,* have also been associated with decreased clearance of efavirenz due to decreased enzyme activity (Desta *et al.*, 2019). *CYP2C9* has approximately 30 different variants reported to date and the enzyme is involved in approximately 9% of all phase I drug metabolism in the liver (Johnson *et al.*, 2011; Rendic and Guengerich, 2015). The variant *CYP2C9*3* causes decreased enzyme function and in African populations and is present at a frequency of 5.6% (Zhou, Ingelman-Sundberg and Lauschke, 2017). It has been linked to impaired metabolism of warfarin which subsequently results in higher risk of excessive bleeding when prescribed to patients with this variant (Johnson *et al.*, 2011). *CYP2C19* is another *CYP* gene that is polymorphic and has more than 48 star allele variants (Fricke-Galindo *et al.*, 2016). The CYP2C19 enzyme is involved in approximately 9% of all phase I drug metabolism in the liver (Rendic and Guengerich, 2015; Zhou, Ingelman-Sundberg and Lauschke, 2017). In African populations, the *CYP2C19*2* variant is present at a

frequency of 18.1% and is associated with an inactive enzyme causing neurological toxicity when the drug mephenytoin is administered (Fricke-Galindo *et al.*, 2016; Zhou, Ingelman-Sundberg and Lauschke, 2017). The *CYP2C19*17* variant is found in African populations at a frequency of 23.5% and is associated with a higher clearance of warfarin affecting the efficacy of warfarin treatment (Fricke-Galindo *et al.*, 2016; Zhou, Ingelman-Sundberg and Lauschke, 2017).

Investigating CYP gene variants is important for developing guidelines for treatments of diseases to ensure better treatment efficacy (de Vries Schultink *et al.*, 2015; Ramaswami, Bayer and Galea, 2018). Also, the ADR burden in South Africa is high, as it is in other populations, and needs to be addressed to resolve with 6.3% of hospital admissions that are the result of ADRs and 6.3% of individuals in hospital who develop an ADR during their admission. ADRs in South Africa are also estimated to be responsible for 2.9% of deaths in adult medical ward admissions, as well as 16% of deaths overall being ADR-related (Terblanche *et al.*, 2017). This can be achieved by screening more in African populations and using bioinformatic tools to highlight CYP variants relevant for pharmaceutical interventions. There is also a need to characterise these variants and link them to *in vivo* functional data using cellular models such as the SC-HLC model propose in this study.

4.2 COMPARISON OF CYP2D6 MODELLING IN SC-HLCs & OTHER CELLULAR MODELLING METHODS

One of the critical components of a successful model is ensuring that it is able to exhibit physiological values that are comparable to *in vivo* hepatocytes. The methodologies used, such as cell culturing techniques and liver differentiation protocols, are important to establishing a functional model. Expression levels of RNA from liver-function genes give insight into gene functioning in comparison to *in vivo* hepatocytes. Protein expression levels are also important to highlight adequate translation of mRNA transcripts into functional protein. Though not highlighted in this study, future attempts should be pursued to differentiated iPSC generated here into SC-HLCs, with mRNA and protein expression in these cells being investigated. It is also important to compare these values to the RNA and protein expression from cells existing literature. Comparisons to the most used cell lines, such PHHs and HepG2s, is essential to exhibiting the utility of this model and its contributions to drug metabolism modelling.

4.2.1 RNA & Protein Expression Levels

RNA expression of *CYP* genes varies between different cell types especially with regards to *CYP2D6*. In PHHs, the literature shows RNA expression levels of the *CYP* genes that are closest to the *in vivo* liver *CYP* gene expression profiles (Bell *et al.,* 2016). Andersson et. al. (2018) showed a relatively high *CYP2D6* gene expression in PHHs 2D cultures (ΔCt = 15.9 – 20.5) compared to SC-HLCs (ΔCt = 22.3 – 24.0) and HepG2 cells (ΔCt = 24.0) 2D cultures. However, *CYP3A4* gene expression was similar in PHHs 2D cultures (ΔCt = 14.0 – 19.3) and SC-HLCs 2D cultures (ΔCt = 14.8 – 16.7), but relatively low in HepG2 cells (ΔCt = 25.4) (Andersson *et al.,* 2018). Bell et. al. (2017) also showed a noticeable difference between PHH spheroids and SC-HLCs in 2D cultures, with the PHHs having a markedly higher gene expression profile. However, *CYP3A5* expression was relatively high in SC-HLCs and comparable to PHHs (Bell *et al.,* 2017). Takayama et. al. (2013) have also reported a higher relative gene expression for *CYP2D6* and *CYP3A4* in 3D SC-HLCs spheroids than the same cells cultured in 2D. These spheroids also had *CYP2D6* and *CYP3A4* gene expression levels similar to that of PHHs (Takayama *et al.,* 2013). Protein levels in a cellular drug metabolism model are critical in determining the amount of RNA expression that results in protein translation. It gives insight into the functional effects of gene variants on the CYP enzymes, and regarding drug metabolism, is one of the most important physiological measurement. Bell et. al. (2018) showed that in PHHs, CYP3A4 protein expression was increased by 5- (Day 0) to 27-fold (Day 14) in 3D culturing condition, and CYP2D6 showed a slight increase in expression in 2D (Day 0) and 3D (Day 14) (Bell *et al.*, 2018). For this study, investigating

protein expression would be useful to determine the effectiveness of the model to test CYP enzyme kinetics.

These values show a similar trend that iPSCs are indeed a better source for creating hepatocyte-like cells. However, they are still not as informative as PHHs and other *in vivo* methods for modelling drug metabolism. Newer technologies to increase RNA and protein expression and mimic *in vivo* liver functioning, such as ECM reconstruction and co-culturing with different liver cell types, are aim to tackled these limitations (Fiorotto *et al.*, 2018). Future research using the iPSCs generated from this study should be used to determine the optimal methodologies and technologies to attain RNA and protein expression profiles similar to *in vivo* liver tissue.

4.3 CYP ENZYME ACTIVITY IN SC-HLCs, HEPG2 CELLS & PHHs

Investigating the effects of genetic variants on enzyme activity is one of the major steps in understanding CYP enzyme kinetics. Bioluminescence assays, such as a luciferase assay, are useful in acquiring insights into the activity of CYP enzymes. LC/MS-MS is another molecular tool that can be used to decipher the chemical structure of compounds and determine the amount of drug substrates metabolised by CYP enzymes. Performing LC/MS-MS analysis and luciferase assays on the SC-HLCs generated here should be pursued in the future.

4.3.1 Luciferase Assay to Investigate Enzyme Activity

Luciferase assays are bioluminescence assays aimed at determining the activity of the CYP enzymes, as well as an indication of the enzyme kinetics. Often luciferase is used to assess the induction of CYP gene expression by using inducers such as rifampicin that induce *CYP3A4* (Niemi *et al.*, 2003). Berger et. al. (2015) showed more than a 2-fold increase in CYP3A4 and CYP2C9 enzymes when incubated with 25 µM of rifampicin for four days (Berger *et al.*, 2015). Takayama et. al. (2013) used luciferase assays to show an increase in expression of both *CYP2C9* and *CYP3A4* when comparing 3D cultured SC-HLCs with 2D cultured SC-HLCs. There was also a substantial increase in both culture conditions when incubated with 10 µM rifampicin for 48 hours. However, these values were significantly lower than those of PHHs including PHHs incubated with rifampicin (Takayama *et al.*, 2013).

For this study, we used a luciferase assay to determine the enzyme activity of CYP3A4 in HepG2 cells. This should be done in the SC-HLCs that will be generated as well to indicate proper CYP3A4 enzyme functioning. Although, this particular luciferase assay method cannot be applied to the *CYP2D6 in vivo* as the luciferin substrate used for CYP2D6 is non-specific and binds to other CYP enzymes (Promega Corporation, 2016). Future research should make use of this and other tools to gain insight into the CYP enzyme activity of the SC-HLCs generated here in both 2D and 3D culturing conditions.

4.3.2 LC/MS-MS & Drug Compound Metabolism

Determining CYP enzymatic function is best achieved by testing the metabolism of these enzymes in the presence of particular drug compound substrates. This allows for better understanding of the interactions between pharmaceutical drugs and CYP enzymes, and what the optimum kinetics would be for these pairings. LC/MS-MS is one of the most common methods to analyse drug metabolism and there are five CYP enzymes that this study would have investigated. These five CYP enzymes are CYP1A2, CYP2B6, CYP2C9, CYP2D6 and CYP3A4/5, with their five substrates being phenacetin, bupropion, diclofenac, bufuralol and

midazolam, respectively. There are numerous different substrates that have been used and Table 4.1 highlights some of the metabolism findings expected when using SC-HLC, HepG2 and PHH cellular models. The majority of the entries into this table indicate that iPSCs are a better source for modelling drug metabolism than cancer cell lines but are not as biologically relevant as PHHs. Nonetheless, the utility of iPSCs in gene editing, drug development and other biological applications makes them an attractive cell source (Goldring *et al.*, 2017). Future research should continue to pursue enzyme kinetic studies to determine the metabolic utility of the SC-HLC model established here. There is also a need to carefully consider the drug compounds used as CYP substrates and the CYP enzymes that are important for drug metabolism modelling.

Table 4.1: Summary of literature on CYP enzyme drug substates used in drug metabolism studies.

CYP ENZYME	SUBSTRATE	METABOLITE	D.O.M	CELL TYPE	CONTROL	RESULTS	DERIVATION	REFERENCE
CYP1A2	**Phenacetin**	**Acetaminophen (Paracetamol)**	Day 4	iCell Hepatocytes	PHHs	19.5 (iHCs)	($) Cellular Dynamics in 2D Monolayer	(Lu et al., 2015)*
						10.6 (PHHs)		
			Day 3 (3D)	PHHs (3D)	PHHs (2D)	1100-1200 (3D)	($) Acquired from 3 Donors (KaLy Cell)	(Bell et al., 2018)***
			Day 7 (2D)+			500-600 (2D)		
			Day 7	PHHs (3D)	*in-vivo* and *in-vitro* references	0.8400 (Day 7)	Unknown (kept of 3T3-Swiss Albino)	(Ohkura et al., 2014)****
			Day 21++	HepG2 (3D & 2D)	HepRG (2D)	5-6 (2D)	($) 2D Monolayer and 3D collagen sandwich, Matrigel, and Alvetex scaffold (ECACC)	(Luckert et al., 2017)
						4-5 (Matrigel)		
						~4 (Collagen)		
						2-3 (Alvetex)		
			Day 18	iPSC-Derived Hepatocytes	PHHs	N/D (2D & Spheroids)	3D Spheroids & Bioreactor (iHCs) and 2D (PHHs)	(Meier et al., 2017)**
						1.0 (Bioreactor)		
						10.0 (PHHs)		
CYP2B6	**Bupropion**	**Hydroxybupropion**	Day 4	iCell Hepatocytes	PHHs	2.3 (iHCs)	($) Cellular Dynamics in 2D Monolayer	(Lu et al., 2015) *
						15.2 (PHHs)		
			Day 18	iPSC-Derived Hepatocytes	PHHs	No Detection	3D Spheroids & Bioreactor (iHCs) and 2D (PHHs)	(Meier et al., 2017)**
			Day 21++	HepG2 (3D & 2D)	HepRG (2D)	>0.1 (2D)	($) 2D Monolayer and 3D collagen sandwich, Matrigel, and Alvetex scaffold (ECACC)	(Luckert et al., 2017)
						>0.1 (Matrigel)		
						>0.1 (Collagen)		
						>0.1 (Alvetex)		
CYP2C9	**Diclofenac**	**4-Hydroxydiclofenac**	Day 4	iCell Hepatocytes	PHHs	26.1 (iHCs)	($) Cellular Dynamics in 2D Monolayer	(Lu et al., 2015)*
						111.5 (PHHs)		
			Day 2	PHHs (3D)	*in-vivo* and *in-vitro* references	0.1823 (Day 2)	Unknown (kept of 3T3-Swiss Albino)	(Ohkura et al., 2014)****
			Day 1 (2D)	HepG2 (3D)	HepG2 (2D)	~20 (Day 2)(3D)	American type tissue culture (ATCC). Monolayer (2D) and Spheroids (3D)	(Ramaiahgari et al., 2014)
	Tolbutamide	4-Hydroxytolbutamide	Day 13 (3D)	PHHs (3D)	PHHs (2D)	90-100 (3D)	($) Acquired from 3 Donors (KaLy Cell)	(Bell et al., 2018)***
			Day 3 (2D)+			70-80 (2D)		
			Day 21++	HepG2 (3D & 2D)	HepRG (2D)	0.05-0.06 (2D)	($) 2D Monolayer and 3D collagen sandwich, Matrigel, and Alvetex scaffold (ECACC)	(Luckert et al., 2017)
						~0.03 (Matrigel)		
						~0.03 (Collagen)		
						~0.0025 (Alvetex)		

For all D.O.M column, the experiments were dated from the day of seeding (PHHs) or the day of differentiation unless stated otherwise (see below). (+) Day 1-14 from spheroid or sandwich formation; Best day for 2D and 3D given. (++) Day 3 or 21 from 3D or 2D culture; Best day for 2D and 3D given. Results were measured multiple different ways, and these are listed as follows: (*) Results measured in pmol/min/mg Protein. (**) Results measured in pmol/h/ug DNA. (***) Results measured in pmol/h/million PHHs. (****) Results measured as a peak area ratio to an internal standard peak area. (*****) Results measured in nmol/h as a rate of formation. Red indicates the drug compounds that were suggested for use in enzyme kinetics studies.

Table 4.1 (Continued): Summary of literature on CYP enzyme drug substates used in drug

CYP2C19	S-Mephenytoin	4-Hydroxymephenytoin	Day 4	iCell Hepatocytes	PHHs	9.9 (iHCs)	($) Cellular Dynamics in 2D Monolayer	(Lu et al., 2015)*
						17.5 (PHHs)		
			Day 21++	HepG2 (3D & 2D)	HepRG (2D)	ND (2D)	($) 2D Monolayer and 3D collagen sandwich, Matrigel, and Alvetex scaffold (ECACC)	(Luckert et al., 2017)
						~0.04 (Matrigel)		
						ND (Collagen)		
						~0.02 (Alvetex)		
CYP2D6	**Bufuralol**	**1-Hydroybufuralol**	Day 2 (2D & 3D)	HepG2 (3D)	HepG2 (2D)	50-60 (Day 2)(3D)	American type tissue culture (ATCC). Monolayer (2D) and Spheroids (3D)	(Ramaiahgari et al., 2014)
						~20 (Day2)(2D)		*****
	Dextromethorphan	Dextrophan	Day 4	iCell Hepatocytes	PHHs	8.7 (iHCs)	($) Cellular Dynamics in 2D Monolayer	(Lu et al., 2015)*
						20.5 (PHHs)		
			Day 3 (3D)	PHHs (3D)	PHHs (2D)	80-90 (3D)	($) Acquired from 3 Donors (KaLy Cell)	(Bell et al., 2018)***
			Day 5 (2D)+			60-70 (2D)		
	Propafenone	4-Hydroypropafenone	Day 21++	HepG2 (3D & 2D)	HepRG (2D)	N/A (2D)	($) 2D Monolayer and 3D collagen sandwich, Matrigel, and Alvetex scaffold (ECACC)	(Luckert et al., 2017)
						N/A (Matrigel)		
						N/A (Collagen)		
						N/A (Alvetex)		
	Propranolol	Hydroxylpropranolol	Day 2	PHHs (3D)	*in-vivo* and *in-vitro* references	0.0022(Day 2)	Unknown (kept of 3T3-Swiss Albino)	(Ohkura et al., 2014)****
CYP3A4/5	**Midazolam**	**1'-Hydroxymidazolam**	Day 4	iCell Hepatocytes	PHHs	62.7 (iHCs)	($) Cellular Dynamics in 2D Monolayer	(Lu et al., 2015) *
						41.9 (PHHs)		
			Day 18	iPSC-Derived Hepatocytes	PHHs	>0.1 (2D)	3D Spheroids & Bioreactor (iHCs) and 2D (PHHs)	(Meier et al., 2017)**
						0.1 (Spheroids)		
						0.4-0.5 (Bioreactor)		
						20.0 (PHHs)		
			Day 3 (3D)	PHHs (3D)	PHHs (2D)	1700-1800 (3D)	($) Acquired from 3 Donors (KaLy Cell)	(Bell et al., 2018)***
			Day 5 (2D)+			400-500 (2D)		
			Day 2	PHHs (3D)	*in-vivo* and *in-vitro* references	2.4884 (Day 2)	Unknown (kept of 3T3-Swiss Albino)	(Ohkura et al., 2014)****
			Day 2 (2D & 3D)	HepG2 (3D)	HepG2 (2D)	35-40 (Day 2)(3D)	American type tissue culture (ATCC). Monolayer (2D) and Spheroids (3D)	(Ramaiahgari et al., 2014)
						5-10 (Day2)(2D)		*****
			Day 21++	HepG2 (3D & 2D)	HepRG (2D)	~0.8 (2D)	($) 2D Monolayer and 3D collagen sandwich, Matrigel, and Alvetex scaffold (ECACC)	(Luckert et al., 2017)
						~1.0 (Matrigel)		
						~1.2 Collagen)		
						~0.8 (Alvetex)		

For all D.O.M column, the experiments were dated from the day of seeding (PHHs) or the day of differentiation unless stated otherwise (see below). (+) Day 1-14 from spheroid or sandwich formation; Best day for 2D and 3D given. (++) Day 3 or 21 from 3D or 2D culture; Best day for 2D and 3D given. Results were measured multiple different ways, and these are listed as follows: (*) Results measured in pmol/min/mg Protein. (**) Results measured in pmol/h/ug DNA. (***) Results measured in pmol/h/million PHHs. (****) Results measured as a peak area ratio to an internal standard peak area. (*****) Results measured in nmol/h as a rate of formation. Red indicates the drug compounds that were suggested for use in enzyme kinetics studies.

4.4 POTENTIAL SOURCES OF CELLS TO DEVELOP IPSC MODELS FOR DRUG METABOLISM

Human fibroblasts are the main cell source for iPSC generation, however, numerous other tissues, such as the heart, lung and in periodontal ligaments and even commercial lines of these are available as alternatives to dermal fibroblasts (Raab *et al.*, 2014). There also are other cell sources such as peripheral blood cells (PBCs) and urine-derived cells (HUCs) that are potential alternatives to dermal fibroblasts (Sacco *et al.*, 2019). Finding the correct cell source is crucial in establishing iPSC lines. Some of the requirements for a good cell source are being easily reprogrammable via virus transduction, easily sourced from patient or commercial means, and cost-effective for model upscaling and cell culturing.

4.4.1 Peripheral Blood & Urine-Derived Cells

Peripheral blood cells (PBCs) are an alternative cell source for generating iPSC lines and are acquired via a technique called PBC harvesting (Raab *et al.*, 2014). One frequently used method for generating PBC-derived stem cell lines is through acquiring mobilised $CD34^+$ PBCs via granulocyte colony-stimulating factor (G-CSF) stimulation in cell donors (Haase *et al.*, 2009). G-CSF is a cytokine that has been identified as a critical regulator of the haematopoietic system and is involved in stimulating the production of haematopoietic stem cells and subsequently increasing the production of white blood cell (Souza *et al.*, 1986). Cell donors are administered G-CSF via injection every day for a period of five days and $CD34^+$ cells are harvest by apheresis. The acquired cells are reprogrammed by viral transductions to generate stem cell lines and differentiated in the tissue of interest afterwards donors (Haase *et al.*, 2009). The benefit of PBCs acquired through this method is that the cells are very similar to human ESCs on both molecular and functional levels, as well as being easier to maintain and having short timespans for reprogramming (Haase *et al.*, 2009). However, some of the shortcomings of using G-CSF for PBC-derived stem cell generation are bone pain, fatigue and nausea (Raab *et al.*, 2014).

Urine has been used as a source of exfoliated renal epithelial cells which frequently detach from the kidney and urinary tract during the tissue turnover. Urine-derived cells (HUCs) are excreted with the urine and are fully functional when found within the urine excretion (Zhou *et al.*, 2011). These cells have a squamous cell morphology and are derived from epithelial tissue, which is important as HUCs do not need to undergo a mesenchymal-to epithelial transition (MET) step (Zhou *et al.*, 2011). The benefits of HUCs as a cell source for the generation of iPSCs are that they are easily acquired by non-invasive methods and do not need any medical personnel to supervise acquiring these cells. HUCs are also regularly

collected in clinical settings as a tool for both screening and diagnosis which highlight the ease of access to these cells (Raab *et al.*, 2014). HUCs also have a fairly high reprogramming efficiency that ranges between 0.1-4.0% and ranks them as a better cell source for iPSC reprogramming, however they do present with poor reprogramming efficiency after five passages (Zhou *et al.*, 2011).

The two cell sources described above are suggested alternatives to acquiring dermal fibroblasts from skin biopsies. They have been proposed as areas of investigation for future research to increase the scalability and remove the need for minimally invasive procedures such as a skin biopsy. However, at the current time of writing the best reprogramming methods still use dermal fibroblasts as a cell source. Though this is the case, it would be of great utility to invest in numerous cell sources to determine which cell source would be best for establishing such cellular models.

4.5 OTHER DRUG METABOLISM MODELLING SYSTEMS FOR USE WITH SC-HLCs

Advances in technology have led to new avenues to explore drug metabolism in more cost effective and efficient ways. From computational models to bioengineering devices, the information era has brought on a new frontier where technology and biology can now work together to solve some of the toughest problems in healthcare. These technologies can also be used to improve the use of iPSCs in drug metabolism modelling. This ranges from highlighting genetic variants associated with drug metabolism, investigating molecular interaction between enzymes and substrates, and microfluidic platforms for emulate *in vivo* organ function. Combining these technologies and methodologies with 3D culturing techniques can increase the accuracy and scope of cellular models established to investigate drug metabolism.

4.5.1 Molecular Dynamics

Molecular dynamics takes advantage of the knowledge about enzyme essential bond forming capabilities, the channels formed, and the overall enzyme plasticity. This is achieved by creating millisecond molecular dynamics simulations of the interactions that the enzyme will undertake with or without a substrate (Don and Smieško, 2018). The process is made possible by the increase in computing power and the great amounts of data pertaining to the protein structures. When applying this to CYP enzymes one can attain insight into the protein structure of the enzymes of choice. Overall investigation of CYP enzymes tend to be similar with changes mainly found within the loops which alter the flexibility of the enzymes, such as in CYP2D6. These changes are often associated with the enzyme activity of the CYP enzyme being studied (Don and Smieško, 2018). For CYP2D6, altered enzyme activity observed from different variants can be attributed to hydrogen bonding changes at the active site, structural differences affecting structure flexibility and binding pocket shape and position. In *CYP2D6*17*, there is a polymorphism that results in a threonine to isoleucine change (T107I) causing problems with the hydrogen bonding and increasing the local hydrophobicity resulting in reduced B-C loop flexibility (Don and Smieško, 2018).

This technology can aid in observing difference between wild-type and variant isozymes and allow *in silico* screening of the substrate-enzyme interactions via molecular dynamic simulations. Combining these technologies with 3D cell culturing, can increase the accuracy of drug metabolism modelling. With regards to CYP enzymes, molecular dynamics can produce better insight into the protein-protein interactions around substrate binding to these enzymes. This can be used to better screen for candidate variants of *CYP* genes that affect

enzyme functioning for *in vitro* cell models, such as with the SC-HLC model we proposed in this study (Don and Smieško, 2018).

4.5.2 *In Silico* Modelling

Using *in silico* and *in vivo* approaches as a preliminary tool to investigate drug compound metabolism and screen for candidate drug compounds for optimal development. The implementation of computational models to further aid in investigating CYP enzyme interactions with drug compounds. This would also assist in understanding data generated from *in vivo* models, such as the SC-HLC model in this study.

There are numerous computational tools that can be used to predict the PD and PK of drug-metabolising enzymes (Ferreira and Andricopulo, 2019). The effective and accurate modelling of these factors are still a serious problem for drug development, even with the use of cellular models. Over the years, the impact of PK and PD attrition rates of drugs has fallen and this is partly due to the implementation of PK control programs in the development process (Ferreira and Andricopulo, 2019). Using *in silico* approaches in drug development can predict the interactions of substrates with enzymes and exclude inappropriate compounds from the development process. Software such as CypReact is able to now give accurate predictions of compounds that will likely undergo metabolism when encountering any of the CYP enzymes (Ferreira and Andricopulo, 2019). This program was able to make predictions of drug-enzyme interactions that in the best fit results ranged from 83-92%. The methods that were best for building the classification models were random forces (RF) (best for CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C19, CYP2E1 and CYP3A4 using program FAME) and ensemble method (best for CYP2C9 and CYP2D6 using CypReact). *In silico* prediction has a large role to play in the drug development process, and can provide very effective and accessible ways to lead-optimisation for drug compounds (Ferreira and Andricopulo, 2019).

4.5.3 Microfluidic Systems

The current drug development processes are subject to high failure rates in transition of candidate drug selection to clinically approved drugs. This is partly due to poor cellular and animal models which are often inaccurate and lack adequacy for pharmacological studies (Bhise *et al.*, 2016). The need for better modelling systems has caused new interest in engineered tissues and tissue constructs that can mimic the cell interaction with other cells and the extracellular matrix. Microfluidics also allows for the creation of perfusion systems that allow cells to experience a continuous perfusion of nutrients and oxygen to the cells. All of these elements allow for the cells to experience numerous physiological stimuli that are able to evoke similar *in vivo* responses from cells (Bhise *et al.*, 2016). This has allowed for

the use of these microfluidic systems, also known as organ-on-a-chip, in drug metabolism and development studies (Bhise *et al.*, 2016). Bioprinting is an important technology that can aid in the creation of such models and gives specificity to the ECM form as well as allowing for mimicking of *in vivo* structure vascularisation. In 3D prints of liver lobule units comprised of spheroids, there is increased expression of CYP enzymes as the structure allows for more *in vivo* biophysical properties. *CYP3A4* expression in this system was seven-fold compared to the 2D monolayer of the same HepRG cells. This shows the power and the utility of such systems to aid in accurately modelling *in vivo* drug metabolism (Grix *et al.*, 2018).

4.6 CONCLUSION

Drug development is a vibrant field in biology, with various new technologies pushing towards the implementation of better models for drug response. Due to high genetic diversity in African populations, it is even more crucial to have accurate models to assist in research of diseases and drug response. This can aid in improving drug development endeavours across the continent. As knowledge grows on the genetic diversity exhibited in African populations, more investigation on the effects of ADME genetic variants should be carried out. Cellular models, such as the SC-HLC model from this study, will be useful in determining the physiological effects of these ADME genetic variants. There is also a need for better systems of pharmacological study, as there is a high burden of disease and adverse side effects on the African continent. With inadequate amounts of R&D occurring on the African continent, it is critical that we find cost-effective ways to investigate drug efficacy and safety. Use of iPSC models for drug metabolism studies can become important in establishing cost-effective and sizeable solutions for ADRs. Translating this into effective pharmaceutical interventions should assist in creating better health outcomes in African populations.

In conclusion, this study aimed to *assess the frequency of the CYP2D6*17 variant in African populations and to evaluate its enzyme activity on the activation of tamoxifen to endoxifen in iPSC-derived hepatocyte-like cells (SC-HLC)*. Establishment of the SC-HLC model was interrupted by unforeseen circumstances, but future endeavours will see that iPSCs are characterised and differentiated into hepatocyte-like cells. Nonetheless, this study sought out to establish one of the first SC-HLC drug metabolism models with the *CYP* gene variant *CYP2D6*17*. Hopefully, more drug metabolism research will be pursued utilising this model, as well as research into emulating the *in vivo* functioning of the liver. Such exciting research will be key to better health outcomes and pharmaceutical R&D on the African continent.

REFERENCES & APPENDICES

5. REFERENCES & APPENDICES

5.1 REFERENCES

Abcam (2015) 'Counting cells using a hemocytometer', pp. 1–3. doi: 10.1038/sj.ejhg.5200954.

Achour, B., Barber, J. and Rostami-Hodjegan, A. (2012) 'Correction to "Cytochrome P450 pig liver pie: Determination of individual cytochrome P450 isoform contents in microsomes from two pig livers using liquid chromatography in conjunction with mass spectroscopy" (Drug Metabolism and Disposition (2011) 39, (21', *Drug Metabolism and Disposition*, 40(1), p. 227. doi: 10.1124/dmd.40.1.227.

Ahern, T. P. *et al.* (2017) 'Cytochrome P-450 2D6 (CYP2D6) Genotype and Breast Cancer Recurrence in Tamoxifen-Treated Patients: Evaluating the Importance of Loss of Heterozygosity', *American Journal of Epidemiology*, 185(2), pp. 75–85. doi: 10.1093/aje/kww178.

Aklillu, E. *et al.* (1996) 'Frequent distribution of ultrarapid metabolizers of debrisoquine in an ethiopian population carrying duplicated and multiduplicated functional CYP2D6 alleles.', *The Journal of pharmacology and experimental therapeutics*, 278(1), pp. 441–446.

Almazroo, O. A., Miah, M. K. and Venkataramanan, R. (2017) 'Drug Metabolism in the Liver', *Clinics in Liver Disease*, 21(1), pp. 1–20. doi: 10.1016/j.cld.2016.08.001.

Altshuler, D. L. *et al.* (2010) 'A map of human genome variation from population-scale sequencing', *Nature*, 467(7319), pp. 1061–1073. doi: 10.1038/nature09534.

Andersson, L. C. *et al.* (2018) 'Critical differences in drug metabolic properties of human hepatic cellular models, including primary human hepatocytes, stem cell derived hepatocytes, and hepatoma cell lines', *Biochemical Pharmacology*. Elsevier, 155(April), pp. 124–140. doi: 10.1016/j.bcp.2018.06.026.

Asha, S. and Vidyavathi, M. (2010) 'Role of Human Liver Microsomes in In Vitro Metabolism of Drugs—A Review', *Applied Biochemistry and Biotechnology*, 160(6), pp. 1699–1722. doi: 10.1007/s12010-009-8689-6.

Ashley, E. A. (2016) 'Towards precision medicine.', *Nature reviews. Genetics*, 17(9), pp. 507–22. doi: 10.1038/nrg.2016.86.

Auton, A. *et al.* (2015) 'A global reference for human genetic variation', *Nature*, 526(7571), pp. 68–74. doi: 10.1038/nature15393.

Bapiro, T. E. *et al.* (2002) 'The molecular and enzyme kinetic basis for the diminished activity of the cytochrome P450 2D6.17 (CYP2D6.17) variant: Potential implications for CYP2D6 phenotyping studies and the clinical use of CYP2D6 substrate drugs in some African populations', *Biochemical Pharmacology*, 64(9), pp. 1387–1398. doi: 10.1016/S0006-2952(02)01351-5.

Bell, C. C. *et al.* (2016) 'Characterization of primary human hepatocyte spheroids as a model system for drug-induced liver injury , liver function and disease', *Nature Publishing Group*. Nature Publishing Group, (May), pp. 1–13. doi: 10.1038/srep25187.

Bell, C. C. *et al.* (2017) 'Transcriptional, functional, and mechanistic comparisons of stem cell-derived hepatocytes, HepaRG cells, and three-dimensional human hepatocyte spheroids as predictive in vitro systems for drug-induced liver injury', *Drug Metabolism and Disposition*, 45(4), pp. 419–429. doi: 10.1124/dmd.116.074369.

Bell, C. C. *et al.* (2018) 'Comparison of hepatic 2D sandwich cultures and 3d spheroids for long-term toxicity applications: A multicenter study', *Toxicological Sciences*, 162(2), pp. 655–666. doi: 10.1093/toxsci/kfx289.

Berger, D. R. *et al.* (2015) 'Enhancing the functional maturity of induced pluripotent stem cell-derived human hepatocytes by controlled presentation of cell-cell interactions in vitro.', *Hepatology (Baltimore, Md.)*, 61(4), pp. 1370–81. doi: 10.1002/hep.27621.

Bernauer, U. *et al.* (2006) 'Characterisation of the xenobiotic-metabolizing Cytochrome P450 expression pattern in human lung tissue by immunochemical and activity determination', *Toxicology Letters*, 164(3), pp. 278–288. doi: 10.1016/j.toxlet.2006.01.007.

Bertholee, D., Maring, J. G. and van Kuilenburg, A. B. P. (2017) 'Genotypes Affecting the Pharmacokinetics of Anticancer Drugs', *Clinical Pharmacokinetics*, 56(4), pp. 317–337. doi: 10.1007/s40262-016-0450-z.

Bhathena, A. *et al.* (2007) 'Frequency of the frame-shifting CYP2D7 138delT polymorphism in a large, ethnically diverse sample population', *Drug Metabolism and Disposition*, 35(8), pp. 1251–1253. doi: 10.1124/dmd.107.014993.

Bhise, N. S. *et al.* (2016) 'A liver-on-a-chip platform with bioprinted hepatic spheroids', *Biofabrication*. IOP Publishing, 8(1), p. 014101. doi: 10.1088/1758-5090/8/1/014101.

Bissig, K.-D. *et al.* (2018) 'P450-Humanized and Human Liver Chimeric Mouse Models for Studying Xenobiotic Metabolism and Toxicity', *Drug Metabolism and Disposition*, 46(11), pp. 1734–1744. doi: 10.1124/dmd.118.083303.

Brodie, B. B. *et al.* (1955) 'Detoxication of Drugs and Other Foreign Compounds by Liver Microsomes', *Science*, 121(3147), pp. 603–604. doi: 10.1126/science.121.3147.603.

Chaudhuri, S. and West, A. (2015) 'Can local producers compete with low-cost imports? A simulation study of pharmaceutical industry in low-income Africa', *Innovation and Development*, 5(1), pp. 23–38. doi: 10.1080/2157930X.2014.921273.

Choi, J. M. *et al.* (2015) 'HepG2 cells as an in vitro model for evaluation of cytochrome P450 induction by xenobiotics', *Archives of Pharmacal Research*. Pharmaceutical Society of Korea, 38(5), pp. 691–704. doi: 10.1007/s12272-014-0502-6.

Choudhury, A. *et al.* (2018) 'African genetic diversity provides novel insights into evolutionary history and local adaptations', *Human molecular genetics*, 27(R2), pp. R209–R218. doi: 10.1093/hmg/ddy161.

Crews, K. R. *et al.* (2012) 'Clinical pharmacogenetics implementation consortium (CPIC) guidelines for codeine therapy in the context of cytochrome P450 2D6 (CYP2D6) genotype', *Clinical Pharmacology and Therapeutics*. Nature Publishing Group, 91(2), pp. 321–326. doi: 10.1038/clpt.2011.287.

Dalton, R. *et al.* (2019) ' Interrogation of CYP2D6 structural variant alleles improves the correlation between CYP2D6 genotype and CYP2D6-mediated metabolic activity ', *Clinical and Translational Science*. doi: 10.1111/cts.12695.

Daly, A. K. *et al.* (1996) 'Nomenclature for human CYP2D6 alleles', *Pharmacogenetics*, 6(3), pp. 193–201. doi: 10.1097/00008571-199606000-00001.

Davies, C. *et al.* (2013) 'Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial', *The Lancet*, 381(9869), pp. 805–816. doi: 10.1016/S0140-6736(12)61963-1.

Dayer, P. *et al.* (1984) 'Characterization of a common genetic defect of cytochrome P-450 function (debrisoquine-sparteine type polymorphism) — Increased michaelis constant (km) and loss of stereoselectivity of bufuralol 1′-hydroxylation in poor metabolizers', *Biochemical and Biophysical Research Communications*, 125(1), pp. 374–380. doi: 10.1016/S0006-291X(84)80378-2.

Denny, J. C. *et al.* (2018) 'The Influence of Big (Clinical) Data and Genomics on Precision Medicine and Drug Development', *Clinical Pharmacology and Therapeutics*, 103(3), pp. 409–418. doi: 10.1002/cpt.951.

Desta, Z. *et al.* (2019) 'Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2B6 and Efavirenz-Containing Antiretroviral Therapy', *Clinical Pharmacology and Therapeutics*, 106(4), pp. 726–733. doi: 10.1002/cpt.1477.

Distlerath, L. M. *et al.* (1985) 'Purification and characterization of the human liver cytochromes P-450 involved in debrisoquine 4-hydroxylation and phenacetin O-deethylation, two prototypes for genetic polymorphism in oxidative drug metabolism', *Journal of Biological Chemistry*, 260(15), pp. 9057–9067.

Dodgen, T. M. *et al.* (2016) 'Pharmacogenetic comparison of CYP2D6 predictive and measured phenotypes in a South African cohort', *Pharmacogenomics Journal*. Nature Publishing Group, 16(6), pp. 566–572. doi: 10.1038/tpj.2015.76.

Don, C. G. and Smieško, M. (2018) 'Microsecond MD simulations of human CYP2D6 wild-type and five allelic variants reveal mechanistic insights on the function.', *PloS one*, 13(8), p. e0202534. doi: 10.1371/journal.pone.0202534.

Dong, A. N. *et al.* (2018) 'Site-Directed Mutagenesis of Cytochrome P450 2D6 and 2C19 Enzymes: Expression and Spectral Characterization of Naturally Occurring Allelic Variants', *Applied Biochemistry and Biotechnology*. Applied Biochemistry and Biotechnology, pp. 1–13. doi: 10.1007/s12010-018-2728-0.

Duffus, J. H., Nordberg, M. and Templeton, D. M. (2007) 'Glossary of terms used in toxicology, 2nd edition (IUPAC Recommendations 2007)', *Pure and Applied Chemistry*, 79(7), pp. 1153–1344. doi: 10.1351/pac200779071153.

Dugger, S. A., Platt, A. and Goldstein, D. B. (2018) 'Drug development in the era of precision medicine', *Nature Reviews Drug Discovery*. Nature Publishing Group, 17(3), pp. 183–196. doi: 10.1038/nrd.2017.226.

Eichelbaum, M. *et al.* (1979) 'Defective N-oxidation of sparteine in man: a new pharmacogenetic defect.', *European journal of clinical pharmacology*, 16(3), pp. 183–7. doi: 10.1007/bf00562059.

Eichelbaum, M. *et al.* (1987) 'Chromosomal assignment of human cytochrome P-450 (debrisoquine/sparteine type) to chromosome 22.', *British Journal of Clinical Pharmacology*, 23(4), pp. 455–458. doi: 10.1111/j.1365-2125.1987.tb03075.x.

Endrizzi, K. *et al.* (2002) 'Discriminative quantification of cytochrome P4502D6 and 2D7/8 pseudogene expression by TaqMan real-time reverse transcriptase polymerase chain reaction', *Analytical Biochemistry*, 300(2), pp. 121–131. doi: 10.1006/abio.2001.5453.

Ferreira, L. L. G. and Andricopulo, A. D. (2019) 'ADMET modeling approaches in drug discovery', *Drug Discovery Today*. Elsevier Ltd. doi: 10.1016/j.drudis.2019.03.015.

Fiorotto, R. *et al.* (2018) 'Liver diseases in the dish: iPSC and organoids as a new approach to modeling liver diseases', *Biochimica et Biophysica Acta - Molecular Basis of Disease*. Elsevier, (July), pp. 1–9. doi: 10.1016/j.bbadis.2018.08.038.

Fisher, B. *et al.* (2005) 'Tamoxifen for the prevention of breast cancer: Current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study', *Journal of the National Cancer Institute*, 97(22), pp. 1652–1662. doi: 10.1093/jnci/dji372.

Flicek, P. *et al.* (2011) 'Ensembl 2011', *Nucleic Acids Research*, 39(SUPPL. 1), pp. 800–806. doi: 10.1093/nar/gkq1064.

Fricke-Galindo, I. *et al.* (2016) 'Interethnic variation of CYP2C19 alleles, "predicted" phenotypes and "measured" metabolic phenotypes across world populations', *Pharmacogenomics Journal*, 16(2), pp. 113–123. doi: 10.1038/tpj.2015.70.

Fritz, A. *et al.* (2019) 'Expression of clinically relevant drug-metabolizing enzymes along the human intestine and their correlation to drug transporters and nuclear receptors: An intra-subject analysis', *Basic and Clinical Pharmacology and Toxicology*, 124(3), pp. 245–255. doi: 10.1111/bcpt.13137.

Gaedigk, A. (2013) 'Complexities of CYP2D6 gene analysis and interpretation', *International Review of Psychiatry*, 25(5), pp. 534–553. doi: 10.3109/09540261.2013.825581.

Gaedigk, A. *et al.* (2015) 'SNP genotyping using TaqMan® technology: The CYP2D6∗17 assay conundrum', *Scientific Reports*, 5, pp. 1–9. doi: 10.1038/srep09257.

Gaedigk, A., Turner, A., *et al.* (2019) 'Characterization of Reference Materials for Genetic Testing of CYP2D6 Alleles', *The Journal of Molecular Diagnostics*. American Society for Investigative Pathology and the Association for Molecular Pathology, 21(6), pp. 1034–1052. doi: 10.1016/j.jmoldx.2019.06.007.

Gaedigk, A., Sangkuhl, K., *et al.* (2019) 'The Evolution of PharmVar', *Clinical Pharmacology and Therapeutics*, 105(1), pp. 29–32. doi: 10.1002/cpt.1275.

Gaedigk, A. and Coetsee, C. (2008) 'The CYP2D6 gene locus in South African Coloureds: Unique allele distributions, novel alleles and gene arrangements', *European Journal of Clinical Pharmacology*, 64(5), pp. 465–475. doi: 10.1007/s00228-007-0445-7.

Gaedigk, A., Gaedigk, R. and Leeder, J. S. (2005) 'CYP2D7 splice variants in human liver and brain: Does CYP2D7 encode functional protein?', *Biochemical and Biophysical Research Communications*, 336(4), pp. 1241–1250. doi: 10.1016/j.bbrc.2005.08.255.

Goetz, M. P. *et al.* (2007) 'The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen', *Breast Cancer Research and Treatment*, 101(1), pp. 113–121. doi: 10.1007/s10549-006-9428-0.

Goh, L. L. *et al.* (2017) 'Analysis of genetic variation in CYP450 genes for clinical implementation', *PLoS ONE*, 12(1). doi: 10.1371/journal.pone.0169233.

Goldring, C. *et al.* (2017) 'Stem cell-derived models to improve mechanistic understanding and prediction of human drug-induced liver injury.', *Hepatology (Baltimore, Md.)*, 65(2), pp. 710–721. doi: 10.1002/hep.28886.

Gomez-Lechon, M. *et al.* (2004) 'Human Hepatocytes in Primary Culture: The Choice to Investigate Drug Metabolism in Man', *Current Drug Metabolism*, 5(5), pp. 443–462. doi: 10.2174/1389200043335414.

Grix, T. *et al.* (2018) 'Bioprinting Perfusion-Enabled Liver Equivalents for Advanced Organ-on-a-Chip Applications', *Genes*, 9(4), p. 176. doi: 10.3390/genes9040176.

Guengerich, F. P. (1994) 'Catalytic selectivity of human cytochrome P450 enzymes: relevance to drug metabolism and toxicity', *Toxicology Letters*, 70(2), pp. 133–138. doi: 10.1016/0378-4274(94)90156-2.

Guengerich, F. P., Waterman, M. R. and Egli, M. (2016) 'Recent Structural Insights into Cytochrome P450 Function', *Trends in Pharmacological Sciences*. Elsevier Ltd, 37(8), pp. 625–640. doi: 10.1016/j.tips.2016.05.006.

Guo, L. *et al.* (2011) 'Similarities and differences in the expression of drug-metabolizing enzymes between human hepatic cell lines and primary human hepatocytes', *Drug Metabolism and Disposition*, 39(3), pp. 528–538. doi: 10.1124/dmd.110.035873.

Gurdasani, D. *et al.* (2015) 'The African Genome Variation Project shapes medical genetics in Africa', *Nature*. Nature Publishing Group, 517(7534), pp. 327–332. doi: 10.1038/nature13997.

H3Africa (2011) 'Harnessing Genomic Technologies Toward Improving Health in Africa : Opportunities and Challenges', *Http://H3Africa.Org/About/White-Paper*, (January). Available at: http://h3africa.org/images/PDF/h3africa_whitepaper.pdf.

Haase, A. *et al.* (2009) 'Generation of Induced Pluripotent Stem Cells from Human Cord Blood', *Cell Stem Cell*, 5(4), pp. 434–441. doi: 10.1016/j.stem.2009.08.021.

Harvey, K. F. *et al.* (2001) 'The Nedd4-like Protein KIAA0439 is a Potential Regulator of the Epithelial Sodium Channel', *Journal of Biological Chemistry*, 276(11), pp. 8597–8601. doi: 10.1074/jbc.C000906200.

Ibidi (2015) 'Live / dead staining with FDA and PI', *Application Note 33*, p. 4. Available at: http://ibidi.com/fileadmin/support/application_notes/AN33_Live_Dead_staining_with_FDA_and_PI.pdf.

Jin, Y. *et al.* (2005) 'CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer

treatment', *Journal of the National Cancer Institute*, 97(1), pp. 30–39. doi: 10.1093/jnci/dji005.

Johnson, J. A. *et al.* (2011) 'Clinical pharmacogenetics implementation consortium guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing', *Clinical Pharmacology and Therapeutics*. Nature Publishing Group, 90(4), pp. 625–629. doi: 10.1038/clpt.2011.185.

Jones, H. M. *et al.* (2015) 'Physiologically based pharmacokinetic modeling in drug discovery and development: a pharmaceutical industry perspective', *Clinical pharmacology and therapeutics*, 97(3), pp. 247–262. doi: 10.1002/cpt.37.

Jones, H. and Rowland-Yeo, K. (2013) 'Basic Concepts in Physiologically Based Pharmacokinetic Modeling in Drug Discovery and Development', *CPT: Pharmacometrics & Systems Pharmacology*, 2(8), p. e63. doi: 10.1038/psp.2013.41.

Jones, K. H. and Senft, J. A. (1985) 'An improved method to determine cell viability by simultaneous staining with fluorescein diacetate-propidium iodide', *Journal of Histochemistry and Cytochemistry*, 33(1), pp. 77–79. doi: 10.1177/33.1.2578146.

Jukic, M. M. *et al.* (2018) 'Functional characterization of CYP2D7 gene variants', *Pharmacogenomics*, 19(12), pp. 931–936. doi: 10.2217/pgs-2018-0065.

Kagimoto, M. *et al.* (1990) 'Multiple Mutations of the Human Cytochrome P450IID6 (CYP2D6) in Poor Metabolizers of Debrisoquine', *The Journal of Biological Chemistry*, 265(26), pp. 17209–17214.

Kim, J. Y. *et al.* (2014) 'Screening for 392 polymorphisms in 141 pharmacogenes.', *Biomedical reports*, 2(4), pp. 463–476. doi: 10.3892/br.2014.272.

Kimura, S. *et al.* (1989) 'The human debrisoquine 4-hydroxylase (CYP2D) locus: Sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene', *American Journal of Human Genetics*, 45(6), pp. 889–904.

Kirchheiner, J. *et al.* (2011) 'CYP2D6 in the brain: Genotype effects on resting brain perfusion', *Molecular Psychiatry*. Nature Publishing Group, 16(3), pp. 333–341. doi: 10.1038/mp.2010.42.

Kiyotani, K. *et al.* (2012) 'Pharmacogenomics of Tamoxifen: Roles of Drug Metabolizing Enzymes and Transporters', *Drug Metabolism and Pharmacokinetics*, 27(1), pp. 122–131. doi: 10.2133/dmpk.DMPK-11-RV-084.

Klingenberg, M. (1958) 'Pigments of rat liver microsomes', *Archives of Biochemistry and Biophysics*, 75(2), pp. 376–386. doi: 10.1016/0003-9861(58)90436-3.

Kola, I. and Landis, J. (2004) 'Can the pharmaceutical industry reduce attrition rates?', *Nature reviews. Drug discovery*, 3(8), pp. 711–5. doi: 10.1038/nrd1470.

Kumar, G. N. and Surapaneni, S. (2001) 'Role of drug metabolism in drug discovery and development.', *Medicinal research reviews*, 21(5), pp. 397–411. doi: 10.1002/chin.200146279.

LeCluyse, E. L. *et al.* (2012) 'Organotypic liver culture models: Meeting current challenges in toxicity testing', *Critical Reviews in Toxicology*, 42(6), pp. 501–548. doi: 10.3109/10408444.2012.682115.

Lennard, M. S. *et al.* (1982) 'Oxidation Phenotype — A Major Determinant of Metoprolol Metabolism and Response', *New England Journal of Medicine*, 307(25), pp. 1558–1560. doi: 10.1056/NEJM198212163072505.

Li, A. P. (2005) 'Preclinical in vitro screening assays for drug-like properties', *Drug Discovery Today: Technologies*, 2(2), pp. 179–185. doi: 10.1016/j.ddtec.2005.05.024.

Li, T. and Chiang, J. Y. L. (2006) 'Rifampicin induction of CYP3A4 requires pregnane X receptor cross talk with hepatocyte nuclear factor 4α and coactivators, and suppression of small heterodimer partner gene expression', *Drug Metabolism and Disposition*, 34(5), pp. 756–764. doi: 10.1124/dmd.105.007575.

Luckert, C. *et al.* (2017) 'Comparative analysis of 3D culture methods on human HepG2 cells', *Archives of Toxicology*. Springer Berlin Heidelberg, 91(1), pp. 393–406. doi: 10.1007/s00204-016-1677-z.

Luo, G. *et al.* (2002) 'CYP3A4 Induction by Drugs: Correlation between a Pregnane X Receptor Reporter Gene Assay and CYP3A4 Expression in Human Hepatocytes', *Drug Metabolism and Disposition*, 30(7), pp. 795–804. doi: 10.1124/dmd.30.7.795.

Lynch, T. and Price, A. (2007) 'The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects', *American Family Physician*, 76(3), pp. 391–396. doi: 10.1046/j.1365-2125.1999.00073.x.

MacRae, C. A. and Peterson, R. T. (2015) 'Zebrafish as tools for drug discovery', *Nature Reviews Drug Discovery*, 14(10), pp. 721–731. doi: 10.1038/nrd4627.

Mahgoub, A. *et al.* (1977) 'Polymorphic hydroxylation of Debrisoquine in man.', *Lancet (London, England)*, 2(8038), pp. 584–6. doi: 10.1016/s0140-6736(77)91430-1.

Marez, D. *et al.* (1997) 'Polymorphism of the cytochrome P450 CYP2D6 gene in a European population: Characterization of 48 mutations and 53 alleles, their frequencies and evolution', *Pharmacogenetics*, pp. 193–202. doi: 10.1097/00008571-199706000-00004.

Masimirembwa, C. *et al.* (1996) 'A novel mutant variant of the CYP2D6 gene (CYP2D6*17) common in a black African population: association with diminished debrisoquine hydroxylase activity.', *British journal of clinical pharmacology*, 42(6), pp. 713–9. doi: 10.1046/j.1365-2125.1996.00489.x.

Matimba, A. *et al.* (2009) 'Novel variants of major drug-metabolising enzyme genes in diverse African populations and their predicted functional effects.', *Human genomics*, 3(2), pp. 169–190. doi: 10.1186/1479-7364-3-2-169.

McGivern, J. V. and Ebert, A. D. (2014) 'Exploiting pluripotent stem cell technology for drug discovery, screening, safety, and toxicology assessments', *Advanced Drug Delivery Reviews*. Elsevier B.V., pp. 170–178. doi: 10.1016/j.addr.2013.11.012.

Medhasi, S. *et al.* (2016) 'Pharmacogenomic Study Reveals New Variants of Drug Metabolizing Enzyme and Transporter Genes Associated with Steady-State Plasma Concentrations of Risperidone and 9-Hydroxyrisperidone in Thai Autism Spectrum Disorder Patients.', *Frontiers in pharmacology*, 7(December), p. 475. doi: 10.3389/fphar.2016.00475.

Meier, F. *et al.* (2017) 'Hepatic differentiation of human iPSCs in different 3D models: A comparative study.', *International journal of molecular medicine*, 40(6), pp. 1759–1771. doi: 10.3892/ijmm.2017.3190.

Meseguer-Ripolles, J. *et al.* (2018) 'Pluripotent Stem Cell-Derived Human Tissue: Platforms to Evaluate Drug Metabolism and Safety', *The AAPS Journal*. The AAPS Journal, 20(1), p. 20. doi: 10.1208/s12248-017-0171-8.

Montagutelli, X. (2015) 'Animal models are essential to biological research : issues and perspectives', 1, pp. 4–6.

Muriel, P. (ed.) (2017) *Liver Pathophysiology: Therapies and Antioxidants*. San Diego: Academic Press (Elsevier). Available at: https://linkinghub.elsevier.com/retrieve/pii/B9780128042748000072.

Muroi, Y. *et al.* (2014) 'Functional Characterization of Wild-type and 49 CYP2D6 Allelic Variants for N-Desmethyltamoxifen 4-Hydroxylation Activity', *Drug Metabolism and Pharmacokinetics*, 29(5), pp. 360–366. doi: 10.2133/dmpk.DMPK-14-RG-014.

Murray, C. J. L. *et al.* (2012) 'Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010', *The Lancet*, 380(9859), pp. 2197–2223. doi: 10.1016/S0140-6736(12)61689-4.

Nakanishi, M. and Otsu, M. (2012) 'Development of Sendai virus vectors and their potential applications in gene therapy and regenerative medicine.', *Current gene therapy*, 12(5), pp. 410–6. doi: 10.2174/156652312802762518.

Niemi, M. *et al.* (2003) 'Pharmacokinetic interactions with rifampicin: Clinical relevance', *Clinical Pharmacokinetics*, 42(9), pp. 819–850. doi: 10.2165/00003088-200342090-00003.

Nofziger, C. *et al.* (2019) 'PharmVar GeneReview: CYP2D6.', *Clinical pharmacology and therapeutics*, p. cpt.1643. doi: 10.1002/cpt.1643.

Olinga, P. and Groothuis, G. M. M. (2001) 'Use of Human Tissue Slices in Drug Targeting Research', in, pp. 309–331. doi: 10.1002/352760006x.ch12.

Oluka, M. N. *et al.* (2014) 'Characterization of inter-ethnic genetic and GSTs in the Bantu and Nilotic populations of Kenya and implications for the chemotherapy of infectious diseases', *African Journal of Pharmacology and Therapeutics*, 3(2), pp. 38–46.

Osborne, C. K. (1998) 'Tamoxifen in the Treatment of Breast Cancer', *New England Journal of Medicine*. Edited by A. J. J. Wood, 339(22), pp. 1609–1618. doi: 10.1056/NEJM199811263392207.

Oscarson, M. *et al.* (1997) 'A combination of mutations in the CYP2D6*17 (CYP2D6Z) allele causes alterations in enzyme function.', *Molecular pharmacology*, 52, pp. 1034–1040. doi: 10.1124/mol.52.6.1034.

Pagani, L. *et al.* (2012) 'Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool', *American Journal of Human Genetics*, 91(1), pp. 83–96. doi: 10.1016/j.ajhg.2012.05.015.

Pai, H. V. *et al.* (2004) 'A frameshift mutation and alternate splicing in human brain generate a functional form of the pseudogene cytochrome P4502D7 that demethylates codeine to morphine', *Journal of Biological Chemistry*, 279(26), pp. 27383–27389. doi: 10.1074/jbc.M402337200.

Palade, G. E. and Siekevitz, P. (1956) 'LIVER MICROSOMES', *The Journal of Biophysical and Biochemical Cytology*, 2(2), pp. 171–200. doi: 10.1083/jcb.2.2.171.

Palma, E., Doornebal, E. J. and Chokshi, S. (2019) 'Precision-cut liver slices: a versatile tool to advance liver research', *Hepatology International*. Springer India, 13(1), pp. 51–57. doi: 10.1007/s12072-018-9913-7.

Pan, X., Ning, M. and Jeong, H. (2017) 'Transcriptional regulation of CYP2D6 expression', *Drug Metabolism and Disposition*, 45(1), pp. 42–48. doi: 10.1124/dmd.116.072249.

Parkinson, A. *et al.* (2004) 'The effects of gender, age, ethnicity, and liver cirrhosis on cytochrome P450 enzyme activity in human liver microsomes and inducibility in cultured human hepatocytes', *Toxicology and Applied Pharmacology*, 199(3), pp. 193–209. doi: 10.1016/j.taap.2004.01.010.

Pawlina, W. and Ross, M. H. (2016) *Histology - Text and atlas*. 7th edn. Philadelphia: Wolters Kluwer Health.

Pearson, P. G. and Wienkers, L. C. (eds) (2019) *Handbook of Drug Metabolism*. 3rd edn. Boca Ratan, Florida: CRC Press. doi: 10.1201/9780429190315.

Pelkonen, O. *et al.* (1974) 'Cytochrome P-450-linked monooxygenase system and drug-induced spectral interactions in human liver microsomes', *Chemico-Biological Interactions*, 9(3), pp. 205–216. doi: 10.1016/S0009-2797(74)80005-0.

Perlman, R. L. (2016) 'Mouse models of human disease: An evolutionary perspective.', *Evolution, medicine, and public health*, 2016(1), pp. 170–6. doi: 10.1093/emph/eow014.

Pratt, V. *et al.* (2018) 'Medical Genetics Summaries [Internet]', *National Center for Biotechnology Information*, (Md), p. 517. doi: 10.1007/s00439-006-0262-6.

Prentis, R., Lis, Y. and Walker, S. (1988) 'Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964-1985).', *British Journal of Clinical Pharmacology*, 25(3), pp. 387–396. doi: 10.1111/j.1365-2125.1988.tb03318.x.

Promega Corporation (2016) 'P450-Glo™ Assays'. Madison, USA, pp. 1–54. Available at: https://www.promega.com/-/media/files/resources/protocols/technical-bulletins/101/p450-glo-assays-protocol.pdf.

Purcell, S. *et al.* (2007) 'PLINK: A tool set for whole-genome association and population-based linkage analyses', *American Journal of Human Genetics*, 81(3), pp. 559–575. doi: 10.1086/519795.

Raab, S. *et al.* (2014) 'A Comparative View on Human Somatic Cell Sources for iPSC Generation', *Stem Cells International*, 2014. doi: 10.1155/2014/768391.

Rajman, I. *et al.* (2017) 'African Genetic Diversity: Implications for Cytochrome P450-mediated Drug Metabolism and Drug Development', *EBioMedicine*. The Authors, 17, pp. 67–74. doi: 10.1016/j.ebiom.2017.02.017.

Ramaswami, R., Bayer, R. and Galea, S. (2018) 'Precision Medicine from a Public Health Perspective.', *Annual review of public health*, 39(1), pp. 153–168. doi: 10.1146/annurev-publhealth-040617-014158.

Rao, M. S. and Malik, N. (2012) 'Assessing iPSC reprogramming methods for their suitability in translational medicine', *Journal of Cellular Biochemistry*, 113(10), pp. 3061–3068. doi: 10.1002/jcb.24183.

Raucy, J. *et al.* (2002) 'A Cell-Based Reporter Gene Assay for Determining Induction of CYP3A4 in a High-Volume System', 303(1), pp. 412–423. doi: 10.1124/jpet.102.038653.agents.

Rendic, S. and Guengerich, F. P. (2015) 'Survey of human oxidoreductases and cytochrome P450 enzymes involved in the metabolism of xenobiotic and natural chemicals', *Chemical Research in Toxicology*, 28(1), pp. 38–42. doi: 10.1021/tx500444e.

Roden, D. M. *et al.* (2011) 'Pharmacogenomics: the genetics of variable drug responses.', *Circulation*, 123(15), pp. 1661–70. doi: 10.1161/CIRCULATIONAHA.109.914820.

Roses, A. D. (2000) 'Pharmacogenetics and the practice of medicine [Review]', *Nature*, 405(6788), pp. 857–865.

Rowland, P. *et al.* (2006) 'Crystal structure of human cytochrome P450 2D6', *Journal of Biological Chemistry*, 281(11), pp. 7614–7622. doi: 10.1074/jbc.M511232200.

Sacco, A. M. *et al.* (2019) 'Diversity of dermal fibroblasts as major determinant of variability in cell reprogramming', *Journal of Cellular and Molecular Medicine*, 23(6), pp. 4256–4268. doi: 10.1111/jcmm.14316.

Saha, K. and Jaenisch, R. (2009) 'Technical Challenges in Using Human Induced Pluripotent Stem Cells to Model Disease', *Cell Stem Cell*. Elsevier Inc., 5(6), pp. 584–595. doi: 10.1016/j.stem.2009.11.009.

Sakuyama, K. *et al.* (2008) 'Functional characterization of 17 CYP2D6 allelic variants (CYP2D6.2, 10, 14A-B, 18, 27, 36, 39, 47-51, 53-55, and 57)', *Drug Metabolism and Disposition*, 36(12), pp. 2460–2467. doi: 10.1124/dmd.108.023242.

Scott, C. W., Peters, M. F. and Dragan, Y. P. (2013) 'Human induced pluripotent stem cells and their use in drug discovery for toxicity testing', *Toxicology Letters*. Elsevier Ireland Ltd, 219(1), pp. 49–58. doi: 10.1016/j.toxlet.2013.02.020.

Shen, H. *et al.* (2007) 'Comparative Metabolic Capabilities and Inhibitory Profiles of CYP2D6.1, CYP2D6.10, and CYP2D6.17', *Drug Metabolism and Disposition*, 35(8), pp. 1292–1300. doi: 10.1124/dmd.107.015354.Caucasians.

Sherwood, L. (2016) *Human Physiology: From Cells to Systems*. 9th edn. Boston: Cengage Learning.

Shukla, S. J. *et al.* (2011) 'Identification of clinically used drugs that activate pregnane X receptors', *Drug Metabolism and Disposition*, 39(1), pp. 151–159. doi: 10.1124/dmd.110.035105.

Skoda, R. C. *et al.* (1988) 'Two mutant alleles of the human cytochrome P-450db1 gene (P450C2D1) associated with genetically deficient metabolism of debrisoquine and other drugs', *Proceedings of the National Academy of Sciences of the United States of America*, 85(14), pp. 5240–5243. doi: 10.1073/pnas.85.14.5240.

Smith, P. K. *et al.* (1985) 'Measurement of protein using bicinchoninic acid', *Analytical Biochemistry*, 150(1), pp. 76–85. doi: 10.1016/0003-2697(85)90442-7.

Souza, L. M. *et al.* (1986) 'Recombinant human granulocyte colony-stimulating factor: Effects on normal and leukemic myeloid cells', *Science*, 232(4746), pp. 61–65. doi: 10.1126/science.2420009.

Spooner, W. *et al.* (2018) 'Haplosaurus computes protein haplotypes for use in precision drug design', *Nature Communications*. Springer US, 9(1). doi: 10.1038/s41467-018-06542-1.

Ssekamwa, J. C. (1986) *A Sketch Map History of East Africa*. 3rd edn. Cheltenham: Hulton Educational Publications.

Strathern, M. (2014) *Optimization in Drug Discovery*. 2nd edn. Edited by G. W. Caldwell and Z. Yan. Totowa, NJ: Humana Press (Methods in Pharmacology and Toxicology). doi: 10.1007/978-1-62703-742-6.

Sullivan, G. J. *et al.* (2010) 'Generation of functional human hepatic endoderm from human induced pluripotent stem cells.', *Hepatology (Baltimore, Md.)*, 51(1), pp. 329–35. doi: 10.1002/hep.23335.

Takahashi, K. *et al.* (2007) 'Induction of pluripotent stem cells from adult human fibroblasts by defined factors.', *Cell*, 131(5), pp. 861–72. doi: 10.1016/j.cell.2007.11.019.

Takahashi, K. and Yamanaka, S. (2006) 'Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors', *Cell*, 126(4), pp. 663–676. doi: 10.1016/j.cell.2006.07.024.

Takayama, K. *et al.* (2013) '3D spheroid culture of hESC/hiPSC-derived hepatocyte-like cells for drug toxicity testing', *Biomaterials*. Elsevier Ltd, 34(7), pp. 1781–1789. doi: 10.1016/j.biomaterials.2012.11.029.

Takayama, K. *et al.* (2014) 'Prediction of interindividual differences in hepatic functions and drug sensitivity by using human iPS-derived hepatocytes', *Proceedings of the National Academy of Sciences*, 111(47), pp. 16772–16777. doi: 10.1073/pnas.1413481111.

Terblanche, A. *et al.* (2017) 'Knowledge, attitudes and perspective on adverse drug reaction reporting in a public sector hospital in South Africa: baseline analysis', *Hospital practice (1995)*. Taylor & Francis, 45(5), pp. 238–245. doi: 10.1080/21548331.2017.1381013.

Thermo Fisher Scientific (2016a) 'Gibco™ Cell Culture Basics Handbook'. Carlsbad, CA: Thermo Fisher Scientific, pp. i–108.

Thermo Fisher Scientific (2016b) 'Invitrogen CytoTune ™ -iPS 2.0 Sendai Reprogramming Kit User Guide'. Carlsbad, CA: Thermo Fisher Scientific, pp. 1–72.

Tracy, T. S. *et al.* (2016) 'Interindividual Variability in Cytochrome P450-Mediated Drug Metabolism', *Drug Metabolism and Disposition*, 44(3), pp. 343–351. doi: 10.1124/dmd.115.067900.

Twist, G. P. *et al.* (2016) 'Constellation: A tool for rapid, automated phenotype assignment of a highly polymorphic pharmacogene, CYP2D6, from whole-genome sequences', *npj Genomic Medicine*, 1, p. 2017. doi: 10.1038/npjgenmed.2015.7.

Vangipuram, M. *et al.* (2013) 'Skin punch biopsy explant culture for derivation of primary human fibroblasts.', *Journal of visualized experiments : JoVE*, (77), pp. 9–11. doi: 10.3791/3779.

de Vries Schultink, A. H. M. *et al.* (2015) 'Effects of Pharmacogenetics on the Pharmacokinetics and Pharmacodynamics of Tamoxifen', *Clinical Pharmacokinetics*. Springer International Publishing, 54(8), pp. 797–810. doi: 10.1007/s40262-015-0273-3.

Walker, J. M. and Walker, J. M. (2003) 'The Bicinchoninic Acid (BCA) Assay for Protein Quantitation', *Basic Protein and Peptide Protocols*, pp. 5–8. doi: 10.1385/0-89603-268-x:5.

Wang, D. *et al.* (2014) 'Common CYP2D6 polymorphisms affecting alternative splicing and transcription: long-range haplotypes with two regulatory variants modulate CYP2D6 activity.', *Human molecular genetics*, 23(1), pp. 268–78. doi: 10.1093/hmg/ddt417.

Wang, Q. *et al.* (2020) 'Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes', *Nature Communications*, 11(1), p. 2539. doi: 10.1038/s41467-019-12438-5.

Wennerholm, A. *et al.* (2002) 'The African-specific CYP2D6*17 allele encodes an enzyme with changed substrate specificity', *Clinical Pharmacology and Therapeutics*, 71(1), pp. 77–88. doi: 10.1067/mcp.2002.120239.

Wolff, T. *et al.* (1985) 'Substrate specificity of human liver cytochrome P-450 debrisoquine 4-hydroxylase probed using immunochemical inhibition and chemical modeling.', *Cancer research*, 45(5), pp. 2116–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3921236.

Wright, R. L. *et al.* (1996) 'Cloning of a potential cytochrome P450 from the Archaeon Sulfolobus solfataricus', 384, pp. 235–239.

Yang, Y. *et al.* (2017) 'Sequencing the CYP2D6 gene: From variant allele discovery to clinical pharmacogenetic testing', *Pharmacogenomics*, 18(7), pp. 673–685. doi: 10.2217/pgs-2017-0033.

Yengi, L. G., Leung, L. and Kao, J. (2007) 'The evolving role of drug metabolism in drug discovery and development',

Pharmaceutical Research, 24(5), pp. 842–858. doi: 10.1007/s11095-006-9217-9.

Zanger, U. M. *et al.* (1988) 'Antibodies against human cytochrome P-450db1 in autoimmune hepatitis type II', *Proceedings of the National Academy of Sciences of the United States of America*, 85(21), pp. 8256–8260. doi: 10.1073/pnas.85.21.8256.

Zanger, U. M., Raimundo, S. and Eichelbaum, M. (2004) 'Cytochrome P450 2D6: Overview and update on pharmacology, genetics, biochemistry', *Naunyn-Schmiedeberg's Archives of Pharmacology*, 369(1), pp. 23–37. doi: 10.1007/s00210-003-0832-2.

Zanger, U. M. and Schwab, M. (2013) 'Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation', *Pharmacology and Therapeutics*. Elsevier Inc., 138(1), pp. 103–141. doi: 10.1016/j.pharmthera.2012.12.007.

Zhang, Z. and Tang, W. (2018) 'Drug metabolism in drug discovery and development', *Acta Pharmaceutica Sinica B*. Elsevier B.V., pp. 721–732. doi: 10.1016/j.apsb.2018.04.003.

Zhou, T. *et al.* (2011) 'Generation of induced pluripotent stem cells from urine', *Journal of the American Society of Nephrology*, 22(7), pp. 1221–1228. doi: 10.1681/ASN.2011010106.

Zhou, Y., Ingelman-Sundberg, M. and Lauschke, V. M. (2017) 'Worldwide Distribution of Cytochrome P450 Alleles: A Meta-analysis of Population-scale Sequencing Projects', *Clinical Pharmacology and Therapeutics*, 102(4), pp. 688–700. doi: 10.1002/cpt.690.

Zuber, R., Anzenbacherová, E. and Anzenbacher, P. (2002) 'Cytochromes P450 and experimental models of drug metabolism.', *Journal of cellular and molecular medicine*, 6(2), pp. 189–98. doi: 10.1111/j.1582-4934.2002.tb00186.x.

5.2 APPENDICES

5.2.1 Appendix A: Ethics Clearance Certificates

Appendix A1: Clearance certificate for approval of ethics application no. M190221 obtained through the Human Research Ethics Committee (HREC) (Medical) at the University of the Witwatersrand.

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Mr Brian Kariithi

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M190221

NAME: (Principal Investigator)	Mr Brian Kariithi
DEPARTMENT:	SBIMB Division of Human Genetics
PROJECT TITLE:	Evaluation of iPSC-derived hepatocyte model to predict functional effects of the CYP2D6*17 genetic variant in activation of tamoxifen to endoxifen
DATE CONSIDERED:	22/02/2019
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Prof Michele Ramsay, Prof Collen Masimirembwa and Dr Janine Scholefield
APPROVED BY:	Dr C Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	29/04/2019

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 301, Third floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. 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 resubmit the application to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed February and will therefore be due in the month of February each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

14/05/2019
Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix A2: Clearance certificate for approval of ethics application for project titled "Development of a cellular model of infectious disease investigating drug toxicity" and application reference no. Ref: 77/2013 obtained through the Research Ethics Committee (REC) at the Council for Scientific and Industrial Research (CSIR).

CSIR Research Ethics Committee
PO Box 395 Pretoria 0001 South Africa
Tel: +27 12 841 4060
Fax: +27 12 841 2476
Email:R&DEthics@csir.co.za

26 February 2019

Modification of Approved Protocol

Dear: **Dr Janine Scholefield**

Title: **Development of a Cellular model of infectious disease investigating drug toxicity (Ref: 77/2013).**

Thank you for submitting your request for modification of the approved protocol Ref: 77/2013 to the CSIR Research Ethics Committee (REC). Your submission was reviewed and discussed in the REC review meeting of the 21st of February 2019. The CSIR REC notes and approves the following protocol modifications:

- The addition of the Master's student, Brian Karithi; and
- The use of the approved existing protocols under Ref: 77/2013 to train the student.

The CSIR REC duly notes the submitted research protocol as approved by the Wits University Research Assessor's Committee and the submitted proof of registration for Brian Kariithi.

The research collaboration between the CSIR and the Sydney Brenner Institute at Wits University for the inclusion of the two skin biopsies from individuals with the necessary genetic variant as a modification under the current research project is noted. This modification cannot be approved at this stage pending receipt of the Wits University Research Ethics Clearance certificate, full research ethics application documentation submitted at Wits and the signed research collaboration agreement between CSIR and Wits University.

We wish you all of the best with your research project.

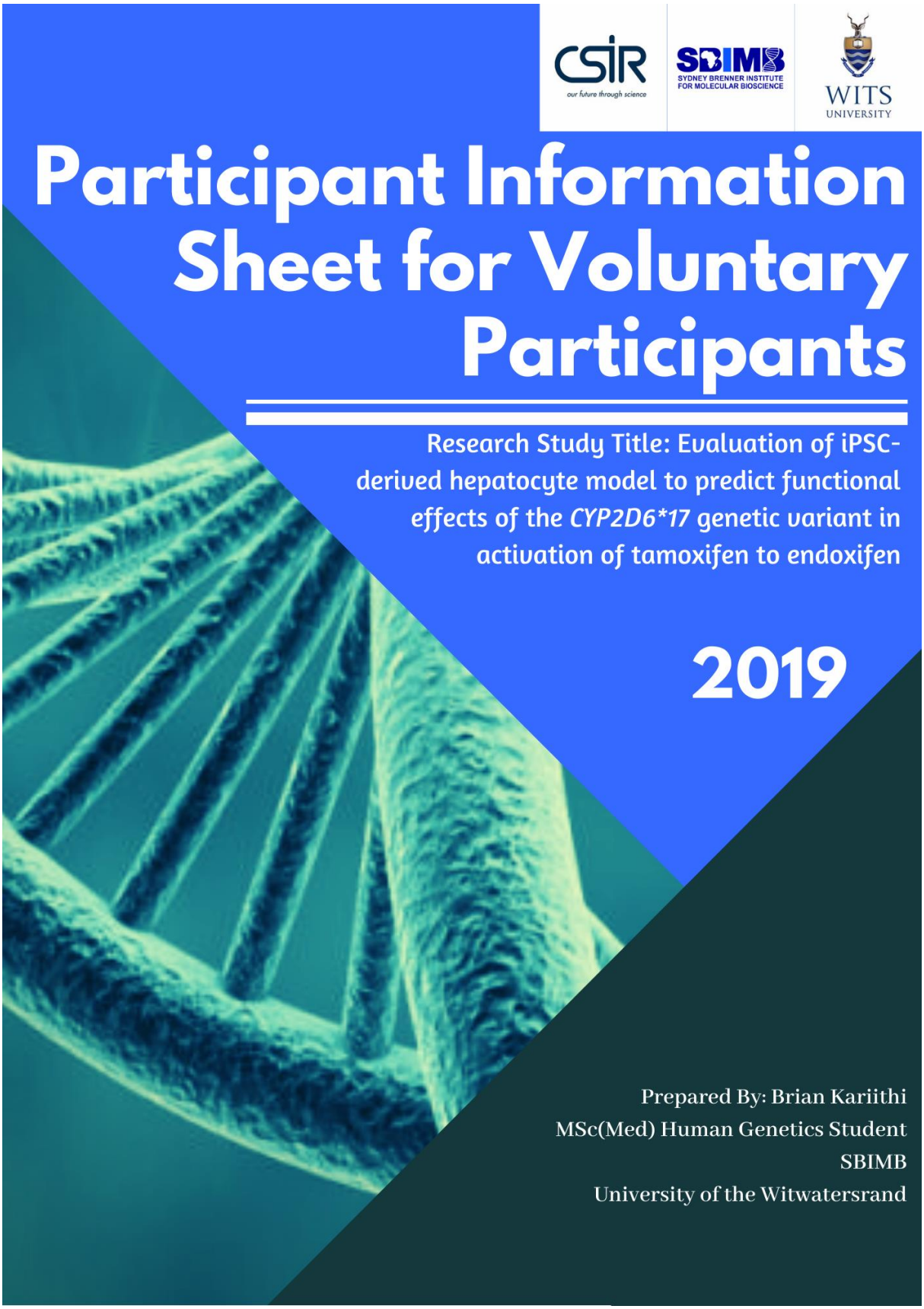
Kind regards

D. S. Jacobs

Prof David Jacobs
CSIR REC Chairperson

5.2.2 Appendix B: Study Recruitment & Participation Forms

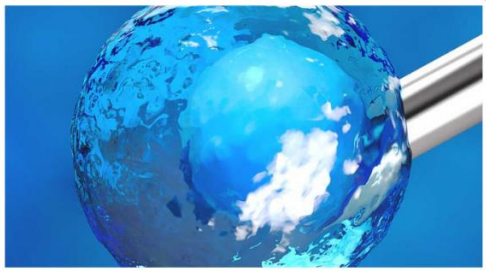
Appendix B1: Information sheet for study participants recruited for buccal swabs & skin biopsies.



Introduction To Study

Hello, my name is Brian Kariithi and I am an MSc student at Wits University. My supervisors are Professor Michele Ramsay, Professor Collen Masimirembwa and Dr Janine Scholefield. We would like to invite you to participate in a genetics study aimed at developing a stem cell-derived hepatocyte-like cell model to mimic the effects of a genetic variant on tamoxifen metabolism to endoxifen in black South African populations. The study is set to start March 2019 and end in June 2020. Participation is entirely voluntary and you will not be disadvantaged in any way if you choose not to participate. You may also withdraw from the study at any time by contacting us and this will not disadvantage you in any way. However, you cannot withdraw from the study once the stem cells are made due to the high cost in making them.





Drugs are metabolised in the liver and can be affected by numerous different factors. CYP2D6 is an important enzyme from the cytochrome P450 superfamily for drug metabolism and assists in the formation of endoxifen from tamoxifen, a drug for breast cancer treatment. The African-specific *CYP2D6*17* allele in the *CYP2D6* gene causes a reduced function effect on the CYP2D6 enzyme and decreases the enzyme substrate binding affinity. Numerous methods have been employed to study such effects in a physiologically relevant *in-vitro* model. In recent times, induced pluripotent stem cells (iPSCs) have become very attractive to develop a cellular model for liver-like tissue and to study the individual differences in drug metabolism and toxicity. **The aim of this study is to identify one or two people who are homozygous for the *CYP2D6*17* allele and then to ask them for a skin biopsy. An iPSC cell line will then be produced from the cells in the biopsy, and will be used to look at the effect of the allele on drug metabolism.**

CYP2D6*17, Tamoxifen Metabolism & iPSC Models

CYP2D6 is an important enzyme for drug metabolism and is responsible for the oxidation of over 70 prescribed drugs, ranging from antidepressants and anti-arrhythmics, to β-blockers and opiates. The *CYP2D6*17* allele causes a reduced function effect on the CYP2D6 enzyme and homozygous individuals would be classified as intermediate metabolisers. Looking at *CYP2D6*17* and its reduced function and activity, it seems that there could be effects seen on important prescription drugs, such as Tamoxifen in breast cancer.

Tamoxifen is frequently used in the treatment of hormone/oestrogen receptor-positive (H+/ER+) breast cancer. Tamoxifen is a prodrug and the pharmacologically relevant compounds are the metabolites that are produced by drug metabolism liver enzymes. The compound that is considered to be the primary active metabolite of tamoxifen is endoxifen. This reaction is catalysed by numerous CYP enzyme isoforms, but the most important and predominating enzymes in this process are CYP2D6 and CYP3A4/5.

iPSCs are derived from dermal fibroblasts and have the potential ability to differentiate into numerous cell types. This ability to differentiate into any cell type in the body, makes iPSCs very attractive as a cellular model for liver-like tissue and to study the individual differences in drug metabolism and toxicity. The iPSCs design for liver tissue differentiation are referred to as stem-cell-derived hepatocyte-like cells (SC-HLCs).



PARTICIPANT INFORMATION SHEET FOR VOLUNTARY PARTICIPANTS 2018/2019

Sampling Methodology

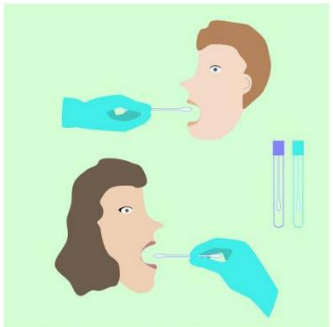
BUCCAL SWAB & SKIN BIOPSY

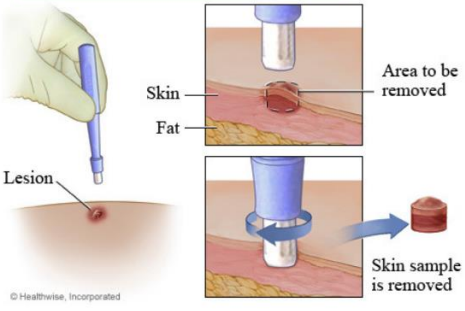


The study will take place in two steps. First, we will enroll about 50 people and ask them for buccal samples. Second, we will then select just one or two people, based on the outcome of our genetic study, and ask them to donate a skin sample at a later stage.

A buccal swab sample will be requested to perform genotyping for the presence of the *CYP2D6*17* haplotype in each participant in the study. A buccal swab is a sterile brush instrument used to scrape off cells on the inner surface of the mouth around the cheek area, this is called a buccal scrape. All participants will be required to consent to a buccal scrape before the sample can be taken for genotyping. The individuals that are found to be homozygous for *CYP2D6*17* will subsequently have their DNA sequenced. The buccal samples will be taken at a suitable venue and stored at the SBIMB biobank facility and will be kept there until the genotyping and sequencing are done. We expect between 1 and 3 individuals to be homozygotes and they are the only ones with whom we will follow up for a skin biopsy. They will then be re-consented for the use of their cells for an iPSC line development.

A skin biopsy procedure is a technique done to remove a portion of skin which is analysed for clinical diagnostic or research purposes. The type of skin biopsy that will be utilised in this study is a punch biopsy. This is when a small circular blade is used to make an incision into your skin retrieving all the layers. This will be a small incision that is roughly about 4mm in diameter. The dermal fibroblasts from the dermis will be harvested to create an iPSC line for a stem cell model. These iPSC-dervied cells are called hepatocyte-like cells.





INFORMED CONSENT & SKIN BIOPSY PROCEDURE 2018/2019

SKIN BIOPSY PROCEDURE

Skin biopsy samples will be taken by a qualified dermatologist. This involves the application of local anaesthetic to the skin. This procedure may cause some discomfort or pain may be felt with the first injection, however once the anaesthetic has taken effect, you should not feel pain when the biopsy is performed. Thereafter the biopsy will be performed producing a skin sample. After the procedure a steri-strip will be placed over the wound to prevent infections and assist with better wound healing. If necessary, the wound from the biopsy will be stitched. Even though precautions will be taken to prevent infections, the risk still exists therefore, the dermatologist will advise you on how to care for the wound. If an infection occurs the participant should contact the physician who performed the biopsy for further assistance at no financial cost to the participant and a travel reimbursement of R 300.00 will be provide. The biopsy procedure will take approximately 20 minutes.

INFORMED CONSENT

Individuals that give consent to voluntarily participate in the study will be required to give consent for a buccal swab for genotyping of *CYP2D6*. The information from the genotyping will be available to the participants pertaining to their *CYP2D6* status. Participants will be required to consent to their samples being sequenced to analyse their *CYP2D6* gene, and we will provide you with your result by email. There are no implications for individuals who do not have the variant and if you have one copy of the variant, this is unlikely to significantly affect your ability to metabolise certain drugs, unless you carry additional variants at affect drug metabolism. Those with two copies of the variant are likely to experience altered drug metabolism for certain drugs. Should you wish to discuss your result with a trained professional, we will provide you with information to make an appointment for genetic counselling, at no cost to you. Participants will be required to give consent to a buccal swab sample and, if found to be homozygous for *CYP2D6*17*, will be ask to reconsent for a skin punch biopsy. This procedure will be done to establish iPSC lines for a stem cell model. The stem cell lines will be stored at CSIR or appropriate biorepository, and may be used in consequent studies or in commercial endeavours. You will not receive any financial or other types of benefits should the iPSC line and its derivatives be used commercially. All data acquired during the study will be kept on a password-protected computer and under restricted access. The identity of each participant will be kept anonymous and every participant will be given a user ID to identify their sample. This user ID will be used to re-contact selected participants for a a skin biopsy. A second consent form will then be signed before the skin biopsy is taken.

POTENTIAL RISKS

PHYSICAL RISK

There are minimal physical risks associated with any of the procedures that will be utilized in this study.

The buccal scrape is a non-invasive procedure and will not result in any scarring or wound to the mouth. No pain or discomfort should be experience during the procedure as the swab exerts a soft touch on the inner surface of the mouth around the cheek area. The skin biopsy is a minimally invasive procedure and will result in a small wound due to the incision made by the punch biopsy tools. The dermatologist will use their discretion to determine if any infection is present, and will decide the amount of care and attention needed for the wound for the healing process. No pain or discomfort should be felt during the incision and removal of the skin sample as local anaesthetic will be given to the participants.

DATA-ASSOCIATED RISK

There are minimal risks associated with the biological data that will be generated from the procedures that will be utilized in this study.

All data will be anonymised and de-indentified from the participant from which it came. There is minimal risk of any data being traced back to any participant in the study. Participants that do not carry the *CYP2D6*17* variant, or have one copy of the variant, have minimal risk of any drug metabolism problems associated with the variant. Participants who are homozygous for *CYP2D6*17* do carry a higher risk of drug metabolism problems linked to drugs metabolised by CYP2D6. Genetic Counselling will be offered to any of the participants in need of such and for guidance and advice on the implications of these findings.

The iPSC lines that will be developed from one or two participants will be anonymised and de-indentified from the participant from which it came. There is minimal risk of any data being traced back to any participant in the study.

CONTACT US



BRIAN KARIITHI

MSc(Med) Human Genetics student at Sydney Brenner Institute for Molecular Bioscience, Division of Human Genetics, School of Pathology, University of the Witwatersrand, Johannesburg.



EMAIL ADDRESS

brian.kariithi@students.wits.ac.za



PHONE NUMBER

+27 76 608 3529
+27 11 717 6635



PHYSICAL ADDRESS

The Mount, 9 Jubilee Road, Parktown, Johannesburg, 2193

SUPERVISOR CONTACTS

PROF MICHELE RAMSAY

Director at Sydney Brenner Institute of Molecular Bioscience, Division of Human Genetics, School of Pathology, University of the Witwatersrand.
michele.ramsay@wits.ac.za

PROF COLLEN MASIMIREMBWA

President and Chief Scientific Officer at African Institute of Biomedical Science and Technology.
collenmasimirembwa@yahoo.com

DR JANINE SCHOLEFIELD

Senior Researcher, Synthetic Biology, Emerging Research Area, Biosciences, Council for Scientific and Industrial Research.
jscholefield@csir.co.za

CONTACT DETAILS OF RESEARCH ETHICS COMMITTEE ADMINISTRATOR - FOR REPORTING OF COMPLAINTS / PROBLEMS.

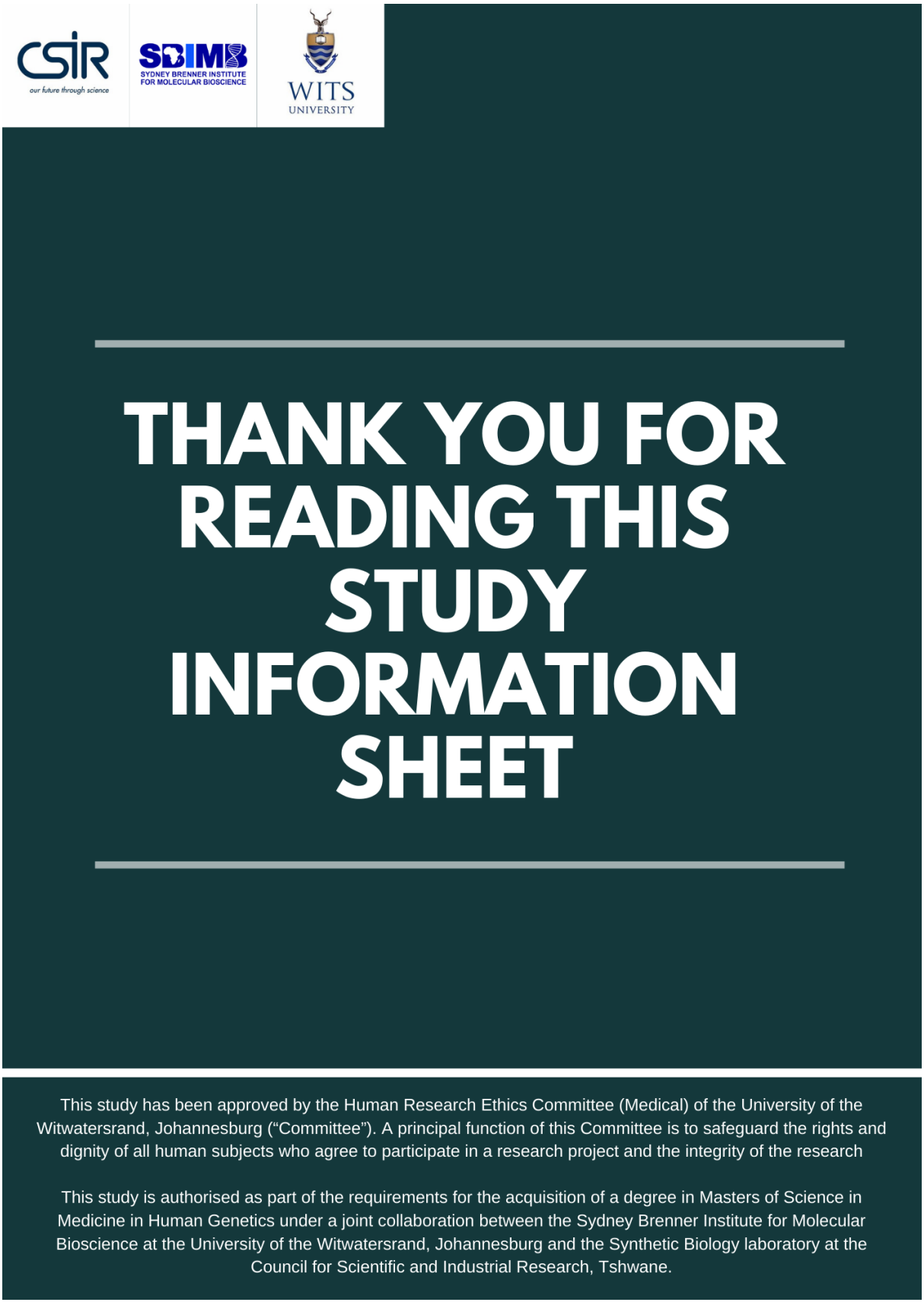
MS ZANELE NDLOVU
MR RHULANI MKANSI

Email Address:
zanele.ndlovu@wits.ac.za
OR
rhulani.mkansi@wits.ac.za.

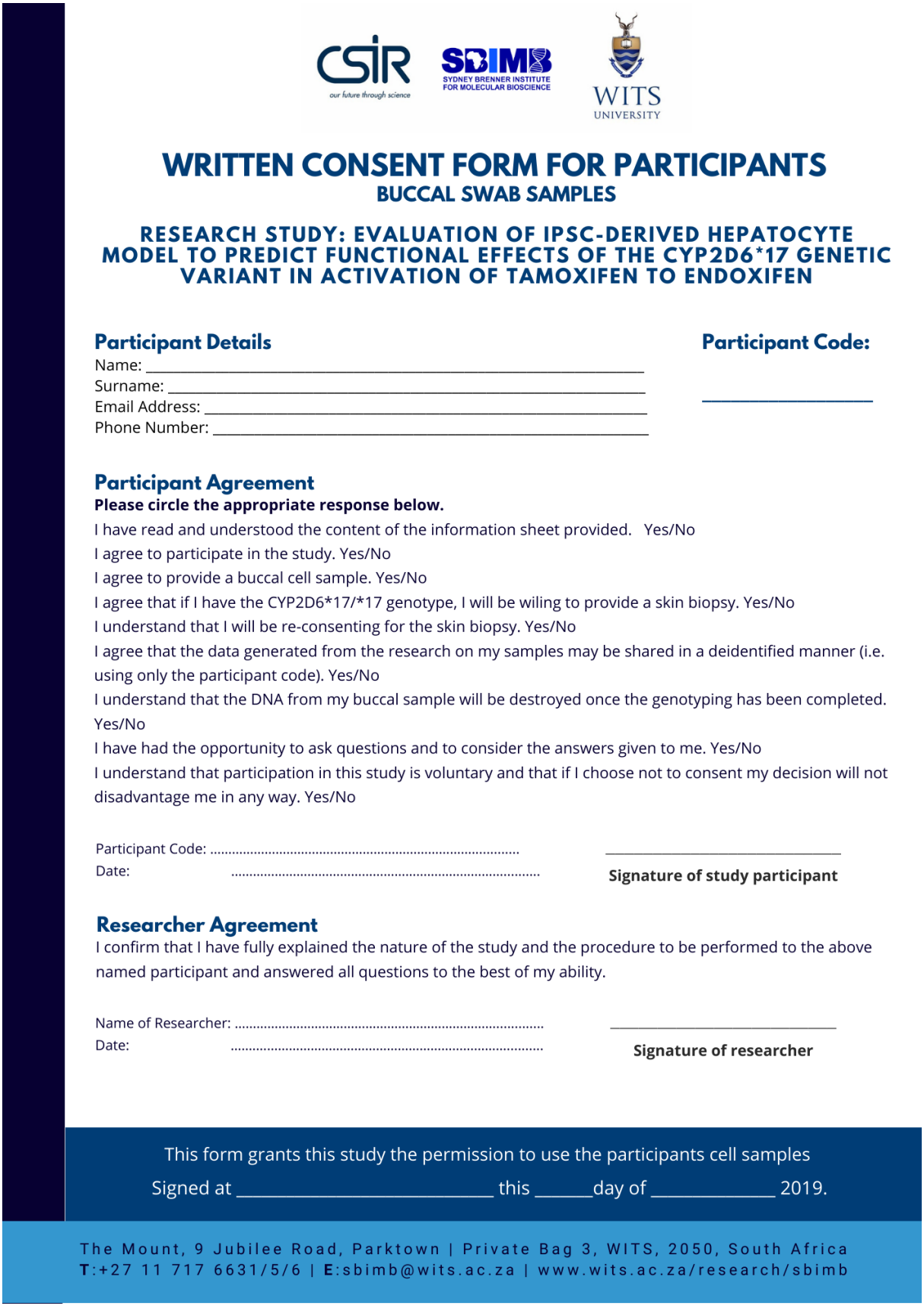
Phillip Tobias Health Sciences Building, 2nd Floor, Cnr York Road and Princess of Wales Terrace,
Tel: 011-717-1234/1252/ 2700
OR
Room SH1005,
10th Floor Senate House, WIts

PROF CLEMENT PENNY
Committee Chairperson

Email Address:
clement.penny@wits.ac.za
Tel: 011-717-2301


CSIR
our future through science
SDIMB
SYDNEY BRENNER INSTITUTE FOR MOLECULAR BIOSCIENCE
WITS
UNIVERSITY
THANK YOU FOR READING THIS STUDY INFORMATION SHEET
This study has been approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand, Johannesburg ("Committee"). A principal function of this Committee is to safeguard the rights and dignity of all human subjects who agree to participate in a research project and the integrity of the research
This study is authorised as part of the requirements for the acquisition of a degree in Masters of Science in Medicine in Human Genetics under a joint collaboration between the Sydney Brenner Institute for Molecular Bioscience at the University of the Witwatersrand, Johannesburg and the Synthetic Biology laboratory at the Council for Scientific and Industrial Research, Tshwane.

Appendix B2: Written consent form for buccal swab sampling for study participants.



CSIR our future through science | SBIMB SYDNEY BRENNER INSTITUTE FOR MOLECULAR BIOSCIENCE | WITS UNIVERSITY

WRITTEN CONSENT FORM FOR PARTICIPANTS
BUCCAL SWAB SAMPLES

RESEARCH STUDY: EVALUATION OF IPSC-DERIVED HEPATOCYTE MODEL TO PREDICT FUNCTIONAL EFFECTS OF THE CYP2D6*17 GENETIC VARIANT IN ACTIVATION OF TAMOXIFEN TO ENDOXIFEN

Participant Details

Name: _____
Surname: _____
Email Address: _____
Phone Number: _____

Participant Code:

_____

Participant Agreement

Please circle the appropriate response below.

I have read and understood the content of the information sheet provided. Yes/No
I agree to participate in the study. Yes/No
I agree to provide a buccal cell sample. Yes/No
I agree that if I have the CYP2D6*17/*17 genotype, I will be wiling to provide a skin biopsy. Yes/No
I understand that I will be re-consenting for the skin biopsy. Yes/No
I agree that the data generated from the research on my samples may be shared in a deidentified manner (i.e. using only the participant code). Yes/No
I understand that the DNA from my buccal sample will be destroyed once the genotyping has been completed. Yes/No
I have had the opportunity to ask questions and to consider the answers given to me. Yes/No
I understand that participation in this study is voluntary and that if I choose not to consent my decision will not disadvantage me in any way. Yes/No

Participant Code: .....
Date: .....

_____
Signature of study participant

Researcher Agreement

I confirm that I have fully explained the nature of the study and the procedure to be performed to the above named participant and answered all questions to the best of my ability.

Name of Researcher: .....
Date: .....

_____
Signature of researcher

This form grants this study the permission to use the participants cell samples

Signed at _____ this _____day of _____ 2019.

The Mount, 9 Jubilee Road, Parktown | Private Bag 3, WITS, 2050, South Africa
T:+27 11 717 6631/5/6 | **E**:sbimb@wits.ac.za | www.wits.ac.za/research/sbimb

Appendix B3: Written consent form for skin biopsy procedure for study participants.

WRITTEN CONSENT FORM FOR PARTICIPANTS
SKIN BIOPSY SAMPLES

RESEARCH STUDY: EVALUATION OF IPSC-DERIVED HEPATOCYTE MODEL TO PREDICT FUNCTIONAL EFFECTS OF THE CYP2D6*17 GENETIC VARIANT IN ACTIVATION OF TAMOXIFEN TO ENDOXIFEN

Participant Details

Participant Code:

Name: _____
Surname: _____
Email Address: _____
Phone Number: _____

Participant Agreement

Please circle the appropriate response below.

I have read and understood the content of the information sheet provided. Yes/No
I agree to participate in the study. Yes/No
I agree to a skin punch biopsy procedure performed by a qualified dermatologist. Yes/No
I agree that the data generated from the research on my samples may be shared in a deidentified manner (i.e. using only the participant code). Yes/No
I understand that a reprogrammed iPSC line will be developed from the cells in my skin biopsy. Yes/No
I understand that the iPSC line will be tested for drug metabolism response. Yes/No
I understand that all studies will be done in an in-vitro laboratory setting. Yes/No
I understand that this iPSC line and its derivatives will be stored, for an indefinite period, at the CSIR, and managed under both CSIR and Wits ethics committees. Yes/No
I understand that, should it become necessary, a back up sample could be stored in a biorepository with appropriate governance structures. Yes/No
I understand that if the iPSC cell line is successfully established, it will be kept for research purposes until it is no longer able to divide successfully. Yes/No
I understand that the cells and their derivatives may be used for commercial purposes, and if this is done it will be managed through the technology transfer offices of CSIR and Wits. Yes/No
I agree that I will not have a claim to any financial or other benefits from potential commercialisation of the modified derivatives of this skin biopsy. Yes/No
I have had the opportunity to ask questions and to consider the answers given to me. Yes/No
I understand that participation in this study is voluntary and that if I choose not to consent my decision will not disadvantage me in any way. Yes/No

Participant Code: .....
Date: .....

_____
Signature of study participant

Researcher Agreement

I confirm that I have fully explained the nature of the study and the procedure to be performed to the above named participant and answered all questions to the best of my ability.

Name of Researcher: .....
Date: .....

_____
Signature of researcher

This form grants this study the permission to use the participants cell samples

Signed at _____ this _____day of _____ 2019.

The Mount, 9 Jubilee Road, Parktown | Private Bag 3, WITS, 2050, South Africa
T:+27 11 717 6631/5/6 | **E**:sbimb@wits.ac.za | www.wits.ac.za/research/sbimb

5.2.3 Appendix C: Bioinformatic Coding Script & Supplementary Data

Appendix C1: Coding script for calling & predicting the CYP2D6 variant frequencies in the AGVP data using Astrolabe 0.8.6.1

Code script for SNP frequency calls for AGVP data

Bagandan Data

```
plink --bfile cyp2d6_Brian --keep baganda --make-bed --out pops/baganda_cyp2d6
plink --bfile baganda_cyp2d6 --freq --out baganda_cyp2d6_freq
```

Ethiopian Data

```
plink --bfile cyp2d6_Brian --keep ethio --make-bed --out pops/ethio_cyp2d6
plink -- bfile ethio_cyp2d6 --freq --out ethio_cyp2d6_freq
```

Zulu Data

```
plink --bfile cyp2d6_Brian --keep zulu --make-bed --out pops/zulu_cyp2d6
plink --bfile zulu_cyp2d6 --freq --out Zulu_freq
```

All Populations

```
plink --bfile ethio_cyp2d6 --bmerge baganda_cyp2d6 --make-bed --out UgEth_cyp2d6
plink --bfile UgEth_cyp2d6 --bmerge zulu_cyp2d6 --make-bed --out AGVP_cyp2d6
plink --bfile AGVP_cyp2d6 --freq --out AGVP_cyp2d6_freq
```

Code script for Haplotype calls for AGVP data

Preparations of VCF filles for Astrolabe

```
for i in APP5117332… EGAN0000107
        vcf-subset -c ${i} cyp2d6_Brian.recode.vcf > ${i}.vcf
done
```

Run for haplotype predictions on Astrolabe

```
for i in APP5117332… EGAN0000107
        astrolabe-0.8.6.1/run-astrolabe.sh -inputVCF ${i}.vcf -conf astrolabe.ini -outFile
        ${i}.log -skipVcfQC -verboseFile ${i}.alleles.txt
done
```

Call haplotype predictions from output file

```
python import os, sys
        in_file = sys.argv[1]
f = open (in_file) #out_file = open ("results.tsv")
for line in f:
```

```
    if line.startswith("##Input"):
        first_coord = line.find("cyp2d6/")
        start_coord = first_coord + 7
        last_coord = line.find(".vcf")
        ID = line[start_coord:last_coord]
    if line.startswith("CYP2D6"):
        req_line = line
print(ID + "\t" + req_line)
f.close()
#out_file.close()
```

Output data in TSV file

```
for file in *.log; do
    python get-allele-calls-astr.py $file >> results.tsv
done
tr "/" "\t" < results.tsv > new_result.tsv && mv new_result.tsv results.tsv
```

5.2.4 Appendix D: ATCC MEF culture and inactivation with mitomycin C

Thawing and bulking up ATCC mouse embryonic fibroblasts (MEF)

1. Add 5 ml of 0.1% gelatin or AF into a T75 flask and incubate in a 37ºC, humidified incubator with 5% CO2 for 1 hr (0.1% gelatin) or 30 min (AF).
2. Pre-warm MEF medium and thaw 1 vial of ATCC MEFs (P2) as described above.
3. Add thawed cells drop-wise into 5 ml MEF medium in a 15 ml conical tube. Close the tube and gently invert to mix the cell suspension.
4. Centrifuge at 300 x *g* for 5 minutes at room temperature.
5. Remove supernatant and resuspend cells in 1 ml of MEF medium.
6. Remove gelatin or AF (Do not rinse flask before adding medium). Add 9 ml of MEF medium into the T75 flask.
7. Add the 1 ml cell suspension into the T75 flask and distribute the cells by tilting the flask sideways. Ensure that the medium does not touch the mouth of flask and lid.
8. Incubate cells in a 37ºC, humidified incubator with 5% CO2 incubator.
9. Culture the cells until they reach 80-90% confluence. Change medium every 2-3 days (twice a week).
10. Passage MEFs into a T150 flask (1/2 split) once confluent.
11. Aspirate spent medium and discard.
12. Wash cell monolayer with 5 mL of Dulbecco's Phosphate Buffered Saline (DPBS) without calcium and magnesium. Aspirate and discard.
13. Add 4 ml of TrypLE™ select in a T75 flask flask. Ensure complete coverage of cell monolayer with TrypLE™ select.
14. Incubate at 37°C for 2-3 minutes, or until cells have detached. Observe cell monolayer using an inverted microscope to ensure complete cell detachment from the surface of the †ask. Gently tap flask to dislodge cells if necessary.
15. Add 5–10 mL of pre-warmed complete MEF medium to flask. Tilt flask in all directions to thoroughly rinse flask. Collect cell suspension to a 15-mL conical tube.
16. Centrifuge at 300 × g for 3-5 minutes.
17. Remove supernatant and triturate cell pellet with 1 ml pre-warmed MEF medium. Resuspend cells in 2 ml of MEF medium.
18. Add 19 ml of MEF medium per T150 flask and add 1 ml of cell suspension per T150 flask (final volume = 20 ml).
19. Incubate cells until 80-90% confluent. Change MEF medium (20 ml) every 2-3 days (twice a week).

20. Passage MEFs onto 3 x T150 flasks (1/3 split). Perform 1/3 split for each T150 flask until cells reach passage 5 (P5).
21. Once MEFs reach P5, inactivate with mitomycin C.

Inactivation of mouse embryonic fibroblasts with Mitomycin C *(modified from amsbio)*

Inactivated primary MEFs are used to promote growth and prevent differentiation of embryonic or induced pluripotent stem cells. MEFs can be inactivated either by chemical treatment or irradiation. When access to an irradiator is limited, treatment with mitomycin C is recommended. Care must be taken when handling mitomycin C as it is genotoxic. Wear gloves, sleeve, and a lab coat when working with the substance. Solutions containing this chemical should be disposed of according to the MSDS. Rinse pipettes with PBS/bleach and place them into separate plastic bag, before discarding them into a biohazard bin.

1. Reconstitute mitomycin C with ddH_2O to make 1 mg/mL mitomycin C solution. Mix thoroughly.
2. Add 200 µl of 1 mg/mL mitomycin C solution to 20 ml media in T150 flask containing 80-90% confluent MEF culture to achieve a final concentration of 10 ug/ml.
3. Incubate the flasks for 2.5–3 hrs at 37^{o}C in humidified incubator with 5% CO2.
4. Carefully remove culture medium containing mitomycin C and **discard into a designated waste bottle that contains 100 ml PBS/bleach for proper disposal.**
5. Wash each flask with 10 mL PBS. Remove the wash and add it to the designated mitomycin C waste bottle. Wash again with 10 mL PBS.
6. Add appropriate amount of 1X TrypLe Select to the flask and incubate for 1 min. in a humidified, 37^{o}C incubator with 5% carbon dioxide. iMEFs lift very quickly, therefore do not incubate with TrypLE select longer than 2 min.
7. While tapping the flask, observe the cells under an inverted microscope until cells detach (1.5–2 min.).
8. Add equal amount of culture medium to inactivate the trypsin and rinse surface of the flask to detach all cells. Gently pipetting up and down will break cell clumps.
9. Transfer the cell suspension into a centrifuge bottle or 50 ml conical tube. Wash the flask with medium.
10. Centrifuge at 300 x *g* for 5 min. Discard the supernatant.
11. Resuspend the pellet in the culture medium.
12. Perform cell count and dilute the cell suspension to twice the final concentration. Freeze down cells at densities of **2 x 10^6**cells/vial and **1 x 10^6** cells/vial.